Department of Computer Science
Technical University of Cluj-Napoca

Breast Cancer Segmentation

Project prepared for the Pattern Recognition Systems course

Developed By: Borbála Fazakas
GitHub
Group: 30441

Laboratory Assistant: Florin Oniga

Contents

Chapter 1

Subject Specification

1.1 Objective

This project has two goals:

1. develop a model that can categorize the cells in breast cancer slides (images), based on the type of the tissue to which they belong. This goal can be phrased as a multi-class semantic segmentation problem.
2. develop a model that can categorize the cells in breast cancer slides (images) as cancerous/non-cancerous. This goal can be phrased as a binary semantic segmentation problem.

1.2 Motivation

Breast cancer is the most common type of Cancer in the UK [3] and is the 2nd most common cancer diagnosed in women in the USA [4]). In the UK, 1 in 8 women are diagnosed with breast cancer during their lifetime [3].

Detecting breast cancer at an early phase can significantly improve the chances of survival, and furthermore, the segmentation of cancerous cells can help physicians quantify the volume of tissue in the breast for treatment planning [5].

Chapter 2

Project Setup

2.1 The Dataset

The dataset contains 151 histology images with segmentation masks created through a crowd-sourcing method, by medical students and experienced pathologists [6]. The images are of different sizes and in each of them, a region of interest (ROI) was defined. The pixels outside the ROI were ignored during the manual annotation for the ground truth masks and will be ignored in the model development as well. The dataset can be downloaded from this GitHub repository.

The cells on the images belong to the classes referenced in Table 2.1 and are marked on the segmentation masks through the colors shown on the color map 2.2.

For the model development and evaluation, the dataset was split into two parts:

- 80% of the data (121 images) were used for training. These are referenced as the "Training Dataset".
- 20% of the data (30 images) were used for evaluation. These are references as the "Test Dataset".

Note that due to the limited amount of available data no validation dataset was defined. Thus, the evaluation is not entirely accurate, the final results may be prone to overfitting (to the test dataset). However, this decision allows us to use more data for training.

The distribution of the pixels among the different tissues is shown on the plot 2.3 for all classes, and on the plot 2.4 for the tumors vs all other classes. Note that all images in the dataset contain tumors, and overall, more than 41% of all pixels belong to tumors.

2.2 The Software and Hardware used for training

The experiments for developing the models were ran on Google Colab's free version, on a Google Compute Engine backend, with a GPU accelerator. Thus, the exact memory, CPU and GPU

