CN-Unet:A Robust Network Based On Deep Convolution For Medical Image Segmentation

Wei Liu[1], Junwei Li [1], Zhiwei Ye [1], Orest Kochan[1,2]

[1] *Hubei University of Technology ,Wuhan, China*
[2] *Lviv Polytechnic National University, 12 S. Bandera Str., Lviv, 79013, Ukraine*

Abstract

With the rapid development of deep learning, traditional medical image segmentation methods are gradually eliminated. The mainstream tasks of medical image segmentation include tumor segmentation, multi-organ segmentation, cardiac segmentation, and retinal segmentation. For the common multi-category segmentation problem in medical segmentation, due to the large differences in individual shapes and textures between categories, it is difficult to achieve segmentation. Taking medical segmentation as an example, we propose an efficient and powerful network architecture for medical segmentation, CN-Unet. CN-Unet is a U-shaped symmetric network based on deep convolution, and its basic unit comes from CN-Block in ConvNeXt. To cope with small object segmentation in medical images, we design a multiple data augmentation module, in which the slice fusion branch can subtly capture the adjacent information of medical slices. Experiments on two public datasets (Synapse and ACDC) show that the segmentation ability of CN-Unet outperforms other state-of-the-art methods.

Keywords

Medical Image Segmentation, Deep learning, Adjacent Information,Data Augmentation

1. Introduction

As technology evolves, various fields of our lives have benefited to varying degrees; deep learning has also begun to penetrate more professional areas, such as the military, medicine, education, transportation, and other regions [1]. As an essential task in the medical field, medical image processing has always received significant attention from medicine and some interdisciplinary experts [2]. Therefore, combining deep learning and medical image processing to solve various subtasks on medical images has become a hot topic in recent years [3]. Due to the rapid development of medical imaging equipment, medical imaging equipment such as magnetic resonance imaging (MRI), computed tomography (CT), and X-ray imaging has gradually become essential in medical image analysis. As the primary method of medical image analysis, medical image segmentation can assist doctors in obtaining information about organs or lesions, which is of significance for a series of medical analysis tasks such as disease observation, treatment plan formulation, and anatomical structure modeling. Due to the characteristics of medical images, it has a series of problems, such as complex image format, difficulty acquiring data sets, and difficulty extracting features. Therefore, medical image segmentation remains a challenging task.

When AlexNet first appeared, it won the ImageNet LSVRC-2010 championship, and its accuracy far exceeded the second place. Since then, the craze of the convolutional neural network has risen to a new height. The convolutional neural network has also indicated that it is about to become the trend in the image field. In 2015, Ronneberger et al. [4] proposed a U-shaped network structure (U-Net) for medical image segmentation, and its excellent segmentation performance and ingenious network structure attracted the attention of scholars. Since then, the U-shaped structure's application in medical

COLINS-2023: 7th International Conference on Computational Linguistics and Intelligent Systems, April 20–21, 2023, Kharkiv, Ukraine
EMAIL:liuwei001@vip.sina.com (W. Liu);lijunwei7800@gmail.com (J. Li); weizhiye121@126.com (Z. Ye); orest.v.kochan@lpnu.ua (O. Kochan)
ORCID: 0009-0003-2912-5405 (W. Liu); 0009-0009-1122-3051 (J. Li); 0009-0006-3234-7377 (Z. Ye); 0000-0002-3164-3821 (O. Kochan)

images has been extensively recognized [5]. As the most classic neural network in medical image segmentation, U-Net is characterized by a symmetrical U-shaped structure and skip connections that can combine features. The U-shaped system consists of an encoder and a decoder. The encoder can extract features from the input original image layer by layer from shallow to deep; the decoder can restore the extracted features to the original size layer by layer through upsampling operations. The skip connection can fuse the features of different levels; the purpose is to reduce the loss of information in the recovery process to achieve a better segmentation effect.

Currently, the medical image segmentation methods with deep learning backgrounds are divided into 2D and 3D segmentation [6]. In the 2D segmentation method, we generally decompose the 3D medical image data into many 2D slices, then input each slice into the segmentation model; the model will output the segmentation effect. 2D segmentation has strong generalization ability, 2D segmentation model also has better transferability, and the training process has the advantages of fewer parameters and faster training speed. In contrast, 2D image segmentation may lose some contextual information in data processing and feature extraction. In the 3D segmentation method, the data is superimposed by multiple layers of slices. Compared with the 2D data, there is one more z-axis, so the (x, y, z) three directions are encoded in the convolution process. This may allow the model to obtain richer feature information, but 3D segmentation will consume more memory and obtain several times the number of parameters of 2D segmentation. Due to the matching problem between the amount of data and the number of model parameters, the 3D segmentation model needs more data to train. Otherwise, it may lead to overfitting problems [7].

