

# Colorectal Cancer Detection Based on Deep Learning

Lin Xu<sup>1</sup>, Blair Walker<sup>2</sup>, Peir-In Liang<sup>3</sup>, Yi Tong<sup>1</sup>, Cheng Xu<sup>1</sup>, Yu Chun Su<sup>1</sup>, Aly Karsan<sup>4</sup>

<sup>1</sup>GenerationsE Software Solutions, Inc., Surrey, Canada, <sup>2</sup>Department of Pathology, St. Paul's Hospital, University of British Columbia, Vancouver, Canada,

<sup>3</sup>Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan, <sup>4</sup>Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, Canada

Submitted: 20-Dec-2019

Revised: 28-Feb-2020

Accepted: 21-Apr-2020

Published: 21-Aug-2020

## Abstract

**Introduction:** The initial point in the diagnostic workup of solid tumors remains manual, with the assessment of hematoxylin and eosin (H&E)-stained tissue sections by microscopy. This is a labor-intensive step that requires attention to detail. In addition, diagnoses are influenced by an individual pathologist's knowledge and experience and may not always be reproducible between pathologists. **Methods:** We introduce a deep learning-based method in colorectal cancer detection and segmentation from digitized H&E-stained histology slides. **Results:** In this study, we demonstrate that this neural network approach produces median accuracy of 99.9% for normal slides and 94.8% for cancer slides compared to pathologist-based diagnosis on H&E-stained slides digitized from clinical samples. **Conclusion:** Given that our approach has very high accuracy on normal slides, use of neural network algorithms may provide a screening approach to save pathologist time in identifying tumor regions. We suggest that this new method may be a powerful assistant for colorectal cancer diagnostics.

**Keywords:** Colorectal cancer, deep learning, digital pathology, medical imaging

## INTRODUCTION

With the development of targeted therapies, many treatments are based on molecular studies, which require sampling tumor tissue from paraffin blocks for sequencing. An automated solution could potentially reduce the workload of pathologists by acting as a screening device and may reduce the subjectivity in diagnosis.<sup>[1]</sup>

In tissue-based diagnostics, most of the work still needs to be done manually by a pathologist using a microscope to examine hematoxylin and eosin (H&E)-stained slides. To determine pN stage in metastatic disease or tumor content for downstream genomic analysis, quantifying the size and numbers of tumor regions is necessary. The foundation of such tasks is to accurately distinguish cancer/malignant cells from normal/benign cells. However, the determination of tumor content is poorly reproducible with significant interpathologist variation.<sup>[1]</sup> Depending on the specific disease entity, pathologist's experience, and the state of the pathologist, results may vary significantly. In addition, the determination of tumor content and circling of tumor regions on H&E slides to identify areas to sample for downstream genomic analysis is a required preanalytical step by accreditation bodies in order

to enrich for tumor content and ensure accurate determination of genomic variants. As the size of tumor regions can be very small, pathologists are often required to use high magnification for detecting tumor cells. This requirement significantly increases the workload for pathologists.

Deep learning has been successfully applied to many tasks including image processing,<sup>[2]</sup> sound/voice processing, and language translation.<sup>[3]</sup> Recent progress shows that deep learning can also be applied to medical image processing, such as magnetic resonance imaging,<sup>[4]</sup> computed tomography,<sup>[5]</sup> biopsy,<sup>[6]</sup> and endoscopy.<sup>[7]</sup> Recently, digital pathology datasets have become publicly available and have opened up the possibility of evaluating the feasibility of applying deep learning techniques to improving the efficiency and quality of histologic diagnosis. For instance, the goal of CAMELYON17<sup>[8]</sup> is automated detection and classification of

**Address for correspondence:** Dr. Aly Karsan,  
Room 9-111, 675 W 10<sup>th</sup> Ave, Vancouver, BC V5Z 1L3, Canada.  
E-mail: akarsan@bcgsc.ca

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** WKHLRPMedknow\_reprints@wolterskluwer.com

**How to cite this article:** Xu L, Walker B, Liang PI, Tong Y, Xu C, Su YC, *et al.* Colorectal cancer detection based on deep learning. J Pathol Inform 2020;11:28.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2020/11/1/28/292721>

