Multi-Magnification Image Search in Digital Pathology

Maral Rasoolijaberi, Morteza Babaei, Abtin Riasatian, Sobhan Hemati, Parsa Ashrafi, Ricardo Gonzalez, and Hamid R. Tizhoosh

***Abstract*—This paper investigates the effect of magnification on content-based image search in digital pathology archives and proposes to use multi-magnification image representation. Image search in large archives of digital pathology slides provides researchers and medical professionals with an opportunity to match records of current and past patients and learn from evidently diagnosed and treated cases. When working with microscopes, pathologists switch between different magnification levels while examining tissue specimens to find and evaluate various morphological features. Inspired by the conventional pathology workflow, we have investigated several magnification levels in digital pathology and their combinations to minimize the gap between AI-enabled image search methods and clinical settings. The proposed searching framework does not rely on any regional annotation and potentially applies to millions of unlabelled (raw) whole slide images. This paper suggests two approaches for combining magnification levels and compares their performance. The first approach obtains a single-vector deep feature representation for a digital slide, whereas the second approach works with a multi-vector deep feature representation. We report the search results of 20×, 10×, and 5× magnifications and their combinations on a subset of The Cancer Genome Atlas (TCGA) repository. The experiments verify that cell-level information at the highest magnification is essential for searching for diagnostic purposes. In contrast, low-magnification information may improve this assessment depending on the tumor type. Our multi-magnification approach achieved up to 11% F1-score improvement in searching among the urinary tract and brain tumor subtypes compared to the single-magnification image search.**

Manuscript received 8 June 2021; revised 16 February 2022; accepted 3 May 2022. Date of publication 10 June 2022; date of current version 9 September 2022. This work was supported by ORF-RE grant from the Ontario Government, Canada. *(Corresponding author: Hamid R. Tizhoosh.)*

Maral Rasoolijaberi, Abtin Riasatian, and Parsa Ashrafi are with the Kimia Laboratory, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: maral.rasoolijaberi@uwaterloo.ca; abtin.riasatian@uwaterloo.ca; pashrafifashi@uwaterloo.ca).

Morteza Babaei and Sobhan Hemati are with the Kimia Laboratory, University of Waterloo, Waterloo, ON N2L 3G1, Canada, and also with the Vector Institute, MaRS Centre, Toronto, ON M5G 1M1, Canada (e-mail: mbabaie@uwaterloo.ca; sobhan.hemati@uwaterloo.ca).

Ricardo Gonzalez is with the Kimia Laboratory, University of Waterloo, Waterloo, ON N2L 3G1, Canada, and also with the Faculty of Health Sciences, McMaster University, Hamilton, ON L8N 3Z5, Canada (e-mail: gonzar5@mcmaster.ca).

Hamid R. Tizhoosh is with the Kimia Laboratory, University of Waterloo, Waterloo, ON N2L 3G1, Canada, also with the Vector Institute, MaRS Centre, Toronto, ON M5G 1M1, Canada, and also with the Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN 14602 USA (e-mail: tizhoosh.hamid@mayo.edu).

Digital Object Identifier 10.1109/JBHI.2022.3181531

***Index Terms*—Digital pathology, image Search, multi-magnification, CBIR.**

I. Introduction

IMAGE search and retrieval in digital pathology can assist pathologists and medical professionals in different diagnostic, research, and educational tasks. Experts users can search through thousands of digital tissue slides through databases, find the cases most similar to a query pathology slide, compare current tissue samples with past patients, and recommend well-informed diagnoses and treatments. Previous studies [1]–[6] primarily defined content-based image search as a technology where the search input (query) is a digital image, not a textual description. A simple text-based search engine may use only some keywords to find the most relevant cases, and might use a retrieval function for estimating the relevance between the query and candidate outputs [7]. However, a content-based image retrieval (CBIR) obtains valuable visual information in digital pathology slides. In CBIR, the output is determined based on the "content" of images, e.g., tissue morphology and cell nuclei distribution. As clinical pathology documents, reports, and other metadata usually apply to the entire digital whole-slide image (WSI) rather than specific locations within the WSI, the content-based WSI search can retrieve invaluable information associated with similar cases. In this fashion, an AI-driven content-based WSI search engine can process a new WSI before any other assessments. Then, this content-based WSI search engine can find and present the most similar cases to this new patient in the dataset, along with additional metadata including previous diagnoses and treatments. A multi-magnification image search engine allows pathologists to benefit from a content-based image search method that actively exploits different magnifications for better search performance.

Pathologists usually use various lenses of a microscope in their detailed inspection of a tissue sample, switching between different magnification as needed. They usually start with low magnifications to identify regions of interest. Then, they look for diagnostically relevant regions to make preliminary diagnoses. Higher magnifications are often needed to confirm or rule out those diagnoses [8]–[10]. The tradition of utilizing the highest magnification power for many diagnosis tasks may be attributed to the fact that in several tumors, a high mitotic rate and atypical mitosis indicate malignancy [11].

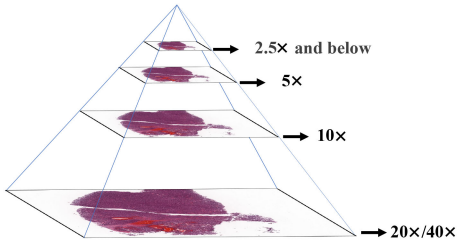


Fig. 1. Pyramidal structures of a whole slide image (WSI) in digital pathology. The base magnification of a WSI is usually $20\times$ or $40\times$.

As a logical expectation, an AI framework should emulate what expert pathologists do. Ultimately, the closer an AI model follows the pathologist's routines, the more realistic and reliable its results may become. Hence, we investigated content-based search encompassing multiple magnification levels to parallel what pathologists do. The preeminent motivation of this research study is to develop the feature of adjustable digital magnification power for a search engine because, conventionally, light microscopy in pathology usually comes with interchangeable several objective lenses.

Digital pathology slides are multi-resolution images. They usually consist of three to five different magnification levels since a typical light microscope conventionally has around four objective lenses. However, the number of slide levels can vary depending on the scanning protocol, and device [12]. Fig. 1 displays the pyramidal data structure of a sample whole slide image (WSI). As it can be seen, the resolution of slide levels, i.e., the number of pixels representing a specimen, increases with the magnification level.

Even though distinct magnification levels are essential for various assessments, and despite the availability of several magnification levels, current methods in computational pathology mainly evaluate digital slides at a single magnification level, most often at $20\times$. Several morphological characteristics can be recognized at low magnifications, including the spatial distribution of normal and abnormal tissue components, tumor growth patterns, and heterogeneity. There are several examples of the low-magnification applications; first, identifying reactive epithelial alterations from dysplasia [13], secondly, estimating the extent of histological patterns for classification (e.g., in pure special type breast carcinomas [14]), and lastly, grading (e.g., in prostatic adenocarcinomas [15]). Hence, exploiting several magnification levels is valuable for different pathology tasks.

A histopathology search tool that can perform at various magnification levels is functionally valuable due to several reasons. First, selecting the right magnification level for digital slide evaluation depends on the type of histopathology assessment. For instance, microinvasions are detectable at the highest magnification [16], whereas distinguishing some well-differentiated malignant tumors from benign tumors or non-tumoral lesions is challenging at high magnifications [17], [18]. Also, each magnification level contains some pieces of histopathology information. At lower magnification levels, the contextual information of glands and tissues is exploited, e.g., architectural patterns, whereas, at higher magnification, cellular contents such as nuclei and cytoplasm are more distinctive.

This research aims to develop adjustable digital magnification power for a WSI search framework in digital pathology by proposing two methods to combine magnification information. Moreover, this paper investigates the effect of magnification on content-based WSI search by comparing the performance of two multi-magnification image search methods at $5\times$, $10\times$, $20\times$ magnifications, and all of their possible combinations.

The rest of this paper is structured as follows. Section II reviews the literature related to either multi-magnification or content-based image search in digital pathology. Section III explains methods and procedures to process and represent WSIs at multiple magnifications for content-based search. Section IV begins by describing the dataset and then reports multi-magnification search experiments. Section V interprets experimental results and discusses findings from the investigation on the effect of magnification on image search. Finally, a summary and conclusions are presented in Section VI.

