DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Biomedical Computing

Deep Learning Based Analysis of Tumor-infiltrating Lymphocytes in H&E Stained Histological Sections for Survival Prediction of Breast Cancer patients

Margaryta Olenchuk

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Biomedical Computing

Deep Learning Based Analysis of Tumor-infiltrating Lymphocytes in H&E Stained Histological Sections for Survival Prediction of Breast Cancer patients

Deep Learning basierte Analyse von tumorinfiltrierenden Lymphozyten in H&E gefärbten histologischen Schnitten zur Überlebensvorhersage von Brustkrebspatienten

Author:	Margaryta Olenchuk
Supervisor:	Prof. Dr. Peter Schüffler
Advisor:	Dr. Philipp Wortmann, Ansh Kapil
Submission Date:	15.12.2022

Contents

1 Introduction **1**

2 Related work **3**
2.1 Deep learning-based semantic segmentation 3
2.1.1 Fully convolutional networks (FCNs) 3
2.1.2 Encoder-decoder networks 3
2.1.3 Generative adversarial networks (GANs) 5
2.1.4 Recurrent neural networks (RNNs) 5
2.1.5 Transformers 6
2.2 Survival analysis 6
2.2.1 TILs as prognostic biomarker 7

3 Methods **9**
3.1 Semantic segmentation 9
3.1.1 DeepLab 9
3.1.2 Transformers 11
3.2 Survival Analysis 14
3.2.1 Kaplan–Meier estimator 15
3.2.2 Cox model 15

4 Data **18**
4.1 Segmentation 18
4.2 Survival Analysis 19

List of Figures **21**

List of Tables **22**

Bibliography **23**

1 Introduction

Breast cancer is the most common form of cancer diagnosed worldwide and the leading cause of cancer-related death among women. [1] It is a heterogeneous disease, consisting of several morphological and molecular subtypes. The molecular subtypes are among the most important factors to characterize breast cancer. Four main groups are defined based on the status of several receptors used in clinical practice [2], namely the Hormonal Receptor (HR, which is positive if either Estrogen Receptor (ER) or Progesterone Receptor (PR) are positive) and of their human epidermal growth factor receptor 2 (Her2):

1. Luminal A (HR positive, Her2 negative)
2. Luminal B (HR positive, Her2 positive)
3. Her2 enriched (HR negative, Her2 positive)
4. Triple Negative (HR negative, Her2 negative)

Regardless of the subtype, for diagnostic confirmation of breast cancer a patient's tissue sample is sectioned onto microscope slides for staining, often with hematoxylin and eosin (H&E), followed by a visual diagnosis by a pathologist. Pathologists examine a tissue specimen for abnormalities that indicate breast cancer. Since cancer causes changes in tissue at the sub-cellular scale, an analysis of normal and tumor tissue can provide novel insights into tissue characteristics and can lead to a better understanding of mechanisms underlying cancer onset, progression and provide valuable information for medical decision making such as treatment choices. [3] While the manual examination continues to be widely applied in clinical settings, it is a subjective and is not scalable to translational and clinical research studies involving a high number of high-resolution whole slide tissue images (WSIs). Hence, there is an increased need for reliable and efficient automated methods to complement the traditional manual examination of tissue samples. Due to advancing technology and access to a large amount of data, deep learning methods have garnered a lot of interest in computer vision for the digital pathology domain. The two most common computerized tasks in WSI analysis are the segmentation of microscopic structures, structures like tumor regions and smaller nuclei and cells, and the classification of image regions or even whole images. And while there are a great number of deep learning based analysis pipelines in the digital pathology domain, automated algorithms also contribute to the development and testing of reliable prognostic and predictive biomarkers to help clinicians choose the best treatment for each patient.

The differentiation between breast cancer types and subtypes is essential. This thesis is focused on Her2 positive and Triple Negative breast cancer (TNBC) subtypes since they have the worst prognosis, and therefore subject to a large corpus of research in prognostic and predictive biomarkers, aiming at improving patient management and prognosis. TNBC

is an aggressive type of cancer and due to the lack of all three receptors, it has a limited response to hormonal and immune treatments. Cancers of this type are known to have high recurrence rates and poor prognoses compared to non-TN breast cancers. [4] There are multiple reported biomarkers for TNBC, such as epidermal growth factor receptor, vascular endothelial growth factor, C-kit, and basal cytokeratins. [5] Moreover, a significant percentage of TNBCs are known to carry BRCA1 mutations, in which tumor cells are defective in homologous recombination DNA repair mechanisms. But this thesis focuses on the tumor-infiltrating lymphocytes (TILs), which could be examined from WSIs therefore directly accessible without the need for any further testing or additional data source.

TILs are mononuclear immune cells that infiltrate tumor tissue and have been detected in almost all solid tumors, including breast cancer. [6] The development and progression of malignant tumors can be characterized by an interaction of the cells in the tumor microenvironment and TILs. In the early stage HER2-positive and TNBC, immune infiltrates are detectable in up to 75% of tumors. [7] Accumulating evidence indicates that tumor-infiltrating lymphocytes are clinically useful biomarkers in TNBC and HER2-positive and that they play an essential role in cancer progression. [8] The patients with a high proportion of TILs in the tumor tissue and high immunogenicity of the tumor were shown to respond better to the chemotherapy. Further research and development of additional TILs related biomarkers would not only grand clinicians important prognostic information but also promote the research focus of novel treatments and therapeutics. Analysis of TILs with exhausted phenotype is associated with loss of antitumor immunity. Single-cell RNA-seq of TILs has been already performed to search for new immune checkpoint blockade targets that enable the precise definition and even novel development of therapeutic strategies to overcome T-cell exhaustion. Therapeutic approaches to influence T-cell exhaustion have been already developed to target proteins CTLA-4, PD-1, and PD-L1 and have proven to be effective in treating melanoma and non-small-cell lung cancer during ongoing trials. [9] TILs in TNBC patients also display immuno-suppressive phenotypes [10] and the number of TILs detected by TNBC patients is higher than in other breast cancer subgroups [11] which makes TNBC a valid target for further TILs sequencing research for its application in the context of TN breast cancer or in search of further targets.

The aim of this work is to experiment with computer algorithms for the automated image segmentation and tumor-infiltrating lymphocytes assessment in Her2 positive and Triple Negative breast cancer histopathology slides based on Tumor InfiltratinG lymphocytes in breast cancER (TiGER) challenge. This works partly relies on publically released TiGER challenge and TCGA dataset to also further experiment with the prognostic significance of computer-generated TILs scores for predicting patients' survival.

2 Related work

2.1 Deep learning-based semantic segmentation

The focus of the following chapter is the existing deep learning-based approaches for semantic image segmentation, particularly for medical image analysis. As Shephard, Adam et al. discuss [12] segmentation of tumor/stroma as well as the detection of TILs can be viewed as a semantic segmentation problem.

The goal of semantic segmentation is to assign each image pixel to a category label corresponding to the underlying object. Due to the success of deep learning models in a wide range of vision applications, various deep learning-based algorithms have been developed and published in the literature [13]. One of the most prominent deep learning architectures used by the computer vision community include fully convolutional networks (FCNs) [14], encoder-decoders [15], generative adversarial networks (GANs) [16] and recurrent neural networks (RNNs) [17].

2.1.1 Fully convolutional networks (FCNs)

FCNs [14] are among the most widely used architectures for computer vision tasks and their general architecture consists of several learnable convolutions, pooling layers, and a final 1×1 convolution. While models based on this architecture perform well on challenging segmentation benchmarks, e.g. applied on scene segmentation [18] and instance aware semantic segmentation [19], they are also used on segmentation problems in histology domain such as colon glands segmentation [20], identification of muscle and messy regions in contexts of inflammatory bowel disease [21] as well as nuclei [22] and TILs [23] segmentation for breast cancer all performed on the Hematoxylin and Eosin (H&E) stained histopathology images. Moreover, the FCN method was applied for semantic segmentation of TCGA [24] breast data set [25], which is also used in this thesis. However, despite its popularity, the conventional FCN model has limitations such as loss of localization and the inability to process potentially useful global context information due to a series of down-sampling and a high sampling rate.