In this paper, we propose a depthwise convolution-based 2D medical image segmentation architecture, CN-Unet. CN-Unet is a U-shaped symmetric architecture based on deep convolution based on CN Block as the basic unit. To take full advantage of the powerful feature extraction capability of CN Block, we divide the encoder and decoder of CN-Unet into four stages, and set the CN Block ratio of several stages to (3, 3, 9, 3). We add skip connections between each symmetry stage to recover the contextual information of the feature maps. To improve the segmentation ability of CN-Unet for small and medium organs in Synapse, we propose a multiple data augmentation module in which the slice fusion branch is presented according to the characteristics of medical images. We use two different datasets, Synapse and ACDC, to evaluate the segmentation performance of CN-Unet. We achieve the best results on the Synapse dataset and break through 80% DSC on the semantic class of gallbladder. To verify the feasibility of our structure, we do in-depth comparative experiments on the Synapse dataset, and our base CN-Unet architecture outperforms ConvNeXt [8] in results.

2. Related works

2.1. Medical image segmentation method based on deep learning

With the advent of convolutional neural networks, deep learning-based segmentation models have been widely used in medical impact. In 2015, Ronneberger et al. pioneered the coder-decoder structure network, U-Net [4]. U-Net is the first deep learning network applied to medical image segmentation. Its unique structure and excellent segmentation performance later became the benchmark for medical image segmentation networks. To make U-Net's skip connections work more fully, U-Net series networks have been proposed successively (U-Net+, U-Net++, U-Net+++). Since 3D methods have become popular in the image field, medical imaging scholars have also begun to apply 3D network models to various tasks in medical image analysis, such as 3D-U-Net and V-Net. With the advent of visual transformers (VITs), researchers began to try to combine Transformers to design medical segmentation models. TransUNet proposed by Chen et al. [15], is the first medical segmentation model that combines CNN [16] and Transformer. The authors used it for multi-organ and heart segmentation, and its performance was better than the state-of-the-art network. In the past two years, Transformer-based medical image segmentation network structures have begun to emerge, such as Swin-Unet [9], MISSFormer [11], UNETR [12], nnFormer [10], etc. These advanced structures have all performed segmentation tasks on datasets such as multiple organs and have achieved positive results in DSC values. Until the emergence of ConvNeXt, Liu et al. [8] designed a convolutional neural network according to the idea of Swin Transformer. They proved through many experiments that the

convolutional neural network is not worse than the transformer-based networks, even better than the Transformer structure in some aspects.

2.2. Data Augmentation Methods in Medical Image Segmentation

Access to medical data has always been challenging due to the specialized nature required to label medical data. In this context, the augmentation of medical data is significant. Methods used for medical image enhancement typically have transformations such as rotation, random cropping, elastic deformation, and inversion, which generate a training image that resembles a specific training example. With the rapid development of deep learning, the effect of these commonly used data augmentation methods on model performance appears to be gradually missing. In 2019, Wang et al. [21] proposed a theoretical formulation of test-time augmentation for deep learning and applied it to medical image tasks. However, many enhancement methods, such as DAGAN augmentation [19], have not been popularized in the downstream tasks of medical images. The method generates semantic maps through labels, and adds semantic maps and labels to the training set in pairs, which is an advanced and effective data augmentation method.

2.3. Method

CN-Unet is an encoder-decoder structure, and its overall architecture is shown in Fig 1. It has the same U-shaped structure as the classic U-Net, mainly composed of encoder, bottleneck, decoder, and skip connections. The basic unit of CN-Unet is the CN block. Specifically, the encoder consists of one embedding layer, three CN modules, and two downsampling layers. Symmetrically, the decoder branch contains three upsampling layers and three CN modules. Furthermore, the bottleneck consists of a downsampling layer and two CN modules to provide a large receptive field to support the decoder. Inspired by U-Net, we design a symmetric encoder-decoder structure and add skip connections between the feature cones of each corresponding CN module. The fusion of multi-scale features helps recover fine-grained details in predictions to compensate for the loss of spatial information caused by downsampling.

