

The WND-CHARM Python API

Pythonic Pattern Recognition for Bioimage Informatics

Christopher E. Coletta, Ilya G. Goldberg

Image Informatics and Computational Biology Unit –

National Institute on Aging/NIH

Baltimore, MD

christopher.coletta@nih.gov, igg@nih.gov

Abstract—We present a Python toolkit based on our existing WND-CHARM pattern recognition software originally developed in C/C++ for analyzing biomedical image data. This toolkit provides modular access to the pattern recognition workflow previously implemented as a monolithic pipeline, including class definitions based on image files, image feature extraction, dimensionality reduction, classification, and evaluation tools for image classification experiments. The Python API facilitates the automation of many types of image classification experiments by providing a simple scripting interface to implement custom workflows as well as rapid prototyping of workflow components. This toolkit performs parallelized, on-demand image feature extraction, and simplifies integration with scientific image repositories such as OMERO and Bisque, machine learning libraries such as Scikit-learn, and interactive tools such as IPython/Jupyter. The 5D nature of image data complicates the definition and construction of a 2D matrix of image features used as input by machine learning algorithms. A new spreadsheet interface offers bench scientists fine-grained control over the definition of an image-based pattern-recognition experiment

without requiring any scripting or programming. This specification allows defining how multi-plane image files are sampled for feature extraction, assembling multiple image planes into multi-dimensional feature vectors, and controlling hierarchical grouping of image samples. This spreadsheet interface also allows the user to easily redefine the class structures in experiments where image samples can be analyzed along multiple experimental axes. Case studies presented to demonstrate this functionality will include a sliding-window classifier to generate heat map overlays and multi-channel feature sets used to analyze histological images stained with Hematoxylin and Eosin. This open-source toolkit and example Jupyter notebooks are available at <http://github.com/wnd-charm>.

Index Terms—computer vision, pattern recognition, machine learning, Python, toolkit

Funding— This research was supported entirely by the Intramural Research Program of the NIH, National Institute on Aging.