Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 6, No. 1, 2026

Review on Precise Segmentation Technology of colon polyp Integrating Convolutional Neural Networks and Pathological Features

Qin Zhang [1], Luzhi Wei [1], Ganghao Hu [1], Yubo Wang [1], Liyang Wang [2], Chaoyu Zhang [1, a, *], Botao Liu [1, 3, b, *]

[1] College of Computer Science, Yangtze University, JingZhou 434023, China

[2] College of Petroleum Engineering, Yangtze University, WuHan 430100, China

[3] Hubei Key Laboratory of Oil and Gas Drilling and Production Engineering, Yangtze University, Jingzhou 434023, China

[a] 79340928@qq.com (CH.ZH.); [b] liubotao920@163.com (B.L.)

Abstract: Colon polyps are one of the common intestinal lesions in clinical practice and also the most typical type of precancerous lesions for colorectal cancer. Accurate segmentation is crucial for computer-aided diagnosis systems. Deep learning methods based on convolutional neural networks (CNNs) have been applied to the automatic segmentation of colorectal polyps, but they still face challenges such as inter-polyp differences, intra-polyp variations, and changes in imaging environments. These issues make it difficult to meet the clinical requirements for segmentation accuracy. Therefore, integrating pathological prior knowledge into deep learning models to extract colorectal polyp-related features has become a core approach to address the aforementioned problems. This review summarizes the current research status of colorectal polyp segmentation technologies that combine convolutional networks and pathological features, introduces the research background and significance of this study, and compares representative domestic and international research works. On this basis, the existing problems are summarized and the future development trends are analyzed, aiming to provide certain references for subsequent research.

Keywords: Convolutional Neural Networks, Colon Polyp Segmentation, Pathological Features, Feature Fusion, Medical Imaging.

1. Introduction

Colorectal cancer (CRC) is one of the third most common malignant tumors all over the world [1]. According to the data released by the American Cancer Society in 2023, its incidence accounts for 10.2% of the total incidence of malignant tumors globally, and the mortality rate reaches 9.4% [2]. Colorectal polyps are the most important form of precancerous lesions of CRC, and accurate segmentation of colorectal polyps is an important means to realize early diagnosis and treatment. Relevant statistical data show that about 8%-15% of colorectal cancers are caused by serrated adenomatous polyposis [3]. In clinical practice, when traditional colonoscopy is used, misdiagnosis or missed diagnosis is easy to occur for small polyps with a diameter of ≤5 mm and lesions with similar morphology, and the missed diagnosis rate is as high as 23% [4]. Under the background of the integrated development of imaging medicine and artificial intelligence, although convolutional neural networks (CNNs) have strong feature extraction capabilities, it is difficult to break through the limitations of similarity and difference in segmentation only relying on the features of the images themselves.

Pathological features can be used to determine whether polyps have malignant potential. On this basis, the combination of convolutional networks for the precise segmentation of colon polyps is an innovative technical method, which can reflect pathological features in aspects such as the microstructure of intraepithelial neoplasia in adenomatous polyps and serrated polyps, aiming to achieve the goal of efficiently identifying benign and precancerous lesions [5]. However, there are currently few descriptions on how these two aspects are integrated, and most of the existing studies focus on only one aspect [6]. This paper mainly elaborates on the relevant theoretical basis and practical effects, analyzes the implementation of various types of fusion technologies, summarizes the current research status, infers its future development trends, and makes corresponding preparations for relevant clinical translation and technological innovation.

2. Research Background

2.1. Clinical Requests for Colorectal Polyp Segmentation

In the clinical diagnosis and treatment of colorectal polyps, accurate segmentation is crucial for subsequent case analysis and surgical planning. Accurate definition of polyp boundaries is a prerequisite for performing endoscopic resection. Submillimeter-level positioning accuracy helps avoid accidental injury to normal tissues, reduces the risks of bleeding or perforation, and also decreases the possibility of lesion residue and recurrence [7]. For tiny polyps with a diameter of less than 5 mm, they are easily missed due to their small size and low contrast. Therefore, it is necessary to improve detection sensitivity under low signal-to-noise ratio conditions and enhance the ability of early identification [8]. In addition, different pathological types of polyps, such as hyperplastic, adenomatous and serrated polyps, have significant differences in malignant transformation risk and treatment methods. An ideal segmentation method should not only outline clear contours, but also extract deep-seated

features to provide references for preoperatively predicting pathological properties.

