Research Article Vol. 14, No. 2 / 1 Feb 2023 / Biomedical Optics Express 667
Biomedical Optics EXPRESS

Computationally efficient adaptive decompression for whole slide image processing

ZHEYU LI,[1,4] BIN LI,[2,3,4] KEVIN W. ELICEIRI,[2,3,4,*] AND VIJAYKRISHNAN NARAYANAN[1,4]

[1]*Department of Computer Science and Engineering, Pennsylvania State University, State College, PA 16801, USA*
[2]*Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA*
[3]*Morgridge Institute for Research, Madison, WI 53715, USA*
[4]*Authors contributed equally*
[*]*eliceiri@wisc.edu*

Abstract: Whole slide image (WSI) analysis is increasingly being adopted as an important tool in modern pathology. Recent deep learning-based methods have achieved state-of-the-art performance on WSI analysis tasks such as WSI classification, segmentation, and retrieval. However, WSI analysis requires a significant amount of computation resources and computation time due to the large dimensions of WSIs. Most of the existing analysis approaches require the complete decompression of the whole image exhaustively, which limits the practical usage of these methods, especially for deep learning-based workflows. In this paper, we present compression domain processing-based computation efficient analysis workflows for WSIs classification that can be applied to state-of-the-art WSI classification models. The approaches leverage the pyramidal magnification structure of WSI files and compression domain features that are available from the raw code stream. The methods assign different decompression depths to the patches of WSIs based on the features directly retained from compressed patches or partially decompressed patches. Patches from the low-magnification level are screened by attention-based clustering, resulting in different decompression depths assigned to the high-magnification level patches at different locations. A finer-grained selection based on compression domain features from the file code stream is applied to select further a subset of the high-magnification patches that undergo a full decompression. The resulting patches are fed to the downstream attention network for final classification. Computation efficiency is achieved by reducing unnecessary access to the high zoom level and expensive full decompression. With the number of decompressed patches reduced, the time and memory costs of downstream training and inference procedures are also significantly reduced. Our approach achieves a 7.2× overall speedup, and the memory cost is reduced by 1.1 orders of magnitudes, while the resulting model accuracy is comparable to the original workflow.

1. Introduction

In recent years, whole slide imaging has emerged as one of the essential tools in modern pathology. Unlike traditional microscopy, where a physical slide is viewed by human experts under a microscope, whole slide imaging scans entire tissue sections on glass slides into digital images with histomorphological details preserved in digital format. The digital files, usually referred to as whole slide images (WSIs), can be used for slide archiving, case examination, patient diagnosis, pathology education, and teleconference [1–4]. WSIs often have extremely large image sizes up to ~100,000×100,000 per slide, and the file is usually saved in a pyramidal structure with tiled image patches stored in multiple magnification levels. Image patches are

#477515 https://doi.org/10.1364/BOE.477515
Journal © 2023 Received 6 Oct 2022; revised 16 Dec 2022; accepted 17 Dec 2022; published 6 Jan 2023

Biomedical Optics EXPRESS

normally compressed with the same compression standard across space, such as JPEG2000 [5–7]. Such a pyramidal file structure allows pathologists to quickly open and view slides at low zoom levels that mimic different optical resolutions without accessing the entire gigapixel image at its finest resolution. If malignant suspect regions are identified in low magnification, pathologists can zoom into those regions and make a diagnosis by accessing tiles at higher zoom levels. In most cases, the portion of diagnostic regions roughly ranges from 20% to 80% of the whole tissue area.

The last two decades have witnessed the rapid development of computational methods for WSI analysis, and computational histopathology has become a fruitful research field that leads to many promising clinical applications with strong commercial prospects. Most WSI analysis methods follow a patch-based scheme—the WSIs are firstly tiled into patches, and features are then extracted and analyzed for more complex downstream analysis. Recent deep learning-based classification methods have achieved success on many challenging tasks such as WSI classification and tumor segmentation [8,9]. Many of the classification models follow a multiple-instance learning (MIL) problem formulation, where classifiers are trained in a weakly-supervised way with slide-level supervision signals [10–13]. Those methods are very amendable to clinical practices because the models do not generally require localized annotations to be made on the slide. Instead, slide-level labels are often sufficient. This aligns with the slide examination routine in clinics where slides are seldom annotated but rather categorized according to whether malignancies are observed in an entire slide.

However, most of these methods require an extensive decompression of the WSI files to generate image patches with very little consideration for optimizing the decompression step involved in the whole workflow. This can cause two potential inefficiencies. First, WSIs need to be fully decompressed to the spatial RGB domain and cropped into patches of the desired size. The decompression requires immense storage and computation due to the large size of WSIs and the complexity of the JPEG2000 decoding algorithm. Second, unlike a pathologist who selectively zooms in from a lower magnification to examine a slide, in many computational workflows, all regions across different magnification levels are extensively decompressed with the same decompression depth. This leads to high computation costs and can make large neural network training prohibitive due to the memory requirement. These issues limit the use of computational workflows for WSI analysis tasks where small processing overheads and high throughput are desired, such as slide retrieval for differential diagnosis [14] and case screening in cytopathology [15]. Several recent studies propose to mimic the selective zooming procedure used by pathologists by injecting region selection mechanisms in the models [16–18]. Nevertheless, the idea of utilizing compression domain processing (CDP) to achieve adaptive decompression for WSI analysis, where image patches are selectively decompressed with different decompression depths, has never been explored to our knowledge.

In this paper, we propose a memory and computation-efficient processing pipeline for WSI classification where the concepts of attention-based zoom-in and adaptive decompression are integrated to minimize the overhead caused by patch decompression and patch creation. Notably, the pipeline design is motivated by the general architecture of attention-based models and can be plugged into existing attention-based MIL WSI classification networks with minor modifications [10,13].

The proposed pipeline consists of two adaptive decompression assignment schemes that can be applied independently or together. The first strategy relies on attention scores computed by an attention-based network trained on low-magnification patches. Firstly, image patches in the low magnification are fully decompressed, and different decompression depths are assigned based on an attention measurement on the patches, resulting in decompression assignments of full decompression, partial decompression, and no decompression for the patches in the next magnification. Secondly, the image tiles in the next magnification level are decompressed

Biomedical Optics EXPRESS

according to the decompression depth assignments. All resulting patches are used for the downstream attention-based WSI classification model [10,13].

The second strategy directly makes use of compression domain features obtained from the raw code stream of WSIs files. This strategy measures the information content in both a low-magnification patch and high-magnification patches that fall inside the region of the low-magnification patch directly from the compressed files. Specifically, the entropy information and wavelet coefficient distributions are directly estimated from the code stream. The method then assigns full decompression to the high-magnification patches with large entropy as well as a large difference in wavelet coefficient distribution. The intuition is that the distribution of wavelet coefficient is an indicator of information in the frequency domain, and a large difference suggests that the patterns in the high-magnification are under-represented in the low-magnification patch, and accessing the decompressed data in the high-magnification patch is necessary. We demonstrate the efficacy of the two designs using the two most commonly used WSI magnifications (20× and 5×), but the designs can be easily extended to more magnifications.

