

Deep learning for colon cancer histopathological images analysis

A Ben Hamida^{a,*}, M Devanne^b, J Weber^b, C Truntzer^c, V Derangère^c, F Ghiringhelli^c, G Forestier^b and C Wemmert^a

^aICube, University of Strasbourg, France

^bIRIMAS, University of Haute-Alsace, France

^cPlatform of Transform in Biological Oncology, Dijon, France

ARTICLE INFO

Keywords:

Digital pathology
Whole Slide Images
Colon cancer
Tumour Segmentation
Deep learning
Artificial Intelligence
Transfer Learning
Semantic Segmentation

ABSTRACT


Nowadays, digital pathology plays a major role in the diagnosis and prognosis of tumours. Unfortunately, existing methods remain limited when faced with the high resolution and size of *Whole Slide Images* (WSIs) coupled with the lack of richly annotated datasets. Regarding the ability of the *Deep Learning* (DL) methods to cope with the large scale applications, such models seem like an appealing solution for tissue classification and segmentation in histopathological images. This paper focuses on the use of DL architectures to classify and highlight colon cancer regions in a sparsely annotated histopathological data context. First, we review and compare state-of-the-art *Convolutional Neural networks* (CNN) including the ALEXNET, VGG, RESNET, DENSENET and INCEPTION models. To cope with the shortage of rich WSI datasets, we have resorted to the use of transfer learning techniques. This strategy comes with the hallmark of relying on a large size computer vision dataset (IMAGENET) to train the network and generate a rich collection of learnt features. The testing and evaluation of such models on our AiCOLO colon cancer dataset ensure accurate patch-level classification results reaching up to 96.98% accuracy rate with RESNET. The CNN models have also been tested and evaluated with the CRC-5000, NCT-CRC-HE-100K and merged datasets. RESNET respectively achieves 96.77%, 99.76% and 99.98% for the three publicly available datasets. Then, we present a pixel-wise segmentation strategy for colon cancer WSIs through the use of both UNET and SEGNET models. We introduce a multi-step training strategy as a remedy for the sparse annotation of histopathological images. UNET and SEGNET are used and tested in different training scenarios including data augmentation and transfer learning and ensure up to 76.18% and 81.22% accuracy rates. Besides, we test our training strategy and models on the CRC-5000, NCT-CRC-HE-100K and WARWICK datasets. Respective accuracy rates of 98.66%, 99.12% and 78.39% were achieved by SEGNET. Finally, we analyze the existing models to discover the most suitable network and the most effective training strategy for our colon tumour segmentation case study.

1. Introduction

Recently, digital pathology has emerged as a key tool for the diagnosis and the prognosis of tumours [1]. Although such concept is no subject of modernity, the lack of available resources has hindered its evolution for years. Therefore, recent technological progresses have tremendously contributed to the proliferation of digital pathology in different applications. Unlike classical glass slides, the novel *Whole Slide Images* (WSI) are numerical copies of stained specimen samples [2]. These images are also playing a major role in the pathological diagnosis process [3, 4, 5] since they enable easy data sharing and storing. In fact, the analysis of WSI provides pathologists with a thorough insight into the data content and enables accurate diagnosis of tumours and cancer sub-types. Over the last few years, a wide range of approaches have been deployed for WSI segmentation and classification in such context. Most of these trials focused on learning shallow features, namely texture and pattern recognition as in [6, 7], gray level co-occurrence matrix [8] or local binary pattern [9]. However, each tumour simultaneously encompasses different texture, shape or color distributions. Therefore, the previously mentioned techniques are often limited when coping with the challenges introduced by recent WSI [10]. As a matter of fact, today's histopathological images are extremely large, including up to billions of pixels. These slides usually introduce high level complex clinical features. Moreover, most of the time,

* This work was supported by the AiCOLO project funded by INSERM/Plan Cancer.

*Corresponding author

 aminabenhameda.bh@gmail.com (A. Ben Hamida)

ORCID(s):

they only represent few available annotated regions. Actually, the manual labeling of each WSI is a time consuming task which requires serious efforts and a great deal of commitment.

The particularities of histopathological images have catalyzed efforts to establish novel automated image analysis methodologies. Such situation can alleviate the workload of pathologists, orchestrate the clinical applications and reduce the processing and handling time. As a matter of fact, AI models have progressively migrated from the expert systems to traditional *Machine Learning* (ML) and eventually to *Deep Learning* (DL). In other words, data analysis tasks were highly dependent from experts knowledge in order to define the key features in the classical hand crafted approaches. Whereas, the current DL networks are capable of automatically learning features from the data itself in an accurate and easier manner [11]. That goes without saying that the arise of powerful computational resources has deflected the interest toward DL models in a wide range of medical image analysis applications. Currently, an important share of the digital pathology research field is dedicated to the exploration and establishment of novel DL models that will enable thorough interpretation, analysis and extraction of pertinent information from WSI.

In this paper, a general overview of the recent digital pathology applications is presented. The main challenges that are disrupting its progress are also presented with a special focus on DL methods for WSIs classification and segmentation. The paper shines the light on the colon cancer WSIs analysis tasks. Therefore, different DL models are introduced and used for colon cancer histopathological image segmentation in a sparsely annotated data scenario. First, we rely on classical CNN models namely ALEXNET [12], VGG [13], RESNET [14], DENSENET [15] and INCEPTION [16] to study the patch-level colon cancer WSIs classification. Then, a pixel-level histopathological image segmentation is executed to highlight colon cancer tissues through the use of both the UNET [17] and SEGNET [18] models. The paper also examines and compares the transfer learning strategies and the from scratch-training. Moreover, both the patch and pixel-level classification models have been tested and evaluated with publicly available WSI datasets namely the CRC-5000, NCT-CRC-HE-100K and WARWICK. The results are compared with state-of-the-art methods. Finally, an evaluation of the different DL techniques for tumour segmentation in colon cancer histopathological images is presented.

2. Deep Learning for Histopathological Image Analysis: An Overview

Until recently, most of the used techniques in medical image analysis did rely on classical ML methods. The workflow of such models usually encompasses different steps including data preprocessing and preparation, *Regions Of Interest* (ROI) selection, feature extraction and eventually feature classification using linear and non-linear models. For example, in the case of colon cancer segmentation, Demir *et al.* used an object graph based approach for the segmentation of colon WSI in [19] while researchers Nayak *et al.* relied on a variation of the Restricted Boltzmann Machine (RBM) in order to learn the key features of the image signature in [20]. Although these methods provide high accuracy rates for colon cancer detection, they still are limited in generalizing their performance levels to different applications and datasets as detailed in [10]. During the last few years, DL models have enabled automated analysis of large scale images unlike the multi-step classical ML techniques. The introduction of DL in histopathology was initially inspired from its success in image analysis and biomedical tasks. One of the first serious trials to deploy DL architectures in a WSI processing context were conducted by Ciresan *et al.* [21]. The paper introduces a deep max-pooling *Convolutional Neural Network* (CNN) to detect mitosis in breast WSI. In fact, CNN are the most popular neural networks for medical imaging analysis. They have played a major role in revolutionizing for example breast, lung and brain cancer detection through the classification and segmentation of histopathological images [22, 23, 24].

Applications to colon cancer Therefore, DL-based approaches are now a gold standard for different colon cancer related applications namely gland segmentation, tumour micro-environment and cells characterization, prognosis prediction and tumour detection, classification and segmentation. A summary of the Gland Segmentation in Colon Histology Images (GlaS) Challenge Contest is presented in [25] where the winning method introduces multi-path CNN. Each path has its own set of convolutional layers and trained to extract and learn features from different views in a local-global fashion. Another method that uses multi-channel CNN is presented in [26] where a foreground segmentation channel is associated with both edge and object detection channels. The segmented outputs are then fused in a convolutional network style in order to obtain accurate colon gland segmentation maps. Such technique was also used for the colon tumour cells classification as depicted in [27] where a CNN is trained in a supervised manner to detect foreground tissues and classify different cells. Tumour cells identification also helps in prognosis and survival scoring. For that purpose, the authors in [28] relied on a 19-layer CNN (VGG-19). In [29], a residual 18-layer CNN

(RESNET18) is used. Long-short Term Memory networks (LSTM) applied on CNN features are also utilized in the same context as detailed in [30]. One of the most depicted uses of CNN in digital pathology is the segmentation and classification of tumour WSI. The deployed techniques range from few classical CNN as the one examined in [31] to more recent models as in [32]. In [31], the authors resorted to a 5-layer CNN to classify tissues into 3 different classes. The same network was used in [33] for the detection of tumour through the classification of preneoplastic and neoplastic lesions. Several reviews have been recently published to discuss DL for histopathological image analysis including [34] and [35]. In [36], the authors represent a comprehensive review of DL applications for colon cancer.

Limitations In [33], the authors shine the spotlight on the importance of richly annotated datasets for tumour segmentation tasks. In fact, classical CNN suffer from gradient vanishing problems which limit their ability to provide generic transferable pathological data representations. Consequently, different enhancement techniques were introduced namely the *Recurrent Neural network* (RNN) and the Inception models [37, 28, 38]. However, the use of large scale images combined with large datasets generates high computational costs.

3. Deep Learning for Histopathological Image Analysis: The Challenges

Regarding the importance of early detection of tumour in colon tissues, fast and accurate classification of the WSI is a crucial step for an early diagnosis of cancers. In order to alleviate the work of pathologists, the segmentation and classification techniques need to perform well in the absence of richly annotated datasets. In what follows, we enumerate and briefly explain the most important obstacles that restrain the evolution of DL for colon cancer WSI classification and segmentation.

Large-Scale images A WSI is a digital scan of a glass slide at a very high resolution. Each WSI is acquired at different magnification levels where a $40\times$ magnification corresponds to $0.25\ \mu\text{m}/\text{pixel}$. Therefore, the size of classical histopathological slides is around $100,000 \times 100,000$ pixels, resulting in gigapixel images. Processing this type of images is often constrained by computer memory and requires huge computational resources. Therefore, most of the used DL models are fed with small WSI tiles. These tiles are small patches extracted from the WSI to ensure effective low-cost learning. An important share of the recent digital pathology literature focuses on ways to benefit from the high resolution images without exploding the computational cost as depicted in [39].

