A Python application programming interface for accessing Philips iSyntax whole slide images for computational pathology

Nita Mulliqi[1]   NITA.MULLIQI@KI.SE
Kimmo Kartasalo[1]   KIMMO.KARTASALO@KI.SE
Henrik Olsson[1]   HENRIK.OLSSON@KI.SE
Xiaoyi Ji[1]   XIAOYI.JI@KI.SE
Martin Eklund[1]   MARTIN.EKLUND@KI.SE
[1] *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden*

Lars Egevad[2]   LARS.EGEVAD@KI.SE
[2] *Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden*

Pekka Ruusuvuori[3,4]   PEKKA.RUUSUVUORI@UTU.FI
[3] *Institute of Biomedicine, Cancer Research Unit and FICAN West Cancer Centre, University of Turku and Turku University Hospital, Turku, Finland*
[4] *Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland*

Editors: Under Review for MIDL 2022

Abstract

Digital pathology has demonstrated its impact in improving diagnostics and prognostics in the field of pathology, through the utilization of deep learning algorithms. However, equipment from different scanner vendors used for digitizing the glass slides impose challenges for researchers due to non-interoperability between their proprietary formats. We have previously published OpenPhi (Open PatHology Interface), a Python Application Programming Interface providing seamless access to the iSyntax format of the Philips Ultra Fast Scanner, and in this short paper, we summarise its key features.

Keywords: Digital pathology, application programming interface, iSyntax.

1. Introduction

Digital pathology enables using Artificial Intelligence (AI) for improved diagnostics and prognostics of various diseases and for reducing inter and intra-observer variability among pathologists (Bulten et al., 2020; Ström et al., 2020). However, Whole Slide Images (WSI) are stored in various proprietary formats, complicating AI development. Previous vendor-neutral libraries, e.g. OpenSlide (Goode et al., 2012), lack support for the iSyntax format of the Philips IntelliSite scanner, gaining in popularity due to FDA approval. While Philips recently released a Software Development Kit (SDK) for accessing iSyntax, challenges remain. Firstly, the SDK represents a relatively low-level interface requiring familiarity with the iSyntax format. Secondly, developers cannot rely on a single, consistent application programming interface (API) due to the lack of integration in vendor-neutral libraries. Thus, we have released a Python API facilitating seamless access to iSyntax data, requiring minimal changes to applications designed to rely on OpenSlide (Mulliqi et al., 2021).
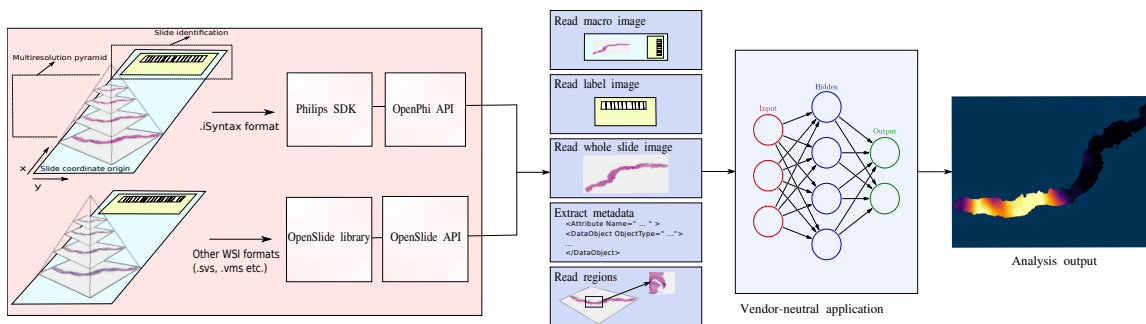
Figure 1: OpenPhi serves as a compatibility layer between Philips SDK and the vendor-neutral applications. It provides streamlined access to the images and metadata through its high-level methods. Figure obtained from (Mulliqi et al., 2021) under CC BY 4.0 (http://creativecommons.org/licenses/by/4.0/).

2. Materials and Methods

OpenPhi follows the structure of the iSyntax data model, where data is represented and accessed in three parts (1) label image with slide identification information, (2) macro image providing a thumbnail, low-resolution version of the entire glass slide, and (3) the WSI providing the regions of interest scanned at high resolution and subjected to the iSyntax compression scheme.