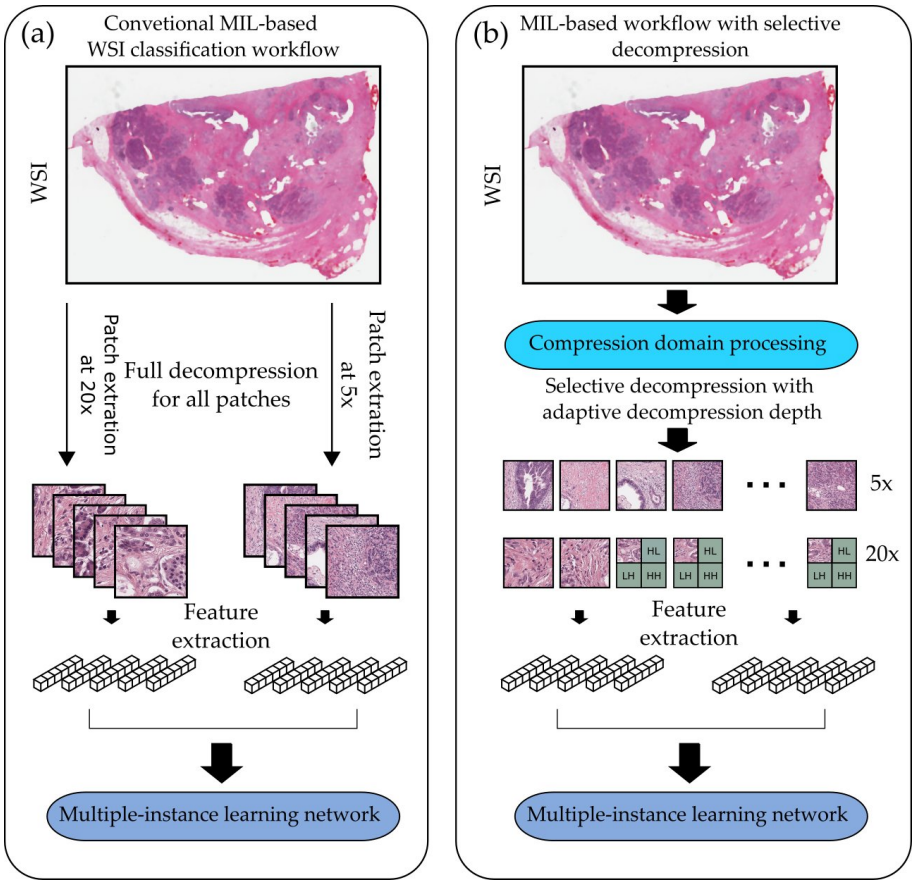

Fig. 1. Comparison between **(a)** the conventional attention-based WSI classification workflow and **(b)** the CDP-based adaptive decompression workflow. The conventional workflow involves extensively accessing decompressed image patches at high magnifications. The proposed CDP-based workflow selectively decompresses high-magnification patches according to features obtained from the compressed domain.

Biomedical Optics EXPRESS

The adaptive decompression schemes can filter out non-informative patches hierarchically and reduce unnecessary accesses to fully decompressed image data which is computationally expensive, leaving the downstream MIL network with a subset of patches 1. We demonstrate the strength of the proposed CDP pipelines on state-of-the-art attention-based neural networks for WSI classification. Evaluation on two large-scale publicly available WSI datasets, including lung cancer and kidney cancer from The Cancer Genome Atlas (TCGA) program, shows that our pipeline is 7.2× faster in the computation time and requires one magnitude less computing memory compared to the baselines, while the degradation in classification accuracy is minor (<2%). The results also show that reducing the redundancies by selectively accessing high-magnification data can potentially improve classification accuracy in some cases.

2. Backgrounds and related works

In this section, we briefly introduce some backgrounds regarding JPEG2000 image coding (the most commonly used standard for WSI coding) [5,6,19], the pyramidal file structure for storing WSIs, recently proposed attention-based MIL approaches for WSI classification and typical applications of CDP in image and video processing.

2.1. JPEG2000

Unlike the discrete cosine transform (DCT)-based JPEG, the JPEG2000 algorithm is based on discrete wavelet transformation (DWT) [20,21]. JPEG2000 was originally proposed as an advanced substitution for JPEG with a better compression ratio and quality. However, compressing and decompressing images with JPEG2000 can be magnitudes slower than with JPEG. This performance issue, along with additional concerns such as reduced software support, has meant JPEG2000 has not yet replaced JPEG in general use. Instead, it has been deployed in some specific technical domains, such as astronomy [22], medical imaging [23], and wireless multimedia [24].

2.2. WSI file structure

The proposed CDP pipelines for WSIs target JPEG2000-based WSIs files (i.e., patches are compressed with JPEG2000). Due to the vastly high resolution, uncompressed WSIs are often very large files with sizes over 20 gigabytes per image. JPEG2000 has become the major standard in WSIs compression in various types of commercial scanners (e.g., 3DHISTECH MRXS, Aperio SVS, and Hamamatsu NDPI) due to its high compression ratio and the ease to create image pyramids [5–7].

Modern WSI scanning can scan tissue sections on glass slides with very high resolution—a typical $2cm \times 2cm$ tissue section results in a WSI with roughly $100,000 \times 100,000$ pixels in 20× magnification, and the number of pixel scales quadratically with the scanning magnification. Most commonly, a WSI is stored as a nested image pyramid with 3 to 4 zoom levels where each level consists of image patches of size around $240 \times 240 \sim 256 \times 256$ pixels. The finest level contains the base magnification while the subsequent levels are reduced magnifications typically with a scaling factor of 4 (e.g., $20\times, 5\times, 1.25\times$), as illustrated in Fig. 2. This file structure is essential to rendering WSIs smoothly for visual inspections, since only the patches and zoom level in the current field-of-view (FOV) need to be read into the memory and displayed, thus, avoiding manipulating the giga-element matrix, which requires immense memory resources.

This file structure with JPEG2000 compression has also been adopted in international biomedical imaging standards such as DICOM [25], OMERO format [26–28] for storing WSIs.

Biomedical Optics EXPRESS

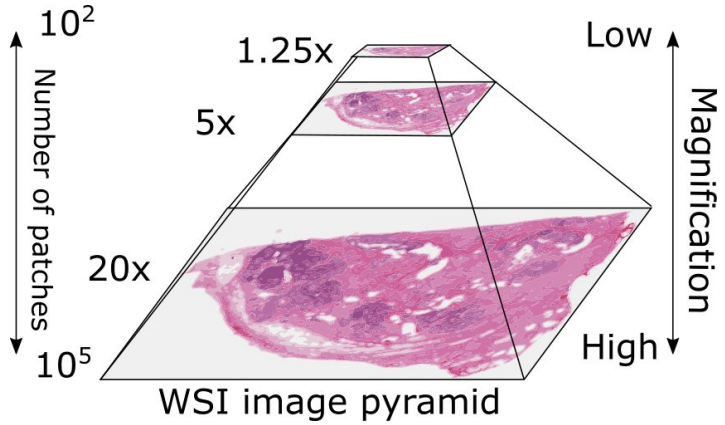

Fig. 2. A WSI image pyramid with multiple magnifications. The number of patches in each magnification level increases exponentially with the magnification level.

2.3. Multiple instance learning for WSI

The major challenges of building computational classification methods for WSIs are the high resolution of WSIs and the lack of localized annotations. The WSIs, though their sizes are in of gigapixel range, are usually labeled globally, i.e., an entire image will be labeled as positive if it contains at least one disease-positive region, and negative otherwise. Consequently, WSI classification is usually cast as a MIL problem, where patches are extracted from a WSI and form "a bag of instances." The slide label is then treated as the bag label. Many MIL-based methods have been proposed for WSI classification, and lesion detection [10,13,29,30], and the majority of them are formulated with attention mechanism [10]. [11] show that a MIL classifier trained on a large amount of unannotated WSIs is favored over a small size annotated WSIs dataset in terms of generalizability, and the former is easy to obtain from clinical routines without labor-intensive annotations. In this study, we evaluate our CDP pipelines on two representative attention-based MIL architectures [10,13] for weakly supervised WSI classification, but the core designs apply to other attention-based models.

2.4. Compression domain learning

Compression domain image processing refers to techniques that perform feature extraction and image manipulation in the compressed images without decoding or with partial decoding [31,32]. CDP is desired in scenarios where full decompression of the image data is expensive and bandwidth-limited with significant overhead, such as video search, online streaming, and image retrieval [33]. CDP benefits from the fact that many compression algorithms already perform some form of feature extraction, either implicitly or explicitly, such as the DCT in JPEG [34], MC-DCT in MPEG [35], and DWT in JPEG2000 [36,37]. Recently deep neural networks (DNNs) are also being trained with compressed domain data for image classification [38,39]. Surprisingly, when some DNNs are optimized for the compression format, (e.g. JPEG2000), the final classification accuracy could surpass a DNN trained on fully decompressed images [40]. In the application of WSI analysis, the extensive full decompression of a large number of image patches leads to a major overhead. Therefore, we leverage the features obtained from the compressed domain to perform selective and adaptive decompression that reduces the unnecessary decompression of redundant and non-informative patches without sacrificing much analysis accuracy.

Biomedical Optics EXPRESS

3. Methodology

3.1. Attention mechanism in MIL-based WSI classification

Following the formulation of MIL, a WSI is represented by a bag of image patches (instances) $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ where $\mathbf{x}_i \in \mathbb{R}^L$ is a low-dimension embedding of an image patch i produced by some feature extractor, y_i is the associated label indicating the malignancy of x_i, and N is the number of patches extracted from the WSI. In the setting of weakly supervised WSI classification, $\{y_i\}_{i=1}^{N}$ is not observed. in our workflow, we used a ResNet18 pre-trained on ImageNet as the feature extractor to compute $\mathbf{x}_i$ As suggested in [12], this is an efficient approach for producing useful embeddings from WSI patches. In general, this feature extractor can be any operator that produces good representations from the image patches.

