Stress Testing Vision Transformers Using Common Histopathological Artifacts

Geetank Raipuria GEETANK.RAIPURIA@AIRAMATRIX.COM
Nitin Singhal NITIN.SINGHAL@AIRAMATRIX.COM
Advanced Technology Group, AIRAMATRIX PVT. LTD., Mumbai, India

Editors: Under Review for MIDL 2022

Abstract

Artifacts on digitized Whole Slide Images like blur, tissue fold, and foreign particles have been demonstrated to degrade the performance of deep convolutional neural networks (CNNs). For prospective deployment of deep learning models in computational histopathology, it is essential that the models are robust to common artifacts. In this work, we stress test multi-head self-attention based Vision Transformer models using 10 common artifacts and compare the performance to CNNs. We discovered that Transformers are substantially more robust to artifacts in histopathological images.

Keywords: Artifacts, Histopathology, Transformers, Robustness

1. Introduction

Automated analysis of high resolution histopathology Whole Slide Images (WSI) using deep learning models has acquired a lot of traction as a way to aid pathologists. However, WSI preparation introduces artifacts such as blur, fold, air bubbles, dirt, which may lead to substantial loss in performance of convolutional deep learning models (Schömig-Markiefka et al., 2021; Wang et al., 2021). For deployment in real-world settings, it is critical to develop models that are robust to common corruptions and provide high accuracy irrespective of the presence of artifacts during testing.

Transformers have demonstrated strong robustness against common image corruptions, occlusion as well as natural adversarial examples (Naseer et al., 2021; Bai et al., 2021), which makes them a suitable candidate for stress testing using histopthological artifacts. Transformers have less texture bias as compared to CNNs as well as benefit from content based context modelling using self-attention modules built into their design. We compare the performance of Transformers and CNNs under the influence of 10 common artifacts on a multi-class tissue type classification dataset. Based on experiments, we show that the performance of Transformers degrades substantially less than that of CNNs on introduction of artifacts on validation images. To best of our knowledge, this is the first work towards identifying and developing deep learning models that are robust to an array of common histopathology artifacts.

2. Training Setup

Dataset: We collect a multi-class tissue classification dataset obtained from Wistar Rat organs (512x512 dimension tiles) in preclinical toxicology studies. The dataset contains 77k training patches from 20 classes and 10k validation patches (500 per class), at 10x magnification. We assess the models' robustness by artificially adding artifacts to the validation

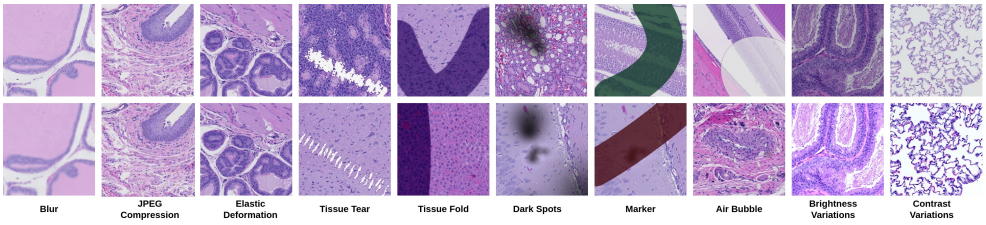

Figure 1: Sample images of artifacts generated to evaluate Transformer & CNN models