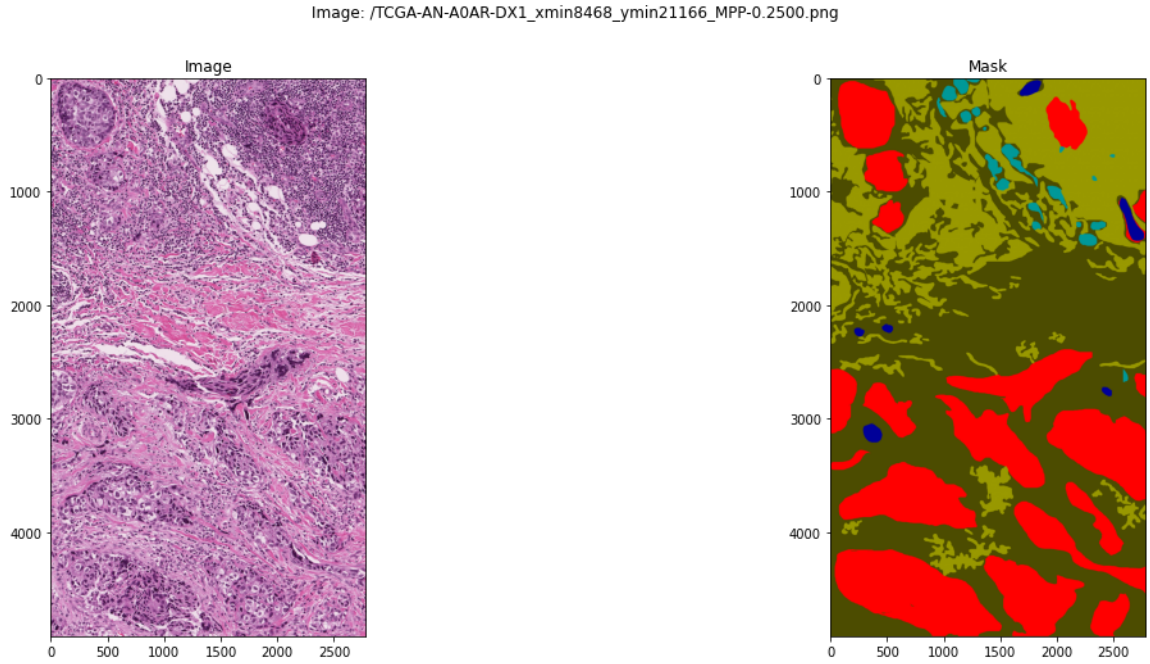

Figure 2.1: Sample Image and the True Mask

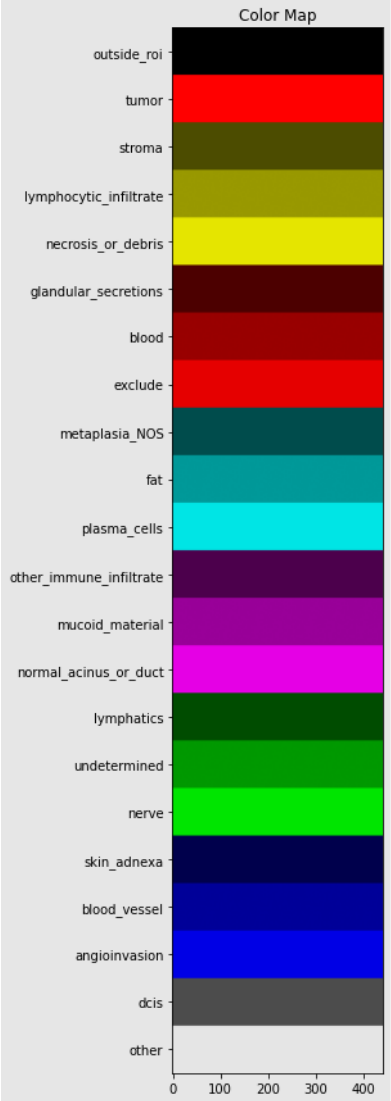

Figure 2.2: Color Map

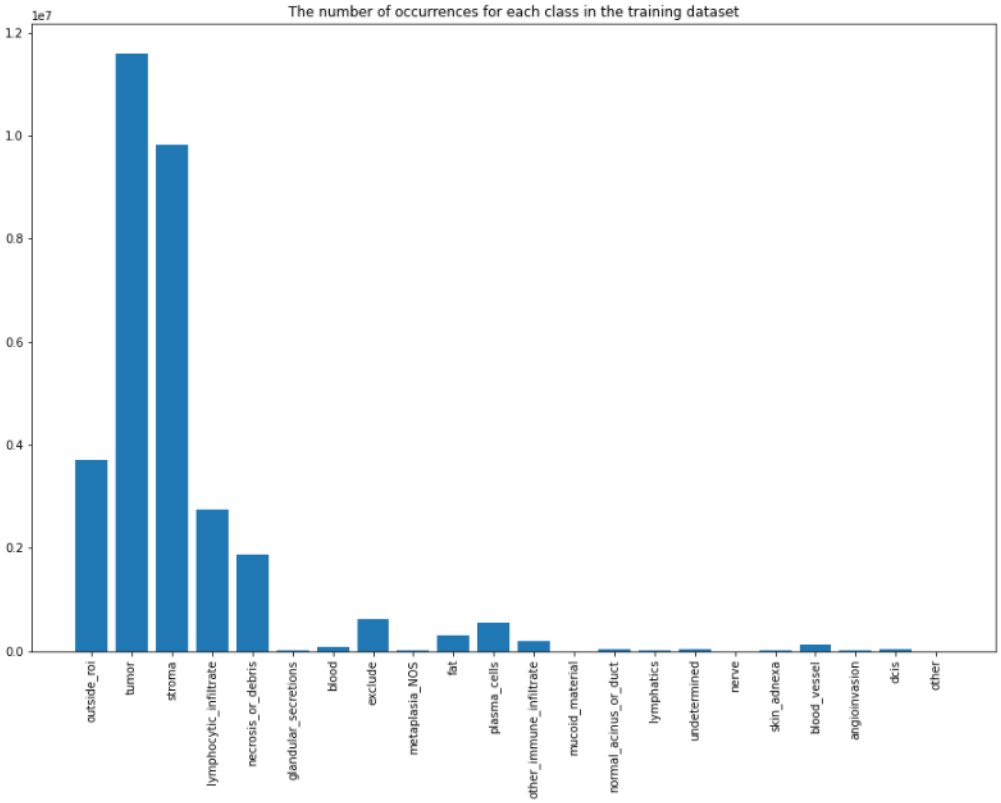

Figure 2.3: The Distribution of the pixels among the tissue types

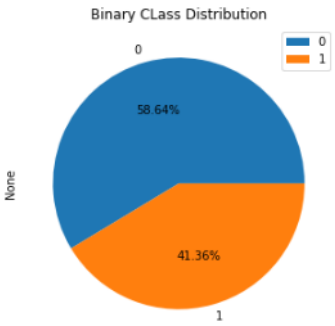

Figure 2.4: The Distribution of cancerous (1) vs non-cancerous pixels (0)

Table 2.1: Types of Tissues in the Dataset

Tissue Label	Class ID
outside ROI	0
tumor	1
stroma	2
lymphocytic infiltrate	3
necrosis or debris	4
glandularsecretions	5
blood	6
exclude	7
metaplasia NOS	8
fat	9
plasma cells	10
other immune infiltrate	11
mucoid material	12
normal acinus or duct	13
lymphatics	14
undetermined	15
nerve	16
skin adnexa	17
lood vessel	18
angioinvasion	19
dcis	20
other	21

usage is unknown.

