Original Research Article

Development and application evaluation of a medical image diagnosis teaching assistant system supported by multi-modal deep learning

Hilzati Kuzati[1], *Baihetiya Mutalipu*[2]

Xinjiang Hetian College, Xinjiang Uygur Autonomous Region, Hotan region, 848000, China

***Abstract**:* With the rapid development of medical imaging technology, medical images collected by different sensors containing different pathological information are also known as multi-modal medical images, which are indispensable tools for modern medical analysis and disease diagnosis. Due to the inability of single-modal images to provide complete pathological information for disease diagnosis, medical staff need to observe multiple images of different modalities simultaneously, and this kind of image examination has errors, which increases the uncertainty of clinical diagnosis. Therefore, some scholars have proposed multi-modal medical image fusion technology, which can not only demonstrate the application advantages of different modal images but also effectively make up for the application problems of single images, providing a reasonable basis for disease diagnosis in the new era. Therefore, this paper mainly discusses the development and application effect of a medical image diagnosis teaching assistant system supported by multi-modal deep learning.

***Keywords**:* Multimodal deep learning; Medical image diagnosis; Teaching assistance system; Image fusion

1. Introduction

With the rapid development of medical imaging technology, multimodal medical images containing various pathological information, such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound, have become indispensable tools for modern medical analysis and disease diagnosis. However, single-modal images cannot provide complete pathological information, leading to the need for healthcare professionals to observe multiple modalities simultaneously. This process is not only time-consuming and prone to errors but also increases the uncertainty in clinical diagnosis. For example, CT clearly shows bone structures but lacks sufficient contrast for soft tissues, while MRI can finely depict soft tissue lesions but struggles to show calcification details. The information complementarity and observational limitations between these modalities urgently require technical means to integrate imaging information.

2. Image fusion framework based on deep learning

In the rapid development of information technology, artificial intelligence has gradually become an important force to promote reform and innovation in various industries. In the medical field, the application of artificial intelligence technology has now been highly valued by people.[1-3]As an important part of medical diagnosis, the application performance of medical imaging diagnosis directly affects the patient's disease judgment and treatment plan. There are limitations in the application of traditional medical imaging diagnosis methods, so research and application The medical imaging diagnosis auxiliary system with artificial intelligence as the core is of great practical significance. After understanding the image fusion framework of deep learning, this paper deeply discusses the development and application evaluation of the medical imaging diagnosis teaching auxiliary system supported by multimodal deep learning.Deep learning algorithms have strong capabilities in feature extraction and information representation and have been widely applied in the field of image fusion, resulting in various fusion methods.[4-6]These methods can be classified into two forms based on the adopted basic framework: one is the method centered on convolutional neural networks, and the other is the method centered on generative adversarial networks. The former, as a feedforward neural network, has the basic characteristics of local connection and weight sharing, and can extract features from source images using multiple convolutional layers, then achieve feature fusion and feature reconstruction, and finally construct a loss function based on the similarity between the generated fusion result and the source image, which can be used for backpropagation to optimize network parameters. The latter, as a generative model, is mainly used to evaluate new samples similar to the target distribution, and the final trained generator can generate samples similar to the target distribution.[7-9]

After understanding the basic framework of image fusion, this study proposes a medical image fusion algorithm centered on saliency perception and generative adversarial networks in response to multiple problems existing in existing algorithms, such as color distortion, contrast deviation, and feature distribution differences that limit fusion performance. This algorithm will comprehensively consider the feature differences of multi-modal images on the basis of the existing convolutional neural network image fusion algorithm process and select the same path to extract features from two different modal images. The structure design of the generative network requires that the size of the first and last convolutional layers be 1×1, and the size of the remaining layers be 3×3. In addition, except for the two convolutions used to calculate spatial attention in each feature fusion module, batch normalization and activation functions are applied after each convolution in the remaining layers. The slope of LeakyReLU is set to 0.01. To avoid losing important information during convolution, the stride size of all convolutional layers is 1, and the padding mode is "SAME". In other words, the size of the feature map remains consistent with the source image throughout the generative network.

3. Design and implementation of a medical imaging diagnosis teaching assistant system

To ensure the orderly conduct of medical imaging diagnosis teaching, a multimodal medical image fusion system is designed based on the above algorithms and framework. See **Table 1** for details.The system mainly features image reading and saving, image preprocessing and registration, image fusion, and simple and efficient interaction. The overall design is divided into three layers, from top to bottom as follows:

Table 1. Architecture hierarchy table.

