Journal of Imaging Informatics in Medicine
https://doi.org/10.1007/s10278-024-01310-8

ORIGINAL PAPER

An Innovative Attention-based Triplet Deep Hashing Approach to Retrieve Histopathology Images

Seyed Mohammad Alizadeh[1] · **Mohammad Sadegh Helfroush**[1] · **M. Emre Celebi**[2]

Received: 12 July 2024 / Revised: 16 October 2024 / Accepted: 19 October 2024

Abstract

Content-based histopathology image retrieval (CBHIR) can assist in the diagnosis of different diseases. The retrieval procedure can be complex and time-consuming if high-dimensional features are required. Thus, hashing techniques are employed to address these issues by mapping the feature space into binary values of varying lengths. The performance of deep hashing approaches in image retrieval is often superior to that of traditional hashing methods. Among deep hashing approaches, triplet-based models are typically more effective than pairwise ones. Recent studies have demonstrated that incorporating the attention mechanism into a deep hashing approach can improve its effectiveness in retrieving images. This paper presents an innovative triplet deep hashing strategy based on the attention mechanism for retrieving histopathology images, called histopathology attention triplet deep hashing (HATDH). Three deep attention-based hashing models with identical architectures and weights are employed to produce binary values. The proposed attention module can aid the models in extracting features more efficiently. Moreover, we introduce an improved triplet loss function considering pair inputs separately in addition to triplet inputs for increasing efficiency during the training and retrieval steps. Based on experiments conducted on two public histopathology datasets, BreakHis and Kather, HATDH significantly outperforms state-of-the-art hashing algorithms.

Keywords Attention mechanism · Hashing algorithms · Content-based histopathology image retrieval · Neural networks · Triplet models

Introduction

Content-based histopathology image retrieval (CBHIR) is a computer vision approach employed in the diagnosis of various diseases [1]. Given a query image, a CBHIR system ranks the database images in order of decreasing similarity based on features extracted from images [2]. An effective ranking process may require high-dimensional features, making the retrieval procedure complex and time-consuming [3]. Therefore, hashing algorithms are used to resolve these issues by converting the feature space into binary values of varying lengths [1]. Deep hashing models often outperform traditional hashing approaches in image retrieval [4]. For deep hashing strategies, however, generating binary codes using the sign function may be challenging due to the vanishing gradient problem [5]. Deep hashing techniques generally employ a pairwise or triplet scheme. Pairwise-based approaches utilize similar and dissimilar pairs to train, while triplet-based methods employ a triplet structure of images, where two images are similar to each other, but distinct from the third [6]. Recent studies have demonstrated that triplet deep hashing models perform better than pairwise-based techniques in image retrieval tasks [7, 8]. Nevertheless, the performance of these models may need to be enhanced to provide high accuracy when retrieving medical images [7].

A number of studies have shown that convolutional neural networks (CNNs) are incredibly powerful for analyzing histopathology images [9]. Several pre-trained CNNs, including VGGNet [10], ResNet [11], and MobileNet [12], have been applied to classify and retrieve histopathological images [3, 9]. Research findings indicate that the effectiveness of CNNs does not significantly increase as they become more complex, leading to the introduction of the attention

✉ Seyed Mohammad Alizadeh
s.alizadeh@sutech.ac.ir

✉ Mohammad Sadegh Helfroush
ms_helfroush@sutech.ac.ir

[1] Department of Electrical Engineering, Shiraz University of Technology, Shiraz, Iran

[2] Department of Computer Science and Engineering, University of Central Arkansas, Conway, AR 72035, USA

Published online: 11 November 2024

mechanism [13]. As human eyes function, the attention mechanism assists CNNs in concentrating on crucial details and information relevant to the defined purpose. Through this mechanism, we can concentrate on key features and exclude irrelevant ones [14]. Hence, the effectiveness of an image retrieval process can be significantly improved by incorporating the attention mechanism into a deep hashing structure [15].

Although histopathology and other images are usually retrieved using a similar process, there are some nuanced differences. Histopathology datasets sometimes contain unequal distributions of samples within classes, which can adversely affect retrieval efficiency [16]. Histopathology images provide a microscopic view of tissues and differ in nature from other medical images, making them challenging to provide and utilize for deep hashing models [17]. In addition, histopathology images can be large and need to be converted into small patches, which complicates a CBHIR model and makes precise retrieval challenging [18]. Moreover, as a CBHIR system can be employed to detect a variety of diseases, it must be capable of retrieving histology images quickly and accurately [3].

Although several CBHIR systems have been developed in recent years, they are not necessarily accurate, especially for histopathology databases with many classes [3]. Our findings indicate that there is a lack of research on evaluating the performance of the attention mechanism incorporated into a deep hashing model for retrieving histopathology images. In addition, while several attention mechanism modules have been introduced over the past few years, designing a simple and practical module remains challenging [13]. Moreover, the implementation of a triplet deep hashing method, which is superior to the current hashing-based approaches, is crucial to the retrieval of histopathology images. We thus propose an effective method for retrieving histopathology images using a novel triplet deep hashing model based on the attention mechanism, called histopathology attention triplet deep hashing (HATDH). As part of the proposed model, we consider pair inputs separately in addition to triplet inputs for improved retrieval performance, especially in multi-class histopathology databases. The designed attention module is also helpful in focusing more on details in a histopathology image to extract features efficiently. Also, an effective hash layer is suggested to overcome the vanishing gradient issue while producing high-precision binary codes. As a result, HATDH performs better than current state-of-the-art hashing-based methods for retrieving histopathology images.

The major contributions of this work are as follows:

- To the best of our knowledge, an effective triplet deep hashing model is proposed to retrieve histopathology images for the first time.
- As part of our deep hashing approach, we design an enhanced attention module, named hybrid coordinate attention module (HCAM), outperforming its alternatives in feature extraction. Although the developed module can be integrated into various CNN architectures to improve both medical and non-medical image retrieval, this study utilizes it to enhance the effectiveness of a histopathology image retrieval process.
- An effective hash layer is suggested to produce binary values of varying lengths with high accuracy, resulting in speeding up the training and retrieval phases as well as addressing the vanishing gradient issue.
- A novel triplet loss function is introduced that takes into account pair inputs separately in addition to triplet inputs for improving the performance of both the training and retrieval phases.
- The presented loss function can also decrease the error between generated hash codes and real ones.
- Our approach enables a better feature extraction stage by concentrating on details in histopathology images thanks to the designed attention module.
- The presented triplet model also improves the performance of an image retrieval system for various histopathological datasets.
- Based on experiments conducted on two public histopathology datasets, HATDH is superior to current cutting-edge hashing-based approaches.

The paper continues as follows. "Related Work" Section provides an overview of recent studies on hashing algorithms, CBHIR systems, and attention mechanism modules. The suggested method is thoroughly explained in the "Methods" Section. "Experimental Results" Section outlines and discusses the results of our study. Finally, a summary of the paper and recommendations for future research are presented in the "Conclusion" Section.

Related Work

Hashing Techniques

Hashing techniques are used to produce binary values that facilitate image retrieval while requiring less storage space [1]. The two main types of hashing approaches are classical and deep learning-based methods [19]. The most popular classical hashing algorithms are locality-sensitive hashing (LSH) [20], iterative quantization (ITQ) [21], and supervised discrete hashing (SDH) [22]. LSH uses random hash functions to encode the feature domain in a data-independent manner. ITQ reduces the quantization gap between the produced binary values and feature space. SDH trains hash codes via a regression-based methodology.

Deep hashing approaches often outperform classical hashing techniques in image retrieval [5]. A deep hash model employs CNNs for feature extraction, followed by a hash layer to create binary values [19]. In the past few years, a variety of deep hashing models have been developed to retrieve images. Deep pairwise-supervised hashing (DPSH), presented by Li et al. [4], uses two similar CNNs to generate hash codes. An advanced loss function is implemented to learn two hash values based on their similarity, leading to promising results for image retrieval on benchmarking datasets. Deep supervised hashing (DSH) is another framework that trains binary codes using a form of contrastive loss function [23]. Deep triplet supervised hashing (DTSH), a triplet version of DPSH, intends to maximize triplet label likelihoods [24]. To deal with the vanishing gradient challenge, HashNet generates hash codes using a scaled tanh function [5]. However, the model may face difficulties when generating highly accurate binary codes. In order to achieve high accuracy in image retrieval tasks, deep triplet quantization (DTQ) utilizes a triplet architecture to learn condensed binary information [8]. Improved deep hashing network (IDHN), proposed by Zhang et al. [25], boosts the performance of image retrieval in multi-class databases. The model compares pair hash codes utilizing a quantitative similarity to yield acceptable results. Attention-based triplet hashing (ATH) leverages the attention mechanism within a triplet hashing scheme to retrieve medical images efficiently [7]. Opponent class adaptive margin (OCAM) suggests a more effective triplet hashing approach for retrieving medical images [26]. The model adaptively selects a margin value according to the dataset. Deng et al. [27] present a triplet-based deep hashing (TDH) model for cross-modal retrieval. They also propose a graph regularization to maintain the initial conceptual similarity between binary values.

Attention Mechanisms

Motivated by the human vision system, several works employed attention mechanisms to boost the effectiveness of CNNs in the feature extraction phase for image retrieval and classification [14, 15]. Using attention strategies, CNN models can be trained to concentrate on what (channel attention) and where (spatial attention) are relevant to the intended aim [28]. The channel and spatial attention mechanisms may be considered separately or in combination when designing CNN models.

In recent years, many attention modules have been developed for improving the performance of CNNs during the feature extraction stage. Woo et al. [14] presented a convolutional block attention module (CBAM) using a sequential combination of channel and spatial attention mechanisms. Experiments have proven that applying CBAM to different CNN models can enhance classification and retrieval accuracy. An efficient channel attention (ECA) module was proposed in [29], focused on improving the performance of CNNs in image analysis with less complexity. A new attention strategy, known as coordinate attention, was introduced to enhance the functionality of MobileNet models in [30]. The authors demonstrated that separately investigating information in two spatial axes can improve CNN performance in image classification and segmentation. Li et al. [13] developed a hybrid attention module (HAM) with a similar design to CBAM. The channel attention phase followed the ECA framework. On the other hand, the spatial attention was implemented similarly to CBAM but with the addition of a channel separation step to improve the overall efficiency.