II. Related Work

A. Multi-Magnification Investigation

Early studies on utilizing multiple magnification levels for digital pathology, before the emerging of deep learning in the computer vision field, are, among others, based on wavelet approaches [19], [20]. Deep-Hipo [21] is among recent multi-magnification studies that combined high and low magnification information to locate cancerous regions. Having multi-magnification receptive fields, Deep-Hipo takes two concentric patches at $20\times$ and $5\times$ magnification and concatenated their features to compute the cancer probability of the central pixel.

Much of the recent literature on histopathology image analysis based on multiple magnification levels pays particular attention to semantic segmentation tasks [22]–[25]. HookNet is a multi-magnification model for histopathology tissue segmentation [23]. Having collected two manually annotated datasets of breast and lung tissues, Van Rijthoven *et al.* trained and evaluated their proposed HookNet model. The model architecture of HookNet included multiple branches to combine contextual information with fine details detectable only at the highest resolution. This network processed a pair of concentric patches at high and low magnification levels to create a segmented map of its input digital whole-slide. Results indicate that multi-magnification models are generally more accurate than single magnification models. Similarly, [24] adopted a multi-magnification approach and proposed deep multi-magnification networks (DMMNs). DMMNs approached the semantic segmentation task in breast tissues via processing sets of patches from $20\times$, $10\times$, and $5\times$ magnifications. The architecture of DMMNs is composed of various combinations of multiple encoders, decoders, and concatenation blocks. In the decoder part, feature maps of $5\times$ and $10\times$ are concatenated with $20\times$ feature maps to recover local and contextual information. In other words, DMMNs enriched feature maps of the high magnification level with feature maps of lower magnification levels to acquire better spatial characteristics in the segmented image. Concerning the dataset,

a pathologist partially annotated 38 breast digital whole slides and classified regions into six tissue subtypes to prepare a private dataset for the DMMNs study. The authors conclude that their proposed deep multi-magnification network outperforms single-magnification counterparts. In another relevant research paper, two multi-magnification networks were presented to segment WSIs using a U-Net-based network architecture [25]. In this method, feature maps of lower resolutions were combined with feature maps of higher resolutions to create accurate binary masks. They used the CAMELYON16 dataset to train and test their proposed networks.

Many other papers in digital pathology have also hinted at combining context and detail information and investigation on multiple magnification levels, including cervical cell segmentation with a multi-scale CNN model [26], detecting regions of interests based on a super-pixel algorithm [27], segmentation of the glands in the colon-rectal digital slides [28], cell classification using a CNN model with multi-scale input and multi-feature layers [29], urothelial carcinoma classification at multi-magnifications using a pre-trained network [30], and cancer subtype classification with a multi-instance CNN [31].

B. Image Search in Digital Pathology

There is a considerable literature on content-based image search in digital pathology ([1], [32]–[42]). *SMILY* [39], allows a user to select a region of interest to obtain matches. A pre-trained network condenses an input image into a feature vector. The network architecture of the SMILY is a deep ranking network that was pre-trained on the 5,000,000 natural images from 18,000 distinct classes. This network learned to extract discriminative features by computing and comparing the embeddings of input images. To evaluate the search performance in finding patches with the same histologic features, SMILY adopted a dataset manually annotated by pathologists. Top-5 scores have been reported for patch-based searches at $40\times$, $20\times$, $10\times$, and $5\times$ magnification levels. In another experiment, pathologists compared the search results of SMILY to random patches.

In another recent paper, [40] introduced a search engine, named *Yottixel*, for real-time whole slide image retrieval in the digital pathology. The authors used an unsupervised color-based clustering method to extract a set of images at $20\times$ magnification from each WSI, then fed images to DenseNet [43] to extract features. Next, the feature vectors were barcoded, i.e., binarized, to create a *Bunch of Barcodes* for indexing digital slides. The barcoding of gigapixel whole-slide images enables Yottixel to perform millions of searches in real-time. KimiaNet [41] is another recent research study that reports image representation for search in digital pathology. In this study, the DenseNet topology was re-trained at several stages by histopathology images extracted at $20\times$ magnification based on a high-cellularity patch selection approach. KimiaNet was tested for image search on three public histopathology datasets for multi-organ whole slide image search.

III. Multi-Magnification Search (MMS)

This section explains the methods employed in our content-based search framework to compare digital slides at multiple magnifications. The framework is based on representations provided by a deep convolutional neural network to find similar digital slides via feature matching.

In the following subsections, the image search framework's data preparation and deep feature extraction steps are described, respectively. Next, Section III-C and Section III-D explain two searching methods based on those deep features. Both searching methods can perform at a single magnification and also any combination of magnifications. Concerning multi-magnification image search, the first method obtains a single multi-magnification feature vector of a WSI without demanding any pixel-level or regional annotation. The primary purpose of the feature vector representation for digital slides is to find the most similar slides to a query slide based on the selected magnification level(s). The second method performs on a multi-feature vector basis and compares patch feature vectors to find similar slides at each magnification, then takes a majority vote among magnifications.

A. Data Preparation

Since WSIs are gigapixel files and too large to be processed by typical CNNs directly, a patch-based approach is usually employed. Therefore, this paper also adopts a patching approach to overcome the computational challenges of training a deep CNN. The term "patch" here refers to a rather small sub-image of a WSI with manageable dimensions. In the first step of image search, RGB images with a manageable image size, e.g., 1000×1000 pixels, are extracted from tissue regions of a WSI at various magnification levels, e.g., $5\times$, $10\times$, and $20\times$. The patch size at all magnification levels stays the same. To identify tissue regions and avoid patch extraction from backgrounds and artifacts, binary masks are used. In a binary mask, pixels belonging to tissue regions have a value of one, whereas pixels belonging to background regions have zero value. This paper uses a pre-trained U-Net [44] for generating the binary masks. Next, we divide the binary mask image into grids to select qualified patches mainly containing tissue.

B. Deep Feature Extraction

Pre-trained deep neural networks (DNNs) can learn to extract content-based features from their input image. An example is KimiaNet [41] which is a customized feature extractor for the digital pathology. The architecture of KimiaNet includes four dense blocks with several convolutional and pooling layers and around seven million parameters. Model parameters of KimiaNet are adjusted to derive histological characteristics from input images. This feature extractor was trained at $20\times$ magnification. Therefore, we employ KimiaNet without any modification to extract tissue features from $20\times$ magnification patches. Note that the term deep feature here indicates the output of the latest pooling layer in the KimiaNet model architecture.

The last DenseNet-121 block [43], i.e., 20 percent of KimiaNet layers, were re-trained with $10\times$ and $5\times$ magnification patches to adjust this feature extractor to those magnifications. In this fashion, we have changed the weights of the KimiaNet according to different magnification levels, attempting to imitate how a pathologist changes their microscope's objective lenses. As a result, different sets of parameters for the KimiaNet model

architecture at three magnification levels were obtained. The model architecture and configurations of our feature extractors are the same for all magnification levels. All versions of KimiaNet were trained to classify patches, with a size of $1000 \times 1000 \times 3$, from 30 primary diagnoses.

After training, in the feature extraction phase, the input of KimiaNet is still a histopathology patch with $1000 \times 1000 \times 3$ in dimension. As the KimiaNet output, we extract a feature vector with a size of 1024×1 representing the input patch. Feature vectors generated by KimiaNet contain histological characteristics and distinct features of the input patches as the network has been trained with high-cellularity patches of carcinoma images [41]. Note that most deep learning approaches in digital pathology, including KimiaNet, are applied on patches rather than the entire WSI due to the extremely large image size of WSIs. This paper utilized two methods to represent an entire WSI based on its patch features. Section III-C and Section III-D explain single-vector and multi-vector methods, respectively. WSI representations can be compared to find and retrieve the most similar WSIs.

C. Single-Vector WSI Representation: Median Aggregation

In the context of WSI feature representation, the goal is to obtain a single feature space so that WSIs with similar histological features are close to each other. Up to this point, patches, i.e., square images acquired from tissue regions of a WSI at a specific magnification level, have been fed to KimiaNet to generate deep feature vectors. In other words, each deep feature vector (the output of KimiaNet as the feature extractor) represents a patch, not the entire WSI. In the median aggregation approach, we create a feature space for each magnification level. To do so, we aggregate all patch feature vectors by taking the median value with respect to feature positions to create a single vector representing of the WSI at a specific magnification. The aggregated vectors of different magnifications are concatenated to create a multi-magnification vector representation. As each WSI is represented by one vector in this method, we called it *Single-Vector*. The distance between WSI feature single-vectors, e.g. Euclidean distance, can be computed to compare WSIs. Fig. 3 illustrates the single-vector method for the content-based WSI search.

