Article

Advancing Content-Based Histopathological Image Retrieval Pre-Processing: A Comparative Analysis of the Effects of Color Normalization Techniques

Zahra Tabatabaei [1,2,*], Fernando Pérez Bueno [3], Adrián Colomer [2], Javier Oliver Moll [1], Rafael Molina [4] and Valery Naranjo [2]

[1] Department of Artificial Intelligence, Tyris Tech S.L., 46021 Valencia, Spain
[2] Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, HUMAN-Tech, Universitat Politècnica de València, 46022 Valencia, Spain
[3] Basque Center on Cognition, Brain and Language, 20009 San Sebastián, Spain; f.perezbueno@bcbl.eu
[4] Department of Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, 18071 Granada, Spain
[*] Correspondence: elec.tabatabaei@gmail.com or zahra.tabatabaei@tyris-software.com

Abstract: Content-Based Histopathological Image Retrieval (CBHIR) is a search technique based on the visual content and histopathological features of whole-slide images (WSIs). CBHIR tools assist pathologists to obtain a faster and more accurate cancer diagnosis. Stain variation between hospitals hampers the performance of CBHIR tools. This paper explores the effects of color normalization (CN) in a recently proposed CBHIR approach to tackle this issue. In this paper, three different CN techniques were used on the CAMELYON17 (CAM17) data set, which is a breast cancer data set. CAM17 consists of images taken using different staining protocols and scanners in five hospitals. Our experiments reveal that a proper CN technique, which can transfer the color version into the most similar median values, has a positive impact on the retrieval performance of the proposed CBHIR framework. According to the obtained results, using CN as a pre-processing step can improve the accuracy of the proposed CBHIR framework to 97% (a 14% increase), compared to working with the original images.

Keywords: color normalization; computer-aided diagnosis (CAD); content-based image retrieval (CBIR); histopathological images; whole-slide images (WSIs)

1. Introduction

Breast cancer is one of the most prevalent cancer types, with 2.3 million women diagnosed with this cancer and 685,000 deaths globally in 2020 [1]. For this large amount of patients, medication should be accurate with little to zero margin for error; otherwise, the consequences of a wrong diagnosis could be fatal [2]. Also, since breast cancer is one of the leading causes of death for women globally, precise detection and timely treatment can enhance the chances of recovery [3]. Computer-aided diagnosis (CAD) provides some deep learning (DL)-based techniques in digital pathology that can assist pathologists to make more accurate cancer diagnoses [4]. These techniques need to be trained by medical images such as those from magnetic resonance imaging (MRI) [5] and histopathological images [6]. A unique feature of histopathological images is that they are typically much larger than other medical images [7].

Histopathological images play a crucial role in the realm of medical image processing, allowing the integration of image information and pathologists' expertise to improve diagnosis [8]. Medical images have witnessed a rapid expansion in quantity, content, and dimensions [9]. Due to an enormous increase in the number of diverse clinical exams and the availability of a wide range of image modalities, the demand for efficient medical

image data retrieval and management has increased [10]. Current approaches to medical image retrieval often rely on alphanumeric keywords assigned by human experts, enabling retrieval at a conceptual level. However, this text-based search methodology falls short in capturing the intricate visual features inherent in image content [11].

Recent Content-Based Histopathological Image Retrieval (CBHIR) methods support full retrieval via visual content and histopathological patterns of the tissue [12]. These advanced CBHIR tools facilitate searching at a perceptual level [13]. It is noteworthy to mention that a single pathology image may contain just basic patterns of tissue such as the epithelium and connective tissue. However, the actual number of patterns in the DL-based technique's point of view is almost infinite.

CBHIR explores a database to find visually similar images to provide clinicians with comparable lesions. In the diagnostic workflow, pathologists utilize this search engine to reach top K similar to their queries to determine if a histological feature is malignant or benign [14].

Mainly, CBHIR tools work on the extracted features of the images, including color, texture, shape, etc. Color is a visual feature that plays an important role in CBHIR techniques due to its invariance with respect to image scaling, translation, and rotation [15]. The use of color improves the capturing of distinctive histopathological features. This provides valuable information about the distribution and arrangement of different tissue components in histopathology [16].

In histopathology, tissues must be stained using various dyes, including hematoxylin and eosin (H&E), in order to be readable for pathologists [9,17]. In digital pathology, these tissues must be scanned as whole-slide images (WSIs) [18]. In addition to the use of different scanners and staining manufacturers, lab conditions and temperatures may cause color variation in WSIs [19]. Color variation in WSIs can arise from both inter- or/and intra-laboratory factors in the acquisition procedure [16]. Figure 1 shows the color variation across five different collaborating hospitals involved in the collection of the CAMELYON17 challenge (CAM17) data set [20]. This diverse color variation misleads the model and potentially deceives the feature extractors (FEs) into extracting erroneous features. Also, the performance of CAD tools may be significantly influenced by these variations in WSIs [21].

In this paper, we propose a CBHIR framework that includes color normalization (CN) as a pre-processing step to address this issue. CN methods have been developed with the aim of transferring the color interval of the WSIs in a data set to a common color range [19,22,23].

| Center 4 | Center 3 | Center 2 | Center 1 | Center 0 |

Figure 1. Image examples from five hospitals that collaborated to collect images into a single data set (CAM17).

2. Related Work

2.1. Content-Based Image Retrieval

There are different approaches to Content-Based Histopathological/Medical Image Retrieval techniques in histopathology [24], which we briefly review here. Ref. [25] proposed a framework for size-scalable query regions of interest (ROIs), including epithelial breast tumors in the Motic data set. This work reached 96% precision in retrieving the top 20 similar images. In [25], the authors applied the proposed technique on the Motic database (Motic (Xiamen) Medical Diagnostic Systems Co. Ltd., Xiamen, China) with the original color of 145 stained WSIs. The authors of [26] applied a supervised kernel hashing technique on several thousand histopathological images from breast microscopic

tissue. The precision result for the top 30 retrieved images was reported to be 77.0%. In this study, they considered the gradient of the pixels around the detected regions to make the model robust to subtle changes in the color of patches. A combination of unsupervised feature learning (UFL) with the classical bag of features (BOF) was introduced in [27]. This proposed method was evaluated in particular histopathological images to show that the learned representation has a positive impact on the retrieval performance. CBHIR based on multi-scale, multichannel decoded local ternary pattern features and VLAD coding was presented in [28]. The authors evaluated their method on the KIMIA Path960 data set [29] by retrieving the top 10 matching images from the data set. The authors of [15] presented a modified convolutional auto encoder (CAE) to extract features of patches from SICAPv2 as the largest pixel-annotated prostate data set [30]. The experimental results of this paper demonstrated 85% and 78% accuracy for the top 7 and top 5, respectively. However, one of the challenges that the authors had with SICAPv2 was the color variation of the images in the data set. To deal with this variation, they applied a simple histogram equalization to the patches, but still, this issue affected the results.

Among the previously mentioned studies, only refs. [15,26] applied CN techniques to tackle color variation, while the rest did not address this issue. Another challenge in CBHIR is the lack of annotated images, which the papers mentioned above struggled with.

