Classifying Breast Histopathology Images with a Ductal Instance-Oriented Pipeline

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Beibin Li*, Ezgi Mercan, Sachin Mehta,
Stevan Knezevich, Corey W. Arnold, Donald L. Weaver, Joann G. Elmore, Linda G. Shapiro
Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle

Introduction

- Ductal Regions are Important for Breast Cancer Diagnosis [4]
- Breast Cancer Often Starts within Ducts or Lobules [6]
- Traditional Pattern Recognition Tools Can Hardly Extract Each Duct from Conglomerated Region
- Deep Learning-based Instance Segmentation Model (e.g. [2]) Could Help
- Instance Segmentation-Labeling is a Tedious and Time-Consuming Task

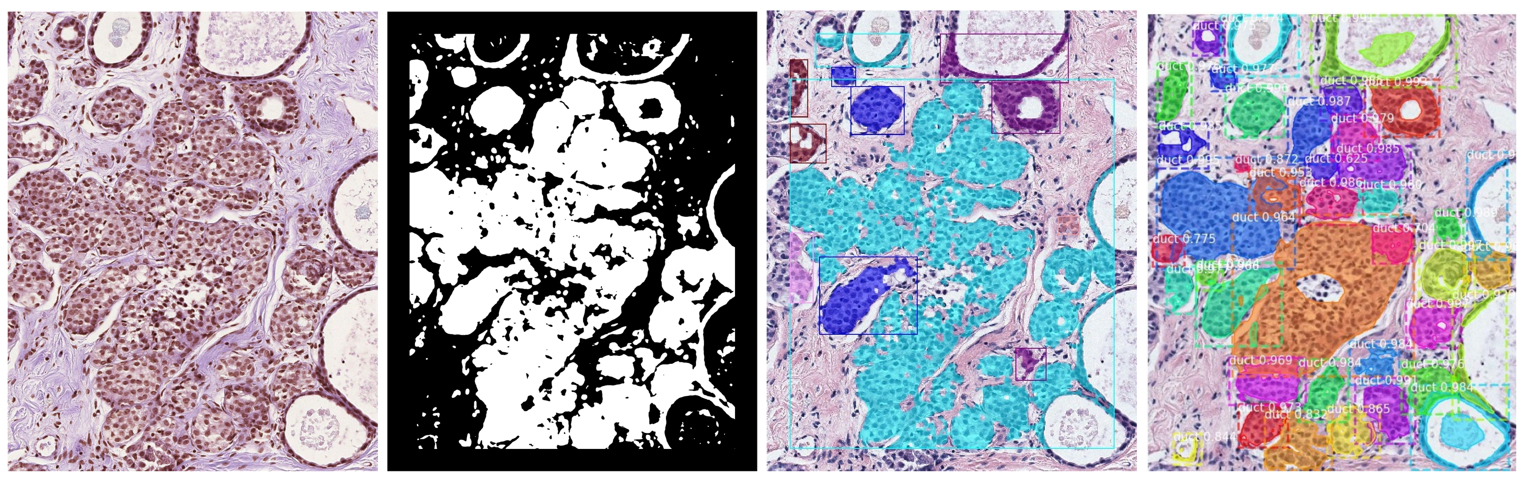

Figure 1: Duct instances: From Left to Right: the input image in RGB color space; (b) the binary image inferred from tissue-level semantic segmentation; (c) duct instances found by mathematical morphology and connected component algorithm; (d) the ducts inferred from our system.

Data and Annotation

- Digital Whole Slide Images from Residual Breast Biopsy Material [5, 7, 1]
- No Instance Segmentation Labels
- Total 428 Histopathological ROIs
- 4 Classes: Benign, Atypia, Ductal Carcinoma in Situ, or Invasive Cancer
- Existing Semantic Segmentation Model [3] for Semantic Segmentation