D. Multi-Vector WSI Representation: Median-of-Mins

The *Multi-Vector* method performs patch-to-patch and then slide-to-slide comparisons to find and retrieve most similar digital slides. The multi-vector method is the same as what is proposed in [40] for single magnification. The procedure is repeated for all magnification levels in an independent manner. Fig. 4 illustrates the three independent search processes for a query WSI at $5\times$, $10\times$, and $20\times$ magnifications. In terms of multi-magnification search, we consider all retrieved WSIs at selected magnification levels for subsequent evaluations. We called this method *Multi-Vector* because it represents a digital slide with multiple feature vectors associated with its patches.

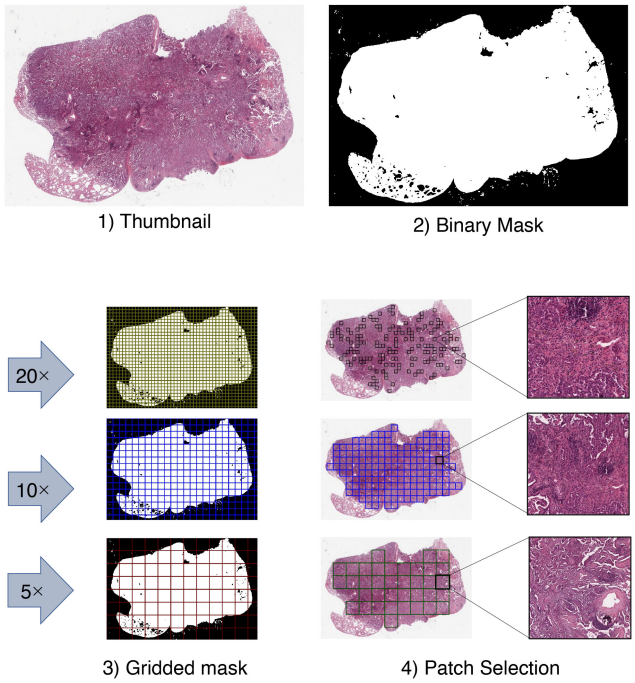


Fig. 2. Patch extraction procedure at $5\times$, $10\times$, and $20\times$ magnification levels. A binary mask is utilized for the tissue specimen detection.

Working with many feature vectors when operating on large WSI archives can be prohibitively time-consuming. The calculation of Hamming distance between binary codes is much faster than calculating Euclidean distance between full-precision feature vectors. In large-scale search problems, reducing the demand for memory resources is another advantage of binary feature vectors. Hence, hashing feature vectors would be required in case of multi-vector WSI representation. Instead of the *Min-Max* barcoding algorithm as proposed before [40], this paper uses the Sequential Non-Rigid Quantization (SNRQ) [45], an unsupervised hashing approach, for barcoding patch feature vectors. The non-rigid quantization (NRQ) method is an extended version of the iterative quantization (ITQ) method [46]. NRQ applies dimensionality reduction and utilizes non-rigid transformations along with rigid transformations such as rotation to further reduce the quantization error caused by hashing. Sequential NRQ (SNRQ) is an efficient implementation of NRQ based on sequential updates.

Let assume the test dataset contains n_p patches from n_s WSIs. First, pairwise Hamming distances between barcoded feature vectors of all patches are calculated, generating a *patch-to-patch-distance* matrix with dimension of $n_p \times n_p$. Note that distances between patches of the same WSI are set to infinity for convenient search purposes. In the next step, the minimum distance from each patch to all patches of each WSI is calculated, generating the *patch-to-slide-distance* matrix of dimension $n_p \times n_s$. Finally, pairwise distances between WSIs are calculated by finding the median of patch-to-slide distances, generating a *slide-to-slide-distance* matrix of dimension $n_s \times n_s$. This *median-of-mins* matching process repeated for all magnification levels.

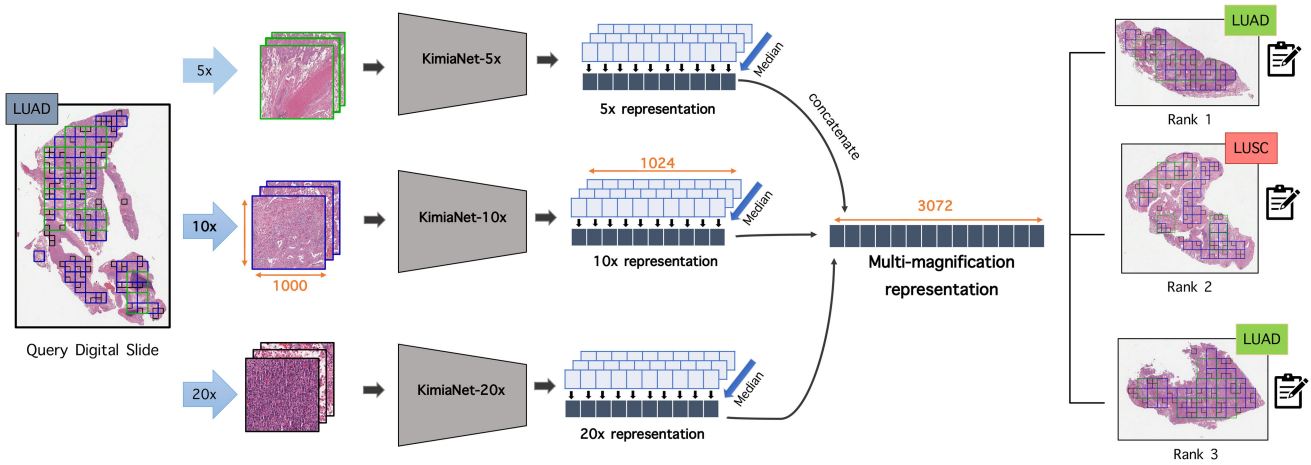


Fig. 3. Searching framework for the single-vector WSI representation method using the median aggregation approach (Section III-C). The pen-and-pad icon stands for metadata such as reports containing primary diagnosis.

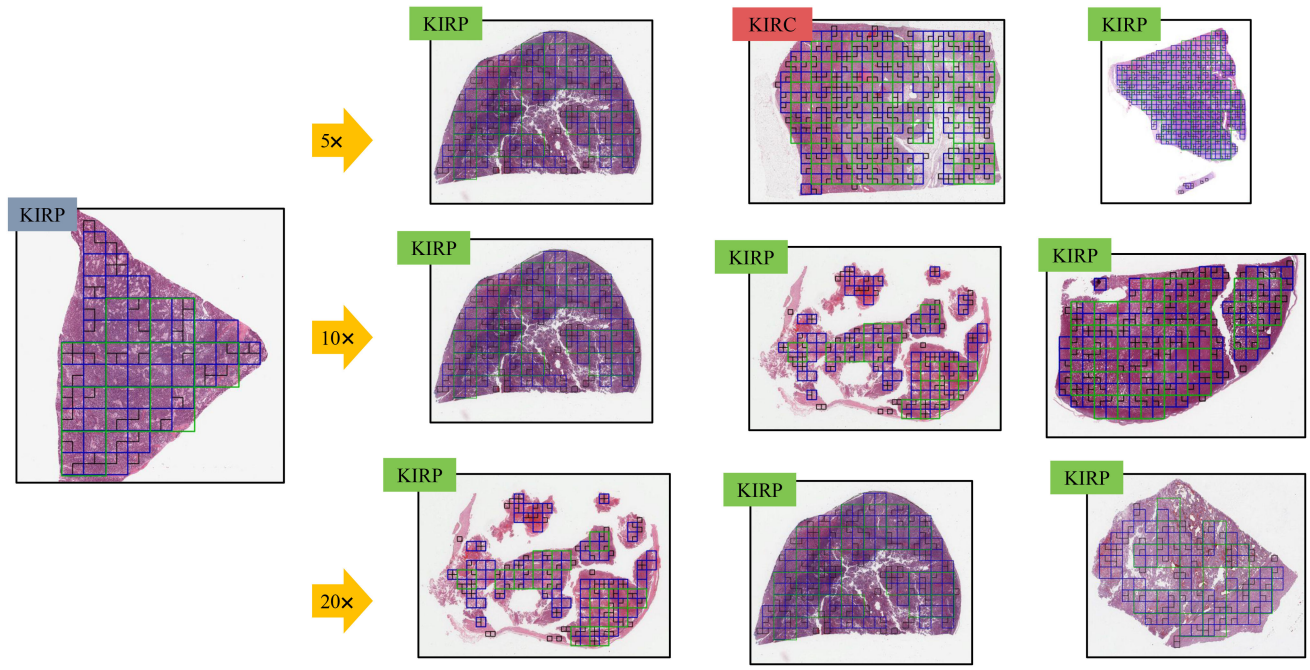


Fig. 4. Image search among urinary tract WSIs using the multi-vector method. The search has been performed three times based on extracted patches at $5\times$, $10\times$, and $20\times$ magnification. The search at a single magnification is successful if at least two of the three search results show correct primary diagnosis. The multi-magnification search (MMS) is successful if the majority of the search results at selected magnifications indicates the same disease type as the query WSI.

