# EndoFinder: Online Image Retrieval for Explainable Colorectal Polyp Diagnosis

Ruijie Yang[1,2,3]†, Yan Zhu[4,5]†, Peiyao Fu[4,5], Yizhe Zhang[6], Zhihua Wang[3]✉, Quanlin Li[4,5], Pinghong Zhou[4,5], Xian Yang[7], and Shuo Wang[1,2]✉

[1]Digital Medical Research Center, School of Basic Medical Sciences, Fudan University, Shanghai, China
[2]Shanghai Key Laboratory of MICCAI, Shanghai, China
[3]Shanghai Institute for Advanced Study of Zhejiang University, Shanghai, China
[4]Endoscopy Center and Endoscopy Research Institute, Zhongshan Hospital, Fudan University, Shanghai, China
[5]Shanghai Collaborative Innovation Center of Endoscopy, Shanghai, China
[6]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, China
[7]Alliance Manchester Business School, The University of Manchester, Manchester, UK

**Abstract.** Determining the necessity of resecting malignant polyps during colonoscopy screen is crucial for patient outcomes, yet challenging due to the time-consuming and costly nature of histopathology examination. While deep learning-based classification models have shown promise in achieving optical biopsy with endoscopic images, they often suffer from a lack of explainability. To overcome this limitation, we introduce EndoFinder, a content-based image retrieval framework to find the 'digital twin' polyp in the reference database given a newly detected polyp. The clinical semantics of the new polyp can be inferred referring to the matched ones. EndoFinder pioneers a polyp-aware image encoder that is pre-trained on a large polyp dataset in a self-supervised way, merging masked image modeling with contrastive learning. This results in a generic embedding space ready for different downstream clinical tasks based on image retrieval. We validate the framework on polyp re-identification and optical biopsy tasks, with extensive experiments demonstrating that EndoFinder not only achieves explainable diagnostics but also matches the performance of supervised classification models. EndoFinder's reliance on image retrieval has the potential to support diverse downstream decision-making tasks during real-time colonoscopy procedures.

**Keywords:** Polyp diagnosis · Content-based image retrieval · Semantic hashing.

† Equal contribution.
✉ Corresponding authors: shuowang@fudan.edu.cn and zhihua.wang@zju.edu.cn

## 1 Introduction

Colorectal cancer (CRC) presents a major public health challenge, accounting for approximately 10% of all cancer incidences worldwide and ranking as the second leading cause of cancer-related deaths [1–3]. Colonoscopy stands as the cornerstone for CRC prevention and early detection, primarily through the identification and subsequent management of polyps. During these procedures, clinical endoscopists face critical decisions on whether to remove potentially malignant polyps or opt for active surveillance of benign ones [24]. While the histopathological analysis of biopsied samples serves as the definitive diagnostic method, it is not immediately available during endoscopic examinations. Consequently, clinicians often rely on optical diagnosis through endoscopic imagery for on-the-spot decision-making regarding small colorectal polyps. Artificial intelligence (AI)-based optical diagnosis of polyps has been developed for augmented decision-making during colonoscopy procedures [29]. However, the predominant AI models, characterized by their supervised learning and "black box" nature, suffer from a lack of interpretability. These inductive models demand extensive annotated image datasets for training and need to be re-trained as new data and annotation are acquired, posing significant challenges for scalability and continuous clinical application.

To mitigate the limitations of existing classifiers, we present EndoFinder (Figure 1), an image retrieval framework enhancing diagnostic explainability for colorectal polyps. Inspired by the 'digital twin' concept, EndoFinder identifies a matching 'digital twin' for new polyps in a reference database containing historical data on similar polyps. This approach facilitates interpretable and informed decision-making by leveraging past diagnostic outcomes, offering a scalable solution for real-time polyp diagnosis.

