

Research on Automated Accurate Segmentation Algorithm of Double Kidney in Renal Dynamic Imaging Based on Improved UNet

Yujie Hao¹, Yuxing Zhang¹, Changbei Shi², Xiaorui Shi², Ping Zhu³, Shuling Zhou⁴

¹School of Electronic Information, Xijing University, Xi'an, China

²Department of Nuclear Medicine, Shaanxi Provincial Cancer Hospital, Xi'an, China

³Shaanxi University of Chinese Medicine, Xianyang, China

⁴Department of Radiological, Xi'an Hospital of Traditional Chinese Medicine, Xi'an, China

Email: 2086393838@qq.com

How to cite this paper: Hao, Y.J., Zhang, Y.X., Shi, C.B., Shi, X.R., Zhu, P. and Zhou, S.L. (2025) Research on Automated Accurate Segmentation Algorithm of Double Kidney in Renal Dynamic Imaging Based on Improved UNet. *Open Journal of Radiology*, 15, 63-75. <https://doi.org/10.4236/ojrad.2025.152007>

Received: May 17, 2025

Accepted: June 24, 2025

Published: June 27, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Objective: Renal dynamic imaging, as an important tool for assessing renal function, is commonly used to test the perfusion and excretory functions of the kidney. In clinical diagnosis, accurate segmentation of renal regions is crucial for subsequent quantitative analysis and functional assessment. Currently, clinical outlining of renal dynamic renal regions still relies on manual labor. The purpose of this study is to construct an automated, accurate segmentation algorithm model for renal dynamic dual kidney regions. **Methods:** In this paper, an automated, accurate segmentation algorithm based on a non-local triple attention UNet network structure is proposed. The algorithm utilizes a deep convolutional neural network and a non-local triple attention mechanism for feature extraction and multi-scale fusion of renal dynamic imaging images to achieve accurate segmentation of renal dynamic imaging dual kidney regions. **Results:** By comparing the segmentation with other segmentation algorithms on the renal dynamic imaging dataset, the experimental results show that the algorithm model of this study is better than the standard Unet and Attention UNet segmentation algorithms in terms of the indicators such as mean Intersection over Union (mIoU), mean Pixel Accuracy (mPA), and so on. **Conclusion:** The algorithmic model in this study is able to automate the accurate segmentation of the double kidney region in renal dynamic images and demonstrates its effectiveness and robustness in automated segmentation of renal dynamic imaging.

Keywords

Renal Dynamic Imaging, Dual Kidney Segmentation, Deep Learning,

1. Introduction

Renal dynamic imaging is an imaging method that assesses the function of the kidneys by continuous imaging of the kidneys after the injection of a radiopharmaceutical. Compared with traditional static imaging methods, renal dynamic imaging can simultaneously observe renal perfusion and excretion functions and thus has significant clinical value in the assessment of renal diseases [1]. However, in renal dynamic imaging, factors such as high background noise, blurred organ boundaries, and interference from surrounding organ tissues pose challenges to automatic segmentation [2]. At present, the clinical renal ROI for renal dynamic imaging is still manually sketched by physicians, which is not only time-consuming and laborious, but also subjectively influenced by physicians. There is no uniform objective standard.

In the field of medical image segmentation [3], deep learning technology, especially the semantic segmentation network represented by UNet, has been widely used in the segmentation task of lungs, heart, liver, and other organs [4]. The UNet structure realizes multi-scale feature fusion by means of an encoder-decoder combined with a jump connection, which enables the model to have a high expressive ability in both semantic and spatial information. However, there are still some limitations of the standard UNet in processing complex medical images, which are mainly manifested in insufficient attention to key regional features and the loss of information in the process of multi-scale feature fusion.

In order to solve these problems, this study proposes an improved model of UNet (TAMUNet) based on a non-local Triple Attention Mechanism (TAM). The model effectively improves the segmentation network feature extraction capability by introducing an innovative attention structure. By simultaneously applying attention to three different dimensions, the mechanism significantly enhances the model's ability to perceive key features and improves the segmentation accuracy, especially when dealing with medical images with blurred boundaries, low contrast, and complex structures. Automated and accurate segmentation of the double kidney region in renal dynamic imaging is realized.