2.2. Color Normalization

CN is the most used technique to deal with color variation [19]. CN can standardize images by referencing an image and simulating a chosen staining procedure. CN methods normalize images with different techniques, including histogram matching, color transfer, and spectral matching [19]. Histogram matching disregards stain separation, color transfer modifies colors based on statistical correspondences between histological regions, and spectral matching estimates stain concentrations and color properties [31].

Among a variety of CN techniques, some are more popular, such as those proposed in [32,33]. Macenko et al. [32] introduced the Mac technique (the proposed method in [32] is named Mac in this paper) in 2009, assuming that the amount of protein or nucleic acids is a random variable. Mac utilizes Singular Value Decomposition (SVD) to separate H&E channels. Within this technique, the concentration intensity of both the source and target images is scaled using the 99th percentile to achieve a robust estimation of the maximum.

Vahadane et al. [33] published the Vah technique (the proposed method in [33] is named Vah in this paper) in 2016 to model the physical phenomena that define tissue structures. In this technique, there is a preferable stain color, based on pathologists' opinion, and a stain density map. Using an unsupervised approach, this method decomposes images into stain density maps. When Vah is applied to a specific image, it combines the relevant stain density maps based on a pathologist's preference for stain color. This process selectively modifies the color while preserving the underlying structure as described by the maps.

The most recent techniques in CN are the auto encoders (AEs), generative adversarial networks (GANs), and Bayesian K-Singular Value Decomposition (BKSVD) [34–36]. Bentaieb et al. [34] applied a GAN to combine color normalization and the classification of WSIs. In this method, a generator is employed as a stain transfer network, while a discriminator separates the classes and the original and normalized images. In order to map unpaired images between two scanners, the cycleGAN was used by StainGAN [31]. In [35], the authors used three different convolutional neural network (CNN) models for CN purposes: the variational auto encoder (VAE), a GAN, and the deep convolutional Gaussian mixture model (DCGMM). In [37], a CN network is fed by a heavily augmented data set and trained to reconstruct the original appearance of WSIs. The Pix2pix conditional GAN framework and CycleGAN are other notable CNN architectures.

BKSVD [36], a cutting-edge technique that was proposed in 2022, utilizes an unsupervised estimation of the stain concentration that preserves histological structures with variational and empirical Bayes.

In this paper, we provide a deep understanding of the effects of CN on the extracted features in the proposed CBHIR tool. Our experiments were conducted by applying three CN techniques in the pre-processing step in order to explore how CN influences the extracted features within the CBHIR tool. Among the CN techniques, Mac [32] and Vah [33] were selected as the two most used CN techniques and BKSVD was applied as a recent CN technique. Furthermore, in this work, we applied an unsupervised FE to tackle the need for a pool of histopathological images to train the deep learning model.

The main contributions of this paper are as follows:

1. Proposal of a new CBHIR framework with an unsupervised feature extractor that includes color normalization as a pre-processing step;
2. Analysis of CBHIR's performance when using normalized images in comparison with original images;
3. We draw attention to the relevance of color variation and its impact on CBHIR;
4. We provide a comprehensive performance assessment of the proposed method. This evaluation employs a large breast cancer database scored from five distinct laboratories. This evaluation has a more restrictive K-top accuracy assessment compared to recent state-of-the-art studies and also involves an in-depth analysis of retrieving patches with the same cancer label.

3. Methodology

This section introduces the three levels of the proposed CBHIR tool in detail. Figure 2 provides an overview of the proposed CBHIR tool, comprising three levels: pre-processing, training, and searching.
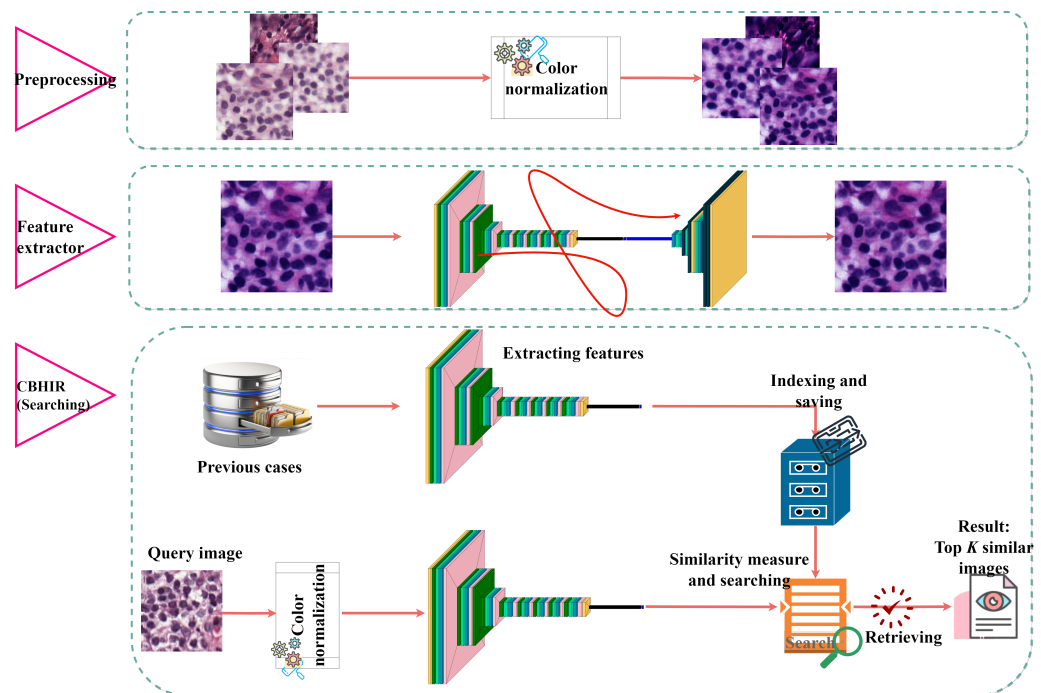
Figure 2. Three main stages of CBHIR, including pre-processing, training, and searching. The pre-processing was performed following [32,33,36]. The training stage consists of CAE training. Here, each layer of the CAE is presented in a different color. Conv2D, Dropout, Dense, and Flatten are shown in green, pink, blue, and black, respectively. The searching stage consists of extracting features, indexing, searching, and displaying the top K similar retrieved images.

3.1. Pre-Processing

In this paper, Mac [32], Vah [33], and BKSVD [36] are the three CN techniques, spanning from 2009 to 2022, that were investigated to normalize the data set. These three CN techniques were applied to the original images of the CAM17 data set, and the results were

collected as a separate data set. The aim is to gain insight into the effectiveness of these CN techniques in tackling the effects of color variation on the results of the CBHIR tool.

It is important to establish and maintain uniformity in color normalization throughout the whole framework. As can be seen in Figure 2, the input query needs to pass through a pre-processing step before being fed to the FE. Furthermore, a crucial consideration is that the CN technique which is applied to the data set in the pre-processing step should match the CN technique employed in the query. Ensuring consistency in the CN technique is imperative for accurate processing and analysis.

