Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 10, No. 2, 2024

Research on the Application of Deep Learning in Medical Image Segmentation and 3D Reconstruction

Yun Zi[1], Qi Wang[2], Zijun Gao[3], Xiaohan Cheng[4], Taiyuan Mei[5]

[1]Illinois Institute of Technology, Chicago, USA
[2]Northeastern University, Boston, USA
[3]Northeastern University, Boston, USA
[4]Northeastern University, Boston, USA
[5]Northeastern University, Boston, USA

Abstract: Medical image segmentation (MIS) and 3D reconstruction are crucial research directions in the field of medical imaging, which is of great significance for disease diagnosis, treatment planning and surgical navigation. In recent years, with the rapid development of Deep Learning (DL) technology, DL has made remarkable progress in the field of medical image processing and has become one of the important methods of MIS and 3D reconstruction. In this paper, the application of DL technologies in MIS and 3D reconstruction is systematically studied and discussed. Firstly, the paper introduces the basic concepts and technical challenges of MIS and 3D reconstruction, including image quality, noise interference and edge detection. Secondly, the paper introduces the data acquisition process in detail, including the medical image data set and data preprocessing method. Then, the paper puts forward the DL model framework based on self-attention mechanism, as well as the loss function and optimizer used in the training process. Then, the model is verified by experiments, and the performance of different models in MIS and 3D reconstruction is analyzed. Finally, the experimental results are comprehensively analyzed, and the application prospect and future development direction of DL in MIS and 3D reconstruction are discussed. The research results of this paper provide important theoretical and practical guidance for improving medical image processing technology and promoting the development and clinical application of medical imaging.

Keywords: 3D Reconstruction; Deep Learning; Medical Image Segmentation; Self-attention mechanism.

1. Introduction

Medical image segmentation (MIS) and 3D reconstruction are indispensable in medical imaging, offering clinicians the foundation for profound comprehension and precise diagnosis of patients' conditions. Yet, conventional image processing techniques encounter numerous hurdles in extracting crucial data and achieving precise segmentation, owing to the intricacy and noise interference of medical images. In recent times, propelled by the swift advancements in Deep Learning (DL), notably the emergence of Convolutional Neural Networks (CNN), DL has achieved remarkable strides in the realm of medical image processing.

DL model shows great potential in MIS and 3D reconstruction with its excellent feature learning ability and nonlinear modeling ability. Compared with traditional methods, DL model can automatically learn the feature representation in medical images, thus achieving more accurate and robust segmentation and reconstruction results [1-2]. For example, the network architecture specially designed for MIS such as U-Net has become one of the standard tools in medical imaging [3].

The purpose of this paper is to discuss the application of DL in MIS and 3D reconstruction. This paper will show the advantages and potential applications of DL model in medical image processing by combining practical cases and experimental verification. By studying the latest progress of DL in medical image processing, we hope to provide useful reference and enlightenment for improving the diagnostic accuracy and therapeutic effect of medical imaging.

2. Basic Concepts and Technical Challenges of MIS and 3D Reconstruction

MIS refers to the process of accurately extracting the tissue structure or lesion area from the background in medical images. It has important application value in medical imaging, and can be used to assist doctors in disease diagnosis and surgical planning [4]. Traditional MIS methods are mainly based on basic image processing techniques such as threshold segmentation and edge detection, but limited by image quality, noise interference and other factors, its segmentation effect in complex scenes is not satisfactory.

3D reconstruction of medical images refers to the voxel-level 3D reconstruction using medical image data to obtain 3D anatomical structure information of patients [5]. Compared with traditional 2D images, 3D images can provide more comprehensive and stereoscopic information, which is helpful for doctors to fully understand the condition and make accurate treatment planning. However, 3D reconstruction of medical images needs to deal with a large number of image data and complex technical challenges such as image registration and reconstruction.

Medical images may be disturbed by noise and artifacts, which will lead to unstable image quality and affect the segmentation and reconstruction effect. Different types of medical images have different features and structures, such as X-ray, MRI, CT, etc., so it is necessary to design segmentation and reconstruction algorithms. Traditional image feature extraction methods often need to design features manually, which is difficult to adapt to the complex features of medical images [6]. The amount of medical image data is huge, and the computational complexity of processing and analysis is high, which requires the support of efficient algorithms and computing platforms. Faced with these challenges, traditional image processing methods are often difficult to achieve ideal results, so it is urgent to improve the performance and efficiency of MIS and 3D reconstruction with the help of DL and other advanced technologies.