Histopathology Image Retrieval

A variety of CBHIR approaches have been developed with or without hashing in recent years. Based on hand-crafted features, Ma et al. [2] introduced an unsupervised approach to retrieve histology images for breast cancer detection. Their system utilized the LSH algorithm for the optimization of the search procedure. Shi et al. [1] designed a pairwise-based deep hashing model for classifying and retrieving histopathology images. The model employed an innovative objective function to optimize the training procedure. In [3], another deep hashing approach was introduced for retrieving images of histopathology focused on breast cancer recognition. The authors applied VGG16 for the feature extraction step, which was then attached to a hash layer for generating binary values. Yang et al. [31] proposed a deep metric learning strategy to retrieve histopathological images. The model performance was enhanced by implementing an attention mechanism. A novel feature learning methodology was suggested for whole-slide histopathology image retrieval in [18]. An attention-based CBHIR system focusing on key regions in whole-slide images was developed by Hashimoto et al. [32], achieving promising results. Our previous work, called histopathology Siamese deep hashing (HSDH), presented a novel deep hashing method based on a Siamese structure for retrieving histopathology images [33]. The approach employed a new hash layer to overcome the vanishing gradient problem. Furthermore, we developed an improved loss function to enhance retrieval performance. Although HSDH could show promising results, the designed hash layer may increase the complexity. Furthermore, pairwise structures may not be as effective as triplet ones for separating samples, especially in multi-class datasets [7]. As a result, it is necessary to develop a triplet deep hashing model for retrieval of histopathology images. In this paper, we develop a novel triplet structure that takes into account pair inputs separately in addition to triplet inputs. Additionally, the presented loss function allows for high-accuracy and efficient binary code generation and training. The proposed

attention mechanism module can also improve the feature extraction process. These contributions aim to increase the retrieval accuracy of histopathology images compared with the existing approaches.

Since a comparison between HSDH and HATDH can assist future researchers in selecting the most effective approach, presenting both models is necessary and important. In comparison with HSDH, HATDH offers several innovations and improvements, as follows:

- Developing a triplet deep hashing scheme to retrieve histology images from various datasets, which is more effective than a pairwise scheme.
- Presenting an efficient hashing layer with reduced complexity to deal with the vanishing gradient problem.
- Improved accuracy in generating and learning binary codes.
- HSDH involves training the distance between binary codes, while HATDH directly accesses binary codes, allowing for more flexibility in retrieval.
- Designing an effective and novel attention module to improve the feature extraction process.
- The separation technique used in designing our module helps to focus better on specific regions and select the best features.
- Proposing an improved triplet loss function considering pair inputs separately in addition to triplet inputs for increasing efficiency during the training and retrieval phases.
- The possibility of decreasing the error between generated hash codes and real ones, which may not be possible in HSDH easily.
- Enhancing the retrieval results on the datasets employed compared with HSDH.

As can be seen, HATDH can offer 9 novelties and improvements over HSDH. The mentioned novelties are so specific and different in detail. It should be noted that some of the innovations mentioned, such as the design of the attention module and hash layer, could each be the subject of an individual paper [5, 13]. However, in this work, we present them together in the context of improving the process of retrieving histopathology images. While HSDH and HATDH appear similar at first glance, they differ substantially in many details, as discussed earlier. HATDH aims to address the challenges and issues associated with HSDH while also improving retrieval performance. Introducing different histopathology image retrieval models can assist researchers in selecting the most suitable approach by understanding the pros and cons of each method for their future research. In light of the numerous similar works to HATDH published in various applications, presenting our method may prove useful in future studies related to histopathology analysis. Furthermore, it should be mentioned that various innovations and advancements incorporated in the design of HATDH are being introduced for the first time within this particular field.

Methods

As illustrated in Fig. 1, HATDH includes three identical deep hashing models with three inputs, namely anchor, positive, and negative. Anchor and positive have the same class label, while negative has a different one. These triplet images are randomly selected using the procedure outlined in [7]. Our first objective is to encode each image x_l employing a simple L-bit binary value $\langle_l \in \{-1, 1\}^L$. Our deep hashing models are composed of a feature extraction stage and a hash layer. A novel attention module is included in the architecture of CNNs to improve the performance of the feature extraction process. The final goal is to learn hash codes via the proposed loss function so that the anchor and positive codes are less distant than the anchor and negative codes. This section describes the structure and development of deep hashing models, the learning process, and how the retrieval process is carried out.

Deep Hashing Model Structure

MobileNet is chosen as the base model because it is easy to implement and effective in analyzing histopathology images [9]. Moreover, several studies have shown that integrating attention modules into the architecture of this network assists in improving an image analysis process [14, 30]. MobileNet also appears to be less sensitive to the number of training samples than other CNN models, such as the VGGNet, ResNet, leading to less complexity [12]. An attention module is developed to enhance the feature extraction procedure. Then, a hash layer is designed for converting the feature space to binary codes. We first explain the proposed attention module, followed by a detailed description of the overall deep hashing structure.

Proposed Attention Module

According to CBAM, HCAM incorporates a sequential arrangement of channel and spatial attention mechanisms. Figure 2 (a) depicts that features are first passed through the channel attention part, and the results are then multiplied by the input to obtain optimized features in the channel attention stage. Following that, the optimized features are divided into two axes for generating horizontal and vertical spatial attention. The following describes the channel and spatial attention components in details.

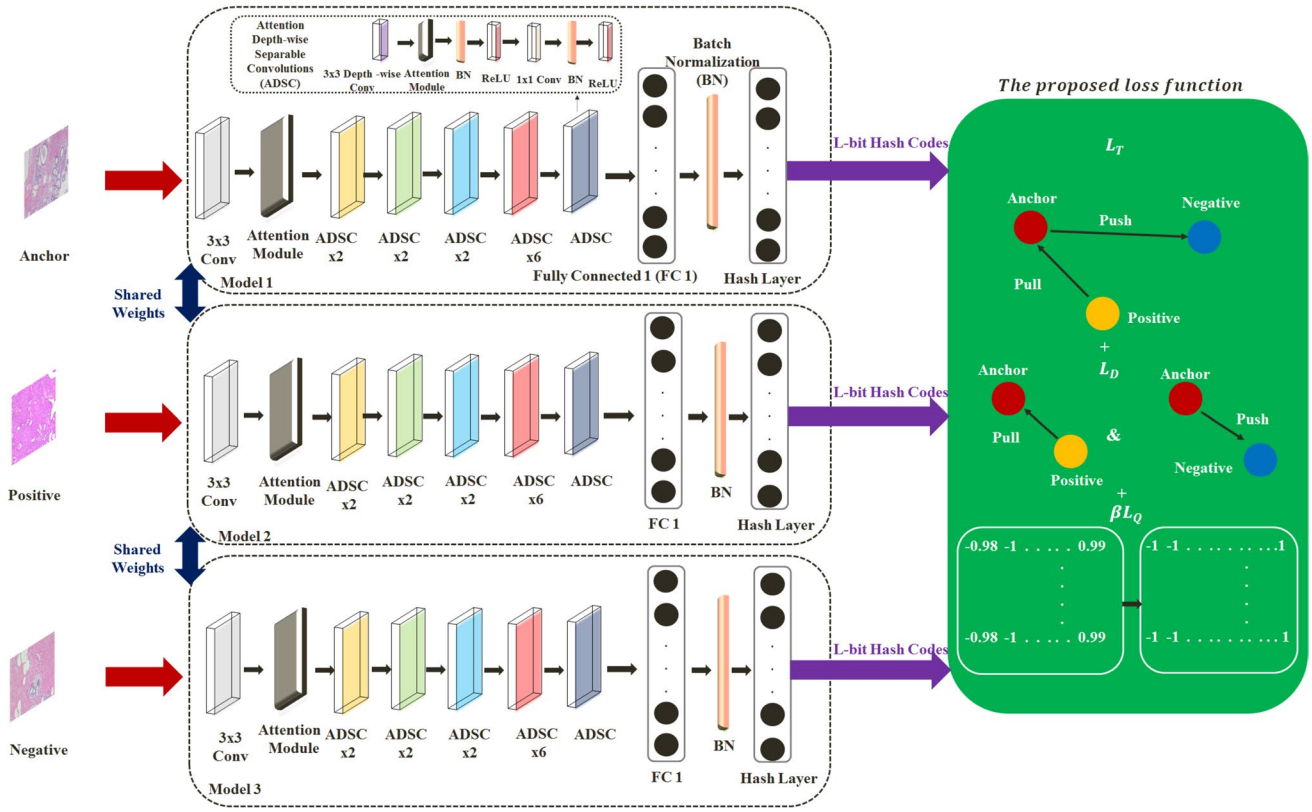


Fig. 1 A general overview of HATDH. HATDH includes three identical deep hashing models with the same weights. The designed attention module can enhance the performance of the models when extracting features. The proposed triplet loss function takes into account pair inputs separately (anchor with positive as well as negative) in addition to triplet inputs to improve the training and retrieval stages

Channel Attention Component As can be seen in Fig. 2 (b), the max-pooling and average-pooling functions are initially employed to gather spatial information on features. The functions produce two feature categories, $\mathcal{F}_C^A$ and $\mathcal{F}_C^M$, representing average-pooled and max-pooled features, respectively. In the next step, both categories are summed adaptively based on two trainable variables ω_1 and ω_2. Inspired by [13], the values of ω_1 and ω_2 are chosen between 0 and 1. The result is then fed into a convolutional network with two 1-D layers. Hang et al. [34] suggested utilizing two convolutional layers with different kernels to improve attention module performance. Therefore, we develop two 1-D convolutional layers with kernels k_1 and k_2. Inspired by ECA, k_1 is adaptively calculated based on Eq. (1).

$$k_1 = \left| \frac{\log_2(number\ of\ channels)}{\gamma} + \frac{\lambda}{\gamma} \right|_{odd} \tag{1}$$

where $|r|_{odd}$ represents the closest odd number to r. The values of γ and λ are 2 and 1, respectively. Moreover, k_2 is set to 7, providing the best experimental results in combination with the first 1-D convolutional layer. In general, the channel attention function can be described by Eq. (2):

$$\mathcal{F}_C = \sigma\left(C1D_{1\times k_2}\left(C1D_{1\times k_1}\left(\omega_1 \bigotimes \mathcal{F}_C^A \bigoplus \omega_2 \bigotimes \mathcal{F}_C^M\right)\right)\right)\ s.t. \omega_1 + \omega_2 = 1 \tag{2}$$

where σ, $C1D$, $\bigotimes$, and $\bigoplus$ indicate the sigmoid function, 1-D convolutional layer with various kernels, element-wise multiplication, and element-wise addition.

