Artifact Removal and FOXP3+ Biomarker Segmentation for Follicular Lymphomas

Francisco Farinha
UBC Engineering Physics

Nicholas Ioannidis
UBC Engineering Physics

***Abstract*— In this paper we present an image segmentation framework for FOXP3+ Biomarkers in Follicular Lymphomas to determine their amount with respect to the rest of the TMA core. Our process consists of first removing artifacts introduced by the TMA building process by passing them through a segmentation U-Net model. Then, the filtered images are passed through a different U-Net model to obtain biomarker segmentation. The segmented area of the biomarkers is calculated and compared relative to the area of the core tissue to determine the Positivity. The Framework proposed achieves good artifact removal results and comparable biomarker segmentation to the currently used Aperio software.**

I. Introduction

Tumor classification systems have been an area of major interest in Machine Learning mainly because clinical outcomes depend on the long-term experience of pathologists, wherein even amongst experts there is disagreement. Deep Learning is a growing technology in the field of machine learning and it has gotten the attention of many researchers, Jiang et al [1] use convolutional neural networks with small SE-ResNet to classify breast cancer histopathological images. Our goal is to create a Computer-aided diagnosis for FOXP3+ Biomarkers in Follicular Lymphomas.

Follicular lymphoma (FL) is the most common indolent non-Hodgkin lymphoma in the world. Survival and progression of FL has been shown to be significantly associated with certain immune cell profiles, one of which is that of regulatory T cells, whose distribution has shown significant impact in the clinical outcome of patients. The forkhead/winged helix transcription factor 3 (FOXP3) is a transcriptional factor shown to be the key control gene in the development and function of regulatory T cells, and FOXP3+ T cells represent the major regulatory T cell population critical for the self-regulation of the immune system.

Currently, Tissue Microarrays (TMA) are the preferred diagnostic and research tool for FL, since they allow the multiple use of often limited biopsy material in a cost time effective manner in which samples are exposed to the same experimental conditions during analysis. As described by Jawhar [2]. Tissue microarrays are composite paraffin blocks constructed by extracting cylindrical tissue core "biopsies" from different paraffin donor blocks and re-embedding these into a single recipient (microarray) block at defined array coordinates. These blocks often contain impurities and flaws called artifacts which can interfere with accurate counts and distribution of markers. Although there exist visualization tools such as Aperio to assist pathologists in analyzing TMA data, there is no automated solution that takes into account artifacts and is able to localize biomarkers, hence introducing additional bias ontop of the inherent bias from the pathologist's assessment. In this project we aim to develop a framework utilizing TMA image data collected by British Columbia Cancer Institute, to perform both artifact removal and biomarker segmentation.

A U-Net architecture is favoured in Biomedical Image Segmentation due to the limited annotated data, as Ronneberger et al [3] describe, it is able to yield more precise segmentation while utilizing fewer training samples. The architecture consists of the usual contracting network with pooling operators, but it is also supplemented with successive layers of upsampling operators instead of pooling. Additionally, concatenation of previous layers allows for the preservation of lower order features in later steps. For these reasons we utilize a UNet architecture to segment both the artifacts and the biomarkers.

II. Artifact Segmentation

A. Network Architecture

In order to accurately conduct biomarker segmentation it is necessary to first remove artifacts from the image, otherwise there is a significant bias introduced which leads to erroneous results. Fig. 1 displays a sample TMA image of a core cell, where the uniformly dark segment is an artifact.

The U-Net architecture used for the segmentation is illustrated in Fig. 2. It consists of a contracting path on the left side and an expansive path on the right side. The contracting path follows the architecture similar to a convolution neural network, of two 3x3 convolutions, a 2x2 max pooling operation and each followed by a rectified linear unit (ReLU). The expansive path consists of an upsampling of the feature map followed by a 2x2 up-convolution and a concatenation with the correspondingly cropped feature map from the contracting path then two 3x3 convolutions, each followed by a ReLU.

