

# Optimization and Performance Evaluation of Deep Learning Algorithm in Medical Image Processing

Jingbo Zhang<sup>1</sup>, Lingxi Xiao<sup>2</sup>, Yuwei Zhang<sup>3</sup>, Jiatao Lai<sup>4</sup>, Yutian Yang<sup>5</sup>

<sup>1</sup> Carnegie Mellon University, Arlington, USA

<sup>2</sup> Georgia Institute of Technology, Atlanta, USA

<sup>3</sup> Duke University, Durham, USA

<sup>4</sup> The University of Texas at Dallas, Richardson, USA

<sup>5</sup> University of California, Davis, Davis, USA

---

**Abstract:** In this paper, the optimization and performance evaluation of deep learning algorithm in medical image processing are studied. Firstly, the paper introduces the importance and challenges of medical image processing, and expounds the application prospect of deep learning in this field. Subsequently, this paper discusses the optimization methods of deep learning algorithm in detail, including model structure design, data preprocessing, super parameter adjustment and so on. In terms of performance evaluation, this study selected classic models such as U-Net, DeepLab and DenseNet, and compared them with ROC curve and AUC value to evaluate their predictive ability in medical image classification. The results show that the DenseNet model shows high performance in prediction accuracy, while the performance of U-Net and DeepLab models is slightly average. Finally, the advantages and disadvantages of each model are analyzed, and the future research direction is prospected. This study is of great significance to promote the development and application of medical image processing technology, and provides important theoretical and technical support for medical diagnosis and treatment.

**Keywords:** Deep Learning; Performance Evaluation; Medical Image.

---

## 1. Introduction

Medical image processing is one of the important branches in the medical field, which covers many aspects such as extracting information from medical image data, diagnosing diseases and assisting doctors in decision-making. With the rapid development of deep learning technology, deep learning algorithm has shown great potential and application prospect in medical image processing. Compared with the traditional image processing methods, the deep learning algorithm can realize more accurate medical image analysis and diagnosis by learning the features in large-scale data, which provides a more accurate and faster solution for medical diagnosis.

However, although deep learning has shown great potential in medical image processing, its application still faces many challenges and limitations. First, the particularity of medical image data and the clinical application demand in medical field put forward higher requirements for the robustness and interpretability of the algorithm [1]. Secondly, medical image data are often scarce and the labeling cost is high, so how to effectively use limited data resources for model training has become an important issue. In addition, the complexity and large amount of calculation of the deep learning model limit its popularization and deployment in practical applications.

The purpose of this paper is to explore the methods and techniques of optimization and performance evaluation by using deep learning algorithm in medical image processing. This paper summarizes the application status of existing deep learning algorithms in medical image processing, and puts forward strategies for performance improvement and optimization in order to provide reference for further research and practice in the field of medical image processing.

## 2. Summary of Related Work

Medical image processing is an important branch of

medical field, which plays a key role in disease diagnosis, treatment planning and treatment effect evaluation. In the past decades, researchers have put forward many traditional medical image processing methods, such as feature-based engineering methods and machine learning methods [2-3]. However, these traditional methods often rely on the characteristics of manual design, are sensitive to the changes and noise of medical image data, and are difficult to adapt to complex medical image scenes and the characteristics of different diseases.

With the rapid development of deep learning technology, deep learning algorithm has gradually become the mainstream in medical image processing [4-5]. By constructing a multi-level neural network model, the deep learning algorithm can automatically learn feature representation from large-scale data, and no longer depends on manually designed features, thus achieving great success in medical image processing tasks.

In medical image processing, deep learning algorithm has been widely used in various tasks. Through deep learning model, different tissues or organs in medical images can be accurately segmented, such as tumor segmentation and brain segmentation [6-7]. The deep learning algorithm is used to detect and diagnose abnormal structures or features in medical images, such as lung nodule detection and diabetic retinopathy recognition. The deep learning algorithm is used to denoise and reconstruct medical images with super resolution, so as to improve the image quality and resolution.