The purpose of an attention operator f^w is assigning a set of attention weights $\{a_i\}_{i=1}^{N}$ to the instances $\{\mathbf{x}_i\}_{i=1}^{N}$ such that a bag embedding $\mathbf{b}_s \in \mathbb{R}^l$ can be obtained via weighted summing the instances:

$$\mathbf{b}_s = \mathbf{A}^\top \mathbf{X} \tag{1}$$

where $\mathrm{X} \in \mathbb{R}^{N\times L}$ is a matrix containing the features of patches in its rows and $\mathrm{A} \in \mathbb{R}^{N\times C}$ is a matrix containing the attention weights for the patches in its columns. For single class binary MIL, $C = 1$, therefore a_i is a real-valued scalar, and $y_i \in \{0, 1\}$ is a one-digit binary scalar.

The attention matrix $\mathbf{A}$ can be computed by a neural network f^ω parameterized by the weight set ω. For example, in ABMIL [10], f^ω is a multi-layer perceptron operating on each $\mathbf{x}_i$ separately and computing $\{a_i\}$. In DSMIL [13], f^ω computes pair-wise similarities between a critical patch $\mathbf{x}_m$ to each $\mathbf{x}_i$ and use the similarity score as $\{a_i\}$. Accounting the varying bag sizes, $\{a_i\}_{i=1}^{N}$ is processed by a SoftMax function to ensure the constraint of $\sum_i^N a_i = 1$. The bag level classification is conducted on $\mathbf{b}_s$ where a linear classification head is often used:

$$\hat{c}_S = \mathbf{w}^\top \mathbf{b}_\mathrm{s} \tag{2}$$

The training procedure then optimizes the following cross-entropy objective:

$$\{\omega, \mathbf{w}\} = \arg\max \sum_{s}^{S} c_s \log \hat{c}_s + (1 - c_s) \log(1 - \hat{c}_s) \tag{3}$$

where $\{c_s\}_{s=1}^{S}$ is the labels of WSIs in the training set with a number of S slides. Under the assumption of MIL, the relation between the hidden instance labels $\{y_i\}_{i=1}^{N}$ of slide s and its bag label c_s satisfies:

$$c_s = \begin{cases} 0, & \text{iff } \sum y_i = 0 \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

The purpose of this work is to introduce an auxiliary operator f^σ that performs instance selection such that only a fraction of $\{\mathbf{x}_i\}_{i=1}^{N}$ is selected from a WSI for model training and inference while keeping the classification accuracy. This is achieved by assigning different decompression depths to the patches based on some measurements obtained using features from the coarser-level patches and information in the raw code stream of the data. The resulting workflow thus avoids accessing all the fully decompressed data in the high-magnification, which can lead to high computation costs and overhead.

We present here two different strategies for decompression depths assignment. 1) A learning-based method that relies on the training of an attention-based network on the low-magnification where different decompression depths are assigned to the higher-magnification patches according to the attention scores of the lower-magnification patches. 2) An low-cost scheme that assigns decompression depths to the patches determined by measurements based on the features directly

Biomedical Optics EXPRESS

obtained from the raw code stream of the data. Two recently proposed attention-based neural networks for MIL are used as the backbones of our design. Both networks cast weakly supervised WSI classification as a MIL problem and generate attention scores for the patches [10,13].

3.2. Decompression depths assignment based on attention scores

This first strategy requires full decompression of the low-magnification patches, but we will see later that the cost of decompression of low-magnification patches is trivial compared to the computation cost saved by selectively accessing the high-magnification patches. Firstly, the attention-based MIL network is trained using the low magnification patches to perform the classification of WSIs and learns to assign attention scores to the patches. Secondly, The attention scores of each WSI are clustered into three classes by a k-mean clustering algorithm. The three clusters correspond to high/mid/low attention scores, resulting in three different decompression depth assignments, namely full decompression/partial decompression/no decompression, respectively. The decompression depth assignments determine the decompression depths of the high-magnification patches that fall inside the low-magnification patch. Fully decompression means the high magnification patches will be fully decompressed, partial decompression means the high magnification patches will be decompressed up to the second last level of the JEPG2000 coding, and no decompression means the high magnification patches are discarded. Thirdly, the high-magnification patches are decompressed based on their decompression depth assignments, and the resulting decompressed high-magnification patches are combined with the low-magnification patches to retrain the classification network. The overall procedure is illustrated in Fig. 3.

Given that the factor between 5× and 20× magnification is 4, full decompression of all 20× patches will theoretically result in overall training and inference time that is 16 times longer. So the extra time cost of the pre-training phase on 5× patches is much smaller compared to the time reduced by the adaptive decompression that only selects a small fraction of high magnification patches to decompress.

One downside of this strategy is that an attention network must be pre-trained using fully decompressed low-magnification data. This adds additional complexity to the overall workflow, and the effectiveness of the decompression depth assignments largely relies on the quality of the pre-training. Besides, this strategy assigns decompression depths according to the attention scores of the 5× patches, which means that all 16 20× patches within the same area of a 5× patch receive the same decompression depth, and there is no differentiation between them. Therefore, we propose below another decompression depth assignment strategy that can be performed using compression domain features directly available in the code stream of the image file and deploys finer-grained decompression depth assignments to the high-magnification patches.

3.3. Decompression depth assignment based on compression domain features

The raw code stream of JPEG2000 in the file may contain useful features for tasks such as texture classification and image retrieval [33,37]. Here we present an adaptive decompression depth assignment strategy for WSI screening based on the features computed from the raw code stream in the compression domain. The goal is to estimate an appropriate decompression depth to a WSI patch by examining the features in the compressed file, such that the resulting selectively decompressed and partially decompressed patches can still lead to satisfactory accuracy for the downstream classification model. We make use of two types of features from the compressed domain.

Type-I feature: Entropy code. The first type of feature is the number of bytes used to encode the patch data, denoted as B. This is sometimes referred to as "header-based" [36] features since it can be directly read from the headers of the file without any decompression. A JPEG2000 code stream consists of a succession of packets, each attached with a header that

Biomedical Optics EXPRESS

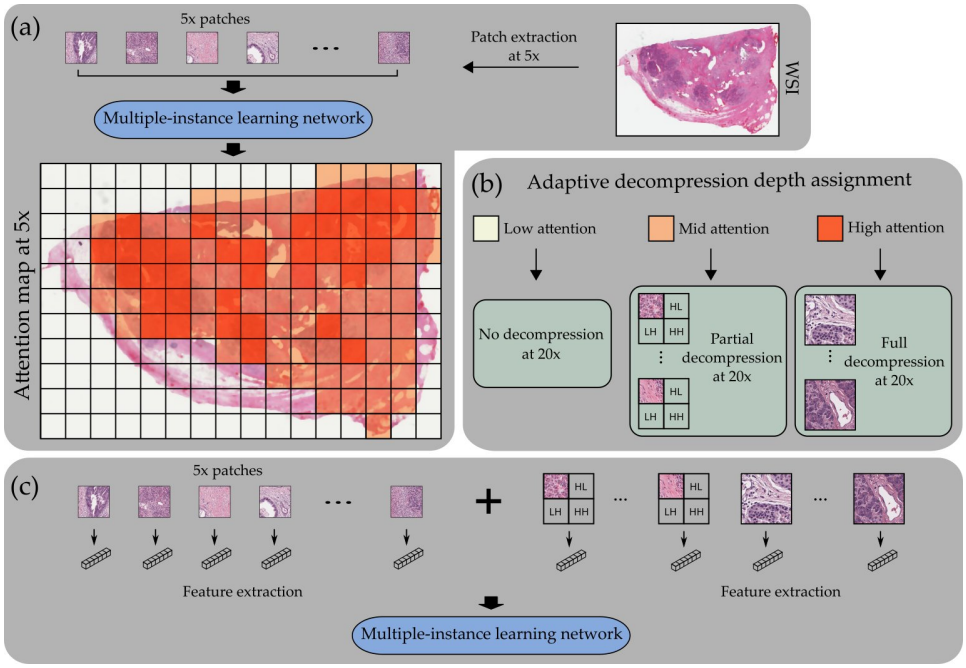

Fig. 3. Decompression depth assignment based on attention scores of low-magnification patches. **(a)**:Patches are first extracted at 5× to train an attention-based MIL network. Attention scores for the patches are obtained. **(b)**: The high-magnification patches within the same field-of-view of a low-magnification patch are decompressed with different decompression depths based on the attention score of the low-magnification patch. **(c)**: 5× patches and selectively decompressed 20× patches are used to train the final network for classification.