As for the software part,

- Tensorflow was used for the model development and training
- OpenCV for the postprocessing of the predicted masks
- NumPy and pandas for pre- and post-processing
- Matplotlib and Seaborn for creating the plots
- Scikit-Learn for the evaluation
- Huggingface for transfer learning (see the details about fine-tuning a SegFormer).

Chapter 3

Model Development

For both image segmentation tasks a deep-learning based approach was planned. Initially, two neural network architectures were tried, the U-Net [1] and the SegFormer [2].

3.1 Background

3.1.1 U-Net [1]

The U-Net is one of the most widely adopted architecture for biomedical image segmentation. It is based on an encoder, a contracting path that captures the context, and a decoder, an expanding path symmetric to the contracting one, that enables precize localization. The encoder consists of repeated convolutional layers with ReLU activation, and max pooling after every 2nd convolutional layers. The decoder is built from inverse convolutional layers for upsampling, and normal convolutional layers with ReLU and skip connections from the encoder convolutions for localization. See Figure 3.1 for more details. In 2015, when the architecture was published, it achieved state-of-the-art performance on several biomedical image segmentation benchmarks, such as the ISBI cell tracking challenge, or the Warping Error of the EM segmentation challenge.

3.1.2 SegFormer [2]

Compared to the U-Net, teh SegFormer is a much more modern, attention-based network architecture, that relies on a positional-encoding-free, hierarchical Transformer-encoder and a lightweight All-MLP decoder. Again, at the moment of publication it achieved SOTA results on common benchmarks, such as ADE20K, COCO-Stuff and Cityscapes.

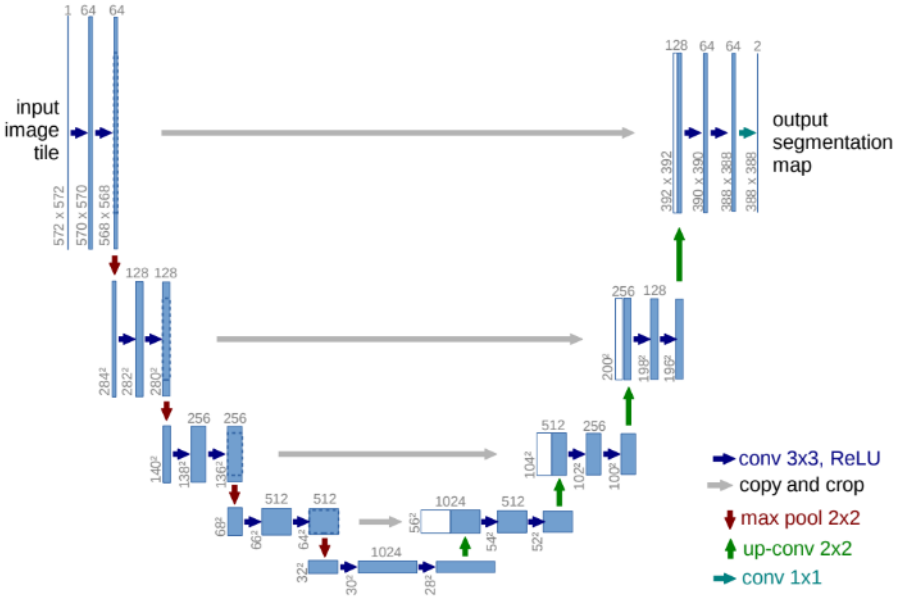

Figure 3.1: The U-Net Architecture

3.2 Model Input. Dataset Preprocessing

3.2.1 Image Sizes

Both of these network architectures can only deal with fixed-sized images. This represents a major drawback of these architectures for our particular breast cancer segmentation problem, given that our dataset contains images of various sizes and shapes.

As a work-around, all images were resized to 512x512 pixels. However, this operation resulted in a distortion of the images, that may have an impact on the final performance of the model.

3.2.2 Class Weights

As previously stated, the pixels outside the ROI should be ignored both during the training and the evaluation. Thus, as a further preprocessing step, sample weights were added to all the pixels of all the images in the training dataset:

- weight 0 for all pixels outside the ROI
- weight 1 for all other pixels

Thus, the model's loss function never considers any pixels outside the ROI. Still, if the model is unsure about a pixel's class, it's technically possible that it would classify it as "ROI".

3.3 Architecture Implementation

3.3.1 U-Net: Trained from Scratch

Because the U-Net is a simpler architecture, it was developed from scratch from basic Tensorflow layers, such as Conv2D, MaxPool2D or Conv2DTranspose.