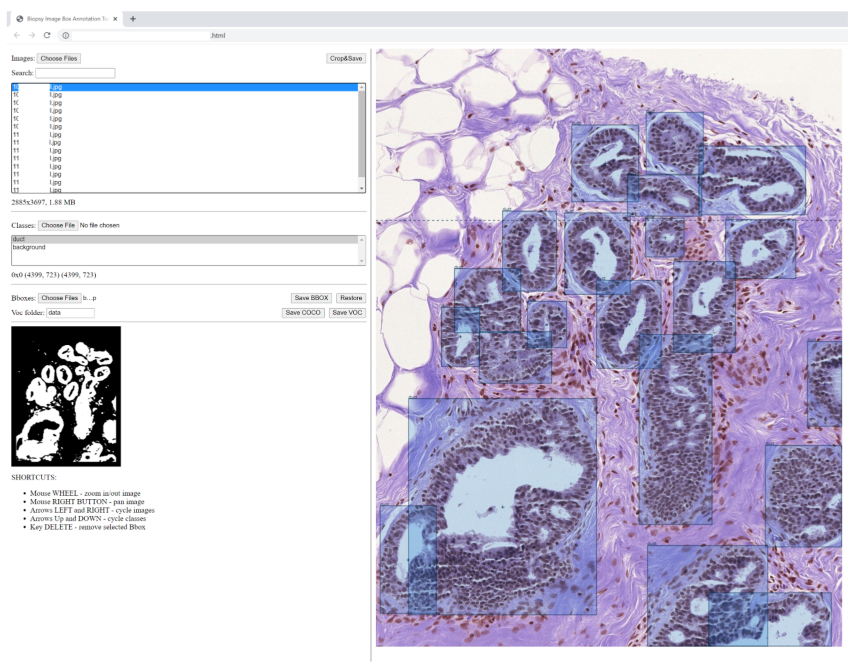

Figure 2: Weakly Supervised Annotation Interface.

- Weakly Supervised Annotation Tool
- Human-AI Collaboration
- AI-Guided Weak Annotation for Human Annotator
- Generate Instance Segmentation Label as Silver Standard
- Labelled 100 ROIs to Train Instance Segmentation Model

DIOP System

- Mask R-CNN for Instance Segmentation
- Y-Net for Semantic Segmentation
- Traditional Feature Extraction: Frequency Features, Co-Occurrence Features
- Features from 3 Different Levels