IV. Experiments on MMS

A. Image Data

This research study has employed a comprehensive publicly available dataset of hematoxylin and eosin (H&E)-stained digital whole slides images. The Cancer Genome Atlas (TCGA) [47] has more than 30,000 digital slides including frozen sections and diagnostic slides from the NIH's pan-cancer analysis project. As frozen sections generally showing somewhat lower image quality, they were eliminated to maintain a more consistent evaluation [48]. At the outset, diagnostic pathology slides of the TCGA dataset were collected for our investigation. Some of the remaining slides did not include $20\times$ or higher magnification levels. We only included diagnostic slides with $20\times$ or higher levels of magnification, and recorded one slide per patient. Next, these slides were categorized based on morphology codes, primary sites, and diagnosis, again omitting the groups with lower than 20 slides due to search purposes in the test phase. This grouping allowed having a minimum of two WSIs from each class in the test dataset. We finally recorded 8,611 permanent H&E stained digital slides from 12 anatomic sites (categorization based on established literature [48]). The anatomical site of a digital

slide indicates the organ or body system biopsied. Anatomical sites of this dataset consists of brain, endocrine, gastrointestinal tract, gynecological, liver/pancreaticobiliary, melanocytic malignancies, prostate/testis, pulmonary, urinary tract, breast, head and neck, and mesenchymal. Next, following the procedure proposed by the KimiaNet paper [41], we split the dataset with a ratio of roughly 80, 10, and 10 percent into the train, test, and validation sets, respectively. The train-valid-test split was with respect to primary diagnoses. As a result, a test dataset of 744 WSIs, a validation dataset of 741 WSIs, and a training dataset of 7,126 WSIs were collected for our investigation. These digital slides were not annotated by a pathologist at any regional or pixel levels.

Next, concerning each WSI, regions containing tissue specimens were segmented to remove background regions. The term background here refers to non-informative pixels in a WSI, including white-coloured backgrounds of a glass slide and artifacts such as ink marker and extra stain. We used a pre-trained U-Net model [44] to generate binary masks at the lowest magnification level for identifying the background and foreground of a WSI. After recognizing tissue regions (foreground), we extracted high-, medium-, and low-resolution patches with a fixed size of 1000×1000 pixels (Fig. 2). Patches of this experiment were extracted without any supervision or delineation of regions of interest.

At $20\times$ magnification, we utilized KimiaNet-IV [41] as the feature extractor without any modification. Therefore, at $20\times$ magnification, the training and validation patches in this paper were identical to those in the KimiaNet paper [41]. As reported in the KimiaNet paper, a total of 242,202 histopathology $20\times$ magnification patches were extracted from those 7,126 WSIs for training. Note that the KimiaNet paper used an algorithm based on cellularity ratio to select training patches at $20\times$ magnification [41]. Generally, cell-level information might not be well recognizable in low magnifications. Patch sampling, on the other hand, is essential for computational convenience and redundancy reduction in training. Therefore, at $10\times$ magnifications, we sampled training and validation patches in such a way that patches were extracted from all across the tissue. In this sampling, patches were uniformly distributed all over the specimen area to optimise the sampling all tissue types in each WSI. Ultimately, we recorded 190,257 training and 25,653 validation patches to fine-tune KimiaNet at $10\times$ magnification. At $5\times$ magnification, we considered all extracted patches for training because there were a manageable collection of patches at $5\times$ magnification (167,746 training, and 29,680 validation).

In the test phase at $20\times$ magnification, due to a large and unmanageable number of training samples, we sampled 20 percent of the tissue regions of each digital slide using the same procedure with preserving the spatial diversity. Previous studies [1], [40] have shown that a small percentage, usually a range between 5 to 20 percent of patches, can be sufficient for representing a TCGA digital slide. In our study, we collected 91,287 patches at $20\times$ magnification from 744 test WSIs. In terms of test patches at $10\times$ and $5\times$ magnifications, we recorded every patch that mainly contains tissue texture (and not much background). As a result, 97,389 and 20,397 test patches were collected for search experiments at $10\times$ and $5\times$ magnifications, respectively. Fig. 5 shows three test WSIs and the location of their extracted patches.

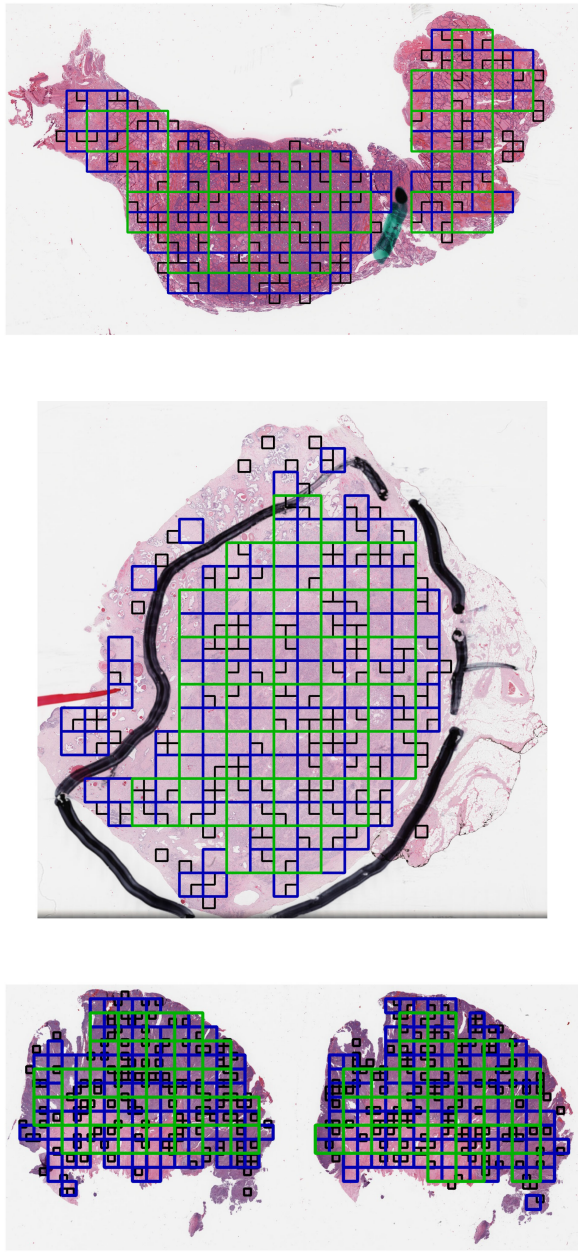


Fig. 5. Visualization of multi-magnification patch selection in three gigapixel WSI samples from different organs. Green, blue, and black bounding boxes indicate patches at $5\times$, $10\times$, and $20\times$ magnification, respectively. All patches are of size 1000×1000 pixels.

