XFibrosis: Explicit Vessel-Fiber Modeling for Fibrosis Staging from Liver Pathology Images
— Supplementary Materials —

Chong Yin[1], Siqi Liu[2], Fei Lyu[1], Jiahao Lu[3], Sune Darkner[3], Vincent Wai-Sun Wong[4], 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 Computer Science, University of Copenhagen, Denmark
[4]Department of Medicine and Therapeutics, Chinese University of Hong Kong, Hong Kong

{chongyin, feilyu, pcyuen}@comp.hkbu.edu.hk, siqiliu@sribd.cn
{lu, darkner}@di.ku, wongv@cuhk.edu.hk

Overview. In the supplementary material, we first give a preliminaries of primal-dual graph. Additionally, we provide detailed descriptions of experimental settings including dataset, evaluation metrics, and implementation details. Furthermore, we present more experimental results to verify the effectiveness of our proposed method.

1. Primal-Dual Graph

In this section, we provide an overview of mathematical notations and associated properties for primal-dual graphs. [7].

- **Primal Graph** Let $G = (\mathcal{V}, \mathcal{E})$ be an undirected graph with node set $\mathcal{V}$ and edge set $\mathcal{E}$. The number of nodes and edges are $|\mathcal{V}|$ and $|\mathcal{E}|$, respectively. An adjacency matrix $A \in R^{|v| \times |v|}$ describes the connections between any two graph nodes in $\mathcal{V}$, where $A(i, j) = 1$ if there is a connection between the node i and node j.

- **Dual Graph** Given the primal graph G, the dual graph is denoted by $\hat{\mathcal{G}} = (\hat{\mathcal{V}}, \hat{\mathcal{E}})$, where $\hat{\mathcal{V}}$ denotes the set of dual graph nodes; $\hat{\mathcal{E}}$ denotes the set of dual graph edges. The associated adjacency matrix is $\hat{A}$. Each dual graph node $\hat{v} \in \hat{\mathcal{V}}$ corresponds to a primal graph edge $e \in \mathcal{E}$. There is a connection between dual graph nodes if the corresponding primal graph edge shares a primal graph node.

2. Experimental setting

Liver-Fibrosis-SR [21] contains 132 WSIs of liver collected from 132 patients with NAFLD. The age ranges from 30-90. These WSIs are scanned under a lens 40×. The average image resolution is $61,000 \times 20,000$ by pixel. All liver tissue sections are stained with Sirius Red (SR).

FibrosisMT-v1.0 contains 47 WSIs of the liver from two age groups. [13] collected 18 WSIs from 16 children with minor fibrosis (FIB-0 ∼ FIB-1). The average image resolution is $30,000 \times 20,000$ by pixel. The other 29 WSIs are collected by [21] from 29 adults. The average image resolution is $61,000 \times 20,000$ by pixel. These WSIs are scanned under a lens 40×. All liver tissue sections are stained with Masson's and Triise (MT).

Each liver biopsy image has been assigned a fibrosis stage (FIB 0-4) by an expert pathologist according to the METAVIR [5] standard: no fibrosis (FIB-0), fibrous portal expansion (FIB-1), few bridges or septa (FIB-2), numerous bridges or septa (FIB-3), and cirrhosis (FIB-4).

Evaluation metrics. For each dataset, we randomly split all patients into 3 groups and validate the method using 3-fold cross-validation. Following [15, 18, 19], we choose the accuracy and area under the receiver operating characteristic values (AUC) as the evaluation matrix. When calculating AUROC, the original multi-classes (FIB-0 ∼ FIB-4) classification is reduced to multiple sets of two-classes classification. It is designed to evaluate the model's ability to discriminate between two adjacent fibrosis stages. Its corresponding evaluation index are $AUROC_{FIB \leq 0 \text{vs} FIB \geq 1}$, $AUROC_{FIB \leq 1 \text{vs} FIB \geq 2}$, $AUROC_{FIB \leq 2 \text{vs} FIB \geq 3}$, $AUROC_{FIB \leq 3 \text{vs} FIB \geq 4}$. The macro-average AUROC is reported.

