Diagnostic Regions Attention Network (DRA-Net) for Histopathology WSI Recommendation and Retrieval

Yushan Zheng, Zhiguo Jiang, Fengying Xie, Jun Shi, Haopeng Zhang,
Jianguo Huai, Ming Cao, and Xiaomiao Yang

Abstract

The development of whole slide imaging techniques and online digital pathology platforms have accelerated the popularization of telepathology for remote tumor diagnoses. During a diagnosis, the behavior information of the pathologist can be recorded by the platform and then archived with the digital case. The browsing path of the pathologist on the WSI is one of the valuable information in the digital database because the image content within the path is expected to be highly correlated with the diagnosis report of the pathologist. In this paper, we proposed a novel approach for computer-assisted cancer diagnosis named session-based histopathology image recommendation (SHIR) based on the browsing paths on WSIs. To achieve the SHIR, we developed a novel diagnostic regions attention network (DRA-Net) to learn the pathology knowledge from the image content associated with the browsing paths. The DRA-Net does not rely on the pixel-level or region-level annotations of pathologists. All the data for training can be automatically collected by the digital pathology platform without interrupting the pathologists' diagnoses. The proposed approaches were evaluated on a gastric dataset containing 983 cases within 5 categories of gastric lesions. The quantitative and qualitative assessments on the dataset have demonstrated the proposed SHIR framework with the novel DRA-Net is effective in recommending diagnostically relevant cases for auxiliary diagnosis. The MRR and MAP for the recommendation are respectively 0.816 and 0.836 on the gastric dataset. The source code of the DRA-Net is available at https://github.com/zhengyushan/dpathnet.

Index Terms

Digital pathology, Recommendation, Gastric cancer, GCN, RNN

I. Introduction

With the development of whole slide imaging and digital pathology, the biopsy sections are well archived, and the frameworks for histopathology whole slide image (WSI) analysis are widely developed [1], [2], [3], [4]. Computer-aided diagnosis (CAD) methods based on histopathology WSIs and artificial intelligent algorithms, especially deep learning techniques [5], [6], [7], have become popular in the last decade. There are two remarkable interests in recent studies on histopathological image analysis (HIA). The one is to develop weak-supervision [8], [9], [10], semi-supervision [11] frameworks, etc., to relieve the annotation workload of the pathologists [12]. Another one is to utilize the resource of large-scale digital pathology platform to improve the information [13], [14], [15], [16] of CAD.

With the increasing application of the telepathology system, abundant diagnosed cases have been accumulated [17], [18], [19]. The cases contain not only the WSIs but also valuable data, including the diagnosis report, meta information, user behavior data, etc. These data are the potential to develop CAD applications that are both light-annotated and informative.

A notable record in the telepathology platform is the browsing path on the WSI during the diagnosis of the pathologist. Theoretically, the pathologist should have reviewed the conclusive regions related to a specific disease before making the diagnosis. Gecer et al. [20] has proven the regions the pathologist focus on during the diagnosis are highly correlated with the diagnosis of the case. And this correlation stably exists within different experts although the browsing habits of the experts are variant [21]. This characteristic motivated us to build a deep learning model to learn the relationship between image content under the diagnosis path and the label of the WSI, and evaluate whether this level of supervision is sufficient to learn pathology knowledge and build computer-assisted cancer diagnosis system.

In this paper, we propose a novel diagnostic regions attention network (DRA-Net) for histopathology image analysis. Correspondingly, we developed a novel CAD approach named session-based histopathology image recommendation (SHIR)

This work was partly supported by the National Natural Science Foundation of China (Grant No. 61901018, 61771031, 61906058 and 61471016), partly supported by the China Postdoctoral Science Foundation (Grant No. 2019M650446), partly supported by the Anhui Provincial Natural Science Foundation (Grant No. 1908085MF210), partly supported by the Fundamental Research Funds for the Central Universities of China (Grant No. JZ2020YYPY0093), and partly supported by the 111 Project (Project B13003). *(Corresponding author: Zhiguo Jiang)*

Y. Zheng is with Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing, 100191, China. (e-mail: yszheng@buaa.edu.cn).

Z. Jiang, H. Zhang, F. Xie are with Image Processing Center, School of Astronautics, Beihang University, Beijing, 102206, China, and also with Beijing Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing, 100191, China. (e-mail: jiangzg@buaa.edu.cn).

J. Shi is with School of Software, Hefei University of Technology, Hefei 230601, China.

J. Huai, M. Cao, and X. Yang are with the department of pathology, the No.1 people's hospital of Wuhu, China. We also thank Yan Jiang, Yanyan Zhu, Man Xu, and Chengxiang Shen for the medical supports to this research.

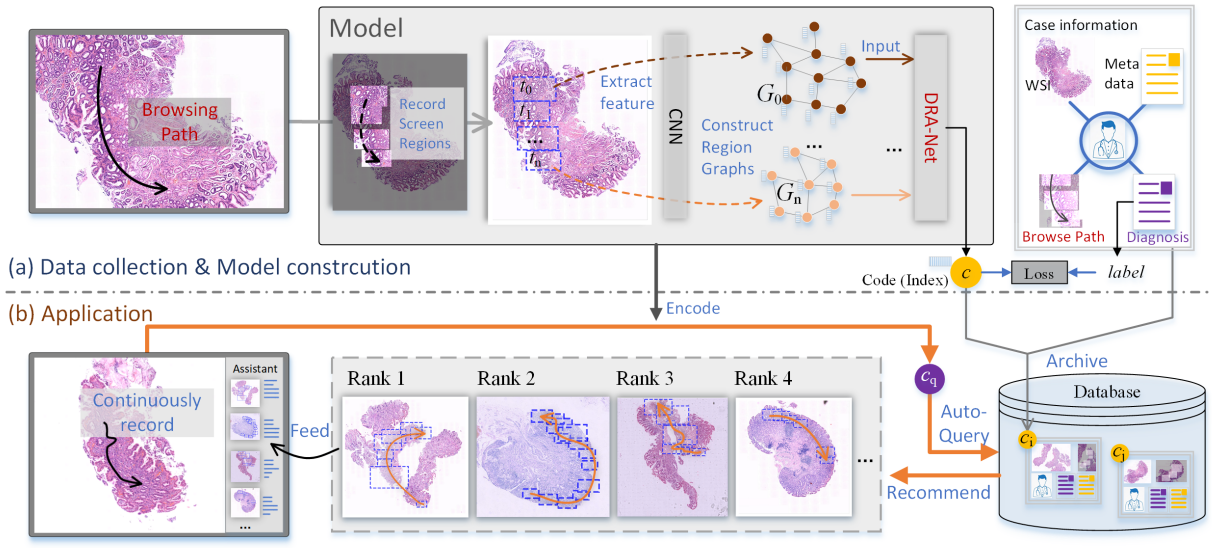

Fig. 1. The framework of the proposed method. (a) illustrates the flowchart of the model construction and database collection, where the ROIs in the diagnosis path are first numbered, then encoded based on the proposed DRA-Net, and the code is used to index the case archived in the recommendation database. (b) illustrates an application instance, where the diagnosis path is encoded using the trained model and then the relevant cases are recommended based on the similarity measurement.

based on the DRA-Net. As shown in Figure 1, the SHIR is designed to monitor the pathologist's browsing path on the WSI during diagnosis and actively query the database to recommend diagnosed cases within a similar path and image content. These cases then feedback to the pathologist to provide assistant information.

The contribution of this paper to this problem is three-fold.

1) We build a novel learning framework for histopathology image analysis based on the diagnosis paths of pathologists on WSIs and propose a novel DRA-Net to learn pathology knowledge from the image content within the diagnosis paths. The training of DRA-Net does not rely on the pixel-level or region-level annotations, for which the workload of manual annotation can be relieved. Moreover, the browsing path can be automatically collected by the platform without interrupting the pathologist's diagnosis. It determines the network is adequate to develop automatic self-training CAD systems based on digital pathology platforms.

2) We develop the session-based histopathology image recommendation (SHIR) application inspired by the concept of session-based recommendation [22], [23]. In the domain of histopathological image analysis, the most related existing application to the proposed SHIR is content-based histopathological image retrieval (CBHIR) [13], [24], [25]. CBHIR applications require the pathologists manually selecting a region of interest (ROI) as the query instance and the retrieval does not consider the regions the pathologists have already viewed. In comparison, the proposed SHIR summarizes the information of a WSI while the pathologist is browsing the WSI, and continuously and actively recommends the relevant cases within similar image content from the database. The application form is more informative and convenient than the CBHIR.

3) We have conducted comprehensive experiments to verify the proposed method and compared it with related methods [26], [27], [28] on a large-scale gastric dataset containing 983 cases. The experimental results have demonstrated the supervision of the diagnosis path on the WSI is sufficient to train a qualified model for gastric histopathology image recognition, and the proposed SHIR is promising to develop systems for computer-aided cancer diagnosis.

A part of the paper has been presented in the conference MICCAI 2020 [29]. In this paper, we optimize the recommendation network with the attention mechanism, presents new findings with extensive experimental results (including the performance for the continuous recommendation, the real-time capability, the training efficiency, the usage of the RNN module, the application for ROI retrieval, ablation study, etc.), provides the details about the methodology and data collection, and gives necessary discussions.

The remainder of this paper is organized as follows. Section II reviews the related works. Section III introduces the methodology of the proposed method. The experiment are presented in Section IV. Section V includes necessary discussions. Section VI summarizes the contributions.

II. Related Works

In this section, we first review the latest development of CBHIR methods, then introduce the methods for session-based recommendation.