Spatial Attention Component According to Fig. 2 (c), in the first step, the max-pooling function is applied along the channel direction to each set, yielding $\mathcal{F}_S^M \in \mathbb{R}^{1\times H\times W}$ and $\mathcal{F}_S^M \in \mathbb{R}^{1\times H\times W}$. The average-pooling operation is then performed separately in the horizontal and vertical directions, leading to $\mathcal{F}_S^C(h) \in \mathbb{R}^{1\times H\times 1}$ and $\mathcal{F}_S^C(w) \in \mathbb{R}^{1\times 1\times W}$, respectively. The results can be expressed as follows:

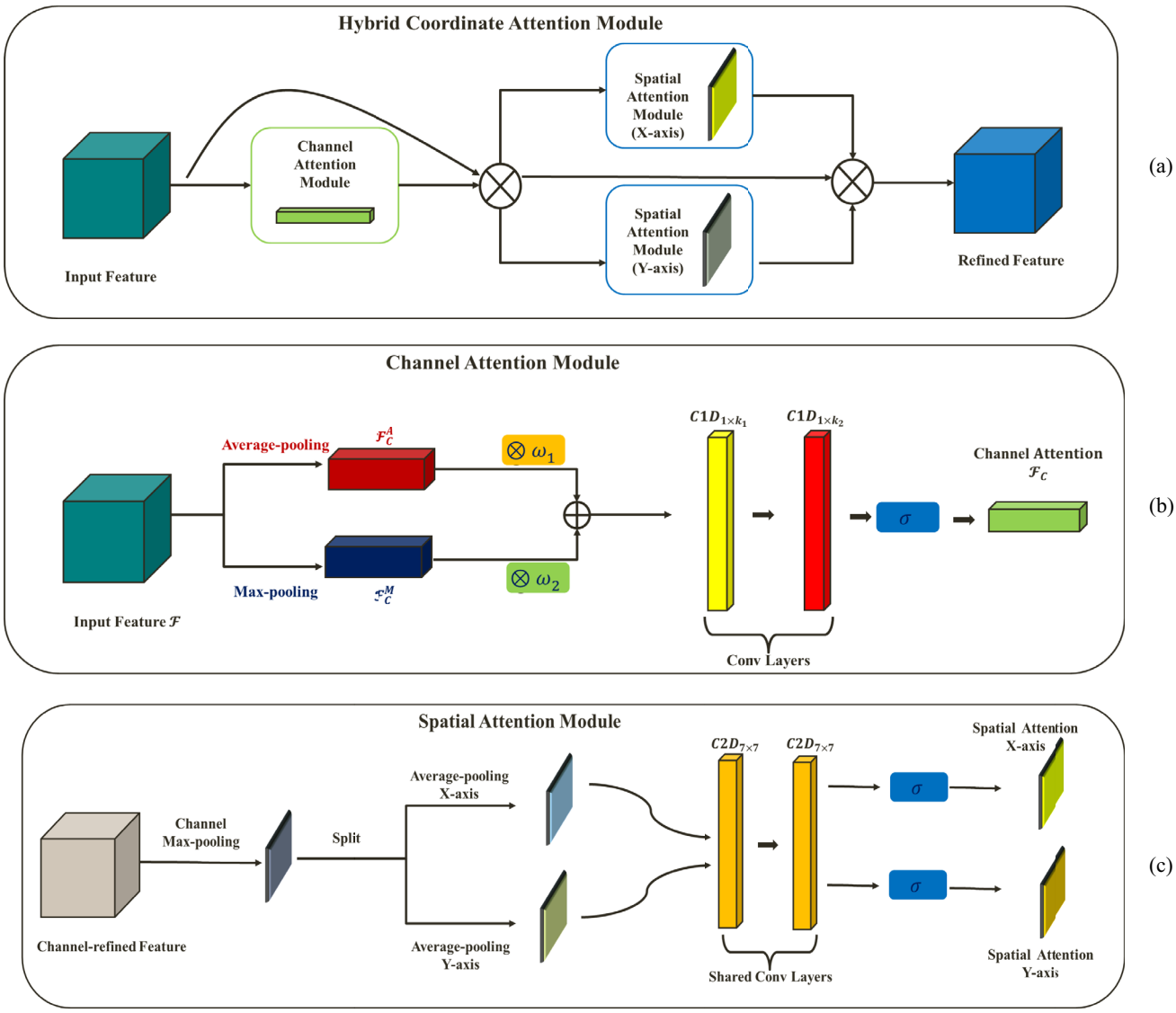


Fig. 2 The overall view of HCAM: **a** the general structure, **b** the channel attention module, and **c** the spatial attention module. Our module consists of two sequential parts: channel and spatial (in two axes). In the channel component, the max-pooling and average-pooling outputs are combined based on two different weights and entered into a convolutional network. The results are fed into the sigmoid function to produce the channel attention. As for the spatial component, the max-pooling operation is applied first in the channel direction, followed by the average-pooling operation in both horizontal and vertical directions. After applying the results into a convolutional network, the sigmoid function is used to generate the spatial attention in the two axes

$$\mathcal{F}_S^C(h) = \frac{1}{W} \sum_{0 \le j < W} \mathcal{F}_S^M(h, j) \tag{3}$$

$$\mathcal{F}_S^C(w) = \frac{1}{H} \sum_{0 \le i < H} \mathcal{F}_S^M(i, w) \tag{4}$$

The recent step collects spatial correlations along one axis while maintaining positional details along the other, which can be beneficial for improving the feature extraction process [30]. Finally, two 2-D convolutional layers with the same kernel size of 7×7 ($C2D_{7\times7}$) are employed to generate the spatial attention representation in two horizontal ($\mathcal{F}_S(h) \in \mathbb{R}^{1\times H\times1}$) and vertical ($\mathcal{F}_S(w) \in \mathbb{R}^{1\times1\times W}$) directions. The proposed spatial attention for a given set can be calculated as follows:

$$\mathcal{F}_S(h) = \sigma(C2D_{7\times7}(C2D_{7\times7}(\mathcal{F}_S^C(h)))) \tag{5}$$

$$\mathcal{F}_S(w) = \sigma(C2D_{7\times7}(C2D_{7\times7}(\mathcal{F}_S^C(w)))) \quad (6)$$

The final enhanced features ($\mathcal{F}_E \in \mathbb{R}^{C\times H\times W}$) can be described according to Eq. (7):

$$\mathcal{F}_E = \mathcal{F}_S(h) \bigotimes \mathcal{F}_S(w) \bigotimes \mathcal{F}_C \quad (7)$$

Deep Hashing Architecture

As mentioned, MobileNet is chosen as the base model for feature extraction. Our attention module is then integrated into the model to optimize performance during the feature extraction phase (Fig. 1). Lastly, the softmax layer is supplanted by a fully connected layer containing L nodes with the tanh activation function, where L indicates the hash code length. We employ the tanh activation function to approximate the sign function when generating binary values to prevent the vanishing gradient problem in back-propagation learning. As a result, an L-bit hash-like value can be created for an image as follows:

$$h = \tanh(WF + v) \quad (8)$$

where $W \in \mathbb{R}^{L\times512}$, $F \in \mathbb{R}^{512\times1}$, and $V \in \mathbb{R}^{L\times1}$ represent a weight matrix, the feature extraction stage outcome, and a bias vector, respectively.

Model Learning

Consider h_A,h_P, and h_N as the hash codes generated for anchor, positive, and negative images, respectively. The aim of a typical triplet deep hashing model is to force the distance between h_A and h_N to be greater than that of h_A and h_P by the margin λ_1. The loss function can therefore be described for a triplet structure i in this way:

$$L_T = \sum_i \left(max\left(\left\| \langle_{A_i} - \langle_{P_i} \right\|_2^2 - \left\| \langle_{A_i} - \langle_{N_i} \right\|_2^2 + \lambda_1, 0 \right) \right) \quad (9)$$

where $\| \ . \ \|_2$ indicates the L2-norm distance.

According to [35] and [36], the anchor distances between positive and negative can also individually affect the performance of triplet-based models. Thus, we present the following equation to limit these distances by λ_2 and λ_3, respectively:

$$L_D = L_{D_1} + L_{D_2} = \sum_i (\max(\| h_{A_i} - h_{P_i} \|_2^2 - \lambda_2, 0) + \max(\lambda_3 - \| h_{A_i} - h_{N_i} \|_2^2, 0)) \quad (10)$$

As our deep hash model may not produce strictly binary values, we introduce a quantization loss term inspired by [4] as follows:

$$L_Q = \sum_i (\| sign(h_{A_i}) - h_{A_i} \|_2^2 + \| sign(h_{P_i}) - h_{P_i} \|_2^2 + \| sign(h_{N_i}) - h_{N_i} \|_2^2) \quad (11)$$

where $sign(x) = 1$ if $x > 0$ and -1 otherwise. Consequently, the final loss function is as follows:

$$L_G = L_T + L_D + \beta L_Q \quad (12)$$

where β is a regulator variable to manage L_Q.