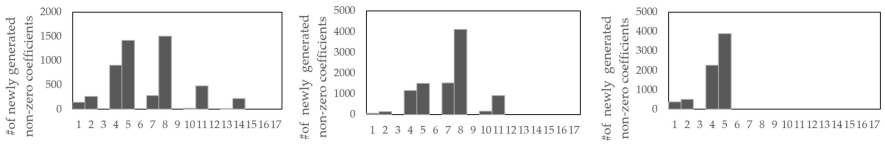

Fig. 4. Illustration of the estimated distribution of wavelet coefficients from the lowest 3 DWT levels. The x-axis is the bitplane from the most significant bitplane to the least significant bitplane. The y-axis is the number of newly generated non-zero coefficients of each bitplane. Newly generated non-zero coefficients at top 9 bit-planes are counted using the openJPEG decoder for all DWT levels. Note that inverse DWT is not needed for this operation.

Biomedical Optics EXPRESS

contains this information. B indicates the efficiency of the entropy encoder compressing the wavelet coefficients of the code blocks in packets; therefore, it serves as a measurement of the source entropy of the data.

We use this feature to exclude low-entropy patches since these patches correspond to out-of-focus low-quality regions of empty backgrounds in WSIs.

Type-II feature: Wavelet coefficient approximations. Wavelets are robust representations of texture-rich images and have been used to characterize histopathological image data [41–43]. Wavelet transform is also central in JPEG2000 compression as the code stream encodes quantized wavelet coefficients. Though extracting wavelet coefficients from JPEG2000 requires decoding the code stream and performing inverse DWT, the distribution of wavelet coefficients can be efficiently estimated from partially decoded code-stream [37].

In the JPEG2000 decoding pipeline, raw code streams are converted to DWT coefficients bit-plane by bit-plane from most significant bit-plane to least significant bit-plane. New non-zero significant coefficients will first be generated from the most significant bit-plane, then further refined when more bit-planes are processed. Inverse DWT needs to be performed in order to obtain the actual wavelet coefficients. However, a very simple wavelet coefficient distribution estimation can be achieved by counting the number of new significant bits generated in the bit-plane decoding process without computing inverse DWT, as suggested in [37]. We follow the method described in [37] and count the number of newly generated non-zero coefficients at the top 9 bit-planes by modifying the openJPEG library to reach the estimation of the wavelet coefficient distribution [Fig. 4].

This feature is used together with the type-I feature to form our second decompression depth assignment strategy.

Joint utilization of type-I and type-II features for decompression depth assignment. Considering a low-magnification (5×) patch and the corresponding 16 high-magnification patches (20×) within the area of the low-magnification patch. We want to assign different decompression depths to the 16 high-magnification patches based on the two types of features retained from the raw code stream.

The first criterion is to use the type-II feature, where a high-magnification patch receives deeper decompression if the wavelet coefficients distribution has a large difference from that of the low-magnification patch. The rationale of this criterion is as follows: 1) Histopathological data is texture-rich, if the high-magnification patch is observed with a similar wavelet spectrum to the low-magnification patch in the coarser level, the frequency components in the two patches are similar. i.e., the effect of aliasing involved in the different magnifications is minor. 2) A large difference in the distributions of the wavelet coefficients between the low-magnification patch and high-magnification patch indicates that the high-resolution details in the low-magnification patch are potentially under-represented, and the high-magnification patch contains very different patterns so it might need to be decompressed and fully observed by the downstream classification model.

The second criterion involves the type-I feature, the indicator of image entropy. This is used because if based solely on the first criterion, a high-magnification patch that contains mostly backgrounds will be chosen to fully decompress, as illustrated in Fig. 5, since the distribution of wavelet coefficient of background will be very different from a non-empty patch. It corresponds to an empty sub-region in the low-magnification patch. Combining the two criteria, a patch will receive deeper compression if its wavelet coefficient distribution is very different from the corresponding low magnification patch and its entropy is not small. In practice, We measure the distribution difference using the cosine similarity metric. Any patches with a similarity smaller than 0.8 (which indicates a large difference) will receive full decompression. Any patches with a similarity between 0.8 and 0.9 (mid difference) will receive partial decompression, and the rest patches are skipped. Meanwhile, a patch is excluded from decompression if its entropy is smaller

Biomedical Optics EXPRESS

than a fixed threshold. The impact of threshold selection on classification accuracy can be found in Appendix III Fig. 10. Some representative selection results are shown in Fig. 6.

Hierarchical decompression depth assignment. Note that this strategy can be integrated with the first strategy described in the above sub-section. The first strategy assigns a constant decompression depth to all 16 high-magnification patches within the area of a low-magnification patch, while the second strategy can perform a finer-grained assignment among the high-magnification patches. For example, instead of assigning full decompression to all 16 high-magnification patches, use the empirical measure based on the compression domain features to select a fraction of the 16 patches for fully decompressing.

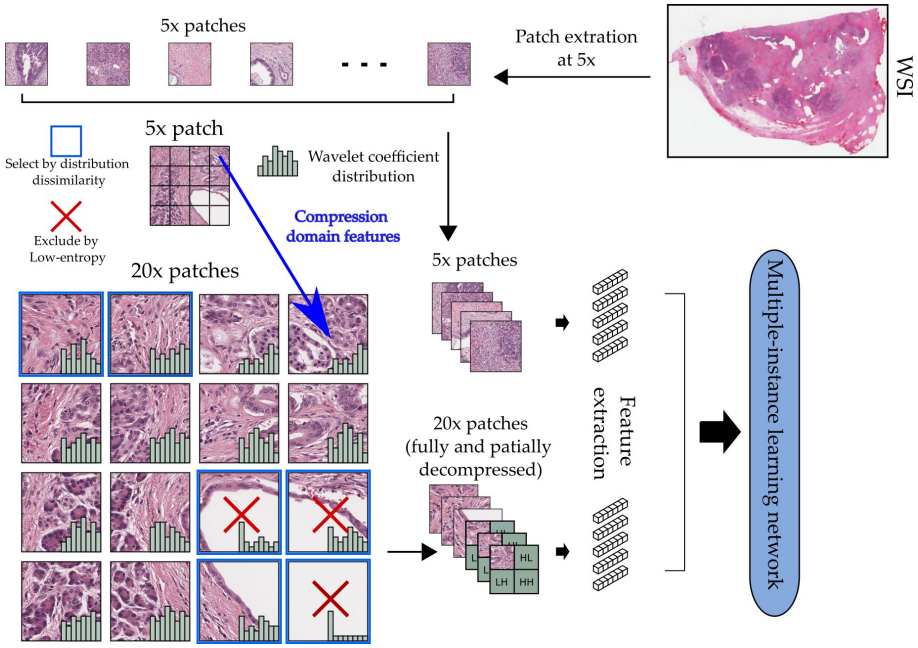

Fig. 5. Decompression depth assignment based on compression domain features obtained from compressed code stream. Type-I feature filters out low-entropy patches. Type-II feature compares the approximate wavelet coefficient distribution of a low-magnification patch to the corresponding high-magnification patches and selects the under-represented high-magnification patches.

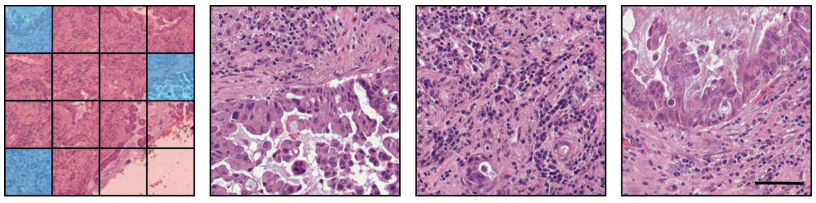

Fig. 6. Non-empty under-represented 20× patches are selected for full decompression in scheme 2. **(a)**: a 5× patch. **(b)-(d)**: 20× patches selected for full decompression. Blue shades in (a) mark the selected fully-decompressed patches. Red shades in (a) mark partially decompressed/discarded patches. The scale bar is 40μm.

Biomedical Optics EXPRESS

3.4. Partial decompression of JPEG2000 encoded images

The encoding pipeline for JPEG2000 is as follows. An input image is first transformed into wavelet representations by applying DWT. The transformation results in four different sub-bands, low-pass in both the horizontal and vertical directions (LL), horizontal low-pass and vertical high-pass (LH), vertical low-pass and horizontal high-pass (HL) and high-pass in both the horizontal and vertical directions (HH), and each with a dimension size half of the original. High-frequency details are decomposed into LH, HL, and HH bands, and the LL band contains low-frequency information, which can be viewed as an approximation of the image with half of the original resolution. The LL band can be applied to another level of DWT transformation which results in another four sub-bands. In practice, the level of DWT transformation is usually set between 4 and 6. Next, quantization is performed, and the quantized DWT coefficients are arranged as small code blocks, usually with a size of 32x32. A block encoder is used to re-arrange code blocks in a bit plane using an arithmetic encoder and EBCOT algorithm, which results in the final JPEG2000 code stream. The decoding can be seen as a reverse procedure of encoding. JPEG2000 code streams are reconstructed to code blocks by EBCOT decoder and arithmetic decoder [44]. The decoded values are de-quantized, and the DWT coefficients are restored. Lastly, inverse DWT is performed from the innermost level's sub-bands (LL, LH, HL, HH) to the outermost level (Fig. 7). The inverse-DWT process can be viewed as progressively reconstructing a lower-resolution image to a higher-resolution image by combining sub-bands. The Inverse-DWT at the outermost level contributes about 50% of total decompressing time and is usually considered as the computational bottleneck (Fig. 8).