3.2. Feature Extractor

In this paper, a CAE is utilized to extract the representative histopathological features of the patches. The CAE aims to solve the back-propagation problem in an unsupervised manner by only relying on the input image as a teacher by itself [38]. It has the in-built ability to compress the data efficiently by extracting the important features and removing noise in the data [39]. These features bring benefits to CBHIR, such as ignoring the noise in WSIs that might be noticeable due to the scanners. Also, they can reduce the demand for annotated images for training an FE, which is expensive and time-consuming [40,41].

Figures 2 and 3 illustrate the structure of the proposed custom-built CAE. In this CAE, a skip connection between a layer in the encoder and its corresponding layer in the decoder can improve the gradient propagation and increase the performance in capturing complex patterns of WSIs. The proposed CAE comprises an encoder with three convolutional layers $[16, 32, 64]$. Moving to the bottleneck, an attention block with a filter size of $[32, 16, 1, 64]$ is introduced to enhance the feature vectors by robustness to noise or occlusions. The decoder is made up of three Conv2DTranspose with $[32, 16, 3]$. The kernel size in this structure was fixed to 3 in all the layers.

The primary objective of the CAE is to reconstruct input images in the output by minimizing the mean squared error (MSE). The CAE workflow involves feeding an input image (X) to the encoder, which compresses the input. Subsequently, the bottleneck compresses the output of the encoder to obtain a feature vector with 200 meaningful features (F). Finally, the decoder reconstructs the output by receiving the feature vector ($Y = Decoder(F)$). In an ideal CAE, the output is identical to the input (*ideal*, $X = Y$). To achieve this ideal goal, the model endeavors to minimize the MSE by comparing the input (X) and the output (Y) [42].

In this paper, the model was trained on a GPU with the NVIDIA GeForce RTX 3090 over 10 epochs, utilizing a batch size of 32 and a learning rate of 0.000001. By discarding the decoder from the CAE, the well-trained FE can be utilized for further experiments.
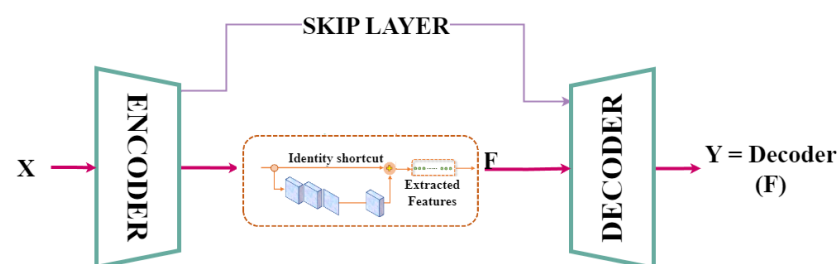
Figure 3. The customized CAE which was utilized as an FE.

3.3. Searching

The search process in CBMIR involves three key steps: similarity calculation, ranking and retrieval, and visualization and presentation. In this paper, at this stage, the well-trained CAE extracts the features of the images of the validation and training sets. These feature vectors are subsequently indexed and saved as the features of the database for the searching process [41].

Upon receiving an input query, the proposed CBHIR technique passes it to the same CN technique, normalizing the whole database. For each query input, a feature vector is

required, encompassing representative features of the query. The feature vector's size is fixed at 200 features, aligning with the number of elements in the feature vectors associated with the database.

After extracting the query's features, the next step involves measuring the similarity between the query feature vector and the indexed feature vectors of the database. To achieve this, a distance function is required. In this paper, the Euclidean distance was selected as one of the most used distance functions in CBHIR [17]. In this context, the distance is inversely related to similarity. Therefore, the most similar images have the smallest distance [15]. Consequently, the top K images with the smallest distance are retrieved and displayed to pathologists for further analysis.

4. Material

CAMELYON17 challenge (CAM17) [20] is a breast cancer metastasis data set in the lymph node sections. It contains WSIs from five different hospitals. Each hospital contributed images from 20 patients, with five slides per patient. The tissues were stained with H&E. Following [36], only the training set of CAM17 was used in the conducted experiments since the WSIs in the test set are not labeled yet.

For the experiments conducted in this paper, images from the first four hospitals that contributed to CAM17 were used as the training and validation set, while images from the fifth hospital, exhibiting a more significant color difference, were used as the test set [36]. The total number of images for training and validating the FE was 48,000 and 12,000, respectively. Then, in the search part, there were 25,406 query images in the test set. Following [43], the experiments in this paper were performed on non-overlapping patches extracted from CAM17, each with a size of 224×224 pixels in the RGB color space. The images were sampled from WSIs containing at least 70% tissue to ensure sufficient histopathological patterns.

5. Experiments and Results

In this section, the procedure is applied for each color version of the data set to report and evaluate the performance of the proposed CBHIR framework.

5.1. Pre-Processing

Figure 4 shows the normalized median intensity (NMI) information for each hospital and method, which is plotted as a violin plot. The NMI is used to assess the effectiveness of each CN technique in normalizing the data set. It involves calculating the median intensity value of pixel values within an image and then normalizing this value. NMI values were computed for individual images within the data set, and metrics such as the standard deviation (NMI SD) and coefficient of variation (NMI CV) were employed. Lower NMI SD and NMI CV values suggest a more consistent normalization across the data set. Ref. [36] presented numerical results for BKSVD, Mac, and Vah across individual clinical centers as well as the entire CAM 17 data set. In this study, Figure 4 visually displays the distribution, variability, and potential outliers of values across different groups, facilitating comparisons between these groups.
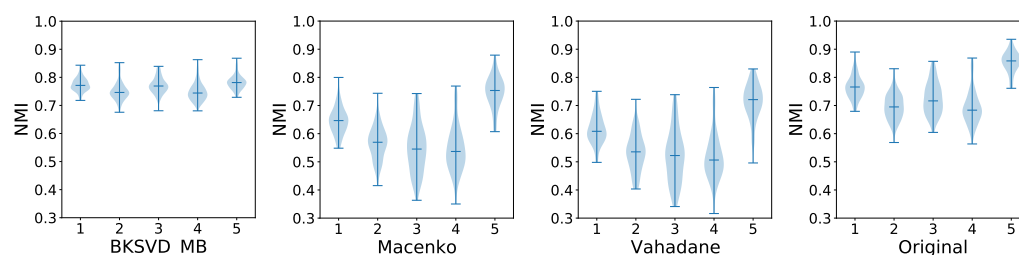

Figure 4. Violin plots of NMI values for each center. The histogram of NMI values for each plot is represented by the blue shadow. Bars mark the maximum, median, and minimum NMI values for each plot. The x-axis corresponds to the hospital numbers.

