ChinaRxiv 中X

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202507.00341

Automatic Segmentation of Colorectal Cancer in Abdominal CT Images Using a Deep Learning Network Combining 3D U-Net and Transformer: A Multi-Center Multi-Device Postprint Study

Authors: Huang Mu, Zhang Daochun, Yang Wei, Zhong Liming, Yuan Wenjing, Jia Ziqi, Tan Xiangliang, Duan Xiaohui, Liu Xian, Chen Weicui, Chen Weicui

Date: 2025-07-17T00:00:00+00:00

Abstract

Background The application of deep learning in medical imaging faces challenges such as time-consuming and labor-intensive data annotation, limiting its clinical translation efficiency. Objective To investigate the feasibility and efficacy of a deep learning network (TransUNet-Cascade) that integrates three-dimensional U-Net (3D U-Net) and Transformer for automatic segmentation of colorectal cancer (CRC) lesions in abdominal CT images. Methods We retrospectively included contrast-enhanced abdominal CT images from 2,180 CRC patients at Guangdong Provincial Hospital of Chinese Medicine (Center 1), Nanfang Hospital of Southern Medical University (Center 2), and Sun Yat-sen Memorial Hospital of Sun Yat-sen University (Center 3) between January 2018 and May 2023, which were divided into training set (n=1,159), validation set (n=289), and external test set (n=732) using weighted random sampling. This study proposes a novel deep learning network model—TransUNet-Cascade, which optimizes segmentation accuracy through a multi-stage learning strategy. Using manual annotation as the ground truth, model performance was evaluated using Dice similarity coefficient (DSC), F1 score, 95% Hausdorff distance (HD95), intersection over union (IoU), precision (PRE), and recall (REC). This study selected three-dimensional no new U-Net (3D nnU-Net) as the comparative baseline model and conducted systematic training and performance comparison with the proposed TransUNet-Cascade network under unified dataset and evaluation criteria to comprehensively validate its effectiveness in CRC segmentation tasks. Results In the independent external test set, the segmentation performance of both deep learning networks based on arterial phase images was overall superior to that of venous phase images; the arterial phase average DSC, F1, HD95, IoU, PRE, and REC of TransUNet-Cascade were 0.740, 0.839, 34.084, 0.656, 0.737, and 0.767 respectively, which were overall superior to those of 3D nnU-Net (average DSC, F1, HD95, IoU, PRE, and REC were 0.724, 0.838, 35.954, 0.642, 0.730, and 0.744 respectively). The model achieved the best segmentation performance for right-sided colon cancer (DSC=0.784), while the segmentation performance for rectal cancer was relatively poorer (DSC=0.622). Conclusion TransUNet-Cascade enhances the automatic segmentation accuracy of CRC lesions by combining the advantages of convolutional neural networks and Transformer, demonstrating certain potential for clinical application.

Full Text

Automatic Segmentation of Colorectal Cancer in Abdominal CT Images Using a Deep Learning Network Based on Fused 3D U-Net and Transformer: A Multicenter, Multi-device Study

Mu Huang[1], Daochun Zhang[2], Wei Yang[1], Liming Zhong[1], Wenjing Yuan[3], Ziqi Jia[3], Xiangliang Tan[4], Xiaohui Duan[5], Xian Liu[3], Weicui Chen[3]*

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[2]Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Taizhou 317000, China
[3]Department of Radiology, Guangdong Provincial Traditional Chinese Medicine Hospital, the Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510120, China
[4]Department of Radiology, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China
[5]Department of Radiology, Sun Yat-Sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510235, China

Corresponding author: Weicui Chen, Associate chief physician; E-mail: weicuichen@126.com