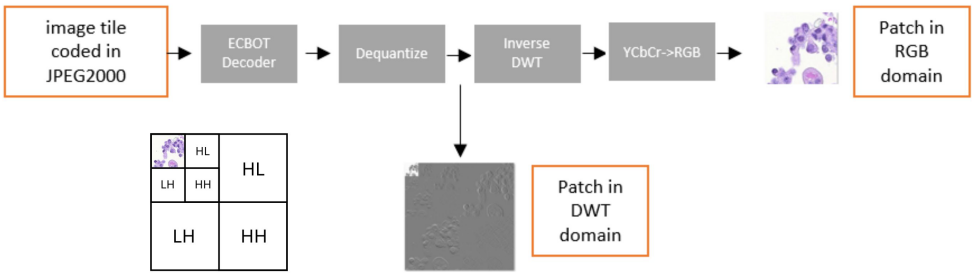

Fig. 7. Decoding Pipeline of JPEG2000. Wavelet coefficients are read from the compressed image file and dequantized, inverse DWT is performed to convert the data in the wavelet domain to the image domain, and the image data is then rendered in an RGB pixel matrix.

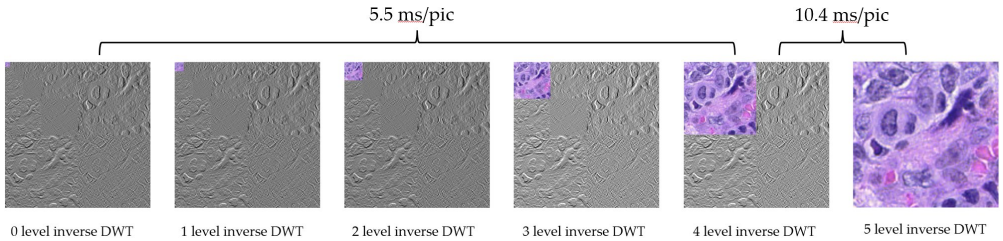

Fig. 8. Hierarchical Inverse DWT transformation from partially decompressed image tile to fully decompressed image tile. Inverse DWT is performed from the innermost level's sub-bands (LL, LH, HL,HH) to the outermost level. The resolution of the decoded image is increased by a factor of 2 after each inverse DWT level. In experiments, decompression takes an average of $5.5ms/pic$ for level 0 to level 4 and $10.4ms/pic$ for level 4 to level 5.

Biomedical Optics EXPRESS

The inverse DWT converts the wavelet coefficients in the compressed domain to the final image representation in the spatial domain. The transformation happens in a hierarchical order, often with 5-6 levels. Figure 8 shows an example of 6 levels (level 0 to level 5). Decoding each level adds more high-frequency details incrementally to the image. This step is a well-known bottleneck in JPEG2000 decompression [37,45], and the outermost level requires the most computation time. For the partial decompression involved in our pipeline, the image domain data is extracted from the 2nd outermost level, which requires significantly less decompression time while still preserving some high-frequency details (Fig. 8). Both fully and partially decompressed patches are fed to a ResNet18 for feature extractions. The resulting feature vectors are then used by the downstream attention-based neural networks.

3.5. Progressive decompression in multi-magnification

We demonstrate the described CDP schemes in the experiment section using two magnifications, but the methods can be applied to multiple magnifications progressively. Starting from the first magnification, the CDP schemes assign different decompression depths to the patches in the second magnification according to their potential importance computed using CDP. Within the ROIs of the patches that receive full decompression in the second magnification, the decompression progresses into the third magnification, assigning different decompression depths to the patches using the same above criteria.

4. Experiments and results

4.1. Experiment setup

We evaluated the proposed CDP pipeline for WSIs classification using two publicly accessible WSI datasets from The Cancer Genome Atlas (TCGA). The first dataset is the non-small cell lung cancer dataset. The tissue specimens contain two sub-type of lung cancer, including lung adenocarcinoma (351 slides) and squamous cell neoplasms (348 slides). The second dataset is the renal cell carcinoma dataset that includes WSIs of two sub-types of kidney cancer, namely kidney renal clear cell carcinoma (380 slides) and kidney renal papillary cell carcinoma (155 slides). Note that a small fraction of slides that are not encoded using JPEG2000 is discarded in both datasets (29.8% in the lung cancer dataset and 30.3% in the kidney cancer dataset). Most of the non-JPEG2000 slides were old, and the latest standard uses JPEG2000.

Most of the tissue specimens were H&E stained, and the slides were scanned by an Aperio scanner (Leica Biosystems, Wetzlar, Germany) at a base magnification of 20× and 40×. The slides are randomly split into 5 equal size sets for 5-fold cross-validation. The cross-validation is repeated for 3 times, each time with a new random split. All patches are extracted in 256×256 or 240×240, which matches the native tile size of JPEG2000 coded WSI image. The patches are either fully decompressed, partially decompressed, or discarded according to the decompression assignment schemes described above. Feature vectors are produced by feeding fully decompressed or partially decompressed patches into a ResNet18 pre-trained on ImageNet.

JPEG2000 decompression and DWT coefficients extraction is done by modifying existing libraries of openJPEG 2.40. For slide access and patch creation, openslide3.4.1 is used. The attention-based network is implemented using PyTorch according to the official implementation available on GitHub. The evaluation is performed on a computer setup with Core i7 8770k CPU, 48 GB memory and NVIDIA RTX3070 GPU.

4.2. WSI classification networks, baseline, and upper-bound

The WSI classification networks used in the evaluation are two attention-based networks proposed for WSI classification, namely ABMIL [10] and DSMIL [13]. Pre-computed feature vectors

Biomedical Optics EXPRESS

(using the CNN backbone) of WSI image patches are fed into the two MIL networks as inputs and are trained to predict the slide label.

The baseline setting involves feeding the attention networks with 5× patches extracted from the WSIs while the upper-bound setting involves feeding the attention networks with all the 5× patches and 20× patches. There is an additional baseline setting where only the 20× patches are fed to the attention networks. For the non-small cell lung cancer dataset, in total 0.13 million patches are extracted at 5× and 2 million patches are extracted at 20×. For the renal cell carcinoma dataset, in total 0.12 million patches are extracted at 5× and 1.7 million patches are extracted at 20×. The total numbers of patches used in the proposed CDP methods lie between the two numbers and are summarized in Table 2 and Table 3 (additional results can be found in Table 1).

4.3. Decompression scheme based on attention scores

The attention networks are firstly trained 20 epochs until convergence using the 5× patches. After the networks are trained, the attention scores are computed for each training slide on the 5× patches. The decompression depths are assigned to the corresponding 20× patches according to the attention scores. The 20× patches are then decompressed based on the assigned decompression depths and combined with the 5× patches to train the attention networks for another 20 epochs until convergence. For the non-small cell lung cancer dataset, in total ~0.05 million 20× patches are fully decompressed, ~0.14 million 20× patches are partially decompressed. For the renal cell carcinoma dataset, in total, ~0.019 million 20× patches are fully decompressed, ~0.084 million 20× patches are partially decompressed. The time spent on decompression and feature computing is significantly shortened (reduced from about 500 minutes to about 30 minutes). The classification accuracy and computation cost are summarized in Table 2 and Table 3 (additional results can be found in Table 1). Though the saving in computation cost is remarkable, there is a certain amount of classification accuracy degradation introduced by this selective decompression scheme on one of the datasets (non-small cell lung cancer dataset). In the renal cell carcinoma dataset, selectively accessing the 20× data results in even higher accuracy compared to the baseline method, where all 20× patches are exhaustively decompressed. The possible reason for this observation is discussed in the next section.

