Visual Transformers for Whole Slide Image Diagnosis

Wenjun Wu
05/18/2022

UNIVERSITY *of* WASHINGTON

Outline

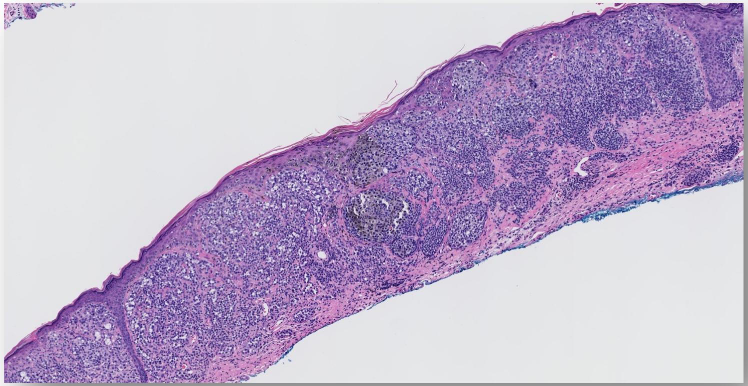
- Background and Goal
- Dataset
- Related Work
- Our Work
  - HATNet
  - ScAtNet
- Next Step

Background

UNIVERSITY *of* WASHINGTON

What is Melanoma?

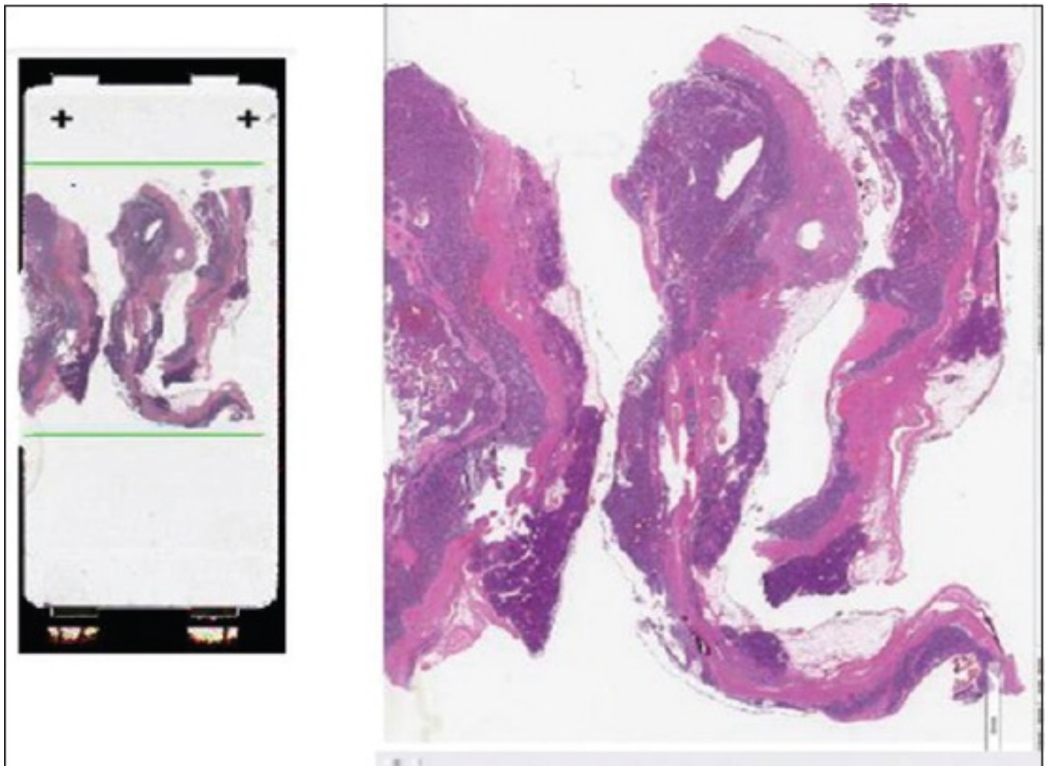
- Melanoma is the most aggressive type of skin cancer.
- Melanoma occurs when UV radiation triggers DNA damages in the melanocytes
- The "gold standard" for diagnosis of invasive melanoma relies on the visual assessments of skin biopsy images by pathologists.