2. Data and Preprocessing

2.1. Data Sources

The renal dynamic imaging data used in this study were obtained from the Department of Nuclear Medicine, Shaanxi Provincial Cancer Hospital, and were used after hospital approval for anonymization. The data format is a static multi-frame DICOM sequence image format, which contains the changes in renal perfusion and excretion under different time phases. In order to simplify the experimental process, the images of 3 - 4 key time phases (optimal contrast time phases) were mainly selected for the segmentation study in this paper.

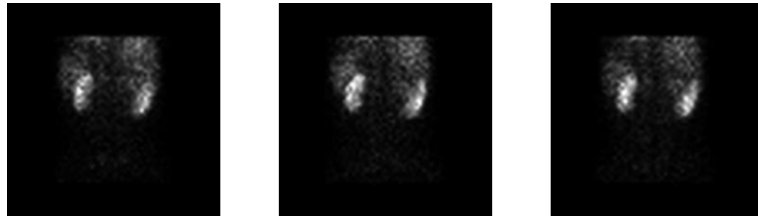


Figure 1. The three-time phases with the best contrast in the renal dynamic image are the 33rd, 34th, and 35th frame images, when the kidney is most clearly visualized and has the best contrast with the background noise.

The criteria for selecting key temporal phases in this study are shown in **Figure 1**. As for the experimental sample size, data from 100 patients were collected, and about 3 - 4 suitable single-frame images were selected for each patient, which ultimately resulted in about 300 - 400 images that could be used for training and testing. In order to ensure the accuracy of segmentation annotation, all kidney regions were manually labeled under the guidance of a professional physician to form the Ground Truth (GT).

2.2. Data Preprocessing

The renal dynamic imaging data used in this study were stored in DICOM standard format, and its SPECT image data can be represented as a four-dimensional matrix $X \in (W, H, C, F)$, in which each dimension characterizes the image's lateral resolution, longitudinal resolution, the number of channels, and the number of frames of the time series, respectively. It contains a total of 90 frames of image information. The first 30 frames are fast dynamic imaging data with acquisition parameters set to 2 s/frame and continuous scanning for 1 minute, and the last 60 frames are slow dynamic imaging data with acquisition parameters set to 20 s/frame and continuous scanning for 20 minutes.

This study mainly analyzes the slow dynamic imaging data, and in order to perform the semantic segmentation task, a single-frame 2D image needs to be extracted from the time-series image. By processing the renal dynamic images of 100 patients, a total of about 400 single-frame images in PNG format with 64×64 resolution were obtained. The annotation of all images after denoising and zooming was done under the guidance of professional physicians and these annotation data will be used for network parameter optimization and segmentation performance evaluation.

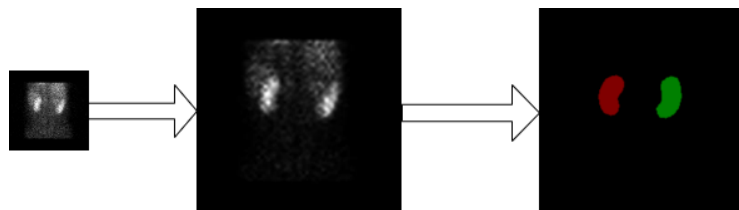


Figure 2. Convert the original DICOM data to a PNG image, select its key time phase to denoise and enlarge it, and then label the enlarged image.

In the model training stage, the data preprocessing process is shown in **Figure 2**. Specifically, the study first scales the original images uniformly to 224×224 resolution and manually annotates them, followed by expanding the samples using spatial transformation (including mirror flip, random rotation, etc.) techniques [3]. By constructing a standardized training set and validation set, not only is the data scale effectively increased, but the generalization performance of the model is also significantly improved.

The allocation ratio of this study for the dataset is 60% training set, 20% validation set, and 20% test set. At the end of the model training, the validation set is used to manually adjust the hyper-parameters of the model to get the final network model, and finally, the test set is used to evaluate the final effect of the model. Images from the same patient (with different phases) are not allowed to appear in different sets.