4.4. Decompression scheme based on compression domain feature

Unlike the first CDP scheme, the second CDP scheme involves only a single training pass and the decompression depth assignments are computed empirically from features computed from the compression domain. Noted that the type-I feature can be used to screen 5× patches prior to the training, to exclude empty or out-of-focus tiles. Same as the first scheme, decompressed 20× patches are combined with 5× patches and used for training the attention networks. The network is trained for 20 epochs until convergence. This scheme results in total ~0.02 million fully decompressed 20× patches and ~0.04 million partially decompressed 20× patches for the non-small cell lung cancer dataset, and ~0.005 million fully decompressed 20× patches and ~0.059 million partially decompressed 20× patches for the renal cell carcinoma dataset. The classification accuracy and computation cost are summarized in Table 2 and Table 3 (additional results can be found in Table 1). Notably, this selective decompression scheme based on compression domain features results in higher classification accuracy compared to the baseline, where all 20× patches are decompressed exhaustively. We give a plausible explanation for this observation in the discussion section.

Biomedical Optics EXPRESS

Table 1. A summary of model classification accuracy, patch numbers, decompression time cost, and model training time cost of different methods (attention MIL network backbone is DSMIL [13]). CDP 1: the first selective decompression scheme described in the above section 3. CDP 2: the second selective decompression scheme in the above section 3. CDP 1+2: apply the first selective decompression scheme followed by the second selective decompression scheme. 20×: fully decompress all 20× patches. 5×: fully decompress all 5× patches. 20×+5×: fully decompress all 20× and 5× patches.

Result on TCGA non-small cell lung cancer dataset								
Method	Accuracy	AUC	Number of 5× patches	Number of 20× patches	Feature extraction	Decompression	Network training	Total
5×	80.4%	0.890	128k	–	15 min	29 min	12 min	56 min
20×	87.7%	0.932	–	2,053k	193 min	464 min	56 min	713 min
5+20×	88.2%	0.926	128k	2,053k	210 min	493 min	65 min	768 min
CDP 1	84.8%	0.915	90k	50k+140k[a]	31 min	47+11 min[b]	12+20 min[c]	121 min
CDP 2	89.4%	0.935	128k	20k+40k	68 min	37+5 min	17 min	127 min
CDP 1+2	86.7%	0.918	90k	20k+35k	36 min	27+5 min	12+16 min	96 min
Result on TCGA renal cell carcinoma dataset								
Method	Accuracy	AUC	Number of 5× patches	Number of 20× patches	Feature extraction	Decompression	Network training	Total
5×	91.6%	0.969	125k	–	15 min	28 min	12 min	55 min
20×	92.5%	0.972	–	1,788k	182 min	450 min	40 min	672 min
5+20×	92.5%	0.976	125k	1,788k	197 min	478 min	42 min	717 min
CDP 1	93.0%	0.971	119k	19k+84k	29 min	30+8 min	12+18 min	97 min
CDP 2	93.5%	0.986	125k	5k+59k	57 min	28+6 min	16 min	107 min
CDP 1+2	93.4%	0.979	118k	4k+17k	39 min	27+2 min	12+14 min	94 min

[a]Include fully decompressed patches and partially decompressed patches
[b]Include full decompression time and partial decompression time
[c]Include 5× attention network training time and final training time with 5× and partially/fully decompressed 20× patches.

4.5. *Hierarchical decompression scheme based on both attention scores and compression domain feature*

As mentioned above, once the attention scores are obtained for the 5× patches, one can use the empirical scheme to perform a finer-grained assignment of decompression depth to the 20× patches falling within the field-of-view of a 5× patch. This results in an aggressive decompression depth assignment where only a very small fraction of 20× patches are selected for full decompressing, thus, largely reducing the computation cost. Combining the two schemes results in total ~0.02 million fully decompressed 20× patches ~0.035 million partially decompressed 20× patches for the non-small cell lung cancer dataset, and ~0.004 million fully decompressed 20× patches ~0.017 million partially decompressed 20× patches for the non-small cell lung cancer dataset. The evaluation results are summarized in Table 3 (additional results can be found in Table 1). The resulting accuracy can be worse than using the second scheme alone (the scheme that relies on the measurement obtained from the compressed code stream without using the attention scores from a pretrained attention-based network in 5×) in the non-small cell lung cancer dataset, indicating that the attention scores obtained for the 5× patches could fail to reflect the importance of some 20× patches. The attention scores are computed by the attention-based network trained

Biomedical Optics EXPRESS

only using the 5× patches, and some important morphological details may be indiscernible in the 5× magnification.

The training of the attention-based classification network can be memory-bounded since it requires all patches (or at least their embeddings) to be loaded to the GPU memory. For large WSIs, this can result in immense GPU memory usage when high-magnification patches are extracted exhaustively. The memory usage (including temporal decompressed data storage and GPU usage for network training) is proportional to the number of patches created from each WSI. Our CDP methods thus alleviate this issue by reducing the number of patches observed by the downstream network, achieving an overall memory usage that is 11× smaller than the original workflow.

Overall, the evaluation results suggest that by carefully selecting a subset of the high-magnification patches, the accuracy of the downstream attention-base classification network can be largely maintained (and sometimes even improved). This suggests that there exist a certain amount of redundancy and noisy (irrelevant) patches in the high magnification, and by making use of the prior knowledge obtained from the compression domain features, redundant and irrelevant high-magnification patches can be efficiently excluded. This benefits the analysis workflow in terms of both computation cost and prediction accuracy, with the decompression time, training/inference time, as well as memory usage reduced at least one order of magnitude while still achieving high classification accuracy (illustrated in Figure 9).

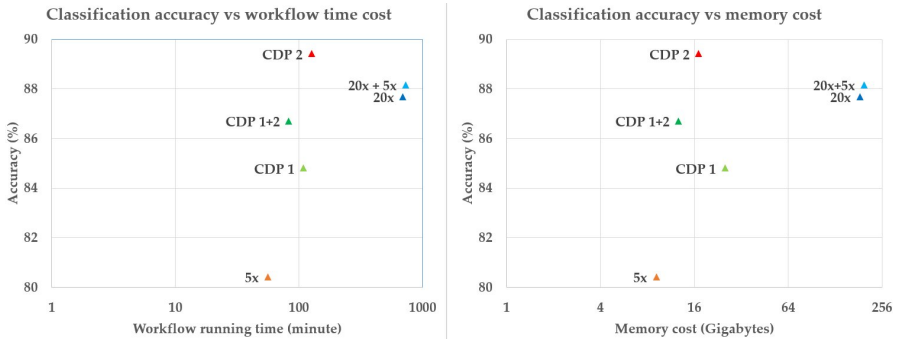

Fig. 9. Computation time cost and memory cost vs classification accuracy. **CDP 1**: the first selective decompression scheme. **CDP 2**: the second selective decompression scheme. **CDP 1+2**: apply the first selective decompression scheme followed by the second selective decompression scheme. **20×**: fully decompress all 20× patches. **5×**: fully decompress all 5× patches. **20×+5×**: fully decompress all 20× and 5× patches.

5. Discussion and conclusion

5.1. Discussion

The central idea of CDP in our proposed WSI classification pipeline is to reduce unnecessary access to fully decompressed image data in high magnifications. The number of image patches increases exponentially with the magnification level. Thus, WSI analysis pipelines could require a significant amount of computation time, and they often become memory bounded when the target magnifications are high. This is because of that training the attention network requires all patches of a WSI to be seen at a single iteration. A common solution to alleviate the issues is to use pre-trained feature extractors and project image patches into low-dimensional vectors before analyzing. For example, a WSI classification model usually operates on pre-computed image patches feature vectors, instead of raw image patches [12,13] However, one downside of using pre-trained feature extractors is that the feature extractors can be sub-optimal regarding

Biomedical Optics EXPRESS

the downstream classification task since it is not a part of the training process. Applying the proposed CDP pipelines to WSI analysis could reduce the number of patches needed to represent a WSI. Therefore, one can potentially train the feature extractor and the downstream attention network end-to-end since the memory requirement for fitting the feature extractor is no longer prohibitive for small numbers of patches.

The other observation is that by selectively decompressing the image patches, the accuracy of the attention network can sometimes be improved. One explanation is that some redundant patches and noisy patches that could lead to smoothing effects in the resulting bag embeddings are removed during the process. In the attention mechanism, though irrelevant patches are generally assigned low attention weights, they still receive positive weights and contribute to the bag embedding, which is a weighted sum of all patch embeddings. Having more noisy and redundant patches in the weighted sum smooths the resulting bag embedding and makes it less discriminative. Therefore, CDP methods can potentially filter out a portion of noisy and redundant patches at high magnification, leading to higher accuracy of the resulting classification network while also reducing computation costs.

Moreover, the idea of adaptive compression can also be applied to other types of classification backbones that are recently gaining favors for WSI classification, including vision transformers [46–48] and graph neural networks [49–51], by selecting a subset of the image patches such that the sequence length for a transformer and the node number for a graph are reduced.

