Proceedings of the Federated Conference on Computer Science and Information Systems pp. 569–576
DOI: 10.15439/2020F175
ISSN 2300-5963 ACSIS, Vol. 21

Instance Segmentation Model Created from Three Semantic Segmentations of Mask, Boundary and Centroid Pixels Verified on GlaS Dataset

Peter Malík, Štefan Krištofík
Institute of Informatics
Slovak Academy of Sciences
Dúbravská cesta 9, 845 07 Bratislava, Slovakia
Email: {p.malik, stefan.kristofik}@savba.sk

Kristína Knapová
Faculty of Informatics and Information Technologies
Slovak University of Technology
Ilkovičova 2, 842 16 Bratislava, Slovakia
Email: knapova.kristina@gmail.com

***Abstract*—Segmentation is the key computer vision task in modern medicine applications. Instance segmentation became the prevalent way to improve segmentation performance in recent years. This work proposes a novel way to design an instance segmentation model that combines 3 semantic segmentation models dedicated for foreground, boundary and centroid predictions. It contains no detector so it is orthogonal to a standard instance segmentation design and can be used to improve the performance of a standard design. The presented custom designed model is verified on the Gland Segmentation in Colon Histology Images dataset.**

I. Introduction

SEMANTIC segmentation is the most important computer vision task in biomedical applications and any improvement of it may result in saved lives [1], [2]. Combining multiple models is a well known technique to improve segmentation. Creating an ensemble of the trained models can significantly increase the single model performance [3], [4], [5], [6], [7]. It is a favorite method of many models which helped them to be placed high in competition leader-boards. The high structural diversity within an ensemble is very beneficial; therefore, varied models are usually used within an ensemble [4], [6]. Another standard method is to use a multiple loss function with at least one element entirely focused on the boundary pixels [8], [9], [10]. The boundary pixels are harder to correctly classify and using a part of the loss function focused on them can significantly improve the overall results. More advanced method is to use a separate model (or at least a separate architectural branch) to learn boundary pixels and its results combined with the standard segmentation model [11], [12].

Instance segmentation is a more complex CV task capable to differentiate classes and objects within classes on the pixel level. The advantage of instance segmentation is a capability to count objects (even objects in contact or partial occlusion) which is very beneficial in many applications [13]. The standard approach to create an instance segmentation model is to combine a semantic segmentation model with a detection model [11], [12]. Joint training of the models improves the overall results. It also improves the single model performance [14] in detection or segmentation tasks. Newer instance segmentation models combine multiple models. One model is usually a detector, one is semantic segmentation and one is dedicated to boundary pixels [15], [16].

Our proposed method is inspired by all the mentioned techniques. We combined three semantic segmentation models into a model capable to perform the instance segmentation task. One model is semantic segmentation of foreground, one is dedicated for boundary pixels and the last model is focused on the most internal pixels (near object centroids) of all segmented objects. Our method is orthogonal to the standard instance segmentation technique because there is no detector. It is also orthogonal to the ensemble technique because the models are dedicated to the different operational tasks. Our method is tested on the GlaS dataset [17]. It is an instance segmentation dataset that provides annotations with the clear differentiation of each object and the background. The presented results are from our custom designed model based on the U-Net general structure [18] incorporating ResNet [19] blocks with spacial [3], [4] and depth-wise [20] separable convolutions. The novelty and contribution of our work is:

- new technique for designing instance segmentation models composed of three semantic segmentation models,
- the custom designed instance segmentation model,
- verification of our techniques on the GlaS dataset,
- verification that our technique improves semantic segmentation in general.

Our motivation lies in finding a novel way to create instance segmentation models that is orthogonal to currently used techniques so it can be used in combination with them to further improve the state-of-the-art models.

The rest of the paper is organized as follows. The GlaS dataset and related biomedical models are discussed in section II, our model is described in section III, training is discussed in section IV, evaluation of the predicted results are presented in section V, and section VI concludes the paper.

This work has been supported by Slovak national project VEGA 2/0155/19

II. The GlaS dataset and related biomedical models

Colorectal adenocarcinoma originating in intestinal glandular structures is the most common form of colon cancer. Patient prognosis and a treatment plan is devised by pathologists based on the morphology of intestinal glands, including architectural appearance and glandular formation. Achieving good inter-observer as well as intra-observer reproducibility of cancer grading is still a major challenge. The Gland Segmentation in Colon Histology Images Challenge Contest (GlaS) held at MICCAI'2015 has been organized with the goal to find and improve an automated approach which quantifies the morphology of glands [17]. The GlaS dataset was made public as part of this challenge. It consists of 165 images derived from 16 H&E stained histological sections (each from different patient) of stage T3 (tumour has grown into the outer lining of the bowel wall) or T4 (tumour has grown through the outer lining of the bowel wall) colorectal adenocarcinoma. The images are divided into 3 parts: training set, test A, test B containing 85, 60 and 20 images respectively.