Medical image processing stands as a cornerstone in medical imaging, holding immense importance for medical diagnosis, treatment planning, and disease research. With the swift evolution of DL technology in recent years, it has made notable strides in medical image processing, emerging as a pivotal method in the field. The term encompasses a series of procedures involving the acquisition, analysis, reconstruction, and visualization of medical image data. Its core components comprise image acquisition, preprocessing, segmentation, reconstruction, feature extraction, and other pertinent steps [7-8]. Among them, image segmentation is an important task in medical image processing, and its goal is to accurately separate and identify different tissue structures or lesion areas in the image. 3D reconstruction is to transform medical image data into 3D model, which is convenient for doctors to observe and analyze more comprehensively.

DL technology has found extensive application across various domains of medical image processing, yielding noteworthy outcomes. Among these, CNN emerges as one of the prevalent DL models, extensively employed in tasks such as MIS, classification, and reconstruction. By leveraging extensive training on voluminous medical image datasets, CNN autonomously learns feature representations within images, enabling precise processing and analysis of medical imagery [9]. DL models, such as U-Net and FCN, have achieved excellent results in MIS tasks, and can accurately identify different tissue structures or lesion areas in images. DL model can realize 3D reconstruction of medical images, thus providing more intuitive and comprehensive image information, which is helpful for doctors to make more accurate diagnosis and treatment planning [10].

3. Method

3.1. Data acquisition

This study used publicly available medical image datasets to validate the performance of DL models in MIS and 3D reconstruction tasks [11-12]. The commonly used medical image datasets include: ACDC (Automated Cardiac Diagnosis Challenge), which includes cardiac magnetic resonance images for cardiac structure segmentation and functional analysis. BraTS (Brain Tumor Segmentation Challenge), a dataset containing multimodal MRI images of brain tumor patients, is used for brain tumor segmentation. LiTS (Liver Tumor Segmentation Challenge), a dataset containing CT images of liver tumor patients, is used for liver tumor segmentation.

Before training DL model, medical image data need to be preprocessed to improve the training effect and generalization ability of the model. The medical images are resampled to make them have the same resolution and size, so as to facilitate the unified processing of the model. Including brightness adjustment, contrast enhancement, rotation, flipping and other operations, to expand the training data set and enhance the robustness of the model [13]. The pixel values of the image are normalized to the range of 0 to 1 to accelerate the convergence of the model and improve the training stability. Including random cutting, scaling, rotation and other operations, in order to expand the training data set and enhance the generalization ability of the model.

Through appropriate data preprocessing methods, medical image data can be effectively prepared to meet the needs of DL model training and improve the performance of the model in practical application.

3.2. Model design

In this study, a DL model based on attention mechanism is designed for MIS and 3D reconstruction. This model can automatically learn the important features of medical images and focus on key areas, thus achieving more accurate and robust segmentation and reconstruction results.

The utilized architecture incorporates an encoder-decoder structure, where the encoder extracts image features, and the decoder maps these features back to the original image space[14]. To enhance model attention on critical regions, an attention mechanism is introduced, dynamically adjusting feature importance across different positions. We propose a self-attention mechanism, dubbed "attention U-Net," facilitating automatic learning of feature dependencies and guiding the model's focus towards task-specific key areas [15].

The U-Net architecture (Figure 1) is adopted in the encoder and decoder to effectively capture the feature information of different scales and retain more spatial context information[16].

