

RESEARCH



# The nnU-Net based method for automatic segmenting fetal brain tissues

Ying Peng<sup>1†</sup>, Yandi Xu<sup>2†</sup>, Mingzhao Wang<sup>1†</sup>, Huiquan Zhang<sup>1</sup> and Juanying Xie<sup>1\*</sup> 

## Abstract

The magnetic resonance (MR) images of fetuses make it possible for doctors to detect out pathological fetal brains in early stages. Brain tissue segmentation is prerequisite for making brain morphology and volume analyses. nnU-Net is an automatic segmentation method based on deep learning. It can adaptively configure itself, so as to adapt to a specific task via preprocessing, network architecture, training, and post-processing. Therefore, we adapt nnU-Net to segment seven types of fetal brain tissues, including external cerebrospinal fluid, gray matter, white matter, ventricle, cerebellum, deep gray matter, and brainstem. With regard to the characteristics of the FeTA 2021 data, some adjustments are made to the original nnU-Net, so that it can segment seven types of fetal brain tissues precisely as far as possible. The average segmentation results on FeTA 2021 training data demonstrate that our advanced nnU-Net is superior to the peers including SegNet, CoTr, AC U-Net and ResUnet. Its average segmentation results are 0.842, 11.759 and 0.957 in terms of Dice, HD95 and VS criteria. Moreover, the experimental results on FeTA 2021 test data further demonstrate that our advanced nnU-Net has obtained good segmentation performance of 0.774, 14.699 and 0.875 in terms of Dice, HD95 and VS, ranked the third in FeTA 2021 challenge. Our advanced nnU-Net realized the task for segmenting the fetal brain tissues using MR images of different gestational ages, which can help doctors to make correct and reasonable diagnoses.

**Keywords:** Fetal brain tissue segmentation, FeTA challenge, Magnetic resonance image segmentation, Image automatic segmentation, nnU-Net

## Introduction

Congenital diseases are the leading cause of infant death in the world [1]. Most of the existing studies on the diagnoses of infant congenital diseases are based on magnetic resonance (MR) images of infants or children's brains [2]. However, MR images of fetal brains have not been fully studied, due to the lack of public, selective and annotated clinical data. MR images of fetal brains can not only enable medical doctors to detect the brain structural abnormalities in the early stages of fetal development, but also can be applied to the study of fetal neurodevelopment.

Brain tissue segmentation is the premise for analyzing brain morphology and volume. But it is a time-consuming and laborious work to segment tissues of fetal brains manually. The segmentation results are often affected by doctors' subjective judgments, so it is of significance to study the automatic segmentation for highly complex and rapidly developing and changing fetal brain morphology before the birth.

The traditional studies of fetal brain segmentation are based on atlas mostly. Habas et al. [3] constructed the MR atlas of the intrauterine fetal brain for the first time, including developing and developed brain tissues. Then, an expectation-maximization model based on atlas was proposed to segment four types of brain tissues including white matter (WM), gray matter (GM), germinal matrix (GMAT) and cerebrospinal fluid (CSF). However, this work focused on a narrow age range of 20.5–22.5 weeks, and four types of fetal brain tissues. After that, Habas et al. [4] further proposed a spatiotemporal atlas of the

<sup>†</sup>Ying Peng, Yandi Xu, and Mingzhao Wang have contributed equally to this work.

\*Correspondence: xiejuany@snnu.edu.cn

<sup>1</sup> School of Computer Science, Shaanxi Normal University, West Chang'an Avenue, Chang'an District, Xi'an 710119, Shaanxi, People's Republic of China

Full list of author information is available at the end of the article

fetal brain with temporal models of MR intensity, tissue probability and shape changes for segmenting the fetal brain tissues of MR images. But this work still focused on a narrow age range from 20.57 to 24.71 weeks, and four types of fetal brain tissues of GM, WM, GMAT, and ventricles (VT). Serag et al. [5] proposed a 4D multi-channel probabilistic atlas for constructing the developing brain based on pairwise non-rigid registration. Compared with the probability atlas generated by affine registration, this method has a better segmentation effect in the cortex, ventricle and brain hemisphere. Although these methods have achieved good results, they all need atlas as the premise. However, up to now the available brain atlas are only from those fetuses with normally developing brains, lacking brain atlas of pathological fetuses.

In recent years, deep learning has been widely used in image classification [6, 7], segmentation [8] and object detection [9–12] because it can automatically learn features related to target tasks from data. Khalili et al. [13] proposed a data augmentation technique by synthetically introducing intensity inhomogeneity images, so as to improve the robustness of U-Net method for artifacts in images. It was demonstrated that this technique can improve the segmentation performance of the model and may replace or supplement pre-processing steps such as bias field correction by testing on MR images of 12 fetal brains aged from 22.9 to 34.6 weeks. Payette et al. [14] proposed to automatically segment the fetal brain tissues via high-resolution MR images with noisy multi-class labels based on transfer learning techniques. This method does not rely on the super-resolution reconstruction techniques used and can quantify the neurodevelopment of normal and pathological fetuses without bias using 15 fetal brains with the age range of 22.6–33.4 weeks.

In order to provide the accurate and automatic segmentation of the cortical plate in the fetal brain, Hong et al. [15] proposed a logarithm and boundary-Dice mixed loss function to gain the segmentation accuracy, and took the multi-view (axial, coronal, and sagittal) aggregation with test-time augmentation to cut down the error using three-dimensional information and multiple predictions. Experimental results show that this method can improve the segmentation accuracy of the fetal cortical plate. For the properties of the cortical plate, Dou et al. [16] proposed the attention module comprising two sets of convolutions with the convolution layer of four convolution kernels of different sizes for each, and developed a deep attentive neural network. This attention module made the network achieving good segmentation results on private datasets by combining full convolutional network and deep supervision mechanism. To improve the morphological consistency of the fetal cortical plate segmentation, Dumast et al. [17] took a topological constraint as

an additional loss function. This method had a significant advantage over the baseline method in segmenting MR images of fetal brains at any gestational age.

The aforementioned methods for segmenting fetal brain tissues are over-fitting to specific tasks to some extent, or may be affected by imperfect validation. Such as the studies in [15–17] focused on segmenting cortical plates in fetal brains while the studies in [3, 4, 13, 14] paid attention to segment four types of tissues of fetal brains covering narrow age range. Therefore, these available methods will fail to produce expected results for new tasks.

