# Research and Implementation of Colorectal Cancer Tumour Segmentation Algorithm based on DRA-UNet

Yuqian Wang [1, a], Aksyonov Sergey Vladimirovich [1], Qilun Li [2]

[1] Tomsk Polytechnic University, Tomsk, Ghent Oblast, Russia

[2] Tomsk State University, Tomsk, Russia

[a] 578985030@qq.com

**Abstract.** Fully convolutional neural networks often suffer from information loss and suboptimal accuracy in medical image segmentation tasks. To mitigate these issues for rectal tumor analysis, we propose an improved U-Net architecture, designated as Deep Residual Attention U-Net (DRA-UNet). First, deep residual modules replace the original convolutional blocks. This substitution enhances feature extraction capabilities while mitigating network degradation. secondly, an attention mechanism is added between each jump connection of the U-Net to focus attention on the features useful for segmentation and suppress redundant features; finally, a DiceLoss loss function is used to alleviate the class imbalance problem. Experimental validation was performed on a colorectal cancer CT image dataset, with segmentation performance quantified using the Dice coefficient. The experimental results demonstrate that the proposed method achieves a segmentation accuracy of 91.38% for colorectal cancer tumors. This represents a significant improvement over existing models, with performance gains of 9.08%, 6.61%, and 4.73% compared to the FCN, U-Net, and Dense-UNet frameworks, respectively. These comparative outcomes confirm the effectiveness and superiority of our approach.

**Keywords:** U-Net; Attentional Mechanisms; Depth Residual Structure; Rectal Tumor.

## 1. Introduction

Colorectal cancer refers to a malignant growth that develops between the dentate line and where the rectum meets the sigmoid colon, making it one of the most prevalent cancers affecting the digestive system[1]. This deadly disease has become a global health crisis, with both its occurrence and death toll climbing steadily while affecting increasingly younger populations—currently standing as the fifth most common cancer worldwide. What's particularly concerning is how this illness has been striking people at younger ages and claiming more lives with each passing year[2]. Colorectal cancer usually progresses slowly and can be prevented if rectal tumors are detected and removed before they develop into cancer[3]. Therefore, localization and identification of rectal tumors is the basis and key to the diagnosis and treatment of colorectal cancer[4]. Generally, the physician views the patient's computed tomography (CT) image to diagnose the presence of tumors and potential colorectal cancer, if present, to mark the area of rectal tumor for surgery[5]. Identifying rectal tumors on CT scans can be quite tricky since they often don't stand out clearly, necessitating the sharp eye of a seasoned medic to pinpoint and demarcate the affected area. When less experienced or inattentive clinicians handle these cases, there's a real risk of either overlooking or incorrectly diagnosing such growths. This predicament underscores the critical need for developing robust algorithms capable of automatically segmenting rectal tumors. These automated systems would not only support medical practitioners in their diagnostic process but also alleviate their workload, enhance the precision of colorectal cancer detection, and ultimately buy patients precious time in receiving appropriate treatment.

Over the past decade, deep learning techniques have made significant inroads across various computer vision applications, ranging from image processing to object detection. The success of these approaches can largely be attributed to Convolutional Neural Networks (CNN), first introduced to image classification by LeCun et al. [6], which quickly stole the spotlight in the machine learning community. Unlike conventional segmentation methods, their deep learning counterparts possess the

unique ability to independently adapt to specific tasks while delivering impressive performance [7]. Deep learning methods have also been widely used in rectal tumor segmentation, such as a region-growth-based colorectal cancer tumor segmentation algorithm proposed in the literature[8], but the algorithm requires human to give the intensity signal of the tumor region during the segmentation process, and the whole process also requires human-computer interaction, and it is easy to cause over-segmentation. A segmentation algorithm for colorectal cancer tumor level set was proposed in the literature[9], which is insensitive to the boundary, and the segmentation effect is not good when the boundary exists ambiguous situation. According to the literature[10], researchers introduced a Fully Convolutional Network (FCN) approach for the automatic segmentation of colorectal cancer tumors in MRI scans. This technique leverages a pre-trained ResNet50 model as its backbone for feature extraction, while incorporating three specialized edge output modules within the network's hidden layers to capture multi-scale information from the medical images. Ultimately, the segmentation results are generated by fusing these three edge outputs, providing a comprehensive analysis of the tumor boundaries. In 2015, Ronneberger et al. [11]introduced the U-Net architecture, which features a symmetric encoder-decoder design. The encoder pathway progressively compresses the input image through a series of convolutional and downsampling operations, thereby capturing hierarchical feature representations. Conversely, the decoder pathway utilizes upsampling and convolution to reconstruct the encoded features into a full-resolution binary segmentation mask. This process ultimately produces the final pixel-wise classification output. The literature introduces an algorithm reminiscent of the 2D U-Net model for rectal tumor segmentation, which hits the ground running with a two-stage training approach to boost efficiency. By capitalizing on 3D image data and feeding the network a 5-channel tensor, this method really hits the nail on the head when it comes to optimizing performance[12]. Although the above model achieves good segmentation results in rectal polyp segmentation, it is not ideal for fast segmentation of fine tumors. Specifically, the segmentation accuracy for rectal tumor regions requires further enhancement. To overcome these limitations, this paper introduces a CT image segmentation method for colorectal cancer that builds upon an improved U-Net architecture.