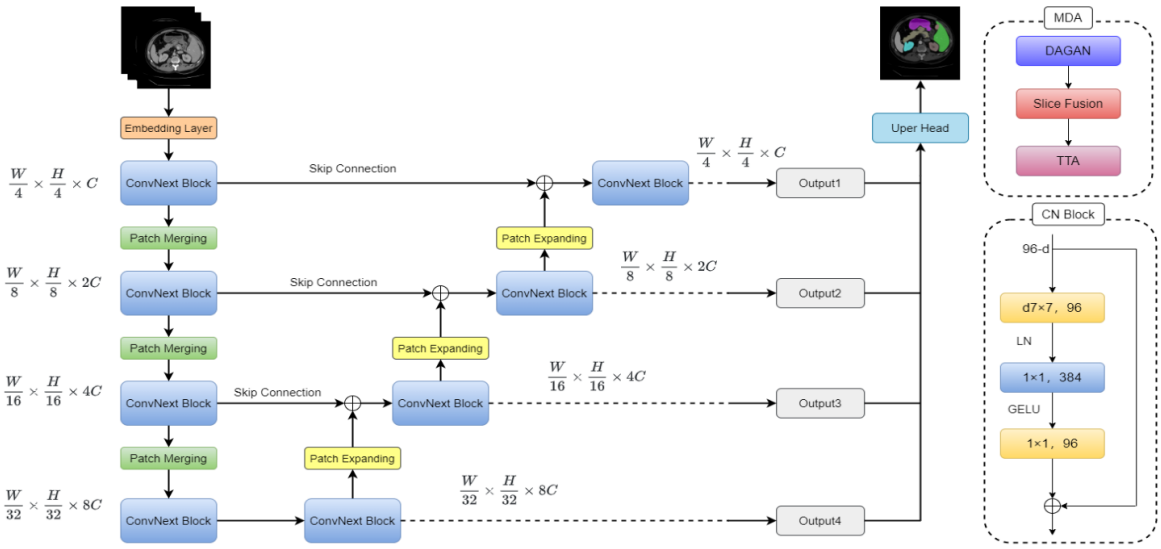

Figure 1: The archietecture of CN-Unet

2.4. Encoder

The input of CN-Unet is the 2D slice $x \in R^{H \times W}$ after slice fusion (x is obtained by scaling, cropping, and rotating the original image), and H and W represent the height and width of each input

scan. Slices of size 512×512 are sent to the Embedding Layer. The Embedding Layer contains a downsampling layer and a normalization layer. Unlike the general average pooling and maximum pooling, our downsampling layer is composed of The convolution operation is completed with a kernel size of 4×4 and a stride of 4. Same as CN Block, Embedding Layer uses layer normalization. After passing through the Embedding Layer, the number of input channels increases to 96, and the normalization is completed. At this time, the data will pass through the first CN Block. On both sides of the entire symmetric structure, we set the stage scale of CN Block to (3, 3, 9, 3), so the first CN Block layer contains three independent CN Block blocks. We call the matrix entering the CN Block x_{in}. Before x_{in} passes the inversion bottleneck, we will perform a 7×7 depthwise separable convolution operation on it, then pass through an LN layer. The process formula is as follows:

$$y_1 = LN\big(SepConv_{7\times7}(x_{in})\big), \tag{1}$$

Then y_1 first performs the permute method to transform the dimension and enters the bottleneck layer. The bottleneck layer contains two fully connected layers and a GELU activation function. The order is FC, GELU, FC. After that, it will continue to use permute to transform the dimension, then go through a DropPath layer to delete the multi-branch structure randomly. Finally, adding x_{in} to the matrix obtained here is a complete CN Block operation, and its formula is as follows:

$$x_{put} = DropPath\Big(FC\Big(GELU\big(FC(y_1)\big)\Big)\Big), \tag{2}$$

The ratio range of DropPath is set between 0 and 0.4. Since the stage ratio of CN Block is (3, 3, 9, 3), we average it into 18 regions in the range of 0 to 0.4, of which the first DropPath of the CN Block is about 0.022, the DropPath of the second CN Block is about 0.044, and so on, the DropPath of the last CN Block of the encoder part is 0.4.

In the U-shaped structure, the encoder usually has a down-sampling operation to increase the receptive field and highlight the features of the image to better extract shallow features. The entire encoder structure of CN-Unet contains three downsampling layers located in the middle of every two CN Block layers. In order to obtain better fusion information and achieve the effect of feature extraction and feature dimensionality reduction, we use 2×2 convolution downsampling in the downsampling stage. We set the step size of convolution downsampling to 2, which can obtain enough receptive field and avoid over-sampling. Since we put four stages for both the encoder and the decoder, the feature size of each stage of the encoder will increase by four times, so each downsampling layer is double downsampling.

2.5. Decoder

The decoder of CN-Unet also contains four stages, maintaining symmetry with the encoder. The ratio of (3, 3, 9, 3) is also maintained in the layout of CN Blocks. After obtaining the shallow feature information, the feature map first passes through a bottleneck containing three CN Blocks, then passes through the first one. Upsampling layer. The traditional upsampling method generally uses a preset interpolation method, so the preset interpolation method cannot bring better learning ability to the network. In CN-Unet, our upsampling layer uses transposed convolution to extract deep features, then concatenates a layer for normalization to make the gradient more stable. Before connecting the next CN Block layer, we use skip connections to connect the feature maps from the upsampling layer with the feature maps from the CN Block at the same stage in the encoder. Assuming that the feature map before entering the first upsampling layer in the decoder is x_1, the input of the next CN Block is x_2, and the output of the corresponding CN Block in the encoder is y_1, we can use the following formula to express the above Operation of sampling layers and skip connections:

$$x_2 = Concat\big(GELU\big(Deconv_{4\times4}(x_1)\big), y_1\big), \tag{3}$$

In the decoder, this procedure is performed three times to restore the feature maps of each stage to the same size as the encoder. In each CN Block layer stage of the decoder, the jump connection between the upsampling layer and the encoder is passed before the input. After that, the number of channels of the feature map changed from C to 2C. To restore the number of channels to C, we add a convolutional layer inside each CN Block layer of the decoder; the convolution kernel size is 3×3, the stride is set to 1, and the padding is set to 1. In this way, we complete the design of the entire decoder. Finally, we

send the output of each CN Block layer of the decoder to the Upper Head to complete the medical image segmentation.