Fortunately, there are some good deep learning models emerged for segmenting images obtained from various scenarios besides the aforementioned ones for segmenting fetal brain tissues. SegNet [18] pays attention to the scene understanding, providing a deep convolutional encoder-decoder architecture network for outdoor and indoor scene segmentation tasks. It was reported that it performed well, having competitive inference time and most efficient inference memory-wise [18]. CoTr [19] efficiently bridges a convolutional neural network and a transformer into a new framework for segmenting 3D medical images accurately. It was reported that CoTr brought a substantial performance improvement over other CNN-based, transformer-based, and hybrid methods on the 3D multi-organ segmentation task [19]. Furthermore, CoTr can be extended to deal with brain structure or tumor segmentation [19]. AC U-Net [20] proposed a new loss function incorporating area and size information into a dense deep U-Net-like learning model. The proposed loss function outperforms the very popular loss function cross-entropy and is robust to hyper-parameter when the AC U-Net is used to segment cardiac MR images [20]. It is reported in [20] that AC U-net can be applied to other challenging segmentation tasks from various real applications. ResUnet [21] combines the strengths of residual learning and U-Net to build a network having similar U-Net architecture for road extraction from high-resolution remote sensing images. It outperformed U-Net and other two state-of-the-art deep learning based road extraction methods [21].

Although these aforementioned four innovative deep learning models including SegNet, CoTr, AC U-Net and ResUnet are not focus on segmenting fetal brain tissues, the strengths and the ideas of them make it possible to use them to the fetal brain tissue segmentation task. That is why we choose these four deep learning models as the baselines to compare with ours.

Furthermore, the recently proposed nnU-Net [22] combines multiple U-Net methods, so that it can automatically adjust its use of multiple U-Net architectures

to model any given types of input images. It outperforms most of the existing methods without any human intervention. Therefore, this paper adapts nnU-Net to study the segmentation of fetal brain tissues. To accommodate this new task, some significant adjustments are made to the original nnU-Net. The results of the FeTA 2021 challenge demonstrate that our advanced nnU-Net has obtained good results in segmenting the fetal brain tissues and ranked the third.

This paper is organized as follows. Section 2 will introduce the data and methods used in this study. Section 3 will display the experimental results and the analyses. Conclusions come in Sect. 4.

### Data and methods

This section will first introduce the data used in this study, that is, the data for FeTA 2021 challenge. Then the advanced nnU-Net adapted in this study will be introduced in detail.

#### Data

This study used the data from FeTA 2021 challenge, that is, from the Fetal Brain Tissue Annotation and Segmentation Challenge in 2021 [23]. This data set was collected from the Children’s Hospital of the University of Zurich, including T2-weighted MR images of normal and pathological fetal brains. Due to the low resolution of clinically acquired image data, the FeTA 2021 Challenge used two different super-resolution reconstruction techniques including mialSRTK [24] and IRTK [25] to build super-resolution images for each case. In order to simulate the real situation, the quality of the super-resolution reconstruction images is different, including poor quality, overall good quality and excellent quality. There are 80 cases in FeTA 2021 training data, where 40 cases are reconstructed using the mialSRTK technique, and the rest using the IRTK technique. Each case is labeled with seven

types of brain tissues, including external cerebrospinal fluid (eCSF), gray matter (GM), white matter (WM), ventricles (VT), cerebellum (CB), deep gray matter (DGM) and brainstem (BS). Table 1 shows the frequencies of the MR images of normal and pathological fetal brains at different gestational ages in FeTA 2021 training data. The MR images from pathological fetal brains are mainly from 22 to 29 weeks of gestational age, while the normal fetal brain MR images are evenly distributed at all gestational ages. There are 40 cases in the test data of the FeTA 2021 challenge, where 20 cases are reconstructed using the mialSRTK technique, and the rest by the IRTK technique. The test data have not been public yet, nor do the labels of them. However, it was reported in [26] that the test data of FeTA 2021 challenge has got similar distribution with that of the training data. Table 2 summarizes the frequencies of the MR images of normal and pathological fetal brains at each gestational age in the test data of FeTA 2021 challenge. Figure 1 shows the axial plane, sagittal plane and coronal plane of the fetal brain MR images and their corresponding labels. The up-left corner three images are the normal fetal brain MR images, and the down-left three images show the labels of these three images. The up-right three images display the pathological fetal brain MR images, and the down-right are the related labels.

#### Network structure

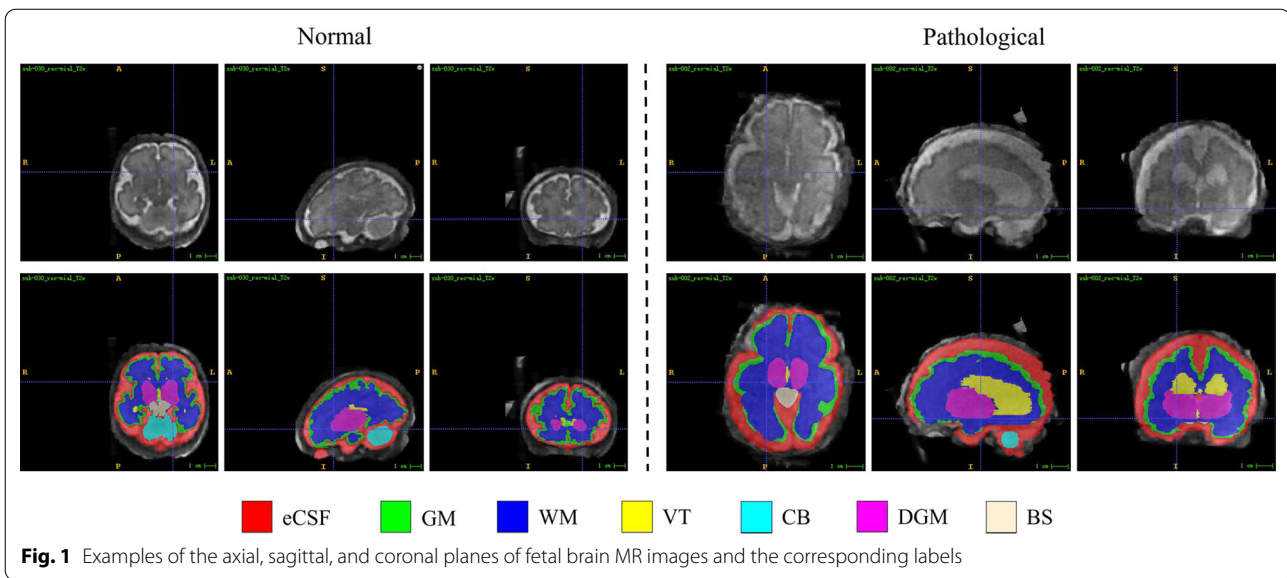
U-Net [27] is composed of encoder, decoder, and skip connection. It combines both low-resolution and high-resolution information, and still has got good generalization even without enough training data. It is widely used in the field of medical image segmentation. However, the available variations of U-Net, such as embedding dense connections or attention mechanism etc., are excessively appropriate for specific tasks to some extent, or may be affected by imperfect validation, so that they cannot

**Table 1** Frequencies of normal and pathological fetal brain MR images at different gestational age in FeTA 2021 training data

Gestation week	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	Sum
Normal cases	0	1	0	3	3	1	2	2	4	3	3	3	3	3	0	2	33
Pathological cases	2	1	3	6	5	3	3	10	6	4	1	1	0	2	0	0	47

**Table 2** Frequencies of normal and pathological fetal brain MR images at different gestational age in FeTA 2021 test data

Gestation week	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	Sum
Normal cases	0	1	1	2	1	1	1	2	1	0	0	2	1	1	0	1	15
Pathological cases	0	0	2	3	2	2	0	5	2	5	1	1	1	1	0	0	25



obtain good performance on new tasks. However, the latest nnU-Net [22] combines multiple U-Net methods, so that it can automatically adjust the use of multiple U-Net architectures to model any types of images input to it. It is reported that nnU-Net outperforms most existing methods without human intervention [22].

