Panoptic Segmentation of Cell-Types Nuclei in Colorectal Adenocarcinoma Histology Images

Laura Acosta
Universidad de los Andes
Bogotá, Colombia
lv.acostac@uniandes.edu.co

Juanita Puentes
Universidad de los Andes
Bogotá Colombia
j.puentes@uniandes.edu.co

Abstract

The classification of cell types is essential for the timely diagnosis of colorectal cancer. We propose an exhaustive experimentation on the original HoverNet architecture implemented by Graham et al [1]. HoVer-Net is a multiple branch network that performs nuclear instance segmentation and classification within a single network. It performs panoptic segmentation by extracting features from histopathology images and differentiating nuclei using vertical and horizontal distance maps and gradient maps to the nuclei centers to separate agglomarated cells. We carry out experimentation by varying the backbone between ImageNet-ResNet50 and ImageNet-ResNet101 to do fine tuning. We modify hyperparameters such as the optimizer, the activation function, the weight decay, the number of epochs and the batch size. We implement ablation studies by removing residual units and modifying the loss function. We obtained the best results using ImageNet-ResNet101 as the backbone with a Panoptic Quality of 0.47438. This result represents an improvement of more than 100% compared to the baseline.

1. INTRODUCTION

Identification and characterization of tumours has been one of the major occupations of the medical and technological field, given that cancer is one of the diseases responsible for a large number of deaths worldwide each year [2]. One of the most relevant areas in the process of analysis and diagnosis of colorectal cancer (CRC). This type of cancer ranks among the three most common cancers in terms of both cancer incidence and cancer-related deaths in most Western countries [3]. It is crucial to determine morphological changes in the cell nucleus because they are considered as an important signal to provide meaningful medical information during diagnosis, especially for colorectar adenocarcinoma [4]. In this field, the evaluation of the disease status is based on cell nuclei information of the tissue images, which is required to develop appropriate treatments depending on the number of malignant cell nuclei. This is why extract the region of cell nuclei accurately is an extremely important task [5]. Depending on the cell types, the response to tumor formation is different. Each of them may have a role related to early stages, invasion and metastasis, and even therapeutic response, which also varies according to the organ in which they occur [6]. Some examples include malignant cells that cause epithelium dysplasia. The presence of colorectal cancer modifies the tumor microenvironment, increasing the proportion of inflammatory cells [7]. Another important factor in the diagnosis of colorectal adenocarcinoma is the shape of the cell nuclei, since their morphology is affected by the occurrence of cancer [8]. It is also important to take into account the association between aberrant nuclear structure and tumour grade [8], especially when analysing the images and drawing conclusions for optimal diagnosis.

Conventional diagnosis of colorectal adenocarcinoma and grades from histopathologic images is performed by specialized pathologists, who are guided by certain morphologic features of the nuclei of the cells. This manual evaluation is tedious and time consuming because there are many aspects of human error such as low sensitivity and low specificity, adding the fact that the results are not entirely reproducible, as they vary from one pathologist to another [4]. In terms of nuclei segmentation this is a slightly more frequent problem, so the approach of automated algorithms arises as a basic need [4].

In addition to human subjectivity, the great variety of cell nuclei types, and their dependence on multiple factors, makes the segmentation process even more complex. Generalization of current methods to all types of histopathological images is difficult to achieve, as the state of the images depends on variables that are not the same in all acquisition contexts. In addition, the nuclei in the given image sometimes overlap each other and are distributed non-uniformly, which affects not only the segmentation, but also prognosis

and diagnosis [5]. This is why the development of more robust methods, is being pursued on a daily basis, in order to provide a much more useful medical service for the subsequent treatment of the disease.

Panoptic segmentation is a type of image segmentation that combines semantic and instance-based segmentation methods. This implies that this algorithm allows to obtain both the features of the background (by pixel assignation to a same category), and to learn the features of the different objects in the foreground (by assigning unique labels to each object belonging to the same class)[9]. This study proposes an algorithm based on machine learning techniques to achieve the panoptical segmentation of the different nuclei present on histopathological images of colorectal adenocarcinoma (malignant/dysplastic epithelium, fibroblast, inflammatory and miscellaneous) included on CoNSeP [1] (Colorectal Nuclear Segmentation and Phenotypes) dataset associated with the cell groups involved in the process of tumour formation and growth.

2. RELATED WORK

Multiple methods associated with image analysis and deep learning have been proposed to solve the problem of cell nuclei segmentation, given its high medical relevance. They range from the separation of nuclei from their background by image binarization, commonly associated with thresholding [5][10], use of markers [5][11], and use of morphological operations [5]; to more sophisticated techniques involving algorithms related to machine learning, the latter being the most recently explored.

2.1. Semantic Segmentation Methods

Deep neural networks are the most general category of neural networks, which seek to perform data processing using mathematical models. To approach the semantic segmentation task, convolutional neural networks (CNNs) are the most frequently used. Some approaches include the use of fully automated deep neural network [12] [10], with some strategies to solve problems like joint kernel segmentation [12], or overlapping nuclei [10]. Other approaches include the execution of CNN models [13] [14], that generates a probability map for each image entered into the network, in such a way that each pixel is assigned a probability of belonging to each particular nucleus, to finally apply an iterative algorithm to perform the final segmentation [14]. CNN methods, however, are usually limited by characteristics such as the size and quality of the input images, requiring specific and above all rather large and diverse datasets [15]. To solve this, some authors have purposed alternatives using conditional generative adversarial network (cGAN) [16], or transformer neural networks[15], since they are capable of capturing the global long-distance dependencies across the entire image as a new deep learning paradigm [15].

2.2. Instance Segmentation Methods

The strategies used to perform instance segmentation are mainly based on the implementation of modified convolutional neural networks. The popular approach for instance segmentation involves object detection using a box followed by object-box segmentation [17]. Mask-R CNN [18], a CNN that allows to generate masks from previously detected bounding boxes, is the baseline network in this task, and many current methods are based on it [19]. This architecture evolves from R CNN networks, by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition [18]. This allows the generalization of network use in different tasks, and a faster and easy training phase [18].

