Interactive classifier training via the new Viewer D5.6.1 – SGA2

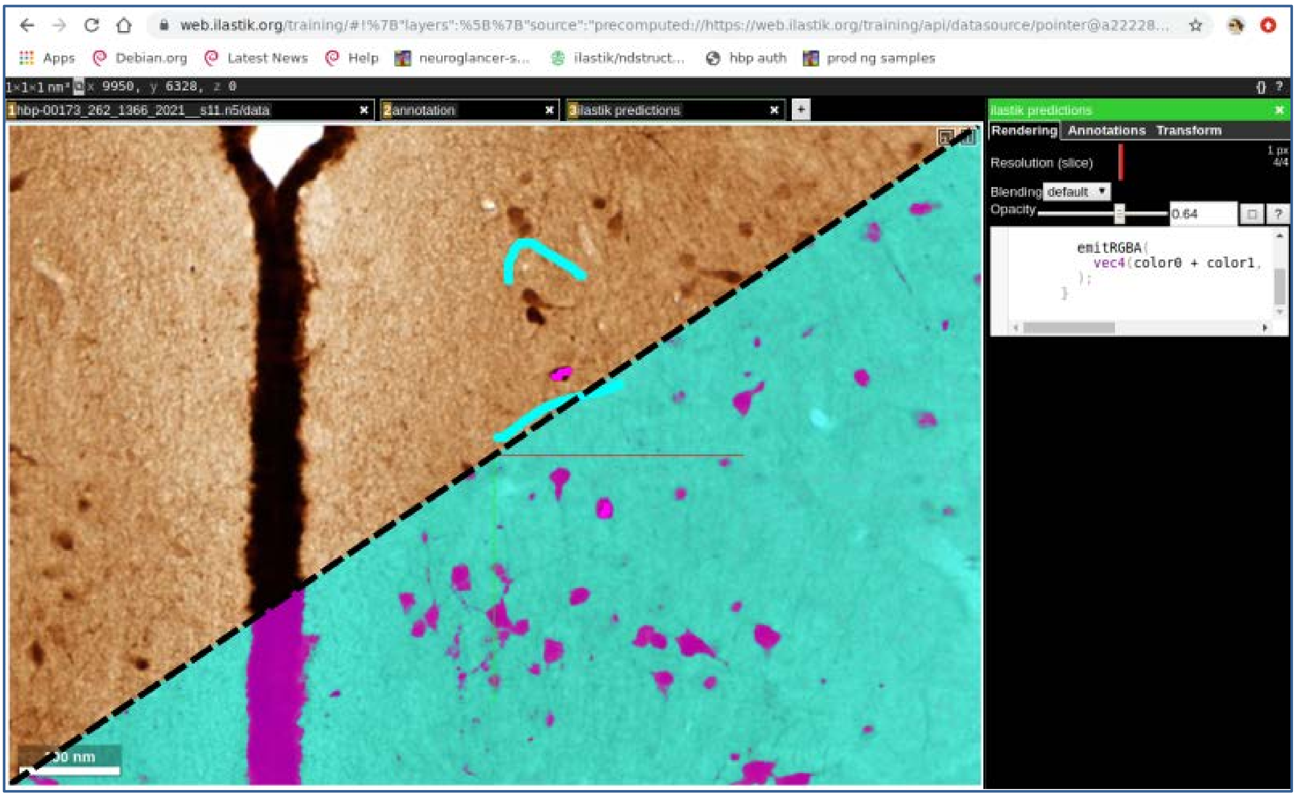

Figure 1: Screenshot of Neuroglancer viewer connected to the ilastik back end

Microscopic image of a histological section from a rat brain stained for Calbindin positive neurons. Ilastik is used to extract the location and area of the labelled neurons. The viewport has been artificially divided along a diagonal to enhance visibility of annotations (coloured scribbles in the upper left half) and predictions (lower right half) by a classifier trained using the annotations defined as "background" and "cells". Note, that annotations (scribbles) have been slightly enhanced to allow for better visibility. (Image from: Boccara, C., Kjonigsen, L., Hammer, I., Bjaalie, J. G., Leergaard, T. B., & Witter, M. (2019). ***Coronal section images series showing neuronal nuclei, calbindin and parvalbumin in the rat hippocampal region*** **[Data set]. Human Brain Project Neuroinformatics Platform. DOI: 10.25493/AMW1-Z16)**

