

# Channel-wise Aggregation with Self-correction Mechanism for Multi-center Multi-Organ Nuclei Segmentation in Whole Slide Imaging

Mohamed Abdel-Nasser<sup>1,3</sup><sup>a</sup>, Adel Saleh<sup>2</sup> and Domenech Puig<sup>1</sup><sup>b</sup>

<sup>1</sup>Computer Engineering and Mathematics Department, University Rovira i Virgili, Tarragona, Spain

<sup>2</sup>Gaist Solutions Ltd., U.K.

<sup>3</sup>Electrical Engineering Department, Aswan University, Aswan, Egypt

**Keywords:** Computational Pathology, Nuclei Segmentation, Whole Slide Imaging, Deep Learning.

**Abstract:** In the field of computational pathology, there is an essential need for accurate nuclei segmentation methods for performing different studies, such as cancer grading and cancer subtype classification. The ambiguous boundary between different cell nuclei and the other objects that have a similar appearance beside the overlapping and clumped nuclei may yield noise in the ground truth masks. To improve the segmentation results of cell nuclei in histopathological images, in this paper, we propose a new technique for aggregating the channel maps of semantic segmentation models. This technique is integrated with a self-correction learning mechanism that can handle noisy ground truth. We show that the proposed nuclei segmentation method gives promising results with images of different organs (e.g., breast, bladder, and colon) collected from medical centers that use devices of different manufacturers and stains. Our method reaches the new state-of-the-art. Mainly, we achieve the AJI score of 0.735 on the Multi-Organ Nuclei Segmentation benchmark, which outperforms the previous closest approaches.

## 1 INTRODUCTION


Currently, digital pathology has an essential role in clinics and laboratories. Thousands of tissue biopsies are taken from cancer patients yearly, and in turn, the whole-slide imaging (WSI) technique permits the acquisition of high-resolution images of slides. Cell segmentation indicates the segmentation of the cell nuclei. Notably, there is a necessity for precise nuclei segmentation techniques in the context of computational pathology for facilitating the extraction of descriptors for the morphometrics of cell nuclei. Several descriptors, such as cell nuclei shape and number of cell nuclei in WSI images, can be used to conduct various studies such as the determination of cancer types, cancer grading, and prognosis (Moen et al., 2019).


Indeed, there is a diverse tissue type, variations in staining and cell type leads to different visual characteristics of WSI images. This makes the segmentation of nuclei segmentation a challenging task. Nuclei segmentation task necessitates a vast effort to manually create the pixel-wise annotations that can be used for training machine learning techniques. There are nuclei segmentation toolboxes available in Cell Pro-

filer (Carpenter et al., 2006) and ImageJ-Fiji (Schindelin et al., 2012). However, the visual characteristics of WSI images makes it very difficult to develop traditional image processing based segmentation algorithms that give acceptable nuclei segmentation results with WSI images taken from several cancer patients and collected at different medical centers for various organs, such as breast, colon, lung, and stomach (Niazi et al., 2019).

Recently, a review and comprehensive comparison are presented in (Vicar et al., 2019) for cell segmentation methods for label-free contrast microscopy. These segmentation methods studied are categorized as follows: 1) single-cell segmentation methods, 2) foreground segmentation methods, such as thresholding, feature-extraction, level-set, graph-cut, machine learning-based, and 3) seed-point extraction methods, namely Laplacian of Gaussians, radial symmetry and distance transform, iterative radial voting, maximally stable extremal region. The study concluded that the machine learning-based methods give accurate segmentation results.

In the last years, deep learning models have been employed for performing different segmentation tasks in biology (Niazi et al., 2019). Naylor et al. (Naylor et al., 2017) introduced a fully automated method for cell nuclei segmenting in histopathology

<sup>a</sup>  <https://orcid.org/0000-0002-1074-2441>

<sup>b</sup>  <https://orcid.org/0000-0002-0562-4205>

images based on three semantic segmentation models: PangNet, a fully convolutional network (FCN) and DeconvNet. They ensembled the three semantic segmentation models and obtained an F1-score of 0.80. In (Al-Kofahi et al., 2018), a three-step cell nuclei segmentation approach is proposed: 1) the detection of the cells using a deep learning-based model to obtain pixel probabilities for nuclei, cytoplasm, as well as background, 2) the separation of touching cells based on blob detection and shape-based watershed techniques that can distinguish between the individual nuclei from the nucleus prediction map and 3) the segmentation of the nucleus and cytoplasm). With four different datasets, they obtained an accuracy of 0.84.