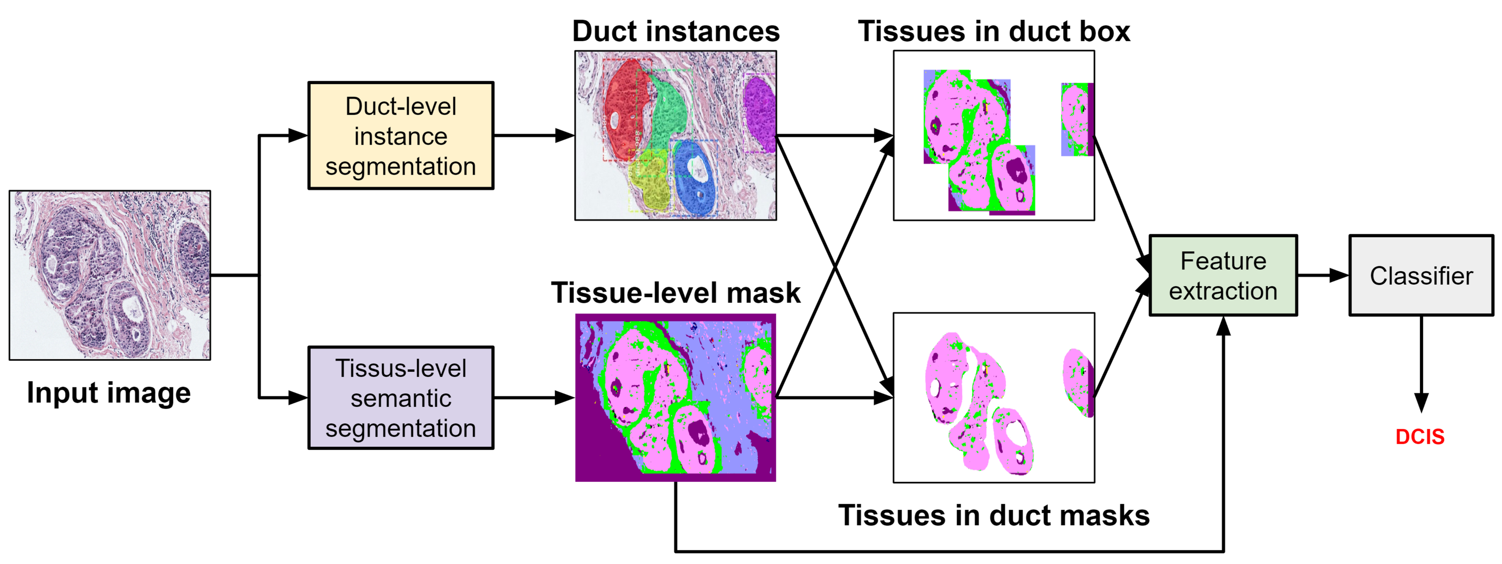

Figure 3: Pipeline

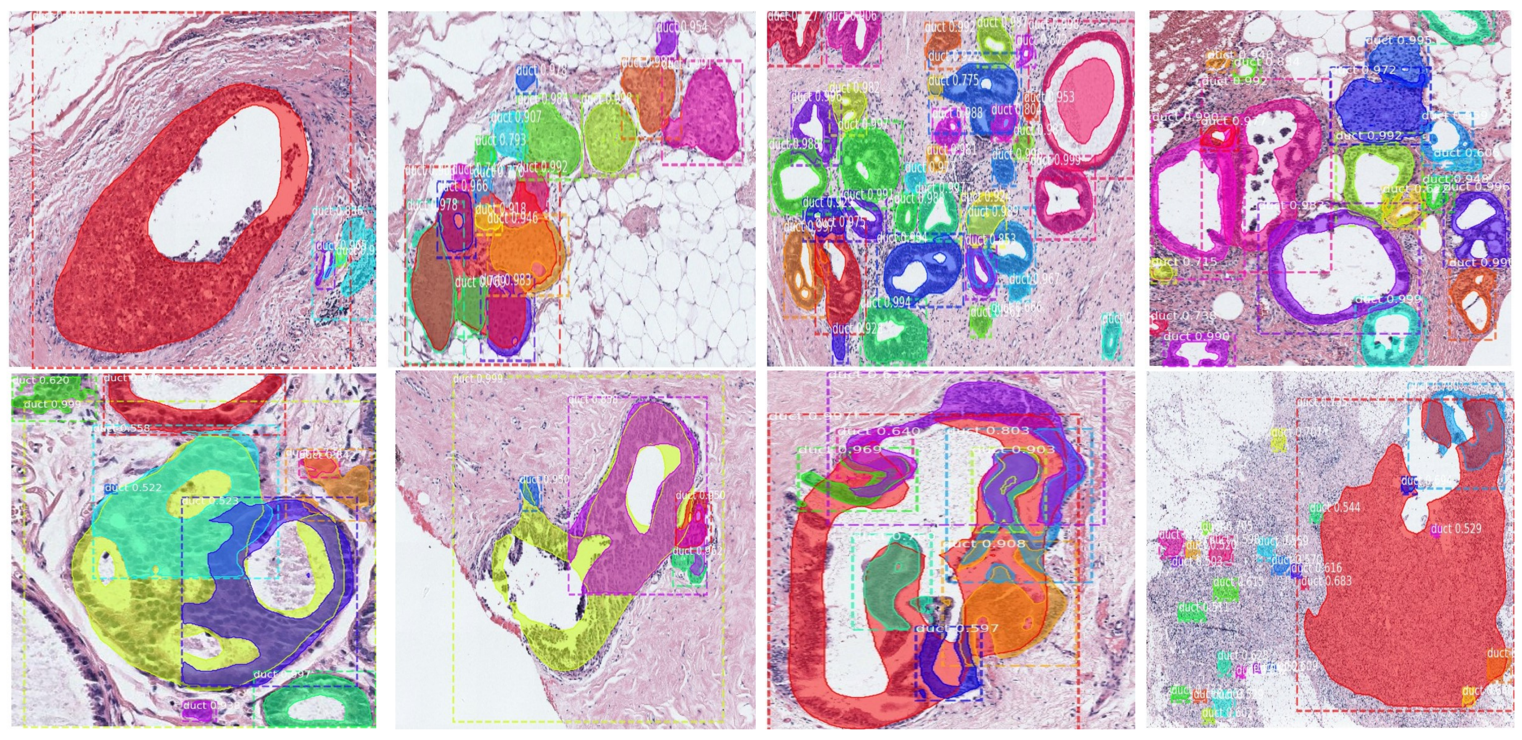

Figure 4: Testing Results for Instance Segmentation. Compare to the Silver Standard, the mIoU is 72%