Project Number:	785907	Project Title:	Human Brain Project SGA2

Document Title:	D5.6.1 (D33.1, D77) Interactive classifier training via the new Viewer
Document Filename:	D5.6.1 (D33.1 D77) SGA2 M22 ACCEPTED 200730.docx
Deliverable Number:	SGA2 D5.6.1 (D33.1, D77)
Deliverable Type:	Report
Work Package(s):	WP5.6
Key Result(s):	KR5.11
Dissemination Level:	PU
Planned Delivery Date:	SGA2 M22 / 31 Jan 2020
Actual Delivery Date:	SGA2 M26 / 06 May 2020; Accepted: 30 Jul 2020
Authors:	Anna KRESHUK, EMBL (P123)
Compiled by:	Anna KRESHUK, EMBL (P123)
Contributors:	Dominik KUTRA, EMBL (P123)
SciTechCoord Review:	Martin TELEFONT, EPFL (P1)
Editorial Review:	Annemieke MICHELS, EPFL (P1)
Description in GA:	Users can train machine learning algorithms interactively by giving annotations in the web-based Viewer
Abstract:	This report describes a new web-based front end of the ilastik toolkit, which allows users to interactively train machine learning algorithms directly in the Neuroglancer 3D viewer. Neuroglancer serves as the base for the HBP atlas viewer NeHuBa. New annotation functionality has been implemented for Neuroglancer along with multiple changes in the ilastik computational back end. The back end runs on the HBP HPC resources.
Keywords:	ilastik, image segmentation, web viewer, interactive, machine learning
Target Users/Readers:	Neuroimaging community, Platform users, computer scientists

Table of Contents

Table of Figures

1. Interactive classifier training via a web-viewer

1.1 Introduction

Over the last years, machine learning has emerged as the strategy of choice for biomedical image analysis. The success of machine learning algorithms on major benchmarks in natural and biomedical image processing tasks is one of the two motivating factors for the work we present here. The other factor is the conceptual simplicity with which such algorithms can be adapted to new datasets: new training data triggers algorithm re-training with no further parameter changes from the user side. These methods are thus particularly attractive for user-facing software which is manipulated by non-experts such as life scientists without extensive computational training.

ilastik[1] is a tool that brings machine learning-based algorithms for image analysis to users without computational expertise. It contains workflows for image segmentation, object classification and tracking as well as object counting without detection. The core of ilastik lies in its lazy processing back end which allows users to interactively explore datasets larger than the available computer memory. Taking the example of segmentation of images, the training proceeds by users painting the annotations directly on the image, using different colours to designate different classes (for example, "background", "cell membrane", "synapse"). Internally, generic image filters are computed for the annotated pixels and the filters values along with the annotation ids are passed to the classifier. The classifier trains a Random Forest, which is then applied to the un-labelled pixels visible by the user. The user can correct the predictions, interactively causing the re-training of the classifier and creating a tight feedback loop with a much steeper classifier learning curve than observed in training from dense annotations.

Ilastik (https://www.ilastik.org) is a desktop application written in Python, with most computationally intensive parts in C++. In the HBP, the ilastik team has now integrated ilastik into the computational pipelines to be made available in the EBRAINS infrastructure. This entails integration with the web. Whereas on the desktop, ilastik uses its own viewer ("volumina") which allows fast navigation and annotation in up to 5-dimensional datasets, ilastik on the web uses the HBP interactive atlas viewer. It can be accessed from https://web.ilastik.org/live_training/ after logging in with your regular HBP account. Our aim is to *enable building of EBRAINS pipelines with ilastik as the image segmentation or feature extraction component* in a transparent way, with full utilization of HBP computational resources and without unnecessary duplication of data.

1.2 Implementation

1.2.1 Neuroglancer viewer extension

The HBP interactive atlas viewer for brain atlases, NeHuBa-UI (https://interactive-viewer.apps.hbp.eu; also see SGA2 Deliverable D5.4.1 (D31.1 D75)) is an extension of the Neuroglancer viewer (https://opensource.google/projects/neuroglancer). Neuroglancer is a viewer for volumetric data, based on WebGL. Besides the regular image planes, it can display arbitrary (not aligned to the axis) slicing of the volume, as well as 3D meshes and line segment markers ("skeletons"). As the viewer is released under the Apache 2.0 license, it can easily be extended for other use cases, including those relevant to researchers in HBP. The viewer is developed by Google (although it is not an official Google product) for the needs of the nanoscale connectomics community.