In (Qu et al., 2019), a weakly supervised deep nuclei segmentation using points annotation in histopathology images is proposed. In this study, the original WSI images and the shape prior of nuclei are employed to obtain two types of coarse labels from the points annotation using the Voronoi diagram and the k-means clustering algorithm. These coarse labels are used to train a deep learning model, and then the dense conditional random field is utilized in the loss function to fine-tune the trained model. With the multi-organ WSI dataset, a dice score of 0.73 is achieved. In (Mahmood et al., 2019), a conditional generative adversarial network (cGAN) model is proposed for nuclei segmentation. A large dataset of synthetic WSI images with perfect nuclei segmentation labels is generated using an unpaired GAN model. Both synthetic and real data with spectral normalization and gradient penalty for nuclei segmentation are used to train the cGAN model.

In (Zhou et al., 2019), a deep learning-based model called contour-aware informative aggregation network (CIA-Net) with a multilevel information aggregation module between two task-specific decoders. Instead of using independent decoders, this model exploits bi-directionally aggregating task-specific features to merge the advantages of spatial and texture dependencies between nuclei and contour. Besides, a smooth truncated loss that modulates losses is utilized to mitigate the perturbation from outliers. As a result, the CIA-Net model is almost built using informative samples, and so its generalization capability could be enhanced (i.e., with multi-organ multi-center nuclei segmentation tasks). With the 2018 MICCAI challenge of the multi-organ nuclei segmentation dataset, they produced a Jaccard score of 0.63.

In addition, in (Wang et al., 2019), a multi-path dilated residual network is proposed for nuclei segmentation and detection. The network includes the following: 1) multi-scale feature extraction based on D-

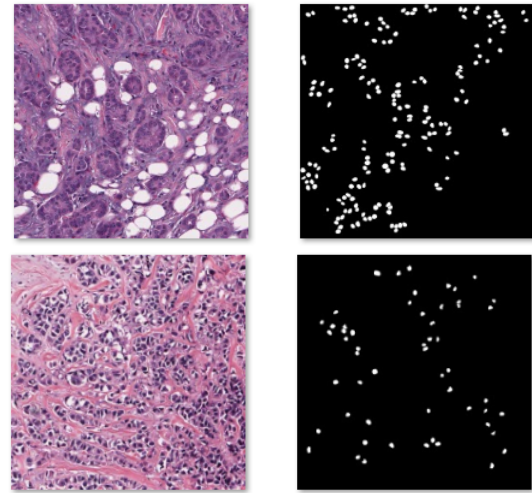


Figure 1: WSI images.

ResNet and feature pyramid network (FPN), 2) candidate region network, and 3) a final network for detection and segmentation. The detection and segmentation network involves three parts: segmentation, regression, and classification sub-networks. With the MonuSeg dataset, an aggregated Jaccard index (AJI) of 0.46 is obtained.

Although methods above achieved promising cell nuclei segmentation results, the ambiguous boundary between different cell nuclei and the other objects that have a similar appearance beside the overlapping and clumped nuclei may yield noise in the ground truth masks (see Fig. 1). To cope with these issues, in this paper, we propose a new technique for aggregating the channel maps of semantic segmentation models. This technique is integrated with a self-correction learning mechanism that can handle noisy ground truth. Notably, we do not claim any novelty of the self-correction mechanism or the channel-wise aggregation mechanism, but only the superiority of the performance of the proposed method with the nuclei segmentation task.

The rest of this paper is organized as follows. Section 2 presents the proposed methods. Section 3 provides the results and discussion. Section 4 concludes the paper and gives some points of future work.

