Prompting Vision Foundation Models for Pathology Image Analysis — Supplementary Materials —

Chong Yin[1], Siqi Liu[2], Kaiyang Zhou[1], Vincent Wai-Sun Wong[3], Pong C. Yuen[1]
[1]Department of Computer Science, Hong Kong Baptist University, Hong Kong
[2]Shenzhen Research Institute of Big Data, Chinese University of Hong Kong, Shenzhen
[3]Department of Medicine and Therapeutics, Chinese University of Hong Kong, Hong Kong
{chongyin, kyzhou, pcyuen}@comp.hkbu.edu.hk, siqiliu@sribd.cn, wongv@cuhk.edu.hk

Overview. In the supplementary material, we delve into the implementation details, providing a comprehensive understanding of our method. Furthermore, we present additional evaluation results on two tissue image classification tasks, showcasing the effectiveness of our approach. Additionally, we enhance the visualization by including attention maps and attribute significance histograms generated by our proposed method.

1. More implementation details.

We have set the length of prompt tokens for each layer to 1. The proposed ACPG module includes a transformer encoder and decoder, both consisting of 1 layer. This configuration ensures efficient and effective processing of the data. We use HoVer-Net [1] for nuclei segmentation. It's a user-friendly tool available at no extra cost and off-the-shelf. In pathology images stained with H&E, fatty cells or vessels are shown as white regions. We consider both the spatial arrangement of white regions and nuclei, as well as the morphological characteristics of white regions such as their area, eccentricity, and perimeters. The morphological attributes are extracted with the scikit-image library [5]. For the feature dimension, we set $d_A = 10$ and $d_{\bar{A}} = 384$.

2. More evaluation results.

As shown in Table 1, it shows the test accuracy for two image classification tasks. By incorporating quantitative attributes, the model becomes more efficient, especially in low-data regimes.

Furthermore, we add a **pairwise t-test** to compare our method's overall performance with other methods, as shown in Table 2 (Note: $p < 0.05$ means significant improvement):

3. Effects of segmentor.

We conduct experiment to compare the performance of different segmentation methods, as shown in Table 3. Enhanced segmentation improves attribution calculation and model performance.

Table 1. Test accuracy on NAFLD abnormalities classification and histological findings recognition tasks.

Methods	NAFLD-anomaly			Liver-NAS
	10% Training	50% Training	100% Training	(recognition)
Fabian et al. [2]	78.91±1.74	94.56±0.68	97.36±0.47	93.76±2.57
SAR [6]	81.06±3.39	99.10±0.35	99.54±0.19	95.12±1.01
VPT-DEEP [3]	96.57±0.61	99.31±0.23	99.49±0.26	95.50±2.05
VQT [4]	95.99±0.76	99.29±0.29	99.47±0.27	95.40±3.60
QAP (Ours)	**97.75**±0.77	**99.38**±0.23	**99.58**±0.17	**96.20**±1.67

Table 2. Statistical significance comparison.

Methods	p-value ↓		
	Task1	Task2	Task3
Fabian et al.	0.0002	0.0661	0.0007
SAR	0.0089	0.1452	0.0096
VPT-DEEP	0.0219	0.0117	0.0002
VQT	0.0225	0.0364	0.0134

Table 3. Effects of segmentor.

Methods	Macro-F1
Base	80.52±6.35
WatershedSeg	81.11±5.80
HoVer-Net	83.37±6.99

4. Other attributes choice

We conduct experiment using the mean value calculated using a different number of samples for quantitative attributes, as shown in Table 4. We can see that the performance can also benefit from the statistical value.

Table 4. Ablation study on attributes choice.

#Samples	Macro-F1
7	80.94±6.46
10	81.97±6.70
14	81.92±6.22

5. Qualitative results.

The introduction of quantitative attributes allow for a more comprehensive interpretation of images. By revealing the reasoning behind the decision-making process, these

attributes help distinguish between common characteristics and unique features in the images. The attributes are:

- a_1^s: The spatial arrangement of nuclei.
- a_2^s: The spatial arrangement of white regions.
- a_3^s: The spatial arrangement of nuclei around white regions.
- a_4^s: The spatial arrangement of white regions around nuclei.
- a_1^m: The area arrangement of white regions.
- a_2^m: The eccentricity of white regions.
- a_3^m: The perimeters of white regions.