Results

- Outperforms Previous Approaches
- Reaches Human Expert's Performance
- Faster than Superpixel-based Approaches
- Combining Three Levels of Features Improves the Results

Task	Features	Sensitivity	Specificity	Accuracy	F_1
Invasive vs Non-invasive	*Pathologists*	*0.84*	*0.99*	*0.98*	*0.86*
	Superpixel Features	**0.70**	0.95	0.94	0.62
	Structure Features	0.49	0.96	0.91	0.51
	Duct-RCNN (Ours)	0.62	**0.98**	**0.95**	**0.73**
Atypia and DCIS vs Benign	*Pathologists*	*0.72*	*0.62*	*0.81*	*0.51*
	Superpixel Features	0.79	0.41	0.70	0.46
	Structure Features	**0.85**	0.45	0.70	0.50
	Duct-RCNN (Ours)	**0.85**	**0.63**	**0.79**	**0.59**
DCIS vs Atypia	*Pathologists*	*0.70*	*0.82*	*0.80*	*0.76*
	Superpixel Features	0.88	0.78	0.83	0.86
	Structure Features	0.89	0.80	0.85	0.87
	Duct-RCNN (Ours)	**0.91**	**0.89**	**0.90**	**0.92**

Figure 5: Comparison with SOTA Methods: Cascade Binary Classification Model

Method	Accuracy
Pathologists	0.70
MIL with max-pooling	0.55
MIL with learned fusion	0.67
Semantic Learning	0.55
Y-Net	0.63
DIOP (Ours)	**0.70** $\pm$ 0.02

Method	Accuracy
Tissue in ROI	0.67
Tissue in Duct box	0.66
Tissue in Duct mask	0.69
Tissue in Duct mask + ROI	0.69
Tissue in Duct box + ROI	0.67
Tissue in Duct box + mask	0.69
Tissue (All)	**0.70**

Figure 6: Comparison with SOTA Methods: Four-Way Classification

Takeaways

- More Clinical Studies are Needed
- Weak Annotation is a Effective Tool for Medical Analysis
- Doctor-AI Collaboration could Benefit Both Communities

Acknowledgement

Research reported in this article was supported by grants R01 CA172343, R01 CA140560, U01CA231782, and R01 CA200690 from the National Cancer Institute of the National Institutes of Health.

References

[1] Patricia A Carney et al. "The New Hampshire Mammography Network: the development and design of a population-based registry." In: *AJR. American journal of roentgenology* 167.2 (1996), pp. 367–372.

[2] Kaiming He et al. "Mask R-CNN". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.

[3] Sachin Mehta et al. "Y-Net: joint segmentation and classification for diagnosis of breast biopsy images". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 893–901.

[4] Ezgi Mercan et al. "Assessment of Machine Learning of Breast Pathology Structures for Automated Differentiation of Breast Cancer and High-Risk Proliferative Lesions". In: *JAMA network open* 2.8 (2019), e198777–e198777.

[5] Natalia V Oster et al. "Development of a diagnostic test set to assess agreement in breast pathology: practical application of the Guidelines for Reporting Reliability and Agreement Studies (GRRAS)". In: *BMC Women's Health* 13.1 (2013), p. 3.

[6] American Cancer Society. "Breast cancer facts & figures 2019-2020". In: *Atlanta: American Cancer Society* (2019).

[7] Donald L Weaver et al. "Predicting biopsy outcome after mammography: what is the likelihood the patient has invasive or in situ breast cancer?" In: *Annals of surgical oncology* 12.8 (2005), pp. 660–673.