Deep learning based quality control of histopathology whole slide images (OR Quality analysis of publicly available whole slide image datasets)

Abhijeet Patil, Harsh Diwakar, Garima Jain, Jay Sawant, Swapnil Rane, Subhash Yadav, Tripti Bameta, Amit Sethi

ABSTRACT:

This research presents a deep learning-based pipeline for quality control of histopathology whole slide images (WSIs) using four different segmentation models. The segmentation models were developed for blur level segmentation, tissue segmentation, tissue fold segmentation, and pen marker segmentation. To minimize annotation efforts for tissue and blur segmentation, the cut-paste method was employed along with the HistoROI predictions to prepare balanced data batches for training the deep neural networks. The pipeline was evaluated on the largest publicly available dataset, containing over 11,000 histopathology images from the TCGA data portal. The segmentation masks generated by the models were evaluated to assess their quality, and the performance of the models was compared to that of the HistoQC tool. The pipeline showed superior performance compared to the HistoQC tool in segmenting different artifacts present in the images. The trained models, training scripts, training data, and inference results are made publicly available, enabling the research community to build more efficient and robust quality control models in the future. The potential impact of this work is vast, as it can enhance the accuracy and reliability of histopathology image analysis and contribute to the development of automated tools that can analyze large-scale histopathology datasets efficiently and accurately. The availability of the trained models and data will help the research community in building better and more efficient quality control models for histopathology WSIs, which will improve the overall quality of cancer diagnosis and treatment.

INTRODUCTION:

In recent years, pathology diagnosis has been undergoing a significant transformation with the integration of cutting-edge image scanning technologies and the availability of advanced IT infrastructure. This digitalization of pathology diagnosis offers a plethora of advantages over the traditional method of carrying glass slides from one place to another. The risk of damage to glass slides during transportation is eliminated, and sharing of images with multiple pathologists in different locations is made possible, leading to enhanced diagnostic accuracy. Additionally, the digitalization of pathology diagnosis offers improved efficiency, faster turnaround time, and increased accessibility to patient data, which ultimately benefits both patients and clinicians. However, this shift towards digital pathology also brings forth a new set of challenges, such as the need for standardized protocols for image acquisition and interpretation, data storage and retrieval, and quality control. Hence, it is essential to explore and evaluate the impact of digital pathology on the field of pathology diagnosis, its advantages, limitations, and future prospects.

The digitization of histopathology images has not only improved the efficiency of pathology diagnosis but also opened new avenues for research and development of computer-aided diagnostic systems. With the help of advanced deep learning algorithms, researchers have been able to utilize these digital images for cancer diagnosis-related tasks. The availability of

large datasets of digitized histopathology images has facilitated the development of automated systems that can accurately detect and classify tumor regions in gigapixel whole slide images, grade the tumor region according to standard protocols, identify subtypes of tumor, predict survival of a patient based on image data, etc. These computer-aided diagnostic systems have the potential to significantly improve the accuracy and speed of cancer diagnosis, especially in cases where manual interpretation of images may be challenging due to the complexity of the tissue structure or the small size of the abnormal cells.

The application of deep learning algorithms in developing computer-aided diagnostic systems has shown promising results in various studies. These algorithms can identify complex patterns and features within histopathology images that are not easily visible to the naked eye. By training these algorithms on large datasets of annotated images, researchers can create highly accurate diagnostic models that can be used to assist pathologists in making more precise diagnoses. The development of such systems has the potential to revolutionize the field of pathology diagnosis, making it more efficient and accessible to a wider population.

Although deep learning algorithms have shown significant potential in assisting with complex diagnosis tasks, most studies to date have utilized clean and curated datasets. In real-world scenarios, histopathology images are often affected by various artifacts that can impact their quality and diagnostic accuracy. Recent research has demonstrated the negative effects of using images with poor quality and artifacts on diagnostic tasks, indicating the need for reliable quality control tools specifically designed for histopathology images.