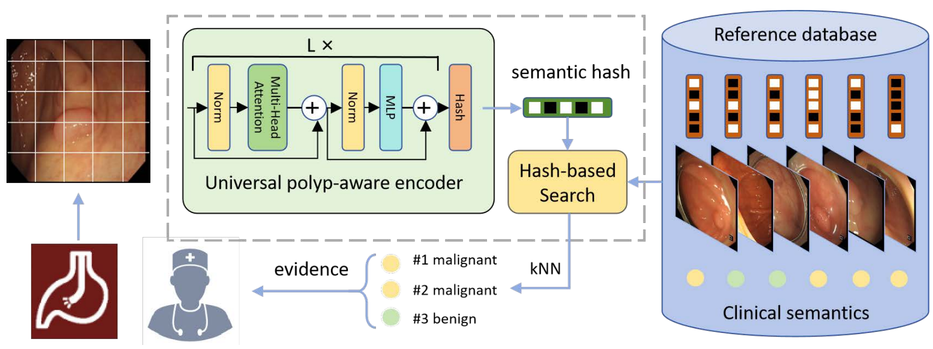


**Fig. 1.** Workflow of the proposed EndoFinder framework. Endoscopic images are encoded into polyp-aware semantic features and discretised into hash codes for fast retrieval. The decision-making is augmented by referring to the historical information of the 'digital twin' polyp in the database.

**Related work.** Here we review the state-of-the-art performance of optical polyp diagnosis and the medical application of Content-Based Image Retrieval (CBIR). **Supervised polyp diagnosis.** Supervised classifiers, particularly those based on deep learning, have matched the expertise of professional endoscopists in optical polyp diagnosis. Ribeiro et al. were pioneers in employing convolutional neural networks (CNNs) for classifying colorectal polyps. Chen et al. developed a system of computer-aided diagnosis system utilizing an Inception v3 architecture to process narrow-band imagery of small colorectal polyps, achieving near-novice doctor accuracy at greater inference speeds. [32]. Yamada et al. developed an AI system based on ResNet152, outperforming expert endoscopists in both internal and external validation [29]. Recently, Krenzer et al. achieved leading accuracy by implementing a method that involves detecting and cropping polyps before classifying them using a Vision Transformer (ViT) [34]. Despite the satisfactory performance of these varied architectural approaches, their clinical applicability is hampered by issues such as limited explainability and vulnerability to data's long-tail distribution.

**Content-based image retrieval for medical image analysis.** Unlike inductive methods that derive general rules from the training set, content-based image retrieval presents a transductive alternative to medical image analysis [20, 23]. Wang et al. pioneered a CBIR system that facilitates the retrieval of pertinent whole-slide images from vast historical databases [9]. Intrator et al. employed the contrastive learning method SimCLR for polyp representation, advancing polyp video re-identification capabilities [31]. A crucial aspect of CBIR involves constructing an effective embedding space and developing efficient search algorithms for identifying nearest neighbors. For natural images, the focus is increasingly shifting towards learning general and robust representations through self-supervised learning (SSL) on extensive datasets. Pizzi et al. enhanced image copy detection by training CNNs using contrastive learning and score normalization to achieve high-quality embeddings [12]. Similarly, El-Nouby et al. harnessed Vision Transformer (ViT) networks, integrating InfoNCE with entropy regularizers for improved learning outcomes [7]. To expedite search speeds, Guan et al. devised a method for training CNNs with attention maps to generate semantic hash codes, enabling rapid image retrieval [8]. However, it remains less explored to construct a universal representation for polyp image retrieval.

**Contributions.** Our contributions are threefold: Firstly, we propose a novel adaptive self-supervised learning method that merges masked image modeling with contrastive learning to create universal polyp-aware representations, significantly improving the precision of polyp re-identification. Secondly, we introduce an image retrieval approach for explainable polyp diagnosis achieving SOTA performance compared to supervised classifiers. Lastly, we developed a hashing technique to realize real-time image retrieval without accuracy loss.

## 2 Methods

### 2.1 Problem Formulation

Let us denote a task-specific collection of a reference database with clinical semantics as $S = \{(I_i, y_i)\}_{i=1}^{N}$, where $I_i$ is the image of the $i$-th polyp with clinical categories $y_i \in \{1, 2, \ldots, C\}$ and $N$ is the database size. The task is to infer the clinical label $y_i$ given an image $I_i$ of a newly detected polyp. In general, a supervised classifier uses the reference database to learn the mapping $f_\theta$ parameterized by $\theta$ such that $y_i = f_\theta(I_i)$. Although this method facilitates an end-to-end diagnostic process, it falls short in terms of explainability.