## 2. DRA-UNet Network

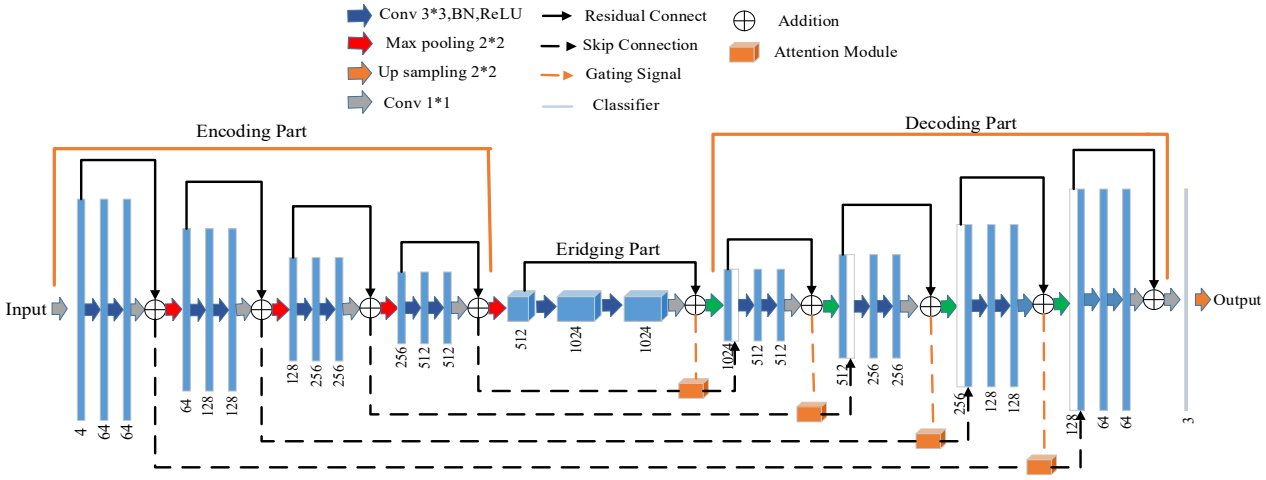



**Fig 1.** DRA-UNet network structure

Based on the U-Net architecture, this paper introduces several key enhancements to improve its performance. First, the original convolutional blocks in both the encoder and decoder are replaced with deep residual modules. This modification serves a dual purpose: it simplifies the training process and mitigates the issue of gradient degradation. Furthermore, an attention mechanism is incorporated into the U-Net framework[13]. This addition enables the network to selectively concentrate on diagnostically relevant features while filtering out non-essential information. Consequently, these

structural improvements collectively enhance the model's segmentation accuracy for rectal tumor images.

Fig. 1 illustrates the DRA-UNet network architecture, which is composed of five key components: the Encoding Section, Decoding Section, Bridging Section, Classifier, and Jump Connection. The first three sections—Encoding, Decoding, and Bridging—are implemented using one or more deep residual modules that feature a specific structure. Each module includes an input layer, followed by a Batch Normalization layer, a ReLU activation function, two separate $3\times3$ convolution operations, a constant mapping unit, and finally the output layer.

The coding part contains 4 depth residual modules and a maximum pooling layer, in which the depth residual modules are used to extract the shallow semantic information of the feature map through the residual connection, and the maximum pooling layer is used for the down-sampling operation of the image to reduce the size of the feature map, and the number of channels of the feature map increases as the size decreases. The decoding section contains 4 deep residual modules and upsampling operation to upsample the feature map from the encoding section. The bridging part is used to connect the encoding part and the decoding part. The classifier module is constructed with two key components: a $1\times1$ convolutional layer followed by a Sigmoid activation layer. The initial $1\times1$ convolution serves to reduce the channel depth of feature maps generated by the decoder. Subsequently, the Sigmoid activation function computes per-pixel class probabilities across these refined feature maps. This sequential processing ultimately transforms the multi-channel feature representations into the final segmentation output. Through skip connections, shallow features from the encoder are directly transferred to corresponding deeper layers in the decoder. This architectural design enables feature map fusion across different network depths. However, the initial feature representations extracted in early encoder stages typically contain limited semantic information. This characteristic introduces substantial redundant data into the fusion process, which may adversely impact the quality of segmentation results.

This paper presents an attention mechanism designed to filter out feature responses in irrelevant areas while minimizing redundant features prior to concatenating and merging the encoding features with their counterparts in the decoding phase.