2.6. Multiple Data Augmentation Modules

In the process of solving downstream tasks in the image field, data augmentation is usually used to increase the number of samples and enrich the diversity of samples. At the same time, data augmentation can also improve the model's generalization ability. Due to the medical professionalism and privacy required for medical image labeling, it is tough to obtain medical data sets, which is also one of the main reasons why the problems in the field of medical image processing are difficult to solve [18]. To avoid letting the lack of data affect the model's efficiency, we propose multiple data augmentation modules specifically for medical data.The MDA module contains the following three branches:

DAGAN data augmentation:Before inputting the images to CN-Unet, we randomly selected the labels of 1280 samples, then fed them into the DAGAN [19] network for a semantic generation. Through the DAGAN network, we can obtain an additional 1280 generated pictures, then expand these generated pictures and the corresponding labels of the input into the data set. The specific steps are shown in Figure 2.

Slice Fusion augmentation:When cutting medical CT images, we have some understanding of the characteristics of medical data. After completing the cutting of a single case, we quickly switch the picture to check whether the cutting effect is ideal roughly. Interestingly, this group of slices is like a video with a timeline, and we can see the internal changes in this case. This means that we can regard this case as a whole, then after cutting it into several slices, each of its adjacent slices has a specific relationship. If we feed the slices into the network structure one by one, we lose this association, the so-called contextual information. Our approach is to merge three adjacent single-channel grayscale slices into a single 3-channel RBG image so that we will effectively preserve the contextual information of the data. Of course, we can also directly convert a single-channel slice into a 3-channel image, but this simply copies the information of a single slice, which not only fails to capture the information between adjacent slices, but also causes data redundancy.

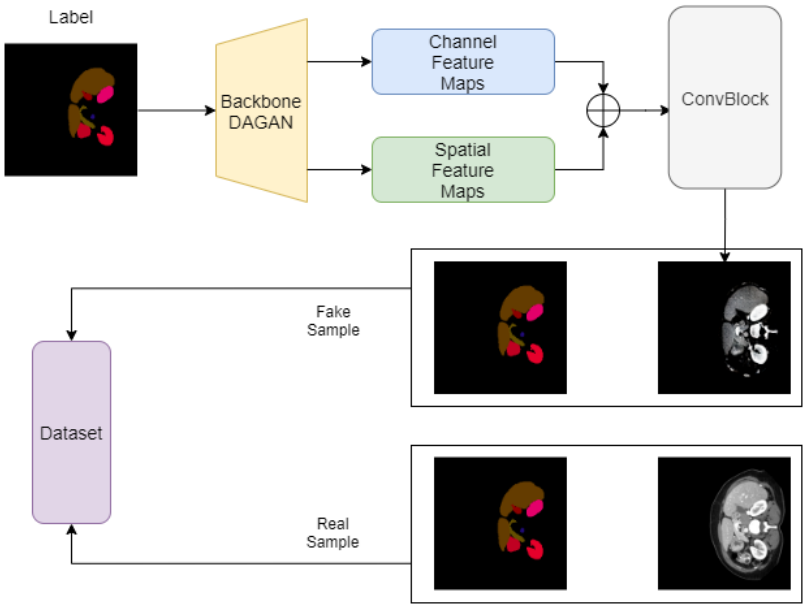

Figure 2: The process of DAGAN data enhancement

Test-Time Augmentation:We introduce test-time data augmentation [21] in the multiple data augmentation module to further reduce the generalization error at training time. The augmentation in Section 4.3 is first performed on the test image, consistent with what was done during training. We let

the model make predictions for each augmented picture and return a set of those predictions, which is the step of de-augmentation. Finally, performing a merge operation completes the test-time data enhancement. The effect of TTA will be discussed later, with the results in Table 4.

2.7. Loss function

This paper uses weighted cross-entropy (WCE) loss to replace the commonly used cross-entropy loss. The WCE loss is a variant of the CE loss. All positive samples are multiplied by a coefficient for weighting. This loss function is widely used in class-imbalanced problems. The formula is as follows:

$$loss_{WCE} = -\sum w * y_i * \log(logitS_i) + (1 - y_i) * \log(1 - logitS_i)\,, \tag{4}$$

The Synapse dataset's training set and the background class have nine categories. The number of samples in each category is n_i, and i ranges from 1 to 9. The median balance method can be used to calculate w.