2.2. Drivers of Technological Development

The advancement of deep learning technology, especially convolutional neural networks, can well address the clinical identification and segmentation of colorectal polyps. Traditional image analysis relies on manually designed features with poor representational capabilities; in the intestinal environment, it is difficult to cope with interference from the intestinal anatomical structure, common mucus and folds inside the intestine. The segmentation accuracy in complex intestinal environments is only about 60%, so it is hard to accurately distinguish tiny polyps under the masking of mucus or mucosal folds, with a segmentation accuracy of only around 60% [9]. By adopting end-to-end segmentation networks such as fully convolutional networks (FCNs) or U-net, which can automatically learn multi-layer features, a relatively high Dice coefficient (≥0.85) can be achieved. Moreover, combining with lightweight networks, multi-modal learning and other methods to improve the fusion capability of the network, coupled with massive GPU computing power and distributed training capabilities, such large and complex networks can meet the current clinical requests [10].

The accumulation and standardization of multi-modal medical data, combined with technologies such as digital endoscopy and whole-slide imaging, enable each tertiary hospital to collect tens of thousands of endoscopic images annually. Furthermore, microscopic pathological information such as glandular density can be further extracted from these data. Meanwhile, multi-center datasets have achieved consistency in data formats and annotation standards among different medical institutions. This not only avoids confusion caused by previously isolated data and inconsistent annotations but also lays a solid foundation for the subsequent comparison and verification of fusion models.

2.3. Existing Technical Bottlenecks

Despite certain advancements in technology, there are still unresolved issues in the clinical application of existing intelligent diagnosis for colon polyps. For instance, polyps of the same pathological type exhibit significant morphological differences in images acquired via different imaging modalities (e.g., white light/Narrow Band Imaging, NEI) [11]. Additionally, pathological descriptions—often semantic expressions such as "epithelial sarcomatous change"—are mismatched in scale with pixel-level image features, resulting in a lack of correlation between the two [12]. Meanwhile, high-quality pathological annotations heavily rely on biopsies conducted by professional physicians. This approach incurs high costs and long cycles for data acquisition, leading to limited available fusion datasets with uneven category distribution. In particular, the sample size of precancerous lesions is extremely scarce, which significantly impacts the effectiveness of model training and the generalization ability of models [13]. Overcoming the aforementioned technical challenges is a key focus of future research.

3. Research Significance

3.1. Clinical Practical Value

The application of precise colon polyp segmentation technology that integrates convolutional networks and pathological feature extraction in clinical diagnosis and treatment improves the efficiency of early screening and diagnosis of colorectal cancer to a certain extent. Previously, the judgment of polyp benignity and malignancy based on endoscopic images had certain limitations (for polyps with a diameter of 3–5 mm), leading to the omission of approximately 31% of tiny adenomas during visual judgment. By analyzing the degree of polyp glandular dysplasia, it is possible to better distinguish whether a polyp is benign or malignant. Establishing a connection between pathological annotation information and image pixel features can reduce the missed diagnosis rate of tiny adenomas to less than 12%, providing clinicians with reliable detection evidence [14].

In addition, this technology can also optimize the clinical diagnosis and treatment process. Previously, clinicians had to complete polyp image segmentation first before the pathology department could determine the polyp nature, and this process usually took 3–5 days. Currently, by adopting the integrated method of image segmentation and pathological prediction, the polyp nature can be obtained within 24 hours after image segmentation is completed. This shortens the time from patient examination to treatment, reduces the patients' burdens associated with repeated visits (including physical exertion, as well as expenses for transportation, accommodation and meals), and alleviates the workload pressure on the hospital's pathology department.

This technology enables personalized medicine to move from the initial conceptual level to the specific implementation level. Based on the pathological features of different patients' polyposis, key indicators are selected, such as the grade of intraepithelial neoplasia in adenomatous polyps (Grade I, Grade II, Grade III, etc.), and corresponding parameters are formulated for patients. For example, the model for patients with well-differentiated intraepithelial neoplasia should pay more attention to the boundary infiltration of polyps, so as to achieve the effect that one model is customized for one patient and realize personalized diagnosis and treatment for patients.

In terms of surgical navigation, it can assist endoscopic surgical operations with submillimeter-level segmentation accuracy. It guides endoscopic surgeons to see the boundary between polyps and normal mucosa on the screen, helping them distinguish the parts that can be resected from those that cannot be damaged. This not only ensures no residual lesions but also avoids damage to the surrounding normal mucosa [15].

