A method integrating multiple detection models for MICCAI CL-Detection 2024 Challenge

junqiangchen[1][1]

Mediworks, Building 7,No 1985, Sanlu Road,Min Hang,Shanghai,China
{chenjunqiang@mediworks.biz

Abstract. This paper discusses the latest advances in key point detection in cephalometric radiographs, aiming to improve detection accuracy and efficiency. In view of the shortcomings of existing methods under complex anatomical structures and individual differences, we propose a multiple detection integration model, which significantly enhances the detection performance of key points by introducing different model detection fusion technologies. Experimental results show that the accuracy and real-time performance of the proposed method on multiple data sets are better than the current mainstream algorithms. Our research not only provides a reliable tool for clinical brain image analysis, but also lays the foundation for subsequent related research and promotes the application of automated detection technology in the medical field. Our method achieved an average MRE error of 1.70mm and an average SDR score of 89.46% for the cephalometric landmark detection on the validation set using a NVIDIA GeForce RTX 3090. The average running time was 10 seconds for one image. The code is available at https://github.com/junqiangchen/Cl-Detection2024.

Keywords: Integrated detection · Unet · Vnet.

1 Introduction

Lateral cephalogram plays an important role in clinical diagnosis, but due to its complex anatomical structure and individual differences among patients, key point detection faces great challenges.

In recent years, with the rapid advancement of deep learning and computer vision technology, many studies have begun to explore automatic detection methods for cephalometric landmarks, such as the combination of convolutional neural networks (CNN) and traditional image processing technology. Although there have been some progress, there are still problems with insufficient accuracy and real-time performance.

The motivation of this study is to improve the accuracy and efficiency of key point detection in lateral cephalograms. We propose a detection model based on the integration of multiple network architectures and verify its effectiveness through a large amount of clinical data. This contribution not only provides a more reliable tool for clinical practice, but also lays the foundation for in-depth exploration of related research fields.

2 Proposed Method

Use multiple U-Net models to detect head key points.

Fig. 1 shows a typical example of U-Net [1].

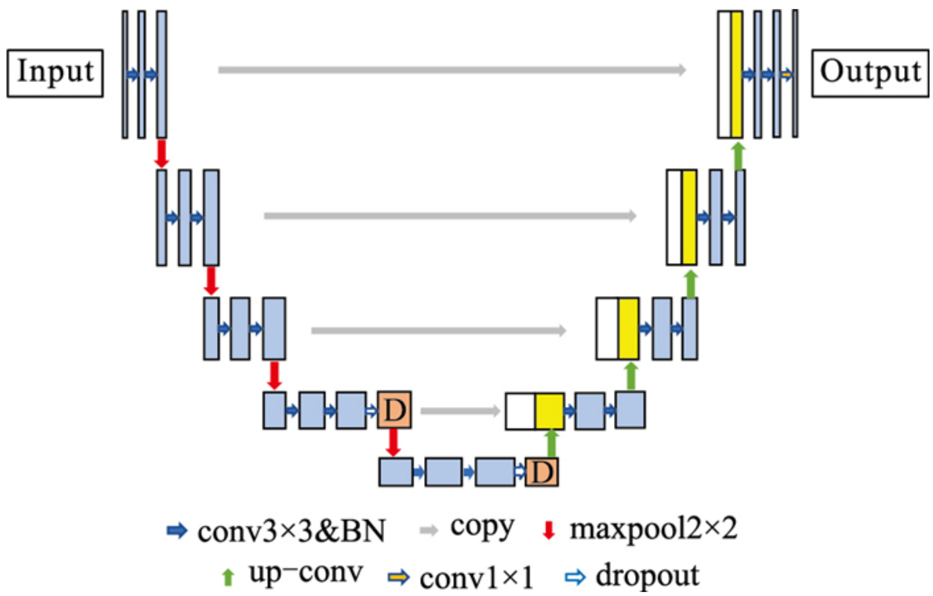

Fig. 1. U-Net Network architecture

2.1 Sub-module

U-Net network structure description Encoder (downsampling) part: consists of multiple convolutional layers and maximum pooling layers. Each convolution block usually includes two convolutional layers (usually 3x3 convolution), followed by an activation function (such as ReLU), and then a maximum pooling layer (usually 2x2). With each layer, the number of feature maps increases and the size is halved. Bottleneck layer: located between the encoder and the decoder, contains two convolutional layers, and the number of feature maps reaches the maximum value. Decoder (upsampling) part: gradually increase the size of the feature map through upsampling layers (such as transposed convolution). Each decoding layer is skipped with the corresponding encoding layer, and the feature map of the encoding layer is concatenated with the feature map of the decoding layer. Each decoding block also contains two convolutional layers and activation functions. The output layer finally compresses the number of channels of the feature map to the required number of categories through 1x1 convolution

2.2 Loss function

we use the summation between $\mathbf{L}_1$ loss because this loss functions have been proven to be robust in various medical image landmark detection tasks.

2.3 Post-processing

Combine the heat map results of multiple detection models to get the final heat map result, and get the key point coordinates based on the result

3 Experiments

3.1 Dataset

The detection targets cover various landmarks. The dataset of CL-Detection 2024 challenge comprises approximately 700 dental X-ray lateral images from four different centers. This dataset is organized into four distinct subsets: training, validation, test, and independent test sets. The training, validation, and test sets all derive from the ISBI challenge dataset, the PKU cephalogram dataset, and the Shenzhen University General Hospital dataset. The training set includes 396 images, the validation set has 50 images, and the test set consists of 150 images. The independent test set features 100 images obtained from a private medical center, offering a unique evaluation benchmark. Dental professionals have annotated 53 landmarks for each image, including 13 soft tissue-related landmarks, 6 tooth-related landmarks, 19 skull-related landmarks, 13 cervical spine-related landmarks and 2 calibration ruler landmarks.