Alternatives include techniques focused on labelling pixels followed by clustering [19] [20] [21], which consist in categorical labelling every pixel of the image, and then group object instances using a clustering algorithm [20]; or classification of mask proposals [19], [22] [23], that start from the generation of candidates for objects or regions of interest to develop the segmentation.

2.3. Panoptic Segmentation Methods

Related to the CoNSeP dataset, multiple algorithms involving deep learning have been proposed to perform panoptic segmentation of the histopathological image nuclei. The participants of the MoNuSAC2020 challenge, focused on this task, used all deep convolutional neural networks to segment and classify the nuclei [24], with variations on the base architectures and the pre-processing methods. SJTU 426 team [24], for example, worked based on performing Reinhard colour normalization, then data augmentation with rescaling transformations, horizontal and vertical flipping, random rotations and normalization of each image channel in its RGB space [24]. The deep neural network consisted of two levels, based on the ResNet34 architecture of the U-Net network in order to obtain probability maps for the input image pixels, combinated with cross-entropy and perceptual loss functions [24][25]. They train a second U-Net (VGG16 architecture) to perform the final classification making use of the cross-entropy related loss [24] [25][26]. Other good approaches are the ones made by the SharifHooshPardaz and The Great Backpropagator teams [24], who based their algorithms on pre-training with ImageNet and training U-Net based architectures (EfficientNetB7 EfficientNetB3 respectively), as well as applying watersheds to the obtained masks for the final separation of the nuclei [24].

2.4. Datasets

Within the existing datasets of histopathological images to carry out the panoptic segmentation of cell nuclei are MIC17[27] and BNS. This first dataset (MIC17) was provided for the MICCAI 2017 Digital Pathology Challenge [27]. It includes 32 annotated squared image patches of sizes 500×500 cropped from HE-stained histopathology Whole Slide Images (WSIs) [27]. It contains images from four types of cancer: glioblastoma (GBM), lower grade glioma (LGG), head and neck squamous cell carcinoma (HNSC) and lung squamous cell carcinoma (LUSC) [27]. As an evaluation metrics, the Dice Score Coefficient (DSC) and Hausdorff distance were used for semantic segmentation. A particular metric to evaluate the performance of this task is the Panoptic Quality (PQ) [28]. The dataset used for MoNuSeg (2018) [29] contains images from seven organs with about 95,000 different nuclei [30] [29]. This dataset was created by downloading HE stained tissue images captured at 40x magnification from TCGA archive [29] [30]. The evaluation methodology used to quantify the performance of the different participants was the Aggregated Jaccard Index (AJI) used to compute the nuclei segmentation accuracy [29][30].

3. APPROACH

3.1. Baseline

For the first approach to solve the panoptic segmentation task we implement HoVer-Net [1]: a deep neural network for feature extraction components. As a baseline, we decided to carry out the training without using pre-trained models. We implement an algorithm for extracting 80x80 dimension patches from the original 1000x1000 train images. Additionally, we implement *Data Augmentation* by performing transformations on patches. For training and validation, we use 25 epochs respectively for each one. In addition, we use a learning rate of 0.001 and Adam as the optimization function. Table 1 shows a comparison between the results obtained through the baseline and the metrics obtained for the original implementation of HoVer-Net [1] proposed by Graham et al. It is important to mention that these two methods differ mainly in the use of pre-trained models to carry out a fine tuning algorithm.

Table 1: Comparison between the results obtained for the baseline and the original implementation of HoVer-Net

Method	Dice	AJI	PQ
Baseline	0.50417	0.14883	0.19375
HoVer-Net [1]	0.8211	0.6321	0.5904

Figure 9 (Annexes) shows the qualitative results obtained for the implementation of the baseline in comparison to the ground truths for images of the test set. Figure 12 (Annexes) shows a variation of the dice coefficient for each of the cell types as a function of the epochs. The complete performed baseline and its discussion can be found in the supplementary material.

3.2. Proposed Method

Figure 8 (Annexes) presents an overview of the evaluated method, proposer for Graham et. al on 2019 [31]. HoverNet architecture [1] consist on the sampling of Patches of 270 x 270 and their normalization; where they apply different transformations to perform data augmentation [24] [31]. The authors generate pixel distance maps of the nuclei to separate the clusters both vertically and horizontally, using the Hover-Net learning model. Then, they perform feature extraction by means of the Resnet50 encoder, followed by three FCN decoders in charge of binary segmentation, distance map predictions and final segmentation respectively [24][31]. The definition of the loss function is the result from the combination of cross entropy, Dice loss and mean squared error [24]. The architecture has three branches that perform simultaneously instance segmentation and nuclei classification [31]. Nuclear pixel (NP) branch, which predicts if a pixel belongs to the nuclei or the background; HoVer branch, that predicts the horizontal and vertical distances of nuclear pixels to their centres of mass; and nuclear classification (NC) branch, which predicts the type of nucleus for each pixel [31]. A previous stage to these branches consist on a block which is composed of four residual blocks and two convolutional layers [31]. Each residual unit includes three convolutional layers, which are interspersed with a batch normalization and an activation stage layer[1]. These are used to down-sampling the patches, in order to reduce the spatial resolution and obtain the general characteristics [31]. Therefore, each of the branches contains a dense decoder unit that will restore different features depending on the branch task [31]

We perform two different studies on the HoverNet network, in order to identify the effect of different parameters on the performance of each branch and as a whole. We start with a variation of multiple hyperparameters such as batch size (8, 16 or 32), optimizer (Adam, AdamW or SGD), learning rate (0.1, 0.01, 0.001 or 1e-8) and the number of epochs, to then conduct an ablation study, where we varied more specific configuration items such as activation function, loss function and the number of residual blocks on the encoder. We also try both studies with ResNet50 and ResNet101 as backbones for fine tunning, and we evaluate the differences of the two of them regarding their properties.