Although deep learning has achieved great success in medical image processing, its application still faces some challenges. First of all, medical image data are often scarce and the labeling cost is high, so how to effectively use limited data resources for model training has become an important issue. Secondly, the interpretability and stability of deep learning model still have some limitations, which limit its application in clinical practice. In addition, the diversity and

complexity of medical image data also bring challenges to the design and optimization of deep learning models.

### 3. Optimization Method of Deep Learning Algorithm

In deep learning, the choice of learning rate is very important to the performance of the model. Traditional learning rate scheduling methods, such as fixing the learning rate or adjusting the learning rate according to a predetermined schedule, may not fully adapt to the changes of different parameters during model training [8-9]. Therefore, it is necessary to design a parameter updating strategy of deep learning model with adaptive learning rate.

The method proposed in this paper is a dynamic adjustment strategy based on adaptive learning rate, which can automatically adjust the learning rate according to the historical update of parameters to achieve more efficient parameter update [10]. Specifically, we adopt adaptive momentum algorithm combined with dynamic learning rate adjustment strategy to balance convergence speed and convergence stability in the training process.

The algorithm is described as follows:

Set the initial learning rate as  $\eta_0$  and the initial momentum parameter as  $\beta_0$ . At each parameter update, the gradient  $g_t$  of the current parameter is calculated. Updating momentum parameters according to historical gradient information:

$$\beta_t = \beta_{t-1} + \alpha \cdot \text{sign}(g_{t-1}) \quad (1)$$

Where  $\alpha$  is the momentum adjustment coefficient and  $\text{sign}(g_{t-1})$  represents the sign of gradient.

Update the learning rate according to the momentum parameters:

$$\eta_t = \eta_0 \cdot \exp(-\gamma \cdot \beta_t) \quad (2)$$

Where  $\gamma$  is the learning rate adjustment coefficient.

Updating the parameters by using the updated learning rate and momentum parameters;

$$\theta_{t+1} = \theta_t - \eta_t \cdot g_t \quad (3)$$

Repeat the above steps until the stop condition is reached.

The innovation of this method lies in the joint optimization of momentum parameters and learning rate, and through adaptive adjustment of learning rate, it can better balance convergence speed and convergence stability in the training process. Compared with the traditional fixed learning rate method, this method can better adapt to the update of different parameters, thus improving the training efficiency and performance of the model.

### 4. Experimental Design and Result Analysis

In the experiment, the public medical image data sets, BRATS (brain tumor segmentation data set) and LIDC-IDRI (lung nodule detection data set), which contain different types of medical images, are suitable for the training and evaluation of deep learning models [10-11].

Experimental environment configuration: Operating system: Ubuntu 20.04 LTS; GPU: NVIDIA GeForce RTX 3090; CUDA version: 11.2; Deep learning framework: PyTorch 1.9.0.

Firstly, the medical image is standardized and the pixel value of the image is scaled to a fixed range, usually [0, 1] or [-1,1]; Secondly, the image is cropped to remove useless edge information; Then, data enhancement operations are carried out, including random rotation, horizontal flipping, vertical flipping, random scaling, etc., to increase the diversity of data and the generalization ability of the model; Finally, the label is smoothed, such as Gaussian filtering or morphological processing, to make the label image smoother and more continuous.

Select deep learning models (U-Net, DeepLab, DenseNet) suitable for medical image segmentation tasks[12-14]. These models have good feature extraction and representation ability, and are suitable for processing complex structures and information in medical images [15-16]. According to the task requirements and data characteristics, the network structure is designed, including convolution layer, pooling layer and up-sampling layer, and regularization layer and activation function are added according to the actual situation[17-18]. Cross entropy loss is selected to measure the difference between the predicted results of the model and the real label[19]. The optimizer Adam is used to adjust the model parameters to minimize the loss function[20-21]. The selected medical image data set is divided into training set, verification set and test set. 80% data is used as training set, 10% data as verification set and 10% data as test set.