Data pre-processing. All variations in staining protocol introduce appearance variability. To alleviate these variances, all WSIs are stain-normalized into standard color following [9, 10]. We further normalize the image by performing channel-wise subtraction and division by its standard
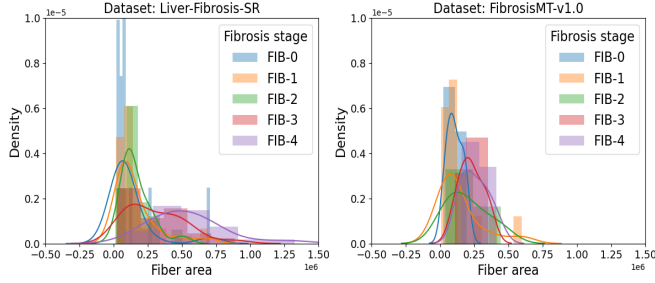
Figure 1. Distribution of detected fibers on WSIs from two datasets.

deviation. The construction of primal-dual graphs is carried out at 4X magnification which reduces the resolution of WSIs at the starting point. Since the fiber and vessels tend to be sparsely distributed in the tissue, we build the primal-dual graphs on the WSI where only the information around the nodes of the two graphs are extracted and aggregated.

Implementation details. Following [16], the algorithm 1 is utilized to implement the vessel/fiber segmentor. The primal-dual graph is built using the NetworkX package [6]. The ResNet-18 [8] is chosen as our feature encoder $\mathcal{F}$ to learn deep features. We set the radius $r = 64$ by default. To prevent over-fitting, we apply Dropout [14] ($ratio = 0.2$) during training. The scorer function S_θ is one full-connected layer followed by a softmax layer. Its input dimension is d_h, the output dimension is 1. The PDGCN is implemented using PyTorch [11] with geometric package [4]. $\mathcal{G}_p$ and $\mathcal{G}_d$ are implemented with one GATv2Conv [3] layer. $\mathcal{G}_e$ is implemented with one Meta-Conv [2] layer. We worked with $l = 1$. The classifier ($\mathcal{M}$LP) consists of three fully-connected layers with a hidden feature dimension $\{512, 128, 5\}$. The feature dimension $d_h = d_M = 512$. The dimension of position features $d_X = d_Y = 14$.

3. Experimental Results

3.1. Generated Graph

Figure 2 shows samples of vessels or fibers regions represented by graphs. It demonstrates the benefit of representing its topology as a graph structure. The contour exhibits a non-convex or circular (left to right) variation. The graph-structured representation shows how the fiber regions are connected and which way the fibers grow. The graphs align well with the shape of the fiber region. This is because each graph node is made up of points that come from the shape of a vessel or fiber segment. Each node in the graph also represents semantic information. It may be located in hepatic cells or in the background. These nodes usually belong to the contour points indicating the end of fiber extension.

Algorithm 1 Vessel/fiber segmentor

Input: The WSI I_i

Step-1: Stain color normalization: apply stain normalization operator $stainN$ [12] to reduce the color and intensity variations present in the stained image

$$I_i^{'} = stainN(I_i)$$

Step-2: Color space transformation:

$$f_i^{'} = RGB2Lab(I_i^{'})$$

Step-3: Apply Mean-Shift algorithm to cluster each pixel

$$m_i = MeanShift(f_i^{'})$$

Step-4: Apply morphological operations Erosion and Dilations $\mathcal{P}$ to the masks to refine the segmentation results.

$$m_i^{'} = \mathcal{P}(m_i)$$

Output: The segmentation mask $m_i^{'}$.

Contour points may be located in tissue or non-tissue regions. The graph nodes also consist of points derived from the centers of vessel segments, representing central veins or portal veins.

3.2. The Robustness of Vessel-Fiber Segmentor

As there are no annotations available for fibers and vessels, an unsupervised segmentation algorithm is the best choice for this task. Furthermore, due to the lack of a large public liver biopsy image dataset and the complexity of tissue structures, traditional machine learning methods still dominate image processing in this domain. To further verify the effect of the vessel/fiber segmentation module on the fibrosis staging task, we implement the vessel/fiber segmentor with three other algorithms HMRF-EM [20], GMM [17], K-Means [1], and report the fibrosis staging performance, as shown in Figure 3. The fibrosis staging can benefit from accurate vessel/fiber segmentation results. As the segmentation accuracy increases from 0.28 to 0.87, the accuracy increases from 38.99 to 53.63. Developing an advanced segmentation algorithm is not the focus of our work, we may leave it as future work.