[1] P2348 *(in validation)* Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M. and Eren, K., 2019. ilastik: Interactive machine learning for (bio) image analysis. *Nature Methods*, pp.1-7

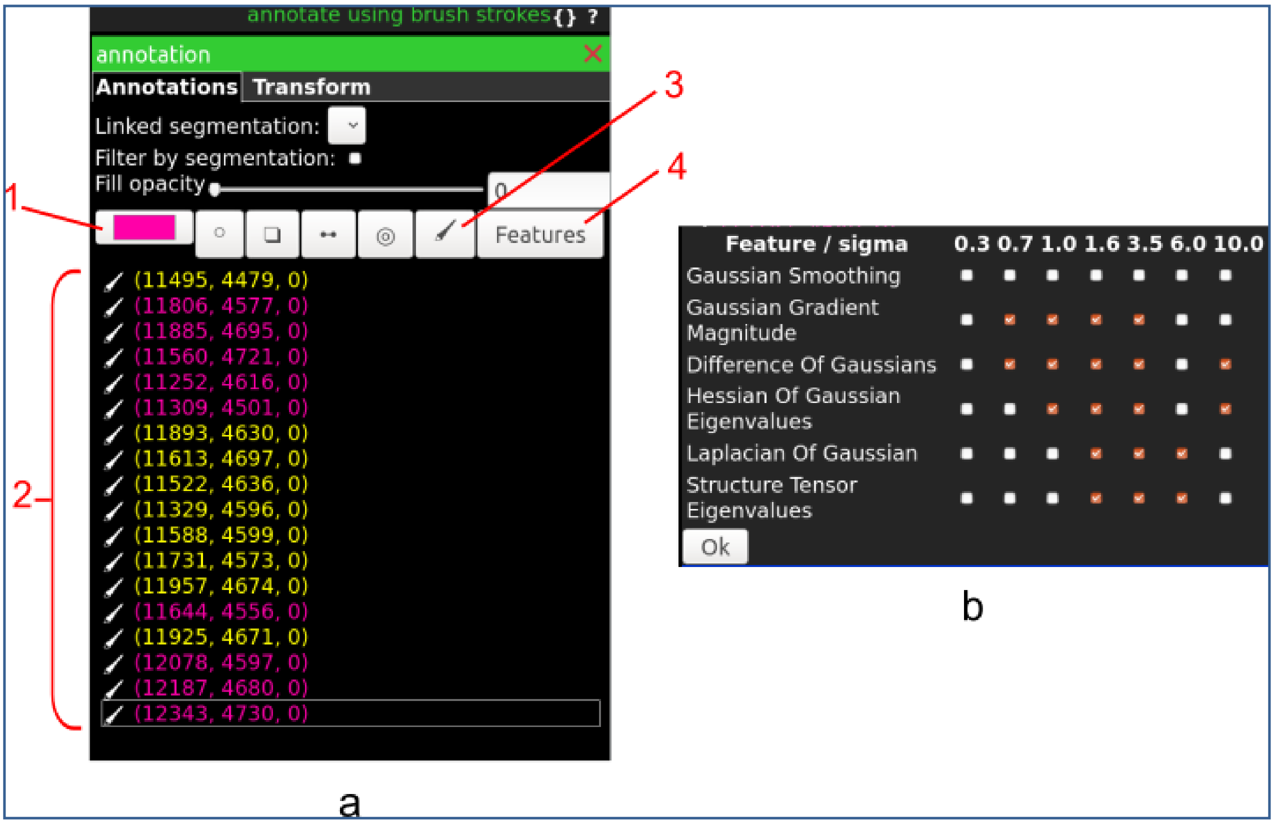

Figure 2: Custom controls in the ilastik web app.

(a) Annotation controls. By pressing on button 1, the user can choose a colour for the annotation. By pressing 3, the user enters the brushing annotation mode. All individual annotations are added to the list 2. Users can click on an annotation to focus on it. Annotations can also be removed individually. Pressing button 4, opens the feature selection control (b). Different features in scales from 0.3 to 10.0 can be selected by clicking the corresponding checkboxes.