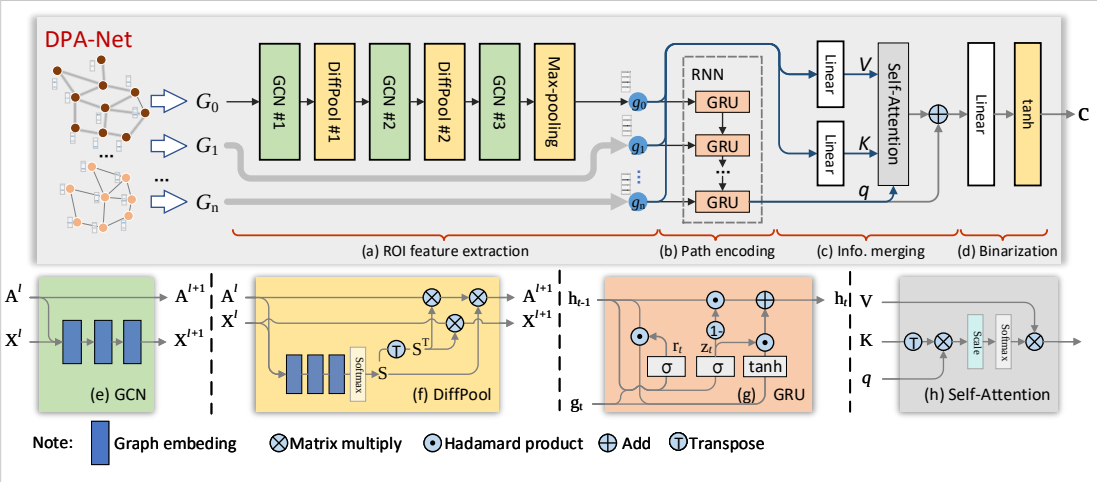

Fig. 2. The structure of the proposed *Diagnostic Regions Attention Network* (DRA-Net), where (a) is the GCN modules for ROI feature extraction, (b) is the RNN where the GRU module is shared for the graph features $\mathbf{g}_t$, (c) is the self-attention module, (d) is the hashing module, and (e-h) are the detailed network structure of these modules, and the notes are interpreted at the bottom of the figure.

A. Content-based histopathology image retrieval

In the last five years, an increasing number of studies focused on the efficiency and scalability of the histopathology image retrieval system.

To improve the efficiency of retrieval, the *Hash* techniques are developed for databases consisting of massive histopathological images. Typically, Zhang et al. [30] and Jiang et al. [31] introduced supervised hashing with kernels (KSH) into the CBHIR. Then, Shi et al. [32] utilized a graph hashing model to learn the similarity relationship of histopathological images. More recently, the end-to-end hashing networks are widely developed [14], [24], [25], [33], which have significantly improved the overall performance of the retrieval. The efficiency is also crucial for the proposed SHIR framework. Therefore, in this paper, we considered using the hashing technique to improve the speed of the SHIR framework.

Another research interest is the retrieval scalability involving the adaption of size and shape variances of query regions and the strategy to indexing whole slide images. In the previous study, Ma et al. [34] proposed dividing the WSIs into sub-regions following the sliding window paradigm and encoding the individual regions to establish the retrieval database. It was a convenient strategy to index WSIs for sub-regions retrieval. However, the tissue appearance was ignored in the division of WSIs. Zheng et al. [15] proposed segmenting a WSI into super-pixels and defining the super-pixels as retrieval instances. Furthermore, The research [35] proposed merging the super-pixels into irregular regions based on selective search [36] to achieve the indexing of WSIs. Meanwhile, the algorithms for measuring the similarity between irregular regions, or even between WSIs were designed [37], [26]. Most recently, Zheng et al. [27] proposed to construct spatial graphs to represent the sub-regions of the WSI and established an end-to-end network based on graph convolution networks (GCNs) to extract the graph features and index the sub-regions. Chen et al. [38] proposed to represent the annotation regions by fusing patch-level features and encoding the region representation by supervised hashing for retrieval. These studies have provided feasible strategies to encode the screen regions within the diagnosis path, which is the basis of the proposed SHIR framework.

B. Session-based recomendation

Recommendation is an important task in online services (e.g., e-commerce, media streaming). The recommendation system can trace the browsing history of a customer and feed the relevant items back to the customer for reference. Session-based recommendation aims to predict the intention of the user based on the user's current behaviors, rather than the historical actions. As the browsing history is a type of sequential data, the Recurrent Neural Networks (RNNs) were thereby applied in the recommendation task [39], [40], [41] and have proven important in the session-based systems [23]. Then, the attention mechanism was widely used in the domain to merge the global and local features within the session [41], [22], for which the recommendation performance was significantly improved. Recently, the graph neural networks (GNNs) were applied to the session-based recommendation by regarding the click connections as a directed graph [23], [42], which have further improved recommendation accuracy. In our study, we regarded the diagnosis on a WSI as a session and traced the screens within the browsing paths for diagnostically relevant region recommendation.

III. Method

A. Overview

DRA-Net is the body in the recommendation framework, which is detailed in this section. As shown in Figure 2, the DRA-Net consists of multiple GCN modules, an RNN module, and a self-attention module. The GCNs in the DRA-Net are used

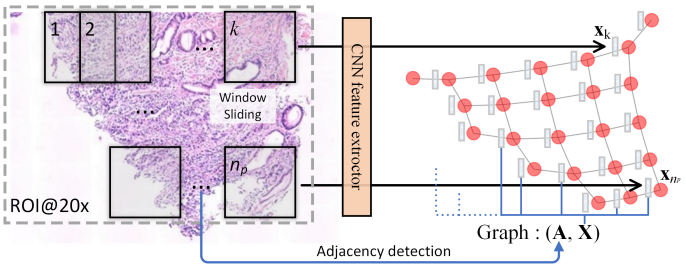

Fig. 3. The flowchart of ROI graph construction, where the patches in the tissue area are extracted with a sliding window and fed into a CNN to obtain the patch features, meanwhile the 4-neighborhood adjacency matrix for these patches are extracted, and the graph is defined based on the patch features and the adjacency matrix.

to extract structural feature for each screen region in a path, the RNN is applied to summarizing the features of the regions, the self-attention module is used to merge the local patterns and the global representation, and finally, the hashing module generates the binary code to indexing the path.

B. ROI feature extraction with GCN

The screen regions in the diagnosis path are referred as ROIs in this section. The ROIs usually contain blank (background) areas that should be ignored in the feature extraction and the size of the ROIs varies to the magnification the pathologists reviewed the slide. Thus, we propose constructing a spatial graph to represent the tissue area in the ROI and then extracting the ROI feature using GCNs [43], [44], [45].

The flowchart of constructing the graph for an ROI is shown in Figure 3. First, the ROI is divided into patches using a sliding window and the patches are fed into a CNN to extract patch features. Letting $\mathbf{x}_k \in \mathbb{R}^{d_f}$ denote the CNN feature of the k-th patch with d_f dimension, the features for the patches in the ROI are represented as a matrix $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_{n_p}]^{\mathrm{T}} \in \mathbb{R}^{n_p \times d_f}$, where n_p is the number of patches involving tissue area. Meanwhile, the 4-neighborhood adjacency of these patches are detected and formulated as an adjacency matrix $\mathbf{A} \in \{0,1\}^{n_p \times n_p}$, where $a_{ij} = 1$ represents the i-th and j-th patch are adjacent and $a_{ij} = 0$, otherwise. Then, the graph for the ROI is defined as $G = (\mathbf{A}, \mathbf{X})$ by regarding the patches as graph vertexes and the graph features as the vertex attributes. Note that the blank (background) patches are filtered by a threshold on the mean intensity of the image and not included in the graph.

The construction of the GCNs refers to Zheng et al.[27]. The network structure is illustrated as Figure 2(a). It includes three GCN modules and two Diffpool [46] modules. The GCN module is used to embed the contextual information of adjacent patches and the Diffpool module is to cluster patches in feature space while maintaining the spatial structure of the ROI. Specifically, the k-th step of graph embedding in a GCN module is defined referring to [44], which is formulated as

$$\mathbf{H}^{(k)} = ReLU(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(k-1)}\mathbf{W}^{(k)}), \tag{1}$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$ with $\mathbf{E}$ denoting the unit matrix, $\tilde{\mathbf{D}} = diag(\sum_j \tilde{\mathbf{A}}_{1j}, \sum_j \tilde{\mathbf{A}}_{2j}, ..., \sum_j \tilde{\mathbf{A}}_{n_p j})$ is the degree of $\tilde{\mathbf{A}}$, and $\mathbf{W}^{(k)}$ is a trainable weight matrix for the k-th step of embedding. $\mathbf{H}^{(k)}$ denotes the node representations after the k-th step of embedding and in specific $\mathbf{H}^{(0)} = \mathbf{X}$. Letting $(\mathbf{A}^{(l)}, \mathbf{X}^{(l)})$ be the graph state before the l-th GCN module in the DRA-Net, the inference of the l-th GCN with K steps of embedding is abbreviated as

$$\mathbf{X}^{(l+1)} = \mathbf{H}^{(K)} = \mathcal{F}^{(l)}_{gcn}(\mathbf{A}^{(l)}, \mathbf{X}^{(l)}). \tag{2}$$

The Diffpool module is also defined based on GCN [46], for which the computation flowchart is illustrated in Figure 2(f). The pooling is achieved by a matrix $\mathbf{S}^{(l)} \in \mathbb{R}_+^{n_l \times n_{l+1}}$ with the constraint $n_{l+1} < n_l$. $\mathbf{S}^{(l)}$ is generated by a GCN based on the current graph state. Specifically,

$$\mathbf{S}^{(l)} = Softmax_r(\mathcal{F}^{(l)}_{pool}(\mathbf{A}^{(l)}, \mathbf{X}^{(l)})) \tag{3}$$

with $Softmax_r(\cdot)$ denoting the row-wise softmax function and $\mathcal{F}^{(l)}_{pool}$ denoting a GCN sharing the same definition of 2. Then, the graph state for the next module is obtained by the pooling functions

$$\begin{aligned} \mathbf{X}^{(l+1)} &= \mathbf{S}^{(l)\mathrm{T}}\mathbf{X}^{(l)}, \\ \mathbf{A}^{(l+1)} &= \mathbf{S}^{(l)\mathrm{T}}\mathbf{A}^{(l)}\mathbf{S}^{(l)}. \end{aligned} \tag{4}$$

Note that the adjacency matrix for the first GCN, i.e. $\mathbf{A}^{(0)}$, is the 4-neighborhood adjacency matrix obtained during the graph construction. Finally, a max-pooling layer is used to quantify the representations of the last GCN module, which is represented as

$$\mathbf{g} = Maxpool_c(\mathcal{F}^{(l)}_{gcn}(\mathbf{A}^{(l)}, \mathbf{X}^{(l)})) \tag{5}$$

with $Maxpool_c(\cdot)$ denoting the column-wise max-pooling operation. The output $\mathbf{g} \in \mathbb{R}^{d_r}$ is regarded as the feature for the ROI. Note that the underlying graph structure for each ROI, which is formulated by $\mathbf{A}(l)$, is fixed in each GCN stage, and the underlying structure is changed by the DiffPool module while the node pooling (Eq. 4) is processed.