Modern biomedical models utilize or are based on some U-Net [18] like architecture. The work [21] uses a structure learning approach to segment instances of glandular structures from colon histopathology images. The authors combined hand-crafted, multi-scale image features with features computed by a U-Net like model trained to map images to segmentation maps. The results are improved with post-processing and they reached better GlaS challenge rank (combined metric) than the challenge winner. The work [15] improves the results further. Authors created their model as a combination of 4 models. The first one segments foreground, the second one with U-Net like structure segments edges, the third one is a detector and the last one fuses these results into the instance segmentation map.

Instance segmentation is very popular in recent years. A novel hierarchical neural network comprising object detection and segmentation modules to accurate cell instance segmentation of neural cells is presented in [22]. Another work oriented to precise instance nuclei segmentation [23] presents a deep multi-scale neural network, with a novel loss function that is sensitive to the Hematoxylin intensity. The work [16] presents an instance segmentation model that segments translucent overlapping objects. Authors combined segmentation and detection models with multiple branches that allowed output transformation from 2D to 3D. The work [10] presents an instance segmentation improvement of cluttered cells by using a novel multiclass weighted loss function. The work [5] uses an ensemble of mask R-CNN models to segment polyps in colonoscopy images.

III. Instance segmentation model design

We were working with the very limited computation power and had to make some compromises. The GlaS dataset contains images in resolutions 574 $\times$ 433, 589 $\times$ 453 but most images are 775 $\times$ 522. Using high resolution inputs is highly computation intensive. Therefore, we transform all these images to 256 $\times$ 256. It is a well known fact that training with higher resolution improves prediction results in general. So, we do not expect to reach the state-of-the-art results with the reduced resolution of inputs. We also made some choices to select architectures with more efficient computation during designing of our model. More details will be mentioned later.

TABLE I
Experiments with different types of input data augmentations

Augmentation types	Loss function	F1	IoU
None	0.5	0.61	0.61
Rotation	0.38	0.84	0.55
Rotation & crop & shear	0.59	0.73	0.36
All 7 types	0.86	0.71	0.26

Medical datasets rarely contain many images. It is also true for the GlaS dataset which contains 165 images. It is a well known fact that using bigger training sets improves prediction results, allows to use higher capacity models and reduces overfitting occurrence in general. We opted for data augmentation which is a standard practice with small datasets. We designed a custom augmentation scheme that uses random combination of rotation, crop, salt & pepper noise, blurring by mean filter, shear deformation in x axis and/or y axis, horizontal and/or vertical flip, and color channel shift. The rotation is in 60 degree steps, the crop size is within 20–80%. We experimented with different combinations. Some results are in table I. All experiments with different augmentation improved the results but the effect varies. We found out that combining many augmentation types in a step is detrimental. For further experiments, we reduced the probability of multiple augmentations in a step and reduced the augmentation types to rotation, crop and salt & pepper noise.

Our design is based on the U-Net plus model. The input resolution is 256 $\times$ 256. The encoder part is composed of 5 blocks. Each block is composed of two 3$\times$3 convolutions and 2$\times$2 max pooling so the input resolution of the next block is halved. Each convolution is followed by the normalization and ReLU. The convolutions in the first block have 32 channels and the number of the channels is doubled with the reduced resolution. The decoder block is a mirrored image of the encoder block. It starts with transposed convolution to two-times increase the resolution and is followed by concatenation that adds outputs of the encoder block with the same resolution. The center block has resolution 8 $\times$ 8, 2 convolutions with 1024 channels and no pooling. It is considered as a part of the encoder. The last encoder block is followed by a 1$\times$1 convolution with the sigmoid. The performance of this model is shown in the first line of table II.

We made experiments with different types of convolutions used instead of standard 2D convolutions and the results are shown in table II. We used space separable convolutions, depth-wise separable convolutions and space and depth-wise separable convolutions. Each convolution was transformed into a sequence of the selected type of separation. There are no normalization and no activation functions between

TABLE II
EXPERIMENTAL RESULTS WITH DIFFERENT CONVOLUTION TYPES

Convolution types	Loss function	F1	IoU	Time	Parameters
Standard	0.58	0.8	0.47	9.04s	31 126 563
Space separable	1.7	0.49	0.02	2.7s	26 923 811
Depth-wise separable	0.5	0.86	0.66	9.3s	14 386 881
Space and depth--wise separable	0.51	0.85	0.6	3.3s	14 386 815

