Progress in the Application of Deep Learning in Medical Image Recognition

Haoyu Tian^{1,a,*}

¹School of International, Guangdong University of Finance, Guangzhou, Guangdong, China a. 17825859072@163.com *corresponding author

Abstract: With the continuous progress of deep learning technology, its application in the field of medical image recognition has made remarkable progress. The purpose of this review is to discuss the effectiveness, application value and future prospects of deep learning in current medical image recognition tasks. The application of deep learning technology in medical image recognition has shown broad prospects, and it is expected to play an important role in more practical scenarios in the future and promote the further development of intelligent vision technology. This paper first introduces the optimization methods, data sets and evaluation indicators of medical image recognition, and finally summarizes the existing limitations and prospects for the future, and makes a summary of this paper. In conclusion, deep learning has shown great potential in medical image recognition and is expected to play a key role in advancing the field, with promising prospects for future applications in clinical practice.

Keywords: Deep learning, Medical image recognition, Convolutional neural network, Feature extraction, Pattern recognition.

1. Introduction

In the past decade, the rapid development of deep learning technology has revolutionized the research landscape of medical image recognition. Traditional medical image recognition methods mainly rely on manual feature extraction and shallow learning models, which are often inadequate in the face of complex scenes and large-scale data. Deep learning, especially the emergence of convolutional neural networks (CNN), enables computers to automatically learn multiple layers of features in images to achieve higher recognition accuracy and accuracy. Medical image recognition, as an important application field of deep learning, has made rapid development in recent years. However, the characteristics of medical images, such as small amount of data, complex data distribution and high cost of annotation, make model optimization a key research direction. By applying optimization methods such as transfer learning, attention mechanism, regularization and reinforcement learning, the performance of the medical image recognition model can be significantly improved and practical problems in medical scenes can be solved. At the same time, the research of medical image recognition largely relies on high-quality data sets and scientific and reasonable evaluation indicators. Data sets provide a basis for training and testing algorithms, while evaluation indicators are used to quantify model performance, thus providing researchers with directions to optimize the model. In the

 $[\]odot$ 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

medical field, due to the complexity of data and the high requirements of clinical applications, it is particularly important to select the right data sets and indicators.

2. Medical image recognition optimization methods

2.1. Transfer learning and pre-training

Transfer learning is a technique in which trained model knowledge is applied to new tasks, especially in scenarios where the amount of data is insufficient or the tasks are similar. In medical image recognition, the cost of data annotation is high and the amount of data is limited, so transfer learning is particularly important. By pre-training the model on large-scale public datasets, researchers are able to take advantage of the general features it learns to provide a good starting point for medical image tasks [1].

The key to transfer learning lies in "transfer", that is, how to effectively apply the knowledge from the original task to the target task. For example, in the lung nodule detection task, researchers can use the ResNet model pre-trained on the ImageNet dataset for feature extraction and then adapt the specific features of the lung CT images by fine-tuning [2]. This approach not only reduces the need for large-scale annotated data but also shortens the training time for the model. In addition, transfer learning has also achieved good results in tasks such as breast cancer screening and brain tumor detection. Some studies have shown that using pre-trained models can achieve similar performance on small-scale datasets as training models on large-scale datasets.

However, transfer learning also faces some challenges. For example, there may be "task bias", or inconsistencies in feature space, between the source task and the target task, which may lead to unsatisfactory transfer effects. To solve this problem, the researchers propose a domain adaptive technique, which uses the feature distribution of the target task to adjust the pre-trained model. In addition, how to balance the general features of the pre-trained model and the specific features of the target task is also the direction that transfer learning needs to focus on.

2.2. Attention Mechanism

Attention mechanism has become an important research direction in deep learning in recent years. Its core idea is to improve the model's ability to focus on important areas by assigning higher weight to key features. In medical image recognition, the attention mechanism can help the model locate the focal area more accurately, and improve the accuracy and efficiency of diagnosis.