C. Diagnosis path representing with RNN

The ROI features for the diagnosis path are represented as $\mathbf{P} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_{n^r}]^{\mathrm{T}}$, where n^r denotes the number of ROIs in the path. In the context of our work, the ROI-level labels are not available. Therefore, we cannot train the ROI feature extraction network independently by an explicit label for each $\mathbf{g}_t$. In this case, we need to establish the connection between the ROI features and the path-level label and then train the network by the path-level labels. Further considering that the number of ROIs in different path varies, we determined to build an RNN module to summarize the ROI features and regarded the output of the last recurrence as the representation of the path. Specifically, we constructed an RNN structure based on the ROI features to obtain the path-level representation. In this paper, the Gated Recurrent Unit (GRU) [47] are employed to build the RNN. Regarding $\mathbf{g}_t$ as the feature for time t, the structure of the GRU is formulated as

$$\begin{aligned}
\mathbf{z}_t &= \sigma(\mathbf{W}_z \mathbf{g}_t + \mathbf{U}_z \mathbf{h}_{t-1}), \\
\mathbf{r}_t &= \sigma(\mathbf{W}_r \mathbf{g}_t + \mathbf{U}_r \mathbf{h}_{t-1}), \\
\tilde{\mathbf{h}} &= \tanh(\mathbf{W} \mathbf{g}_t + \mathbf{U}(\mathbf{r}_t \odot \mathbf{h}_{t-1})), \\
\mathbf{h}_t &= (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}},
\end{aligned} \tag{6}$$

where $\mathbf{z}_t$ and $\mathbf{r}_t$ serve as the update gate and the reset gate, respectively, the notations involving $\mathbf{W}$ and $\mathbf{U}$ are trainable parameters, the $\odot$ represents the Hadamard product, and $\mathbf{h}_t \in \mathbb{R}^{d_r}$ is the activation at time t ($\mathbf{h}_0 = \mathbf{0}$).

D. Representative regions mining with self-attention

In the proposed network, the output of the final recurrence of the RNN ($\mathbf{h}_{n^r}$) is regarded as the representation of the path in the length of n^r. $\mathbf{h}_{n^r}$ is obtained by recurrently computing the activation from the first ROI to the last ROI. The information of beginning ROIs need traverse the network in both the forward and backward computation to contact with the final output, and the long-range traverse may affect the ability of the network[48]. To alleviate the problem, we proposed building a self-attention (SA) module referring to [48] to provide an equally short connection between each ROI feature $\mathbf{g}_t$ and the final path representation. It enables the network to highlight the representative ROIs during the encoding. The formulation of the SA module is

$$\tilde{\mathbf{z}} = Softmax(\frac{\mathbf{q}\mathbf{K}^{\mathrm{T}}}{\sqrt{d_r}})\mathbf{V}, \tag{7}$$

where $\mathbf{K} = \mathbf{P}\mathbf{W}_{key}$ and $\mathbf{V} = \mathbf{P}\mathbf{W}_{val}$ are the Key and $Value$ input for the SA module [48], respectively, $\mathbf{W}_{key}, \mathbf{W}_{val} \in \mathbb{R}^{d_r \times d_r}$ are the weights for linear projection and $\mathbf{q} = \mathbf{h}_{n^r}$ is regarded as the *Query* input. The output $\tilde{\mathbf{z}}$ is the softmax-weighted sum of the local patterns. Afterward, a residue layer $\mathbf{z} = \tilde{\mathbf{z}} + \mathbf{h}_{n^r}$ is applied to merge the local patterns into the path representation.

Here, we define the attention score

$$\mathbf{a} = Softmax(\frac{\mathbf{q}\mathbf{K}^{\mathrm{T}}}{\sqrt{d_r}}), \tag{8}$$

which indicates the contribution of the local regions to the global decision. The score will be used to enhance the visual performance of the recommendation.

E. Recommendation with hashing

SHIR aims to feed back the diagnostically relevant cases based on the WSI regions the pathologists have browsed, rather than to predict the next item the medical doctors would be interested in. In this context, we proposed to achieve the recommendation by actively retrieving the most similar cases from the database. The retrieval is completed by hashing search with binary codes to ensure the high efficiency of the application. Specifically, a hashing layer is connected to the residue layer to converting the path representation $\mathbf{z}$ into a binary code. The hash function is defined as

$$\mathbf{y} = \tanh(\mathbf{W}^h \mathbf{z} + \mathbf{b}^h), \tag{9}$$

where the $\mathbf{W}^h$ and $\mathbf{b}^h$ are the weight and bias for the hash function and $\mathbf{y} \in \mathbb{R}^{d_h}$ is the binary-like code which can be converted into binary code by the sign function $\mathbf{c} = sign(\mathbf{y})$.

Letting $\mathbf{c}_q$ be the code of the path the pathologists has browsed and $\mathbf{c}_j$ be the j-th code in the database, the similarity of the two codes is measured by the inner product

$$\theta_{\mathbf{c}_q \mathbf{c}_j} = \frac{1}{2} \mathbf{c}_q^{\mathrm{T}} \mathbf{c}_j, \tag{10}$$

where $\theta_{\mathbf{c}_q\mathbf{c}_j} \in \{i|i = -d_h/2, ..., d_h/2\}$. The larger the $\theta_{\mathbf{c}_q\mathbf{c}_j}$, the similar the codes $\mathbf{c}_q$ and $\mathbf{c}_j$. By calculating the similarity between the current path code and those in the database, the top-similar cases are obtained and then recommended to the pathologist.

F. Training the DRA-Net

The explicit label for each ROI in the path is unavailable in the practical application of the SHIR. Therefore, we train the DRA-Net by only using the path-level labels. Specifically, the loss function is built based on the negative log triplet label likelihood [49], which is formulated as

$$\begin{aligned} L = &- \frac{1}{M}\sum_{m=1}^{M} \log \sigma(\frac{1}{2}\mathbf{y}_{a_m}^{\mathrm{T}}\mathbf{y}_{p_m} - \frac{1}{2}\mathbf{y}_{a_m}^{\mathrm{T}}\mathbf{y}_{n_m} - \alpha) \\ &+ \lambda\frac{1}{M}\sum_{m=1}^{M}\sum_{k\in\{a_m,p_m,n_m\}} \|\mathbf{y}_k - \mathbf{c}_k\|_2^2, \end{aligned} \tag{11}$$

where $(\mathbf{y}_{a_m}, \mathbf{y}_{p_m}, \mathbf{y}_{n_m})$ denotes the codes of (anchor, positive, negative) for the m-th triplet, M is the number of training paths, α is defined as the margin in the triplet loss, $\sigma(\cdot)$ is the sigmoid function to generate the probability, and $\mathbf{c}_k = sign(\mathbf{y}_k)$.

All the modules of the DRA-Net, including the GCNs, the GRU, the self-attention, and the hashing module, were trained end-to-end by backward propagation. The trainable parameters in the network were initialized using the uniform distribution following [50]. The gradient optimization algorithm was mini-batch Stochastic Gradient Descent (SGD) with the momentum. The mini-batch data for each step of training was generated by weighted sampling to ensure a balanced distribution of category labels. Moreover, the batch-hard strategy was used to generate triplets. Specifically, the triplet for each sample was dynamically constructed in the mini-batch, where the sample itself was set as the *anchor*, the intra-class (i.e., relevant to the anchor) sample within the batch that has the farthest distance to the anchor was assigned as the *positive* sample and the inter-class (i.e., irrelevant to the anchor) sample that has the nearest distance to the anchor was regarded as the *negative* sample. As a result, the number of triplets for each step of optimization was the same as the size of the mini-batch.

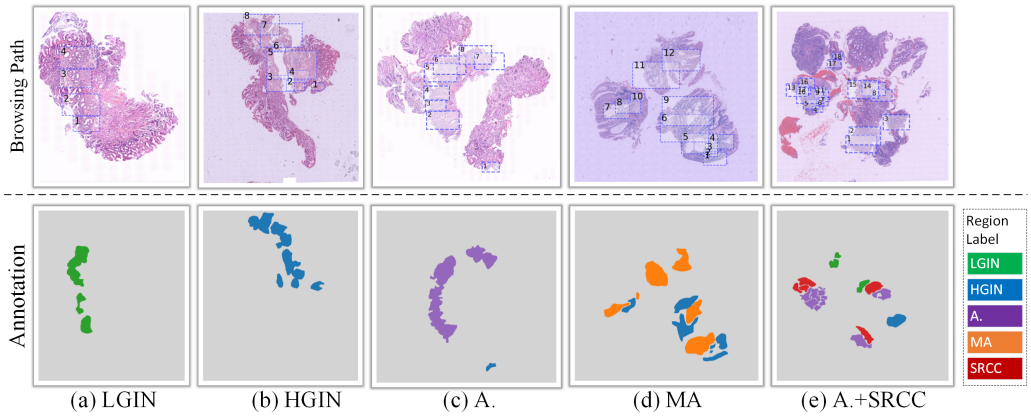

Fig. 4. Instances from the gastric database, where first row displays the screen regions recorded by telepathology platform, the second row provides the annotations with the notes on the right, and the path labels are provided under the annotations.

IV. Experiment

A. Dataset

To study the feasibility and effectiveness of the SHIR proposal, we collected a gastric dataset containing 983 gastric cases and including 5 category of gastric lesions, namely *Low-grade intraepithelial neoplasia* (LGIN), *High-grade intraepithelial neoplasia* (HGIN), *Adenocarcinoma* (A.), *Mucinous adenocarcinoma* (MA), and *Signet-ring cell carcinoma* (SRCC). One conclusive WSI was selected from each case to build the WSI dataset. We invited pathologists to make diagnoses on the WSIs using the digital pathology platform[1]. During the diagnoses, the browsing paths of the pathologists were recorded by the platform and then supplied for this research. Specifically, the screens *focused* on by the pathologists under lenses from $10\times$ (0.96μm/pixel) to $80\times$ (0.12μm/pixel)[2] were regarded as ROIs and recorded to the sequential path data. A screen is recognized as *focused* when the rendering is completed for display after all the image data within the screen are downloaded from the

[1]https://gallery.motic.com

[2]The upper bound of a WSI being browsed in the platform.

cloud. This process costs around 0.5 to 2 second, depending on the performance of the computer and the internet condition. The length of the paths ranges from 1 to 61 with an average number of 10.20.