Drawing inspiration from the K-Nearest Neighbors (KNN) algorithm, the proposed EndoFinder framework builds on the hypothesis that *polyps in close proximity within the embedding space are likely to share similar clinical semantics.* EndoFinder identifies a set of 'digital twins' from the reference database given a test polyp image, leveraging the clinical semantics of the 'digital twins' for transductive reasoning. Formally, the clinical label of a test image can be determined by

$$y_i = \underset{c \in \{1,\ldots,C\}}{\text{argmax}} \sum_{k \in \mathcal{N}(I_i)} \mathbf{1}_{\{y_k = c\}}. \tag{1}$$

where $\mathcal{N}(I_i)$ denotes the set of indices corresponding to the K nearest neighbors of the query image $I_i$. Here, $\mathbf{1}_{\{y_k = c\}}$ is the indicator function whether the class label $y_k$ of the $k$th nearest neighbors is equal to the class $c$.

### 2.2 Overview of the EndoFinder design

The core of EndoFinder is to construct a plausible embedding space for polyp image retrieval, denoted as $z = E_\phi(I)$, where $E_\phi$ represents the feature extractor. Our approach involves learning a universal representation from extensive polyp image datasets through self-supervised learning (SSL) (Figure 2) and subsequently converting this representation into semantic hash codes to enable rapid retrieval.

**Universal polyp-aware image encoder:** Drawing inspiration from the effectiveness of masked autoencoder (MAE) and contrastive learning approaches, we integrate these two SSL techniques to pre-train a ViT encoder.

On one hand, the image encoder is trained under the MAE framework to reconstruct masked image patches from the embedding features. In particular, we introduce an adaptive masking strategy that leverages the available polyp segmentation masks. This is realized by masking a larger proportion of background patches compared to foreground patches inversely proportional to the ratio of pixels within the segmentation mask (supplementary material), enabling the encoder to focus on the most informative regions of the image and to generate so-called polyp-aware representation. The MAE reconstruction loss for a batch of N images is the mean square error between the reconstructed image and the

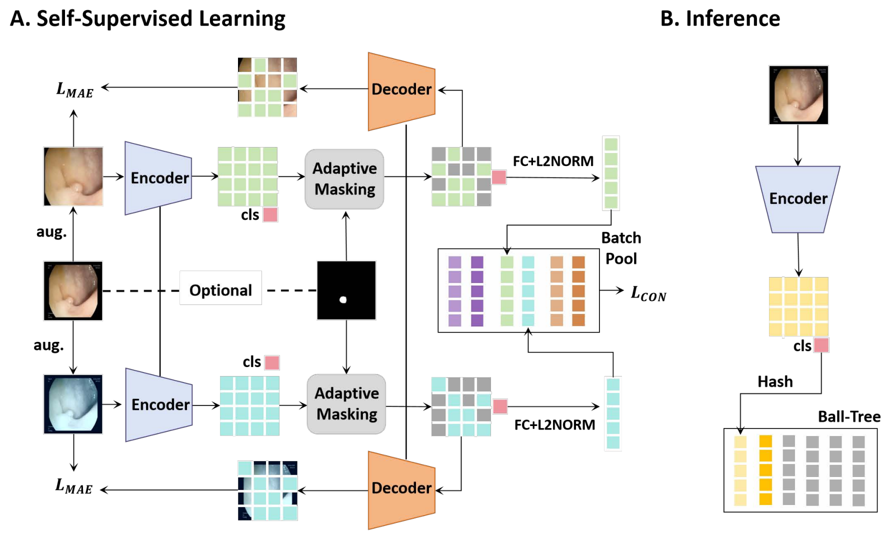


**Fig. 2.** Polyp-aware self-supervised representation learning and inference.

original image, focusing solely on the masked regions:

$$L_{MAE} = \frac{1}{2N} \sum_{i=1}^{2N} \frac{1}{|M_i|} \sum_{k \in M_i} (\hat{I}_{i,k} - I_{i,k})^2. \tag{2}$$

where $M_i$ is a set of non-zero pixels in the masked image $i$, $\hat{I}_{i,k}$ and $I_{i,k}$ refer to the pixel $k$ in the reconstructed and original image $i$, respectively. For a set of $N$ images, we generated $2N$ transformed images through repeated augmentations.

