High-rate emphasized DeepLabV3Plus for Semantic Segmentation of Breast Cancer-related Hematoxylin and Eosin-stained Images

Sanghoon Lee
Department of Computer Science
Kennesaw State University
Marrietta, GA
slee297@kennesaw.edu

Yanjun Zhao
Department of Computer Science
Troy University
Troy, AL
yjzhao@troy.edu

Wookjin Choi
Sidney Kimmel College & Cancer Center
Thomas Jefferson University
Philadelphia, PA
wookjin.choi@jefferson.edu

***Abstract*—Many deep learning algorithms have been successfully adopted to extract meaningful information from histopathology images, but they have been untapped in semantic image segmentation. In this paper, we propose a deep convolutional neural network model that strengthens Atrous separable convolutions with a high rate within spatial pyramid pooling for histopathology image segmentation. We adopted DeepLabV3Plus for the encoder and decoder process. ResNet50 was used as the encoder block of the model for taking advantage of attenuating the problem of the increased depth of the network by using skip connections. Three Atrous separable convolutions with higher rates were added to the existing Atrous separable convolutions. We conducted a performance evaluation on three tissue types: tumor, tumor-infiltrating lymphocytes, and stroma for comparing the proposed model with the eight state-of-the-art deep learning models: DeepLabV3, DeepLabV3Plus, LinkNet, MANet, PAN, PSPnet, UNet, and UNet++. The performance results show that the proposed model outperforms the eight models on *mIOU* (0.8058/0.7792) and *FSCR* (0.8525/0.8328) for both tumor and tumor-infiltrating lymphocytes.**

Keywords— Deep learning, Image Segmentation, Tumor, Histopathology, Hematoxylin and eosin-stained images.

I. Introduction

Advances in artificial intelligence (AI) have provided evidence regarding the efficiency and effectiveness of data-driven predictive modeling in many fields of research [1]. Recently, deep learning models as part of AI have received much attention in biomedical image analysis tasks because of their highly accurate performance and versatilities for a comprehensive analysis through multimodal integration [2]. These deep learning models enable the exploration of complex patterns providing clinical decision support and contributing to additional insights based on the biomedical image data.

Many deep-learning models have been proposed for image classification, object detection, and image segmentation. LeNet is the earliest deep convolutional neural network (CNN) with a 7-layered neural network [3] and AlexNet is a deeper version of CNNs with an 8-layered neural network demonstrating the effectiveness of the network on two GPUs on Large Scale Visual Recognition Challenge (LSVRC) [4]. VGG network series as a very deep convolutional network for large-scale image recognition has been highlighted as a leading deep CNN [5].

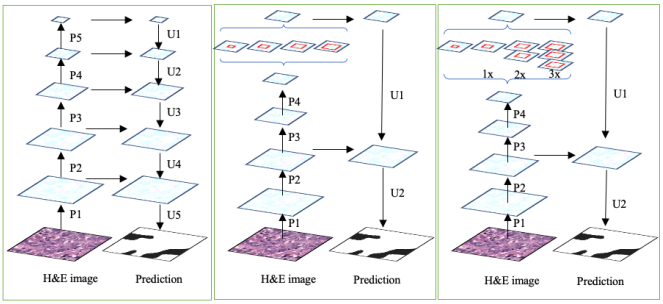

Fig. 1. The proposed rate emphasized DeepLabV3Plus. Traditional encoder-decoder model reserves the spatial information through skip connection for each output of the encoder (Left) [8]. DeepLavV3Plus reserves the spatial information through Atrous convolution reducing skip connections (Middle) [8]. Rate emphasized DeepLabV3Plus emphasizes the spatial information through Atrous convolution reducing skip connections (right).

However, there has been a potential issue associated with increasing the depth of neural networks. ResNet series has been considered as a solution for the problem of increased depth of the network by using skip connections [6]. While these deep CNNs mainly focus on classifying images into a set of categories, some models have been segmenting images into categories. Image segmentation is crucial in biomedical image analysis because segmentation helps in accurately identifying regions of interest within medical images and the segmented regions can be used for understanding the size, shape, and volume of different structures, facilitating the analysis of the target region for treatment. DeepLabV3 is one of the advanced versions of deep CNNs and is mainly used to perform semantic segmentation of images [7]. DeepLabV3 uses dilated convolution called Atrous spatial pyramid pooling to capture multi-scale contextual information in images with reduced computational burden. DeepLabV3Plus is the extended version of DeepLabV3 [8], providing an encoder-decoder architecture that captures more detailed contextual information.