In this study, we introduce hand-annotated datasets and a novel technique for training segmentation models for histopathology. The datasets include annotations for pen marker segmentation and tissue fold segmentation, and we have developed four segmentation models for blur level segmentation, tissue segmentation, tissue fold segmentation, and pen marker segmentation. Our study aims to address the challenges associated with histopathology image analysis, particularly in identifying different tissue components and artifacts. The hand-annotated datasets and segmentation models we introduce can assist in improving the accuracy and efficiency of histopathology image analysis. To develop our models, we utilized a novel technique for training segmentation models, which takes into account the unique characteristics of histopathology images. The four segmentation models we developed were designed to address different aspects of histopathology image analysis, such as identifying tissue components, tissue folds, and pen markers. We conducted a thorough analysis of the largest publicly available dataset to validate the accuracy and effectiveness of our models. Main contributions of our work are

- We present a novel method to train a segmentation model for histopathology images which utilizes domain knowledge generated from histoROI for better sampling of data.
- We make hand-annotated datasets for pen marker segmentation and tissue fold segmentation publicly available.
- We release our training and inference codes along with weights of trained models publicly available. These datasets and models can help the research community to develop better and efficient quality control tools.

- We also release the predictions of our models on more than 11,000 WSIs. These predictions can be used to develop weakly supervised learning algorithms for quality control of histopathology images. Also, researchers can utilize these predictions to prepare and choose their datasets keeping quality of images in check.

RELATED WORK:

This section examines studies related to quality control in histopathology images. These studies can be categorized into two broad categories: traditional image processing-based methods and deep learning-based methods. Most of the studies focus on addressing specific types of tissue defects, while some attempt to solve multiple defects with a single model or separate models.

HistoQC is by far the most widely used histopathology QC tool. It performs multiple tasks based on conventional image processing, machine learning techniques and deep learning based models to identify regions such as useful tissue region, adipose-like region, background, tissue folds, out-of-focus region, etc. It also provides an easy to use HTML based UI to adjust parameters depending on staining protocols used for slide preparation. They have shown HistoQC can identify foreground regions correctly for more than 95% WSIs.

Wide variety of work, inspired from natural image settings has been applied to pathology images to identify out-of-focus regions. This problem has seen lots of traction in the research community as well as in industrial settings. Identification of lens parameters is of paramount importance while capturing images through digital scanners, which can be guided by automatic focus assessment algorithms. The quality of histopathology patches at 20X is accesses by comparing metrics like SSIM and IL-NIQE to show pristine histopathology image patches can be distinguished from patches affected with artifacts such as out-of-focus regions applied image quality metrics used in natural imaging settings to histopathology images. They have calculated several full reference and no-reference image quality metrics and validated these metrics with pathologists.

Image gradient based features have also been used for identification of out-of-focus regions in WSIs using linear regression on image features to quantify focus quality of WSIs. takes advantage of image features generated from synthetically blurred images to build histopathology patch classifiers. builds random forest based classifiers on image features to identify out-of-focus patches in WSI to reduce the error from 17% to 4% for WSI quality assessment task. Authors explored multiple features for determining focus at the time of image acquisition such as color, brightness, contrast, focus sharpness, etc. to determine necessity of re-capture of image. have released a python and java based programming library to assess focus quality of WSIs with the help of various parameters. uses physics based domain knowledge to develop novel deep cascade networks to improve an autofocusing quality. uses logistic regression based classifiers on texture features of a patch to identify patches affected by artifacts.

Human visual system based kernels are also utilized for out-of-focus region detection. proposed computation efficient, no reference image sharpness assessment metric called

HVS-MaxPol. They also released a FocusPath dataset, which contains around 8000 histopathology patches captured at various focus settings (z-planes). Further, they proposed no-reference focus quality assessment metric specifically for digital pathology images that operate by using a sum of even-derivative filter bases to synthesize a human visual system-like kernel. Similarly, FocusLiteNN proposes a very efficient image kernel learnt using CNNs to identify patches with out-of-focus regions in histopathology image patch. They also compiled a dataset of around 14000 patches from TCGA data portal out of which around 3000 patches are labeled as out-of-focus patch.

Deep learning based approaches have also been tried to distinguish out-of-focus patches. CNN based classifiers have been trained on synthetic as well as systematically captured dataset varying the focus settings on scanners. Senaras presents a patch based CNN classifier trained with systematically captured data from various focal planes. reported 93% accuracy on binary classification tasks. Albuquerque utilizes synthetic images to build a multi-class classifier trained with novel ordinal loss functions. Kohlberger proposes ConvFocus, trained on semi-synthetic data and shown Spearman rank coefficients of 0.81 and 0.94 on two different datasets when compared with assessment by pathologists. They have also shown a decrease in performance on CAMELYON when blurred images are used for testing.