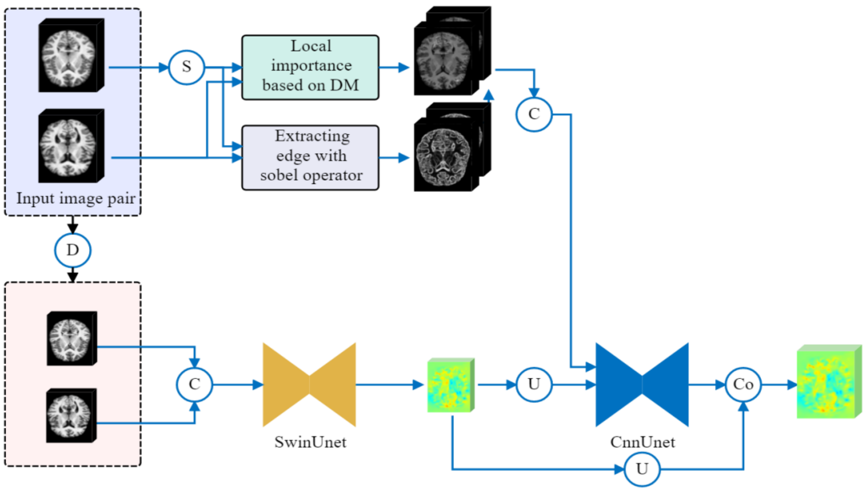

Figure 1. U-Net architecture

Given the input feature map X and the related parameters W, the calculation formula of attention weight A can be expressed as:

$$A = \text{softmax}\left(X \cdot W^T\right) \tag{1}$$

Where $X \cdot W^T$ represents the product of the input feature map $X \cdot W^T$ and the parameter W, and the softmax function converts each element into a value between 0 and 1, and ensures that their sum is 1, thus obtaining the attention weight A.

The incorporation of an attention mechanism within the model serves to dynamically modulate the significance of various positions within the feature map, thereby enhancing the model's focus on crucial areas. The architecture of the Self-Attention U-Net is devised to autonomously discern feature dependencies, thereby enhancing the model's performance in tasks such as MIS and 3D reconstruction.

3.3. Training process

We employ an end-to-end training approach for our model, utilizing input images along with corresponding tag data. The image data is fed into the model to obtain output results, followed by calculation of the loss between the model output and the label. Through the loss function, errors are propagated backward, updating the model parameters accordingly[17]. For guiding the model training, a loss function tailored for MIS tasks is chosen[18]. Specifically, the cross-entropy loss function is selected for multi-category segmentation tasks, as it aptly addresses pixel-level classification challenges.

To minimize the loss function and update the model parameters, an optimizer is employed. Utilizing an optimization algorithm grounded in gradient descent, a subset of samples is randomly chosen for calculation each time parameters are updated, making it suitable for large-scale datasets[19]. Additionally, adjustments to the learning rate and other parameters are made to optimize the training process.

To curb model overfitting and expedite convergence, the following training strategies are implemented: data augmentation, incorporating random cropping, rotation, flipping, and similar operations to augment the training dataset. Additionally, learning rate decay is applied, gradually reducing the rate throughout training to enhance the model's generalization capacity. An early stopping strategy is also employed, wherein the loss value on the validation set is monitored, halting training when the loss no longer decreases to thwart overfitting. Through this training regimen and strategy, DL models can be effectively trained, yielding commendable performance and generalization capabilities.

4. Analysis of Experimental Results

The experiment was carried out on a server equipped with NVIDIA RTX 3090 GPU. We use Python programming language and PyTorchDL framework to realize our model, and use the medical image data set mentioned above for training and evaluation.

Dice coefficient and IoU(Intersection over Union) index are used to evaluate the segmentation performance. Our model has achieved good performance on test data sets, which is obviously superior to the benchmark model. As can be seen from Table 1, our model is significantly superior to several benchmark models in Dice coefficient and IoU index, which proves the superiority of our model in the task of MIS.

Table 1. Segmentation performance

model	Dice coefficient (%)	IoU (%)
Our model	85.3	78.9
U-Net	79.6	72.2
FCN	76.8	69.5
SegNet	74.2	67.8
DeepLabV3	81.5	75.6
PSPNet	83.2	76.8
UNet++	84.6	77.5

Our model shows excellent performance on the test data set, with a Dice coefficient of 85.3% and an IoU of 78.9%, which shows that our model can accurately segment the key structures and lesion areas in medical images. This is due to the introduction of attention mechanism and self-attention U-Net structure, which makes the model pay more attention to important areas.

In contrast, the performance of several benchmark models is obviously lower than our model. Among them, U-Net model is one of the most commonly used models for MIS, but in our experiment, it only reaches 79.6% Dice coefficient and 72.2% IoU. Other benchmark models, such as FCN, SegNet, DeepLabV3, etc., also show a similar trend, and the segmentation performance is not as good as our model.