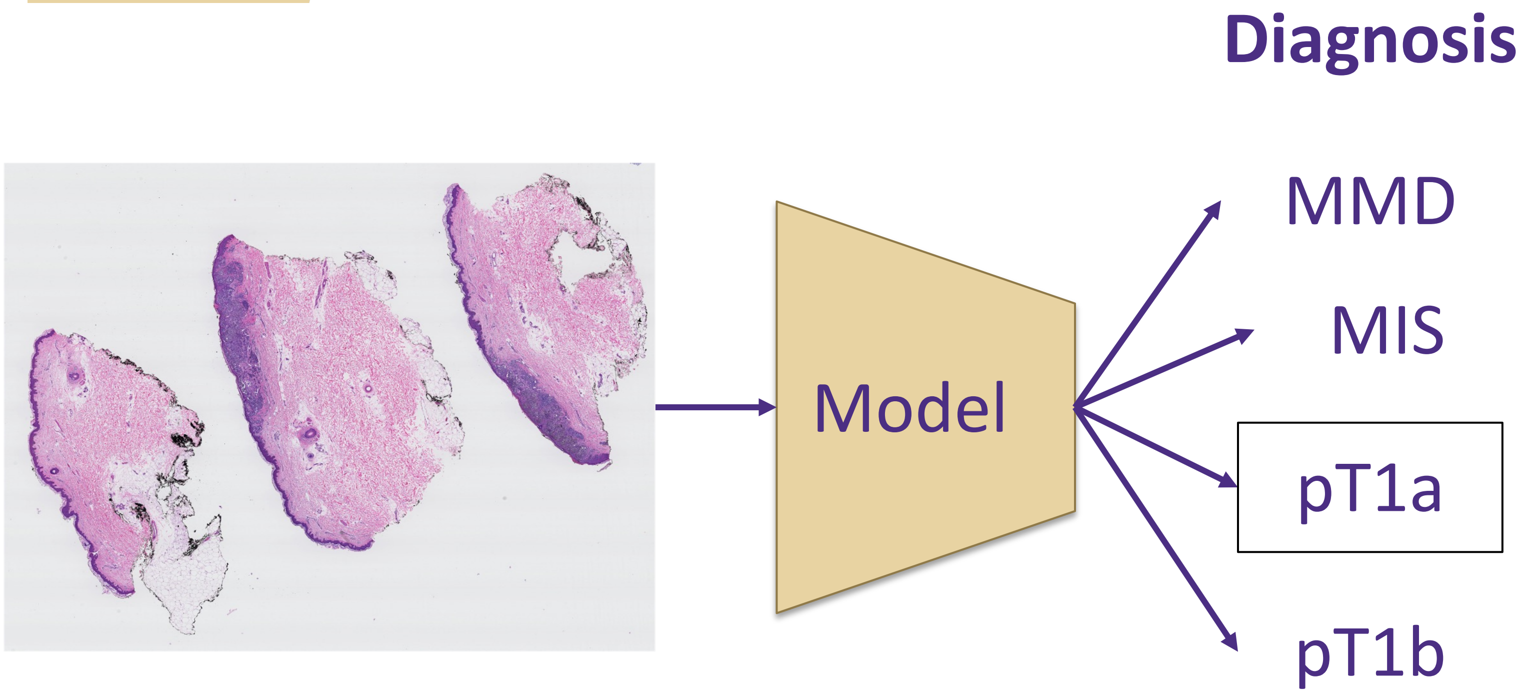
An example of an Invasive Melanoma T1b in M-Path dataset.

Why melanoma diagnosis?

- Unfortunately, diagnostic errors are common
- Computer-aided diagnostic system can be a second reader and help reduce uncertainties



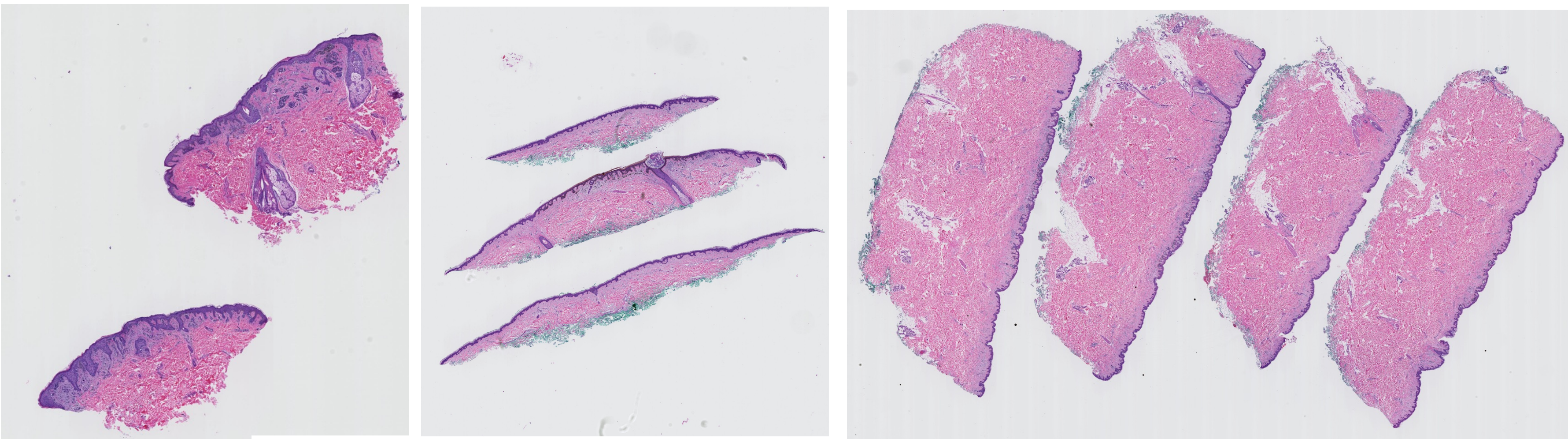
Goal



Dataset

UNIVERSITY *of* WASHINGTON

Melanoma Dataset

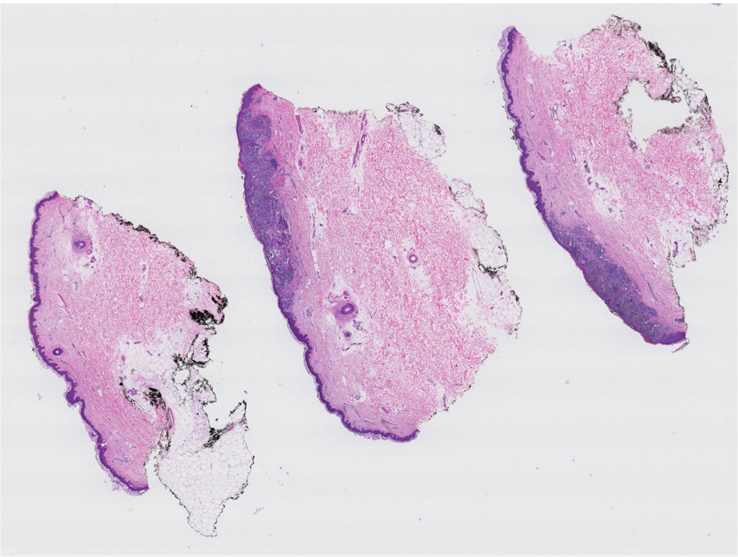


Diagnostic Category	Number of WSIs				Average WSI size (in pixels)
	Training	Validation	Test	Total	
MMD	26	6	29	61	11843×10315
MIS	25	5	30	60	9133×8501
pT1a	33	6	34	73	9490×7984
pT1b	18	6	22	46	14858×12154
Total	102	23	115	240	11130×9603