B. Assessment of the Search Accuracy

Manual comparing and contrasting WSIs is infeasible and at best highly time-consuming and costly due to the complexity of image search and also the large size of each WSI. A typical WSI in the TCGA dataset has around 4 to 5 slide levels (magnifications), and in highest magnification, the image size can be more than $100,000 \times 100,000$ pixels with countless

TABLE I
3-NEAREST NEIGHBORS ACCURACY (%) FOR THE HORIZONTAL SEARCH AMONG TEST DATASET

Tumor Type	Patient#	KimiaNet	*Single-Vector*		*Multi-Vector*	
			20x	TM	20x	TM
Brain	74	99	99	95	99	100
Breast	91	91	83	82	84	86
Endocrine	72	92	97	95	90	93
Gastro.	88	84	80	76	77	77
Gynaeco.	30	57	85	79	71	74
Head/Neck	32	88	68	81	66	90
Liver/panc.	51	88	83	87	86	90
Melanocytic	28	86	89	86	81	79
Mesenchymal	13	69	62	73	69	75
Prostate/testis	53	96	89	88	92	96
Pulmonary	86	86	80	74	75	76
Urinary tract	123	89	84	81	84	76

details. Moreover, hand-operated assessment of content-based WSI retrieval by pathology experts is prone to variability, and therefore is impractical for a large-scale dataset. Hence, we conducted algorithmic experiments to assess the performance of our multi-magnification search framework.

In these experiments, we used the "*leave-one-patient-out*" approach in the test dataset, i.e., excluding one WSI, finding similar cases in the search space to this WSI, and repeating this process for all WSIs in the test dataset. The search space of the leave-one-out experiments is the test dataset since the training and validation WSIs were already used for training the feature extractor networks. We investigated the top-three most similar WSIs to the query slide where the term query slide indicates that excluded WSI. Next, we examined the majority vote among those three most similar WSIs to the query slide to make a prediction. One may examine the top-n search results for $n > 3$ if larger test datasets are available. Finally, the prediction via majority vote was compared to the actual label, e.g., primary diagnosis or tumor type associated with that query WSI. In other words, these experiments consider a search successful if two out of three matched cases have the same label as the query slide. Taking the majority vote is much more rigorous than the top-n accuracy in computer vision that assumes correctness if at least one of the search results is correct. In this paper, two types of search experiments, horizontal and vertical searching, were performed to systematically assess the performance of our multi-magnification search framework.

The first experiment, horizontal searching, measures how accurate a WSI search method can find WSIs with similar anatomical sites. In horizontal searching, the search space is the entire test dataset. Table I reports the results of horizontal search. This table compares the performance of proposed MMS methods, single-vector and multi-vector, at $20\times$ magnification and Tri-Magnifications (TM) with the horizontal search results reported in KimiaNet paper [41]. Note that the anatomical site (organ) of any WSI is given since the organ of a biopsy is known before any assessment. Therefore, horizontal search is a basic algorithmic validation with limited clinical applications. This paper performed horizontal searches as a fundamental evaluation and sanity check. Further discussion on the horizontal search results are provided in Section V.

As the primary site is a pivotal knowledge about a WSI, the vertical searching experiment limits the search space for each query WSI to WSIs with the same anatomical organ. Therefore, we would be interested in identifying primary diagnosis for each primary site. Accordingly, each query WSI was compared to all other WSIs of the same primary site, with the intention of finding cases with the same subtype of malignancy. In other words, the search space in vertical searching is limited to one anatomical site at each time. Note that among the 12 anatomical sites, the classification (majority voting) experiments concern 9 anatomical sites with more than one subtype. The tumour subtypes, i.e., primary diagnoses, and the number of WSIs associated with each tumour subtype are presented in Table II. For instance, there are 34, 11, 50, 28 WSIs in our test dataset with a primary diagnosis of BLCA, KICH, KIRC, KIRP, respectively, which all can be considered as a *urinary tract* tumor.

Since the vertical search has more clinical value than horizontal searches, this paper thoroughly investigated the vertical search experiment at three magnifications and all their combinations. The results of vertical search experiments are reported in Table II and Fig. 6. Seven experiments at 1) $5\times$, 2) $10\times$, 3) $20\times$, 4) $5 \times 10\times$, 5) $5 \times 20\times$, 6) $10 \times 20\times$, and 7) TriMagnification (TM), i.e., $5 \times 10 \times 20\times$ magnifications, using single-vector and multi-vector methods, were conducted independently on nine anatomical sites. Table II reports the F1-score for all tumor subtypes in nine different anatomical sites. F1-score represents both specificity and sensitivity of the search. Fig. 6(a) and (b) show the overall accuracy achieved at each magnification using the *single-vector* method and the *multi-vector*, respectively.

V. ANALYSIS AND DISCUSSION

This section interprets the WSI search results and describes our findings. First, the results of the horizontal search, as a preliminary evaluation with limited clinical applications, are briefly discussed. Next, we thoroughly analyze the vertical search results to investigate the effect of magnification in content-based WSI search. The goal is to suggest the best search method, and above that, to suggest the most effective magnification level (or combination of levels) relevant to each anatomical site.

As Table II shows, our WSI search framework could accurately find and retrieve WSIs with a similar tumor type in the majority of cases, and our results are in line with previous studies in the literature. The base of our multi-vector method is the median-of-min approach which makes it somehow similar to the method used in KimiaNet especially at single magnification ($20\times$). However, there are two significant differences between multi-vector at $20\times$ magnification and the method used in KimiaNet. First, the multi-vector method used the SNRQ hashing technique to convert each deep feature into 128 bits, while the KimiaNet paper used the *min-max* hashing approach to convert each deep feature into 1023 bits. Plus, our patching approach is different from KimiaNet in some respects.

TABLE II
PERCENTAGE OF F1-SCORE VERTICAL SEARCH RESULTS

			single-vector							*multi-vector*						
Site	Subtype	#samples	20×	10×	5×	5 × 10×	5 × 20×	10 × 20×	*TM*	20×	10×	5×	5 × 10×	5 × 20×	10 × 20×	*TM*
Brain	GBM	35	83	80	78	76	83	83	89	84	79	59	74	84	82	82
Brain	LGG	39	78	75	78	74	78	78	87	83	79	66	76	83	82	82
Endocrine	ACC	6	50	0	0	0	29	29	0	0	0	20	0	0	0	0
Endocrine	PCPG	15	82	72	67	64	85	85	79	85	77	69	77	77	88	85
Endocrine	THCA	51	98	90	93	94	98	98	98	96	95	95	95	96	98	97
Gastro.	COAD	32	65	66	68	61	66	66	63	73	63	74	73	77	73	71
Gastro.	ESCA	14	64	57	56	60	64	64	57	61	55	53	53	55	55	55
Gastro.	READ	12	40	30	17	17	40	38	32	29	22	38	30	33	32	32
Gastro.	STAD	30	79	61	65	64	78	76	77	76	62	54	66	76	72	71
Gynaeco.	CESC	17	92	74	80	75	89	92	89	91	86	87	85	89	89	89
Gynaeco.	OV	10	89	67	84	76	82	89	89	82	75	82	78	71	71	75
Gynaeco.	UCS	3	80	0	0	0	80	80	50	75	80	80	80	86	86	67
Liver/panc.	CHOL	4	55	29	25	0	55	60	60	22	44	20	25	20	36	25
Liver/panc.	LIHC	35	96	88	99	97	97	96	96	92	88	88	91	93	90	91
Liver/panc.	PAAD	12	80	69	67	79	86	86	87	67	72	50	67	78	73	67
Melanocytic	SKCM	24	92	92	94	94	92	92	92	94	92	89	92	94	92	92
Melanocytic	UVM	4	0	0	40	40	0	0	0	40	0	29	33	40	0	0
Prostate/testis	PRAD	40	99	96	95	95	99	99	99	99	96	95	98	98	99	98
Prostate/testis	TGCT	13	96	89	82	82	96	96	96	96	89	85	92	92	96	92
Pulmonary	LUAD	38	78	64	56	48	78	78	75	68	72	59	60	65	74	66
Pulmonary	LUSC	43	82	69	67	60	82	82	80	76	77	71	72	72	80	76
Pulmonary	MESO	5	67	29	0	0	67	67	50	80	0	0	0	29	67	29
Urinary tract	BLCA	34	94	91	90	88	94	94	99	94	96	91	96	96	99	99
Urinary tract	KICH	11	90	67	62	59	90	90	86	91	87	47	76	91	95	95
Urinary tract	KIRC	50	93	87	85	83	93	93	96	93	93	85	90	93	94	92
Urinary tract	KIRP	28	83	64	65	57	83	83	93	82	82	61	80	84	93	86

"*TM*" Stands for Tri-Magnification, Indicates Analysis Based on Three Magnification Levels. See Table III for Subtype Abbreviation Codes.

