Applications of Visual Transformers for Whole Slide Skin Biopsy Image Diagnosis

Wenjun Wu

Winter 2023
What is Melanoma?

- Melanoma is the most aggressive type of skin cancer.
- Melanoma occurs when UV radiation triggers DNA damages in 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

An example of Skin Biopsies of pT1a
Goal

Model

Diagnosis
Melanoma Dataset

Statistics of skin biopsy whole slide image (WSI) dataset. 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 (WSI)

An example image from ImageNet [500 x 375]

An example WSI at 10x [15264 x 19824]
Difficulties in diagnosis

Size of whole slide images (WSIs)

Dataset size

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 (WSIs)

Dataset size

cancerous structure vs. normal structure
Related Work
Related Work

> Multiple Instance Learning

Negative Bag

Positive Bag
Related Work

> Multiple Instance Learning

An example of Skin Biopsies of Benign

Negative Bag

An example of Skin Biopsies of Invasive T1a

positive Bag
Related Work

> Multiple Instance Learning (MIL)
  + 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

> End-to-End Learning

[Diagram showing a model leading to diagnosis]
Related Work

> Visual Transformers

Vaswani et al. Attention is all you need. Advances in neural information processing systems, 30
Related Work

> Self-attention

Vaswani et al. Attention is all you need. *Advances in neural information processing systems*, 30
Our Work
Holistic Attention Network (HATNet)

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
Scale-Aware Transformer Network (ScAtNet)

**Experimental Result: baseline methods**

<table>
<thead>
<tr>
<th>Row #</th>
<th>Method</th>
<th>Accuracy</th>
<th>F1</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>Patch-based (SSC)</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.79</td>
<td>0.67</td>
</tr>
<tr>
<td>R2</td>
<td>Patch-based (MSC)</td>
<td>0.40</td>
<td>0.40</td>
<td>0.40</td>
<td>0.80</td>
<td>0.68</td>
</tr>
<tr>
<td>R3</td>
<td>Penultimate-weighted (SSC)</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
<td>0.81</td>
<td>0.67</td>
</tr>
<tr>
<td>R4</td>
<td>Hypercolumn-weighted (SSC)</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.43</td>
<td>0.67</td>
</tr>
<tr>
<td>R5</td>
<td>Streaming CNN (SSC)</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
<td>0.77</td>
<td>0.58</td>
</tr>
<tr>
<td>R6</td>
<td>ChikonMIL (SSC)</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
<td>0.85</td>
<td>0.74</td>
</tr>
<tr>
<td>R7</td>
<td>MS-DA-MIL (SSC)</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
<td>0.83</td>
<td>0.68</td>
</tr>
<tr>
<td>R8</td>
<td>MS-DA-MIL (MSC*)</td>
<td>0.58</td>
<td>0.58</td>
<td>0.58</td>
<td>0.86</td>
<td>0.75</td>
</tr>
<tr>
<td>R9</td>
<td>ScAtNet (SSC)</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.87</td>
<td>0.77</td>
</tr>
<tr>
<td>R10</td>
<td>ScAtNet (MSC)</td>
<td><strong>0.64</strong></td>
<td><strong>0.64</strong></td>
<td><strong>0.64</strong></td>
<td><strong>0.88</strong></td>
<td><strong>0.79</strong></td>
</tr>
</tbody>
</table>

**TABLE 2:** Comparison of overall performance with state-of-the-art WSI classification methods across different metrics on the test set. Here, SSC denotes single input scale (10×). MSC denotes multiple input scales (7.5×, 10×, 12.5×). MSC* denotes multiple input scales (10×, 20×)
Experimental Result: baseline methods

![Bar chart showing accuracy for different diagnostic categories and methods.](chart.png)
Experimental Result: single vs. multiple input scales

(a) Overall performance of ScAtNet

(b) Class-wise accuracy of ScAtNet
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

<table>
<thead>
<tr>
<th>Input scales</th>
<th>Accuracy</th>
<th>F1</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5x</td>
<td>0.55</td>
<td>0.55</td>
<td>0.55</td>
<td>0.85</td>
<td>0.75</td>
</tr>
<tr>
<td>10x</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.87</td>
<td>0.77</td>
</tr>
<tr>
<td>12.5x</td>
<td>0.61</td>
<td>0.61</td>
<td>0.61</td>
<td>0.87</td>
<td>0.78</td>
</tr>
<tr>
<td>7.5x 10x</td>
<td>0.64</td>
<td>0.64</td>
<td>0.64</td>
<td>0.88</td>
<td>0.79</td>
</tr>
<tr>
<td>7.5x 12.5x</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.88</td>
<td>0.80</td>
</tr>
<tr>
<td>10x 12.5x</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.88</td>
<td>0.79</td>
</tr>
<tr>
<td>7.5x 10x 12.5x</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.88</td>
<td>0.79</td>
</tr>
</tbody>
</table>

(a) Overall performance of ScAtNet

Limitations

- Limited study on whole slide skin biopsy images (lack of public datasets)
- Limited in-house dataset size
- Mostly binary classification
  - This study covers the full spectrum of melanocytic skin biopsy
- Small test set
  - We have independent test set of 115 WSIs (50%)
Future Work

- Other types of image and cancer
- Learnable scale
- Wider range of scales
- Interpreting choice of scale, class and diagnosis accuracy
- Comparing viewing behavior with pathologists
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

Collaborators: Shima Nofallah, Dr. Sachin Mehta
Reference


Q&A