With regard to the properties of the fetal brain tissues in MR images, we adapt the nnU-Net by advancing it to do the segmentation for the fetal brain tissues. The specific structure of the nnU-Net used in this paper is shown in Fig. 2. The residual structure is adopted in the encoder to enhance the information propagation in the shallow layers, and a deep supervision mechanism is introduced to accelerate the effective training for the network. The basic network module of our nnU-Net is  $3 \times 3 \times 3$  convolution plus Instance Normalization and Leaky ReLU together.

In the encoder of our nnU-Net shown in Fig. 2, one residual structure is initially adopted, and the number of residual structures is increased by 1 for each down-sampling, up to the maximum of 4. If the number of input/output channels is different in residual structure, a convolution layer with a convolution kernel size of  $1 \times 1 \times 1$  is inserted into the skip connection for non-linear mapping, so as to match the input and output image size. The down-sampling is done in the encoder by using the convolutions with a stride of 2.

The decoder of our nnU-Net shown in Fig. 2 comprises the stacked basic modules consisting of  $3 \times 3 \times 3$  convolution and Instance Normalization and Leaky ReLU. The up-sampling is done to the output of each layer using transposed convolution with a stride of  $2 \times 2 \times 2$ .

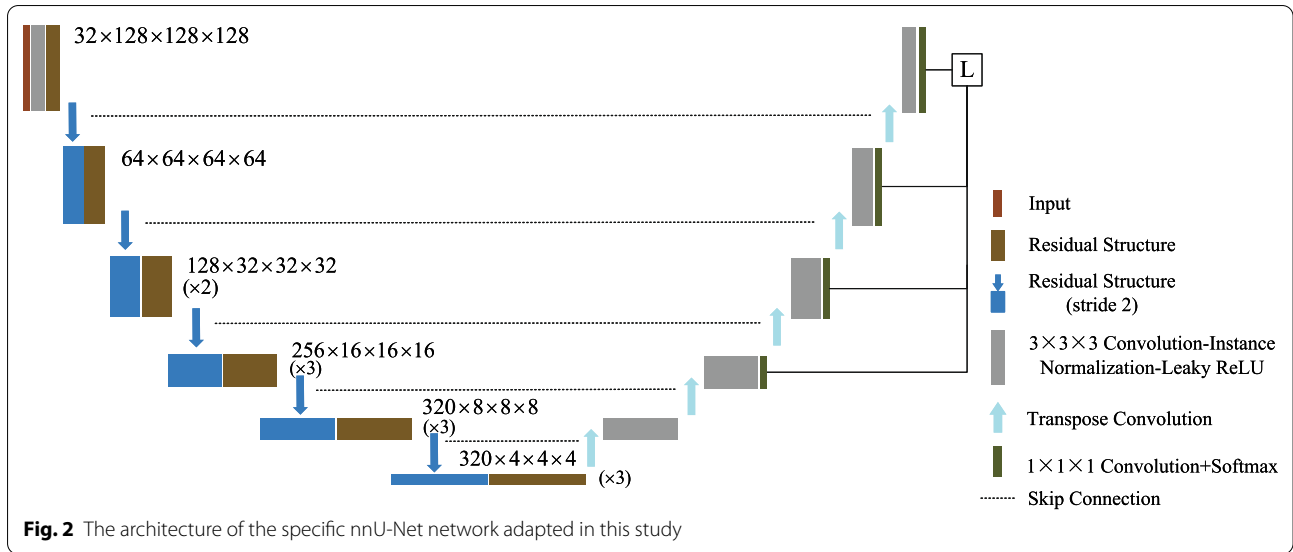
There are initial 32 feature channels in our nnU-Net. The number of channels is doubled with each down-sampling, up to the maximum of 320, while the number of channels will be halved with each transposed convolution of up-sampling, so as to back to the initial channels at last.

Our nnU-Net calculates the loss function using the output of the last four layers of the decoder, due to the big noises in the feature maps of the shallow layers of the network. Specifically, down-sampling is done to the labels with respect to the ratio of [1, 2, 4, 8] in the process of online data augmentation. Then, the outputs of the last four layers of the decoder are respectively input into a convolution layer with 8 filters and  $1 \times 1 \times 1$  convolution kernel size, and the segmentation probability map is obtained by using the activation function of Softmax. Finally, the segmentation errors are calculated for the last four layers, and the loss function  $L$  will be obtained by summing these errors with the weights of  $\frac{8}{15}, \frac{4}{15}, \frac{2}{15}, \frac{1}{15}$ , respectively. The loss function  $L$  adopted in this paper is the sum of cross-entropy loss and Dice loss. It is calculated in (1).

$$L = L_{CE} + L_{Dice}. \tag{1}$$

where,  $L_{CE}$  and  $L_{Dice}$  are the cross-entropy loss and Dice loss, respectively. They are calculated in (2) and (3), respectively.

$$L_{CE} = -\frac{1}{N} \sum_{c=1}^C \sum_{i=1}^N y_i^c \log(p_i^c). \tag{2}$$



$$L_{Dice} = -\frac{2}{C} \sum_{c=1}^C \frac{\sum_{i=1}^N p_i^c y_i^c}{\sum_{i=1}^N p_i^c + \sum_{i=1}^N y_i^c + \epsilon} \quad (3)$$

where,  $N$  is the number of pixels in the segmentation map encoded using one-hot, and  $C$  is the number of classes and it is 8 in this study. The  $y_i^c$  is a sign function. Its value is 1 when the true category of the  $i$ th pixel of the region labeled  $y$  is  $c$ ; otherwise, its value is 0. The  $p_i^c$  is the probability of the  $i$ th pixel of the model predictive segmentation map  $p$  belonging to the category  $c$ .

The cross-entropy loss  $L_{CE}$  calculates the loss of each pixel. The positive and negative samples contribute equally to this loss. The cross-entropy loss can effectively avoid the problem of the gradient disappearance in training network. But it will prefer the majority class when segmenting the image with class imbalance, thus affecting the optimization direction of the network.

Dice loss  $L_{Dice}$  only calculates the proportion of the overlap between the ground truth and model predictive segmentation result for the foreground of an image, ignoring the background region, thus alleviating the problem of category imbalance in the training data. The  $\epsilon$  in (3) is a very small constant, and is  $1e - 8$  in this paper, so as to avoid the zero denominator of the Dice loss function in a special case.