There are relatively less studies concentrating on other types of defects than out-of-focus artifacts. Major challenge in addressing these defects is unavailability of annotated datasets and difficulty to create these datasets. These defects are also hard to generate synthetically. Few studies dealing with tissue fold segmentation utilize features such as color and connectivity properties of tissue structures color saturation and luminance to detect tissue folds. Reducing tissue fold artifacts in analysis has also been shown to improve the performance of cancer-grade prediction models and trains a support vector machine based classifier to identify patches affected by artifacts such as tissue folds and air bubbles which can not be identified by similarity metrics like SSIM or laplacian based methods. Deep learning based classifiers have also been used to classify patches with tissue fold.

(Bándi et al., 2019) trains a tissue segmentation algorithm using fully convolutional neural networks on a dataset containing multiple stains and showed Dice score of 0.98 on test dataset. They have compared their algorithm against HistoQC and demonstrated improved performance. Subsequently, they have annotated artifacts in these WSIs to train a DeepLabV3 based segmentation network to train a semantic segmentation network to identity region with tissue folds, out-of-focus region, dust on slide, pen marker, etc. and showed more than 90% pixel level accuracy on most of the classes. To mitigate unavailability of annotated dataset, methods like weakly supervised learning, few-shot learning and training with noisy annotations, etc. also have been explored.

Ali identifies patches with pen markers using convolutional neural network based classifiers and tries to recover information using cycle-GAN uses deep neural networks to access tissue stain quality and also identifies artifacts like dirt, damaged or crystallized tissue slides, etc. trains a patch based CNN classifier to identify defects like tissue folds, staining related artifacts, out-of-focus regions and assigns quality score to WSI based on a regression model trained on top of multi-class classification model. Alongwith identification of staining related artifacts, there have been multiple studies focusing on stain correction of H&E stained

pathology images based on generative models, self-supervised learning and domain knowledge driven optimisation methods.

There are few studies concentrating on the effect of quality control on downstream diagnosis tasks. Introduction of synthetically generated defects at test time shown to deteriorate the performance of prostate cancer detection and kidney tissue segmentation. Filtering out the region corresponding to artifacts and adipose-like regions is also shown to increase performance on CAMELYON dataset and TCGA-Lung cancer dataset for the task of subtype classification. Similarly, the importance of identifying artifact affected regions is highlighted in few studies to improve automatic diagnostics algorithms. Computer aided quality control tools have shown to have capabilities to improve WSI curation process (Chen et al., 2021). Also, there are multiple studies relating image compression and hence image quality to diagnosis.

METHODS:

The accurate identification and classification of artifacts is essential for effective analysis of medical imaging and scientific data. In this study, we developed four distinct artifact segmentation models, each tailored to a different type of artifact: blur level, tissue fold, pen marker, and tissue segmentation. We utilized annotated datasets for training the pen marker and tissue fold segmentation models, and incorporated a semi-supervised approach for the blur level segmentation and tissue segmentation models, using the predictions generated by the recently published HistoROI. In this section, we provide a detailed account of our methodology, including the architecture of our models, the datasets used for training and testing. Additionally, we describe the techniques we employed to optimize our models, including data augmentation and fine-tuning.

Cut-paste method: Training segmentation models for WSI level tasks can be a challenging task, as it requires a large amount of annotated data. However, in most cases, there is a lack of annotated data, which makes it difficult to train accurate segmentation models. Even if annotated data is available, the data distribution is often skewed, which can lead to a biased model. Training deep neural networks requires each batch of data to contain representative data, but this is not always possible with WSIs since ROIs generally span thousands of pixels.

To address this challenge, we have used a novel cut-paste method for training segmentation networks. This method involves cutting patches from WSIs and pasting them into smaller patches of a fixed size, which are then used for training. This ensures that each batch of data contains representative data and helps to overcome the issue of skewed data distribution.