Difficulties in diagnosis

Size of whole slide images

An example image from ImageNet [500 x 375]



An example WSI at 10x [15264 x 19824]

Difficulties in diagnosis

Size of whole slide images

Dataset size

Diagnostic Category	Number of WSIs				Average WSI size (in pixels)
	Training	Validation	Test	Total	
MMD	26	6	29	61	11843×10315
MIS	25	5	30	60	9133×8501
pT1a	33	6	34	73	9490×7984
pT1b	18	6	22	46	14858×12154
Total	102	23	115	240	11130×9603

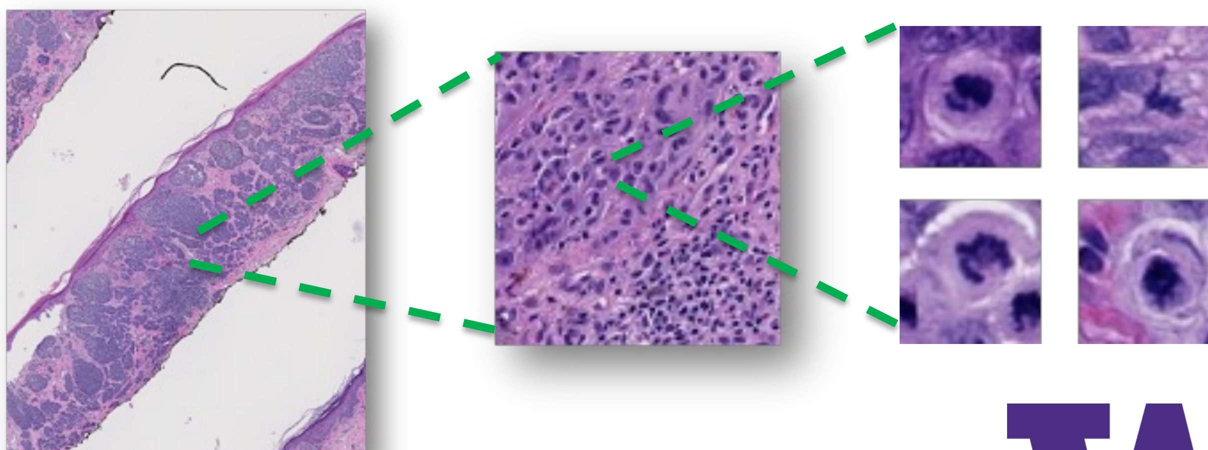
TABLE 1: Statistics of skin biopsy whole slide image (WSI) dataset. The average WSI size is computed at a magnification factor of x10. Diagnostic terms for the dataset used in this study are as follows: mild and moderate dysplastic nevi (MMD), melanoma in situ (MIS), invasive melanoma stage pT1a (pT1a), invasive melanoma stage ≥ pT1b (pT1b).