## 2 METHODOLOGY

The ambiguous boundary between different cell nuclei and the other objects that have a similar appearance beside the overlapping and clumped nuclei may yield noise in the ground truth masks. To improve the segmentation results of cell nuclei in histopathological images, we propose a new technique for ag-

gregating the channel maps of semantic segmentation models. This technique is integrated with a self-correction learning mechanism that can handle noisy ground truth.

## 2.1 Channel-wise Aggregation Mechanism

The aggregation operators reduces a set of numbers  $(x_1, x_2, \dots, x_n)$  into a single representative number  $y$ . This operation can be expressed as follows (Lucca et al., 2017):

$$y = A(x_1, x_2, \dots, x_n) \quad (1)$$

here  $A : [0, 1]^n \times [0, 1] \rightarrow [0, 1]$  is said to be aggregation function **Iff** it has the following conditions:

- **Identity in the Case of Unary:**  $A(x) = x$
- **Boundary Conditions:**  $A(0_1, \dots, 0_n) = 0$  and  $agg(1_1, \dots, 1_n) = 1$
- **Non Decreasing:**  $A(x_1, \dots, x_n) \leq A(y_1, \dots, y_n)$  when  $(x_1, \dots, x_n) \leq (y_1, \dots, y_n)$

For the sake of simplicity, in this paper, we use the maximum operator ( $\max\{x_1, \dots, x_n\}$ ). In the decoder part of the segmentation model, we apply a element-wise max. aggregation function on the channel maps. If we have channel maps  $\{C_1, C_2, \dots, C_n\}$ , the aggregation function will produce one channel map ( $C_{agg}$ ). Then, we concatenate the  $C_{agg}$  with the original feature maps and feed all together to the next layers as follows  $\{C_{agg}, C_1, C_2, \dots, C_n\}$ .

## 2.2 Self Correction Training Mechanism

The self-correction training strategy (Li et al., 2019) can be employed to aggregate a segmentation model and labels, which can enhance the performance of the segmentation model and the ground-truth labels in an iterative manner. The improvement that can be achieved by the self-correction training strategy depends on the initial segmentation results of the basic segmentation model. It is important to note that if the initial segmentation results are not accurate, they may worsen the self-correction training strategy. Thus, the self-correction strategy should be begun after the training loss begins to take a flat shape.

After we get good segmentation results with the basic segmentation model, a cyclically learning scheduler with warm restarts is used. Here, a cosine annealing learning rate scheduler with cyclical restart

(Loshchilov and Hutter, 2016) that can be mathematically formulated as follows:

$$\phi = \phi_{\min} + \frac{1}{2}(\phi_{\max} - \phi_{\min}) \left( 1 + \cos \left( \frac{EP_{res}}{EP} \pi \right) \right) \quad (2)$$

where  $EP$  is the number of epochs in each cycle,  $EP_{res}$  indicates the number of epochs gone since the previous restart,  $\phi_{\max}$  is the initial learning rate, and  $\phi_{\min}$  is the final learning rate.

After each cycle of the self-correction mechanism, we obtain a set of weights (models)  $\theta = \{\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_T\}$ , and the corresponding predicted labels  $Y = \{\hat{y}_0, \hat{y}_1, \dots, \hat{y}_T\}$ , where  $T$  is the number of training cycles. After each training cycle, the current model weights  $\hat{\theta}$  are aggregated with the weights of previous cycle  $\hat{\theta}_{t-1}$  in order to obtain new weights  $\hat{\theta}_t$  as follows:

$$\hat{\theta}_t = \frac{t}{t+1} \hat{\theta}_{t-1} + \frac{1}{t+1} \hat{\theta} \quad (3)$$

Similarly, the the ground-truth labels are be aggregated as follows:

$$\hat{y}_t = \frac{t}{t+1} \hat{y}_{t-1} + \frac{1}{t+1} \hat{y} \quad (4)$$

where  $t$  refers to the number of the current cycle ( $0 \leq t \leq T$ ), and  $\hat{y}$  is the generated pseudo-labels (pseudo-masks) with the model  $\hat{\theta}_t$ .