2.1.2 Encoder-decoder networks

A popular group of deep learning models for semantic image segmentation that aims to solve the aforementioned issues of FCNs is based on the convolutional encoder-decoder architecture [15]. Their model consists of two parts, an encoder consisting of convolutional layers and a deconvolution network that consists of deconvolution and unpooling layers that take

the feature vector as input and generate a map of pixel-wise class probabilities. An example of such a convolutional encoder-decoder architecture for image segmentation is SegNet [26]. The SegNet's encoder network has 13 convolutional layers with corresponding layers in the decoder. The final decoder output is fed to a multi-class soft-max classifier to produce class probabilities for each pixel independently. The main feature of SegNet is that the decoder uses pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear upsampling. This allows it to achieve high scores for road scene understanding problems [26], COVID-19 lung computed tomography image segmentation [27], liver tumor segmentation in computed tomography scans [28] and colon cancer histopathological images analysis [29]. There are several encoder-decoder models initially developed for biomedical image segmentation. Ronneberger et al. [30] proposed the U-Net model for segmenting biological microscopy images that can train with few annotated images effectively. U-Net has an FCN-like down-sampling part that extracts features with 3×3 convolutions and an up-sampling part. Feature maps from the encoder are copied to the corresponding decoder part of the network to avoid losing pattern information. Besides the segmentation of neuronal structures in electron microscopic recordings demonstrated in the original paper [30], U-Net was applied for numerous further tasks such as nuclei segmentation in histology images [31, 32], segmenting individual colon glands in histopathology images [33], epidermal tissue segmentation in histopathological images of skin biopsies [34] and cell segmentation on histopathology triple-negative breast cancer patients dataset [35]. A further example of an encoder-decoder model for semantic segmentation of histopathology images is HookNet [36]. The architecture consists of two encoder-decoder branches to extract contextual and fine-grained detailed information and combine it (hook up) for the target segmentation. The model showed improvement compared with single-resolution models and was applied to segment different histopathologies like breast cancer tissue sections [36], lung squamous cell carcinoma [36], invasive melanoma tumor [37] and cervical cancer [38] slides.

Another widely used group of deep learning models for semantic segmentation are the atrous (or dilated) convolutional models that include the DeepLab family [39, 40]. The use of atrous convolutions addresses the decreasing resolution caused by max-pooling and striding and Atrous Spatial Pyramid Pooling analyzes an incoming convolutional feature layer with filters at multiple sampling rates allowing to capture objects and image contexts at multiple scales to robustly segment objects at multiple scales. DeepLabv3+ [41] uses encoder-decoder architecture including atrous separable convolution, composed of a depthwise convolution (spatial convolution for each channel of the input) and pointwise convolution (1×1 convolution with the depthwise convolution as input). Authors [41] demonstrated the effectiveness of DeppLabv3+ model with modified Xception backbone at the recognition of visual object classes in realistic scenes, but it also found multiple applications such as skin lesion segmentation [42], segmentation of H&E stained breast cancer [43] and colorectal carcinoma [44] histopathology images. Despite all the efforts, even this popular architecture has constraints in learning long-range dependency and spatial correlations due to the inductive bias of locality and weight sharing [45] that may result in the sub-optimal segmentation of complex structures.

2.1.3 Generative adversarial networks (GANs)

GANs [16] have been applied to a wide range of computer vision tasks, and have been adopted for image segmentation as well. The general architecture of GANs consists of the discriminator and the generator. The generator learns the training data distribution and produces similar data, while the discriminator discriminates between real data and simulated data. Hence the task of the generator is to learn to generate the best images to fool the discriminator. There are many extended models such as conditional GAN (cGAN) [46] where the additional information is added to both the generator and the discriminator as a condition. This architecture was used for semantic segmentation of brain tumor in magnetic resonance imaging [47] and nuclei segmentation in histopathology images [48]. Further extended version of cGAN, pix2pix [49] was developed for conversion between different types of images but also found use cases in medical setting such as cell image segmentation on the fluorescence liver images [50] and retinal blood vessel segmentation [51]. A further GAN extension originally developed for image transformation between two domains but also applicable for segmentation is CycleGAN [52]. The architecture has two mirror-symmetric GANs to form a ring network to find the mapping between domains. For instance, CycleGAN was applied to kidney tissue [53] segmentation. Some GAN-based models were specifically developed for semantic segmentation in the medical domain, such as Domain Adaptation and Segmentation GAN (DASGAN) [54] that performs image-to-image translation and semantic tumor epithelium segmentation. It has an extended CycleGAN architecture with discriminator networks adjusted to predict pixel-wise class probability maps on top of predicting the correct source of an image. As a further example the proposed architecture consisting of a pyramid of GAN structures [55], each responsible for generating and segmenting images at a different scale, was applied to segment prostate histopathology images.

2.1.4 Recurrent neural networks (RNNs)

RNNs [17] have proven to be useful in modeling the short/long-term dependencies among pixels to generate segmentation maps. Pixels can be linked together and processed sequentially to model global contexts and improve semantic segmentation. ReSeg [56] is an RNN-based model for semantic segmentation. Each layer is composed of four RNNs that go through the image horizontally and vertically in both directions to provide relevant global information, while convolutional layers extract local features that are then followed by upsampling layers to recover the predictions at original image resolution. Another important development is a pixel-level segmentation of scene images using a long-short-term-memory (LSTM) network [57]. Segmentation is then carried out by 2D LSTM networks, allowing texture and spatial model parameters to be learned within a single model. But despite all further developments that showcase the potential even for histopathology image segmentation: RACE-net [58] applied for segmentation of the cell nuclei in H&E stained breast cancer slides, Her2Net [59] segmenting cell membranes and nuclei from human epidermal growth factor receptor-2 (HER2)-stained breast cancer images, etc., an important limitation of RNNs is that, due to their sequential nature, they are comparably slower, since this sequential calculation

cannot be easily parallelized.

2.1.5 Transformers

The Transformer in Natural Language Processing is an architecture that aims to solve sequence-to-sequence problems based on encoder-decoder architecture. These models rely on self-attention mechanisms and capture long-range dependencies among tokens (words) in a sentence without using RNNs or convolution. Transformers have also emerged in image semantic segmentation. Recent studies have shown that the Transformers can achieve superior performance than CNN-based approaches in various semantic segmentation applications [60]. The state-of-the-art Transformer-based semantic segmentation methods can be often applied either as convolution-free models or/and as CNN-Transformer hybrid models. Swin-Transformer [61] for instance is a pure hierarchical Transformer that can serve as a backbone for various computer vision tasks including semantic segmentation. To tokenize the image, it brakes the image into windows that further consist of patches. It constructs a hierarchical representation of an image by starting from small-sized patches and gradually merging neighboring patches into deeper Transformer layers. Swin-Transformer or its slightly modified successors found its application in the medical domain as well, often as a backbone, for example for colon cancer segmentation in H&E stained histopathology images [62] or gland segmentation [63]. A further popular fully transformer-based model for semantic segmentation is Segmenter [64]. The encoder consists of Multi-head Self Attention and Multi-Layer Perceptron (MLP) blocks, as well as two-layer norms and residual connections after each block and a linear decoder that bilinearly up-samples the sequence into a 2D segmentation mask. While performing well on scene segmentation [64], is not particularly used in the medical domain. In the field of medical image segmentation, TransUNet [65] was the first attempt to establish self-attention mechanisms by combining transformer with U-Net and proved that transformers can be used as powerful encoders for medical image segmentation. A novel positional-encoding-free Transformer SegFormer [66] set new state-of-the-art in terms of efficiency and accuracy in publicly available semantic segmentation datasets and applied for instance in gland and nuclei segmentation [63]. This architecture remains promising also for semantic segmentation in medical applications due to the positional-encoding-free encoder and lightweight MLP decoder.

2.2 Survival analysis

The following chapter focuses on conducted research for the development of TILs scores as a prognostic biomarker for survival analysis in breast cancer. Since the overall survival (OS) is the primary endpoint for prognostic analysis in this thesis, the survival methods are well established and include the Kaplan–Meier method [67] to estimate OS and Cox proportional hazard models [68] to quantify the hazard ratio (HR) for the effects of biomarker groups.

2.2.1 TILs as prognostic biomarker

Tumor-Infiltrating Lymphocytes (TILs) have strong prognostic and predictive value in breast cancer [69, 70]. Amgad, M. et al. [70] assessed three variants of the TILs score:

1. Number of TILs / Stromal area
2. Number of TILs / Number of cells in stroma
3. Number of TILs / Total Number of cells

The results performed on the BCSS and NuCLS breast carcinoma datasets [25, 71] showed the most prognostic TILs score to be the number of TILs divided by the total number of cells within the stromal region. A further breast cancer study [72] showed that the binarized tumor TILs infiltration fraction is predictive of survival, by analyzing the proportion of pixels in the image that were predicted as containing tumor as well as lymphocytes (number of pixels predicted as lymphocyte and tumor divided by the number of pixels predicted as tumor). Bai, et al. [73] also found associations of clinical outcomes in breast cancer with TILs scores based on the number of TILs divided by the number of TILs and tumor cells detected.

