tissueloc: Whole slide digital pathology image tissue localization

Pingjun Chen[1] **and Lin Yang**[1]

1 J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida

Background

`tissueloc` is an open-source Python package for fast and accurate tissue localization on whole slide image (WSI). Automatic pathology diagnosis using WSI gradually becomes a research hotspot in biomedical imaging domain (Cruz-Roa et al., 2014, Barker, Hoogi, Depeursinge, & Rubin (2016)). Because of the gigabyte size of WSI, instead of directly taking the WSI as input, patch-based strategy is commonly used to deal with WSI (Cruz-Roa et al., 2014, Hou et al. (2016)). As there are large amounts of background regions that are useless for diagnosis, researchers working on automatic WSI diagnosis can utilize `tissueloc` to locate genuine tissue regions and focus their analysis on these regions.

Overview

`tissueloc` mainly contains two functionalities: selecting proper low level image from WSI and tissue localization based on the selected low level WSI.

The width and height of WSI are far larger than 10,000 pixels. Locating tissue regions directly on WSI image is computationally expensive. However, based on the pyramid storage structure of WSI, we can select a proper low level image from the WSI for following tissue localization. The low level slide image can have much smaller size, thus can speed up the tissue localization process. Based on the setting of maximum width or height of the low level image, we select the level that its corresponding image has size smaller but closest to the setting.

Tissue localization is applied on the selected low level image based on a series of basic image processing techniques. The main procedures include: 1) Low level WSI loading. 2) Color space conversion from RGB to gray. 3) Inverse binarization to generate binary image. 4) Hole filling of the binary image. 5) Small object removal. 6) Contour finding.

The proposed WSI tissue localization is very efficient as it is entirely based on basic image processing techniques and applied on low level image, which could act as a preprocessing step for automatic WSI analysis. Researchers can focus their analysis on those patches inside the located tissue regions and avoid those irrelevant background regions.

Acknowledgement
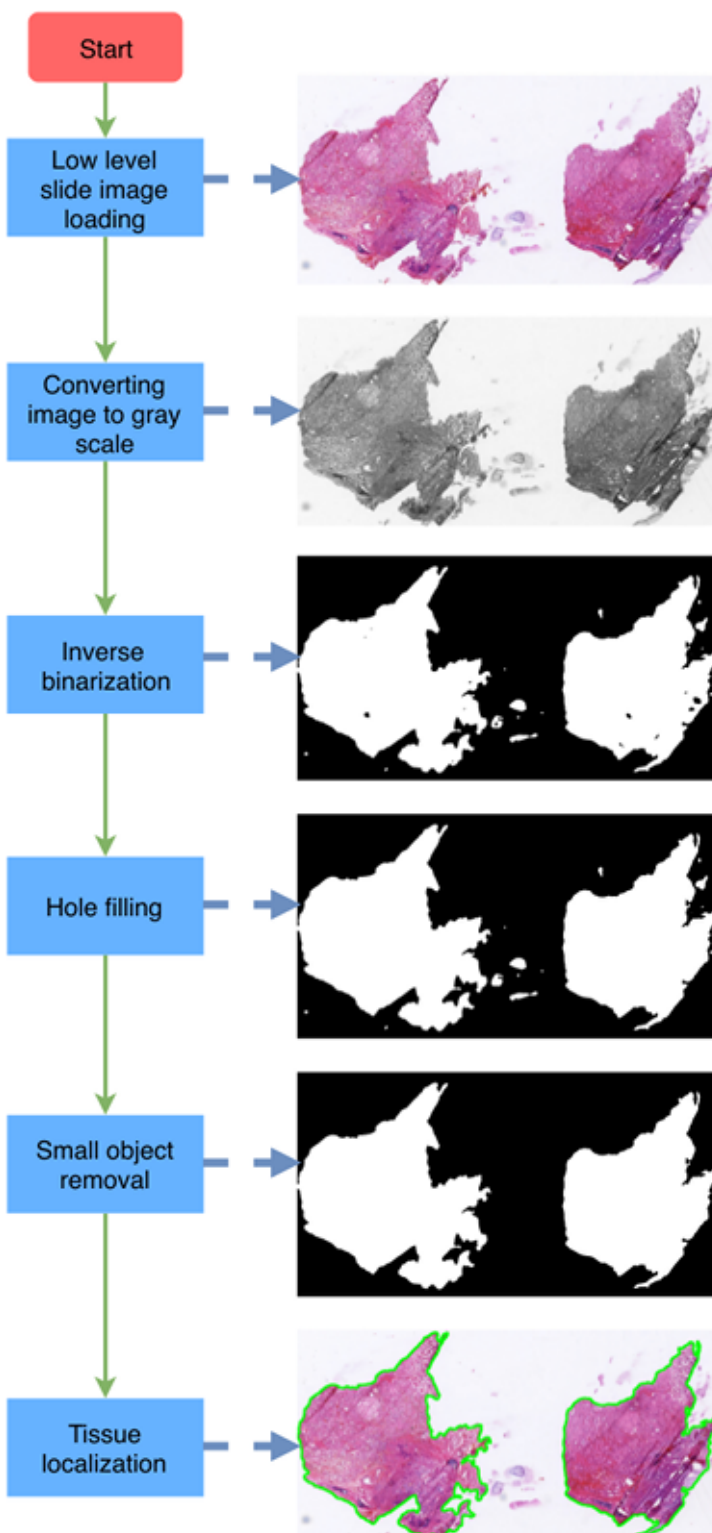
Figure 1: Tissue localization pipeline for whole slide image. The main procedures include: 1) Low level image loading. 2) Color space conversion. 3) Inverse binarization. 4) Hole filling. 5) Small object removal. 6) Contour finding.

References

Barker, J., Hoogi, A., Depeursinge, A., & Rubin, D. L. (2016). Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles. *Medical image analysis*, *30*, 60–71. doi:10.1016/j.media.2015.12.002

Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., Shih, N., et al. (2014). Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In *Medical imaging 2014: Digital pathology* (Vol. 9041, p. 904103). International Society for Optics; Photonics. doi:10.1117/12.2043872

Hou, L., Samaras, D., Kurc, T. M., Gao, Y., Davis, J. E., & Saltz, J. H. (2016). Patch-based convolutional neural network for whole slide tissue image classification. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2424–2433). doi:10.1109/CVPR.2016.266