Images artifacts In digital pathology, the trained algorithm needs to perform well for a variety of pathologies and patient populations and expand over WSI artifacts and color variability in staining [40]. In fact, each histopathological image requires a multi-step workflow for acquisition, digitization and processing. Consequently, different image artifacts can appear throughout the whole process including uneven hue and illumination, resolution and focus problems, cutting and tears in slides and sometimes the presence of foreign objects. Color and staining related artifacts can also appear in histopathological slides as seen in Figure 1. The presence of such defects in the images can tremendously distort the extracted data and mislead the tumour diagnosis process. As a remedy to this problem, Two main strategies have been introduced : Image pre-processing as detailed in [41] and data augmentation techniques [42].

Lack of annotated samples When dealing with DL models in a histopathological context, the training data must be richly annotated. Otherwise stated, pathologists need to highlight the ROI and specify the tissue types presented in each WSI according to the desired application. However, the elaboration of detailed annotations for large amounts of data is a hard task when pathologists deal with images at different resolutions and colorization techniques which adds ambiguity to the features of the dataset. The available WSI datasets suffer from limited samples with unbalanced classes and samples for each application. As a matter of fact, many of the recent research relied on data with restricted access as seen in [36]. The limited datasets used for AI models training usually result in restricted utility and poor generalization. One of the key solution is to take benefit from the pre-trained deep models for other computer vision applications. In other words, transfer the already acquired knowledge from different datasets to the WSIs context. This procedure appears in many state-of-the-art methods [43, 44, 45].

4. Deep learning for Colon Histopathological Image Analysis: Methodology

In this section we present two main colon WSI analysis applications. First, we consider patch-level image classification into eight different tissue classes including tumour, stroma, tissue, necrosis, immune, fat, background and trash.

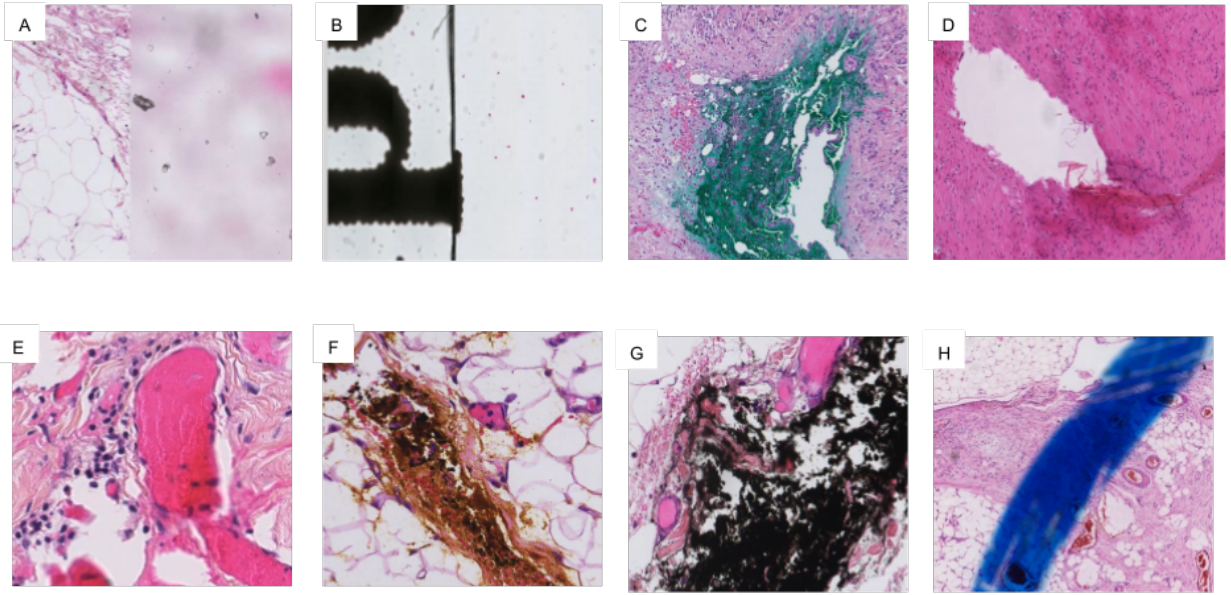


Figure 1: (A-H) Examples of artifacts from our HES stained dataset. (A) Out of focus region.(B) and (H) Typing/Pen marks. (C), (F) and (G)Pigment artifacts. (D) and (E) Tear /cutting in Tissues

	Key Feature	Nbr of Layers	Nbr of params	Size	Time
ALEXNET[12]	Deep	8	62M	233MB	12
VGG-16 [13]	Fixed size Kernels	16	138M	527MB	16
RESNET[14]	Shortcut Connections	18	12M	44MB	6
DENSENET[15]	Dense connections	161	8M	33MB	8
INCEPTIONV3 [16]	Wide/Parallel kernels	48	27M	103MB	10

Table 1

Summary of the studied architectures and their corresponding computational costs. Time corresponds to the tuning time in minutes.

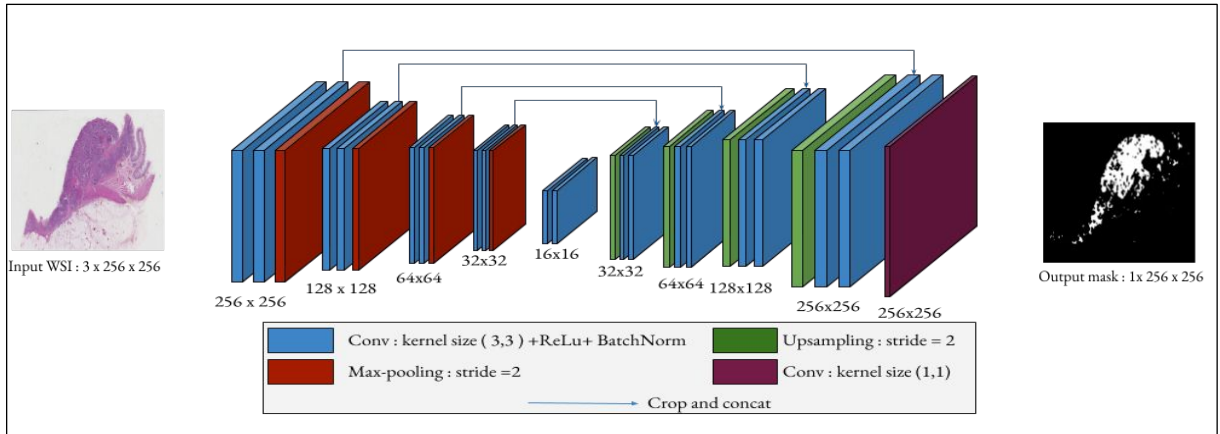
For this purpose we have used and assessed different DL architectures in a pre-trained and a from scratch strategies. Then, a pixel-wise segmentation of colon WSI is introduced. Two deep models are deployed and compared in order to segment and highlight the presence of tumour tissues in our colon WSI dataset.

4.1. Patch-level classification of colon cancer histopathological images

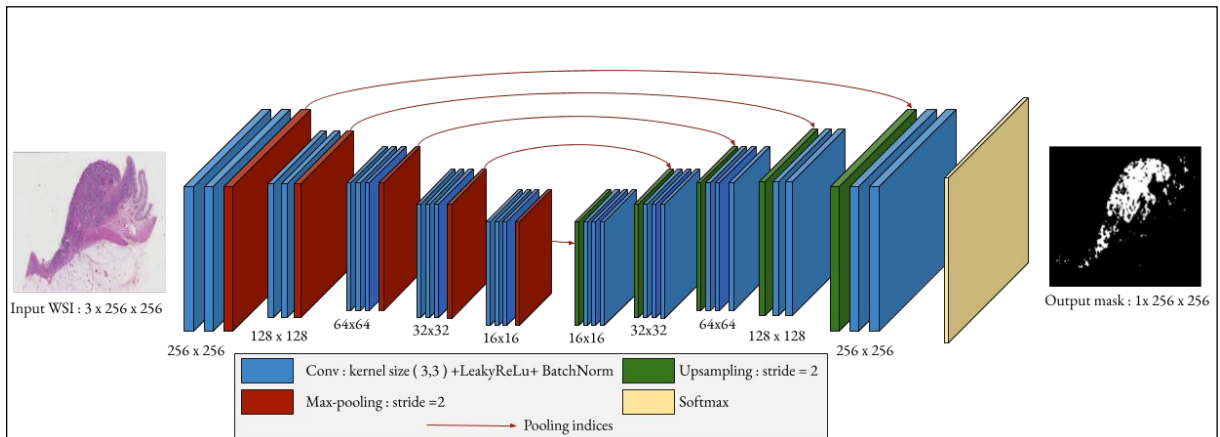
Here, we resort to state-of-the-art methods in computer vision applications namely image classification tasks. Recently, a wide range of DL architectures have been developed ranging from large basic models to light more enhanced networks. ALEXNET, VGG, DENSENET, RESNET and INCEPTION networks are selected and deployed to classify WSI patches into different semantic classes as detailed in Table 1. The last fully connected layer in the original network include 1,000 neurons with a softmax classifier. To cope with our 8-class AiCOLO dataset content, the last layer of each model is replaced by a fully connected layer of size 8 (the number of classes in our dataset) with a softmax classifier. At each training, we adjust the size of the last layer to fit with the number of classes of each dataset including the CRC-5000 and NCT-CRC-HE-100K.

4.2. Pixel-level segmentation of colon cancer histopathological images

Semantic segmentation is currently one of the most crucial tasks in the field of digital pathology towards complete histopathological slides understanding. Here, we focus on DL approaches for the segmentation of tumorous tissues in WSI. The purpose is to generate a binary mask that highlights the colon cancer regions in the slides. Therefore, we



(a) Architecture of the used UNET



(b) Architecture of the used SEGNET

Figure 2: Architectures of the used UNET and SEGNET models.

resort to state-of-the-art methods SEGNET and UNET as deep semantic segmentation networks.

SEGNET This architecture has been introduced in [18] and relies on the idea of "reflecting" a Fully Connected Network (FCN). In other words, creating the mirror of each layer and each kernel to perform an up-sampling of the down-sampling FCN output. Consequently, deconvolutional layers substitute the convolutional ones while the classical pooling layers are reflected by unpooling layers. In order to cope with the high dimensionality of WSI, both the encoder and the decoder are inspired from the VGG-16 network as presented in [46]. The first part of the network or the so-called "encoder" uses the 13 first layers of the VGG-16. It alternates two different blocks of convolutional layers: A set of 2 convolutional layers and another set of 3 convolutions where both rely on 2D filters of size 3×3 . The deployment of convolutions with different filters generates a set of representation maps which undergo a batch normalization step along with a ReLU non-linearity function $f(x) = \max(0, x)$. For sub-sampling purposes, max-pooling layers are inserted in the network with non-overlapping 2×2 filters. This technique adds robustness to the network allowing invariance to minor spatial shifts. Despite its robust classification abilities, max-pooling and sub-sampling degrade the resolution of the classification map. The decoder is the "mirrored" version to the encoder. The main difference is the substitution of the classical convolutions and poolings with sparse $2 \times$ up-sampling. The up-sampling produces sparse feature maps using the memorized max-pooling indexes from the respective encoder feature maps. The final decoder output is fed to a multi-class softmax classifier to pixel wise class probabilities as seen in Figure 2.

