Physical Color Calibration of Digital Pathology Scanners for Deep Learning Based Diagnosis of Prostate Cancer

Xiaoyi Ji[1] XIAOYI.JI@KI.SE
[1] *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.*
Richard Salmon[2] RICHARD.SALMON@FFEI.CO.UK
[2] *Department of Life Sciences, FFEI Ltd, Hemel Hempstead, UK.*
Nita Mulliqi[1] NITA.MULLIQI@KI.SE
Henrik Olsson[1] HENRIK.OLSSON@KI.SE

Lars Egevad[3] LARS.EGEVAD@KI.SE
[3] *Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden.*
Pekka Ruusuvuori[4,5] PEKKA.RUUSUVUORI@UTU.FI
[4] *Institute of Biomedicine, Cancer Research Unit and FICAN West Cancer Centre, University of Turku and Turku University Hospital, Turku, Finland.*
[5] *Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.*
Martin Eklund[1] MARTIN.EKLUND@KI.SE
Kimmo Kartasalo[1] KIMMO.KARTASALO@KI.SE

Editors: Under Review for MIDL 2022

Abstract

Variation in whole slide image (WSI) across different scanners poses a problem for deep learning algorithms. We apply a color calibration slide to standardize WSIs from different sites and evaluate the effect of calibration on a deep learning model for prostate cancer diagnosis. We show that calibration can significantly improve the accuracy of the model.
Keywords: Color Calibration, Cancer Detection, Artificial Intelligence

1. Introduction

Artificial intelligence (AI) algorithms applied to whole slide images (WSI) in digital pathology have demonstrated high performance in cancer detection (Campanella et al., 2019; Ström et al., 2020). However, decreased performance caused by technical variability across WSIs acquired using different scanners remains a key problem, as observable mismatches in brightness, contrast and sharpness often affect the generalization performance of AI systems.

The proposed solutions, including computational stain normalization and data augmentation, unfortunately either have produced mixed results or are hard to design to cover the real world variance. It has been shown that utilizing a physical color calibration slide can effectively normalize the color variations in WSIs caused within scanning processes (Bautista et al., 2014). With the help of this color calibration slide, one can estimate the International Color Consortium (ICC) profile of scanners, with which the color of each WSI could be calibrated to a chosen ICC profile. In this study, we aim to apply a commercial color calibration slide for standardizing WSIs of prostate biopsies, and to assess the effect of the calibration on the diagnostic performance of an AI system for detecting prostate cancer.

2. Materials and Methods

The AI model is based on a previously developed one, where each WSI is processed in patch-wise manner (Ström et al., 2020). Ensembles of Inception V3 deep neural networks (DNNs) implemented in TensorFlow are used for patch-level training and prediction, and gradient boosted trees implemented in Xgboost for WSI-level prediction of: 1) probability of cancer presence, 2) cancer length in millimeters. In comparison to the original model, we add color calibration for each patch as a pre-processing step for DNN training and prediction. OpenSlide was used for accessing the WSIs in native *.svs* and *.ndpi* formats.

For the training data set, we scanned 3,651 prostate needle biopsies from 957 patients on an *Aperio ScanScope AT2* scanner at SciLifeLab, Uppsala, Sweden. The tuning (100 WSIs) and testing (230 WSIs) sets were scanned with a *Hamamatsu NanoZoomer S360 C13220-01* at Karolinska University Hospital, Stockholm, Sweden. The ICC profiles of the scanners were obtained by scanning the Sierra color calibration slide (FFEI Ltd., Hemel Hempstead, UK). We chose sRGB as the calibration standard, whose ICC profile can be found on the website of International Color Consortium. The approximate annotations of malignant regions of the WSIs were performed by a single pathologist (L.E.). Approximately 1000 image patches per WSI were extracted from benign and malignant regions.

