

Supplementary Material: 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

1. RELATED WORK

1.1. 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 [1], with variations on the base architectures and the pre-processing methods. Fourteen teams used variants of U-Net with different base architectures, while three teams used feature pyramid network and EfficientNetB7 as an encoder; while pixel-wise cross entropy loss, Dice loss, or a hybrid of the two were the most common loss functions used [1]. Pre-processing techniques included image color normalization or color jitter, and data augmentation was also performed over the training set. Post-processing allowed to solve the overlapping nuclei problem, to which Watersheds was used for binary segmentation and neural networks on the other cases. Markers were based on either maps of the nucleus centers, or maps of the distance from or the direction towards the nearest nucleus center [1].

SJTU 426 team [1], 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 [1]. 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, combined with cross-entropy and perceptual loss functions [1][2]. They train a second U-Net (VGG16 architecture) to perform the final classification making use of the cross-entropy related loss [1] [2][3]. Other good approaches are the ones made by the SharifHooshPardaz and The Great Backpropagator teams [1], 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 [1].

1.2. Datasets

Within the existing datasets of histopathological images to carry out the panoptic segmentation of cell nuclei are MIC17[4] and BNS. This first dataset (MIC17) was provided for the MICCAI 2017 Digital Pathology Challenge [4]. It includes 32 annotated squared image patches of sizes 500×500 cropped from HE-stained histopathology Whole Slide Images (WSIs) [4]. 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) [4]. As an evaluation metrics, the Dice Score Coefficient (DSC) and Hausdorff distance were used for semantic segmentation. Some challenges, such as the MoNuSAC, propose a particular metric to evaluate the performance of this task. Panoptic Quality (PQ) is one of the most frequent metrics [5]. Expanding this evaluation scheme, others authors used metrics of *Sensitivity* and *Specificity* [4]. These metrics allowed to determine potential over- or under-segmentations of the tumor sub-regions [4]. The dataset used for MoNuSeg (2018) [6] contains images from seven organs with about 95,000 different nuclei [7] [6]. This dataset was created by downloading HE stained tissue images captured at 40x magnification from TCGA archive [6] [7]. 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 [6][7]. The AJI is equal to the ratio of the sums of the cardinals of intersection and union of these matched ground truth and predicted nuclei [7].

2. APPROACH

2.1. Baseline

For the first approach to solve the panoptic segmentation task we implement HoVer-Net [8]: a deep neural net-

work 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 [8] 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
<i>Baseline</i>	0.50417	0.14883	0.19375
<i>HoVer-Net [8]</i>	0.8211	0.6321	0.5904

Figure 1 shows the qualitative results obtained for the implementation of the baseline in comparison to the ground truths for images of the test set.

The images show the results of panoptic segmentation since they differentiate between cell nuclei and a marker with a different color is used for each type of cell. The qualitative results show that the network proposed in the baseline does not correctly segment the different nuclei, that is, its performance is not favorable when segmenting instances. This is evidenced by the image in the first row of Figure 1, where the cell nuclei of the *Malignant/Dysplastic epithelium* type are segmented together. In the image of the second row, it is clearly observed how the baseline classifies cells of the *Inflammatory* type as *Malignant/Dysplastic epithelium*. However, the *Fibroblasts* were classified correctly for the most part. Finally, for the image present in the third row of Figure 1, we observe that the model does not correctly segment the cells of type *Miscellaneous*. This may be the result of being the class with fewer cell nuclei, therefore, when training and validating, the network does not learn correctly to segment this cell line.

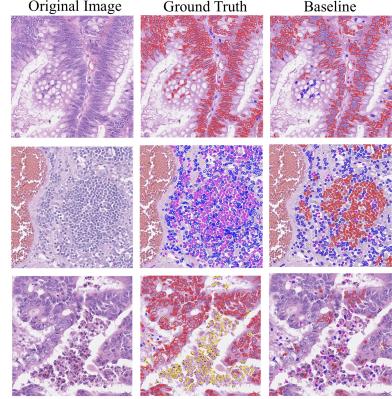


Figure 1: Predictions for panoptic segmentation obtained for the baseline. Three images of the test set are included in the figure, along with their ground truths and prediction.

Figure 2 shows a variation of the dice coefficient for each of the cell types as a function of the epochs. This graph shows how class 0 (belonging to areas of the images to which a label is not assigned) has a dice coefficient close to 0.9. For the other classes, variable metrics are observed throughout the epochs. Similar metrics are observed for the *Malignant/Dysplastic epithelium*, *Fibroblast* and *Inflammatory* cell types, ranging mainly between 0.3 and 0.6. It is important to mention that the *Malignant/Dysplastic epithelium* class presents the best metrics, probably since it is the most abundant class in all the histopathological images of this dataset. The *Miscellaneous* class was not segmented correctly and has metrics ranging between 0 and 0.1. As mentioned above, this is the result of being the minority class in the dataset. Based on the conclusions of the baseline, we choose the parameters to experiment in our final method.

