

An ImageJ Plugin for Whole Slide Imaging

Daniel Haak¹, Yusuf Z. Filmwala¹, Eric Heder¹, Stephan Jonas¹, Peter Boor²,
Thomas M. Deserno¹

¹Department of Medical Informatics, Uniklinik RWTH Aachen

²Institute of Pathology, Uniklinik RWTH Aachen

dhaak@mi.rwth-aachen.de

Abstract. Whole slide imaging (WSI) has become important in medicine and pathology, and challenges processing and management of high-resolution images with up to 10 GB of data. Open source tools such as ImageJ do not sufficiently support high volume data and require manual interaction in otherwise automatic workflows. We present an open source ImageJ plugin for automatic processing of Nanozoomer Digital Pathology Images (NDPI). In a batch-orientated workflow, the plugin provides an image processing pipeline including data conversion, segmentation, tiling, region of interest detection, thresholding, and quantification. The plugin is exemplarily applied to quantitative analysis of renal histology images. However, the general design supports other WSI file formats and analysis tasks.

1 Introduction

Whole slide imaging (WSI) plays an increasingly important role in medical diagnostics, educations and research [1]. Particularly in pathology, WSI is used in routine diagnostics and telemedicine. WSI yields high-resolution images of up to 10 GB uncompressed data, scanning the entire glass slide. Since this volume is exceeding common memory size of computers, specialized requirements have to be met by WSI tools and formats. Hence, WSI is separately specified in standards such as Digital Imaging and Communications in Medicine (DICOM) [2], which, however, are not sufficiently supported yet by the vendors.

Particularly for image processing in research, open sources tools such as the Insight Segmentation and Registration Toolkit (ITK), the Visualization Toolkit (VTK), the Medical Imaging Interaction Toolkit (MITK) [3], and ImageJ [4] have been established. ITK and VTK provide methods for image processing and visualization, respectively, and MITK combines their functionality for image-based medical interaction. However, all f these tools do not aim at WSI-analysis and corresponding WSI-formats stay unsupported. Although – in particular for digital pathology – various macros and plugins are provided [5], ImageJ still lacks in seamless handling of WSI. Researchers and physicians have to manually crop snapshots, iteratively perform analysis on each part separately, and then join the results, which is time-consuming and error-prone.

In this work, we present an open source plugin for ImageJ providing an automatic analysis pipeline for WSI. Our workflow follows the basic idea of

indeed performing analysis on large images but outsourcing processing steps to lower-resoluted images, if possible.

2 Material and methods

We demonstrate our plugin by quantifying Nanozoomer Digital Pathology Images (NDPI) to quantitatively assess the scarring tissue (fibrosis) of mice kidneys.

2.1 Application domain

Assessment of fibrosis in mice kidneys requires the quantification of immunohistochemical stained WSI. This is done by pixel counting of stained areas in renal tissue (Fig. 1). Here, the slide is stained for collagen type 3, a typical extracellular matrix protein deposited in exaggerated manner in kidney fibrosis. The extent of this deposition correlates with the extent of fibrosis and thereby with the extent of kidney injury and loss of kidney function. For automatization of this method, an WSI processing pipeline consisting of conversion, tissue segmentation, tiling, region of interest (ROI) and threshold setting, analysis and joining of results is provided.

2.2 NDPI format

WSI are acquired with specialized scanners for high-resolution recordings. The vendors provide proprietary file formats. For example, the NanoZoomer 2.0-HT (Hamamatsu Corporation, Japan)¹ acquires data in the pyramidal NDPI format. Specifications are unknown to the public, but an excerpted view has been published by the OpenSlide project². An NDPI file consists of a single Tagged Image File Format (TIFF) files with customized tags, including a pyramid of Joint Picture Expert Group (JPEG)-compressed TIFF-files. The various magnification levels are stored as a multi-plane image stack forming a so-called pyramid. A viewer is provided, which is limited for data processing or analysis only supporting navigation and zooming through the image.

¹ <http://www.hamamatsu.com/>

² <http://openslide.org/>

Fig. 1. Histological WSI of renal tissue after coloration of starring tissue area.

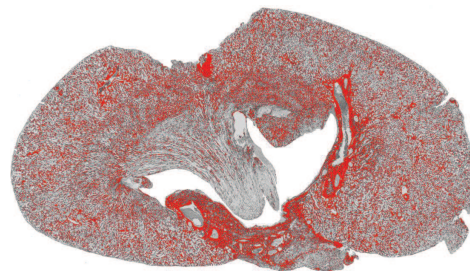


Table 1. Resolutions and data volume sizes and of a typical NDPI file.