The goal of a learning procedure is the minimization of a loss function using the back-propagation method. Thus, we have to find the derivative of L_G to W, F, and V, i.e., the parameters of our hash layer. Initially, the gradient of L_G is computed relative to h_{A_i} as follows:

$$\frac{\partial L_G}{\partial h_{A_i}} = \frac{\partial L_T}{\partial h_{A_i}} + \frac{\partial L_D}{\partial h_{A_i}} + \beta\frac{\partial L_Q}{\partial h_{A_i}} = \frac{\partial L_T}{\partial h_{A_i}} + \frac{\partial L_{D_1}}{\partial h_{A_i}} + \frac{\partial L_{D_2}}{\partial h_{A_i}} + \beta\frac{\partial L_Q}{\partial h_{A_i}} \quad (13)$$

where:

$$\frac{\partial L_T}{\partial h_{A_i}} = \begin{cases} \sum_i 2\times\left(h_{A_i} - h_{P_i}\right) - 2\times\left(h_{A_i} - h_{N_i}\right), \| h_{A_i} - h_{P_i} \|_2^2 > \| h_{A_i} - h_{N_i} \|_2^2 - \lambda_1 \\ 0, otherwise \end{cases} \quad (14)$$

$$\frac{\partial L_{D_1}}{\partial h_{A_i}} = \begin{cases} \sum_i 2\times\left(h_{A_i} - h_{P_i}\right), \| h_{A_i} - h_{P_i} \|_2^2 < \lambda_2 \\ 0, otherwise \end{cases} \quad (15)$$

$$\frac{\partial L_{D_2}}{\partial h_{A_i}} = \begin{cases} \sum_i -2\times\left(h_{A_i} - h_{N_i}\right), \| h_{A_i} - h_{N_i} \|_2^2 < \lambda_3 \\ 0, otherwise \end{cases} \quad (16)$$

$$\frac{\partial L_Q}{\partial h_{A_i}} = \sum_{i,j} -2\left(sign(h_{A_i}) - h_{A_i}\right) \quad (17)$$

The chain rule can then be applied to compute the gradients of L_G relative to the target parameters:

$$\frac{\partial L_G}{\partial W} = \frac{\partial L_G}{\partial h_{A_i}}\frac{\partial h_{A_i}}{\partial \mathcal{W}} = (\frac{\partial L_G}{\partial h_{A_i}})(F)(sech^2(WF + V)) \quad (18)$$

$$\frac{\partial L_G}{\partial F} = \frac{\partial L_G}{\partial h_{A_i}} \frac{\partial h_{A_i}}{\partial F} = (\frac{\partial L_G}{\partial h_{A_i}})(W)(sech^2(WF + V)) \quad (19)$$

$$\frac{\partial L_G}{\partial V} = \frac{\partial L_G}{\partial h_{A_i}} \frac{\partial h_{A_i}}{\partial V} = \left(\frac{\partial L_G}{\partial h_{A_i}}\right)\left(sech^2(WF + V)\right) \quad (20)$$

Additionally, the above stages can be performed for the hash codes generated for positive and negative samples. All parameters can then be optimized using the standard back-propagation strategy. Algorithm 1 provides a general description of the overall HATDH learning procedure.

Input:

Initial histopathology samples $\boldsymbol{X} = \{\boldsymbol{x_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{M}}$

Output:

The parameter values of $\boldsymbol{\mathcal{W}}$, $\boldsymbol{\mathcal{F}}$, and $\boldsymbol{\mathcal{V}}$

Initialization:

Initialize $\boldsymbol{\mathcal{W}}$ and $\boldsymbol{\mathcal{V}}$

Repeat

1. Construct triplet structures;
2. Obtain $\boldsymbol{\mathcal{F}}$ using our attention-based model;
3. Produce hash-like values for images of a triplet structure via (8);
4. Compute the derivatives as per (13) - (20);
5. Update the parameters $\boldsymbol{\mathcal{W}}$, $\boldsymbol{\mathcal{F}}$, and $\boldsymbol{\mathcal{V}}$ using the back-propagation learning approach;

Up to a specified number of times

Algorithm 1: The HATDH training procedure.

Searching Stage

Using the forward propagation technique, a query sample is compared with the entire training set to identify the most similar images in the Hamming space. The sign function can be utilized directly at this stage to calculate a binary value for an image x_t:

$$\langle_{x_t} = sign\left(\mathcal{W}\mathcal{F}_{x_t} + \mathcal{V}\right) \quad (21)$$

Experimental Results

Datasets

We compare our model with currently available hashing techniques using two public histopathology datasets, Kather [37] and BreakHis [16]. Both datasets employ small patches and slides, which may not effectively represent the multi-scale aspect of digital pathology. Nonetheless, they have shown the potential for content-based histopathological image retrieval in multiple previous experiments [3, 33].

Kather: This database includes 5000 RGB histology images with a pixel size of 150×150 for colorectal cancer detection, categorized into 8 classes containing 625 different regions of tissue, comprising simple stroma (including homogeneous content, smooth muscle, tumor stroma, and extra-tumoral stroma), complex stroma (incorporating single tumor cells and few immune cells), tumor epithelium, immune cells (having submucosal lymphoid follicles and immune-cell conglomerates), normal mucosal, background (no tissue), adipose tissue, and debris (comprising hemorrhage, mucus, and necrosis).

BreakHis: The dataset comprises 7909 images with a size of 700×460 pixels, including two classes of benign and malignant breast tumors. The images have been collected from biopsy slides stained with hematoxylin and eosin. This dataset has four magnification factors: 40×(652 benign and 1370 malignant samples), 100×(644 benign and 1437 malignant samples), 200×(623 benign and 1390 malignant samples), and 400×(588 benign and 1232 malignant samples). Since the purpose of this research is not to diagnose breast cancer at different magnification levels, we utilize the 40×version, as the typical magnification factor in the analysis of histology images.

According to [16, 33, 38], for the two datasets, 70% and 30% of each class are randomly chosen for training and testing, respectively. We report the average results of five experiments.

Based on [39, 40], all images are normalized before entering our model as part of preprocessing. Furthermore, Kather

images do not change in size, but BreakHis samples are resized to 224 × 224 to improve the training stage according to previous research [33, 41].

Implementation Details

The experiments are carried out on a computer with the 2.20 GHz Intel(R) Xeon(R) central processing unit (CPU), 32 GB of RAM, and Tesla T4 graphics processing unit (GPU) via the Python deep learning library Keras based on the TensorFlow framework. We apply the adaptive moment estimation (Adam) optimizer [42] with a learning rate of 0.001 and batch size of 64.

We compare our model with classical, such as LSH and ITQ, and deep hashing approaches, including DPSH, DSH, DTSH, HashNet, DTQ, IDHN, ATH, OCAM, and HSDH. DTSH, DTQ, ATH, OCAM, and are triple-based deep hashing methods, while the rest employ a pairwise structure. With the aim of providing a fair comparison, this paper follows the same strategy as previous similar studies [25]. Thus, both classical and deep hashing approaches employ our CNN attention-based model to extract features. Moreover, we implement state-of-the-art approaches based on public source codes. It was necessary, however, to modify some parameters of the models to achieve better results on the datasets utilized.

Table 1 Results of MAP on Kather for various lengths of binary values

Methods	*32-bits*	*64-bits*	*128-bits*
HATDH (ours)	0.9718 ± 0.001	0.9846 ± 0.012	0.9890 ± 0.008
HSDH	0.9523 ± 0.012	0.9679 ± 0.006	0.9795 ± 0.009
OCAM	0.8957 ± 0.028	0.9104 ± 0.008	0.9174 ± 0.007
ATH	0.8928 ± 0.024	0.9016 ± 0.014	0.9126 ± 0.013
IDHN	0.8609 ± 0.023	0.8861 ± 0.019	0.9002 ± 0.029
DTQ	0.8750 ± 0.016	0.8835 ± 0.007	0.8917 ± 0.012
HashNet	0.8802 ± 0.030	0.8859 ± 0.026	0.8894 ± 0.025
DTSH	0.8861 ± 0.029	0.8870 ± 0.029	0.8897 ± 0.030
DSH	0.8503 ± 0.031	0.8537 ± 0.027	0.8677 ± 0.021
DPSH	0.8685 ± 0.038	0.8749 ± 0.035	0.8839 ± 0.034
ITQ	0.7231 ± 0.075	0.7418 ± 0.055	0.7592 ± 0.043
LSH	0.7193 ± 0.079	0.7414 ± 0.056	0.7505 ± 0.048

Metrics

The following evaluation metrics are employed in this research based on similar works [25].

Mean average precision (*MAP*): measures the mean of the average precision (*AP*) for a query, where *AP* is calculated in the following way:

$$AP = \frac{1}{N_p}\sum_{l}^{N_t}\frac{N_l}{l} \times S_l \tag{22}$$

in this case, N_t indicates the size of the dataset; N_p indicates the number of pertinent samples; and N_l relates to how many pertinent samples appear in the top l results. In addition, S_l is 1 if the retrieved image matches the query, otherwise it is 0.

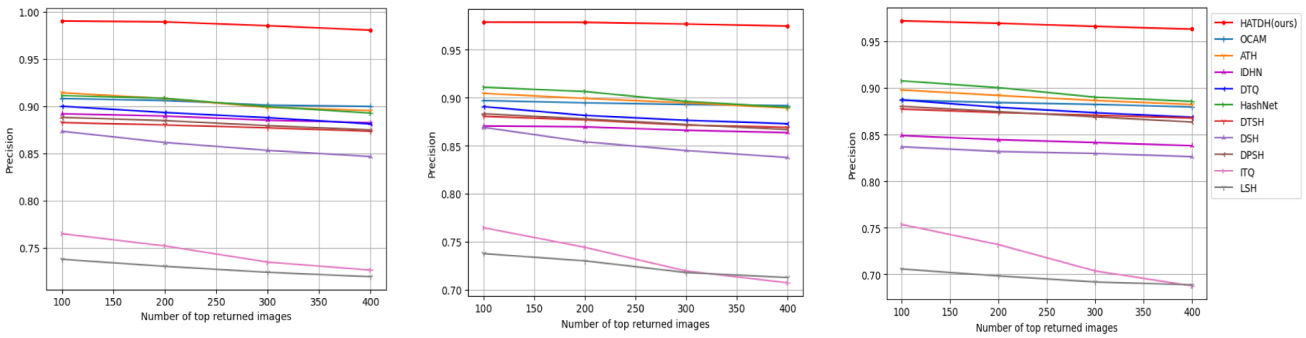
Precision@n: The precision of retrieving the first n samples matching the query image is expressed as the following:

$$Precision@n = \frac{1}{n}\sum_{l=1}^{n} S_l \tag{23}$$

Results

Kather Retrieval Results

According to Table 1, the HATDH method is superior to other existing hash-based approaches in terms of MAP



(a) (b) (c)

Fig. 3 A comparison of the precision curves for different numbers of images retrieved on Kather based on various lengths of binary codes (**a** 128, **b** 64, and **c** 32)