In specific applications, attention mechanisms can be divided into two forms: channel attention and spatial attention. Channel attention finds the most distinguishing feature by analyzing the different channels of the feature map; While spatial attention focuses on key areas in the image. For example, the Squeeze-and-Excitation(SE), a typical channel attention mechanism, significantly improves the model's classification and segmentation performance by adjusting the weights of different channels in the feature map [3]. SE modules have been successfully applied to tasks such as lung nodule detection and breast cancer classification.

In addition to channel and spatial attention, self-attention mechanisms have also made significant advances in medical image analysis in recent years. For example, the Transformer architecture addresses the difficulty of convolutional neural networks in capturing long distance dependencies by modeling relationships between features at a global scale through self-attention mechanisms. In the brain tumor segmentation task, the model based on the Transformer can effectively separate different tumor regions and improve segmentation accuracy [4]. In addition, the attention mechanism is also used to generate more explanatory models, enabling the model to provide a clearer diagnostic basis.

Although attention mechanisms have achieved good results in medical image recognition, they are computationally expensive and may not be suitable for resource-constrained scenarios. Future

research could explore lightweight attentional modules while further enhancing their adaptability to complex medical tasks.

2.3. Regularization

Regularization techniques are widely used in deep learning to solve overfitting problems by limiting the complexity of a model to improve its generalization ability. In medical image recognition tasks, regularization techniques are particularly important due to the small amount of training data.

Common regularization methods include L2 regularization, Dropout, and data enhancement. L2 regularization improves the stability of the model by limiting the size of the weight values and preventing the model from overfitting on small sample data sets. Dropout is a technique that randomly removes a portion of neurons during training, which can effectively reduce the network's overreliance on specific features. For example, in a lung nodule classification task, using Dropout was able to significantly reduce the risk of overfitting the model and improve performance on the test set.

Data enhancement technology generates diverse training samples by rotating, flipping, cropping and other operations on the original image, which is an important means to deal with the problem of small samples. For example, in the task of skin lesion detection, more diversified training data can be generated by adjusting the Angle and scaling ratio of the lesion area, thus improving the robustness of the model [5]. In addition, the hybrid data enhancement method further improves the generalization ability of the model by fusing features from multiple images.

However, the application of regularization techniques also requires balance. For example, overregularization can cause the model to underfit, which degrades its performance. In addition, how to design regularization strategies for specific medical tasks is also an important problem to be solved in the future.

2.4. Reinforcement Learning

Reinforcement learning is a technique that optimizes decision-making strategies by interacting with the environment and getting reward signals. In recent years, reinforcement learning has gradually shown important application potential in medical image recognition. Different from traditional deep learning methods, reinforcement learning can dynamically adjust strategies to adapt to complex task requirements.

In medical image recognition, the application of reinforcement learning is mainly focused on automated detection and segmentation tasks. For example, in tumor detection tasks, reinforcement learning models can guide the model to more accurately locate the lesion area by setting a reward mechanism. Studies have shown that reinforcement learning can significantly improve the detection accuracy of the model while reducing unnecessary computing overhead. In addition, strategy optimization methods combined with Deep learning, such as deep Q-Learning and strategy gradient method, also show good performance in endoscopic image analysis and real-time medical systems [6].

Reinforcement learning has also been widely used in robot-assisted diagnosis and surgical navigation systems. For example, in surgical navigation tasks, reinforcement learning algorithms can dynamically adjust surgical paths according to real-time input medical image data to improve surgical accuracy and safety. At the same time, reinforcement learning can also combine multi-modal data (such as CT and MRI) to optimize the diagnostic process, thereby improving the overall efficiency of diagnosis and treatment.

Despite the promising application of reinforcement learning in medical image recognition, its development still faces some challenges. For example, the training of reinforcement learning algorithms usually requires a lot of computational resources and time, and may not be suitable for

resource-limited scenarios. In addition, how to design a reasonable reward mechanism to make the model learn the optimal strategy is also a difficult point in reinforcement learning research.