4. EXPERIMENTS

4.1. DATASET

The colorectal nuclear segmentation and phenotypes (CoNSeP) dataset was acquired via Hover-Net paperwork [31], where it was first introduced. The dataset contain 41 H&E stained images from 16 colorectal adenocarcinoma WSIs, each of them belonging to an individual patient [31], and distributed in training and testing sets. All the images have a size of 1,000×1,000 pixels at 40× objective magnification, and were scanned with an Omnyx VL120 scanner within the department of pathology at University Hospitals Coventry and Warwickshire, UK [31]. In order to minimize the omission of individual nuclei segmentation, every nucleus was annotated by one of two expert pathologists, and then, each annotated sample was reviewed by both of the pathologists; therefore refining their own and each others' annotations [31]. The authors focused on a single cancer type - colorectal adenocarcinoma (CRA) - in order to to detail the variation of cell types in tissue, rather than focusing on a small number of visual fields for different types of cancer. Within the dataset, stroma, glandular, muscular, collagen, fat and tumour regions can be observed [31]. Beside incorporating different tissue components, the 41 images were also chosen such that different nuclei types were present, including: normal epithelial; tumour epithelial; inflammatory; necrotic; muscle and fibroblast [31]. The histological images were stained using hematoxylin and eosin (HE). Together, HE enhance the contrast between nuclei, epithelium, and stroma to facilitate the discrimination of structures under a microscope [32]. There is also a significant amount of overlapping nuclei with indistinct boundaries and there exists various artifacts. Given the diversity present on the images, it is likely for the ConNSeP trained models to have a good performance for unseen CRA cases [31].

4.1.1 Training and Testing sets

This dataset contains 41 images from 16 colorectal adenocarcinoma patients, which include 24,319 exhaustively annotated nuclei adenocarcinoma image tiles. The training set was obtained by cropping whole slide images (WSIs) and scanning them at 40x magnification. For performing the annotations, the pathology slide viewing software OMNYX VL120 SCANNER was used. The semantic segmentation annotations contain each type of nucleus colored differently (Red: Malignant/dysplastic ephitelium, Yellow: Miscellaneous, Blue: Fibroblast, Magenta: Inflammatory) [31]. An example of the training set annotations is shown in Figure 1. Annotations were saved as .mat files. For overlapping nuclei, each multi-nuclear pixel was assigned to the largest nucleus containing that pixel [32]. In addition to nuclear boundary annotations, nucleus class labels were provided for each annotated nuclei included in CoNSeP dataset.

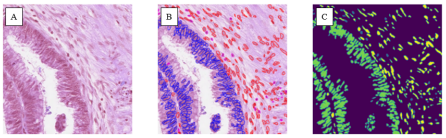


Figure 1: Annotations for images on the training set (A) A sub-image cropped from whole slide image of a patient included in the training set, (B) Boundary annotations of different cell-types done using unique marker colors and (C) Masks generated from the annotations - epithelial cells are shown in red, lymphocytes in yellow, macrophages in green and neutrophils in blue.

CoNSeP dataset training set contains 27 image tiles, each of them with a semantic and instance map (Figure 1 2), stored in the .mat file. Instance map allows to know the number of nuclei present in an image, their size and the area they comprise. The annotations all follow a consistent color map, although each instance is not segmented under a particular color [6]. These are marked and well differentiated, although overlapping of cell groups is observed, where they give greater importance to the one perceived as the one located in the front by the annotator.

Testing set contains 14 image tiles, with its respective semantic and instance segmentation map, as well as a boundary overlay in .png format, where it can be seen the borders surrounding each segmented nuclei with the corresponding color.

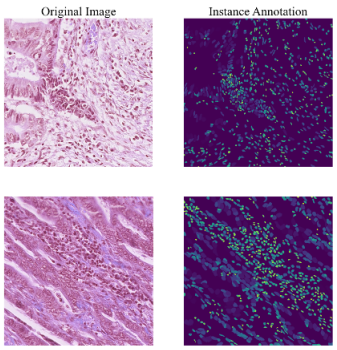


Figure 2: Example of instance annotations for images on the training set

4.1.2 Evaluation Metodology

In order to evaluate the results of the algorithm, the metrics related to segmentation problems used were: Aggregated Jaccard Index (AJI), Dice Score and Panoptic Quality (PQ).

AJI is equal to the ratio of the sums of the cardinals of intersection and union of these matched ground truth and predicted nuclei. The Ensemble Dice Score (DICE2) [31], like Jaccard Index, quantifies the pixel-wise degree of similarity between the model predicted segmentation mask and the ground truth, and ranges from 0 to 1 [33] and it is commonly used to evaluate performance on binary segmentation. PQ measures the quality of a predicted panoptic segmentation relative to the ground truth for each class independently and average over classes [28]. PQ can be seen as the product between segmentation quality (SQ) and detection quality (DQ) (Equation 1) [11]. The detailed description of the metrics used can be found in the supplementary material.

$$PQ = \underbrace{\frac{\sum_{(p,g)\in TP} IoU(p,g)}{|TP|}}_{Segmentation\,Quality\,(SQ)} \; X \; \overbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}^{Detection\,Quality\,(DQ)} \tag{1}$$

4.2. Experiments and results

Table 2 shows the different hyperparameter variations performed on HoverNet with ResNet50 and ResNet101 backbones pretrained with ImageNet. At first, we set our batch size as 16 and try a learning rate (lr) of 0.1, considering that learning rates have an acceptable performance with values between 0.01 and 0.1 [34]. Due to memory availability, the number of epochs was set to 25, and we let the optimization function as default (Adam). This first attempt resulted on empty maps that lead to nan values, to which we decided to decrease the value of the lr. We followed a similar procedure in order to increase the segmentation metric values, to be able to evaluate the one with higher performance. For ResNet50, we obtained that SGD function did not allow the convergence of the model, so we did not use it for ResNet101 experimentation. The best model for ResNet50 was the one obtained by using a batch size of 16, lr of 0.001, Adam optimizer and 50 epochs, which gave a dice score of 0.826, an AJI of 0.501 and a PQ value of 0.465. For ResNet101, we set the parameters to the same values of the best ResNet50 model, and the performances showed an increase of between 0.1 and 0.2. Figure 11 (Annexes) shows the change of dice for each of the cell types using ResNet-50 and ResNet-101 as backbones.