The recursive DWT employed in the iSyntax format uses the raw pixel data to construct a multi-resolution pyramid, where the pixel data can be extracted at multiple resolutions efficiently. Methods for extracting the pixel data are read_wsi, get_thumbnail and read_region. The read_wsi method reads the entire WSI at a desired level of resolution, and the read_region extracts uncompressed image data of a WSI in rectangular regions at any desired pyramid level. The label image associated with an iSyntax file contains identification numbers specific to the respective laboratory, and can be e.g. extracted using optical character recognition algorithms. Metadata attributes are extracted through the DICOM standard (4-byte DICOM tag) (Singh et al., 2011) or adhering to the OpenSlide generic properties format.

Since the highest resolution pyramid level 0 is the base level, all coordinates of the subsequent levels are specified with respect to the coordinate system of level 0, but with different downsampling factors on each level. In addition, image data is premultiplied by ARGB (Alpha, Red, Green and Blue) format. The names and input parameters of the methods adhere to the OpenSlide API, allowing developers to use existing code developed on top of OpenSlide without modifications.

3. Results and Conclusions

Images used for testing the API are scanned needle biopsies from patients in the prospective Stockholm-3 (STHLM3) study (Grönberg et al., 2015). Around 50,000 biopsies were scanned

with the Philips UFS and stored as iSyntax. During the experiments, we verified that our API can successfully interface the iSyntax format with a previously developed deep learning framework (Ström et al., 2020). OpenPhi serves as a powerful interface between the AI algorithms and the iSyntax format, thus scientific researchers can easily build computer vision methods and apply analysis modules with minimal effort on the proprietary format. OpenPhi is freely available at https://gitlab.com/BioimageInformaticsGroup/openphi.

Acknowledgments

References

Wouter Bulten, Hans Pinckaers, Hester van Boven, Robert Vink, Thomas de Bel, Bram van Ginneken, Jeroen van der Laak, Christina Hulsbergen-van de Kaa, and Geert Litjens. Automated deep-learning system for gleason grading of prostate cancer using biopsies: A diagnostic study. *The Lancet Oncology*, 21:233–41, 2020.

Adam Goode, Benjamin Gilbert, Jan Harkes, Drazen Jukic, and Mahadev Satyanarayanan. Openslide: A vendor-neutral software foundation for digital pathology. *Journal of pathology informatics*, 27, 2012.

Henrik Grönberg, Jan Adolfsson, Markus Aly, Tobias Nordström, Peter Wiklund, Yvonne Brandberg, James Thompson, Fredrik Wiklund, Johan Lindberg, Mark Clements, Lars Egevad, and Martin Eklund. Prostate cancer screening in men aged 50-69 years (sthlm3): a prospective population-based diagnostic study. *The Lancet Oncology*, 16:1667–76, 2015.

Nita Mulliqi, Kimmo Kartasalo, Henrik Olsson, Xiaoyi Ji, Lars Egevad, Martin Eklund, and Pekka Ruusuvuori. Openphi: An interface to access philips isyntax whole slide images for computational pathology. *Bioinformatics (Oxford, England)*, 37(21):3995–3997, 2021.

Rajendra Singh, Lauren Chubb, Liron Pantanowitz, and Anil Parwani. Standardization in digital pathology: Supplement 145 of the dicom standards. *Journal of pathology informatics*, 23, 2011.

Peter Ström, Kimmo Kartasalo, Henrik Olsson, Leslie Solorzano, Brett Delahunt, Daniel M Berney, David G Bostwick, Andrew J Evans, David J Grignon, Peter A Humphrey, Kenneth A Iczkowski, James G Kench, Glen Kristiansen, Theodorus H van der Kwast, Katia R M Leite, Jesse K McKenney, Jon Oxley, Chin-Chen Pan, Hemamali Samaratunga, John R Srigley, Hiroyuki Takahashi, Toyonori Tsuzuki, Murali Varma, Ming Zhou, Johan Lindberg, Cecilia Lindskog, Pekka Ruusuvuori, Carolina Wählby, Henrik Grönberg, Mattias Rantalainen, Lars Egevad, and Martin Eklund. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: A population-based, diagnostic study. *The Lancet Oncology*, 21:222–32, 2020.