3. Experiments

In this section, to verify the effectiveness of CN-Unet, we conduct experiments on two commonly used datasets: Synapse Multi-Organ Segmentation and Automatic Cardiac Diagnosis Challenge (ACDC). We chose to compare with current state-of-the-art ConvNets and transformer-based architectures, using the reported results to explore the research space in the field of medical image segmentation and demonstrate the superiority of CN-Unet.

3.1. Datasets

Synapse for multi-organ CT segmentation.: The dataset consists of abdominal clinical CT scan images of 30 patients, which contains 3779 axial abdominal images. After using the split in [15], 18 sample cases were constructed as the training set, and the remaining 12 sample cases were divided into the test set. Using the mean Dice Similarity Coefficient (DSC) and mean Hausdorff distance (HD) [28] as evaluation metrics, we assessed CN- Unet performance.

ACDC for automated cardiac diagnosis: The ACDC challenge obtained examinations of 100 patients using an MRI scanner. The MR image is a series of short-axis slices, the heart area is covered from the left atrium to the apex, and the thickness of the slices is 5 to 8 mm. Each data was manually labeled for left ventricle (LV), right ventricle (RV), and myocardium (MYO). The entire dataset was split into 70 training samples (1930 axial slices), ten validation samples, and 20 test samples. Like [9], we evaluate our method using the average DSC index.

3.2. Evaluation metrics

We use Dice score and 95% Hausdorff distance (HD95) [28] to evaluate the accuracy of segmentation results. The Dice score is calculated based on the accuracy and sensitivity of the test samples and is a result of balancing the two criteria. Hausdorff distance is often used as a segmentation indicator, mainly used to measure the segmentation accuracy of the boundary. For a given semantic class, we denote the ground truth and predicted values of each pixel as Xi and Yi, respectively, and X' and Y' represent the surface point set of ground truth and predicted values. The evaluation formula of Dice score and HD is as follows:

$$Dice = \frac{2\sum_{i=1}^{I} X_i Y_i}{\sum_{i=1}^{I} X_i + \sum_{i=1}^{I} Y_i} \tag{5}$$

$$\mathrm{HD} = \max\{\mathrm{d_{XY}}, \mathrm{d_{YX}}\} = \max\left\{\max_{x\in X}\min_{y\in Y}(x,y), \max_{y\in Y}\min_{x\in X}(x,y)\right\}, \quad (6)$$

95% HD is similar to Max HD. However, it is based on calculating the 95th percentile of the distance between boundary points in X and Y. The purpose of using this measure is to remove the effect of a minimal subset of outliers.

3.3. Experimental details

CN-Unet is implemented on Python 3.6, PyTorch 1.8.1, and Ubuntu 20.04. For data preprocessing on Synapse, we adopt the method introduced in Swin-Unet. For all samples in the training set, data augmentation methods such as random cropping, scaling, and rotation are used to increase the diversity of samples. In the training phase, we use a batch size of 2, and the input image size is 256×256×3. We used the AdamW optimizer to train the model, set the initial learning rate to 0.0001, and set the weight decay to 0.05 to prevent overfitting during gradient descent; exponential decay of the first and second moment estimates The rates were set to 0.9 and 0.999, respectively. All training procedures are performed on an NVIDIA 3080ti GPU with 12GB memory.

Table 1
Compared with other advanced methods

Method	HD95	DSC	Aor	Gall	Lkid	RKid	Liv	Pan	Spl	Sto
TransUNet	31.69	77.48	87.23	63.13	81.87	77.02	94.08	55.86	85.08	75.62
CoTr	27.38	78.08	85.87	61.38	84.83	79.36	94.28	57.65	87.74	73.55
Swin-Unet	21.55	79.13	85.47	66.53	83.28	79.61	94.29	56.58	90.66	76.60
U-NETR	22.97	79.56	89.99	60.56	85.66	84.80	94.46	59.25	87.81	73.99
MISSFormer	18.20	81.96	86.99	68.65	85.21	82.00	94.41	65.67	91.92	80.81
nnFormer	15.80	86.56	92.13	70.54	86.50	86.21	96.88	**83.32**	90.10	86.83
nnUNet	13.69	86.79	**93.20**	71.50	84.39	88.36	97.31	82.89	91.22	85.47
ConvNeXt	**12.43**	87.14	89.86	68.73	**94.65**	**93.32**	96.48	73.03	93.62	87.43
CN-Unet	12.53	**90.23**	91.74	**81.23**	94.64	93.13	**96.63**	78.08	**94.69**	**91.70**
P-values	<1e-2(HD95),<1e-2(DSC)									