Once we completed the first experimental stage, we manteined the hyperparameters constant and begin to vary other relevant parameters. Table 3 shows the results of the network segmentation performance using three different activation functions. ReLU function is the one the network uses as default, so we were interested to contrast this results to Sigmoid and Tanh functions. With Sigmoid function, we evidenced a decreasing behaviour of the model on the five aspects evaluated, but it was specially drastic on the Dice score, the AJI and the DQ. There was not convergence of the method when Tanh was used.

Table 3: Experimentation with different activation functions for the best models found varying hypermarameters

Activation Function	Dice	AJI	DQ	SQ	PQ
ReLU	**0.82663**	**0.5101**	**0.62786**	**0.75364**	**0.47438**
Sigmoid	0.45378	0.16941	0.27578	0.71103	0.19705
Tanh	-	-	-	-	-

We mantained our best model to continue with the experimentation. The next step was to try a direct modification of the architecture, particularly focused on the down-sampling segment where the residual blocks are located. We removed the last residual block - that is, three residual units, each of them comprising three convolutional layers, for a total of 9 covolutions removed - and watched the performance for both backbones. There was an evident decrease on the first three metrics, which directly affected the PQ value.

Table 4: Ablation study removing residual block on HoverNet's architecture

Residual Blocks	Backbone	Batch Size	Dice	AJI	DQ	SQ	PQ
4	**ResNet-50**	**16**	**0.82683**	**0.50142**	**0.61532**	**0.73362**	**0.46513**
4	**ResNet-101**	**16**	**0.82663**	**0.5101**	**0.62786**	**0.75364**	**0.47438**
3	ResNet-50	32	0.39876	0.10869	0.19932	0.7371	0.14849
3	ResNet-101	12	0.40287	0.18452	0.12561	0.75239	0.09451

We also attempted to vary the value of the optimizer function weight decay, taking the two most common values used for neural networks. The effect of increasing the value from 0.01 to 0.1 in both cases (ResNet50 and ResNet101) was similar, and involved lower metric values in all the five metrics evaluated, with a similar effect on each of them.

Table 5: Weight decay expermentation on the best models found varying hyperparameters

Backbone	Weight Decay	Dice	AJI	DQ	SQ	PQ
ResNet-50	**0.01**	**0.82683**	**0.50142**	**0.61532**	**0.73362**	**0.46513**
ResNet-101	**0.01**	**0.82663**	**0.5101**	**0.62786**	**0.75364**	**0.47438**
ResNet-50	0.1	0.64483	0.31846	0.40029	0.69267	0.27992
ResNet-101	0.1	0.65123	0.24458	0.27811	0.66167	0.18675

Finally, we evaluated the performance of the method by using different variations of the loss function, and obtained that the most drastic decrease of the metrics was when using just $\lambda_c V_c + \lambda_d V_d$ for ResNet50, and $\lambda_b V_b + \lambda_d V_d + \lambda_f V_f$ for ResNet101, with a variation of -0.46 and 0.40 respectively. The best method in both cases was when using the complete combination of λs.

After varying different parameters, we observed that the best metrics were obtained by using ImageNet-ResNet101 as backbone, Adam as optimization function and ReLU as activation function. The comparative results between the

Table 2: Experimentation varying hyperparameters

Backbone	Batch Size	Learning Rate	Optimization Function	Epochs	Dice	AJI	DQ	SQ	PQ
ImageNet-ResNet50	16	0.1	Adam	25	-	-	-	-	-
	8	1.00E-07	Adam	25	0.30934	0.02228	0.00443	0.28502	0.00254
	16	0.01	SGD	7	-	-	-	-	-
	16	0.1	Adam	20	0.80467	0.43847	0.54374	0.73121	0.39968
	16	**0.001**	**Adam**	**50**	**0.82683**	**0.50142**	**0.61532**	**0.73362**	**0.46513**
	16	0.001	ASGD	25	0.19361	0.06197	0.10300	0.75995	0.07795
ImageNet-ResNet101	16	0.01	Adam	25	0.65642	0.33531	0.42735	0.70947	0.30663
	8	1.00E-0.7	Adam	20	-	-	-	-	-
	16	**0.001**	**Adam**	**50**	**0.82663**	**0.5101**	**0.62786**	**0.75364**	**0.47438**
	16	0.1	Adam	25	-	-	-	-	-
	16	0.0001	Adam	15	0.23783	0.04863	0.09075	0.77019	0.07081
	16	0.001	ASGD	25	0.16487	0.06617	0.11028	0.74808	0.08295

Table 6: Ablation study varying the loss strategy

Backbone	Loss Function	Dice	AJI	DQ	SQ	PQ
ImageNet-ResNet-50	$\lambda_a\nu_a + \lambda_b\nu_b + \lambda_c\nu_c + \lambda_d\nu_d + \lambda_e\nu_e + \lambda_f\nu_f$	**0.82683**	**0.50142**	**0.61532**	**0.73362**	**0.46513**
	$\lambda_c\nu_c + \lambda_d\nu_d$	0.00365	0.00181	0	0	0
	$\lambda_a\nu_a + \lambda_c\nu_c + \lambda_e\nu_e$	0.46011	0.14055	0.23919	0.71509	0.17224
	$\lambda_b\nu_b + \lambda_d\nu_d + \lambda_f\nu_f$	0.5519	0.21995	0.31621	0.76119	0.23237
ImageNet-ResNet-101	$\lambda_a\nu_a + \lambda_b\nu_b + \lambda_c\nu_c + \lambda_d\nu_d + \lambda_e\nu_e + \lambda_f\nu_f$	**0.82663**	**0.5101**	**0.62786**	**0.75364**	**0.47438**
	$\lambda_a\nu_a + \lambda_c\nu_c + \lambda_e\nu_e$	0.52146	0.11606	0.21784	0.72194	0.15840
	$\lambda_b\nu_b + \lambda_d\nu_d + \lambda_f\nu_f$	0.23783	0.04863	0.09075	0.77019	0.07081

baseline, our final method and the original hovernet architecture are shown in the Table 7.

Table 7: Comparative results of our final method in comparison with baseline al HoverNet's original results

Method	Dice	AJI	PQ
Baseline	0.50417	0.14883	0.19375
Final Method	**0.82663**	0.5101	0.47438
HoVer-Net Results	0.8211	**0.6321**	**0.5904**

To support the quantitative results obtained through experimentation, we performed quantitative experimentation to visually observe the results. Figure 3 illustrates the predictions obtained for an image of the test set using different backbones (ResNet-50 and ResNet-101) with the best hyperparameters on Table 2. Regarding the ground truth, an incorrect classification of the miscellaneous type is evident.