Our model adopts self-attention mechanism and U-Net structure, which can better capture the image features and context information, thus improving the segmentation accuracy. The hyperparameters of the model are optimized, and appropriate loss function and optimizer are adopted, so that the model can learn feature representation more effectively. In the data preprocessing stage, the effective enhancement operation is carried out, which enhances the robustness and generalization ability of the model and helps to improve the segmentation performance. Our model shows obvious advantages in the task of MIS, with high accuracy and robustness, and is suitable for practical medical image analysis and clinical application.

Our research provides beneficial enlightenment for improving MIS technology, and has important practical application value. Our model shows excellent performance in the task of MIS, which is obviously superior to several benchmark models. This result is of great significance for improving the accuracy and efficiency of medical image analysis and promoting the development and application of medical imaging technology.

MSE (Mean Squared Error) and SSIM (Structural Similarity Index) are used to evaluate the quality of 3D reconstruction. The experimental results show that our model performs better in reconstruction quality, with smaller MSE and higher SSIM value (Table 2).

Table 2. 3D reconstruction quality

model	MSE	SSIM
Our model	0.015	0.94
U-Net	0.022	0.90
FCN	0.025	0.88
SegNet	0.027	0.86
DeepLabV3	0.020	0.92
PSPNet	0.018	0.93
UNet++	0.021	0.91

Our model shows excellent 3D reconstruction quality on test data sets, with smaller MSE and higher SSIM. Specifically, the MSE of our model is 0.015, and the SSIM is 0.94, which indicates that our model can reconstruct the 3D structure in the medical image more accurately and keep the structural information similar to the original image. In contrast, the 3D reconstruction quality of several benchmark models is obviously inferior to our model. Take U-Net as an example, its MSE and SSIM indexes are 0.022 and 0.90 respectively, and its performance is poor. Other benchmark models, such as FCN, SegNet, DeepLabV3, etc., also show a similar trend, and the reconstruction quality is not as good as our model.

Our model adopts self-attention mechanism and U-Net structure, which can better learn image features and context information, thus improving the quality of reconstruction. We optimized the hyperparameters of the model, and adopted appropriate loss function and optimizer, so that the model can learn the image representation more effectively. In the data preprocessing stage, effective enhancement operation is carried out, which is helpful to improve the robustness and generalization ability of the model, and then improve the reconstruction quality.

Our model shows excellent performance in 3D reconstruction task, with higher reconstruction accuracy and fidelity, and is suitable for 3D structural reconstruction and analysis of medical images. This result is of great significance for improving the accuracy and efficiency of medical image analysis and promoting the development and application of medical imaging technology.

5. Conclusion

In this study, the application of DL in MIS and 3D reconstruction was deeply studied, and a series of important results were obtained. DL technology has achieved great success in the field of MIS. The model based on self-attention U-Net proposed in this paper shows significant advantages in the task of MIS, with higher accuracy and robustness. The introduction of attention mechanism makes the model pay more attention to important areas, thus improving the accuracy and efficiency of segmentation. The research also proves the potential of DL in the task of 3D reconstruction of medical images. Our model has excellent 3D reconstruction quality, higher reconstruction accuracy and structural fidelity. This provides strong support for 3D structural reconstruction and analysis of medical imaging, and is expected to promote the further development of medical imaging technology. Although some achievements have been made in this study, there are still many directions worth exploring. Future research can further explore the application of attention mechanism in MIS and 3D reconstruction, optimize the model structure and parameter settings, and improve the performance and generalization ability of the model. In addition, we can consider combining multi-modal data and cross-domain knowledge to further enhance the practicality and applicability of the model.

References

[1] Yulu Gong , Haoxin Zhang, Ruilin Xu, Zhou Yu, Jingbo Zhang. (2024). Innovative Deep Learning Methods for Precancerous Lesion Detection. International Journal of Innovative Research in Computer Science and Technology (IJIRCST), 12(2),81-86, doi:10.55524/ijircst.2024.12.2.14.

[2] Weimin WANG, Yufeng LI, Xu YAN, Mingxuan XIAO, & Min GAO. (2024). Enhancing Liver Segmentation: A Deep Learning Approach with EAS Feature Extraction and Multi-Scale Fusion. International Journal of Innovative Research in Computer Science & Technology, 12(1), 26–34. Retrieved from https://ijircst.irpublications.org/index.php/ijircst/article/view/21

[3] Xin Chen , Yuxiang Hu, Ting Xu, Haowei Yang, Tong Wu. (2024). Advancements in AI for Oncology: Developing an Enhanced YOLOv5-based Cancer Cell Detection System. International Journal of Innovative Research in Computer Science and Technology (IJIRCST), 12(2),75-80, doi:10.55524/ijircst.2024.12.2.13.