Furthermore, the idea of selective decompression based on CDP could potentially be applied to other image analysis tasks that target multi-resolution images with hierarchical zoom levels, such as remote sensing image analysis.

5.2. Conclusion

In conclusion, this paper presents designs of computationally efficient attention-based WSI classification pipelines utilizing CDP. The methods progressively assign different decompression depths to high-magnification image patches according to the information measurements derived from compression domain features or partially decompressed image features. Our methods can be applied to the widely adopted MIL-based WSI analysis pipelines with a small number of modifications. Computation profiling and accuracy evaluation results show that the proposed methods can reduce the processing time, including image decompression and model training, achieving 7.2× speed up, and reducing memory cost by 1.1 orders of magnitudes, with minimum compromise in the model accuracy. Moreover, in some circumstances, reducing the amount of fully decompressed high-magnification patches can even be beneficial to the model's accuracy.

The computation efficiency of WSI analysis pipelines can still be improved by leveraging compatible hardware designs. As the next step, we will explore a hardware acceleration strategy based on near-storage computing [52]. Regular image processing operations such as decompression and patching can be efficiently offloaded to a near-storage hardware accelerator. The near storage accelerator can exploit massive parallelism and internal bandwidth of fast storage, which could potentially reduce computing time as well as costly data transfer to CPU/GPU. The hardware acceleration strategy is also well suited to our proposed method since the compressed domain feature analysis and early selection process can be done optimally close to the data storage.

Biomedical Optics EXPRESS

Appendix I

Table 2. A summary of model classification accuracy and model training time cost of different methods (attention MIL network backbone is ABMIL [10]). Note that the number of patches and time for feature extraction and decompression are the same as shown in Table 3. CDP 1: the first selective decompression scheme described in the above section 3. CDP 2: the second selective decompression scheme in the above section 3. CDP 1+2: apply the first selective decompression scheme followed by the second selective decompression scheme. 20×: fully decompress all 20× patches. 5×: fully decompress all 5× patches. 20×+5×: fully decompress all 20× and 5× patches.

Result on TCGA non-small cell lung cancer dataset			
Method	Accuracy	Network training time	Total time
5×	79.0%	10 min	56 min
20×	84.1%	45 min	702 min
5+20×	84.1%	52 min	865 min
CDP 1	81.2%	10+16 min	115 min
CDP 2	84.1%	14 min	124 min
CDP 1+2	82.6%	10+13 min	81 min
Result on TCGA renal cell carcinoma dataset			
Method	Accuracy	Network training time	Total time
5×	88.8%	10 min	53 min
20×	89.7%	32 min	664 min
5+20×	89.7%	34 min	699 min
CDP 1	91.6%	10+15 min	92 min
CDP 2	91.6%	13 min	194 min
CDP 1+2	93.5%	10+11 min	90 min

Biomedical Optics EXPRESS

Appendix II

Table 3. A summary of model classification accuracy, patch numbers, decompression time cost, and model training time cost of different methods (attention MIL network backbone is DSMIL [13]). CDP 1: the first selective decompression scheme described in the above section 3. CDP 2: the second selective decompression scheme in the above section 3. CDP 1+2: apply the first selective decompression scheme followed by the second selective decompression scheme. 20×: fully decompress all 20× patches. 5×: fully decompress all 5× patches. 20×+5×: fully decompress all 20× and 5× patches.

Result on TCGA brain cancer (low grade glioma and glioblastoma multiforme) dataset								
Method	Accuracy	AUC	Number of 5× patches	Number of 20× patches	Feature extraction	Decompression	Network training	Total
5×	84.4%	0.857	59k	–	7 min	16 min	8 min	31 min
20×	91.0%	0.951	–	382k	50 min	103 min	29 min	182 min
5+20×	91.0%	0.954	59k	382k	57 min	118 min	36 min	211 min
CDP 1	86.4%	0.895	57k	62k+21k[a]	17 min	14+2 min[b]	8+13 min[c]	54 min
CDP 2	89.8%	0.932	59k	81k+40k	22 min	29+4 min	15 min	70 min
CDP 1+2	85.6%	0.878	57k	30k+31k	14 min	13+3.0 min	8+12 min	50 min

[a]Include fully decompressed patches and partially decompressed patches
[b]Include full decompression time and partial decompression time
[c]Include 5× attention network training time and final training time with 5× and partially/fully decompressed 20× patches.

Appendix III

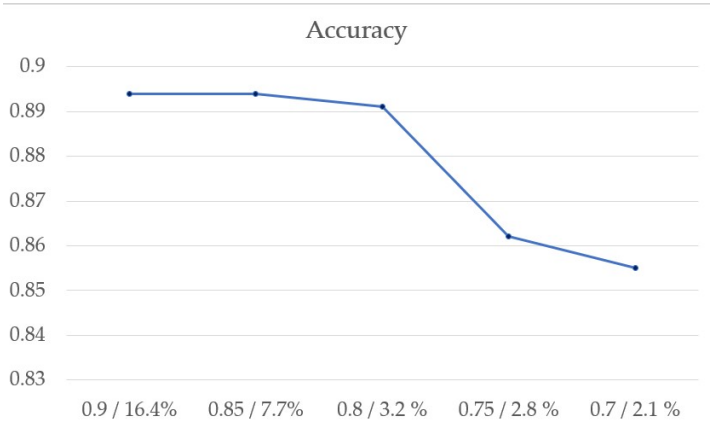

Fig. 10. The impact of different thresholds for selective decompressing patches in scheme CDP2 (TCGA non-small cell lung cancer dataset). **Y axis:** accuracy; **X axis:** wavelet coefficients similarity threshold/fraction of patches selected for decompression in the high magnification.

Funding. National Institutes of Health (GM135019); Morgridge Institute for Research; Semiconductor Research Corporation.

Acknowledgments. We thank the members of our labs for useful technical discussions.

Disclosures. The authors declare no conflicts of interest.

Biomedical Optics EXPRESS

Data availability. Data used in the experiments presented in this paper are available at [53].

References

1. L. Pantanowitz, P. N. Valenstein, A. J. Evans, K. J. Kaplan, J. D. Pfeifer, D. C. Wilbur, L. C. Collins, and T. J. Colgan, "Review of the current state of whole slide imaging in pathology," J. Pathol. Inf. **2**(1), 36 (2011).
2. F. Ghaznavi, A. Evans, A. Madabhushi, and M. Feldman, "Digital imaging in pathology: whole-slide imaging and beyond," Annu. Rev. Pathol.: Mech. Dis. **8**(1), 331–359 (2013).
3. N. Farahani, A. V. Parwani, and L. Pantanowitz, "Whole slide imaging in pathology: advantages, limitations, and emerging perspectives," Pathol. Lab. Med. Int. **7**, 4321 (2015).
4. R. Huss and S. E. Coupland, "Software-assisted decision support in digital histopathology," J. Pathol. **250**(5), 685–692 (2020).
5. M. García-Rojo, "International clinical guidelines for the adoption of digital pathology: a review of technical aspects," Pathobiology **83**(2-3), 99–109 (2016).
6. T. Kalinski, R. Zwönitzer, F. Grabellus, S.-Y. Sheu, S. Sel, H. Hofmann, and A. Roessner, "Lossless compression of jpeg2000 whole slide images is not required for diagnostic virtual microscopy," Am. J. Clin. Pathol. **136**(6), 889–895 (2011).
7. A. Huisman, A. Looijen, S. M. van den Brink, and P. J. van Diest, "Creation of a fully digital pathology slide archive by high-volume tissue slide scanning," Hum. Pathol. **41**(5), 751–757 (2010).
8. C. L. Srinidhi, O. Ciga, and A. L. Martel, "Deep neural network models for computational histopathology: a survey," Med. Image Anal. **67**, 101813 (2021).
9. N. Dimitriou, O. Arandjelović, and P. D. Caie, "Deep learning for whole slide image analysis: an overview," Front. Med. **6**, 264 (2019).
10. M. Ilse, J. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," in *International Conference on Machine Learning*, (PMLR, 2018), pp. 2127–2136.
11. G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. Werneck Krauss Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, and T. J. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," Nat. Med. **25**(8), 1301–1309 (2019).
12. M. Y. Lu, D. F. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, "Data-efficient and weakly supervised computational pathology on whole-slide images," Nat. Biomed. Eng. **5**(6), 555–570 (2021).
13. B. Li, Y. Li, and K. W. Eliceiri, "Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021), pp. 14318–14328.
14. S. Kalra, H. R. 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," Med. Image Anal. **65**, 101757 (2020).
15. E. A. El-Gabry, A. V. Parwani, and L. Pantanowitz, "Whole-slide imaging: widening the scope of cytopathology," Diagn. Histopathol. **20**(12), 456–461 (2014).
16. A. Katharopoulos and F. Fleuret, "Processing megapixel images with deep attention-sampling models," in *International Conference on Machine Learning*, (PMLR, 2019), pp. 3282–3291.
17. J. Zhang, K. Ma, J. Van Arnam, R. Gupta, J. Saltz, M. Vakalopoulou, and D. Samaras, "A joint spatial and magnification based attention framework for large scale histopathology classification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021), pp. 3776–3784.
18. N. Dong, M. Kampffmeyer, X. Liang, Z. Wang, W. Dai, and E. Xing, "Reinforced auto-zoom net: towards accurate and fast breast cancer segmentation in whole-slide images," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (Springer, 2018), pp. 317–325.
19. H. Helin, T. Tolonen, O. Ylinen, P. Tolonen, J. Napankangas, and J. Isola, "Optimized jpeg 2000 compression for efficient storage of histopathological whole-slide images," J. Pathol. Inf. **9**(1), 20 (2018).
20. G. K. Wallace, "The jpeg still picture compression standard," IEEE Trans. Consumer Electron. **38**(1), xviii–xxxiv (1992).
21. A. Skodras, C. Christopoulos, and T. Ebrahimi, "The jpeg 2000 still image compression standard," IEEE Signal Process. Mag. **18**(5), 36–58 (2001).
22. J. Bobin, J.-L. Starck, and R. Ottensamer, "Compressed sensing in astronomy," IEEE J. Sel. Top. Signal Process. **2**(5), 718–726 (2008).
23. D. H. Foos, E. Muka, R. M. Slone, B. J. Erickson, M. J. Flynn, D. A. Clunie, L. Hildebrand, K. S. Kohm, and S. S. Young, "Jpeg 2000 compression of medical imagery," in *Medical Imaging 2000: PACS Design and Evaluation: Engineering and Clinical Issues*, vol. 3980 (SPIE, 2000), pp. 85–96.
24. D. Taubman and M. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards and Practice*, vol. 642 (Springer Science & Business Media, 2012).
25. M. D. Herrmann, D. A. Clunie, A. Fedorov, S. W. Doyle, S. Pieper, V. Klepeis, L. P. Le, G. L. Mutter, D. S. Milstone, and T. J. Schultz, "Implementing the dicom standard for digital pathology," J. Pathol. Inf. **9**(1), 37 (2018).
26. I. G. Goldberg, C. Allan, J.-M. Burel, D. Creager, A. Falconi, H. Hochheiser, J. Johnston, J. Mellen, P. K. Sorger, and J. R. Swedlow, "The open microscopy environment (ome) data model and xml file: open tools for informatics and quantitative analysis in biological imaging," Genome Biol. **6**(5), R47 (2005).