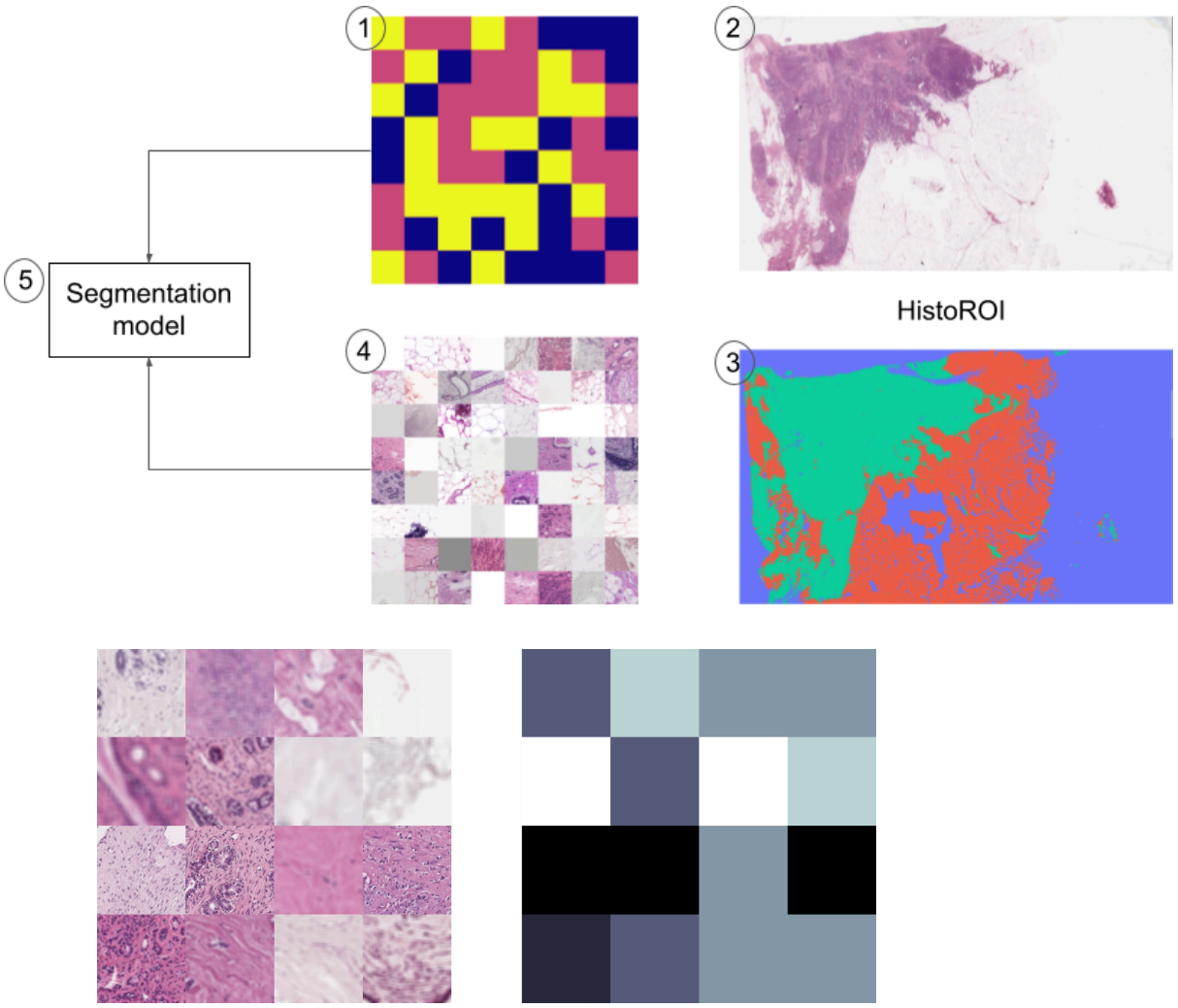

Figure: Cut-paste method to generate uniform data distribution across all classes. (1) Random grid of patches of size 64x64 generated on the empty array of size 512x512. (2),(3) A WSI alongwith its HistoROI predictions is sampled from a pool of WSIs selected from training dataset. (4) A patch of required class is randomly selected from WSI. Index of patch is obtained by HistoROI predictions. Patches from 2.5X magnification are used to train this model. (5) A CNN segmentation model is trained using initially generated random masks and images created by pasting patches from WSI. Generated images can have patches from multiple WSIs, making the segmentation model more robust. (6) Sample input label pair for training blur level segmentation model. We generate patches of size 128x128 at 5X magnification level for this task.

To further enhance the accuracy of the segmentation model, we have also used the HistoROI patch classification model. This model predicts the class of a patch of a WSI, which can be epithelial, stroma, lymphocytes, adipose, miscellaneous, or artifact. By using this model, we can select patches or regions from specific classes, which ensures that the balance between different types of tissue regions is maintained during the training of the segmentation model.

The HistoROI model uses a deep neural network trained on patch-level data, which makes it highly accurate in predicting the class of a patch. By combining the cut-paste method and HistoROI model, we were able to overcome the challenge of lack of annotated data and

skewed data distribution. This allowed us to train highly accurate segmentation models for WSI level tasks.