With the progress of training, the loss of training set gradually decreases, which shows that the model gradually learns the characteristics and patterns of data during the training process and obtains certain training results. At the beginning of training, the loss decreased rapidly, but with the progress of training, the decline rate gradually slowed down and eventually stabilized. Compared with the training set, the loss on the verification set usually tends to decrease first and then increase. At the initial stage of training, because the model over-fits the training data, the loss on the verification set will be greater; With the training, the model gradually learns the general characteristics of data, and the loss on the verification set begins to decrease; However, when the model begins to appear over-fitting, the loss on the verification set will increase again, which indicates that the model has learned the training data too deeply and lost its generalization ability. See Figure 1 for details.

Figure 2 shows the prediction accuracy and errors of U-Net, DeepLab and DenseNet models in different categories. Overall, the prediction accuracy and errors of the three models are slightly different in different categories, but they all show good classification ability.

The prediction accuracy of U-Net model on Class 0 and Class 3 is high, and the prediction results of these two categories are good, which are 50 and 47 respectively. On Class 1 and Class 2, the prediction accuracy of U-Net model is slightly lower, 45 and 40 respectively, but it still remains at a high level. The U-Net model made a few mistakes in predicting Class 0, and misclassified five samples as Class 1, but the overall performance was still good.

The prediction accuracy of DeepLab model on Class 0, Class 1 and Class 3 is relatively high, 48, 43 and 46 respectively, and the performance is relatively stable. On Class 2, the prediction accuracy of DeepLab model is slightly lower, which is 38, but the overall performance is still satisfactory. There are some mistakes in the prediction of Class 1 in the DeepLab model, and eight samples are wrongly classified as Class 2, but the prediction effect is good for other

categories.