TABLE III
EXPERIMENTAL RESULTS OF INCREASING THE ENCODER CAPACITY

Number of Convolution sequences in the encoder block	Loss function	F1	IoU	Time	Parameters
2	0.5	0.85	0.6	3.3s	14 386 881
4	0.57	0.83	0.58	10.13s	17 200 623
6	0.98	0.8	0.44	10.05s	20 032 431

separated convolutions. Space separable convolution uses a sequence of 3×1 and 1×3 convolutions. Depth-wise separable convolution uses a sequence of 3×3 depth-wise (applied to each channel separately) and 1×1 convolutions. The space and depth-wise separable convolution uses a sequence of 3×1 depth-wise, 1×3 depth-wise and 1×1 convolutions. The best results were reached by depth-wise separable convolutions, but they consume the most computation time to train an epoch. Therefore, we decided to use space and depth-wise separable convolution instead which has only slightly worse results but is significantly faster.

We also experimented with increasing the model capacity by doubling and tripling the number of convolutions used in the encoder block. To reduce the computational requirement, we focused only on the encoder. The results are in table III. The table shows that increasing the encoder worsened the results.

Our instance segmentation model is composed of a single encoder and 3 decoders dedicated to segment foreground, boundaries and centroid pixels of all objects (glands). Its block architecture is shown in Fig. 1. Segmentation of the boundaries helps to improve the overall segmentation and allows to separate the glands that are in a contact. Segmentation of the centroids allows to filter out the noise and to focus on the true glands.

IV. TRAINING

Our earlier experiments were done with training from the scratch. We used default random seeding offered by Tenforflow and Keras libraries. It is well known that pretraing improves the overall results. Due to our limited computation power, we selected small biomedical datasets for pretrainig. At the beginning, we used one dataset (95 images) from Nuclei Segmentation In Microscope Cell Images dataset composition [24]. Later, we used a combination of Colorectal Adenocarcinoma Gland (CRAG) dataset (173 images) [25] and PATH-DT-MSU dataset (120 images) [26], [27] which both contain images with the cervical glands. Pretraining slightly improved the results by approximately 1 % and using slightly bigger and topically close datasets improved the results slightly further.

We used the weighted binary cross-entropy loss function for the most of our experiments because it produced the best results. We experimented with our custom designed loss function that allowed more precise weight control and focus on the boundary and centroid pixels, but it was always outperformed by weighted binary cross-entropy.

Segmentation of the boundary and centroid pixels required to create extra annotations from the ground truth masks. The annotation boundaries were separated by canny algorithm and the centroids were calculated as the center positions of tight bounding boxes. To improve the imbalance of foreground and background pixels, we increased the width of boundaries and centroids by a dilatation filter. We varied the size of the dilatation filter. The experiments showed that the best prediction results were produced when the width of the annotation boundaries was approximately 11 pixels and the width of the annotation centroids was approximately 14 pixels.

As the main metric was selected F1 score and as the second evaluation score was used Intersection over Union (IoU). F1 score correlated more with visual quality inspection of segmentation results in comparison to IoU. F1 and IoU were also used for the evaluation of boundary and centroid segmentation results. However, they were calculated from their respective annotations.

Our instance segmentation model produces 3 separate output maps that have to be combined into the final instance segmentation map. It is done by simple postprocessing. The first step uses threshold values to transform predicted values to binary numbers. The second step tightens the boundary and centroid prediction by erosion filters. To improve prediction, the wider annotations were used. The erosion transforms the prediction into tight boundaries and centroids. The third step slightly denoises the foregrounds masks by using the dilatation and erosion filter in a sequence. The fourth step subtracts the boundaries from foreground masks to find the true separation between glands in contact. The fifth step removes the objects that do not have segmented centroids. This step significantly removes the noise. The thresholds of the first step are set to lower values (approximately 0.33) so most of the true foreground pixels are segmented. It can be done this way because the fifth step removes most of false objects still present in the mask.

Hyperparameters overview. The combination of grid and line searches was used to find the optimal parameters. We selected more computation efficient solutions due to the limited computational power. Selected experiments were discussed in section III. We used Adam optimizer and the default setting achieved sufficiently good results in our experiments. The default setting is represented by beta_1 = 0.9 (the exponential decay rate for the 1st momentum estimate), beta_2 = 0.999