In this paper, we propose a rate-emphasized DeepLabV3Plus that emphasizes depth-wise convolutions with higher rates in spatial pyramid pooling, improving the performance of the semantic segmentation in breast cancer-related hematoxylin and eosin images. The proposed method uses ResNet50 as a backbone network of the encoder and utilizes DeepLabV3Plus

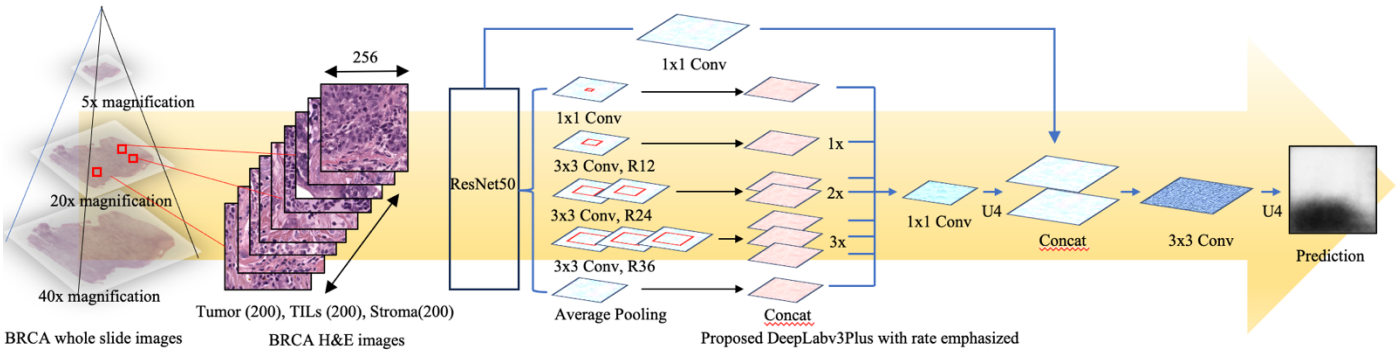

Fig. 2. The overall process of proposed model. This paper obtained 600 hematoxymin and eosin stained images of tumor, tumer-infiterating lymphocytes, and strom extracted from whole slide images for image segmentation. ResNet50 was used as a encoder of the model. Atrous separable convolution was performed on the top of the ReaNet50, emphasizing the features with the larger field of view filters with rate 12 (R12), 24 (R24), and 36 (R36). Binlear upsampling with a factor of 4 (U4) was performed after completing the 1x1 convolution. Another binlear upsampling with a factor of 4 (U4) was performed to be activated for the prediction.

equipped with the proposed rate-emphasized module on the decoder. The detail of the proposed model is shown in Figure 1.

II. Higher Rate-Emphasized DeepLabV3Plus for Semantic Segmentation

The proposed model uses ResNet50 as a backbone network of the encoder, facilitating skip connections and reducing the deep network burden. Since ResNet50 including a 50-layer deep residual neural network has demonstrated its effectiveness in the architecture of biomedical image analysis, we used ResNet50 to capture complex patterns of spatial information within images during the encoding. Atrous separable convolution involving both the depthwise convolution and the pointwise convolution was employed in the last layer of ResNet50.

Rate emphasized Atrous separable convolution. While DeepLabV3Plus uses Atrous separable convolution with different rates at multiple scales, our paper emphasizes spatial information of Atrous separable convolution with large rates where the field-of-view of the filter is large. The intuition of the idea of the proposed method is that the larger Atrous separable convolution with a large rate is, the less its effect will be. We performed six Atrous separable convolutions using 3x3 kernels with 12, 24, 24, 36, 36, and 36 rate values. After completing Atrous separable convolutions, we applied 1x1 convolution to perform bilinear upsampling with a factor of 4, and the unsampled features were concatenated with the corresponding features in the ResNet50 followed by 3x3 convolutions and one more bilinear upsampling with a factor of 4 to be activated for the prediction of the image. We followed the decoder process of DeepLabV3Plus. The details of the overall process of the proposed model are shown in Figure 2.