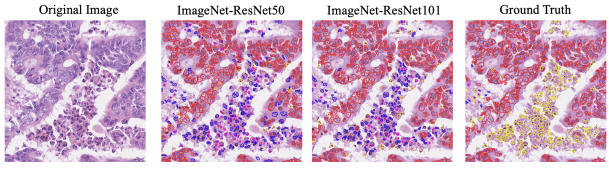


Figure 3: Qualitative results obtained for the best hyperparameters varying the backbone between ImageNet-ResNet-50 and ImageNet-ResNet-101

Figure 4 shows the results in the predictions on the histological images by varying the activation function of ReLU to sigmoid. The qualitative results allow to show the decrease in the metrics (shown in Table 3) when using the sigmoid function.

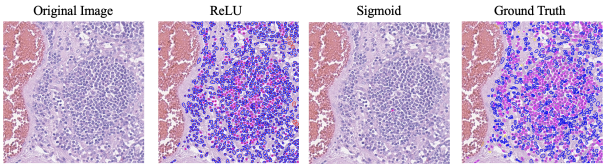


Figure 4: Qualitative results obtained varying the activation function from ReLU to Sigmoid

Figure 5a illustrates the effect of modifying the weight decay from 0.01 to 0.1 on the predictions in the test set. Figure 5b shows how removing a residual unit from the network architecture significantly impairs panoptic segmentation.

Finally, when carrying out all the experimentation, the best model was obtained (Table 7). The qualitative results of this model in comparison with the baseline and with the ground truth are presented in Figure 6.

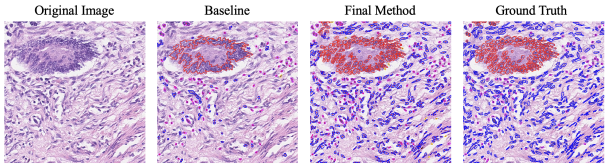
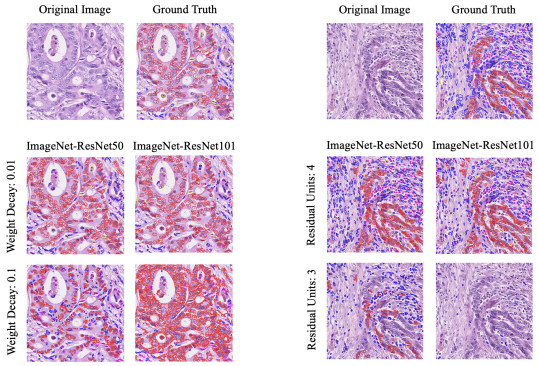


Figure 6: Comparative results of the best model with the baseline and the ground truth



(a) Weight Decay (b) Residual Unit

Figure 5: Comparison of qualitative results when experimenting with weight decay and when removing a residual unit from the architecture.

Graph 7 describes the behavior of the different models implemented with an ImageNet-ResNet101 backbone.

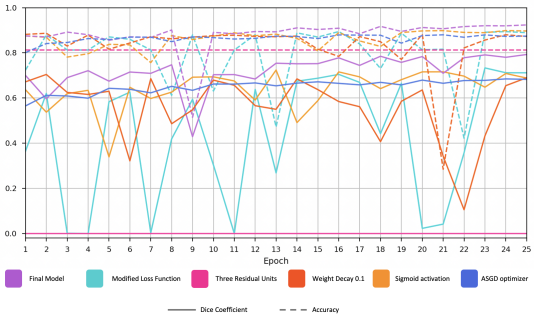


Figure 7: Dice coefficient and accuracy for the different models implemented with ImageNet-ResNet101 backbone.

5. DISCUSSION

This study was developed to evaluate the variance on the behaviour of the HoverNet method [1] according to the modification of particular parameters. The experimental stages were carefuly carried out, but the results obtained can still be improved. Some main limitations detected during the experimental process were associated with the size of the dataset and the dataset used for the pretraining process of the ResNet backbones. Even though the HoverNet method [1] involves data augmentation and other strategies to overcome the absence of enough information to train the network, this is still very biased and could probably lead to overfitting, which is very serious taking into account the problem we are trying to solve. Regarding the backbones pretraining dataset, ImageNet [35], this is a dataset which contains images belonging to various general categories (such as cars, food and landscapes) is a benchmark on machine learning tasks nowadays [35], but it does not represent the categories covered in this study. The use of another dataset to train the mentioned backbones could be useful to increase the general performance of HoverNet on this task. Another general reason is associated with the unbalanced classes present on the dataset, specially when it comes to colorectal cancer nuclei, which can directly affect the performance if the test images contain this low represented classes. Figure 11 (Annexes) shows how the dice and the accuracy vary as the epochs pass with the different backbones. A similar behavior is observed for both. Especially for the *Inflammatory class*, the metrics improve when using ResNet101. For both models, approximately at epoch 8 we observe a decrease of all the metrics since all cells are classified as the same type to improve instance segmentation.

Performance was significantly decreased when using extreme values for the learning rate, when varying the optimizer, and when changing the default activation function. Regarding the learning rate, described as the parameter responsible for the adjustment of the network with respect the loss gradient [36], we can conclude that, when using higher values, the network will not be able to retain the enough information related to the task, while when using very low values, the convergence time could be longer, and it could even get the network stuck on a plateau region [36]. This explains why, for most ResNet101 experiments and some ResNet50, the model never converge and therefore the metrics are not valid numbers. For the type of optimization function, we found that SGD didn't allow the model convergence, but ASGD did. This might be due to the capability of ASGD to adapt the batch size for each parameter in the model and to then use the mean effective gradient as the actual gradient for parameter updates [37]. Adam optimizer, however, showed best performance, which can be attributed to its stochastic optimization technique [38] merged with an adaptive learning rate, and well associated with the ReLU activation function. In relation to the activation function, the best performance was the one that used ReLU function, which is as expected given that ReLU has one of the best performances among all the possible functions. Compared to sigmoid - the other function that gave numeric results - it appears that ReLU correct a problem known as the vanishing gradient in regions where the y values have a minimum reaction to x values, and therefore learning is minimal. When slow learning occurs, the optimization algorithm that minimizes error can be attached to local minimum values and cannot get maximum performance from the model [39].