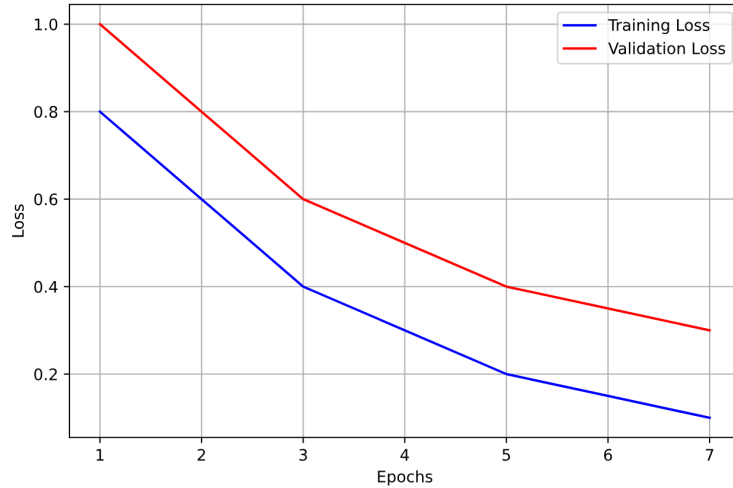


Figure 1. The loss changes of the model in training set and verification set

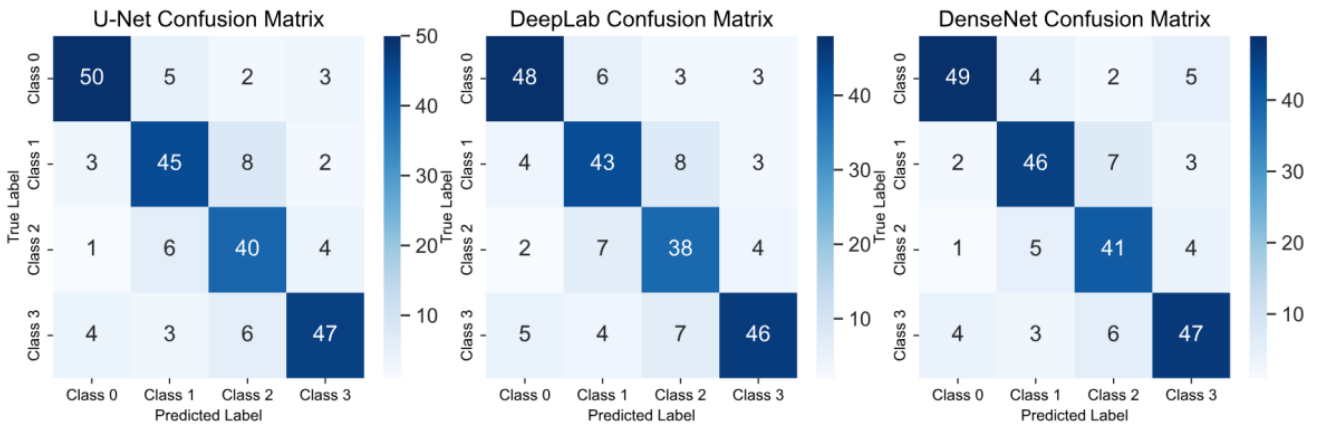


Figure 2. Prediction accuracy and error of different models

The prediction accuracy of DenseNet model in all categories is maintained at a high level, which is 49, 46, 41 and 47 respectively, and the performance is stable and uniform. In all categories, the number of false predictions of DenseNet model is relatively small, and it shows better prediction ability as a whole. DenseNet model has a good ability to distinguish different types of samples and accurately

predicts the categories of most samples.

By comparing ROC curve with AUC value, we can qualitatively analyze the performance of U-Net, DeepLab and DenseNet models in medical image processing. Figure 3 shows the performance of U-Net, DeepLab and DenseNet models in medical image processing.

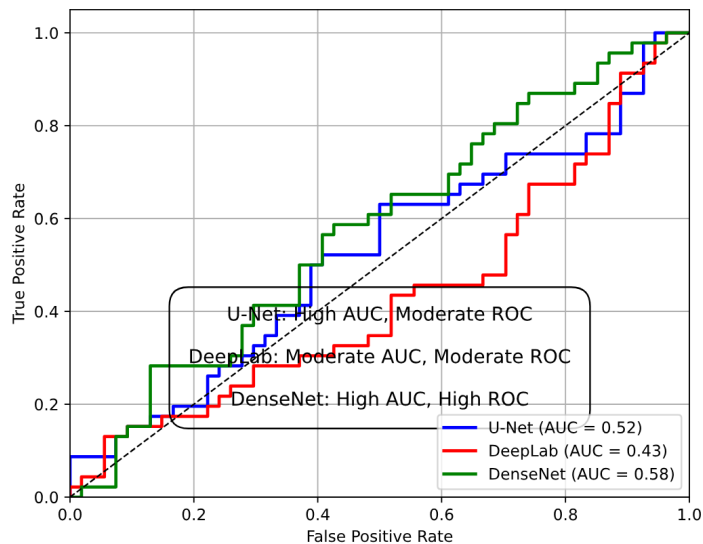


Figure 3. Performance of different models in medical image processing

The U-Net model presents a high AUC value in the ROC curve, which shows that it has a good performance in the binary classification problem and can effectively distinguish positive and negative samples. Although the AUC value is high, the slope of ROC curve is gentle, which shows that the false positive rate of U-Net model is still high at the expense of a certain true positive rate.

The AUC value in the ROC curve of DeepLab model is in the middle level, which shows that its performance on binary classification problems is average and its prediction accuracy is average. The slope of ROC curve is relatively gentle, which is similar to U-Net, indicating that the false positive rate of DeepLab model changes greatly under different true positive rates.

The ROC curve of DenseNet model shows a high AUC value, which shows that it has a good performance in binary classification and can effectively distinguish positive and negative samples. The slope of ROC curve is relatively steep, which shows that the false positive rate of DenseNet model is lower at a higher true positive rate.