In refs. [32,33], the color of images for each hospital could be transferred to a similar distribution, but with a larger inter/intra-center than the original images' distribution. Among these three CN methods, BKSVD exhibited the best performance in approximately normalizing the images of all five centers to the same NMI interval. This means that BKSVD transforms the images from each hospital with the lowest intra-center variance and most similar median values, which makes its outputs more interpretable and reliable. To provide a more detailed statistical analysis to strengthen the evidence for the effectiveness of BKSVD compared with Vah and Mac, a quantitative comparison was conducted, Table 1 [36], based on the peak signal to noise ratio (PSNR). The table shows that BKSVD outperforms the rest of the methods, obtaining a higher mean PSNR while requiring a similar time to that required by the Vah method.

Table 1. PSNR for the normalized CAM17 data set.

CN Techniques	BKSVD [36]	Vah [33]	Mac [32]
PSNR	19.54	12.74	13.80

Figure 5 provides visual information about the impact of each CN technique on the color distribution of five random images. Randomly chosen images from all five centers exhibit noticeable color variations, as depicted in the initial line showing the original images. Notably, BKSVD effectively mitigated this color variation, achieving a harmonized and consistent color version that surpassed the performance of Vah and Mac as the classical methods.
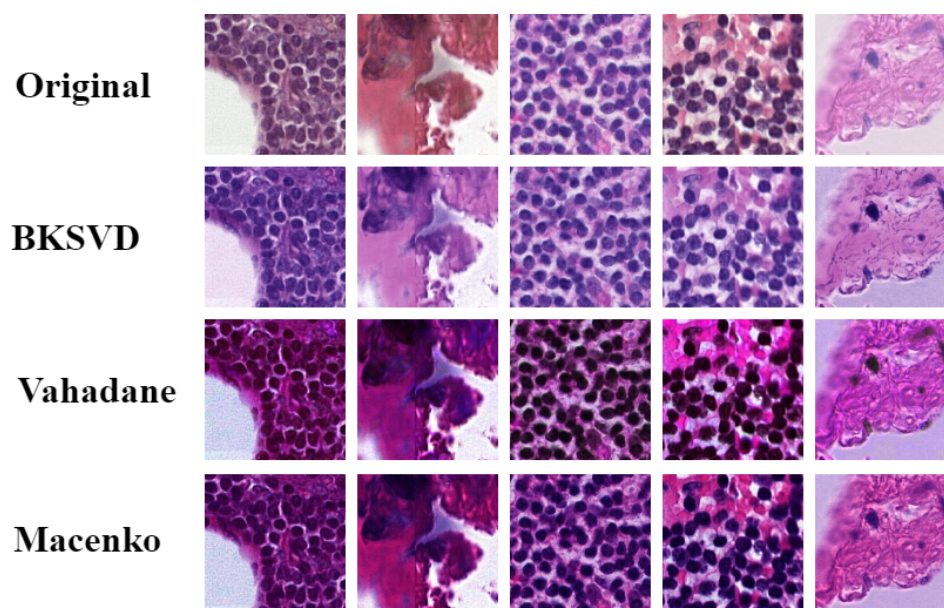
Figure 5. Line 1 illustrates the original color version of CAM17. Lines 2–4 contain five random images of CAM17 as a result of the CN techniques BKSVD, Vah, and Mac. These images correspond to all hospitals.

5.2. CBMIR Results

The most usable strategy to measure the performance of the CBHIR tool is the "top K accuracy". In this strategy, the CBHIR tool displays the top K-matched patches. If there are one or more correct retrieved images among them, the CBHIR tool performs well [7].

$$\begin{cases} ACC = 1, & \text{If any of the K-top retrieved images match with the query,} \\ ACC = 0, & \text{Otherwise} \end{cases}$$

It is noteworthy to mention that in this paper, the results are reported using the top 3, 5, while in recent papers [25,44,45], K was 20, 100, 200, or 400. While the amount of K in this paper is notably lower compared to the other studies, it still shows the model achieved an impressive accuracy. This highlights the model's reliability for pathologists, as it can retrieve similar patches even with a smaller set of retrieved images.

Table 2 reports the results when K = 3 and 5 for CAM17 in the four color versions. After comparing Table 2 and Figure 4, a lower intra-center variance can improve the search performance. According to Table 2, CN as a pre-processing technique might have a negative impact on the final results. The obtained accuracy for the top 3 and 5 in the experiment on the original images is higher than the accuracy of the experiments with the Vah and Mac versions of CAM17. This means that utilizing a insufficient CN technique not only does not improve the final results, but can also decrease the performance of the main model. Figure 6 illustrates four confusion matrices for the four color versions of CAM17 for the top 5.

Table 2. The achieved top K accuracy in the proposed CBIR method on color-normalized images from the CAM17 data set. K = 3 and 5.

K	Original	BKSVD	Vah [33]	Mac [32]
3	0.73	0.91	0.66	0.68
5	0.83	0.97	0.79	0.81
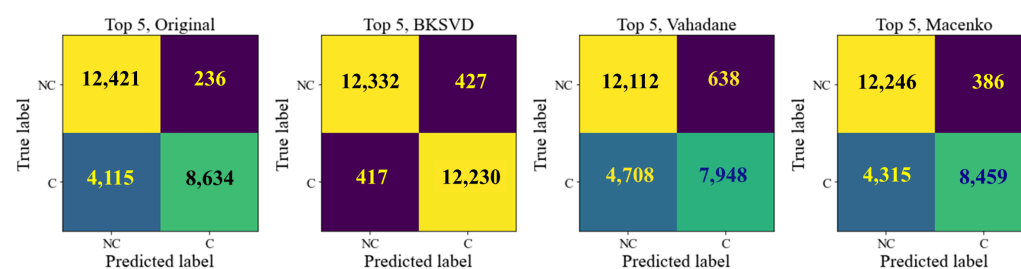
Figure 6. Confusion matrices show the effects of BKSVD, Vah, and Mac on the performance of CBHIR in retrieving the patches with the most similarity. "C" and "NC" stand for cancerous and non-cancerous tissue.

According to the obtained results, thanks to CN techniques, by reducing the negative impacts of color variation, the texture, shape, and histologic features of each patch are highlighted. Therefore, the Euclidean function ranks the images which are more similar in the sense of relevant features.

5.3. Visual Evaluation

A visual evaluation with some sample queries and their retrieved patches provides transparency and insight into the functioning of the CBMIR system. This allows pathologists to understand how the CBMIR framework responds to different queries. The results of feeding the proposed framework with the original and normalized images are presented in the four figures below.

Figure 7 illustrates four random queries in their original color space and the top five retrieved images. As can be seen, for each of the queries, the retrieved images have different colors, not only to the corresponding query, but also among themselves. This highlights the need for color normalization as a pre-processing step for the searching framework.

Figures 8 and 9 show the same random queries as Figure 7, but they were normalized by Mac and Vah, respectively. These figures clarify the numerically reported results and confirm that the effects of Vah as a CN technique can decrease the accuracy of retrieval.