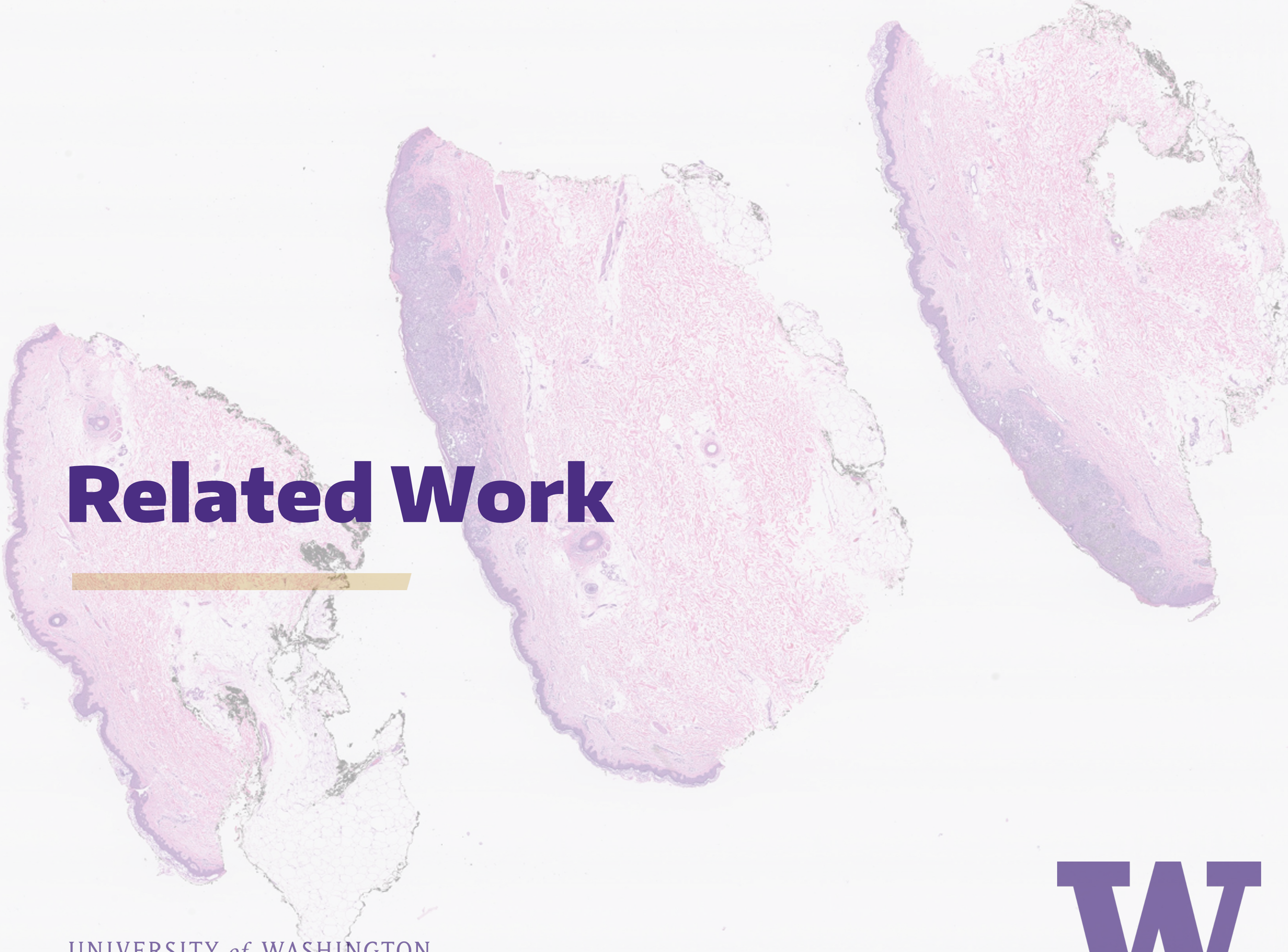
Difficulties in diagnosis

Size of whole slide images

Dataset size

cancerous structure vs. normal structure



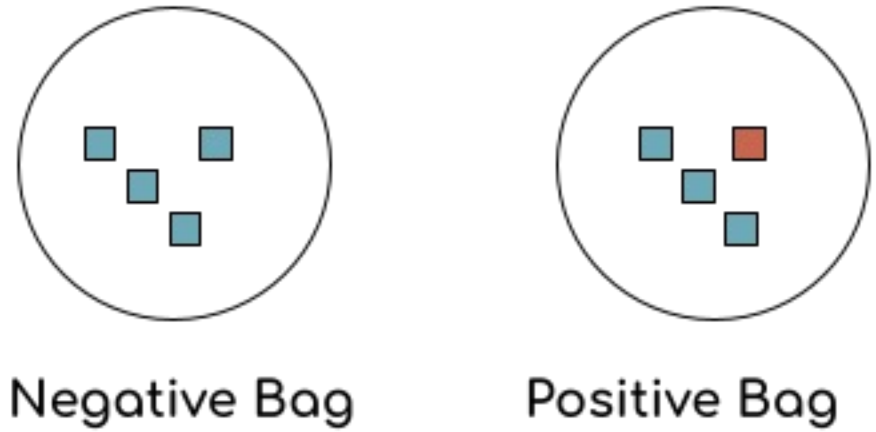


Related Work

UNIVERSITY *of* WASHINGTON

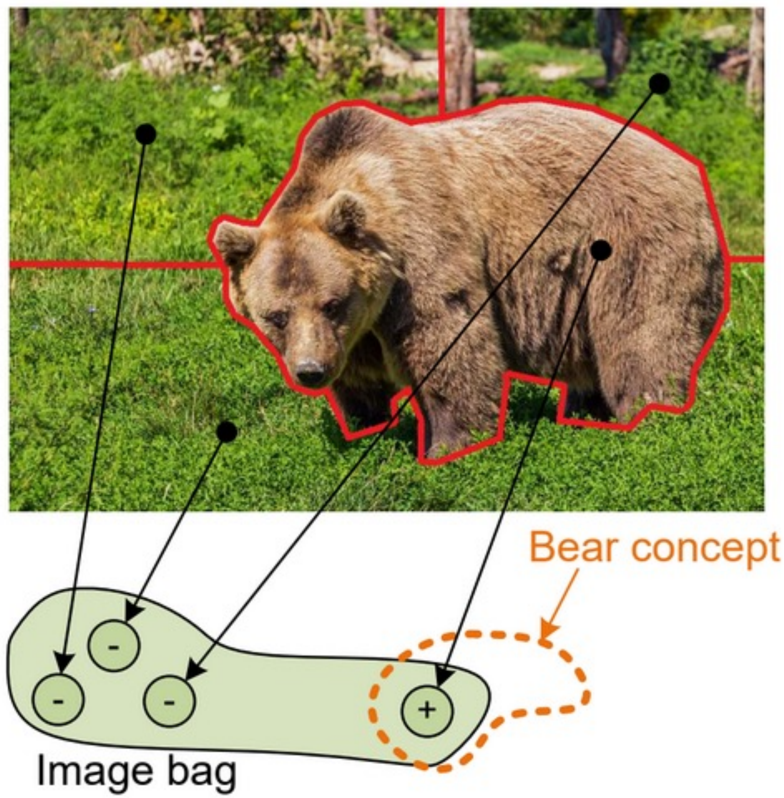
Related Work

> Multiple Instance Learning



Related Work

> Multiple Instance Learning



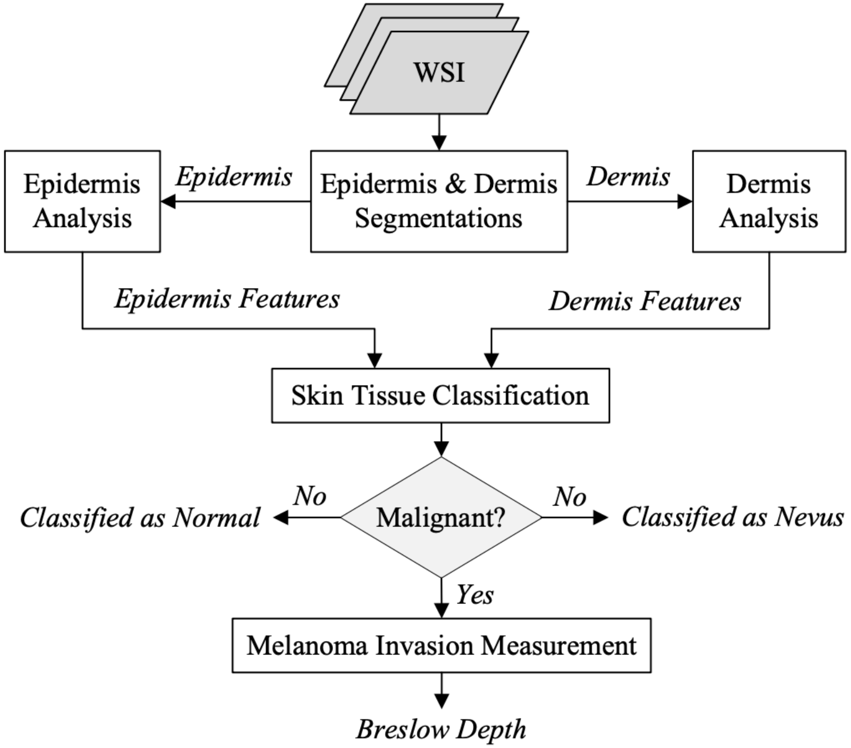
Related Work

> **Multiple Instance Learning**

+ reduce high computational cost
+ effective in learning instance/bag-wise representation
- Does not allow long-range/global feature interaction
- Prone to label ambiguity/noise

