课程设计报告书

题目： A spine segmentation model based on an UNet combining 2D and 3D modules

学　院　生物医学科学与工程学院

专　业　生物医学工程

学生姓名　周钰斌

学生学号　202164030408

指导教师　吴凯

课程编号　081100161

课程学分　1

起始日期　2023-12-18

教 师 评 语	教师签名： 日期：
成 绩 评 定	
备 注	

Automatic spine segmentation based on an UNet combining 2D and 3D modules

1. Background

The spine is the most complex load-bearing structure in the human body, and its health affects the quality of life.MRI is a well-established imaging method to examine spinal diseases, but image segmentation is often required to identify the lesion site and analyze it quantitatively. Due to the poor resolution and contrast of MR images, they are susceptible to noise, artifacts and local volume effects, which makes automatic segmentation of MR spine images a challenging task.

The mainstream image segmentation method is to train neural network models using labeled data. Some 3D MRI images are characterized by thick-slice scanning, which will reflect discontinuity in a certain dimension, making the data be anisotropic. The normal 2D/3D convolutional neural network (CNN) cannot well synthesize and extract the sparse inter-slice information and dense intra-slice information. This poses a challenge to the design of neural networks.

In this course design, this design tries to construct a neural network using 2D/3D convolutional blocks, and proposes to balance the extraction of information by extracting fine 2D features in the encoder-decoder structure of UNet, and 3D inter-slice information in the residual connections in the middle. After ablation experiments, the method proposed in this design can be well adapted to the provided thick-slice scanning spine MRI dataset.

2. Design concepts

Considering that the data is anisotropic and has both 2D and 3D attributes, when designing the network we should take into account both the model's ability to segment individual slices and the model's ability to integrate information between slices. Observing the dataset and reviewing the literature focusing on this dataset, we can get that the slices have richer information and can be segmented as a single 2D image, while the differences between the slices are larger. Based on this consideration, this design proposes a UNet built by 2D and 3D modules, using 2D modules for fine feature extraction in the encoder-decoder structure that performs down-sampling and up-sampling in the UNet, and adding 3D modules in the connection between the encoder and the decoder in order to exchange information between slices. In practical tests, this scheme can consider both intra-slice and inter-slice information, and has higher accuracy than the model based on pure 2D neural network.

All the experiments were conducted on a 2 x NVIDIA Geforce GTX 3080 Graphics Processing Unit (GPU) setup. The deep learning model was implemented using the PyTorch API version 1.12.1, in CUDA version 11.3, compatible with the GPU, and scripted in Python. Details of the runtime environment are provided in a later link.

3. Process of this design

This design uses the N4BiasFieldCorrection function of the SimpltITK library to correct the bias field of MRI images for preprocessing.

This design proposes to build the UNet from 2D and 3D modules. for the 2D module, the UX-Net Block proposed by Ho Hin Lee et al. is chosen[1], which borrows the idea of swin-transform and uses a larger convolutional kernel to provide a larger sensory field, which enhances the global modeling capability. This design changes its convolution kernel to 2D to fit the dataset used in this design. A 3D-ResBlock is designed in the downsampling layers after the UX-Net Blocks to integrate the 3D information, and it also be used in upsampling layers. 3D-ResBlock is designed based on the ResNet model by He et al.[2] In the residual connection section, the HANCBlock proposed by Nabil Ibtehaz and Daisuke Kihara is chosen for this design[3], which improves the expressive power of the convolutional block by including the reverse bottleneck in the convolutional block and also provides an approximate notion of neighborhood comparisons by calculating the mean and maximum of neighboring pixel features to simulate the attentional mechanism. This design applies HANCBlock on 3*3*n neighborhoods of different thicknesses to integrate 3D information between slices. In this design, this block has been modified.

The network structure in use is shown as Figure 1.