### Experiments and analyses

This section will first introduce the metrics used in this paper to evaluate the performance of our specific nnU-Net on the FeTA 2021 challenge. Then is the experimental design followed by the experimental results and the analyses.

### Evaluation metrics

The Dice [28], HD95 (Hausdorff 95 Distance) [29], and VS (Volume Similarity) metrics are adopted to quantify the segmentation performance of a model.

Dice is one of the most commonly used segmentation performance evaluation indexes. It is defined as the proportion of the overlap between the segmentation map predicted by the model and the region labeled by experts to the sum of the two regions [28]. Its value range falls in [0, 1]. The bigger the Dice, the better the segmentation results are obtained by the model; otherwise, the worse is the model segmentation effect. We define  $P$  as the segmentation map predicted by the model, and  $GT$  as the region labeled by experts, that is, the ground truth, then the Dice can be calculated in (4), where  $|\cdot|$  means absolute value.

$$Dice(P, GT) = \frac{2 \times |P \cap GT|}{|P| + |GT|} \quad (4)$$

Given sets  $A = \{a_1, a_2, a_3, \dots\}$ ,  $B = \{b_1, b_2, b_3, \dots\}$ , the HD (Hausdorff Distance) is used to value the similarity between them, and is defined in (5). In the following experiments, these two sets will indicate the segmentation result predicted by the model and the ground truth labeled by experts.

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (5)$$

where,  $h(A, B) = \max_{a \in A} (\min_{b \in B} \|a - b\|)$ , and  $h(B, A) = \max_{b \in B} (\min_{a \in A} \|b - a\|)$ , and  $\|\cdot\|$  indicates the  $L_2$  norm or any other distance. We adopt the Euclidean distance, that is, the  $L_2$  norm is used in this paper.  $H(A, B)$  is called the bidirectional HD, and it is the most

**Table 3 Experimental environment settings**

Names	Descriptions
OS	Linux
CPU	Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz
GPU	NVIDIA GeForce RTX 3090
CUDA Version	11.1
Programming Language	Python 3.7
Deep Learning Framework	Pytorch 1.8.0

basic form of HD. The  $h(A, B)$  is called the one-way HD from set  $A$  to set  $B$ , so is the  $h(B, A)$  from set  $B$  to set  $A$ .

In order to eliminate the unreasonable distance caused by some outliers and preserve the stability of the overall values, HD95 is adopted in this paper instead of the original HD, that is, the 95th quantile of all the HD values between pixels of the segmentation map and the ground truth. The smaller the HD95 value, the more similar between the two sets, and the better the segmentation result is. On the contrary, the greater the difference between the two sets, the worse is the segmentation effect.

VS is an indicator used to compare the volumes of two segmentation maps. There are many definitions for this indicator, and the one we adopted in this paper is as follows.

$$VS(P, GT) = 1 - \frac{|GT_{vol} - P_{vol}|}{|GT_{vol}| + |P_{vol}|}. \quad (6)$$

where,  $GT_{vol}$  is the volume of the Ground Truth of an image to be segmented, and  $p_{vol}$  the volume of the image predicted by the segmentation model, and  $|\cdot|$  means the absolute value. It can be seen that the bigger the VS value, the better segmentation result is obtained. The best case is 1, while the worst is 0.

However, it should be noted that although VS is a commonly used indicator, it cannot be treated alone, because there may exist the situation that the segmentation map predicted by a model is equal to the region marked by experts in volumes but there is not any overlap between the predictive segmentation map and the ground truth.

### Experiment design

Experiments are carried out in the environment with Linux operating system, and the NVIDIA GeForce RTX 3090 GPU with 24GB memory to accelerate the training process. The program code is implemented based on the nnU-Net [22]. The library of nnU-Net can be found at <https://github.com/MIC-DKFZ/nnUNet>. Other experimental environment configurations are shown in Table 3.

The SGD (Stochastic Gradient Descent) optimizer with momentum is used to optimize the network in the

experiment, and the momentum size is set to 0.99. The initial learning rate is 0.01 for training network. The weight decay coefficient is  $3e-5$ . The patch size is  $128 \times 128 \times 128$ , and batch size is 2. The total iterations are 1000. The polyLR [30] is adopted to adjust the learning rate dynamically, and the updating method is as follows.

$$\alpha = \alpha_0 \times \left(1 - \frac{t}{T}\right)^{0.9}. \quad (7)$$

where,  $\alpha_0$  is the initial learning rate, and  $t$  is the current iteration number, and  $T$  is the total iteration number.

Data augmentation technology refers to process original images, so as to augment samples in the dataset. It can effectively remedy the insufficiency of the existing training data, so as to avoid the over-fitting of the model, and enhance the generalization of the model. In our experiments, we adopt the following data augmentation techniques, including rotation, scaling, mirroring, brightness transforming, contrast enhancement, gamma correction, Gaussian noise and Gaussian blurring. The rotation angle is set to one of the degrees in  $-30$  to  $+30$ . The scaling ratio is set to one of the values in  $0.7-1.4$ . The brightness transforming is set to one of the values in  $0.75-1.25$ . The contrast enhancement is set to one of the values in  $0.75-1.25$ . The gamma correction is set to one of the values in  $0.7-1.5$ . Gaussian noise is set to one of the values from 0 to 0.1, and Gaussian blurring is set to one of the values from 0.5 to 1.0.

### Experimental result analyses

This subsection will first display the extensive experimental results we did on FeTA 2021 training data, then show the results we obtained on the test data of FeTA 2021 challenge, followed by the comparison between our method and other methods within Top 12 in FeTA 2021 challenge.

#### Experiments on FeTA 2021 training data

This subsection will carry out experiments on FeTA 2021 training data. The FeTA 2021 training data is partitioned into training and test subsets in a ratio of 7:3. The model is trained using the training subset, and its capability is verified on the test subset. The performance of our specific nnU-Net is compared to that of SegNet [18], CoTr [19], AC U-Net [20] and ResUnet [21] models in terms of Dice, HD95, and VS criteria.