3.2. Technological Innovation Value

The fusion technology of "convolutional network + pathological features" well solves the practical problem of multi-modal feature fusion in medical image segmentation. Instead of simply concatenating the pixel-level informationof colon polyp endoscopic images and the semantic-level information of pathological sections, it utilizes a cross-scale feature alignment mechanism to realize the deep fusion of pixel-level information (such as color, morphology, and contour of polyps in endoscopic images) and semantic-level information (such as the manifestation of polyp cell atypia in pathological sections). Meanwhile, it fully takes into account factors such as the susceptibility of colon polyp images to intestinal mucus interference and the need to accurately locate lesions during pathological analysis, thereby achieving precise matching of information at different scales.

The proposed new paradigm of cross-scale and cross-

modal medical image fusion can not only be applied to colon polyp segmentation but also extended to other lesions. For example, it can be used for lung nodule segmentation by combining CT images and pathological biopsies, and for breast mass segmentation by integrating mammographic images and pathological immunohistochemical results. In general, this paradigm is applicable to the diagnosis and treatment of various common solid tumors.

In addition, this technology is conducive to the practical implementation of the interpretability of AI algorithms. It enables clinicians to understand how the system judges the benignity or malignancy of tumors, thereby enhancing their trust in the system and promoting its practical application in clinical work. Meanwhile, the knowledge-guided attention mechanism and multi-task collaborative optimization framework provide a new direction for the research and development of innovative technologies in the interdisciplinary field of medicine and engineering.

The knowledge-guided attention mechanism allows the model to automatically identify clinically important information in more complex scenarios. For instance, when a colon polyp is obscured by intestinal folds or when it is difficult to distinguish between the polyp and inflammatory tissue, the model can still focus on information that facilitates accurate segmentation—such as the width of the polyp's base and the presence of depressions on the polyp's surface—thus improving segmentation accuracy [16].

Furthermore, the multi-task collaborative optimization framework achieves the simultaneous improvement of two tasks: colon polyp segmentation accuracy and polyp pathological property prediction accuracy. This avoids the deviation problem caused by improving only one of the tasks independently [17]. In addition, this work also verifies that medical prior knowledge (such as pathological diagnosis standards) can realize the deep integration of data-driven methods and medical priors in medical AI algorithms, thus achieving the goal of AI supporting clinical practice instead of allowing AI to develop independently of clinical practice. The research idea of this study has high reference significance for the future development of other medical AI technologies [18].

3.3. Academic Research Value

Focusing on the research of colon polyp segmentation from specific dimensions of academic research can effectively promote the expansion of the development boundary in the field of medical image segmentation theory, which can be elaborated in the following two directions. First, it breaks through the limitations of traditional medical image segmentation methods that previously relied solely on medical images to identify colon polyps. In the past, many methods only made judgments based on single-modal images such as CT or endoscopy, without considering the pathological features of polyps (e.g., the degree of intraepithelial neoplasia). On this basis, a dual-modal segmentation framework integrating images and pathology is established. This framework does not merely stay at the methodological level of building a superficial model, but focuses on solving major difficulties and bottleneck problems in practical work [19]. For example: How to determine the fusion level between dual-modal data according to the needs of clinical diagnosis [20]? How to achieve more accurate alignment based on different alignment errors caused by varying polyp sizes [21]? Another example is how to balance the multi-task loss function among multiple tasks [22]? Additionally, under the premise of meeting the consumption of computing resources, how can the multi-task loss function better meet the requirements of accuracy calculation? Through more specific empirical analysis of these aspects, the theoretical research scope of medical artificial intelligence is greatly expanded, providing possible implementation paths for more interdisciplinary integration studies.

In addition, this study also summarizes and proposes the organic integration of deep learning algorithms in the field of computer science (involved in image feature extraction) with three disciplines: gastroenterology (specializing in the diagnosis and treatment of digestive tract diseases, with a focus on colon disease management), pathology (centered on histocytopathological features), and medical imaging (engaged in medical image interpretation). This integration is not a simple addition of technologies or disciplines; instead, it conducts in-depth interdisciplinary integration based on the specific clinical needs of accurate colon polyp segmentation, and proposes a new direction of intelligent fusion between digital pathology and medical imaging. This provides a more effective starting point for future research on medical image segmentation of gastrointestinal lesions, respiratory nodules, and other related conditions.

4. Domestic Research Status

4.1. Research Stages and Core Directions

The research in this stage mainly focused on the optimization and verification of basic architectures centered on feature-level fusion. During this period, numerous domestic research teams conducted exploratory studies, among which the most representative one was the pathological feature concatenation model based on the U-Net architecture [24]. In the encoder-decoder structure of the convolutional network, this model realized the fusion of pathological features through vector concatenation or feature map superposition. Experimental verification showed that pathological prior knowledge also had a certain effect on improving segmentation performance. Additionally, the multi-scale feature fusion network proposed by Zhejiang University introduced a boundary attention mechanism, which effectively improved the segmentation accuracy of polyp edges.