For the residuals removal experiment, it is evident how the absence of multiple convolutional layers on the downsampling process the information that enters the branches is not so general as to carry out in a specific way each one of the tasks separately, so the performance of each one of

the tasks is significantly reduced. Future ablation studies should include additional experimental stages to also evaluate the effect of the abscense of layers in each one of the branches, in order to see how this may affect the final instance segmentation and classification process. Secondly, weight decay is defined as a way to penalize complexity on deep learning. Generally in the literature the use of wd=0.1 is reported. However, more conservative studies recommend the use of this parameter at 0.01. The reason to choose this value is because if the weight decay is too large (0.1 or more), despite of the training time, the model will never fit well enough. This behavior is clearly observed in the results reported in Table 4 and Figure 5a. When using a weight decay of 0.01, the metrics improve and a better segmentation and classification of the cell nuclei is observed. When increasing the weight decay, the model does not fit correctly and agglomerates different cells as if they were one.

The results presented significant variations when changing the loss function. This function is a measure of how good the prediction model does in terms of being able to predict the expected outcome. The loss function that allowed to obtain the best results is the one that considers the three branches of the architecture (Hover Branch, NP Branch and NC Branch). In it, the different interbranch parameters represent the regression loss. λ parameters are scalars that assign a weight to each function. The loss function values ν_a, ν_c and ν_e denote the mean squared error between the predicted horizontal and vertical distances in comparison to the GT [1]. And ν_b, ν_d and ν_f calculate the mean squared error between the horizontal and vertical gradients of the horizontal and vertical maps respectively and the GT [1]. Therefore, in the Table 6 we experimented considering both groups of parameters in the loss function. It is clearly observed how both the horizontal and vertical distance maps and the gradient maps have a contribution to improve the performance of the network. When implementing them independently, low metrics are obtained that are an indicator of an incorrect fit in the model predictions. For future studies, we propose to use different maps (distance or gradient) in more directions, which allow better differentiation of cell types. Thus, modifying the loss function based on these maps will have a contribution to the performance of the network. When analyzing the effect that modifying the loss function has on the metrics, it is observed in Figure 7 how this generates peaks close to 0 with many occurrences throughout the epochs. This allows to show that the network is not learning correctly to extract the features of the images.

Low metrics are observed in the results when classifying the *Miscellaneous* cell type. This is mainly because it is the minority class in the dataset. Therefore, throughout the training there are not enough patches corresponding to this type of cells for the network to learn them correctly. Therefore, for future studies using this dataset, the classes should be balanced equally to avoid this type of inconvenience. Figure 3 shows how the network does not correctly classify the cells of the Miscellaneous type, but on the contrary, assigns them the labels of the predominant classes in the datset: *Malignant/Dysplastic epithelium* and *Fibroblast*.

Finally, Figure 6 shows how the final method performs the panoptic segmentation task satisfactorily compared to the baseline. The panoptic quality obtained for the final method was 0.47438 and for the baseline 0.19375. In other words, our method supposes an improvement of more than 100% with respect to the baseline. In order to obtain results comparable to those obtained by Graham et al.[1] it is necessary to have computational resources that allow increasing the batch size to increase the number of examples that are introduced in the network to train. In addition, a balance of classes must be carried out.

6. ETHICAL CONSIDERATIONS

Panoptic segmentation of cell nuclei in histological images, despite being a highly relevant task in the medical field for playing an important role in the diagnosis of various diseases, has certain considerations that must be taken into account at the time of its application. One of the aspects that presents some restrictions associated to these methods is related to the way in which the annotations, both semantic and instance, are generated. Although this process is often performed by trained histopathologists, human error is an unavoidable factor that, when using automatic training, can be exponentially increased if an optimal parameter selection process is not performed. This can lead to an incorrect diagnosis of a pathology in a certain number of cases which, depending on the values obtained for the metrics, can be representative. Even if machine learning methods save time and resources, the dependence they still have on human indications, especially in the early stages of training, can have a negative impact on the results and thus - in this particular case - on the diagnosis and treatment of a patient. The accuracy with which the annotations used to train the networks are taken is also a factor that must be carefully reviewed, since histopathological image datasets are very small, and if a very detailed segmentation is performed this will be replicated exponentially in the data augmentation process, which will cause the network to not adjust correctly to new data, leading again misdiagnoses. For this reason, ablation processes that include an extensive experimentation process and a detailed supervision of each step taken by the architecture are essential to reduce these risks, although this does not imply that they are no longer latent. It is also important that certified health entities have a detailed control over the way in which the annotations are made, since supervision will help to reduce the risks that may arise.

References

[1] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical Image Analysis*, p. 101563, 2019.

[2] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012," *International Journal of Cancer*, vol. 136, no. 5, pp. E359–E386, 2015.

[3] M. Mäkinen, "Colorectal serrated adenocarcinoma," *Histopathology*, vol. 50, no. 1, pp. 131–150, 2007.

[4] Y. Kong, G. Z. Genchev, X. Wang, H. Zhao, and H. Lu, "Nuclear segmentation in histopathological images using two-stage stacked u-nets with attention mechanism," *Frontiers in Bioengineering and Biotechnology*, vol. 8, 2020.

[5] T. Hayakawa, V. B. S. Prasath, H. Kawanaka, B. J. Aronow, and S. Tsuruoka, "Computational nuclei segmentation methods in digital pathology: A survey," *Archives of Computational Methods in Engineering*, vol. 28, pp. 1–13, Jan 2021.

[6] N. Kumar, R. Verma, D. Anand, Y. Zhou, O. F. Onder, E. Tsougenis, H. Chen, P. A. Heng, J. Li, and Z. Hu, "Monusac 2020 - grand challenge," 2020.

[7] L. M. Coussens and Z. Werb, "Inflammation and cancer," *Nature*, vol. 420, no. 6917, pp. 860–867, 2002.