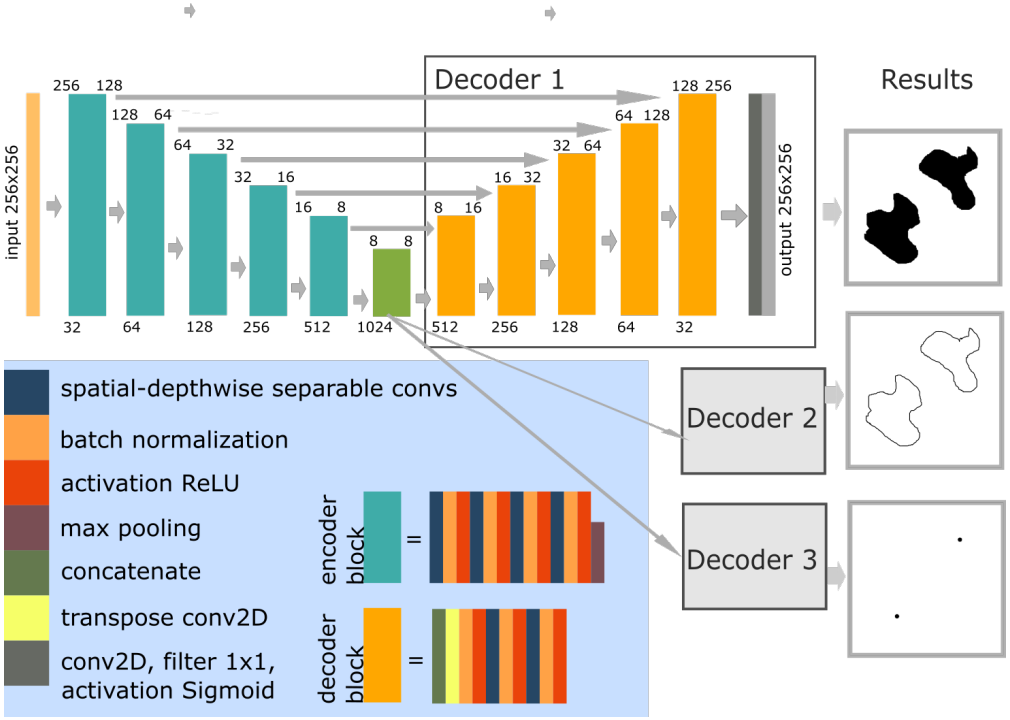

Fig. 1. Block architecture of our final instance segmentation model

(the exponential decay rate for the 2nd momentum estimate), epsilon = 1e-7 (a small constant for numerical stability). Most of our training is done with default learning rate of 0.001. We used batch size = 8 and max epochs = 150 but usually training was stopped sooner. We used early stopping with patience = 20. Input resolution = output resolution = 256 $\times$ 256, the main evaluation metric is F1 score, the additional metric is IoU and the loss function is weighted binary cross-entropy.

V. Evaluation of the Predicted Results

The GlaS dataset was used in Colon Histology Images Challenge Contest held at MICCAI'2015 and therefore there are a lot of great performing models listed in the challenge leader-board, see table IV. Our model reaches the 7th best place in F1 score while using only 256 $\times$ 256 input resolution. With reduced input resolution we did not expect to improve the state-of-the-art. Our results prove that precise instance segmentation can be done with only segmentation models, no detector is necessary. Our presented instance segmentation designed technique can be used also with a detector to improve instance segmentation further and push the state-of-the-art.

Our instance segmentation design technique can be also used to improve standard semantic segmentation. As was described in the previous section, the boundaries improve the object separation and the centroids improve the noise reduction (in the form of reduction of false predictions). The comparison of our best instance segmentation model with its only foreground (mask) segmentation branch is shown in table V. Adding boundary and centroid segmentation branches can significantly improve the performance of standard semantic segmentation.

TABLE IV
Comparison the state-of-the-art models

Model name	F1
CUMedVision2	0.912
ExB3	0.896
Work [15]	0.893
ExB2	0.892
Work [21]	0.892
ExB1	0.891
Our model	**0.874**
Freiburg2	0.870
CUMedVision1	0.868
CVIP Dundee	0.863
Xu et al.	0.858
Freiburg1	0.834
LIB	0.777
CVML	0.652
vision4GlaS	0.635

Visual evaluation of predicted results of our best instance segmentation model and its only foreground segmenting branch can be seen in Fig. 2, Fig. 3, Fig. 4 and Fig. 5. Fig. 2 and Fig. 3 show easy samples represented by regular structure and good contrast. Fig. 4 and Fig. 5 show hard samples represented by more complex structure (irregularities, high deformations) and less contrasted texture details. Instance

TABLE V
EVALUATION OF THE IMPACT OF BOUNDARY AND CENTROID SEGMENTATION AS AN ADDITION TO THE FOREGROUND PREDICTION

Model	F1	IoU
Only foreground segmentation branch of our best IS model	0.737	0.601
Our best IS model	0.874	0.784

segmentation model clearly improves the separation between glands and helps to remove false positives.

VI. CONCLUSION

This work presents a novel way to design an instance segmentation model that is composed of 3 semantic segmentation models. Because it does not include a detector, it is orthogonal to standard instance segmentation design methods and can be used together with them to further improve the state-of-the-art. The presented results clearly show that adding 2 segmentation branches with foreground segmentation improves the segmentation results significantly. The boundary and centroid segmentation branches improve the separation between objects and remove false positives.

