

Segmentation of Breast Cancer Tissue Using Deep Learning

Oliver Johnson
Department of Electrical Engineering
Stanford University
johnsono@stanford.edu

Andrew Su *
Department of Medicine
Stanford University
suandrew@stanford.edu

Abstract

This project applied deep learning methods to the tissue segmentation of breast cancer tissue, as captured in whole slide images. Encoder-decoder models were used and they gave a successful segmentation prediction that quantitatively and qualitatively agreed well with the ground truths. The best performing model was a Feature Pyramid Network (FPN) with a ResNet encoder, which gave an IoU score of 0.864. This shows supports the proposition that deep learning can be an effective method to segment breast cancer tissue, and contribute to the calculation of a Tumor Infiltrating Lymphocytes score.

1. Introduction

Breast cancer is the second most common cancer among women in the United States. Correctly segmenting whole slide images of breast cancer tissue is important for many pathology workflows. Generating a segmentation is also an important prerequisite to automatically calculating the Tumor Infiltrating Lymphocytes (TILs) score. Tumor Infiltrating Lymphocytes (TILs) are immune cells that infiltrate into tumors and their proximity to tumor cells predict treatment response and survival in many cancers, including breast cancer. TILs have been proposed as biomarkers for routine assessment of tumors, however, it is not feasible for pathologists to manually make quantitative assessments. Tumor tissues can contain thousands to millions of cells and it is time-consuming and impracticable for pathologists to assess each one. Current methods classify tumor infiltration in broad categories and estimates but do not quantitatively measure TIL infiltration, nor do they capture the complexity of interactions between TILs and the tumor. In this project we will apply different methods of image segmentation to breast cancer tissue slide images, which can then be used downstream to help calculate the TILs score.

In this project we will be using Whole Slide Image

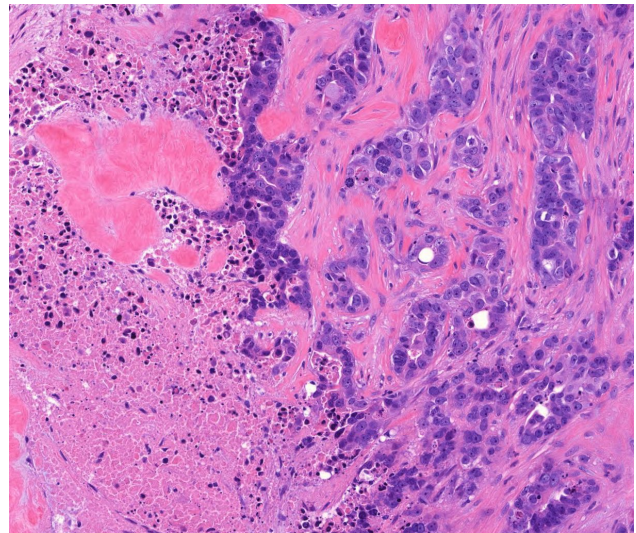


Figure 1. Example patch of a WSI from the TIGER [1] dataset.

(WSIs). Most WSIs are stained with haematoxylin and eosin (H&E) and captured using brightfield illumination, as it gives good visibility and detail for a broad range of features within an image [7], as shown in Figure 1. However, since WSIs can be resolutions of up to 100,000 x 100,000 pixels, feeding them into a neural network and establishing long range relationships across the entire image can be challenging. In addition, some artifacts can present and the morphological features in may not be consistent throughout the whole (very large) image. Artifacts include tissue tears, light bleed through uneven focusing and variability of the staining reactant. There is also an additional question about what magnification to do the analysis at (or whether to combine them different scales). Typically WSIs are broken into patches before being passed into a neural network [7].

2. Related Work

Machine learning techniques have been used for medical image analysis, and have been reasonably successful,

*External mentor; visiting scholar; not enrolled in CS231n

however they still do not enjoy widespread use clinically. Traditional machine learning techniques use hand-crafted features, whereas deep learning approaches use automatic feature extraction using large datasets. CNNs are the most common architecture for medical imaging related tasks.