Tissue segmentation model: We are using predictions of HistoROI to generate training data for a segmentation model. A 2D array of 512x512 is created, and a grid of size 8x8 is overlaid on the array. Each grid cell is 64x64 in size. A label is randomly assigned to each grid cell, corresponding to either foreground tissue, adipose, or background. This process generates a segmentation mask, which serves as the ground truth for training the tissue detection model. To create a corresponding image for each segmentation mask, patches are strategically mined from a pre-selected pool of WSIs. We have created a pool of twenty WSIs for each class. These WSIs were carefully selected so that enough variation for each class is captured by a segmentation model. For each grid cell, we selected a random WSI from a pool of WSIs and mined a patch randomly corresponding to that class. We are using patches at 2.5X magnification to perform this task. A sample input image and segmentation mask as shown in Figure.

This approach leverages the predictions of HistoROI, allowing us to take advantage of its accuracy while also creating a diverse and representative training set. The use of patches from WSIs at 2.5X magnification helps to balance the computational cost of tissue detection with the need for high-resolution Images. This magnification level provides sufficient detail for detecting tissue structures. Additionally, randomly mining patches from a pool of WSIs helps to introduce variability into the training set, making the model more robust and able to generalize to new data. This approach allows for the creation of a balanced and representative training set, as the labels are randomly assigned, ensuring that each class (foreground tissue, adipose, and background) is represented in the training data. In the process of tissue detection on whole slide images, the distinction between background and adipose tissue can present a significant challenge. Adipose tissue is commonly present in tissue samples and is not always important for diagnostic purposes, as it does not possess relevant features for many diagnosis tasks. However, it cannot be assigned to the background class, as it is still considered a part of the tissue region. To address this challenge, we have decided to keep adipose tissue as a separate class in the tissue detection task. This approach allows us to better differentiate between the background and adipose tissue, improving the accuracy of the tissue detection algorithm. Additionally, to diversify patches belonging to the background class and make the model more robust in identifying artifacts such as coverslip-related artifacts and pen markers outside tissue regions, heavy color jitter augmentation is applied to the background patches. This augmentation technique randomly applies changes to the color and intensity values of the pixels in the background patches, creating a diverse set of training data that helps the model to learn to identify and distinguish between tissue and non-tissue regions. A sample of model prediction is shown in Figure. In this manner, the use of HistoROI predictions and color jitter augmentation creates a balanced and diverse training dataset that can improve the accuracy and robustness of tissue detection algorithms on WSIs.

Blur level segmentation: Detecting blur levels in WSI can be challenging as they have a texture that varies according to the region of interest. Eg. A normal stroma has a smoother texture than normal cellular region. The laplacian-based methods often fail to provide consistent performance in detecting blur. However, by using patch mining with HistoROI, the model can learn the normal texture of a particular type of region in the WSI. This helps to

mitigate the problems that laplacian-based methods often face. The HistoROI model assists in learning the texture of specific regions, enabling the segmentation model to identify the degree of blur more accurately. This synthetic data approach, along with data mining strategy from HistoROI is useful in situations where obtaining real-world annotated data is difficult or impossible.

To develop a segmentation model for detecting blur levels, we introduce synthetic blur to patches identified as foreground by the HistoROI model. This creates a six-class segmentation problem using synthetic data, with Class 0 representing no blur. To introduce blur, we use the Boxblur method available in the PIL library, gradually increasing the blur level from level 1 to level 5 using the parameters of Boxblur. The model is trained using patches from a magnification level of 5X. To generate the input and ground truth pair, we started by creating a mask of size 512x512. This mask was then divided into a grid of 4x4 equal-sized cells. Each cell was randomly assigned a label between 0 and 5, indicating the level of blur. The corresponding input image was created using the HistoROI predictions and the Boxblur function of the PIL library. A sample input and segmentation mask is shown in Figure.

Pen marker segmentation: A dataset of thumbnails of 350 whole slide images from TCGA was created to train and validate the segmentation model to detect pen markers on whole slide images. Each image in a dataset was annotated at a magnification of 0.625X. The annotation process was a time-consuming task, with a single annotation taking approximately 5 minutes to complete. The dataset includes images annotated with pen markers of various colors, including black, red, blue, green, yellow, etc. This was done to ensure that the dataset represents a wide range of pen marker colors that are commonly used in pathology. A sample WSI thumbnail with annotation for pen marker is shown in Figure X. Our data sampling strategy for training a segmentation model involved selecting a mask randomly, then sampling a positive pixel (a pixel with a pen marker) randomly. A patch of size 512x512 was then taken around the randomly sampled pixel, and this patch was used as input to the model during training. The model was trained with a batch size of 8 and optimized using a combination of cross-entropy, focal and dice loss. This data sampling strategy was designed to ensure that the model was exposed to a diverse set of examples during training, thereby helping to prevent overfitting and improve its generalization performance.