Related Work

> Segmentation-based methods



Hongming Xu, Cheng Lu, Richard Berendt, Naresh Jha, and Mrinal Mandal. Automated analysis and classification of melanocytic tumor on skin whole slide images. Computerized medical imaging and graphics, 66:124–134, 2018.

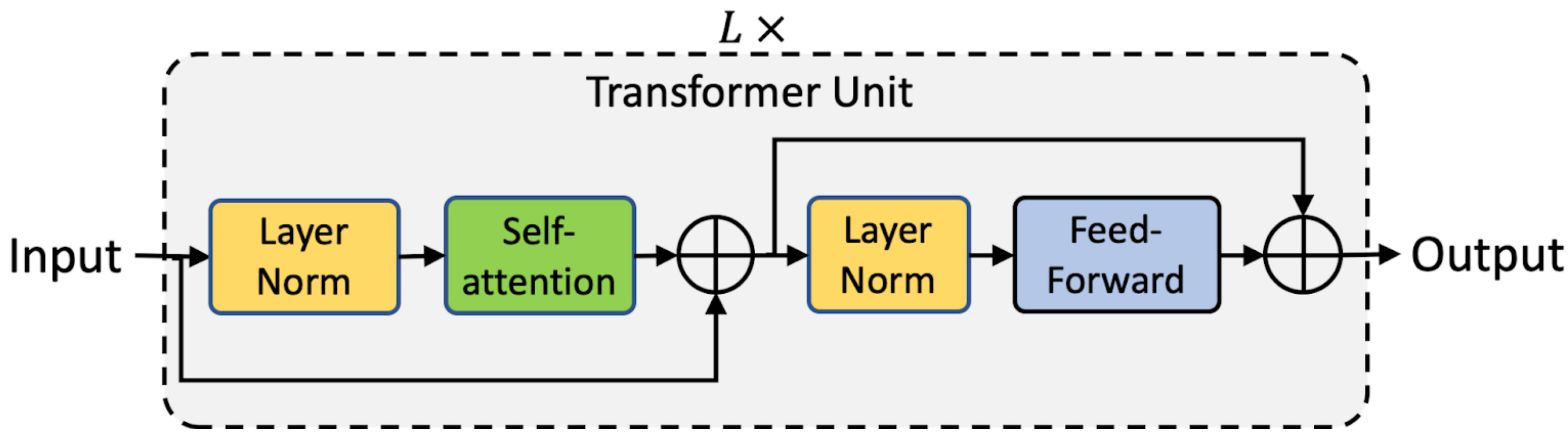
Related Work

> **Segmentation-based methods**

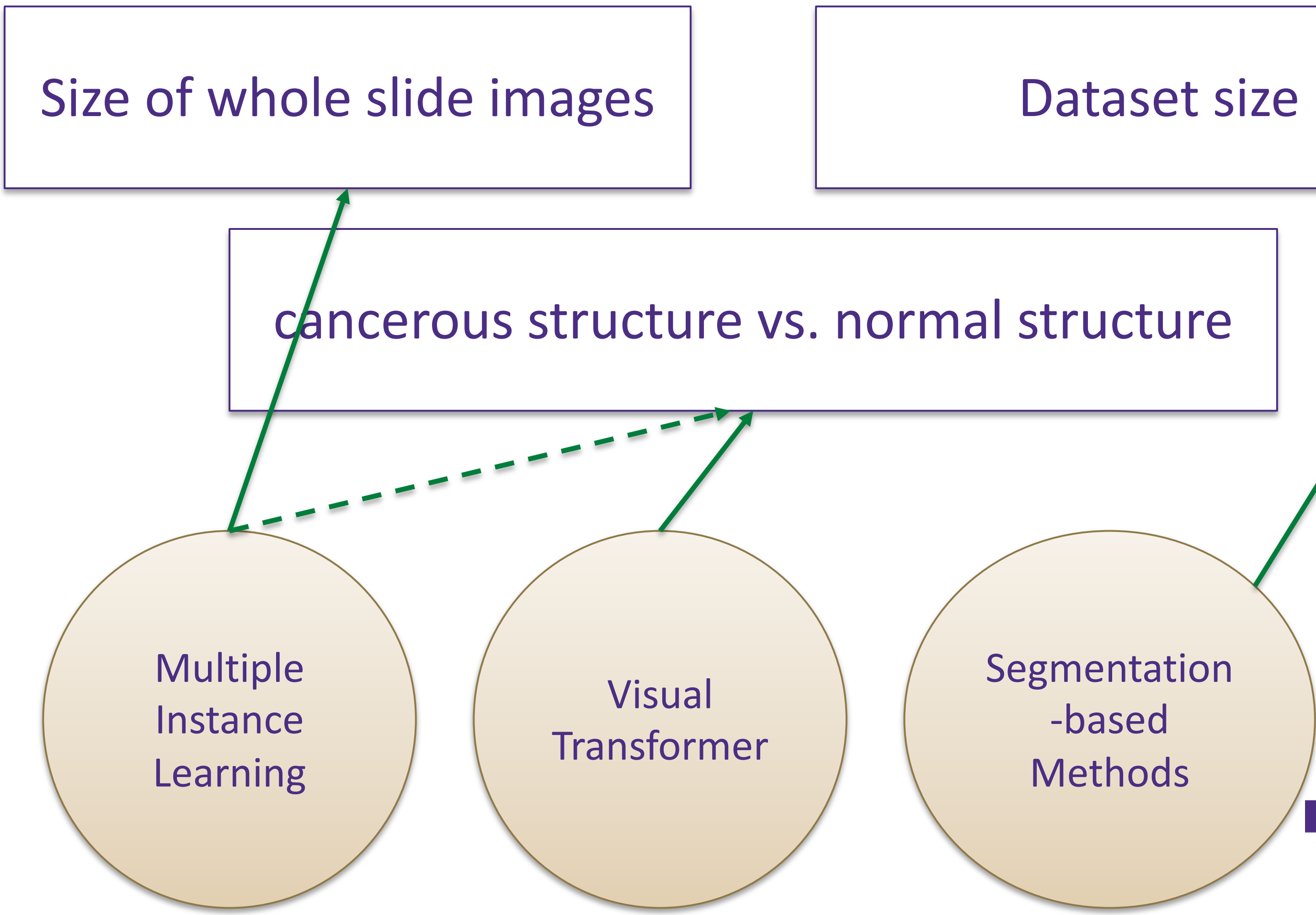
+ Learns global representation
+ More effective (better performance) on small dataset
- Require fine tissue-level segmentation masks
- Diagnostic performance highly dependent on segmentation quality