Figure 10 corresponds to the same four queries as Figure 9 and their top five retrieved patches from the data set. In this figure, images were normalized using BKSVD. As can be understood from the figure, normalizing the images using BKSVD can enhance the

performance of the search engine in finding similar histopathological patterns without the negative impact of color variation.
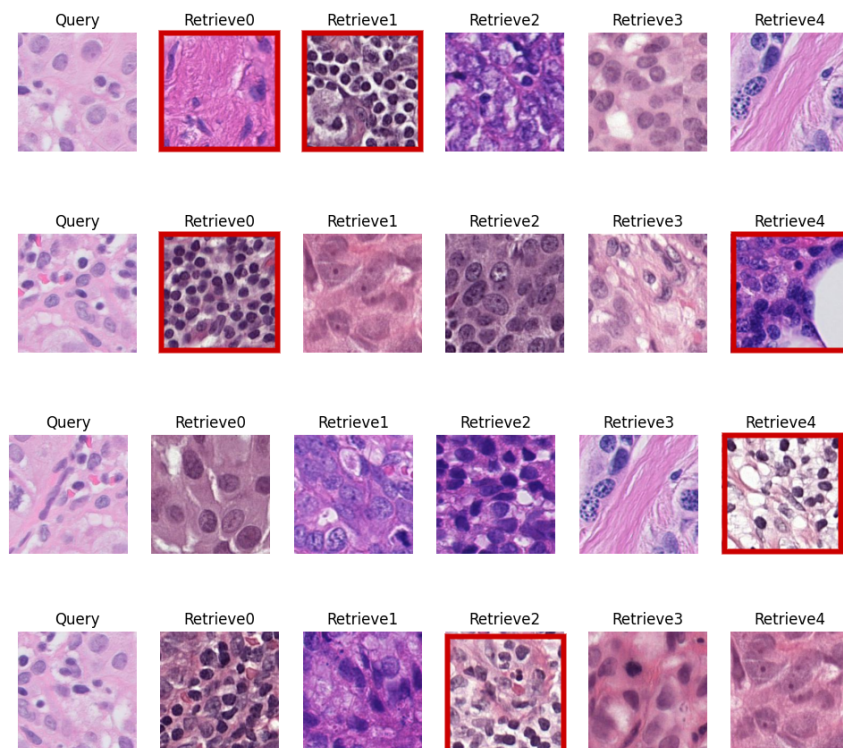
Figure 7. For each random query, the five top similar patches are presented. The red lines show the miss-retrieved patches. The patches are in their original colors.
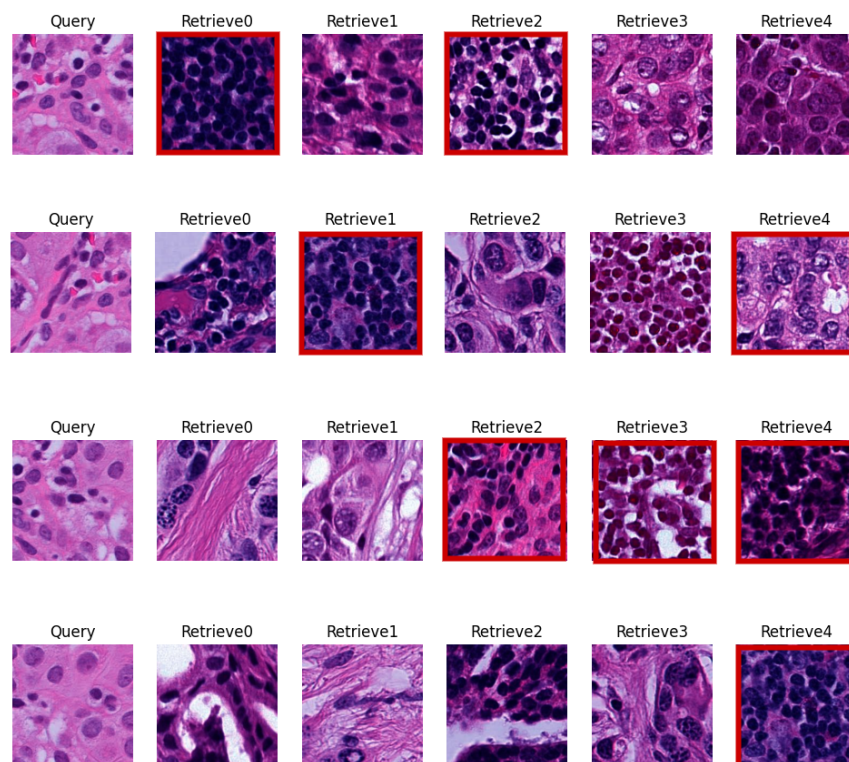
Figure 8. Four random patches of breast cancer data set with their corresponding top five retrieved images. The images were normalized by Macenko and the red boxes show the non-similar patches based on their labels.
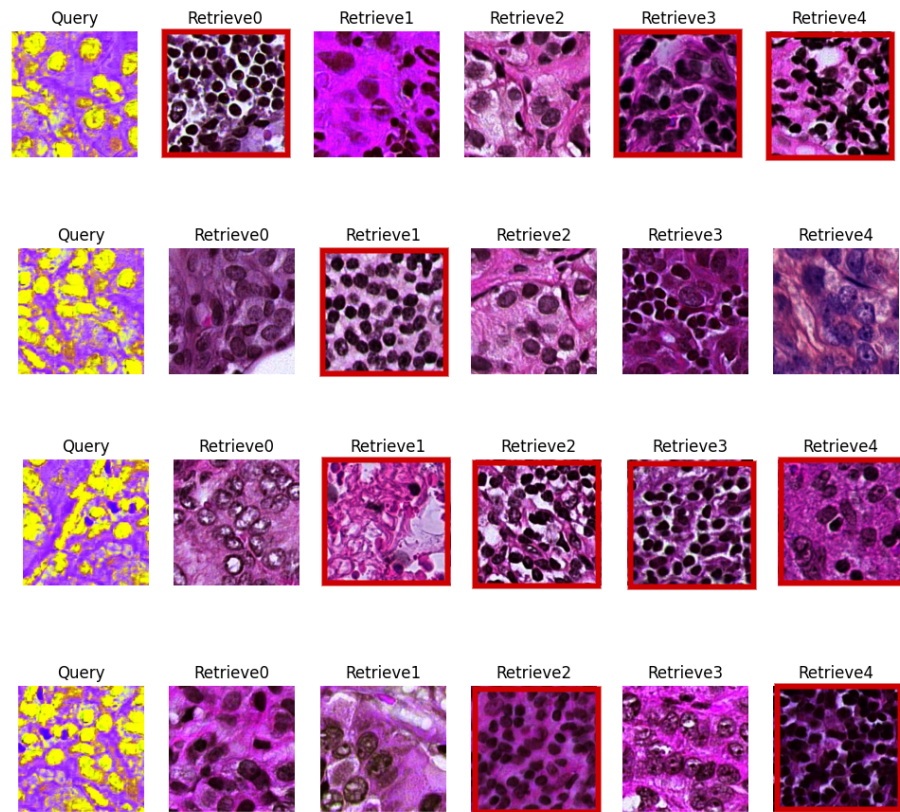
Figure 9. Four random breast queries with their five top retrieved images. The images were normalized by Vahadane. The miss-retrieved images are marked in red.
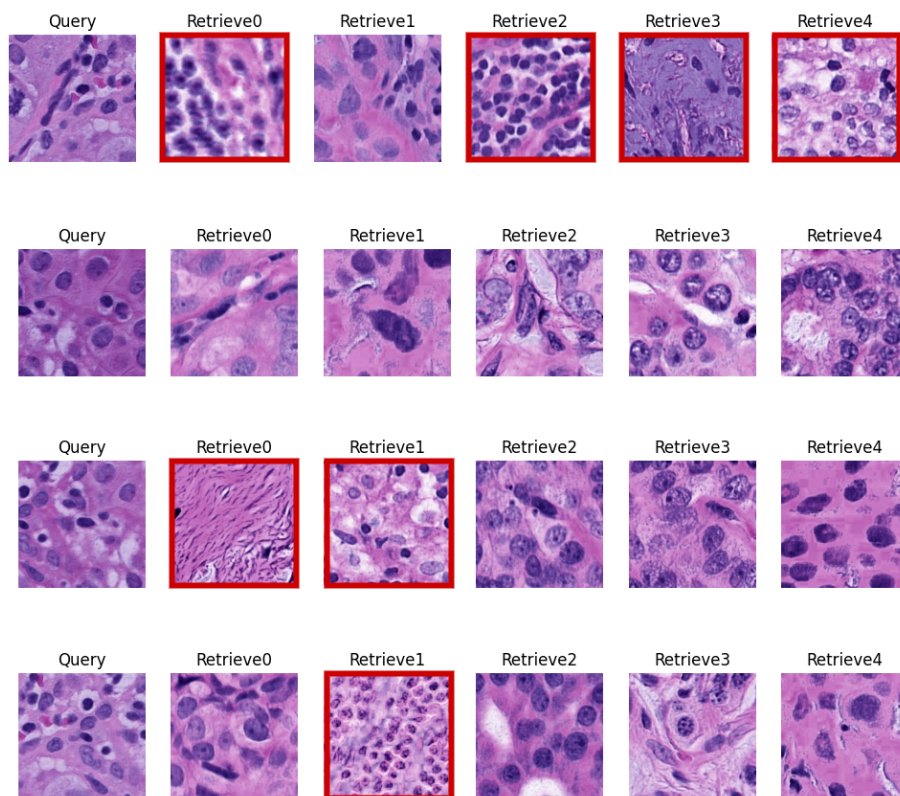
Figure 10. Four random queries from the normalized data set by BKSVD with their five top retrieved images.