3. Methods

3.1. UNet Network Structure

UNet was proposed by Ronneberger *et al.* in 2015 and designed for medical image segmentation, which is characterized by a symmetric encoder-decoder structure and jump connections.

The classical UNet consists of two main parts, the encoder (down-sampled branch) and the decoder (up-sampled branch), and splices the features in the corresponding layer of the encoder with the corresponding scales in the decoder through jump connections. The structure of the network is shown in 3. This allows the network to fully integrate the low-level spatial information with the high-level semantic information in the process of up-sampling, thus improving the segmentation accuracy.

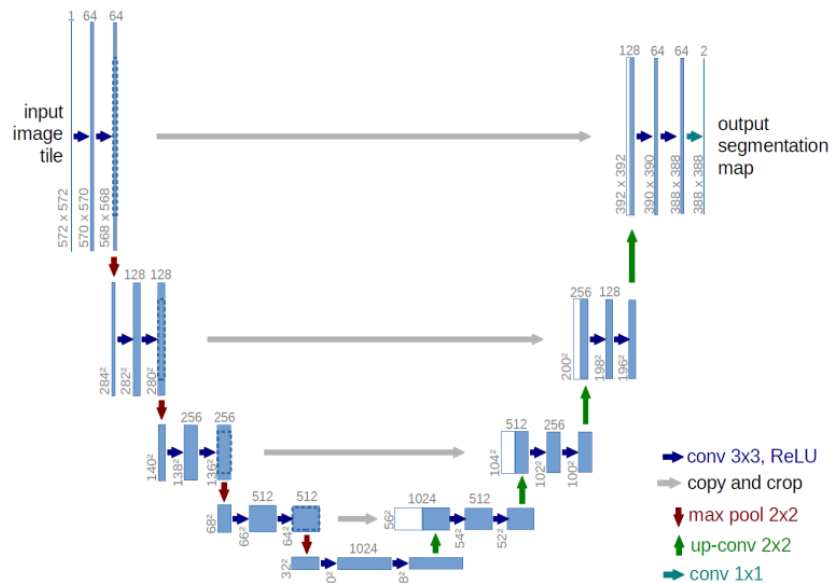


Figure 3. Structure of classical UNet network.

As shown in **Figure 3**, the UNet network architecture adopts the classical encoder-decoder structure. The encoder module implements feature extraction by stacking convolutional layers, batch normalization layers, and maximum pooling layers and gradually captures high-level semantic information during downsampling; accordingly, the decoder module consists of multiple up-sampling operations (or transposed convolution) with convolutional units and reconstructs the spatial details through level-by-level up-sampling. In particular, the network introduces a jump-joining mechanism in the decoding process, which cascades and fuses the features of each layer of the encoder with the corresponding decoding layer, thus effectively preserving the fine-grained feature information.

The output layer of the network uses a 1×1 convolutional kernel for feature dimensionality reduction to compress the number of channels to the number of target categories (this study contains 3 categories: left kidney, right kidney, and background region). Eventually, the model outputs a segmentation result map that is consistent with the size of the input image, realizing accurate classification at the pixel level.

3.2. Improvement Strategy

3.2.1. Adopt VGG16 as the Backbone Network

The traditional UNet model mainly consists of a simple convolutional layer and a pooling layer and usually uses a 3×3 convolutional kernel for feature extraction. The architecture includes an encoder (down-sampling part) and a decoder (up-sampling part), where the encoder gradually reduces the spatial dimensions of the feature maps, and the decoder gradually restores these dimensions. The feature maps are passed directly to the decoder via jump connections to help recover detailed information. However, due to the small number of convolutional layers, the feature extraction capability of traditional UNet is limited and suitable for handling relatively simple tasks, which mainly rely on the network depth and the number of convolutional layers to extract features.