III. Experiments

Datasets. Our experiment was performed by the Breast Cancer Semantic Segmentation (BCSS) dataset. BCSS dataset is a large breast cancer-related image dataset that consists of 151 hematoxylin and eosin-stained images extracted from whole slide images in the Genomic Data Commons Data Portal [9]. These images include various tissue types such as tumors, stroma, lymphocytic infiltrate, necrosis, glandular secretions, blood, fat, and plasma cells. All the images in the BCSS dataset were annotated by senior residents, junior residents, and medical students of pathology through the annotation review process. In this paper, we randomly selected 600 tumors, stroma, and tumor-infiltrating lymphocytes (TILs) related images with 256x256-sized cropped from the 151 images, creating three datasets: 200 images for tumor, 200 images for stroma, and 200 images for TILs. Since it is critical to reduce the impact of variation in color balance that affects the appearance of images, we performed color normalization by using Reinhard normalization ensuring that similar spatial information is represented consistently. Each dataset was split into training, validation, and test sets with 60%, 20%, and 20% of images respectively.

Baselines and Metrics. The proposed model was compared with eight state-of-the-art deep learning-based segmentation models including UNet [10], UNet++ [11], MANet [12], LinkNet [13], PSPNet [14], PAN [15], DeepLabV3 [7], and DeepLabV3Plus [8]. We used Adam optimizer with a 0.0001 learning rate for training. The evaluation metrics used in this experiment include *Precision (PREC.)*, *Recall (RECL.)*, *Accuracy (ACC.), F score (FSCR), and mean Intersection Over Union* (m*IoU*) to evaluate the performance of the deep learning-based segmentation models. A confusion matrix (true positive: TP, false positive: FP, false negative: FN, and true negative: TN) was created for the evaluation matrix as follows: IoU = TP/(TP+FP+FN), $PREC.$ = TP/(TP+FP), $REC.$= TP/(TP+FN), $ACC.$ = (TP+TN) / (TP+FP+FN+TN), $FSCR$ =$((1+\beta^2) \times TP)/(((1+\beta^2) \times TP)+ (\beta^2 \times FN)+FP)$.

Results. The experiment results on tumor, tumor-infiltrating lymphocytes, and stroma using deep learning models are shown in Table 1. For tumor, Table 1 shows that the proposed model outperforms traditional deep CNNs in terms of *mIoU* (0.8058) *and FSCR.* (0.8525). For TILs, Table 1 shows that the proposed model outperforms traditional deep CNNs in terms of *mIoU* (0.7792) *and FSCR.* (0.8328). For stroma, Table 1 shows that the proposed model outperforms traditional deep CNNs in terms of *RECL.* (0.7686). Our experiment indicates that DeepLabV3Plus outperforms other deep learning models on *ACC.* (0.9153) for both tumor and stroma. *ACC.* typically refers to the ratio of correctly predicted pixels to the total pixels in the image segmentation and it may be sensitive to label noise and mislabeled pixel information. Therefore, it will be more

informative to consider other metrics such as *mIOU, FSCR., RECL.,* and *PREC. UNet++* shows better performance results on both *mIOU* (0.6069) and *FSCR.* (0.6830) for stroma, *RECL.* (0.9441*)* for tumor, *PREC.* (0.9275) for TILs. However, the overall performance results on UNet++ are very vulnerable (i.e., for tumor, *RECL.* is 0.9441 but *PREC.* is 0.5824) suggesting the model is not robust or reliable compared with the proposed model. Examples of the predicted images based on eight deep learning models as well as the proposed model are shown in Figure 3.

IV. Conclusion

In this paper, we presented a high rate emphasized DeepLabV3Plus that strengthens Atrous separable convolutions with a high rate in spatial pyramid pooling for histopathology image segmentation. ResNet50 was used as the encoder block of the model for taking advantage of attenuating the problem of the increased depth of the network by using skip connections. Three Atrous separable convolutions with higher rates were added to the existing Atrous separable convolutions. We followed the process of the decoder block in DeepLabV3Plus. The performance evaluation of the proposed model was conducted based on tumor, TILs, and stroma, comparing with the eight state-of-the-art deep learning models: DeepLabV3, DeepLabV3Plus, LinkNet, MANet, PAN, PSPnet, UNet, and UNet++. The performance results show that the proposed model outperforms the eight models on *mIOU* and *FSCR* for both tumor and TILs, as well as *RECL.* for stroma. We plan to extend the proposed model not only to radiology images but also to other types of cancer types in the future.

Acknowledgment

This research was supported by the National Science Foundation under Grant No. 2138260 and Grant No. 2153063.