In these figures, the images were compared with their queries based on the labels provided by expert pathologists in the ground truth of the data set. The returned patches with the same label as the query were considered as the correct retrieved patches. However, images with different labels to the query were highlighted by a red square in order to clarify the miss-retrieved patches based on their labels.

5.4. Comparing the Results of CBHIR with a Classifier

CBHIR and classification are two different CAD tools for pathologists. CBHIR provides similar patches for pathologists based on the content features of the query and the database. However, the main objective of classification is categorizing images into pre-defined labels. In CBHIR, in addition to the corresponding label of the query, the top K similar patches are shown to the pathologists. This enhances the reliability of CBHIR for pathologists as it avoids being a completely black box. CBHIR gives pathologists an opportunity to compare the histological patterns of the query with similar patches in addition to knowing the corresponding label.

Table 3 reports the area under the curve (AUC) of the unsupervised CBHIR as a result of applying all CN techniques. These results were compared with the results of applying VGG19 on CAM17 as a classifier, which is reported in [36]. Following [46], the main objective of this comparison with the supervised classifier is to evaluate the unsupervised CBHIR performance in terms of retrieving images belonging to the same cancer grade [46]. According to Table 3, BKSVD-VGG19, with a 98.17% AUC, had the highest performance among (Vah, Mac, Original)-VGG19. The unsupervised BKSVD-CBHIR method, with a 96.31% AUC, has a comparable performance to the fully supervised method.

Table 3. Comparison of the performance of VGG19 as a classifier and the unsupervised CBHIR regarding the top 5. The reported metric is AUC.

	Original	BKSVD	Vah [33]	Mac [32]
VGG19 (supervised)	0.941	0.9817	0.7985	0.9499
CBHIR (unsupervised)	0.9631	0.9754	0.9429	0.9632

6. Conclusions

In this paper, we have proposed a novel CBHIR framework for histopathological images based on an unsupervised feature extractor and color normalization. We utilized a custom-built convolutional auto encoder (CAE) to extract the features in an unsupervised manner to tackle the challenges of a lack of annotated data sets. In this feature extractor, a skip connection between a layer in the encoder, its corresponding layer in the decoder, and a residual block in the bottleneck provided the meaningful features of the data set for the search engine.

The proposed framework is designed to work with data sets with intra- or inter-laboratory color variations since it solves the problem of the dependency of CBHIR on the color variation of the data set. We analyzed the effect of using color-normalized images to feed a CBHIR tool. We observed that as the effectiveness of color normalization techniques in reducing intra-center variance improved, the CBHIR results were better.

In this paper, we provided a visual evaluation in order to illustrate the results of the proposed framework visually. With this type of assessment, users can visually evaluate the quality of the retrieved patches, which might help identify potential areas for improvement. Furthermore, reporting these figures can be valuable for educational purposes, enabling users to grasp the functionality and the potential applications. It also facilitates communication between developers, researchers, and end-users, fostering collaboration and improving CBMIR technologies.

Finally, we compared the results of the proposed unsupervised CBHIR framework with VGG19 classifiers to evaluate the performance of the proposed unsupervised CBHIR

framework in order to retrieve images of the same cancer type. The proposed framework was found to be highly effective in discriminating between the grades of tissues. This observation clarifies the success of the proposed unsupervised CBHIR framework in identifying the correct histopathological features in images.

7. Future Work

CBMIR, a recent framework in digital pathology linked to CAD, plays a vital role in reducing the incidence of human errors and provides an inclusive worldwide platform for pathologists with varying levels of expertise. Acting as a bridge between medicine and engineering, CBMIR fills the gaps between these two fields, offering future opportunities in both realms. From an engineering perspective, despite achieving a high predictive accuracy, this method has its limitations. DL-based FEs should be trained on an extensive database, prompting the need for future work with more data using a prospective approach.

On the medical front, the entire framework can be tested in various hospitals and integrated into traditional cancer diagnosis workflows to analyze its pros and cons in real-world practice. These tests can provide a clearer understanding of the CBMIR framework and serve as a guide for subsequent steps in training DL-based methods.