Our best performing instance segmentation model reached the 7th best result in F1 score when compare to recent works and the MICCAI'15 contest leader-board while using only 256 $\times$ 256 resolution. The model is custom designed with space and depth-wise separable convolutions and basic U-Net like structure. The segmentation models share single encoder while they use their own separate decoders.

Our model can be improved by better postprocessing in the form of fusing neural network model which we are planing to add in the future.

REFERENCES

[1] T. Adams, J. Dörpinghaus, M. Jacobs, and V. Steinhage, "Automated lung tumor detection and diagnosis in ct scans using texture feature analysis and svm," in *Communication Papers of the 2018 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 17. PTI, 2018, pp. 13–20. [Online]. Available: http://dx.doi.org/10.15439/2018F176

[2] M. Li, Q. Yin, and M. Lu, "Retinal blood vessel segmentation based on multi-scale deep learning," in *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 15. IEEE, 2018, pp. 117–123. [Online]. Available: http://dx.doi.org/10.15439/2018F127

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1409.1556

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7–12, 2015*. IEEE Computer Society, 2015, pp. 1–9. [Online]. Available: https://doi.org/10.1109/CVPR.2015.7298594

[5] J. Kang and J. Gwak, "Ensemble of instance segmentation models for polyp segmentation in colonoscopy images," *IEEE Access*, vol. 7, pp. 26 440–26 447, 2019. [Online]. Available: http://dx.doi.org/10.1109/ACCESS.2019.2900672

[6] A. O. Vuola, S. U. Akram, and J. Kannala, "Mask-rcnn and u-net ensembled for nuclei segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, April 2019, pp. 208–212. [Online]. Available: http://dx.doi.org/10.1109/ISBI.2019.8759574

[7] L. Podlodowski, S. Roziewski, and M. Nurzyński, "An ensemble of deep convolutional neural networks for marking hair follicles on microscopic images," in *Position Papers of the 2018 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 16. PTI, 2018, pp. 23–28. [Online]. Available: http://dx.doi.org/10.15439/2018F389

[8] B. D. Brabandere, D. Neven, and L. V. Gool, "Semantic instance segmentation with a discriminative loss function," *CoRR*, vol. abs/1708.02551, 2017. [Online]. Available: http://arxiv.org/abs/1708.02551

[9] J. Dai, K. He, Y. Li, S. Ren, and J. Sun, "Instance-sensitive fully convolutional networks," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 534–549. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-46466-4_32

[10] F. A. Guerrero-Peña, P. D. Marrero Fernandez, T. Ing Ren, M. Yui, E. Rothenberg, and A. Cunha, "Multiclass weighted loss for instance segmentation of cluttered cells," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens, October 2018, pp. 2451–2455. [Online]. Available: http://dx.doi.org/10.1109/ICIP.2018.8451187

[11] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, April 2018. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2017.2699184

[12] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, June 2018, pp. 8759–8768. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2018.00913

[13] J. Respondek and W. Westwańska, "Counting instances of objects specified by vague locations using neural networks on example of honey bees," in *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 18. IEEE, 2019, pp. 87–90. [Online]. Available: http://dx.doi.org/10.15439/2019F94

[14] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 2980–2988. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2017.322

[15] Y. Xu, Y. Li, Y. Wang, M. Liu, Y. Fan, M. Lai, and E. I. Chang, "Gland instance segmentation using deep multichannel neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 12, pp. 2901–2912, December 2017. [Online]. Available: http://dx.doi.org/10.1109/TBME.2017.2686418

[16] A. Böhm, A. Ücker, T. Jäger, O. Ronneberger, and T. Falk, "Isoodl: Instance segmentation of overlapping biological objects using deep learning," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC, April 2018, pp. 1225–1229. [Online]. Available: http://dx.doi.org/10.1109/ISBI.2018.8363792

[17] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Böhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. Snead, and N. M. Rajpoot, "Gland segmentation in colon histology images: The glas challenge contest," *Medical Image Analysis*, vol. 35, pp. 489 – 502, January 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361841516301542

