Image and Vision Computing 168 (2026) 105924

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

HOPE: Histopathological image Organization and Processing Environment ☆

Daniel Riccio *, Mara Sangiovanni, Francesco Longobardi, Andrea Francesco Scalella, Vincenzo Manfredi

Department of Electrical Engineering and Information Technology, via Claudio, 21, Naples, 80125, Italy

ARTICLE INFO

Keywords:
Digital pathology
Whole slide images
Fractal compression
Image retrieval
Progressive transmission

ABSTRACT

In disciplines such as digital pathology, the management of vast amounts of data, primarily ultra-high-resolution images, remains a significant barrier to the widespread adoption and seamless sharing of knowledge. Current research efforts are heavily focused on image encoding, often overlooking equally critical aspects such as indexing and efficient content transmission. Traditional compression methods, such as JPEG2000, prioritize reconstruction quality but do not inherently support direct retrieval or progressive transmission, both of which are essential for applications like telemedicine and large-scale digital pathology archives. To bridge this gap, we introduce a novel framework that integrates fractal compression, deep learning-based retrieval, and adaptive transmission, optimizing not only storage efficiency but also accessibility and scalability in histopathological imaging.

The Histopathological image Organization and Processing Environment (HOPE) framework here proposed exploits Partitioned Iterated Function Systems for image compression, achieving high compression ratios while preserving essential structural details. To mitigate the inherent artifacts of fractal compression, a U-Net autoencoder is integrated, refining decompressed images and enhancing visual quality. Additionally, a residual encoding mechanism is employed, allowing for lossless reconstruction when necessary. Unlike conventional methods, this framework enables direct retrieval from the compressed domain by extracting discriminative features from the fractal encoding coefficients. Another key innovation is its progressive transmission capability, which allows an initial low-bitrate preview to be sent, followed by incremental quality refinements based on diagnostic needs. This significantly reduces network load and enables real-time access to high-resolution histopathological images on resource-limited devices. Experimental results demonstrate that the proposed framework achieves compression performance comparable to JPEG2000, while simultaneously enabling efficient indexing, high-accuracy retrieval, and scalable transmission.

1. Introduction

Medical images pose significant challenges for healthcare storage and retrieval systems, not only due to their sheer number but also to their size, as in the case of histopathological (HP) data or 3D acquisition modalities such as Digital Breast Tomosynthesis (DBT). It is essential to efficiently store and process images, and, with the advent of telemedicine, it is crucial to be able to rapidly search, share, and transfer images within different Picture Archiving and Communication Systems (PACS) or among experts worldwide. This is particularly necessary in the digital pathology field, where images are captured at high spatial resolution and are highly complex in terms of color and pattern. HP images are the gold standard for diagnosing several diseases, including almost all types of cancers. Apart from efficiently storing and rapidly retrieving images, pathologists also need tools to effectively search for similar images among already classified ones, in order to facilitate and support the diagnostic process.

Here we present HOPE, Histopathological image Organization and Processing Environment, a framework that leverages a combination of fractal compression and a U-Net autoencoder to allow search, indexing, and retrieval of histopathological images. More specifically, HOPE is composed of three distinct components: (i) the image compression/decompression module, which relies on Partitioned Iterated Function Systems (PIFS) for compressing, coupled with a neural autoencoder to address the reconstruction of lost information during the decompression phase; (ii) an indexing module, in which the fractal compression coefficients are used to perform search and retrieval without fully

☆ This article is part of a Special issue entitled: 'Smart Medicine' published in Image and Vision Computing.

* Corresponding author.
E-mail addresses: daniel.riccio@unina.it (D. Riccio), mara.sangiovanni@unina.it (M. Sangiovanni), francesco.longobardi3@unina.it (F. Longobardi), a.scalella@studenti.unina.it (A.F. Scalella), vi.manfredi@studenti.unina.it (V. Manfredi).

https://doi.org/10.1016/j.imavis.2026.105924
Received 1 March 2025; Received in revised form 11 November 2025; Accepted 29 January 2026
Available online 30 January 2026

decompressing the images; (iii) a transmission module, which manages the information transmission based on the needed detail. We demonstrate that HOPE's fractal codec achieves high compression ratios, thereby significantly reducing storage requirements while maintaining high reconstruction quality. Moreover, HOPE allows for the initial transmission of a low-bitrate compressed representation, followed by progressive quality refinement based on diagnostic needs, significantly reducing network load. The paper is organized as follows: Section 2 introduces the related works; Section 3 describes the three main HOPE modules; Section 4 deals with the experiments designed to test HOPE's performance; lastly, we address the strengths and limitations of the HOPE framework in Section 5.

2. Related work

HP images are obtained by scanning thin slices of biological tissue with an optical microscope to obtain whole-slide images (WSI). To better identify the biological features, images are optically magnified, typically at 20x although, for some sample types, the magnification might reach 100x. These high spatial resolutions correspond to digital uncompressed files that easily reach sizes of several Gigabytes. To allow for a smooth visualization at different magnification levels and in different spatial regions, HP images are usually stored in compressed formats coupled with efficient compressed-data reorganization based on tiling strategies, such as in the Digital Imaging and Communications in Medicine (DICOM) format [1], now a standard in the field. The DICOM standard includes support for several standard lossless compression methods including JPEG-2000, JPIP, HEVC/H.265, with JPEG-2000 achieving the best compression on HP images [2].

Ideally, medical images should be compressed using lossless methods to prevent degradation of details that may be significant for diagnosis. However, the compression performances obtained are very modest, especially on HP images. Moreover, compression might be performed not only to efficiently store and retrieve medical images, but also with the purpose of reducing their size and complexity before different downstream analysis tasks. The constraint of lossless compression has been relaxed by attempting to define the limits of content loss that would permit high performance while maintaining diagnostic accuracy unchanged. For instance, in [3], the authors investigate the potential of applying lossy compression to medical images using the High Efficiency Video Coding (HEVC) format, exploiting the already consolidated medically acceptable compression range for JPEG 2000. The authors successfully tested this approach, obtaining higher compression rates on several image types, including CT and MRI. However, they do not take into account HP images. New directions also investigate the possibility of hybrid approaches in which regions of interest (ROI) are losslessly compressed while extra-ROI portions are subjected to lossy compression [4]. In [5] several compression strategies, including lossless and hybrid approaches, are reviewed. Recently, deep learning (DL) has been widely used to implement compression methods. For instance, in [6], an asymmetric image compression system is proposed: compression is very fast, whereas the decompression phase, obtained with a CNN, is heavier and slower than the compression phase. Although the authors say that the methodology is designed to deal with medical images also, it is not tested on this kind of data. In [7], a hybrid approach is proposed: the authors use the Fuzzy C-Means clustering approach on Magnetic Resonance (MR) images to identify and separate the ROI. An autoencoder method is used for compressing the non-ROI portion, whereas Discrete Cosine Transform with Huffman Run-length encoding is used to compress the region of interest. Many other approaches have been proposed, as thoroughly revised for images in general in [8,9], and with a focus on lossy compression in [10]. In [11], DL methods specifically devised for medical images are reviewed. To the best of our knowledge, only a few works address the challenging problem of compressing HP images. The problem is further amplified by the lack of a common and accepted baseline for comparing different methods [12]. Tellez et al. [13] propose a lossless approach that focuses on preserving the semantically relevant parts of histological gigapixel medical images, with the aim of extracting a reduced but significant representation of the input and then performing classification on it. Hence, no decompression method is proposed. Nagavi et al. [14] use Long Short Term Memories (LSTM) Networks for compression, but only on images at very low resolution, thereby leaving unknown the actual usability of their approach on standard high spatial resolution HP images. Fisher et al. propose Stain Quantized Latent Compression (SQLC), a deep learning approach based on compressing staining and RGB channels before passing it through a compression autoencoder (CAE) [15]. The authors test the method's performance on a downstream classification task, without considering the decompression part.

Another interesting research path involves fractal-based lossy compression/decompression approaches. Fractal Image Compression (FIC) is based on the concept of self-similarity, essentially involving a block-wise search for local similarities within the image using affine transformations, as described in [16,17]. The encoding part is computationally prohibitive if exhaustively performed. Hence, several methods have been proposed to accelerate the process while maintaining high reconstruction accuracy, such as those in [18,19]. The decompression process, however, is very fast. FIC has been successfully applied to medical image compression, as reviewed by [20]. The authors of [21] compare the performances of different fractal compression methods on grayscale medical images (CT, MRI, RX, Ultrasound). The obtained accuracy is very good, although at the price of long compression times, with the best-performing approach being the Fractal Dimension-based fractal image compression (FDFP). In [22], a fast compression algorithm for MR images is proposed, whereas in [23], the authors apply a hybrid approach in which ROIs are losslessly compressed with Context tree weighting and fractal lossy compression is used on non-ROI parts. The method is tested on grayscale medical images. The authors of [24] also apply a hybrid approach in which they exploit fractal compression for DL-obtained ROIs of CT images.

Lastly, a crucial aspect of digital pathology is the retrieval of images similar in content to a given WSI, used to support pathologists on several tasks, including diagnosis and prognosis. Basically, a Content-based image retrieval (CBIR) system is composed of an indexing part, in which features are extracted from the images, and a retrieval part, in which the same features are extracted from the query and used to search among the indexed ones to find similar entries. Zhang et al. [6] were the first to propose a CBIR for pathology images. The authors evaluated similarity based on four image feature types, namely color histogram, image texture, Fourier coefficients, and wavelet coefficients. As a distance metric, they used the vector dot product. However, this approach was tested on small images and is not suitable for large WSIs. More recently, the authors of [25] proposed Yottixel, which uses an indexing algorithm based on dividing the WSI into patches at the lower resolution, clustering them based on their color, and then converting information of representative patches into barcodes using deep encoding. Lastly, the encodings are converted into barcodes to speed retrieval. The work of Lahr et colleagues [26] provides a comprehensive survey of the histopathological image retrieval systems and analyzes the performance, strengths, and limitations of different search methods, including Yottixel and newer deep learning based approaches. The authors push for further research to investigate the dual aspects of minimal storage requirements and accuracy of the returned results: while the latter is still too low for clinical application, the former is often too high to permit a democratic usage of these tools.

