Automated removal of pen ink on whole slide images using weakly-supervised deep neural networks

SA Kohn, Ph.D.,[1] S Sankarapandian,[1] D Ayyagari,[1] RV Chamarthi,[1] W Shon, D.O.,[2] Z Laszik, M.D., Ph.D.,[3] S Bowman,[3] E Chan, M.D., Ph.D.,[3] MJ Bonham, M.D., Ph.D.,[1] RE Soans, Ph.D.,[1] JD Ianni, Ph.D.[1*]

[1]Proscia Inc., Philadelphia, Pennsylvania, USA
[2]Department of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, California, USA
[3]Department of Pathology, University of California San Francisco, San Francisco, California, USA
[*] julianna@proscia.com

Introduction

Artificial Intelligence has proven to be a useful tool in tackling several problems in computational histopathology [2,3]. In spite of its success, it has been recently identified that artifacts on Whole Slide Images (WSIs) adversely affect machine learning models [2,8]. Several deep learning based solutions have been proposed in the literature to tackle this problem [1,4,6,7] which either require hand crafted features or finer labels than slide-level labels.

Pen ink is often used by pathologists to indicate a region of interest; often of malignancy. If not removed, weakly supervised models can erroneously learn this signal as evidence of malignancy, and falsely predict a slide as malignant if pen ink is present, regardless of the underlying pathology. We trained an attention-based neural network under a multiple-instance learning (MIL) framework[4] to detect whether or not ink was present on a slide. Our MIL ink detector treats a slide as a bag and tiles from the slide as its instances. Attention-based MIL models learn weights for individual instances in a bag that determine the importance of an instance in arriving at a prediction. We could then use those weights (which we refer to as *attention values*) to determine the essential instances for a prediction of "pen ink is present". Removing these instances is equivalent to removing the pen ink from the slide, which can then be used for training downstream weakly supervised models.

Materials & Methods

If we used both benign and malignant tissue types, a weakly supervised model used to detect ink might instead have learned to identify patterns of malignancy. To avoid this, we gathered WSIs with and without pen ink from a dataset of skin biopsies specifically melanocytic tissues (240 WSIs; 236 melanomas [in situ: 118, invasive: 118], 3 dysplastic, 1 Spitz) scanned on Ventana DP-200 and from a dataset of prostate biopsies (465 WSIs; 182 Gleason Grade 6, 201 Grade 7, 40 Grade 8, 42 Grade 9) scanned on Epredia (3D Histech). WSIs were drawn from both source datasets such that 50% of WSIs had ink present and 50% did not.

Each WSI was first passed through tissue segmentation stage and the tissue regions are divided into a bag of 128x128 pixel tiles to train the MIL model. The MIL model consisted of five convolutional layers, two fully connected layers, a single attention head and a single sigmoid-activated output head. Our MIL ink detector was trained only on WSI-level labels without requiring pixel level annotations. If the output was greater than 0.5, it was interpreted as a positive prediction.

Critical Region of Interest Detection

If ink was detected (i.e. the output value was greater than 0.5), we were able to use the attention values for each tile to steadily remove highly-attended tiles by iteratively performing inference on subsets of tiles, until the decision of the model changed to "no ink" (i.e. output dropped beneath 0.5). This search process can be performed efficiently by sorting the tiles by attention values and performing a binary search over the attention values. Figure 1 illustrates this process. After the tiles that contributed to the decision of ink being present were identified, they could be removed from the bag, and the cleaned WSI could be used for downstream training of weakly supervised models.

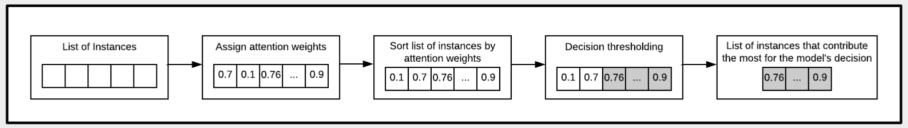

Figure 1: Process of selecting the tiles that contribute the most for the model's decision. 1) Obtain list of tiles in a WSI 2) Get attention values for each tile after passing through our MIL ink detector 3) Sort tiles in ascending order of their attention values 4) Iteratively perform inference on subset of tiles while steadily removing highly attended tiles 5) Obtain the list of tiles that contributed to the decision of ink being present