3. Data sets and evaluation indicators

3.1. Data Sets

Medical image recognition tasks cover a wide range of disease types and imaging modes, so a rich and diverse dataset is needed to support them. Here are some typical datasets that are widely used in medical image recognition research:

The ILung Image Database Consortium and Image Database Resource Initiative (LDC-IDR) is a dataset of lung CT images for the detection of lung nodules [7]. This dataset was annotated by multiple radiologists, including the size, shape and malignancy score of the nodules. The LIDC-IDRI data are of high quality and detailed labeling information, providing important support for the early detection of lung cancer. The study shows that this dataset is of great value for the development of deep learning-based lung nodule detection models.

Digital Database for Screening Mammography (DDSM) is a mammography dataset focused on breast cancer screening [8]. DDSM contains a large number of mammography images while providing detailed lesion location and categorical labeling information, including mass boundaries and pathologic findings. Breast cancer screening is a task of clinical importance, and DDSM provides researchers with high-quality training and test data, which promotes the development of breast cancer detection technology.

The Brain Tumor Segmentation Challenge (BraTS) is an MRI data set mainly used for the segmentation of brain tumors [9]. This dataset covers multimodal MRI images, including T1, T2, and FLAIR sequences, while providing detailed segmentation annotation for different regions of the tumor (e.g., enhanced and necrotic regions). The BraTS dataset is widely used in brain tumor segmentation and classification tasks due to its comprehensiveness and high quality labeling.

International Skin Imaging Collaboration (ISIC) is a publicly available dataset for the detection and classification of skin lesions [10]. This dataset contains images of multiple types of skin lesions, particularly skin cancer cases, and provides corresponding labeling information. The emergence of the ISIC dataset has greatly promoted the research of skin cancer detection technology and spurred the development of related competitions and tools.

The NIH Chest X-ray Dataset (NCXD) is a large dataset of chest X-ray images containing over 100,000 images accompanied by label information for 14 common chest diseases, including pneumonia, pneumothorax, and tuberculosis [11]. This dataset is suitable for anomaly detection and multi-label classification tasks and provides an important resource for automated diagnostic studies of chest diseases.

Although these datasets provide rich support for medical image recognition research, they also have certain limitations. For example, the cost of tagging data sets is high, and the quality of tagging may be affected by human factors. In addition, differences in distribution between different datasets can lead to reduced generalization performance of the model, limiting the ability of the model to be applied to tasks across datasets. Therefore, more standardized, high-quality open data sets need to be developed in the future, and data sharing combined with technologies such as federated learning is needed to promote the further development of medical image recognition research.

3.2. Evaluation indicators

The evaluation index is an important tool to measure the performance of the medical image recognition model, and selecting the appropriate index can fully reflect the advantages and disadvantages of the model. Due to the diversity of medical image recognition tasks, different

evaluation indexes need to be used for different tasks to ensure that the model performance can be accurately described.

Accuracy is the most basic evaluation index, which is used to measure the proportion of samples predicted correctly by the model. It is intuitive and easy to understand, but in medical images, category distribution is often uneven, and relying on accuracy alone may not fully reflect the actual performance of the model [1]. For example, if most of the images in a dataset fall into the "normal" category, the accuracy may be high even if the model consistently predicts "normal", but the rate of missed diagnoses will be extremely high. Therefore, using accuracy alone as a single evaluation criterion is not sufficient.

Sensitivity and Specificity are indispensable indicators for a medical diagnostic task. Sensitivity reflects the ability of the model to correctly identify lesion samples, that is, the importance of avoiding missed diagnosis; Specificity measures the ability of the model to correctly exclude non-pathological samples, which is of great significance in reducing misdiagnosis [12]. For example, in breast cancer detection tasks, high sensitivity ensures that as many diseased patients are identified as possible, while high specificity reduces the likelihood that healthy patients will be misdiagnosed as having the disease. These two metrics are often used in combination to fully evaluate model performance.

The Dice coefficient is one of the most commonly used metrics in image segmentation tasks and is used to evaluate the degree of overlap between the segmentation results predicted by the model and the actual annotations. The higher the Dice coefficient is, the closer the segmentation effect of the model is to the real situation [13]. In medical tasks, such as brain tumor segmentation or lung nodule segmentation, the Dice coefficient can effectively reflect the model's ability to capture lesion areas. There may be other evaluation dimensions involved in segmentation tasks, such as the accuracy of segmentation boundaries, but the Dice coefficient is the most core index.