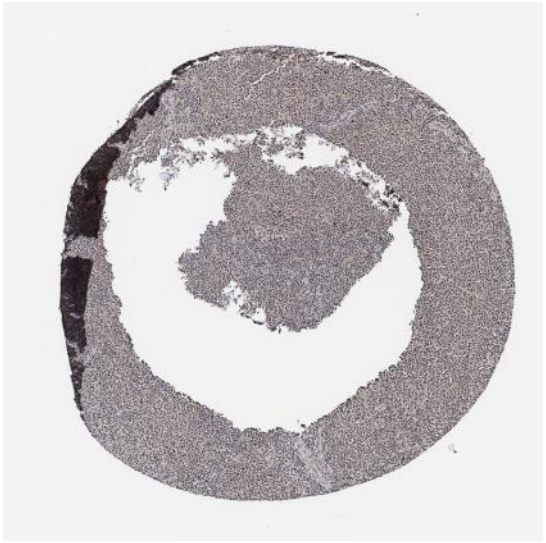


Fig. 1. TMA Image of Cell with Artifact

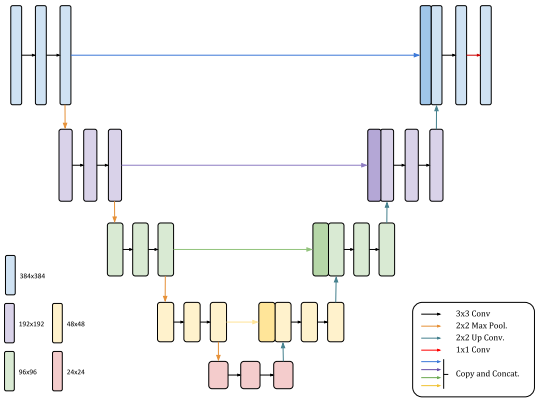


Fig. 2. Artifact Segmentation U-Net Architecture

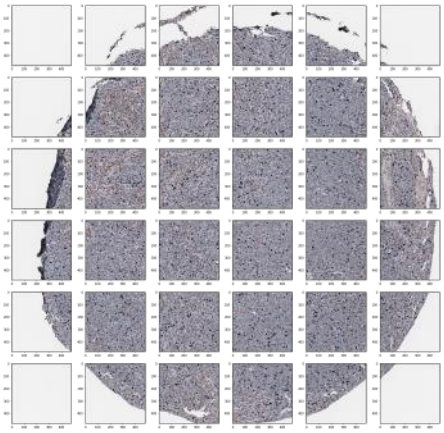


Fig. 3. Patched Sample Image

B. Training

The dataset received from BC Cancer Research Institute contained high resolution images (2886x2886) without annotations, and training the data on a U-Net architecture in contrast to a traditional deep neural network was chosen to reduce the amount of annotated pictures needed for accurate predictions and to make the model less computationally expensive and accelerate experimentation with different hyperparameters. One thing noticed from the first mini-batches trained through the model, was that due to the high resolution of the images the time complexity was not efficient leading to a very long training time. Splitting the images into patches of equal size (481x481) allowed to both reduce the time complexity of the model and increase the dataset of annotated images since one original image was now transformed to 36 patches of equal size (see Fig. 3). When reconstructing the original image an issue arose while fusing the masks from the segmented patches outputted from the neural network, this was due to the inability of the model to extrapolate dependent data from neighbouring patches, and was resolved by adding padding to the patched images. The hyperparameters used for the neural network can be found in more detail in Table 1.

TABLE I
ARTIFACT SEGMENTATION HYPERPARAMETER SELECTION

Hyperparameter	**Value**
Learning Rate	0.0001
Activation Function (Convolution)	ReLU
Activation Funtion (Output)	Sigmoid
Loss Funtion	Binary Cross Entropy
Optimizer	Adam

C. Results

Fig. 4 shows artifact segmentation and removal for three samples. The image on the left corresponds to the raw TMA core, followed by the TMA core with the artifact removed.