UNET A histopathological image can contain different types of tissues including tumorous and healthy epithelium regions. The presence of different detailed patterns in the slide hardens the task of detecting and highlighting the tumour. In order to cope with the different level of features representation presented in WSI, we resort to the use of skip architectures as proposed by Long *et al.* [47]. One of the main networks that relies on skip connections for image segmentation is the UNET. As detailed in [17], the UNET model combines the high-level representation from deep decoding layers with representations from shallow encoding layers to generate detailed segmentation maps. The original UNET model relies on the same encoder-decoder duality as in the SEgNET network. Our UNET model is inspired from the original architecture which consists of 5 convolutional blocks for each path with 2 convolutional layers for each block. Each branch is composed of 5 unities where each unity encompasses 2 convolutional layers with kernel size of 3×3 , stride of 1×1 without padding. Each convolutional layer is associated to a ReLU activation, a batch normalization and a max-pooling to reduce the size of the feature maps. Whereas, the up-sampling path inserts a deconvolutional layer with filter size of 3×3 and a stride of size 2×2 in order to increase the number of the output features maps while decreasing their size by a 2 factor. Unlike the original UNET model, our proposed architecture presented in Figure 2 takes as input 3 channel images and generates binary masks of the same size. Therefore, we use zero padding in our convolutional layers of both the decoding and encoding paths. Each block begins with a convolution using a stride of size 2×1 to halve the spatial size of the input. Finally, a Convolutional layer of kernel size 1×1 allows to output binary masks with 2 classes. No fully connected layer is invoked in the network. Along the network, we compute and select the width of each block in order to generate down and up-sampled features maps without exploding the computational cost. Therefore, a number of filters per convolutional layer is initialized to N_w which denotes the number of the outputted feature maps per layer. For each Convolutional layer in the decoding path the number of filters is calculated according to the following equation which helps to extract more complex features from the image: $2^{(N_w + layer_{index})}$

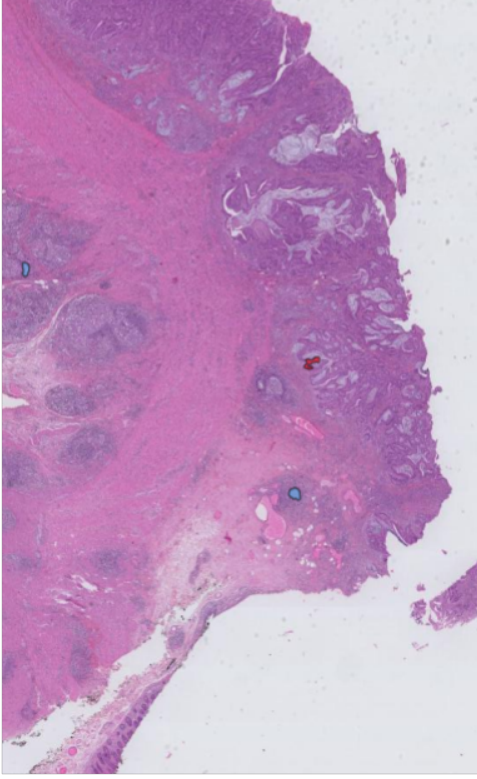
4.3. Learning from sparsely annotated histopathological data

In digital pathology, dense annotation of histopathological images is required to ensure accurate supervised learning of the data content and to enable efficient tumour segmentation. In such context, all pixels of certain areas in the data are manually labeled by experts as seen in Figure 3. However, as detailed in previous sections, this task is extremely expensive and tiring for pathologists. Therefore, we adapt our networks training process to the available sparsely annotated data. In fact, sparse data labeling enables rapid annotation of an important amount of tissues which makes it an attractive solution for pathologists. However, the absence of expert indications about the localization and boundaries of the classes represents an important shortage of information for the learning process. We operate in a weakly supervised mode while presenting solutions to cope with the problem of class imbalance and lack of annotated pixels in WSI. Therefore, we use and combine two loss function enhancement methods. First, we apply a mask of valid patches to learn from. Then, we apply a weighted cross entropy loss to cast the unbalanced representation of tumour and non-tumour classes. Finally, we add a boundary-aware loss function to ensure accurate separation between the different tissues presented in the data.

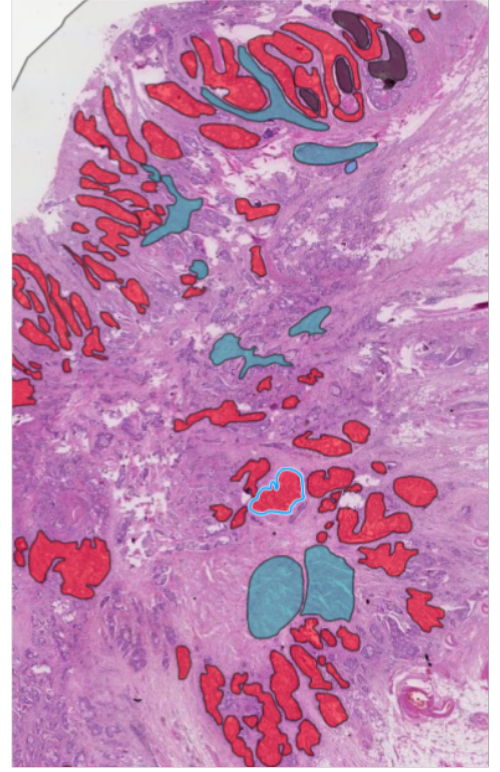
Mask of valid patches The learning process of the network relies on two steps: a forward pass and a backward pass. The second step enables the model to evaluate and adjust its knowledge of the data content. Therefore, we only take into account the annotated pixels from the training set in order to contribute in the optimization process of the network. Therefore, we check the content of our dataset training patches. If a patch contains a majority of annotated pixels (more than 90%) then it is kept for the training process while the weights of the non-annotated pixels are set to 0. Otherwise, the patch is discarded and cannot contribute in the optimization process. Although this step is crucial to avoid confusions between classes, it does not guarantee balance between the presented classes in the data.

Weighted cross entropy loss When dealing with sparse annotation, the classes are usually unequally presented. Since we deal with a binary segmentation task, the presence of the class tumour is remarkably less presented than the non-tumour tissues. In fact, the non-tumour class encompasses different tissue types including stroma, necrosis, immune, background and trash. Therefore, we resort to a weighted cross entropy loss in order to adjust and evaluate our model. Classically, for each class prediction probability y and a groundtruth expected value \hat{y} we assign a loss value computed as in equation (1).

$$Loss(y, \hat{y}) = - \sum_{i=1}^n \hat{y}_i \log y_i \quad (1)$$



(a) Sample of sparse annotation



(b) Sample of dense annotation

Figure 3: Samples from our colon WSI dataset

This function does not take into account the unbalanced representation of the different classes. Therefore, we add a weighted factor to tackle such issue according to equation (2).

$$Loss(y, \hat{y}) = -W_i \sum_{i=1}^n \hat{y}_i \log y_i, \quad \sum_{i=1}^n W_i = 1 \quad (2)$$

$Loss(y, \hat{y})$ is the cross entropy loss function (equation 2) that represents representing the error between the estimated probability $y_i \in [0, 1]$ and the groundtruth class $\hat{y}_i \in [0, 1]$, n is the number of classes, $W_i \in [0, 1]$ is the weighting factor assigned to the class i .

Boundary-aware loss In order to focus on the annotated edges, we add a penalty for errors made when predicting tissue borders. In fact, regarding the dependency of bulk continuous tissue regions in the data, the network usually deflects its attention from the infrequent edge pixels. To ensure that all parts of the data contribute in the learning process, we need to make sure that activation functions are assigned to all filters of the network. Therefore, we compute a new weights feature map where we highlight the edges. In UNET original paper [17] the authors rely on sophisticated morphological functions to generate the edge-aware weights. Here, we simplify the task by using a binary morphological dilatation of the border pixels. In other words, we dilate each edge pixel (i,j) and set it to the maximum value at each region centred at (i,j) . The new weighted boundary-aware loss function $Loss_{wba}(y, \hat{y})$ is then computed as follows:

$$Loss_{wba}(y, \hat{y}) = W_e W_i \sum_{i=1}^n \hat{y}_i \log y_i, \quad (3)$$

where W_e is the new edge dependent weight map and $W_i \in [0, 1]$.

	Training samples	Testing samples
Tumour	976	478
Necrosis	387	193
Immune	301	150
Stroma	642	320
Fat	75	37
Tissue	477	238
Trash	280	139
Background	326	162

Table 2

Number of samples in training and testing sets in our AiCOLO-8 patch-based datasets.

	Training samples	Testing samples
Tumour	976	478
Non-Tumour	2488	1239

Table 3

Number of samples in training and testing in our AiCOLO-2 pixel-wise datasets.

5. Materials and Methods

In order to thoroughly study the performance of the proposed networks, we have used different histopathological datasets. We explore the hallmarks and limitations of the DL networks for both a from scratch training strategy and a transfer learning use case.

5.1. Data

The AiCOLO dataset: eight classes The AiCOLO-8 dataset contains a total of 396 HES stained colorectal histopathological slides digitized with a Hamamatsu photonics scanner at a spatial resolution of $0.454\mu\text{m}/\text{pixel}$. Each image encompasses around 4 to 5 billion pixels. Pathologists from the Centre Georges Francois Leclerc, Cancer and Adaptive Immune Response team (Dijon, France) annotated 60 slides from the dataset. The labelling process highlighted sparse regions of tumour, stroma, fat, necrosis, immune cells clusters, normal tissue, trash and background as seen in Figure 4.

Regarding the important size of the histopathological slides, we have created a patch-based AiCOLO dataset from the few available annotated regions. Therefore, we have used the Cytomine [48] content-based image retrieval algorithms to extract 256×256 patches from the labeled polygons in both the WSI and their corresponding masks, as detailed in Table 2. The obtained patch-based dataset was split into two subsets: training and testing sets. The WSI that have not been annotated are used as validation slides.