On the other hand, the class token (CLS) from the MAE encoder is subject to a linear projection and L2 normalization, resulting in the embedding feature $z_i \in R^d$. At this stage, contrastive learning is applied, leveraging InfoNCE and Entropy loss to evaluate the distance between augmented images of samples [12]. The positive pairs of matching images are $P = \{(i, i+N), (i+N, i)\}_{i \in \{1,\ldots,N\}}$. We denote positive matches for image $i$ as $P_i = \{j | (i,j) \in P\}$. The contrastive InfoNCE loss maximizes the similarity between copies relative to the similarity of non-copies. Entropy loss will push away the nearest neighbor who does not belong to the positive pair. The temperature-adjusted cosine similarity $s_{i,j}$ is computed between the feature embeddings $z_i$ and $z_j$. The loss $L_{CON}$ of contrastive learning is the weighted sum of the infoNCE (first term) and entropy loss (second term), with entropy loss weighted by hyper-parameter $\gamma$:

$$L_{CON} = -\frac{1}{|P|} \sum_{(i,j) \in P} \log \frac{\exp(s_{i,j})}{\sum_{v \neq i} \exp(s_{i,v})} + \gamma \left( -\frac{1}{N} \sum_{i=1}^{N} \log(\min_{j \notin \hat{P}_i} ||z_i - z_j||) \right) \tag{3}$$

where $\hat{P}_i = P_i \cup \{i\}$. The overall loss $L$ is a weighted sum of the aforementioned components, with MAE loss modulated by its weight parameter $\lambda$:

$$L = L_{CON} + \lambda L_{MAE}. \tag{4}$$

**Semantic hashing for image retrieval:** To accelerate the image retrieval speed, we transform the features into hash codes through a hashing layer.

The quantization process is defined by:

$$\bar{z}_{i,k} = \begin{cases} 1 & \text{if } z_{i,k} \geq 0, \\ -1 & \text{if } z_{i,k} < 0. \end{cases} \tag{5}$$

This function assigns a binary code of 1 if the feature value $z_{i,k}$ for image $i$ pixel $k$ is non-negative and $-1$ if $z_{i,k}$ is negative. Upon obtaining the binary codes, the next step involves retrieving the images most similar to the query image. Using binary codes for constructing a ball tree retrieval system significantly boosts retrieval speed [36]. The retrieval process is based on the similarity of these binary codes to those of the reference images. Once the most similar images are retrieved, a voting mechanism is employed to determine the category of the query image. This mechanism takes into account the categories of the $k$-nearest reference images, thereby leveraging the collective information of the retrieved set for accurate image categorization.

## 3 Experiments and Results

We first train the image encoder on Polyp-18k and then test the utility of EndoFinder in polyp re-identification and optical polyp diagnosis on Polyp-Twin and Polyp-Path, respectively. It is noted that polyps in the datasets do not overlap. We implement two versions of EndoFinder using hashed features (EndoFinder-Hash) or raw features (EndoFinder-Raw). The implementation details and hyper-parameter studies can be found in Supplementary Material.

### 3.1 Datasets

**Polyp-18k:** An in-house dataset of 17,969 polyp images with corresponding polyp segmentation masks for the training of the image encoder of EndoFinder.
**Polyp-Twin:** A curated set of 200 images representing various angles of 100 distinct polyps (two images for each polyp) from colonoscopy video recordings.
**Polyp-Path:** A dataset of of 147 images with pathological classification [37]. 57% are malignant and 43% are benign according.

### 3.2 Polyp Re-Identification

The first task is to retrieve the other paired image of the polyp given one polyp from the Polyp-Twin. We compared our methods to ImageNet pre-trained feature extractors or SSL methods (MAE [11], ViT-SimCLR and CNN-SimCLR[12])

**Table 1.** Comparison of Polyp Re-identification Performance.

| | | uAP | Acc@1 | Recall@90% | time(s) | FPS |
|---|---|---|---|---|---|---|
| ImageNet Features | Resnet50 | 0.365 | 0.495 | 0.128 | 0.485 | 2.06 |
| | VGG19 | 0.338 | 0.564 | 0.118 | 0.545 | 1.83 |
| | Densenet121 | 0.377 | 0.514 | 0.128 | 0.471 | 2.12 |
| | ViT-L16 | 0.243 | 0.386 | 0.059 | 0.451 | 2.21 |
| | SSCD | 0.581 | 0.673 | 0.326 | 0.434 | 2.30 |
| SSL on Polyp-18k | MAE | 0.470 | 0.554 | 0.227 | 0.453 | 2.22 |
| | ViT-SimCLR | 0.591 | 0.623 | 0.415 | 0.454 | 2.18 |
| | CNN-SimCLR | 0.672 | **0.693** | 0.495 | 0.433 | 2.30 |
| | EndoFinder-Raw | **0.695** | **0.693** | 0.495 | 0.456 | 2.19 |
| | EndoFinder-Hash | 0.693 | **0.693** | **0.524** | **0.009** | **108.57** |

pre-trained on Polyp-18k. As evidenced in Table 1, our model surpasses other models across all metrics.