### Access this article online

#### Quick Response Code:



**Website:**  
[www.jpathinformatics.org](http://www.jpathinformatics.org)

**DOI:**  
10.4103/jpi.jpi\_68\_19

breast cancer metastases in whole-slide images of histological lymph node sections. By applying a convolutional neural network (CNN) architecture, previous research shows 92.4% sensitivity on the CAMELYON dataset, which is significantly higher than a human pathologist attempting an exhaustive search, which resulted in a sensitivity of 73.2%.<sup>[6]</sup> In addition, deep learning can be used to predict clinical outcomes directly from histological images. Based on more than 100,000 H&E image patches, a CNN was trained with a nine-class accuracy of >94%. Such information has proven useful for improving survival prediction compared to the Union for International Cancer Control staging system.<sup>[9]</sup>

To evaluate whether CNNs could be used as an adequate assistant in a clinical setting, we applied a deep learning-based approach to identify tumor regions in a model of colorectal cancer and compared the results to that performed by a pathologist. Our approach achieves an accuracy of 99.9% for normal slides and 94.8% for cancer slides on H&E-stained histology slides compared to pathologists. Automation of this task would save highly trained pathologist resources in a high-volume molecular laboratory.

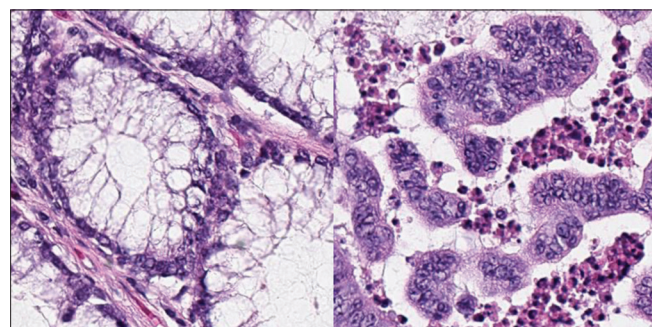
## METHODS

### Dataset

We obtained 322 colorectal cancer-related digital slides from St. Paul's Hospital with all patient identifiers removed. Among the collected images, 15 slides were excluded because we could not obtain complete annotations due to poor image quality. The remaining 307 slides were exhaustively annotated by a pathologist where the boundary between cancer and normal tissue was drawn manually. It was assumed that the neoplastic epithelium above invasive cancer was also cancer. Among the 307 slides, 85 were from normal colorectal tissue and 222 slides contained various proportions of colorectal cancer. We randomly selected 275 slides (76 normal controls and 199 containing colorectal cancer) for training; the remaining slides were used for testing (9 controls and 23 colorectal cancer). The study was approved by the UBC/BC Cancer Research Ethics Board.

### Data preprocessing

We applied patch-based approaches in the current study for handling gigapixel pathology images, as the size of a digital



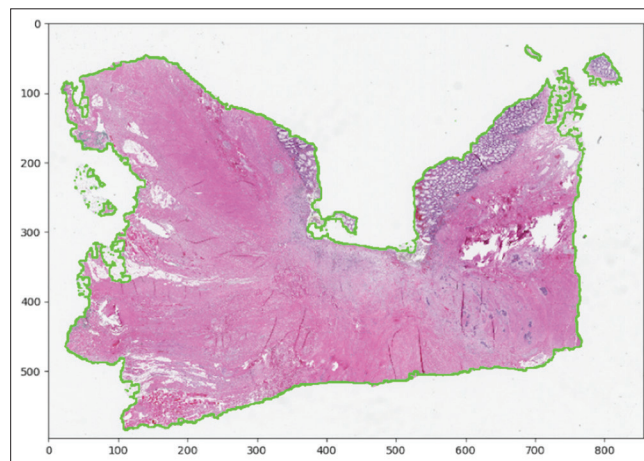
**Figure 1:** Example of tissue content in a patch comprising 768 × 768 pixels (left: tumor negative; right: tumor positive)