Biomedical Optics EXPRESS

27. M. Linkert, C. T. Rueden, C. Allan, J.-M. Burel, W. Moore, A. Patterson, B. Loranger, J. Moore, C. Neves, and D. MacDonald, "Metadata matters: access to image data in the real world," J. Cell Biol. **189**(5), 777–782 (2010).
28. S. Besson, R. Leigh, M. Linkert, C. Allan, J.-M. Burel, M. Carroll, D. Gault, R. Gozim, S. Li, and D. Lindner, "Bringing open data to whole slide imaging," in *European Congress on Digital Pathology*, (Springer, 2019), pp. 3–10.
29. J. Yao, X. Zhu, J. Jonnagaddala, N. Hawkins, and J. Huang, "Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks," Med. Image Anal. **65**, 101789 (2020).
30. S. Wang, Y. Zhu, L. Yu, H. Chen, H. Lin, X. Wan, X. Fan, and P.-A. Heng, "Rmdl: Recalibrated multi-instance deep learning for whole slide gastric image classification," Med. Image Anal. **58**, 101549 (2019).
31. S.-F. Chang, "Compressed-domain techniques for image/video indexing and manipulation," in *Proceedings., International Conference on Image Processing*, vol. 1 (IEEE, 1995), pp. 314–317.
32. F. Arman, A. Hsu, and M.-Y. Chiu, "Image processing on compressed data for large video databases," in *Proceedings of the first ACM international conference on Multimedia*, (1993), pp. 267–272.
33. H. Wang, A. Divakaran, A. Vetro, S.-F. Chang, and H. Sun, "Survey of compressed-domain features used in audio-visual indexing and analysis," J. Vis. Commun. Image Represent. **14**(2), 150–183 (2003).
34. B. C. Smith and L. A. Rowe, "Algorithms for manipulating compressed images," IEEE Comput. Grap. Appl. **13**(5), 34–42 (1993).
35. S.-F. Chang and D. G. Messerschmitt, "Manipulation and compositing of mc-dct compressed video," IEEE J. Select. Areas Commun. **13**(1), 1–11 (1995).
36. A. Tabesh, A. Bilgin, K. Krishnan, and M. W. Marcellin, "Jpeg2000 and motion jpeg2000 content analysis using codestream length information," in *Data Compression Conference*, (IEEE, 2005), pp. 329–337.
37. A. Descampe, C. De Vleeschouwer, P. Vandergheynst, and B. Macq, "Scalable feature extraction for coarse-to-fine jpeg 2000 image classification," IEEE Trans. on Image Process. **20**(9), 2636–2649 (2011).
38. L. Gueguen, A. Sergeev, B. Kadlec, R. Liu, and J. Yosinski, "Faster neural networks straight from jpeg," Advances in Neural Information Processing Systems **31** (2018).
39. M. Ehrlich and L. S. Davis, "Deep residual learning in the jpeg transform domain," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), pp. 3484–3493.
40. L. D. Chamain, S.-c. S. Cheung, and Z. Ding, "Quannet: Joint image compression and classification over channels with limited bandwidth," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*, (IEEE, 2019), pp. 338–343.
41. T. Wan, X. Liu, J. Chen, and Z. Qin, "Wavelet-based statistical features for distinguishing mitotic and non-mitotic cells in breast cancer histopathology," in *2014 IEEE International conference on image processing (ICIP)*, (IEEE, 2014), pp. 2290–2294.
42. C. M. Lopez and S. Agaian, "A new set of wavelet-and fractals-based features for gleason grading of prostate cancer histopathology images," in *Image Processing: Algorithms and Systems XI*, vol. 8655 (International Society for Optics and Photonics, 2013), p. 865516.
43. R. Karthiga and K. Narasimhan, "Automated diagnosis of breast cancer using wavelet based entropy features," in *2018 Second international conference on electronics, communication and aerospace technology (ICECA)*, (IEEE, 2018), pp. 274–279.
44. D. Taubman, "High performance scalable image compression with ebcot," IEEE Trans. on Image Process. **9**(7), 1158–1170 (2000).
45. L. D. Chamain and Z. Ding, "Improving deep learning classification of jpeg2000 images over bandlimited networks," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2020), pp. 4062–4066.
46. X. Wang, S. Yang, J. Zhang, M. Wang, J. Zhang, J. Huang, W. Yang, and X. Han, "Transpath: Transformer-based self-supervised learning for histopathological image classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Springer, 2021), pp. 186–195.
47. C. Nguyen, Z. Asad, R. Deng, and Y. Huo, "Evaluating transformer-based semantic segmentation networks for pathological image segmentation," in *Medical Imaging 2022: Image Processing*, vol. 12032 (SPIE, 2022), pp. 942–947.
48. N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European conference on computer vision*, (Springer, 2020), pp. 213–229.
49. M. Adnan, S. Kalra, and H. R. Tizhoosh, "Representation learning of histopathology images using graph neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, (2020), pp. 988–989.
50. P. Pati, G. Jaume, L. A. Fernandes, A. Foncubierta-Rodríguez, F. Feroce, A. M. Anniciello, G. Scognamiglio, N. Brancati, D. Riccio, and M. D. Bonito, "Hact-net: A hierarchical cell-to-tissue graph neural network for histopathological image classification," in *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis*, (Springer, 2020), pp. 208–219.
51. B. Li, M. S. Nelson, O. Savari, A. G. Loeffler, and K. W. Eliceiri, "Differentiation of pancreatic ductal adenocarcinoma and chronic pancreatitis using graph neural networks on histopathology and collagen fiber features," J. Pathol. Inf. **13**, 100158 (2022).
52. J. H. Lee, H. Zhang, V. Lagrange, P. Krishnamoorthy, X. Zhao, and Y. S. Ki, "SmartSSD: FPGA accelerated near-storage data analytics on ssd," IEEE Comput. Arch. Lett. **19**(2), 110–113 (2020).
53. NIH, "GDC — portal.gdc.cancer.gov," GDC Data Portal, 2022, https://portal.gdc.cancer.gov/.