Hence, the major challenges in digital pathology are tightly interconnected: the high spatial image resolution and the complex patterns and colors make the compression of HP images a computationally expensive task, in which a delicate balance must be found between elaboration time and diagnostic accuracy of the decompressed image; at the same time, CBIR systems are needed to search and retrieve similar images efficiently, but this need is again hampered by the complexity

of processing these type of data. An equilibrium is also needed between fast indexing and searching and the usefulness of the retrieved images. Last but not least, the large size of HP images also strongly influences the bandwidth requirements needed to ensure their fast transfer, thus hindering the easy retrieval and sharing of such data. Solutions tailored to address all these problems in the specific context of HP images are urgently needed.

3. The proposed framework

The proposed Histopathological image Organization and Processing Environment (HOPE) is designed as a modular architecture. The system is organized into three primary components: a compression module based on *Partitioned Iterated Function Systems* (PIFS), coupled with a neural autoencoder for reconstructing lost information, described in Section 3.1; a retrieval module – based on a combination of a clustering phase and a convolutional neural network trained with the siamese protocol – that operates directly on the fractal compression coefficients, enabling image indexing and retrieval without requiring full decompression, described in Section 3.2; a progressive transmission module that minimizes data traffic by deferring the transmission of additional information until explicitly requested, described in Section 3.3.

More specifically, in the first module each WSI is partitioned into *tiles* of dimension $H \times W$, allowing each tile to be processed independently through the pipeline described below. This initial partitioning not only significantly reduces the computational cost of fractal encoding but also enables localized processing of different image regions, facilitating selective compression, indexing, and transmission of relevant data. During fractal encoding, each tile is segmented into *range* and *domain* blocks and represented through affine transformations, producing a sequence of coefficients that describe the relationship between the blocks. Image decoding is performed by inverting these transformations, generating an initial decompressed version that may suffer from a loss of detail. To mitigate this limitation, HOPE integrates a reconstruction module based on a *CNN autoencoder*, trained to refine the decompressed image and reduce artifacts introduced by fractal compression. However, in some cases, the recovery of information is incomplete, and part of the original content cannot be reconstructed by the model. To ensure *lossless* compression when necessary, the system includes an additional phase where the residual information is identified and encoded separately, allowing its recovery on demand.

In the second module, HOPE introduces an advanced indexing system that directly operates on the fractal encoding coefficients, eliminating the need to decode images for similarity-based retrieval. This approach leverages the fact that PIFS coefficients, in addition to representing compressed information, contain structural descriptors of the image that can be exploited for indexing. The retrieval module in HOPE utilizes these coefficients to construct compact 2D fractal representations of archived images, which are then organized through clustering techniques to optimize search efficiency. A siamese deep neural network, named *PIFSnet*, is trained with a triplet loss to efficiently perform the search step. The result is a scalable and computationally efficient retrieval system that can identify similar images in the database without requiring full decompression.

The last key component of the HOPE architecture is the transmission module, which leverages the decoupling between the fractal codec and the autoencoder module to optimize bandwidth usage during image transfer. In a conventional transmission scenario, the entire image must be sent to the recipient, resulting in significant network resource consumption, particularly for high-resolution images such as HP slides. In contrast, HOPE adopts a progressive transmission strategy: initially, only a compressed version of the image, composed exclusively of PIFS coefficients, is transmitted. If the user requires higher quality, the reconstruction module refines the visualization without transmitting additional data. Only when full lossless quality is requested are the separately encoded residuals transmitted, thereby minimizing the volume of data sent over the network. This adaptive transmission paradigm significantly reduces network load, enhancing the efficiency of image sharing in telemedicine and remote archiving applications.

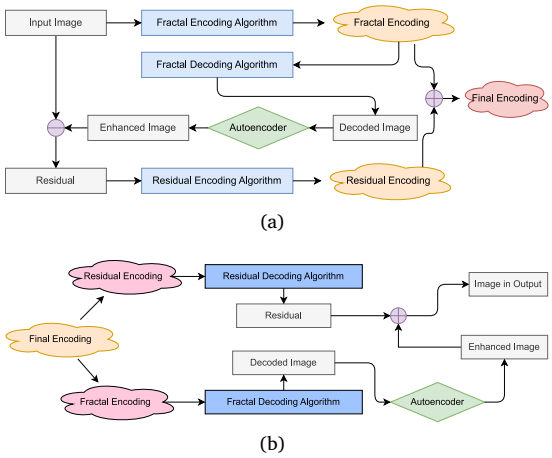

Fig. 1. (a) PIFS-based compression pipeline: the input image undergoes fractal encoding, autoencoder refinement, and residual encoding; (b) PIFS-based decompression pipeline: fractal and residual decoding reconstruct the image, which is further enhanced by an autoencoder.

3.1. Image compression

The HOPE encoder targets high compression ratios, while preserving diagnostically relevant structure and enabling indexing in the compressed domain and progressive, ROI-driven transmission. A schematic of the HOPE encoder and decoder is reported in Fig. 1. As shown in Fig. 1(a), given an input image, the baseline PIFS encoder produces a compact fractal bitstream that serves both compression and compressed-domain indexing. A provisional reconstruction is then obtained by invoking the corresponding fractal decoder within the encoding loop and is subsequently processed by a trained autoencoder to restore structures attenuated by the basic PIFS representation. The difference between the original image and this enhanced reconstruction defines a residual, which is encoded to capture details not recoverable from the fractal stream alone. The final file multiplexes the fractal bitstream, the residual bitstream, and the required headers, thereby supporting efficient storage and progressive delivery.

Fig. 1(b) illustrates the corresponding decoding procedure. The fractal bitstream is first decoded to produce a baseline reconstruction that preserves global appearance and maintains a consistent, indexable structure. The same autoencoder is then applied to enhance this reconstruction and recover diagnostically relevant detail. When higher fidelity is requested, the residual bitstream is decoded and added to the enhanced image, yielding progressively finer reconstructions and, in the limit, lossless quality. This organization preserves interoperability with compressed-domain indexing and retrieval while exposing a controllable trade-off between bitrate and fidelity at decode time.

3.1.1. Fractal codec

Fractal compression is an effective approach for achieving high compression ratios by exploiting the self-similarity properties of images. The method is based on Iterated Function Systems (IFS) [27], which provide a compact representation through contractive transformations that converge to a fractal attractor. An IFS consists of a set of affine geometric transformations that map a figure onto smaller copies of itself, repeating the process iteratively until an autosimilar structure is obtained. This principle allows an image to be encoded using a limited set of parameters rather than storing pixel values directly, making compression highly efficient. However, the classical IFS model is challenging to apply to general image compression, as it requires manually defining the transformations. To overcome this limitation, Jacquin introduced Partitioned IFS (PIFS) [28], where the image is

divided into small square regions, namely range blocks R and domain blocks D, each of size $n \times n$ pixels. For each range $r \in R$, encoding seeks a domain block $d \in D$, an isometry $T \in \mathrm{T}$, and contrast/brightness parameters $\alpha, \beta \in \mathbb{R}$ such that an affine model

$$w(d) = \alpha\, T(d) + \beta \tag{1}$$

approximates r with minimal error. The set T contains the eight planar isometries generated by rotations of multiples of $\pi/2$ and reflections about the horizontal/vertical axes and the two diagonals. Under this formulation, fractal encoding amounts to solving, for each range block r,

$$\min_{\alpha,\beta\in\mathbb{R}} \min_{d\in D, T\in\mathrm{T}} \left\| r - \left(\alpha\, T(d) + \beta\right)\right\|_2^2. \tag{2}$$

For any fixed pair (d, T), the optimal α and β admit closed forms:

$$\alpha^* = \frac{\sum_p \left(r(p) - \bar{r}\right)\left(T(d)(p) - \overline{T(d)}\right)}{\sum_p \left(T(d)(p) - \overline{T(d)}\right)^2}, \qquad \beta^* = \bar{r} - \alpha^* \overline{T(d)}, \tag{3}$$

where the sums run over pixels p in the block, and $\bar{r}$ and $\overline{T(d)}$ denote the mean intensities of r and $T(d)$, respectively. The main computational challenge is the search over $d \in D$ and $T \in \mathrm{T}$ for each r. To keep the bit rate low, the set of range blocks R is defined as a tiling of the image with disjoint blocks (e.g., 16 × 16 pixels), so that only one transform must be stored per range. By contrast, reconstruction quality improves when the candidate domain set D is large: domain blocks are typically extracted with spatial overlap, and thus D can be extremely numerous. An exhaustive search over all $d \in D$ for every $r \in R$ quickly becomes computationally prohibitive on large images, making naive encoding times impractically long. To reduce this cost, heuristic search strategies are adopted. In particular, the Saupe criterion projects each block into a compact feature space and performs nearest-neighbor search in that space rather than directly in the pixel domain. Each block b (either range or domain) is mapped to a normalized feature vector

$$f(b) = \left(\frac{x_1 - \mu_b}{\sigma_b}, \frac{x_2 - \mu_b}{\sigma_b}, \ldots, \frac{x_N - \mu_b}{\sigma_b}\right), \tag{4}$$

with mean and standard deviation

$$\mu_b = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad \sigma_b = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_b)^2}, \tag{5}$$

so that contrast and brightness are normalized out. Because range/domain blocks have a 16 × 16 pixels size, the raw descriptor would be 256-dimensional, which is too expensive for fast search. For this reason, the feature vector is downsampled to a much lower dimension (e.g., $M = 8$ or $M = 16$). All domain descriptors $f(d)$ are then indexed in a KD-Tree [29]. For each range block r, its descriptor $f(r)$ is queried against the KD-Tree to retrieve only a small set of nearest candidate domains,

$$d_{\mathrm{opt}} = \arg\min_{d_i \in D} \left\| f(r) - f(d_i)\right\|, \tag{6}$$

and the full error minimization is performed only on those candidates. In this way, the encoder avoids a full exhaustive search over D while still selecting, for each r, a domain d that closely matches it under an affine transform. This strategy preserves the compression benefits of disjoint range blocks while keeping encoding time tractable even when the domain pool is dense and highly redundant. Most fractal image encoding techniques in the literature focus solely on optimizing the quality of the decoded image. These methods typically employ complex partitioning schemes, such as quad-tree decomposition, to minimize approximation error and enhance visual fidelity. However, in the case of HOPE, fractal encoding is just one of multiple functionalities centered around fractal-based processing. To ensure compatibility with image indexing and selective data transmission, a fixed block size of 16×16 pixels has been adopted for both range and domain blocks. While this choice results in lower initial image quality compared to adaptive partitioning approaches, the degradation is mitigated and corrected through a different mechanism. More importantly, this design enables efficient image retrieval and progressive transmission, which are fundamental aspects of the HOPE system. To improve encoding accuracy, the number of transformations applicable to domain blocks has been increased from 8 to 24. In addition to 3 rotations (90°, 180°, 270°) and 4 reflections (along the axes and diagonals), all possible permutations of the four quadrants of a domain block are considered, yielding a total of 24 transformations. This expansion enhances the approximation of ellipsoidal structures, which are characteristic of histopathological images, improving the perceptual quality of decompressed images.

