Identifying contaminants in astronomical images using convolutional neural networks

Maxime Paillassa[1], Emmanuel Bertin[2], Hervé Bouy[1]

[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.*

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.

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 [5] 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 adress 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.

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 [3]. 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 astronomical objects have been separated in an additional class (magenta). Black pixels are pixels that belong to several classes.

Neural network architecture

The model used for the semantic segmentation is similar to Segnet [2] 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 elu activations, except the last one that uses sigmoid to produce the probability maps for each class. This choice instead of the more renown softmax function enables the network to assign several classes to the same pixel, which is a behaviour that we desire. The architecture is represented in figure 1. It was implemented using the TensorFlow library [1].

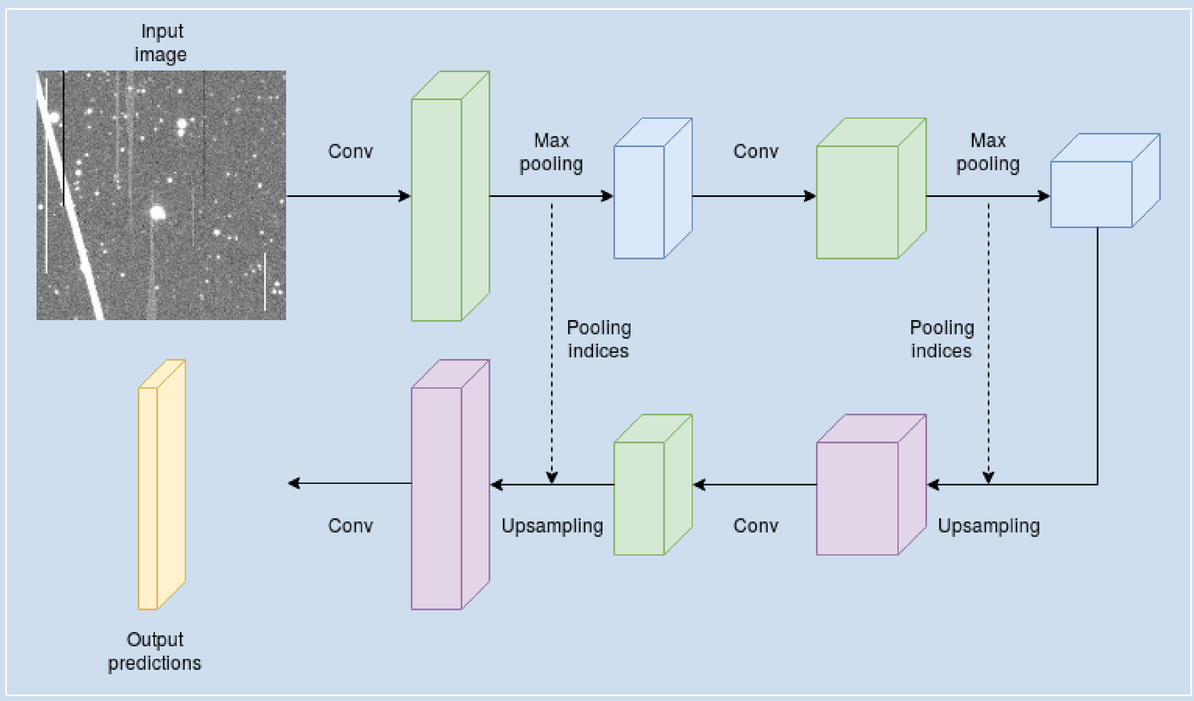

Figure 1. Architecture of the neural network

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: a first weight map is computed where each pixel is assigned a weight inversely proportional to its class fraction in the training set. Then this weight map is smoothed using a 3x3 gaussian kernel. This gives the final weights used in the cost function.