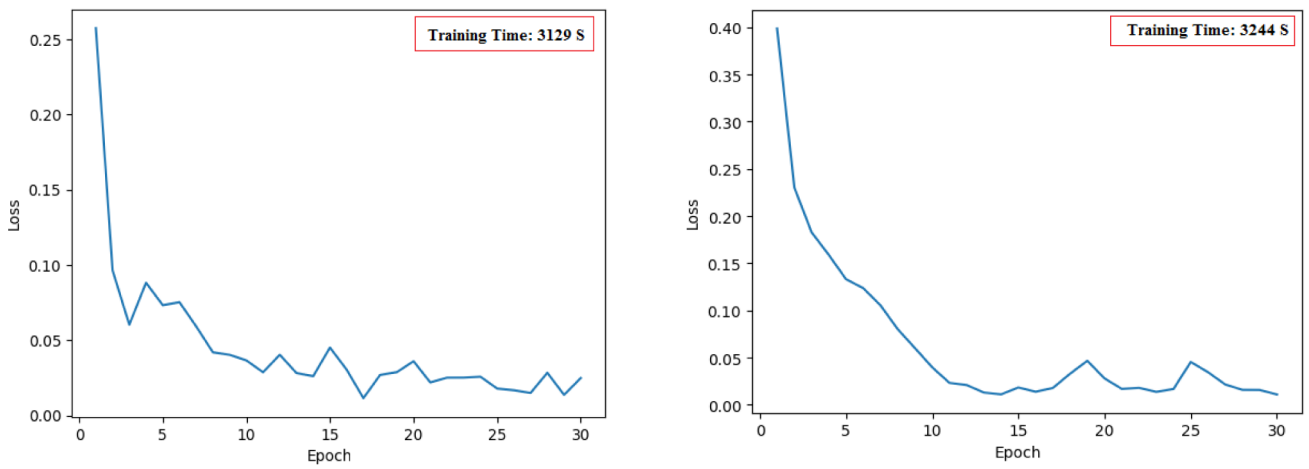
for various binary code lengths. As compared with IDHN, which performs best among pairwise-based methods, our model provides an approximately 8% enhancement in MAP. HATDH can also raise MAP by roughly 7% in comparison with OCAM, the most efficient triplet-based algorithm. Compared to HSDH, our previous work, HATDH can increase MAP by approximately 2%. To further evaluate the efficacy of the mentioned methods, we plot the precision curves for a variety of numbers of retrieved images based on multiple binary code lengths. As illustrated in Fig. 3, HATDH achieves outstanding accuracy in retrieving histopathology images for the studied binary value lengths. Our designed attention module improves the feature extraction phase, resulting in increased retrieval accuracy. Besides three input binary codes, our developed loss function provides better training and retrieval results by considering pairs of input binary codes separately. Moreover, the quantization loss term can contribute to more accurate binary values, enhancing the final result. Figure 4 shows a sample of the convergence curves and training times for the training process of HATDH on the Kather and BreakHis datasets for 30 epochs. Nevertheless, some experiments may require more epochs. It should be noted that the mentioned training times do not include loading images, installing Python packages, etc.

Table 2 Results of MAP on BreakHis for various lengths of binary values

Methods	*32-bits*	*64-bits*	*128-bits*
HATDH (ours)	0.9853 ± 0.018	0.9967 ± 0.002	0.9975 ± 0.002
HSDH	0.9706 ± 0.025	0.9782 ± 0.016	0.9845 ± 0.008
OCAM	0.9259 ± 0.024	0.9357 ± 0.023	0.9412 ± 0.018
ATH	0.9209 ± 0.015	0.9259 ± 0.015	0.9301 ± 0.017
IDHN	0.9374 ± 0.047	0.9418 ± 0.045	0.9457 ± 0.045
DTQ	0.8951 ± 0.087	0.9047 ± 0.088	0.9230 ± 0.071
HashNet	0.9081 ± 0.066	0.9137 ± 0.061	0.9275 ± 0.051
DTSH	0.9198 ± 0.061	0.9249 ± 0.057	0.9366 ± 0.047
DSH	0.8729 ± 0.034	0.8976 ± 0.018	0.9131 ± 0.020
DPSH	0.8782 ± 0.031	0.8964 ± 0.017	0.9105 ± 0.015
ITQ	0.7303 ± 0.080	0.7491 ± 0.059	0.7664 ± 0.045
LSH	0.7265 ± 0.084	0.7487 ± 0.060	0.7577 ± 0.051

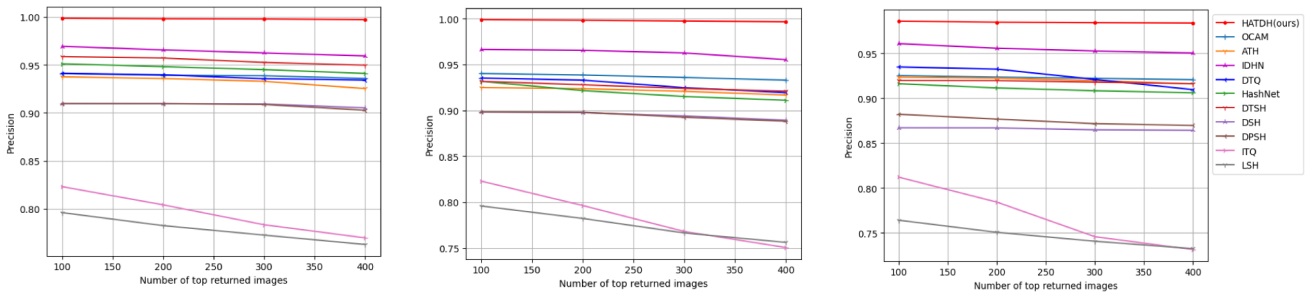
BreakHis Retrieval Results

Table 2 compares HATDH with several state-of-the-art hashing techniques on the BreakHis dataset. According to the findings, our approach is more effective than other hashing-based retrieval strategies for histopathological images. Based on a particular comparison, HATDH boosts MAP by roughly 5% for various binary value lengths relative to IDHN. In addition, our method can increase MAP by nearly 6% when compared with OCAM. Compared to HSDH, our previous work, HATDH can increase MAP by approximately 2%. Figure 5 illustrates that HATDH provides excellent accuracy for examined binary value lengths on this dataset. The BreakHis dataset is not balanced, affecting the efficiency of a retrieval model, but HATDH can deal with this issue and achieve outstanding results. Figure 4 (b) illustrates an example of the convergence curve and training time for the training procedure of our model on BreakHis for 30 epochs.



(a) (b)

Fig. 4 The convergence curves and training times for the training process of HATDH on the datasets used (**a** Kather; **b** BreakHis)



(a) (b) (c)

Fig. 5 A comparison of the precision curves for different numbers of images retrieved on BreakHis based on various lengths of binary codes (**a** 128, **b** 64, and **c** 32)

Table 3 A comparison of MAP results for HATDH and related types based on different hash code lengths from the Kather dataset

Methods	*32-bits*	*64-bits*	*128-bits*
HATDH-V1	0.9298	0.9397	0.9443
HATDH-V2	0.9391	0.9407	0.9448
HATDH-V3	0.9416	0.9479	0.9490
HATDH-V4	0.9385	0.9432	0.9534
HATDH-V5	0.9281	0.9324	0.9330
HATDH-V6	0.8958	0.9219	0.9224
HATDH-V7	0.9177	0.9333	0.9375
HATDH-V8	0.9345	0.9489	0.9503
HATDH-V9	0.8552	0.8807	0.9052
HATDH-V10	0.8573	0.9016	0.9227
HATDH-V11	0.9374	0.9443	0.9472
HATDH	0.9756	0.9823	0.9842

Table 4 A comparison of MAP results for HATDH and related types based on different hash code lengths from the BreakHis dataset

Methods	*32-bits*	*64-bits*	*128-bits*
HATDH-V1	0.9173	0.9453	0.9508
HATDH-V2	0.9231	0.9461	0.9521
HATDH-V3	0.9048	0.9303	0.9511
HATDH-V4	0.9182	0.9274	0.9431
HATDH-V5	0.9013	0.9106	0.9301
HATDH-V6	0.8917	0.9001	0.9094
HATDH-V7	0.9178	0.9285	0.9391
HATDH-V8	0.9160	0.9309	0.9416
HATDH-V9	0.8972	0.9260	0.9394
HATDH-V10	0.9149	0.9422	0.9465
HATDH-V11	0.9212	0.9435	0.9543
HATDH	0.9596	0.9938	0.9963

Analyzing Hyperparameters

The subsection examines the impact of various loss function hyperparameter values on MAP results. For simplicity, we present only the results of one study on Kather and BreakHis for a 128-bit binary code. The hyperparameters for our loss function are λ_1, λ_2, λ_3, and β. To begin with, we evaluate how λ_1 affects the results. A step-by-step examination is then carried out on the remaining hyperparameters. When the other hyperparameters are ignored, MAP reaches its highest point around $\lambda_1 = 0.3$ and 0.4 for Kather and BreakHis, respectively (Fig. 6(a) and (d)). As shown in Fig. 6 (b) and (e), choosing suitable λ_2 and λ_3 contributes to improving the performance of our model. Moreover, adding the quantization term with an appropriate coefficient can further boost the MAP value (Fig. 6(c) and (f)). For the Kather dataset, the best MAP is achieved when λ_1, λ_2, λ_3, and β are around 0.3, 0.2, 0.2, and 0.01, respectively. On the other hand, For the BreakHis dataset, the best MAP is achieved when λ_1, λ_2, λ_3, and approximately β are 0.4, 0.3, 0.4, and 0.1, respectively.

Ablation Study

This subsection examines eleven subtypes of HATDH. Tables 3 and 4 present MAP results for HATDH and the variants based on several binary code lengths for the two datasets used. These tables present the findings of only one study across both datasets to simplify analysis. HATDH-V1 to HATDH-V4 utilize ECA, CBAM, HAM, and coordinate attention modules in place of our attention module. As indicated in Tables 3 and 4, applying attention modules to a deep hashing structure can improve retrieval accuracy, but the complexity may increase slightly. Additionally, modules employing both channel and spatial attention mechanisms, such as CBAM and HAM, yield better results. HATDH-V5, HATDH-V6, and HATDH-V7 extract features using various CNN models, including MobileNet, VGG19, and ResNet101, respectively. According to the results, although the CNN models alone perform well for feature extraction, they are not as effective as when the attention modules are incorporated. Rather than our designed loss function, HATDH-V8, HATDH-V9, and HATDH-V10 employ those proposed in

Table 5 A detailed description of HATDH and its subtypes

Methods	*CNN model*	*Attention module*	*Loss function*	*The best MAP*
HATDH-V1	MobileNet	ECA	Ours	0.9508
HATDH-V2	MobileNet	CBAM	Ours	0.9521
HATDH-V3	MobileNet	HAM	Ours	0.9511
HATDH-V4	MobileNet	Coordinate	Ours	0.9534
HATDH-V5	MobileNet	-	Ours	0.9330
HATDH-V6	VGG19	-	Ours	0.9224
HATDH-V7	ResNet101	-	Ours	0.9391
HATDH-V8	MobileNet	Ours	proposed in [7]	0.9503
HATDH-V9	MobileNet	Ours	proposed in [8]	0.9394
HATDH-V10	MobileNet	Ours	proposed in [26]	0.9465
HATDH-V11	MobileNet	Ours (without the separation technique)	Ours	0.9543
HATDH	MobileNet	Ours (with the separation technique)	Ours	0.9963