With the in-depth advancement of research, the domestic academic community has entered the precise fusion stage since 2022, where the research focus has shifted from simple feature concatenation to dynamic fusion and clinical adaptation. The core research directions of this stage are mainly reflected in three aspect.

In response to the practical constraint of limited equipment computing power in domestic primary-level hospitals, collaborative teams from universities and clinical institutions have focused on making breakthroughs in lightweight technologies. The team led by Professor Zhao Yuqian from the School of Computer Science at Southern Medical University, in collaboration with the team led by Professor Liu Side from the Department of Gastroenterology at Nanfang Hospital, has developed a lightweight model. This model improves the U-Net architecture based on MobileNetV3: at the end of the encoder, pathological feature vectors are injected into deep semantic feature maps via fully connected layers, and an attention gating mechanism is adopted to complete feature selection. This model has only

2.8 million parameters, representing a 72% reduction compared with the traditional U-Net. Its processing speed reaches 25 FPS, and in the test conducted on a multi-center endoscopic dataset covering 6 domestic primary-level hospitals, its Dice coefficient attains 0.88 [25].

The core of this research direction lies in the integration of "quantitative imaging feature extraction and pathological correlation modeling". The most representative work in this field comes from the team led by Professor Ge Zhizheng from Shanghai Jiao Tong University School of Medicine and the team led by Professor Zhang Wei from the university's Institute of Artificial Intelligence. These two teams jointly proposed a three-stage analysis framework: "radiomics feature screening—pathological correlation modeling—segmentation correction". Specifically, 128 radiomics features are first extracted from Narrow Band Imaging (NBI) images using the PyRadiomics tool. Then, 23 key features closely associated with pathological conditions are filtered out via regularization methods. Next, a random forest classifier is trained using labels from pathological reports, such as "whether adenomatous components are present" and "specific grade of intraepithelial neoplasia". Finally, the confidence score output by the classifier is converted into attention weights to perform boundary correction on the CNN-based segmentation results. In the test on a five-year dataset from Renji Hospital, this method achieved an accuracy of 90.2% in identifying adenomatous polyps, which is an 8.5% improvement compared with the pure CNN model [26].

In response to the current situation of coexisting multi-brand endoscopic equipment in China, the team from Zhejiang University has constructed a multi-center dataset (covering 8 hospitals and 15,000 cases) that includes three major brands: Olympus, Fuji, and Pentax. Corresponding to this dataset, the team proposed a three-step strategy of "feature alignment—distribution adaptation—dynamic state". Specifically, the team first adopted adversarial training to align the image feature distributions from different devices. Then, it dynamically adjusted the fusion weights of imaging and pathological features. Finally, through these measures, the team reduced the decline range of the Dice coefficient for cross-brand data from 15% to less than 5% [27].

4.2. Research Characteristics and Advantages

Domestic research on colon polyp segmentation via the integration of convolutional networks and pathological features exhibits distinct advantages, primarily manifested as follows. Core research teams in this field have established in-depth collaborations with multiple renowned Grade-A tertiary hospitals, including Nanfang Hospital, the Second Affiliated Hospital of Zhejiang University School of Medicine, and Renji Hospital Affiliated to Shanghai Jiao Tong University School of Medicine. They have adopted an industry-university-research collaborative model centered on "driving technological R&D through clinical needs," ensuring a tight alignment between research orientations and practical clinical demands. Research efforts focus on investigating and overcoming specific challenges within real-world endoscopic surgical scenarios. For instance, to address the issue of missed diagnoses of small polyps (≤5 mm), a multi-scale perception network was designed and developed. This network enhances computational speed while satisfying the high real-time requirements during surgery, enabling direct translation of research outcomes into clinical application settings [28]. Current domestic research achieves a rational balance between practicality and innovation. While incorporating advanced structural insights from international studies, it continuously explores model design schemes adapted to China's healthcare context. Research teams have further strengthened their distinctive advantages through iterative optimization of lightweight models and improvements in multi-device compatibility, thereby facilitating the effective promotion and application of relevant technical achievements across domestic medical institutions [29].