To assess the related fully supervised methods, we invited the pathologists to annotate the exact lesion areas associated with the paths. Then, the paths were labeled following the priority A.=MA=SRCC>HGIN>LGIN according to the annotated regions under the path. Particularly, A path was assigned multiple labels if and only it contained more than one malignant tumors (A., MA, and SRCC), and otherwise, it was assigned a single label. As a result, the total number of area annotations is 9916, and 34 of the total 983 paths have two labels (the percentage is 3.45%). There are no paths with more than two labels. Several instances from the dataset are presented in Figure 4. Note that the ROI labels and the pixel-level annotations were only used to do comparison experiments. The ROI labels or the pixel-level annotations were not used in the training of the DRA-Net. Unless otherwise noted, all the models involving the notation *DRA-Net* in the experiments were trained only under the path-level supervision.

B. Experimental settings

The CNN structure for patch feature extraction was the EfficientNet [51] for its good performance in the image classification task. Specifically, the EfficientNet-b0 structure pre-trained on the ImageNet was used. The resolution to extract the features is 0.48μm/pixel (under $20\times$ lens). The size of ROIs ranges from 3.02×10^4 to 1.62×10^6 μm^2. Correspondingly, the pixel resolution of the ROIs in the path ranges from 512×256 to 3360×2100. The size of patches is 224×224 and the dimension of patch feature is $d_f = 1280$ as defined in EfficientNet-b0. The step of window sliding is half of the window side length. As a result, the average number of patches in each ROI is 165.2. The node reduction factor $\gamma = n_{l+1}/n_l$ of the Diffpool module was set to 0.2 referring to [27]. The margin α in the loss function was set as half of the hash bits, e.g. $\alpha = 16$ for $d_h = 32$, referring to [52].

In both the training and the evaluation phases, a pair of paths were considered as *relevant* if the intersection set of their labels is nonempty and as *irrelevant*, otherwise. The precision of top N instances (P@N), mean reciprocal rank (MRR) and mean average precision (MAP) for recommendation were used to evaluate the recommendation performance, which are defined as follows

$$\begin{aligned} \text{P@N} &= \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|} p_i(N), \\ \text{MAP} &= \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|} \frac{\sum_{k=1}^{|\mathcal{D}|} p_i(N)\cdot r_{ik}}{\sum_{k=1}^{|\mathcal{D}|} r_{ik}}, \\ \text{MRR} &= \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|} \frac{1}{rank_i}, \end{aligned} \tag{12}$$

where $p_i(N) = \sum_{j=1}^{N} r_{ij}/N$ with $r_{ij} = 1$ denoting the i-th query instance and the j-th recommended instance are relevant and $r_{ij} = 0$, otherwise, $rank_i$ is the position of the first relevant instance to the i-th query instance, and $\mathcal{T}$ and $\mathcal{D}$ denote the collections of the query paths and database cases, respectively.

In the experiment, 295 WSIs (30%) were randomly sampled from the dataset as the query set $\mathcal{T}$ and the remainder were used for training. Furthermore, the training WSIs were randomly divided into five parts with the distribution of $(138, 137, 138, 138, 137)$ for five-fold cross-validation. The five parts took turns as the validation data with the other parts as the training data and also as the recommendation database $\mathcal{D}$. In each turn, the model corresponding to the least validation loss was recorded, then used to encode the database $\mathcal{D}$ and query dataset $\mathcal{T}$, and the metrics for this turn were calculated by Eq. 12. Finally, the average and standard deviation values of the metrics for the five turns were used to assess the performance of the network.

The initial learning rate was set at 0.01. The momentum was set as 0.9. The network was trained by 300 epochs, where the learning rate decayed by 10 times at the epoch of 150 and 225.

All the algorithms were implemented in python with PyTorch and run on a computer cluster with 10 available GPUs of Nvidia Geforce 2080Ti. The source code of the proposed method is available at https://github.com/zhengyushan/dpathnet.

C. Structure verification of DRA-Net

1) Ablation study: The DRA-Net ensembles GCN, RNN and attention module, etc. to achieve the path encoding. We first conducted an ablation study to verify the necessity of these components. The considered degraded models are introduced as follows

- *DRA-Net w/o GCN.* The adjacency matrix $\mathbf{A}$ in the graph is replaced as a zero matrix, for which the spatial adjacency information within the ROI is abandoned in the feature extraction.
- *DRA-Net w/o RNN.* The RNN is replaced as point-wise linear transformation followed by a global average pooling operation.

TABLE I
RESULTS FOR THE ABLATION STUDY, WHERE THE MEAN VALUES FOR THE FIVE-CROSS VALIDATION WITH THE STANDARD DEVIATIONS (IN PARENTHESES) ARE COMPARED AND THE BEST VALUES ARE SHOWN IN BOLD.

Networks	P@5	MRR	MAP
DRA-Net w/o GCN	0.782 (0.032)	0.812 (0.039)	0.812 (0.017)
DRA-Net w/o RNN	0.773 (0.035)	0.779 (0.052)	0.813 (0.019)
DRA-Net w/o Attention	0.778 (0.027)	0.791 (0.021)	0.799 (0.021)
DRA-Net w/o Triplet loss	0.777 (0.025)	0.782 (0.012)	0.825 (0.010)
DRA-Net	**0.810** (0.022)	**0.816** (0.022)	**0.836** (0.010)

- *DRA-Net w/o Attention.* The attention model is removed, and the output of the RNN $\mathbf{q}$ directly feeds to the hashing module.
- *DRA-Net w/o Triplet loss.* The triplet loss function is replaced with a common hashing loss function used in [14].

The experimental results are presented in Table I. It shows that the performance of recommendation decreases when discarding any of these components. The experiment has demonstrated that the contextual information modeled by the GCN and RNN is necessary to learn pathology knowledge from histopathology images, and the merging of the local patterns and the global representation by the attention module is important to improve the recommendation performance. The result has also proven the advantage of the triplet loss to the traditional hashing loss function for this problem.

2) Verification the RNN module: The function of the RNN module in the DRA-Net is to summary the information of the collection within different number of ROIs. There are other strategies besides the RNN module that can achieve the function in the DRA-Net. In this experiment, we implemented another two trainable modules based on interpolation and collection distance, which are defined as follows.

- *Interpolation* The collection of ROI features $\mathbf{P} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_{n^r}]^{\mathrm{T}}$ are resized to the same length through *Nearest/Linear interpolation* as $\mathbf{P}_{inter} = [\tilde{\mathbf{g}}_1, \tilde{\mathbf{g}}_2, ..., \tilde{\mathbf{g}}_{n_{max}}]^{\mathrm{T}}$, where n_{max} denotes the maximum path length in the dataset. The RNN module is removed and the hashing module acts on $\tilde{\mathbf{g}}_t$, converting it into binary-like code by equation $\tilde{\mathbf{y}}_t = \tanh(\mathbf{W}^h_{inter}\tilde{\mathbf{g}}_t + \mathbf{b}^h_{inter})$. Then, the path is represented by the concatenation of the collection of $\tilde{\mathbf{y}}_t$, which is formulated as $\mathbf{y}_{cat} = [\tilde{\mathbf{y}}_1^{\mathrm{T}}; \tilde{\mathbf{y}}_2^{\mathrm{T}}; ...; \tilde{\mathbf{y}}_{n_{max}}^{\mathrm{T}}]^{\mathrm{T}}$. Afterwards, $\mathbf{y}_{cat}$ is used as the path index in both the calculation of similarity measurement and the loss function. The training settings are the same with the *DRA-Net w/o Attention*.
- *Collection distance* The RNN module is removed and the hashing module directly acts on the ROI feature, converting it into binary-like code by equation $\mathbf{y}_t = \tanh(\mathbf{W}^h_{cd}\mathbf{g}_t + \mathbf{b}^h_{cd})$. The similarity of two paths is directly measured by the mean distance of all possible pairs of ROI codes across the two paths. Specifically, letting $\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_{|\mathcal{Y}|}\}$ and $\mathcal{Y}' = \{\mathbf{y}'_1, \mathbf{y}'_2, ..., \mathbf{y}'_{|\mathcal{Y}'|}\}$ denote the code collections of two diagnosis paths, the similarity measurement of the two paths are defined as

$$\theta_{\mathcal{Y},\mathcal{Y}'} = \frac{1}{|\mathcal{Y}|}\sum_{i=1}^{|\mathcal{Y}|}\frac{1}{|\mathcal{Y}'|}\sum_{j=1}^{|\mathcal{Y}'|}\frac{1}{2}sign(\mathbf{y}_i)^{\mathrm{T}}sign(\mathbf{y}'_j).$$

  Correspondingly, the item $\frac{1}{2}\mathbf{y}^{\mathrm{T}}_{a_m}\mathbf{y}_{p_m}$ in the loss function (Eq.11) is replaced as the collection distance format

$$\frac{1}{|\mathcal{Y}_{a_m}|}\sum_{\mathbf{y}_{a_m,i}\in\mathcal{Y}_{a_m}}\frac{1}{|\mathcal{Y}_{p_m}|}\sum_{\mathbf{y}_{p_m,j}\in\mathcal{Y}_{p_m}}\frac{1}{2}\mathbf{y}^{\mathrm{T}}_{a_m,i}\mathbf{y}_{p_m,j},$$

  where $\mathcal{Y}_{a_m}$ and $\mathcal{Y}_{p_m}$ denote the code collections of the anchor path and the positive path, and the same modification is done to the item $\frac{1}{2}\mathbf{y}^{\mathrm{T}}_{a_m}\mathbf{y}_{n_m}$. The regularization for binarization is adjusted to act on each binary-like code in $\mathcal{Y}_{a_m}$ and $\mathcal{Y}_{p_m}$. Then, the whole network is trained end-to-end.

