ISSN: 2414-1895 DOI: 10.6919/ICJE.202601_12(1).0007

Medical Image Segmentation Using SegResNet Integrated with Non-local Neural Networks

Jun Liao[1], Jing Zhang[2], Tingjie He[3], and Yujie Zhang[4,*]

[1] School of Big Data and Artificial Intelligence, Chengdu Technological University, Chengdu 610000, China

[2] School of Computer Engineering, Chengdu Technological University, Chengdu 610000, China

[3] School of Computer Engineering, Chengdu Technological University, Chengdu 610000, China

[4] School of Big Data and Artificial Intelligence, Chengdu Technological University, Chengdu 610000, China

[*]Corresponding Author

Abstract

Pancreas segmentation is a challenging task in medical image analysis due to factors like ambiguous pancreatic boundaries and inter-individual variability. Current methods need improvement in balancing accuracy and efficiency. This study designs a lightweight pancreas segmentation model based on SegResNet, enhanced with Non-local Neural Networks. Trained on the Task07_Pancreas dataset, the model achieves a Dice coefficient of 91.51% and a Recall rate of 91.97% on the test set, with only 6.3M parameters. It outperforms advanced models such as UMambaEnc and UNETR, effectively balancing segmentation precision and computational efficiency.

Keywords

Pancreas Segmentation; SegResNet; Medical Image Segmentation; CT Imaging.

1. Introduction

The pancreas is a crucial organ in the human body, performing both digestive and endocrine functions. Pancreatic cancer, which originates from the epithelial cells of the pancreatic ducts or acinar cells, is a highly malignant tumor that presents significant challenges in diagnosis and treatment. In recent years, its global incidence has been increasing, and it shows a trend of affecting younger populations. The five-year survival rate is typically below 9% [1].

In clinical practice, abdominal CT imaging helps physicians examine pancreatic anatomy, identify tissue lesions, and evaluate their spatial correlation with surrounding vasculature. Manually delineating the pancreatic region in CT image sequences requires expertise and is subject to subjective experience. Given the limitations of manual segmentation, automated and precise pancreas segmentation has become a significant research area in medical image analysis. Early automated methods mainly used traditional techniques such as atlas registration [2], region growing [3], and simple linear iterative clustering [4]. These methods required manual feature design and parameter adjustment, were sensitive to image quality and noise, and lacked robustness against interference.

In recent years, the field of medical image segmentation has experienced revolutionary advancements, propelled by the emergence of Convolutional Neural Networks (CNNs) and Transformer architectures. Based on U-Net [5], researchers have [6] incorporated attention mechanisms into the skip connections, enabling the model to adaptively learn and assign weights to task-relevant feature regions. Models based on Transformers, such as SegFormer [7], utilize hierarchical Transformer

ISSN: 2414-1895 DOI: 10.6919/ICJE.202601_12(1).0007

encoders to produce multi-scale features. They make up for the shortcomings of CNNs in modeling long-range dependencies with their powerful ability to capture global context, offering a promising novel approach for pancreas segmentation.

2. Related Work

Reference [8] proposed the SMBC-Net, which utilizes SegFormer as the encoder and integrates three key modules: a Multi-scale Attention Module (MAM), a Boundary Awareness Module (BAM), and a Channel Aggregation Module (CAM). On the public Medical Segmentation Decathlon (MSD) dataset [9], this method achieved remarkable performance, with a pancreatic segmentation Dice of 80.34% and a tumor segmentation Dice of 69.95%.

The complex and variable morphology of the pancreas makes it difficult to characterize fully using single-scale features. Reference [10] incorporated a global feature module at the base of the U-Net architecture. Their Dice Similarity Coefficient reached 87.13% on the NIH dataset, representing an improvement of 7.43% over the original U-Net. Reference [11] addressed the issue of easily lost shallow feature information in the encoding path by proposing a dual downsampling module.