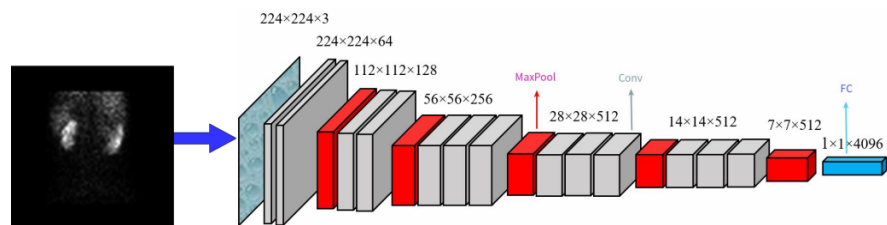


Figure 4. Model diagram of VGG16 network structure.

The UNet model proposed in this study uses the VGG16 network as a feature extractor in the encoder part. VGG16 is a deep convolutional neural network with multiple convolutional layers and pooling layers, which is capable of extracting richer features, and its network structure is shown in **Figure 4**. Compared with the simple convolutional layers of traditional UNet, VGG16 has a

stronger feature extraction capability and is able to capture more complex image features. By using the pre-trained weights of VGG for migration learning, the model performance is significantly improved. The UNet architecture based on the VGG16 backbone network is capable of extracting multi-level features and is suitable for handling more complex image segmentation tasks, while the multiple convolutional layers of VGG16 capture richer contextual information and detailed features.

3.2.2. Introduction of Attention Mechanisms

Attention mechanisms originally originated from the field of natural language processing and were later introduced into computer vision tasks. In the field of image segmentation, attention mechanisms are divided into three main categories: spatial attention, channel attention, and hybrid attention.

Spatial attention focuses on learning the importance of spatial regions of an image, such as non-local Neural Networks [5] and Spatial Attention Modules [6]. Channel attention, on the other hand, focuses on the inter-channel relationships of feature maps, and representative works include Squeeze-and-Excitation Networks (SENet) [7]. Hybrid attention considers both spatial and channel dimensions, such as the Convolutional Block Attention Module (CBAM) [6] and Efficient Channel Attention (ECA) [8].

Traditional attention mechanisms usually focus on a single dimension (e.g., channel or spatial dimension) and cannot fully capture the complex relationships in the feature map. In medical image segmentation tasks, the target structures often have complex morphology and fuzzy boundaries, requiring the model to focus on feature relationships in multiple dimensions simultaneously.

Based on this, this study proposes a triple-attention mechanism to model feature relationships simultaneously from three orthogonal dimensions: Channel-Width dimension, Height-Channel dimension, and Height-Width dimension, which achieves an all-round perception of the feature map.

The working principle of the triple attention mechanism can be seen in **Figure 5**.

1) Channel-width attention: Transform the input tensor from [B, C, H, W] to [B, H, C, W] by permute operation so that the attention gate pays attention to the relationship between the channel and width dimensions

2) Height-channel attention: transform the input tensor to [B, W, H, C] through the permute operation so that the attention gate focuses on the relationship between the height and channel dimensions.

3) Height-width attention: Apply directly on the original input [B, C, H, W] to focus on the feature distribution of the spatial dimension.

The results of the three dimensions of attention are averaged and fused to obtain a fully enhanced feature representation. This design enables the model to capture feature relationships from three different dimensions simultaneously, which significantly improves the feature representation capability and is particularly suitable for dealing with complex structures and fuzzy boundaries in medical images.