Related Work

> Visual Transformers



Difficulties in diagnosis



Our Work

UNIVERSITY *of* WASHINGTON

HATNet

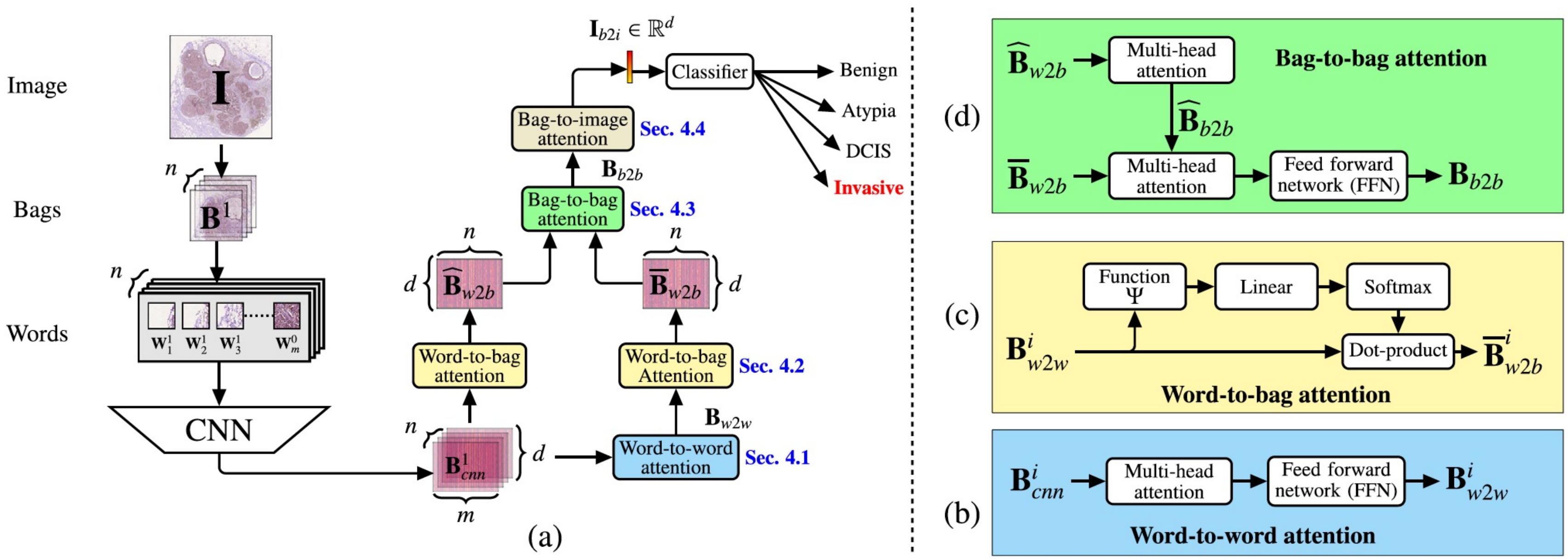
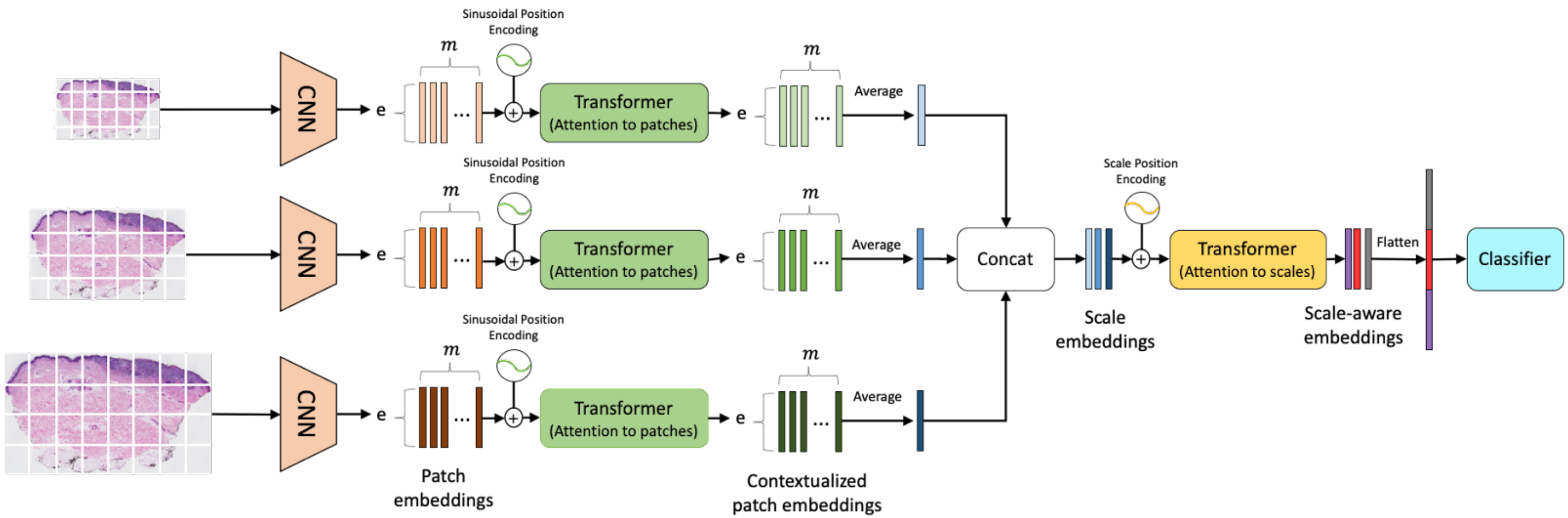