References

[1] Liu, P., Wang, L., Ranjan, R., He, G. and Zhao, L., 2022. A survey on active deep learning: from model driven to data driven. *ACM Computing Surveys (CSUR)*, *54*(10s), pp.1-34.

[2] He, J., Baxter, S.L., Xu, J., Xu, J., Zhou, X. and Zhang, K., 2019. The practical implementation of artificial intelligence technologies in medicine. *Nature medicine*, *25*(1), pp.30-36.

[3] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), pp.2278-2324.

[4] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

[5] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[6] 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* (pp. 770-778).

[7] Chen, L.C., Papandreou, G., Schroff, F. and Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

[8] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818).

[9] Amgad, M., Elfandy, H., Hussein, H., Atteya, L.A., Elsebaie, M.A., Abo Elnasr, L.S., Sakr, R.A., Salem, H.S., Ismail, A.F., Saad, A.M. and Ahmed, J., 2019. Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics*, *35*(18), pp.3461-3467.

[10] Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234-241). Springer International Publishing.

[11] Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4* (pp. 3-11). Springer International Publishing.

[12] Fan, T., Wang, G., Li, Y. and Wang, H., 2020. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, *8*, pp.179656-179665.

[13] Chaurasia, A. and Culurciello, E., 2017, December. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE visual communications and image processing (VCIP)* (pp. 1-4). IEEE.

[14] Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J., 2017. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2881-2890).

[15] Li, H., Xiong, P., An, J. and Wang, L., 2018. Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180*.

TABLE I. Results of Tumor Image Classification

Models	Tumor/ Tumor-infiltrating Lymphocytes / Stroma				
	mIoU	***FSCR.***	***ACC.***	***RECL.***	***PREC.***
DeepLabV3 [7]	0.7877/ 0.7736/ 0.5727	0.8324/ 0.8198/ 0.6519	0.9143/ **0.8901**/ 0.8436	0.8712/ 0.8886/ 0.7196	**0.9093**/ 0.8460/ 0.7145
DeeLabV3Plus [8]	0.8005/ 0.7441/ 0.5702	0.8498/ 0.7835/ 0.6450	**0.9153**/ 0.8654/ **0.8596**	0.8943/ 0.8250/ 0.6949	0.8985/ 0.8756/ 0.7299
LinkNet [13]	0.5419/ 0.4605/ 0.5809	0.5905/ 0.5069/ 0.6611	0.8944/ 0.8456/ 0.8356	0.9250/ **0.9362**/ 0.7525	0.5649/ 0.4882/ 0.7164
MANet [12]	0.5642/ 0.4885/ 0.5870	0.6092/ 0.5290/ 0.6669	0.8968/ 0.8617/ 0.8470	0.9325/ 0.8086/ 0.7503	0.5765/ 0.6343/ 0.6550
PAN [15]	0.7474/ 0.7731/ 0.5820	0.7892/ 0.8187/ 0.6539	0.9052/ 0.8817/ 0.8562	0.8746/ 0.8543/ 0.7059	0.8554/ 0.8818/ 0.7456
PSPNet [14]	0.7562/ 0.7305/ 0.5225	0.8021/ 0.7755/ 0.5960	0.8959/ 0.8732/ 0.8162	0.8938/ 0.8740/ 0.6496	0.8154/ 0.8275/ **0.7529**
UNet [10]	0.7622/ 0.6745/ 0.5600	0.8080/ 0.7238/ 0.6453	0.9129/ 0.8543/ 0.8458	0.9072/ 0.8439/ 0.6838	0.8198/ 0.7639/ 0.7159
UNet++ [11]	0.5582/ 0.7747/ **0.6069**	0.6085/ 0.8260/ **0.6830**	0.8879/ 0.8582/ 0.8521	**0.9441**/ 0.8204/ 0.7462	0.5824/ **0.9275**/ 0.6816
Proposed model	**0.8058**/ **0.7792**/ 0.5905	**0.8525**/ **0.8328**/ 0.6585	0.9006/ 0.8781/ 0.8357	0.9255/ 0.8757/ **0.7686**	0.8750/ 0.8682/ 0.7257