Results

We show here some qualitative results:

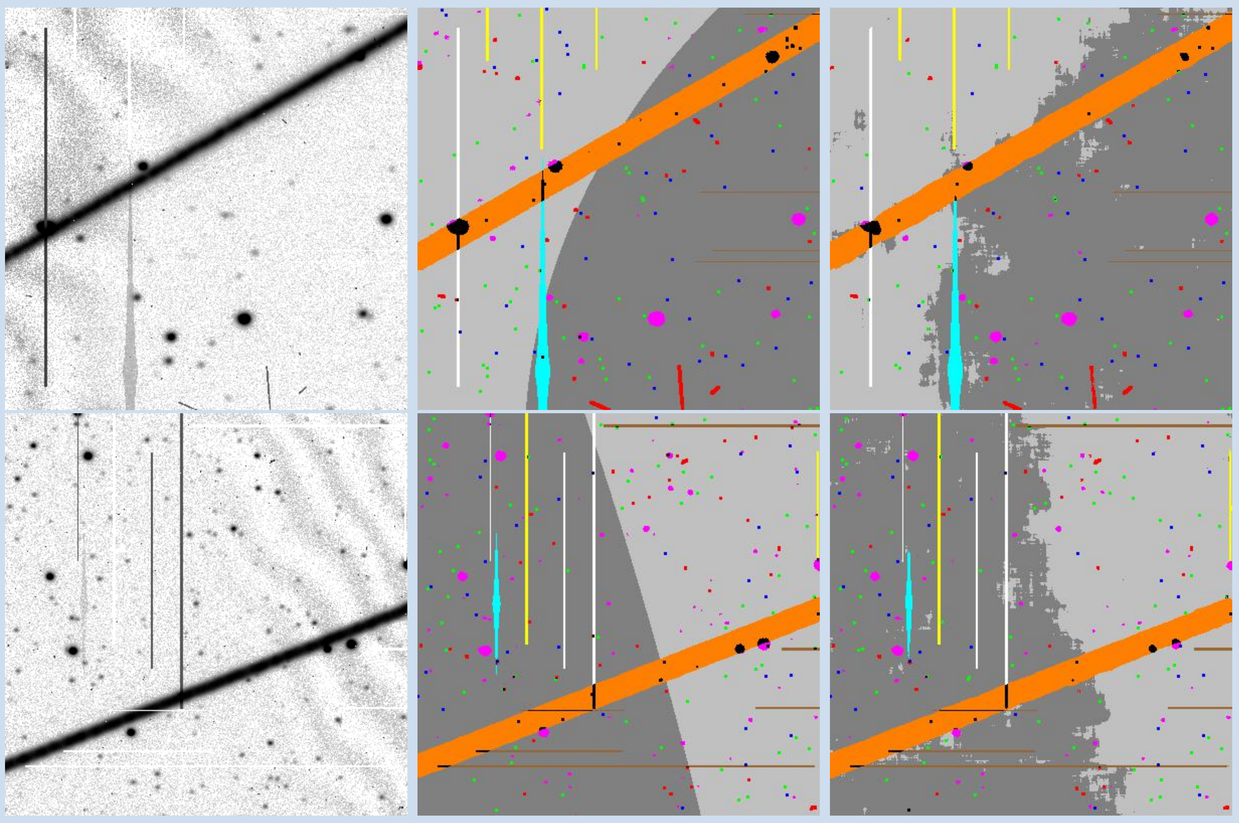

Figure 2. Example of qualitative results: left: input image, center: ground truth, right: prediction

The prediction maps were built by assigning to class c all the pixels whose probabilities to belong to class c are higher than some threshold. This threshold was chosen as the best threshold in the sense of the MC coefficient [4] which is an accuracy score reliable even with class imbalance. Though, one is free to use different thresholds to constrain a particular true positive or false positive rate on its desired class.

Conclusion

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), explore more ways to resolve the strong class imbalance that biases the training procedure and explore ways to modify the output probabilities of the neural network to adapt to different expected class proportions in the data.

References - Acknowledgements

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283, 2016.

[2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[3] H Bouy, E Bertin, E Moraux, J-C Cuillandre, J Bouvier, D Barrado, E Solano, and A Bayo. Dynamical analysis of nearby clusters-automated astrometry from the ground: precision proper motions over a wide field. *Astronomy & Astrophysics*, 554:A101, 2013.

[4] Brian W Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451, 1975.

[5] Pieter G van Dokkum, J Bloom, and Malte Tewes. La cosmic: Laplacian cosmic ray identification. *Astrophysics Source Code Library*, 2012.

CNES CENTRE NATIONAL D'ÉTUDES SPATIALES

INITIATIVE D'EXCELLENCE Université de Bordeaux

European Commission

European Research Council
Established by the European Commission