

# A Comparative Study of U-Net Topologies for Background Removal in Histopathology Images

Abtin Riasatian, Maral Rasoolijaberi, Morteza Babaei, H.R. Tizhoosh

*Kimia Lab, University of Waterloo, Canada*

{abtin.riasatian, mrasooli, mbabaie, tizhoosh}@uwaterloo.ca

**Abstract**—During the last decade, the digitization of pathology has gained considerable momentum. Digital pathology offers many advantages including more efficient workflows, easier collaboration as well as a powerful venue for telepathology. At the same time, applying Computer-Aided Diagnosis (CAD) on Whole Slide Images (WSIs) has received substantial attention as a direct result of the digitization. The first step in any image analysis is to extract the tissue. Hence, background removal is an essential prerequisite for efficient and accurate results for many algorithms. In spite of the obvious discrimination for human operator, the identification of tissue regions in WSIs could be challenging for computers mainly due to the existence of color variations and artifacts. Moreover, some cases such as alveolar tissue types, fatty tissues, and tissues with poor staining are difficult to detect. In this paper, we perform experiments on U-Net architecture with different network backbones (different topologies) to remove the background as well as artifacts from WSIs in order to extract the tissue regions. We compare a wide range of backbone networks including MobileNet, VGG16, EfficientNet-B3, ResNet50, ResNext101 and DenseNet121. We trained and evaluated the network on a manually labeled subset of The Cancer Genome Atlas (TCGA) Dataset. EfficientNet-B3 and MobileNet by almost 99% sensitivity and specificity reached the best results.

**Index Terms**—Histopathology, Convolutional Networks, Tissue Segmentation, U-Net, artifact removal.

## I. INTRODUCTION

In the recent decade, the image digitization has recently become more popular in the pathology practice. Improvement in this technology has led to the manufacturing of high-resolution whole-slide scanners which can produce WSIs in a short time. The digital scan of the biopsy glass slides can be explored by image viewers rather than the conventional microscope. Also, despite the large size of scans (a typical WSI file is usually at least several hundred megabytes), new storage and network sharing progress make it possible to share these files much faster than mailing glass samples for the purpose of consultations and acquiring second opinions [1]. An important benefit of digital pathology is that AI and computer vision methods can be applied on tissue scans to help pathologists create more accurate reports [2]. Due to the large size of WSIs, most pathology image processing methods divide the slides into small tiles (patches) before feeding them to the CAD systems. Unquestionably foreground segmentation

is a necessary prerequisite for almost every tile-based method to decrease the time complexity and possibility of making mistakes by the algorithms due to analyzing irrelevant parts [3]. Thus, one has to remove irrelevant pixels from WSIs as much without removing any tissue pixels [4], [5]. Since in medical imaging, histopathology image analysis is generally the last step for cancer diagnosis [6], it is crucial to avoid losing tissue pixels. Therefore, the expected segmentation sensitivity has to be very high.

Another application of tissue foreground segmentation is in whole slide scanners which digitize glass slides containing tissue specimens to generate WSI files. The focus depth of whole slide scanners must be adjusted for different tissue regions due to variable tissue thickness. Hence, scanners need to identify all areas which contain tissue. If an error occurs during digitizing glass slides, there is no way to fix the error in the following steps of the digital pathology workflow. Currently, a technician manually checks every slide after scanning, which is a tedious and expensive procedure [2], [7].

Some of the challenges in tissue segmentation in histopathology images are related to the tissue type. For instance, air sacs in the lung, and fat which could appear in many tissue types, may confuse algorithms due to their resemblance with the background color while they can be easily segmented as tissue by an expert. Another important challenge is the presence of artifacts including bubbles, tissue folds, extra stain, broken glass, debris, and marker traces [4], [8]. Moreover, mistakes in tissue preparation such as weak staining raise difficulties for tissue identification algorithms [2], [9]. Examples of some of the mentioned challenging cases are indicated in Fig.1.

In this paper, we propose a novel method to identify tissue areas in the WSI thumbnail images. The main contributions of this paper are: (1) Releasing a publicly available dataset consisting of 244 thumbnails of TCGA WSIs along with their segmentation masks, (2) proposing a deep learning topology using U-Net for reliable, accurate, and automatic tissue segmentation, and (3) comparing the performance of different encoders as the backbone of U-Net in the tissue segmentation task. The manifest to download the data from the GDC website, manually refined labels, as well as codes to run the proposed U-Net, is available for download<sup>1</sup>.

