Type: Bachelor Thesis

Title: Efficient WSI Storage and Streaming with WebP

Supervisor: Prof. Dr. Peter Schüffler (TUM), Computational Pathology

Keywords: digital pathology, resource saving programming

Goal

To write a tool for WebP conversion and reading of digital whole slide images (WSI) using Libvips and Openslide and to compare the performance of WebP to JPEG.

Description

Problem: Digital pathology is based on digital representations of high-resolution tissue images, scanned at 0.25µm/px (Figure 1). These whole slide images (WSI) can exceed 100.000 x 100.000 px, and are usually saved using JPEG compressed image tiles. Typically, and supported by all major scanner vendors, WSI are saved in a pyramidal format at full resolution and lower resolution levels. Each level consists of hundreds or thousands of small image tiles (e.g. 256x256px), that stitched together form the corresponding image layer. At the TUM, we are using .SVS files from Leica Scanners and .MRXS files from 3DHistech Scanners. With these image formats, pathologists can navigate through the WSI using slide viewers that stream the corresponding region of the image to the client (similar to Google Maps).

A WSI is typically 1-2 GB in size. Digital pathology archives thus can require petabytes for data storage, forming a challenge for many labs. Further, as a user navigates quickly through a WSI, many JPEG-compressed image tiles have to be loaded, which might be a problem for a consumer bandwidth *(1)*. As an efficient alternative to JPEG, WebP compression has emerged but has not yet been implemented in digital slide scanners.

Goal: In this project, we aim to investigate the potential of WebP for WSI, with following subgoals:

1. Develop a (Windows) conversion tool for conventional .SVS files into a new WebP BigTiff file format at a given tile size. (Optionally extend this tool for MRXS files). Libvips[1] can be used for conversion. This will require command line programming or scripting (R, Python…)
2. Extend Openslide (www.openslide.org)*(2)* to read WebP Generic TIFs. This requires C programming.
3. Using 1), quantify difference in File size for JPEG/WebP WSI. This requires a scripting language (R, Python, Matlab,…) and data visualization.
4. Using 1), quantify DeltaE for JPEG/WebP WSI (difference in pixelwise image appearance). This requires scripting (R, Paython, Matlab, …) and data visualization.
5. OPTIONAL: Using 2), quantify difference in Viewing / Streaming time for JPEG/WebP WSI in our pathology viewer*(3)*. This requires data base (Redis) parsing and data visualization.

Data: 60 WSI of a benchmark data set, scanned at 40x with Leica GT450 Dx, 3DHistech P1000 and Precipoint Scanner. More slides (>100k) are available if needed.

Computing: The student will use the LRZ high performance computing cluster for AI (GPU-based), and/or the group's computing infrastructure and/or a workstation.

[1] https://www.libvips.org/

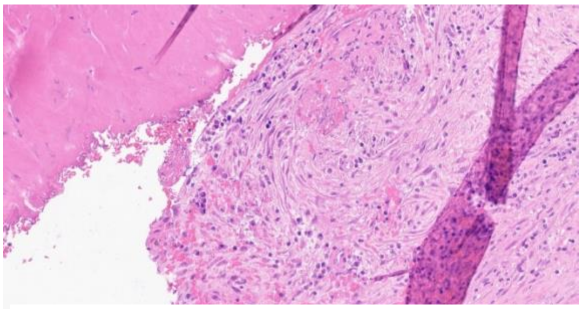

Figure 1: Example image of a WSI.

References

1. P. J. Schüffler, E. Stamelos, I. Ahmed, D. V. K. Yarlagadda, M. G. Hanna, V. E. Reuter, D. S. Klimstra, M. Hameed, Efficient Visualization of Whole Slide Images in Web-based Viewers for Digital Pathology. *Archives of Pathology & Laboratory Medicine* (2022), doi:10.5858/arpa.2021-0197-OA.

2. A. Goode, B. Gilbert, J. Harkes, D. Jukic, M. Satyanarayanan, OpenSlide: A vendor-neutral software foundation for digital pathology. *J Pathol Inform* **4**, 27 (2013).

3. P. J. Schüffler, L. Geneslaw, D. V. K. Yarlagadda, M. G. Hanna, J. Samboy, E. Stamelos, C. Vanderbilt, J. Philip, M.-H. Jean, L. Corsale, A. Manzo, N. H. G. Paramasivam, J. S. Ziegler, J. Gao, J. C. Perin, Y. S. Kim, U. K. Bhanot, M. H. A. Roehrl, O. Ardon, S. Chiang, D. D. Giri, C. S. Sigel, L. K. Tan, M. Murray, C. Virgo, C. England, Y. Yagi, S. J. Sirintrapun, D. Klimstra, M. Hameed, V. E. Reuter, T. J. Fuchs, Integrated digital pathology at scale: A solution for clinical diagnostics and cancer research at a large academic medical center. *Journal of the American Medical Informatics Association* **28**, 1874–1884 (2021).

TUM is an equal opportunity employer. TUM aims to increase the proportion of women, therefore, we particularly encourage applications from women. Applicants with severe disabilities will be given priority consideration given comparable qualifications.

Data Protection Information: As part of your application for a position at the Technical University of Munich (TUM), you submit personal data. Please note our privacy policy in accordance with Art. 13 General Data Protection Regulation (DSGVO).

Prof. Dr. Peter Schüffler · Trogerstraße 18 · 81675 Munich, Germany · https://schuefflerlab.org
TUM School of Computation (CIT) · TUM School of Medicine and Health · Munich Data Science Institute (MDSI)