Segmentation methods in medical imaging aim to partition an image into different regions based on tissue types or cell properties [5]. Classical image segmentation can be achieved using simple methods such as thresholding and edge detection, however these methods cannot account for the inhomogeneity of tissue images and more complex structures. Following the success of CNNs in computer vision tasks, U-Net [13] was developed to segment biomedical images, and was a very successful method. U-Net comprises of a U-shaped network structure with a contracting path and an expansive path. Inspired by U-Net, many U-shaped network structures have emerged, such as Res-UNet [6], U-Net++ [17] and H-Dense-UNet [10], which have all improved on the original U-Net in various ways.

More recently, there has been an effort to combine transformer and CNN methods to improve performance in image segmentation. For example, TransUNet [3] overcomes the weakness of the U-net in modelling long range dependency by using Transformers. Typically transformers exhibit poorer spatial localization, but better long range dependency. TransUNet upsamples the encoded features from a CNN feature map which allows for more precise localization from the CNN and long range dependency from the transformer. Other examples include TransFuse [16], which uses CNNs and Transformers in parallel and then employs a BiFusion module to effectively fuse outputs from each branch. Also more recently convolutional-free networks have been used with promising results [9]. While in this project we will focus on the CNN based methods, in a future work transformer based methods would possibly give improved performance, especially given the size of the image dimensions of breast cancer images.

For the application of these methods to WSIs, most split the WSI into patches and then feed that into the model. However some approaches have used patch extraction from different magnification levels [12], which could help overcome the lack of long range dependency information.

3. Data

We plan to use public training data released as part of the Tumor Infiltrating Lymphocytes in Breast Cancer (TIGER) Grand Challenge [1]. This includes images of Her2 positive and Triple Negative (TNBC) breast cancer images and TILs in these images have been annotated by pathologists. This data has a pixel size of approximately 0.5 micro-meters per pixel. More specifically, it contains three separate sections, labelled WSIROIS, WSIBULK and WSITILS. The WSIROIS section contains 195 images of breast cancer that

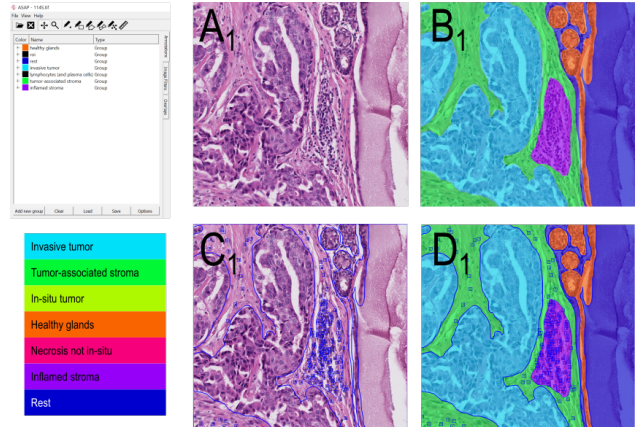


Figure 2. Example of ground truth segmentation map from sample 114S in the dataset [1]. The labels in the bottom left show the possible tissue types.

include manual annotations of 7 different labels of different tissue types. WSIBULK contains 93 images of breast cancer that include annotations of the tumor bulk. The WSI-TILS section contains 82 images where the visual assessment of the TIL has been completed, giving a TILs score for each image [1]. With only 370 WSIs, this dataset may seem to small to train a neural network at first glance, however given how detailed and large the WSIs are, they provide many examples of different cell types to effectively train a model. The entire dataset is 180GB.

For this project we will use the data from WSIBULK, as it contains data for the interesting cases of segmentation. The dataset is also small enough that a model can feasibly trained within the resources available for this course. WSI-BULK consists of whole slide images that include patches where the image is labelled into different tissue types. Most of the slide is unlabelled, there are only a few labelled patches per slide. Each labelled patch is a patch with a side dimension on the order of a few thousand pixels. To process this data, these were split into the 512 x 512 patches to feed into the neural network. Where the patches did not align with a dimension of the original image, it was zero-padded. After empty patches were removed this resulted in a dataset size of 2064, which was split a training set of 1651 images (80%), a validation set of 206 images (10%), and a test set of 207 images (10%). This validation set was used to cross validate when training the model.