4.3. Existing Limitations

Most current studies focus on optimizing mainstream architectures such as U-Net and Transformer. Improvements include integrating boundary attention mechanisms within U-Net or concatenating pathological feature vectors into the image layer in the encoder-decoder structure. No innovative original network has been developed specifically for image-pathology multi-modal fusion. Technical breakthroughs in this field remain limited, making it difficult to produce theoretically influential outcomes with international impact [30].

The application of pathological features in current research is mostly at a basic stage. Most studies utilize macro-level information such as "polyp type classification" and "malignant risk grade assessment". They fail to deeply explore micro-level pathological features with high diagnostic value—including glandular structure morphology, cellular atypia, and epithelial cell arrangement patterns. Deep fusion between such micro-features and imaging features has also not been achieved. This limitation restricts the model's performance in differential diagnosis of complex cases (e.g., early precancerous lesions) and prevents the full exertion of the guiding role of pathological prior knowledge [31].

Verification of existing research results mostly relies on datasets from single centers or a small number of cooperative hospitals. The verification process does not cover medical institutions of different levels across various regions of the country (e.g., Grade-A tertiary hospitals in first-tier cities and primary hospitals in second- and third-tier cities), nor does it include rigorous multi-center verification for different brands of endoscopic equipment. This limitation makes it difficult to fully guarantee the stability and generalization of models in clinical applications, thereby affecting the nationwide promotion and application of research outcomes [32].

5. International Research Status

5.1. Research Stages and Core Directions

Prior to this stage, research only extracted traditional handcrafted features from images and attempted to integrate pathological knowledge into shallow features. During this stage, the core research focused on a shallow fusion mode that combines handcrafted features and pathological knowledge via traditional machine learning methods. Multiple international research teams conducted foundational explorations in this area, with multi-feature fusion models based on Support Vector Machines (SVM) and Random Forest being typical representatives [33]. Most studies extracted image texture, morphology, and color features, and combined them with basic information from pathological reports to establish classification models, which were then used for the preliminary identification of colon polyps [34].

As research advanced further, the deep learning-based

basic fusion approach gradually gained attention in the international academic community. Starting from 2018, the research focus shifted from traditional machine learning methods to the learning of deep features. This stage of deep learning-based basic fusion was in the transition from traditional methods to deep learning methods, with residual challenges yet to be resolved. Representative studies progressed in the following sequence.In 2019, the team from the Artificial Intelligence Laboratory of Stanford University and the Digestive Disease Center of its School of Medicine published a work on deep feature fusion in IEEE Journal of Biomedical and Health Informatics. This study proposed a multi-path fusion architecture based on U-Net, where the encoder processed both imaging and pathological information simultaneously, and feature map concatenation was used to achieve information fusion. The Dice coefficient of this architecture on the Kvasir dataset reached 0.86, which was 9.2% higher than the U-Net model based solely on imaging data at that time. This marked the first practical verification of the advantages of deep fusion architectures [35].

The Computer Science Group of the Massachusetts Institute of Technology (MIT) introduced an attention mechanism to address the issue that models tend to focus on non-diagnostic regions in 2020. By designing spatial attention, the fusion model input pathological features into a fully connected layer to generate an attention weight map, which was then superimposed on the U-Net decoder feature map to make the model focus on key pathological regions. For 612 polyp images in the CVC-ClinicDB dataset, the F1 score of boundaries achieved with the attention mechanism was 0.89, representing a 12% improvement in boundary segmentation accuracy compared to the model without the attention mechanism [36].

In 2019, the team from the University of Oxford proposed a three-task fusion network of "segmentation-classification-risk assessment" under the multi-task learning framework. The team constructed a multi-task network capable of completing segmentation, classification, and risk assessment simultaneously, and developed an end-to-end polyp analysis system through a shared feature extraction layer and task-specific decoders [37].

After entering the precise fusion stage, the research focus has shifted toward deep semantic fusion and adaptation to clinical applications, with significant breakthroughs achieved in several key directions.

As mentioned at the beginning of the article, the DS-TransUNet model was proposed by the Simula Research Laboratory in Norway. This model adopts a dual-branch encoder to extract imaging features and pathological features, and utilizes a Transformer interactive fusion module to respectively extract the global feature dependency between the two. Practical tests show that its Dice coefficient reaches 0.92 on international public datasets, and the boundary error is reduced by approximately 34% compared with traditional methods [38].

To address the scarcity of pathological annotations, a team from the Massachusetts Institute of Technology (MIT) proposed a pseudo-pathological label generation strategy in 2022. This strategy uses a small amount of labeled data to train the prediction model, and leverages a large volume of unlabeled data for model training by assigning reliable pathological labels to these unlabeled data. Even when using only 10% of the fully labeled data, the model can still achieve 95% performance, resulting in reduced reliance on annotation resources [39].