Tissue folds segmentation: To train a segmentation model for tissue folds detection, the dataset was created by selecting 250 patches from the BRIGHT dataset. These patches were carefully selected to ensure that they contain enough variations in tissue structures and tissue folds to enable the model to generalize. The patches were annotated with binary masks indicating the presence of tissue folds. The annotation process was performed by experienced pathologists who manually traced the tissue folds in each patch. The annotations were performed at a 5x magnification level to ensure that the tissue folds were accurately captured. The annotation of a single patch took around 10 minutes, indicating that the annotation process is time-consuming and requires a high level of attention to detail. Each patch in the dataset is of a different size, with the average size of a patch being 2800 x 2800 pixels. A similar training strategy was used to train a segmentation model as used for pen marker detection. A sample annotated patch is shown in Figure.

Model architectures: The four segmentation models under consideration are all based on the UNet++ architecture. The pen marker segmentation model uses Resnet34 as its backbone, while the other three models use EfficientNet-b0. These backbones are pre-trained on the Imagenet dataset, which ensures that the models can extract useful features from the input images. The blur Segmentation, tissue fold segmentation, and tissue segmentation models are designed to run on higher magnifications, and thus they need to be lightweight to run efficiently. Therefore, we have used a lighter--ResNet18 backbone for these three models. To train these models, we have used PyTorch, which is a popular deep learning framework. We have also utilized the segmentation-models-pytorch library available in PIP for model architectures and loss functions. This library provides a range of pre-defined architectures and loss functions that are commonly used in image segmentation tasks, making it easy to implement these models without having to build them from scratch.

Task	Data	Labels	Magnification	Details
Tissue segmentation	60 WSIs from BRIGHT dataset	Generated using cut-paste method	2.5X	Three class segmentation model. 20 WSIs for each class.
Blur level segmentation	50 WSIs from BRIGHT dataset	Generated using cut-paste method	5X	Six class segmentation mode. 10 WSIs for each class of HistoROI.
Tissue fold segmentation	250 patches from BRIGHT dataset	Hand annotated	5X	Patches are of various sizes. Average size is 2800 x 2800 pixels.
Pen marker segmentation	300 WSI thumbnails from TCGA dataset	Hand annotated	0.625X	Selected WSIs contains pen markers of various colors.

RESULTS:

Results 1:

To evaluate the performance of our pipeline, we compared it with HistoQC, a widely used quality control tool. We ran all 11,000 WSIs available on the TCGA data portal through HistoQC's default parameters to generate a foreground tissue mask, and performed a similar procedure to predict foreground masks through our proposed pipeline. We then computed the Dice score between the masks to measure the agreement between the two predictions. A high Dice score indicates a high agreement between HistoQC and our pipeline, indicating that the masks generated by both strategies are either equally good or equally bad. However, in cases of lower agreement (low Dice score), there may be instances where

HistoQC performs better, our pipeline performs better, or it may not be possible to determine which strategy is better.

We sorted the Dice scores of the foreground tissue masks generated by HistoQC and our pipeline and divided them into buckets according to the dice scores. We sampled 20 WSIs from each bucket, resulting in a total of 100 WSIs, which were then manually analyzed for quality by pathologists. Proportion of WSIs in each bucket is observed to be 60%, 20%, 10%, 5% and 5% for the bins of dice 0.8 to 1, 0.6 to 0.8, 0.4 to 0.6, 0.2 to 0.4 and 0 to 0.2 respectively. The pathologists were presented with two anonymized masks in QuPath compatible annotation format, along with the WSI. For each pair of masks, the pathologists were asked to select the better one.