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-24574-4_28

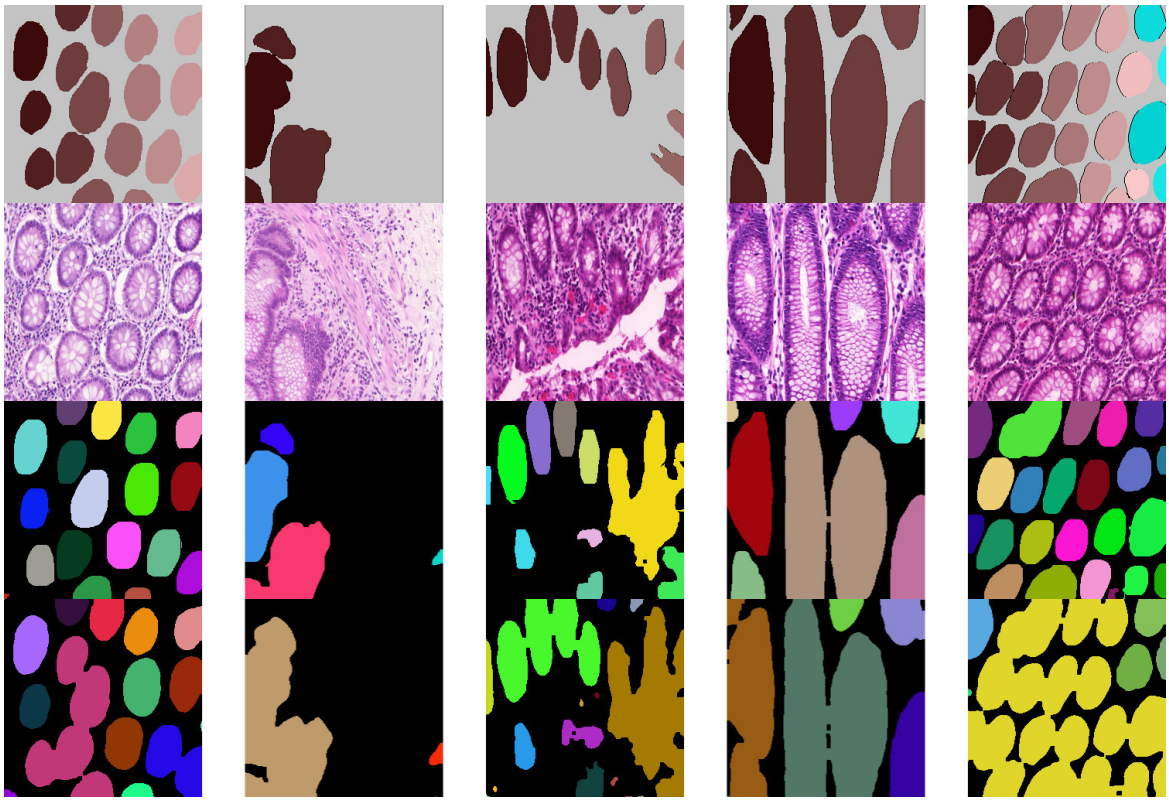

Fig. 2. Visual evaluation of easy samples. From top to bottom: annotation, input, IS prediction, only foreground prediction branch.

