

Histopathology Whole Slide Image Analysis with Heterogeneous Graph Representation Learning

Supplementary Materials

Tsai Hor Chan^{1,*}, Fernando Julio Cendra^{1,2*}, Lan Ma², Guosheng Yin^{1,3}, Lequan Yu¹

¹Department of Statistics and Actuarial Science, The University of Hong Kong

²TCL Corporate Research Hong Kong

³Department of Mathematics, Imperial College London

{hchanth, fcendra}@connect.hku.hk, rubyma@tcl.com, guosheng.yin@imperial.ac.uk, lqyu@hku.hk

Overview. In the supplementary material, we first present the visualization results for quantitative analysis. Additionally, we provide detailed descriptions of evaluation metrics defined in the main text. We also provide more information on the adopted datasets and their class distributions. Furthermore, we describe our proposed pseudo-label pooling in more detail. Finally, we present more visualization examples of whole slide images (WSIs) and their predicted patch types.

1. Visualization with Causality-based Explanations

Figure 1 presents the visualization results of our proposed explanation approach based on Granger causality versus the existing approach based on association (GNNExplainer).

2. More Description of Evaluation Metrics

We provide more details on the evaluation metrics used in the experiments.

- Accuracy: the fraction of correct predictions to the total number of ground truth labels.
- F-1 score: The F-1 score for each class is defined as

$$\text{F-1 score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

where ‘recall’ is the fraction of correct predictions to the total number of ground truths in each class and precision is the fraction of correct predictions to the total number of predictions in each class. The F-1 score is then the weighted average for computing the final F-1 score of the classification.

- AUC: the area under the receiver operating curve (ROC) which is the plot of the true positive rate (TPR/Recall) against the false positive rate (FPR).

3. More Information of Adopted Datasets

Table 1 presents the detailed information on each of the datasets used in our experiments.

Table 1. Summary of datasets [5]

<i>Classification</i>	Tumor		Normal	
TCGA-COAD	1325		109	
TCGA-BRCA	1365		347	
<i>Staging</i>	Stage I	Stage II	Stage III	Stage IV
TCGA-COAD	226	518	365	195
TCGA-BRCA	217	780	301	30
<i>Typing</i>	Type I		Type II	
TCGA-ESCA	127		104	

4. More Details on Pseudo-label Graph Pooling

Figure 3 presents the detailed mechanism of the pseudo-label (PL) pooling. After the propagation of each layer, we obtain the updated node embeddings. Node embeddings are clustered according to the pre-defined node types instead of similarities. The node types are defined according to a pre-trained patch classifier. We adopt the HoverNet [2] classifier since it is able to detect the nuclei feature in a patch. We pool the node features in each cluster by an aggregation method (e.g., mean aggregation) and obtain a single embedding vector for each cluster. Other aggregation methods such as the attention-based approach [3] also works. We then aggregate the cluster-level embeddings again to obtain the graph-level representations.

Pooling by predefined node types can ensure the semantic consistency of the feature embeddings. For instance, node embeddings under the “neoplastic” type share more

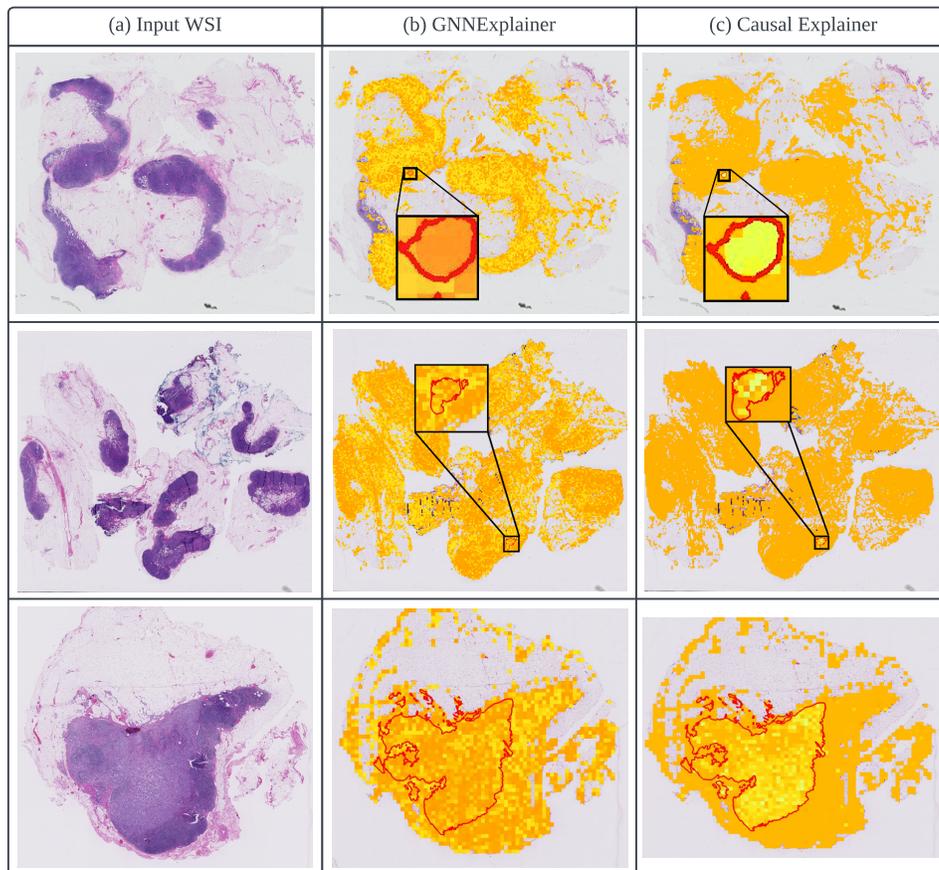


Figure 1. The input WSI (left) and the explanation heatmaps generated by GNNExplainer (middle) and our causal-driven explanation method (right). Ground truth regions are outlined with red boundaries. Lighter yellow indicates a higher importance score.