Inflammation. When identifying inflammation, the distribution of nuclei plays a crucial role, as emphasized in both the attention map and attribute significance histogram, as shown in Figure 1. The attention map specifically highlights the model's focus on nuclei, indicating their importance in the analysis. The attribute a_1^s indicates the internuclear distance and tends to be high in all three sample cases. Additionally, the histogram of attribute significance indicates that morphological attributes $\{a_4^m, a_5^m, a_6^m a_7^m\}$ are also considered in diagnosis when a white area is observed in the image (3rd rows in Figure 1).

Ballooning. When it comes to identifying ballooning, there are distinct differences in behavior compared to inflammation. The attention map provides insights into the model's focus on nuclei and white regions, indicating their importance in identifying ballooning. Moreover, the attribute significance histogram provides additional evidence to support this observation. It emphasizes the importance of considering the spatial arrangement and morphological features of these regions during the diagnosis procedure.

Steatosis. The analysis of nuclei and white regions is crucial for accurately identifying steatosis. This is supported by the attention map and attribute significance histogram, which both highlight the importance of these features. The attention map shows that white regions and nuclei have a high attention score, indicating their relevance in the identification process. Furthermore, the analysis of liver pathology image samples reveals that certain attributes have a significant impact on the diagnosis process. In the first two rows, the spatial arrangement of white regions, referred to as a_2^s, has shown to play a crucial role and receives high scores. Similarly, for the third sample case, the displacement of white regions around nuclei, denoted as a_3^s, is also found to be crucial for accurate diagnosis. Refer to Figure 3 for a visual demonstration of this.

References

[1] 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*, page 101563, 2019. 1

[2] Fabian Heinemann, Gerald Birk, and Birgit Stierstorfer. Deep learning enables pathologist-like scoring of nash models. *Scientific reports*, 9(1):1–10, 2019. 1

[3] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *European Conference on Computer Vision (ECCV)*, 2022. 1

[4] Cheng-Hao Tu, Zheda Mai, and Wei-Lun Chao. Visual query tuning: Towards effective usage of intermediate representations for parameter and memory efficient transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7725–7735, 2023. 1

[5] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014. 1

[6] Chong Yin, Siqi Liu, Rui Shao, and Pong C Yuen. Focusing on clinically interpretable features: selective attention regularization for liver biopsy image classification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pages 153–162. Springer, 2021. 1