## 2.3 Basic Framework for Conducting Cell Nuclei Segmentation

The self-correction training used in his paper utilizes the A-CE2P model (Ruan et al., 2019) as the basic framework for conducting cell nuclei segmentation. As shown in Fig. 2, the CE2P model comprises three branches: segmentation branch (top), fusion branch (middle), and edge branch (bottom). The operation of A-CE2P can be mathematically formulated as follows:

$$\mathcal{E} = \alpha_1 \mathcal{E}_{edge} + \alpha_2 \mathcal{E}_{parsing} + \alpha_3 \mathcal{E}_{consistent} \quad (5)$$

where  $\alpha_1, \alpha_2$  and  $\alpha_3$  are hyper-parameters to control the contribution among these three losses. The CE2P model is jointly trained in an end-to-end manner by minimizing  $\mathcal{E}$ .

For an input WSI image  $img$ , assume that the cell nuclei ground truth label is  $\hat{y}_i^{cn}$  and the predicted mask is  $y_i^{cn}$ , where  $cn$  refers to the number of pixels for class  $i$ . The pixel-level supervised nuclei segmentation task can be expressed using the cross-entropy loss as follows:

$$\mathcal{E}_e = -\frac{1}{N} \sum_i \sum_{cn} \hat{y}_i^{cn} \log p(y_i^{cn}) \quad (6)$$

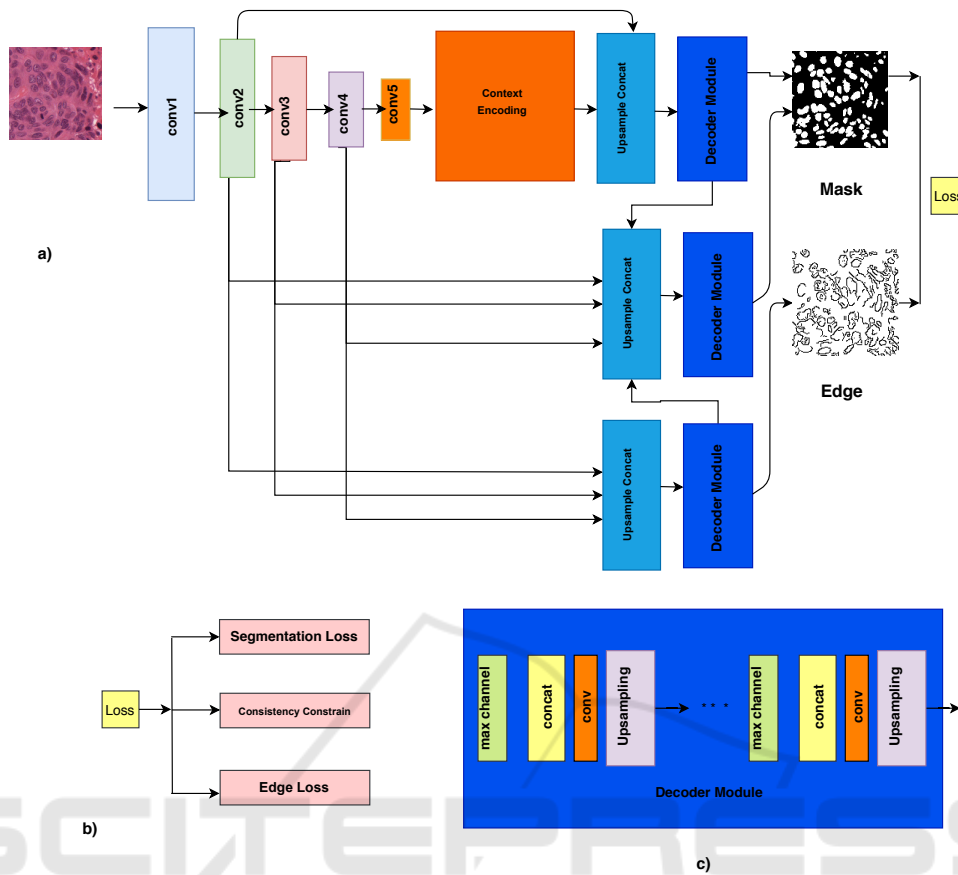


Figure 2: Self correction training mechanism.