\* This work was funded by a NSERC-CRD grant on “Design and Development of Devices and Procedures for Recognizing Artefacts and Foreign Tissue Origin for Diagnostic Pathology”

<sup>1</sup><https://kimialab.uwaterloo.ca/kimia/index.php/data-and-code/>

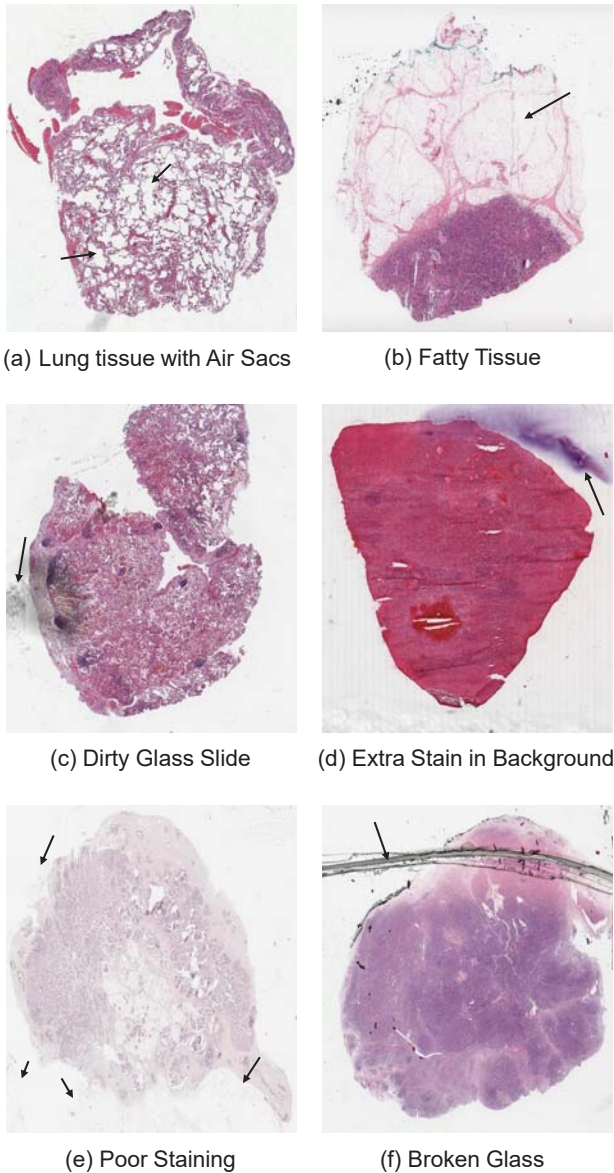


Fig. 1: Some examples for challenges in tissue extraction (images selected from the TCGA dataset).

## II. RELATED LITERATURE

The identification of regions containing tissue is usually the first step in histopathology image analysis. However, this problem is often treated as a trivial part of research mostly solved via threshold-based methods. Most research papers have used empirical rules to set the threshold for different image specifications such as gradient, intensity, color, etc. [9]–[13].

### A. Machine vision based methods

Estimation of the texture complexity in small neighbors has been used by Oswal et al. to detect the foreground [10]. Babaie et al. [11] used homogeneity and gradient values to estimate the patch complexity. Bentaieb et al. [12] used a threshold

on the pixel intensity values to detect the tissue. As another example, Kothari et al. [9] removed blank regions by setting a threshold on saturation and intensity of pixels. Other works used homogeneity criteria to only select patches containing a considerable part of the tissue [14].

The Otsu algorithm [15] as a robust iterative thresholding method has been widely used to compute the optimal threshold. Mohit employed the Otsu method on HSV transformed image for background removal [16]. Nguyen et al. applied Otsu’s method on the b channel of the LAB color space to obtain tissue regions in WSIs [17].

One of the most well known and widely used open-source libraries in digital histopathology, Histomics Toolkit (HistomicsTK)<sup>2</sup> can also perform tissue detection on the thumbnail of a WSI. The process contains a series of Gaussian smoothing and Otsu thresholding. Also, another threshold is used to filter regions smaller than a preset size.

In contrast to the mentioned works, which treat background detection as a small part of the entire WSI processing, there are few studies which have addressed the foreground/background detection in histopathology slides as a major problem [4], [7], [13], [18]. Similar to the previous vision-based methods, FESI [13] used a combination of basic methods, such as median filtering, thresholding, erosion and dilation to address this problem. Calculation of absolute value of the Laplacian based on gray-scale image, and then applying Gaussian filter is used in their work. Recently Chen et al. [19] introduced tissue localization method by applying inverse binarization on the gray-scale images followed by erosion and dilation.

### B. Network based Methods

Neural network based methods are a rather recent trend in the literature to address the tissue segmentation. Raja et al. [4] have extracted four different features including color, appearance, texture and spatial features. They fed the selected features to a two-layer neural network to classify the patches into background and foreground pixels. Bandi et al. [7] trained FCN and U-Net networks for tissue segmentation with patches with a single label. They used patches with the size of  $892 \times 892$  pixels for U-Net and  $128 \times 128$  pixels for FCNN. Their patches were randomly extracted from 54 WSIs. They assigned only one label to each patch based on its central pixel which means the same labels are allocated to roughly 800,000 pixels in the U-Net case. It seems that all network-based methods are working on the highest usually available magnification (namely,  $20\times$  magnification). As a result, for a whole slide processing, a large number of small patches must be fed to their network which is a time-consuming task. However, a more efficient way of segmentation is to assign a label to each pixel in a thumbnail to save time and also to avoid losing tissue parts (especially borders) as much as possible. Therefore, in this paper we provide manually labeled WSI thumbnails (low magnification) to train U-Net models (Fig. 2). We have compared the most commonly used network architectures to find the best backbone for proposed U-Net.