3.3.2 SegFormer: Transfer Learning

The SegFormer has a far more complex architecture that would make an implementation from scratch hard to obtain. Thus, I opted for a transfer-learning-based approach instead: I took a checkpoint of SegFormer from **Huggingface** (TFSegformerForSemanticSegmentation's checkpoint "nvidia/mit-b0"), which was already trained on another dataset, and fine-tuned the model to allow it to learn the specific of the breast cancer dataset applied in our project.

3.4 Solving the Multiclass Segmentation Problem

3.4.1 Initial Experiments

As an initial step, one experiment with both the U-Net and the SegFormer was performed, without any further tuning. The findings were the following:

- U-Net: converged in 30 epochs, a test accuracy of 56.29% is achieved.
- SegFormer: converged in 10 epochs, a test accuracy of 20.84% is reached

Based on these initial results, and given the limited time-frame and computational resources, further experiments were only carried out on the U-Net.

3.4.2 Further Experiments to improve the U-Net-based Model

3.4.2.1 Balanced Class Weighting

Based on the observation that the initial U-Net based model classified all pixels as "tumor", "stroma" or "lymphocytic_infiltrate", seems like the heavily unbalanced representation of the classes in the training dataset causes an issue. Thus, a new model was trained in which the pixels were given balanced weight, based on how over-/under-represented their class is.

3.4.2.2 Data Augmentation

Given that there are relatively few images available for training, the model was prone to overfitting. To overcome the issue of the model learning the particularities of the training dataset, the dataset was extended with images generated from the original ones via simple image transformation operations:

- random brightness adjustments
- random contrast adjustments
- random rotations (max angle: 15 degrees)
- random horizontal and vertical flips

In the augmented dataset, each original image appeared 5 times, with random augmentation operations applied on each instance. The final dataset had 605 images. A U-Net was trained on the augmented dataset.

3.4.2.3 Regularization

As another approach to overcome overfitting, regularization was added to the U-Net architecture:

- dropout, with a factor of 0.003, after every double convolutional block
- L2-regularization on every convolutional layer, with a factor of 0.03

3.4.2.4 Data Augmentation with Regularization combined

As a last experiment, the previously presented regularized model was trained on the augmented dataset.

3.4.3 Results

The overall accuracies achieved via each model after the model converged are summarized in table 3.1.

The confusion matrix for the best model (U-Net trained on the Augmented Dataset) is visible on Figure 3.2. Note that the model achieves a 86.1% recall and an 50.1% precision in recognising tumors, but is very low-performing when it comes to recognising classes other than "tumor" and "stroma".

Table 3.1: U-Net Results

Model	Test Accuracy
Simple U-Net	56.29%
U-Net trained with balanced Sample Weights	1.34%
U-Net trained on the Augmented Dataset	**57.76%**
Regularized U-Net	56.33%
Regularized U-Net trained on the Augmented Dataset	51.86%

3.4.3.1 Observations

- Applying balanced class weights destroyed the model's capacity to learn anything. Possibly because for the underrepresented classes have simply to few pixels for the model to learn aything about them, but with the weighting, the overrepresented classes were not relevant enough for the model to learn about them.
- Based on the predictions made on the test dataset, the U-Net trained on the augmented dataset was the only one capable of detecting multiple types of tissues, not just "tumor", "stroma" or "lymphocytic_infiltrate". For example, it successfully recognizes fat cells and some plasma cells. Overall, the Intersection-over-Union metric still shows very low (<10%) values for all models.
- While both data augmentation and regularisation improved the initial model's accuracy, the two methods combined didn't achieve a good result. The explanation for this is unknown at the moment.

