Ink Marker Segmentation in Histopathology Images Using Deep Learning *

Danial Maleki¹, Mehdi Afshari¹, Morteza Babaie¹[0000–0002–6916–5941], and H.R. Tizhoosh¹[0000–0001–5488–601X]

Kimia Lab, University of Waterloo, Canada
(dmaleki, m4afshar, mbabaie and tizhoosh)@uwaterloo.ca

Abstract. Due to the recent advancements in machine vision, digital pathology has gained significant attention. Histopathology images are distinctly rich in visual information. The tissue glass slide images are utilized for disease diagnosis. Researchers study many methods to process histopathology images and facilitate fast and reliable diagnosis; therefore, the availability of high-quality slides becomes paramount. The quality of the images can be negatively affected when the glass slides are ink-marked by pathologists to delineate regions of interest. As an example, in one of the largest public histopathology datasets, The Cancer Genome Atlas (TCGA), approximately 12% of the digitized slides are affected by manual delineations through ink markings. To process these open-access slide images and other repositories for the design and validation of new methods, an algorithm to detect the marked regions of the images is essential to avoid confusing tissue pixels with ink-colored pixels for computer methods. In this study, we propose to segment the ink-marked areas of pathology patches through a deep network. A dataset from 79 whole slide images with 4,305 patches was created and different networks were trained. Finally, the results showed an FPN model with the EfficientNet-B3 as the backbone was found to be the superior configuration with an F1 score of 94.53%.

Keywords: Histopathology, whole slide images, convolutional neural Networks, Ink Marker Segmentation, U-Net, FPN, artifact removal.

1 Introduction

The definitive diagnosis of numerous diseases is possible through meticulous visual inspection by a trained pathologist, an examination that requires considerable time and effort while inevitably being prone to error. Computer vision and machine learning can assist the pathologists to reduce their workload and to increase the diagnostic accuracy. Computer-assisted diagnosis necessitates the digitization of biopsy glass slides. In some cases, pathologists use ink markers

* This work was supported by an NSERC-CRD Grant on “Design and Development of Devices and Procedures for Recognizing Artefacts and Foreign Tissue Origin for Diagnostic Pathology”
to highlight parts of glass slides for a variety of purposes such as educational and diagnosis hints. Once the slides are digitized, marker signs naturally appear in the digital images as well. Consequently, computer methods are potentially prone to mistaking marker colored regions for tissue information [1].

The pathology slides contain valuable information in high resolution; therefore, rejection of an entire slide because of some ink markings may not be a feasible approach. To enable reliable processing with images with marker signs, we have used deep learning models to detect and removed different marker colors. Several deep learning strategies have been explored to generate a mask that extracts the ink marker areas automatically. The aim of this task is to distinguish between the tissue areas and the areas that are colorized through the marker’s ink. Elimination of areas that have marker signs is useful since in most cases areas that are covered by marker are of less importance (therefore overpainted by the pathologists to show something more important). As a result, discarding these areas may generally not affect the tissue relevant for diagnosis. Moreover, removing manual markings can help to generate a considerable number of proper patches without the presence of the marker artifacts for the benefit of many machine learning algorithms.

Smart software for digital scanners is another potential application of this research. The focus depth of whole slide scanners must be adjusted for different tissue regions due to variable tissue thickness. Focus points on the marker area will considerably affect the scan quality. Avoiding to set focus points on the marker places from a pre-scan image, could boost the automated scanning process and hence improve the lab workflow in digital pathology [3].

Three main challenges concerning the detection of ink-marking areas should be addressed. The first one is that these markings can be drawn by pens with different colors. The worst case in this regard occurs when ink markers have a color similar to the tissue such as red or pink. The second problem is the composition of inconsistent patterns and ink markers’ shapes/symbols. They can be letters, circles, or other shapes such as arrows and dots. The third issue is that ink markers can be created/seen at different transparency levels. In addition, markers may cross the tissue regions, mark outside or around the tissue, which can affect the observed color of the marker. A combination of these scenarios in a single image makes the discovery of marked regions a challenging task [23].