<sup>2</sup><https://github.com/DigitalSlideArchive/HistomicsTK>

### C. U-Net

U-Net is a convolutional neural network which firstly was proposed for the segmentation of neural structures in electron microscopic images in 2015 [20]. Since then, this network has shown impressive performance in various segmentation tasks in medical imaging. Dong et al. [21] proposed an automatic method to detect and segment brain tumors in MRI by using U-Net. Bulten et al. [22] utilized U-Net for epithelial tissue segmentation to assist pathologists in prostate cancer diagnosis. Naylor et al. [23] proposed a method for cell nuclei segmentation by formulating this task as the regression of the distance map. They compared results of three different architectures: (1) the pre-trained VGG16 with fine-tuning as the FCN approach, (2) U-Net, and (3) Mask R-CNN with the pre-trained ResNet 101 as its backbone. U-Net can be trained end-to-end using a small number of images [24]. This is the most significant advantage of the U-Net, especially in applications such as biomedical domain where usually only a few annotated images are available.

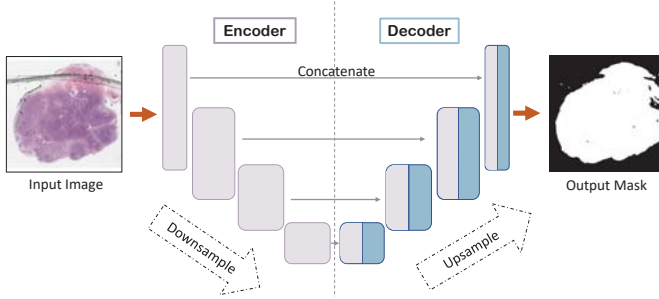


Fig. 2: Network Architecture: Each block shows a feature map.

## III. METHODOLOGY

### A. Data Annotation

For tissue segmentation, a label must be assigned to each pixel to indicate whether it belongs to a tissue region or not. A mask is a binary image where every pixel is either zero (black) or one (white) where the white pixels generally mark the region of interest.

A typical WSI contains more than several thousand pixels in each image axis (i.e., a typical WSI may easily be  $50,000 \times 50,000$  or larger). Thus, assigning a label to each pixel of the WSI is not a feasible task. To overcome this challenge, we first work with thumbnails instead of WSIs, that is generally the image at  $1\times$  magnification. Working with thumbnails has the advantage of fast computation. Also, tissue regions at higher magnifications can be constructed from their corresponding segmented thumbnails by simple calculations commonly known for the pyramidal structures of whole slide images. We developed a handcrafted image processing approach, details in Alg. 1, to produce initial masks from thumbnails [25], [26].

Masking was performed in  $2.5\times$  magnification to preserve details. Note that based on our practical experiments, in

---

### Algorithm 1: Handcrafted Masking Method

---

```

Input : The rgb thumbnail of the WSI
Output: Thumbnail binary mask with the same size

chosenContours  $\leftarrow$  [];
binThmb  $\leftarrow$  binaryThresholding(rgbThmb);
contours, hierarchy  $\leftarrow$  findContours(binThmb);

fatherContours  $\leftarrow$  getContours(contours,
    hierarchy, 0);
append(chosenContours, fatherContours);

firstLevelChildren  $\leftarrow$  getContours(contours,
    hierarchy, 1);
firstLevelChildren  $\leftarrow$  sort(firstLevelChildren,
    'area');
append(chosenContours, firstLevelChildren[0]);
i  $\leftarrow$  1;
while firstLevelChildren[i].area
    > min(firstLevelChildren[i-1].area *
    ratioThreshold, areaThreshold) do
    | append(chosenContours, firstLevelChildren
    | [i]);
    | i  $\leftarrow$  i + 1
end

foreach x in firstLevelChildren do
    | distCond  $\leftarrow$  distanceToClosest(x,
    | chosenContours) < distThreshold;
    | if distCond and x not in chosenContours then
    | | append(chosenContours, x);
end

foreach x in firstLevelChildren do
    | areaCond  $\leftarrow$  getArea(x) > areaThreshold;
    | if areaCond and x not in chosenContours then
    | | append(chosenContours, x);
end
drawContours(chosenContours, finalMask,
    'white');

foreach hole in invert(binThmb) do
    | if hLowerThresh < hole.area < hUpperThresh
    | then
    | | drawContours(hole, finalMask, 'black');
end

```

---

challenging cases, the  $2.5\times$  magnification was the smallest size which still could distinguish the tissue parts from artifacts such as extra staining. Thereafter, initial masks were refined manually to make sure that all tissue regions are selected, and noise and artifacts are removed as much as possible. An example of the mentioned steps can be found in Fig. 3.