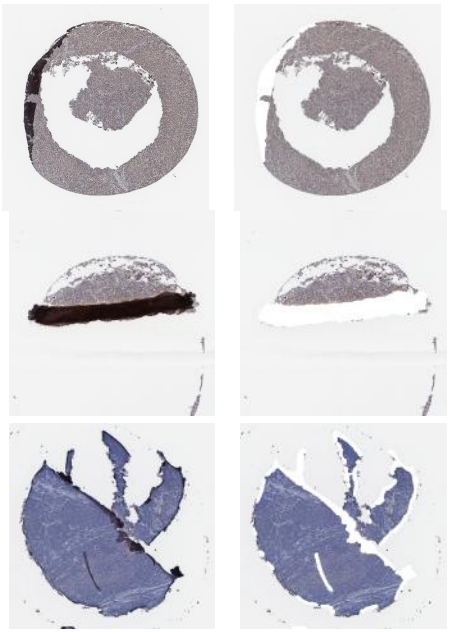


Fig. 4. Examples of Artifact Removal

III. Biomarker Segmentation

A. Network Architecture

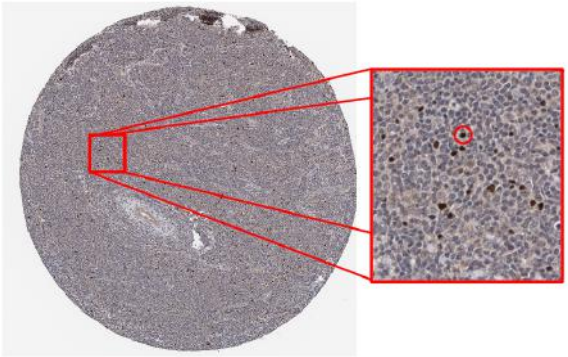


Fig. 5. Biomarker from sample TMA Image

The objective of the analysis is to segment biomakers from TMA cores (see Fig. 5). For reasons similar to the Artifact Segmentation section (efficient allocation of limited annotated images, lower time complexity) the neural network architecture chosen was U-Net. Initially, the same architecture as artifact segmentation was used, and through experimentation it was noted that the output collapsed to zero due to excessive convolutions and up-sampling for the given resolution. Hence, a scaled down version with one less contracting and expansive layer was used (see Fig. 6).

Additionally, upon initial experimentation, checkerboarding was heavily present in the output. Checkboarding refers to the grid-like artifact that can be seen in Fig. 7.

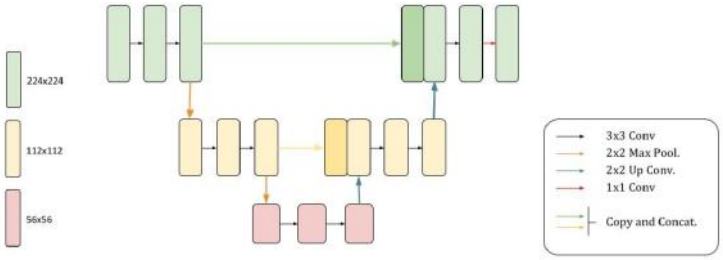
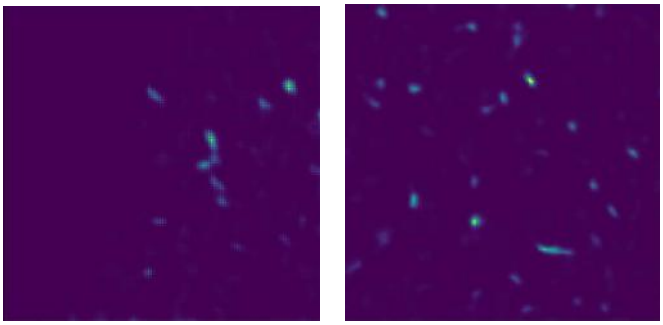


Fig. 6. Biomarker Segmentation U-Net Architecture



(a) Checkerboarding (b) No Checkerboarding

Fig. 7. An example of model output with a) Checkboarding and b) no Checkerboarding. Note these are predictions of different inputs.

This was found to arise from the use of Conv2DTranspose, which was solved by utilizing UpSampling2D with bilinear interpolation.

To save computation time, any TMA cores with less than 10% of tissue present were disregarded before being fed into the model.

B. Training

The dataset for the biomarker neural network consisted of artifactless imaged produced by the artifact removal pipeline. Patches as shown in the previous section were used to preserve global and local structure with uniform size of 224x224.

TABLE II
Biomarker Hyperparameter Selection

Hyperparameter	Value
Learning Rate	0.0001
Activation Function (Convolution)	ReLU
Activation Funtion (Output)	Sigmoid
Loss Funtion	Dice Loss
Optimizer	Adam

For the Loss function, Dice Loss was used to combat class imbalance. Since the positive markers occupy a much smaller area than the rest of the patch, it was necessary to weight the markers more. After 150 Epochs, the Dice Loss for the training dataset was Dice$= 0.00144$, with a validation loss of Dice$=$ 0.00165, suggesting a good fit to the labelled distribution.