Traditional methods, such as thresholding, may not be accurate enough because of the difficulties mentioned above. An automatic approach that is capable of overcoming these issues is the goal of this study. The advent of deep learning has led to multiple breakthroughs in different domains. One of these domains is medical imaging which consists of numerous tasks including classification, image search, and segmentation [21][3]. In this work, a comparative study of deep network segmentation models (U-net and FPN) that can generate a mask of areas that contain ink markers is presented. The areas that include ink marker are determined, and then a binary mask of the detected regions is generated. By having this mask, no trace of markers may appear in the patch selection phase avoiding the loss of tissue information. The method is accurate, tolerant and
2 Related Work

Ink marker segmentation could be considered as a part of the pre-processing of histopathology images. The main goal of pre-processing methods is to prepare high-quality images for the training procedure. Most of the pre-processing methods for histopathology images ignore the presence of ink marker in images. Taqi et al. [21] investigated various types of artifacts that may appear in histopathological images and how could one differentiate an artifact and a tissue segment. The presence of these artifacts may render images useless for computer aided diagnosis when accessing existing repositories. Rastogi et al. [17] demonstrated that the presence of artifact can be a major pitfall. The artifacts may occur in different stages of the process. Therefore, it is essential to identify the presence of artifacts. Janowczyk et al. [8] mentioned that variability of histopathology images can mislead the diagnostic algorithms. Stain normalization using sparse autoencoders has been proposed to avoid the problem. However, the presence of ink markers as an artifact has not been discussed in these studies.

A few techniques are developed for the detection and removal of the ink marker signs in histopathology images. These methods could be categorized into two following types. The first type is the image processing methods based on traditional algorithms. Most of the techniques that are proposed to remove marker signs use thresholding techniques combined with mathematical morphology operations. In this approach, different types of color space, filters, and thresholds for different ink marker colors have been used. HistoQC is a tool presented for the quality control of WSIs which discovers artifacts and outputs outliers. The method was validated using open-access dataset of TCGA and the results were verified by pathologists. However, the precision of the outlier detection of the method is not high enough which can result in loss of valuable data. Mostly, these methods are not fully automated. Due to the high variation of the markers’ color and intensity, thresholds fail in different images. As a result, the manual setting of the thresholds or other parameters for each image may be required.

The second method type is deep learning. Recent improvements in deep learning enable the modeling of complex data distributions. One of the well-known tasks that are done using deep learning techniques is image segmentation and reconstruction. Venkatesh et al. [23] used CycleGan for the reconstruction of a marker removed WSI. Sharib et al. [1] separated tiles that were contaminated with markers with a binary classification network and fed the tiles with marker signs to a YOLO3 network [18] to detect the marker areas. Finally, for the restoration of the original non-marked tiles, CycleGAN was used [24].

1 https://github.com/deroneriksson/python-wsi-preprocessing/blob/master/docs/wsi-preprocessing-in-python/index.md
2.1 Segmentation models

U-Net has been developed to segment biomedical images [19]. The network architecture has two major parts. The first part is the encoding path which is used to capture the image context. In this part, convolutional layers are applied which followed by max-pooling layers. The second part uses transposed convolution layers and up-sampling layers to construct an image with the same size as the input image. The main goal is to add skip connections from the down-sampling part to the second part as an up-sampling part which can help the model to have better localization information. The down-sampling layers output is then concatenated with the input of up-sampling layers. Figure 1 shows the proposed configuration when U-Net architecture is used. For the pathology practice, U-Net has been utilized for a wide range of applications, for instance for epithelial tissue segmentation [4] and cell nuclei segmentation [16].

