SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY - 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

SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY - 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

I confirm that this master's thesis in biomedical computing is my own work and I have documented all sources and material used.

Munich, 15.12.2022 Margaryta Olenchuk

Contents

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 prime factors to characterize breast cancer. There are four main clinically used [2] groups defined based on the status of several receptors, namely the Hormonal Receptor (HR, which is positive if either Estrogen Receptor (ER) or Progesterone Receptor (PR) is positive) and the 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, breast cancer is primarily classified by its histological appearance. Thus for diagnostic confirmation a patient's biopsy or surgical resection samples are sectioned onto microscope slides for staining, often with hematoxylin and eosin (H&E), followed by a visual diagnosis by a pathologist. Pathologists examine tissue for abnormalities that indicate breast cancer. Cancer causes changes in tissue at the sub-cellular scale, hence an analysis of normal and tumor tissue can provide novel insights into tissue characteristics, lead to a better understanding of mechanisms underlying cancer progression and provide valuable information for medical decision-making such as tumor grading and treatment choices. [3]

One of the characteristics of histological images that can be visually assessed by pathologists is lymphocytic infiltration. There are a number of publications that emphasize the prognostic value of tumor-infiltrating lymphocytes (TILs), especially in triple negative (TNBC) and human epidermal growth factor receptor 2 (HER2+) breast cancer [4, 5]. TILs are mononuclear immune cells that infiltrate tumor tissue. They 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 between the cells in the tumor microenvironment and TILs. In the early stage HER2+ and TNBC, TILs are detectable in up to 75% of tumors. [7] Studies have shown that an increased degree of lymphocytic infiltration is predictive of better long-term control of the disease. 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. Accumulating evidence indicates that tumor-infiltrating lymphocytes are clinically useful biomarkers in TNBC and HER2+ and that they play an essential role in cancer progression. [8] Further research and development of TILs related biomarkers would grant clinicians essential prognostic information and promote the research on novel treatments and therapeutics.

For instance, since TILs with exhausted phenotype are 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 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 one of the highest of all breast cancer subgroups [11] which makes TNBC a valid target for further TILs research.

A valuable contribution to TILs research and any task involving visual analysis of histological images would be method automatization. Because while the manual examination continues to be widely applied in a clinical setting, it is subjective and not scalable to translational and clinical research studies involving large datasets of high-resolution whole slide tissue images (WSIs). Hence, there is a raised demand for reliable and efficient automated methods to complement the traditional manual examination of tissue samples.

With advancing technology and access to a large amount of data, deep learning methods have garnered an interest in computational pathology. There are multiple deep learning-based methods and pipelines that have been proposed for detection and segmentation tasks of WSIs. To stimulate the development of algorithms for automatic TILs evaluation, a special Tumor InfiltratinG lymphocytes in breast cancER (TiGER) [12] challenge was formed. Within this competition, various algorithms were evaluated for the automated assessment of TILs in H&E stained histopathology WSIs that resulted in automatically acquired TILs scores. Those were later internally checked for significance as prognostic values and the concordnace was reported. The clinical focus of the TiGER challenge is on Her2+ and TNBC. It is motivated by research and clinical data that show that Her2+ and TNBC have the worst prognosis making them an intense target of prognostic and predictive biomarker research aimed at improving patient management and prognosis.

This work is closely linked to the TiGER challenge. The goal is to develop a pipeline for HER2 positive and triple negative breast cancer H&E slides that segments tumor and stroma regions, detects TILs, and produces TILs scores as pictured in Figure 1.1 block 1-3.

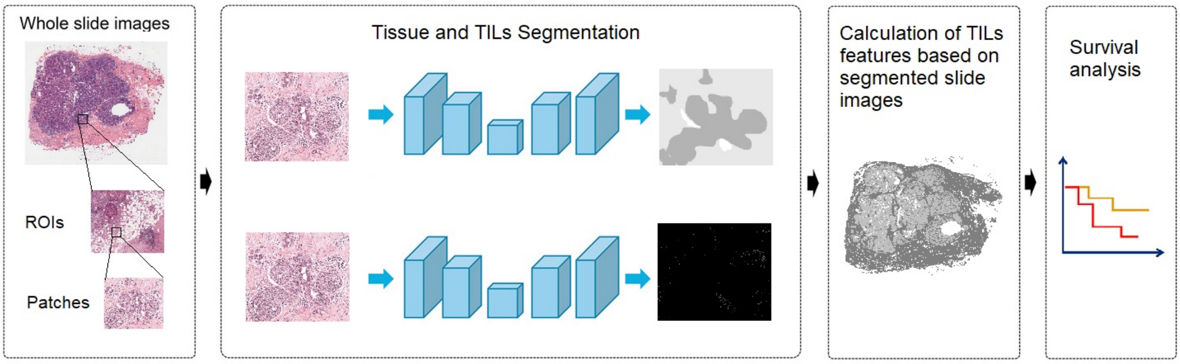

Whole slide images
ROIs
Patches
Tissue and TILs Segmentation
Calculation of TILs features based on segmented slide images
Survival analysis

Figure 1.1: Abstract scheme to visually introduce the flow of this thesis work.