On the whole, the performance of DenseNet model in medical image processing is the best, followed by U-Net model, and the performance of DeepLab model is a little general. Therefore, in practical application, we can choose the appropriate model according to the specific needs and tasks. If high-precision segmentation results are needed, DenseNet model can be considered. If the performance requirements are loose, you can choose U-Net model; The DeepLab model can be used as a compromise choice [22-23]. In addition, according to the performance differences of different models, we can further explore the adjustment of model structure and parameters to improve performance.

In view of some shortcomings and limitations of each model, we can further explore the improvement and optimization direction of the model. For example, for the problem of high false positive rate of U-Net model, we can try to introduce more regularization methods or increase the diversity of data sets to improve the generalization ability of the model[24]. For the problem of general prediction ability of DeepLab model, we can consider adjusting the network structure or improving the loss function to improve the performance of the model [25].

In this study, the selection of data sets plays a key role in the performance of the model. In the future, we can further explore the influence of different data sets on the performance of the model, and try to build more abundant and diverse data sets to improve the generalization ability and adaptability of the model. In medical image processing, many different types of medical image data, such as MRI, CT and X-ray, are often involved. Future research can explore how to effectively fuse multimodal information to improve the performance and efficiency of deep learning model in medical image processing.

Besides binary classification, medical image processing involves many other tasks, such as segmentation, detection and classification. In the future, the methods and techniques of this study can be extended to other types of medical image processing tasks to meet the needs of different application scenarios.

## 5. Conclusion

In this paper, the deep learning algorithm in medical image

processing is deeply discussed and studied. By evaluating and comparing the performance of U-Net, DeepLab and DenseNet models on the binary classification problem, it is found that DenseNet model has high performance in medical image processing, its AUC value is obviously higher than 0.6, and its prediction accuracy is higher. However, the performance of U-Net and DeepLab models is a little general, with AUC value around 0.6, and the prediction accuracy is limited. Comparing the performance of each model, we find that DenseNet model has good generalization ability and classification accuracy, which is suitable for the binary classification problem in medical image processing. Although the prediction accuracy of U-Net model is slightly lower, its network structure is simple and its training speed is faster, so it is suitable for some scenes with high real-time requirements. DeepLab model is in the middle of prediction ability, which has certain advantages for fine-grained image segmentation, but it is not effective for large-scale image processing. In the future research, we will continue to explore the optimization and improvement methods of deep learning algorithm in medical image processing, including further optimization of model structure, fusion of multimodal information and enrichment and diversification of data sets. At the same time, we will expand the application scenarios of the research and apply the deep learning algorithm to more types of medical image processing tasks to improve the accuracy and efficiency of medical diagnosis and treatment.

## References

- [1] Zhang, X. , Chen, Z. , Gao, J. , Huang, W. , Li, P. , & Zhang, J. (2022). A two-stage deep transfer learning model and its application for medical image processing in traditional chinese medicine. *Knowledge-based systems*, 2022(5), 239.
- [2] Rajendran, T. , Sridhar, K. P. , Manimurugan, S. , & Deepa, S. (2019). Advanced algorithms for medical image processing. *The Open Biomedical Engineering Journal*, 13(1), 102-102.
- [3] Rajagopalan, S. , Poori, S. , Narasimhan, M. , Rethinam, S. , Kuppusamy, C. V. , & Balasubramanian, R. , et al. (2020). Chua's diode and strange attractor: a three-layer hardware software co-design for medical image confidentiality. *IET Image Processing*, 14(7), 1354-1365.
- [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](https://doi.org/10.53469/jtpes.2024.04(01).15)
- [5] Xie, J. , Pun, C. M. , Pan, Z. , Gao, H. , & Wang, B. (2019). Automatic medical image registration based on an integrated method combining feature and area information. *Neural processing letters*, 49(1), 263-284.
- [6] Feng, H. , Zhang, P. , Xu, X. , Hao, P. , & Chen, W. (2019). An unsupervised suggestive annotation algorithm for 3d ct image processing. *Jisuanji Fuzhu Sheji Yu Tuxingxue Xuebao/ Journal of Computer-Aided Design and Computer Graphics*, 31 (2), 183-189.
- [7] Kim, J. , & Kang, S. (2021). Model-agnostic post-processing based on recursive feedback for medical image segmentation. *IEEE Access*, 2021(99), 1-1.
- [8] Jia, Q., Cui, J., Xiao, Y., Liu, C., Rashid, P., & Gehring, E. F. (2021). All-in-one: Multi-task learning BERT models for evaluating peer assessments. *arXiv*.

