Effects of Scanner Variability on Deep Learning based Lymph Node Segmentation

AUTHORS:

Institution/Hospital, Department, City, Country.

- Amjad Khan, Institute of Pathology, University of Bern, Switzerland

- Dr. med. Annika Blank, Institute of Pathology, University of Bern, Switzerland

- MMed Felix Müller, Institute of Pathology, University of Bern, Switzerland

- Dr. Huu Giao Nguyen, Institute of Pathology, University of Bern, Switzerland

- Prof. Dr. med. Alessandro Lugli, Institute of Pathology, University of Bern, Switzerland

- PD Dr. med. Heather Dawson, Institute of Pathology, University of Bern, Switzerland

- Prof. Dr. Jean-Philippe Thiran, Signal Processing Laboratory (LTS5), Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland

- Prof. Dr. phil. nat. Inti Zlobec, Institute of Pathology, University of Bern, Switzerland

BACKGROUND:

Stain color and contrast variations are common challenges faced by digital analysis of whole slide images (WSI) when glass slides are scanned with different scanners in an institute of pathology. These variations can further affect the outcome of computational pathology. We propose a novel study to assess such effects on lymph node segmentation.

METHODS:

Four different scanners were used to scan the same one hundred glass slides of lymph node tissues stained with hematoxylin and eosin. Two experienced pathologists annotated all four hundred WSI by manually drawing polygons around each lymph node. We applied a U-Net based lymph node segmentation pipeline with cross-validation. To avoid overfitting, training samples were augmented by flipping, rotating, shearing, zooming, and cropping. Finally, segmented regions were quantitatively evaluated with annotations by DSC (Dice Similarity Coefficient) on a scale between zero (no overlap) and one (100% overlap). Similarly, HD (Hausdorff Distance) measures the boundary loss on the lymph node capsule where zero HD means no loss.

RESULTS:

In a first analysis, the trained model on one scanner was used to predict lymph node regions on test sets of all scanners. Variability was evident from the outcome, where *Scanner1* (DSC: 0.786±0.266 and HD: 6.158±3.942), *Scanner2* (DSC: 0.745±0.269 and HD: 8.228±4.141) and *Scanner3* (DSC: 0.738±0.272 and HD: 5.307±2.490) showed poor performance. In a second analysis, pre-trained weights of one scanner were used to fine-tune other scanners. Fine-tuning minimized the variations of the results observed in the first analysis, where *Scanner1* (DSC: 0.847±0.215 and HD: 5.509±3.729), *Scanner2* (DSC: 0.847±0.210 and HD: 6.870±3.760) and *Scanner3* (DSC: 0.837±0.236 and HD: 4.704±2.563) showed improvement.

CONCLUSIONS:

Our analysis indicates that scanner variability can be optimized by fine-tuning. This technique will be valuable for institutes of pathology using different scanner types.