Vertical search results are presented in Table II and Fig. 6. We analyze the results to recommend the most effective magnification level(s) as well as the optimal vertical search method for searching among each anatomical site.

1) Brain: Using TriMagnification (TM) instead of 20× magnification improved the single-vector method performance, increasing the F1-score for Glioblastoma Multiforme (GBM) and Low-grade Gliomas (LGG) from 0.83 and 0.78 to 0.89 and 0.87, respectively. Results indicate that the single-vector algorithm at TM may be the most appropriate configuration for searching among brain tumors.

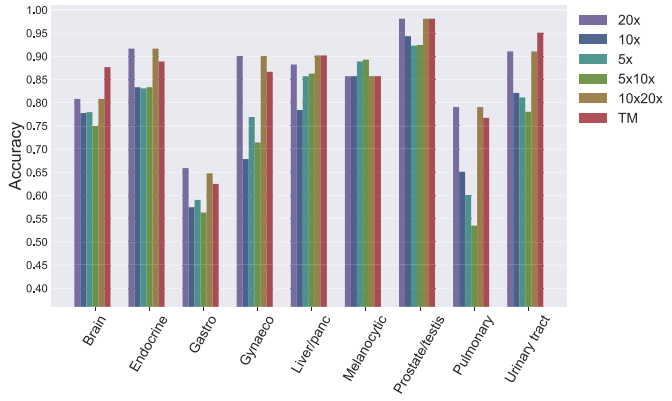
2) Endocrine: The class imbalance challenge in this dataset is noticeable. Adrenocortical carcinoma (ACC), pheochromocytoma and paraganglioma (PCPG), and thyroid carcinoma (THCA) have 6, 15, and 51 samples, respectively. This disparity in the number of subtype samples markedly affects the search performance. There are not enough samples of ACC tumors, causing confusion in finding two or more similar cases and resulting in zero F1-scores in the evaluation using some magnifications/methods. Overall, results indicate that utilizing 10× magnification along with 20× magnification has been beneficial for the multi-vector method, improving the F1-score accuracy of PCPG and THCA by three and two percent, respectively, compared to 20× magnification. These results support using the single-vector method at 20× magnification or at the 10 × 20× magnifications.

3) Gastrointestinal Tract: The single-vector method at 20× magnification achieved the most accurate results concerning the image search among gastrointestinal tract tumors. However, the multi-vector method at 5 × 20× magnification surprisingly showed 12 percent F1-score improvement compared to the previously suggested configuration for searching for colon adenocarcinoma (COAD) cases.

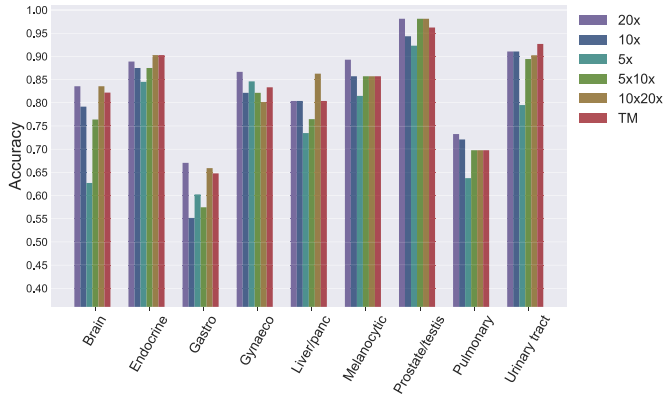
4) Gynecological Tumors: The limited number of samples for uterine carcinosarcoma tumors (UCS) negatively affected the performance of the single-vector method at 10× and 5× magnifications and their combination. Interestingly, the multi-vector method showed satisfactory performance across all magnifications. However, more samples are required for a more reliable assessment. Overall, the single-vector method at 20× magnification is the best configuration for diagnosis-related searches among gynecological WSIs.

5) Liver, Pancreaticobiliary: The single-vector method at TM achieved the most accurate results. Interestingly, single-vector method 5× achieved the F1-score of 0.99 concerning liver hepatocellular carcinoma (LIHC).

6) Melanocytic Malignancies: Since there are only four samples of uveal melanoma (UVM) tumors, only three samples were left in the *leave-one-out* approach. Therefore, finding two or more UVM matches within three retrieved samples can be considered slightly unlikely. Overall, the single-vector method at 5×, 5 × 10×, and the multi-vector method at 20× and 5 × 20× magnifications achieved the most accurate results. Since the computational complexity using the single-vector method at 5× magnification is less than other configurations, representing

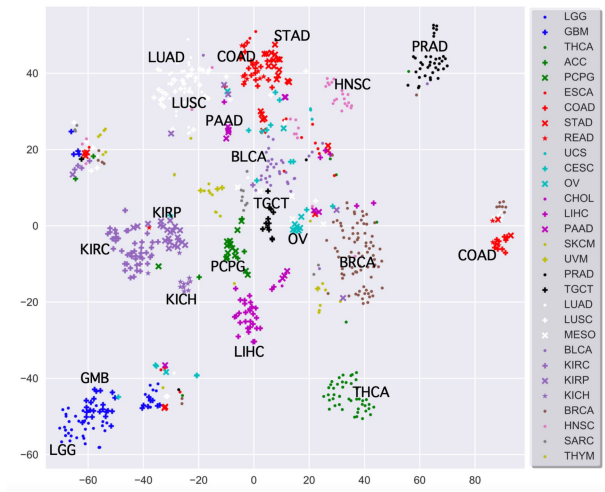


(a) Single-Vector (*Median Aggregation*)

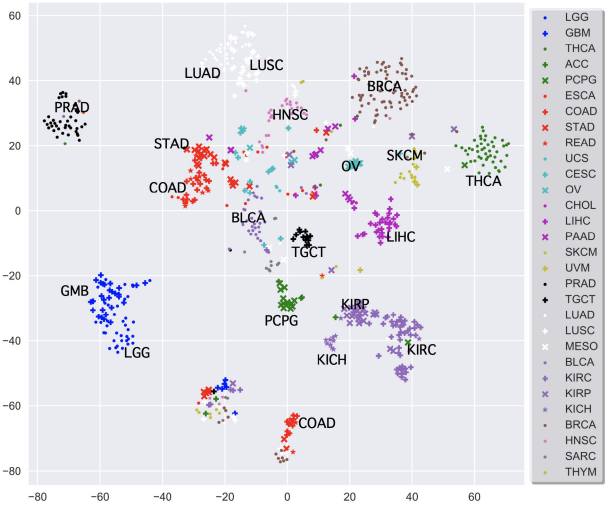


(b) Multi-Vector method (*median-of-mins*)

Fig. 6. Accuracy histogram for vertical search experiments with respect to the magnification level and anatomical site. The vertical axis shows fraction of correct subtype classifications. For each anatomical site, the total image search accuracy at different magnification levels is visualized.



(a) TriMagnification (TM)



(b) 20× magnification

Fig. 7. T-SNE visualizations of digital slide embeddings (Single-Vector method). Each point displays a feature vector associated with one digital slide in the test dataset. Each color is associated with one anatomical site. Tumor types of the same anatomical are indicated by different markers, e.g., plus, x, point, and star. T-SNE visualizations show a clear class discrimination of single-vector representations.

WSIs with melanocytic malignancies using the single-vector method at 5× magnification is preferred.

7) Prostate: Image search at 20× achieved accurate results with 0.99 and 0.96 F1-score for prostate adenocarcinoma (PRAD) and testicular germ cell tumors (TGCT), respectively. These results are regardless of the searching method. Combining lower magnification levels information with 20× magnification information using the single-vector method also achieved similar results as for the 20× magnification.

8) Pulmonary: Top results concerning lung tumors, i.e., Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC), were achieved using the single-vector method at 20× magnification, and also the combination of 20× magnification with one lower magnification, i.e., 10 × 20× and 5 × 20×. Concerning Mesothelioma (MESO) diseases, the multi-vector method showed an unanticipated improvement in the F1-score compared to the single-vector method at 20× magnification, improving from 67 to 80 percent.