To optimize storage efficiency, all encoding parameters are quantized and packed into fixed-length records. The image tile of size $H \times W$ is partitioned into disjoint range blocks of side r, yielding

$$n_{\mathrm{range}} = \frac{H \times W}{r^2} \tag{7}$$

rows in a two-dimensional parameter table. Each row stores, with fixed bit budgets, the following fields: (i) contrast coefficient α (5 bits), (ii) brightness coefficient β (6 bits), (iii) domain block coordinates (x_d, y_d) (12 bits total), and (iv) transformation index (5 bits). Fractal decoding reconstructs the image through an iterative application of the contractive maps recorded at encoding time. Starting from an initial flat image of the same size as the original, the decoder uses the current iterate as a source of domain content and, for each range block, applies the stored geometric transform and the associated photometric adjustment before writing the result into the spatial support of the range. Because range blocks are disjoint in the proposed layout, updates proceed without conflicts and the image is refreshed block by block in a stable manner. The first iterations rapidly recover the global appearance, while subsequent iterations stabilize local contrasts and sharpen boundaries as the process is driven toward the fixed point implied by the contractive transforms.

Convergence is monitored by tracking the variation between successive iterates and halting when the change falls below a small tolerance, or when a maximum number of iterations is reached to bound latency. The stopping criterion is defined as:

$$\|I^{(i)} - I^{(i-1)}\| < \epsilon_{\mathrm{decoding}} \tag{8}$$

where $\epsilon_{\mathrm{decoding}}$ is a predefined threshold determining the tolerance for the error between two consecutive iterations.

3.1.2. Image enhancement with the CNN autoencoder

Fixing the range and domain block sizes in the fractal encoding process results in suboptimal image quality upon decoding due to compression-induced artifacts. To mitigate these effects and improve the perceptual quality of reconstructed images, a neural autoencoder is employed [30]. An autoencoder is a neural architecture structured as an Encoder–Decoder (ENC-DEC) model, where the input and output dimensions are identical. This architecture consists of two primary components: the encoder, which reduces the dimensionality of the image while extracting relevant features, and the decoder, which reconstructs the image using this compact representation. The bottleneck layer compresses the information into a latent space, retaining only the most significant characteristics.

In medical imaging, deep learning-based image enhancement is a well-established research field. Fully Convolutional Networks (FCN) are particularly effective for artifact removal and visual quality optimization. Popular models include AlexNet, VGG, GoogLeNet, ResNet, GAN, and U-Net [31,32]. However, given the computational constraints of the proposed system, U-Net was chosen due to its lower complexity, ensuring real-time decoding performance without compromising efficiency. U-Net is characterized by a symmetric "U"-shaped structure, consisting of a contraction path (encoder) and an expansion path (decoder). The encoder applies successive convolution and pooling operations, progressively reducing the spatial resolution while extracting high-level features. The decoder, in contrast, employs transpose

convolutions and upsampling layers to restore the image to its original resolution. The U-Net architecture implemented in this work is a customized version of the original model, where the softmax loss function has been replaced with an error function better suited for intensity regression.

The choice of loss function is a critical aspect of training, as it dictates how network weights are updated. Traditional error metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) evaluate pixel-wise errors independently, disregarding the global structure of the image. Similarly, Peak Signal-to-Noise Ratio (PSNR), a common metric for image quality assessment, suffers from similar limitations, as it measures pixel intensity fidelity without considering structural preservation.

To overcome these limitations, we introduce a loss function that incorporates gradient-based weighting, referred to as weighted PSNR (wPSNR). The underlying idea is to prioritize errors in structurally significant regions, penalizing contour loss more heavily than errors in homogeneous regions.

The construction of this loss function begins with the extraction of the spatial gradient of the image I. The gradient, computed along the x and y directions, provides a measure of local intensity variations, highlighting edges and fine details:

$$\Delta_{x,y}(I) = \left|\frac{\partial I}{\partial x}\right| + \left|\frac{\partial I}{\partial y}\right| \tag{9}$$

This step ensures that regions with significant contrast changes, such as tissue boundaries, are identified and assigned greater relevance in the loss computation.

Once the gradient information is extracted, a normalized weight function is applied to modulate the contribution of different image regions. The weight function is defined as:

$$W = \frac{\Delta_{x,y}(I)}{\max_{x,y}(\Delta_{x,y}(I))} \tag{10}$$

This normalization ensures that areas with the highest gradient magnitudes (i.e., regions rich in structural detail) receive greater emphasis during training, while relatively homogeneous regions have a reduced impact. As a result, the loss function dynamically prioritizes areas where fine-grained texture and contours must be preserved, mitigating the typical smoothing effect introduced by standard MSE-based optimization.

With the weight map defined, the next step involves computing a weighted MSE, where the squared error between the original image I and the reconstructed image I^* is scaled by the corresponding weight values:

$$M = \sum W(I - I^*)^2 / \|I\| \tag{11}$$

Unlike traditional MSE, which treats all pixel errors equally, this formulation ensures that errors in highly structured regions contribute more significantly to the total loss, effectively guiding the network to focus on preserving fine details rather than optimizing for overall intensity fidelity.

Finally, the loss function is reformulated as a modified PSNR metric that accounts for image structure:

$$J(I, I^*) = 100 - 10\log_{10}\left(\frac{\max(I^2)}{M}\right) \tag{12}$$

By integrating a weighted error term, wPSNR prioritizes high-frequency spatial details, ensuring structural accuracy that is an essential requirement in medical imaging. Its dynamically weighted gradient formulation adapts to image structures, enhancing training effectiveness for histopathological image enhancement. This is particularly advantageous where contour sharpness is critical, optimizing the network for both perceptual quality and numerical accuracy.

3.1.3. Residual encoding

Despite the improvements introduced by the autoencoder network, the quality of the reconstructed image may still be insufficient, particularly in regions containing highly complex structural details. To address this limitation, a residual encoding strategy has been implemented, where the residual is defined as the channel-wise difference between the original image and the one reconstructed by the network. This residual preserves critical information necessary for reconstruction fidelity and, unlike standard grayscale or RGB images, can contain negative values. Proper handling of this information is crucial to prevent a significant increase in the file size after compression.

To achieve efficient residual compression, an encoding strategy inspired by the JPEG standard has been developed, leveraging a frequency-domain transformation followed by quantization and entropy coding. Specifically, the residual encoding process involves applying the Discrete Cosine Transform (DCT) [33] to blocks of variable size (8, 16, 32, or 64 pixels), followed by quantization of the coefficients using the same quantization matrix employed for luminance in the JPEG standard. Finally, entropy compression is performed using Run-Length Encoding (RLE) and Huffman coding. The quantization matrix selection allows for control over the compression level and image quality through a scaling factor that adjusts the precision of the quantized coefficients.

To ensure truly lossless encoding, an additional parameter has been introduced to bypass quantization, optimizing bit allocation through a statistical analysis of the minimum and maximum residual values. This approach minimizes the number of bits required for representation while maintaining lossless information fidelity.

The introduction of residual encoding enhances the framework's flexibility, enabling a more dynamic trade-off between image quality and compression efficiency compared to pure fractal encoding. The latter was inherently constrained by a limited number of parameters, such as the number of transformations and the quantization of coefficients. By handling the residual separately, the system can better adapt to application-specific requirements, balancing reconstruction accuracy with storage efficiency while preserving the essential structural details of the decompressed image.

3.1.4. The compressed file structure

The compressed file systematically stores all the necessary information to reconstruct the entire Whole Slide Image (WSI) from its fractal representation, enhanced by the autoencoder and residual encoding. Its structure is designed to reflect the adopted encoding protocol, which involves partitioning the WSI into tiles of dimensions $H \times W \times 3$, separating the R, G, and B channels, applying fractal compression, performing decoding and enhancement through a U-Net architecture, and subsequently encoding and storing the residual information. The resulting file is organized into three main sections: a global header, a fractal encoding section, and a residual encoding section.

The global header contains general information about the WSI and the compression parameters. Specifically, it records the original image dimensions, the subdivision into tiles, and the total number of tiles, along with the parameters related to the chosen compression mode. In the case of lossy encoding, the header includes the quantization parameters used in both fractal compression and residual calculation, whereas in lossless mode, it also stores the minimum residual value required for exact differential reconstruction.

Following the header, the file stores the fractal representation of the image for each R, G, and B channel. Each channel is compressed separately, and its parameters are written sequentially for every tile. The fractal representation consists of a sequence in which each item represents a range block and stores the necessary parameters for reconstructing the original image. Each item contains the fractal transformation coefficients, including the contrast factor α and brightness term β, encoded with 5 bits and 6 bits, respectively. Additionally, for each range block, the coordinates of the associated domain block are stored,

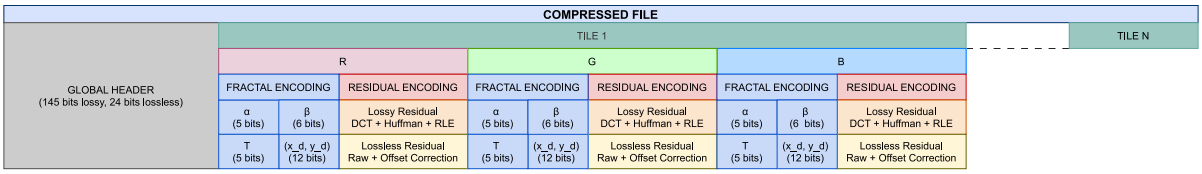

Fig. 2. Schematic representation of the internal structure of the compressed file.

encoded with 6 bits per coordinate, along with the transformation index, represented with 5 bits.

After storing the fractal encoding, the file registers the residual information, computed as the difference between the image reconstructed by the U-Net autoencoder and the original image. The residual data is structured to optimize both progressive transmission and selective retrieval. After the PIFS coefficient section, the residual encoding data is recorded, divided into the residual header and the residual bitstream. The residual header includes: (i) the compression mode (lossy/lossless), (ii) the DCT block size used, (iii) the number of bitstreams generated for each channel, (iv) the minimum residual value (for lossless encoding), and (v) the entropy coding dictionary, containing Huffman tables used for compressing the quantized coefficients. The residual bitstream stores the compressed DCT coefficients for each R, G, and B channel, organized sequentially to ensure optimized access.