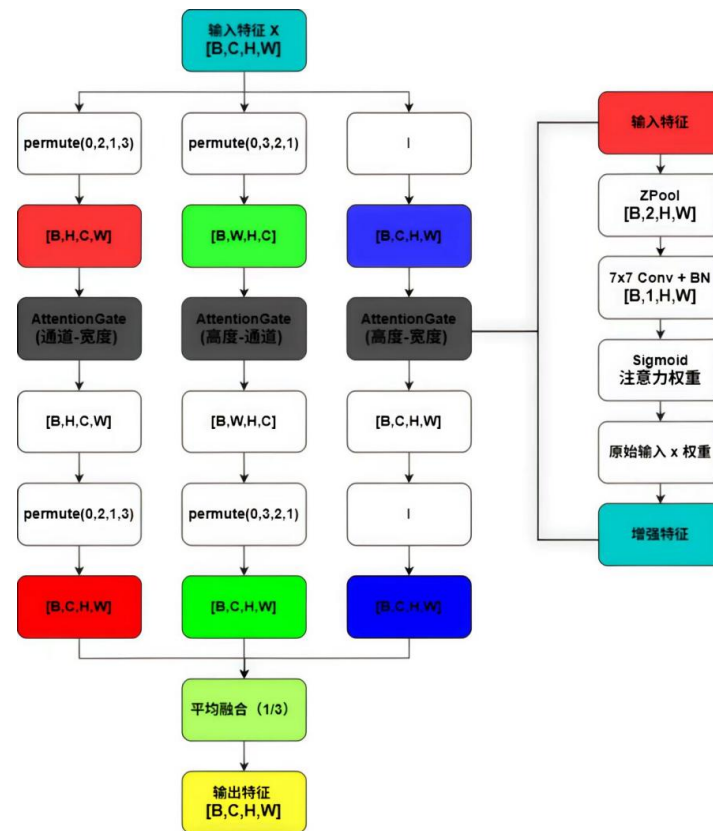


Figure 5. Structure of the triple attention mechanism module.

3.2.3. Deeply Separable Convolution with Mixed-Accuracy Training

In this study, the optimization strategy of combining depth-separable convolution with mixed-accuracy training is adopted. Depth separable convolution significantly reduces the parameter scale and computational complexity while ensuring the model performance by decomposing the standard convolution into two independent operations: depth convolution and point-by-point convolution. Experiments show that the method can reduce about 70% of the computation.

During the training process, we innovatively introduce a mixed-precision training mechanism, which achieves training acceleration by dynamically allocating the computational tasks of 16-bit floating point (FP16) and 32-bit floating point (FP32). Among them, FP16 is mainly used to reduce memory consumption and improve computational efficiency, while FP32 precision is retained for critical computation to ensure numerical stability. This mixed-precision strategy can increase the training speed by about 40% while maintaining model convergence stability.

By organically combining the above two techniques, this study achieves a significant effect of reducing the computational resource consumption by more than 50% and improving the training efficiency by 35% while ensuring segmentation accuracy. This optimization scheme is particularly suitable for task scenarios such as medical image segmentation that require processing high-resolution images.

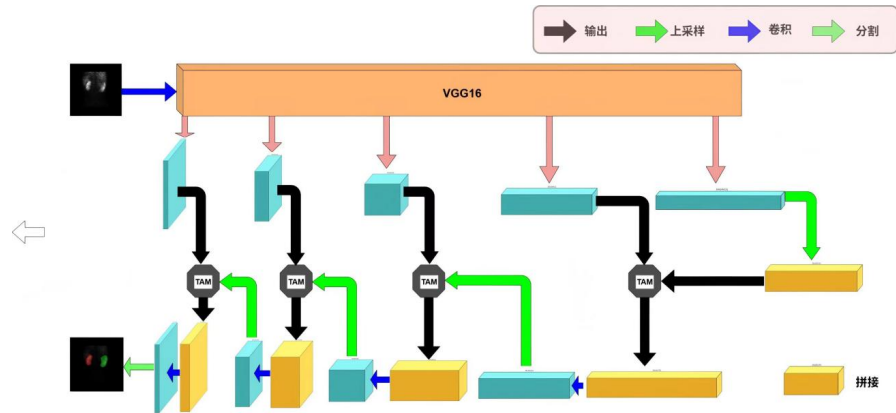


Figure 6. TAMUNet network structure.

The TAMUNet network model is shown in **Figure 6**. From the figure, it can be seen that the TAMUNet network follows the encoder-decoder structure of the classical UNet, but there are many changes, choosing VGG16 as the backbone network for feature extraction, using depth-separable convolution instead of the standard convolution to reduce the number of references and the computational complexity; and adding triple-attention mechanism at the jump connection to enhance the feature representation capability and focus on the multidimensional feature relationships.