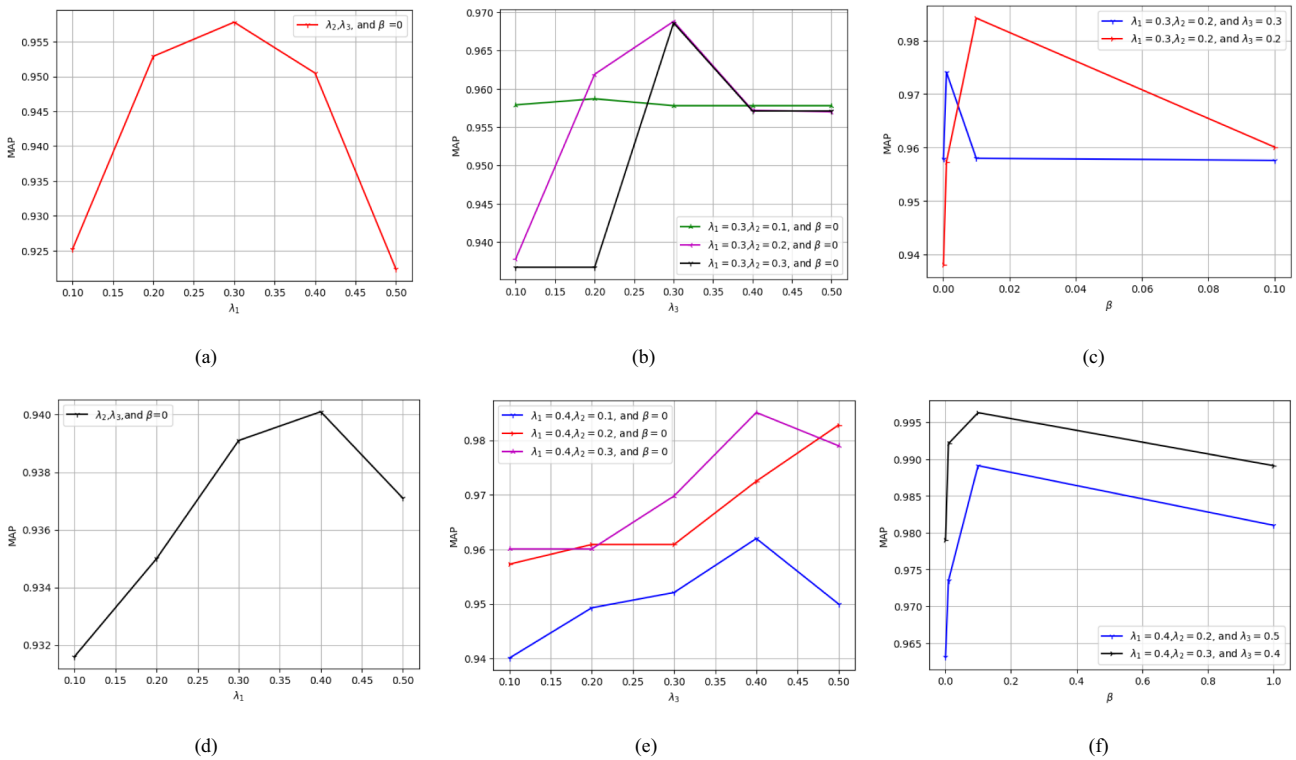


Fig. 6 Investigating how different hyperparameters of the proposed loss function affect retrieval performance (high row: Kather; low row: BreakHis)

[7, 8], and [26]. The performance of these loss functions is acceptable when dealing with large hash codes but may not be satisfactory for short binary values. In addition, our triplet loss function outperforms the mentioned triplet loss functions in retrieving histopathology images. In HATDH-V11, we apply our attention module without separating the axes in the spatial component. The findings indicate that the separation strategy improves the effectiveness of our attention module for feature extraction. A detailed description of HATDH and its subtypes can be found in Table 5.

In summary, the reported results reveal that using the attention modules can improve the performance of our model in the feature extraction phase, leading to increased image retrieval accuracy. Also, the designed attention module outperforms the others. Further investigation reveals that the introduced loss function can increase MAP by approximately 3–12% and 5–7% over the other types of triplet loss functions for the Kather and BreakHis datasets, respectively. Additionally, the results indicate that the separation of axes in the spatial component enhances MAP by around 2–5%.

Effects of Our Triplet Loss Function on the Results

Compared with prior triplet loss functions, such as [7, 8], and [26], the suggested triplet loss function takes into account pair inputs separately (anchor with positive and negative) in addition to triplet inputs, which can improve retrieval performance in various histopathology databases. Figure 6 (b) shows that having the hyperparameters λ_2 (for anchor and positive pairs) and λ_3 (for anchor and negative pairs) improves MAP by approximately 1.2% when compared with the conventional triplet loss function with only one hyperparameter (λ_1). Moreover, the quantization term can assist in the generation of binary codes with a high level of accuracy utilizing β, resulting in an increase of around 1.6% in MAP (Fig. 6(c)). On the other hand, based on the experimental results of HATDH-V8, HATDH-V9, and HATDH-V10 (Tables 3 and 4), the designed triplet loss function, employing different hyperparameters (λ_1, λ_2, λ_3, and β), outperforms the recently proposed triplet loss functions, including [7, 8], and [26], for the used histopathological datasets.

Effects of Our Attention Module on the Results

While our novel attention module follows a similar structure to CBAM, it differs in a few ways. One of the main differences is that the spatial attention component uses the separation in two axes technique. The technique captures direction-aware and position-sensitive information, enabling us to focus on the main objective in complicated images [30]. HATDH and HATDH-V2 results in Tables 3 and 4 demonstrate that utilizing our attention module instead of CBAM increases MAP by 3.92% and 4.28% for Kather and BreakHis, respectively, on average. Additionally, the proposed module provides superior performance in feature extraction for histopathological images when compared to other attention modules, such as ECA, HAM, and coordinate. Further, based on the experimental results of HATDH and HATDH-V11, the applying separation strategy to the model can improve its performance in feature extraction, leading to an increase in MAP of 3.8–5.03%. This strategy captures spatial correlations along one axis while retaining positional details along the other, facilitating the feature extraction process by accurately locating a desired target in histopathology images (Table 5).

Discussion

Histopathological databases may include asymmetrical distributions of samples within classes, which can negatively impact retrieval efficiency [16]. Histopathological images present a microscopic view of tissues and differ from other kinds of medical images, which poses challenges in collecting and employing them for deep hashing models [17]. Furthermore, histopathological images might be large and have to be divided into small patches, complicating a CBHIR model and making precise retrieval challenging [18]. In addition, since a CBHIR system can be used to diagnose various types of diseases, it must be able to retrieve samples rapidly and reliably [3]. Choosing and focusing on optimum features is critical while studying histopathology images [33]. To assess image similarity, a CBHIR model compares extracted features of samples using distance metrics, including Hamming and Euclidean [1]. One factor that can reduce the accuracy of a CBHIR system is the feature extraction stage [43]. Since low-level features may not properly describe a histopathological image, deep learning approaches are frequently utilized for automated feature extraction [44]. However, these strategies need a large number of training datasets to function effectively [33]. On the other hand, an effective retrieval process may require high-dimensional features, making a CBHIR model complex and slow [1]. The vanishing gradient problem in deep hashing models can also cause difficulties in generating precise binary codes during a retrieval process [5]. In summary, the errors during a histopathology image retrieval process can be caused by issues with datasets, the feature extraction stage, vanishing gradient, etc.

A histopathology image retrieval process often requires high-dimensional features, making a CBHIR model complex and time-consuming [1]. On the other hand, histopathology databases may be unbalanced and include several classes with limited samples, which can have a negative impact on retrieval [3]. Also, the nature of histopathology images further complicates choosing the best features to represent them [17]. To overcome these problems, for the first time, we propose a novel attention-based triplet deep hashing model for histopathology image retrieval. The hashing concept can help to speed up the training and retrieval stages [33]. Also, the triplet structure is effective in solving the dataset issues [7]. Additionally, the attention mechanism assists in selecting the optimal features from images by focusing on the best region [14]. A hash layer is suggested to address the vanishing gradient problem [5], capable of generating and training binary codes simultaneously. Moreover, the proposed loss function improves training and retrieval performance by considering both pair and triplet inputs. Combined, these contributions make a powerful CBHIR model that outperforms similar ones. It should be noted that this paper aims to utilize and enhance the existing neural network concepts to present an effective CBHIR model.

This paper proposes an attention-based innovative triplet deep hashing model to retrieve histopathology images. One of the significant limitations of studies on histopathology images is associated with datasets. Histopathology databases are typically organized as either whole slide or patch-based images. Using whole slide images can increase the

complexity, especially in deep learning models. On the other hand, patch-based images may not effectively represent the multi-scale characteristics of digital pathology, but they have shown the potential for content-based retrieval [3, 33]. Thus, we opted to use popular patch-based databases to reduce complexity and facilitate our model training.

The next challenge is to select the best features to convert them into binary codes. To achieve this goal, we design a novel attention module that utilizes both channel and spatial information. Despite its positive effects on retrieval results, the presented module may add a bit of complexity to the CNN architecture. Hence, designing a less complex attention module, which focuses only on spatial or channel information, may become a consideration for the next study as well.

Another problem in deep hashing models is producing binary values accurately. Since the derivative of the sign function is zero and undefined for non-zero and zero inputs, respectively, using that directly in deep learning models is not possible (the vanishing gradient issue). Therefore, we suggest an effective hash layer to produce high-accuracy binary codes using a proper approximation of the sign function. Furthermore, the presented loss function can aid in reducing errors when generating binary codes. Nevertheless, the vanishing gradient problem remains an open challenge for deep hashing research.

In this paper, we proposed a novel attention module to improve the feature extraction process, but the use of different aspects of the attention mechanism can still be a significant subject in histopathology analysis. Recently, Song et al. [45] presented a fusion approach using local and global attention methods to enhance an image retrieval process. It is possible to use the idea of this study as a motivation for designing a fusion attention module in retrieving histopathology images, especially since details are very important when analyzing these images. As mentioned, the vanishing gradient issue can be a serious challenge in generating binary codes. Inspired by [5], one of the simple solutions can be to apply similar functions to the sign function with automatic parameters. The pairwise or triplet deep hashing methods can effectively retrieve images from histopathology databases, but applying them to multi-class databases with various samples may pose a challenge. Based on the recent research findings, utilizing quadrupled structures can offer a solution to this issue [46]. Designing a quadruple model with adjustable margins to retrieve histology images can be an interesting subject for future research.