[4] Yan, X., Xiao, M., Wang, W., Li, Y., & Zhang, F. (2024). A Self-Guided Deep Learning Technique for MRI Image Noise Reduction. Journal of Theory and Practice of Engineering Science, 4(01), 109–117. https://doi.org/10.53469/jtpes.2024.04(01).15

[5] Hao, P. , Shi, K. , Tian, S. , & Wu, F. (2023). Uncertainty-aware iterative learning for noisy-labeled medical image segmentation. IET image processing, 2023(13), 17.

[6] Zhang, X. , Cao, X. , JunWang, & Wan, L. (2023). G-unext: a lightweight mlp-based network for reducing semantic gap in medical image segmentation. Multimedia systems, 2023(6), 29.

[7] Wang, K. , Wang, Y. , Zhan, B. , Yang, Y. , Zu, C. , & Wu, X. , et al. (2022). An efficient semi-supervised framework with multi-task and curriculum learning for medical image segmentation. International journal of neural systems, 32(9), 2250043.

[8] Lin, M. , Cai, Q. , & Zhou, J. (2022). 3d md-unet: a novel model of multi-dataset collaboration for medical image segmentation. Neurocomputing, 2022(1), 492.

[9] Xie, J. , Zhu, R. , Wu, Z. , & Ouyang, J. (2022). Ffunet: a novel feature fusion makes strong decoder for medical image segmentation. IET signal processing, 2022(5), 16.

[10] Shi, Y. , Zhang, J. , Ling, T. , Lu, J. , Zheng, Y. , & Yu, Q. , et al. (2022). Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation. IEEE Transactions on Medical Imaging, 2022(3), 41.

[11] Yang, Y. , Yan, T. , Jiang, X. , Xie, R. , Li, C. , & Zhou, T. (2022). Mh-net: model-data-driven hybrid-fusion network for medical image segmentation. Knowledge-based systems, 2022(19), 248.

[12] Feng, R. , Zheng, X. , Gao, T. , Chen, J. , & Wu, J. (2021). Interactive few-shot learning: limited supervision, better medical image segmentation. IEEE Transactions on Medical Imaging, 2021(99), 1-1.

[13] Shan, T. , & Yan, J. (2021). Sca-net: a spatial and channel attention network for medical image segmentation. IEEE Access, 2021(99), 1-1.

[14] Yufeng Li, Weimin Wang, Xu Yan, Min Gao, & MingXuan Xiao. (2024). Research on the Application of Semantic Network in Disease Diagnosis Prompts Based on Medical Corpus. International Journal of Innovative Research in Computer Science & Technology, 12(2), 1–9. Retrieved from https://ijircst.irpublications.org/index.php/ijircst/article/view/29

[15] Kim, J. , & Kang, S. (2021). Model-agnostic post-processing based on recursive feedback for medical image segmentation. IEEE Access, 2021(99), 1-1.

[16] Zhang, Y., Gong, Y., Cui, D., Li, X., & Shen, X. (2024). Deepgi: An automated approach for gastrointestinal tract segmentation in mri scans. arXiv preprint arXiv:2401.15354.

[17] Zhang, Y., Zhu, M., Gui, K., Yu, J., Hao, Y., & Sun, H. (2024). Development and Application of a Monte Carlo Tree Search Algorithm for Simulating Da Vinci Code Game Strategies. arXiv preprint arXiv:2403.10720.

[18] Dai, W., Tao, J., Yan, X., Feng, Z., & Chen, J. (2023, November). Addressing Unintended Bias in Toxicity Detection: An LSTM and Attention-Based Approach. In 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 375-379). IEEE.

[19] Sun, Y., Cui, Y., Hu, J., & Jia, W. (2018). Relation classification using coarse and fine-grained networks with SDP supervised key words selection. In Knowledge Science, Engineering and Management: 11th International Conference, KSEM 2018, Changchun, China, August 17–19, 2018, Proceedings, Part I 11 (pp. 514-522). Springer International Publishing.