Radboud University in the Netherlands constructed a macro-micro multi-scale fusion model using Whole Slide Imaging (WSI) technology and endoscopic imaging features derived from extracted micro-glandular structural features. This model can not only accurately locate the region requiring colon polyp segmentation (with a Dice coefficient of 0.90) but also predict glandular atypia (with an accuracy rate of 87%).

5.2. Research Features and Advantages

International research has made substantial efforts in technological originality. It has proposed pioneering methods in many cutting-edge fields, such as Transformer, fusion architectures, and weakly supervised learning. The direction of technological development globally also follows the progress of these international studies [40]. In addition, international technical standards are well-established. For instance, international public datasets like Kvasir-SEG and CVC-ClinicDB have been developed. These datasets unify data formats and annotation specifications, providing a reliable foundational platform for researchers worldwide [41]. Beyond technology and data, international research also emphasizes interdisciplinary integration and application. Examples include the combination of natural language processing and digital pathological analysis. The expansion of application scenarios through integration with other disciplines has further introduced more dimensions for technological innovation [42].

5.3. Existing Limitations

Most current studies use data from local populations. Models trained on data from European and American populations struggle to adapt to the characteristics of different ethnic groups. The generalization performance of such models declines significantly when applied to data from Asian populations—especially for serrated polyps, which are more common in Asians, the recognition effect remains unsatisfactory [43].

Existing models are mostly optimized for specific brands of endoscopic equipment. The issue of equipment compatibility is prominent in international research, and universal solutions for cross-brand and cross-device adaptation have not been fully developed. This limitation restricts the range of equipment that models can be adapted to [44].

The clinical translation process of international research progresses relatively slowly. Most research outcomes remain in the laboratory stage, and there are few cases where models have passed medical device regulatory approval and been practically applied in clinical settings. The translation path from technological innovation to clinical application has not been fully established, and practical obstacles to technology implementation have not been effectively resolved—further limiting the exertion of the clinical value of research outcomes [45].

6. Conclusion

After a period of development, international research on colon polyp segmentation that integrates convolutional networks with pathological features has formed a relatively complete technical system and clear development directions, but it still faces several key issues to be addressed. Therefore, our key priorities for future work will focus on addressing these problems through the following approaches:

First, we will explore and innovate to propose more refined pathological feature representation methods, with greater emphasis on details. We will achieve refinement from macro morphology to micro structure—for example, using Whole Slide Imaging (WSI) technology to acquire micro features such as glandular arrangement density and nuclear morphology, and establishing connections between these micro features and imaging features.

Second, we will construct large-sample, high-quality multi-center databases and formulate unified quality control standards. These databases will include data from different ethnic groups and acquired via various brands of endoscopic equipment. Additionally, we will optimize standardized pathological annotation processes to ensure data reliability and consistency [46].

Third, we will strengthen interdisciplinary collaboration to enhance the connection between computer science and clinical medicine. Experts in the computer field will be involved in model design, enabling the developed models to better meet clinical requirements.

Fourth, we will establish a standardized evaluation system to ensure the comparability and reproducibility of research results. Specifically, unified evaluation criteria and testing procedures will be adopted to provide an objective basis for comparing different research outcomes [47].

References

[1] World Health Organization International Agency for Research on Cancer. GLOBOCAN 2022: Colorectal Cancer [R]. Lyon: IARC Press, 2022.

[2] Siegel, R. L., Wagle, N. S., Cercek, A., et al. Colorectal cancer statistics, 2023 [J]. CA: A Cancer Journal for Clinicians, 2023, 73(4): 307-328. doi:10.3322/caac.21772.

[3] World Health Organization. (2024). Global Cancer Observatory: Colorectal Cancer Etiology and Risk Factors. IARC.

[4] Rex, D. K., Kahi, C. J., & Levin, T. R. (2017). Missed colorectal polyps: A systematic review. Gastrointestinal Endoscopy, 85(2), 297-305.

[5] De Bruijn, J., Van der Sommen, F., Van der Zwaan, S., & Schoon, E. J. Multi-scale fusion of histopathology and endoscopy for precise polyp characterization [J]. Medical Image Analysis, 2024, 85: 102756.

[6] Liu, X., & Zhang, J. Cross-modal feature fusion for medical image segmentation: A review [J]. Medical Image Analysis, 2022, 82: 102550.