Since the scanning directions of MR images in FeTA 2021 training data are not consistent completely, the ITK toolkit is used to adjust all of MR images and their labels to RAI (Right, Anterior, Inferior) direction. The resolutions of MR images and their labels in the training subset are resampled to the isotropic MR images of  $0.5 \times 0.5 \times 0.5$  mm by using the third-order spline

**Table 4 Performance of our nnU-Net and other peer models for segmenting fetal brain tissues of MR images**

Models	Metrics	eCSF	GM	WM	VT	CB	DGM	BS	Average $\pm\sigma$
SegNet	Dice	0.795	0.694	0.893	0.848	0.833	0.856	0.820	0.820 $\pm$ 0.122
	HD95	17.896	15.533	15.714	11.794	<b>6.343</b>	8.019	<b>8.785</b>	12.012 $\pm$ 6.846
	VS	0.962	0.963	0.980	0.965	<b>0.920</b>	0.945	0.948	0.955 $\pm$ 0.071
CoTr	Dice	0.806	0.742	0.907	0.884	0.835	<b>0.871</b>	0.840	0.841 $\pm$ 0.128
	HD95	16.962	15.106	14.959	11.167	10.399	7.951	9.133	12.240 $\pm$ 9.212
	VS	0.942	0.966	0.979	0.974	0.902	<b>0.951</b>	0.946	0.951 $\pm$ 0.098
AC U-Net	Dice	0.796	<b>0.746</b>	0.906	0.881	<b>0.850</b>	<b>0.871</b>	<b>0.843</b>	<b>0.842</b> $\pm$ 0.131
	HD95	17.418	18.882	<b>13.729</b>	11.026	8.420	8.012	9.265	12.393 $\pm$ 7.997
	VS	0.921	0.956	0.977	0.970	0.911	0.949	<b>0.959</b>	0.949 $\pm$ 0.106
ResUNet	Dice	0.798	0.734	0.901	0.877	0.818	0.866	0.842	0.834 $\pm$ 0.140
	HD95	<b>16.782</b>	<b>14.982</b>	14.070	11.974	8.159	<b>7.823</b>	9.079	11.838 $\pm$ 7.357
	VS	0.954	<b>0.975</b>	0.974	0.969	0.886	0.946	0.951	0.951 $\pm$ 0.107
Our nnU-Net	Dice	<b>0.810</b>	0.740	<b>0.909</b>	<b>0.886</b>	0.839	0.868	<b>0.843</b>	<b>0.842</b> $\pm$ 0.125
	HD95	16.824	15.193	14.470	<b>10.125</b>	7.746	8.837	9.115	<b>11.759</b> $\pm$ 7.619
	VS	<b>0.963</b>	0.967	<b>0.983</b>	<b>0.978</b>	0.906	0.947	0.953	<b>0.957</b> $\pm$ 0.090

Bold values indicate the best results of the specific metric on the specific tissue among all models

interpolation and the linear interpolation, respectively. It is a general way that the MR images are usually sampled by using the third-order spline interpolation while the labels usually by the linear interpolation or the nearest neighbor interpolation method. Finally, each MR image is normalized using Z-score.

Experiments are done under same environment configuration. The models are trained on the training subset, and the performance of each model is tested on the test subset. Table 4 shows the experimental results of our specific nnU-Net and other four peers in comparison for segmenting seven types of fetal brain tissues using the MR images from the test subset in terms of Dice, HD95, and VS metrics.

To test whether there is a difference between the performance of each model in segmenting the normal and pathological fetal brain MR images, the normal and pathological fetal brain images from test subset are, respectively, tested on each model. The predictive segmentation results of each model are calculated for the normal and pathological fetal brain MR images, respectively. The experimental results of each model on normal and pathological fetal brain MR images from test subset are shown in Tables 5 and 6, respectively. The bold fonts in Tables 4, 5 and 6 represent the best results of the specific metric on the specific tissue among all models.

The results in Table 4 show that our advanced nnU-Net has obtained the best results in segmenting VT tissue of the fetal brain MR images in terms of Dice, HD85 and VS metrics when comparing to the peers including SegNet, CoTr, AC U-Net, ResUNet. In addition, it performs best in segmenting eCSF and WM tissues in terms

of Dice and VS criteria among the five models, so do in segmenting BS tissue in terms of Dice criterion. Furthermore, our improved nnU-Net and the peers including SegNet, CoTr, AC U-Net and ResUNet have simultaneously obtained good segmentation results on WM, VT, CB, DGM and BS tissues of fetal brain MR images, with the Dice values over 0.8, while the segmentation results on eCSF and GM tissues are not good with the Dice value less than 0.8, except for our nnU-Net and the peer CoTr having a comparatively good Dice values over 0.8 on eCSF tissue.

The results in Table 4 demonstrate that our nnU-Net has obtained the best average segmentation results in segmenting seven types of fetal brain tissues with 0.842, 11.759 and 0.957 in terms of Dice, HD95 and VS, though it is not the most stable model. The SegNet model is the most stable model among the peers in comparison for segmenting seven types of fetal brain tissues of MR images, though it has obtained the worst segmentation results in terms of Dice. AC U-Net obtained the same performance as our nnU-Net did in terms of Dice criterion. The results in Table 4 also disclose that there is an apparent difference in the segmentation results between all models on seven types of fetal brain tissues of MR images in terms of HD95 while few difference in segmentation results in terms of VS metric.

The above analyses to the segmentation results of all five models in terms of Dice, HD95 and VS demonstrate that our improved nnU-Net is superior to the other four peers in comparison when segmenting the seven types of fetal brain tissues in MR images.

**Table 5 Segmentation results of our nnU-Net and other peer models on normal fetal brain MR images**

Models	Metrics	eCSF	GM	WM	VT	CB	DGM	BS	Average±σ
SegNet	Dice	0.860	0.704	0.906	0.831	0.910	0.876	0.871	0.851 ± 0.074
	HD95	14.511	15.593	15.873	11.892	5.016	7.956	8.970	11.402 ± 5.759
	VS	0.974	0.969	0.983	0.962	0.979	0.959	0.943	0.967 ± 0.025
CoTr	Dice	<b>0.878</b>	0.759	<b>0.925</b>	<b>0.875</b>	0.927	0.882	0.884	0.876 ± 0.061
	HD95	13.878	15.506	15.518	12.202	4.457	8.303	<b>8.743</b>	11.230 ± 6.218
	VS	0.978	0.973	0.983	0.982	0.982	0.961	<b>0.958</b>	0.974 ± 0.022
AC U-Net	Dice	<b>0.878</b>	<b>0.765</b>	0.924	0.873	0.927	<b>0.884</b>	<b>0.886</b>	<b>0.877</b> ± 0.059
	HD95	16.893	16.522	<b>14.370</b>	10.705	<b>4.229</b>	7.782	9.074	11.368 ± 6.077
	VS	0.974	0.971	0.983	0.981	0.980	<b>0.971</b>	0.957	0.974 ± 0.022
ResUnet	Dice	0.875	0.758	0.924	0.873	0.925	0.880	0.878	0.873 ± 0.061
	HD95	13.671	<b>14.780</b>	14.641	10.529	4.274	7.875	9.306	10.725 ± 5.626
	VS	<b>0.981</b>	<b>0.975</b>	<b>0.987</b>	0.983	<b>0.984</b>	0.963	0.948	<b>0.975</b> ± 0.024
Our nnU-Net	Dice	0.877	0.763	<b>0.925</b>	0.874	<b>0.928</b>	0.882	0.883	0.876 ± 0.060
	HD95	<b>13.583</b>	15.243	14.834	<b>10.067</b>	4.328	<b>7.696</b>	9.222	<b>10.710</b> ± 5.845
	VS	0.979	0.971	<b>0.987</b>	<b>0.984</b>	0.983	0.964	0.952	0.974 ± 0.024