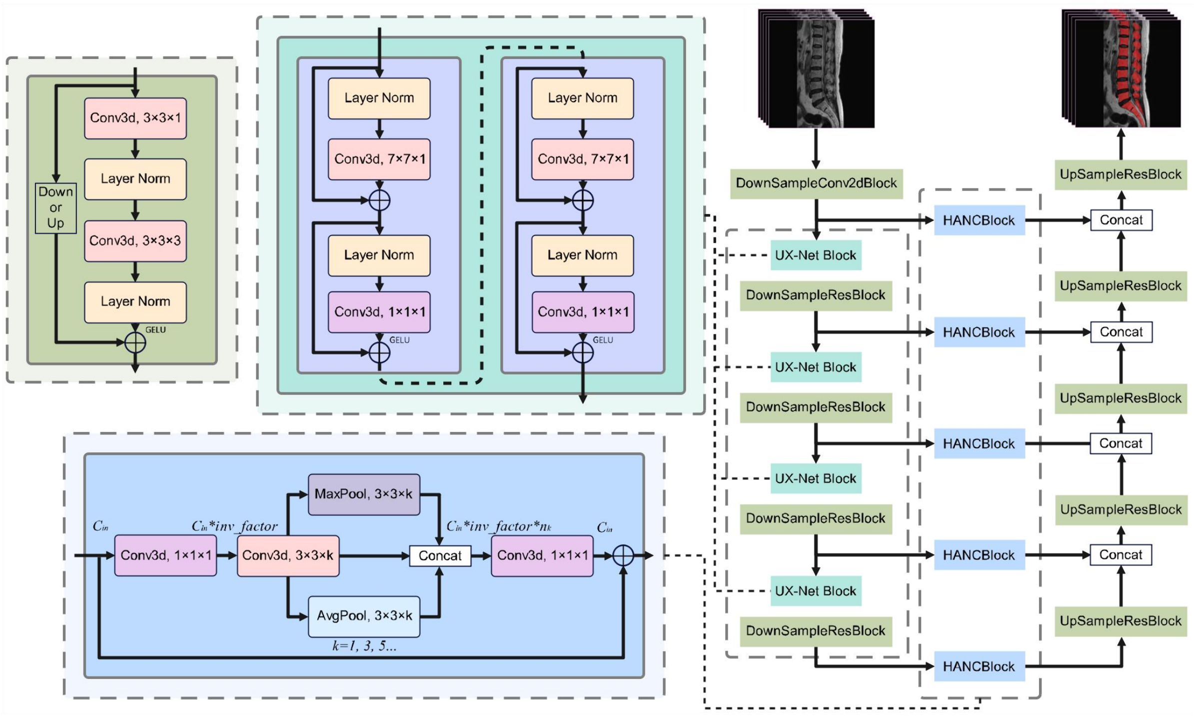

Figure 1. The architecture of the neural network used in this design. The left side of the figure shows the structure of 3D-ResBlock, UX-Net Block and HANCBlock and the right side of the figure shows the architecture of the complete neural network

This design uses the Mean Square Error Loss Function (MSE) for image segmentation tasks. At the same time, to prevent the loss from fluctuating up and down when converging to the global optimum due to excessive learning rate, the training model adopts the Adam optimizer and learning rate exponential decay method, so that the learning rate decreases as the training progresses. The initial learning rate is set to 0.001, the decay step is 10, and the decay rate is 0.9.

Due to the large dimension of the data (mostly 880 * 880 * 12), the graphics memory of the computing device used in this design cannot support the model to calculate the complete image. Meanwhile, due to the small sample size in the dataset, the model may experience overfitting. Therefore, the model training in this design adopts a random cropping regularization method, applying random 128 * 128 * 12 windows to the original image to obtain training data. At the same time, 40% of the samples are randomly selected in each epoch, and 16 windows are randomly extracted from each sample as training data to prevent overfitting. Considering the time factor, this design sets the batch size for training to 8, which is the maximum value that the computing device can support. After verification, the model converges when there are about 150 Epochs. We set the Epoch to 200, which means that after 200 rounds of training, the model will be saved and the performance indicators of the model will be tested.

4. Result

In addition to the proposed network structure, this design also tests the performance metrics of the model after replacing UX-Net Block and HANCBlock with ResBlock, ablation experiments are performed to verify the superiority of UX-Net Block and HANCBlock. In addition to this the performance of the classical 2D image segmentation network UNet++ on this dataset is also tested. The performance of each model as determined by the five-fold cross-validation are presented in Table 1.

Table 1. Comparison of the performance of different models

	UNet++	Our Model (only use ResBlock)	Our Model (without UX-Net Block)	Our Model (without HANCBlock)	Our Model
DSC	0.9058	0.9071	0.9031	0.9116	**0.9119**
PPV	0.9301	0.9370	0.9258	**0.9386**	0.9361
Sensitivity	0.8828	0.8790	0.8815	0.8862	**0.8890**

*Note: DSC: Comparison of all models. The best scores are styled as bold.

The confusion matrix for ACC-UXNet is as follows

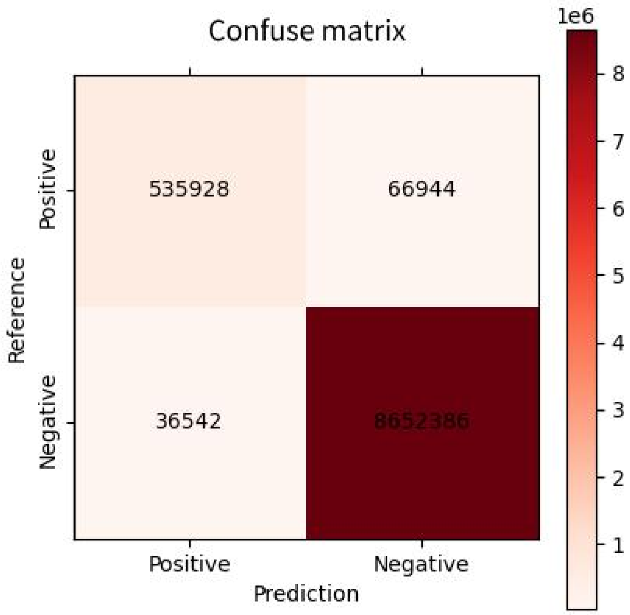

Figure 2. Confusion matrix of our model

5. Summary of the course design

In this course design, I fully realized the superiority of UNet in medical image segmentation: while exploring the UNet method, I also tried to fine-tune SAM (Segment Anything Model) to classify the dataset used in this course design, however, the result was not satisfactory, with the DSC around 0.88, which is lower than UNet++. The most rewarding aspect of this course design was the experience of fine-tuning a large model and a better understanding of the powerful performance of UNet.

The code for this course design can be found at https://github.com/MurasameDaisuki/mip_scut_course_design.

References

[1] Lee H H, Bao S, Huo Y, et al. 3d ux-net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation[J]. arXiv preprint arXiv:2209.15076, 2022. https://arxiv.org/abs/2209.15076

[2] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C] Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[3] Ibtehaz N, Kihara D. ACC-UNet: A Completely Convolutional UNet Model for the 2020s[C] International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 692-702.