image of a tissue section can be very large and there is no way to segment the entire image in a single pass. In this approach, the whole slide is divided into multiple patches, and each patch is called independently. The final segmentation is obtained by assembling all patch predictions. There are two common approaches for making the patch prediction. Classification-based approaches<sup>[6,10]</sup> divide a whole slide into multiple overlapping patches. For cancer diagnosis, patches with the majority of cells comprised cancer are labeled as cancer patches (tumor positive), while those with no cancer as normal patches (tumor negative). Therefore, the final segmentation can be generated by averaging results of patch classification from all patches. The second approach is based on patch segmentation.<sup>[11,12]</sup> This approach still generates many patches, but then segments cancer cells from normal ones for each patch. In this study, we used the former patch classification-based approach as the majority of patches in our dataset were either normal or contained mostly cancer.

We trained our model on a Dell T630 server with 128G memory and 4 Titan X GPUs for 20 epochs. The learning rate was set to  $3 \times 10^{-4}$  with Adam optimizer. The batch size was set as 92. We chose a patch size of 768 × 768 pixels in order to preserve tissue architecture information and reduce computational cost. We also tried a smaller patch size of 384 × 384 pixels; the results were similar, but the computational cost was 4 times higher.

Figure 1 shows examples of the structural information that is available with a patch size of 768 × 768 pixels. In order to further improve the efficiency of training and test sets, we removed nontissue regions and generated regions of interest (ROI) by first applying Otsu's threshold method<sup>[13]</sup> for automatically separating foreground and background, then using OpenCV 4.2 findContours function with approximation method CHAIN\_APPROX\_SIMPLE for contour detection.<sup>[14]</sup> The sample result is shown in Figure 2.

For the training data, we randomly extracted patches within ROI of all 275 training slides. All patches having more than 60% tumor were considered as tumor positive, and <40% tumor



**Figure 2:** Regions of interest detecting based on contours and Otsu's method

were considered as tumor negative, while patches showing between 40% and 60% tumor were not considered, as such a measurement may not be reliable due to the precision of the annotation. Since there were a lot more negative patches (from normal slides and normal regions of cancer slides) than positive patches, we set upper bounds for the number of positive patches and negative patches from each slide. Overall, we obtained similar amounts of positive and negative patches in the training and test data [Table 1].

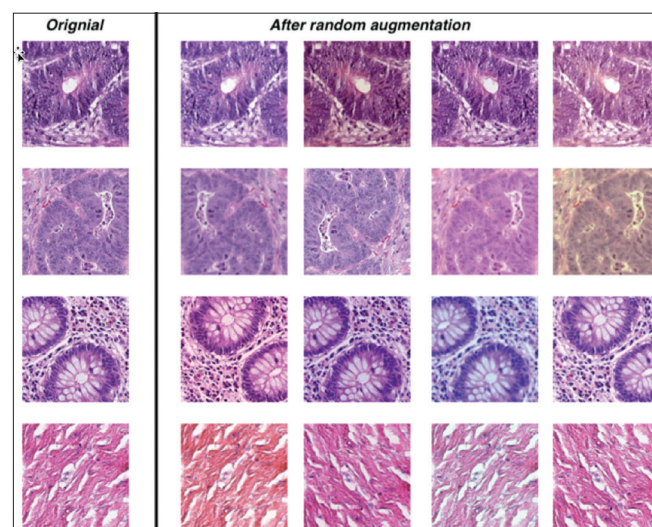
### Model training and evaluation

We chose to use PyTorch implementation of Inception V3 architecture<sup>[15]</sup> pretrained on the ImageNet dataset<sup>1</sup>. All positive and negative patches were sampled in advance. In order to increase the robustness of the model, heavy data augmentation was performed during training on the fly (e.g., flip and HSV color argumentation) in a random way and combined [Figure 3]. Before sending image data into the training process, all patches were resized to  $299 \times 299$  pixels as the standard input size of Inception V3.

In the training set, we found patch-level accuracy to be 99.1%. In order to obtain final tumor segmentation results for images in the test set, we first extracted overlapping patches (step size 384 pixels) from ROI, which means that the overlap between two patches is 50% (side adjacent), 25% (diagonal adjacent), and 0% (nonadjacent). Then, we made the prediction and produced a probability of cancer for each patch. The final heatmap of the whole slide was obtained by averaging the corresponding overlapped patch predictions. Finally, a threshold  $t$  was selected to determine the tumor regions from the heatmap ( $>t$  was considered as a tumor).