or less similar features of neoplastic cells. However, conventional graph pooling methods [1, 4] based on similarity clustering cannot address this. Clusters formed by these methods have less cosine distances, but such distances cannot represent the semantic similarities between the embeddings. Hence, the embedding distributions among graphs have inconsistent semantic meanings, which leads to ineffective information pooling via these methods.

5. More Visualization Examples

Figure 2 presents more examples of WSIs and their predicted patch types.

References

- [1] Moshe Eliasof and Eran Treister. Diffgcn: Graph convolutional networks via differential operators and algebraic multi-grid pooling. *Advances in neural information processing systems*, 33:18016–18027, 2020. 2
- [2] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58:101563, 2019. 1
- [3] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel. Gated graph sequence neural networks. *Proceedings of ICLR’16*, 2015. 1
- [4] Ekagra Ranjan, Soumya Sanyal, and Partha Talukdar. Asap: Adaptive structure aware pooling for learning hierarchical graph representations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5470–5477, 2020. 2
- [5] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013. 1

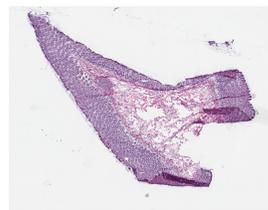
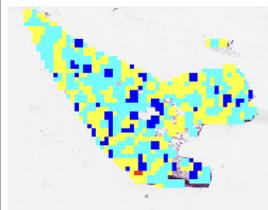
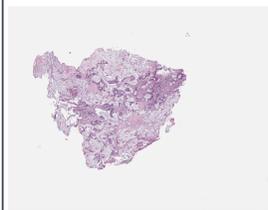
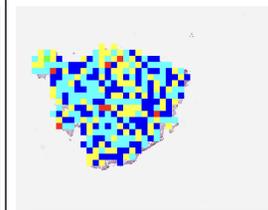
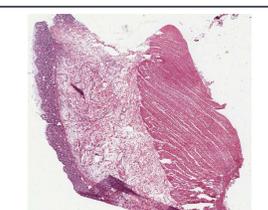
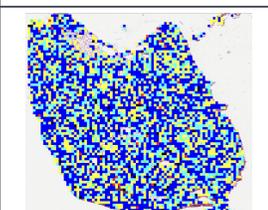
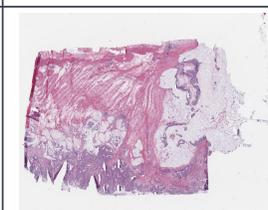
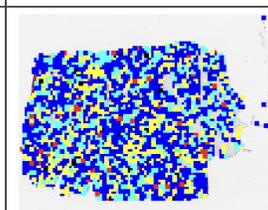
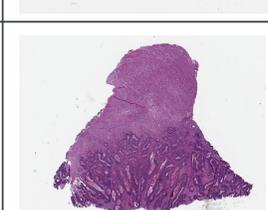
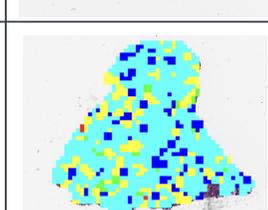
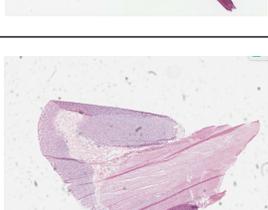
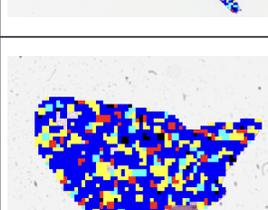
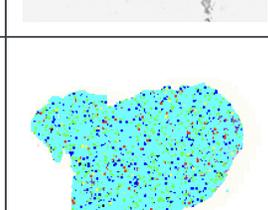
Normal WSI		Tumor WSI	
Input WSI	WSI with patch types	Input WSI	WSI with patch types
			
			
			
			

Figure 2. Additional visualization examples for each input WSI (normal and tumor) and its predicted patch types.

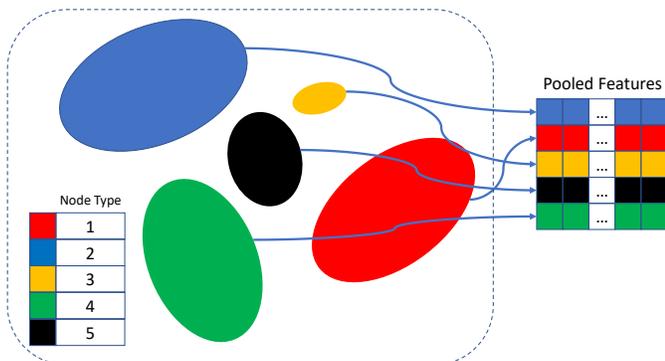


Figure 3. The mechanism of pseudo-label Pooling.