This work takes benefit of the annotated ROIs provided by the TiGER challenge for the development of patch-based automated tissue and TILs segmentation. As a step beyond the challenge, the scores based on the degree of lymphocytic infiltration are evaluated on the breast cancer TCGA-BRCA dataset generated by the TCGA Research Network and not on the hidden TiGER dataset which is not available after the end of the competition. The TCGA-BRCA clinical data enables independent broad survival analysis of different experimental TILs characteristics that can be calculated solely based on histological images (Figure 1.1 block 4). As a result, this work aims not only to develop a computational approach to compute TILs score on H&E images of Her2+ and TNBC but also experiment with different TILs scores and show detailed survival analysis based on publically available TCGA-BRCA dataset together with their predictive value for overall patient survival.

2 Related work

2.1 Deep learning-based semantic segmentation

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]. As tissue segmentation, TILs detection can also be viewed as a semantic segmentation problem, since detection bounding boxes can be transformed into pseudo-segmentation masks. The focus of the following chapters is to superficially introduce the existing deep learning-based approaches for the histopathological tasks, that can be adapted or extended for tissue and TILs segmentation of breast cancer WSIs.

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. Such models are used on segmentation problems in histology domain such as colon glands segmentation [18], as well as nuclei [19] and TILs [20] 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 [21] breast data set [22], 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 [23]. 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 architecture also find a use in histopathology, e.g. colon cancer analysis [24]. There are several encoder-decoder models initially developed for biomedical image segmentation. Ronneberger et al. [25] 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 [25], U-Net was applied for numerous histopathology tasks such as nuclei segmentation [26, 27], individual colon glands segmentation [28], epidermal tissue segmentation of skin biopsies [29] and cell segmentation on triple-negative breast cancer patients dataset [30]. A further development of an encoder-decoder model for semantic segmentation of histopathology images is HookNet [31]. 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 breast cancer tissue sections [31].

Another widely used group of deep learning models for semantic segmentation are the atrous (or dilated) convolutional models that include the DeepLab family [32, 33]. 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+ [34] 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 [34] demonstrated the effectiveness of DeepLabv3+ model on segmentation of H&E stained breast cancer [35]. 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 [36] that may result in the sub-optimal segmentation of complex structures.

2.1.3 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 [37] 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 up-sampling layers to recover the predictions at original image resolution. But despite all further developments that showcase the potential for histopathology image segmentation: RACE-net [38] applied for segmentation of the cell nuclei in H&E stained breast cancer slides, Her2Net [39] 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.

2.1.4 Transformers

The Transformer in Natural Language Processing is an architecture that aims to solve sequence-to-sequence problems. 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 [40]. 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 [41] 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, often as a backbone, for example for colon cancer segmentation in H&E stained histopathology images [42] or gland segmentation [43]. A further popular fully transformer-based model for semantic segmentation is Segmenter [44]. 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 [44], is not particularly used in the medical domain. In the field of medical image segmentation, TransUNet [45] 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 [46] set new state-of-the-art in terms of efficiency and accuracy in publicly available semantic segmentation datasets and applied for gland and nuclei segmentation [43]. This architecture remains promising also for semantic segmentation in medical applications due to the positional-encoding-free encoder and lightweight MLP decoder.

2.2 TILs as prognostic biomarker

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 [47] to estimate OS and Cox proportional hazard models [48] to quantify the hazard ratio (HR) for the effects of biomarker groups. The following chapter focuses on conducted research for the development of TILs scores as a prognostic biomarker for survival analysis in breast cancer based solely in histological slides.