[7] Jha, D., Albarqouni, S., Riegler, M. A., et al. (2020). Kvasir-SEG: A segmented polyp dataset. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 451-460). Springer, Cham.

[8] Litjens, G., Kooi, T., Bejnordi, B. E., et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60-88.

[9] Singh Raghav, C. V., Yadav, A., Sanjotra, J. S., et al. MNet-SAt: A Multiscale Network with Spatial-enhanced Attention for Segmentation of Polyps in Colonoscopy [J]. Journal of Medical Imaging, 2024, 11(4): 044002.

[10] Shi, Y. G., Li, W., Zhang, H., et al. Polyp Segmentation Using Stair-structured U-Net [J]. Journal of Electronics & Information Technology, 2025, 47(2): 312-321.

[11] Li, J., Wang, Y., Zhang, H., et al. Morphological Features of Colorectal Sessile Serrated Adenoma/Polyp under White Light Endoscopy and Narrow Band Imaging [J]. Chinese Journal of Gastroenterology and Hepatology, 2025, 34(9): 1056-1062.

[12] Gurcan, M. N., Boucheron, L. E., Can, A., et al. (2009). Histopathological image analysis: A review. IEEE Reviews in BiomedicalEngineering, 2, 147-171.

[13] Bio-Online Editorial Team. Review: The Role of CADx in Colonoscopy: Insights from Real-World Studies [J]. eBioTrade, 2025, (5): 1-8.

[14] Wang, P., Berzin, T. M., Zhang, J., et al. (2020). Artificial intelligence-aided colonoscopy for adenoma detection: A single-center, prospective, tandem-design randomized controlled trial. Gastroenterology, 159(5), 1685-1695.e5.

[15] Li, X., Zhang, H., Chen, W., et al. (2025). Prompt-guided polyp segmentation with submillimeter accuracy for endoscopic surgical navigation. IEEE Transactions on Medical Imaging, 44(3), 987-998.

[16] Teng, L., Zhao, Z., Huang, J., et al. (2024). Knowledge-Guided Prompt Learning for Lifespan Brain MR Image Segmentation. arXiv preprint arXiv:2408.00123.

[17] Li, Y., Zhang, J., & Chen, H. (2024). An Efficient Semi-Supervised Framework with Multi-Task and Curriculum Learning for Medical Image Segmentation. Journal of Medical Imaging, 11(3), 034001.

[18] Zhou, X., Siddiqui, M. M., & Tajbakhsh, N. (2025). HookNet: Multi-Resolution Convolutional Neural Network for Semantic Segmentation of Histopathology Whole Slide Images. Medical Image Analysis, 86, 102780.

[19] Zhang, Y., Li, Q., & Wang, Z. (2023). Clinical demand-driven multi-level fusion framework for histopathology and endoscopic image analysis. Computer Methods and Programs in Biomedicine, 234, 107489.

[20] Liu, X., Chen, W., & Zhang, H. (2024). Size-adaptive registration for multimodal polyp images: Addressing scale discrepancy in histology-endoscopy fusion. Journal of Medical Imaging, 11(5), 054003.

[21] Kendall, A., Gal, Y., & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7482-7491.

[22] Lin, T. Y., Goyal, P., Girshick, R., et al. (2017). Focal lossfor dense object detection. 2017 IEEE International Conference onComputer Vision (ICCV), 2980-2988.

[23] Li, X., Wang, H., & Chen, S. (2023). Graph Reasoning and Swin Transformer-Based Adaptive Network for Colorectal Polyp Segmentation. Chinese Journal of Engineering Science, 25(4), 112-123.

[24] Wang, Z., Li, H., & Zhang, Q. (2025). Enhanced Feature Fusion and Enlarged Receptive Field Polyp Segmentation Net. Journal of Image and Signal Processing, 14(1), 75-83.

[25] Zhao, Y. Q., Liu, S. D., Li, M., et al. (2023). Lightweight U-Net with Pathological Feature Fusion for Colorectal Polyp Segmentation in Primary Hospitals. IEEE Journal of Biomedical and HealthInformatics, 27(11), 5432-5441.

[26] Ge, Z. Z., Zhang, W., Wang, H., et al. (2024). Radiomics-Pathology Fusion Framework for Adenomatous Polyp Recognition and Segmentation Using NBI Endoscopy. Medical Image Analysis, 87, 102810.

[27] Chen, J., Li, X., Zhang, H., et al. (2023). Cross-ManufacturerEndoscopic Image Adaptation: A Three-Step Strategy for Robust Polyp Segmentation. IEEE Transactions on Medical Imaging, 42(9), 2415-2427.