[8] K.-H. Chow, R. E. Factor, and K. S. Ullman, "The nuclear envelope environment and its cancer connections," *Nature Reviews Cancer*, vol. 12, pp. 196–209, Mar 2012.

[9] D. Liu, D. Zhang, Y. Song, C. Zhang, F. Zhang, L. O'Donnell, and W. Cai, "Nuclei segmentation via a deep panoptic model with semantic feature fusion," pp. 861–868, 08 2019.

[10] Y. Cui, G. Zhang, Z. Liu, Z. Xiong, and J. Hu, "A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images," *Medical & Biological Engineering & Computing*, vol. 57, pp. 2027–2043, Sep 2019.

[11] R. Ahasan, A. U. Ratul, and A. S. M. Bakibillah, "White blood cells nucleus segmentation from microscopic images of strained peripheral blood film during leukemia and normal condition," in *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 361–366, 2016.

[12] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Nuclei segmentation in histopathology images using deep neural networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 933–936, 2017.

[13] Y. Wu, M. Cheng, S. Huang, Z. Pei, Y. Zuo, J. Liu, K. Yang, Q. Zhu, J. Zhang, H. Hong, D. Zhang, K. Huang, L. Cheng, and W. Shao, "Recent advances of deep learning for computational histopathology: Principles and applications," *Cancers*, vol. 14, no. 5, 2022.

[14] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 550–566, 2016.

[15] C. Nguyen, A. Zuhayr, and H. Yuankai, "Evaluating transformer-based semantic segmentation networks for pathological image segmentation," *arXiv*, 2021.

[16] F. Mahmood, D. Borders, R. J. Chen, G. N. Mckay, K. J. Salimian, A. Baras, and N. J. Durr, "Deep adversarial training for multi-organ nuclei segmentation in histopathology images," *IEEE transactions on medical imaging*, vol. 39, pp. 3257–3267, Nov 2020. 31283474[pmid].

[17] A. G. S. A. K. R. G. Bowen Cheng, Ishan Misra1, "Masked-attention mask transformer for universal image segmentation," 2021.

[18] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," 2017.

[19] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: state of the art," *International Journal of Multimedia Information Retrieval*, vol. 9, pp. 171–189, jul 2020.

[20] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.

[21] G. Neuhold, T. Ollmann, S. Rota Bulò, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in *International Conference on Computer Vision (ICCV)*, 2017.

[22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[23] J. Pont-Tuset, P. Arbelaez, J. T.Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping for image segmentation and object proposal generation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 128–140, jan 2017.

[24] R. Verma, N. Kumar, A. Patil, N. C. Kurian, S. Rane, S. Graham, Q. D. Vu, M. Zwager, S. E. A. Raza, N. Rajpoot, X. Wu, H. Chen, Y. Huang, L. Wang, H. Jung, G. T. Brown, Y. Liu, S. Liu, S. A. F. Jahromi, A. A. Khani, E. Montahaei, M. S. Baghshah, H. Behroozi, P. Semkin, A. Rassadin, P. Dutande, R. Lodaya, U. Baid, B. Baheti, S. Talbar, A. Mahbod, R. Ecker, I. Ellinger, Z. Luo, B. Dong, Z. Xu, Y. Yao, S. Lv, M. Feng, K. Xu, H. Zunair, A. B. Hamza, S. Smiley, T.-K. Yin, Q.-R. Fang, S. Srivastava, D. Mahapatra, L. Trnavska, H. Zhang, P. L. Narayanan, J. Law, Y. Yuan, A. Tejomay, A. Mitkari, D. Koka, V. Ramachandra, L. Kini, and A. Sethi, "Monusac2020: A multi-organ nuclei segmentation and classification challenge," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3413–3423, 2021.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.

[26] e. Russakovsky, O., "Imagenet large scale visual recognition challenge," 2015.

[27] L. Putzu and G. Fumera, "An empirical evaluation of nuclei segmentation from hamp;e images in a real application scenario," *Applied Sciences*, vol. 10, no. 22, 2020.

[28] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar, "Panoptic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[29] N. Kumar, R. Verma, and A. Sethi, "Monuseg - grand challenge," *MoNuSeg Grand Challenge*, 2018.

[30] N. Kumar, R. Verma, D. Anand, Y. Zhou, O. F. Onder, E. Tsougenis, H. Chen, P.-A. Heng, J. Li, Z. Hu, Y. Wang, N. A. Koohbanani, M. Jahanifar, N. Z. Tajeddin, A. Gooya, N. Rajpoot, X. Ren, S. Zhou, Q. Wang, D. Shen, C.-K. Yang, C.-H. Weng, W.-H. Yu, C.-Y. Yeh, S. Yang, S. Xu, P. H. Yeung, P. Sun, A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, O. Smedby, C. Wang, B. Chidester, T.-V. Ton, M.-T. Tran, J. Ma, M. N. Do, S. Graham, Q. D. Vu, J. T. Kwak, A. Gunda, R. Chunduri, C. Hu, X. Zhou, D. Lotfi, R. Safdari, A. Kascenas, A. O'Neil, D. Eschweiler, J. Stegmaier, Y. Cui, B. Yin, K. Chen, X. Tian, P. Gruening, E. Barth, E. Arbel, I. Remer, A. Ben-Dor, E. Sirazitdinova, M. Kohl, S. Braunewell, Y. Li, X. Xie, L. Shen, J. Ma, K. D. Baksi, M. A. Khan, J. Choo, A. Colomer, V. Naranjo, L. Pei, K. M. Iftekharuddin, K. Roy, D. Bhattacharjee, A. Pedraza, M. G. Bueno, S. Devanathan, S. Radhakrishnan, P. Koduganty, Z. Wu, G. Cai, X. Liu, Y. Wang, and A. Sethi, "A multi-organ nucleus segmentation challenge," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1380–1391, 2020.

[31] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical Image Analysis*, vol. 58, p. 101563, 2019.

[32] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1550–1560, 2017.