![Figure 1](image1.png)

**Fig. 1.** Overall structure of the proposed method for segmentation of ink marked images using U-Net architecture network. Blue boxes show the CNN layers.

Feature Pyramid Network (FPN) [14] like U-Net has two parts. The first part which is a bottom-up pathway has a convolutional network as backbone. The output of each stage in this part will be used as a skip connection for enriching the second part layers. The second part of the model, which is named top-down pathway, uses transposed convolution and up-sampling layers to up-sample features. Afterward, FPN concatenates each stage in this pathway with skip connections from the bottom-up pathway. In the end, the network concatenates the output of each stage after feeding each of them to the convolutional layer with a kernel size of three and predicts the segmentation for the input image. Fig. 2 shows the FPN architecture used in this study.
2.2 Transfer learning

Providing large sets of labeled data in medical imaging is expensive [10]. Recently, transfer learning has become a feasible in the medical imaging domain [11]. Transfer learning helps to build new models in a shorter time requiring a rather small number of training samples. In computer vision, transfer learning is mainly implemented through the use of pre-trained models. A pre-trained model is a network that is trained on a large dataset for a similar task. A pre-trained model which was trained on the ImageNet dataset [5] with more than 14 million images for the encoding pathway that can extract features by using a few images from a new domain for the segmentation task. For instance, in EfficientNet [20] the main idea is to find an optimal baseline network using Neural Architecture Search (NAS) with respect to a trade-off between accuracy and FLOPS (Floating-point Operations Per Second) and then scale the depth, width and resolution through an equation. By changing the parameter in the formula, different network settings can be achieved that have different FLOPS and accuracy levels.

In ResNet [7], using skip connections is the main idea. Skip connections between ResNet blocks help the network to perform the backward propagation of the loss better and to prevent gradient vanishing that is a common problem in very deep networks.

3 Method

The goal of this study is a fully automatic Convolutional Neural Network (CNN) that generates a binary mask with the same size of the input image. Two of the most popular architectures for the medical image segmentation namely U-Net
Accepted for publication in the 15th International Symposium on Visual Computing (ISVC 2020)

and FPN, are chosen and investigated in this paper. These networks achieved success in many segmentation tasks mainly due to skipping connections from the down-sampling pathways to the up-sampling pathways. On the other hand, due to the limited number of training samples, transfer learning approach is utilized by using a pre-trained network as the backbone of segmentation models. Different backbone networks are compared to discover a suitable pre-trained network for the segmentation task. Selected backbones are EfficientNet and ResNet.

In this study, our network uses a combination of two types of cost functions. One of them is the Dice loss function and the other one is the Focal loss function [15]. The dice loss function principally measures the overlap between two binary samples. The measurement range is 0 to 1 where a dice coefficient value of 1 denotes a perfect match between two samples which in this case would be the target and the network output. The dice loss function can be calculated as follows:

\[
\text{Dice} = \frac{2|\text{Output} \cap \text{Target}|}{|\text{Output}| + |\text{Target}|},
\]

where \(|\text{output} \cap \text{target}|\) denotes the common pixels between output and target and \(|\text{output}|\) shows the number of pixels in output (and likewise for target).

The second term of the cost function is the focal loss. The distribution of each class pixels in the data is imbalanced which serves the definition of focal loss term. The term avoids the complication of correct pixels number and the total number of pixels. Relatively speaking, uncomplicated examples are well classified; therefore, these examples contribute to the loss rather in a minor fashion. The focal loss function can be described as:

\[
\text{FL}(p_t) = -(1 - p_t)^\gamma \log (p_t).
\]

In this formula, \(\gamma\) is a tunable parameter. This controlling element defines the steepness of the function shape. Parameter \(p_t\) is the probability of the background class. The term puts more emphasis on hard samples while reduces loss for well-classified examples.