9) Urinary Tract: The most apparent advantage of utilizing multi-magnification can be seen in searching among urinary tract tumor cases. Both MMS methods achieved higher-accuracy results for most urinary tract tumor cases compared to single magnification counterparts. For instance, 10 and 4 percent F1-accuracy improvement in the single-vector method and the multi-vector method were observed for KIRP, respectively. These results indicate that employing low and medium magnification along with high magnification can improve the assessment of urinary tract tumors.

Moreover, the following important findings were observed:

- The results confirmed that the high magnification information played an essential role in diagnosis-based evaluations. The searching at 20× magnification or its combination with other magnifications outperformed searching at other magnification levels. As seen in Table II and

Fig. 6, searching at $20\times$ magnification, the combination of $10 \times 20\times$ magnifications, or TriMagnifications (TM) reported the highest accuracy results in most anatomical sites. The general trend of successful searches is also visible among searching at the mentioned magnifications regardless of the searching methods. In 7 out of 9 anatomical sites, both search methods achieved more than 80 percent classification accuracy at the mentioned magnifications (Fig. 6). The $20\times$ magnification (both searching methods) acquired above 80 percent F1-score accuracy in the classification of 13 tumor subtypes. This value for $10 \times 20\times$ and TM is 12 and 13 subtypes, respectively. Considering a lower threshold of 65 percent, both search methods could achieve an F1-score above this threshold tumor subtypes with at $20\times$, $10 \times 20\times$, and TM is 20, 21, and 18 (out of 26 primary diagnosis types), respectively. It must be pointed out that the highest magnification is usually used by pathologists for confirming a diagnosis [8]–[11].

- Many of the "primary diagnosis" labels obtained from the TCGA (see Table III) referred to groups of tumors with different morphologies (e.g., LGG, THCA, ESCA, CESC, OV, CHOL, SKCM, PRAD, and TGCT). It is possible that multiple magnifications improved the assessment of urinary tract tumors because only specific types of tumors were included in this anatomical site (i.e., BLCA, KICH, KIRC, and KIRP). However, this improvement was not seen for the specific types of tumors included in "Pulmonary".
- Using the single-vector method for the brain, liver/pancreas, and urinary tract tumor cases, TM achieved the highest overall classification accuracy (Fig. 6(a)), and the highest F1-score in most subtypes (Table II) in comparison with all other magnification settings. Results show that exploiting single-vector method at TriMagnification (TM) for representing WSIs can improve morphology assessments. Accordingly, if one of the anatomical sites mentioned earlier is a diagnostic consideration, evaluation based on multiple magnification levels using the single-vector method is highly recommended.
- Concerning the multi-vector method, $20\times$ magnification has achieved slightly higher accuracy compared to the TriMagnification (TM) in many cases. This method at TriMagnification votes between retrieved WSIs of single-magnifications. As a potential solution to improve the accuracy of TM, a weighted majority voting approach may emphasize the $20\times$ magnification. The suggested approach may be more successful with trainable weights.
- In comparison between methods at TriMagnification (TM), the single-vector method tends to be more successful in combining all magnifications than the multi-vector method. That means the single-vector WSI representation method performed more accurately than the multi-vector method at TM in finding cases with the same tumor subtypes in most anatomical sites. The reported search accuracy associated with the brain, gynecological, liver/pancreaticobiliary, prostate/testis, and urinary tract tumors supports this finding. Another advantages of the single-vector method is less computational complexity, more efficient storage, and faster searching.
- Our results provide further support for the hypothesis that there is a direct relationship between the number of WSIs in the test dataset and the search accuracy [1]. ACC, USC, CHOL, UVM, and MESO are tumor subtypes with less than six samples in the test dataset. Due to the limited number of samples, the search framework was unsuccessful in finding two or more similar WSIs at some magnifications. Besides, top accuracy results are

TABLE III
TCGA CANCER SUBTYPES [48]

TCGA code	Primary Diagnosis
ACC	Adrenocortical Carcinoma
BLCA	Bladder Urothelial Carcinoma
BRCA	Breast Invasive Carcinoma
CESC	Cervical Squamous Cell Carcinoma and Endocervical Adenoc.
CHOL	Cholangiocarcinoma
COAD	Colon Adenocarcinoma
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
ESCA	Esophageal Carcinoma
GBM	Glioblastoma Multiforme
HNSC	Head and Neck Squamous Cell Carcinoma
KICH	Kidney Chromophobe
KIRC	Kidney Renal Clear Cell Carcinoma
KIRP	Kidney Renal Papillary Cell Carcinoma
LGG	Brain Lower Grade Glioma
LIHC	Liver Hepatocellular Carcinoma
LUAD	Lung Adenocarcinoma
LUSC	Lung Squamous Cell Carcinoma
MESO	Mesothelioma
OV	Ovarian Serous Cystadenocarcinoma
PAAD	Pancreatic Adenocarcinoma
PCPG	Pheochromocytoma and Paraganglioma
PRAD	Prostate Adenocarcinoma
READ	Rectum Adenocarcinoma
SARC	Sarcoma
SKCM	Skin Cutaneous Melanoma
STAD	Stomach Adenocarcinoma
TGCT	Testicular Germ Cell Tumors
THCA	Thyroid Carcinoma
THYM	Thymoma
UCEC	Uterine Corpus Endometrial Carcinoma
UCS	Uterine Carcinosarcoma
UVM	Uveal Melanoma

associated with BLCA, PRAD, and THCA, with 34, 40, and 51 samples, respectively.

VI. Summary and Conclusion

We proposed a novel multi-magnification search tool to find similar digital slides without demanding any user annotation. The magnification level can be selected in our content-based search framework to find similar tumor types and type of malignancy more accurately. This multi-magnification search framework works based on the deep feature vectors and is empowered with two independent search methods to measure content-based similarities between WSIs.

Our main contribution was to investigate three magnification levels and their combinations to retrieve similar WSIs. Also, we proposed two content-based image searching methods for multi-magnification searching. A subset of the TCGA dataset with 12 anatomical sites and 30 tumor subtypes was employed for our evaluations. Patches were collected on a unsupervised basis – neither annotation provided by pathologists nor segmentation of tissue types were used.

We achieved a significant accuracy improvement for diagnosis tasks concerning kidney tumors, e.g., F1-score of 0.93 for Kidney renal papillary cell carcinoma (KIRP) using the TriMagnification search compared to 0.83 F1-score using the single magnification ($20\times$) counterpart. The results confirmed that enriching high magnification search information with low magnification search information is a promising way to increase search accuracy in some cases. Also, the results showed that the highest magnification, containing the cell-level information and detailed morphological features, is an essential resolution for diagnostic tasks (the most accurate results were achieved using the $20\times$ magnification or a combination of $20\times$ magnification with other magnifications).

In conclusion, the investigation of various magnification levels in digital slide search has shown promising improvements in many cases. These experiments implied that depending on the tumor type, we need different magnifications and combinations of magnifications for an accurate and reliable search and classification. Therefore, magnification level selection is a valuable option for a digital pathology search framework.

References

[1] S. Kalra *et al.*, "Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–15, 2020.

[2] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications-clinical benefits and future directions," *Int. J. Med. Informat.*, vol. 73, no. 1, pp. 1–23, 2004.

[3] M. Jain and D. Singh, "A survey on CBIR on the basis of different feature descriptor," *J. Adv. Math. Comput. Sci.*, vol. 14, no. 6, pp. 1–13, 2016.

[4] T. M. Lehmann *et al.*, "Content-based image retrieval in medical applications," *Methods Inf. Med.*, vol. 43, no. 4, pp. 354–361, 2004.

[5] L. R. Long, S. Antani, T. M. Deserno, and G. R. Thoma, "Content-based image retrieval in medicine: Retrospective assessment, state of the art, and future directions," *Int. J. Healthcare Inf. Syst. Informat.*, vol. 4, no. 1, pp. 1–16, 2009.

[6] D. Markonis *et al.*, "A survey on visual information search behavior and requirements of radiologists," *Methods Inf. Med.*, vol. 51, no. 6, pp. 539–548, 2012.

[7] O. Kolomiyets and M.-F. Moens, "A survey on question answering technology from an information retrieval perspective," *Inf. Sci.*, vol. 181, no. 24, pp. 5412–5434, 2011.

[8] T. Jaarsma, H. Jarodzka, M. Nap, J. J. van Merriënboer, and H. P. Boshuizen, "Expertise in clinical pathology: Combining the visual and cognitive perspective," *Adv. Health Sci. Educ.*, vol. 20, no. 4, pp. 1089–1106, 2015.