Receiver Operating Characteristic Curve (ROC) and Area Under the Curve (AUC) are common evaluation methods in classification tasks [4]. The ROC curve shows the performance of the model under different classification thresholds, while the AUC value summarizes the overall effect of the ROC curve through a numerical value. The closer the AUC value is to 1, the better the classification performance of the model. For example, in a lung nodule detection task, an AUC value can help researchers determine whether the model can accurately distinguish between healthy and diseased samples.

Mean square error (MSE) and mean absolute error (MAE) are mainly used for regression tasks in medical images. Such tasks often require the model to predict continuous values, such as calculating the volume of a lesion or the dimensions of an organ. MSE emphasizes large prediction errors, while MAE focuses more on the average error size per sample. Selecting the appropriate indicators in the task enables a more accurate assessment of the model's performance and guidance for improvement.

In the medical image recognition task, different indexes have their own applicable scenarios. For example, in the classification task, sensitivity is more concerned with the recognition of lesions, while specificity is more concerned with reducing misdiagnosis. For segmentation tasks, the Dice coefficient is the key metric, but the accuracy of boundary details is equally important. In the future, researchers will need to develop more comprehensive evaluation methods to more truly reflect the reliability and stability of the model in practical applications.

4. Common techniques of deep learning in medical image processing

4.1. Multi-task learning

Multitasking learning is a technique that improves generalization by integrating details from various tasks (which can be seen as soft constraints on parameters). Multitask learning can be a useful problem-solving method when one task has a large amount of labeled input data and can be shared

with another task with much less labeled data. For example, a multitasking learning problem can contain the same input patterns that can be used for many different output or supervised learning problems. In this configuration, each output can be predicted by a different part of the model, allowing the core of the model to generalize the same input for each task.

4.2. Active Learning

Active learning is usually an approach in which a model asks a human user operator questions to resolve uncertainties during the learning process. Active learning is a form of supervised learning that aims to produce the same or better results as so-called passive supervised learning, even if the model's data is more efficient. The core principle behind active learning is that allowing a machine learning algorithm to choose the data it learns allows it to achieve greater accuracy with fewer training labels. Active learners ask questions, often in the form of unlabeled instances of information that will be labeled by a prophecy machine. Active learning is a valuable tool when there is little data available and it is costly to collect or label new data. The active learning process allows domain sampling to be done in a way that reduces the number of samples while improving the validity of the model.

4.3. Online Learning

Machine learning is often done offline, meaning the operator has a batch of data and then improves on it. However, if there is streaming data, online learning is needed so that previous estimates can be updated as each new data point arrives, rather than waiting until the end (which may never happen). Online learning is useful because the data can change quickly over time. It is also useful for applications that involve a wide range of knowledge, even if the changes are gradual and increasing. In general, online learning aims to eliminate inconsistencies in how well the model performs relative to what it would have done if all available knowledge had been provided in batch form. So-called stochastic or online gradient descent used for artificial neural networks is an example of online learning. Stochastic gradient descent minimizes generalization error, and this is most easily seen in the case of online training, where examples or small batches are taken from a data stream.

4.4. Ensemble learning

Ensemble learning is a method of fitting at least two models into similar data and integrating the predictions of each model. Unlike any individual model, the goal of ensemble learning is to achieve better performance through groups of models. This includes a view of how to build the models used in the team and how to blend the models in the team's predictions together as best as possible.

5. Existing limitations and future perspectives

5.1. Existing Limitations

The application and development of medical image recognition technology face multiple challenges, among which insufficient data and data quality problems are particularly prominent. There are often great difficulties in obtaining medical image data, mainly due to legal requirements for the protection of patient privacy and the high cost of access, which makes the number of high-quality public data sets available for research limited. In addition, the annotation of medical images usually requires the participation of professional doctors, which is not only time-consuming and laborious, but also may lead to an insufficient amount of annotated data, limiting the training effect of the model. Even the existing public data sets still face the problem of category imbalance, for example, the number of healthy samples is usually much more than the pathological samples, which affects the performance of the model on the small sample category. Differences in imaging equipment between different

medical institutions can also lead to inconsistent image quality, which in turn affects the model's ability to generalize.