Furthermore, We evaluated the speed enhancement achieved using binary codes for image retrieval on a dataset with over 12000 images, as shown in Table 1. The use of binary codes to construct a ball tree retrieval system significantly enhances retrieval speed. Fig. 3 illustrates a comparative analysis of retrieval outcomes using different feature extractors.

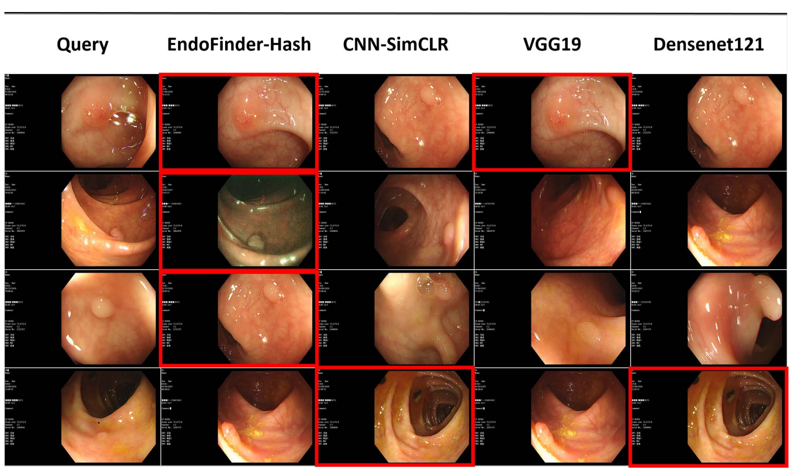


**Fig. 3.** Examples of polyp re-identification results. Each row depicts a polyp, showing the query image followed by the first retrieval results from EndoFinder, pre-trained SSCD, VGG19 and Densenet121, respectively. Correct retrievals are bounded in red.

### 3.3 Optical Polyp Diagnosis

After validating the performance of our universal polyp-aware representation, we evaluated the proposed image retrieval-based classification in a more clinically relevant task - determining the pathological malignancy on the Polyp-Path dataset. The outcomes of EndoFinder are illustrated in Fig.4, demonstrating the model's effectiveness. We compared the performance of image retrieval-based classification using different feature embeddings with supervised classifiers fine-tuned on Polyp-Path with ImageNet pre-trained weights. The performance was evaluated using 5-fold cross-validation, where 4 folds were used as the reference database and the remaining fold was used for testing. The average results are shown in Table 2.