[9] T. T. Brunye, P. A. Carney, K. H. Allison, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "Eye movements as an index of pathologist visual expertise: A pilot study," *PLoS One*, vol. 9, no. 8, 2014, Art. no. e103447.

[10] D. Treanor, C. H. Lim, D. Magee, A. Bulpitt, and P. Quirke, "Tracking with virtual slides: A tool to study diagnostic error in histopathology," *Histopathol.*, vol. 55, no. 1, pp. 37–45, 2009.

[11] D. W. Molavi, *The Practice of Surgical Pathology: A Beginner's Guide to the Diagnostic Process*. Berlin, Germany: Springer, 2017.

[12] C.L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: A survey," *Med. Image Anal.*, vol. 67, 2020, Art. no. 101813.

[13] N. A. Sangle, S. L. Taylor, M. J. Emond, M. Depot, B. F. Overholt, and M. P. Bronner, "Overdiagnosis of high-grade dysplasia in Barrett's esophagus: A multicenter, international study," *Modern Pathol.*, vol. 28, no. 6, pp. 758–765, 2015.

[14] E. Rakha *et al.*, "Invasive breast carcinoma: General overview," in *Breast Tumours WHO Classification of Tumours*, 5th Ed., Lyon, France: World Health Organization, 2019, pp. 82–101.

[15] R. Montironi *et al.*, "Narrative review of prostate cancer grading systems: Will the Gleason scores be replaced by the Grade groups?," *Transl. Androl. Urol.*, vol. 10, no. 3, pp. 1530–1540, 2021.

[16] P. Fitzgibbons *et al.*, "Protocol for the examination of resection specimens from patients with invasive carcinoma of the breast," *J. Breast Cancer*, vol. 24, no. 1, pp. 1–21, 2021.

[17] A. Mathur, M. T. Olson, and M. A. Zeiger, "Follicular lesions of the thyroid," *Surg. Clin.*, vol. 94, no. 3, pp. 499–513, 2014.

[18] A. Chaux and A. L. Cubilla, "Diagnostic problems in precancerous lesions and invasive carcinomas of the penis," *Seminars Diagn. Pathol.*, vol. 29, no. 2, pp. 72–82, May 2012.

[19] J. Z. Wang, J. Nguyen, K.-k. Lo, C. Law, and D. Regula, "Multiresolution browsing of pathology images using wavelets," in *Proc. AMIA Symp.*, 1999, pp. 430–434.

[20] J. Z. Wang, "Pathfinder: Multiresolution region-based searching of pathology images using IRM," in *Proc. AMIA Symp.*, 2000, pp. 883–887.

[21] S. C. Kosaraju, J. Hao, H. M. Koh, and M. Kang, "Deep-Hipo: Multi-scale receptive field deep learning for histopathological image analysis," *Methods*, vol. 179, pp. 3–13, 2020.

[22] H. Tokunaga, Y. Teramoto, A. Yoshizawa, and R. Bise, "Adaptive weighting multi-field-of-view CNN for semantic segmentation in pathology," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12597–12606.

[23] M. van Rijthoven, M. Balkenhol, K. Siliņa, J. van der Laak, and F. Ciompi, "Hooknet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images," *Med. Image Anal.*, vol. 68, 2021, Art. no. 101890.

[24] D. J. Ho *et al.*, "Deep multi-magnification networks for multi-class breast cancer image segmentation," *Computerized Med. Imag. Graph.*, vol. 88, 2021, Art. no. 101866.

[25] F. Gu, N. Burlutskiy, M. Andersson, and L. K. Wilén, "Multi-resolution networks for semantic segmentation in whole slide images," in *Computational Pathology and Ophthalmic Medical Image Analysis*. Berlin, Germany: Springer, 2018, pp. 11–18.

[26] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 10, pp. 2421–2433, Oct. 2015.

[27] B. E. Bejnordi, G. Litjens, M. Hermsen, N. Karssemeijer, and J. A. van der Laak, "A multi-scale superpixel classification approach to the detection of regions of interest in whole slide histopathology images," *Medical Imaging 2015: Digital Pathology*, vol. 9420, 2015, Art. no. 94200H.

[28] Y. Gao *et al.*, "Multi-scale learning based segmentation of glands in digital colonrectal pathology images," *Medical Imaging 2016: Digital Pathology*, vol. 9791, 2016, Art. no. 97910M.

[29] K. Zormpas-Petridis, H. Failmezger, I. Roxanis, M. Blackledge, Y. Jamin, and Y. Yuan, "Capturing global spatial context for accurate cell classification in skin cancer histology," in *Computational Pathology and Ophthalmic Medical Image Analysis*. Cham, Switzerland: Springer, 2018, pp. 52–60.

[30] R. Wetteland, K. Engan, T. Eftestøl, V. Kvikstad, and E. A. Janssen, "Multiscale deep neural networks for multiclass tissue classification of histological whole-slide images," 2019, *arXiv:1909.01178*.

[31] N. Hashimoto *et al.*, "Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with unannotated histopathological images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3852–3861.

[32] N. Mehta, A. Raja'S, and V. Chaudhary, "Content based sub-image retrieval system for high resolution pathology images using salient interest points," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2009, pp. 3719–3722.

[33] X. Qi *et al.*, "Content-based histopathology image retrieval using cometcloud," *BMC Bioinf.*, vol. 15, no. 1, pp. 1–17, 2014.

[34] A. Sridhar, S. Doyle, and A. Madabhushi, "Content-based image retrieval of digitized histopathology in boosted spectrally embedded spaces," *J. Pathol. Informat.*, vol. 6, 2015, Art. no. 41.

[35] M. Babaie *et al.*, "Classification and retrieval of digital pathology scans: A new dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 8–16.

[36] D. Komura *et al.*, "Luigi: Large-scale Histopathological Image Retrieval System Using Deep Texture Representations," 2018, *bioRxiv: 345785*.

[37] Y. Zheng *et al.*, "Histopathological whole slide image analysis using context-based CBIR," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1641–1652, Jul. 2018.

[38] R. Schaer, S. Otálora, O. Jimenez-del Toro, M. Atzori, and H. Müller, "Deep learning-based retrieval system for gigapixel histopathology cases and the open access literature," *J. Pathol. Informat.*, vol. 10, 2019, Art. no. 19.

[39] N. Hegde *et al.*, "Similar image search for histopathology: SMILY," *NPJ Digit. Med.*, vol. 2, no. 1, pp. 1–9, 2019.

[40] S. Kalra *et al.*, "Yottixel-an image search engine for large archives of histopathology whole slide images," *Med. Image Anal.*, vol. 65, 2020, Art. no. 101757.

[41] A. Riasatian *et al.*, "Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides," *Med. Image Anal.* vol. 70, 2021, Art. no. 102032.

[42] S. Hemati, S. Kalra, C. Meaney, M. Babaie, A. Ghodsi, and H. Tizhoosh, "CNN and deep sets for end-to-end whole slide image representation learning," in *Proc 4th Conf. Med. Imag. Deep Learn.*, 2021, pp. 301–311.

[43] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "Densenet: Implementing efficient convnet descriptor pyramids," 2014, *arXiv:1404.1869*.

[44] A. Riasatian, M. Rasoolijaberi, M. Babaei, and H. R. Tizhoosh, "A comparative study of U-net topologies for background removal in histopathology images," in *Proc. Int. Joint Conf. Neural Netw.*, 2020, pp. 1–8.

[45] S. Hemati and H. Tizhoosh, "Beyond neighbourhood-preserving transformations for quantization-based unsupervised hashing," *Pattern Recognit. Lett.*, vol. 153, 2021, pp. 44–50.

[46] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2916–2929, Dec. 2013.

[47] J. N. Weinstein *et al.*, "The cancer genome atlas pan-cancer analysis project," *Nature Genet.*, vol. 45, no. 10, pp. 1113–1120, 2013.

[48] L. A. Cooper, E. G. Demicco, J. H. Saltz, R. T. Powell, A. Rao, and A. J. Lazar, "Pancancer insights from the cancer genome atlas: The pathologist's perspective," *J. Pathol.*, vol. 244, no. 5, pp. 512–524, 2018.