OXFORD

Bioimage informatics

Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification

Ignacio Arganda-Carreras[1,2,3,*], Verena Kaynig[4], Curtis Rueden[5], Kevin W. Eliceiri[5], Johannes Schindelin[5], Albert Cardona[6] and H. Sebastian Seung[7]

[1]Ikerbasque, Basque Foundation for Science, Bilbao 48013, Spain, [2]Computer Science and Artificial Intelligence Department, Basque Country University, San Sebastian 20018, Spain, [3]Donostia International Physics Center, San Sebastian 20018, Spain, [4]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA, [5]Laboratory for Optical and Computational Instrumentation, University of Wisconsin, Madison, WI 53706, USA, [6]Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, VA 20147, USA and [7]Neuroscience Institute and Computer Science Department, Princeton University, NJ 08544, USA

*To whom correspondence should be addressed.
Associate Editor: Robert Murphy

Abstract

Summary: State-of-the-art light and electron microscopes are capable of acquiring large image datasets, but quantitatively evaluating the data often involves manually annotating structures of interest. This process is time-consuming and often a major bottleneck in the evaluation pipeline. To overcome this problem, we have introduced the Trainable Weka Segmentation (TWS), a machine learning tool that leverages a limited number of manual annotations in order to train a classifier and segment the remaining data automatically. In addition, TWS can provide unsupervised segmentation learning schemes (clustering) and can be customized to employ user-designed image features or classifiers.

Availability and Implementation: TWS is distributed as open-source software as part of the Fiji image processing distribution of ImageJ at http://imagej.net/Trainable_Weka_Segmentation.

Contact: ignacio.arganda@ehu.eus

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

With the progress of microscopy techniques and the fast growing amounts of acquired imaging data, there is an increased need for automated image analysis solutions in biological studies. Prior to analysis, structures of interest must be detected and defined according to a representation suitable for quantification by the computer. This is achieved through segmentation, the process of partitioning an image into multiple homogeneous regions or segments. Segmentation constitutes a major transition in the image analysis pipeline, replacing intensity values by region labels.

Most traditional segmentation methods are based on the intensity and spatial relationships of pixels, or constrained models found by optimization. Nonetheless, humans use much more knowledge when performing manual segmentation. For that reason, in recent years,

trainable machine learning methods have emerged as powerful tools to include part of that knowledge in the segmentation process and improve the accuracy of the labeled regions. Just a few software platforms partially provide both machine learning and image processing tools. These include commercial platforms (e.g. MATLAB, MathWorks, Natick, MA) and open-source platforms, e.g. the Konstanz Information Miner (KNIME) (Dietz and Berthold, 2016) and CellProfiler (Kamentsky *et al.*, 2011). Commercial platforms usually target inexperienced users and a wide range of image types. However, the details of the algorithms are hidden, which is undesirable for use in scientific research. Conversely, those details are available in open-source platforms but many of them are developed primarily by and for the machine learning community, provide only a minimal set of image tools or are focused on
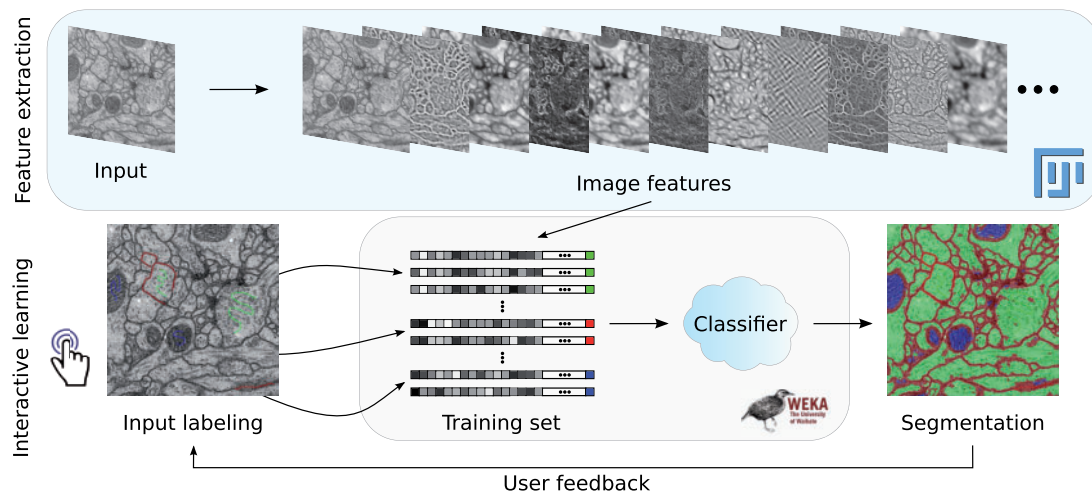
Fig. 1. TWS pipeline for pixel classification. Image features are extracted from an input image using Fiji-native methods. Next, a set of pixel samples is defined and represented as feature vectors, and a WEKA learning scheme is trained on those samples and finally applied to classify the remaining image data. The user can then interactively provide feedback by correcting or adding labels. The input image in this example pipeline is a serial section from a transmission electron microscopy dataset from the Drosophila first instar larva ventral nerve cord; its pixels are divided into three classes: membrane, mitochondria and cytoplasm

algorithms and data structures and do not provide visualization tools or user-friendly interfaces. Among the bioimage informatics tools using machine learning for image segmentation we find Ilastik (Sommer *et al.*, 2011), which contains a powerful interface to supply user feedback, although is limited to a small set of classifiers; the Vaa3D plugin for interactive cell segmentation (Li *et al.*, 2015); and the Cytomine data mining module (Marée *et al.*, 2016).