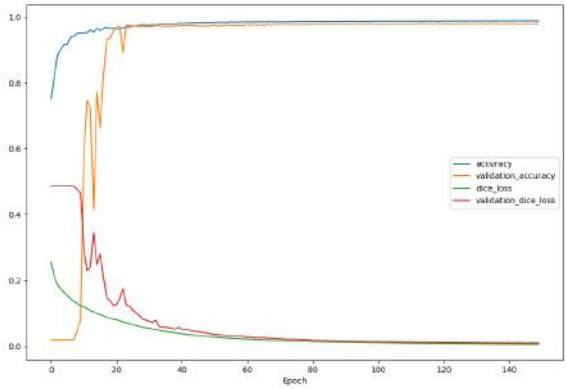


Fig. 8. Training History

C. Results

Fig 9. Show some results from marker segmentation. Of interest is the last row of Fig. 9. where the raw TMA core (left) show lymphoid follicles (brown tint), and the model predictions clearly show clustering within those follicles. This kind of mask could be useful for predicting distributions of markers within follicles which have clinical significance [4]

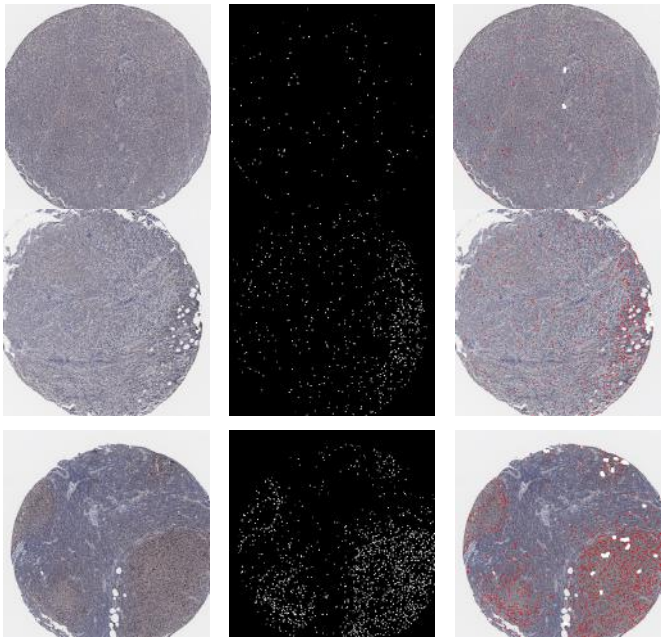


Fig. 9. Marker Segmentation Results. Each row corresponds to a different core. The first image is the raw TMA Core, followed by the predicted mask and then the segmented image.

With the trained artifact segmentation and marker segmentation models, the entire dataset was processed through the pipeline. Processing times averaged $1.51s/it$ for artifact removal and $1.2s/it$ for marker segmentation on a GTX 1050 Ti Max-Q. Processing times for a TMA core in Aperio were reported to be between 10-15 minutes for a 4×12 array without any artifact removal. This places the model at around $4.6-6.9$ times faster with artifact removal (2.2 minutes for a TMA).

Afterwards, data for Aperio was obtained from BC Cancer Agency. The metric of interest is Positivity, which refers to the ratio of positive pixels to total tissue. After dropping nan values from the dataset, there were 984 point of comparison between the model and Aperio. Fig. 10 shows a log-log plot of Aperio vs Model Positivity, which shows obvious correlation. The clear skewness is potentially explained by Aperio's expected bias towards predicting larger values of positivity since it does not have artifact removal, this is discussed further in the *Comments* section. A linear regression was computed which yielded $R^2 = 0.4688$.

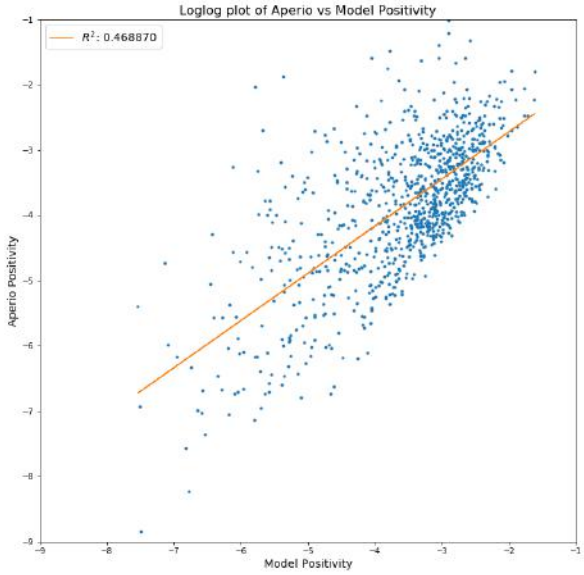


Fig. 10. Model and Aperio Comparison

D. Comments

It is difficult to interpret the comparison between the model and Aperio since the underlying distribution inherently has a high uncertainty. This devolves to comparing two methods of fitting an uncertain distribution, each of which introduces additional bias and uncertainty. Going through some of the outliers in the Mean Squared Error comparison of Positivity ($\frac{NPositive}{NTotal}$) between the model and Aperio reveals some insights into the discrepancy observed. Fig. 9 demonstrates that Aperio often labels artifacts as positive samples.