3.3. Ablation Study On Feature Embedding

The image patch centered at graph nodes provides essential information for graph representation. The size r of image patches is important for building feature embedding. We vary the tile size $r = \{16, 32, 48, 64, 80\}$ and report the performance. As shown in Figure 4, larger r means captur-
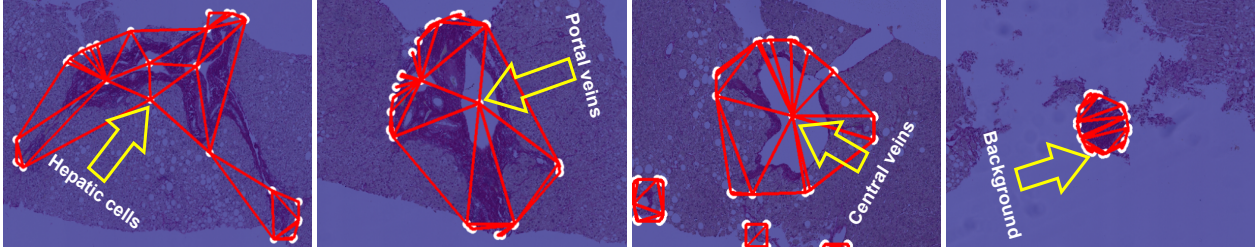
Figure 2. Samples of graph representation on liver tissues. The node in the primal graph may be located at hepatic cells, portal veins, central veins, or background.
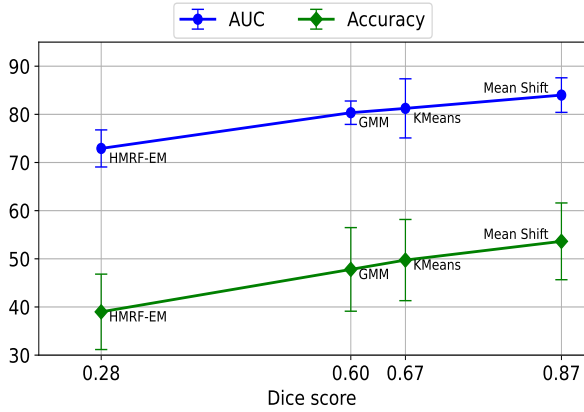

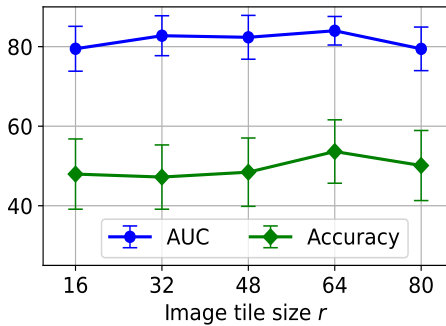Figure 3. Fibrosis staging performance under different vessel/fiber segmentation results

Figure 4. Performance with varying size of image tile on Liver-Fibrosis-SR dataset.

Table 1. Variants of aggregation methods used for generating WSI-level features.

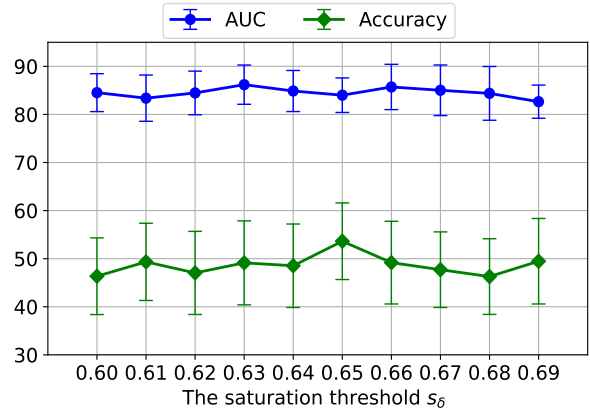| Aggregation methods	AUC	Accuracy
MaxPooling	85.01 ± 5.46	47.02 ± 8.07
MeanPooling	$\mathbf{86.52} \pm 3.78$	50.87 ± 8.12
SumPooling	86.18 ± 3.81	48.57 ± 8.00
AttentionPooling	82.76 ± 5.80	41.27 ± 7.93
TopKPooling	84.00 ± 3.59	$\mathbf{53.63} \pm 7.97$

Figure 5. Performance with varying saturation threshold s_δ.

3.4. Ablation Study on Feature Aggregation

We implement different feature aggregation methods that are commonly used in previous works on WSI-level feature generation. The feature aggregation methods includes *MaxPooling*, *MeanPooling*, *SumPooling*, *AttentionPooling*, and *TopKPooling*.

Table 1 summarizes the performance with different feature aggregation methods. TopKPooling achieves the highest accuracy of 53.63. TopKPooling enables the model to focus on important fiber regions and discard the less important regions. It works as a kind of regularization technique, preventing the model from overfitting meaningless features. As for other aggregation methods, *MaxPooling*, *MeanPooling*, *SumPooling*, *AttentionPooling*, the performance is not

ing context information from a larger field. The AUC and accuracy increase with image size r, reaching peaks of 84.0 and 53.63 at $r = 64$, respectively. A large field means providing efficient morphological information in a neighboring region to distinguish locations and topology. When we further increase the image size r, we observe the performance decrease. Increasing the patch size can result in overlapping adjacent graph nodes that contain similar information, thereby reducing the discriminability of the graph topology.

very competitive with TopKPooling. The reason behind this could be that these methods integrate information from single or all image tiles.