[33] L. Baskaran, S. Al'Aref, G. Maliakal, B. Lee, J. Choi, S.-E. Lee, J. Sung, F. Lin, S. Dunham, B. Mosadegh, Y.-J. Kim, I. Gottlieb, B. Lee, E. Chun, F. Cademartiri, E. Maffei, H. Marques, S. Shin, and L. Shaw, "Automatic segmentation of multiple cardiovascular structures from cardiac computed tomography angiography images using deep learning," *PLOS ONE*, vol. 15, p. e0232573, 05 2020.

[34] Y. B. Ian Goodfellow and A. Courville, *Deep Learning (Adaptive Computation and Machine Learning series)*. 2019.

[35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[36] H. Zulkifli, "Understanding learning rates and how it improves performance in deep learning," 2018.

[37] H. Shi, N. Yang, H. Tang, and X. Yang, "asgd: Stochastic gradient descent with adaptive batch size for every parameter," *Mathematics*, vol. 10, p. 863, 03 2022.

[38] M. Alom, “Adam optimization algorithm,” 06 2021.

[39] A. Kızrak, “Comparison of activation functions for deep neural networks,” 2019.

7. CREDITS

Both authors contributed equally to the selection of the project problem, to the searching of information and the redaction of the introduction and the state of the art and other methods described in the related work. The understanding of the method, the structure of the base repository, and the logic behind the steps suggested was also equally carried out by both of the members. Juanita Puentes was in charge of experimentation with ResNet50 backbone, and the model obtained for training without a backbone (corresponding to the baseline); while Laura Acosta was responsible for the ResNet101 weights search and corresponding experimentation. The description of the results, discussion and ethical considerations were done by both of the authors.

We certify that all the members of the group had an equivalent contribution to the project

8. Annexes

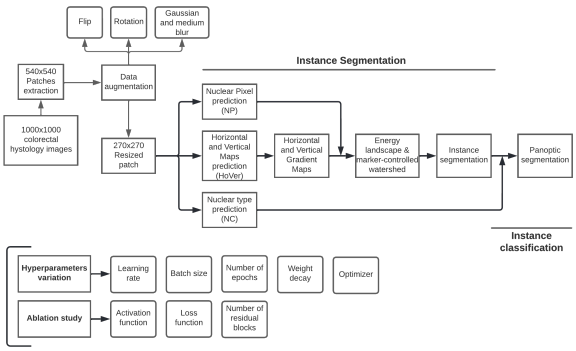


Figure 8: Proposed method overview for nuclei instance segmentation and classification

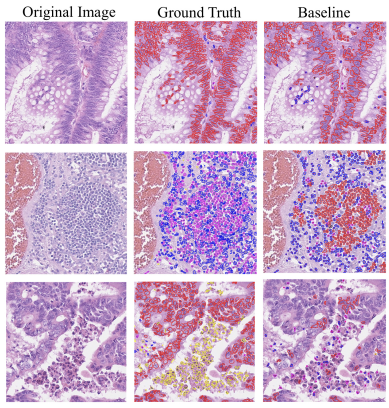


Figure 9: Predictions for panoptic segmentation obtained for the baseline. Three images of the test set are included in the figure, along with their ground truths and prediction.

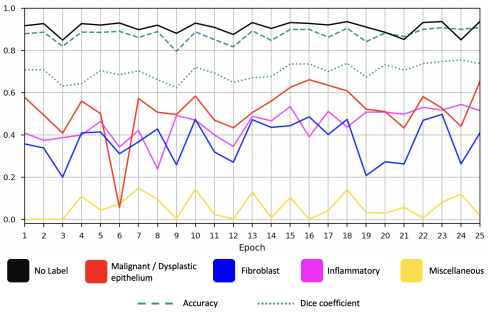


Figure 10: Variation of the dice coefficient metrics for each of the cell types as a function of the number of epochs using the model implemented for the baseline. The general accuracy is also included.

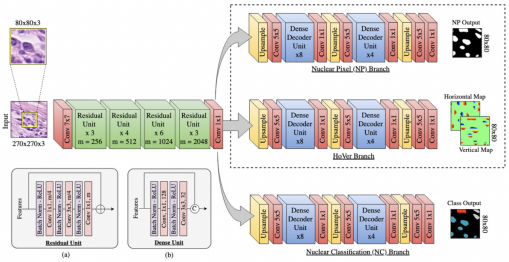


Figure 12: HoVer-Net architecture [1]

Made on LaTeX

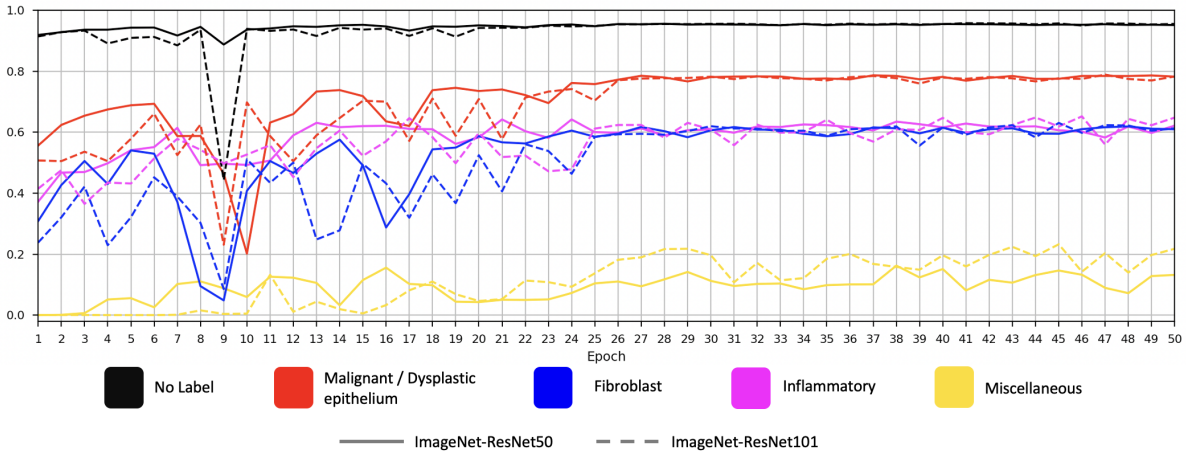


Figure 11: Variation of the dice coefficient for each type of cell using different backbones as a function of the number of epochs.