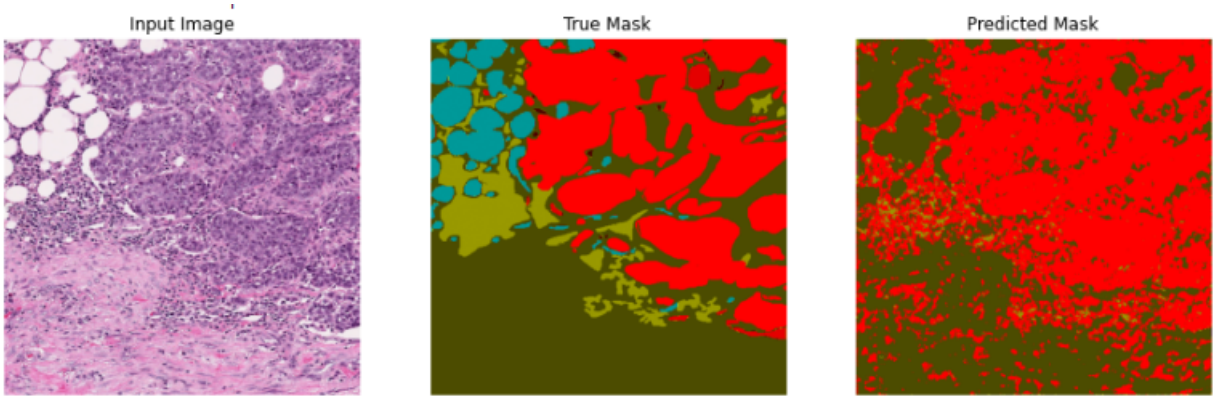

Figure 3.3: Sample Prediction by the Simple U-Net

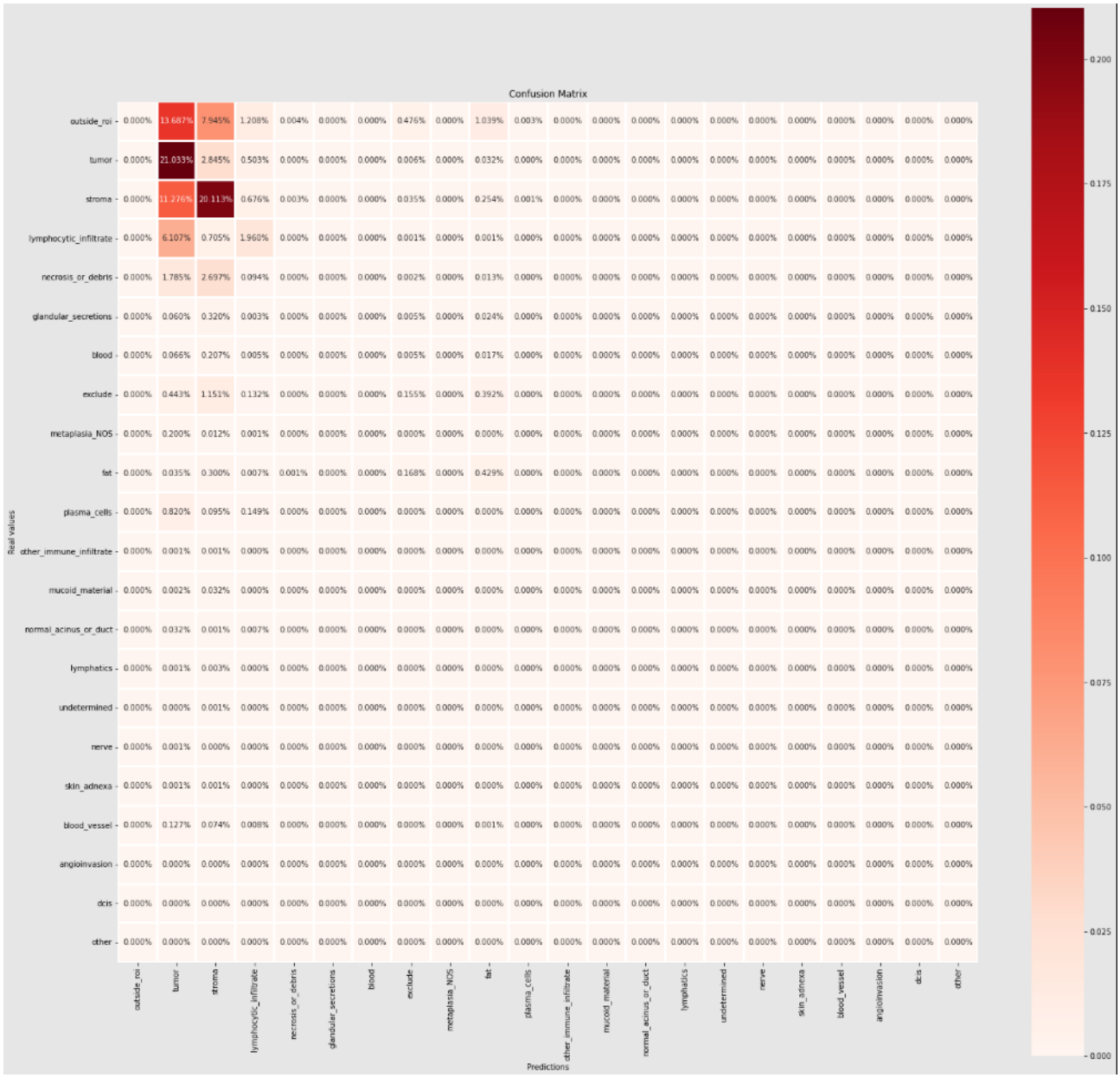

Figure 3.2: Confusion Matrix for the best model

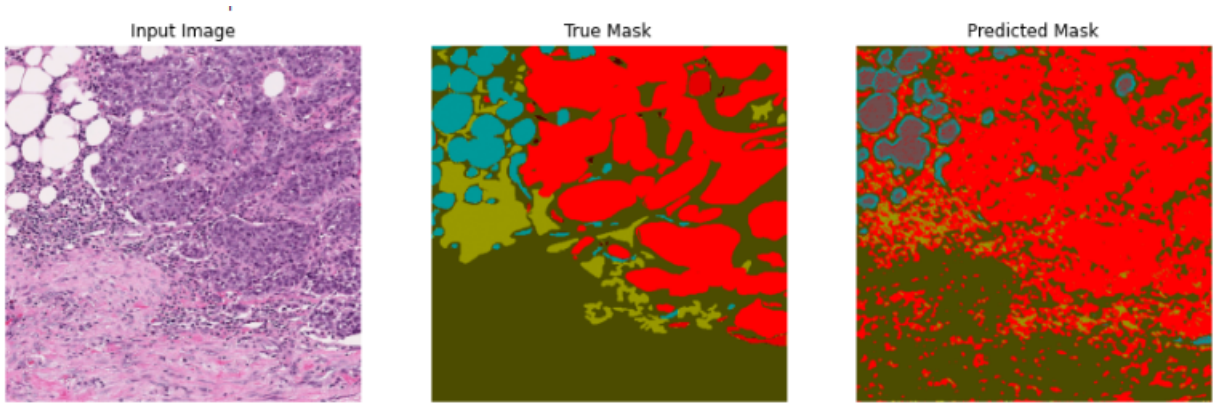

Figure 3.4: Sample Prediction by the U-Net trained on the Augmented Dataset

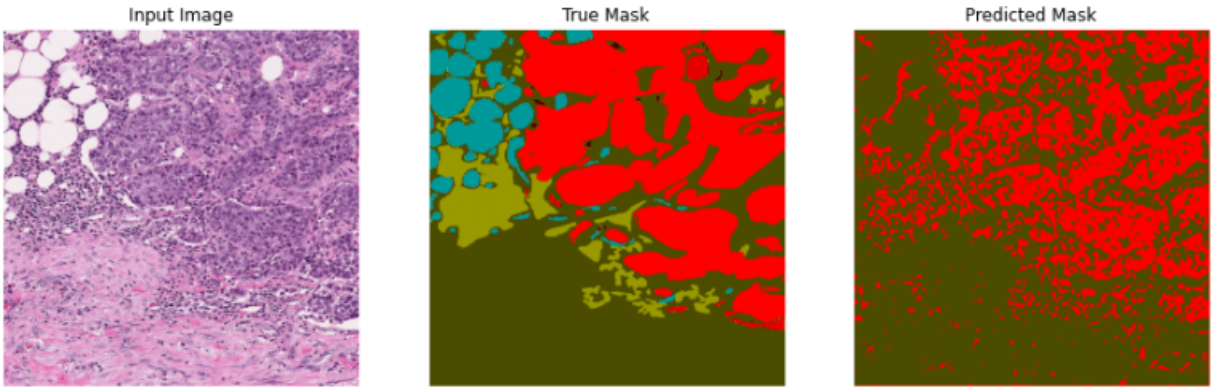

Figure 3.5: Sample Prediction by the Regularized U-Net

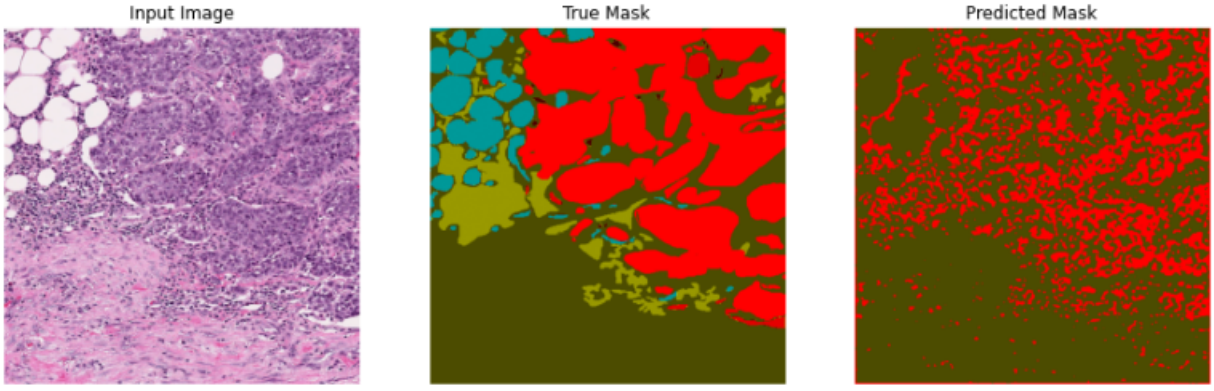

Figure 3.6: Sample Prediction by the Regularized U-Net trained on the Augmented Dataset

3.5 Solving the Binary Segmentation Problem

3.5.1 Pre-processing

As a preprocessing step, the class labels in the masks were mapps as follows:

- class 0 (ROI) →class 2. It was given weight 0 in the training.
- class 1 (tumor) →class 1 (positive)
- all other classes →class 0 (negative)

3.5.2 Model Training. Initial results

A simple U-Net was trained until convergence, achieving the results summarized in 3.2. The confusion matrix can be seen on Figure 3.7.