3.4. Results

To evaluate our CN-Unet, we compare current state-of-the-art medical segmentation methods with CN-Unet, including TransUNet, CoTR [24], Swin-Unet, UNETR, MISSFormer, nnFormer, and nnUNet [25]. The segmentation results of all models are shown in Table 1; our CN-Unet achieves the best scores of 90.23 and 10.53 on the average Dice score and the average HD95 score, respectively. The Dice score of 90.23 shows that our CN-Unet segmentation's accuracy is the best among all models. In contrast, the HD95 score of 10.53 represents the superior performance of CN-Unet on organ edge segmentation. Excluding ConvNeXt, the average Dice score and average HD95 score of the second nnUNet are 86.79 and 13.69, respectively. In contrast, our CN-Unet is 3.44 and 3.16 higher than nnUNet, a remarkable achievement for Synapse Improve. Regarding semantic categories, our CN-Unet achieves the best Dice scores on the three categories of aorta, spleen, and stomach, outperforming the second-place Dice scores by 9.73, 2.77, and 4.87, respectively. The gallbladder breaks the 80% level on Synapse for the first time with a Dice score of 81.23, which shows that our CN-Unet has significantly improved the segmentation of small organs. From Figure 4, we can see that CN-Unet outperforms other methods in segmenting small objects and is comparable to nnFormer in edge detection. The boxplots in Figure 5 show that CN-Unet has higher upper and lower quartiles than the rest of the methods, which validates the average superiority of CN-Unet over Synapse's categories.

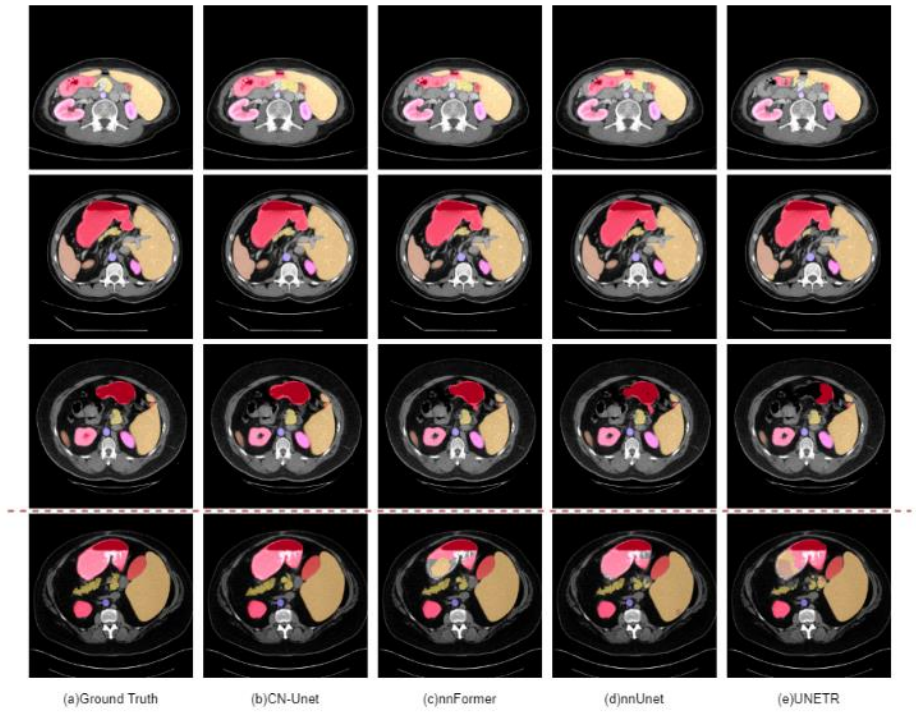

Figure 3: Visual comparison with current state-of-the-art methods on the Synapse dataset.

3.5. Ablation experiment

As shown in Table 1, the U-shaped symmetric structure based on CN Block is 0.9 higher than the tiny version of ConvNeXt. In each semantic category, our base structure outperforms ConvNeXt in five categories. The line graph in Figure 4 shows that CN-Unet reaches the optimal value range faster than ConvNeXt. The experimental results demonstrate the feasibility of our CN-Unet infrastructure in general.

After determining the feasibility of CN-Net's infrastructure, we conducted in-depth research on the MDA module. The results of each branch on the ACDC dataset are shown in Table 2. Each branch has improved the performance of CN-Unet, and slice fusion has improved the average DSC by 1.13, which is the branch with the most contribution in MDA.Overall, the progress of the MDA module on CN-Unet is considerable, and each branch of MDA contributes to the whole module to varying degrees.