4 Experiments and Results

To train the model, 89 ink marked TCGA [22] slides were selected visually through diverse body parts. The ground truth masks (ink marker areas) of selected slides are created manually. Then, each WSI has been patched into images with the size of 256 × 256 pixels at 1× magnification. The total number of 4,305 patches are extracted from 89 annotated slides to train the model. We applied 8-fold cross-validation to train and test our model. Therefore, training, validation, and test data are set to 70%, 15% and 15% from the randomly selected patches, respectively. The encoder pathway is the pre-trained model and different models are trained using the Adam optimizer [12] with the learning rate set to 0.00001. Each model is trained for 60 epochs with one Nvidia Tesla V100 GPU. The maximum training time of a network is roughly 160 seconds per epoch. On the other
The hand, the minimum training time is 60 seconds per epoch for the smallest size network as U-Net architecture and EfficientNet-B0 backbone. Two evaluation criteria are used. The first one is the Jaccard index defined as:

\[
\text{Jaccard index} = \frac{|Output \cap Target|}{|Output \cup Target|}.
\] (3)

The second one is the F1-score which measures the sensitivity and specificity together. As discussed in the method section, the proposed model can adapt to different architectures and backbones. FPN and U-Net are used for the architecture and EfficientNet-B0, EfficientNet-B3 and ResNet-18 are used for the encoder part. Fig. 3 (left) shows the Jaccard index for all training scenarios over the training process. The vertical axis shows trained epochs and the horizontal axis shows the Jaccard index for the validation data. The Jaccard index trend shows the fastest learning trend of FPN architecture with ResNet backbone with respect to number of epochs. Our study shows that ResNet-18 with FPN architecture has the steepest slope at the starting epochs, but at the end, other backbones show better performance. U-Net architecture with EfficientNet-B0 and EfficientNet-B3 show similar trends as the slowest learners with respect to the number of training epochs. All networks reach a considerably high Jaccard index at the end of training process. Comparison among the validation loss values are shown in Fig. 3 (right) for all architectures and backbones. As expected, the loss value over epochs has correlation with Jaccard criterion. Fig. 3 (left) shows the importance of architecture based on loss. Different architectures show varied behaviour in the course of loss value trends. FPN architecture has lower loss value in comparison with U-Net architecture through all epochs.

Table 1 shows the evaluation on the test set (average of 8-fold cross validations) of the trained models on the presented dataset. In the ink marker segmentation problem, FPN architecture shows better performance comparing U-Net architecture. FPN architecture with ResNet-18 backbone performed less accurate compared with other networks. The reason of lower F1-score is the lower capacity of the network. However, the smaller size of network results in favorable
computation time. The results show that FPN architecture is a fast learner for the desired task compared with the U-Net architecture. The reason of this faster learning trend is the additional part that appears after $1 \times 1$ convolution that FPN architecture has.

Fig. 4 shows some sample input-output pairs of trained networks. The samples include different colors, intensities, and overlaps with the tissue.

![Sample Input-Output Pairs](image)

**Fig. 4.** WSIs (top) and ground-truths (bottom): The network segmentation is applied on individual patches and then combined to a complete image.

| Metrics | Base | Backbone | IoU  | F1 Score | Time (s) |
|---------|------|----------|------|----------|----------|
|         | U-Net| EfficientNet-B0 | 0.8244 | 0.9128 | 0.0238 |
|         |      | EfficientNet-B3 | 0.8251 | 0.9227 | 0.029  |
|         |      | ResNet-18     | 0.8234 | 0.9091 | 0.0195 |
| FPN     |      | EfficientNet-B0 | 0.8326 | 0.9312 | 0.028  |
|         |      | EfficientNet-B3 | 0.8354 | 0.9453 | 0.0352 |
|         |      | ResNet-18     | 0.8154 | 0.8912 | 0.0217 |
| Baseline|      | -           | 0.6512 | 0.7312 | 0.0189 |

2 https://github.com/deroneriksson/python-wsi-preprocessing/blob/master/docs/wsi-preprocessing-in-python/index.md
5 Conclusions

A fully automatic deep model to generate a binary mask of areas that have ink markers on archived WSIs is proposed and evaluated in this study. The investigated method is capable of the extraction of ink markers with different colors and different shapes and intensities. Calculation of two evaluation metrics including the Jaccard index and F1-score showed the efficiency of the method. Removing ink markings helps to use all archived WSIs that have ink highlights for research instead of discarding them. As well, this method can be added to digital scanners to apply the ink removal during the image acquisition.