The AiCOLO dataset: two classes In order to detect the tumorous tissues, we performed a pixel-wise segmentation of the images from AiCOLO-8. A binary mask has been generated to highlight the targeted class pixels. The task covers the presence of two tissue types: tumour and non tumour classes as seen in Figure 5. Thus, we re-arrange the dataset patches as originally depicted in 5.1. The patches that belong to the class "Tumour" are kept aside. The rest of the crops are joint into one specific class that we call "Non Tumour". We end up with a new dataset; AiCOLO-2, composed of 256×256 patches as shown in Table 3.

Breast Cancer dataset The BREASTHIS dataset is composed of 42 images of estrogen receptor-positive (ER+) breast cancer images scanned at $20\times$ magnification level as depicted in [10]. Each image is a 1000×1000 pixels slide. The epithelium regions were manually annotated by an expert pathologist. To alleviate the computational costs, we created a collection of 256×256 images from all the original slides.

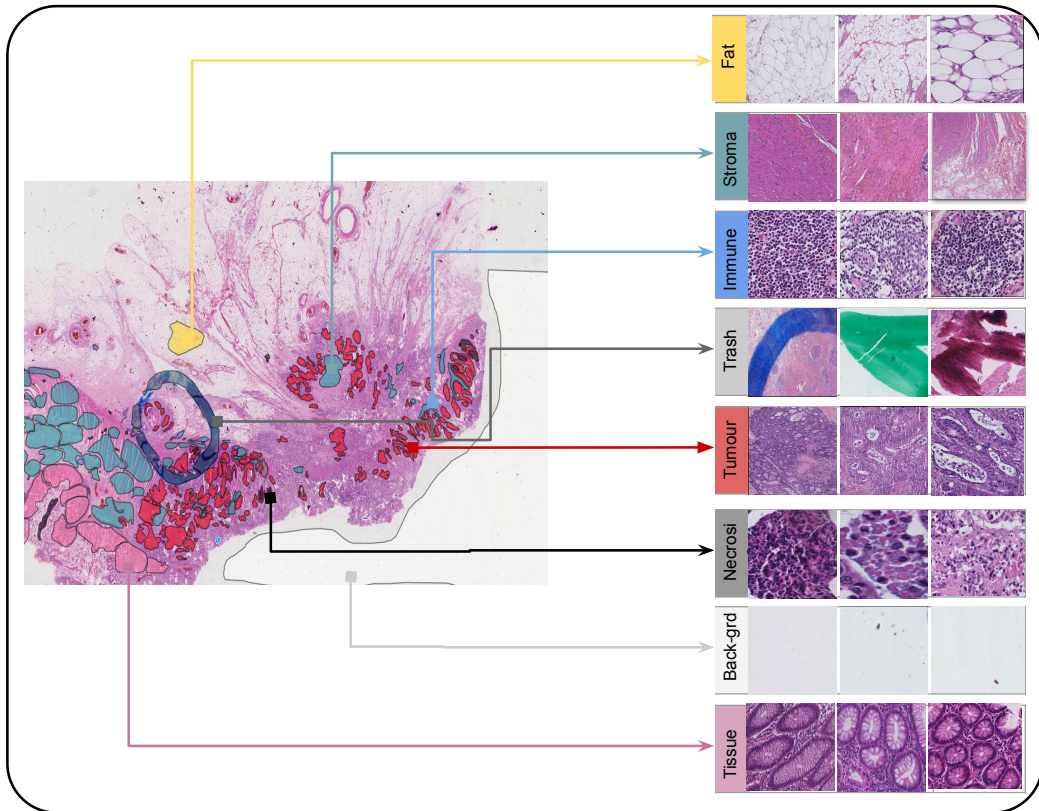
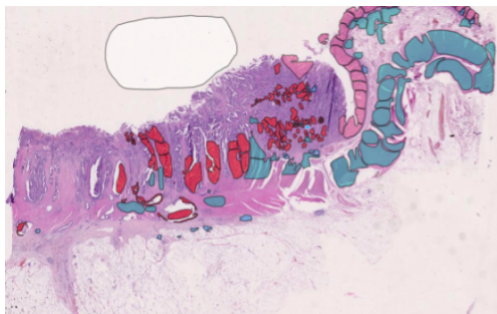


Figure 4: AiCOLO histopathological slide into 256×256 patches extraction from the different annotated regions.



(a) Annotated AiCOLO WSI



(b) Binary mask ("Tumour" tissues in white)

Figure 5: Sample of WSI/mask from the AiCOLO-2 dataset.

ImageNet dataset The IMAGENET dataset was first developed for visual object recognition built with WordNet which is a lexical database for the English language. A total of 1000 classes are presented by over 14 million images. Since 2010, an annual IMAGENET [49] Large Scale Visual Recognition Challenge (ILSVRC) has been taking place. It is a competition where research teams evaluate their algorithms on the given dataset, and compete to achieve higher accuracy on several visual recognition tasks. Besides, IMAGENET gives researchers a common set of images to benchmark their models and algorithms. It's fair to say that IMAGENET has played an important role in the advancement of computer vision.

The 100,000 histological images dataset The NCT-CRC-HE-100K dataset is a set of 100,000 non-overlapping image patches from *H&E* stained histological images of human colorectal cancer (CRC) and normal tissue. All images are 224×224 pixels at 0.5 microns per pixel. All images are color-normalized using Macenko’s method [50]. These images were manually extracted from 86 *H&E* stained human cancer tissue slides from the NCT Biobank (National Center for Tumor Diseases, Heidelberg, Germany) and the UMM pathology archive (University Medical Center Mannheim, Mannheim, Germany). The dataset covers nine tissue classes which are: adipose, background, debris, lymphocytes, mucus, smooth muscle, normal colon mucosa, cancer-associated stroma and colorectal adenocarcinoma epithelium.

The Colorectal Histology MNIST The CRC-5000 dataset contains 5,000 histological images of 150×150 pixels each. All slides were digitized with an Aperio ScanScope (Aperio/Leica biosystems) at a $20\times$ magnification level. Histological samples show human colorectal adenocarcinomas (primary tumours) from our pathology archive (Institute of Pathology, University Medical Center Mannheim, Heidelberg University, Mannheim, Germany). The data has 8 different classes of tissue. Each category has 625 images of *H&E*.

The GlaS (Gland Segmentation in Colon Histology Images Challenge): Warwick The WARWICK dataset used in this challenge consists of 165 images derived from 16 *H&E* stained histological sections of stage T3 or T4 colorectal adenocarcinoma. Each section belongs to a different patient, and sections were processed in the laboratory on different occasions. Thus, the dataset exhibits high inter-subject variability in both stain distribution and tissue architecture.

The AiCOLO dataset VS available datasets As detailed above, our AiCOLO dataset represents only few annotated samples. Unlike the rest of the used datasets, only 15% of the AiCOLO slides have been annotated by the pathologists in a sparse manner. Besides, the WSIs contain many artifacts ranging from out of focus regions to tears and cuts in the tissues as seen in Figure 1. No data cleaning or filtering was applied to the slides. Although the AiCOLO dataset introduces different tissue types, it still suffers from unbalanced representation of the different classes. As seen in Table 2, some classes are poorly presented namely fat and Trash and Immune. The problem arises with the AiCOLO-2 binary segmentation tasks where the "Tumour" tissues are around $2.5\times$ less presented than the "Non Tumour" Tissues. That goes without saying that the BREASTHIS, CRC-5000, NCT-CRC-HE-100K and WARWICK datasets encompass a rich collection of balanced tissue types. The used patches of these WSIs are free of artifacts and staining issues. Our AiCOLO dataset introduces a challenging level of difficulty to train the DL models.

5.2. Experimental settings

The operating system is Linux with an Intel(R) Xeon(R) Bronze 3204 1.9 GHz processors with 62GB RAM. Training and testing processes of the proposed architectures for all experiments are implemented in Python using Pytorch DL framework. The experiments were run on Nvidia Quadro RTX 5000 with 16GB memory.

Data preprocessing and augmentation DL models are in certain need for an important amount of data to avoid over-fitting problems. In other words, the images fed into the network has to present a wide range of data varieties and features. The more rich is the training data, the best are the network generalization performances. In our use case we apply data augmentation techniques to increase the robustness of our network to data variations and to make sure that our trained models can be transferable to different contexts and histopathological images. The augmentation techniques are applied to the original dataset images without increasing the total number of images. For each image, we apply both spatial and color transformations. A series of spatial transformation techniques are applied to each image and its corresponding binary mask. Random vertical, horizontal flips and random zoom crops are applied to the images. Then we utilize rotations of 90 and 180 degrees to simulate the different orientations from a pathologists point of view. Eventually, we randomly change the brightness, contrast and saturation of each slide and convert images to grayscale with a random probabilities.

Patch-level classification of colon cancer histopathological images According to the size of each pre-trained model input, all AiCOLO-8 crops were resized to 224×224 patches. In order to guarantee rich feature representation all along the learning process, we select a batch size of 32. Each model was trained for 25 epochs. Here, we use a Stochastic Gradient Descent (SGD) optimizer with a momentum. The learning rate is first set to 0.001 with a 0.9 momentum value and a L2 penalty of 0.0005. Each 5 epochs we decay the learning rate by a 0.1 factor to cope with the different hierarchy levels of the data content. Regardless from the network architectures we insert a final fully

connected layer with a softmax to ensure patches classification. The size of this final layer is always equal to the number of the targeted classes. A probabilistic representation is then produced by the softmax and the class with the highest score is attributed to each patch according to equation (4) where x_i denotes output of the final convolutional layer, i is the index of each element of the output vector and n is the number of classes.

$$P(x_i) = \frac{e^{x_i}}{\sum_j^n e^{x_j}} \quad (4)$$

Pixel-level Segmentation of colon cancer histopathological images The images taken into account for this experiment are the 256×256 AiCOLO-2 patches as detailed in Table 3. The batch size is set to 32. Since the annotations are sparse and only cover a part of the dataset, we set a weight edge value to 1.1. Each of the UNET and SEGNET models are trained for 100 epochs. The Adam optimization algorithm is employed with a learning rate value of 0.0001. Then, each 20 epochs, learning rate is divided by 10 in order to converge to a lower local loss value and then increase the accuracy. For all experiments and models the convolutional layers are trained with the rectified linear unit (ReLU) activation function. In order to evaluate and compare our models with state-of-the-art methods, we use the following classical metrics:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP}$$

where TP is the total number of true positives, FP the number of false positives, TN the number of true negatives and FN is the number of false negatives.