---

Abstract

Background: The application of deep learning in medical imaging faces challenges such as time-consuming and labor-intensive data annotation, which hinders clinical translation efficiency. **Objective:** To investigate the feasibility and efficacy of a deep learning network (TransUNet-Cascade) that integrates 3D U-Net and Transformer for automatic segmentation of colorectal cancer (CRC) lesions in abdominal CT images. **Methods:** We retrospectively included contrast-enhanced abdominal CT images from 2,180 CRC patients across three centers: Guangdong Provincial Traditional Chinese Medicine Hospital (Center 1), Nanfang Hospital of Southern Medical University (Center 2), and Sun Yat-sen Memorial Hospital of Sun Yat-sen University (Center 3) between January 2018 and May 2023. The dataset was divided into training (n=1,159), validation (n=289), and external test sets (n=732) using weighted random sampling. We proposed a novel deep learning network model—TransUNet-Cascade—that optimizes segmentation accuracy through a multi-stage learning strategy. Using manual annotations as the gold standard, model performance was evaluated using Dice similarity coefficient (DSC), F1 score, 95% Hausdorff distance (HD95), intersection over union (IoU), precision (PRE), and recall (REC). We selected 3D no new U-Net (3D nnU-Net) as the baseline comparison model and conducted systematic training and performance comparison under unified dataset and evaluation criteria to comprehensively validate the effectiveness of TransUNet-Cascade for CRC segmentation. **Results:** In the independent external test set, both deep learning networks demonstrated superior segmentation performance on arterial phase images compared with venous phase images. TransUNet-Cascade achieved average DSC, F1, HD95, IoU, PRE, and REC values of 0.740, 0.839, 34.084, 0.656, 0.737, and 0.767, respectively, on arterial phase images, outperforming 3D nnU-Net (average DSC, F1, HD95, IoU, PRE, and REC of 0.724, 0.838, 35.954, 0.642, 0.730, and 0.744). The model achieved the best segmentation performance for right-sided colon cancer (DSC=0.784), while rectal cancer segmentation was relatively poorer (DSC=0.622). **Conclusion:** By combining the strengths of convolutional neural networks and Transformers, TransUNet-Cascade improves the accuracy of automatic CRC lesion segmentation and demonstrates potential for clinical application.

Keywords: Colorectal neoplasms; Tomography, X-ray computed; Deep learning; Artificial intelligence; Self-attention mechanism; Convolutional neural network

---

1. Materials and Methods

1.1 Study Subjects

This cross-sectional study was approved by the Ethics Committee of Guangdong Provincial Traditional Chinese Medicine Hospital (BE2023-142). Informed consent was waived due to its retrospective nature. We retrospectively enrolled CRC patients from three centers: Guangdong Provincial Traditional Chinese Medicine Hospital (Center 1), Nanfang Hospital of Southern Medical University (Center 2), and Sun Yat-sen Memorial Hospital of Sun Yat-sen University (Center 3) between January 2018 and May 2023. Inclusion criteria were: (1) pathologically confirmed primary CRC by postoperative histology, and (2) complete preoperative CT imaging data covering both arterial and venous phases. Exclusion criteria were: (1) poor image quality or severe artifacts, and (2) lesions not visually identifiable on CT images. Based on these criteria, 2,180 patients were ultimately included (Center 1: 777, Center 2: 732, Center 3: 671), comprising 1,285 males and 895 females aged 34-87 years with a mean age of 59$±$11 years. Data from Centers 1 and 3 were combined and split into training (n=1,159) and validation (n=289) sets at a 4:1 ratio, while Center 2 data served as an independent external test set (n=732). Stratified weighted random sampling was used for training/validation splitting, with tumor location (right/left colon, rectum, etc.) and scanner model as stratification variables to ensure balanced subgroup proportions. The external test set retained the original distribution from Center 2 to evaluate model performance in real clinical scenarios. The patient selection flowchart is shown in [Figure 1: see original paper].

1.2 CT Examination Methods

Abdominal CT scan data were acquired from four brands and seven models of CT scanners. Detailed equipment information and scanning parameters are provided in . For contrast agents, Centers 1 and 2 used iopromide (Bayer, Germany) at 370 mgI/mL, while Center 3 used ioversol or iomeprol (Bracco, Italy) at 350 mgI/mL. All centers administered contrast via elbow vein injection using a power injector at 2.5-3.0 mL/s with a dose of 1.2-1.5 mL/kg. Arterial phase images were acquired at 25-30 s post-injection, and venous phase images at 55-70 s.