## RESULTS

We first evaluated the performance of the patch classification model on sampled patches from test slides [Table 1]. The



**Figure 3:** Patch image augmentation examples, where the first column is shows the original patches, and all other columns are images produced after random augmentation

overall classification accuracy was 95.1% on test patches. Figure 4 shows the receiver operating characteristic (ROC) curve for binary cancer/normal classification. The area under the ROC curve was 0.99.

In terms of whole-slide predictions, we used sensitivity (E1), specificity (E2), accuracy (E3), and dice coefficient (E4) to evaluate the consistency of prediction compared to annotations performed by experts. Note that normal slides do not contain tumor cells; therefore, the sensitivity and dice coefficient were not defined [Table 2].

The segmentation produces a heat mask with probability for each pixel being cancer/normal for each slide. We empirically selected a threshold value of 0.65 to separate tumor cells from normal. We computed the prediction performance for each slide and summarized the overall performance across all slides

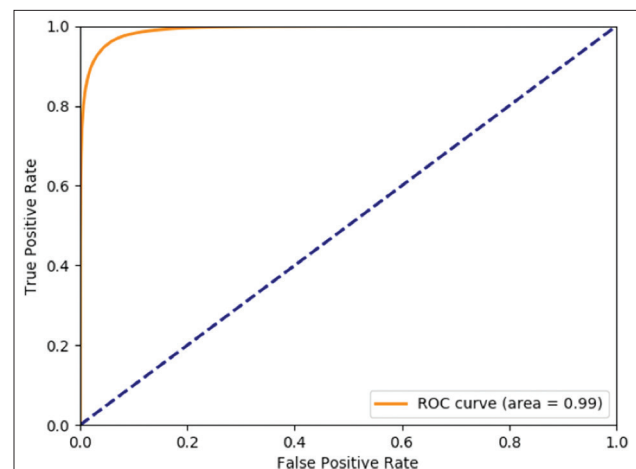
**Table 1: Data for training and test**

Data Sets	Normal Slides	Cancer Slides	Positive Patches	Negative Patches
Training set	76	199	2,481,335	2,378,267
Test set	9	23	308,643	311,690

**Table 2: Definition of the performance metrics**

Metric	Label
$Sensitivity = \frac{TP}{TP + FN}$	E 1
$Specificity = \frac{TN}{TN + FP}$	E 2
$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$	E 3
$Dice\ coefficient = \frac{2TP}{2TP + FP + FN} = \frac{2 X \cap Y }{ X  +  Y }$	E 4

$|X|$  and  $|Y|$  are the cardinalities of annotated and predicted cancer regions. Where TP: True positive, TN: True negative, FP: False positive, FN: False negative



**Figure 4:** Patch-level classification performance



in Table 3. An example of the final segmentation is shown in Figure 5, which has an accuracy of 96.9%.

To better estimate the robustness of our model, we test its performance on a completely independent dataset of 50 CRC slides obtained from Kaohsiung Medical University Chung-Ho Memorial Hospital. Each slide was annotated by a different group of certified pathologists from this hospital. We did not perform any fine-tuning with those slides and only applied our models to make the prediction. The performance is reported in Table 4.