6. Results and Discussion

We present and examine the results of both patch and pixel level histopathological image classification. The network performances are evaluated in terms of accuracy and computational cost. Each model is trained and evaluated 10 times using different non overlapping train and test random splits. The impact of data preprocessing is also explored and presented in different tasks. To make sure that the presented networks do not overfit, we tackle both from scratch and fine tuning training strategies.

6.1. Patch-level classification of colon cancer histopathological images

The purpose of this task is to ensure accurate classification of histopathological images relying on state-of-the-art DL models. To study the behaviour of the different CNN we deploy three training strategies. Strategy 1 represents a from scratch training. Strategy 2 introduces the fine tuning of the networks and strategy 3 relies on the CNN as fixed feature extractors. The IMAGENET dataset is used to train the networks in both transfer learning strategies. In fact, classical CNN models like ALEXNET, RESNET, DENSENET and VGG require huge amount of thoroughly annotated data. Unfortunately, the histopathological imaging field suffers from a serious lack of richly labeled slides. In our case of study, the AiCOLO-8 dataset is small in size and only represents few sparsely annotated pixels. Therefore, we propose the use of available multimedia datasets to train CNN and fulfill the shortage of histopathological data.

Learning from scratch vs. Transfer Learning We test the ability of CNN architectures to recognize features learnt from computer vision datasets in a specific colorectal histopathological context. As seen in Table 4, the from scratch training has noticeable low accuracy compared to the transfer learning performance rates. Although we have applied spatial and color augmentation to the training data, the networks have tendency to rapidly over-fit and fail to classify features that have not been presented in the training phase. Using the CNN as pre-trained fixed feature extractors enhances the model performances but fails to ensure accurate classification rates. As a matter of fact, when using the pre-trained CNN as fixed feature extractors we only update the weights of the final layer. In other words, we expect high level features learnt from IMAGENET to be highly similar to our AiCOLO slides. However, regarding the difference in both

	From Scratch	Feature Extract.	Fine-Tun/no data augmentation		Fine-Tun/with data augmentation	
			w/o norm.	img norm.	w/o norm.	img norm.
ALEXNET	64.12 ± 0.05	72.23 ± 0.01	76.23 ± 0.02	81.12 ± 0.10	84.21 ± 0.03	89.42 ± 0.01
VGG16	76.48 ± 0.03	79.11 ± 0.03	76.89 ± 0.03	82.02 ± 0.06	90.11 ± 0.09	95.25 ± 0.02
RESNET	79.05 ± 0.02	89.18 ± 0.01	89.98 ± 0.07	90.76 ± 0.02	92.45 ± 0.00	96.98 ± 0.08
DENSENET	76.86 ± 0.00	86.36 ± 0.01	89.24 ± 0.00	90.58 ± 0.05	91.28 ± 0.02	95.86 ± 0.03
INCEPTIONV3	75.29 ± 0.08	83.14 ± 0.01	83.01 ± 0.05	86.74 ± 0.08	90.07 ± 0.01	92.43 ± 0.04

Table 4

Accuracy rates (in %) for ALEXNET, VGG, RESNET, DENSENET and INCEPTION with and without color normalization and data augmentation trained with the AiCOLO-8 dataset.

Method	CRC-5000	NCT-CRC-HE-100K	Merged(NCT-CRC-HE-100K+ CRC-5000)
ALEXNET	91.53 ± 0.04	96.21 ± 0.06	97.56 ± 0.12
VGG	95.15 ± 0.12	98.88 ± 0.03	98.42 ± 0.08
RESNET	96.77 ± 0.04	99.76 ± 0.06	99.98 ± 0.04
DENSENET	95.89 ± 0.09	99.26 ± 0.08	99.76 ± 0.08
INCEPTION	93.65 ± 0.23	97.65 ± 0.08	98.73 ± 0.17
ENSEMBLE DNN[51]	92.83	96.16	99.13
TEXTURE ANALYSIS[52]	87.40	-	-
SQUEEZNET[53]	96.67	-	-

Table 5

Accuracy rates (in %) for the fine-tuned ALEXNET, VGG, RESNET, DENSENET, INCEPTION, SQUEEZNET, ENSEMBLE DNN and TEXTURE ANALYSIS for the CRC-5000, NCT-CRC-HE-100K and merged datasets.

data contents, the pre-trained CNN are limited in recognizing the same high level IMAGENET features in the colorectal histopathological images. The fine-tuning of ALEXNET, RESNET, DENSENET, VGG and INCEPTION models enable the best performances among the three training strategies. As detailed in Table 4, the CNN networks respectively ensure 89.42%, 95.25%, 96.98%, 95.86% and 92.43% accuracy rates. In fact, fine-tuning the different CNN layers enable rapid scanning and updating of the parameters to cope with our AiCOLO data content. Therefore, we could benefit from the early low level generic features and update the final high level layers to fit to our colorectal tissues context.

Raw input data vs. pre-processed input data One main solution for the lack of histopathological data is the use of different data augmentation techniques as detailed earlier in Section 5.2. The application of color and spatial augmentation methods on our AiCOLO slides creates a dataset with richer feature representation. That goes without saying that the use of spatial translations and rotations on the data introduces robustness to the networks and prevent them from over-fitting to a restrained context. Since we are dealing with a fine-tuning task, the training and validation datasets usually need to have the same range. Therefore, we study the role of image normalization in our use case by applying the same mean and deviation of the IMAGENET dataset to our AiCOLO patches before feeding them to the CNN. In Table 4, we compare different pre-processing scenarios for the AiCOLO inputs of ALEXNET, VGG, RESNET, DENSENET and INCEPTION-v3 models. Applying color normalization to the input patches enhances the ability of the DL architectures to learn and detect features in AiCOLO-8 dataset. Thus, when histopathological images share the same mean and deviation as IMAGENET data, it is often easier to recognize the common features in both datasets. Although the color normalization step seems important in enhancing the performance rates of the CNN models, the absence of data augmentation limits the networks ability to accurately classify the histopathological images. AiCOLO-8 data pre-processing raises the accuracy rates by values ranging from about 2% to more than 5% in the case of ALEXNET.

The different CNN models In order to evaluate the ability of the pre-trained DL models to cope with a specific histopathological context, we resort to the use of state-of-the-art methods as detailed in Section 4. The early CNN

models like ALEXNET and VGG networks introduce hierarchical learning of the data content with high computational costs (more than 60M trainable parameters). The fine-tuning of such models on AICOLO-8 dataset ensures relatively high accuracy rates ranging from 89.42% to 95.25%. INCEPTION-v3 architecture comes with the hallmark of using auxiliary networks to better learn the data content without exploding the computational cost. Here, it is remarkably surpassed by both RESNET and DENSENET models. RESNET takes the lead with **96.98%** accuracy rate as seen in Tables 4 and 5. As detailed in Table 1, RESNET also provides lower computational cost while ensuring a few minutes long fine-tuning process. In fact, such network benefits from its feature-reuse training strategy. Thus, it enables the low level features to contribute along the learning process. Therefore, the fine-tuning of RESNET easily recognizes the low level features presented in AICOLO-8. The DENSENET model follows the same pattern as RESNET learning process which explains its similar classification performances.

The overall accuracy rate for the RESNET model is **96.98%**. Delving into the details of its classification results gives a more detailed vision of its performances. Figures 6 and 7 represent the accuracy of each class for both models. The class "Fat" is **100%** recognized by the fine-tuned RESNET and the "Tumour" tissues successfully detected in **99%** of the cases. Its is sometimes confused with "Necrosis" and "Immune" tissues. The confused classes actually introduce high inter-class visual variability and share common visual features (color, shape...). These classes are even harder to detect in the absence of the feature re-use strategy. As seen in Figure 7, the ALEXNET model confuses the "Tumour" class with "Immune", "Necrosis", "Stroma" and "Tissue" with higher rates than the RESNET model.

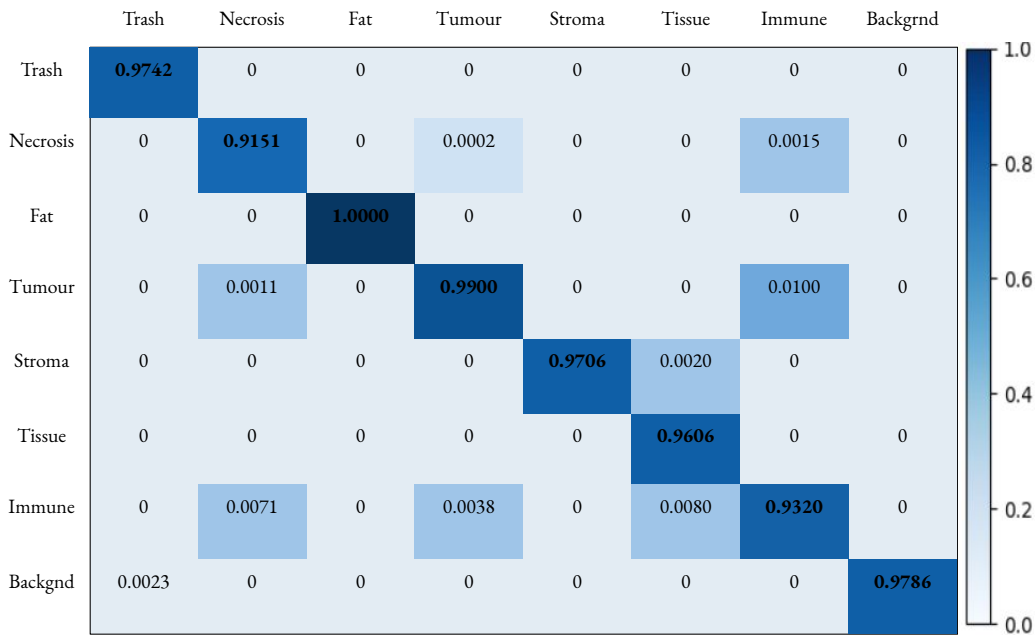


Figure 6: Confusion matrix of the fine-tuned RESNET model.

Comparison with state-of-the-art approaches In order to evaluate our models, we resorted to publicly available colon cancer histopathological datasets namely the CRC-5000 (5,000 images) and NCT-CRC-HE-100K (107,180 images) datasets. In [51], the authors propose a new merged dataset that combines both the NCT-CRC-HE-100K and the CRC-5000 into one 10-class dataset. The merged 112,180 images represent the classes: stroma, debris, adipose, mucosa, tumour, lymphocytes, complex, background, muscle and normal tissue.