For a fair comparison, all the methods did not use attention modules. The results are presented in Table II. The results of Rows 1&2 in the table show that the strategy based on interpolation cannot meet the requirement of the application. Theoretically, the representation $\mathbf{p}_{cat}$ obtained by concatenating the ROI representations implies a strict restriction of the order the pathologists browsing the ROIs. It will give a low similarity score when two paths are diagnostically relevant but the positions of the conclusive regions in the two paths are staggered, which contradicts to the labels we use in the training. This label noise affected the converge of the network, and meanwhile led to poor recommendation performance. The method *Collection-distance* achieved a reasonable result. The cross-region measurement strategy ensures all pairs of conclusive regions across two paths can be reflected in the similarity measurement. However, this strategy indiscriminately calculates the similarities between all possible ROI pairs. The values from inconclusive ROIs will weaken the discrimination of the similarity measurement. In contrast, the network equipped with the RNN module performs significantly better. The *update gate* and *reset gate* in the RNN module makes the network be able to maintain the information of important ROIs while restraining the effect of inconclusive ROIs during the recurrent encoding. It is the major advantage of the RNN strategy to the *Collection-distance*. Moreover, the recurrent computation of RNN keeps a linear time complexity ($\mathcal{O}(n)$) to the length of the collections during the running of the recommendation algorithm, which is more applicable compared to *Collection-distance* whose complexity is $\mathcal{O}(n^2)$. Finally, we adopted the RNN strategy in the proposed method.

TABLE II
PERFORMANCE COMPARISON OF RNN MODULE AND THE ALTERNATIVE STRATEGIES, WHERE THE TIME COMPLEXITY OF THE RECOMMENDATION TO THE LENGTH OF THE PATH IS MEANWHILE EVALUATED BY $\mathcal{O}$ NOTATION.

No.	Method	P@5	MRR	MAP	Time Complexity
1	Nearest-interpolation	0.596 (0.022)	0.602 (0.027)	0.544 (0.026)	$\mathcal{O}(n)$
2	Linear-interpolation	0.626 (0.032)	0.621 (0.037)	0.571 (0.027)	$\mathcal{O}(n)$
3	Collection-distance	0.731 (0.031)	0.748 (0.032)	0.706 (0.021)	$\mathcal{O}(n^2)$
4	RNN (DRA-Net w/o Attention)	**0.778** (0.027)	**0.791** (0.021)	**0.799** (0.021)	$\mathcal{O}(n)$

TABLE III
RECOMMENDATION PERFORMANCE FOR DIFFERENT RNN MODULES.

RNN	P@5	MRR	MAP
Simple	0.768 (0.039)	0.775 (0.041)	0.807 (0.012)
LSTM	0.797 (0.021)	0.804 (0.026)	0.818 (0.025)
GRU	**0.810** (0.022)	**0.816** (0.022)	**0.836** (0.010)

We additionally assessed another two RNN structures for the complete DRA-Net. The one is the simple RNN defined as $\mathbf{h}_t = \tanh(\mathbf{W}[\mathbf{g}_t, \mathbf{h}_{t-1}] + \mathbf{b})$ and another is LSTM. The results in Table III indicates that the GRU structure is the most appropriate for the DRA-Net.

D. Sufficiency assessment of the path-level supervision

The DRA-Net is expected to learn pathology knowledge for cancer diagnosis without manual annotation on the WSI. Generally, the performance of a learning system will decrease as the supervision becomes weaker. In this experiment, we quantified the gap of the proposed training strategy to those by stronger supervisions. One optional supervision is for the CNN. Specifically, the patches in the training set were labeled as six classes (the 5 lesion types plus a type of normal tissue) according to the pixel-level annotation and used to train the CNN via image patch classification task. Then, in the graph construction stage, the CNN trained by the patch classification task substituted the CNN trained on the ImageNet dataset to extract the patch features. Another optional supervision is for the GCN structure. The ROIs were labeled via majority voting of the patch labels. Then, a hashing layer was built on the ROI features, which was formulated as

$$\mathbf{y}_i^r = \tanh(\mathbf{W}_{rh}\mathbf{g}_i + \mathbf{b}_{rh}), \tag{13}$$

where $\mathbf{g}_i$ denotes the feature of the i-th ROI within the training set and $\mathbf{W}_{rh}$ and $\mathbf{b}_{rh}$ are the weight and bias for the ROI-level hash function, respectively. Correspondingly, an additional loss function was defined as

$$\begin{aligned} L = &- \frac{1}{M^r}\sum_{i=1}^{M^r} \log\sigma(\frac{1}{2}\mathbf{y}_{a_i}^{r\mathrm{T}}\mathbf{y}_{p_i}^r - \frac{1}{2}\mathbf{y}_{a_i}^{r\mathrm{T}}\mathbf{y}_{n_i}^r - \alpha) \\ &+ \lambda\frac{1}{M^r}\sum_{i=1}^{M^r}\sum_{k\in\{a_i,p_i,b_i\}} \|\mathbf{y}_k^r - \mathbf{c}_k^r\|_2^2, \end{aligned} \tag{14}$$

where $(\mathbf{y}_{a_i}^r, \mathbf{y}_{p_i}^r, \mathbf{y}_{n_i}^r)$ denotes codes of (anchor, positive, negative) for the i-th ROI, M^r is the number of training ROIs and $\mathbf{c}_k^r = sign(\mathbf{y}_k^r)$. The additional loss was added to the path-level loss (Eq. 11) in the training of DRA-Net.

The models trained by different supervision combination are compared in Table IV. Generally, the model trained with full supervision (the first row in Table IV) achieved the best recommendation performance. The MAP for this model has reached to 0.851, which can be regarded as the upper bound of the DRA-Net structure for the SHIR task. The DRA-Net trained simply by the path-level labels has a MAP of 0.836, which is only 0.015 inferior to the upper bound. Meanwhile, the accuracy for the top-recommended items (P@5 and MRR) shows little difference to the upper bound. The results have indicated that the path-level (i.e., the WSI-level) supervision is sufficient to train the DRA-Net. It means the workload of pixel-level annotations for pathologists can be relieved in the cost of a slight decrease in the recommendation accuracy. The result is promising for building an automatic learning system based on large scale telepathology databases.

We also considered the rate of convergence of the assessed models. The loss curves for training are drawn in Figure 5. It shows the proposed DRA-Net converged in 300 epochs, which cost about 2 times longer than the models utilizing pixel-level supervision. The speed of convergence is acceptable.

E. Effect of the session length

The recommendation application is expected to continuously running from the first view to the last view during the diagnosis, and the recommendation results should be updated when a new view is appended to the path. In this situation, we conducted experiments to assess the running performance of the system. Specifically, for each path in the testing set, we evaluated the

TABLE IV
COMPARISON OF THE DRA-NET MODELS TRAINED BY DIFFERENT SUPERVISION STRATEGIES, WHERE *Pixel*, *ROI*, AND *Path* RESPECTIVELY DENOTE THE SUPERVISION OF PIXEL-LEVEL ANNOTATIONS, ROI-LEVEL LABELS, AND PATH-LEVEL LABELS CONSIDERED IN THE TRAINING. THE DETAIL DESCRIPTION FOR THE SUPERVISION PLEASE REFER TO SECTION IV-D

No.	Supervision			P@5	MRR	MAP
	Pixel	*ROI*	*Path*			
1	✓	✓	✓	**0.813** (0.057)	0.815 (0.057)	**0.851** (0.021)
2	✓	×	✓	0.806 (0.028)	0.808 (0.023)	0.843 (0.010)
3	×	✓	✓	0.805 (0.005)	0.809 (0.018)	0.847 (0.008)
4	×	×	✓	0.810 (0.022)	**0.816** (0.022)	0.836 (0.010)

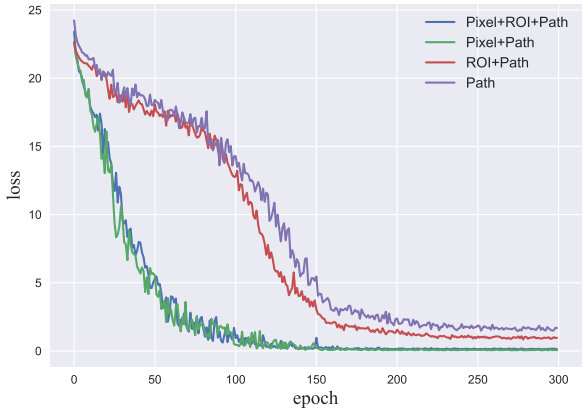

Fig. 5. The loss values as a function of training epochs for DRA-Net trained by different supervision strategies.

recommendation accuracy for time $t = 1, ..., max(20, n^r)$. The path label for time t was simulated based on the annotation of the pathologist. The average metrics as a function of t are displayed in Figure 6.

Overall, the precision for recommendation apparently improved as the running length of the path (i.e. the number of ROIs have been reviewed by the pathologists) increasing. Especially, the P@5 is above 93% when the length reached to 17 and then changes little. The results have demonstrated that the DRA-Net has successfully summarized the local pattern of the current ROI into the global path representation as the path growing. Furthermore, the increasing monotony trend of the curve indicates that the information of the conclusive ROIs has been well reserved by the attention module for the final decision.

F. Efficiency of the recommendation

The proposed SHIR application is desired to provide real-time assistant to pathologists during the diagnosis. Therefore, the running time is an important property for the application. The entire recommendation procedure can be divided into four stages: 1) patch feature extraction based on the CNN, 2) ROI feature extraction based on the GCN, 3) ROI feature fusion based on the RNN and attention module, and 4) binarization and retrieval. The first two stages are processed only once as a new ROI appended to the path, and the extracted ROI feature will be shared in the follow-up processing.

The factors that significantly affects the speed of the proposed method are the magnification to extract the ROI features and the step of patch sampling. In this experiment, we evaluated the different combination of the two factors. The results are presented in Table V. Generally, the accuracy of recommendation increases when the magnification becomes larger and the step of patch sampling becomes smaller. Correspondingly, the FLOPs significantly grow as the average number of patches in each ROI increases. To meet the real-time requirement and meanwhile maintain the high accuracy of the proposed framework, we determined to extract ROI features under the magnification of 20 lenses and set the step of sampling patches to be half

TABLE V
EFFECT OF THE ROI MAGNIFICATION AND PATCHES SAMPLING IN THE GRAPH CONSTRUCTION.