This organization enables the selective decoding of only the necessary image portions, eliminating the need to read the entire file. Furthermore, in cases where higher quality refinement is requested, the residual data can be retrieved without requiring redundant fractal decompression, improving the efficiency of progressive image reconstruction.

The data layout in the file follows a precise order, reflecting both the WSI tiling structure and the separation of the three color channels. For each tile, the fractal encoding and residual encoding data for the red channel are stored first, followed by the corresponding data for the green and blue channels. This organization enables efficient access to specific image portions, facilitating the selective decoding of individual regions without requiring the processing of the entire file. Fig. 2 shows the internal structure of the compressed file layout.

The decoding process follows the inverse of the encoding steps and is applied independently for each tile and each channel. Image reconstruction starts with fractal decompression, generating an initial version of the image through the iterative application of the transformations stored in the file. The fractal output is then refined by the U-Net autoencoder, which reduces the artifacts introduced by compression. Finally, the residual information is restored by performing entropy decompression, dequantization, and inverse DCT, which is then added to the autoencoder-enhanced image to obtain the final reconstruction.

The file structure is designed to ensure a high level of modularity and adaptability. The tiling approach enables efficient management of compression and decoding for specific portions of the WSI, optimizing data access and allowing for the selective transmission of regions of interest. The integration of fractal encoding with independent residual handling provides finer control over the trade-off between quality and compression, allowing the system to be adapted to specific application requirements. Compared to pure fractal compression, this strategy better preserves structural information, reduces artifacts, and improves the perceived quality of the reconstructed image.

3.2. Retrieval of fractally encoded histopathological images

The retrieval of histopathological images is a crucial task in computer-assisted diagnosis, as it allows pathologists to access previously diagnosed slides stored in hospital or diagnostic center databases that exhibit similarities to the case under examination. The objective is to facilitate comparisons with past cases, thereby improving diagnostic accuracy and supporting clinical decision-making. However, traditional image indexing and retrieval methods require full decompression of stored images, leading to high computational costs and inefficient resource utilization. In this context, fractal encoding-based retrieval offers a significant advantage by enabling direct search operations in the compressed domain, eliminating the need for explicit image decompression.

An additional benefit of the proposed method is the ability to perform retrieval on arbitrary image regions, even if they were not initially treated as independent tiles during compression. This provides greater flexibility in analyzing archived images. Furthermore, fractal feature extraction enables the construction of rotation-invariant representations, improving robustness to geometric variations in the spatial organization of histopathological structures.

The implementation of the retrieval module is based on three key objectives: operating locally on image data without decompression, extracting spatial features invariant to rotation, and indexing images directly in their compressed format. Over time, fractal compression-based retrieval techniques have evolved from early indexing models, such as FIRST (Fractal Indexing and Retrieval System for Image Databases) [34] and FIRE (Fractal Indexing with Robust Extensions for Image Databases) [35], to more advanced methodologies incorporating deep neural networks. The FIRST system employs a multi-level indexing algorithm that computes the center of mass for all range and domain blocks, while FIRE builds upon FIRST by enhancing retrieval robustness. However, a fundamental limitation of both systems is their inability to preserve spatial information, as they generate histograms of fractal parameters without maintaining the correlation between different regions of the image.

To overcome this limitation, the retrieval system proposed in HOPE reorganizes fractal encoding coefficients and transformations into feature maps, from which descriptors are extracted using a Deep Convolutional Network (DCN). The construction of these maps begins with subdividing the image into range blocks of 16×16 pixels, where each block is characterized by a transformation described by: (i) the contrast coefficient α, (ii) the brightness coefficient, and (iii) the geometric transformation index applied. These parameters, originally derived from fractal compression, are converted into bidimensional representations, which can then be processed by convolutional neural networks.

To ensure that a DCN can extract meaningful features, it must be trained on a large number of annotated images, allowing it to distinguish between similar and dissimilar structures. However, training a deep convolutional network (DCN) to produce similarity-preserving descriptors (i.e., small distances for similar tiles and large distances for dissimilar ones) requires a reference set of images annotated into similarity classes that are distinct from diagnostic categories (e.g., tumor). Obtaining such similarity judgments at the tile level would demand that pathologists manually group images according to perceived resemblance, a task that is prohibitively tedious and time-consuming at the scale considered. To overcome this limitation, an automatic annotation procedure is adopted, organized in three steps and detailed below, which approximates similarity classes without relying on exhaustive human labeling: (i) Extracting features from RGB tiles using a pre-trained network, (ii) clustering these features to generate automatic labels, and (iii) training a new DCNN on the feature maps, using the generated labels.

3.2.1. Feature extraction and clustering

To obtain compact vector representations that enable efficient comparison, 2048-dimensional descriptors are extracted with ResNet50 [36] by removing the classification layer and using the penultimate activations as features. These embeddings place tiles in a metric space where similarity can be approximated by distances between feature vectors.

However, while this representation provides a practical basis for measuring proximity, applying off-the-shelf clustering methods (e.g., k-means or standard agglomerative clustering with Euclidean distance) does not consistently align clusters with human judgments of tile similarity. To bridge feature extraction and grouping more coherently, we first compute an affinity matrix over the embeddings and then apply an ad hoc agglomerative procedure that operates on these affinities. This tailored clustering stage leverages the structure captured by the descriptors while better reflecting perceptual similarity within clusters.

Given a collection of tiles for which descriptors have been extracted with ResNet50, let $\{v_i\}_{i=1}^{N} \subset \mathbb{R}^{2048}$ denote the resulting feature vectors. To compute an affinity matrix over these descriptors, a radius-based neighborhood search is first performed: for each v_i, the algorithm identifies the index set

$$\mathcal{N}_r(i) \;=\; \left\{\, j \neq i \;\middle|\; \|v_i - v_j\|_2 < r \right\},$$

i.e., all vectors within a Euclidean distance r of v_i. This yields, for every i, a list of neighboring vectors capturing local proximity in feature space.

A symmetric affinity matrix $A \in \mathbb{R}^{N\times N}$ is constructed, where each entry quantifies the frequency with which two vectors co-occur in a radius-defined neighborhood:

$$A(i,j) \;=\; \mathbb{1}\big[\, j \in \mathcal{N}_r(i) \big] \;+\; \mathbb{1}\big[\, i \in \mathcal{N}_r(j) \big], \qquad A(i,j) = A(j,i).$$

Equivalently, $A(i,j)$ can be accumulated across all neighborhoods as the number of times (i,j) are jointly retrieved under the range search. The matrix A thus summarizes the local similarity structure induced by the ResNet50 embeddings and serves as input to the subsequent clustering stage and to retrieval refinement procedures.

The agglomerative clustering process iterates over the affinity matrix A using a predefined threshold th_A to progressively merge clusters until the maximum affinity in the matrix falls below th_A. Initially, each sample is assigned to its own cluster. In each iteration, the pair of clusters corresponding to the highest affinity in A is merged, and affinities between clusters are updated based on the median affinity of the merged elements. This process continues until no affinity value exceeds th_A, producing the final clustering assignment C.

The obtained clusters and affinity relationships can be leveraged for two primary purposes. The first is retrieval performance evaluation: tiles that fall into an affine class in response to a query are not considered entirely distinct but weighted according to their affinity degree in matrix A. This allows the computation of a query affinity index, providing a more accurate measure of image similarity. The second objective is training a deep convolutional neural network (DCNN), named PIFSNet, on the feature maps.

3.2.2. PIFSNet training

PIFSNet is a convolutional backbone trained in a siamese setting [37] with a triplet loss [38]. Two identical branches, sharing all weights, process a pair (or triplet) of inputs and produce fixed-length embeddings; the loss pulls together embeddings of tiles deemed similar and pushes apart those of dissimilar tiles. This yields a metric representation suitable for nearest-neighbor retrieval.

Each branch follows the residual design illustrated in Fig. 3. An Input Block ingests the multi-channel feature maps derived from the fractal codec and applies a 3×3 convolution, batch normalization, and ReLU. This is followed by a stack of six Residual Blocks, each comprising 3×3 convolution → batch normalization → ReLU with an identity skip connection summed at the block output. An Output Block completes the backbone with global average pooling, dropout, and a fully-connected layer that maps to the embedding space of dimension d (optionally ℓ_2-normalized at test time). The residual topology stabilizes optimization and preserves spatial coherence across the tile grid.

This structure is optimized to learn the relationships among fractal transformations applied to tiles, preserving spatial coherence and improving retrieval quality. As a result, the model significantly reduces computational costs associated with search operations, maintaining high efficiency in indexing and accessing compressed histopathological image databases.

3.3. Efficient transmission of histopathological images

The sharing of histopathological images among different healthcare institutions is a key factor in improving diagnostic quality and ensuring equitable access to specialized resources. Central institutions, such as major hospitals and research centers, often possess extensive datasets and advanced image analysis tools. In contrast, peripheral facilities, which typically handle less diverse cases, could greatly benefit from the ability to compare their cases with those stored in specialized databases. However, transferring Whole Slide Images (WSI), which can reach sizes of GB per image, over bandwidth-limited networks or managing them on computationally constrained devices, such as tablets or smartphones, presents significant challenges in terms of transmission efficiency and usability.

The HOPE system addresses these challenges by implementing an optimized transmission protocol, as presented in Fig. 4, which allows for progressive access to compressed image data, thereby eliminating the need for full image transfer at the outset. This approach significantly reduces access times, bandwidth usage, and computational load on remote devices. The protocol consists of the following key steps:

- Initial Query – The remote device sends a request in the form of a numerical vector, which is used to identify the most similar images in the database without requiring the transmission or decompression of entire WSIs.
- Fractal Encoding (PIFS) Transmission – Instead of transmitting the entire image, the system sends a sequence of coefficients derived from fractal encoding, enabling the generation of a compressed preview of the image.
- Preliminary Decoding and Enhancement with U-Net – The receiving device generates a preliminary decompressed image and applies a convolutional autoencoder (U-Net) to correct fractal compression artifacts, improving visual quality.
- Refinement Request (ROI Selection) – The user can select Regions of Interest (ROI) requiring higher quality. At this stage, the device sends a request containing the indices of the blocks for which residual information is needed.
- Selective Residual Block Transmission – Only the compressed residual coefficients corresponding to the requested blocks are transmitted, allowing high-quality reconstruction of the selected portion without requiring the transfer of the entire WSI.