The training dataset was also augmented to include random 90 degree rotations and random horizontal/vertical flips. It was found including this improved validation performance on the order of a few percent.

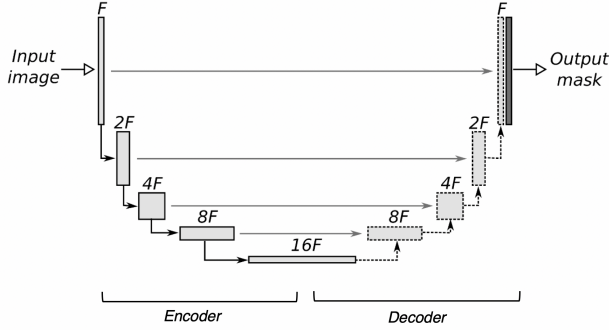


Figure 3. General U-Net Architecture showing encoder and decoder. Diagram adapted from [4].

4. Methods

4.1. Semantic Segmentation

For the segmentation problem, after splitting the images into patches as described above, they were fed in to a variety of encoder-decoder networks to assess performance. This network structure was chosen because it has been used successfully on many image segmentation problems. We used the PyTorch library `segmentation_models` (source code available on Github) [15] to define our network structure and train on the dataset. In this library the example iPython scripts in the Github repository were used as a base for defining appropriate classes and to train our models. While in the WSIROIS dataset there were seven classes for different tissue types, we split this into a binary task as this was one of the tasks in the TIGER challenge. This was to categorize inflamed/tumor-associated stroma (mask value 6 and 2) against all other classes (mask value 1, 3, 4, 5, 7). The motivation behind this that recognizing invasive tumor and stroma are important for calculating the TILs score.

4.2. Network Architecture

We compared several different segmentation models. As a baseline we used the highly successful U-Net [13] model a VGG16 [14] encoder with a encoder that was pretrained on Imagenet. We compared this with the same network, but with a more modern ResNet34 [8] encoder, also trained on U-Net. The U-Net architecture consists of an encoder and decoder stage as shown in Figure 3.

The second architecture that was considered was the the Feature Pyramid Network (FPN) [11]. Normally feature pyramids have been avoided in deep learning for computer vision due to their computation cost, however the architecture developed by the authors only added marginal additional cost despite adding a pyramid structure [11]. It is expected that learning at different scales such as in a pyramid structure might help with uncovering some of the more long

range dependencies within a medical image. The FPN applied to a Faster R-CNN and predictions at different levels are combined to output the segmentation map [2].

In both cases the dice loss was used to determine the loss for the network.

4.3. Evaluation

We evaluated the methods using the intersection over union (IoU) score as this is a natural metric for binary segmentation. We also evaluated the methods qualitatively so see what kinds of images are difficult for the models to predict labels.

5. Experiments and Discussion

The various network architectures that were discussed in the previous section were trained on the dataset. All models were trained on an AWS Ubuntu 18.04 instance using PyTorch with a NVIDIA Tesla T4 GPU. The Adam optimizer was used to train all the models, and learning rates were experimented with to find a learning rate that worked well. Each model was trained for 25 epochs, and for all the the models, the learning rate was set to 1×10^{-4} for the first 15 epochs, then reduced to 1×10^{-5} for the last 10 epochs. In all cases a batch size of 16 was used, as this was the highest available batch size that fit onto the memory of the GPU. The dice loss and IoU score was captured as the model trained. The highest scoring model on the validation set was used as the final model to output predictions.

The quantitative results show that ResNet with FPN was the best performing model, with the highest IoU score. However the analysis done here was not exhaustive so it possible with more finetuning that the U-Net network could perform equivalently.