On the other hand, the interpretability of the model is also a problem to be solved. Although many deep learning-based models perform well on indicators such as accuracy, their decision-making process is still not transparent enough and is often seen as a "black box". For example, when the model identifies a certain lesion area, doctors may not understand why the model made that judgment, which makes the model lack credibility and transparency. This situation seriously affects the application of the model in clinical practice, especially in high-risk diagnostic tasks, where doctors need to have a clear basis for the model's results before they can adopt its recommendations.

In terms of computational resources and efficiency, the high resolution and multi-dimensional nature of medical image recognition put forward higher computational requirements. Threedimensional data such as CT and MRI images require a lot of computational resources, especially in the training and reasoning phase of deep learning models, which usually require high-performance hardware devices. However, in many resource-constrained healthcare Settings, such as hospitals in remote areas, high computing costs and hardware requirements limit the widespread deployment of these models. In addition, the reasoning speed of models is also particularly important in real-time medical scenarios, but the reasoning speed of many current high-precision models is slow, which makes it difficult to meet the needs of practical applications.

Cross-field application and standardization are also the bottleneck of medical image recognition technology development. Due to differences in format, resolution and imaging methods of image data generated by different hospitals and medical devices, models perform well in one particular environment but may be less effective in others. When training and testing with different datasets, model performance may decrease significantly. At the same time, the lack of a unified technical standard and evaluation system in the field of medical image recognition makes it difficult to directly compare different research results and also hinders the promotion and clinical application of the technology to a certain extent.

Finally, ethical and privacy issues have become increasingly important in medical image recognition technology. Medical image data involves sensitive private information of patients, and its use is subject to strict legal and ethical constraints. Although privacy-protecting technologies such as federal learning have made some progress, how to effectively use data while ensuring patient privacy remains a difficult problem that needs to be solved. At the same time, there may be biases in the model, especially in the case of unevenly distributed data, which may lead to poor diagnostic results for certain groups, leading to ethical controversies.

5.2. Future Outlook

The future development of medical image recognition technology faces many challenges, but it also provides broad prospects for improving the quality of medical services and diagnostic efficiency.

Firstly, the improvement of data acquisition and annotation methods is the key to solving the problem of insufficient medical image data. In the future, semi-supervised learning and unsupervised learning methods can reduce the dependence on labeled data, and high-quality medical image data can be generated by generative adversarial networks (GAN) and other technologies to further expand the scale of data sets. At the same time, the use of crowd-sourced annotation platforms and auxiliary annotation tools will also help improve the efficiency and accuracy of data annotation, and provide more annotation resources for researchers.

To solve the problem of insufficient explainability of the model, future research can integrate attention mechanism and visualization technology to make the model more intuitive and display its decision-making process. Combining multi-modal data (such as image data and electronic medical

records) further enhances the model's explanatory ability and provides doctors with more comprehensive diagnostic support.

When it comes to computing resources and deployment efficiency, the design of lightweight models is critical. Through techniques such as model compression, pruning and quantization, the complexity of models can be reduced to make them more suitable for resource-constrained environments. At the same time, by combining edge computing technology, some computing tasks can be assigned to the medical device end, which can reduce the burden of the central server and improve the real-time performance of the system. The exploration of distributed computing framework will also help medical image recognition technology run efficiently in large-scale environments.

In order to improve the generalization ability and application scope of the model, cross-field collaboration needs to be strengthened in the future. For example, data sharing between different medical institutions can be achieved through federal learning technology, more adaptable models can be developed, and a unified medical image recognition technology standard and evaluation system can be established to promote the promotion and application of the technology.

With the development of medical image recognition technology, the importance of enhancing privacy protection and ethical compliance has become increasingly prominent. In the future, technologies such as federal learning and encrypted computing can be combined to ensure that data is effectively utilized without exposing privacy. At the same time, it is necessary to establish an ethical review mechanism involving multiple parties to supervise the development and application process of the model to avoid potential ethical risks and social disputes.