The adoption of this protocol offers several advantages over conventional histopathological image transmission. First, bandwidth consumption is significantly reduced, as progressive transmission allows the system to send only fractal coefficients initially (few MegaBytes per image, compared to GigaBytes for a full WSI). Only specific regions requested by the user are subsequently transmitted, further optimizing data transfer. Second, the protocol enables real-time image visualization on computationally constrained devices, such as smartphones or tablets. The preview generated by PIFS with the U-Net decoding provides a diagnostically acceptable quality level without requiring the download of the entire image. From a practical perspective, the HOPE protocol allows a pathologist in a small hospital to remotely access a specialized database, identifying images similar to the case under examination without waiting for large file transfers. Similarly, a clinician

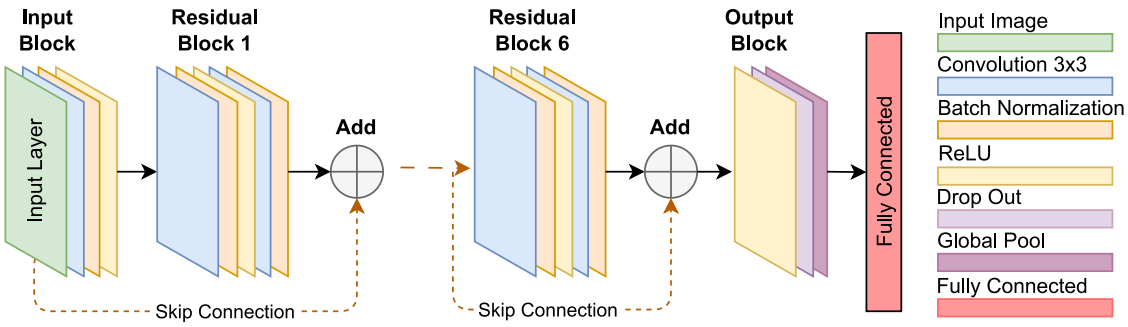

Fig. 3. Schematic representation of the residual PIFSNet backbone.

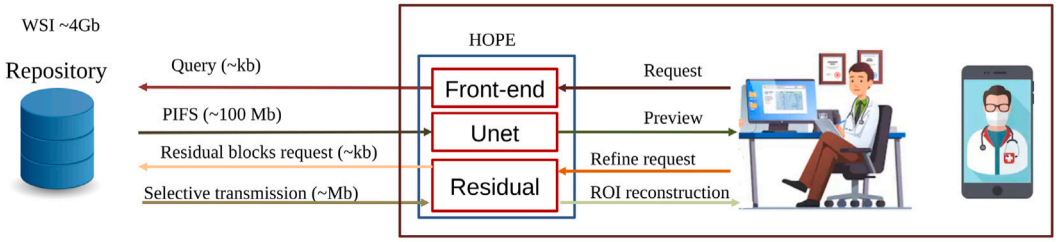

Fig. 4. Schematic representation of the image transmission protocol implemented in HOPE.

using a mobile device can access a preview of archived images and request full-resolution details only when necessary. By implementing this progressive and selective transmission strategy, HOPE effectively addresses the challenges associated with histopathological image transfer, achieving an optimal balance between transmission efficiency and visualization quality. This approach ensures that the system remains scalable and accessible, even in low-bandwidth environments or on computationally limited devices.

4. Experiments

The HOPE system was designed to ensure computational efficiency and scalability in the management of high-resolution histopathological images. The experiments conducted aimed to evaluate the performance of the compression, retrieval, and transmission pipeline, comparing the proposed approach with reference techniques available in the literature.

The system was developed using a high-performance computing infrastructure, leveraging x86-64 architecture servers with 20 cores and 128 GB of RAM, equipped with Intel Core i9-7900X processors operating at a maximum frequency of 4.5 GHz. The entire framework was implemented in Python, utilizing advanced libraries for image processing, including OpenCV, NumPy, and SciPy for preprocessing operations, as well as PyTorch and TensorFlow for training and inference of deep neural networks.

For fractal compression, custom-optimized libraries were developed to handle Partitioned Iterated Function Systems (PIFS). The search within the fractal coefficients was accelerated using KD-Tree data structures [29], significantly improving retrieval times. Additionally, the storage and management of HP images were performed using HDF5 databases, allowing efficient access to large volumes of data.

4.1. The BACH image dataset

The BACH (Breast Cancer Histology) dataset, used for the experimental evaluation of the HOPE system, was released as part of the ICIAR 2018 Grand Challenge and contains annotated histopathological images for the analysis and classification of breast carcinoma [39]. The images are categorized into four diagnostic classes: normal, benign, in situ, and invasive. The normal class represents breast tissue without abnormalities, while the benign class includes alterations in normal mammary development processes that are not indicative of tumor pathology. The in situ category identifies pre-invasive tumors where cancerous cells remain confined within ducts or lobules without infiltrating the surrounding tissue, whereas the invasive class encompasses malignant tumors with the ability to spread to adjacent tissues and other parts of the body. The dataset comprises 400 training images, evenly distributed across the four classes, and 100 test images. All images were acquired in TIFF format with a resolution of 2048×1536 pixels and a spatial scale of $0.42\ \mu m \times 0.42\ \mu m$ per pixel. The acquisitions were performed between 2014 and 2017 using a Leica DM 2000 LED microscope and a Leica ICC50 HD camera. The samples originate from three hospitals in Portugal: Hospital CUF Porto, Centro Hospitalar do Tâmega e Sousa, and Centro Hospitalar Cova da Beira. The images were annotated by two expert pathologists from the Institute of Molecular Pathology and Immunology (IPATIMUP) and the Institute for Research and Innovation in Health (i3S), with ambiguous cases being further reviewed through immunohistochemical analysis. To optimize memory usage during the fractal encoding phase, each image was subdivided into 12 tiles of 512×512 pixels. This partitioning facilitated a more efficient data management strategy, preserving the original resolution while ensuring improved processing during compression and indexing.

4.2. Evaluation of the compression method

Assessing the quality of compressed images is a fundamental step in evaluating the effectiveness of the HOPE system. In the conducted experiments, the objective metrics adopted include the Structural Similarity Index (SSIM), the Peak Signal-to-Noise Ratio (PSNR), the Root Mean Squared Error (RMSE), and a variant of PSNR, referred to as Weighted-PSNR (WPSNR), which accounts for image structure to provide a more perceptually accurate quality assessment.

The SSIM index is designed to evaluate image quality by considering three key components: luminance, contrast, and structure. Given two images A and B, SSIM is defined as:

$$SSIM(A,B) = \frac{(2\mu_A\mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)} \tag{13}$$

where μ_A and μ_B represent the mean intensity values of A and B, σ_A^2 and σ_B^2 denote their variances, and σ_{AB} is the covariance between the two images. The constants C_1 and C_2 prevent numerical instability and are defined as $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$, where $L = 65536$ for 16-bit images and $K_1, K_2 \ll 1$.

The Root Mean Squared Error (RMSE) quantifies the average squared intensity differences between the pixels of two images and is computed as:

$$RMSE(A, B) = \sqrt{\frac{1}{nm} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [A(i, j) - B(i, j)]^2} \tag{14}$$

where m and n represent the image dimensions, while $A(i, j)$ and $B(i, j)$ denote the pixel intensity values at location (i, j) in images A and B. RMSE provides a direct measure of the average error between pixels but does not account for perceptual differences. The Peak Signal-to-Noise Ratio (PSNR) is a widely used metric for assessing the quality of compressed images and is defined as:

$$PSNR(A, B) = 20 \log_{10} \left(\frac{L - 1}{RMSE(A, B)} \right) \tag{15}$$

where $\max(A)$ represents the maximum pixel intensity in image A, and A is considered the original image. Higher PSNR values indicate better reconstruction quality; however, PSNR is limited in its correlation with human visual perception. To address the limitations of PSNR, which tends to underestimate the perceptual impact of structural detail loss, a modified metric called Weighted-PSNR (WPSNR) has been developed. This metric incorporates a weighting factor based on image gradient variations, emphasizing errors in high-contrast regions while reducing their impact in homogeneous areas. The image gradient is computed along the x and y directions as:

$$WI = 0.5 \times (|\nabla_x I| + |\nabla_y I|) \tag{16}$$

where $\nabla_x I$ and $\nabla_y I$ represent the gradients along the horizontal and vertical axes. The weighted RMSE (WRMSE) is then defined as:

$$WRMSE(A, B) = \sqrt{\frac{1}{nm} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} W_A(i, j) W_B(i, j) [A(i, j) - B(i, j)]^2} \tag{17}$$

The final WPSNR formulation is:

$$WPSNR(A, B) = 20 \log_{10} \left(\frac{L - 1}{WRMSE(A, B)} \right) \tag{18}$$

This approach provides a more perceptually accurate evaluation by penalizing the loss of structural details more significantly than traditional PSNR. Finally, the Compression Ratio (CR) was used to assess the efficiency of the compression method. CR is defined as the ratio between the original file size and the compressed file size. Higher CR values indicate greater storage savings, but they must be balanced with reconstruction quality to ensure sufficient diagnostic accuracy. The adoption of these metrics enables a comprehensive analysis of the trade-off between reconstruction quality and compression efficiency, allowing for a comparison of the HOPE system with traditional compression techniques. Furthermore, it facilitates the evaluation of the contribution of the autoencoder and residual coding in improving the quality of decompressed images.

Evaluation of the Fractal Codec. To assess the performance of the fractal codec implemented in the HOPE system, an experimental evaluation was conducted using a subset of the BACH dataset. A total of 120 original images were selected, evenly distributed across the four diagnostic classes (Benign, InSitu, Invasive, Normal), with 30 images per class. Each image was further divided into 12 tiles of 512×512 pixels, resulting in a total of 1440 tiles subjected to fractal encoding.

The average results, along with their standard deviations, are presented in Table 1. The Compression Ratio (CR) confirms the effectiveness of the fractal approach in significantly reducing image size, while the values of WPSNR and SSIM reflect the loss of quality associated with compression. The relatively high standard deviation of WRMSE suggests variability in reconstruction quality depending on the structural complexity of the original images.

Table 1
Average performance of the fractal codec with standard deviation.