where  $N$  is the number of pixels,  $K$  is the number of classes. To let the model facilitate the mean intersection-over-union (mIoU) directly, the cross-entropy loss and the mIoU loss are combined as follows:

$$\mathcal{E}_{\text{cells}} = \mathcal{E}_e + \mathcal{L}_{\text{mIoU}} \quad (7)$$

To preserve the consistency between the predicted nuclei segmentation masks and the boundary prediction, the following constraint term is exploited:

$$\mathcal{E}_{\text{consistent}} = \frac{1}{|N|} \sum_{n \in N^{\text{pos}}} |\tilde{edg}^n - edg^n| \quad (8)$$

In this expression,  $N$  is the number of positive edge pixels,  $edg^{cn}$  refers to the edge maps produced by the edge branch and  $\tilde{edg}^{cn}$  is the edge maps produced by the nuclei segmentation branch  $y_i^{cn}$ . When computing the loss, we should avoid that non-edge pixels dominate the loss. To do so, the non-edge pixels are suppressed, and only positive edge pixels  $cn \in N^{\text{pos}}$  are allowed to contribute to the consistency term. Here, the ResNet-101 (He et al., 2016) is used as the backbone of the feature extractor with the ImageNet pre-trained weights.

## 3 RESULTS AND DISCUSSION

### 3.1 Evaluation Metrics

Aggregated Jaccard Index (AJI) is proposed in (Kumar et al., 2017) assessing the performance of nuclei segmentation methods. If AJI equals 1, it means that we obtain perfect nuclei segmentation results. AJI is an modified version of the Jaccard index that divides the aggregated intersection cardinality by the aggregated union cardinality in the ground truth and segmented masks. AJI can be expressed as follows:

$$AJI = \frac{\sum_{i=1}^L |GT_i \cap NP_j^*(i)|}{\sum_{i=1}^K |GT_i \cup NP_j^*(i)| + \sum_{k \in \text{Ind}} |NP_k|} \quad (9)$$

In this expression,  $GT = \bigcup_{i=1,2..K} GT_i$  is the ground-truth of the nuclei pixels,  $NP = \bigcup_{j=1,2..L} NP_j$  are the prediction nuclei segmentation results,  $NP_j^*(i)$  is the connected component from the prediction result that maximize the Jaccard index, and  $\text{Ind}$  is the list of indices of pixels that do not belong to any component in the GT.

The F1-score is the harmonic mean between precision and recall, which can be formulated as follows:

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (10)$$

where TP, FP, FN are the true positive, false positive, and false negative rates, respectively. The true negative (TN) rate is defined as  $TN = \overline{GT} \cap \overline{NP}$ , which is the area not belonging to any of the two masks  $GT$  and  $NP$ .

### 3.2 Experimental Results and Discussion

The dataset used in this study has been obtained from (Kumar et al., 2017). This dataset includes 30 WSI images with annotations from 7 organs (breast, kidney, colon, stomach, prostate, liver, and bladder) collected at different medical centers. The size of each image is  $1000 \times 1000$ .

The test data includes one image from every organ that was not exposed to the network. The rest of images is used for training. Every image in the training and testing data was scaled to  $1024 \times 1024$  and divided into four non-overlapping patches of size  $512 \times 512$ . Further random cropping of  $512 \times 512$  from every image was applied, as well. The overall training data had 4906 of  $512 \times 512$  patches. A batch size of one image was used due to the limitation in resources, namely the GPU memory. Proposed model was trained for 50 epoch. The stochastic gradient descent (SGD) was used as an optimizer with an initial learning rate as  $1e-1$ , momentum as 0.99 and weight decay as  $1e-8$ . A Titan X GPU was used to run the experiments. Table 1 shows the results of the proposed method and ones of five state-of-the-art semantic segmentation based on deep learning models nuclei segmentation. The five models are Fully Convolutional Network (FCN), U-Net, Mask R-CNN, and conditional GAN (cGAN). As shown the proposed method achieves an F1-score of 0.876 and AJI score of 0.735. These results are better than the ones of the previous approach (Mahmood et al., 2019). We also show that the addition of the channel wise aggregation improves the performance of the baseline framework (self attention mechanism with CE2P).