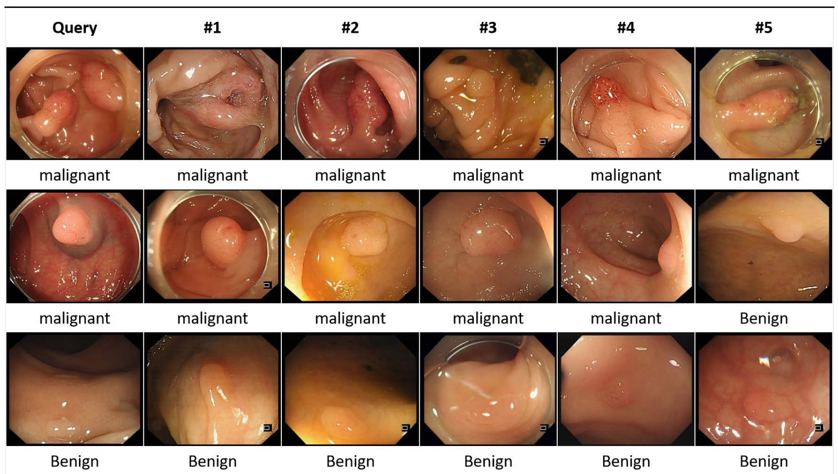


**Fig. 4.** Examples of image-retrieval based classification by EndoFinder.

**Table 2.** Comparison of optical polyp diagnosis performance.

| | | ACC | SEN | SPE | F1 |
|---|---|---|---|---|---|
| Supervised classifier | Resnet50 | 74.482 | 79.095 | 68.988 | 77.880 |
| | VGG19 | 76.550 | 77.954 | 76.259 | 78.848 |
| | Densenet121 | 75.864 | 79.212 | 71.082 | 79.062 |
| | ViT-L16 | 75.862 | 74.286 | **77.362** | 78.34 |
| Retrieval using ImageNet features | Resnet50 | 66.896 | 78.910 | 53.090 | 73.055 |
| | VGG19 | 68.275 | 71.203 | 66.320 | 71.719 |
| | Densenet121 | 73.793 | 80.815 | 64.796 | 77.562 |
| | ViT-L16 | 68.965 | 76.641 | 59.073 | 73.816 |
| Retrival using SSL features | MAE | 66.896 | 70.182 | 65.437 | 70.939 |
| | ViT-SimCLR | 67.586 | 70.753 | 62.588 | 71.593 |
| | EndoFinder-Raw | **77.241** | 81.239 | 73.748 | **80.445** |
| | EndoFinder-Hash | 73.793 | **81.916** | 63.922 | 78.213 |

## 4 Discussion and Conclusion

By combining advanced SSL techniques, EndoFinder has achieved outstanding performance in polyp image retrieval and pathological classification. Our experimental findings highlight EndoFinder's proficiency in identifying polyp-specific features, as demonstrated by its superior accuracy and F1 scores compared to traditional classification models. Image retrieval performance using EndoFinder features outperforms that of features pre-trained solely through MAE or contrastive learning techniques. This superiority highlights the effectiveness of the adaptive masking strategy and the synergistic benefits of combining SSL techniques. It should be noted that the EndoFinder features were not fine-tuned on the downstream classification task, demonstrating the power of universal representation learned from large datasets in a self-supervised manner. The polyp-aware semantic hash could serve as a unique identification (UID) to be explored in future studies. By employing hashing-based retrieval methods, EndoFinder ensures scalability to extensive reference datasets. Beyond merely enhancing optical polyp diagnosis performance, EndoFinder has the potential to facilitate various decision-making processes, such as determining the optimal approach for polyp removal by searching and matching similar cases in historical records.

In conclusion, the EndoFinder framework establishes a universal representation for endoscopic images and delivers exceptional performance in real-time polyp diagnosis, complete with explainability.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

**Acknowledgement.** This study was supported in part by the Shanghai Sailing Program (22YF1409300), International Science and Technology Cooperation Program under the 2023 Shanghai Action Plan for Science (23410710400) and National Natural Science Foundation of China (No.62201263).

## References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022; 72: 7-33. https://doi.org/10.3322/caac.21708
2. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020; 70: 145-164. https://doi.org/10.3322/caac.21601
3. Biller LH, Schrag D. Diagnosis and treatment of metastatic colorectal cancer: a review. JAMA. 2021; 325: 669-685. https://doi.org/10.1001/jama.2021.0106
4. Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//International conference on machine learning. PMLR, 2020: 1597-1607.
5. Oord A, Li Y, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv preprint arXiv:1807.03748, 2018.
6. Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.
7. El-Nouby A, Neverova N, Laptev I, et al. Training vision transformers for image retrieval[J]. arXiv preprint arXiv:2102.05644, 2021.