No.	Configuration		P@5	MRR	MAP	#FLOPs
	Mag.	Step				
1	10×	224	0.760	0.767	0.796	**10.14B**
2	10×	112	0.781	0.791	0.819	21.31B
3	20×	224	0.789	0.807	0.810	24.85B
4	20×	112	0.810	0.816	**0.836**	72.79B
5	40×	224	0.806	0.811	0.822	76.95B
6	40×	112	**0.817**	**0.822**	0.835	258.49B

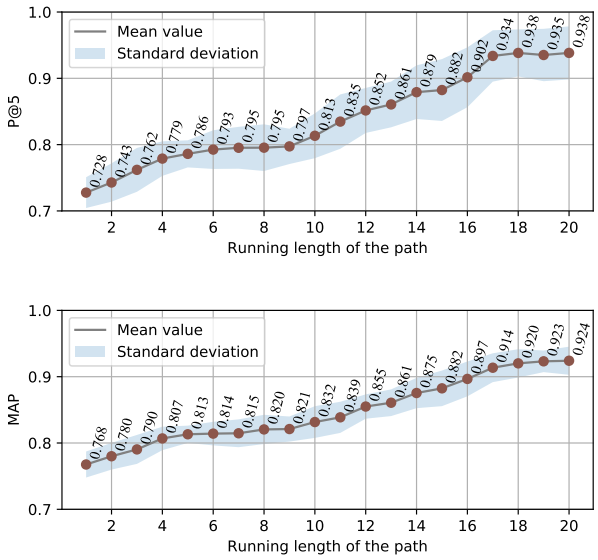

Fig. 6. The P@5 and MAP as function of the running length of the testing paths, where the red points are the mean values of the five-cross validation and the corresponding standard deviation values are represented by blue shadow.

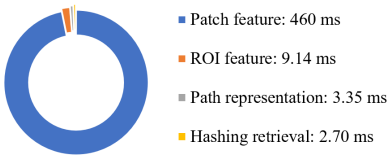

Fig. 7. The proportion of time consumption of the framework, for which the total time for a step of recommendation is an average of 475 ms over the test set.

of the patch size (The No. 4 in Table V). As a result, the average time was 475 ms by using one GPU in our experiment environment. The speed is promising to develop real-time AI assistant for cancer diagnosis. The specific times for the four stages are illustrated in Figure 7. Obviously, the inference of CNN costs 460 ms, which is about 96% of the total processing time. The inference of the DRA-Net and the following retrieval process cost about 15 ms, which is very fast and will not significantly slow down the recommendation speed as the number of the ROIs increases.

The time for retrieval is in an average of 2.70 ms. It benefits from the binary encoding of the path representation and hashing search based on the Hamming distance. The number of hash bits can be set larger as the scale of the dataset increasing. The comparison for different bit number on the gastric dataset is given in Table VI. It presents the recommendation performance is robust to the number of hash bits and 32 is the most appreciated to this dataset.

G. Visualization

Figure 8 illustrates 2 typical instances of the recommendation results by our method. The path in Figure 8(a) starts with adenocarcinoma regions, and thereby the recommended WSIs are all from adenocarcinoma cases. Then, the WSIs within SRCC region appears to the recommendation queue at time $t = 4$ where an SRCC ROI is appended to the input path (as directed by red arrows). And more WSIs within both adenocarcinoma and SRCC regions are returned when more SRCC ROIs are appended into the path as the time flowing. The results indicate that the proposed SHIR framework is sensitive to the change of image content under the path and can provide appropriate information to the pathologists for aided diagnosis. Furthermore, the recommendation queue changes little after $t = 7$. The main reason is that the attention score $\mathbf{a}$ defined in Eq. 8 (visualized as the attentive regions in Figure 8) tends to be stable as sufficient regions are provided. It also demonstrates that the attention module designed in the DRA-Net is effective and robust to locate conclusive regions for lesion recognition. The path in Figure 8(b) displays a hard sample from adenocarcinoma cases. The first ROI in the path ($t = 1$) was confused with HGIN and, as a consequence, the recommendation queue was occupied by WSIs within HGIN. While the mistake was quickly corrected when more ROIs were appended to the path.

TABLE VI
THE EFFECT OF THE HASH BITS FOR THE PROPOSED RECOMMENDATION FRAMEWORK.

Hash bits	P@5	MRR	MAP
16	0.795 (0.025)	0.800 (0.024)	0.835 (0.021)
32	**0.810** (0.022)	**0.816** (0.022)	0.836 (0.010)
48	0.805 (0.026)	0.807 (0.026)	**0.841** (0.021)
64	0.791 (0.018)	0.800 (0.019)	0.832 (0.019)

TABLE VII
COMPARISON OF THE RECOMMENDATION METRICS FOR RELATED APPROACHES AND THE PROPOSED METHOD, WHERE THE BEST SCORE FOR EACH COLUMN IS DISPLAYED IN BOLD.

No.	Method	Supervision			P@5	P@20	MRR	MAP
		Pixel	*ROI*	*Path*				
1	Jimenez-del-Toro et al. [26]	×	×	×	0.647 (0.042)	0.635 (0.046)	0.691 (0.025)	0.663 (0.036)
2	Yan et al. [28]	×	✓	×	0.701 (0.036)	0.684 (0.034)	0.732 (0.018)	0.721 (0.025)
3	Zheng et al. [27]	×	✓	×	0.712 (0.032)	0.693 (0.036)	0.716 (0.041)	0.735 (0.018)
4	DPath-Net[29]	×	×	✓	0.778 (0.027)	0.760 (0.026)	0.791 (0.021)	0.799 (0.021)
5	DRA-Net (Ours)	×	×	✓	**0.810** (0.022)	**0.800** (0.012)	**0.816** (0.022)	**0.836** (0.010)

H. Comparison with related methods

We compared the proposed method with related works. Since we are the first to deal with the recommendation problem for histopathology WSIs, there are no complete frameworks that can be directly compared. In this situation, we modified three related methods [26], [27], [28] to meet the requirement of the proposed recommendation application. The methods were detailed below.

- *Jimenez-del-Toro et al.* [26] The method deals with the retrieval task of WSIs. Referring to [26], we measured the similarity of two WSIs by the mean cosine distance of all possible pairs of patch features across the two paths and obtained the recommendation results by similarity ranking.
- *Yan et al.* [28] The method is designed for histopathology ROI classification. The CNN features of patches in an ROI are ordered as sequential data from left to right and then top to bottom based on the patch locations in the ROI. Then, the sequential data is fed into a 4-layer bidirectional LSTM to obtain the ROI representation.
- *Zheng et al.* [27] The method is designed for irregular-shape ROI retrieval. The ROIs are encoded as graphs and the features of the ROIs are extracted by GCNs with Diffpool modules. The approach to constructing the graphs and the structure of the GCN-Diffpool network are the same as those in the DRA-Net.
- *DPath-Net* [29] The recommendation network finalized in the conference version of the paper, which is also the method *DRA-Net w/o Attention* discussed in the ablation study.

The methods by *Yan et al.* and *Zheng et al.* rely on ROI labels. In this experiment, we trained the networks in the two methods based on the ROI labels. Meanwhile, we adapted the patch-based similarity measurement proposed in [37] to the ROI features to realize the recommendation. The comparison results are given in Table VII.

Overall, the proposed DRA-Net performed significantly better than the other methods. The level of supervision (ROI labels) in [28] and [27] was stronger than that in DRA-Net, but the two methods considered the ROIs as individual samples and did not utilize the contextual information among the ROIs. In comparison, DRA-Net modeled the contextual information with the RNN. The higher level of pathology knowledge for lesion recognition was learned from the relationship within the collection of ROIs. Therefore, the recommendation performance of DRA-Net is superior to those by [28] and [27].

I. Extended evaluation for ROI retrieval

It is noted that the GCN module in the trained DRA-Net is able to extract ROI features, which are potential to achieve the task of ROI retrieval. Therefore, we conducted extended experiments to assess the performance of the DRA-Net for ROI retrieval. In this case, the DRA-Net should be regarded as a weakly supervised solution since the training of the DRA-Net did not use ROI labels. The dataset split in this experiment was the same as that in the WSI recommendation experiment. The ROIs in the paths were regarded as individual samples in the retrieval. The P@5, MRR, and MAP are also used as the metrics. Furthermore, we additionally calculated the Recall@N (R@N) [16], [53], [54], [55] in this experiment to assess the success rate that the doctors obtain useful information by only examining N returned items. The R@N is defined as

$$\mathrm{R@N} = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|}\max_{1\le j\le N}(r_{ij}) \tag{15}$$

In addition, we implemented an unsupervised framework, multi-binary-code (MBC) [37] and a fully supervised method, GCN-Hash [27] for comparison purpose. The results are presented in Table VIII.