### 2.1 Depth Residual Module

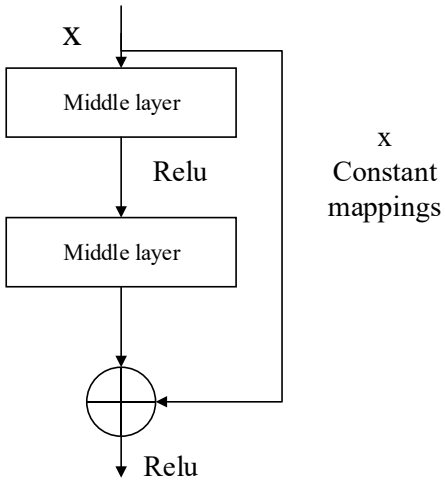
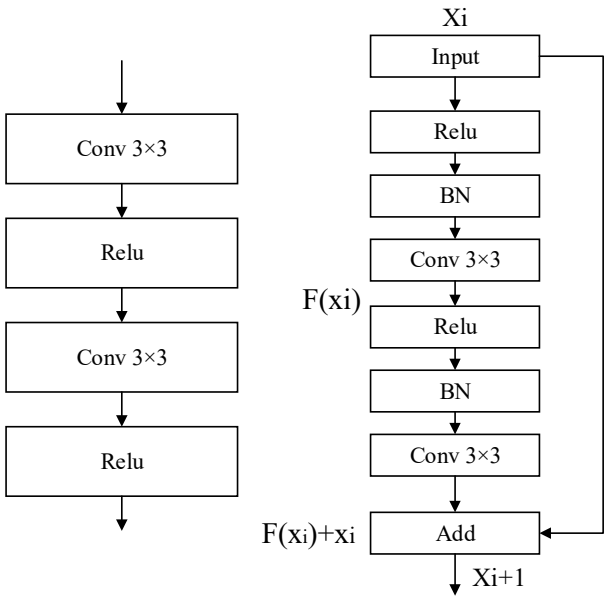



**Fig 2.** Residual unit

In computer vision tasks, network depth significantly impacts model performance. Conventional wisdom suggested that deeper architectures yield superior recognition and classification results. In computer vision tasks, network depth significantly impacts model performance. Conventional wisdom suggested that deeper architectures yield superior recognition and classification results. To

resolve this fundamental dilemma, He et al[14] developed the residual network, which consists of multiple residual modules connected together.

As shown in Fig. 2, consider a segment of a neural network where the input is $x$ and the desired underlying mapping is $H(x)$. A residual unit simplifies learning by introducing an identity mapping, or shortcut connection, which bypasses one or more layers. This connection allows the input xx to be directly added to the output of the layered transformations. Consequently, the network is not required to learn the complete mapping $H(x)$, but instead learns a residual function $F(x)=H(x)-x$. In this formulation, $x$ and $H(x)$ represent the input and output vectors of the considered layers, respectively. The residual formula in the network can be expressed as: $y=w_s x+F(x,\{w_i\})$, where $x$ and $y$ are the input and output of this network, $w_x$ is the convolution operation. The function $F(x,\{w_i\})$ denotes the residual mapping to be learned, typically implemented through a series of operations such as convolution, batch normalization, and a non-linear activation function (e.g., ReLU). For a basic two-layer residual block, this can be expressed as $F(x,\{w_i\})=W_2\sigma(W_1 x)$, where W represents the convolutional weights and σ is the activation function. when the dimensionality of the feature map is different, the zero-filling method can be used to increase the dimensionality, or the $1\times1$ convolution operation can be used to achieve consistent dimensionality.




(a) Standard Convolution Module (b) Depth Residual Module

**Fig 3.** Comparison of standard convolution module and depth residual module structures

Conventional convolutional neural networks leverage multiple convolutional operations to extract features from images, as illustrated in Fig. 3(a). Nevertheless, as the network architecture becomes increasingly deep, training the model turns into a challenging endeavor. In our work, we adopt the concept of residual connections and substitute the original convolutional layers within both the encoder and decoder components of U-Net with an in-depth residual module, depicted in Fig. 3(b). By incorporating a direct pathway that enables input data to bypass intermediate layers and reach subsequent network sections, we effectively mitigate the gradient degradation problem that typically arises during network training. This approach facilitates smoother information flow throughout the network, thereby addressing the issue of information loss commonly encountered in traditional architectures to a considerable degree.