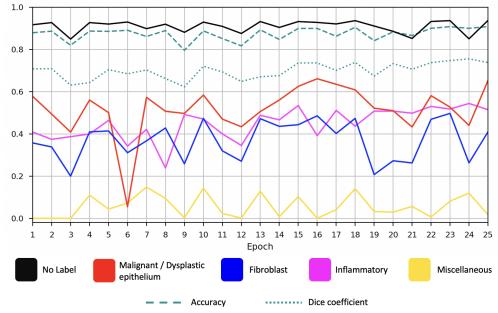


Figure 2: Variation of the dice coefficient metrics for each of the cell types as a function of the number of epochs using the model implemented for the baseline. The general accuracy is also included.

2.2. Proposed Method

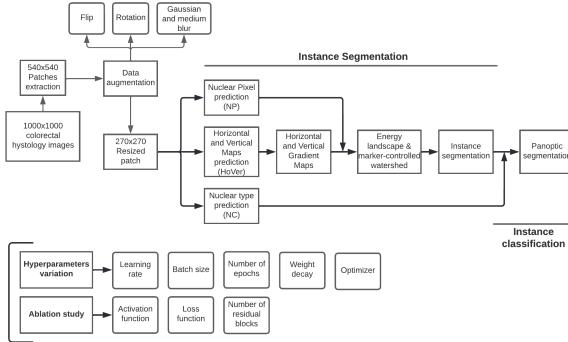


Figure 3: Proposed method overview for nuclei instance segmentation and classification

Figure 3 presents an overview of the evaluated method, proposed by Graham et al. in 2019 [9]. The detailed description of the method and architecture can be found on the main paper.

3. EXPERIMENTS

3.1. DATASET

3.1.1 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). The Aggregated Jaccard Index (introduced by Kumar et al.) is more appropriate to evaluate algorithms for this instance segmentation problems. It involves matching every ground truth nuclei to one detected nuclei by maximizing the Jaccard index. The AJI is then 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) [9], 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 [10] and it is commonly used to evaluate performance on binary segmentation. The associated expression to the DICE2 metric is seen on 1

$$\text{Dice Score} = \frac{2 * TP}{2 * TP + FP + FN} \quad (1)$$

PQ is used with the aim of evaluating both semantic segmentation and instance classification accuracy [5]. PQ measures the quality of a predicted panoptic segmentation relative to the ground truth for each class independently and average over classes [5]. PQ represents the average Jaccard Index for every image (p) and every class nuclei (g), where non-matching segments are penalized [11].

$\frac{1}{|TP|} \sum_{(p,g) \in TP} IoU(p, g)$ represents the average IoU of matched segments, and the denominator $\frac{1}{2}|FP| + \frac{1}{2}|FN|$ penalizes segments without matches [11].

$$PQ = \frac{\sum_{(p,g) \in TP} IoU(p, g)}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|} \quad (2)$$

PQ can be seen as the product between segmentation quality (SQ) and detection quality (DQ) (Equation 3) [11]. DQ is the used for detection settings and SQ is the average IoU of matched segments. There are two sources of void labels in the ground truth: (a) out of class pixels and (b) ambiguous or unknown pixels represented specially on the testing dataset [5].

$$PQ = \underbrace{\frac{\sum_{(p,g) \in TP} IoU(p, g)}{|TP|}}_{\text{Segmentation Quality (SQ)}} \times \overbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}^{\text{Detection Quality (DQ)}} \quad (3)$$

The value of Jaccard Index and PQ ranges between 0 to 1. Obtaining a value of 0 means all nucleus instances were segmented incorrectly (no-overlap between the annotation and the segmented nuclei) at the pixel-level, while 1 means a perfect algorithm output [11]. For each image in the test set, the weighted panoptic quality will be computed as the weighted sum of the class specific panoptic quality, where the weights for the four cell-types are given according to their grade of representation on the training and testing sets [11].

References

- [1] 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.
- [2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014.

- [3] e. Russakovsky, O., “Imagenet large scale visual recognition challenge,” 2015.
- [4] L. Putzu and G. Fumera, “An empirical evaluation of nuclei segmentation from histology images in a real application scenario,” *Applied Sciences*, vol. 10, no. 22, 2020.
- [5] 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.
- [6] N. Kumar, R. Verma, and A. Sethi, “Monuseg - grand challenge,” *MoNuSeg Grand Challenge*, 2018.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.

- [11] 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.

4. CREDITS

Both authors contributed equally to the development of the supplementary material file.

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

5. Annexes

Made on L^AT_EX