1.3 CRC Localization

Referring to the *Chinese CRC Diagnosis and Treatment Guidelines (2023 Edition)* issued by the National Health Commission, CRC was classified by tumor location into right-sided colon cancer (including cecum, ascending colon, and hepatic flexure), transverse colon cancer, left-sided colon cancer (including splenic flexure, descending colon, and sigmoid colon), and rectal cancer.

1.4 Manual Tumor Annotation

Arterial and venous phase images were saved in DICOM format and converted to NIfTI format using Python. Manual tumor annotation was performed by one annotator (junior physician) and one reviewer (senior physician). The reviewer provided training to ensure annotation quality, and was responsible for reviewing, modifying (if necessary), and finalizing all annotations. The 3D-Slicer software (Version 4.10.2, http://www.slicer.org) was used for annotation, following the expert consensus on CRC CT and MRI annotation. The final annotated dataset was used to train network models and evaluate segmentation performance.

1.5 Automatic Segmentation Using Fused 3D U-Net and Transformer DL Network

We proposed a two-stage CRC segmentation algorithm, with the overall workflow illustrated in [Figure 2: see original paper].

1.5.1 Image Preprocessing We implemented the following standardized preprocessing pipeline for raw CT images. First, to eliminate spatial resolution differences caused by varying CT equipment and scanning parameters, all raw images were resampled to a uniform slice thickness of 3 mm. Subsequently, a two-stage intensity normalization method was applied. The first stage involved truncated normalization: based on the voxel intensity distribution of each image, we calculated its mean and standard deviation and truncated intensity values to the 0.5%-99.5% percentile range to eliminate extreme outliers. The second stage applied Z-score normalization by subtracting the image mean and dividing by its standard deviation, transforming the intensity distribution to a standard normal distribution (mean=0, SD=1) and ensuring consistency across CT images from different sources.

1.5.2 Fused 3D U-Net and Transformer DL Network Framework We designed a novel deep learning network framework called TransUNet-Cascade, built upon a 3D U-Net backbone with integrated Swin Transformer modules to enhance segmentation accuracy through a multi-stage learning strategy (Figure 2). The Swin Transformer module consists of two basic blocks, each including normalization steps. Input features are first normalized through layer normalization, then processed by a 3D window-based multi-head self-attention mechanism to extract local features via interactions between Query, Key, and Value. Features are subsequently fed into a multi-layer perceptron module for nonlinear mapping to enhance feature representation. In the second block, the conventional window mechanism is replaced with a 3D shifted window multi-head self-attention mechanism that integrates global contextual information through window shifting strategies, significantly improving global feature modeling capability. Each step incorporates layer normalization and residual connections.

TransUNet-Cascade employs a cascaded network architecture with a multi-stage learning strategy for progressive segmentation optimization. In the first stage, the network processes input images at low resolution, combining CNN for local feature extraction with Transformer for global modeling to generate coarse regions of interest (ROIs). These preliminary results serve as input for the second stage, which crops and focuses on high-resolution ROIs. In the second stage, the network performs fine-grained segmentation on high-resolution ROIs to further optimize boundary delineation and structural detail preservation.

To comprehensively evaluate TransUNet-Cascade' s effectiveness, we introduced the classic segmentation model 3D nnU-Net for comparison. 3D nnU-Net is a highly automated 3D medical image segmentation framework based on standard 3D U-Net with a symmetric encoder-decoder design and skip connections for effective fusion of low-level features and high-level semantic information. The encoder extracts multi-scale features through multiple 3D convolution and downsampling layers, while the decoder gradually restores spatial resolution through upsampling and convolution for fine structural reconstruction. We selected 3D nnU-Net as the baseline model for systematic training and performance comparison under unified dataset and evaluation criteria to comprehensively validate TransUNet-Cascade' s effectiveness for CRC segmentation.