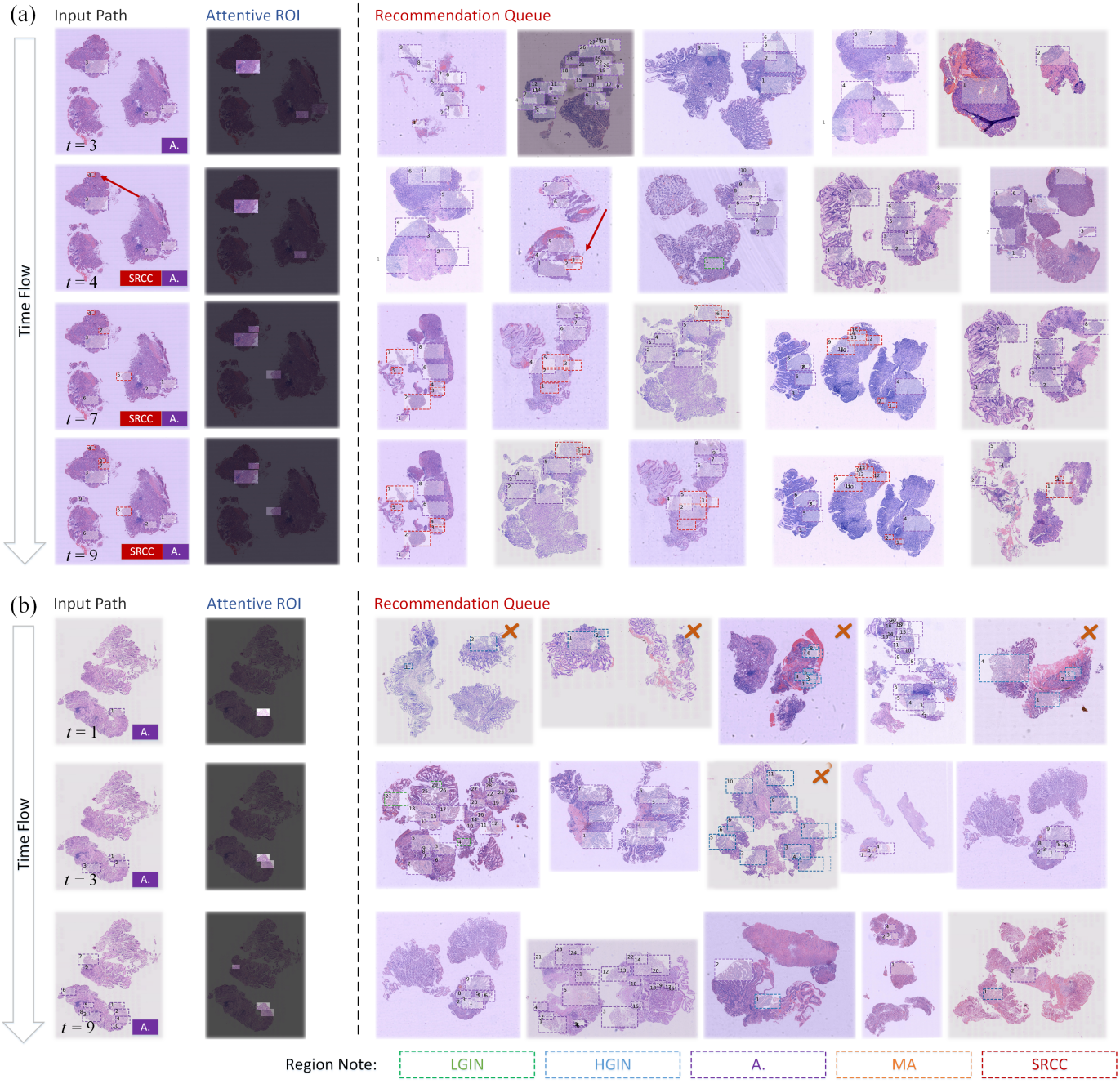

Fig. 8. Instances of visual performance of the proposed recommendation framework, where (a) displays the result for a path within A. and SRCC regions, (b) displays a path within only A. regions, the first column in each instance gives the input path in different time t, the second column shows the top-3 attentive regions according to the attention scores (Eq. 8), and the right columns present the recommended slides (the incorrect results are marked by red crosses). For a clear display, only the path screens within abnormal tissue are counted and drawn in the figure.

Overall, The retrieval performance by the DRA-Net is significantly better than the unsupervised method (MBC) and is also comparable with the fully supervised method (GCN-Hash) under the metrics P@5, R@5, and MRR. Especially, our method achieved an R@5 of 0.849 and an MRR of 0.755, which are the best in the comparison. The results have indicated the proposed DRA-Net is potential to develop an efficient ROI retrieval system. Moreover, the comparable retrieval performance with the fully supervised framework has demonstrated that the DRA-Net learned discriminative ROI features to identify the regions containing different tissue types although only the path-level labels were provided.

V. Discussion

In the modeling phase, we took the browsing path and employed an RNN module to extract the feature of the path, and found the design was reasonable based on experimental evaluation. The usage of the RNN module here is quite similar to the

TABLE VIII
COMPARISON OF ROI RETRIEVAL PERFORMANCE ON THE GASTRIC DATASET, WHERE THE CLASSIFICATION ACCURACY BY MAJORITY VOTING OF THE TOP-5 RETURNED SAMPLES, I.E. THE K-NEAREST NEIGHBOR CLASSIFIER WITH K=5, IS ALSO EVALUATED.

Method	Retrieval				Classification
	P@5	R@5	MRR	MAP	Accuracy
MBC [37]	0.610	0.805	0.709	0.658	0.671
GCN-Hash [27]	**0.725**	0.735	0.743	**0.798**	**0.731**
DRA-Net-ROI	0.692	0.**849**	**0.755**	0.739	0.695*

* The accuracy was calculated as a reference only. Theoretically, the ROI classification cannot be realized here because the ROI labels should not be available in this experiment.

TABLE IX
EXPLICIT ASSESSMENT OF THE CONTRIBUTION OF THE SEQUENTIAL DATA.

Operation	P@5	MRR	MAP
Rand-erase 20% ROIs	0.755 (0.033)	0.781 (0.026)	0.791 (0.018)
Rand-erase 50% ROIs	0.676 (0.029)	0.660 (0.037)	0.751 (0.022)
Shuffle ROIs	0.809 (0.015)	0.812 (0.020)	0.832 (0.007)
Original	**0.810** (0.022)	**0.816** (0.022)	**0.836** (0.010)

studies [8], [28] in the domain of histopathology image/WSI analysis. Here, we further assess the effect of ROI order in the path by conducting the following experiments.

- *Rand-erase n% ROIs*: $n\%$ of the ROIs in each path are randomly erased during the training and testing stages. (The path contains at least one ROI.) The order of the remainder ROIs and the label for the path are not changed.
- *Shuffle ROIs*: The order of the ROIs within each path is randomly shuffled.

The results are provided in Table IX. The performance significantly reduced when half of the ROIs were removed from each path (see No.2 in Table IX). Especially, the precision of the top-5 recommended cases decreased by 15.4%. The most likely reason is the conclusive ROIs were lost in some paths and the semantics within these path were changed, which had misled the DRA-Net in learning diagnostic patterns. Whereas, the performance was almost unchanged when the ROIs in a path were randomly shuffled. It indicates the order the pathologists review the screens has limited effect on the diagnostic information within the path data. Based on these results, we can conclude that 1) the complete record of the browsing path for the diagnosis of the pathologist is important to our framework, and 2) the main contribution of the RNN module in the DRA-Net is to summarize the key information within the path rather than depict the order the ROI appearing in the path. Moreover, the comparable results achieved by *Shuffle ROIs* have shown possible benefits to the practical application that 1) the ROIs do not have to be fed into the network strictly in chronological order which makes the application of the DRA-Net more flexible, and 2) the shuffle operation could be used as a type of data augmentation which would improve the robustness in the modeling of some histopathology types.

The context of our research is online telepathology platforms, where the data mainly come from remote consultations. There will be multiple paths from different users of the platform for a WSI if the WSI is publicly available. However, only the path generated by the assigned pathologist during the consultation is rigorous related to the diagnostic report of the pathologist. The paths generated by other users have indeterminate meanings because many of them browse the WSI for purposes other than diagnosis. Therefore, only the browsing record generated during the consultation period can be considered for the SHIR application. In the construction of the gastric dataset, we assigned the WSIs we collected to pathologists working for the telepathology platform to make diagnoses. During the diagnosis, we recorded the browsing paths of the pathologists. A WSI was assigned to a single pathologist. Therefore, there is a single path for each WSI in the gastric dataset.

We have tried to detect focused screens by monitoring the duration of stay and the movement distance on the slide as suggested in [56]. However, the time of loading a screen from the cloud for clear display varies a lot to the internet condition and computer performance. And the meaning of the movement distance is also different as the screen resolution of the computer changes. These issues had made it hard to design robust thresholds for the path collection task. Instead, we directly used the tag of *Rendering Complete* (The rendering includes downloading image tiles within the screen from the cloud and mosaicking the tiles for clear display) of the platform as the trigger of ROI recording. The principle is as follows. 1) Before the rendering is completed, the screen will display a blurred thumbnail of the region. The pathologist should wait for the rendering to be completed if the region is important in the diagnosis. 2) The rendering starts when the screen stops moving. Therefore, quick sliding on the WSI will not trigger the rendering, and thereby the passing area will not be recorded. 3) Owning to the cache mechanism of the browsing system, small movement and zoom or the revisit to a specific screen will not repeatably trigger the rendering and thereby ROIs with high overlap rate will not be repeatably recorded to the path. Benefiting from these properties of the digital platform, the ROIs were robustly recorded without excessive redundancy.

The reason we decided to extract the ROI features using graph neural networks (GNNs) is that the ROI varies in size and shape. Theoretically, there are no special restrictions on GNN structures in our framework. The reason we used the graph

convolution network (GCN) structure in [1] is the GCN structure has been proven robust and effective in a large number of studies on graph data processing. We have tried to substitute the GCN embedding formulation in the DRA-Net with the naive GNN embedding $\mathbf{H}^{(k)} = ReLU(\mathbf{A}\mathbf{H}^{(k-1)}\mathbf{W}^{(k)})$, and observed a slight decrease in the metric P@5 by 0.009 and MAP by 0.007 for the recommendation performance task. This change is reasonable, which is in line with the expectations.

Although we found the RNN performs better than a naive mean pooling layer, we also realized the discrimination of the path representation can be further improved by the attention mechanism referring to the recent studies on the relevant recommendation [23], [42]. That was one of the reasons we built the self-attention module at the end of the DRA-Net. The experimental results have shown that the self-attention module has improved the P@5 by 3.2%, MRR by 2.5%, and the MAP by 3.7%. The improvement is significant.

The labels in the standard cross-entropy loss function (the widely used class-exclusive loss function) are required to be one-hot, i.e., only one entry in the label vector is allowed as non-zero. Whereas the path labels considered in our problem sometimes contain more than one non-zero entries. Therefore, the cross-entropy loss is inappropriate to the DRA-Net. That is one of the reasons that we used the triplet loss function to train the DRA-Net.

The methods we compared in the manuscript were either similar application or related methodology to our work. The paper [26] deals with the retrieval task of WSIs, of which the application is similar to ours. The paper [28] proposes encoding the structural information of an ROI using the RNN. It provides an alternative solution to encode an ROI besides the GCNs we used.