With regard to difficult cases, we used image dilation with  $3 \times 3$  kernels to make sure every pixel of tissue, especially

at borders, are preserved. Finally, each pair of mask and thumbnail is resized (preserving the aspect ratio) in a way that each image dimension does not exceed 1024 pixels to make the images small enough to be processed by the network. It is noteworthy that the thumbnails have various dimensions necessitating background padding to have the unified size  $1024 \times 1024$  for all images.

### B. Model Architecture

U-Net, which is a fully convolutional network with a U-shape architecture, has two parts, called encoder and decoder. The first sub-network, known as the encoder, extracts high-level features to capture the image content. The decoder sub-network, also known as the expansion part, creates the desired segmentation map [20]. Fig. 2 shows the proposed network architecture. U-Net-based deep networks, the same as U-Net, include two encoder and decoder sub-networks. As the input image passes through the first sub-network, higher-level features are extracted. In the next sub-network, deep feature maps are combined with low-level feature maps from the encoder sub-network. The spatial resolution of feature maps are increased in the second sub-network to achieve an output mask with the same size as the input image. The connections between the encoder and decoder in U-Net architecture facilitate information propagation. In terms of connections in the U-Net architecture, feature maps from the encoder part are cropped and concatenated to feature maps in the decoder sub-network to retrieve local information. These connections enable the network to learn from a few samples [20].

To improve the performance of U-Net, we applied custom backbones on its architecture using Segmentation\_Models library<sup>3</sup>. The encoder part of these customized networks are the feature extractor, i.e., complete network architecture except the last dense layer, of a chosen network, e.g., MobileNet. The decoder part consists of 5 decoder blocks with filters of size 256, 128, 64, 32 and 16 as it gets deeper. The structure of each decoder block is made up of one 2d-upsampling layer and two repetitions of 2d-convolution, batch-normalization and ReLU activation. Four skip connections connect layers from the encoder part, usually the output of ReLU activation at a certain layer of each encoder block, to the last four decoder blocks, after the up-sampling layer. The last layers of the network is a 2d-convolution layer with Sigmoid activation.

We experimented with six different backbones (topologies) for U-Net-based solutions for tissue segmentation which are introduced in Section IV-B.

## IV. EXPERIMENTS

### A. Data

We used 244 WSIs selected from different organs such as brain, breast, kidney, and lung. All WSIs were randomly selected from The Cancer Genoum Atlas (TCGA) dataset [27], [28]. TCGA is one of the largest publicly available datasets with histopathology whole slide images.

<sup>3</sup>[https://github.com/qubvel/segmentation\\_models](https://github.com/qubvel/segmentation_models)

### B. Topologies and Training Process

We have experimented with various network topologies including MobileNet, VGG16, EfficientNetB3, ResNet50, ResNext101, and DenseNet121 as the backbone of U-Net model to find the most suitable ones for tissue segmentation [29]–[33]. All networks were trained for 50 epochs, no early stopping, using Adam optimizer [34] with the learning rate of  $1e - 4$  on one NVIDIA Tesla V100 GPU with 32GB memory. After running the experiments with two loss functions, namely Jaccard Index and sensitivity plus specificity, we chose the latter so the network tries to come up with an approximation which avoids the misclassification of tissue parts as background while having a good performance at recognizing background. The drawback of using Jaccard Index as the loss function was the relatively low sensitivity of the results. The networks were initialized with ImageNet weights and trained and evaluated with five-fold cross validation. For each fold, 195  $1024 \times 1024$  RGB images were used as the input and binary masks with the same size as the label in which pixel value 1 (positive) meant tissue and pixel value 0 (negative) meant background. Input images and their corresponding masks were augmented by three transformations: (1) Random rotation within the range of -180 and 180 degrees, (2) random horizontal flipping, and (3) random vertical flipping. The validation dataset contained 49 images for each fold.

### C. Comparison of Methods

To compare our results against other methods, we used the same input images fed to our networks as their input and calculated their performance against the ground-truth masks. All methods were checked to be able to work with the given inputs. We compared our results against four traditional computer vision methods:

(1) FESI algorithm [13] is improved by changing the color space of the input image from BGR to LAB and the value of the first two channels, lightness and red/green value, are changed to maximum intensity value<sup>4</sup>. Color space of the resulting image is changed to gray-scale and binerized using the mean value of the image as threshold. This binary image is passed to the Gaussian filter instead of using the absolute value of the Laplacian of the gray-scale image as done in the original paper.