4. Experiments and Results

4.1. Experimental Environment

The network model is trained with a freeze training strategy for the first 50 Epochs. This is because the features extracted from the feature extraction part of the neural network backbone are generic, and freezing this part can improve the training efficiency and prevent the weights from being corrupted. In the freezing phase, the backbone of the model remains unchanged, and only the feature extraction network is fine-tuned, so the memory footprint is small.

After entering the unfreezing phase, the backbone of the model will no longer be frozen, and the feature extraction network will change accordingly, at which time the memory footprint increases, and all parameters in the network are updated.

In the training task, Dice loss is used as the loss function of the model for optimization. The choice of Dice loss is usually based on its high fit with the task characteristics. Dice loss directly optimizes the segmentation objective: maximization of the overlap region. The core metrics of the segmentation task (e.g., Dice coefficients, IoUs) directly measure the overlap region of the predicted and the real masks. The Dice loss is consistent with the evaluation metrics, and there is no need to optimize the proxy objective indirectly. Boundaries are often unclear in nuclear medicine images, and the gradient computation of Dice loss relies on both predicted and true masks, which is more stable to fuzzy boundaries.

As shown in **Table 1**, these are the hardware devices used in this research experiment and some training parameter Settings of the algorithm.

Table 1. Configuration of the experimental environment for training with the TAMUNet network model.

Experimental environment	Conditions and settings
Hardware environment	GPU P100
Software environment	Python 3.11 + PyTorch deep learning framework
Parameter settings	Learning rate's initial value is 1e-4, and the minimum value is set to 1e-6
Batch size	Adjust the batch size according to the GPU memory, usually between 4 and 16.
Number of training rounds	300 Epoch
Loss function	Dice Loss

4.2. Evaluation Metrics

In the classification task of the semantic segmentation model, the prediction results can be categorized into four basic cases: true positive examples (TP), false positive examples (FP), true negative examples (TN), and false negative examples (FN). Specifically, TP denotes the samples that the model correctly predicts as positive cases, FP is the negative samples that the model misjudges as positive cases, FN refers to the positive case samples that the model fails to recognize, and TN is the negative samples that the model correctly judges.

An important metric for assessing the performance of the model is the mean Intersection over Union (mIoU), which is calculated as the average of the ratio of the intersection and concatenation of the prediction results and the true values for each category. Mathematically, mIoU can be expressed as the ratio of true cases (TP) to the sum of prediction error (FP+FN) and true cases, i.e., $\text{IoU} = \text{TP}/(\text{FP} + \text{FN} + \text{TP})$. This metric can effectively reflect the accuracy of the model in pixel-level classification tasks.

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}$$

p_{ij} denotes the number of true values of i that are predicted to be j . $K + 1$ is the number of categories (including the empty category). p_{ii} is the true number, p_{ij} and p_{ji} denote false positive and false negative, respectively. mIoU is generally computed based on classes, and the global-based evaluation is obtained by accumulating the IOUs of each class after computation and then averaging them. Larger values represent better segmentation accuracy.

Mean Pixel Accuracy (mPA): the proportion of the number of correctly categorized pixels within each class, after which the average of all classes is sought. That is, $\text{PA} = (\text{TP} + \text{TN})/(\text{FP} + \text{FN} + \text{TP} + \text{TN})$.

$$mPA = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij}}$$

In semantic segmentation model evaluation, Precision and Recall are two key metrics. Precision reflects the accuracy of the positive samples in the prediction results, which is expressed as the category pixel accuracy (CPA), Precision = TP/(TP + FP) or TN/(TN + FN), and is used to measure the probability of correct prediction for a certain category. Recall, on the other hand, characterizes how well the model covers the true positive sample and is calculated as Recall = TP/(TP + FN) or TN/(TN + FP).

Accuracy, as a global evaluation metric, characterizes the proportion of correct predictions to the total samples, accuracy = (TP + TN)/(TP + TN + FP + FN), which is equivalent to pixel accuracy (PA). However, in the case of uneven sample distribution, a single reliance on accuracy may lead to assessment bias, and thus, a combination of other metrics is needed for a comprehensive evaluation of the model.