This paper proposes a deep hashing model using pattern recognition methods for histopathology image retrieval. In this case, as with similar works [1, 3], we focus more on the mathematical and machine learning aspects of the problem than on the pragmatic ones, which may negatively affect its cognitive and practical value. Although this problem can often be ignored in academic research [1], we will strive to address and consider it in our next studies.

Table 6 Results of the training and retrieval times for one iterative in seconds

Methods	*Training*	*Retrieval*
HATDH (ours)	168.06	0.83
ATH	290.22	2.14
HashNet	249.50	1.52
DPSH	209.31	1.38

Assume efficiency equals the ratio of our outputs to inputs. For this study, inputs can include the CNN models, training images, and used computers. The outputs are the retrieved images, evaluated using MAP and AC. According to previous research [4, 5], since we use identical inputs for a fair comparison, MAP and AP can serve as appropriate metrics to measure the model efficiency. However, since the structures and strategies of the models may differ, training and retrieval times can also vary. Therefore, we compare our model to three unique deep hashing models, including ATH, HashNet, and DPSH, in terms of the training and retrieval times in the same situation (Table 6). We have not included other approaches since their structures are often similar to those of these models. Furthermore, since we do not take into consideration the loading time of the images, the results are not greatly different for the two datasets. The results demonstrate that our model performs better in training and retrieval than other models.

Our method is a CAD system whose goal is to assist clinicians in making the best decision in a treatment process. Our study simulates a real-world experiment using two popular patch-based datasets, Kather and BreakHis, which have demonstrated their potential for content-based retrieval. According to the results, HATDH is an effective tool for helping clinicians detect diseases such as colorectal and breast cancers with high accuracy. In summary, our study, as well as related works such as [1, 3] and [33], only suggests a CAD model that can be used effectively in retrieving histopathology images based on academic evaluation metrics. However, using this model and similar methods directly in hospitals is neither the goal of these works nor may it be feasible.

The proposed CBHIR system can aid in the early detection of various cancers, including breast and colorectal, resulting in decreased mortality rates. Although final diagnosis is still carried out by doctors using biopsy and other clinician tools, HATDH can identify suspect cases rapidly and accurately to facilitate further evaluation. The model also enables doctors to search and find similar cases for a query sample, allowing them to select the best treatment option. Moreover, the proposed model can help minimize

human errors in the detection process. All techniques and novelties used in designing HATDH, such as using the triple deep hashing structure and attention mechanism, have been employed to increase the accuracy and reliability of the model. In the future, however, HATDH should be tested more in clinical settings to evaluate and improve its performance in practical studies.

In this paper, we propose a novel attention-based triplet deep hashing model to retrieve histopathology images. As previous studies have shown [7, 8], a triplet structure may yield better results in multi-class datasets compared to pairwise methods. Furthermore, the process of extracting features from histopathology images can be challenging due to their special nature and different sizes. Therefore, we introduce a new attention module integrated into the CNN architecture to improve feature extraction. The separation concept employed in designing our module helps to focus better on specific regions and select the best features. In addition, since high-dimensional features are often required to represent histology images, an effective hash layer is suggested to generate binary codes of varying lengths, resulting in speeding up the training and retrieval phases as well as addressing the problem of vanishing gradients. Also, we introduce an improved triplet loss function considering pair inputs separately in addition to triplet inputs for enhancing efficiency during the training and retrieval phases. Moreover, the presented loss function can decrease the error between generated hash codes and real ones. As a result of these novelties, our model performs better than other hashing-based approaches when retrieving histopathology images.

Generally, in contrast to most medical images that often focus on one part of the body, histopathology images can be used to detect various diseases in different parts of the body [18]. Despite this, retrieving them by traditional methods can be challenging because of the structural complexity, uneven distribution, and other problems [1]. Furthermore, an effective retrieval procedure may involve high-dimensional features, making the process time-consuming and complex [3]. On the other hand, choosing optimized features and focusing on the considered region in images can be important [14]. Therefore, we present a novel attention-based triplet deep hashing model for tackling these issues. MobileNet is chosen as the base model because it is easy to implement and effective in the analysis of histopathology images [9]. MobileNet also appears to be less sensitive to the number of training samples as compared to other CNN models, such as the VGGNet and ResNet [12]. Tables 3 and 4 demonstrate that our CNN model performs better than other models, such as VGG19 and ResNet101, in feature extraction. For improving feature extraction, an attention module is proposed. The presented module may add some complexity to the CNN architecture. Therefore, developing a less complex attention module, which focuses only on spatial or channel information, may be considered for the next work. However, its effectiveness in comparison to other modules can be restricted.

Several experiments are conducted to assess different components of our model, resulting in the introduction of multiple subtypes of HATDH (Table 5). HATDH-V1 to HATDH-V4 utilize ECA [29], CBAM [14], HAM [13], and coordinate attention modules in place of our attention module. Other parts of these subtypes, such as the datasets, CNN model, and loss function, are identical to HATDH. As indicated in Tables 3 and 4, using our attention module can lead to better results than others. To evaluate the effect of the concept of the attention mechanism on retrieval performance, we use three popular CNN models, including MobileNet [12], VGG19 [10], and ResNet101 [11], instead of our attention-based CNN model (HATDH-V5, HATDH-V6, and HATDH-V7). As per our model, the softmax layer of these three models is supplanted by the suggested hash layer. In this study, the datasets and the loss function remain unchanged. The results show that applying attention modules to a deep hashing structure can improve retrieval accuracy. In HATDH-V8, HATDH-V9, and HATDH-V10, only our loss function is replaced by [7, 8], and [26] without changing other parts. The results demonstrate that our loss function outperforms other recently proposed ones.

Content-based histopathology image retrieval can aid in the detection of numerous diseases, including breast and colorectal cancers [1]. For a query sample, several database images are ranked, and the class can be determined based on the ranked samples [3]. For example, if the top ten images retrieved correspond to the cancer category, we can classify the query as a cancer case. Human investigation of all images in a dataset may be error-prone and time-consuming [1]. By using the techniques mentioned, HATDH can rank a wide range of similar images for a query accurately and automatically. As a result, the detection process can be carried out with high efficiency and minimal error rates. By utilizing the proposed model, various diseases can be diagnosed based on similar cases without undergoing heavy surgery or paying high prices. Furthermore, clinicians can make the best treatment decisions based on similar samples. The evaluation metrics used in this study, including MAP and Precision, can demonstrate the reliability of HATDH during a diagnosis process. The high MAP and precision of our model indicate that if a query sample is labeled as cancer, doctors can be fairly confident in the classification and take action to save the patient. In particular, a high MAP shows that HATDH has effectively searched and retrieved numerous samples, enabling clinicians to make more informed decisions with increased data. On the other hand, our model saves and visualizes results, which can be used to evaluate treatment progress. HATDH can ensure a more accurate investigation of suspicious cases, especially

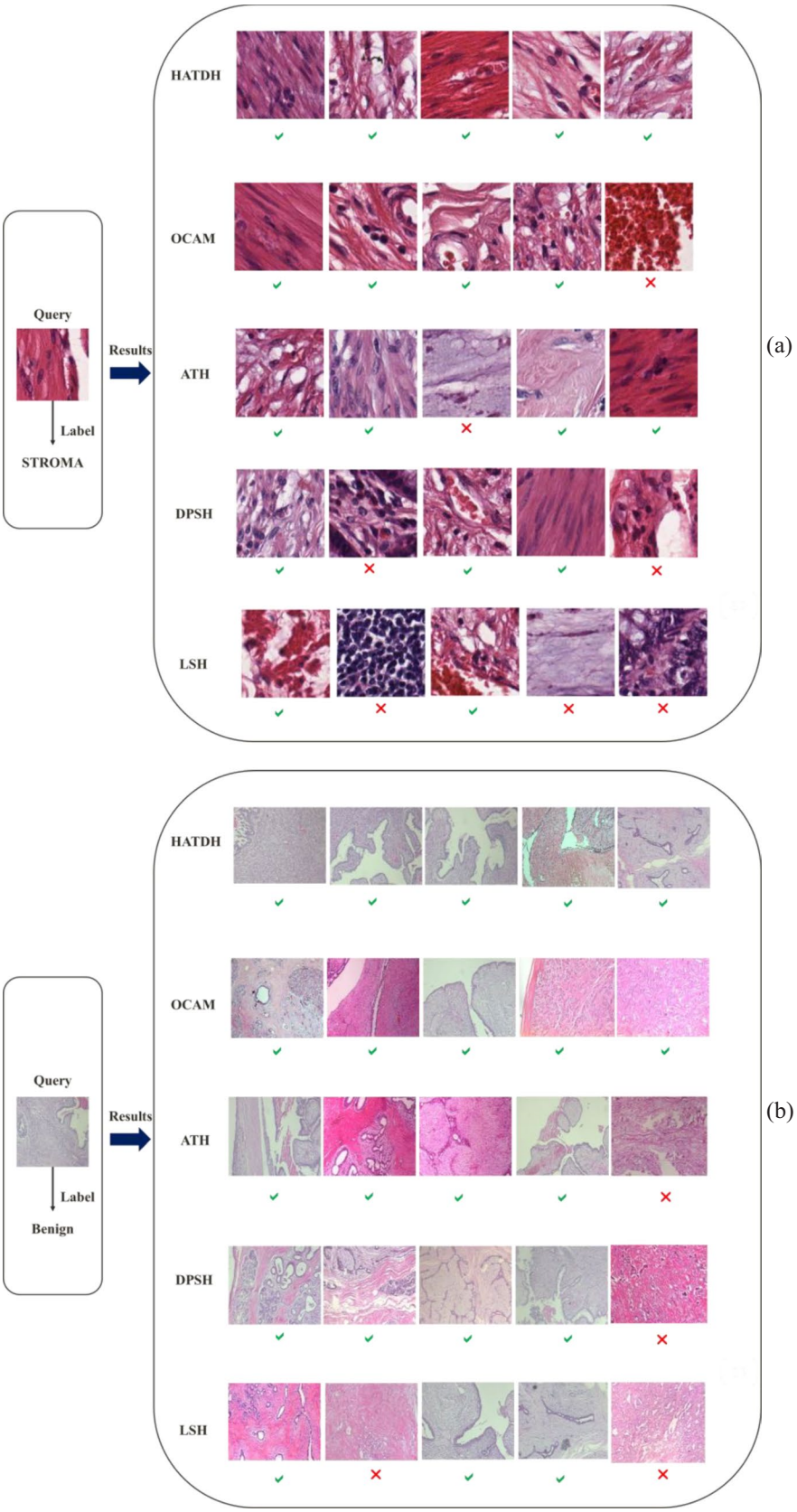


Fig. 7 A selection of 5 retrieved samples of the two studied databases by HATDH and some hashing approaches for a randomly selected query sample (**a** Kather; **b** BreakHis)

in remote areas and emergencies. In addition, HATDH is a very flexible model that can even be applied to other medical images in particular scenarios. The purpose of this study is to introduce HATDH as a powerful CBHIR model with the potential to aid in accurate and rapid disease recognition. Hence, using this model in a clinical setting for more accurate practical evaluation could be an interesting future study topic.