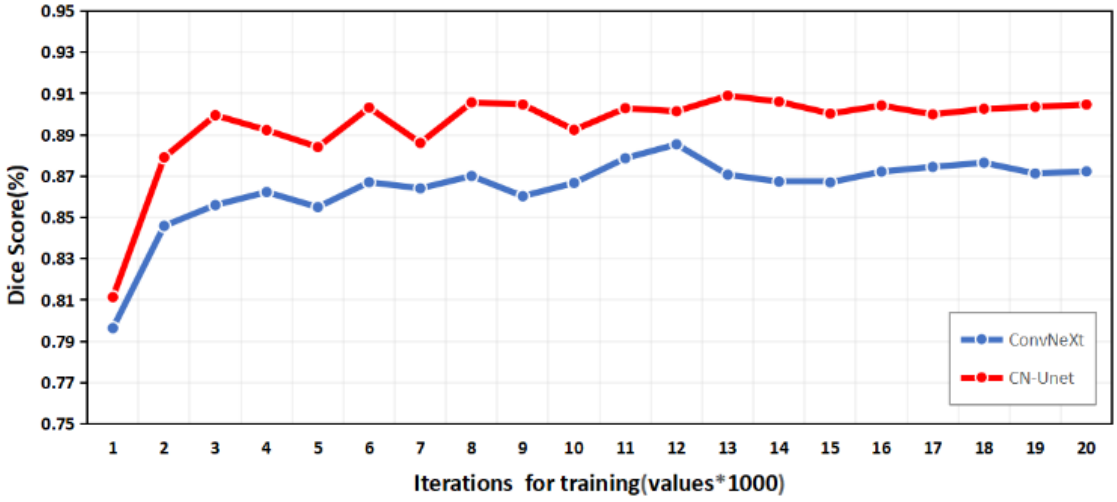

Figure 4: DSC line graphs per 1000iterations for CN-Unet and ConvNeXt.

Table 2
Ablation experiments

Method	Average	RV	Myo	LV
CN-Unet*	87.37	87.26	83.35	91.52
CN-Unet*+SF	88.50	88.92	84.45	92.13
CN-Unet*+SF+DAGAN	89.35	89.75	84.62	93.68
CN-Unet*+MDA	90.43	90.89	86.74	93.66
P-value		<1e-2(DSC)		

4. Conclusion

In this paper, we propose a 2D medical segmentation network, CN-Unet. CN-Unet is a robust deep convolutional network, we design it as a U-shaped symmetrical structure, and the basic design is the same as Swin-Unet. CN Block is an advanced convolution block proposed in ConvNeXt. CN Block absorbs the advantages of Swin Transformer Block and ResNet Block simultaneously. It can not only accurately obtain encoded spatial information but also build standard hierarchical objects. To give full play to the feature learning ability of CN Block, we divide the encoder and decoder into four layers and use skip connections to connect each corresponding layer to recover contextual information. To improve the small object segmentation ability of CN-Unet, we propose the MDA module. Experiments on Synapse and ACDC datasets show that our CN-Unet achieves promising results on 2D medical segmentation, which shows that CN-Unet has good segmentation performance and robustness. Especially on Synapse, the HD95 and DSC scores obtained by CN-Unet exceed the current state-of-the-art methods, and DSC, for the first time in the semantic class of gallbladder breaks through 80%, reflecting CN-Unet's performance in small object segmentation. We also conduct comparative experiments with the original ConvNeXt on the Synapse dataset, demonstrating the effectiveness of our encoder-decoder structure.

5. Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No.41901296, the Fujian Provincial Key Laboratory of Data Intensive Computing and Key Laboratory of Intelligent Computing and Information Processing under Grant No. BD201801, and the Wuhan Science and Technology Bureau Knowledge Innovation Dawning Plan Project: Detection and Optimization Method of GNSS Hybrid Attacks for Connected and Autonomous Vehicles und-er Grant No. 2022010801020270.

References

[1] P. Malhotra, S. Gupta, D. Koundal, A. Zaguia, W. Enbeyle, Deep neural networks for medical image segmentation, Journal of Healthcare Engineering 2022 (2022) 9580991 doi: 10.1155/2022/9580991

[2] M. H. Hesamian, W. Jia, X. He, P. Kennedy, Deep learning techniques for medical image segmentation: achievements and challenges, Journal of digital imaging 32 (2019) 582-596. doi: 10.1007/s10278-019-00227-x

[3] J. Peng, Y. Wang. Medical image segmentation with limited supervision: a review of deep network models, IEEE Access 9 (2021) 36827-36851. doi: 10.1109/ACCESS.2021.3062380

[4] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015, pp. 234-241. doi: 10.1007/978-3-319-24574-4_28