### 2.2 Attentional Mechanisms

The attention mechanism module completes the following equations:

$$F' = \alpha_i \otimes F_x \tag{1}$$

Where $F_x \in R^{C_x \times H_x \times W_x \times D_x}$ is the input feature map from the encoded part of the network structure in Fig. 1, $F' \in R^{C_x \times H_x \times W_x \times D_x}$ idenotes the corresponding output after attention enhancement. The dimensions of this feature map are defined by its channel count $C$, height $H$ and width $W$. The attention mechanism operates by performing an element-wise multiplication (denoted by $\otimes$) between $F_x \in R^{C_x \times H_x \times W_x \times D_x}$ and a learned attention coefficient. This coefficient, which assumes values between 0 and 1, functions to accentuate regions contain $\alpha_i$ ing critical target information while suppressing features irrelevant to the segmentation task. The computation of the attention coefficient $\alpha_i$ is defined by the following formula:

$$a_i = \sigma_2\left(q_{att}\left(F_{xi}, F_{gi}, \theta_{att}\right)\right) \tag{2}$$

where $\sigma_2$ is the Sigmoid activation function, $q_{att}$ is the additive attention calculation, and the attention operation is characterized by a set of parameters $\theta_{att}$, including the convolution op erations $W_g \in R^{C_g \times C_{\text{int}}}$, $W_x \in R^{C_x \times C_{\text{int}}}$, $\Psi \in R^{C_{\text{int}} \times 1}$ and the bias terms at the corresponding positions, $b_g \in R^{C_{\text{int}}}$, $b_\Psi \in R$. $q_{att}$ is calculated as follows:

$$q_{att} = \Psi^T\left(\sigma_1\left(W_x^T F_{xi} + W_g^T F_{gi} + b_g\right)\right) + b_\Psi \tag{3}$$

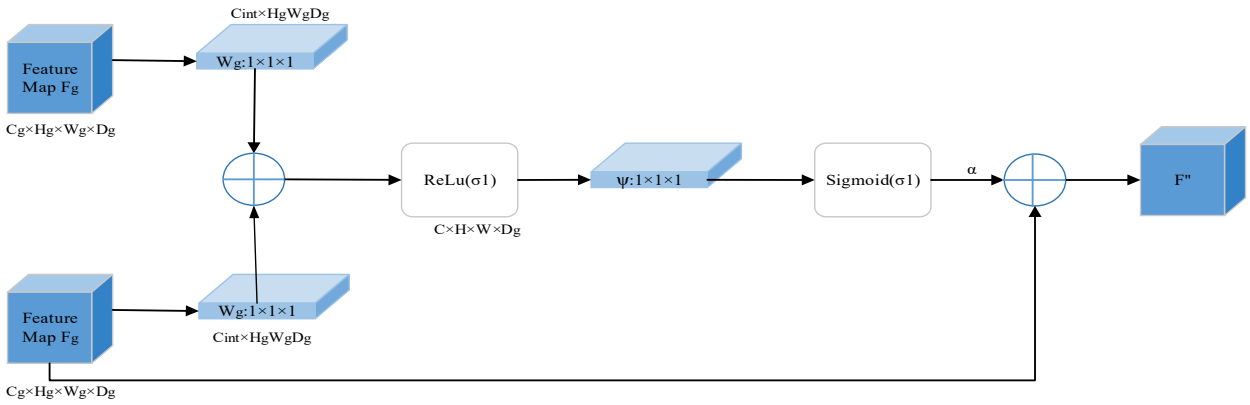



**Fig 4.** Door Control Attention Module

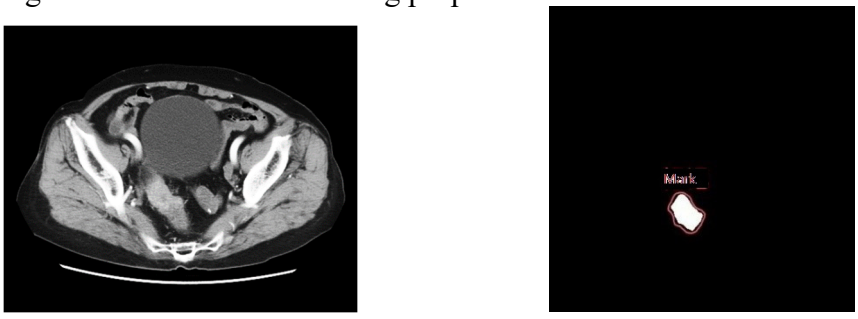
As shown in Fig. 4, $F_g \in R^{C_g \times H_g \times W_g \times D_g}$ is a gating signal ( Gating Signal) from the feature vector of the layer before the upsampling corresponding to the encoding part, which contains contextual information. The input features $F_g$ and $F_x$ are linearly transformed along the channel direction using a $1 \times 1 \times 1$ conventional layer. This operation projects them along the channel dimension. Subsequently, the transformed features are combined using an element-wise summation. Throughout this process, the channel depth of feature $F_g$ is progressively reduced from $C_g$ to $C_{\text{int}}$, and finally to a single channel. Similarly, the channel count for feature $F_x$ is adjusted from $C_x$ to $C_{\text{int}}$, and then again to 1. A ReLU (Rectified Linear Unit) activation function is applied after these operations to increase the network's nonlinear representational capacity. Since $F_g$ contains richer feature information, the additive attention operation enables $F_x$ to learn the gap between the feature information and $F_g$, thus

focusing attention on the target region and suppressing the activation of non-target region information, which improves the segmentation ability of the network.