Layer	Function Category	Function Module	Dependent Lower-Level Resources
User Interface Layer	Interaction Function	Visualization	
		Interaction	
		Parameter Configuration	
Server Layer	Technical Support	Support Software (PyTorch, VTK, xhboost, etc.)	Data, CPU, GPU
	Data Processing	Data Storage	Data, CPU, GPU
		Rendering	Data, CPU, GPU
		Configuration	Data, CPU, GPU
		Preprocessing	Data, CPU, GPU
		Integration	Data, CPU, GPU
		Evaluation	Data, CPU, GPU
Resource Layer	Data Resources	Data	
	Computing Resources	CPU	
		GPU	

First, the user layer. This layer is responsible for the interaction between the user and the system and is implemented using PyQt5 GUI. After the user clarifies the basic functions to be executed, they can complete parameter configuration and function invocation through input and selection operations. The system then returns the results to the user in a visual manner, providing an effective bridge for users to use the system architecture.[10, 11]

Second, the service layer. This layer effectively connects the user layer and the resource layer and executes various business functions based on the PyTorch deep learning framework and VTK visualization tools. After receiving the user's parameter and function selection instructions, the service layer sends requests to the resource layer to use computer resources and then runs the corresponding algorithms to execute the functions. After data processing, the results are fed back to the user layer for presenting the image data.

Finally, the resource layer. This layer provides display data resources for the user layer and includes the physical hardware resources used by the service layer when executing business functions.

From a practical application perspective, the overall system design uses Python as the development language and Windows 10 as the development platform. Multiple libraries such as PyQt5 and VTK are selected for system development and design. After testing, it was piloted in a hospital in a certain area for feedback. The final results prove that it can assist doctors in completing disease analysis and treatment through medical imaging diagnosis and can also be used for professional medical education and guidance. From a practical application perspective, the fused images generated by the algorithm and system architecture in this study have high-quality color and contrast that conform to human visual perception, with clear overall structure and rich texture details. The final presentation effect meets the real-time requirements of clinical application scenarios.

4. Conclusion

In summary, multimodal medical images play a significant role in clinical diagnosis and related teaching. The auxiliary system based on multimodal medical image fusion technology can solve the inconvenience for doctors to observe multiple images and present comprehensive pathological information, thereby improving the efficiency and quality of medical diagnosis.

References

[1] Juanjuan Wang, Jie Ma, Chunyan Yi, et al. Diagnostic performance analysis of ductal carcinoma in situ of the breast based on deep learning multimodal imaging examination [J]. Journal of Medical Imaging, 2024, 34(4):49-53.

[2] Shan Xiong, Wenze Wu, Sibin Liu, et al. Application value of deep learning-assisted diagnostic system for rib fractures in emergency elderly patients [J]. Imaging Science and Photochemistry, 2024, 42(5):443-450.

[3] Huichang Yu, Shiyuan Liu. Application of deep learning in super-resolution reconstruction of magnetic resonance images [J]. Chinese Journal of Medical Physics, 2024, 41(10):1243-1248. DOI:10.3969/j.issn.1005-202X.2024.10.008.

[4] Chengyi Qian, Yuanjun Wang. Research progress on deep learning-based classification of Alzheimer's disease imaging [J]. Journal of Spectroscopy, 2023, 40(2):220-238. DOI:10.11938/cjmr20223013.

[5] Dingyang Lv, Weibing Shuang. Research progress on precision diagnosis and treatment of renal cancer using radiomics and deep learning [J]. Journal of Urology (Electronic Edition), 2024, 16(2):50-54.

[6] Yiman Liu, Xiaoxiang Han, Yuqi Zhang, et al. Application of deep learning in automatic intelligent recognition of standard sections in pediatric cardiac ultrasound [J]. Journal of Naval Medical University, 2023, 44(7):822-829. DOI:10.16781/j.CN31-2187/R.20220936.

[7] Shiwei Zhong, Dayou Wei, Chaojun Wu, et al. Application value of deep learning ultrasound radiomics in distinguishing between mass breast inflammation and breast cancer [J]. Chinese Journal of Ultrasound

Medicine, 2023, 39(12):1341-1344. DOI:10.3969/j.issn.1002-0101.2023.12.006.

[8] Wenjiang Wang, Jiaojiao Li, Zimeng Wang, et al. Application of deep learning integrated T2WI and DCE-MRI model in classification of breast lesions [J]. Clinical Radiology Journal, 2024, 43(11):1871-1876.

[9] Skaramagkas V, Pentari A, Kefalopoulou Z, Tsiknakis M. Multi-Modal Deep Learning Diagnosis of Parkinson's Disease-A Systematic Review. IEEE Trans NeuralSyst Rehabil Eng. 2023;31:2399-2423.

[10] Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deeplearning in cancer diagnosis, prognosis and treatment selection. Genome Med.2021 Sep 27;13(1):152.

[11] Pfänder L, Schneider L, Büttner M, Krois J, Meyer-Lueckel H, Schwendicke F.Multi-modal deep learning for automated assembly of periapical radiographs. JDent. 2023 Aug;135:104588.