[5] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: redesigning skip connections to exploit multiscale features in image segmentation, IEEE transactions on medical imaging, 39 (2019) 1856-1867. doi: 10.1109/TMI.2019.2959609
[6] T. Hassanzadeh, D. Essam, R. Sarker, 2D to 3D evolutionary deep convolutional neural networks for medical image segmentation, IEEE Transactions on Medical Imaging 40 (2020) 712-721. doi: 10.1109/TMI.2020.3035555
[7] Z. Wang, E. Wang E, Y. Zhu, Image segmentation evaluation: a survey of methods, Artificial Intelligence Review 53 (2020) 5637-5674. doi: 10.1007/s10462-020-09830-9
[8] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A convnet for the 2020s, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022, pp. 11976-11986. doi: 10.1109/CVPR52688.2022.01167
[9] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, M. Wang, Swin-unet: Unet-like pure transformer for medical image segmentation, in: L. Karlinsky, T. Michaeli, K. Nishino (Eds.), Computer Vision – ECCV 2022 Workshops. ECCV 2022. Lecture Notes in Computer Science, vol 13803. Springer, Cham, 2023. doi: 10.1007/978-3-031-25066-8_9
[10] H. Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, Y. Yu, nnformer: Interleaved transformer for volumetric segmentation. arXiv preprint arXiv:2109.03201, 2021.
[11] X. Huang, Z. Deng, D. Li, X. Yuan, Missformer: An effective medical image segmentation transformer[J]. arXiv preprint arXiv:2109.07162, 2021.
[12] A. Hatamizadeh, Y. Tang, V. Nath, et al. Unetr: Transformers for 3d medical image segmentation, in: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 574-584.
[13] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. Roth, D. Xu, Swin unetr: Swin transformers for semantic segmentation of brain tumors in MRI images, in: Crimi, A., Bakas, S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2021. Lecture Notes in Computer Science, vol 12962. Springer, Cham, 272-284. doi: 10.1007/978-3-031-08999-2_22
[14] H. H. Lee, S. Bao, Y. Huo, B. Landman, 3D UX-Net: A large kernel volumetric ConvNet modernizing hierarchical transformer for medical image segmentation, arXiv preprint arXiv:2209.15076, 2022. doi: 10.48550/arXiv.2209.15076
[15] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, arXiv preprint arXiv:2102.04306, 2021. doi: 10.48550/arXiv.2102.04306
[16] M.-T. Fang, K. Przystupa, Z.-J. Chen, et al, Examination of abnormal behavior detection based on improved YOLOv3, Electronics, 10 (2021) 197. doi: 10.3390/electronics10020197.
[17] Li J, Chen J, Tang Y, et al. Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives, Medical Image Analysis 85 (2023) 102762. doi: 10.1016/j.media.2023.102762
[18] A. Antoniou, A. Storkey, H. Edwards, Data augmentation generative adversarial networks, arXiv preprint arXiv:1711.04340, 2017. doi: 10.48550/arXiv.1711.04340
[19] H. Tang, S. Bai, N. Sebe, Dual attention GANs for semantic image synthesis, in: Proceedings of the 28th ACM International Conference on Multimedia. 2020, pp. 1994-2002. doi: 10.1145/3394171.3416270
[20] Z. Eaton-Rosen, F. Bragman, S. Ourselin, M. J. Cardoso, Improving data augmentation for medical image segmentation, 2018. URL: https://openreview.net/pdf?id=rkBBChjiG
[21] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks, Neurocomputing 338 (2019) 34-45. doi: 10.1016/j.neucom.2019.01.103
[22] Y. Gao, M. Zhou, D. Liu, Z. Yan, S. Zhang, D. N. Metaxas, A multi-scale transformer for medical image segmentation: architectures, model efficiency, and benchmarks, arXiv preprint arXiv:2203.00131, 2022. doi: 10.48550/arXiv.2203.00131
[23] Y. Zhou, Z. Li, G. Bai, X. Chen, Prior-aware neural network for partially-supervised multi-organ segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019, pp. 10672-10681. doi:10.1109/ICCV.2019.01077
[24] Y. Xie, J. Zhang, C. Shen, Y. Xia, Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation, in: Proceedings of International conference on medical image computing and

computer-assisted intervention. Springer, Cham, 2021, pp. 171-180. doi: 10.1007/978-3-030-87199-4_16
[25] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, K. H. Maier-Hein, Nnu-net: Self-adapting framework for u-net-based medical image segmentation, In: H. Handels, T. Deserno, A. Maier, K. Maier-Hein, C. Palm, T. Tolxdorff, (eds.), Bildverarbeitung für die Medizin 2019. Informatik aktuell. Springer Vieweg, Wiesbaden. doi:10.1007/978-3-658-25326-4_7
[26] S. Chen, Z. S. Gamechi, F. Dubost, G. van Tulder, M. de Bruijne, An end-to-end approach to segmentation in medical images with CNN and posterior-CRF, Medical Image Analysis 76 (2022) 102311. doi: 10.1016/j.media.2021.102311
[27] K. B. Girum, G. Créhange, A. Lalande, Learning with context feedback loop for robust medical image segmentation, IEEE Transactions on Medical Imaging 40 (2021) 1542-1554. doi: 10.1109/TMI.2021.3060497
[28] D. Karimi, S. E. Salcudean, Reducing the hausdorff distance in medical image segmentation with convolutional neural networks, IEEE Transactions on medical imaging 39 (2019) 499-513. doi: 10.1109/TMI.2019.2930068