Technically speaking, the methodology of the proposed SHIR application is with the domain of content-based image retrieval. The recommendation is achieved by searching for diagnostically relevant cases from the established database. The main difference of the proposed method from the general CBIR framework is that the query data is a sequence of ROIs with different lengths, rather than a single ROI. The main reason we drew an analogy with the recommendation systems is that the user experience of the proposed application is similar to the recommendation. The system feeds back relevant merchandise/tumor cases based on the context the users/medical doctors have browsed. The major difference of the proposed SHIR application to the common session-based recommendation is that the goal of the application is to predict and retrieve the diagnostically relevant cases from the historical archives, rather than predict which regions the pathologists would review next.

VI. Conclusion

In this paper, we contributed a novel DRA-Net for modeling the browsing path of the pathologist and a novel computer-aided cancer diagnosis framework named session-based histopathological image recommendation (SHIR). The WSI label is validated to be sufficient to train the DRA-Net, for which pixel-level or ROI-level annotations of pathologists can be relieved. The SHIR based on the DRA-Net can actively recommend diagnostically relevant cases from the database of the telepathology platform while the pathologists are browsing the WSI. The experiments have shown that the DRA-Net has successfully learned the pathology knowledge for lesion recognition by only using the WSI labels, and the SHIR framework achieves a good accuracy in the application for gastric cancer database. The time for a step of recommendation is less than 0.5 seconds, which is very efficient and is adequate to develop real-time applications. The future work will focus on training automatic cancerous region detection models based on the supervision of the diagnosis paths.

References

[1] B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.

[2] P. Bandi *et al.*, "From detection of individual metastases to classification of lymph node status at the patient level: The camelyon17 challenge," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 550–560, 2019.

[3] M. Veta *et al.*, "Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge," *Medical image analysis*, vol. 54, pp. 111–121, 2019.

[4] T. C. Hollon *et al.*, "Near real-time intraoperative brain tumor diagnosis using stimulated raman histology and deep neural networks," *Nature Medicine*, vol. 26, no. 1, pp. 52–58, 2020.

[5] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[6] H. Lin, H. Chen, S. Graham, Q. Dou, N. Rajpoot, and P.-A. Heng, "Fast scannet: Fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1948–1958, 2019.

[7] T. Falk *et al.*, "U-net: deep learning for cell counting, detection, and morphometry," *Nature methods*, vol. 16, no. 1, p. 67, 2019.

[8] G. Campanella *et al.*, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature medicine*, vol. 25, no. 8, pp. 1301–1309, 2019.

[9] H. Le *et al.*, "Pancreatic cancer detection in whole slide images using noisy label annotations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 541–549.

[10] G. Xu, Z. Song, Z. Sun, C. Ku, Z. Yang, C. Liu, S. Wang, J. Ma, and W. Xu, "Camel: A weakly supervised learning framework for histopathology image segmentation," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 10 681–10 690.

[11] H. Su, X. Shi, J. Cai, and L. Yang, "Local and global consistency regularized mean teacher for semi-supervised nuclei classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 559–567.

[12] J. van der Laak, F. Ciompi, and G. Litjens, "No pixel-level annotations needed," *Nature Biomedical Engineering*, vol. 3, no. 11, pp. 855–856, 2019.

[13] Z. Li, X. Zhang, H. Müller, and S. Zhang, "Large-scale retrieval for medical image analytics: A comprehensive review," *Medical image analysis*, vol. 43, pp. 66–84, 2018.

[14] X. Shi, M. Sapkota, F. Xing, F. Liu, L. Cui, and L. Yang, "Pairwise based deep ranking hashing for histopathology image classification and retrieval," *Pattern Recognition*, vol. 81, pp. 14–22, 2018.

[15] Y. Zheng *et al.*, "Histopathological whole slide image analysis using context-based cbir," *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1641–1652, 2018.

[16] S. Kalra, H. Tizhoosh, C. Choi, S. Shah, P. Diamandis, C. J. Campbell, and L. Pantanowitz, "Yottixel an image search engine for large archives of histopathology whole slide images," *Medical Image Analysis*, p. 101757, 2020.

[17] M. S. Hosseini *et al.*, "Atlas of digital pathology: A generalized hierarchical histological tissue type-annotated database for deep learning," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11 747–11 756.

[18] G. Litjens *et al.*, "1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset," *GigaScience*, vol. 7, no. 6, 2018.

[19] Z. Li *et al.*, "Deep learning methods for lung cancer segmentation in whole-slide histopathology images - the acdc@lunghp challenge 2019," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2020.

[20] B. Gecer, S. Aksoy, E. Mercan, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks." *Pattern Recognition*, vol. 84, pp. 345–356, 2018.

[21] E. Mercan, L. G. Shapiro, T. T. Bruny, 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.

[22] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, "Stamp: Short-term attention/memory priority model for session-based recommendation," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 1831–1839.

[23] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, "Session-based recommendation with graph neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 346–353.

[24] M. Sapkota, X. Shi, F. Xing, and L. Yang, "Deep convolutional hashing for low-dimensional binary embedding of histopathological images," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 805–816, 2018.

[25] T. Peng *et al.*, "Multi-task learning of a deep k-nearest neighbour network for histopathological image classification and retrieval," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 676–684.

[26] O. Jimenez-del Toro, S. Otálora, M. Atzori, and H. Müller, "Deep multimodal case–based retrieval for large histopathology datasets," in *MICCAI 2018 Workshop on Patch-based Techniques in Medical Imaging*. Springer, 2017, pp. 149–157.

[27] Y. Zheng, B. Jiang, J. Shi, H. Zhang, and F. Xie, "Encoding histopathological wsis using gnn for scalable diagnostically relevant regions retrieval," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 550–558.

[28] R. Yan, *et al.*, "Breast cancer histopathological image classification using a hybrid deep neural network," *Methods*, vol. 173, pp. 52–60, 2019.

[29] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, and J. Shi, "Tracing diagnosis paths on histopathology wsis for diagnostically relevant case recommendation," in *Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 459–469.

[30] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards Large-Scale Histopathological Image Analysis: Hashing-Based Image Retrieval," *IEEE Transactions on Medical Imaging*, vol. 34, no. 2, pp. 496–506, 2015.

[31] M. Jiang, S. Zhang, J. Huang, L. Yang, and D. N. Metaxas, "Scalable histopathological image analysis via supervised hashing with multiple features," *Medical Image Analysis*, vol. 34, pp. 3–12, 2016.

[32] X. Shi, F. Xing, K. Xu, Y. Xie, H. Su, and L. Yang, "Supervised graph hashing for histopathology image retrieval and classification," *Medical Image Analysis*, vol. 42, p. 117, 2017.

[33] D. Hu, Y. Zheng, H. Zhang, S. Sun, F. Xie, J. Shi, and Z. Jiang, "Informative retrieval framework for histopathology whole slides images based on deep hashing network," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 244–248.

[34] Y. Ma *et al.*, "Breast histopathological image retrieval based on latent dirichlet allocation," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 4, pp. 1114–1123, July 2017.

[35] ——, "Generating region proposals for histopathological whole slide image retrieval," *Computer methods and programs in biomedicine*, vol. 159, pp. 1–10, 2018.

[36] J. R. R. Uijlings, K. E. A. V. D. Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.

[37] Y. Zheng, Z. Jiang, H. Zhang, F. Xie, Y. Ma, H. Shi, and Y. Zhao, "Size-scalable content-based histopathological image retrieval from database that consists of wsis," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1278–1287, 2018.

[38] P. Chen, X. Shi, Y. Liang, Y. Li, L. Yang, and P. D. Gader, "Interactive thyroid whole slide image diagnostic system using deep representation." *Computer Methods and Programs in Biomedicine*, vol. 195, p. 105630, 2020.

[39] Y. K. Tan, X. Xu, and Y. Liu, "Improved recurrent neural networks for session-based recommendations," in *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, 2016, pp. 17–22.

[40] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," in *ICLR 2016 : International Conference on Learning Representations 2016*, 2016.

[41] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, "Neural attentive session-based recommendation," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1419–1428.

[42] C. Xu *et al.*, "Graph contextualized self-attention network for session-based recommendation," in *Proc. 28th Int. Joint Conf. Artif. Intell.(IJCAI)*, 2019, pp. 3940–3946.

[43] H. Dai, B. Dai, and L. Song, "Discriminative embeddings of latent variable models for structured data," in *International conference on machine learning*, 2016, pp. 2702–2711.

[44] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR 2017 : International Conference on Learning Representations 2017*, 2017.

[45] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *arXiv preprint arXiv:1901.00596*, 2019.

[46] Z. Ying *et al.*, "Hierarchical graph representation learning with differentiable pooling," in *Advances in neural information processing systems*, 2018, pp. 4800–4810.

[47] K. Cho *et al.*, "Learning phrase representations using rnn encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734.

[48] A. Vaswani *et al.*, "Attention is all you need," in *Neural information processing systems*, 2017, pp. 6000–6010.

[49] X. Wang, Y. Shi, and K. M. Kitani, "Deep supervised hashing with triplet labels," in *Asian conference on computer vision*. Springer, 2016, pp. 70–84.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026–1034.

[51] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*, 2019, pp. 6105–6114.

[52] C. Deng, Z. Chen, X. Liu, X. Gao, and D. Tao, "Triplet-based deep hashing network for cross-modal retrieval," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 3893–3903, 2018.

[53] D. Manandhar, M. Bastan, and K.-H. Yap, "Semantic granularity metric learning for visual search," *Journal of Visual Communication and Image Representation*, vol. 72, p. 102871, 2020.

[54] F. Cakir, K. He, X. Xia, B. Kulis, and S. Sclaroff, "Deep metric learning to rank," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 1861–1870.

[55] H. Jegou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 117–128, 2011.

[56] E. Mercan, S. Aksoy, L. G. Shapiro, D. L. Weaver, T. T. Bruny, and J. G. Elmore, "Localization of diagnostically relevant regions of interest in whole slide images: a comparative study." *Journal of Digital Imaging*, vol. 29, no. 4, pp. 496–506, 2016.