Top Retrieval Results

Figure 7 represents the top five retrieved images by HATDH and some hashing approaches for a query image. Based on its methodology, each model returns the five most similar images. Therefore, the priority of retrieved images may be different. For example, the first similar image in our model may be the sixth in another model. In fact, it is important to determine how many images are retrieved correctly among these five images. In summary, this study shows the accuracy of various models in retrieving histopathology images visually. As a result of the comparison, HATDH appears to be the most effective approach.

Conclusion

This paper proposed an innovative triplet deep hashing model based on the attention mechanism to retrieve histopathology images, called histopathology attention triplet deep hashing (HATDH). Three deep attention-based hashing models with identical architectures and weights were employed to produce binary values. We designed an attention module to enhance the performance of CNNs during the feature extraction process. Additionally, a new triplet loss function was introduced to increase the efficiency of our model in the training and retrieval stages by incorporating pair inputs along with triplet inputs. Based on experiments conducted on two public histopathology databases, HATDH was shown to be superior to the cutting-edge hashing methods.

This study aims to present an effective CBHIR model to assist clinicians in the rapid and precise detection of various diseases. Although the suggested model serves as a support tool, clinicians make the final decision. In this regard, we believe the publication of this study would be beneficial in bridging the gap between academic and practical settings, assisting in the treatment process, decreasing patient mortality rates, and minimizing human error. As with other works, our model may have some inherent limitations, which are inevitable. As mentioned in the "Discussion" Section previously, while some challenges were acknowledged and addressed in this study, others will require further research in the future and may not align with the current study.

This study analyzed patch-based histopathology databases. Therefore, in future research, additional histopathology datasets should be examined, specifically large images of whole slides, to obtain a better understanding of the sizes, categories, and other aspects of various databases.

Data Availability The data that support the findings of this study are available on request from the corresponding author.

Declaration

Conflict of Interest The authors declare no competing interests.

References

1. X. Shi, M. Sapkota, F. Xing, F. Liu, L. Cui, and L. Yang, Pairwise based deep ranking hashing for histopathology image classification and retrieval, *Pattern Recognition,* vol. 81, pp. 14-22, 2018.
2. Y. Ma *et al.*, Breast histopathological image retrieval based on latent dirichlet allocation, *IEEE journal of biomedical and health informatics,* vol. 21, no. 4, pp. 1114-1123, 2016.
3. Y. Gu and J. Yang, Densely-connected multi-magnification hashing for histopathological image retrieval, *IEEE journal of biomedical and health informatics,* vol. 23, no. 4, pp. 1683-1691, 2018.
4. W.-J. Li, S. Wang, and W.-C. Kang, Feature learning based deep supervised hashing with pairwise labels, *arXiv preprint arXiv:1511.03855,* 2015.
5. Z. Cao, M. Long, J. Wang, and P. S. Yu, Hashnet: Deep learning to hash by continuation, in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5608–5617.
6. Y. Liang, Y. Pan, H. Lai, W. Liu, and J. Yin, Deep Listwise Triplet Hashing for Fine-Grained Image Retrieval, *IEEE Transactions on Image Processing,* vol. 31, pp. 949-961, 2021.
7. J. Fang, H. Fu, and J. Liu, Deep triplet hashing network for case-based medical image retrieval, *Medical image analysis,* vol. 69, p. 101981, 2021.
8. B. Liu, Y. Cao, M. Long, J. Wang, and J. Wang, Deep triplet quantization, in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 755–763.
9. N. Dif, M. O. Attaoui, Z. Elberrichi, M. Lebbah, and H. Azzag, Transfer learning from synthetic labels for histopathological images classification, *Applied Intelligence,* vol. 52, no. 1, pp. 358-377, 2022.
10. K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* , 2014.
11. K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
12. A. G. Howard *et al.*, Mobilenets: Efficient convolutional neural networks for mobile vision applications, *arXiv preprint* arXiv: 1704.04861, 2017.
13. G. Li, Q. Fang, L. Zha, X. Gao, and N. Zheng, HAM: Hybrid attention module in deep convolutional neural networks for image classification, *Pattern Recognition,* vol. 129, p. 108785, 2022.
14. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, Cbam: Convolutional block attention module, in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.
15. X. Li *et al.*, Image retrieval using a deep attention-based hash, *IEEE Access,* vol. 8, pp. 142229-142242, 2020.

16. F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, A dataset for breast cancer histopathological image classification, *IEEE Transactions on Biomedical Engineering,* vol. 63, no. 7, pp. 1455-1462, 2015.
17. E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan, Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review, *Expert Systems with Applications,* p. 114161, 2020.
18. X. Wang *et al.*, RetCCL: Clustering-guided contrastive learning for whole-slide image retrieval, *Medical Image Analysis,* vol. 83, p. 102645, 2023.
19. X. Luo *et al.*, A survey on deep hashing methods, *ACM Transactions on Knowledge Discovery from Data,* vol. 17, no. 1, pp. 1-50, 2023.
20. A. Gionis, P. Indyk, and R. Motwani, Similarity search in high dimensions via hashing, in *Vldb*, 1999, vol. 99, no. 6, pp. 518-529.
21. Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval, *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, no. 12, pp. 2916-2929, 2012.
22. F. Shen, C. Shen, W. Liu, and H. Tao Shen, Supervised discrete hashing, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 37–45.
23. H. Liu, R. Wang, S. Shan, and X. Chen, Deep supervised hashing for fast image retrieval, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2064–2072.
24. X. Wang, Y. Shi, and K. M. Kitani, Deep supervised hashing with triplet labels, in *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20–24, 2016, Revised Selected Papers, Part I 13*, 2017, pp. 70–84: Springer.
25. Z. Zhang, Q. Zou, Y. Lin, L. Chen, and S. Wang, Improved deep hashing with soft pairwise similarity for multi-label image retrieval, *IEEE Transactions on Multimedia,* vol. 22, no. 2, pp. 540-553, 2019.
26. Ş. Öztürk, E. Çelik, and T. Çukur, Content-based medical image retrieval with opponent class adaptive margin loss, *Information Sciences,* p. 118938, 2023.
27. C. Deng, Z. Chen, X. Liu, X. Gao, and D. Tao, Triplet-based deep hashing network for cross-modal retrieval, *IEEE Transactions on Image Processing,* vol. 27, no. 8, pp. 3893-3903, 2018.
28. Z. Niu, G. Zhong, and H. Yu, A review on the attention mechanism of deep learning, *Neurocomputing,* vol. 452, pp. 48-62, 2021.
29. Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11534–11542.
30. Q. Hou, D. Zhou, and J. Feng, Coordinate attention for efficient mobile network design, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13713–13722.
31. P. Yang *et al.*, A deep metric learning approach for histopathological image retrieval, *Methods,* vol. 179, pp. 14-25, 2020.
32. N. Hashimoto *et al.*, Case-based similar image retrieval for weakly annotated large histopathological images of malignant lymphoma using deep metric learning, *Medical Image Analysis,* vol. 85, p. 102752, 2023.
33. S. M. Alizadeh, M. S. Helfroush, and H. Müller, A novel Siamese deep hashing model for histopathology image retrieval, *Expert Systems with Applications,* vol. 225, p. 120169, 2023.
34. R. Hang, Z. Li, Q. Liu, P. Ghamisi, and S. S. Bhattacharyya, Hyperspectral image classification with attention-aided CNNs, *IEEE Transactions on Geoscience and Remote Sensing,* vol. 59, no. 3, pp. 2281-2293, 2020.
35. Q. Zhou *et al.*, Fine-grained spatial alignment model for person re-identification with focal triplet loss, *IEEE Transactions on Image Processing,* vol. 29, pp. 7578-7589, 2020.
36. G. He, F. Li, Q. Wang, Z. Bai, and Y. Xu, A hierarchical sampling based triplet network for fine-grained image classification, *Pattern Recognition,* vol. 115, p. 107889, 2021.
37. J. N. Kather *et al.*, Multi-class texture analysis in colorectal cancer histology, *Scientific reports,* vol. 6, no. 1, pp. 1-11, 2016.
38. M. Yazdi and H. Erfankhah, Multiclass histology image retrieval, classification using Riesz transform and local binary pattern features, *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,* vol. 8, no. 6, pp. 595-607, 2020.
39. M. Liu *et al.*, A deep learning method for breast cancer classification in the pathology images, *IEEE Journal of Biomedical and Health Informatics,* vol. 26, no. 10, pp. 5025-5032, 2022.
40. H. Erfankhah, M. Yazdi, M. Babaie, and H. R. Tizhoosh, Heterogeneity-aware local binary patterns for retrieval of histopathology images, *IEEE Access,* vol. 7, pp. 18354-18367, 2019.
41. R. Karthik, R. Menaka, and M. Siddharth, Classification of breast cancer from histopathology images using an ensemble of deep multiscale networks, *Biocybernetics and Biomedical Engineering,* vol. 42, no. 3, pp. 963-976, 2022.
42. D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv preprint* arXiv:1412.6980, 2014.
43. X. Li, J. Yang, and J. Ma, Recent developments of content-based image retrieval (CBIR), *Neurocomputing,* vol. 452, pp. 675-689, 2021.
44. J.-M. Chen *et al.*, Computer-aided prognosis on breast cancer with hematoxylin and eosin histopathology images: A review, *Tumor Biology,* vol. 39, no. 3, p. 1010428317694550, 2017.
45. C. H. Song, H. J. Han, and Y. Avrithis, All the attention you need: Global-local, spatial-channel attention for image retrieval, in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2754–2763.
46. Q. Qin, L. Huang, K. Xie, Z. Wei, C. Wang, and W. Zhang, Deep adaptive quadruplet hashing with probability sampling for large-scale image retrieval, *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 33, no. 12, pp. 7914-7927, 2023.