To address this gap in the field, we started the open-source software project Trainable Weka Segmentation (TWS). The project combines the popular image processing toolkit Fiji (Schindelin *et al.*, 2012), with the state-of-the-art machine learning algorithms provided in the latest version of the data mining and machine learning toolkit Waikato Environment for Knowledge Analysis (WEKA) (Hall *et al.*, 2009).

2 Materials and methods

2.1 Machine learning approach

To segment the input image data (2D/3D grayscale or color), TWS transforms the segmentation problem into a pixel classification problem in which each pixel can be classified as belonging to a specific segment or class. A set of input pixels that has been labeled is represented in the feature space and then used as the training set for a selected classifier. Once the classifier is trained, it can be used to classify either the rest of the input pixels or completely new image data (see Fig. 1). All methods available in WEKA can be used. These include a large variety of supervised classification and regression algorithms and clusterers.

2.2 Image features

TWS includes a wide range of image features, most of which are extracted using common filters or plugins available as part of Fiji. The user has complete freedom to select features and tune their scales and optional parameters using either the settings dialog in the GUI or specific library methods. Based on their purpose, the features available in TWS can be categorized as: *edge detectors*, which aim at indicating boundaries of objects in an image (e.g. Laplacian and Sobel filters, difference of Gaussians, Hessian matrix eigenvalues and Gabor filters); *texture filters*, to extract texture information (including filters such as

minimum, maximum, median, variance, entropy, structure tensor, etc.); *noise reduction filters*, such as Gaussian blur, bilateral filter, Anisotropic diffusion, Kuwahara and Lipschitz; and *membrane detectors*, which localize membrane-like structures of a certain size and thickness. In addition, TWS allows users to customize features. As described in its online wiki, a very simple script is needed to include user-defined features in the segmentation process, alone or in combination with the existing filters. This opens the door to all kinds of linear and nonlinear features that users can externally create.

2.3 GUI and library use

One possibility is to use the GUI following an active learning approach with a small number of annotations. The user is allowed to interactively provide training samples while navigating the data, obtain on-the-fly test results, and retrain the classifier as many times as needed. In this way, the user can fine-tune the parameters of the classifier and select labels until achieving satisfactory results. More classical (not interactive) approaches are also available via the library methods, allowing training on arbitrarily large labeled data. Both GUI and library methods supply two types of outputs: a segmentation and a class-probability map.

3 Conclusion and perspectives

TWS is a versatile tool for pixel classification. As a pixel classifier, it has a wide range of applications such as boundary detection, semantic segmentation, or object detection and localization. The software contains a library of methods and a GUI that makes it easy to use without any programming experience (see Supplementary Material). In particular, this toolbox is an important addition to the growing arsenal of segmentation plugins in Fiji for analyzing biological and nonbiological image data. TWS is designed to help developers as well by facilitating the integration of machine learning schemes with image processing modules into a pipeline. Researchers can easily prototype segmentation algorithms using TWS methods with any of the scripting languages available in Fiji. The usefulness of TWS has already been demonstrated by its utilization in many scientific publications (more than 100 according to Google Scholar) since its first release,

and its future is guaranteed by the thriving community of ImageJ/Fiji users and developers.

Conflict of Interest: none declared.

References

Dietz,C. and Berthold,M.R. (2016) Knime for open-source bioimage analysis: a tutorial. In: DeVos,W.H. *et al.* (eds) *Focus on Bio-Image Informatics*. Springer-Verlag Berlin, pp. 179–197.

Hall,M. *et al.* (2009) The weka data mining software: an update. *ACM SIGKDD Explor. Newslett.*, **11**, 10–18.

Kamentsky,L. *et al.* (2011) Improved structure, function and compatibility for cell-profiler: modular high-throughput image analysis software. *Bioinformatics*, **27**, 1179–1180.

Li,X. *et al.* (2015). Interactive exemplar-based segmentation toolkit for biomedical image analysis. In: *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pp. 168–171. IEEE.

Marée,R. *et al.* (2016) Collaborative analysis of multi-gigapixel imaging data using cytomine. *Bioinformatics*, **32**, 1395–1401.

Schindelin,J. *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods*, **9**, 676–682.

Sommer,C. *et al.* (2011) ilastik: Interactive learning and segmentation toolkit. In: *8th IEEE International Symposium on Biomedical Imaging (ISBI 2011)*.