Self-supervised Learning for Astronomical Image Classification

Ana Martinazzo   Mateus Espadoto   Nina Hirata

University of São Paulo

Intro

- Modern Astronomy is a data-intensive discipline
- Astronomical data are abundant, rich and unique, but mostly unlabeled
- Labelling such data is costly, since it requires domain knowledge
- Automated data analysis techniques, such as Machine Learning and Deep Learning, are needed to extract useful information from these data

Astronomical Data

Science-ready data products may come in various formats:

- raw images (may have dozens of channels)
- RGB composite images (generated from raw images)
- tabular catalogues of properties (generated from raw images)
- spectra

The core idea of our work is combining images and astronomical properties into a single representation.

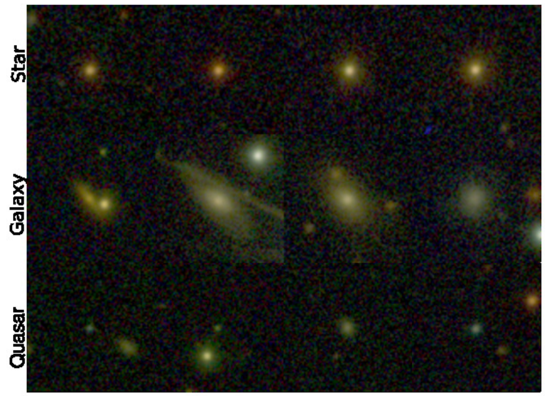

Figure 1:Sample images from the SGQ dataset, which contains 54k images.

Our Self-Supervised Approach

Pretext (regression) task

Downstream (classification) task

Results

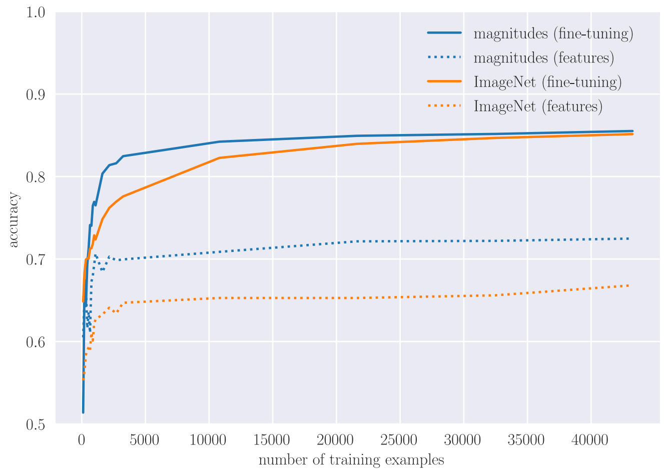

Figure 2:Accuracy curves as a function of the number of training examples for the SGQ dataset.

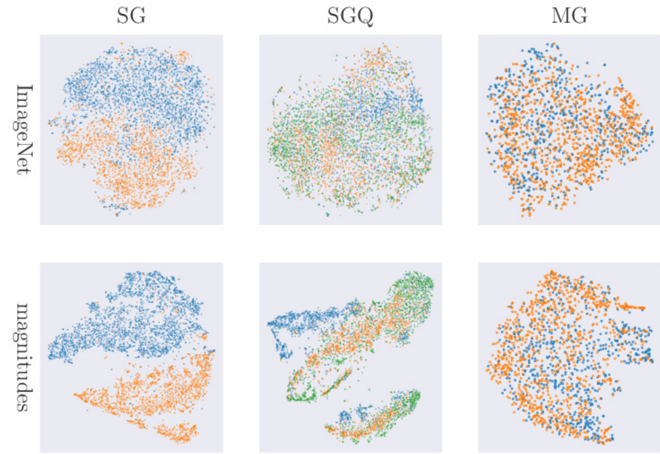

Figure 3:t-SNE projections of learned representations.

Conclusions

Our method is advantageous when:

- there are few labeled examples
- ImageNet weights for the CNN architecture of interest are not available
- extracting features for unsupervised tasks, such as clustering

Future Work

Possible directions for future work are: extending our analyses for raw images, using different properties as targets for self-supervised learning, and evaluating learned representations in unsupervised tasks.

Acknowledgments. This study was financed in part by FAPESP (2015/22308-2, 2017/25835-9 and 2018/25671-9) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.