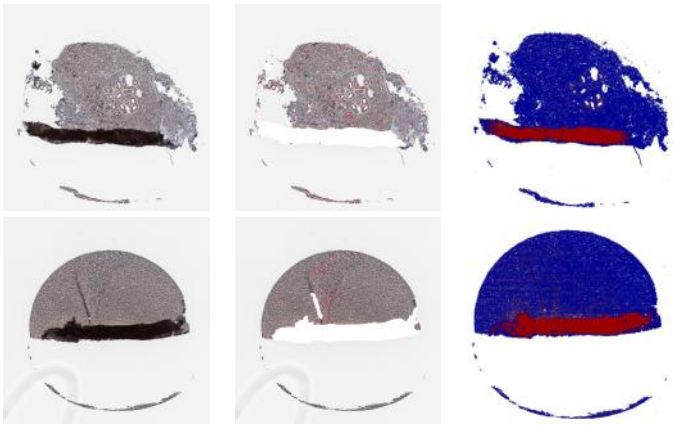


Fig. 11. The first row corresponds to TMA Core D_1_3_12. The first image is the raw core, followed by the model prediction and Aperio's prediction. The model predicted Positivity= 0.026182, whereas Aperio predicted Positivity= 0.143681.
The second row corresponds to TMA Core D_1_5_12. In this case, the model prediction was Positivity= 0.026182 and Aperio's prediction was Positivity= 0.143681.

Additionally, there is also uncertainty regarding the background. Close contrasts make it difficult to segment markers properly. Fig. 11. Shows an example where Aperio wrongly classifies the background.

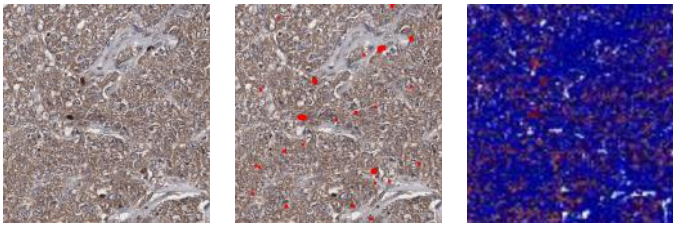


Fig. 12. Background Outliers of section of TMA Core B_1_8_8. The first image is the raw core, followed by the model prediction and Aperio's prediction. The model predicted Positivity= 0.022071 and Aperio predicted Positivity= 0.174185. As shown, Aperio classifies sections of the background incorrectly.

These two categories of outliers - Artifact and Background - make up most of the samples falling outside one standard deviation of the Mean Squared Error between Aperio and Model Positivity. This suggests a large part of the variation between both predictions is due to Aperio failing with artifact and background predictions.

IV. Conclusion and Future Work

The model developed through this project is able to predict Positivity of FOXP3+ given a TMA core while removing artifacts with satisfactory results. Regardless, further improvements are proposed in order to better access the accuracy of the model and expand its diagnostics capabilities.

The output results have yet to be validated by experts and pathologists although performance testing was done using the Aperio software results. This is significant since Aperio also tries to aide diagnosis, but comes with inaccuracies and limitations of its own as stated in the *Comments* section. Furthermore, the problem that this project aims to solve is of high inherent uncertainty, and hence there can be major discrepencies between the output and the expected result depending on the variance of the input image. The model was trained on limited annotated data, therefore an expansion of the labelled dataset would most probably improve its overall performance by capturing more of the inherent variation. Finally to improve the capabilities of this diagnostic tool, future steps include analyzing the marker distribution to identify patterns which are clinically relevant and proven to have impact on survival as analyzed by Farinha et al [4].

References

[1] Jiang Y, Chen L, Zhang H, Xiao X (2019) Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module. PLoS ONE 14(3): e0214587. https://doi.org/10.1371/journal.pone.0214587

[2] Jawhar, Nazar M T. "Tissue Microarray: A rapidly evolving diagnostic and research tool." Annals of Saudi medicine vol. 29,2 (2009): 123-7. doi:10.4103/0256-4947.51806

[3] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." arXiv preprint arXiv:1505.04597 (2015).

[4] Farinha P, Al-Tourah A, Gill K, Klasa R, Connors JM, Gascoyne RD. "The architectural pattern of FOXP3-positive T cells in follicular lymphoma is an independent predictor of survival and histologic transformation." Blood. 2010;115:289–95.