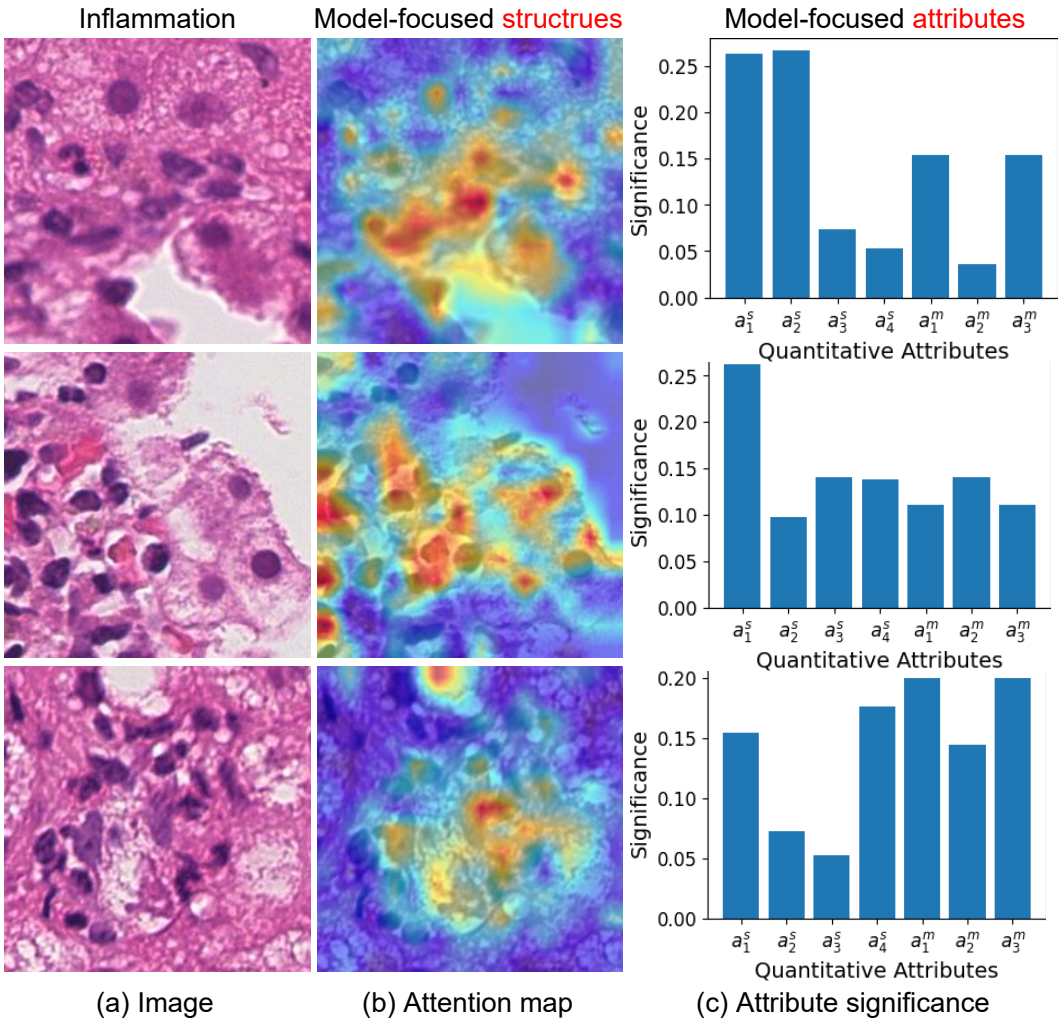

Figure 1. Image samples with its attention map and attribute significance histogram when identifying inflammation.