Figure 3 shows the segmentation results of the proposed method with different organs: breast, kidney, liver, prostate, bladder, colon, and stomach. As shown, our method produces good segmentation results with bladder and stomach histopathological images with AJI scores of 0.85 and 0.83, respectively. The proposed method gives a segmentation results lower than 0.67 with the liver image because of the apparent overlap between several cell nuclei.

Table 1: Comparison between the proposed model and the related methods: FCN, U-Net, Mask R-CNN, and cGAN.

Method	F1-score	AJI
FCN (Long et al., 2015)	0.35	0.35
U-Net (Ronneberger et al., 2015)	0.41	0.41
Mask R-CNN (He et al., 2017)	0.50	0.50
cGAN (Mahmood et al., 2019)	0.87	0.72
Baseline	0.88	0.73
<b>Proposed</b>	<b>0.89</b>	<b>0.74</b>

Figure 4 shows a comparison between the number of cell nuclei in the predicted masks and the corresponding ground truth. As shown, the number of cell nuclei obtained by the proposed method is a bit higher than the ones of the ground-truth.

The proposed model gives promising segmentation results when we have noisy ground truth masks because of the ambiguous boundary between different cell nuclei and the other objects that have a similar appearance beside the overlapping and clumped. If the cell nuclei ground-truth is almost clean, the channel-wise aggregation, label refinement, and the self-correction training mechanisms can be seen as an ensembling of clones of the basic segmentation model (i.e., CE2P), which would improve the cell nuclei segmentation results and produce a generalized model that can be used with images of different organs acquired at different medical centers.

## 4 CONCLUSION

In this paper, we propose a new technique for aggregating the channel maps of semantic segmentation models in order to improve the segmentation results of cell nuclei in histopathological images. This technique is integrated with a self-correction learning mechanism that can handle noisy ground truth. We show that the proposed nuclei segmentation method gives promising results with images of different organs (e.g., breast, bladder, and colon) collected from medical centres that use devices of different manufacturers and stains. Our method achieves the new state-of-the-art. Particularly, we achieve the AJI score of 0.735 on the Multi-Organ Nuclei Segmentation benchmark, which outperforms the previous closest approaches. In the future work, we will explore the use of different aggregation functions to improve the segmenting cell nuclei results. We will also use the proposed segmentation model to segment breast masses in other modalities, such as thermography (Abdel-Nasser et al., 2016a; Abdel-Nasser et al., 2016b).