References

1. Ali, S., Alham, N.K., Verrill, C., Rittscher, J.: Ink removal from histopathology whole slide images by combining classification, detection and image generation models. arXiv preprint arXiv:1905.04385 (2019)
2. Babaie, M., Kalra, S., Sriram, A., Mitcheltree, C., Zhu, S., Khatami, A., Rahnamayan, S., Tizhoosh, H.R.: Classification and retrieval of digital pathology scans: A new dataset. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (July 2017)
3. Babaie, M., Tizhoosh, H.R.: Deep features for tissue-fold detection in histopathology images. arXiv preprint arXiv:1903.07011 (2019)
4. Bulten, W., Hulsbergen-van de Kaa, C.A., van der Laak, J., Litjens, G.J., et al.: Automated segmentation of epithelial tissue in prostatectomy slides using deep learning. In: Medical Imaging 2018: Digital Pathology. vol. 10581, p. 105810S. International Society for Optics and Photonics (2018)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
6. Fuchs, T.J., Buhmann, J.M.: Computational pathology: Challenges and promises for tissue analysis. Computerized Medical Imaging and Graphics 35(7-8), 515–530 (2011)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
8. Janowczyk, A., Basavanhally, A., Madabhushi, A.: Stain normalization using sparse autoencoders (stanosa): Application to digital pathology. Computerized Medical Imaging and Graphics 57, 50–61 (2017)
9. Janowczyk, A., Zuo, R., Gilmore, H., Feldman, M., Madabhushi, A.: Histoqc: An open-source quality control tool for digital pathology slides. JCO clinical cancer informatics 3, 1–7 (2019)
10. Kalra, S., Tizhoosh, H., Shah, S., Choi, C., Damaskinos, S., Safarpoor, A., Shafiei, S., Babaie, M., Diamandis, P., Campbell, C.J., et al.: Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence. arXiv preprint arXiv:1911.08736 (2019)
11. Kieffer, B., Babaie, M., Kalra, S., Tizhoosh, H.R.: Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. In: 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA). pp. 1–6. IEEE (2017)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Kumar, M.D., Babaie, M., Tizhoosh, H.R.: Deep barcodes for fast retrieval of histopathology scans. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (July 2018). https://doi.org/10.1109/IJCNN.2018.8489574
14. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
15. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
16. Naylor, P., Laclé, M., Reyal, F., Walter, T.: Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE transactions on medical imaging 38(2), 448–459 (2018)
17. Rastogi, V., Puri, N., Arora, S., Kaur, G., Yadav, L., Sharma, R.: Artefacts: a diagnostic dilemma—a review. Journal of clinical and diagnostic research: JCDR 7(10), 2408 (2013)
18. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
20. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946 (2019)
21. Taqi, S.A., Sami, S.A., Sami, L.B., Zaki, S.A.: A review of artifacts in histopathology. Journal of oral and maxillofacial pathology: JOMFP 22(2), 279 (2018)
22. Tomczak, K., Czerwińska, P., Wiznerowicz, M.: The cancer genome atlas (tcga): an immeasurable source of knowledge. Contemporary oncology 19(1A), A68 (2015)
23. Venkatesh, B., Shah, T., Chen, A., Ghaforian, S.: Restoration of marker occluded hematoxylin and eosin stained whole slide histology images using generative adversarial networks. arXiv preprint arXiv:1910.06428 (2019)
24. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)