1.5.3 Model Training and Testing During training, image patches of size 192\$×192×\$48 (length × width × slices) were randomly sampled and fed into the network in batches of 2. To mitigate overfitting, spatial and intensity data augmentation was applied, including random flipping, rotation, scaling, Gaussian noise addition, and Gaussian blurring. Stochastic gradient descent was used as the optimizer with an initial learning rate of 0.01, gradually decaying to 5.5\$×10^{-4}.*Themodelwastrainedfor*1,000*epochsusingthePyTorchframework.Thelossfunctioncombinedbinarycross*
*entropylossandDicelosswithequalweightsof*0.5.*ThemodelwiththehighestaverageDicecoefficientonthevalida*
Network output probabilities were binarized at a threshold of 0.5, and the largest connected component was retained as the final segmentation result.

1.6 Evaluation of Segmentation Model Performance

Using manual annotations as the gold standard, we evaluated TransUNet-Cascade and 3D nnU-Net performance using Dice similarity coefficient (DSC), F1 score, 95% Hausdorff distance (HD95), intersection over union (IoU), precision (PRE), and recall (REC). The formulas are as follows:

$$\mathrm{DSC} = \frac{2 \cdot |X \cap Y|}{|X + Y|}$$

$$\text{F1-score} = \left[\frac{\mathrm{PRE}^{-1} + \mathrm{REC}^{-1}}{2}\right]^{-1}$$

$$\mathrm{HD95} = \max\left[95\text{percent} \sup_{a\in A} \inf_{b\in B} d(a,b),\ 95\text{percent} \sup_{b\in B} \inf_{a\in A} d(a,b)\right]$$

$$\mathrm{PRE} = \frac{V_{TP}}{|V_{TP} + V_{FP}|}$$

$$\mathrm{IoU} = \frac{V_{TP}}{|V_{TP} + V_{FP} + V_{FN}|}$$

$$\mathrm{REC} = \frac{V_{TP}}{|V_{TP} + V_{FN}|}$$

where X represents the ground truth set and Y the prediction set; V_{TP} denotes voxels correctly classified as tumor region (consistent with manual annotation); V_{FN} represents tumor region voxels classified as background; and V_{FP} denotes background voxels classified as tumor region. $d(a, b)$ is the distance from point a to point b; inf is the infimum (minimum possible value); sup is the supremum (maximum possible value); and 95percent sup is the supremum calculated based on the 95th percentile of boundary point distances between sets A and B.

1.7 Statistical Analysis

Statistical analysis was performed using SPSS 19.0. All measurement data were normally distributed and expressed as mean ± standard deviation. Inter-group comparisons were conducted using one-way ANOVA. Count data were expressed as percentages and compared using chi-square tests. Statistical significance was set at $P < 0.05$.

---

2. Results

2.1 Patient Clinical Data

There was no statistically significant difference in age among the training, validation, and external test sets ($P > 0.05$). However, significant differences were observed in gender and tumor location distribution among the three sets ($P < 0.001$), as shown in .

2.2 Comparison of Overall Segmentation Performance Between Two Models

To validate TransUNet-Cascade’ s segmentation performance, we compared it with 3D nnU-Net, with results illustrated in [Figure 3: see original paper]. In the external test set, TransUNet-Cascade achieved average DSC, F1, HD95, IoU, PRE, and REC values of 0.740, 0.839, 34.084, 0.656, 0.737, and 0.767, respectively, on arterial phase images, outperforming 3D nnU-Net (average DSC, F1, HD95, IoU, PRE, and REC of 0.724, 0.838, 35.954, 0.642, 0.730, and 0.744). On venous phase images, TransUNet-Cascade achieved average DSC, F1, HD95, IoU, PRE, and REC values of 0.688, 0.812, 32.364, 0.601, 0.666, and 0.743, respectively, while 3D nnU-Net achieved 0.660, 0.810, 35.412, 0.576, 0.625, and 0.730. Both DL models demonstrated superior segmentation performance on arterial phase compared with venous phase images, with TransUNet-Cascade consistently outperforming 3D nnU-Net.