(2) We used *locate\_tissue\_cnts* function available in the open-source Python package<sup>5</sup>, TissueLoc [19], as a recently developed method for comparative purposes. We modified the function in a way that it uses the thumbnail image as input. Also all of input parameters of the function are set to default values except *min\_tissue\_size* which is set to 50 to make sure the algorithm would detect all tissue parts.

(3) Histomics Toolkit Python library is one of the most popular libraries in the histopathology domain. *saliency.tissue\_detection.get\_tissue\_mask* function was used as tissue segmentation method. We set the input parameters

<sup>4</sup>[https://github.com/alexander-rakhlin/he\\_stained\\_fg\\_extraction](https://github.com/alexander-rakhlin/he_stained_fg_extraction)

<sup>5</sup><https://github.com/PingjunChen/tissueloc>

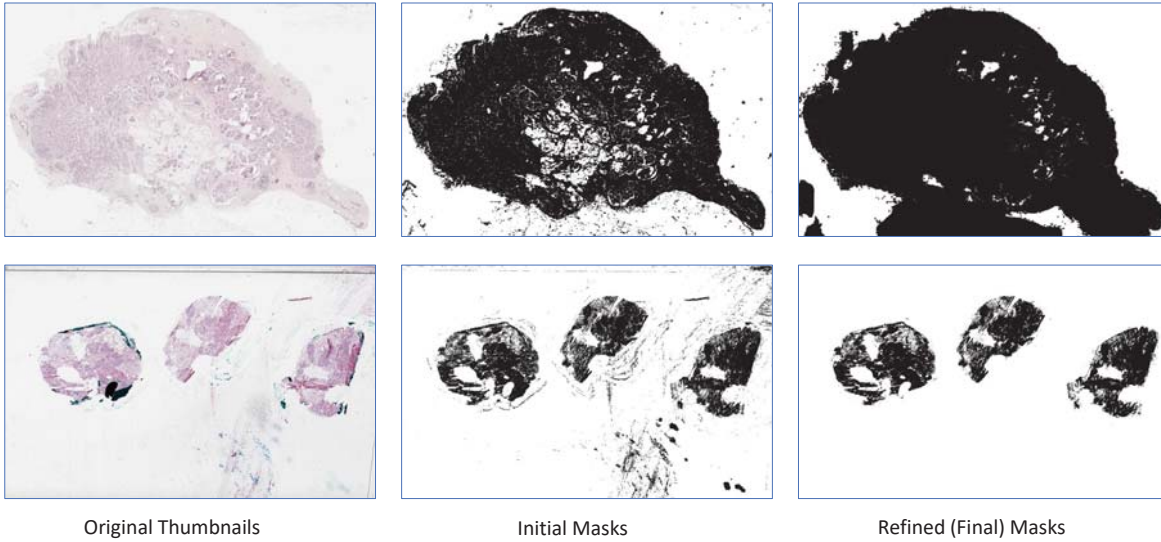


Fig. 3: Two samples of sequential steps for generating masks.

TABLE I: Summary of results: Comparing our networks with image-processing methods (gray rows).

Method	Time (s)	Jaccard Index	Dice Coeff.	Sensitivity	Specificity
MobileNet	0.11	0.95	0.97	0.99	0.99
EfficientNet-B3	0.18	0.95	0.97	0.99	0.98
ResNet50	0.16	0.94	0.97	0.99	0.98
DenseNet121	0.16	0.93	0.96	0.99	0.98
ResNext101	0.50	0.93	0.96	0.99	0.98
VGG16	0.11	0.75	0.82	0.99	0.81
Improved FESI [13]	0.11	0.86	0.92	0.91	0.97
TissueLoc [19]	0.26	0.81	0.88	0.88	0.97
Otsu algorithm	0.02	0.81	0.89	0.82	0.99
Histomics-TK	0.13	0.78	0.87	0.79	0.99

*deconvolve\_first* to False, *n\_thresholding\_steps* to 1 and *min\_size* threshold to 50.

(4) Otsu binarization method as one of the well-known algorithms to classify pixels into foreground and background. The RGB thumbnail images are first converted to gray-scale and then the Otsu method is applied.

#### D. Performance Evaluation

In the test phase, we compared the result of all methods with the ground-truths via 5-fold cross validation. In addition to the processing time, four different performance measurements, i.e., Jaccard index, Dice coefficient, sensitivity, and specificity, were conducted which are defined as:

$$\text{Jaccard} := \frac{TP}{TP + FP + FN}, \quad (1)$$

$$\text{Dice} := \frac{2 * TP}{2 * TP + FP + FN}, \quad (2)$$

$$\text{Sensitivity} := \frac{TP}{TP + FN}, \quad (3)$$

$$\text{Specificity} := \frac{TN}{TN + FP}, \quad (4)$$

where TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives,

respectively. The segmented pixels are considered as positive and negative where they are labeled as tissue and non-tissue pixels, respectively. Note that the sensitivity is more important than the specificity in the tissue segmentation task because sensitivity penalizes wrong labeling of tissue region as background while specificity gives a penalty to the wrong labeling of the background region as tissue. In histopathology, it is paramount to do not miss any part of the tissue.