## 3. Experimental Results and Analysis

### 3.1 Experimental Dataset

This study utilized abdominal CT imaging data from colorectal cancer patients treated between 2018 and 2021. The dataset was obtained from the Upper Radiology Department of the First Affiliated Hospital of Henan University of Science and Technology in Luoyang, Henan Province. The data collection process fully complied with ethical standards and protected all patients' personal information throughout the research. In this paper, the acquired CT images of colorectal cancer are two-dimensional images. Compared with three-dimensional CT images, the study of two-dimensional CT images will be more conducive to assisting diagnosis. Moreover, the research technology of two-dimensional images is more mature than three-dimensional images, and it is better than three-dimensional models in terms of model generalisation ability. The dataset comprises portal venous phase CT images from 75 colorectal cancer patients and 30 normal abdominal CT scans. The patient cohort includes 48 males and 27 females, with ages ranging from 36 to 92 years. A significant majority (70.67%) of patients fall within the 50-80 age range, totaling 53 individuals. In addition, of the 75 patients, 59 (78.67 per cent) had colon cancer and 16 (21.33 per cent) had rectal cancer. Among the colon cancers, there were 28 cases of sigmoid colon cancer, accounting for 47.45% of the total number of colon cancers. CT images of patients' colorectal cancer foci areas mostly showed irregular thickening of the bowel wall, and patients' colorectal cancer foci areas occurred in the rectum, sigmoid colon, ascending colon, hepatic flexure of the colon, transverse colon, descending colon, and sigmoid colon and the junction of rectal cancer, among which the foci areas occurred in the rectum and sigmoid colon with a high proportion of the foci areas, which accounted for 37.34% and 21.34% of the total patients, and foci areas showed a diversity. regions showed diversity. The dataset in this paper contains 3057 CT images and corresponding mask images. All images in the dataset have a resolution of 512×512 pixels. For the experimental setup, 2,700 images were randomly selected as the training set, while the remaining 357 images were allocated for testing purposes.




(a) Transformation and normalization (b) mask

**Fig 5.** Mask mark

The segmentation data set needs to use the mask containing the target area as the original image label, and use the annotation software to manually label the nodule position to generate the mask image. There are only 2 categories in the mask, 0 and 1, respectively. 1 represents the pixel where the nodule is located, and 0 represents the pixel where the nodule is located. The white area circled in Fig. 5(b) is the shape generated according to the position of the nodule. The dataset was further categorized based on tumor size, containing 1,194 images of small tumors, 1,125 of medium tumors, and 738 of large tumors.

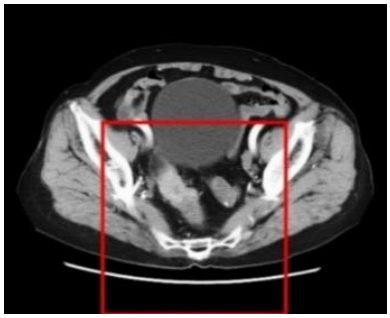


**Fig 6.** Schematic diagram of CT image cropping

Since there are fewer training set images after medical image division, it is far from enough for deep learning. Therefore, it is necessary to expand the sample size of the training set through image enhancement technology. To enhance sample diversity and improve model robustness, this study employs several data augmentation techniques during the training phase, as illustrated in Fig. 6. The original dataset was expanded through transformations including image flipping, random cropping, and elastic deformations.

### 3.2 Evaluation Index and Loss Function

To quantitatively assess tumor segmentation performance, this study employs two key evaluation metrics. The first metric is the Dice Similarity Coefficient, which measures the spatial overlap accuracy between segmented and ground truth regions. The second metric evaluates the precision of tumor boundary delineation.

Semantic segmentation is fundamentally a task of classifying each pixel in an image. For this purpose, the cross-entropy loss function has been widely adopted as the standard choice. However, a significant challenge in medical imaging is the frequent class imbalance between target structures and background. The conventional cross-entropy loss is inadequate in handling this issue effectively. To overcome this limitation, our work employs the Dice Loss function as a replacement for the traditional cross-entropy loss.

$$L_{Dice} = 1 - \sum_{k=1}^{K} \frac{2w_k \sum_{i=1}^{N} p(k,i) g(k,i)}{\sum_{i=1}^{N} p(k,i) + \sum_{i=1}^{N} g(k,i)} \tag{4}$$

In this context, N denotes the total pixel count, while K represents the number of semantic categories. The value of K is set to 2, corresponding to the lesion area and the background. $P_{(k,i)} \in [0,1]$ represents the probability that the pixel is predicted to be a category, and $g_{(k,i)} \in \{0,1\}$ represents the label value of the pixel i belonging to category k.