Author Contributions: Z.T.: Conceptualization, Methodology, Analyzing, Writing—Reviewing and Editing, Formal Analysis, Visualization; F.P.B.: Review and Editing; A.C.: Supervision and Review; J.O.M.: Supervision; R.M.: Supervision and Review; V.N.: Supervision. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: The paper has been applied on a public data set.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AE	Auto Encoder
AUC	Area Under the Curve
BKSVD	Bayesian K-Singular Value Decomposition
BOF	Bag Of Features
CAD	Computer-Aided Diagnosis
CAE	Convolutional Auto Encoder
CAM17	CAMELYON17 challenge
CBHIR	Content-Based Histopathological Image Retrieval
CBMIR	Content-Based Medical Image Retrieval
CN	Color Normalization
CNN	Convolutional Neural Network
DCGMM	Deep Convolutional Gaussian Mixture Model
DL	Deep Learning
FE	Feature Extractor
GAN	Generative Adversarial Network

H&E	Hematoxylin and Eosin
MRI	Magnetic Resonance Imaging
Mac	Macenko
NMI	Normalized Median Intensity
NMI SD	Normalized Median Intensity Standard Deviation
NMI CV	Normalized Median Intensity Coefficient of Variation
PSNR	Peak Signal to Noise Ratio
ROI	Regions Of Interest
SVD	Singular Value Decomposition
UFL	Unsupervised Features Learning
VAE	Variational Auto Encoder
Vah	Vahadane
WSIs	Whole-Slide Images

References

1. Ibrahim, A.; Gamble, P.; Jaroensri, R.; Abdelsamea, M.M.; Mermel, C.H.; Chen, P.H.C.; Rakha, E.A. Artificial intelligence in digital breast pathology: Techniques and applications. *Breast* **2020**, *49*, 267–273. [CrossRef]
2. Al-Hussaeni, K.; Karamitsos, I.; Adewumi, E.; Amawi, R.M. CNN-Based Pill Image Recognition for Retrieval Systems. *Appl. Sci.* **2023**, *13*, 5050. [CrossRef]
3. Khalil, S.; Nawaz, U.; Zubariah; Mushtaq, Z.; Arif, S.; ur Rehman, M.Z.; Qureshi, M.F.; Malik, A.; Aleid, A.; Alhussaini, K. Enhancing Ductal Carcinoma Classification Using Transfer Learning with 3D U-Net Models in Breast Cancer Imaging. *Appl. Sci.* **2023**, *13*, 4255. [CrossRef]
4. Fuster, S.; Khoraminia, F.; Eftestøl, T.; Zuiverloon, T.C.; Engan, K. Active Learning Based Domain Adaptation for Tissue Segmentation of Histopathological Images. In Proceedings of the 2023 31st European Signal Processing Conference (EUSIPCO), Helsinki, Finland, 4–8 September 2023; pp. 1045–1049. [CrossRef]
5. Shahdoosti, H.R.; Mehrabi, A. Multimodal image fusion using sparse representation classification in tetrolet domain. *Digit. Signal Process.* **2018**, *79*, 9–22. [CrossRef]
6. Alrowais, F.; Alotaibi, F.A.; Hassan, A.Q.; Marzouk, R.; Alnfiai, M.M.; Sayed, A. Enhanced Pelican Optimization Algorithm with Deep Learning-Driven Mitotic Nuclei Classification on Breast Histopathology Images. *Biomimetics* **2023**, *8*, 538. [CrossRef]
7. Hegde, N.; Hipp, J.D.; Liu, Y.; Emmert-Buck, M.; Reif, E.; Smilkov, D.; Terry, M.; Cai, C.J.; Amin, M.B.; Mermel, C.H.; et al. Similar image search for histopathology: SMILY. *NPJ Digit. Med.* **2019**, *2*, 56. [CrossRef] [PubMed]
8. Strittmatter, A.; Caroli, A.; Zöllner, F.G. A Multistage Rigid-Affine-Deformable Network for Three-Dimensional Multimodal Medical Image Registration. *Appl. Sci.* **2023**, *13*, 13298. [CrossRef]
9. Kanwal, N.; Fuster, S.; Khoraminia, F.; Zuiverloon, T.C.; Rong, C.; Engan, K. Quantifying the effect of color processing on blood and damaged tissue detection in whole slide images. In Proceedings of the 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Nafplio, Greece, 26–29 June 2022; pp. 1–5.
10. Shahdoosti, H.R.; Tabatabaei, Z. MRI and PET/SPECT image fusion at feature level using ant colony based segmentation. *Biomed. Signal Process. Control* **2019**, *47*, 63–74. [CrossRef]
11. Long, F.; Zhang, H.; Feng, D.D. Fundamentals of content-based image retrieval. In *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 1–26.
12. Tabatabaei, Z.; Colomer, A.; Moll, J.O.; Naranjo, V. Siamese Content-based Search Engine for a More Transparent Skin and Breast Cancer Diagnosis through Histological Imaging. *arXiv* **2024**, arXiv:2401.08272.
13. Smeulders, A.W.; Worring, M.; Santini, S.; Gupta, A.; Jain, R. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1349–1380. [CrossRef]
14. Qi, X.; Wang, D.; Rodero, I.; Diaz-Montes, J.; Gensure, R.H.; Xing, F.; Zhong, H.; Goodell, L.; Parashar, M.; Foran, D.J.; et al. Content-based histopathology image retrieval using CometCloud. *BMC Bioinform.* **2014**, *15*, 287. [CrossRef]
15. Tabatabaei, Z.; Colomer, A.; Engan, K.; Oliver, J.; Naranjo, V. Residual block convolutional auto encoder in content-based medical image retrieval. In Proceedings of the 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Nafplio, Greece, 26–29 June 2022; pp. 1–5.
16. Bianconi, F.; Kather, J.N.; Reyes-Aldasoro, C.C. Experimental assessment of color deconvolution and color normalization for automated classification of histology images stained with hematoxylin and eosin. *Cancers* **2020**, *12*, 3337. [CrossRef]
17. Tabatabaei, Z.; Wang, Y.; Colomer, A.; Oliver Moll, J.; Zhao, Z.; Naranjo, V. WWFedCBMIR: World-Wide Federated Content-Based Medical Image Retrieval. *Bioengineering* **2023**, *10*, 1144. [CrossRef]
18. Fuster, S.; Khoraminia, F.; Kiraz, U.; Kanwal, N.; Kvikstad, V.; Eftestøl, T.; Zuiverloon, T.C.; Janssen, E.A.; Engan, K. Invasive cancerous area detection in Non-Muscle invasive bladder cancer whole slide images. In Proceedings of the 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Nafplio, Greece, 26–29 June 2022; pp. 1–5.
19. Tosta, T.A.A.; de Faria, P.R.; Neves, L.A.; do Nascimento, M.Z. Computational normalization of H&E-stained histological images: Progress, challenges and future potential. *Artif. Intell. Med.* **2019**, *95*, 118–132.

20. Bandi, P.; Geessink, O.; Manson, Q.; Van Dijk, M.; Balkenhol, M.; Hermsen, M.; Bejnordi, B.E.; Lee, B.; Paeng, K.; Zhong, A.; et al. From detection of individual metastases to classification of lymph node status at the patient level: The camelyon17 challenge. *IEEE Trans. Med. Imaging* **2018**, *38*, 550–560. [CrossRef] [PubMed]

21. Salvi, M.; Acharya, U.R.; Molinari, F.; Meiburger, K.M. The impact of pre-and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Comput. Biol. Med.* **2021**, *128*, 104129. [CrossRef]

22. Vijh, S.; Saraswat, M.; Kumar, S. A new complete color normalization method for H&E stained histopatholgical images. *Appl. Intell.* **2021**, *51*, 7735–7748.