The stromal TILs (sTILs) have been shown to have prognostic value in HER2+ breast cancer and TNBC [69]. sTIL density was found significantly prognostic for OS not only while applied on H&E slides but IHC as well. [74] Applied on the TCGA-BRCA mixed with non publicaly available dataset, Thagaard, J. et al. [75] tried to mimic the approach of the pathologist and therefore defined tumor-associated stroma. Tumor-associated stroma includes a margin of 250µm from the border of the tumor into the surrounding stroma. The sTIL density was calculated as the number of TILs within the tumor-associated stroma per mm^2. The patient cohort was then stratified into two groups: high and low sTIL density by using maximally selected rank statistics for cutpoint selection. As a result sTIL density stratified the patients significantly into two distinct prognostic groups. For continuous variables, the sTIL density was divided by 300 and higher sTILs scores were associated with significantly prolonged overall survival. For the TCGA-BRCA dataset, a further TIL score was found significant as the overlapping area between lymphocyte-dense regions and stromal regions divided by the size of the stromal regions. [76] Whereas a study, that focused on TNBC cases of TCGA, did not observe any differences in OS neither while using a continuous variable of manually annotated TILs (scored by a pathologist and partitioned into eight different groups, e.g. < 1%, 10-20%, etc.) nor after applying the log-rank test [77]. On the other hand, Fassler, D. J., et al. [78] confirmed correlation of intratumoral TIL infiltration with increased OS in breast cancer in the TCGA-BRCA cohort. TIL infiltrate percentage was calculated as the number of predicted patches that were classified as positive for tumor and lymphocyte divided by total number of cancer patches. Another used definition of sTILs was the percentage of tumor stroma area containing a lymphocytic infiltrate without direct contact with tumor cells [79]. Hence, there is no canonic method for the automatic determination of TILs score based on the H&E breast cancer tissue samples.

Furthermore, studies found a three-scale grading system for reporting TILs status to be applicable, instead of continuous or binary grouped TILs densities [80]. More advanced

TILs-based features such as the Ball-Hall Index of spatially connected TILs regions (clusters) also showed association with survival, particularly within the BRCA dataset of TCGA [81].

3 Methods

3.1 Semantic segmentation

3.1.1 DeepLab

One of the challenges in semantic segmentation using standard CNNs is that as the input feature map goes through the network it gets smaller and the information about objects of a smaller scale can be lost. DeepLab family introduces atrous convolutions that extract more dense features which help to preserve the object's information. Compared to standard convolutions, atrous convolutions have an additional parameter, atrous rate, which is the stride at which the input is sampled (Figure 3.1 a). The atrous convolution is used in the last few blocks on features that were extracted from the backbone network (e.g. ResNet [82]).

One of the latest models in this family, DeepLabv3 [40], applies several parallel atrous convolutions with different atrous rates (Atrous Spatial Pyramid Pooling, or ASPP, Figure 3.1 b) to effectively capture multi-scale information. Image-level features, or image pooling, are also applied to incorporate global context information. Those are calculated by applying global average pooling on the last feature map of the backbone. After applying all the operations in parallel, the results of each operation along the channel is concatenated and 1×1 convolution is applied to get the output. The addition of atrous convolutions allows the enlargement of the field of view without increasing the size of the filtering kernel, therefore no increase in the computation time.

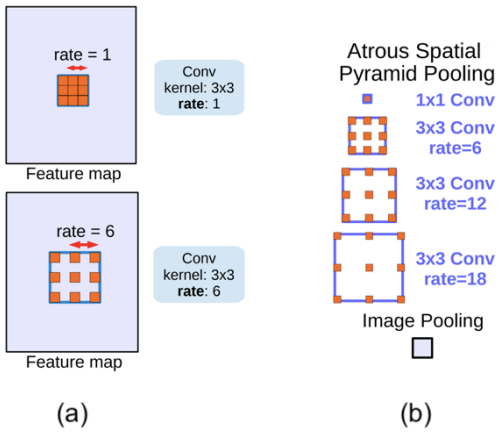
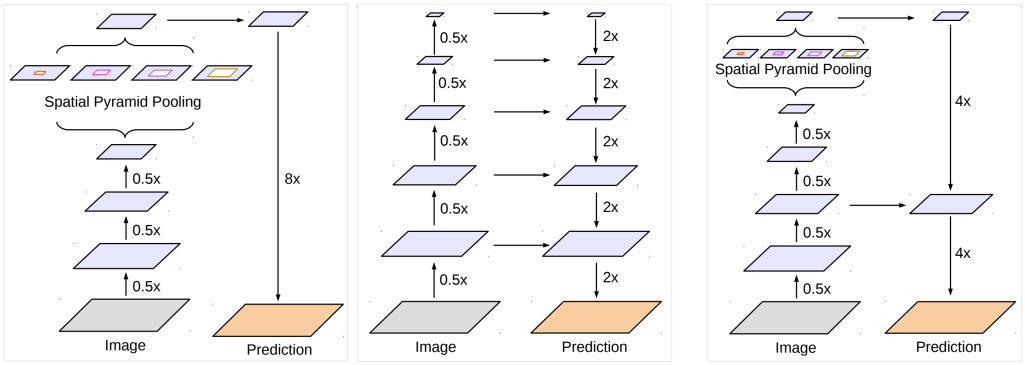


Figure 3.1: (a) Atrous convolution, (b) ASPP augmented with Image Pooling (or Image-level features) [40]

DeepLabv3+

The reproduction of shape contours during semantic image segmentation remained difficult with DeepLabv3 [41]. DeepLabv3 bilinearly upsamples the logits both during training and evaluation (Fig. 3.2 a), hence the improvements were made to employ the encoder-decoder structure (Figure 3.2) to avoid using a naive decoder. DeepLabv3+ [41] adds the decoder



(a) Spatial Pyramid Pooling (b) Encoder-Decoder (c) Encoder-Decoder with Atrous Conv

Figure 3.2: The spatial pyramid pooling module of DeepLabv3 (a), the encoder-decoder structure (b) and DeepLabv3+ adaptation (c) [41]

module on top of the encoder output, as shown in Fig. 3.3. In the decoder module, the 1×1 convolution reduces the channels of the low-level feature map from the encoder module which is then concatenated with the DeepLabv3 feature map and the 3×3 convolution obtains sharper segmentation results. As a result, DeepLabv3+ holds rich semantic information from the encoder module, while the detailed object boundaries are recovered by the decoder module and the spatial information is retrieved.

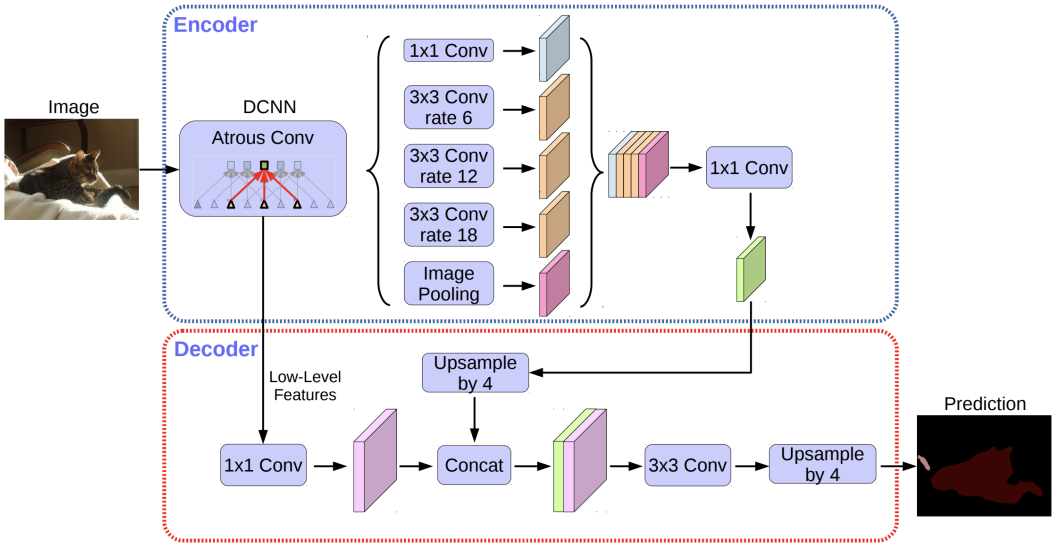


Figure 3.3: DeepLabv3+ architecture. DeepLabv3 as encoder and proposed decoder structure for semantic image segmentation. [41]

3.1.2 Transformers

Transformers [83] were originally designed for the neural machine translation problem in NLP to capture long-range dependencies among words in a sentence. Their architecture converts one sequence into another one based on encoder-decoder architecture, but it differs from the previously existing sequence-to-sequence models because it does not imply any Recurrent Networks.