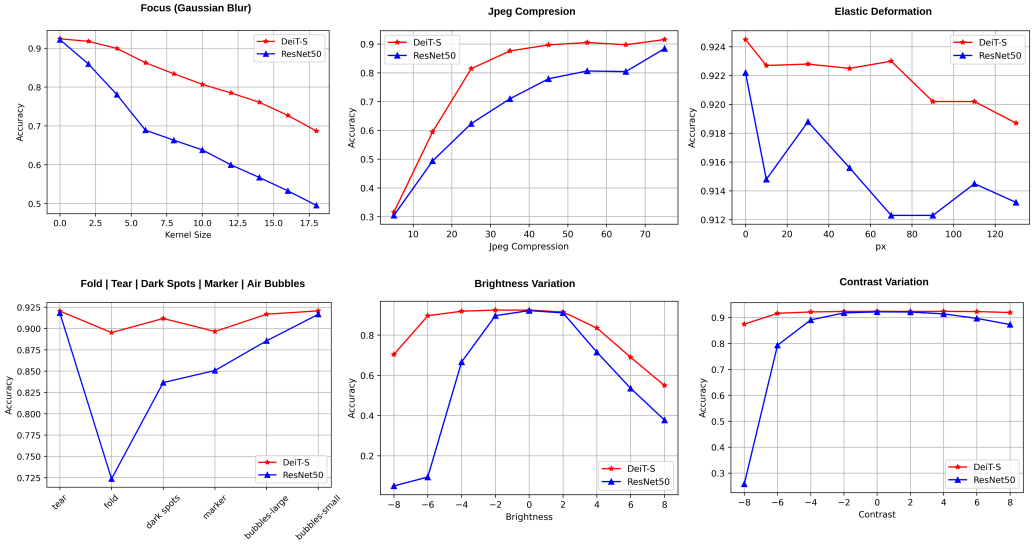

Figure 2: Stress testing Transformer and CNN models using histopathological artifacts.

data. Data with artifacts for 10 corruptions were created using code provided by (Wang et al., 2021), as seen in figure 1. We fix seed values to ensure consistent artifact generation for evaluating models.

Training Details: ResNet-50 and DeiT-S (Touvron et al., 2021) are trained as CNN and Transformer models, initialized with ImageNet pre-trained weights, which have a similar number of trainable parameters (22M vs 23M respectively). We performed a grid search for number of epochs {10, 15, 20} and the learning rate {1e-3, 5e-4, 1e-4}. Both models were trained using cosine learning rate decay with warm-up, for a batch size of 128. ResNet-50 was trained using Adam optimizer with a maximum learning rate of 5e-4, whereas DeiT-S was trained with a maximum learning rate of 1e-4 using AdamW optimizer. Weight decay of 5e-4 and 5e-2 was used for ResNet-50 and DeiT-S models, respectively. Only rotation and flipping were employed as augmentations for all models. The model with the lowest validation loss from the training run is chosen for evaluation. NVIDIA RTX A6000 GPU was used for training all models, and the code was written in PyTorch.

3. Results and Discussion

Both DeiT-S and ResNet-50 models achieve similar clean image (without artifacts) accuracy of 92.45% & 92.22%, respectively. Figure 2 plots the performance on the introduction of 10 types of artifacts, for both the deep learning models. The accuracy of ResNet-50

considerably deteriorates due to the presence of artifacts, based on the severity of the corruption, as also observed by (Wang et al., 2021; Schömig-Markiefka et al., 2021). DeiT-S, a Transformer model, deteriorates to a far lesser extent, especially in case of high severity.

We attribute Transformer's increased robustness to common artifacts to two key aspects a) Transformers have been demonstrated to be less sensitive to local texture information than CNNs, exhibiting stronger shape bias instead (Naseer et al., 2021). This makes Transformers more robust to image texture distortions like blur, JPEG compression, foreign particles (dark spots), and air bubble; b) Transformers are capable of modeling global interactions among image regions. Multi-head self-attention, in particular, finds content-dependent long-range interactions between embeddings, incorporating global information from distant image regions. This provides transformers the ability to adjust their attention to cope with corruptions in input data (Naseer et al., 2021).

4. Conclusion and Future work

In this study, we compared the robustness of the Transformers model to Convolutional Neural Networks under the impact of histopathological artifacts. We show that Transformers are less sensitive to common artifacts like blur, fold, dark spots and retain performance much better as compared to CNNs, based on an evaluation of a multi-class tissue classification dataset. We intend to add more types of artifacts in future work, as well as evaluate the models using classification and segmentation tasks on multiple datasets. We hope our work encourages further research to develop models that are robust to common histopathological artifacts.

References

Yutong Bai, Jieru Mei, Alan L Yuille, and Cihang Xie. Are transformers more robust than cnns? *Advances in Neural Information Processing Systems*, 34, 2021.

Muhammad Muzammal Naseer, Kanchana Ranasinghe, Salman H Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Intriguing properties of vision transformers. *Advances in Neural Information Processing Systems*, 34, 2021.

Birgid Schömig-Markiefka, Alexey Pryalukhin, Wolfgang Hulla, Andrey Bychkov, Junya Fukuoka, Anant Madabhushi, Viktor Achter, Lech Nieroda, Reinhard Büttner, Alexander Quaas, et al. Quality control stress test for deep learning-based diagnostic model in digital pathology. *Modern Pathology*, 34(12):2098–2108, 2021.

Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International Conference on Machine Learning*, pages 10347–10357. PMLR, 2021.

Nicholas Chandler Wang, Jeremy Kaplan, Joonsang Lee, Jeffrey Hodgin, Aaron Udager, and Arvind Rao. Stress testing pathology models with generated artifacts. *Journal of Pathology Informatics*, 12, 2021.