Authors in [52] deploy a series of texture-analysis filters namely the Grey Level Co-occurrence Matrix (GLCM) and the Local Binary patterns (LBP). The combination of such tools was able to differentiate multiple tissue types in colon cancer WSI. The paper reports a 87.40% accuracy rate when using the CRC-5000 dataset. However, more recent papers propose methods to automate the task of tissue classification within colorectal cancer WSI using DL. In [51], an Ensemble Deep Neural network was used to classify colorectal histopathological images. The DenseNet-121, InceptionResNetV2, Xception and a custom feed forward CNN are fed into the ensemble architecture, where the

Deep learning for colon cancer

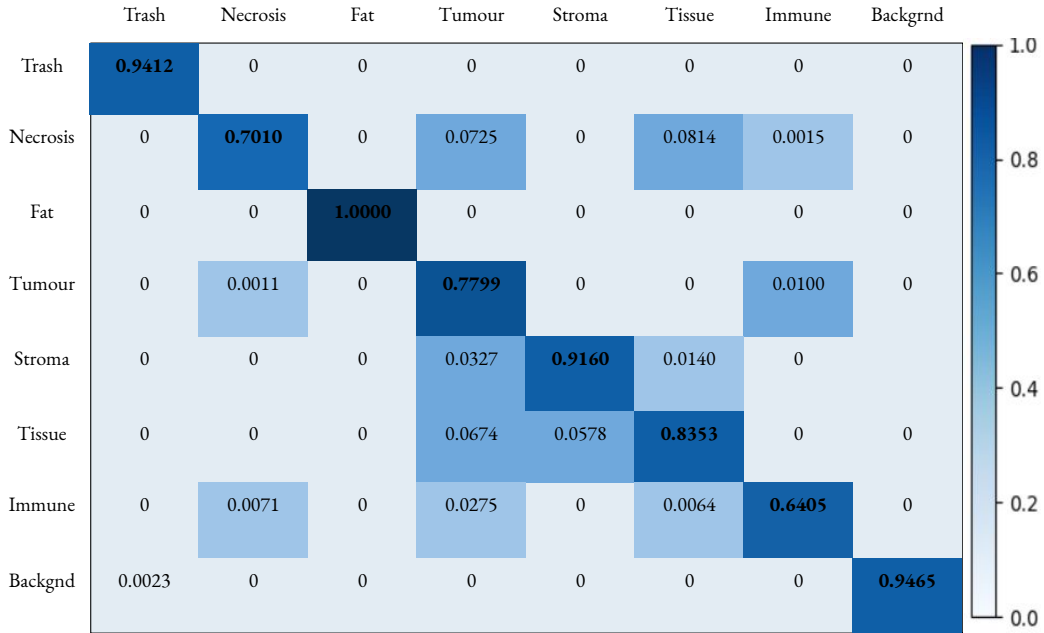


Figure 7: Confusion matrix of the fine-tuned ALEXNET model.

finest feature of each model is extracted and implemented upon the input data, followed by evaluation of ensemble learning model. The ENSEMBLE DNN approach achieves accuracies of 96.16%, 92.83% and 99.13% on respectively NCT-CRC-HE-100K, CRC-5000 and merged datasets. Another DL training strategy was proposed in [53]. The authors resort to a pre-trained SQUEEZNET model to classify the CRC-5000 dataset. First, the model is trained using the IMAGENET dataset. Then, SQUEEZNET is fine-tuned to the colorectal WSI in a one cycle policy. This approach ensures a 96.67% accuracy. As detailed in Table 5, the training strategy that we propose to fine-tune the ALEXNET, VGG, RESNET, DENSENET and INCEPTION models ensures high accuracy rates with the CRC-5000, NCT-CRC-HE-100K and the merged datasets. Except ALEXNET, all proposed fine-tuned networks ensure state-of-the-art accuracy rates > 93%. Here again, RESNET ensures the highest accuracy with respectively 96.77%, 97.65% and 99.98% when fine-tuned with the CRC-5000, NCT-CRC-HE-100K and merged datasets. We also report an accuracy which is > 98% in a Tumor-Stroma tissue classification scenario for all the used datasets. That goes without saying that the fine-tuned RESNET model is capable of thoroughly recognizing the different colon tissues presented in the WSI. The proposed approach also ensures low computational cost and tuning time compared with the state-of-the-art methods.

6.2. Pixel-level segmentation of colon cancer histopathological images

We restrain here the issue to a binary segmentation of the AICOLO-2 dataset using UNET and SEGNET models which are introduced in section 4.2. The learning process is executed at pixel level, which means that the model weights are progressively adjusted according to the class value of each pixel.

Learning from scratch vs. Transfer Learning We test and evaluate two training strategies. First, a training from scratch is performed, based on images and masks from AICOLO-2 dataset. Then, we have resorted to the use of a publicly available BREASTHIS dataset to fine-tune our network by updating all weights of the models. Both UNET and SEGNET perform much better when trained and tested on the same dataset (no fine-tuning) as seen in Table 6. In fact, fine-tuning the networks from BREASTHIS data to AICOLO-2 enables limited pixel-level segmentation. When thoroughly exploring the content of the training data, one can easily notice the difference in tissue annotations. All epithelium tissues in BREASTHIS images are considered as tumorous tissues whereas only malignant epithelium tissues in AICOLO-2 slides are considered as tumorous. Therefore, the confusion between the "Benign" and "Tumour" pixels inhibits accurate pixel-wise segmentation in a fine-tuning strategy. As seen in Figure 8, the fine-tuning of the SEGNET model enables accurate segmentation of the "Tumour" pixels, but the generated masks cover an important

	From Scratch				Fine-Tuning				Parameters	Size	Time	Convergence
	Acc.	Spec.	Sens.	Dice.	Acc.	Spec.	Sens.	Dice.				
UNET	76.18 ± 0.13	74.31 ± 0.20	76.05 ± 0.08	71.89 ± 0.04	32.17 ± 0.08	30.41 ± 0.21	31.86 ± 0.08	28.82 ± 0.02	0.12M	4GB	4260	20/100 Epochs
SEgNET	81.22 ± 0.02	80.70 ± 0.06	81.40 ± 0.02	75.53 ± 0.03	46.23 ± 0.07	45.88 ± 0.4	46.12 ± 0.03	43.52 ± 0.02	7.6M	31GB	11480	40/100 Epochs

Table 6

Accuracy rates, precision, sensitivity (in %) and computational cost for SEgNET and UNET with 1) the training from scratch strategy using AiCOLO-2 dataset and 2) the fine-tuning training strategy using BREASTHis dataset. Time corresponds to the tuning time in minutes.

	CRC-5000				NCT-CRC-HE-100K				Warwick			
	Acc.	Spec.	Sens.	Dice.	Acc.	Spec.	Sens.	Dice.	Acc.	Spec.	Sens.	Dice.
UNET	97.43 ± 0.15	97.96 ± 0.08	96.69 ± 0.12	96.78 ± 0.03	98.73 ± 0.05	98.89 ± 0.02	97.66 ± 0.012	97.99 ± 0.02	77.83 ± 0.17	74.33 ± 0.07	78.87 ± 0.09	74.20 ± 0.02
SEgNET	98.66 ± 0.08	99.02 ± 0.12	98.14 ± 0.08	98.38 ± 0.04	99.12 ± 0.08	99.56 ± 0.07	98.36 ± 0.13	98.73 ± 0.04	78.39 ± 0.24	76.09 ± 0.16	81.93 ± 0.08	74.48 ± 0.05
TEXTURE ANALYSIS[52]	98.60	-	-	-	-	-	-	-	-	-	-	-
ENSEMBLE DNN[51]	-	-	-	-	96.16	-	-	-	-	-	-	-
CNN[54]	-	-	-	-	-	-	-	-	-	57.00	73.00	61.0

Table 7

Accuracy rates, precision and sensitivity (in %) for UNET and SEgNET trained from scratch.

amount of "False positive" pixels. This phenomenon is also visible when using the UNETmodel in the same context. Therefore, the training from scratch strategy increases a lot the accuracy rates from 32.17% to 76.18% in the case of the UNET model and from 46.23% to 81.22% in the case of the SEgNET network.

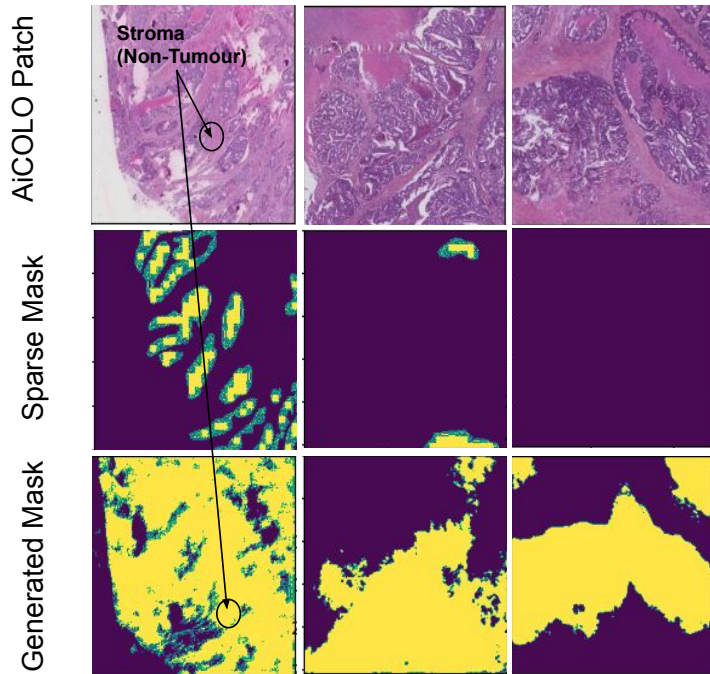


Figure 8: Samples of the generated masks from fine-tuning SEgNET.