5.1. Qualitative Analysis

The results showed reasonable success in segmenting the images. As shown in Figure 4, the images are shown to be reasonably close to the ground truths, although there is some differences in between each model. The wide spectrum of tissue types that the model needs to account for is also demonstrated. In particular, it seems that the VGG16 + U-Net images have have more precise lower level features, whereas Resnet seems to have smoother edges. This is particularly apparent in the first row of images in Figure 4. It is also noted that sometimes the predictions from the networks give more precise boundaries between regions than the human annotations. This may be because the pathologists may not spend excess time zooming into the image to find the precise boundary between regions. It is also interesting to note that in the second row of images, some of the models do not always recognize some of the smaller fat cells.

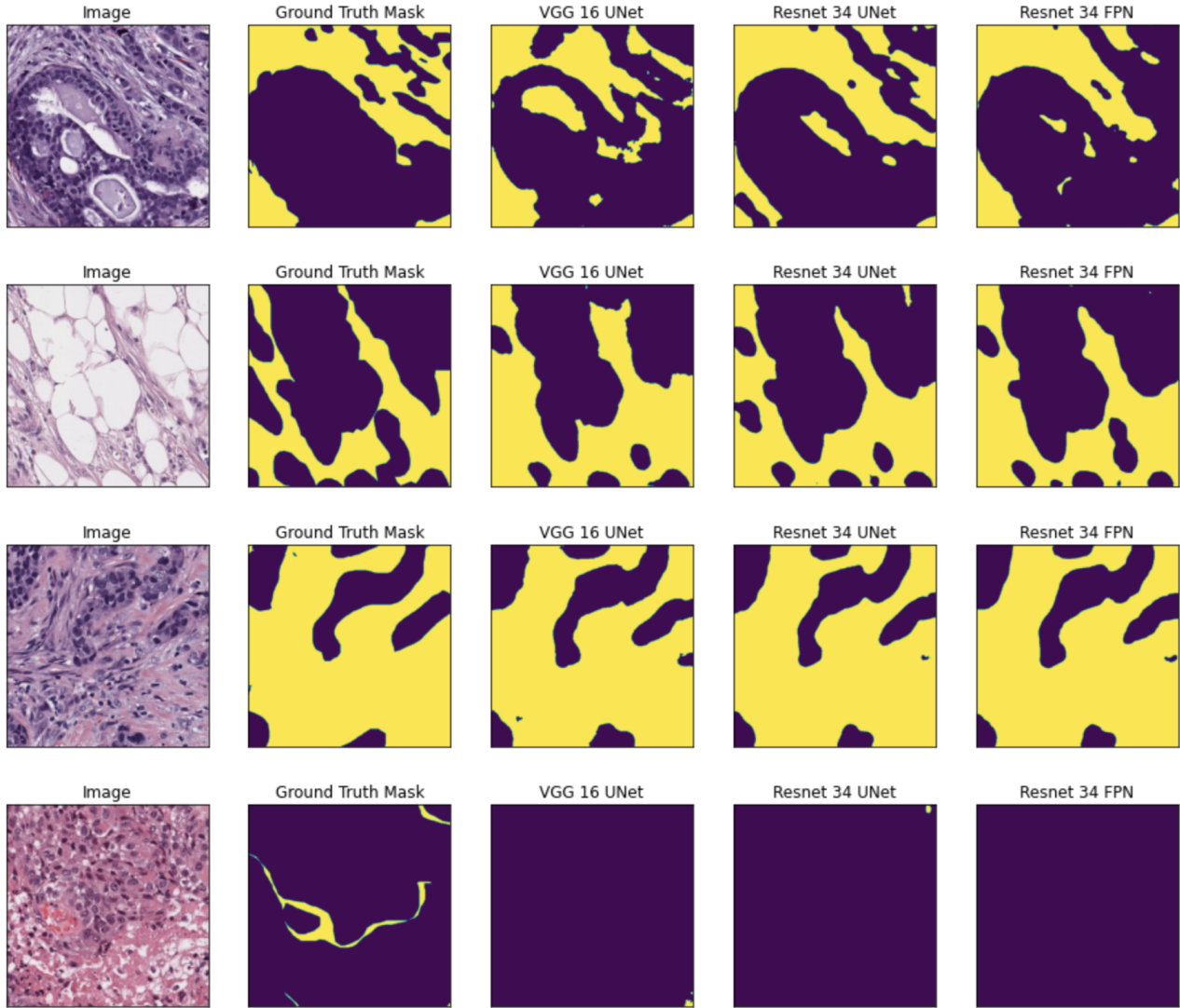


Figure 4. Example Segmentation outputs.