The input and output are first embedded into an n-dimensional space. Since the network and the self-attention are permutation invariant, the positional encoding is added to create a representation of the position of the word in the sentence. The following modules consist mainly of Multi-Head Attention and Feed Forward layers. Encoder (Figure 3.4, left) and decoder (Figure 3.4, right) are composed of those modules that can be stacked on top of each other $N\times$ times.

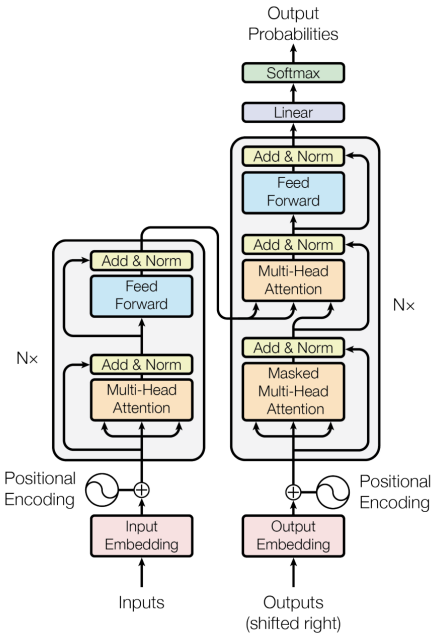


Figure 3.4: Transformer model architecture. [83]

Self-attention is a sequence-to-sequence operation. It takes a weighted average over all the input vectors using dot product. Scaled Dot-Product Attention (Figure 3.5, left) can be described by the following equation:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (3.1)$$

where in the context of the translation problem, Q is a matrix of vector representation of one word in the sequence, K contains vector representations of all the words in the sequence and V contains again the vector representations of all the words in the sequence. For the multi-head attention modules in the encoder and decoder, V consists of the same word sequence as Q. However, for the attention module that is taken into account, the encoder and the decoder sequences, V, and Q are different. Q, K, and V matrices are used to calculate the attention scores. These scores measure how much attention needs to be placed on words of the input sequence with respect to a word at a certain position. The scaling factor $\sqrt{d_k}$ is applied to avoid large values that after applying softmax would lead to vanishing gradients.

While Scaled Dot-Product Attention focuses on the whole sentence, Multi-Head Attention approaches different segments of the words. The word vectors are divided into a fixed number (number of heads) of parts, and then within Multi-Head Attention (Figure 3.5, right) the attention mechanism is repeated multiple times on those separate parts with linear projections of Q, K, and V. Since the Feed-Forward layer is expecting just one matrix, a vector for each word, the outputs are linearly concatenated. This allows the system to learn from different representations of Q, K, and V.

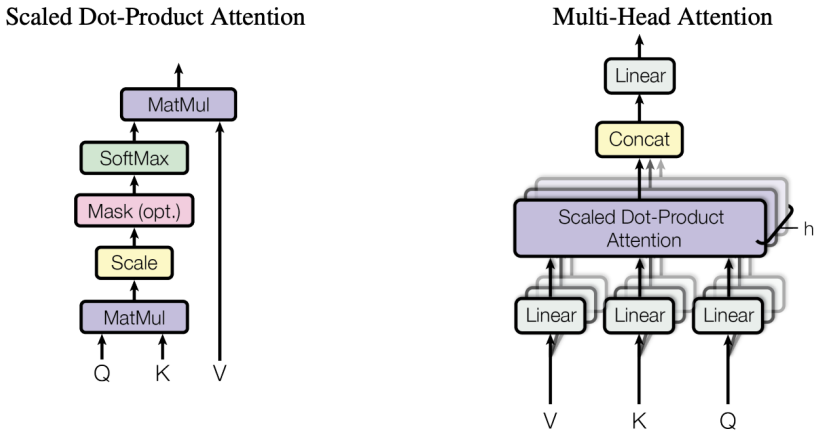


Figure 3.5: Scaled Dot-Product Attention (left). Multi-Head Attention consists of several attention layers running in parallel (right). [83]

To add element-wise non-linearity transformation of incoming vectors, the transformer includes feed-forward networks. It processes the output from one attention layer so that it fits better for the next attention layer. Each of the layers in the encoder and decoder contains a fully connected feed-forward network, which is applied to each position separately and identically. These feed-forward layers can be described as a separate, identical linear transformation of each element from the given sequence.

Naive application of the transformers approach into the image domain would require evaluation of relations between each pixel and every other pixel, which is obviously not scalable. The Visual transformer (ViT) [84] is the first work to prove that a pure Transformer can achieve state-of-the-art performance in image classification. ViT converts the input image into a 1D series by cutting it into patches and feeding it to a linear layer. It yields a patch embedding. Position embeddings are added to the image patch embeddings. Adding the learnable position embeddings to each patch allows the model to learn the structure of the image. The rest of the pipeline is a standard encoder and decoder blocks of the transformer. The decoder learns to map patch-level encodings coming from the encoder to patch-level class scores. Next, these patch-level class scores are upsampled by bilinear interpolation to pixel-level scores.

SegFormer

SegFormer [85] is a positional-encoding-free transformer based semantic segmentation method. As depicted in Figure 3.6, it consists of two main modules: a hierarchical Transformer encoder to generate high-resolution coarse features and low-resolution fine features, and a lightweight All-MLP decoder to fuse these multi-level features and produce the final semantic segmentation mask.

The $H \times W \times 3$ input image is forwarded to the hierarchical Transformer encoder to obtain multi-level features at $\frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}$ resolution after passing through four transformer blocks.

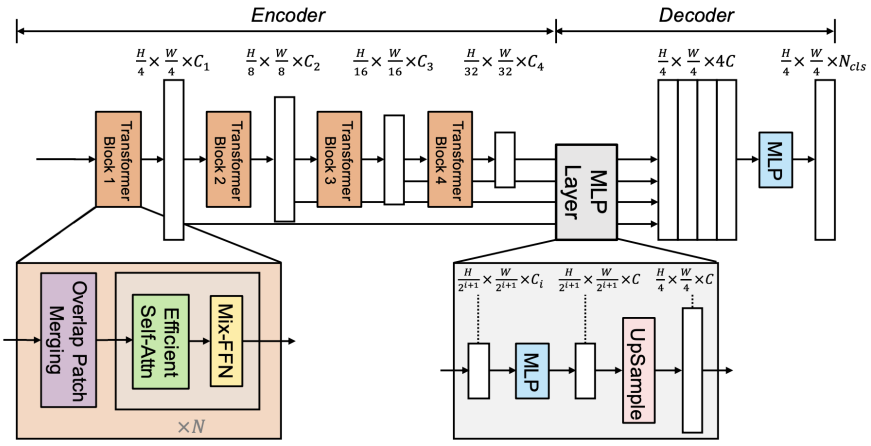


Figure 3.6: SegFormer consists of two main modules: A hierarchical Transformer encoder to extract coarse and fine features; and a lightweight All-MLP decoder to directly fuse these multi-level features and predict the semantic segmentation mask. "FFN" indicates feed-forward network. (modified image [85] according to the official implementation)

Each transformer block consists of three modules: Overlap Patch Merging, and classical transformer building blocks: Self-Attention and Feed-forward network.

The standard transformer receives input as a 1D sequence (such as word embeddings in the previous chapter 3.1.2). To handle images, those need to be reshaped into a sequence of flattened 2D patches. Overlapped Patch Merging produces features given an image and parameters: patch size K, stride between two adjacent patches S, and padding size P. In the original paper [85] those are set to $K = 7$, $S = 4$ and $P = 3$. Therefore the input is split into fixed-size patches, which then go through a linear projection. The result is a hierarchical feature map F_i with a resolution $\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times C_i$ where $i \in \{1, 2, 3, 4\}$ and C_{i+1} is larger than C_i. By performing this with overlapped patches SegFormer aims to preserve the local continuity around those patches.