The image is scaled to a fixed size of 1024x1024, and the image is normalized with a mean of 0 and a variance of 1. The 53 key point coordinates are scaled to a scale of 1024x1024 according to the image ratio, and then the 53 key point coordinates are used to generate a 53-channel Gaussian heat map with a Gaussian sigma parameter of 10.

3.2 Evaluation metrics

For evaluation metrics, check out the Challenge's official online introductory page, which clearly gives the evaluation metrics, as well as the final score calculation. The address of the competition is: CL-Detection 2024 Challenge

3.3 Implementation details

Environment settings The development environments and requirements are presented as follows: The system is running Ubuntu 18.04.5 LTS as the operating system. The CPU in use is an Intel(R) Core(R) Core(TM)i9-14900K CPU with a clock speed of 3.20GHz. The system has a total of 64GB RAM, divided into 16 modules of 4GB each, operating at a speed of 2.67MT/s. The system is equipped with four NVIDIA 3090 24G GPUs. The CUDA version installed on the system is 11.4. The programming language used for development is Python 3.8.3. The deep learning framework employed includes torch 1.12.1, torchvision 0.13.1.2. These specifications provide insight into the hardware and software setup used for the development of a specific project or application.

Training protocols Please describe at least the following aspects: Unet2d network uses AdamW optimizer, learning rate is 0.001, batch size is 2, epoch is 300, and loss function uses focal loss and L2.

4 Results and discussion

Table 1. Quantitative evaluation results. **The results should correspond to your final algorithm submission. The online validation denotes the leaderboard results. The Testing set and Independent Testing set results will be released during MICCAI. Please leave them blank at present.** You can use a similar Table format to present the ablation study results of the public and online validation. A useful online tool to create latex table https://www.tablesgenerator.com/latex_tables.

Method	Online Validation		Hidden Test		Independent Test	
	MRE (mm)	SDR@2mm (%)	MRE (mm)	SDR@2mm (%)	MRE (mm)	SDR@2mm (%)
U-Net	3.141 ± 3.141	31.415 ± 3.145	3.141 ± 3.141	31.415 ± 3.141	3.141 ± 3.141	31.415 ± 3.141
Ohter algorithm						
Your baseline model	2.856	0.73				
Your final model	1.65	0.779				

Note to Table 1: if you have multiple solutions, such as a faster model with lower SDR or a slower model with higher SDR, you can use a similar Table format to report the performance on the online validation set.

4.1 Quantitative results on validation set

the MRE value is 1.65mm and SDR metrics is 0.779 on the validation set.

4.2 Results on hidden testing set and independent testing set

This is a placeholder. We will send you the testing results during MICCAI conference.

4.3 Limitation and future work

The defects of current deep learning methods

Strong data dependence: The performance of deep learning models is highly dependent on high-quality, large-scale training data. The labeling of lateral skull radiographs is complex and time-consuming, and it is difficult to obtain high-quality labeled data.

Limited generalization ability: The trained model may not be able to generalize well when facing images of different devices, different patients, and different disease states.

Poor interpretability: The deep learning model is a black box, and it is difficult to explain the specific reasons why the model makes decisions, which is particularly important in the field of medical diagnosis.

Insufficient sensitivity to detect small lesions: For some subtle lesions, deep learning models may be difficult to accurately identify.

High computing resource consumption: Deep learning models usually require a lot of computing resources, which limits their application in actual clinical practice.

Future work plan

Data enhancement and synthesis: Through data enhancement technology and generative adversarial networks (GAN) and other methods, expand the training data set and improve the robustness of the model.

Model optimization and improvement: Explore new network structures, loss functions, and optimization algorithms to improve the performance and efficiency of the model.

Multimodal fusion: Fusion of lateral head radiographs with other medical imaging data (such as CT and MRI) to improve the accuracy of diagnosis.

Interpretability research: Enhance the interpretability of the model through visualization, attention mechanism and other methods to help doctors better understand the decision-making process of the model.

Federated learning: Under the premise of protecting patient privacy, use federated learning technology to integrate multi-center data and train more powerful models.

Model miniaturization: Develop lightweight deep learning models for mobile medical devices to reduce computing resource consumption.

Multi-task learning: Combine multiple related medical imaging tasks (such as segmentation, classification, and detection) for multi-task learning to improve the overall performance of the model.

Adversarial training: Improve the robustness of the model to adversarial samples and enhance the security of the model through adversarial training.

Clinical verification and application: Combine deep learning models with clinical practice, conduct large-scale clinical verification, and promote their application in actual medical care.

5 Conclusion

Deep learning has broad application prospects in the field of lateral head radiographs. By solving the above problems, we can develop more accurate, reliable and efficient deep learning models to provide strong support for clinical diagnosis. In the future, with the continuous development of deep learning technology and its cross-integration with other disciplines, lateral head radiographs will achieve more breakthrough progress.

Acknowledgements. The authors of this paper declare that the segmentation method they implemented for participation in the CL-Detection 2024 challenge has not used any additional datasets other than those provided by the organizers. The proposed solution is fully automatic without any manual intervention. We thank all the data owners for making the X-ray images and CT scans publicly available and Codebench [2] for hosting the challenge platform.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241 (2015) 2
2. Xu, Z., Escalera, S., Pavao, A., Richard, M., Tu, W.W., Yao, Q., Zhao, H., Guyon, I.: Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. Patterns **3**(7) (2022) 6