Image
Bags
Words
CNN
Classifier
Benign
Atypia
DCIS
Invasive
Bag-to-image attention
Sec. 4.4
Bag-to-bag attention
Sec. 4.3
Word-to-bag attention
Word-to-bag Attention
Sec. 4.2
Word-to-word attention
Sec. 4.1
(a)
(d)
Multi-head attention
Bag-to-bag attention
Feed forward network (FFN)
(c)
Function Ψ
Linear
Softmax
Dot-product
Word-to-bag attention
(b)
Multi-head attention
Feed forward network (FFN)
Word-to-word attention

HATNet (on a breast dataset)

- Outperforms CNN-based methods by a large margin
- Significant overlap between top bags, words and annotations of clinical biomarkers
- Learned representations from clinically relevant tissue structures without any supervision

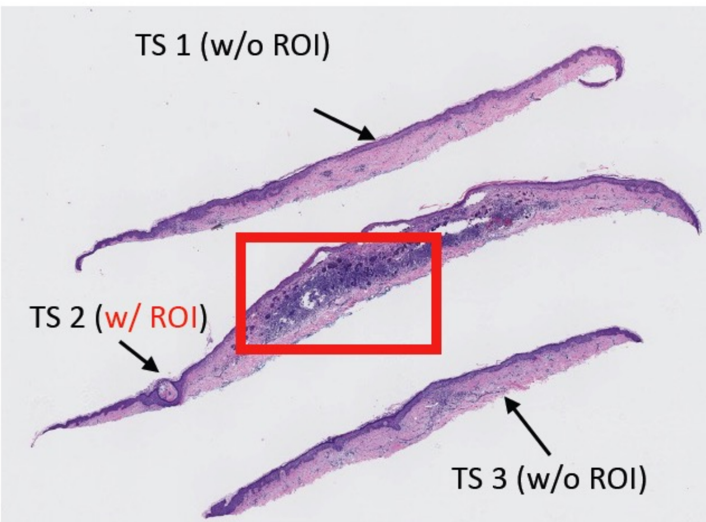
ScAtNet



Sinusoidal Position Encoding
CNN
e
m
Transformer (Attention to patches)
Average
Concat
Scale Position Encoding
Transformer (Attention to scales)
Flatten
Classifier
Patch embeddings
Contextualized patch embeddings
Scale embeddings
Scale-aware embeddings

ScAtNet: Soft Label

Invasive T1a Skin Biopsy Image (or Class 3)



Hard Label (one-hot encoding)

TS 1	0	0	1	0
TS 2	0	0	1	0
TS 3	0	0	1	0

Label smoothing (smoothing=0.1)

TS 1	0.033	0.033	0.9	0.033
TS 2	0.033	0.033	0.9	0.033
TS 3	0.033	0.033	0.9	0.033

Constrained label smoothing

TS 1	0.5	0.5	0	0
TS 2	0	0	1	0
TS 3	0.5	0.5	0	0

Soft labels (ours)

TS 1	0.54	0.46	0	0
TS 2	0	0	1	0
TS 3	0.28	0.72	0	0

ScAtNet

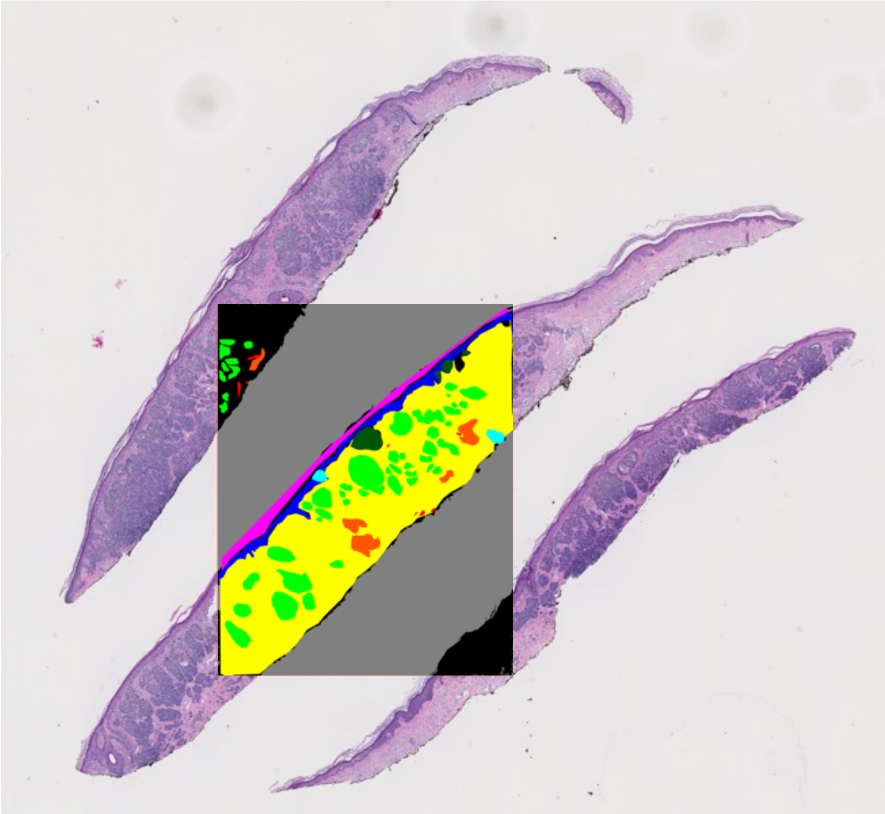
- > Outperforms MIL and CNN based methods
- > Achieves comparable performance to 187 practicing U.S. pathologists
- > Saliency analysis shows that ScAtNet learns to weigh features from different scales