The main computation bottleneck of each transformer block in encoder is the self-attention layer. In SegFormer, before applying the self-attention according to the formula 3.1, the sequence K is reduced by ratio R:

$$\hat{K} = Reshape(\frac{N}{R}, C \cdot R)(K)$$

$$K = Linear(C \cdot R, C)(\hat{K})$$

where $N = H \times W$, $Reshape(\frac{N}{R}, C \cdot R)(K)$ refers to reshaping K to the the shape of $\frac{N}{R} \times (C \cdot R)$, and $Linear(C \cdot R, C)(\hat{K})$ refers to a linear layer taking a $(C \cdot R)$-dimensional tensor as input and generating a C-dimensional tensor as output. Therefore, the new K has dimensions $\frac{N}{R} \times C$. In original experiments, R was set to $[64, 16, 4, 1]$ from stage-1 to stage-4 and resulted in a reduction of the complexity of the self-attention mechanism.

Mix-FFN (feed-forward network) can be formulated as:

$$x_{out} = MLP(GELU(Conv3 \times 3(MLP(x_{in})))) + x_{in}$$

where x_{in} is the feature from the self-attention module. By using 3×3 convolution and zero padding in a feed-forward network SegFormer aims to leak pixel location information since it is a positional-encoding-free method.

The multi-level features are then passed to All-MLP decoder to predict the segmentation mask at $\frac{H}{4} \times \frac{W}{4} \times N_{cls}$ resolution, where N_{cls} is the number of classes. The proposed All-MLP decoder consists of four main steps. First, multi-level features from the encoder go through an MLP layer to unify the channel dimension (3.2). Then, features are up-sampled to $\frac{1}{4}$th of the original image (3.3). Third, an MLP layer is adopted to fuse the concatenated features (3.4). Finally, another MLP layer takes the fused feature to predict the segmentation mask (3.5).

$$\hat{F}_i = MLP(C_i, C)(F_i), \forall i \tag{3.2}$$

$$\hat{F}_i = Upsample(\frac{H}{4} \times \frac{W}{4})(\hat{F}_i), \forall i \tag{3.3}$$

$$F = MLP(4C, C)(MLP(\hat{F}_i)), \forall i \tag{3.4}$$

$$M = MLP(C, N_{cls})(F) \tag{3.5}$$

where F_i is the the feature and M is the final mask.

3.2 Survival Analysis

The overall survival (OS) is the primary endpoint for prognostic analysis in this thesis, hence time to the event (death) is of interest. Survival data are generally described and modeled in terms of two related probabilities, namely survival and hazard. [86] This thesis focuses on non-parametric models to avoid making any additional assumptions about the distributions. The survival probability $S(t)$ is the probability that an individual survives from the time origin (in our case diagnosis of breast cancer) to a specified future time t. It can be denoted as:

$$S(t) = Pr(T > t) = 1 - F(t) = \int_t^\infty f(x)dx = \text{ Probability of surviving past time } t$$

where T is a random variable that indicates the time until the event of interest (death). $F(t)$ and $f(t)$ are the cumulative distribution function and probability density function of T. The hazard is the probability that an individual who is under observation at a time t has an event at that time:

$$h(t) = lim_{\delta t \to 0} \frac{Pr(t \leq T \leq t + \delta t | T > t)}{\delta t} = \frac{f(t)}{1 - F(t)}$$

In contrast to the survivor function, which focuses on not having an event, the hazard function focuses on the event occurring. So if hazard probability describes the intensity of death [87] at the time t given that the individual has already survived past time t, then the cumulative hazard is the cumulative amount of hazard up to time t. The cumulative hazard $H(t)$, defined as the integral of the hazard, can be calculated using the survival probability with help of the Laplace transform:

$$H(t) = \int_0^t h(x)dx = \int_0^t \frac{f(x)}{1-F(x)}dx = -ln(1-F(t)) = -\log(S(t))$$

The cumulative hazard can be interpreted as the number of events that would be expected for each individual by time t if the event was a repeatable process. [86]

3.2.1 Kaplan–Meier estimator

The survival probability can be estimated nonparametrically from observed survival times, both censored and uncensored, using the Kaplan–Meier method. The estimated probability of surviving past time t is calculated as:

$$\hat{S}(t) = \prod_{i;t_i \leq t} (1 - \frac{d_i}{n_i})$$

where n_i is the number of patients alive before t_i (and not censored) and d_i is the number of observed events at t_i. $t_0 = 0$ and $S(0) = 1$. The estimated probability is a step function that changes value only at the time of an event. To characterize the survival in a homogeneous group often the empirical survival function is visualized with Kaplan–Meier plot.

3.2.2 Cox model

Additionally to the event time, there is often access to other covariates of individuals (e.g. age, gender, BMI, etc.). Often the goal is to understand how the covariates affect the survival function of the event. [87] Let C denote those covariates. The conditional survival function can be formulated as followed:

$$S(t|c) = Pr(T > t|C = c) = \text{ Probability of surviving past time } t \text{ given } c$$

Hence, the conditional hazard function and conditional cumulative hazard are:

$$H(t|c) = -\log(S(t|c)), \text{ hence } h(t|c) = -\frac{\partial \log S(t|c)}{\partial t}$$

The Cox proportional hazard model models the hazard function $h(t|C = c)$ as:

$$h(t|C = c) = h_0(t)\exp(c^T\beta)$$

where β is the vector of coefficients for each of the covariates and $h_0(t)$ is the baseline hazard function. The hazard ratio, or ***risk***, is the exponential of β_i value $\eta_i = \exp(\beta_i)$ and the baseline

hazard describes how the risk of event per time unit changes over time at baseline levels of covariates. The Cox model assumes that the covariates have a linear multiplication effect on the hazard function and the effect stays the same over time.

$$\frac{h(t|c_i)}{h(t|c_j)} = \frac{h_0(t)\exp(c_i^T\beta)}{h_0(t)\exp(c_j^T\beta)} = \frac{\exp(c_i^T\beta)}{\exp(c_j^T\beta)} = \exp((c_i - c_j)^T\beta)$$

The ratio of the hazard function between two individuals with different covariates c_i and c_j is a constant over time since $h_0(t)$ was canceled out. Hence the name, proportional hazard model. The conditional hazard function is:

$$H(t|c) = \exp(c^T\beta)\int_0^t h_0(s)ds = \exp(c^T\beta)H_0(t)$$

It yields a conditional survival function:

$$S(t|c) = \exp(-H(t|c)) = \exp(-\exp(c^T\beta)H_0(t)) = \exp(-H_0(t))^{\exp(c^T\beta)} = S_0(t)^{\exp(c^T\beta)}$$

Estimation of the parameter β is often done by maximizing the partial likelihood function $\hat{\beta}_n = argmax_\beta \hat{L}_n(\beta)$, where:

$$\hat{L}(\beta) = \prod_{i=1}^{n} \frac{h(T_i|C_i)}{\sum_{j:T_j \geq T_i} h(T_j|C_j)} = \prod_{i=1}^{n} \frac{\exp(C_i^T\beta)}{\sum_{j:T_j \geq T_i} \exp(C_j^T\beta)}$$

A positive sign of β_i indicates a higher risk of an event, hence the probability for the event for that particular subject is higher. Likewise for a negative signed β_i, lower risk, and lower probability. The actual value of β_i plays a role as well. Values less than one will reduce the hazard and values greater than one, increase it.

A model's accuracy can be quantified based on concordance. [88] It is a measure of the rank correlation between predicted risks and observed time points. It is defined as the ratio of correctly ordered (concordant) patient pairs to all concordant and discordant patient pairs. Let i, j be a patient pair. If a model predicts a higher risk for the first patient ($\eta_i > \eta_j$), for it to be a concordant pair first patient should have a shorter survival time in comparison with the other patient ($T_i < T_j$) and similarly if lower risk then longer survival time, $\eta_i < \eta_j \ \ \& \ \ T_i > T_j$. If both patients are censored the pair is discarded. If only one patient is censored, the pair is not discarded only if the other patient experienced the event before the censoring time. By construction, concordance must be between 0 and 1, with 1 representing the perfect agreement between model and observation and 0.5 representing random guesses.

Additionally, to estimate the goodness-of-fit the p-value is determined. The Wald test is typically used to evaluate the significance of a variable in the model estimated with the maximum likelihood function. The null hypothesis is that the model does not fit the data well. The Wald statistic tests, whether β_i coefficient is statistically significantly different from 0 and is defined as:

$$W = \frac{(\hat{\beta}_n - \beta_0)^2}{var(\hat{\beta}_n)}$$

If the true coefficient was β_0, then the sampling distribution of the Wald test statistic should be approximate $\mathcal{N}(0,1)$. The p-value gives the probability of observing a test statistic as extreme as the one observed if the sampling distribution was $\mathcal{N}(0,1)$. If the p-value is small, the observed test statistic is very unlikely under the null hypothesis. And the significance level of 0.05 indicates that there is a 5% risk of being wrong by concluding that the model fits the data well when it doesn't.