8. Guan A, Liu L, Fu X, et al. Precision medical image hash retrieval by interpretability and feature fusion[J]. Computer Methods and Programs in Biomedicine, 2022, 222: 106945.
9. Wang X, Du Y, Yang S, et al. RetCCL: clustering-guided contrastive learning for whole-slide image retrieval[J]. Medical image analysis, 2023, 83: 102645.
10. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
11. He K, Chen X, Xie S, et al. Masked autoencoders are scalable vision learners[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 16000-16009.
12. Pizzi, E., Roy, S. D., Ravindra, S. N., Goyal, P., & Douze, M. (2022). A self-supervised descriptor for image copy detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14532-14542).
13. Hirsch R, Caron M, Cohen R, et al. Self-supervised learning for endoscopic video analysis[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 569-578.
14. Shen C, Zhang J, Liang X, et al. Forensic Histopathological Recognition via a Context-Aware MIL Network Powered by Self-supervised Contrastive Learning[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 528-538.
15. Grill J B, Strub F, Altché F, et al. Bootstrap your own latent-a new approach to self-supervised learning[J]. Advances in neural information processing systems, 2020, 33: 21271-21284.
16. Caron M, Misra I, Mairal J, et al. Unsupervised learning of visual features by contrasting cluster assignments[J]. Advances in neural information processing systems, 2020, 33: 9912-9924.
17. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
18. Wu Z, Xiong Y, Yu S X, et al. Unsupervised feature learning via non-parametric instance discrimination[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 3733-3742.
19. Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification[J]. arXiv preprint arXiv:1703.07737, 2017.
20. Öztürk Ş. Class-driven content-based medical image retrieval using hash codes of deep features[J]. Biomedical Signal Processing and Control, 2021, 68: 102601.
21. Liu C, Ma J, Tang X, et al. Deep hash learning for remote sensing image retrieval[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(4): 3420-3443.
22. Li T, Zhang Z, Pei L, et al. HashFormer: Vision transformer based deep hashing for image retrieval[J]. IEEE Signal Processing Letters, 2022, 29: 827-831.
23. Chen Y, Tang Y, Huang J, et al. Multi-scale Triplet Hashing for Medical Image Retrieval[J]. Computers in Biology and Medicine, 2023, 155: 106633.
24. Chandran S, Parker F, Lontos S, et al. Can we ease the financial burden of colonoscopy? Using real-time endoscopic assessment of polyp histology to predict surveillance intervals[J]. Internal medicine journal, 2015, 45(12): 1293-1299.
25. van den Broek F J C, Reitsma J B, Curvers W L, et al. Systematic review of narrow-band imaging for the detection and differentiation of neoplastic and nonneoplastic lesions in the colon (with videos)[J]. Gastrointestinal endoscopy, 2009, 69(1): 124-135.

26. Ladabaum U, Fioritto A, Mitani A, et al. Real-time optical biopsy of colon polyps with narrow band imaging in community practice does not yet meet key thresholds for clinical decisions[J]. Gastroenterology, 2013, 144(1): 81-91.
27. Togashi K, Osawa H, Koinuma K, et al. A comparison of conventional endoscopy, chromoendoscopy, and the optimal-band imaging system for the differentiation of neoplastic and non-neoplastic colonic polyps[J]. Gastrointestinal endoscopy, 2009, 69(3): 734-741.
28. Kuiper T, Marsman W A, Jansen J M, et al. Accuracy for optical diagnosis of small colorectal polyps in nonacademic settings[J]. Clinical Gastroenterology and Hepatology, 2012, 10(9): 1016-1020.
29. Yamada M, Shino R, Kondo H, et al. Robust automated prediction of the revised Vienna classification in colonoscopy using deep learning: development and initial external validation[J]. Journal of Gastroenterology, 2022, 57(11): 879-889.
30. Ribeiro E, Uhl A, Häfner M. Colonic polyp classification with convolutional neural networks[C]//2016 IEEE 29th international symposium on computer-based medical systems (CBMS). IEEE, 2016: 253-258.
31. Intrator Y, Aizenberg N, Livne A, et al. Self-Supervised Polyp Re-Identification in Colonoscopy[J]. arXiv preprint arXiv:2306.08591, 2023.
32. Chen P J, Lin M C, Lai M J, et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis[J]. Gastroenterology, 2018, 154(3): 568-575.
33. Björnsson B, Borrebaeck C, Elander N, et al. Digital twins to personalize medicine[J]. Genome medicine, 2020, 12: 1-4.
34. Krenzer A, Heil S, Fitting D, et al. Automated classification of polyps using deep learning architectures and few-shot learning[J]. BMC Medical Imaging, 2023, 23(1): 59.
35. Ribeiro E, Uhl A, Häfner M. Colonic polyp classification with convolutional neural networks[C]//2016 IEEE 29th international symposium on computer-based medical systems (CBMS). IEEE, 2016: 253-258.
36. Brearley B J, Bose K R, Senthil K, et al. Knn Approaches By Using Ball Tree Searching Algorithm With Minkowski Distance Function on Smart Grid Data[J]. Indian J. Comput. Sci. Eng, 2022, 13(4): 1210-1226.
37. Wang S, Zhu Y, Luo X, et al. Knowledge Extraction and Distillation from Large-Scale Image-Text Colonoscopy Records Leveraging Large Language and Vision Models[J]. arXiv preprint arXiv:2310.11173, 2023.