Magnification	Resolution (in pixels)	Compressed (in MB)	Uncompressed (in MB)
×40	77.824 × 44.032	676,10	9.574,22
×10	19.456 × 11.008	44,30	627,45
×2.5	4.864 × 2.752	2,50	39,21
×0.625	1.216 × 688	0,17	2,45
×0.15625	304 × 172	0,01	0,15

In our application, the dimension of the highest resolution in the pyramid is 77.824×44.032 pixels, resulting in a data volume of approximately 670 MB and 9.5 GB in compressed JPEG and TIFF formats, respectively (Tab. 1).

2.3 Conversion

A couple of tools for handling of WSI file formats, such as NDPI, are offered by the open source community [6]. Bio-Formats³ of the Open Microscopy Environment consortium is a Java library, which supports up to 127 scientific image formats and their conversion to common formats, such as TIFF or JPG. Using Bio-Formats, the NDPI file is converted into multiple TIFF files, each file containing the data of one NDPI magnification level.

2.4 Tissue segmentation

Usually, tissue regions are relevant for analysis. Hence, the highest-resolved image is segmented into tissue and background areas. Several tissue segments can be processed consecutively. Nevertheless, the segmentation boundaries are calculated on a lower-resolved image and upscaled, since the highest resolution may exceed the free memory. This is done with ImageJ segmentation and masking functionality. The desired tissue boundary is extracted using likelihood features, which have been trained on a series of WSI. Then, the upscaled ROI boundaries of tissue areas (t-ROIs) are stored.

2.5 Tiling

Due to their large volume, tissue segmentation and further processing steps cannot be performed directly on the highest-resolution images. Hence, the images have to be tiled, which means building a physical tessellation of the original image on the hard disk, which allows reading it in parts instead of entirely. For this, the free memory space of the Java Virtual Machine (JVM) is determined and a maximum image size calculated. Using this information, valuable tiles are created and stored on the file system. Meanwhile, a Java object is created, which holds references of the tiled parts, tile sizes and (x, y) -coordinates

³ <http://www.openmicroscopy.org/site/support/bio-formats4/>

of tile placements. This information allows interpretation of tiles as one whole image. Among others, the Java Advanced Imaging (JAI) application programming interface (API)⁴ contains methods for tiling of images and processing of tiles. During tiling, the t-ROIs are considered and irrelevant background area is directly removed.

2.6 Input parameters

Usually, specific tissue ROIs need analysis rather than the entire area. Hence, tools for ROI selection are provided. The ImageJ ROI Manager⁵ provides extensive methods for user-interactive ROI drawing. Considering that according to tissue segmentation ROIs have to be defined on lower-resolution images and upscaled, the ImageJ ROI Manager has been modified with the user-selected ROI referred to as s-ROI.

Beside a ROI, a threshold is required for each image. Based on the gray scale, the threshold defines the pixels which have to be counted during analysis. The optimal threshold depends on specific image characteristics, which is visualized to the user by colorizing pixels in a lower resolution using ImageJ Threshold⁶ (Fig. 1).

2.7 Quantitative analysis and batch processing

For analysis, the entire image information is needed. Thus, analysis is performed on the highest resolution. Due to limited memory capacity, this image may have been tiled and tiles are processed sequentially. Recorded tile size and the (x, y) -coordinates support whole image assessment, representing the complete ROI. If the s-ROI has been set, a final ROI of each tile is computed, which is then processed with ImageJ standard methods.

Automated batch processing is provided by an image queue. Since batch processing in ImageJ is only supported by macro scripting, and in addition a fixed workflow with user-interaction is needed here at all, an graphical user interface (GUI) has been developed, which visualizes progress information and analysis results. For each tile or each tile-specific ROI, the ratio between the number of pixels with a gray value above the threshold and all ROI pixels is calculated, and the average ratio of all tiles is determined.