To alleviate severe imbalance between the target organ and the background from the very beginning, two-stage strategy has been extensively adopted. Self-attention mechanism of Transformers effectively addresses this drawback. The emergence of Vision Transformers has further expedited advancements in computer vision tasks [12]. A reference [13] introduced TransUNet, a hybrid CNN-Transformer architecture devised for multi-organ segmentation in medical images.

The nnU-Net [14], with its self-configuring approach, performs well across diverse datasets and is a prominent option for many segmentation benchmarks. However, the locality of Convolutional Neural Networks (CNNs) may limit their effectiveness in segmenting complex anatomical structures like the pancreas. To incorporate global information, Vision Transformers and their medical - imaging adaptations such as TransUNet and UNETR [15] have emerged.

3. Method

3.1 Dataset

This research uses the Task07_Pancreas dataset from the Medical Segmentation Decathlon (MSD). It includes annotated scan images of the pancreas and related tumors, with each case having labels for the pancreatic structure. Due to the small volume of the pancreas, its irregular morphology, and the sparse and subtle edges in images, the training faces a severe class-imbalance challenge. Dataset has 281 contrast-enhanced abdominal CT scans with manual pancreas annotations. 257 cases were for training and 24 for testing.

3.2 Network Architecture

This study utilizes the SegResNet architecture for the pancreas segmentation task. This network is derived from the 3D MRI brain tumor segmentation model [16]. It adopts a classic U-Net style encoder-decoder structure, which is optimized for medical image segmentation. SegResNet effectively resolves the vanishing gradient problem in deep network training by integrating residual connections, multi-scale feature fusion, and skip connections.

In the realm of medical image segmentation tasks, the segmentation accuracy of small target structures holds paramount importance for clinical diagnosis. To enhance the model's sensitivity towards small lesion regions, two cascaded Non-local Blocks are integrated [17] into the bottleneck layer of the SegResNet encoder-decoder architecture. This mechanism is capable of effectively capturing long-range dependencies within the feature maps. The network structure is depicted in Figure 1. Structure of the proposed SegResNet model.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202601_12(1).0007