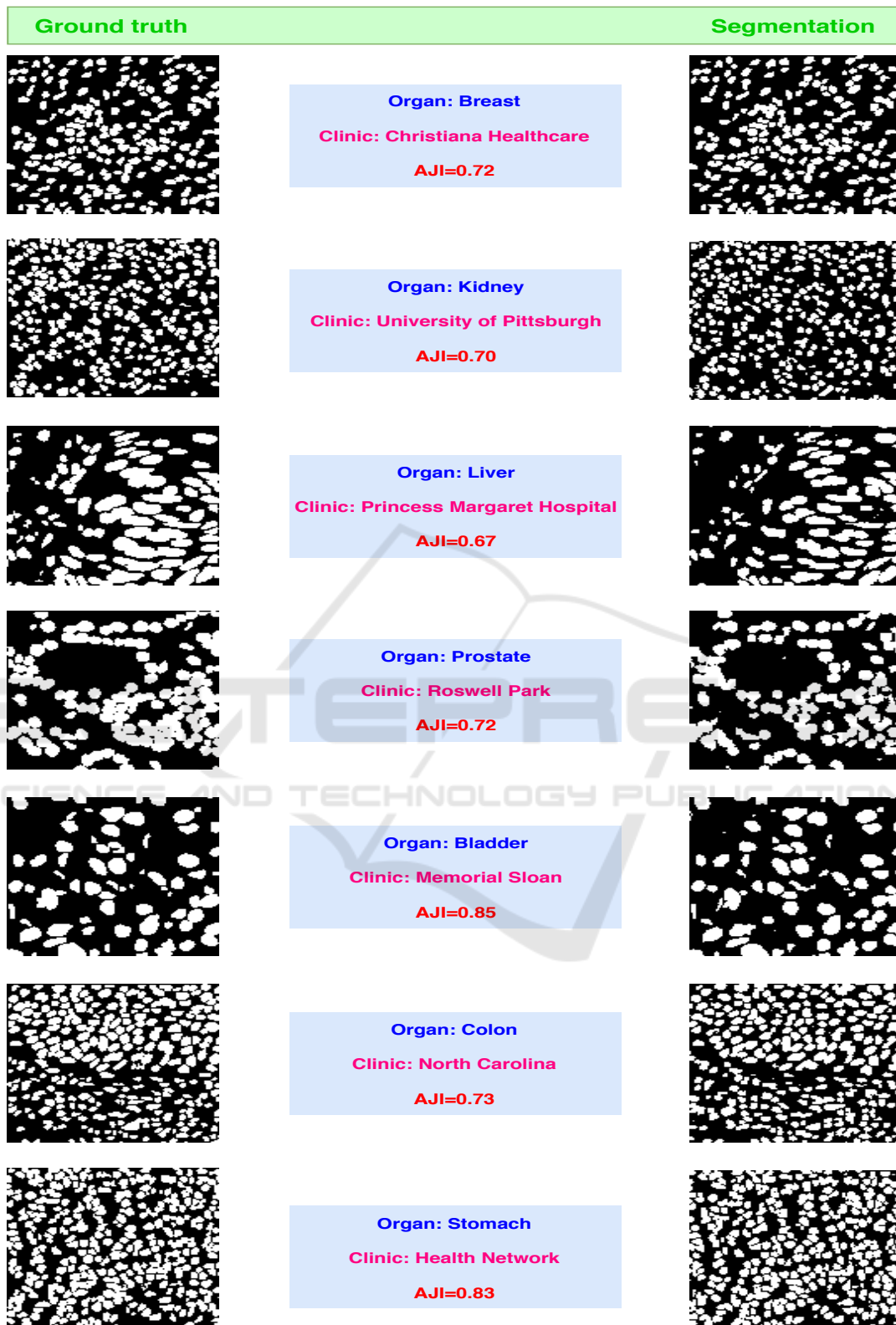


Figure 3: Segmentation results of the proposed model with different organs: breast, kidney, liver, prostate, bladder, colon, and stomach.