23. Roy, S.; kumar Jain, A.; Lal, S.; Kini, J. A study about color normalization methods for histopathology images. *Micron* **2018**, *114*, 42–61. [CrossRef]

24. Ionescu, B.; Müller, H.; Drăgulinescu, A.M.; Popescu, A.; Idrissi-Yaghir, A.; García Seco de Herrera, A.; Andrei, A.; Stan, A.; Storås, A.M.; Abacha, A.B.; et al. ImageCLEF 2023 Highlight: Multimedia Retrieval in Medical, Social Media and Content Recommendation Applications. In Proceedings of the European Conference on Information Retrieval, Dublin, Ireland, 2–6 April 2023; pp. 557–567.

25. Zheng, Y.; Jiang, Z.; Zhang, H.; Xie, F.; Ma, Y.; Shi, H.; Zhao, Y. Size-scalable content-based histopathological image retrieval from database that consists of WSIs. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 1278–1287. [CrossRef]

26. Zhang, X.; Liu, W.; Dundar, M.; Badve, S.; Zhang, S. Towards large-scale histopathological image analysis: Hashing-based image retrieval. *IEEE Trans. Med. Imaging* **2014**, *34*, 496–506. [CrossRef]

27. Vanegas, J.A.; Arevalo, J.; González, F.A. Unsupervised feature learning for content-based histopathology image retrieval. In Proceedings of the 2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI), Klagenfurt, Austria, 18–20 June 2014; pp. 1–6.

28. Sukhia, K.N.; Riaz, M.M.; Ghafoor, A.; Ali, S.S.; Iltaf, N. Content-based histopathological image retrieval using multi-scale and multichannel decoder based LTP. *Biomed. Signal Process. Control* **2019**, *54*, 101582. [CrossRef]

29. Riasatian, A.; Babaie, M.; Maleki, D.; Kalra, S.; Valipour, M.; Hemati, S.; Zaveri, M.; Safarpoor, A.; Shafiei, S.; Afshari, M.; et al. Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides. *Med. Image Anal.* **2021**, *70*, 102032. [CrossRef]

30. Silva-Rodríguez, J.; Colomer, A.; Sales, M.A.; Molina, R.; Naranjo, V. Going deeper through the Gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection. *Comput. Methods Programs Biomed.* **2020**, *195*, 105637. [CrossRef]

31. Shaban, M.T.; Baur, C.; Navab, N.; Albarqouni, S. Staingan: Stain style transfer for digital histological images. In Proceedings of the 2019 Ieee 16th international symposium on biomedical imaging (Isbi 2019), Venice, Italy, 8–11 April 2019; pp. 953–956.

32. Macenko, M.; Niethammer, M.; Marron, J.S.; Borland, D.; Woosley, J.T.; Guan, X.; Schmitt, C.; Thomas, N.E. A method for normalizing histology slides for quantitative analysis. In Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, USA, 28 June–1 July 2009; pp. 1107–1110.

33. Vahadane, A.; Peng, T.; Sethi, A.; Albarqouni, S.; Wang, L.; Baust, M.; Steiger, K.; Schlitter, A.M.; Esposito, I.; Navab, N. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Trans. Med. Imaging* **2016**, *35*, 1962–1971. [CrossRef]

34. BenTaieb, A.; Hamarneh, G. Adversarial stain transfer for histopathology image analysis. *IEEE Trans. Med. Imaging* **2017**, *37*, 792–802. [CrossRef]

35. Zanjani, F.G.; Zinger, S.; Bejnordi, B.E.; van der Laak, J.A.; de With, P.H. Stain normalization of histopathology images using generative adversarial networks. In Proceedings of the 2018 IEEE 15th International symposium on biomedical imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 573–577.

36. Pérez-Bueno, F.; Serra, J.G.; Vega, M.; Mateos, J.; Molina, R.; Katsaggelos, A.K. Bayesian K-SVD for H and E blind color deconvolution. Applications to stain normalization, data augmentation and cancer classification. *Comput. Med. Imaging Graph.* **2022**, *97*, 102048. [CrossRef]

37. Tellez, D.; Litjens, G.; Bándi, P.; Bulten, W.; Bokhorst, J.M.; Ciompi, F.; Van Der Laak, J. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med. Image Anal.* **2019**, *58*, 101544. [CrossRef] [PubMed]

38. Chen, M.; Shi, X.; Zhang, Y.; Wu, D.; Guizani, M. Deep feature learning for medical image analysis with convolutional autoencoder neural network. *IEEE Trans. Big Data* **2017**, *7*, 750–758. [CrossRef]

39. Ahn, E.; Kumar, A.; Fulham, M.; Feng, D.; Kim, J. Unsupervised domain adaptation to classify medical images using zero-bias convolutional auto-encoders and context-based feature augmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 2385–2394. [CrossRef] [PubMed]

40. Park, S.; Gach, H.M.; Kim, S.; Lee, S.J.; Motai, Y. Autoencoder-inspired convolutional network-based super-resolution method in MRI. *IEEE J. Transl. Eng. Health Med.* **2021**, *9*, 1800113. [CrossRef] [PubMed]

41. Daoud, M.I.; Saleh, A.; Hababeh, I.; Alazrai, R. Content-based Image Retrieval for Breast Ultrasound Images using Convolutional Autoencoders: A Feasibility Study. In Proceedings of the 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART), Paris, France, 24–26 April 2019; pp. 1–4. [CrossRef]

42. Tabatabaei, Z.; Colomer, A.; Engan, K.; Oliver, J.; Naranjo, V. Self-supervised learning of a tailored Convolutional Auto Encoder for histopathological prostate grading. In Proceedings of the 2023 31st European Signal Processing Conference (EUSIPCO), Helsinki, Finland, 4–8 September 2023; pp. 980–984.
43. Zheng, Y.; Jiang, Z.; Zhang, H.; Xie, F.; Shi, J.; Xue, C. Adaptive color deconvolution for histological WSI normalization. *Comput. Methods Programs Biomed.* **2019**, *170*, 107–120. [CrossRef] [PubMed]
44. Kalra, S.; Tizhoosh, H.R.; Choi, C.; Shah, S.; Diamandis, P.; Campbell, C.J.; Pantanowitz, L. Yottixel–an image search engine for large archives of histopathology whole slide images. *Med. Image Anal.* **2020**, *65*, 101757. [CrossRef]
45. Cao, Z.; Long, M.; Wang, J.; Yu, P.S. Hashnet: Deep learning to hash by continuation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5608–5617.
46. Tabatabaei, Z.; Colomer, A.; Moll, J.O.; Naranjo, V. Toward More Transparent and Accurate Cancer Diagnosis With an Unsupervised CAE Approach. *IEEE Access* **2023**, *11*, 143387–143401. [CrossRef]