UNET vs. SEgNET for histopathological image pixel-wise segmentation We evaluate the performance rates of both UNET and SEgNET models in a histopathological image segmentation context. As detailed in Table 6, SEgNET enables ≈ 5% higher accuracy rate than UNET. As represented in Figure 9, UNET is capable of avoiding false positive pixels but generates false negatives. In other words, the UNET architecture successfully classifies "non tumorous" pixels but fails in accurately detecting sparse "Tumour" regions. SEgNET learns and detects "Tumour" pixels in

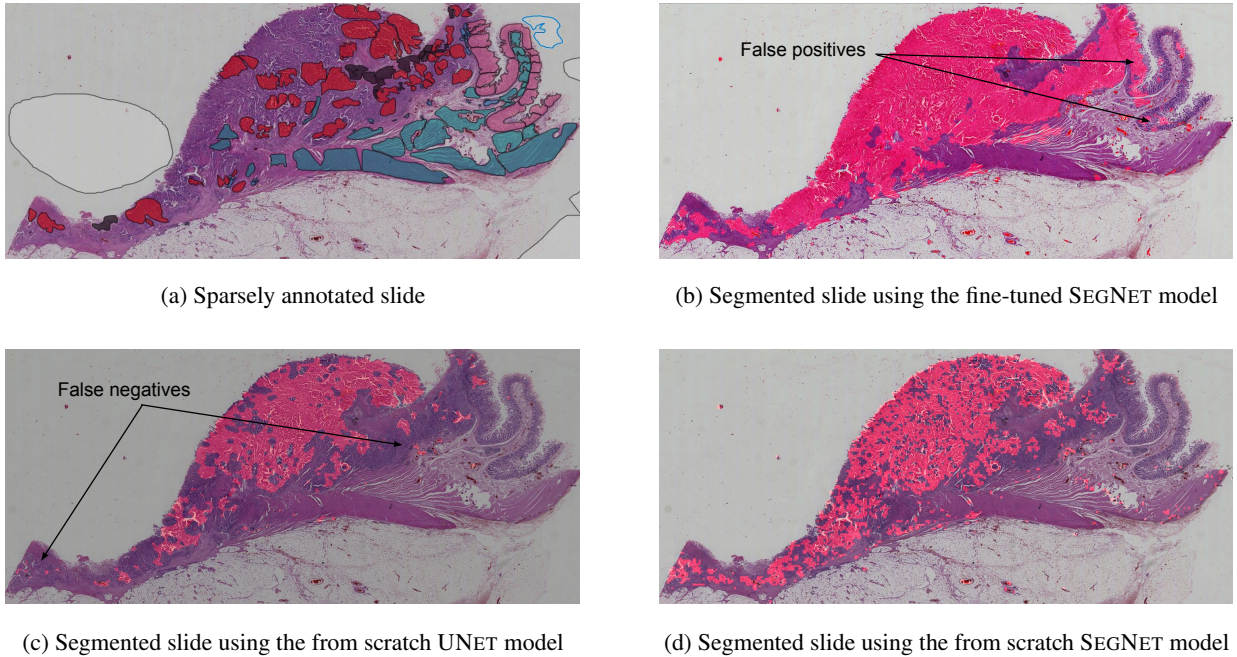


Figure 9: Segmentation results for UNET and SEGNET models trained from scratch and fine-tuned.

both large and narrow regions as seen in Figure 9. The errors made by the network usually cover edges of "Tumour" and "Necrosis" regions. As a matter of fact, the UNET model relies on the skip connections to dissipate the loss of spatial features along the encoding/decoding process. The network merges low-level features from the encoder with high deep features from the decoder. Consequently, the semantic gap between the two types of features may adversely affect the prediction procedure. In other words, the fusion of these incompatible sets of features could cause some discrepancy throughout the learning. Besides, the use of low level features in the learning process encourages UNET to focus on edges and boundaries of large regions. Whereas, The deployment of convolutions with different filters enables SEGNET to inspect points of interest in images from different scales. The use of learnt weights in the pooling layers adds robustness to the network and ensure effective feature learning at different semantic levels.

For a complete comparison between UNET and SEGNET models, we summarize the computational parameters of both networks in Table 6. Actually, SEGNET architecture represents a heavy model that trains 7 times more parameters than UNET and takes 2 times more epochs to converge toward the final accuracy. Despite its high accuracy rate, SEGNET remains limited by its high computational cost.

Comparison with state-of-the-art approaches In order to evaluate our models, we resorted to publicly available colon cancer histopathological datasets namely the CRC-5000 (5,000 images) and WARWICK (165 images) datasets. As detailed in Section 6.1, authors in [52] rely on the combination of different texture analysis filters in a binary classification scenario. An accuracy of 96.60% was reached with the CRC-5000 dataset. However, the use of classical filters remain limited when dealing with larger datasets. Therefore, the paper [51] relies on automatic classification DL models in a tumour vs. non tumour classification task on the NCT-CRC-HE-100K histopathological dataset (100,000 images). The ENSEMBLE DNN generates a 96.16% accuracy. As detailed in Table 7, the proposed SEGNET model ensures the highest accuracy rates: 98.66% and 99.12% for the respective CRC-5000 and NCT-CRC-HE-100K datasets. The network demonstrates good performance level when detecting both tumour and non-tumour tissues. A specificity $> 99\%$ in both cases shows the ability of SEGNET to correctly identify tissues without tumour. Moreover, a sensitivity $> 98\%$ proves that SEGNET correctly identifies tumour tissues. We have also trained SEGNET for a colorectal WSI segmentation task using the WARWICK dataset. The authors in [54] introduce a novel DL model to segment benign and malignant tissues in colorectal cancer WSI. This approach was developed to participate to the GlaSMICCAI2015 colon gland segmentation challenge. In order to generate a pixel-wise binary segmentation of the WARWICK dataset,

two CNN were used. First, the classifier separates glands from background. Then, a different classifier identifies gland-separating structures. Finally, a figure-ground segmentation based on weighted total variation produces the final segmentation result by regularizing the CNN predictions. The CNN models generate both a precision and sensitivity $> 50\%$. However, our proposed training strategy with SEGNET enhances the segmentation results with the WARWICK dataset. As detailed in Table 7, our proposed approach generates a 78.39% accuracy with a 76.09% True Negative ratio detection and a 81.93% as a true positive ratio detection.

Over-fitting vs. Good-fitting When dealing with DL for a segmentation task, good accuracy rates do not provide a complete picture of the network performances. In fact, a successful model ensures high accuracy rates without over-fitting to the training data. Therefore, we need to make sure that our from scratch trained SEGNET did not over-learn the AiCOLO features. For that purpose, we plot the training and validations losses for the from scratch SEGNET model in Figure 10. According to the loss curves, both the training and validation losses follow a similar pattern. The dataset is then capturing good statistical representations in both training and validation. Moreover, both losses decrease to a stability point with a minimal difference between final values. The training loss is slightly higher than the validation loss which validates the "Good fitting" theory. That is to say that the network recognizes unseen data slightly less than training images. The learning stops before the error on the validation dataset increases. Consequently, the network did not learn irrelevant details and noise in the training dataset and acquired a good ability to generalize to the validation dataset.

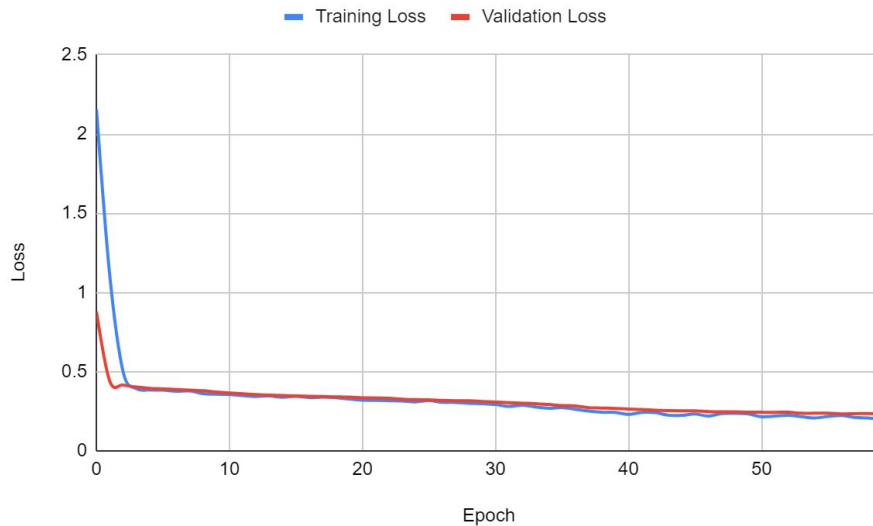


Figure 10: Training vs. validation loss for SEGNET training from scratch model.

7. Conclusion

In this paper, we have presented and evaluated state-of-the-art DL models for patch and pixel level classification of a sparsely annotated colorectal histopathological dataset. We have introduced the use of transfer learning from a generalized multimedia dataset to a specific histopathological image context. A "from scratch" training has also been presented in a pixel-wise tumour segmentation task in histopathological data. The models were also evaluated with different colorectal histopathological datasets and compared to state-of-the-art methods.

The patch-level colon cancer histopathological image classification combined with a fine-tuning training strategy provides a key path for time gaining in tumour classification tasks. However, this method is only highly performing when the patches are pre-selected to represent one class each. In other words, if the WSI encompasses small regions of neighboring classes at each patch, the models fail to generate accurate classification. Therefore, the pixel-wise segmentation represents a more accurate tool for the analysis of colorectal histopathological images. Both UNET and SEGNET models perform well in our case of study. But, SEGNET architecture comes with the hallmark of accurately

detecting tumorous pixels without adding "false positives". That goes without saying that SEGNET generates more important computational cost than UNET. As a remedy, future work will focus on optimization techniques for accuracy against computational cost balance.

While its has been first dedicated for computer visions tasks, SEGNET seems like an appealing solution for cancer segmentation in colon histopathological images. However, this model is greedy for richer training data as seen in the different results. As a matter of fact, SEGNET reaches very high performance rates (accuracy >99.5%) when trained with the NCT-CRC-HE-100K dataset (100,000 images). Although the proposed training scenario enhances the results when dealing with a sparsely annotated dataset, the DL models still confuse tumour tissues with immune and necrosis at many occasions. Actually, both AiCOLO and CRC-5000 datasets almost encompass the same number of training patches in total. However, CRC-5000 represents a balanced number of artifact-free patches per class unlike AiCOLO data. Therefore, the performance rates of the SEGNET model witnesses an increase by a noticeable margin of > 15%. As a matter of fact, the proposed study is capable of coping with the sparse unbalanced noisy data to a certain extent.

Acknowledgment

This work was supported by the AiCOLO project funded by INSERM/Plan Cancer. The authors would like to thank the Cytomine project for their tool used to annotate the images: <http://www.cytomine.org>.