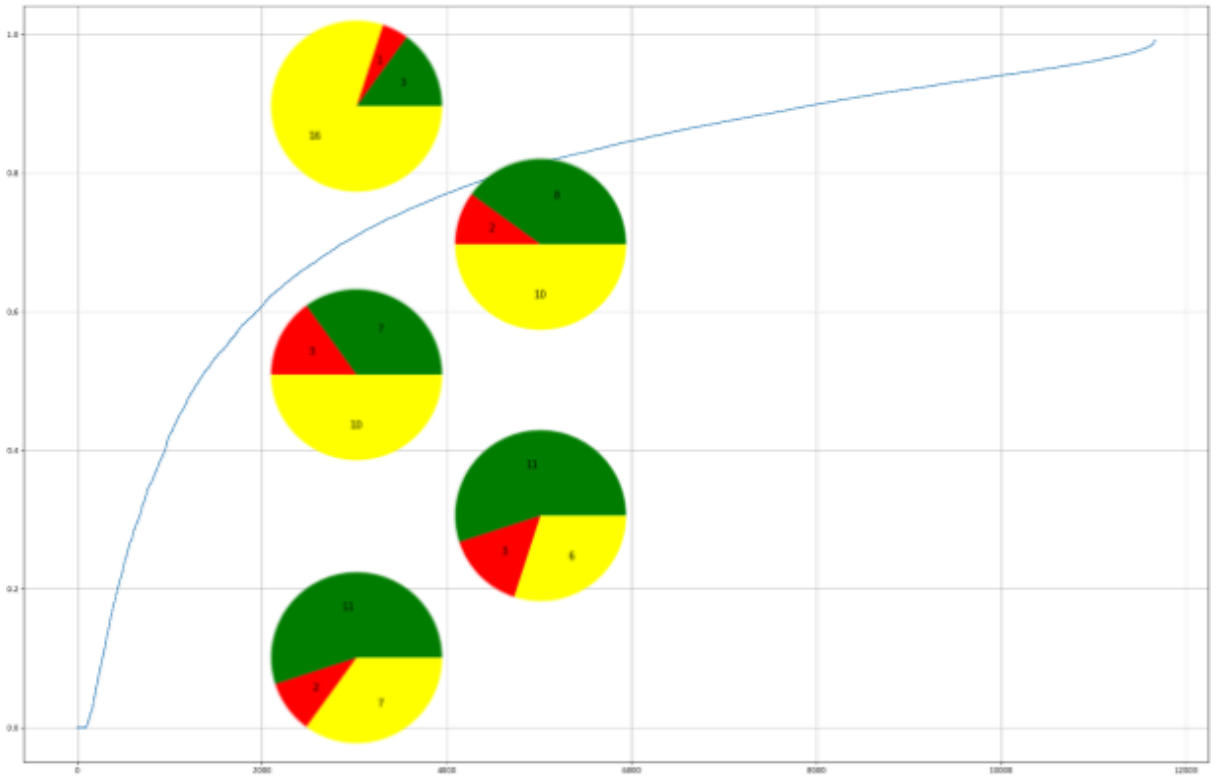

Figure: Agreement between HistoQC and our pipeline for 11,650 WSIs in TCGA dataset. Pie charts show subjective comparison between histoQC and our predictions. In case of disagreement, our predictions perform better than HistoQC.

From the bucket of the most agreement, the outcome of comparison to most of the WSIs was not conclusive, which was expected because of a higher Dice score. Out of 20 WSI mask pairs in this bucket, comparison for 16 was non conclusive, for 3 WSIs our pipeline performed better than HistoQC and for one WSI, HistoQC performed better. Results were more conclusive on the other end, where our pipeline performed better than HistoQC for 12 and 13 WSIs for the dice bucket of 0.2 to 0.4 and 0 to 0.2 respectively. Few samples were not conclusive in these buckets because both the models performed equally bad for few WSIs. Most of these bad performances were because of discrepancies in WSI metadata. Though our predictions are better than HistoQC, comparison based on WSIs predictions becomes a hectic task and can add too much subjectivity and bias while comparing different

masks. Therefore we have carried out patch level mask comparison to address minute differences in HistoQC and our pipeline.

Results 2:

In this experiment, we have systematically sampled patches from WSIs along with foreground masks predicted by our pipeline and by HistoQC. We have selected 100 WSIs from the TCGA dataset for extracting patches in this experiment. We first identify the regions of disagreements between both the masks, and then sample 5 patches from a region where HistoQC predicted a foreground but our pipeline predicted background and vice-versa. With this strategy, we sample 10 patches from each of 100 WSIs, generating a dataset of 1,000 patches along with their predictions by HistoQC and our pipeline. We call a set of 500 patches for which the center pixel of prediction is foreground for HistoQC but background for our pipeline, SET 1 and for another scenario, SET 2. Each of these sets contains 500 patches.