Amgad, M. et al. [49, 50]. 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 [22, 51] (the source datasets for TCGA part of TiGER dataset) 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 [52] 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. [53] 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 [49]. sTIL density was found significantly prognostic for OS not only while applied on H&E slides but IHC as well. [54] Applied on the TCGA-BRCA mixed with non publicaly available dataset, Thagaard, J. et al. [55] 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. [56] 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 [57]. On the other hand, Fassler, D. J., et al. [58] 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 [59]. Furthermore, studies found a three-scale grading system for reporting TILs status to be applicable, instead of continuous or binary grouped TILs densities [60]. 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 [61]. Hence, there is no canonic method for the automatic determination of TILs score based on the H&E breast cancer tissue samples but number of TILs per mm^2 of stromal area is used most frequently.

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 [62]).

One of the latest models in this family, DeepLabv3 [33], 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.

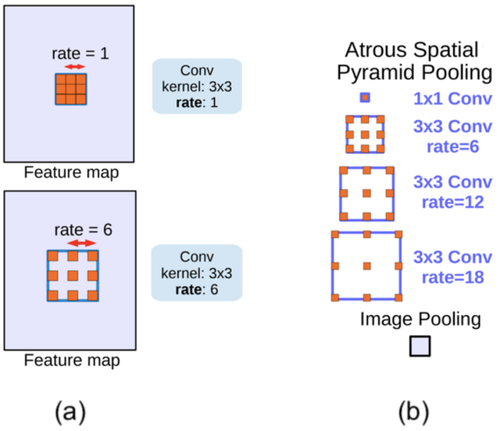
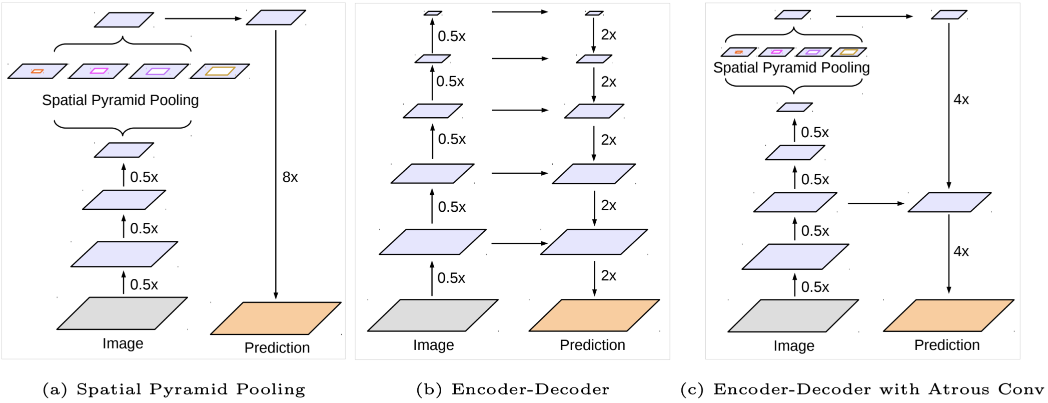


Figure 3.1: (a) Atrous convolution, (b) ASPP augmented with Image Pooling (or Image-level features) [33]

DeepLabv3+

The reproduction of shape contours during semantic image segmentation remained difficult with DeepLabv3 [34]. 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+ [34] 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) [34]

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.

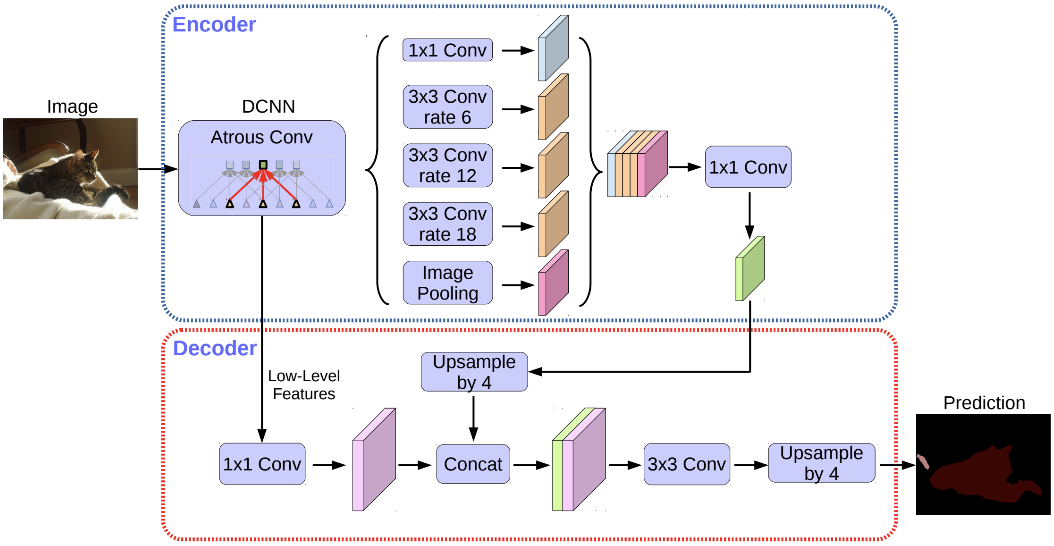


Figure 3.3: DeepLabv3+ architecture. DeepLabv3 as encoder and proposed decoder structure for semantic image segmentation. [34]

3.1.2 Transformers

Transformers [63] 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.