#### E. Analysis of Results

In addition to Improved FESI and TissueLoc methods, we chose HistomicsTK tissue segmentation and Otsu algorithm to compare our networks' results against well-known methods for histopathology image analysis. Table.I shows that all networks, except VGG16, outperform all four handcrafted methods with respect to all performance measurements. The most important advantage of networks over the handcrafted methods is their high sensitivity ( $\approx 99\%$ ). In addition networks are as fast as handcrafted methods such as Improved FESI and TissueLoc while achieving considerably higher Jaccard Index and Dice Coefficients. It can be seen that Jaccard Index for the networks with best performance, namely MobileNet and EfficientNet-B3, is 9% higher than the best handcrafted method, namely Improved FESI. Considering the changes in the validation

loss for two networks, ResNext101 with around 51 million parameters and EfficientNet-B3 with less than 18 million parameters, through 50 epochs, Fig.5, it seems that both have the same pattern; 20 epochs appeared to be enough for proper network training. This would take around 20 minutes for a medium-size network and 40 minutes for a large network which is a negligible cost considering the benefits of using networks.

To compare network backbones, it can be seen that MobileNet showed the best performance. Also EfficientNet-B3 shows very high performance. The poor performance of VGG16 could be due to several reasons. First of all, this network has a large number of parameters (more than 23 millions) while it only has 66 layers compared to other networks such as MobileNet with more than 8 million parameters and 128 layers and EfficientNet-B3 with around 18 million parameters and 418 layers. Also, the use of batch normalization and ReLU activation layers in the convolution blocks in other architectures have the benefits of avoiding internal covariate shift, which results in faster convergence, and keeping the network sparse, causing the generalization error to decrease, respectively. Since VGG16 lacks these layers in its architecture, it converges with difficulty. Fig. 6 depicts a visual overview of the proposed network results versus image-processing methods. As we can see, the proposed networks outperform in fatty tissue (second column) or tissue with an air bubble (third column) considerably.

## V. CONCLUSION

In this paper, we have compared the performance of U-Net with various custom topologies (backbones) for the identification of tissue regions in whole slide images. Using different networks combines the strength of current state-of-the-art CNNs with the custom architecture of the U-Net model for image segmentation. Whereas U-Net topologies can generate segments with 99% sensitivity, handcrafted methods struggled to approach high 80%. Both MobileNet and EfficientNet-B3 appeared to be the best backbone topology for the U-Net.

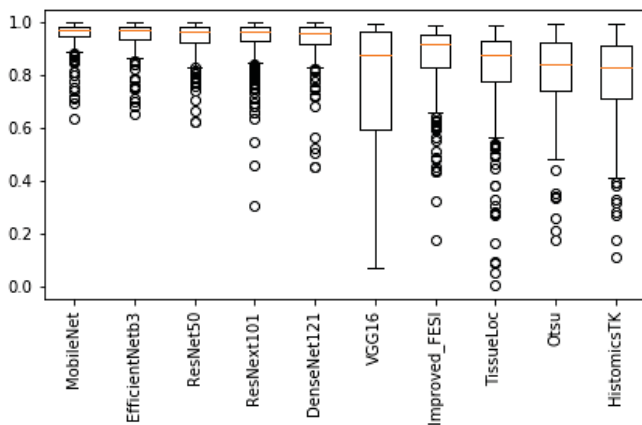


Fig. 4: Jaccard Index

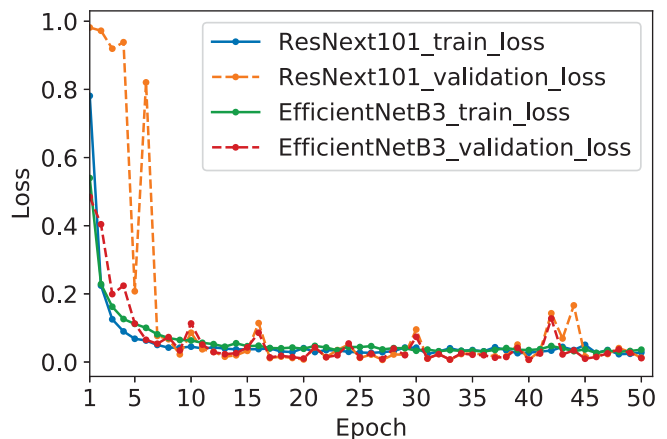


Fig. 5: Losses for ResNext101 and EfficientNet-B3

The next step for this research would be changing the current binary masking network to a multi-class one which could label each pixel as classes such as marker trace, dirt and tissue fold, fat and informative tissue. Authors have made the segmented images publicly available for sake of reproducibility.

