Galaxy classification using convolutional networks

1 Description

Convolutional neural networks [1, 2, 3] are ones of the most employed DNN architectures. They are particularly efficient for computer vision tasks such as image classification [4, 5].

Galaxy classification [6, 7] consists of, given an image, predict the probability that it belongs in a particular galaxy class (generally determined by its morphology).

2 Objectives

The goal of the project is to design a convolutional network that outputs the probability that a given galaxy image belongs to one of the possible categories. This is a supervised classification problem. The dataset was used for one of the Kaggle challenges[1], as commented in class.

The student should: 1) Preprocess the images. 2) Design the network architecture and train it. 3) Validate the network.

Not all images in the training set have to be necessarily used for the project. The student can decide whether to use all or a subset of them. Around 10000 should be sufficient.

As in other projects, a report should describe the characteristics of the design, implementation, and results. A Jupyter notebook should include calls to the implemented function that illustrate the way it works.

3 Suggestions

- Review the description of the galaxy morphology classification problem `https://arxiv.org/pdf/1308.3496.pdf`.
- You could review a number of previous approaches to the galaxy classification problem with convolutional networks as the following
  `https://arxiv.org/abs/1709.02245`
  `https://arxiv.org/pdf/1711.04573.pdf`
  `https://arxiv.org/abs/1711.05744`
- Implementations can use any Python library that implements DNNs.

References

[1] Dan C Ciresan, Ueli Meier, Jonathan Masci, Luca Maria Gambardella, and Jürgen Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 1237. Barcelona, Spain, 2011.

[2] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature em-

[1] The data is available from `https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge/data`

bedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678. ACM, 2014.

[3] Yoon Kim. Convolutional neural networks for sentence classification. *CoRR*, abs/1408.5882, 2014.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[6] Jorge De La Calleja and Olac Fuentes. Machine learning and image analysis for morphological galaxy classification. *Monthly Notices of the Royal Astronomical Society*, 349(1):87–93, 2004.

[7] Kyle W Willett, Chris J Lintott, Steven P Bamford, Karen L Masters, Brooke D Simmons, Kevin RV Casteels, Edward M Edmondson, Lucy F Fortson, Sugata Kaviraj, William C Keel, et al. Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 435(4):2835–2860, 2013.