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.

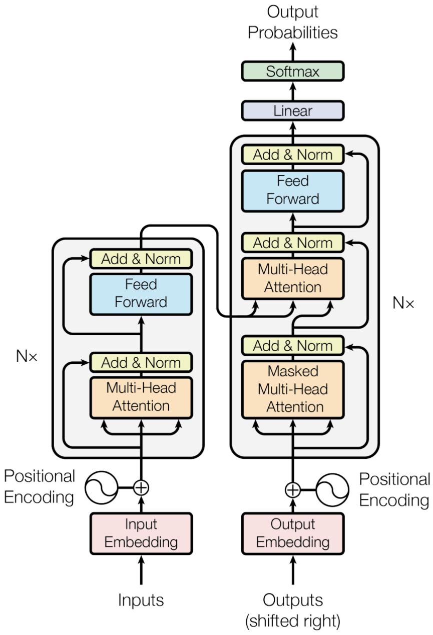


Figure 3.4: Transformer model architecture. [63]

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.

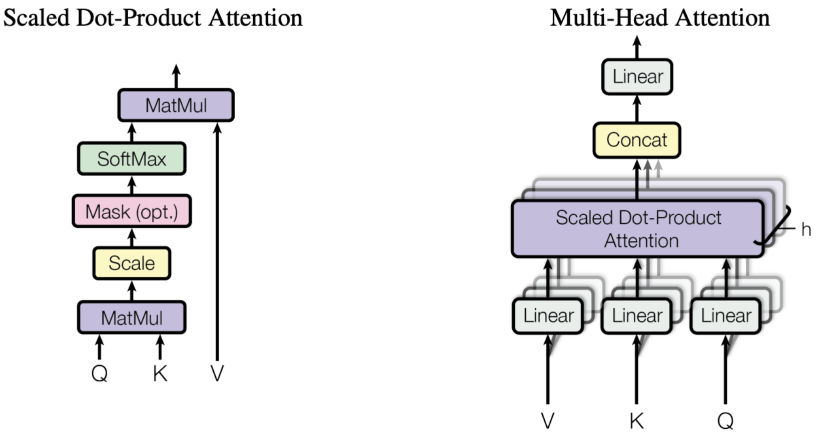


Figure 3.5: Scaled Dot-Product Attention (left). Multi-Head Attention consists of several attention layers running in parallel (right). [63]

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) [64] 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 [46] 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.

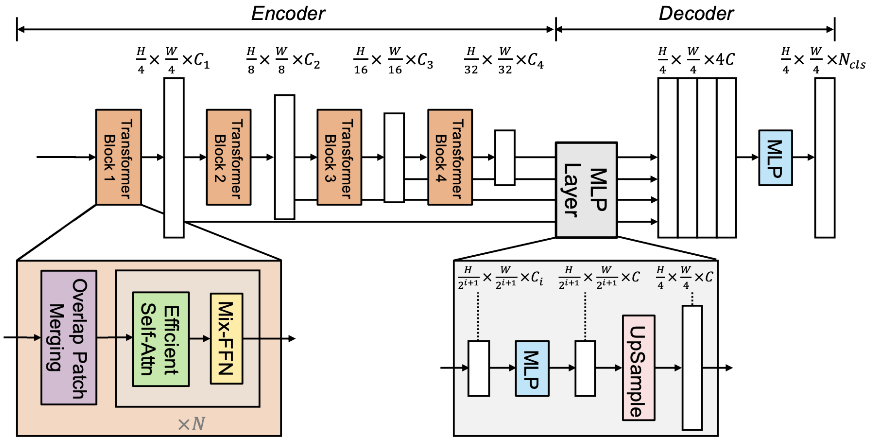

Encoder
Decoder
$\frac{H}{4}\times\frac{W}{4}\times C_1$
$\frac{H}{8}\times\frac{W}{8}\times C_2$
$\frac{H}{16}\times\frac{W}{16}\times C_3$
$\frac{H}{32}\times\frac{W}{32}\times C_4$
$\frac{H}{4}\times\frac{W}{4}\times 4C$
$\frac{H}{4}\times\frac{W}{4}\times N_{cls}$
Transformer Block 1
Transformer Block 2
Transformer Block 3
Transformer Block 4
MLP Layer
MLP
Overlap Patch Merging
Efficient Self-Attn
Mix-FFN
×N
$\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}\times C_i$
$\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}\times C$
$\frac{H}{4}\times\frac{W}{4}\times C$
MLP
UpSample

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 [46] 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 [46] 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 $\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}} \times \mathrm{C}_i$ 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.