Bold values indicate the best results of the specic metric on the specific tissue among all models

**Table 6 Segmentation results of our nnU-Net and other peer models on pathological fetal brain MR images**

Models	Metrics	eCSF	GM	WM	VT	CB	DGM	BS	Average±σ
SegNet	Dice	0.749	0.687	0.883	0.861	0.778	0.842	0.784	0.798 ± 0.143
	HD95	20.314	15.491	15.601	11.725	<b>7.291</b>	8.064	<b>8.652</b>	<b>12.448</b> ± 7.525
	VS	<b>0.953</b>	0.958	0.978	0.968	<b>0.878</b>	0.935	0.951	<b>0.946</b> ± 0.089
CoTr	Dice	0.755	0.729	0.894	0.891	0.770	<b>0.864</b>	0.809	0.816 ± 0.155
	HD95	19.165	<b>14.820</b>	14.559	10.427	14.643	<b>7.700</b>	9.411	12.961 ± 10.833
	VS	0.916	0.961	0.976	0.969	0.846	<b>0.943</b>	0.938	0.935 ± 0.125
AC U-Net	Dice	0.737	<b>0.733</b>	0.893	0.887	<b>0.794</b>	0.861	0.812	0.817 ± 0.160
	HD95	<b>17.793</b>	20.567	<b>13.271</b>	11.255	11.414	8.176	9.402	13.125 ± 9.085
	VS	0.884	0.946	0.973	0.962	0.862	0.933	<b>0.960</b>	0.931 ± 0.135
ResUnet	Dice	0.742	0.717	0.885	0.880	0.741	0.856	<b>0.816</b>	0.805 ± 0.171
	HD95	19.004	15.126	13.663	13.006	10.934	7.785	8.917	12.634 ± 8.315
	VS	0.935	0.975	0.965	0.959	0.817	0.934	0.952	0.934 ± 0.136
Our nnU-Net	Dice	<b>0.761</b>	0.723	<b>0.898</b>	<b>0.894</b>	0.775	0.859	0.815	<b>0.818</b> ± 0.152
	HD95	19.139	15.157	14.209	<b>10.166</b>	10.188	9.653	9.040	12.507 ± 8.618
	VS	0.952	<b>0.964</b>	<b>0.981</b>	<b>0.973</b>	0.851	0.935	0.954	0.944 ± 0.115

Bold values indicate the best results of the specic metric on the specific tissue among all models

The experimental results in Table 5 for normal fetal brain tissue segmentation show that AC U-Net model has obtained the best average Dice of 0.877 for segmenting the seven types of fetal brain tissues of MR images of the normal fetal brains. Our nnU-Net has obtained the best average HD95 of 10.710 for segmenting the seven types of fetal brain tissues of the MR images of normal fetal brains. The ResUnet model defeats the others in comparison in terms of the average VS. The peer model SegNet in comparison performs worst, with the average Dice of 0.851, and HD95 of 11.402, and VS of 0.967.

Furthermore, its variances in terms of Dice and VS is the worst too among all peers in comparison, which means it is not a stable model. However, the AC U-Net not only performs best in terms of Dice, it is also the most stable model in terms of Dice metric. The stable of our nnU-Net ranks the second in terms of Dice metric which are the most significant metric among the Dice, HD95 and VS metrics.

The results in Table 5 also show that all five models in comparison have obtained good segmentation results in terms of Dice criterion with values over 0.8 nearly



0.9 when segmenting 6 types of tissues including eCSF, WM, VT, CB, DGM and BS of normal fetal brains in MR images. However, they all cannot obtain the Dice values over 0.8 in segmenting the GM tissue.

The above analyses to the segmentation results of five models for the normal fetal brain tissues of MR images demonstrate that our improved nnU-Net model in this paper has obtained the comparative good effect in segmenting seven types of normal fetal brains of MR images. It is the second stable model following AC U-Net.

In addition, comparing the experimental results in Tables 4 and 5, it can be seen that except for segmenting the VT tissue of fetal brains of MR images, all models in comparison performed better in segmenting eCSF, GM, WM, CB, DGM and BS fetal brain tissues of MR images of normal fetal brains than they did in segmenting same fetal brain tissues of the MR images comprising both normal and pathological fetal brains.

The segmentation results of pathological fetal brain MR images in Table 6 show that our adjusted nnU-Net model is superior to other peer models in terms of Dice metric of the average segmentation results for seven types of pathological fetal brain tissues of MR images, with the mean Dice of 0.818. The SegNet model has achieved the best average segmentation results in terms of HD95 and VS indexes, with the mean HD95 of 12.448 and mean VS of 0.946, respectively. It is followed by our advanced nnU-Net model.

Although the average segmentation results of our adjusted nnU-Net model are not the optimal in terms of HD95 and VS when segmenting seven types of pathological fetal brain tissues in MR images, it performs best for segmenting the VT tissue in terms of Dice, HD95 and VS metrics, and its Dice, HD95 and VS values are 0.894, 10.166 and 0.973, respectively. Furthermore, it defeats other peer models in segmenting WM tissue of pathological fetal brain MR images in terms of Dice and VS, and its Dice is 0.898 and VS is 0.981. It is also superior to the peer models in segmenting eCSF tissue of MR images of pathological fetal brains in terms of Dice metric, with the value of 0.761. Meanwhile, it is also superior to the peer methods in comparison in segmenting GM tissue of MR images of the pathological fetal brains in terms of VS metric of 0.964.

As we all known, Dice is the most significant metric among the Dice, HD95 and VS ones. This indicates that our advanced nnU-Net is the best one among the five models in comparison including SegNet, CoTr, AC U-Net, ResUnet and our nnU-Net for segmenting seven types of brain tissues of pathological fetal brain MR images. Though it has the second stable property.

Comparing the experimental results in Tables 5 and 6, it can be found that there is a significant difference in

the segmentation results of each model's for segmenting eCSF, CB and BS tissues of MR images of normal and pathological fetal brains. Furthermore, the segmentation results show the rule that all methods in comparison have obtained better performance on normal fetal brain MR images than on pathological fetal brain MR images when segmenting the six types of fetal brain tissues including eCSF, GM, WM, CB, DGM and BS, though the segmentation results of all models in comparison show the opposite rule on segmenting VT tissues, that is, all models have achieved better segmentation results when segmenting the VT tissue of pathological fetal brain MR images than of normal fetal brain MR images.