Metric	Mean value	Standard deviation
WPSNR	25.09	2.76
WRMSE	3790.93	1438.66
SSIM	0.66	0.07
CR	157.53	12.75

Table 2
Performance comparison between U-Net and Pix2Pix on the test set.

Model	WPSNR (↑)	WRMSE (↓)	SSIM (↑)
U-Net	29.73 ± 1.82	2124.88 ± 478.79	0.82 ± 0.03
Pix2Pix	28.61 ± 1.73	2411.80 ± 519.70	0.77 ± 0.03

Evaluation of the Autoencoder. Further experiments were conducted to evaluate the performance of the UNet autoencoder, which was employed to enhance the quality of images reconstructed after fractal decoding. Since UNet requires supervised training, the dataset was partitioned into a training and a testing set to assess the network's ability to restore lost details. The training set was built using the tiles obtained from fractal decoding as input and the corresponding original tiles as target images. Specifically, 20 images from each of the four classes of the BACH dataset were selected, totaling 80 images. Each image was divided into 12 tiles, generating 960 input-target pairs for training, while the test set consisted of 480 tiles from an additional 40 images, ensuring a balanced class distribution.

To evaluate UNet's performance, it was compared with Pix2Pix [40], a well-established image enhancement model in the literature. Pix2Pix is based on Generative Adversarial Networks (GANs) [41], where a generator produces realistic images from degraded inputs, while a discriminator assesses image quality, iteratively improving the reconstruction. Unlike UNet, which optimizes pixel-wise reconstruction through error minimization, Pix2Pix introduces an adversarial component, which better preserves image structures.

To ensure effective training, a careful data preparation strategy was applied. Both models require input sizes smaller than the original tiles, necessitating a patching process. Additionally, data augmentation was performed, including random 90-degree rotations and vertical reflections applied to both input and target images. Finally, data normalization was required, and training was performed in a batch mode, where network weights were updated after processing a complete original image. After training, the effectiveness of UNet and Pix2Pix was evaluated on a test set consisting of 480 tiles extracted from 40 original images, with 10 images per class. The numerical results of this experiment are presented in Table 2. UNet consistently outperforms Pix2Pix across all considered metrics. Notably, UNet achieves a WPSNR that is approximately 1.1 dB higher than that of Pix2Pix, indicating better fidelity in reconstructed images. Additionally, WRMSE is lower for UNet, suggesting a reduced reconstruction error. Finally, the SSIM confirms the advantage of UNet, achieving 0.82 compared to 0.77 for Pix2Pix.

The analysis of these results suggests that UNet is more effective in restoring information lost during fractal compression. This can be attributed to its symmetric architecture, which better preserves local features, whereas Pix2Pix, leveraging a GAN-based approach, tends to generate realistic reconstructions but introduces higher variability. The higher standard deviation in Pix2Pix's results across all metrics suggests that the network is more prone to introducing artifacts that were not present in the original image. Overall, UNet proves to be the more suitable choice for enhancing compressed histopathological

Table 3
Comparison of WPSNR, SSIM, and WRMSE across different compression methods.

CR	WPSNR (dB)			SSIM			WRMSE		
	JPEG	JPEG2000	HOPE	JPEG	JPEG2000	HOPE	JPEG	JPEG2000	HOPE
5	–	67.1	61.78	–	0.99	0.99	–	29	53
10	44.8	55.1	50.73	0.98	0.98	0.96	377	115	191
20	39.2	42.5	39.13	0.96	0.97	0.94	719	491	724
50	34.9	36.7	33.79	0.92	0.95	0.92	1179	958	1340
100	30.2	34.9	32.13	0.86	0.91	0.88	2025	1179	1622
130	28.8	33.7	31.03	0.84	0.89	0.86	2379	1354	1841
160	27.5	32.5	30.01	0.79	0.87	0.84	2764	1554	2070

images, ensuring a more stable and faithful reconstruction compared to Pix2Pix.

Evaluation of the whole compression pipeline. To assess the effectiveness of the complete pipeline (i.e. PIFS,UNet, and Residual Coding), we conducted a comparative evaluation against JPEG [42] and JPEG2000 [43], two widely adopted image compression standards in medical imaging. The numerical results in terms of WRMSE, WPSNR, and SSIM at varying Compression Ratios are reported in Table 3. JPEG represents a baseline method known for its computational efficiency but is limited in terms of quality preservation at high compression rates. JPEG2000, on the other hand, is an advanced wavelet-based codec that achieves superior rate-distortion performance and is extensively used in DICOM-compliant medical imaging systems.

The comparative analysis between JPEG, JPEG2000, and HOPE shows that JPEG2000 achieves slightly superior performance in terms of WPSNR, SSIM, and WRMSE. However, HOPE closely approaches JPEG2000's results, maintaining high-quality metrics suitable for histopathological analysis. JPEG, on the other hand, exhibits lower performance, highlighting its limitations in preserving structural image details. Considering the balance between reconstruction quality, transmission efficiency, and retrieval capabilities, HOPE emerges as a competitive solution for large-scale histopathological image management. While JPEG2000 maintains a slight advantage in quality, HOPE's additional functionalities make it a promising choice for applications where efficient access and intelligent image handling are crucial.

Progressive ROI refinement experiment. To complement the full-pipeline evaluation, we analyze the behavior of the systems in a progressive Region of Interest (ROI) delivery setting. In this protocol, a low-bitrate preview is first transmitted to the client. Subsequently, a square ROI of size 512 × 512 pixels is requested at increasing fidelity targets measured by WPSNR. For HOPE (PIFS + UNet + residual), refinement is obtained by transmitting only a residual stream for the ROI computed with respect to the already available preview. For comparison purposes, JPEG2000 receives a full-image preview at the same compression ratio as the initial PIFS preview, whereas for the selected ROI only the residual is transmitted to estimate the additional payload required to reach the target WPSNR. The additional payload required to refine the ROI, expressed as bit per pixel (bpp), on the ROI pixels, i.e., (bpp_{ROI}) and the corresponding SSIM are reported, for targets from 40 to 64 dB WPSNR, in Table 4 (mean ± std over the test set). Fig. 5 presents a qualitative comparison on a representative ROI (a). The top row (b) displays the HOPE pipeline, which includes PIFS decoding and UNET enhancement, alongside a JPEG2000 preview of the same input tile. The bottom panel (c) displays the reconstructions refined to target WPSNRs of 35, 40, 50, and 60 dB for HOPE (top line) and JPWG2000 (bottom line). For each panel, WPSNR and SSIM are reported. The example indicates sharper nuclear boundaries and tissue textures in the HOPE preview and comparable or superior detail recovery at higher targets, consistent with the quantitative payload trends in Table 4. Three observations follow from Table 4. First, in the clinically relevant range 40–54 dB, HOPE requires consistently less ROI payload than JPEG2000 to achieve the same target, while attaining a higher SSIM at every operating point. This suggests that the PIFS base layer, further enhanced by the UNet, already conveys most of the low/mid-frequency morphology, so the residual is compact and concentrated on the remaining high-frequency corrections. Second, as the target approaches the near-lossless regime (62–64 dB), JPEG2000 becomes slightly more bit-efficient on average. This is expected due to the fine-grain scalability of its wavelet progression and entropy coding; nevertheless, HOPE remains close and maintains SSIM on par with or above JPEG2000. Third, the standard deviations increase with the target for both methods, reflecting the heterogeneity of tissue micro-structures: ROIs with dense, high-contrast detail require more refinement bits regardless of the codec.

Table 4
ROI refinement payload per target WPSNR (ROI: 512 × 512). Values are mean ± std on the test set.

Target	PIFS		JPEG2000	
WPSNR (dB)	bpp_ROI	SSIM	bpp_ROI	SSIM
40	0.006 ± 0.018	0.960 ± 0.016	0.049 ± 0.046	0.923 ± 0.028
42	0.022 ± 0.041	0.962 ± 0.014	0.074 ± 0.087	0.931 ± 0.026
44	0.053 ± 0.095	0.965 ± 0.011	0.115 ± 0.143	0.940 ± 0.025
46	0.129 ± 0.192	0.970 ± 0.010	0.183 ± 0.226	0.950 ± 0.026
48	0.226 ± 0.323	0.973 ± 0.011	0.301 ± 0.359	0.962 ± 0.022
50	0.367 ± 0.463	0.977 ± 0.010	0.435 ± 0.446	0.973 ± 0.015
52	0.569 ± 0.588	0.981 ± 0.010	0.694 ± 0.635	0.980 ± 0.012
54	0.919 ± 0.883	0.985 ± 0.009	0.955 ± 0.783	0.985 ± 0.010
56	1.306 ± 1.093	0.989 ± 0.008	1.552 ± 1.656	0.989 ± 0.009
58	1.779 ± 1.572	0.992 ± 0.007	1.906 ± 1.552	0.992 ± 0.006
60	2.257 ± 1.394	0.994 ± 0.005	2.352 ± 1.364	0.994 ± 0.005
62	3.613 ± 3.028	0.996 ± 0.003	3.461 ± 1.494	0.997 ± 0.003
64	4.586 ± 2.699	0.998 ± 0.002	3.997 ± 1.243	0.998 ± 0.002

Beyond the small advantage of JPEG2000 at very high targets, HOPE offers two practical benefits. (i) At typical operating points, it achieves lower incremental ROI bitrate for the same WPSNR while delivering higher structural fidelity (SSIM). (ii) The PIFS bitstream preserves a compressed-domain structure amenable to indexing and retrieval without re-encoding the base layer. Consequently, HOPE supports efficient preview, search, and progressive refinement within a unified framework, making it attractive for large-scale histopathology workflows where rapid access and interactive inspection of diagnostically relevant regions are crucial.