The ultimate goal of the development of medical image recognition technology is to serve clinical practice. In the future, more attention should be paid to the deep integration of algorithms and clinical processes, and auxiliary diagnostic tools compatible with doctors' workflows should be developed, so that the results of model output can be directly applied to clinical decision-making. In addition, combined with robotics and augmented reality technology, medical image recognition can also play an important role in fields such as surgical navigation to provide patients with more accurate treatment plans.

All in all, there are still many challenges in the development of medical image recognition technology. However, by solving these problems and combining them with future research directions, medical image recognition technology is expected to become an indispensable and important tool in the medical industry.

6. Conclusion

The research and application of deep learning in the field of medical image recognition has made remarkable progress, showing strong performance and broad application prospects. By using deep learning models such as convolutional neural networks, the accuracy and efficiency of medical image recognition have been significantly improved, especially in complex scenes and large-scale data processing. Deep learning can automatically extract multi-level features and significantly improve the effectiveness of tasks such as classification, detection and segmentation. Despite the many successes of current technologies, challenges remain, such as insufficient training data, interpretability of models, high demands on computational resources, and overfitting in some cases. In the future, with the improvement of computing power and the continuous optimization of algorithms, deep learning will further play a role in medical image recognition. The research can foresee that deep learning will be combined with other technologies in fields such as augmented reality, driverless driving and intelligent manufacturing, leading to the birth of more innovative applications. Exploring more efficient training methods and model structures, as well as promoting

the development of small sample learning and self-supervised learning, are important directions for future research.

References

- [1] Zheng Yuan-Pan, Li Guang-yang, & Li Ye. (2019). Application of deep learning in image recognition. Computer Engineering and Applications, 55(12), 17.
- [2] Liu Fei, Zhang Junran, & Yang Hao. (2018). Research progress of medical image recognition based on deep learning. Chinese Journal of Biomedical Engineering, 37(1), 9.
- [3] Zhang Qianyu. (2018). Research and Application of deep learning in Medical image recognition (Master Dissertation, Taiyuan University of Technology). M.S.
- [4] Zhang Shuo, Zhang Rong, & Zhang Ybo. (2020). Research review on medical image recognition based on deep learning. Chinese Journal of Health Statistics, 37(1), 7.
- [5] Cao Zhenguan, Li Rui, & Zhang ZT. (2022). Deep learning-based lung medical image recognition. Journal of *Qiqihar University: Natural Science Edition*.
- [6] Zhang L F. (2020). Research on pathological image recognition technology based on deep learning (Master dissertation, Hangzhou Dianzi University). M.s.
- [7] Li, J., Jiang, P., An, Q., Wang, G. G., & Kong, H. F. (2024). Medical image identification methods: A review. Computers in Biology and Medicine, 169, 107777.
- [8] Thakur, N., Kumar, P., & Kumar, A. (2024). A systematic review of machine and deep learning techniques for the identification and classification of breast cancer through medical image modalities. Multimedia Tools and Applications, 83(12), 35849-35942.
- [9] Wu, X., Chen, H., Wu, X., Wu, S., & Huang, J. (2024). Retraction Note: Burn Image Recognition of Medical Images Based on Deep Learning: From CNNs to Advanced Networks.
- [10] Xiao, R., Zhang, Y., & Li, M. (2024). Automated High-Throughput Atomic Force Microscopy Single-Cell Nanomechanical Assay Enabled by Deep Learning-Based Optical Image Recognition. Nano Letters, 24(39), 12323-12332.
- [11] Liu, Y., Huang, S., & Wang, Z. (2024). Medical Image Recognition Based on Multiscale Cascade Segmentation Network MCSnet. Academic Journal of Computing & Information Science, 7(9), 60-66.
- [12] Li H. (2024). Research on Medical image recognition method based on Bayes neural network. (Doctoral dissertation, Chongqing University of Technology).
- [13] Zhang Dezhou, Wang Chaoqun, Shi Jun, Li Kun, Zhang Shaojie, & Ma Yuan et al. (2024). Characteristics and significance of age-related changes in uncinate process of cervical spine. Journal of Tissue Engineering, 28(36), 5766-5772.