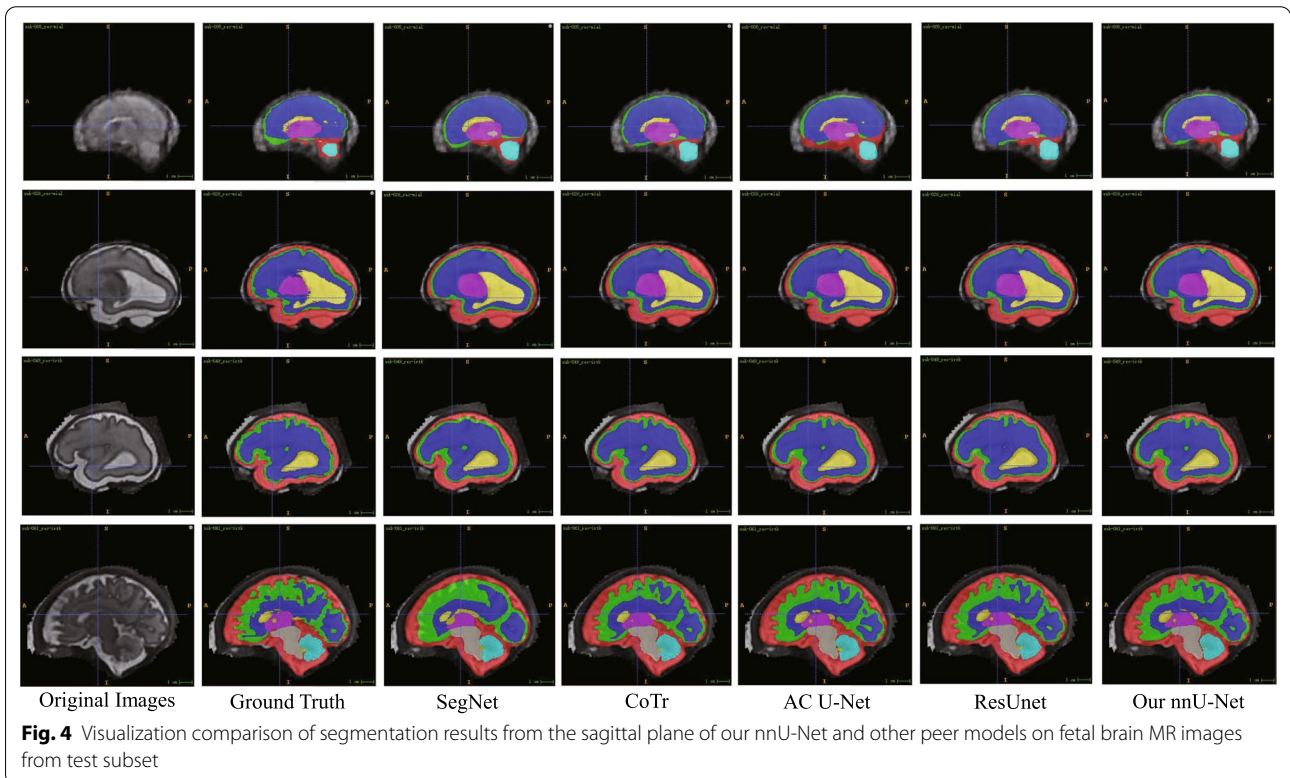
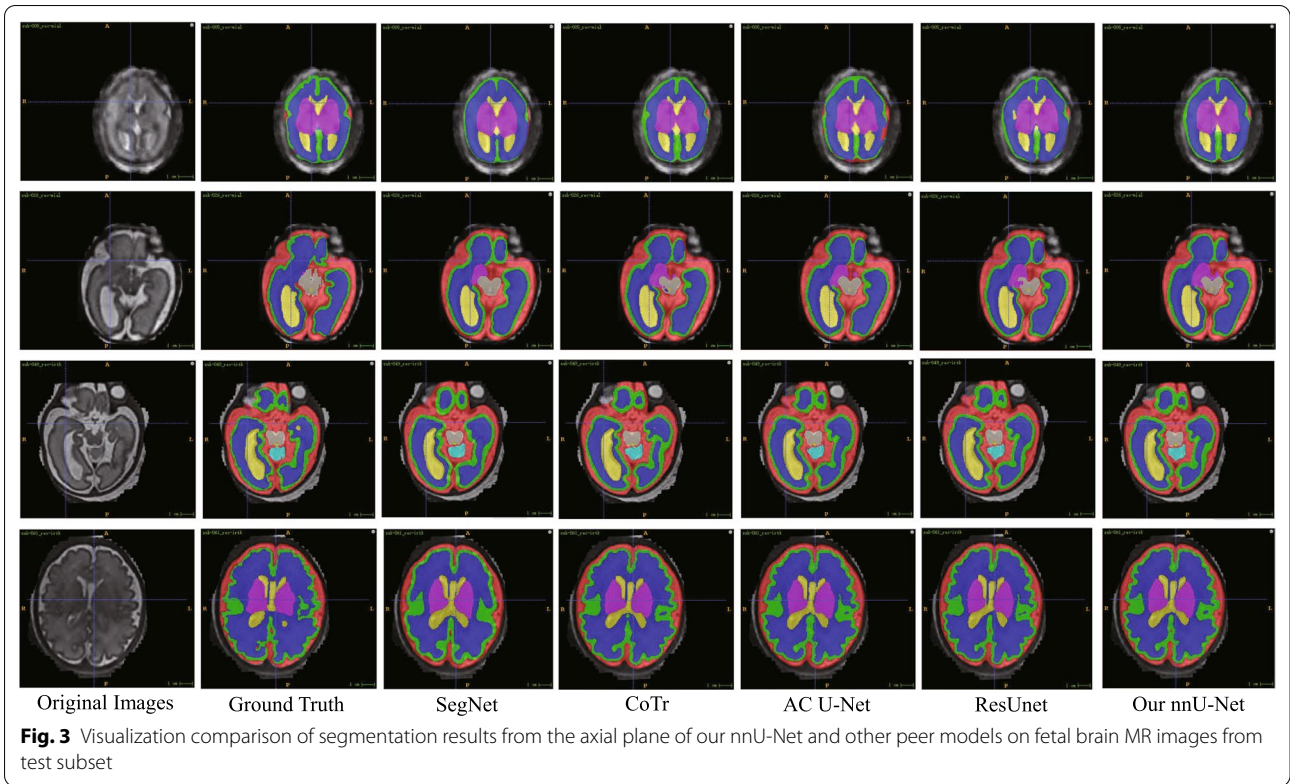
The segmentation results shown in Tables 4, 5 and 6 demonstrate that our adjusted nnU-Net demonstrates the advantages over the peers in comparison no matter when segmenting the tissues of normal or pathological fetal brains of MR images. Though its stable is not the best with the second place instead.