2.3 Comparison of Segmentation Performance by Tumor Location

Further analysis of TransUNet-Cascade’ s performance by tumor location revealed optimal segmentation for right-sided colon cancer, with DSC, F1, HD95, IoU, PRE, and REC values of 0.730, 0.833, 36.539, 0.650, 0.728, and 0.757, respectively. Performance was slightly lower for left-sided colon cancer and transverse colon cancer, while rectal cancer segmentation was the least effective, with DSC, F1, HD95, IoU, PRE, and REC values of 0.622, 0.835, 43.832, 0.547, 0.611, and 0.659, respectively, as detailed in .

Discussion

Accurate segmentation of CRC lesions is a critical prerequisite for building robust artificial intelligence algorithms. The colon and rectum are flexible hollow organs, and CRC exhibits significant heterogeneity in imaging, with variable size, diverse morphology, and irregular margins. Additionally, the complex anatomical surroundings with adjacent organs further increase the difficulty of automatic CRC segmentation from imaging data.

Over the past decade, convolutional neural networks (CNNs) have achieved remarkable progress in medical image segmentation. U-Net, with its symmetric encoder-decoder architecture and skip connections, effectively addresses challenges of small sample sizes and high-precision localization in medical image segmentation, becoming a milestone in the field. Our previous work has preliminarily explored 3D nnU-Net for automatic CRC segmentation in abdominal CT images. However, CNNs have limited receptive fields and typically use small convolution kernels to balance model precision and computational complexity, emphasizing local feature extraction while performing poorly in capturing long-range dependencies.

Transformer, a deep learning model based on self-attention mechanisms, was first proposed by Vaswani et al. in 2017. Transformer abandons traditional convolutional and recurrent structures, capturing long-range dependencies through parallel sequence processing. Its core features, including multi-head attention and positional encoding, have demonstrated excellent performance in natural language processing tasks. The encoder-decoder architecture and stackable multi-layer structure provide powerful feature extraction capabilities, forming the foundation for advanced models such as BERT and GPT. Recently, Transformer has been increasingly applied to medical image segmentation with significant advantages. For example, Ghazouani et al. combined Transformer with CNN modules in an encoder-decoder structure to segment MR images of 1,251 brain tumor patients from the BraTS 2021 dataset, achieving an average DSC of 0.8977. Sun et al. proposed a network algorithm fusing Vision Transformer with edge-guided encoding-decoding for automatic spine segmentation in 195 T2WI images, attaining a DSC of 0.9015. Liu et al. designed a multi-scale edge optimization algorithm based on Swin Transformer for bladder cancer MRI segmentation, achieving an average DSC of 0.9373 in 100 patients. However, to our knowledge, no study has applied Transformer to CRC lesion segmentation in contrast-enhanced abdominal CT images. This study aimed to explore Transformer' s potential and clinical value for CRC segmentation.

Our dataset comprised contrast-enhanced abdominal CT images from 2,180 CRC patients acquired using different scanner brands and models across three independent medical centers. With a substantial training set (n=1,159) and significant heterogeneity, the data well reflect real clinical scenarios. We proposed a novel CNN-Transformer hybrid TransUNet-Cascade model, compared its performance with 3D nnU-Net, and validated generalization capability through an independent external test set. Results showed that both TransUNet-Cascade and 3D nnU-Net achieved superior tumor segmentation on arterial phase images compared with venous phase. CRC contains abundant neovascularization with defective endothelial structure, incomplete smooth muscle, and pericyte coverage, leading to significantly increased permeability. During the arterial phase, tumor tissue rapidly takes up contrast agent and reaches peak enhancement, creating optimal contrast between tumor and normal bowel wall that facilitates segmentation.

Comparative analysis revealed that TransUNet-Cascade effectively identified and segmented CRC lesions, achieving average DSC, F1, HD95, IoU, PRE, and REC values of 0.740, 0.839, 34.084, 0.656, 0.737, and 0.688, respectively, on arterial phase images in the external test set, overall outperforming 3D nnU-Net. This advantage may be attributed to TransUNet-Cascade' s architectural design and its ability to capture multi-scale features. CRC lesions are distributed across different abdominal locations with complex spatial relationships to surrounding tissues and organs. While 3D nnU-Net excels at local feature extraction through adaptive hyperparameter and architecture optimization, it lacks sufficient global context modeling, potentially leading to inaccurate boundary and spatial distribution segmentation when contrast between lesions and surrounding tissue is low. In contrast, TransUNet-Cascade combines the strengths of CNNs and Transformers: CNNs effectively extract local detail features, while Transformer modules capture long-range dependencies through self-attention mechanisms, better understanding spatial distributions between lesions and surrounding organs. This balance between local details and global context enables more effective handling of CRC' s complex morphology and blurred boundaries, thereby improving segmentation accuracy.