## REFERENCES

- [1] S. Al-Janabi, A. Huisman, and P. J. Van Diest, "Digital pathology: current status and future perspectives," *Histopathology*, vol. 61, no. 1, pp. 1–9, 2012.
- [2] L. Pantanowitz, P. N. Valenstein, A. J. Evans, K. J. Kaplan, J. D. Pfeifer, D. C. Wilbur, L. C. Collins, and T. J. Colgan, "Review of the current state of whole slide imaging in pathology," *Journal of pathology informatics*, vol. 2, 2011.
- [3] M. D. Kumar, M. Babaie, and H. R. Tizhoosh, "Deep barcodes for fast retrieval of histopathology scans," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.
- [4] R. S. Alomari, R. Allen, B. Sabata, and V. Chaudhary, "Localization of tissues in high-resolution digital anatomic pathology images," in *Medical Imaging 2009: Computer-Aided Diagnosis*, vol. 7260. International Society for Optics and Photonics, 2009, p. 726016.
- [5] A. M. Khan, H. El-Daly, and N. Rajpoot, "Ranpec: Random projections with ensemble clustering for segmentation of tumor areas in breast histology images," in *Medical Image Understanding and Analysis*, 2012, pp. 17–23.
- [6] L. He, L. R. Long, S. Antani, and G. R. Thoma, "Histology image analysis for carcinoma detection and grading," *Computer methods and programs in biomedicine*, vol. 107, no. 3, pp. 538–556, 2012.
- [7] P. Bándi, R. van de Loo, M. Intezar, D. Geijs, F. Ciompi, B. van Ginneken, J. van der Laak, and G. Litjens, "Comparison of different methods for tissue segmentation in histopathological whole-slide images," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 591–595.
- [8] M. Babaie and H. R. Tizhoosh, "Deep features for tissue-fold detection in histopathology images," *arXiv preprint arXiv:1903.07011*, 2019.
- [9] S. Kothari, J. H. Phan, and M. D. Wang, "Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade," *Journal of pathology informatics*, vol. 4, 2013.
- [10] V. Oswal, A. Belle, R. Diegelmann, and K. Najarian, "An entropy-based automated cell nuclei segmentation and quantification: application in analysis of wound healing process," *Computational and mathematical methods in medicine*, vol. 2013, 2013.
- [11] M. Babaie, S. Kalra, A. Sriram, C. Mitcheltree, S. Zhu, A. Khatami, S. Rahnamayan, and H. R. Tizhoosh, "Classification and retrieval of digital pathology scans: A new dataset," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

- [12] A. BenTaieb and G. Hamarneh, "Predicting cancer with a recurrent visual attention model for histopathology images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 129–137.
- [13] D. Bug, F. Feuerhake, and D. Merhof, "Foreground extraction for histopathological whole slide imaging," in *Bildverarbeitung für die Medizin 2015*. Springer, 2015, pp. 419–424.
- [14] H. Erfankhah, M. Yazdi, M. Babaie, and H. R. Tizhoosh, "Heterogeneity-aware local binary patterns for retrieval of histopathology images," *IEEE Access*, vol. 7, pp. 18354–18367, 2019.
- [15] T. Otsu and M. Yoshida, "Role of initiator-transfer agent-terminator (iniferter) in radical polymerizations: Polymer design by organic disulfides as iniferters," *Die Makromolekulare Chemie, Rapid Communications*, vol. 3, no. 2, pp. 127–132, 1982.
- [16] M. Mohit, "Automated histopathological analyses at scale," Ph.D. dissertation, Massachusetts Institute of Technology, 2017.
- [17] K. Nguyen, A. K. Jain, and B. Sabata, "Prostate cancer detection: Fusion of cytological and textural features," *Journal of pathology informatics*, vol. 2, 2011.
- [18] S. Fouad, D. Randell, A. Galton, H. Mehanna, and G. Landini, "Unsupervised morphological segmentation of tissue compartments in histopathological images," *PLoS one*, vol. 12, no. 11, p. e0188717, 2017.
- [19] P. Chen and L. Yang, "tissueloc: Whole slide digital pathology image tissue localization." *J. Open Source Software*, vol. 4, no. 33, p. 1148, 2019.
- [20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [21] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using u-net based fully convolutional networks," in *annual conference on medical image understanding and analysis*. Springer, 2017, pp. 506–517.
- [22] W. Bulten, C. A. Hulsbergen-van de Kaa, J. van der Laak, G. J. Litjens *et al.*, "Automated segmentation of epithelial tissue in prostatectomy slides using deep learning," in *Medical Imaging 2018: Digital Pathology*, vol. 10581. International Society for Optics and Photonics, 2018, p. 105810S.
- [23] P. Naylor, M. Laé, F. Reyat, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 448–459, 2018.
- [24] P. Yakubovskiy, "Segmentation models," [https://github.com/qubvel/segmentation\\_models](https://github.com/qubvel/segmentation_models), 2019.
- [25] J. S. and S. Prasanna, "Measuring distinct regions of grayscale image using pixel values," *International Journal of Engineering Technology*, vol. 7, p. 121, 12 2017.
- [26] A. Belsare and M. Mushrif, "Histopathological image analysis using image processing techniques: An overview," *Signal & Image Processing*, vol. 3, no. 4, p. 23, 2012.
- [27] "The cancer genome atlas (tcga) dataset," <https://portal.gdc.cancer.gov/>, accessed: November 2019.
- [28] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "The cancer genome atlas (tcga): an immeasurable source of knowledge," *Contemporary oncology*, vol. 19, no. 1A, p. A68, 2015.
- [29] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [31] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [33] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