4 Data

4.1 Segmentation

The data comes from publicly released Tumor InfiltratinG lymphocytes in breast cancER (TiGER) challenge dataset containing digital pathology images of Her2 positive (Her2+) and Triple Negative (TNBC) breast cancer whole-slide images (WSIs), regions of interest (ROIs), and manual annotations. More specifically, the WSIROIS dataset was used for model training, validation, and testing (see Table 4.1). TiGER data, both at WSI and ROI level, was released at a spacing (pixel size) of approximately 0.5 µm/px, for more information please refer to the original challenge website[1]. The TiGER tissue annotations include eight labels that

Source	Tissue			TILs		
	#slides	#ROIs	median ROI size #pixels [k]	#slides	#ROIs	median ROI size #pixels [k]
TCGA-BRCA	151	151	4 983	124	1744	20
RUMC	26	81	1 312	26	81	1 312
JB	18	54	1 465	18	54	1 465
	195	286		168	1879	

Table 4.1: TiGER data overview. Sources: Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA), Radboud University Medical Center (RUMC) and Jules Bordet Institute (JB). Tissue slides and ROIs refer to the segmentation images and annotations whereas TILs prefix specifies the data for TILs detection provided by the challenge.

were reduced to three (see Table 4.2). The training masks were generated using available XML files. In the provided mask images, in certain cases, regions not included in ROIs and non-annotated regions in ROIs where marked with the same label, which could not be directly used for training.

While for tissue segmentation the images and their masks could be used as directly extracted from the dataset, the data for TILs segmentation required some preprocessing. The TiGER fixed-size bounding box annotation for lymphocytes and plasma cells (see Table 4.3) was adapted for segmentation by transforming each bounding box into an annotation of the center pixel with dilatation of three.

[1] `https://tiger.grand-challenge.org/Data/`

TiGER Tissue Label	Share	ID	new ID
Invasive tumor	0.283	1	1
In-situ tumor	0.029	3	1
Tumor-associated stroma	0.286	2	2
Inflamed stroma	0.096	6	2
Necrosis not in-situ	0.048	5	0
Healthy glands	0.0008	4	0
Background	0.231	0	0
Rest	0.026	7	0

Table 4.2: Reduction of labels provided in TiGER challenge dataset. Resulting labels include three classes: Tumor (1), Stroma (2) and Rest (0) with shares of 0.312, 0.382 and 0.306. Shares were calculated by dividing the number of pixels belonging to some label by the number of the pixel in the current image and averaged over all images.

Source				Number of cells per ROI		
	#slides	#ROIs	#cells	min	max	median
TCGA-BRCA	124	1 744	19 115	0 (44.3%)	206	1
RUMC	26	81	4 728	0 (7.4%)	657	19
JB	18	54	5 523	0 (7.4%)	608	51.5
	168	1 879	29 366			

Table 4.3: Data overview for TILs detection. Sources: Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA), Radboud University Medical Center (RUMC) and Jules Bordet Institute (JB). Number of cells here refers to the number of bounding boxes that were assigned for lymphocytes and plasma cells, further named TILs.

4.2 Survival Analysis

TiGER challenge aims to assess the prognostic significance of computer-generated TILs scores for predicting survival by applying the Cox proportional hazards model. The survival analysis is done using a large independent test dataset that includes cases from both clinical routine and from a phase 3 clinical trial, which is not directly accessible by participants. The survival analysis within this thesis is done exclusively on publicly available TCGA-BRCA data. Where death (`vital_status` = 1) is considered as an event, and the time until the event or censoring is taken either from `days_to_death` (number of days to death from the first diagnosis) or `days_to_followup` (number of days to last follow-up from first diagnosis).

vital_status	#cases	median age at diagnosis [years]	median time to event [months]
Dead	146	62	37.8
Alive	919	58	26.3
	1065	58	28.7

Table 4.4: Survival data overview.

List of Figures

3.1 (a) Atrous convolution, (b) ASPP augmented with Image Pooling (or Image-level features) [40] 9
3.2 The spatial pyramid pooling module of DeepLabv3 (a), the encoder-decoder structure (b) and DeepLabv3+ adaptation (c) [41] 10
3.3 DeepLabv3+ architecture. DeepLabv3 as encoder and proposed decoder structure for semantic image segmentation. [41] 10
3.4 Transformer model architecture. [83] 11
3.5 Scaled Dot-Product Attention (left). Multi-Head Attention consists of several attention layers running in parallel (right). [83] 12
3.6 SegFormer consists of two main modules: A hierarchical Transformer encoder to extract coarse and fine features; and a lightweight All-MLP decoder to directly fuse these multi-level features and predict the semantic segmentation mask. “FFN” indicates feed-forward network. (modified image [85] according to the official implementation) 13

List of Tables

4.1 TiGER data overview. Sources: Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA), Radboud University Medical Center (RUMC) and Jules Bordet Institute (JB). Tissue slides and ROIs refer to the segmentation images and annotations whereas TILs prefix specifies the data for TILs detection provided by the challenge. 18
4.2 Reduction of labels provided in TiGER challenge dataset. Resulting labels include three classes: Tumor (1), Stroma (2) and Rest (0) with shares of 0.312, 0.382 and 0.306. Shares were calculated by dividing the number of pixels belonging to some label by the number of the pixel in the current image and averaged over all images. 19
4.3 Data overview for TILs detection. Sources: Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA), Radboud University Medical Center (RUMC) and Jules Bordet Institute (JB). Number of cells here refers to the number of bounding boxes that were assigned for lymphocytes and plasma cells, further named TILs. 19
4.4 Survival data overview. 20

Bibliography

[1] B. S. Chhikara and K. Parang. "Global Cancer Statistics 2022: the trends projection analysis". In: (2022).

[2] *Tiger - Grand Challenge*. URL: https://tiger.grand-challenge.org/Home/.

[3] Q. D. Vu, S. Graham, T. Kurc, M. N. N. To, M. Shaban, T. Qaiser, N. A. Koohbanani, S. A. Khurram, J. Kalpathy-Cramer, T. Zhao, et al. "Methods for segmentation and classification of digital microscopy tissue images". In: *Frontiers in bioengineering and biotechnology* (2019), p. 53.

[4] L. Cao and Y. Niu. "Triple negative breast cancer: special histological types and emerging therapeutic methods". In: *Cancer biology & medicine* 17.2 (2020), p. 293.

[5] B. S. Yadav, P. Chanana, and S. Jhamb. "Biomarkers in triple negative breast cancer: A review". In: *World journal of clinical oncology* 6.6 (2015), p. 252.

[6] A.-V. Laenkholm, G. Callagy, M. Balancin, J. Bartlett, C. Sotiriou, C. Marchio, M. Kok, C. H. Dos Anjos, and R. Salgado. "Incorporation of TILs in daily breast cancer care: how much evidence can we bear?" In: *Virchows Archiv* (2022), pp. 1–16.

[7] M. V. Dieci, N. Radosevic-Robin, S. Fineberg, G. Van den Eynden, N. Ternes, F. Penault-Llorca, G. Pruneri, T. M. D'Alfonso, S. Demaria, C. Castaneda, et al. "Update on tumor-infiltrating lymphocytes (TILs) in breast cancer, including recommendations to assess TILs in residual disease after neoadjuvant therapy and in carcinoma in situ: a report of the International Immuno-Oncology Biomarker Working Group on Breast Cancer". In: *Seminars in cancer biology*. Vol. 52. Elsevier. 2018, pp. 16–25.

[8] G. Gao, Z. Wang, X. Qu, and Z. Zhang. "Prognostic value of tumor-infiltrating lymphocytes in patients with triple-negative breast cancer: a systematic review and meta-analysis". In: *BMC cancer* 20.1 (2020), pp. 1–15.

[9] M. A. Postow, M. K. Callahan, and J. D. Wolchok. "Immune checkpoint blockade in cancer therapy". In: *Journal of clinical oncology* 33.17 (2015), p. 1974.

[10] W. Chung, H. H. Eum, H.-O. Lee, K.-M. Lee, H.-B. Lee, K.-T. Kim, H. S. Ryu, S. Kim, J. E. Lee, Y. H. Park, et al. "Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer". In: *Nature communications* 8.1 (2017), pp. 1–12.