3.5. Investigation of Hyper-parameters

The saturation threshold s_δ is a hyper-parameter. We conduct experiments on the larger Liver-Fibrosis-SR dataset to evaluate the performance under different saturation thresholds. We vary the saturation threshold $s_\delta \in [0.60, 0.69]$. As shown in Figure 5, the accuracy increases with a higher score saturation threshold s_δ, reaching peaks of 53.63 at $s_\delta = 0.65$. A larger s_δ means that more image tiles are selected and efficient information is provided for diagnosis. When we further increase the saturation threshold s_δ, we observe the performance decreasing. Selecting too many image tiles would distract the model by introducing those with lower scores.

References

[1] David Arthur and Sergei Vassilvitskii. K-means++ the advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035, 2007. 2

[2] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Tacchetti, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018. 2

[3] Marc Brockschmidt. Gnn-film: Graph neural networks with feature-wise linear modulation. In *ICML*, pages 1144–1152. PMLR, 2020. 2

[4] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019. 2

[5] Zachary D Goodman. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. *Journal of hepatology*, 47(4):598–607, 2007. 1

[6] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008. 2

[7] Frank Harary and Robert Z Norman. Some properties of line digraphs. *Rendiconti del circolo matematico di palermo*, 9(2):161–168, 1960. 1

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 2

[9] Guillaume Jaume, Pushpak Pati, Valentin Anklin, Antonio Foncubierta, and Maria Gabrani. Histocartography: A toolkit for graph analytics in digital pathology. In *MICCAI Workshop on Computational Pathology*, pages 117–128. PMLR, 2021. 1

[10] Marc Macenko, Marc Niethammer, James S Marron, David Borland, Woosley, et al. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE international symposium on biomedical imaging: from nano to macro*, pages 1107–1110. IEEE, 2009. 1

[11] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 2

[12] Pushpak Pati, Guillaume Jaume, Antonio Foncubierta-Rodriguez, Florinda Feroce, Anna Maria Anniciello, Giosue Scognamiglio, Nadia Brancati, Maryse Fiche, Estelle Dubruc, Daniel Riccio, et al. Hierarchical graph representations in digital pathology. *Medical image analysis*, 75:102264, 2022. 2

[13] Mousumi Roy, Fusheng Wang, Hoang Vo, Dejun Teng, George Teodoro, Alton B Farris, Eduardo Castillo-Leon, Miriam B Vos, and Jun Kong. Deep-learning-based accurate hepatic steatosis quantification for histological assessment of liver biopsies. *Laboratory Investigation*, 100(10):1367–1383, 2020. 1

[14] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 2

[15] Marta Wojciechowska, Stefano Malacrino, Natalia Garcia Martin, Hamid Fehri, and Jens Rittscher. Early detection of liver fibrosis using graph convolutional networks. In *MICCAI*, pages 217–226. Springer, 2021. 1

[16] Geming Wu, Xinyan Zhao, Shuqian Luo, and Hongli Shi. Histological image segmentation using fast mean shift clustering method. *Biomedical engineering online*, 14(1):1–12, 2015. 2

[17] Lei Xu and Michael I Jordan. On convergence properties of the em algorithm for gaussian mixtures. *Neural computation*, 8(1):129–151, 1996. 2

[18] Koichiro Yasaka, Hiroyuki Akai, Akira Kunimatsu, Osamu Abe, and Shigeru Kiryu. Deep learning for staging liver fibrosis on ct: a pilot study. *European radiology*, 28:4578–4585, 2018. 1

[19] Koichiro Yasaka, Hiroyuki Akai, Akira Kunimatsu, Osamu Abe, and Shigeru Kiryu. Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid–enhanced hepatobiliary phase mr images. *Radiology*, 287(1):146–155, 2018. 1

[20] Yongyue Zhang, Michael Brady, and Stephen Smith. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20(1):45–57, 2001. 2

[21] Yu-Jie Zhou, Feng Gao, Wen-Yue Liu, Grace Lai-Hung Wong, Sanjiv Mahadeva, Nik Raihan Nik Mustapha, Xiao-Dong Wang, Wah-Kheong Chan, Vincent Wai-Sun Wong, and Ming-Hua Zheng. Screening for compensated advanced chronic liver disease using refined baveno vi elastography cutoffs in asian patients with nonalcoholic fatty liver disease. *Alimentary pharmacology & therapeutics*, 54(4):470–480, 2021. 1