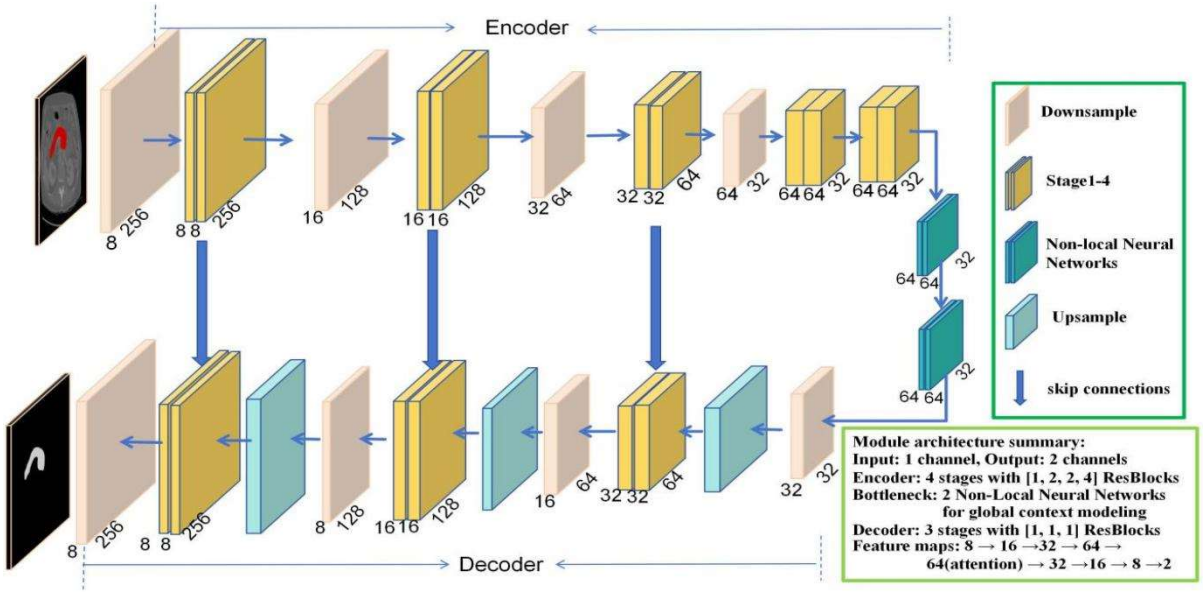


Figure 1. Structure of the proposed SegResNet model

The encoder consists of four downsampling stages, each having [1, 2, 2, 4] Residual Blocks (ResBlocks). Each ResBlock is made up of two convolutional layers with Group Normalization and ReLU activation, which helps reduce internal covariate shift. Downsampling is achieved via convolutional layers with a stride of 2, gradually decreasing spatial resolution and increasing the number of feature channels. This process builds a multi - scale feature pyramid, allowing the network to use both local details and global context.

The decoder has three upsampling stages, each with one ResBlock. Upsampling uses non-trainable layers for model reproducibility. In each stage, the channel dimension is adjusted 1×1 by convolution first, then upsampled by a factor 2×, and finally refined by a Residual Block. Skip connections between the encoder and decoder enable feature fusion, combining high-resolution spatial info from the encoder with high-level semantic info from the decoder. This reduces spatial info loss during upsampling and benefits accurate segmentation of organs with complex structures like the pancreas.

3.3 Non-local Neural Networks

Non-local Neural Networks can capture long-range dependencies within each feature map independently through Non-local computations. The core idea of the Non-local attention module is to establish global contextual connections by calculating correlations between any two positions in the feature map. An illustration of a Non-local Block is shown in Figure 2. Structure of the Non-local Block.

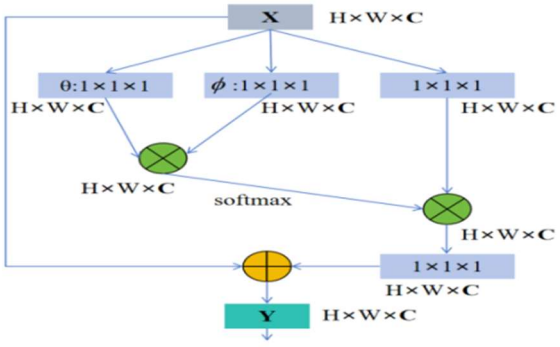


Figure 2. Structure of the Non-local Block

ISSN: 2414-1895 DOI: 10.6919/ICJE.202601_12(1).0007

Given an input feature map $X \in \mathbb{R}^{H \times W \times C}$, where C denotes the number of channels, and H and W represent the height and width respectively, the Non-local operation is defined as:

$$Y_i = X_i + W_z \sum_{j=1}^{N} \frac{f(X_i, X_j)}{C(X)} (W_v X_j) \tag{1}$$

Here, i and j are indices representing the output position and all possible input positions, respectively. The function $f(X_i, X_j)$ computes the affinity between positions i and j. The normalization factor is given by $C(X) = \sum_{j=1}^{N} f(X_i, X_j)$. X_i denotes the input feature vector at position i. W_v and W_z are learnable weight matrices. Y_i represents the output feature vector at position i.

In our implementation, we employ the embedded Gaussian function as the pairwise function:

$$f(X_i, X_j) = e^{\theta(X_i)^T \phi(X_j)} \tag{2}$$

where $\theta(X_i) = W_{\theta X_i}$ and $\phi(X_j) = W_{\phi X_j}$ represent the query and key embedding transformations, respectively.

4. Experimental Setup

4.1 Environment and Hyperparameters

All experiments were carried out within a Linux environment. The specific hardware and software configurations are presented in Table 1. Experimental environment configuration.