4.3. Evaluation of the indexing method

One of the main challenges in organizing tiles for histopathological image retrieval is the lack of explicit annotations defining the similarity and dissimilarity between tiles in publicly available datasets. The BACH dataset, for example, is annotated for only four diagnostic classes at the whole-slide image level. These labels do not specify localized relationships between different regions within an image or across different images, limiting the applicability of supervised retrieval methods. Without explicit ground truth annotations for tile similarity, it is not possible to directly train a model to map visually similar regions into a common feature space. To overcome this limitation, we introduce an automated annotation strategy based on an affinity-based agglomerative clustering algorithm, as described in Section 3.2.1. The objective is to group tiles based on their visual content and construct an affinity matrix, which

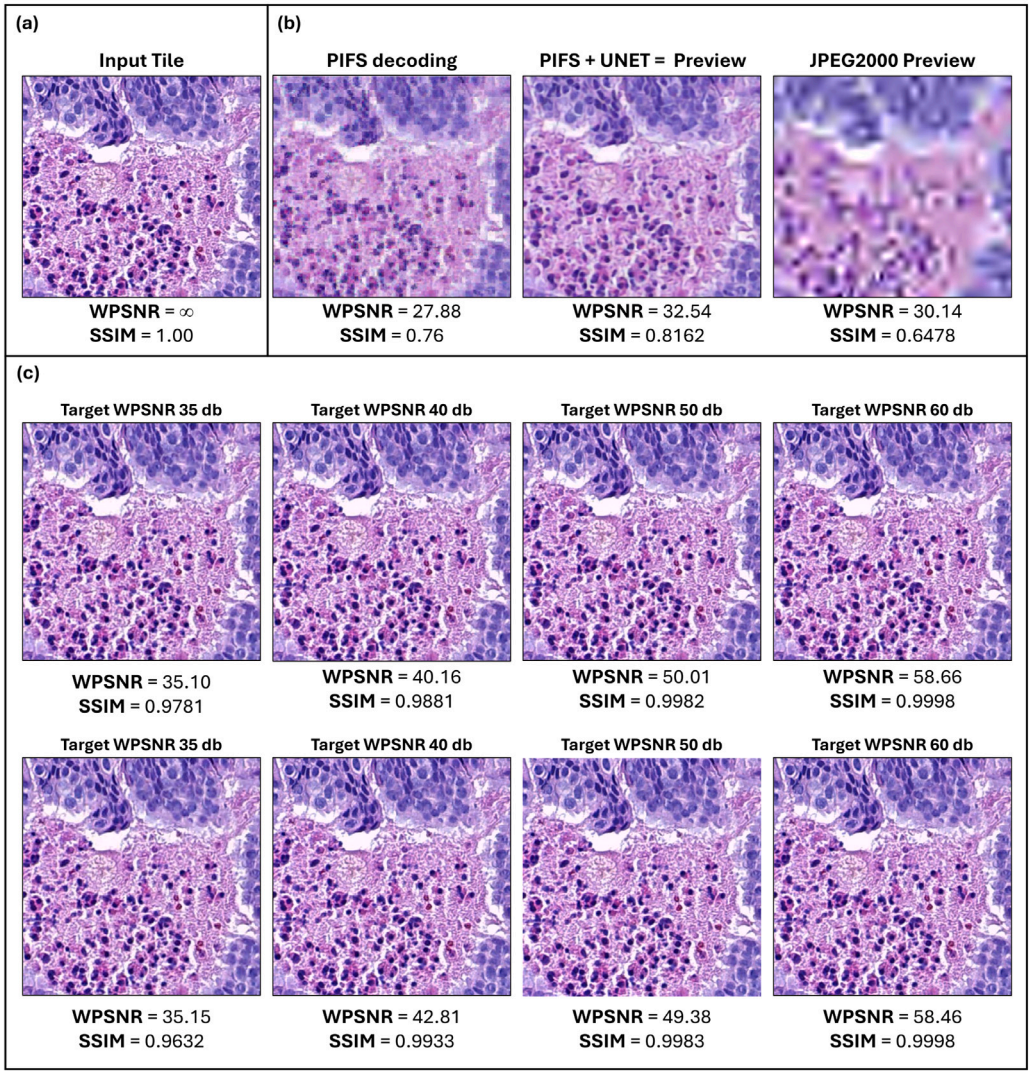

Fig. 5. Qualitative illustration of preview and ROI refinement. (a) Input ROI (512 × 512 24 bpp) cropped from the WSI. (b) Left-to-right: PIFS decoding; PIFS+UNet (i.e., the HOPE client-side preview reconstructed from the fractal stream); and the JPEG2000 preview obtained by transmitting the same preview bitrate as HOPE. (c) Top row: examples of progressive refinement with HOPE, obtained by transmitting only the residual needed to reach target WPSNR levels of 35, 40, 50, and 60 dB. Bottom row: corresponding refinements for JPEG2000 at the same targets, achieved by ROI re-encoding. For each panel, the achieved WPSNR and SSIM are reported below the image.

serves as a proxy for human-labeled similarity. This matrix enables us to evaluate how well a retrieval system can recover images belonging to perceptually coherent clusters. Our ultimate goal is to train a model capable of embedding images in the fractal domain such that the distance between embeddings reflects the visual similarity observed in the original pixel space. Specifically, if two tiles are visually similar in terms of pixel content, the model should produce embedding vectors that are close in feature space, even though they are derived from the fractal transformation coefficients rather than raw pixel intensities.

To construct a robust evaluation framework, we partitioned the tiles extracted from the BACH dataset into two disjoint subsets, ensuring that tiles extracted from a given image are either entirely in the training set or entirely in the test set. Out of the 400 images in BACH, we allocated 280 images for training and 120 images for testing, yielding 3360 training tiles and 1440 test tiles. For feature extraction, we employ a ResNet50 model [36], modified by removing the softmax and fully connected layers, so that each input tile is mapped to a feature vector (embedding) that captures its structural and morphological characteristics. To establish a structured embedding space, we apply agglomerative clustering to the training embeddings generated by ResNet50, obtaining both the cluster assignments and the affinity matrix for the training data. The PIFSNet model is then trained using the fractal encodings of the tiles as input while leveraging the ResNet50 embeddings as the target feature space. As described in Section 3.2.2, the model was trained using a triplet-loss protocol, where each triplet of tiles (x_a, x_p, x_n) consists of an anchor tile x_a, a positive tile x_p (similar to x_a), and a negative tile x_n (dissimilar to x_a). During training, the network is optimized to produce embeddings that preserve the same similarity relationships observed in the original image domain. For evaluation, we apply the same agglomerative clustering process to the test set embeddings generated by ResNet50, yielding clusters and an affinity matrix for the test set.

To assess the quality of the features extracted by PIFSNet, we compared them against two well-established techniques — Generic Fourier Descriptor (GFD) [44], which uses the Fourier transform to capture geometric and structural properties, and Bag of Visual Words (BoVW) [45], which represents images as histograms of quantized local descriptors — and against Yottixel [25], a large-scale content-based retrieval engine for histopathology whole-slide images that indexes WSIs with compact binary "barcodes" to enable efficient nearest-neighbor search.

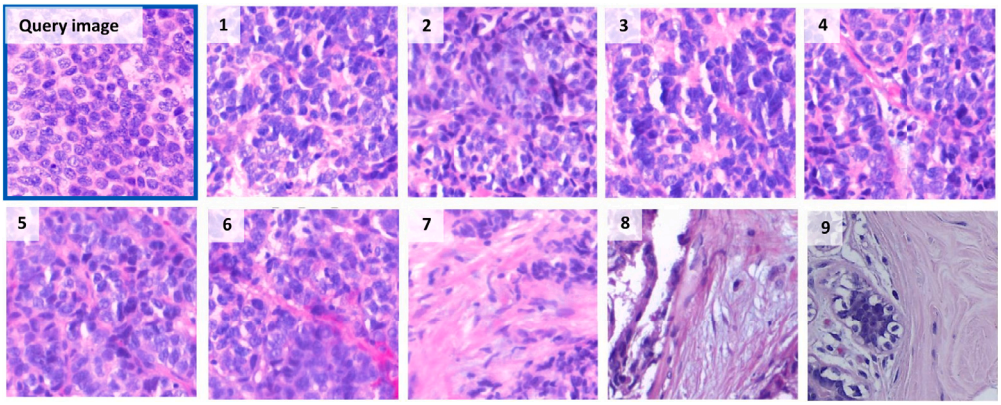

Fig. 6. Example of content-based image retrieval using features extracted by PIFSNet. The first image represents the query, followed by the retrieved images ranked by similarity.

Notice that, in this case, accuracy was not considered as a retrieval quality metric, as it fails to describe the relationship between clusters. Accuracy does not account for similarity between different clusters; it simply evaluates whether two tiles belong to the same cluster or not, implementing an overly rigid selection criterion unsuitable for retrieval tasks. To address this limitation, we leverage affinity-based clustering to define a pairwise score, hereafter referred to as *affinity*, computed from the neighborhood structure induced by the feature embeddings. Let $v_k \in \mathbb{R}^d$ denote the feature vector of tile k, and let $N(k)$ be the index set of its neighbors in feature space (e.g., k-NN or all vectors within a radius r). We use $\mathrm{sim}(\cdot,\cdot)$ to denote a dissimilarity (distance) between feature vectors and $\theta > 0$ a fixed distance threshold. The indicator $\mathbb{1}[\cdot]$ equals 1 if its condition is true and 0 otherwise.

With these definitions, the affinity between tiles i and j is

$$A(i,j) = \frac{1}{|N(i)| + |N(j)|} \sum_{v_i \in N(i)} \sum_{v_j \in N(j)} \mathbb{1}(\mathrm{sim}(v_i, v_j) < \theta) \tag{19}$$

This metric allows cases where a test tile is assigned to a cluster different from its training counterparts to be considered, as long as they still share a high structural similarity. Using the test set affinity matrix, we calculate the affinity scores achieved by each of the three retrieval techniques, providing a quantitative comparison of their ability to effectively group structurally similar tiles. A comparative analysis of features extracted with different methods confirmed the superiority of PIFSNet, which achieved an average affinity score of 0.71, outperforming Yottixel (0.65), Bag of Visual Words (BoVW, 0.52), and Generic Fourier Descriptor (GFD, 0.45). Fig. 6 illustrates an example of a query tile submitted to the HOPE system and the corresponding retrieved tiles.

To evaluate the diagnostic relevance of retrieved images, a pathologist study was conducted. Using Yottixel, GFD, BoVW, and PIFSNet descriptors, queries were submitted to the system, consisting of a reference image and nine retrieved images. Pathologists were asked to assign a relevance score from 1 to 10 to each retrieved image based on its diagnostic utility. The pathologists' evaluation indicated that the PIFSNet-based retrieval returns images that are more diagnostically useful than the alternatives. Average usefulness scores (higher is better) were: PIFSNet 6.13, Yottixel 4.10, GFD 3.86, and BoVW 2.22. The comparatively low scores of GFD and, especially, BoVW highlight difficulties in forming diagnostically coherent clusters, whereas PIFSNet attains the highest rating, reflecting its superior retrieval of histopathologically relevant images.