4.3. Experimental Results

In order to verify the effectiveness of the proposed model, the following baseline methods are selected for comparison experiments in this study:

Threshold Segmentation Method: as a traditional method in the field of image segmentation, it realizes segmentation by setting a threshold combined with post-processing. Although it has the advantage of simple computation, it is often difficult to obtain ideal results when dealing with complex images.

Standard UNet [9]: as a benchmark model in the field of medical image segmentation, this architecture has become the de facto standard for all kinds of medical image segmentation tasks, and its encoder-decoder structure provides an important reference for subsequent research.

Attention UNet [10] (AttUNet): this model innovatively introduces the attention gating mechanism on the basis of UNet, optimizes the jump connections through the attention gate, effectively solves the noise problem in medical images, and is one of the earlier models that apply the attention mechanism to medical image segmentation.

The method in this paper (TAMUNet): the improved model proposed in this study adopts the VGG network as the encoder feature extractor and introduces a triple attention module at the jump connection, which significantly improves the segmentation accuracy by adaptively adjusting the feature channel weights. The model realizes the effective integration of feature extraction and attention mechanism while maintaining the basic architecture of UNet.

From **Table 2**, it can be seen that this paper's method has obvious improvement in core indexes such as mPA and mIoU compared with other models. Meanwhile, compared with the traditional method, which is prone to wrong segmentation in the edge region, this paper's method can visually fit the kidney contour better, indicating that the introduction of the VGG16 backbone network as well as the

non-local triple attention mechanism, positively helps the segmentation performance.

Table 2. Comparative evaluation results of renal segmentation indexes.

Method	mIoU (%)	mPA (%)	mPrecision (%)	mRecall (%)	Accuracy (%)
Threshold	70.13	80.24	85.56	78.30	87.22
UNet	81.47	89.92	90.33	92.37	98.07
AttUNet	85.32	93.56	94.41	94.12	99.32
TAMUNet	89.17	94.58	93.68	94.58	99.56

AttUNet only increases the inference time by a small amount compared to the original UNet. The TAMUNet triple attention mechanism increases the inference time accordingly due to the need to compute the attention in three dimensions (channel-width, height-channel, and height-width), but reduces the computation amount to a certain extent due to the use of ZPool for feature compression. In practice, if the inference speed requirement is high, partial attention gates can be selectively disabled, and the use of the complete triple attention mechanism can be weighed according to the specific task requirements and hardware conditions.

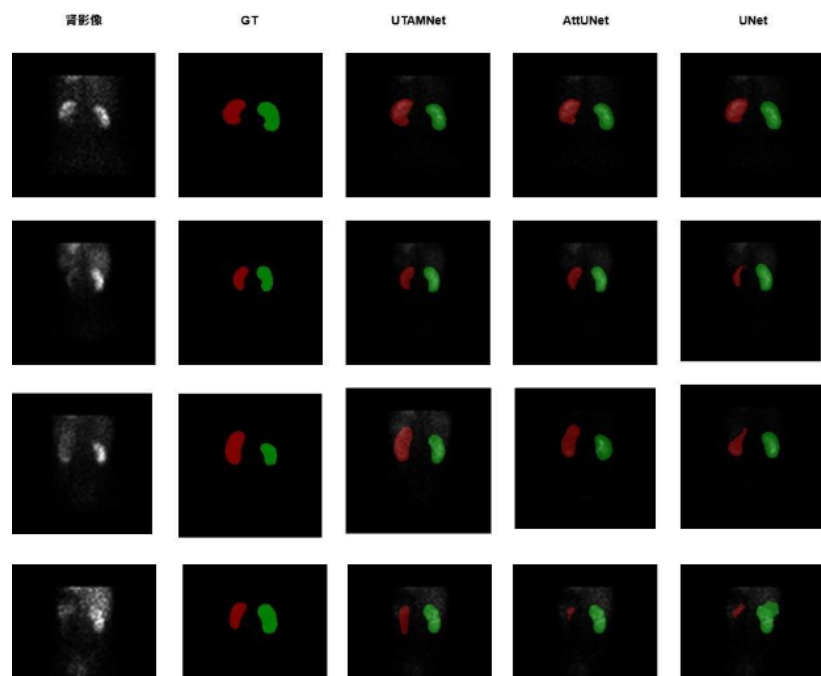


Figure 7. Segmentation effect of renal dynamic imaging on TAMUNet, AttUNet, and standard UNet.