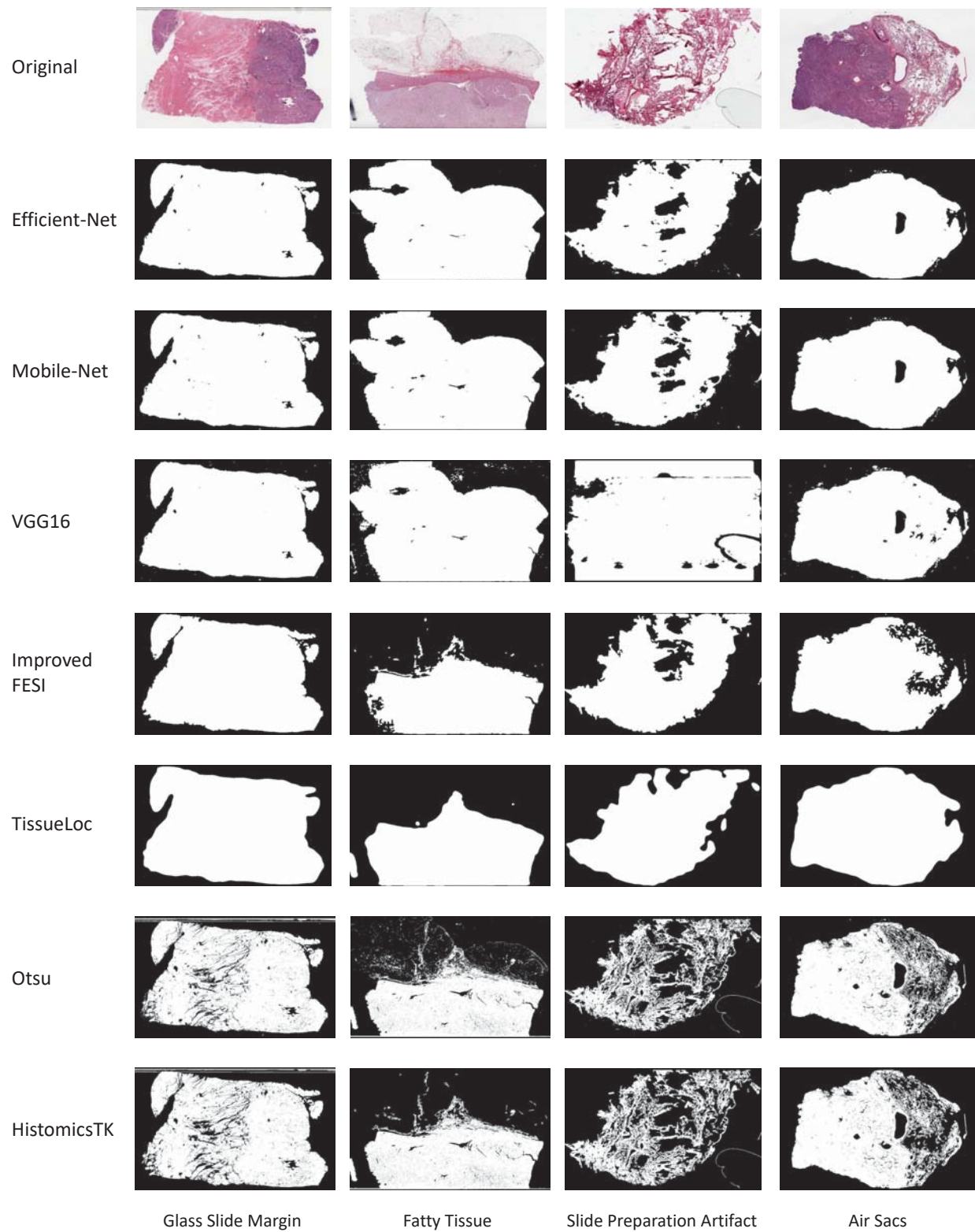


Fig. 6: Example results of different methods in tissue segmentation task for challenging cases.