3.1.3 TILs segmentation postprocessing

Non-maximum Suppression (NMS)

Non-maximum Suppression is a class of algorithms to select one entity out of many overlapping entities. In terms of detection models, used to find the best-fitting bounding box out of all predicted bounding boxes for the same object [65]. Each proposal comes with a confidence score. One by one the bounding box with the highest score is selected to keep and compared to all bounding boxes by calculating Intersection Over Union (IoU). If a comparison scores the IoU higher than some defined threshold, those bounding boxes out of the pool are eliminated. The process is repeated by picking the highest confidence bounding box out of the remaining pool of proposals until there are any.

In terms of TILs segmentation, there are no proposed bounding boxes and no scores. Nonetheless, to obtain good centroids for each TILs prediction, the segmentation posteriors need to be condensed to local maxima. The NMS application will in this case appropriately filter the segmentation posteriors with a kappa threshold and the segmentation borders refined to get a clear centroid annotation with a defined kernel size used for dilation.

Hungarian algorithm

To evaluate the match of predicted TILs and ground truth annotations, it is not enough to compare the coordinates. Each TIL should be annotated with a point annotation of one pixel, it is aimed to be placed in the center but really can be placed anywhere within a TIL area. The goal is to find the pairs of coordinates that indicate the same TIL object. Matching between the ground truth set and predictions can be solved as an assignment problem with the Hungarian algorithm [66]. It is a combinatorial optimization algorithm that solves the problem in polynomial time complexity. The distance matrix $n \times m$ between all annotated (n) versus predicted (m) TILs is the cost matrix of each of the prediction coordinates to match any of the ground truth coordinates. The goal is to assign the ground truth coordinates to the prediction coordinates to minimize the total distance. The algorithm can be introduced as step-by-step instruction:

- Step 1: For each row, subtract the smallest element of the row from each of its elements.
- Step 2: For each column, subtract the smallest element of the column from each of its elements.
- Step 3: Cover all zeros with a minimum number of lines

  In the resulting matrix cover all zeros using a minimum number of horizontal and vertical lines. If there are n lines, Step 5.
- Step 4: Find the smallest element that is not covered by a line in Step 3. Subtract it from all uncovered elements, and add it to all elements that are covered twice. Step 3.
- Step 5: Assignment pairs are indicated by the positions of the zeros in the cost matrix.

As a result, the assignment pairs include the matching of ground truth TILs coordinates to the prediction coordinates that minimize the total distance. If desired, the pairs can be filtered by minimal allowed distance. This work uses the `scipy.optimize.linear_sum_assignment` implementation of the Hungarian algorithm.

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. [67] 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 survival 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 [68] 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. [67]

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. [68] 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. [69] 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.

List of Figures

List of Tables

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] R. Salgado, C. Denkert, S. Demaria, N. Sirtaine, F. Klauschen, G. Pruneri, S. Wienert, G. Van den Eynden, F. L. Baehner, F. Pénault-Llorca, et al. “The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014”. In: *Annals of oncology* 26.2 (2015), pp. 259–271.

[5] C. Denkert, G. von Minckwitz, S. Darb-Esfahani, B. Lederer, B. I. Heppner, K. E. Weber, J. Budczies, J. Huober, F. Klauschen, J. Furlanetto, et al. “Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy”. In: *The lancet oncology* 19.1 (2018), pp. 40–50.

[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] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

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

[29] 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.

[30] 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.

[31] 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.

[32] 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.

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

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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).

[43] 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).

[44] 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.

[45] 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).

[46] 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.

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

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

[49] 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.

[50] 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).

[51] 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).

[52] 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.

[53] 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.

[54] 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).

[55] 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.

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] 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.

[61] 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.

[62] 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.

[63] 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).

[64] 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).

[65] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis. "Soft-NMS–improving object detection with one line of code". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 5561–5569.

[66] H. W. Kuhn. "The Hungarian method for the assignment problem". In: *Naval research logistics quarterly* 2.1-2 (1955), pp. 83–97.

[67] 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.

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

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