In **Figure 7**, some examples from renal dynamic imaging are used to demonstrate the actual segmentation effect of the model. From the figure, it can be seen that the TAMUNet model constructed in this study achieves good segmentation results under different conditions, while the segmentation results of AttUNet and

standard UNet are less satisfactory when the development is not clear (the second and the third renal images) and when the background noise is serious and the contrast is low (the fourth renal image). It can be seen that the model performs well on injured or postoperatively changed kidneys.

The experimental results show that the TAMUNet segmentation model exhibits excellent performance in kidney dynamic image processing, and its segmentation results are highly consistent with the real labeling. This effect is mainly attributed to the designed non-local triple attention module, which effectively suppresses the background noise interference and enhances the attention to the target region of the kidney through the adaptive feature weighting mechanism, and finally realizes the automated and accurate segmentation of the double kidney region.

5. Discussion

When the TAMUNet segmentation model handles case images with obvious noise, weak kidney rendering, and blurred contours, the introduction of the TAM attention mechanism can suppress the background noise interference and allow the network to pay more attention to the key regions of the kidneys, with a better ability to capture the edge information. The model structure is relatively simple, and easy to generalize its use in other organs or other time series data. Only minor modifications to the network input channels, the number of output splits, and the data preprocessing part are required. The model maintains good segmentation results in the face of a variety of imaging conditions, such as different imaging time phases or weak renal imaging, which, to some extent, proves the generalization ability of the algorithm.

However, the current implementation of the model mainly relies on the feature extraction capability of the model itself to deal with image quality issues without a dedicated quality assessment or frame restoration module, which may have limitations in dealing with severe motion blur or artifacts.

Since the data used in this study came from a single device in a single hospital, the effect needs to be further verified if applied to multi-center data or different types of dynamic imaging. Although this study only utilized the frame images of key temporal phases for segmentation, more stable and accurate segmentation results may be obtained if the complete dynamic timing information is combined and combined with temporal convolution or 3D convolutional networks [11]. Deep learning models, especially those that use VGG16 as the backbone, require a large amount of computational resources and video memory in the training and inference process, which is more demanding on the device.

Later, more lightweight model structures should be investigated to reduce the computational cost while incorporating interpretable methods to enhance the interpretability and credibility of the models in clinical applications.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Chen, J.F. (2002) The Value of Renal Dynamic Imaging in the Diagnosis of Renal Function. *Journal of Nanhua University*, **30**, 39-40.
- [2] Li, Y. (2021) Progress in the Application of Imaging Histology in Nuclear Medicine Imaging. *Intelligent Health*, **7**, 36-38.
- [3] Zhang, Y. (2018) Image Engineering. Tsinghua University Press.
- [4] Li, X., Zhang, J., Liu, P., et al. (2021) Current Status and Prospect of Deep Learning Application in Medical Imaging. *Journal of Clinical Radiology*, **40**, 2423-2429.
- [5] Wang, X., Girshick, R., Gupta, A. and He, K. (2018) Non-Local Neural Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7794-7803. <https://doi.org/10.1109/cvpr.2018.00813>
- [6] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [7] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/cvpr.2018.00745>
- [8] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q. (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 11531-11539. <https://doi.org/10.1109/cvpr42600.2020.01155>
- [9] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [10] Oktay, O., Schlemper, J., Folgoc, L.L., et al. (2018) Attention U-Net: Learning Where to Look for the Pancreas. arxiv:1804.03999.
- [11] Mahmoudi, S.E., Akhondi-Asl, A., Rahmani, R., Faghih-Roohi, S., Taimouri, V., Sabouri, A., et al. (2010) Web-Based Interactive 2D/3D Medical Image Processing and Visualization Software. *Computer Methods and Programs in Biomedicine*, **98**, 172-182. <https://doi.org/10.1016/j.cmpb.2009.11.012>