References

- [1] Metin N Gurcan, Laura E Boucheron, Ali Can, Anant Madabhushi, Nasir M Rajpoot, and Bulent Yener. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2:147–171, 2009.
- [2] Liron Pantanowitz. Digital images and the future of digital pathology. *Journal of pathology informatics*, 1, 2010.
- [3] David RJ Snead, Yee-Wah Tsang, Aisha Meskiri, Peter K Kimani, Richard Crossman, Nasir M Rajpoot, Elaine Blessing, Klaus Chen, Kishore Gopalakrishnan, Paul Matthews, et al. Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology*, 68(7):1063–1072, 2016.
- [4] Liron Pantanowitz, John H Sinard, Walter H Henricks, Lisa A Fatheree, Alexis B Carter, Lydia Contis, Bruce A Beckwith, Andrew J Evans, Avtar Lal, and Anil V Parwani. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the college of american pathologists pathology and laboratory quality center. *Archives of Pathology and Laboratory Medicine*, 137(12):1710–1722, 2013.
- [5] Saiful Amin, Taro Mori, and Tomoo Itoh. A validation study of whole slide imaging for primary diagnosis of lymphoma. *Pathology international*, 69(6):341–349, 2019.
- [6] Rozita Rastghalam and Hossein Pourghassem. Breast cancer detection using mrf-based probable texture feature and decision-level fusion-based classification using hmm on thermography images. *Pattern Recognition*, 51:176–186, 2016.
- [7] Ilias Theodorakopoulos, Dimitris Kastaniotis, George Economou, and Spiros Fotopoulos. Hep-2 cells classification via sparse representation of textural features fused into dissimilarity space. *Pattern Recognition*, 47(7):2367–2378, 2014.
- [8] Akira Saito, Yasushi Numata, Takuya Hamada, Tomoyoshi Horisawa, Eric Cosatto, Hans-Peter Graf, Masahiko Kuroda, and Yoichiro Yamamoto. A novel method for morphological pleomorphism and heterogeneity quantitative measurement: Named cell feature level co-occurrence matrix. *Journal of pathology informatics*, 7, 2016.
- [9] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.
- [10] Andrew Janowczyk and Anant Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of pathology informatics*, 7, 2016.
- [11] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- [13] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
- [15] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016.
- [16] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567, 2015.
- [17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [18] Vijay Badrinarayanan, Ankur Handa, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *arXiv preprint arXiv:1505.07293*, 2015.
- [19] Cigdem Gunduz-Demir, Melih Kandemir, Akif Burak Tosun, and Cenik Sokmensuer. Automatic segmentation of colon glands using object-graphs. *Medical image analysis*, 14(1):1–12, 2010.
- [20] Nandita Nayak, Hang Chang, Alexander Borowsky, Paul Spellman, and Bahram Parvin. Classification of tumor histopathology via sparse feature learning. In *2013 IEEE 10th international symposium on biomedical imaging*, pages 410–413. IEEE, 2013.
- [21] Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with

- deep neural networks. In *International conference on medical image computing and computer-assisted intervention*, pages 411–418. Springer, 2013.
- [22] Fabio Alexandre Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. Breast cancer histopathological image classification using convolutional neural networks. In *2016 international joint conference on neural networks (IJCNN)*, pages 2560–2567. IEEE, 2016.
- [23] Matko Šarić, Mladen Russo, Maja Stella, and Marjan Sikora. Cnn-based method for lung cancer detection in whole slide histopathology images. In *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–4. IEEE, 2019.
- [24] Justin Ker, Yeqi Bai, Hwei Yee Lee, Jai Rao, and Lipo Wang. Automated brain histology classification using machine learning. *Journal of Clinical Neuroscience*, 66:239–245, 2019.
- [25] Korsuk Sirinukunwattana, Josien PW Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J Matuszewski, Elia Bruni, Urko Sanchez, et al. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35:489–502, 2017.
- [26] Wenqi Li, Siyamalan Manivannan, Shazia Akbar, Jianguo Zhang, Emanuele Trucco, and Stephen J McKenna. Gland segmentation in colon histology images using hand-crafted features and convolutional neural networks. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1405–1408. IEEE, 2016.
- [27] Mary Shapcott, Katherine J Hewitt, and Nasir Rajpoot. Deep learning with sampling in colon cancer histology. *Frontiers in Bioengineering and Biotechnology*, 7:52, 2019.
- [28] Jakob Nikolas Kather, Johannes Krisam, Pornpimol Charoentong, Tom Luedde, Esther Herpel, Cleo-Aron Weis, Timo Gaiser, Alexander Marx, Nektarios A Valous, Dyke Ferber, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS medicine*, 16(1):e1002730, 2019.
- [29] Jakob Nikolas Kather, Alexander T Pearson, Niels Halama, Dirk Jäger, Jeremias Krause, Sven H Loosen, Alexander Marx, Peter Boor, Frank Tacke, Ulf Peter Neumann, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature medicine*, 25(7):1054–1056, 2019.
- [30] Dmitrii Bychkov, Nina Linder, Riku Turkki, Stig Nordling, Panu E Kovanen, Clare Verrill, Margarita Walliander, Mikael Lundin, Caj Haglund, and Johan Lundin. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific reports*, 8(1):1–11, 2018.
- [31] Hawraa Haj-Hassan, Ahmad Chaddad, Youssef Harkouss, Christian Desrosiers, Matthew Toews, and Camel Tanougast. Classifications of multispectral colorectal cancer tissues using convolution neural network. *Journal of pathology informatics*, 8, 2017.
- [32] Osamu Iizuka, Fahdi Kanavati, Kei Kato, Michael Rambeau, Koji Arihiro, and Masayuki Tsuneki. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Scientific Reports*, 10(1):1–11, 2020.
- [33] Paola Sena, Rita Fioresi, Francesco Faglioni, Lorena Losi, Giovanni Faglioni, and Luca Roncucci. Deep learning techniques for detecting preneoplastic and neoplastic lesions in human colorectal histological images. *Oncology Letters*, 18(6):6101–6107, 2019.
- [34] Daisuke Komura and Shumpei Ishikawa. Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*, 16:34–42, 2018.
- [35] Chetan L Srinidhi, Ozan Ciga, and Anne L Martel. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, page 101813, 2020.
- [36] Ishak Pacal, Dervis Karaboga, Alper Basturk, Bahriye Akay, and Ufuk Nalbantoglu. A comprehensive review of deep learning in colon cancer. *Computers in Biology and Medicine*, page 104003, 2020.
- [37] Bruno Korbar, Andrea M Olofson, Allen P Mirafior, Catherine M Nicka, Matthew A Suriawinata, Lorenzo Torresani, Arief A Suriawinata, and Saeed Hassanpour. Deep learning for classification of colorectal polyps on whole-slide images. *Journal of pathology informatics*, 8, 2017.
- [38] Osamu Iizuka, Fahdi Kanavati, Kei Kato, Michael Rambeau, Koji Arihiro, and Masayuki Tsuneki. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Scientific Reports*, 10(1):1–11, 2020.
- [39] Jiayun Li, Wenyuan Li, Arkadiusz Gertych, Beatrice S Knudsen, William Speier, and Corey W Arnold. An attention-based multi-resolution model for prostate whole slide image classification and localization. *arXiv preprint arXiv:1905.13208*, 2019.
- [40] Peter D Caie, Klaas Schuur, Anca Oniscu, Peter Mullen, Paul A Reynolds, and David J Harrison. Human tissue in systems medicine. *The FEBS journal*, 280(23):5949–5956, 2013.
- [41] Derek Magee, Darren Treanor, Doreen Crellin, Mike Shires, Katherine Smith, Kevin Mohee, and Philip Quirke. Colour normalisation in digital histopathology images. In *Proc Optical Tissue Image analysis in Microscopy, Histopathology and Endoscopy (MICCAI Workshop)*, volume 100, pages 100–111. Citeseer, 2009.
- [42] Nicolas Brieu, Christos G Gavriel, David J Harrison, Peter D Caie, and Günter Schmidt. Context-based interpolation of coarse deep learning prediction maps for the segmentation of fine structures in immunofluorescence images. In *Medical Imaging 2018: Digital Pathology*, volume 10581, page 105810P. International Society for Optics and Photonics, 2018.
- [43] Francesco Ponzio, Enrico Macii, Elisa Ficarra, and Santa Di Cataldo. Colorectal cancer classification using deep convolutional networks. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*, volume 2, pages 58–66, 2018.
- [44] Francesco Ponzio, Gianvito Urgese, Elisa Ficarra, and Santa Di Cataldo. Dealing with lack of training data for convolutional neural networks: The case of digital pathology. *Electronics*, 8(3):256, 2019.
- [45] Srinath Jayachandran and Ashlin Ghosh. Deep transfer learning for texture classification in colorectal cancer histology. *arXiv preprint arXiv:2004.01614*, 2020.
- [46] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.
- [47] Patrice Y Simard, David Steinkraus, John C Platt, et al. Best practices for convolutional neural networks applied to visual document analysis. In *Icdar*, volume 3, 2003.
- [48] Raphaël Marée, Loïc Rollus, Benjamin Stévens, Renaud Hoyoux, Gilles Louppe, Rémy Vandaele, Jean-Michel Begon, Philipp Kainz, Pierre Geurts, and Louis Wehenkel. Collaborative analysis of multi-gigapixel imaging data using cytomine. *Bioinformatics*, 32(9):1395–1401, 2016.
- [49] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE*

- conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [50] Marc Macenko, Marc Niethammer, James S Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1107–1110. IEEE, 2009.
- [51] Sourodip Ghosh, Ahana Bandyopadhyay, Shreya Sahay, Richik Ghosh, Ishita Kundu, and KC Santosh. Colorectal histology tumor detection using ensemble deep neural network. *Engineering Applications of Artificial Intelligence*, 100:104202, 2021.
- [52] Jakob Nikolas Kather, Cleo-Aron Weis, Francesco Bianconi, Susanne M Melchers, Lothar R Schad, Timo Gaiser, Alexander Marx, and Frank Gerrit Zöllner. Multi-class texture analysis in colorectal cancer histology. *Scientific reports*, 6(1):1–11, 2016.
- [53] Srinath Jayachandran and Ashlin Ghosh. Deep transfer learning for texture classification in colorectal cancer histology. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*, pages 173–186. Springer, 2020.
- [54] Philipp Kainz, Michael Pfeiffer, and Martin Urschler. Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization. *PeerJ*, 5:e3874, 2017.