[11] A. Schneeweiss, C. Denkert, P. A. Fasching, C. Fremd, O. Gluz, C. Kolberg-Liedtke, S. Loibl, and H.-J. Lück. "Diagnosis and therapy of triple-negative breast cancer (TNBC)–recommendations for daily routine practice". In: *Geburtshilfe und Frauenheilkunde* 79.06 (2019), pp. 605–617.

[12] A. Shephard, M. Jahanifar, R. Wang, M. Dawood, S. Graham, K. Sidlauskas, S. A. Khurram, N. Rajpoot, and S. E. A. Raza. "TIAger: Tumor-Infiltrating Lymphocyte Scoring in Breast Cancer for the TiGER Challenge". In: *arXiv preprint arXiv:2206.11943* (2022).

[13] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos. "Image segmentation using deep learning: A survey". In: *IEEE transactions on pattern analysis and machine intelligence* (2021).

[14] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

[15] H. Noh, S. Hong, and B. Han. "Learning deconvolution network for semantic segmentation". In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1520–1528.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. "Generative adversarial nets". In: *Advances in neural information processing systems* 27 (2014).

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors". In: *nature* 323.6088 (1986), pp. 533–536.

[18] C. Yu, J. Wang, C. Gao, G. Yu, C. Shen, and N. Sang. "Context prior for scene segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 12416–12425.

[19] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei. "Fully convolutional instance-aware semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2359–2367.

[20] A. BenTaieb and G. Hamarneh. "Topology aware fully convolutional networks for histology gland segmentation". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2016, pp. 460–468.

[21] J. Wang, J. D. MacKenzie, R. Ramachandran, and D. Z. Chen. "A deep learning approach for semantic segmentation in histology tissue images". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016, pp. 176–184.

[22] V. A. Natarajan, M. S. Kumar, R. Patan, S. Kallam, and M. Y. N. Mohamed. "Segmentation of nuclei in histopathology images using fully convolutional deep neural architecture". In: *2020 International Conference on computing and information technology (ICCIT-1441)*. IEEE. 2020, pp. 1–7.

[23] M. Amgad, A. Sarkar, C. Srinivas, R. Redman, S. Ratra, C. J. Bechert, B. C. Calhoun, K. Mrazeck, U. Kurkure, L. A. Cooper, et al. "Joint region and nucleus segmentation for characterization of tumor infiltrating lymphocytes in breast cancer". In: *Medical Imaging 2019: Digital Pathology*. Vol. 10956. SPIE. 2019, pp. 129–136.

[24] D. A. Gutman, J. Cobb, D. Somanna, Y. Park, F. Wang, T. Kurc, J. H. Saltz, D. J. Brat, L. A. Cooper, and J. Kong. "Cancer Digital Slide Archive: an informatics resource to support integrated in silico analysis of TCGA pathology data". In: *Journal of the American Medical Informatics Association* 20.6 (2013), pp. 1091–1098.

[25] M. Amgad, H. Elfandy, H. Hussein, L. A. Atteya, M. A. Elsebaie, L. S. Abo Elnasr, R. A. Sakr, H. S. Salem, A. F. Ismail, A. M. Saad, et al. "Structured crowdsourcing enables convolutional segmentation of histology images". In: *Bioinformatics* 35.18 (2019), pp. 3461–3467.

[26] V. Badrinarayanan, A. Kendall, and R. Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation". In: *IEEE transactions on pattern analysis and machine intelligence* 39.12 (2017), pp. 2481–2495.

[27] A. Saood and I. Hatem. "COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet". In: *BMC Medical Imaging* 21.1 (2021), pp. 1–10.

[28] S. Almotairi, G. Kareem, M. Aouf, B. Almutairi, and M. A.-M. Salem. "Liver tumor segmentation in CT scans using modified SegNet". In: *Sensors* 20.5 (2020), p. 1516.

[29] A. B. Hamida, M. Devanne, J. Weber, C. Truntzer, V. Derangère, F. Ghiringhelli, G. Forestier, and C. Wemmert. "Deep learning for colon cancer histopathological images analysis". In: *Computers in Biology and Medicine* 136 (2021), p. 104730.

[30] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi. Cham: Springer International Publishing, 2015, pp. 234–241.

[31] A. Lagree, M. Mohebpour, N. Meti, K. Saednia, F.-I. Lu, E. Slodkowska, S. Gandhi, E. Rakovitch, A. Shenfield, A. Sadeghi-Naini, et al. "A review and comparison of breast tumor cell nuclei segmentation performances using deep convolutional neural networks". In: *Scientific Reports* 11.1 (2021), pp. 1–11.

[32] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu. "RIC-Unet: An improved neural network based on Unet for nuclei segmentation in histology images". In: *Ieee Access* 7 (2019), pp. 21420–21428.

[33] H. Pinckaers and G. Litjens. "Neural ordinary differential equations for semantic segmentation of individual colon glands". In: *arXiv preprint arXiv:1910.10470* (2019).

[34] K. R. Oskal, M. Risdal, E. A. Janssen, E. S. Undersrud, and T. O. Gulsrud. "A U-net based approach to epidermal tissue segmentation in whole slide histopathological images". In: *SN Applied Sciences* 1.7 (2019), pp. 1–12.

[35] M. E. Bagdigen and G. Bilgin. "Cell segmentation in triple-negative breast cancer histopathological images using U-Net architecture". In: *2020 28th Signal Processing and Communications Applications Conference (SIU)*. IEEE. 2020, pp. 1–4.

[36] M. Van Rijthoven, M. Balkenhol, K. Siliņa, J. Van Der Laak, and F. Ciompi. "HookNet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images". In: *Medical Image Analysis* 68 (2021), p. 101890.

[37] A. Shah, A. Mehta, M. Wang, N. Neumann, T. McCalmont, and A. Zakhor. "Deep Learning Segmentation of Invasive Melanoma". In: *International Conference on Image Processing, Bordeaux, France*. 2022.

[38] Z. Meng, Z. Zhao, B. Li, F. Su, and L. Guo. "A cervical histopathology dataset for computer aided diagnosis of precancerous lesions". In: *IEEE Transactions on Medical Imaging* 40.6 (2021), pp. 1531–1541.

[39] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs". In: *IEEE transactions on pattern analysis and machine intelligence* 40.4 (2017), pp. 834–848.

[40] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. "Rethinking atrous convolution for semantic image segmentation". In: *arXiv preprint arXiv:1706.05587* (2017).

[41] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. "Encoder-decoder with atrous separable convolution for semantic image segmentation". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 801–818.

[42] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera. "Attention deeplabv3+: Multi-level context attention mechanism for skin lesion segmentation". In: *European conference on computer vision*. Springer. 2020, pp. 251–266.

[43] B. M. Priego-Torres, D. Sanchez-Morillo, M. A. Fernandez-Granero, and M. Garcia-Rojo. "Automatic segmentation of whole-slide H&E stained breast histopathology images using a deep convolutional neural network architecture". In: *Expert Systems With Applications* 151 (2020), p. 113387.

[44] H. Xu, Y. J. Cha, J. R. Clemenceau, J. Choi, S. H. Lee, J. Kang, and T. H. Hwang. "Spatial analysis of tumor-infiltrating lymphocytes in histological sections using deep learning techniques predicts survival in colorectal carcinoma". In: *The Journal of Pathology: Clinical Research* (2022).

[45] Y. Xie, J. Zhang, C. Shen, and Y. Xia. "Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2021, pp. 171–180.

[46] M. Mirza and S. Osindero. "Conditional generative adversarial nets". In: *arXiv preprint arXiv:1411.1784* (2014).

[47] M. Rezaei, K. Harmuth, W. Gierke, T. Kellermeier, M. Fischer, H. Yang, and C. Meinel. "A conditional adversarial network for semantic segmentation of brain tumor". In: *International MICCAI Brainlesion Workshop*. Springer. 2017, pp. 241–252.

[48] F. Mahmood, D. Borders, R. J. Chen, G. N. McKay, K. J. Salimian, A. Baras, and N. J. Durr. "Deep adversarial training for multi-organ nuclei segmentation in histopathology images". In: *IEEE transactions on medical imaging* 39.11 (2019), pp. 3257–3267.

[49] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. "Image-to-image translation with conditional adversarial networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1125–1134.

[50] H. Tsuda and K. Hotta. "Cell Image Segmentation by Integrating Pix2pixs for Each Class". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2019.

[51] D. Popescu, M. Deaconu, L. Ichim, and G. Stamatescu. "Retinal blood vessel segmentation using pix2pix gan". In: *2021 29th Mediterranean Conference on Control and Automation (MED)*. IEEE. 2021, pp. 1173–1178.