Further analysis by tumor location showed that TransUNet-Cascade performed best on right-sided colon cancer (DSC, F1, HD95, IoU, PRE, and REC of 0.730, 0.833, 36.539, 0.650, 0.728, and 0.757), followed by left-sided colon cancer and transverse colon cancer, while rectal cancer segmentation was relatively poor. This pattern is consistent with our previous 3D nnU-Net study and may be related to anatomical characteristics of different colorectal segments. Right-sided colon is located in the right abdomen with relatively fixed anatomy and high surrounding tissue contrast. In contrast, the rectum lies within the pelvis with complex anatomy adjacent to multiple organs (bladder, prostate, uterus, etc.), and smaller density differences among fat, muscle, and bone tissues in the pelvis result in lower contrast between rectal cancer and surrounding tissue on CT images, increasing segmentation difficulty. Although TransUNet-Cascade employs a cascaded structure with multi-stage segmentation strategy for progressive optimization, the unclear lesion boundaries and complex spatial relationships in rectal cancer may limit the model' s ability to capture sufficient detail information, leading to reduced segmentation accuracy.

This study has several limitations. First, model diversity is limited: only 3D nnU-Net and TransUNet-Cascade were used for CRC segmentation, without comparison to other advanced DL models, which may restrict the comprehensiveness and generalizability of our findings. Second, rectal cancer segmentation performance was suboptimal, likely due to complex anatomy, blurred boundaries, and low contrast on CT images. The model' s capability for handling low-contrast and complex anatomical structures requires further optimization. Third, despite inclusion of data from three independent centers with different scanner models, the dataset sample size remains limited, particularly for certain tumor locations (e.g., transverse colon cancer), which may affect segmentation performance for these subsets.

In summary, the TransUNet-Cascade model, by combining CNN and Transformer advantages, employing a cascaded architecture, and enhancing global context modeling, outperforms traditional 3D nnU-Net for CRC segmentation. These results provide a new technical approach for precise CRC segmentation and show promise for future clinical application.

Author Contributions: Mu Huang, Daochun Zhang, and Weicui Chen conceived the study and designed the methodology. Mu Huang developed and validated the deep learning models. Wei Yang and Liming Zhong provided methodological support. Wenjing Yuan and Ziqi Jia collected and curated the data. Xiangliang Tan and Xiaohui Duan provided external validation data. Xian Liu performed statistical analysis. Weicui Chen revised the final manuscript and is responsible for the overall content.

Funding: National Natural Science Foundation of China (82202259); 13th Chaoyang Talent Project of Guangdong Provincial Traditional Chinese Medicine Hospital (ZY2022YL05)

ORCID IDs:
Mu Huang: https://orcid.org/0009-0007-9289-2366
Weicui Chen: https://orcid.org/0000-0002-1814-8295

References

[1] HAN B F, ZHENG R S, ZENG H M, et al. Cancer incidence and mortality in China, 2022[J]. J Natl Cancer Cent, 2024, 4(1): 47-53. DOI:10.1016/j.jncc.2024.01.006.
[2] ZHENG R S, CHEN R, HAN B F, et al. Analysis of cancer incidence and mortality in China, 2022[J]. Chinese Journal of Oncology, 2024, 46(3): 221-231. DOI:10.3760/cma.j.cn112152-20240119-00035.
[3] Medical Administration Department of National Health Commission, Chinese Society of Oncology, Chinese Medical Association. Chinese guidelines