Table 1. Experimental environment configuration

GPU	CPU	Python	CUDA	Operating System
NVIDIA RTX 5060Ti	Intel Core i5-12600KF	3.10	12.9	Ubuntu20.4

4.2 Evaluation Metrics

To assess the segmentation performance of the proposed model, we predominantly utilized the following metrics: the Dice coefficient (Dice) and Recall. These metrics evaluate the accuracy and comprehensiveness of the segmentation results from diverse perspectives.

Recall measures the proportion of actual positive samples that are correctly identified as positive by the model. It is calculated as shown in Equation (3):

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

The Dice coefficient quantifies the spatial overlap between the ground truth and the predicted segmentation. Its calculation is given by Equation (4):

$$Dice = \frac{2TP}{FN + 2TP + FP} \tag{4}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202601_12(1).0007

The quantitative assessment model's capability in addressing severe category imbalance between small pancreatic regions and broad backgrounds was evaluated using Specificity, as shown in Equation (5):

$$Specificity = \frac{TN}{TN+FP} \tag{5}$$

In these formulas,TPrepresents pixels accurately predicted as positive (pancreas). FP denotes pixels wrongly predicted as positive (background misclassified as pancreas). FN signifies positive samples the model failed to detect (pancreas predicted as background). TN indicates pixels correctly predicted as negative (background).

5. Results and Discussion

To validate the effectiveness of Non-local Blocks, an ablation experiment was conducted. The test used 0, 1, 2, 3, and 4 modules respectively and evaluated Dice score and Specificity. The results in Table II show that increasing the number of cascaded Non-local Blocks improves Dice score, but the performance gains are limited when the number reaches 3 or 4. After evaluating model complexity and computational overhead, we adopted two cascaded Non-local Blocks as our final model to ensure operational efficiency while preserving robust feature representation.The SegResNet model achieved 99.98% specificity (Table 2), which reduces false positives, enhances target feature discrimination, suppresses background responses, and addresses class imbalance.

Table 2. Ablation Study Design

Model	Non-local Blocks number	Specificity(%)	Dice(%)	Recall(%)
SegResNet	-	99.98	87.83	88.84
	1	99.98	89.20	89.81
	2	99.98	91.51	91.97
	3	99.98	91.30	91.85
	4	99.98	91.78	92.11

Under the same experimental conditions, a series of comprehensive comparative experiments were carried out between the proposed method and several state-of-the-art medical image segmentation models, specifically UNETR, SwinUNETR [18], and UMambaEnc [19]. These models perform well in segmentation tasks, which helps with a fair performance assessment of the pancreas segmentation task. Table III presents the Dice coefficient, Recall, and the number of parameters of each model on the Task07_Pancreas dataset for 2D pancreas segmentation.

Table 3. Comparative results of different models

Method	Dice(%)	Recall(%)	Parameters(M)
UMambaEnc	91.3	91.6	39.26
UNETR	91.10	90.50	87.56
SwinUNETR	88.89	89.46	25.14
Ours	91.51	91.97	6.30

ISSN: 2414-1895 DOI: 10.6919/ICJE.202601_12(1).0007

The quantitative comparison presented in Table 3 indicates that the proposed model attains state of the art segmentation accuracy (91.51% Dice, 91.97% Recall) and features a lightweight structure (6.3M parameters). This superiority stems from a well considered architectural design. By incorporating cascaded Non-local Blocks into the bottleneck of SegResNet, a potent and compact global context modeling mechanism is established. In comparison with transformer based models (UNETR, SwinUNETR), the proposed approach circumvents the quadratic complexity of self - attention on high resolution feature maps and reduces the number of parameters by over 75%. When compared to UMambaEnc, the model exhibits slightly higher accuracy while using only 16% of its parameters, demonstrating superior parameter efficiency.The high Recall rate underscores the model's robustness against class imbalance, as the Non local Blocks modules enhance the salience of sparse pancreatic pixels against a complex background.