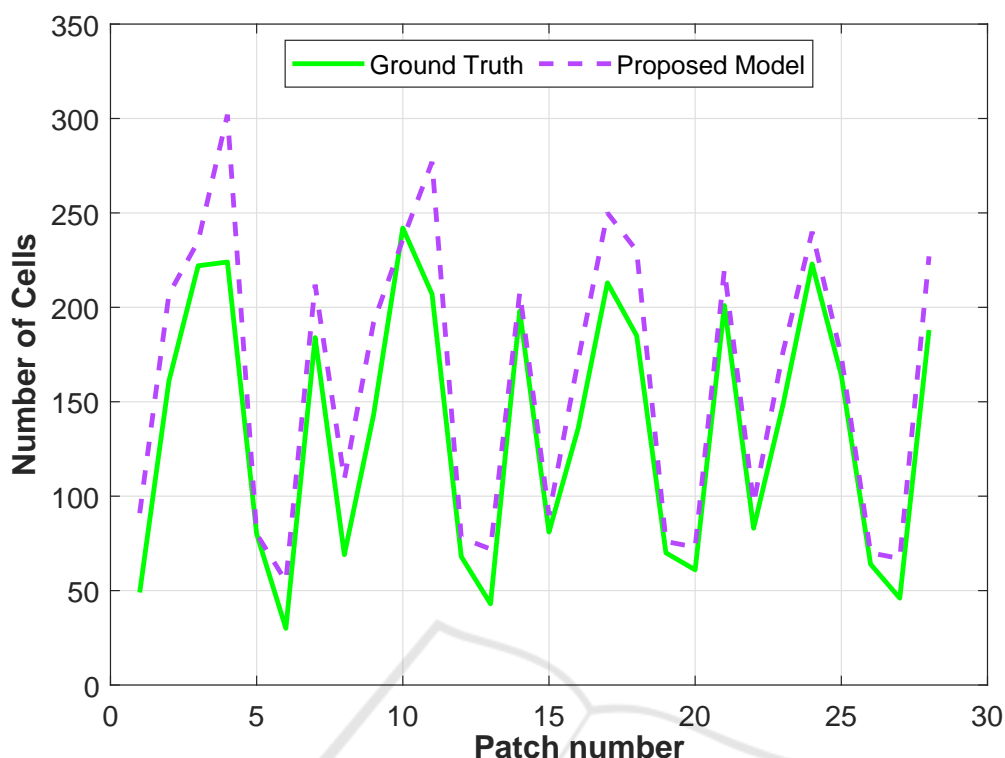


Figure 4: Number of cell nuclei in the predicted masks and the corresponding ground truth.

## ACKNOWLEDGEMENTS

This research was partly supported by the Spanish Government through project DPI2016-77415-R.

## REFERENCES

- Abdel-Nasser, M., Moreno, A., and Puig, D. (2016a). Temporal mammogram image registration using optimized curvilinear coordinates. *Computer methods and programs in biomedicine*, 127:1–14.
- Abdel-Nasser, M., Saleh, A., Moreno, A., and Puig, D. (2016b). Automatic nipple detection in breast thermograms. *Expert Systems with Applications*, 64:365–374.
- Al-Kofahi, Y., Zaltsman, A., Graves, R., Marshall, W., and Rusu, M. (2018). A deep learning-based algorithm for 2-d cell segmentation in microscopy images. *BMC bioinformatics*, 19(1):365.
- Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A., Chang, J. H., Lindquist, R. A., Moffat, J., et al. (2006). Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10):R100.
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., and Sethi, A. (2017). A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging*, 36(7):1550–1560.
- Li, P., Xu, Y., Wei, Y., and Yang, Y. (2019). Self-correction for human parsing. *arXiv preprint arXiv:1910.09777*.
- Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.
- Loshchilov, I. and Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.
- Lucca, G., Sanz, J. A., Dimuro, G. P., Bedregal, B., Asiaín, M. J., Elkano, M., and Bustince, H. (2017). Cc-integrals: Choquet-like copula-based aggregation functions and its application in fuzzy rule-based classification systems. *Knowledge-Based Systems*, 119:32–43.
- Mahmood, F., Borders, D., Chen, R., McKay, G. N., Salmian, K. J., Baras, A., and Durr, N. J. (2019). Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE transactions on medical imaging*.

- Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep learning for cellular image analysis. *Nature methods*, page 1.
- Naylor, P., Laé, M., Reyat, F., and Walter, T. (2017). Nuclei segmentation in histopathology images using deep neural networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 933–936. IEEE.
- Niazi, M. K. K., Parwani, A. V., and Gurcan, M. N. (2019). Digital pathology and artificial intelligence. *The Lancet Oncology*, 20(5):e253–e261.
- Qu, H., Wu, P., Huang, Q., Yi, J., Riedlinger, G. M., De, S., and Metaxas, D. N. (2019). Weakly supervised deep nuclei segmentation using points annotation in histopathology images. In *International Conference on Medical Imaging with Deep Learning*, pages 390–400.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Ruan, T., Liu, T., Huang, Z., Wei, Y., Wei, S., and Zhao, Y. (2019). Devil in the details: Towards accurate single and multiple human parsing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4814–4821.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676.
- Vicar, T., Balvan, J., Jaros, J., Jug, F., Kolar, R., Masarik, M., and Gumulec, J. (2019). Cell segmentation methods for label-free contrast microscopy: review and comprehensive comparison. *BMC bioinformatics*, 20(1):360.
- Wang, E. K., Zhang, X., Pan, L., Cheng, C., Dimitrakopoulou-Strauss, A., Li, Y., and Zhe, N. (2019). Multi-path dilated residual network for nuclei segmentation and detection. *Cells*, 8(5):499.
- Zhou, Y., Onder, O. F., Dou, Q., Tsougenis, E., Chen, H., and Heng, P.-A. (2019). Cia-net: Robust nuclei instance segmentation with contour-aware information aggregation. In *International Conference on Information Processing in Medical Imaging*, pages 682–693. Springer.