The performance of the proposed segmentation method is evaluated using three standard metrics: the Dice Similarity Coefficient (DSC), Precision, and Recall. The formulas for calculating these metrics are provided below.

$$DSC = \frac{2TP}{FP + 2TP + FN} \quad \Pr ecision = \frac{TP}{TP+FP} \quad \mathrm{Re} call = \frac{TP}{TP + FN} \tag{5}$$

The Dice coefficient serves as a metric to quantify the similarity between the model's predictions and the ground truth labels. In this context, True Positives (TP) represent the count of positive samples that were correctly identified. Conversely, False Positives (FP) occur when negative samples are erroneously classified as positive, while False Negatives (FN) refer to positive samples that are incorrectly predicted as negative. Based on these core components, the precision metric is defined as the proportion of correctly predicted tumor pixels out of all pixels the model identified as tumor. Meanwhile, the recall metric is calculated as the ratio of correctly predicted tumor pixels to the total number of actual tumor pixels present in the data.

### 3.3 Experimental Environment

Experimental platform hardware configuration: Intel(R) Core(TM) i5-9300H CPU @ 2.40GHzCPU; 16 GB memory; GTX1660Ti graphics card, 6G video memory; 64-bit Windows operating system. Software: Anaconda3 is selected as the development platform, and the network model architecture is based on the Keras environment of the Tensorflow background to realize the construction of the DRA-UNet network.

### 3.4 Results and Analysis

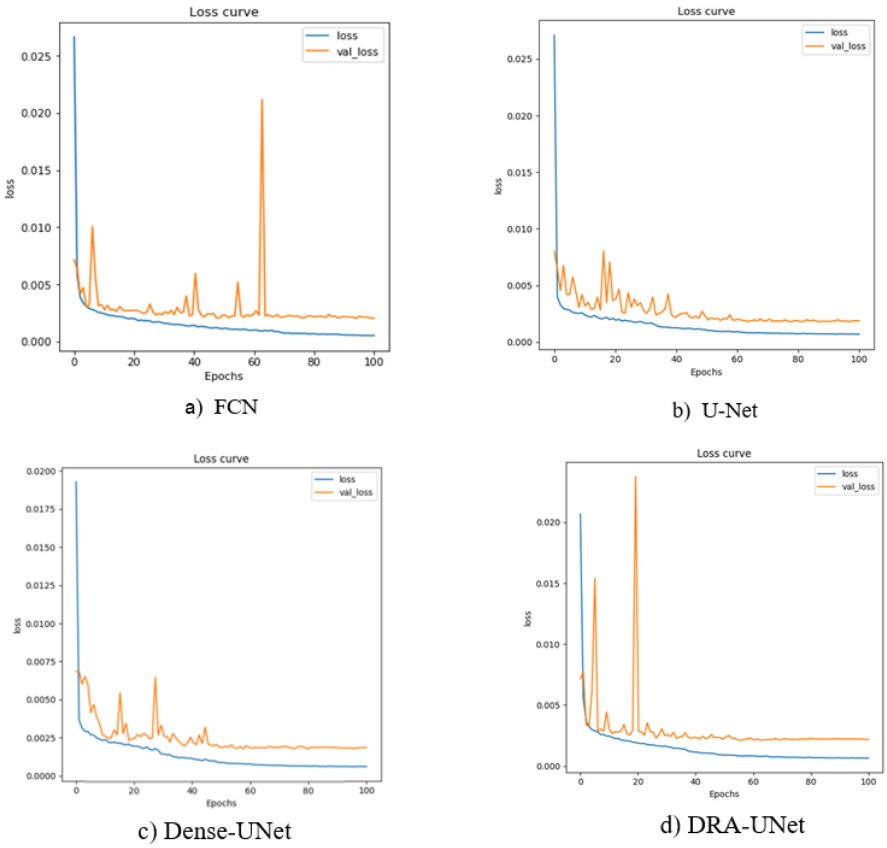



**Fig 7.** Loss function variations for the four network models

Fig. 7 presents a comparative analysis of the loss function curves for four network models—FCN, U-Net, Dense-UNet[15], and the proposed DRA-UNet—evaluated on both their training and validation sets. As seen in the figure, the loss profile of the validation set of DRA-UNet converges after 30 time periods, but the validation sets of the other three models converge less well. Therefore, compared with the three network models, FCN, U-Net and Dense-UNet, DRA-UNet can effectively solve the overfitting problem in classification.