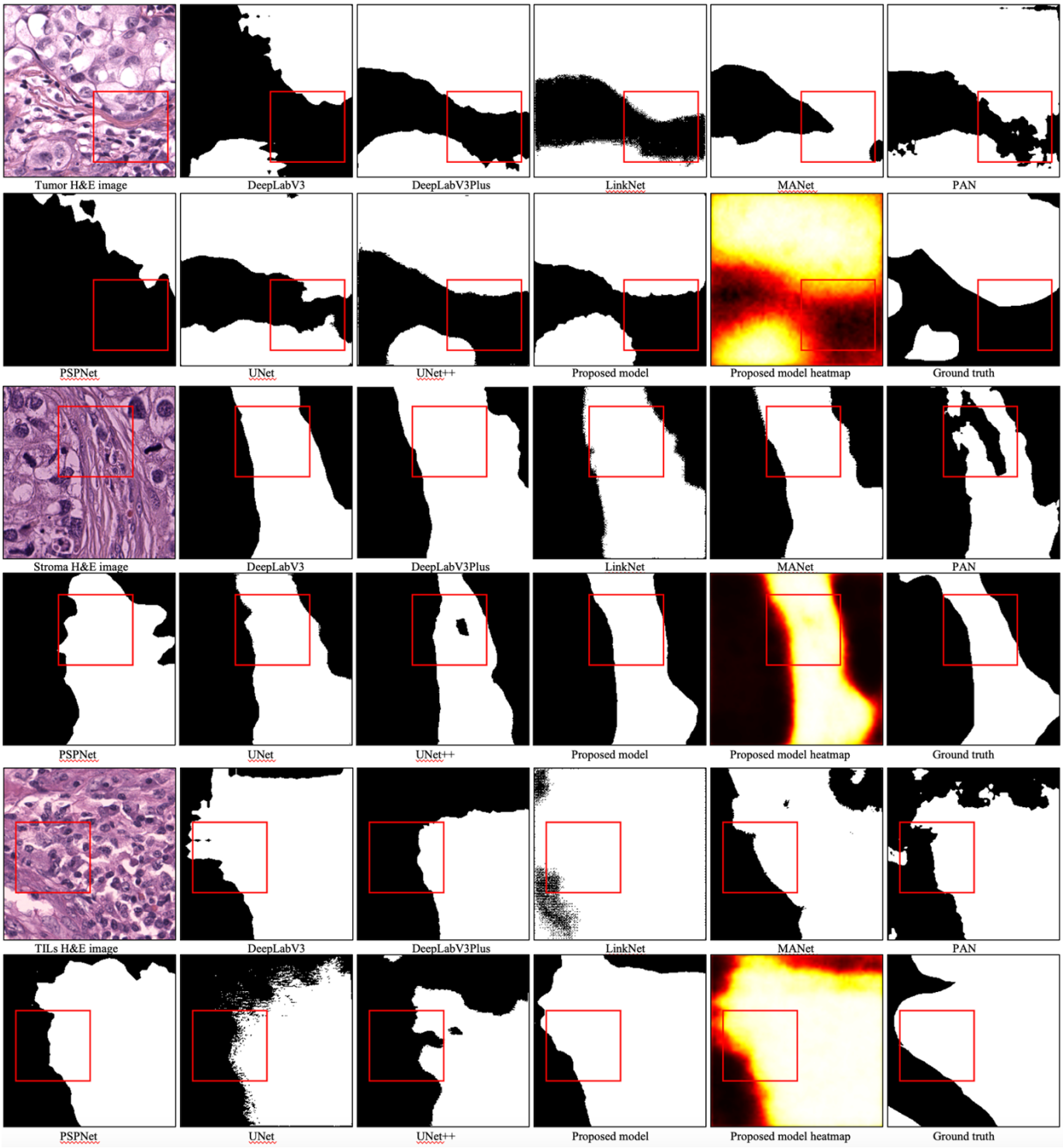

Fig. 3. Examples of the predicted images based on eight deep learning models: DeepLabV3, DeepV3Plus, LinkNet, MANet, PAN, PSPNet, UNet, and UNet++, as well as the proposed model. A color normalized tumor-related hematoxylin and eosin stained image, the predicted images on DeepLabV3, DeepV3Plus, LinkNet, MANet, PAN, PSPNet, UNet, UNet++, and the proposed model, the probability heatmap of the proposed model, and the ground truth image (Left to right on the top). A color normalized stroma-related hematoxylin and eosin stained image, the predicted images on the eight models and the proposed model, and the probability heatmap of the proposed model, and the ground truth image (Left to right on the middle). A color normalized tumor-infiltrating lymphocyte-related hematoxylin and eosin stained image, the predicted images on the eight models and the proposed model, and the probability heatmap of the proposed model, and the ground truth image (Left to right on the bottom)