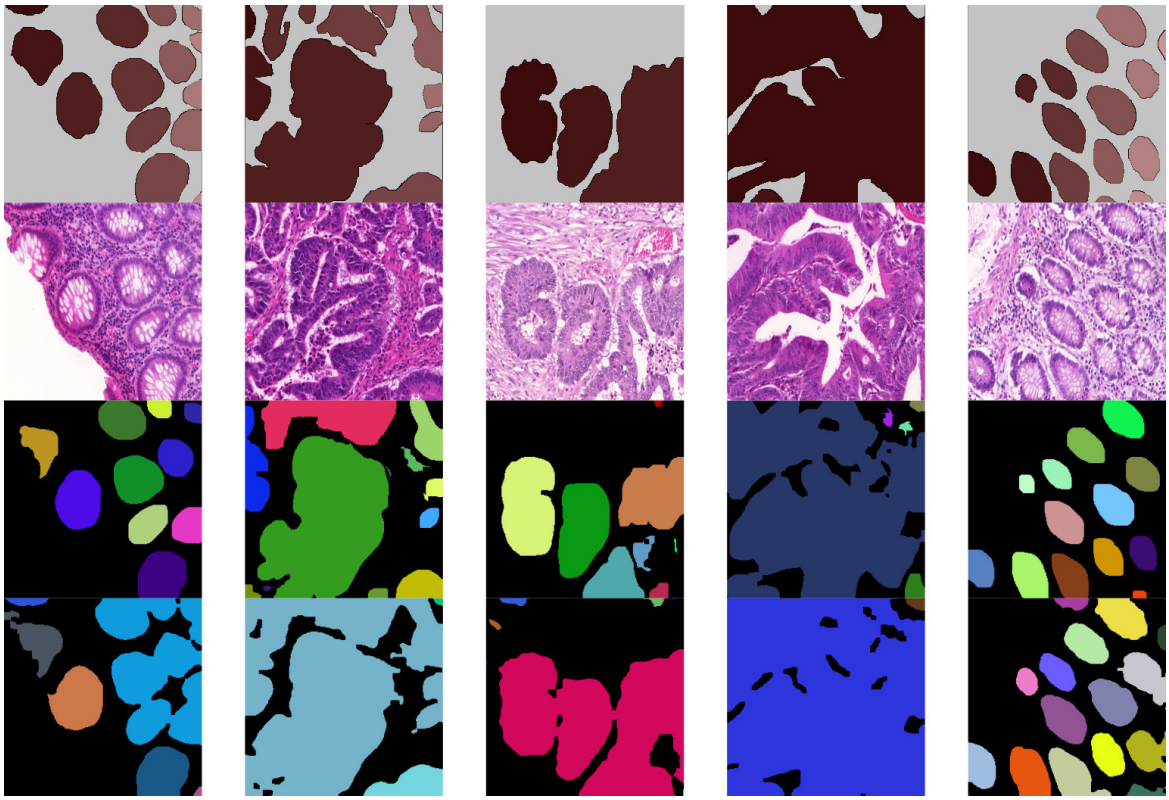

Fig. 3. Visual evaluation of easy samples. From top to bottom: annotation, input, IS prediction, only foreground prediction branch.

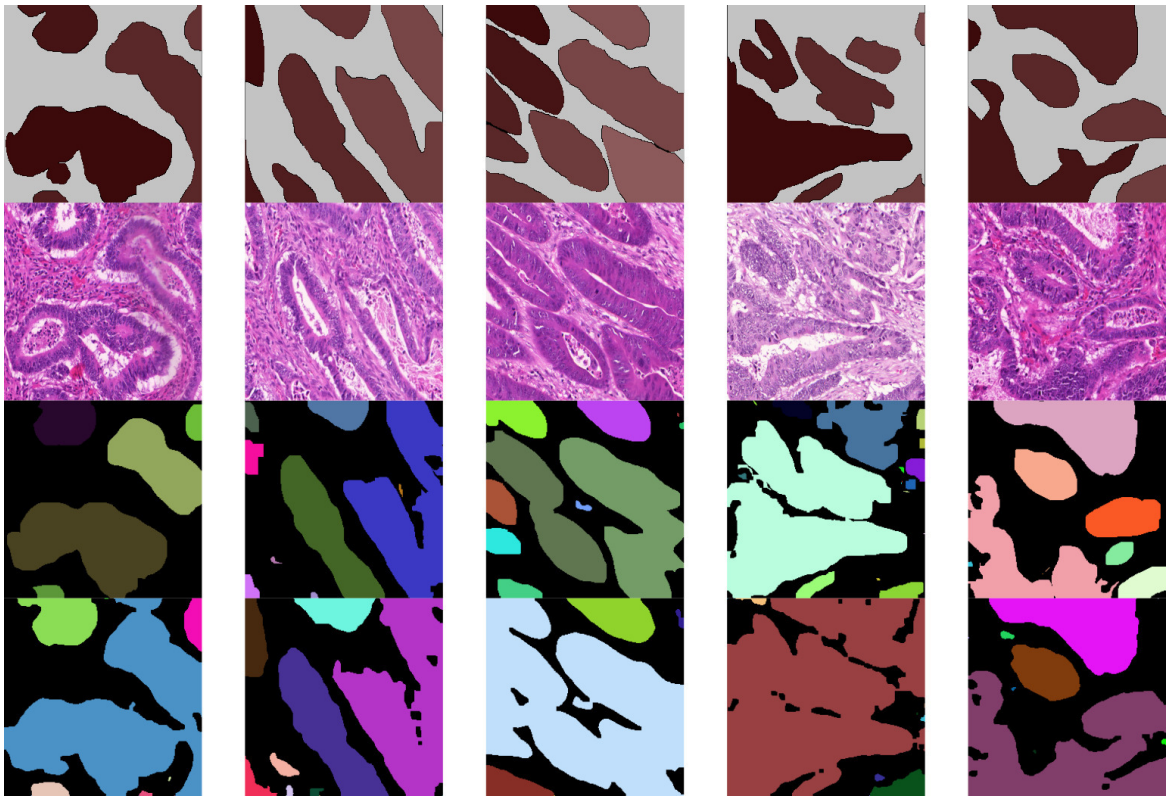

Fig. 4. Visual evaluation of hard samples. From top to bottom: annotation, input, IS prediction, only foreground prediction branch.