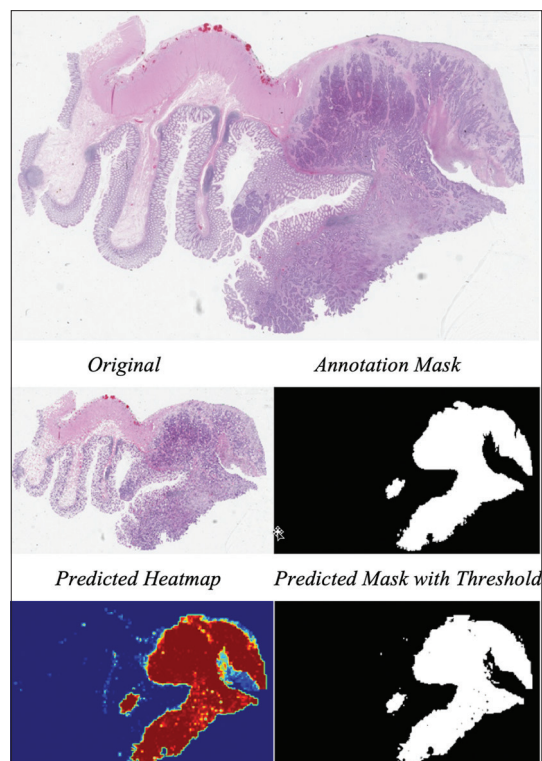
## DISCUSSION

Overall, tumor prediction using our algorithm aligns well with annotation by a pathologist. The median prediction accuracy

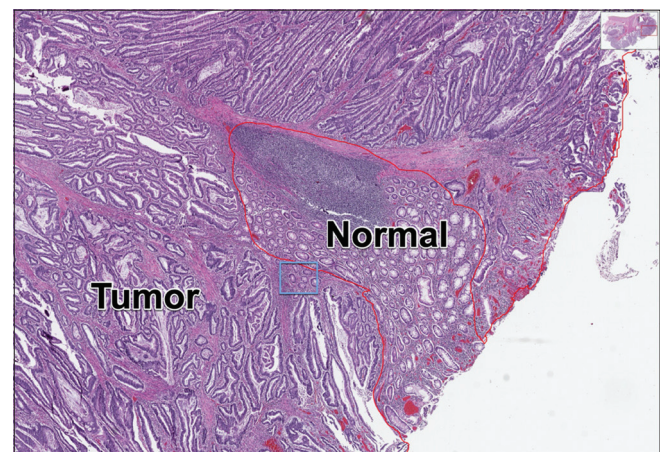
for normal slides was 99.9% and for cancer slides was 94.8%. We believe that the performance of our predictive model can be further improved by adding more representative digital slides as training data.

For normal slides, there is no ambiguity in calling ground truth and all patches are labeled as tumor negative. The model successfully identifies the pattern of normal tissue and makes accurate predictions. Our model achieves nearly perfect prediction as the accuracy and specificity approach 100%.

For cancer slides, the median slide-level prediction accuracy is 94.8% comparing to the pathologist's annotation. This observation can be explained by two reasons in addition to requiring more training data. First, we argue that the degree of detail in the annotation has a big impact on the final performance. Obtaining detailed annotation is very time-consuming. Occasionally, the pathologist's annotation does not follow the tumor boundary exactly [Figure 6], which introduces noise during model training and validation. The second possible reason is related to the classification approach we used for training models. For cancer patches,



**Figure 5:** Sample of whole-slide segmentation results



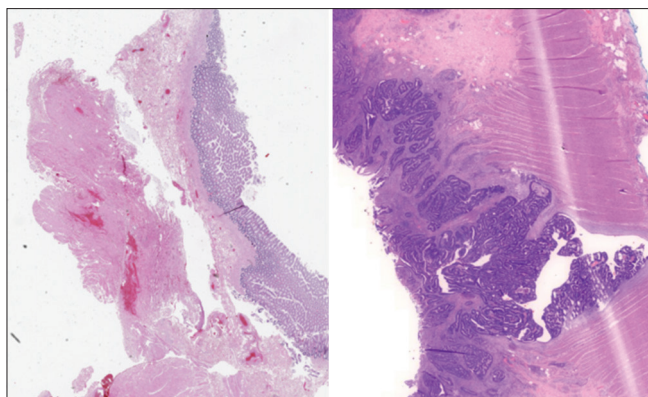
**Figure 6:** Example of an imperfect annotation by a pathologist (the region in the blue box is normal tissue but was annotated as tumor by the pathologist [red outline])

**Table 3: Summary of the prediction performance across all slides (the mean and median over the statistics of all slides)**