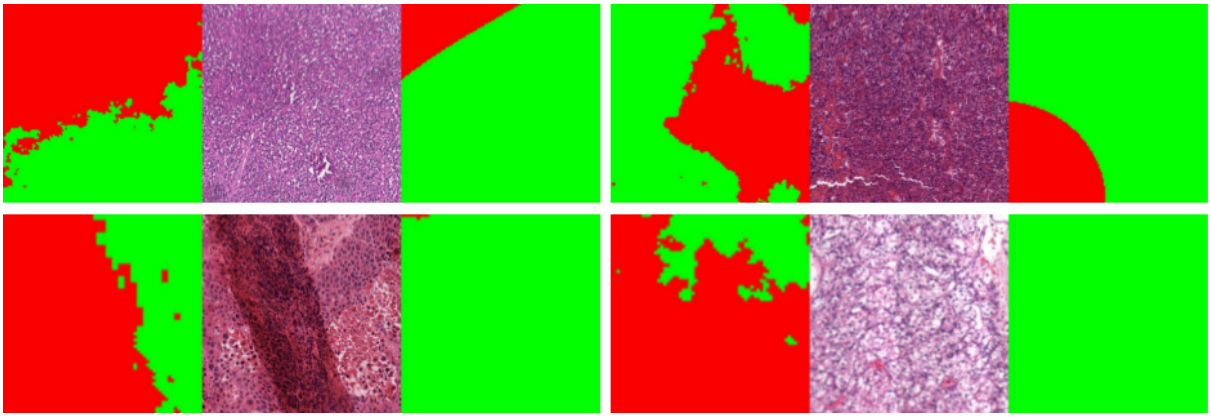

Figure: Sample patches for experiment 2. (Middle) Patch from a WSI. (Left) Prediction from our pipeline. (Right) Prediction from HistoQC. Green: foreground, Red: background.

Out of 500 patches from SET 1, comparison for 127 patches is non conclusive. From the remainder of 373 patches, our pipeline performed better on 289 patches and HistoQC performed better on 94 patches. In the majority of cases, HistoQC can not identify tissue folds as background but our pipeline can predict tissue folds correctly. On the other hand, our model predicts background for debris, etc, which HistoQC correctly predicts as foreground. Analysis of SET 2 uncovers false positives predicted by our pipeline for background class. Overall, performance of our pipeline is better than that of HistoQC in this set as well. Out of 500 patches, 158 were not conclusive, whereas 278 patches from our pipeline were observed to be better than HistoQC. Out of 64 patches where HistoQC performed better, most of these patches correspond to small bubbles or wipes on glass slides.

CONCLUSION:

In this study, we have presented a pipeline for quality control of histopathology images that utilizes four different segmentation models. We have also leveraged the prediction of HistoROI and the cut-paste method to reduce annotation efforts for tissue segmentation and blur segmentation. Our approach has shown superior performance compared to available solutions, as evidenced by our detailed analysis on the largest publicly available dataset to validate our models. Furthermore, we have made all our models, training scripts, training

data, and inference results publicly available for the research community to build more efficient and robust quality control models in the future.

Despite the promising results of our pipeline, there is still room for improvement. One of the major limitations of our approach is that we have not targeted artifacts such as dirt, tissue tear, bubble, etc., which are difficult to annotate and simulate. Currently, these artifacts are predicted as either blur or tissue fold, depending on the semantic closeness of the artifact. However, in the future, we can explore the correlation between the presence of artifacts and develop models specifically targeting these artifacts. For example, in the case of tissue folds, we can almost certainly observe a blurry region around it, and these kinds of observations can help us to build more accurate models.

Another avenue for future work is to explore the use of generative adversarial networks (GANs) to simulate and augment the dataset with artifacts that are difficult to obtain in real-world scenarios. This can help in improving the robustness and effectiveness of our models in identifying and handling artifacts. GANs can be used to generate images with different levels of artifacts, which can be used to augment our existing dataset and improve the performance of our models.

Furthermore, our pipeline currently requires the use of multiple models, which can be computationally expensive and time-consuming. In the future, we can explore the development of a single model that can perform all the tasks efficiently and effectively. This can be achieved by leveraging the strengths of different models and combining them to create a more powerful and efficient model. We can explore the use of multi-task learning or knowledge distillation to train a single model that can perform all the tasks required for quality control of histopathology images. However, this would be a challenging task, as different artifacts are visible or distinguishable at certain magnification levels, and data balancing between multiple classes while training a single model would be a challenge.

In conclusion, our study has presented a novel approach to histopathology image analysis that utilizes hand-annotated datasets and segmentation models to improve the accuracy and efficiency of image analysis. Our pipeline has shown promising results and has the potential to significantly improve the quality control of histopathology images. However, there is still room for improvement, and we hope that our work will inspire further research and development in this area, leading to more accurate and efficient histopathology image analysis techniques in the future.