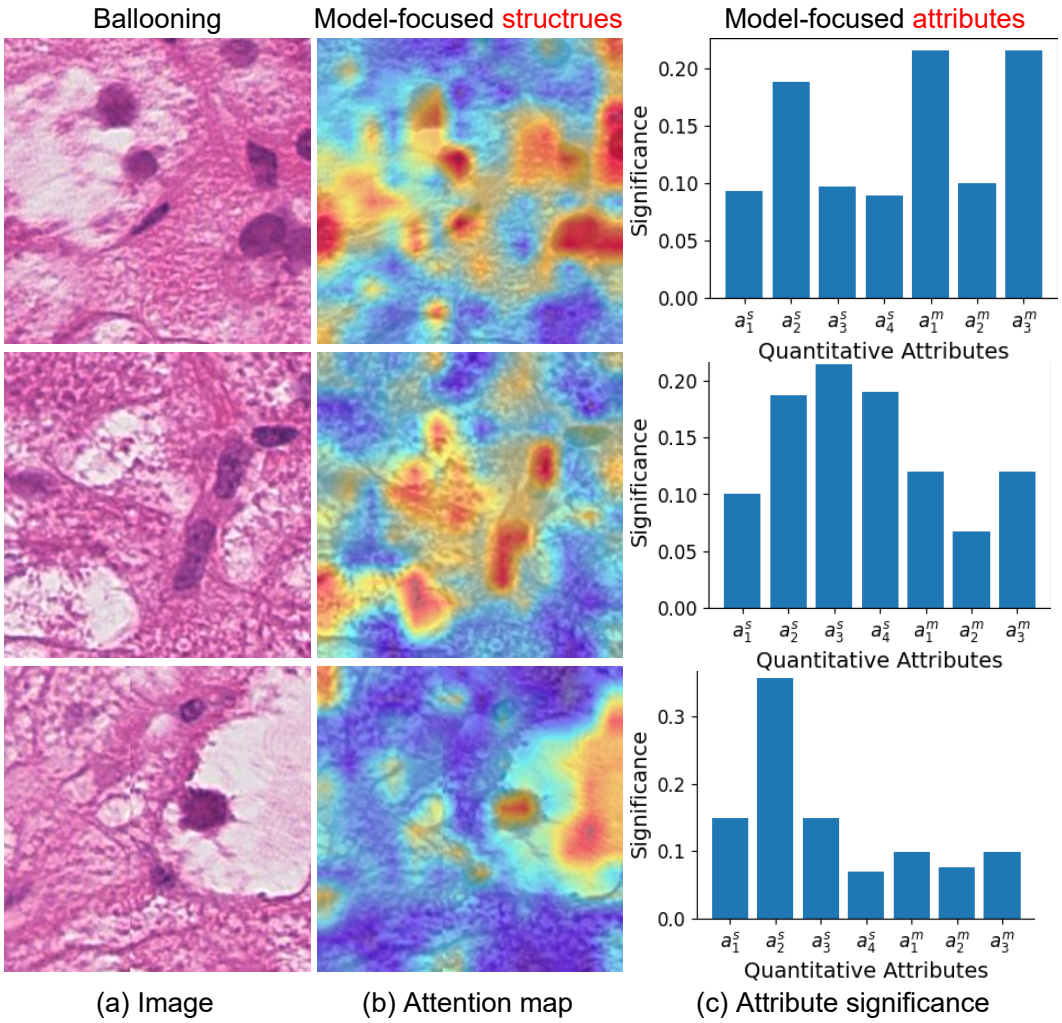

Figure 2. Image samples with its attention map and attribute significance histogram when identifying ballooning.

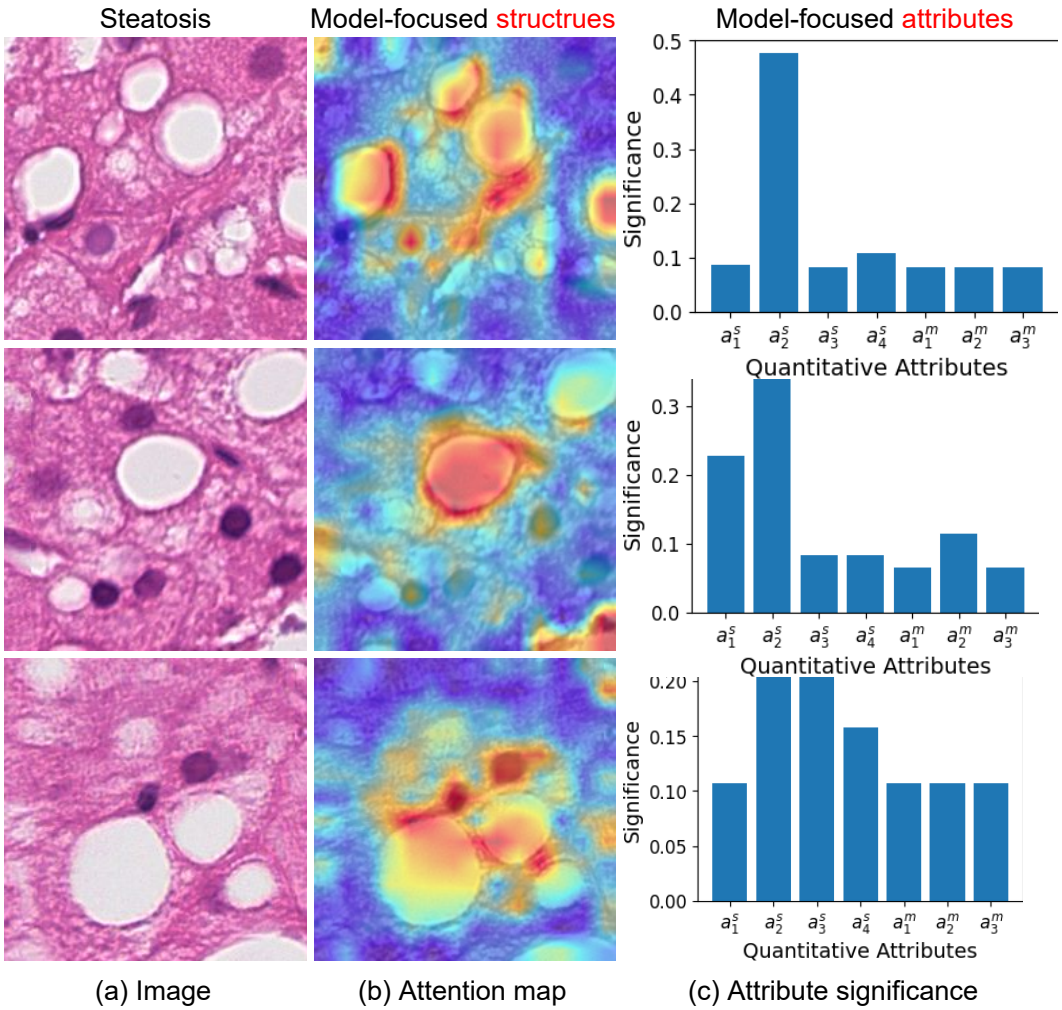

Figure 3. Image samples with its attention map and attribute significance histogram when identifying steatosis.