As shown by the last row where the ground has little slithers of positive class, none of the networks are able to correctly predict this region. It is possible that there are few examples of this exact situation in the dataset so that the network is not able to easily classify it. It is also possible that some pathologists are more precise than others and may not deem a small region like the above significant enough to label.

6. Conclusion

This project demonstrated neural network segmentation methods could successfully segment breast cancer tissue, with the best performing method (Resnet34 encoder, FPN) obtaining a IoU score of 0.864. In this project only su-

Architecture	Encoder	IoU Score (Test set)
U-Net	VGG16	0.830
U-Net	Resnet34	0.862
FPN	Resnet34	0.864

Table 1. Quantitative results

pervised learning was used but for in the future semi-supervised models could provide a large boost in performance. This is because most of the dataset is unlabelled, and these additional examples could allow the the model to generalize to more situations. In addition, more modern encoder-decoder networks that have been successful on other medical imaging datasets could be used for this prob-

lem, such as methods that utilize transformers. Also with more time a more detailed comparison of the models could also be performed where more hyperparameters were varied.

7. Contributions and Acknowledgements

Oliver wrote the manuscript, wrote the code, trained models and collated results. Andrew (external mentor, not enrolled in CS231n) formulated the original idea, gave advice and contributed to some code debugging.

References

- [1] Tiger challenge, <https://tiger.grand-challenge.org/home/>, online resource. 1, 2
- [2] Ross Girshick Alexander Kirillov, Kaiming He and Piotr Dollar. A unified architecture for instance and semantic segmentation. <http://presentations.cocodataset.org/COCOL7-Stuff-FAIR.pdf>. Accessed: 2022-06-02. 3
- [3] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation, 2021. 2
- [4] Nikolaos Kosmas Chlis. Image segmentation with monte carlo dropout unet and keras. https://nchlis.github.io/2019_10_30/page.html. Accessed: 2022-06-02. 3
- [5] Chiranji Lal Chowdhary and D.P. Acharjya. Segmentation and feature extraction in medical imaging: A systematic review. *Procedia Computer Science*, 167:26–36, 2020. International Conference on Computational Intelligence and Data Science. 2
- [6] Foivos I. Diakogiannis, François Waldner, Peter Caccetta, and Chen Wu. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:94–114, apr 2020. 2
- [7] Neofytos Dimitriou, Ognjen Arandjelović, and Peter D. Caie. Deep learning for whole slide image analysis: An overview. *Frontiers in Medicine*, 6, 2019. 1
- [8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 3
- [9] Davood Karimi, Serge Didenko Vasylechko, and Ali Gholipour. Convolution-free medical image segmentation using transformers. In Marleen de Bruijne, Philippe C. Cattin, Stéphane Cotin, Nicolas Padoy, Stefanie Speidel, Yefeng Zheng, and Caroline Essert, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 78–88, Cham, 2021. Springer International Publishing. 2
- [10] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: Hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE Transactions on Medical Imaging*, 37(12):2663–2674, 2018. 2
- [11] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection, 2016. 3
- [12] Yun Liu, Krishna Gadepalli, Mohammad Norouzi, George E Dahl, Timo Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Philip Q Nelson, Greg S Corrado, et al. Detecting cancer metastases on gigapixel pathology images. *arXiv preprint arXiv:1703.02442*, 2017. 2
- [13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. 2, 3
- [14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014. 3
- [15] Pavel Yakubovskiy. Segmentation models pytorch. https://github.com/qubvel/segmentation_models.pytorch, 2020. 3
- [16] Yundong Zhang, Huiye Liu, and Qiang Hu. Transfuse: Fusing transformers and cnns for medical image segmentation, 2021. 2
- [17] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation, 2018. 2