5. Conclusions

This work presented HOPE, an innovative framework for the compression, indexing, and efficient transmission of histopathological images. The system leverages fractal compression based on Partitioned Iterated Function Systems (PIFS) and integrates a neural autoencoder to enhance reconstruction quality, along with a residual encoding technique that enables lossless compression when necessary. Additionally, HOPE incorporates a deep learning-based retrieval mechanism, allowing for the retrieval of similar images directly in the compressed domain.

The experimental evaluation demonstrated that the fractal codec in HOPE achieves high compression ratios, significantly reducing storage requirements without substantially compromising reconstruction quality. The integration of the U-Net autoencoder further improved the visual quality of decompressed images.

From an indexing and retrieval perspective, HOPE introduced an innovative strategy that directly utilizes the fractal compression coefficients to extract representative features, eliminating the need for full decompression. The PIFSNet model, trained with a triplet-loss strategy, outperformed not only traditional feature extraction methods such as Generic Fourier Descriptor (GFD) and Bag of Visual Words (BoVW), but also more recent approaches like Yottixel, so demonstrating superior capability in preserving the cellular structure of histopathological images. Another key advantage of the HOPE framework is its ability to support progressive transmission. Unlike conventional compression methods, HOPE enables the initial transmission of a low-bitrate compressed representation, followed by progressive quality refinement based on diagnostic needs, thereby significantly reducing network load. Although the WPSNR and SSIM values achieved by HOPE are slightly lower than those of JPEG2000, its ability to integrate compression, direct retrieval, and adaptive transmission makes it a highly competitive and functional system for telemedicine applications and the management of large-scale histopathological image archives.

The results obtained confirm that HOPE represents a promising approach for optimizing the management of large-scale biomedical image databases. Future improvements may include further optimization of the retrieval process, the integration of more advanced neural networks for image enhancement, and the extension of the framework to support multi-modal medical imaging analysis.

CRediT authorship contribution statement

Daniel Riccio: Software, Methodology, Conceptualization. **Mara Sangiovanni:** Writing – review & editing, Writing – original draft. **Francesco Longobardi:** Writing – review & editing, Writing – original draft, Software. **Andrea Francesco Scalella:** Software. **Vincenzo Manfredi:** Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We acknowledge financial support from the PNRR MUR project PE0000013-FAIR

Data availability

The authors do not have permission to share data.

References

[1] M.D. Herrmann, D.A. Clunie, A. Fedorov, S.W. Doyle, S. Pieper, V. Klepeis, L.P. Le, G.L. Mutter, D.S. Milstone, T.J. Schultz, et al., Implementing the DICOM standard for digital pathology, J. Pathol. Informatics 9 (1) (2018) 37.
[2] F. Liu, M. Hernandez-Cabronero, V. Sanchez, M.W. Marcellin, A. Bilgin, The current role of image compression standards in medical imaging, Information 8 (4) (2017) 131.
[3] S.S. Parikh, D. Ruiz, H. Kalva, G. Fernández-Escribano, V. Adzic, High bit-depth medical image compression with HEVC, IEEE J. Biomed. Health Informatics 22 (2) (2017) 552–560.
[4] S. Ab Aziz, S.M. Sam, N.H. Hassan, H. Abas, S.Z.A. Rasid, M.F. Yusof, N. Mohamed, A performance review for hybrid region of interest-based medical image compression, IEEE Access 11 (2023) 98025–98038.
[5] P. Kumar, A. Parmar, Versatile approaches for medical image compression: A review, Procedia Comput. Sci. 167 (2020) 1380–1389.
[6] X. Zhang, X. Wu, Ultra high fidelity deep image decompression with $l\infty$-constrained compression, IEEE Trans. Image Process. 30 (2020) 963–975.
[7] P. Bindu, J. Afthab, Region of interest based medical image compression using DCT and capsule autoencoder for telemedicine applications, in: 2021 Fourth International Conference on Electrical, Computer and Communication Technologies, ICECCT, IEEE, 2021, pp. 1–7.
[8] D. Mishra, S.K. Singh, R.K. Singh, Deep architectures for image compression: a critical review, Signal Process. 191 (2022) 108346.
[9] S. Jamil, M.J. Piran, M. Rahman, O.-J. Kwon, Learning-driven lossy image compression: A comprehensive survey, Eng. Appl. Artif. Intell. 123 (2023) 106361.
[10] Y. Hu, W. Yang, Z. Ma, J. Liu, Learning end-to-end lossy image compression: A benchmark, IEEE Trans. Pattern Anal. Mach. Intell. 44 (8) (2021) 4194–4211.
[11] N.E.H. Bourai, H.F. Merouani, A. Djebbar, Deep learning-assisted medical image compression challenges and opportunities: systematic review, Neural Comput. Appl. (2024) 1–42.
[12] M. Fischer, P. Neher, P. Schüffler, S. Ziegler, S. Xiao, R. Peretzke, D. Clunie, C. Ulrich, M. Baumgartner, A. Muckenhuber, et al., Unlocking the potential of digital pathology: Novel baselines for compression, J. Pathol. Informatics (2025) 100421.
[13] D. Tellez, G. Litjens, J. Van der Laak, F. Ciompi, Neural image compression for gigapixel histopathology image analysis, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2) (2019) 567–578.
[14] T.C. Nagavi, P. Mahesha, et al., Medical image lossy compression with LSTM networks, in: Histopathological Image Analysis in Medical Decision Making, IGI Global, 2019, pp. 47–68.
[15] M. Fischer, P. Neher, T. Wald, S.D. Almeida, S. Xiao, P. Schüffler, R. Braren, M. Götz, A. Muckenhuber, J. Kleesiek, et al., Learned image compression for he-stained histopathological images via stain deconvolution, in: International Workshop on Medical Optical Imaging and Virtual Microscopy Image Analysis, Springer, 2024, pp. 97–107.
[16] M.F. Barnsley, L.P. Hurd, Fractal Image Compression, AK Peters, Ltd., 1993.
[17] A.E. Jacquin, Fractal image coding: A review, Proc. IEEE 81 (10) (1993) 1451–1465.
[18] R. Distasi, M. Nappi, D. Riccio, A range/domain approximation error-based approach for fractal image compression, IEEE Trans. Image Process. 15 (1) (2005) 89–97.
[19] J. Wang, N. Zheng, A novel fractal image compression scheme with block classification and sorting based on Pearson's correlation coefficient, IEEE Trans. Image Process. 22 (9) (2013) 3690–3702.
[20] S. Bhavani, K.G. Thanushkodi, Comparison of fractal coding methods for medical image compression, IET Image Process. 7 (7) (2013) 686–693.
[21] A.K. Biswas, S. Karmakar, S. Sharma, Performance analysis of a new fractal compression method for medical images based on fixed partition, Int. J. Inf. Technol. 14 (1) (2022) 411–419.
[22] S. Liu, W. Bai, N. Zeng, S. Wang, A fast fractal based compression for MRI images, IEEE Access 7 (2019) 62412–62420.
[23] M. Kaur, V. Wasson, ROI based medical image compression for telemedicine application, Procedia Comput. Sci. 70 (2015) 579–585.
[24] N. Baranwal, K.N. Singh, A.K. Singh, et al., YOLO-based ROI selection for joint encryption and compression of medical images with reconstruction through super-resolution network, Future Gener. Comput. Syst. 150 (2024) 1–9.
[25] S. Kalra, H.R. Tizhoosh, C. Choi, S. Shah, P. Diamandis, C.J. Campbell, L. Pantanowitz, Yottixel–an image search engine for large archives of histopathology whole slide images, Med. Image Anal. 65 (2020) 101757.
[26] I. Lahr, S. Alfasly, P. Nejat, J. Khan, L. Kottom, V. Kumbhar, A. Alsaafin, A. Shafique, S. Hemati, G. Alabtah, et al., Analysis and validation of image search engines in histopathology, IEEE Rev. Biomed. Eng. (2024).
[27] M.F. Barnsley, S. Demko, Iterated function systems and the global construction of fractals, Proc. R. Soc. A 399 (1817) (1985) 243–275.
[28] S. Mitra, C. Murthy, M. Kundu, Partitioned iterative function system: A new tool for digital imaging, IETE J. Res. 46 (5) (2000) 279–298.
[29] J.L. Bentley, Multidimensional binary search trees used for associative searching, Commun. ACM 18 (9) (1975) 509–517.
[30] D. Bank, N. Koenigstein, R. Giryes, Autoencoders, in: Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook, Springer, 2023, pp. 353–374.
[31] M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, B.C. Van Esesn, A.A.S. Awwal, V.K. Asari, The history began from alexnet: A comprehensive survey on deep learning approaches, 2018, arXiv preprint arXiv:1803.01164.
[32] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.
[33] G. Strang, The discrete cosine transform, SIAM Rev. 41 (1) (1999) 135–147.
[34] M. Nappi, G. Polese, G. Tortora, First: Fractal indexing and retrieval system for image databases, Image Vis. Comput. 16 (14) (1998) 1019–1031.
[35] R. Distasi, M. Nappi, M. Tucci, FIRE: Fractal indexing with robust extensions for image databases, IEEE Trans. Image Process. 12 (3) (2003) 373–384.
[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[37] S. Chopra, R. Hadsell, Y. LeCun, Learning a similarity metric discriminatively, with application to face verification, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'05, Vol. 1, IEEE, 2005, pp. 539–546.
[38] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: A unified embedding for face recognition and clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823.
[39] G. Aresta, T. Araújo, S. Kwok, S.S. Chennamsetty, M. Safwan, V. Alex, B. Marami, M. Prastawa, M. Chan, M. Donovan, et al., Bach: Grand challenge on breast cancer histology images, Med. Image Anal. 56 (2019) 122–139.
[40] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
[41] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[42] G.K. Wallace, The JPEG still picture compression standard, Commun. ACM 34 (4) (1991) 30–44.
[43] D.S. Taubman, M.W. Marcellin, M. Rabbani, JPEG2000: Image compression fundamentals, standards and practice, J. Electron. Imaging 11 (2) (2002) 286–287.
[44] D. Zhang, G. Lu, Shape-based image retrieval using generic Fourier descriptor, Signal Process., Image Commun. 17 (10) (2002) 825–848.
[45] Sivic, Zisserman, Video Google: A text retrieval approach to object matching in videos, in: Proceedings Ninth IEEE International Conference on Computer Vision, IEEE, 2003, pp. 1470–1477.