Table 3.2: U-Net Results for the Binary Segmentation

Metrics	Result
Accuracy	77.05%
F1-score	62.8%
Precision	65.8%
Recall	60.1%

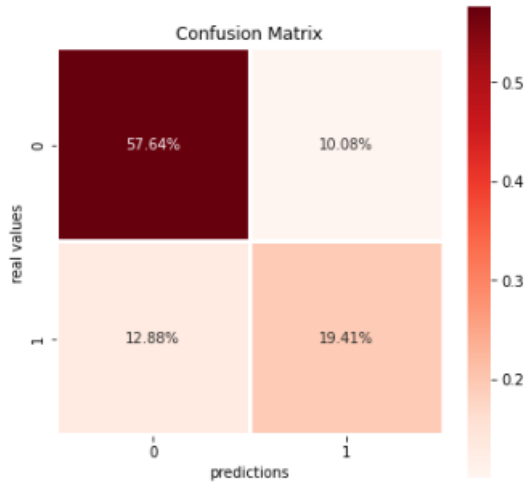

Figure 3.7: Confusion Matrix for the Simple U-Net on the Binary Segmentation Problem

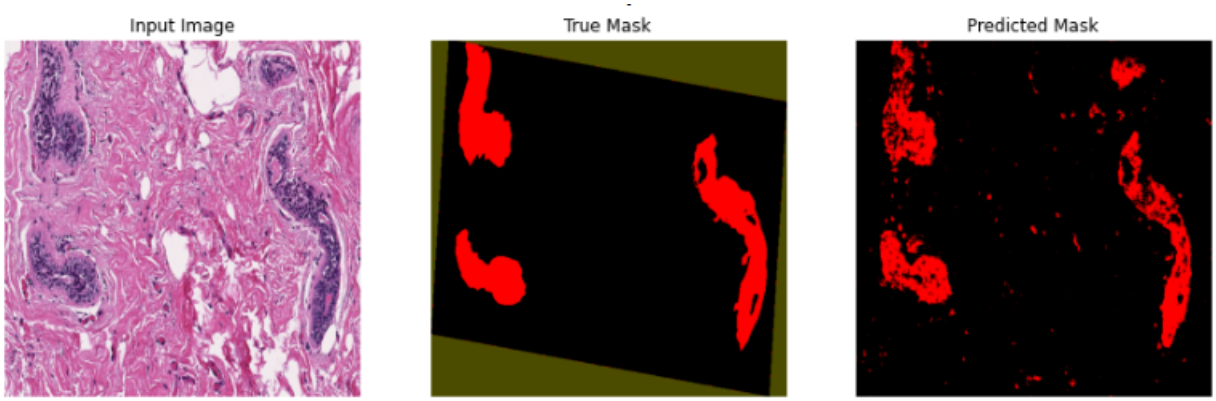

Figure 3.8: Sample Prediction by the Binary U-Net. (Green Class is outside ROI, Red is Tumor)

Figure 3.9: Sample Prediction by the Binary U-Net. (Green Class is outside ROI, Red is Tumor)

Figure 3.10: Sample Prediction by the Binary U-Net, (Red is Tumor)

Figure 3.11: Sample (Bad) Prediction by the Binary U-Net, (Red is Tumor)

3.5.3 Post-Processing via Morphological Operations. Improved Results

From the predictions made on the test dataset, it is clear that the prediction's are noisy: there are small "islands" of tumor regions predicted in the non-tumor regions, and vice-versa. To solve this issue, simple morphological operations were applied on the predictions of the model, and then the metrics were recomputed. As post-processing operations, an alternating sequence of closing and opening operations were applied with an increasing kernel size (the kernel is a matrix of 1s), in order to filter out the noise. The following sequences of operations were tried (c = closing, o = opening, the numbers in the brackets represent the kernel's size):

- S1: closing (5) + opening (5)
- S2: closing (5) + opening (9)
- S3: closing (5) + opening (5) + closing (11) + opening (11)
- S4: closing (5) + opening (9) + closing (11) + opening (15)
- S5: closing (5) + opening (5) + closing (11) + opening (11) + closing (17) + opening (17)

The results are summarized in table 3.3. It can be seen that the morphological operations can bring some improvements on the accuracy, but there's a trade-of between the improvements in the precision and the recall. However, the results filtered by these morphological operations have an other significant advantage as well: they can be more easily understood by humans, ad the regions for tumors and other classes are clearly separated, not at a pixel-level anymore.

