Analysis and Archiving of High-Dimensional Bio-Imaging Data: Towards an Open, Extensible Framework.

Ahmed E. Fetit[1], Sebastien Besson[2], Ewan McDowall[1] and Ann Wheeler[1]
[1]Institute of Genetics and Molecular Medicine, University of Edinburgh, Scotland, UK
[2]Centre for Gene Regulation & Expression, University of Dundee, Dundee, Scotland, UK

Motivation and Aim Advances in imaging systems have meant that microscopy is increasingly becoming a big data science. High content screening and digitisation of pathology mean that it is now possible to acquire hundreds of thousands on images in multiple colours and 3D. This gives rise to the need to ensure that imaging information are paragmatically mined, correlated and archived, without relying on expensive licenses. The aim of this work is to demonstrate how a bio-image archiving and analysis framework, with automated learning functionalities, can be designed using open-source solutions.

Methods The presented framework is based on the Open Microscopy Environment Remote Objects (OMERO 5.1) [1], ImageJ 1.50b (Fiji distribution) [2] and Weka's Trainable Segmentation tool [3]. To set up a test environment, we used an OMERO virtual appliance, running on Ubuntu, which was virtualised on an Oracle VirtualBox workstation. The use of OMERO's open client-server platform allows for rapid visualisation of microscopy imaging data and organising them under project categories, through various layers of access control. The platform enables integration of the core archive with ImageJ through the *bioformats* and *OMERO.insight-ij* .jar plugins.

Fig 1. Proposed analysis and archiving framework.

Integration with ImageJ means that users do not need to download imaging files on their workstations for analysis, thus addressing data duplication or loss issues that are common in research workflows. To achieve automated machine learning functionalities, such as rapid classification of tissue regions using prior training knowledge, Weka's trainable segmentation tool, a plug-in that can be integrated with ImageJ, can be used. This is particularly useful if users are looking to automate processing of large datasets through manual training of a small subset and using supervised classification on the remaining images. Figure 1 shows how a histopathological image can be rapidly used to train a Weka classifier, with the results being sent for archiving in OMERO.
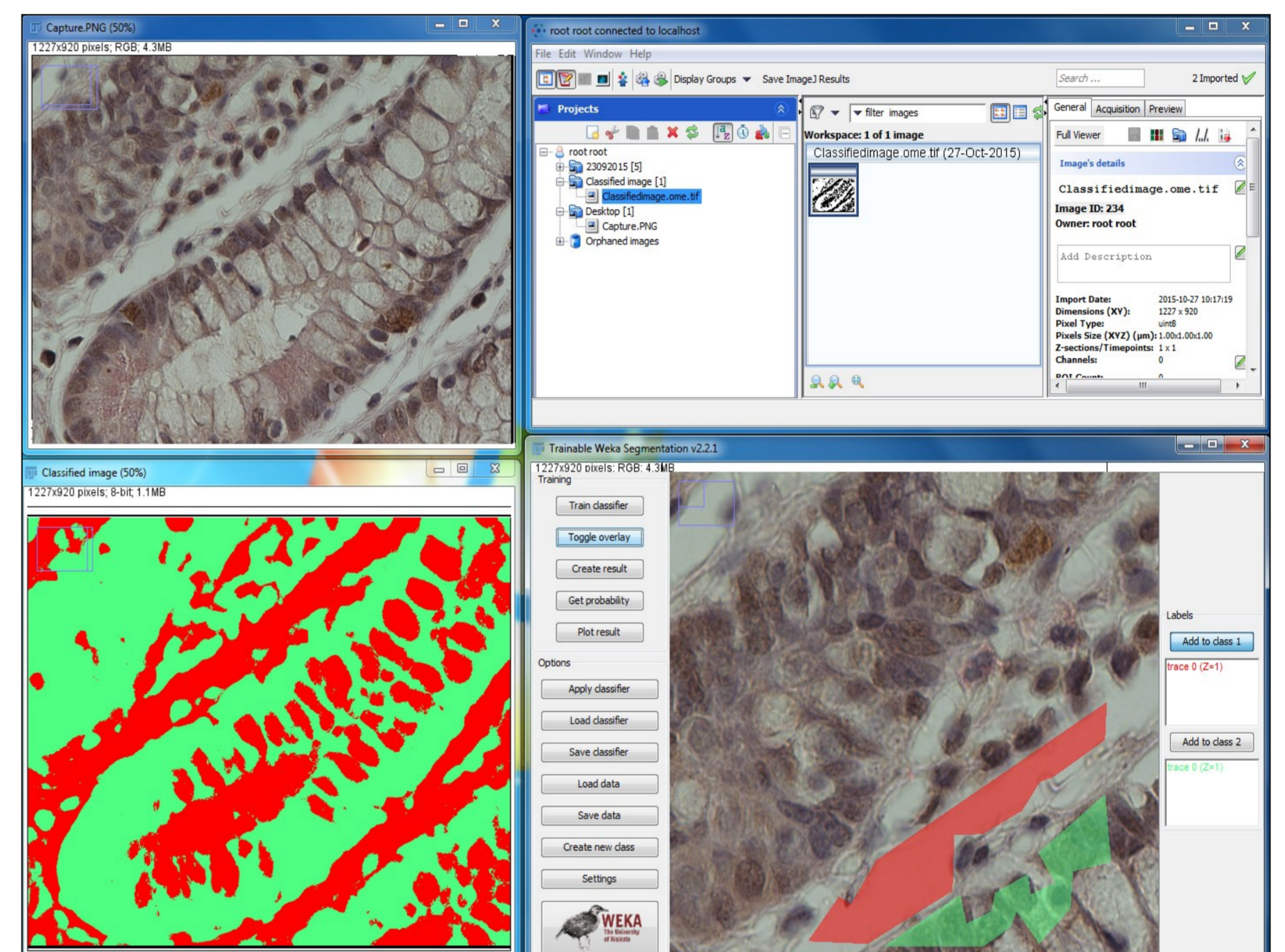
Fig 2. Supervised classification and archiving of a histopathological image.

Work in Progress A separate environment has been set up on a computer cluster, to allow for extensive testing of both the core archive and the learning functionalities. In doing this, getting usability metrics and feedback from potential users will be necessary.

1. OMERO, The University of Dundee and Open Microscopy Environment, Available from: http://www.openmicroscopy.org/site/products/omero
2. Schindelin, J.; Rueden, C. T. & Hiner, M. C. et al. (2015), "The ImageJ ecosystem: An open platform for biomedical image analysis", *Molecular Reproduction and Development*, PMID 26153368.
3. Trainable Weka Segmentation, Available from: http://fiji.sc/Trainable_Weka_Segmentation