To evaluate the performance of our enhanced model, we conducted a comparative analysis against three established networks: FCN, U-Net, and Dense-UNet. As summarized in Table 1, the FCN model yields the least favorable segmentation outcomes. This limitation stems from its architectural design, which underutilizes multi-level feature information. The FCN network primarily relies on upsampling the final convolutional features to the original image resolution, a process that overlooks the crucial spatial relationships between pixel localization and classification, consequently producing coarser results. In contrast, U-Net-based architectures, including U-Net and Dense-UNet, integrate skip connections to merge shallow and deep features, thereby capturing a richer set of characteristics. However, the shallow features from the encoder often contain limited semantic information and significant redundancy, which can adversely impact the final

segmentation quality. The method proposed in this paper achieves superior results across multiple evaluation metrics compared to the three benchmark models. This performance improvement can be primarily attributed to our strategic enhancements to the network structure.

**Table 1.** Comparison of segmentation performance of four models

| Model | Precision (%) | Recall (%) | Dice coefficient(%) |
|---|---|---|---|
| FCN | 82.30% | 79.14% | 80.85% |
| U-Net | 84.79% | 80.68% | 84.67% |
| Dense-UNet | 86.65% | 83.66% | 87.59% |
| DRA-UNet | 91.38% | 84.08% | 90.68% |

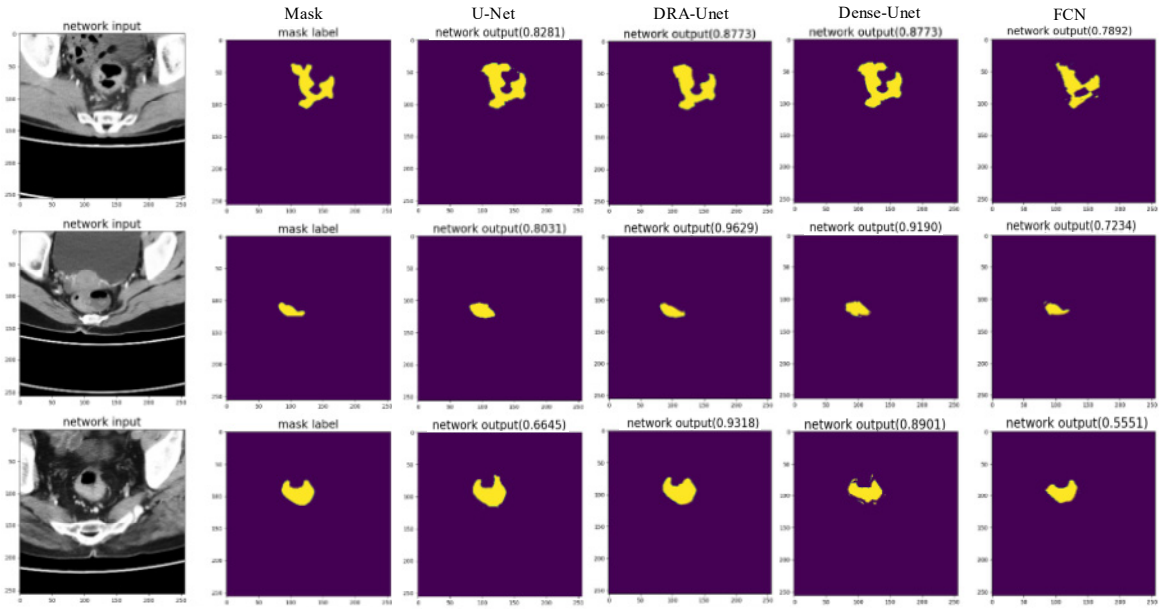



**Fig 8.** Graph of segmentation results for four models

Fig. 8 presents a visual comparison of colorectal cancer tumor segmentation results generated by four different methods. The qualitative analysis reveals that the FCN model produces suboptimal segmentation outcomes, primarily because it fails to effectively integrate multi-scale feature information. Consequently, this architectural limitation restricts the FCN approach to merely identifying the approximate tumor location rather than achieving precise boundary delineation. In terms of segmentation performance, both U-Net and Dense-UNet generate contour boundaries for rectal tumors that are relatively well-defined. However, their outputs are often marred by scattered noise and a lack of precision. When segmenting large tumor regions, FCN and U-Net demonstrate notably lower accuracy than the other two models. For medium-sized tumors, U-Net achieves better results than FCN. The most significant challenge lies in segmenting small tumors, which typically rely solely on subtle grayscale features and possess ambiguous boundaries. The proposed DRA-UNet model in this paper addresses this issue effectively, achieving superior accuracy for small tumor segmentation compared to the other three networks. Experimental results confirm that DRA-UNet produces segmentation maps that closely resemble the ground truth annotations. This demonstrates a clear performance enhancement and confirms the model's robust segmentation capability.