3 Results

The resulting workflow is composed of interactive and automatic modules (Fig. 2). The user adds multiple NDPI files to the batch. Conversion, tissue segmentation, and – depending on the free memory – tiling is performed. The tissue segments

⁴ <http://tinyurl.com/nevbveo>

⁵ <http://tinyurl.com/ldveabw>

⁶ <http://tinyurl.com/kwhof2m>

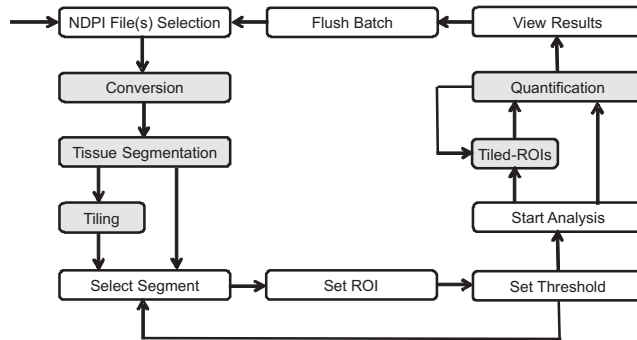


Fig. 2. Workflow of the proposed ImageJ plugin. White and gray boxes denote user interaction and automatic modules, respectively.

and associated NDPI file names are listed to the user, who can select individual segments and sets ROIs and thresholds repeatedly for any number of images in the batch. As soon as one image of the batch is ready for processing, the user starts automatic analysis of all images in the queue. The required parameters are provided automatically. Considering tiling, quantification is done on each tile and the ratio of stained area for all tiles is visualized. At any time, the user is allowed to flush the batch and to remove images and results.

Figure 3 shows the NDPI pyramid including the highest and second-highest magnifications levels, which are – after conversion – used by our ImageJ plugin for user-interaction and quantitatively analysis, respectively. The microscopy slide contains two tissue segments (t-ROIs), each including one s-ROI, which has been set by the user. For analysis, the s-ROIs are upscaled and projected on the highest-resolution image, resulting in a final ROI for all affected tiles.

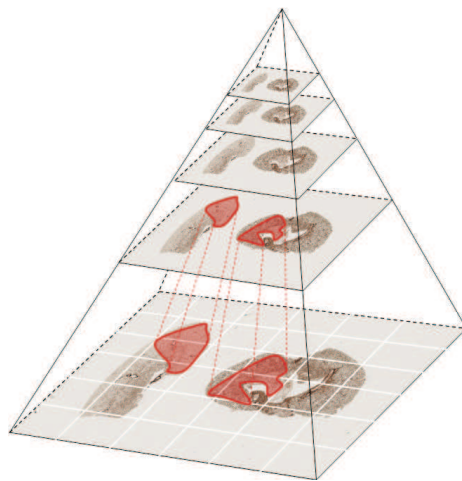


Fig. 3. NDPI pyramid with five magnification levels and user-selected s-ROIs on two tissue segments.

4 Discussion

In this paper, we presented an ImageJ plugin for automatic analysis of WSI and applied it to renal histology. Our method is based on a conversion and tiling of WSI in NDPI format and processing in ImageJ standard components. Compared to the physician's manual workflow, our automatized method is time-effective and less error-prone, in particular with respect to inter-individual observer variability. The batch-orientated workflow allows selection of input parameters for multiple tissue segments before starting the analysis.

Since the included Bio-Formats library supports conversion of various WSI file formats, also formats of other scanners, such as Leica LAS AF LIF, should be processable. However, tissue segmentation is currently optimized for renal histology data, but should be applicable on other tissue types, such as liver, by training of specific likelihood features. Furthermore, various analysis algorithm can be integrated. However, analysis is performed on each tile individually, which limits analyzing methods.

In future, our plugin should be extended by further integration of analysis algorithms for automatic cell and point counting on histological WSI. Support of DICOM's WSI standard will be integrated, too. The pyramidal structure (Fig. 3) is similar that of the DICOM Supplement 145, which will simplify DICOM compliance.

References

1. Daniel C, Rojo MG, Klossa J, et al. Standardizing the use of whole slide images in digital pathology. *Comput Med Imaging Graph.* 2011;35(7):496–505.
2. Singh R, Chubb L, Pantanowitz L, et al. Standardization in digital pathology: supplement 145 of the DICOM standards. *Pathol Inf J.* 2011;2(1):23.
3. Nolden M, Zelzer S, Seitel A, et al. The medical imaging interaction toolkit: challenges and advances. *Int J Comput Assist Radiol Surg.* 2013;8(4):607–20.
4. Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012;9(7):671–5.
5. Rangan GK, Tesch GH. Quantification of renal pathology by image analysis (Methods in Renal Research). *Nephrology.* 2007;12(6):553–8.
6. Deroulers C, Ameisen D, Badoual M, et al. Analyzing huge pathology images with open source software. *Diagn Pathol.* 2013;8(1):92.