Type	Aggregation	Accuracy	Specificity	Sensitivity	Dice
Normal	Mean	99.9%	99.9%	NA	NA
	Median	99.9%	99.9%	NA	NA
Cancer	Mean	93.6%	94.3%	91.3%	88.5%
	Median	94.8%	95.9%	94.3%	93.1%

**Table 4: Summary of the prediction performance across all slides on the independent dataset (the mean and median over the statistics of all slides)**

Type	Aggregation	Accuracy	Specificity	Sensitivity	Dice
Cancer	Mean	87.8%	90.0%	85.2%	87.2%
	Median	88.1%	91.1%	88.4%	88.2%



**Figure 7:** H&E staining differences. (Left: Training/test dataset; Right: Independent test dataset)

the labels are determined by the percentage of cancer regions in patches. Therefore, small variations may cause the label to flip from tumor positive to tumor negative or vice versa. This ambiguity may cause the model to be less accurate at the edge of a tumor region.

Except for both sets of H&E slides coming from patients with CRC, our independent test dataset had nothing in common with the training data. The patients were all from Asia, the slides had different tinctorial properties of the H&E staining [Figure 7], and the annotations were performed by separate group of pathologists. Despite these significant differences, our model still performed well on this independent dataset. The mean dice coefficient was 87.2% [vs. 88.5% from Table 1]. In terms of accuracy (specificity and sensitivity), we observed a 5.8% (4.3% and 6.1%) performance difference in the prediction accuracies.

## CONCLUSION

Computer-aided diagnosis systems may become one of the most useful applications of deep learning. A technology that improves the detection of tumors for colorectal cancer can be easily adapted for detecting many other types of diseases. Given the shortage of pathologists, this type of software may be ideal for screening cases to select the best regions for pathologists to focus on when making diagnoses of cancer. It can also be integrated into genomic analytic pipelines to reduce pathologist workload and improve turnaround times of genomic profiling.

## Financial support and sponsorship

The study was financially supported by the NRC-IRAP research grants.

## Conflicts of interest

This work was funded in part by GenerationsE Software Solutions, Inc. Authors (L. Xu, Y. Tong, C. Xu and Y.C. Su) are employees of GenerationsE.

## REFERENCES

1. Brunnström H, Johansson A, Westbom-Fremer S, Backman M, Djureinovic D, Patthey A, *et al.* PD-L1 immunohistochemistry in clinical diagnostics of lung cancer: Inter-pathologist variability is higher than assay variability. *Mod Pathol* 2017;30:1411-21.
2. Deng J, Dong W, Socher R, Li L.J, Li K, Fei-Fei L. ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami; 2009.
3. Sutskever I, Vinyals O, Le QV. Sequence to Sequence Learning with Neural Networks, in *Advances in Neural Information Processing Systems* 27. New York, United States: Curran Associates, Inc.; 2014. p. 3104-12.
4. Sarraf S, Tofighi G. Deep Learning-Based Pipeline to Recognize Alzheimer's Disease Using fMRI Data, in *Future Technologies Conference*, San Francisco; 2016.
5. Shan H, Padole A, Homayounieh F, Kruger U, Khera RT, Nitiwarangkul C, *et al.* Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction. *Nat Mach Intell* 2019;1:269-76.
6. Liu Y, Gadepalli K, Norouzi M, Dahl GE, Kohlberger T, Boyko A, *et al.* Detecting Cancer Metastases on Gigapixel Pathology Images," *arXiv:1703.02442v2*; 2017.
7. Min JK, Kwak MS, Cha JM. Overview of deep learning in gastrointestinal endoscopy. *Gut Liver* 2019;13:388-93.
8. Bandi P, Geessink O, Manson Q, Van Dijk M, Balkenhol M, Hermesen M, *et al.* From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 challenge. *IEEE Trans Med Imaging* 2019;38:550-60.
9. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis CA, *et al.* Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med* 2019;16:1-22.
10. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep learning for identifying metastatic breast cancer. *arXiv:1606.05718v1*; 2016.
11. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans Pattern Anal Mach Intell* 2018;40:834-48.
12. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Med Image Comput Comput Assist Interv* 2015;9351:234-41.
13. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 1979;9:62-6.
14. Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell* 1986;8:679-98.
15. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016.