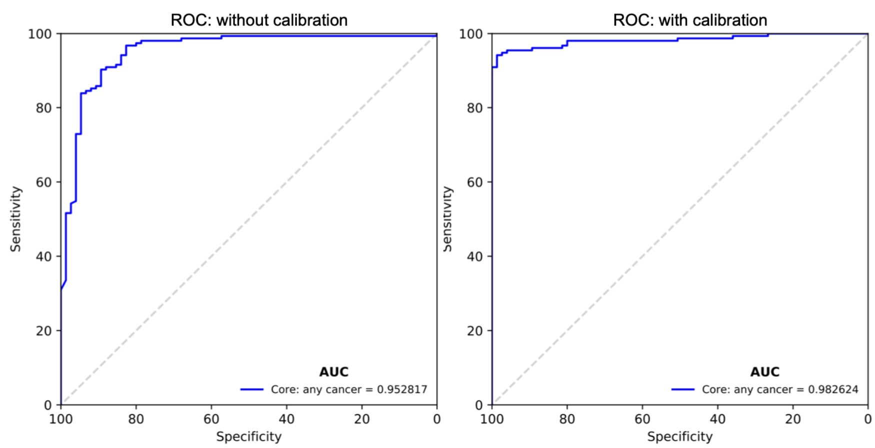

Figure 1: ROC curves indicating slide-level cancer detection performance with (right) and without (left) color calibration.

3. Results

We successfully integrated color calibration into the AI system, and qualitatively observed standardized colors between images from different scanners. For a quantitative assessment, we trained and evaluated the AI model with and without applying color-calibration on the input WSIs. In the independent test set, Receiver Operating Characteteistics (ROC) analysis showed Area Under the Curve (AUC) values for cancer detection on slide-level of 0.953 and 0.983 without and with calibration, respectively (Figure 1). The linear correlation coefficients between cancer lengths estimated by the AI system and the pathologist were 0.78 and 0.84 without and with calibration, respectively (Figure 2). The AI model with calibration shows improved performance particularly for the small cancer lengths.

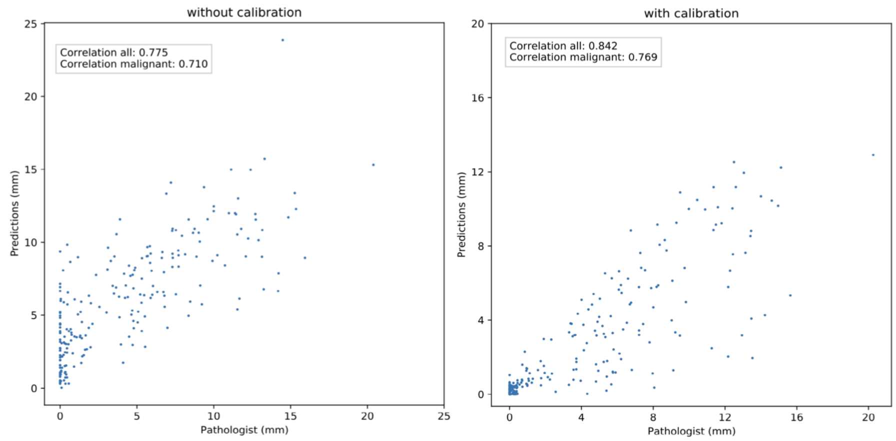

Figure 2: Linear correlation between cancer lengths estimated by the AI system (y-axis) and the pathologist (x-axis), with (right) and without (left) color calibration. Each data point represents a single slide in an independent test set.

4. Conclusion

Our results show that spectrophotometric color calibration using a physical slide can improve the robustness of AI models for prostate cancer detection across different scanner systems. Next, we expect to refine our result by more validations to further confirm the feasibility of this approach in practical applications.

References

Pinky A Bautista, Noriaki Hashimoto, and Yukako Yagi. Color standardization in whole slide imaging using a color calibration slide. *Journal of pathology informatics*, 5, 2014.

Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019.

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, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *The Lancet Oncology*, 21(2):222–232, 2020.