Input scales			Accuracy	F1	Sensitivity	Specificity	AUC
7.5×	10×	12.5×					
✓			0.55	0.55	0.55	0.85	0.75
	✓		0.60	0.60	0.60	0.87	0.77
		✓	0.61	0.61	0.61	0.87	0.78
✓	✓		0.64	0.64	0.64	0.88	0.79
✓		✓	0.63	0.63	0.63	0.88	0.80
	✓	✓	0.63	0.63	0.63	0.88	0.79
✓	✓	✓	0.63	0.63	0.63	0.88	0.79

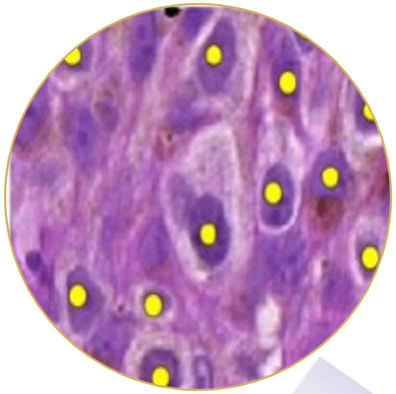
(a) Overall performance of `ScAtNet`

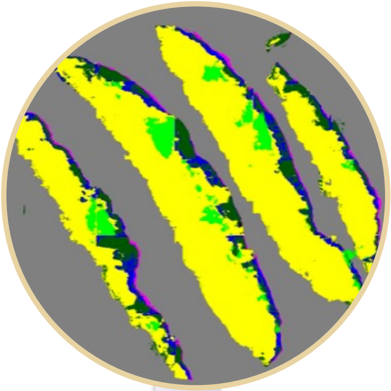
Next Step

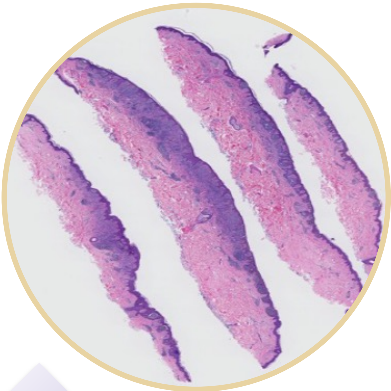
Semantic Segmentation-based Method



How do we combine everything?







Diagnoses:

Mild Dysplastic

Moderate Dysplastic

Melanoma in situ

Invasive T1a

Invasive T1b

Acknowledgement

Research reported in this study was supported by grants R01CA200690 and U01CA231782 from the National Cancer Institute of the National Institutes of Health, 622600 from Melanoma Research Alliance, and W81XWH-20-1-0798 from the United States Department of Defense.

Advisor:

Dr. Linda Shapiro

PI:

Dr. Joann Elmore

Pathologists:
Dr. Stevan Knezevich
Dr. Caitlin May
Dr. Oliver Chang
Dr. Mojgan Mokhtari
Dr. Donald Weaver

Collaborators:
Shima Nofallah
Ximing Lu
Dr. Sachin Mehta

Reference

C. Mercan, B. Aygunes, S. Aksoy, E. Mercan, L. G. Shapiro, D. L.Weaver, and J. G. Elmore, “Deep feature representations for variable-sized regions of interest in breast histopathology,” IEEE Journal of Biomedical and Health Informatics, 2020.

E. Mercan, L. G. Shapiro, T. T. Brunyé, D. L. Weaver, and J. G. Elmore, “Characterizing diagnostic search patterns in digital breast pathology: scanners and drillers,” Journal of digital imaging, vol. 31, no. 1, pp. 32–41, 2018.

H. Pinckaers, W. Bulten, J. Van der Laak, and G. Litjens, “Detection of prostate cancer in whole-slide images through end-to-end training with image-level labels,” IEEE transactions on medical imaging, vol. PP, March 2021.

Elmore et al., “Diagnostic concordance among pathologists interpreting breast biopsy specimens,” JAMA, 2015.

J. G. Elmore, R. L. Barnhill, D. E. Elder, G. M. Longton, M. S. Pepe, L. M. Reisch, P. A. Carney, L. J. Titus, H. D. Nelson, T. Onega et al., “Pathologists’ diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study,” Bmj, vol. 357, 2017.

K. H. Allison, L. M. Reisch, P. A. Carney, D. L. Weaver, S. J. Schnitt, F. P. O’Malley, B. M. Geller, and J. G. Elmore, “Understanding diagnostic variability in breast pathology: lessons learned from an expert consensus review panel,” Histopathology, vol. 65, no. 2, pp. 240–251, 2014.

Mehta, Sachin, et al. "End-to-End Diagnosis of Breast Biopsy Images with Transformers." Medical Image Analysis (2022): 102466.

Nofallah, Shima, et al. "Segmenting Skin Biopsy Images with Coarse and Sparse Annotations using U-Net." Journal of Digital Imaging (2022): 1-12.