for diagnosis and treatment of colorectal cancer (2023 edition)[J]. Medical Journal of Peking Union Medical College Hospital, 2023, 14(4): 706-733. DOI:10.12290/xhyxzz.2023-0315.
[4] LI S L, YUAN L, LU T, et al. Deep learning imaging reconstruction of reduced-dose 40 keV virtual monoenergetic imaging for early detection of colorectal cancer liver metastases[J]. Eur J Radiol, 2023, 168: 111128. DOI:10.1016/j.ejrad.2023.111128.
[5] YAO L S, LI S Y, TAO Q, et al. Deep learning for colorectal cancer detection in contrast-enhanced CT without bowel preparation: a retrospective, multicentre study[J]. EBioMedicine, 2024, 104: 105183. DOI:10.1016/j.ebiom.2024.105183.
[6] MIAO S D, SUN M Z, ZHANG B B, et al. Multimodal deep learning: tumor and visceral fat impact on colorectal cancer occult peritoneal metastasis[J]. Eur Radiol, 2025. DOI:10.1007/s00330-025-11450-2.
[7] WANG X, SHI C, YUAN Z Y. Fused regional attention UNet model for segmentation of knee cartilage in MRI[J]. Chinese Journal of Medical Imaging Technology, 2024, 40(5): 764-768. DOI:10.13929/j.issn.1003-3289.2024.05.027.
[8] ZHENG K Y, WU H, YUAN W J, et al. Feasibility study of automatic colorectal cancer segmentation from abdominal CT images using 3D nnU-Net deep learning network[J]. Chinese Journal of Radiology, 2024, 58(8): 829-835.
[9] XIAO L, ZHANG L, TANG Y, et al. Segmentation of thoracic and lumbar vertebrae in sagittal spine X-ray images using dual attention U-Net network based on global and contextual information[J]. Chinese Journal of Medical Imaging Technology, 2025, 41(1): 128-132. DOI:10.13929/j.issn.1003-3289.2025.01.027.
[10] PU Q M, XI Z X, YIN S, et al. Advantages of transformer and its application for medical image segmentation: a survey[J]. Biomed Eng Online, 2024, 23(1): 14. DOI:10.1186/s12938-024-01228-6.
[11] Medical Imaging Big Data and Artificial Intelligence Working Committee of Chinese Society of Radiology, Abdominal Group of Chinese Society of Radiology, Magnetic Resonance Group of Chinese Society of Radiology. Expert consensus on CT and MRI annotation of colorectal cancer (2020)[J]. Chinese Journal of Radiology, 2021, 55(2): 111-116. DOI:10.3760/cma.j.cn112149-20200706-00894.
[12] CARIN L, PENCINA M J. On deep learning for medical image analysis[J]. JAMA, 2018, 320(11): 1192-1193. DOI:10.1001/jama.2018.13316.
[13] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[M]//Medical Image Computing and Computer-Assisted Intervention -MICCAI 2015. Cham: Springer International Publishing, 2015: 234-241. DOI:10.1007/978-3-319-24574-4_{28}.
[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017: 5998-6008.
[15] SHAMSHAD F, KHAN S, ZAMIR S W, et al. Transformers in medical imaging: a survey[J]. Med Image Anal, 2023, 88: 102802. DOI:10.1016/j.media.2023.102802.

[16] GHAZOUANI F, VERA P, RUAN S. Efficient brain tumor segmentation using Swin transformer and enhanced local self-attention[J]. Int J Comput Assist Radiol Surg, 2024, 19(2): 273-281. DOI:10.1007/s11548-023-03024-8.
[17] SUN H, MO G P, XU G H, et al. Segmentation of spine MRI using fused Vision Transformer and edge-guided encoder-decoder network (RET-Net) algorithm[J]. Chinese Journal of Medical Imaging Technology, 2023, 39(4): 577-581. DOI:10.13929/j.issn.1003-3289.2023.04.021.
[18] LIU L B, LI X, WEI B Z, et al. Multi-scale edge optimization algorithm for bladder cancer magnetic resonance imaging segmentation based on Swin Transformer[J]. Journal of Biomedical Engineering Research, 2023, 42(1): 43-49. DOI:10.19529/j.cnki.1672-6278.2023.01.07.

Received: March 14, 2024; Revised: June 18, 2025
Edited by: Jia Mengmeng

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.