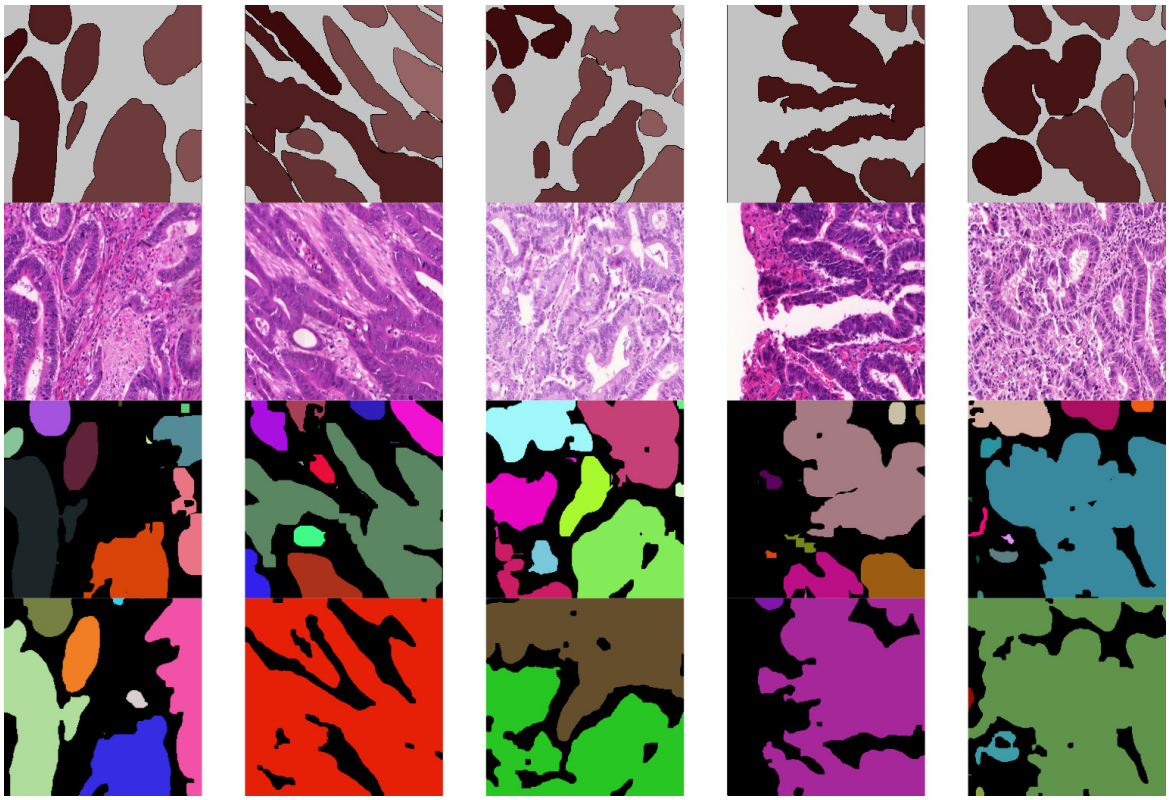

Fig. 5. Visual evaluation of hard samples. From top to bottom: annotation, input, IS prediction, only foreground prediction branch.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, June 2016, pp. 770–778. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2016.90

[20] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, July 2017, pp. 1800–1807. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2017.195

[21] S. Manivannan, W. Li, J. Zhang, E. Trucco, and S. J. McKenna, "Structure prediction for gland segmentation with hand-crafted and deep convolutional features," *IEEE Transactions on Medical Imaging*, vol. 37, no. 1, pp. 210–221, January 2018. [Online]. Available: http://dx.doi.org/10.1109/TMI.2017.2750210

[22] J. Yi, P. Wu, D. J. Hoeppner, and D. Metaxas, "Pixel-wise neural cell instance segmentation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC, April 2018, pp. 373–377. [Online]. Available: http://dx.doi.org/10.1109/ISBI.2018.8363596

[23] S. Graham and N. M. Rajpoot, "Sams-net: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC, April 2018, pp. 590–594. [Online]. Available: http://dx.doi.org/10.1109/ISBI.2018.8363645

[24] G. Payyavula, "Nuclei segmentation in microscope cell images," 2018. [Online]. Available: https://www.kaggle.com/gangadhar/nuclei-segmentation-in-microscope-cell-images/

[25] S. Graham, H. Chen, Q. Dou, P.-A. Heng, and N. M. Rajpoot, "Mild-net: Minimal information loss dilated network for gland instance segmentation in colon histology images," *Medical Image Analysis*, vol. 52, pp. 199–211, 2019. [Online]. Available: http://dx.doi.org/10.1016/j.media.2018.12.001

[26] A. Khvostikov, A. Krylov, I. Mikhailov, O. Kharlova, N. Oleynikova, and P. Malkov, "Automatic mucous glands segmentation in histological images," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W12, pp. 103–109, 2019. [Online]. Available: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W12/103/2019/

[27] N. Oleynikova, A. Khvostikov, A. Krylov, I. Mikhailov, O. Kharlova, N. Danilova, P. G. Mal'kov, N. Ageykina, and E. Fedorov, "Automatic glands segmentation in histological images obtained by endoscopic biopsy from various parts of the colon," *Endoscopy*, vol. 51, 04 2019. [Online]. Available: http://dx.doi.org/10.1055/s-0039-1681188