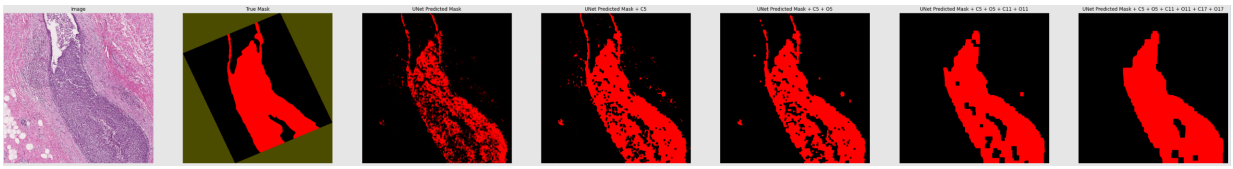

Figure 3.12: Sample Post-processing

Table 3.3: U-Net Results after the Morphological Post-Processing, for the Binary Segmentation

Post-processing sequence on the U-Net Prediction	Accuracy	F1	Precision	Recall
None	77.05%	62.8%	**65.8%**	60.1%
S1	76.9%	67.2%	61.9%	73.5%
S2	**78.2%**	66.9%	65.7%	68.3%
S3	76.3%	**68.0%**	60.2%	78.0%
S4	78.1%	66.8%	65.3%	68.4%
S5	75.3%	67.2%	58.9%	**78.3%**

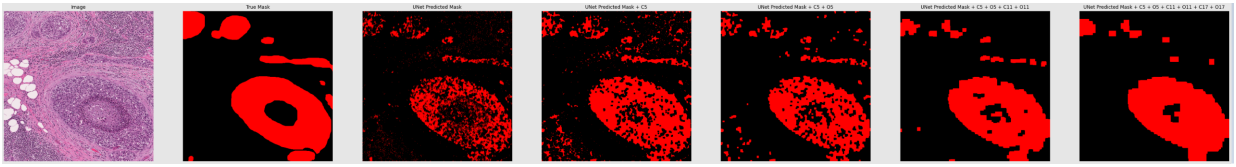

Figure 3.13: Sample Post-processing

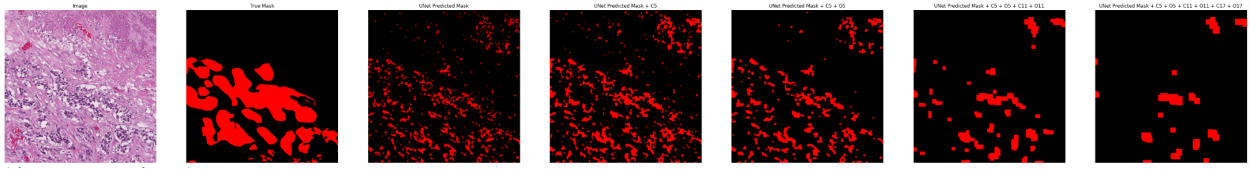

Figure 3.14: Sample Post-processing

Chapter 4

Conclusions

For the multi-class segmentation problem, the U-Net showed more promising results than the SegFormer, and the initial accuracy of 56.29% could be improved by adding data augmentation and regularization to the model, leading to an over 57.7% accuracy. This is a significant achievement for a 21-class segmentation problem. Still, most classes are never "predicted" by the model. Data augmentation brings a significant improvments towards detecting the under-represented classes, but the results are still not satisfactory.

For the binary segmentation problem, the U-Net shows much better results than for the multi-class one, and it achieves a 77.05% accuracy without any tuning. A post-processing on the predicted masks, consisting of alternating opening and closing operations to filter out the noise in the prediction, can further improve the accuracy, to 78.2%, without any relevant reductions in the precision or the recall.

Overall, the results achieved are promising and they demonstrate how different tuning and post-processing operations can bring improvements to a model. However, the results achieved are stil far from the SOTA.

Chapter 5

Further Work

The following experiments seem promising for improving the current results:

- Trying Fully Convolutional networks, which can handle images of different sizes. With those, the distortions in the pre-processing phase could be avoided.
- Applying higher-weights for underrepresented classes, but not fully balanced weights.
- Trying other attention-based models.
- ...

Bibliography

[1] P. F. Olaf Ronneberger and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, 2015.

[2] Z. Y. A. A. . J. M. A. P. L. Enze Xie1, Wenhai Wang2, *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*, 2021.

[3] "Nhs - breast cancer in women." [Online]. Available: https://www.nhs.uk/conditions/breast-cancer/

[4] "Mayo clinic - breast cancer." [Online]. Available: https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470

[5] H. L. . F. K. . J. L. Epimack Michael 1, He Ma 1, "Breast cancer segmentation methods: Current status and future potentials," *BioMed research international*.

[6] H. H. L. A. A. M. A. T. E. L. S. A. E. R. A. S. H. S. E. S. A. F. I. A. M. S. J. A. M. A. T. E. M. R. I. A. R. N. M. E. Y. A. M. H. O. A. M. A. M. M. K. A.-A. F. Y. A. A. D. M. Y. A. M. G. A. M. E. S. Y. F. B. M. Z. J. B. D. R. C. D. M. D. A. G. Mohamed Amgad 1, Habiba Elfandy2 and L. A. D. Cooper, "Structured crowdsourcing enables convolutional segmentation of histology images," *Bioinformatics*.