To comprehensively assess the proposed pancreas segmentation model's performance, the Precision - Recall curve was plotted . This curve shows the Precision - Recall trade - off in the segmentation results and suits medical image segmentation tasks with imbalanced sample distributions. As shown in Figure 3. Precision-Recall Curve, the AP value is 0.976, indicating that the model can effectively detect true pancreatic pixels, reduce misclassification of non pancreatic pixels, and accurately identify most pancreas pixels.

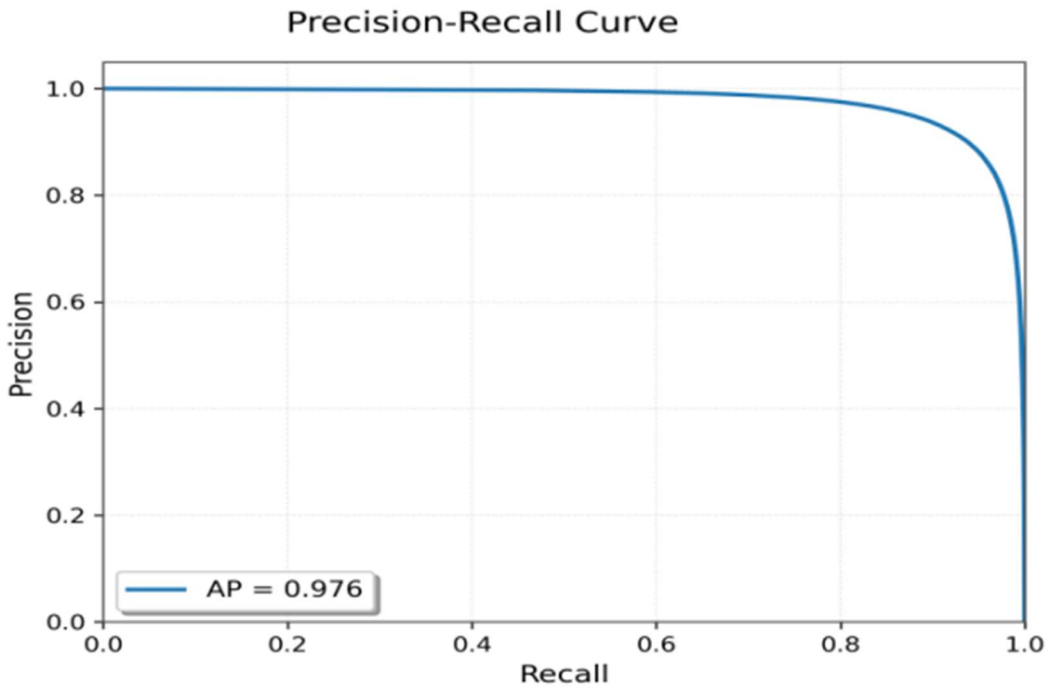
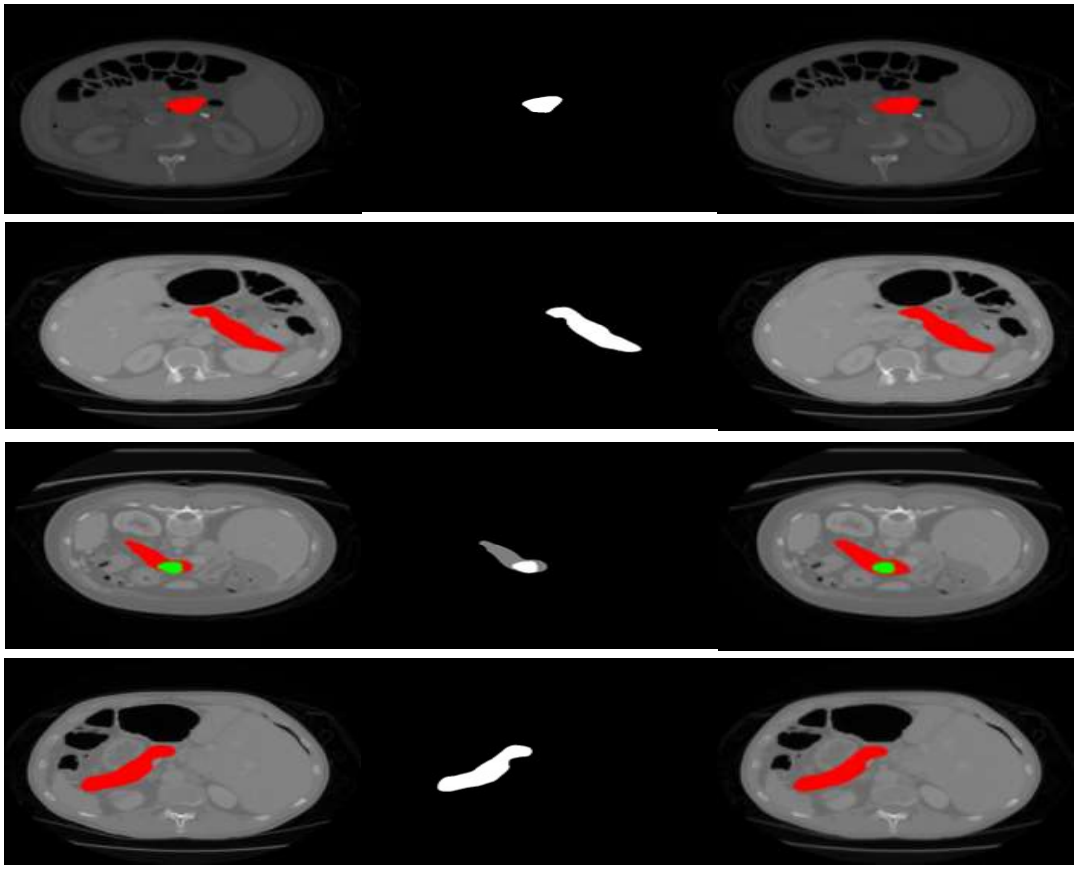