## 4. Summary

Accurate rectal tumor segmentation is crucial. It enables the early diagnosis of colorectal cancer. This paper introduces an improved U-Net model for segmenting rectal tumors in CT scans. Current methods like FCN and U-Net often face two main problems. They can be computationally complex

and lack accuracy in medical images. We made three key improvements to the network. First, we replaced standard layers with a deep residual module. This helps information flow better during training. Second, we added an attention mechanism to the skip connections. This lets the network focus on the most important features. Third, we used the Dice Loss function. It solves class imbalance by focusing on hard-to-outline areas. Our method achieves high accuracy for rectal tumor segmentation. The final accuracy reaches 91.38%. This result is a major improvement over older models. Our model is 9.08% more accurate than FCN. It also outperforms U-Net by 6.61% and Dense-UNet by 4.73%. These results prove our approach works well. We show that our method clearly improves segmentation accuracy. In the end, our model provides better performance.

## References


[1] F. A. Sinicrope, "Increasing incidence of early-onset colorectal cancer," New England Journal of Medicine, vol. 386, no. 16, pp. 1547-1558, 2022. DOI: [10.1056/NEJMra2200869] (https:// doi.org/ 10. 1056/ NEJMra2200869).

[2] R. L. Siegel, K. D. Miller, A. G. Sauer, et al., "Colorectal cancer statistics," CA: A Cancer Journal for Clinicians, vol. 70, no. 3, pp. 145-164, 2020. DOI: [10.3322/caac.21601] (https:// doi.org/ 10. 3322/ caac. 21601).

[3] F. T. Kolligs, "Diagnostics and epidemiology of colorectal cancer," Visceral Medicine, vol. 32, no. 3, pp. 158-164, 2016. DOI: [10.1159/000446488](https://doi.org/10.1159/000446488).

[4] J. Dong, T. S. Ma, Y. H. Xu, et al., "Characteristics and potential malignancy of colorectal juvenile polyps in adults: a single-center retrospective study in China," BMC Gastroenterology, vol. 22, no. 1, p. 256, 2022. DOI: [10.1186/s12876-022-02564-8](https://doi.org/10.1186/s12876-022-02564-8).

[5] M. Bretthauer, M. Løberg, P. Wieszczy, et al., "Effect of colonoscopy screening on risks of colorectal cancer and related death," New England Journal of Medicine, vol. 387, no. 17, pp. 1547-1556, 2022. DOI: [10.1056/NEJMoa2208375] (https://doi.org/10.1056/NEJMoa2208375).

[6] Y. Lecun, L. Bottou, Y. Bengio, et al., "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998. DOI: [10.1109/5. 726791] (https:// doi. org/ 10.1109/5.726791).

[7] K. Trebing, T. Stańczyk, and S. Mehrkanoon, "SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture," Pattern Recognition Letters, vol. 145, pp. 178-186, 2021. DOI: [10. 1016/ j. patrec. 2021.01.027] (https://doi.org/10.1016/j.patrec.2021.01.027).

[8] M. M. Van Heeswijk, D. M. J. Lambregts, J. J. M. Van Griethuysen, et al., "Automated and semiautomated segmentation of rectal tumor volumes on diffusion-weighted MRI: can it replace manual volumetry?" International Journal of Radiation Oncology Biology Physics, vol. 94, no. 4, pp. 824-831, 2016. DOI: [10. 1016/j.ijrobp.2015.12.017](https://doi.org/10.1016/j.ijrobp.2015.12.017).

[9] M. H. Soomro, G. Giunta, A. Laghi, et al., "Segmenting MR images by level-set algorithms for perspective colorectal cancer diagnosis," in Proceedings of the 2017 European Congress on Computational Methods in Applied Sciences and Engineering, 2017, pp. 396-406. DOI: [10.1016/j. phro. 2022. 05.001] (https://doi.org/10.1016/j.phro.2022.05.001).

[10] Z. Ran, J. M. Jian, M. M. Wang, et al., "Automatic segmentation method based on full convolution neural network for rectal cancer tumors in magnetic resonance image," Beijing Biomedical Engineering, vol. 38, no. 5, pp. 465-471, 2019.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234-241. DOI: [10.1007/978-3-319-24574-4_28](https: //doi. org/ 10.1007/978-3-319-24574-4_28).

[12] J. Wang, J. Lu, G. Qin, et al., "Technical note: A deep learning-based auto segmentation of rectal tumors in MR images," Medical Physics, vol. 45, no. 5, pp. 2560-2564, 2018. DOI: [10.1002/ mp. 12887] (https: //doi. org/10.1002/mp.12887).

[13] O. Oktay, J. Schlemper, L. L. Folgoc, et al., "Attention U-Net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018. [Online]. Available: [https://arxiv.org/ abs/1804. 03999] (https: // arxiv.org/abs/1804.03999).

[14] K. He, X. Zhang, S. Ren, et al., "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778. DOI: [10.1109/ CVPR. 2016. 90] (https://doi.org/10.1109/CVPR.2016.90).

[15] A. Kaku, C. Hegde, J. Huang, et al., "Darts: DenseUnet-based automatic rapid tool for brain segmentation," 2019. [Online]. Available: https://arxiv.org/abs/1911.05567.