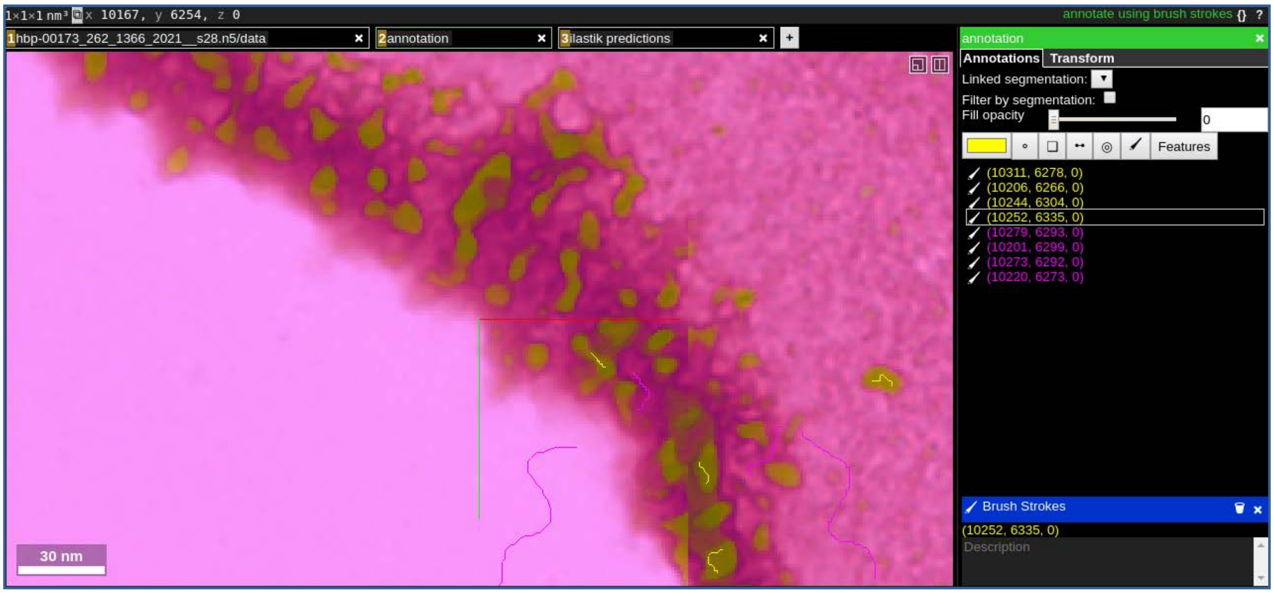

Figure 3: Live predictions overlaid over raw data in the ilastik web app.

Classifier trained on the back end with the user-added annotations.

Neuroglancer delivers a substantial set of features to serve as a web viewer for ilastik. The main missing component is the pixel-level annotation ilastik needs for training the classifier. In particular, to mimic the functionality of the ilastik volumina viewer, we needed to develop a tool to add brushstrokes of different colour and delete them afterwards. This has been implemented by extending the native Neuroglancer annotation tools with a possibility to draw directly on image pixels instead of vector graphics-based approach. Pixel coordinates are sent directly to the ilastik back end, which returns a ready-to-render image to the viewer. Selection of features is performed in another widget. We advise users to select all features first and to de-select them one by one once a reasonable prediction accuracy has been reached as evaluated by visual inspection in areas without annotations. Figure 2 shows the user controls for (a) brush annotations and (b) features selection. Figure 3 shows the result of user annotations as an overlay image showing predictions in colours corresponding to the annotations that were used to train the classifier.

The annotation and prediction process can be seen in this video: https://youtu.be/bQ0G83ucWCk.

1.2.2 ilastik as a server

The original ilastik desktop application is implemented tightly coupled to the ilastik viewer, volumina. In the course of separating front and back end, a subset of the ilastik back end functionality has been re-implemented in a lightweight architecture. This comes with several advantages over the original implementation. The advantages include a simplified API in order to connect to different front ends, the ability to use ilastik Pixel Classification as a library and streamlined scalability strategy. Furthermore, having ilastik as a web-based tool increases accessibility, by eliminating the need to install it on the client. We foresee that the new back end implementation for Pixel Classification will be extended to other workflows in the next funding phase.

1.2.3 Connection to the HPC infrastructure

The ilastik back end is hosted on CSCS OpenStack virtual machines. A storage back end has been implemented to make HBP data available. Users are authenticated against the EBRAINS authentication infrastructure. The developed ilastik-HPC connection allows both for interactive and batch jobs to be run at the CSCS resources.

1.2.4 Validation and user feedback

The implementation of the new server back end has been discussed with the core developers in the open source image analysis community, building on the collective expertise for the design of the API between imaging tools. After extensive internal validation, the online training functionality was presented at the 4th HBP student conference and the HBP Summit (Athens, Greece, 2020) and discussed with multiple users. It is already available publicly on the ilastik website: web.ilastik.org, but is not yet advertised widely, pending implementation of the current user requests. Next validation steps are planned as follows:

- Once own data upload functionality is implemented: recruitment of beta users in the wider ilastik community, initiation of discussion on the common image analysis forum, and active feedback solicitation.
- Presentation at the HBP summer schools where data analysis practical sessions are planned.

1.3 Future steps

While the current version of web-based ilastik is already providing value to users, further development is necessary for it to reach its full potential. Based on our experience with user-facing tools and the feedback of the first users of the web version, we envision the following steps:

- Better control of the brush width and more intuitive controls in general (such as adding an eraser)
- Full integration with the EBRAINS portal (https://ebrains.eu), including tutorial and guidelines
- Keyboard shortcuts, coordinated with the interactive atlas viewer (NeHuBa-UI)
- Overall interactive atlas viewer (NeHuBa-UI) integration