[52] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.

[53] M. Gadermayr, L. Gupta, V. Appel, P. Boor, B. M. Klinkhammer, and D. Merhof. "Generative adversarial networks for facilitating stain-independent supervised and unsupervised segmentation: a study on kidney histology". In: *IEEE transactions on medical imaging* 38.10 (2019), pp. 2293–2302.

[54] A. Kapil, T. Wiestler, S. Lanzmich, A. Silva, K. Steele, M. Rebelatto, G. Schmidt, and N. Brieu. "DASGAN–Joint Domain Adaptation and Segmentation for the Analysis of Epithelial Regions in Histopathology PD-L1 Images". In: *arXiv preprint arXiv:1906.11118* (2019).

[55] W. Li, J. Li, J. Polson, Z. Wang, W. Speier, and C. Arnold. "High resolution histopathology image generation and segmentation through adversarial training". In: *Medical Image Analysis* 75 (2022), p. 102251.

[56] F. Visin, M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci, and A. Courville. "Reseg: A recurrent neural network-based model for semantic segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2016, pp. 41–48.

[57] W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki. "Scene labeling with lstm recurrent neural networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3547–3555.

[58] A. Chakravarty and J. Sivaswamy. "RACE-net: a recurrent neural network for biomedical image segmentation". In: *IEEE journal of biomedical and health informatics* 23.3 (2018), pp. 1151–1162.

[59] M. Saha and C. Chakraborty. "Her2Net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation". In: *IEEE Transactions on Image Processing* 27.5 (2018), pp. 2189–2200.

[60] C. Nguyen, Z. Asad, R. Deng, and Y. Huo. "Evaluating transformer-based semantic segmentation networks for pathological image segmentation". In: *Medical Imaging 2022: Image Processing*. Vol. 12032. SPIE. 2022, pp. 942–947.

[61] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. "Swin transformer: Hierarchical vision transformer using shifted windows". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 10012–10022.

[62] Z. Qian, K. Li, M. Lai, E. I. Chang, B. Wei, Y. Fan, Y. Xu, et al. "Transformer based multiple instance learning for weakly supervised histopathology image segmentation". In: *arXiv preprint arXiv:2205.08878* (2022).

[63] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang. "Ds-transunet: Dual swin transformer u-net for medical image segmentation". In: *IEEE Transactions on Instrumentation and Measurement* (2022).

[64] R. Strudel, R. Garcia, I. Laptev, and C. Schmid. "Segmenter: Transformer for semantic segmentation". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 7262–7272.

[65] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou. "Transunet: Transformers make strong encoders for medical image segmentation". In: *arXiv preprint arXiv:2102.04306* (2021).

[66] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo. *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*. 2021. DOI: 10.48550/ARXIV.2105.15203. URL: https://arxiv.org/abs/2105.15203.

[67] E. L. Kaplan and P. Meier. "Nonparametric estimation from incomplete observations". In: *Journal of the American statistical association* 53.282 (1958), pp. 457–481.

[68] D. R. Cox. "Regression Models and Life-Tables". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2 (1972), pp. 187–202.

[69] Z. Kos, E. Roblin, R. S. Kim, S. Michiels, B. D. Gallas, W. Chen, K. K. van de Vijver, S. Goel, S. Adams, S. Demaria, et al. "Pitfalls in assessing stromal tumor infiltrating lymphocytes (sTILs) in breast cancer". In: *NPJ breast cancer* 6.1 (2020), pp. 1–16.

[70] M. Amgad, R. Salgado, and L. A. Cooper. "MuTILs: Explainable, multiresolution computational scoring of Tumor-Infiltrating Lymphocytes in breast carcinomas using clinical guidelines". In: *medRxiv* (2022).

[71] M. Amgad, L. A. Atteya, H. Hussein, K. H. Mohammed, E. Hafiz, M. A. Elsebaie, A. M. Alhusseiny, M. A. AlMoslemany, A. M. Elmatboly, P. A. Pappalardo, et al. "Nucls: A scalable crowdsourcing, deep learning approach and dataset for nucleus classification, localization and segmentation". In: *arXiv preprint arXiv:2102.09099* (2021).

[72] H. Le, R. Gupta, L. Hou, S. Abousamra, D. Fassler, L. Torre-Healy, R. A. Moffitt, T. Kurc, D. Samaras, R. Batiste, et al. "Utilizing automated breast cancer detection to identify spatial distributions of tumor-infiltrating lymphocytes in invasive breast cancer". In: *The American journal of pathology* 190.7 (2020), pp. 1491–1504.

[73] Y. Bai, K. Cole, S. Martinez-Morilla, F. S. Ahmed, J. Zugazagoitia, J. Staaf, A. Bosch, A. Ehinger, E. Nimeus, J. Hartman, et al. "An Open-Source, Automated Tumor-Infiltrating Lymphocyte Algorithm for Prognosis in Triple-Negative Breast CancerMachine-Read TIL Variables in TNBC". In: *Clinical Cancer Research* 27.20 (2021), pp. 5557–5565.

[74] A. Kapil, A. Meier, A. Shumilov, S. Haneder, H. Angell, and G. Schmidt. "Breast cancer patient stratification using domain adaptation based lymphocyte detection in HER2 stained tissue sections". In: (2021).

[75] J. Thagaard, E. S. Stovgaard, L. G. Vognsen, S. Hauberg, A. Dahl, T. Ebstrup, J. Doré, R. E. Vincentz, R. K. Jepsen, A. Roslind, et al. "Automated quantification of stil density with h&e-based digital image analysis has prognostic potential in triple-negative breast cancers". In: *Cancers* 13.12 (2021), p. 3050.

[76] P. Sun, J. He, X. Chao, K. Chen, Y. Xu, Q. Huang, J. Yun, M. Li, R. Luo, J. Kuang, et al. "A computational tumor-infiltrating lymphocyte assessment method comparable with visual reporting guidelines for triple-negative breast cancer". In: *EBioMedicine* 70 (2021), p. 103492.

[77] K. E. Craven, Y. Gökmen-Polar, and S. S. Badve. "CIBERSORT analysis of TCGA and METABRIC identifies subgroups with better outcomes in triple negative breast cancer". In: *Scientific reports* 11.1 (2021), pp. 1–19.

[78] D. J. Fassler, L. A. Torre-Healy, R. Gupta, A. M. Hamilton, S. Kobayashi, S. C. Van Alsten, Y. Zhang, T. Kurc, R. A. Moffitt, M. A. Troester, et al. "Spatial Characterization of Tumor-Infiltrating Lymphocytes and Breast Cancer Progression". In: *Cancers* 14.9 (2022), p. 2148.

[79] S. Meng, L. Li, M. Zhou, W. Jiang, H. Niu, and K. Yang. "Distribution and prognostic value of tumor-infiltrating T cells in breast cancer". In: *Molecular medicine reports* 18.5 (2018), pp. 4247–4258.

[80] V. Kotoula, K. Chatzopoulos, S. Lakis, Z. Alexopoulou, E. Timotheadou, F. Zagouri, G. Pentheroudakis, H. Gogas, E. Galani, I. Efstratiou, et al. "Tumors with high-density tumor infiltrating lymphocytes constitute a favorable entity in breast cancer: a pooled analysis of four prospective adjuvant trials". In: *Oncotarget* 7.4 (2016), p. 5074.

[81] J. Saltz, R. Gupta, L. Hou, T. Kurc, P. Singh, V. Nguyen, D. Samaras, K. R. Shroyer, T. Zhao, R. Batiste, et al. "Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images". In: *Cell reports* 23.1 (2018), pp. 181–193.

[82] 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.

[83] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. "Attention is all you need". In: *Advances in neural information processing systems* 30 (2017).

[84] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. "An image is worth 16x16 words: Transformers for image recognition at scale". In: *arXiv preprint arXiv:2010.11929* (2020).

[85] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo. "SegFormer: Simple and efficient design for semantic segmentation with transformers". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 12077–12090.

[86] T. G. Clark, M. J. Bradburn, S. B. Love, and D. G. Altman. "Survival analysis part I: basic concepts and first analyses". In: *British journal of cancer* 89.2 (2003), pp. 232–238.

[87] Y.-C. Chen. "Lecture 5: Survival Analysis". In: *STAT 425: Introduction to Nonparametric Statistics, University of Washington* (Winter 2018).

[88] T. Therneau and E. Atkinson. "1 The concordance statistic". In: (2020).