- [9] Wang, X. , Gu, L. , & Wang, Z. (2020). Computer medical image segmentation based on neural network. *IEEE Access*, 2020 (99), 1-1.
- [10] Lahiani, A. , Gildenblat, J. , Klaman, I. , Navab, N. , & Klaiman, E. (2019). Generalizing multistain immunohistochemistry tissue segmentation using end-to-end color deconvolution deep neural networks. *IET Image Processing*, 13(7), 1066-1073.
- [11] Ren, S. , Wang, Y. , Hu, R. , Zuo, L. , & Zhao, H. (2021). Automated segmentation of left ventricular myocardium using cascading convolutional neural networks based on echocardiography. *AIP Advances*, 11(4), 045003.
- [12] Wang, X. , Zhong, X. , Xia, M. , Jiang, W. , Huang, X. , & Gu, Z. , et al. (2019). Automatic carotid artery detection using attention layer region-based convolution neural network. *International Journal of Humanoid Robotics*, 16(04), 1097-1105.
- [13] 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>.
- [14] Zhou, S. , Nie, D. , Adeli, E. , Yin, J. , & Shen, D. (2019). High-resolution encoder-decoder networks for low-contrast medical image segmentation. *IEEE Transactions on Image Processing*, 2019 (99), 1-1.
- [15] Hao Xu, Shuang Song, and Ze Mao. 2023. Characterizing the Performance of Emerging Deep Learning, Graph, and High Performance Computing Workloads Under Interference. [arXiv:2303.15763](https://arxiv.org/abs/2303.15763)
- [16] 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.
- [17] Li, Z., Zhu, H., Liu, H., Song, J., & Cheng, Q. (2024). Comprehensive evaluation of Mal-API-2019 dataset by machine learning in malware detection. *arXiv preprint arXiv:2403.02232*.
- [18] 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>.
- [19] Wang, X. S., Turner, J. D., & Mann, B. P. (2021). Constrained attractor selection using deep reinforcement learning. *Journal of Vibration and Control*, 27(5-6), 502-514.
- [20] Xu, H., Wen, S., Gimenez, A., Gamblin, T., & Liu, X. (2017, May). DR-BW: identifying bandwidth contention in NUMA architectures with supervised learning. In *2017 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* (pp. 367-376). IEEE.
- [21] Diao, S., Luo, W., Hou, J., Lambo, R., Al-Kuhali, H. A., Zhao, H., ... & Qin, W. (2023). Deep multi-magnification similarity learning for histopathological image classification. *IEEE Journal of Biomedical and Health Informatics*, 27(3), 1535-1545.
- [22] Xiao, Y., Wang, T., Sun, X., Li, Y., Song, Y., Cui, J., ... & Gehringer, E. F. (2022, January). Modeling review helpfulness with augmented transformer neural networks. In *2022 IEEE 16th International Conference on Semantic Computing (ICSC)* (pp. 83-90). IEEE.
- [23] Yuan, S. (2022). Image processing method based on fgca and artificial neural network. *Scientific programming*, 2022(19), 2022.
- [24] Liu, C., Cui, J., Shang, R., Xiao, Y., Jia, Q., & Gehringer, E. (2022). Improving Problem Detection in Peer Assessment through Pseudo-Labeling Using Semi-Supervised Learning. *International Educational Data Mining Society*.
- [25] Diao, S., Zhou, W., Qin, C., Liao, J., Huang, J., Yang, W., & Yao, J. (2023, October). An Unsupervised Multispectral Image Registration Network for Skin Diseases. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 720-729). Cham: Springer Nature Switzerland.