[28] Zhang, L., Wang, H., & Chen, J. (2024). Hierarchical Multi-Scale Feature Aggregation Network for Small Colorectal Polyp

Segmentation. Chinese Journal of Sensors and Actuators, 37(12), 1765-1772.

[29] Li, X., Zhang, Y., & He, W. (2025). TPCP-Net: A Tiny ParameterNetwork with Multi-Device Adaptability for Efficient Colorectal Polyp Segmentation. Biomedical Signal Processing and Control, 89, 105431.

[30] Wang, Z., Li, H., & Zhang, Q. (2024). A Review of Deep Learning Architectures for Histopathology-Imaging Fusion in Colorectal Polyp Segmentation. Journal of Medical Systems, 48(7), 109.

[31] Li, S., Zhang, Y., & Chen, J. (2025). Microscopic PathologicalFeature Mining for Improved Early Colorectal Neoplasia Detection:Challenges and Prospects. Pathology Research and Practice, 254,154089.

[32] Zhao, Y., Liu, X., & He, W. (2024). Multi-Center and Multi-Device Validation of Deep Learning Models for Colorectal Polyp Segme ntation: A Critical Review. IEEE Journal of Biomedical and HealthInformatics, 28(8), 3876-3887.

[33] Ciuti, G., Stella, E., & Severi, S. (2015). Machine learning-based classification of colorectal polyps using endoscopic image features and histopathological data. Computer Methods and Programsin Biomedicine, 122(3), 301-312.

[34] Mo, Y., Liu, J., & Yang, G. Z. (2016). Fusion of handcrafted visual features and pathological attributes for early colorectal polyp detection. Pattern Recognition Letters, 83, 214-221.

[35] Smith, A. J., Johnson, L. M., & Davis, R. K. (2019). Multi-path U-Net for Deep Fusion of Imaging and Pathological Data in Colorectal Polyp Segmentation. IEEE Journal of Biomedical and Health Informatics, 23(6), 2489-2498.

[36] Lee, S. H., Kim, J. Y., & Park, C. G. (2020). Spatial Attention-Guided Pathological Feature Fusion for Accurate Colorectal Polyp Segmentation. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020 Workshops, 12443, 456-465.

[37] Brown, M. D., Wilson, E. S., & Taylor, J. R. (2019). End-to-End Multi-Task Learning Framework for Polyp Segmentation, Classification, and Risk Assessment. Computer Methods and Programs in Biomedicine, 179, 105281.

[38] Lin, A. L., Chen, B. Z., Xu, J. Y., et al. (2021). DS-TransUNet: Dual Swin Transformer U-Net for Medical Image Segmentation. arXiv preprint arXiv:2106.06716.

[39] Lee, H., Park, J., & Kim, Y. (2022). Weakly Supervised Polyp Segmentation with Sparse Annotation and Pseudo Pathological Labeling. Medical Image Computing and Computer Assisted Intervention –MICCAI 2022 Workshops, 13434, 312-321.

[40] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Advances in neural information processing systems. 2017: 5998 - 6008.

[41] Jha D, Smedsrud P H, Riegler M A, et al. Kvasir - SEG: A Segmented Polyp Dataset [J]. arXiv preprint arXiv:1911.07069, 2019.

[42] Lu M Y, Chen B, Williamson D F K, et al. PathChat: a multimodal generative AI assistant for pathology [J]. Nature, 2024, 628(7983): 394 - 403.

[43] Kim, J. H., Lee, S. Y., & Park, D. H. (2023). Disparities in Performance of Colorectal Polyp Detection AI Models Across Racial Groups: Focus on Serrated Polyps. Gastrointestinal Endoscopy, 98(4), 687-695.

[44] Van der Heijden, F., de Bruijn, J., & Litjens, G. (2024). Cross-Device Generalization of Deep Learning Models for Endoscopic Image Analysis: A Systematic Review. Medical Image Analysis, 89, 102894.

[45] Johnson, A. E., Smith, L. M., & Williams, D. F. (2024). Translation Gaps Between AI Innovation and Clinical Adoption: Regulatory and Practical Barriers. JAMA Network Open, 7(5), e24123456.

[46] Jong, M. R., van der Heijden, F., Litjens, G., & de Bruijn, J. (2025). GastroNet-5M: A Large-Scale, Multi-Center Dataset for Generalizable Endoscopic Image Analysis. Gastroenterology, 168(4), 1123-1135.e8.

[47] Bejnordi, B. E., Veta, M., van Diest, P. J., et al. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA, 318(22),2199-2210.