Figure 3. Precision-Recall Curve

Figure 4. Visualization of segmentation resultsdepicts a qualitative assessment of the pancreas segmentation outcomes achieved by the proposed model. The initial column (a) exhibits the original abdominal CT slices, revealing the intrinsic difficulties, including low contrast and the intricate anatomical context in the vicinity of the pancreas region. The second column (b) illustrates the corresponding ground-truth annotations, which were manually delineated by clinical experts. The third column (c) presents the segmentation predictions produced by the model.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202601_12(1).0007



(a) (b) (c)

Figure 4. Visualization of segmentation results

Visually, the predictions of our model exhibit remarkable consistency with the ground - truth annotations. The segmentation boundaries precisely demarcate the intricate morphological structures of the pancreas, encompassing its head, body, and tail. Notably, the model tackles the challenging aspects of pancreas segmentation: it sustains high precision at the interfaces between pancreatic tissue and adjacent organs and accurately identifies the slender, curved segments of the pancreatic tail. The consistent performance across slices validates the model's robustness in dealing with variations in the pancreas's shape and size. The segmentation results maintain anatomical continuity without significant over or under segmentation, demonstrating the effectiveness of integrating Non - local Blocks into the SegResNet architecture to capture long - range dependencies and enhance boundary awareness.

6. Conclusion

This research devised a lightweight pancreas segmentation model that balances accuracy and efficiency on a public dataset. Compared with existing state - of - the - art approaches, the proposed model, with only 6.3M parameters, achieved a Dice coefficient of 91.51% and a Recall of 91.97%, outperforming larger parameter models like UMambaEnc (39.26M) and UNETR (87.56M). Also, the analysis of the Precision - Recall curve showed a high AP of 0.976, indicating the model can effectively capture real pancreatic tissue and reduce false positives, which is crucial for clinical diagnostic reliability.