Qualitative results

We show an example ink detection in Figure 2. On the left panel is a region of a WSI -- a dermatopathology slide containing residual melanoma in situ -- with blue pen ink present, likely indicating the depth of the tumor. On the right, we show attention values from the ink detection model for each tile overlaid on the original WSI. Most are transparent grey: that is, they have low attention values, whereas pink tiles indicate high attention. A green outline shows the region identified by the critical region detection method described above. The inked region is completely isolated. Removing those tiles from the WSI allows it to be used for downstream weakly supervised models without risk of ink producing biased, false signals of malignancy.

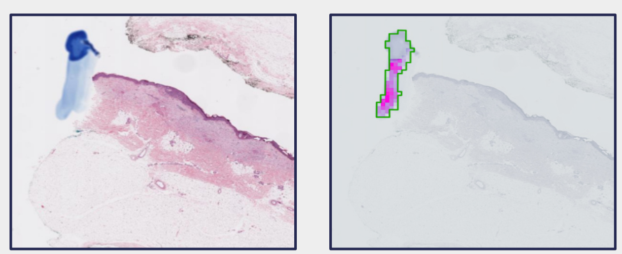

Figure 2: *(Left)* Example of a region of ink on a WSI. *(Right)* Attention overlay on the same region, pink regions indicate high attention, transparent gray regions indicate low attention. Green outlined region indicates critical region of interest

Quantitative results

The MIL model used for pen ink detection obtained 98% balanced accuracy (F_1 Score = 0.98) on 106 withheld test WSIs, where roughly two thirds were prostate biopsies and one third were skin biopsies and excisions.

To demonstrate the efficacy of removing inked tiles in downstream weakly supervised models, we trained a binary MIL model to determine whether a given prostate biopsy was graded with a Gleason score ≥ 6, with and without removing inked tiles via our critical region detection method.

The model trained *without removing* pen ink erroneously focussed on ink tiles to achieve 92% balanced accuracy. The model trained *after removing* pen ink yielded better performance - 95% balanced accuracy - focusing on regions of malignancy.

Discussion and Conclusions

Our use of weakly-supervised, multiple-instance learning coupled with deep features allows us to remove pen ink automatically without annotations on both prostate and skin. It is not color-dependent, requires no annotations to train, and doesn't need handcrafted or heuristic features to select inked regions. In this work, we have demonstrated the importance of removing seemingly innocuous artifacts from machine learning datasets. We showed evidence that downstream deep learning models could see an improvement in performance when pen ink regions are removed from WSIs. However, pen ink is one of many commonly occurring quality issues. More broadly, our work emphasizes the significance of detection and removal of any such artifacts when using weakly supervised learning which could adversely bias models if ignored.

References:

[1] Ali, S. & May, C. V. Ink removal from histopathology whole slide images by combining classification, detection and image generation models. ISBI. 928–932 (2019)
[2] Campanella, Gabriele, et al. "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images." Nature medicine 25.8 (2019)
[3] Ianni JD, Soans RE, Sankarapandian S, Chamarthi RV, Ayyagari D, Olsen TG, et al. Tailored for real-world: a whole slide image classification system validated on uncurated multi-site data emulating the prospective pathology workload. Sci Rep. 2020
[4] Janowczyk, Andrew, et al. "HistoQC: an open-source quality control tool for digital pathology slides." JCO clinical cancer informatics 3 (2019)
[5] Ilse, M., Tomczak, J.M., Welling, M.: Attention-based deep multiple instance learning, ICML 2018
[6] Maleki, Danial, et al. "Ink Marker Segmentation in Histopathology Images Using Deep Learning." International Symposium on Visual Computing. Springer, Cham, 2020.
[7] Smith, Byron, et al. "Developing image analysis pipelines of whole-slide images: Pre-and post-processing." Journal of Clinical and Translational Science 5.1 (2021).
[8] Wright, Alexander Ian, et al. "The effect of quality control on accuracy of digital pathology image analysis." IEEE Journal of Biomedical and Health Informatics (2020)

PROSCIA