Acknowledgments

This research was supported by the project "Research on Medical Image Processing and Object Detection Based on Deep Learning" (Project No: 2025ZR018).

ISSN: 2414-1895 DOI: 10.6919/ICJE.202601_12(1).0007

References

[1] R.L. Siegel, K.D. Miller and A. Jemal: Cancer Statistics, 2019, Ca: A Cancer Journal For Clinicians, Vol. 69 (2019) No.1, p.7-34.

[2] K. Karasawa, M. Oda, T. Kitasaka, et al.: Multi-Atlas Pancreas Segmentation: Atlas Selection Based On Vessel Structure, Medical Image Analysis, Vol. 39 (2017), p.18-28.

[3] T.D. Tam and N.T. Binh: Efficient Pancreas Segmentation in Computed Tomography Based on Region-Growing, Proc. International Conference on Nature of Computation and Communication (Springer, 2015), p.332-340.

[4] A. Farag, L. Lu, B. Turkbey, et al.: A Bottom-Up Approach for Automatic Pancreas Segmentation in Abdominal CT Scans, Proc. Abdominal Imaging. Computational and Clinical Applications: 6th International Workshop (Springer, 2014), p.103-113.

[5] O. Ronneberger, P. Fischer and T. Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation, Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer, 2015), p.234-241.

[6] 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).

[7] E.Z. Xie, W.H. Wang, Z.D. Yu, et al.: SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers, Advances in Neural Information Processing Systems, Vol. 34 (2021), p.12077-12090.

[8] P.R. Liang, G.J. Xin and C.S. Ding: Pancreatic Image Segmentation Method Based on Improved SegFormer, Computer and Modernization, (2025) No.6, p.71-78.

[9] M. Antonelli, A. Reinke, S. Bakas, et al.: The Medical Segmentation Decathlon, ArXiv Preprint ArXiv:2106.05735, (2021).

[10] Z.T. Xiang, J.C. Liu, L. Wei, et al.: Pancreatic Image Segmentation Based on Global Feature U-Net, Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), Vol. 34 (2022) No.2, p.216-222.

[11] J.B. Ji, S. Chen and Y.Y. Yang: Pancreatic CT Segmentation Using Dual Dimensionality Reduction Channel Attention Gated U-Net, Chinese Journal of Biomedical Engineering, Vol. 42 (2023) No.3, p.281-288.

[12] B.Y. Zhou, G.J. Xin, H. Liang, et al.: Pancreatic Segmentation Based on Two-Stage Multi-Attention Mechanism Network, Computer and Modernization, (2025) No.10, p.67-72.

[13] H. Ma, Y. Liu and J.R. Zhang: Pancreatic Segmentation Based on Model Compression and Reconstruction U-Net, Computer Engineering and Design, Vol. 43 (2022) No.7, p.1998-2006.

[14] Information on: https://ai.googleblog.com/2020/12/transformers-for-image-recognitionat.html

[15] A. Hatamizadeh, Y. Tang, V. Nath, et al.: UNETR: Transformers for 3D Medical Image Segmentation, Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (2022), p.1748–1758.

[16] A. Myronenko: 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization, Proc. International MICCAI Brainlesion Workshop (Springer, Cham 2018), p.311-320.

[17] X. Wang, R. Girshick, A. Gupta, et al.: Non-Local Neural Networks, Proc. the IEEE Conference on Computer Vision and Pattern Recognition (2018), p.7794-7803.

[18] A. Hatamizadeh, V. Nath, Y. Tang, et al.: Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images, Proc. International MICCAI Brainlesion Workshop (Cham: Springer International Publishing, 2021), p.272-284.

[19] J. Ma, F. Li and B. Wang: U-Mamba: Enhancing Long-Range Dependency for Biomedical Image Segmentation, ArXiv Preprint ArXiv:2401.04722, (2024).