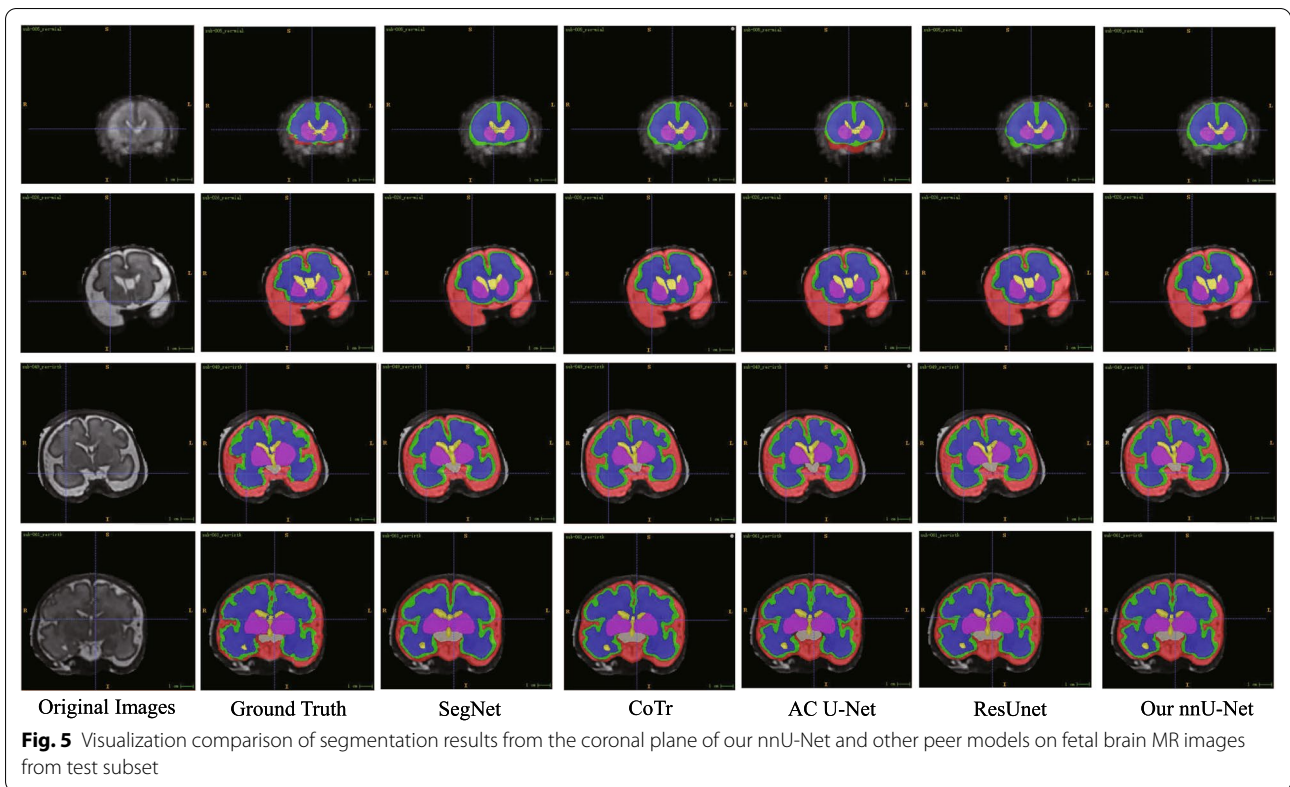
In order to intuitively compare the differences between each model in segmenting the fetal brain tissues, some typical segmentation results are selected from the test subset for qualitative analyses. Figure 3 shows the segmentation results from the axial plane of each model on several fetal brain MR images from the test subset, in which the first column is the original MR images of fetal brains, and the second column is the true labels, that is, the ground truth, and the third to seventh columns are the predictive results of SegNet, CoTr, AC U-Net, ResUnet and our nnU-Net adjusted in this paper, respectively. Figures 4 and 5 show the corresponding segmentation results of Fig. 3 of each model from the sagittal plane and the coronal plane, respectively.

It can be seen from the segmentation results shown in Figs. 3, 4 and 5 that SegNet, CoTr, AC U-Net, ResUnet and our nnU-Net can locate the basic target locations of seven types of fetal brain tissues, but there are over-segmentation and under-segmentation phenomena in all the results of them. The structures of eCSF and GM tissues are complex, and their shapes are various greatly, which make it more difficult to segment these two types of tissues. The overall segmentation results demonstrate that our nnU-Net model adjusted in this paper outperforms the peer methods in comparison, its segmentation results are most consistent to the true labels. The SegNet model performs worst, which is clearly seen from the visualization of the segmentation results shown in Fig. 4 from the sagittal plane and that in Fig. 5 from the coronal plane.

#### **Experiments on FeTA 2021 test data**

In this subsection, the 10-fold cross-validation experiments are carried out on the FeTA 2021 training data. That is, the FeTA 2021 training data are randomly





partitioned into 10 subsets, and one subset is selected as the validation subset, and the rest 9 subsets are used as the training subset to train our improved nnU-Net model, till each one among 10 subsets has been selected as the validation subset. The average results of the 10 validation subsets are the result of the 10-fold cross-validation experiment. Table 7 displays the results of the 10-fold cross-validation experiment of the nnU-Net adjusted in this paper on the FeTA 2021 training data. The 10 models trained on the FeTA 2021 training data are then used jointly to predict the segmentation results for each fetal brain MR image in the FeTA 2021 test data. To improve the segmentation accuracy, TTA (Test Time Augmentation) technique is applied to mirror all of the axial directions. Table 8 shows the segmentation results of seven types of brain tissues of FeTA 2021 test data by our nnU-Net adapted in this paper. The segmentation results of the improved nnU-Net model's in this paper

and that of the other teams' methods within top 12 for FeTA 2021 test data are compared in Table 9. The bold fonts in Table 9 mean the optimal results of the specific metric.

The average experimental results in Table 7 of the 10-fold cross-validation experiments of our nnU-Net method on FeTA 2021 training data demonstrate that this nnU-Net can achieve good segmentation results on WM, VT, CB, DGM, and BS fetal brain tissues, with Dice over 0.84. However, its segmentation results on eCSF and GM fetal brain tissues are not comparatively good, with Dice less than 0.8 but above 0.74. The best Dice of 0.914 is obtained by our adapted nnU-Net model when segmenting the WM tissue of the fetal brains while the worst Dice of 0.742 is obtained for segmenting the GM tissue of the fetal brains.

The segmentation results in Table 7 also show that the nnU-Net adapted in this paper has obtained the best

**Table 7** The average results of 10-fold cross-validation experiments of our nnU-Net on training data of FeTA 2021

Metrics	eCSF	GM	WM	VT	CB	DGM	BS	Average±σ
Dice	0.794	0.742	0.914	0.886	0.878	0.867	0.843	0.846 ± 0.120
HD95	19.161	15.907	13.870	10.773	8.278	7.994	9.165	12.164 ± 10.022
VS	0.922	0.959	0.979	0.968	0.944	0.940	0.943	0.951 ± 0.084

**Table 8 Segmentation results of our nnU-Net on FeTA 2021 test data**

Metrics	eCSF	GM	WM	VT	CB	DGM	BS	Average $\pm\sigma$
Dice	0.744	0.698	0.896	0.895	0.767	0.705	0.715	0.774 $\pm$ 0.182
HD95	24.350	15.673	14.769	11.389	7.692	13.115	15.902	14.699 $\pm$ 10.049
VS	0.856	0.962	0.967	0.959	0.806	0.746	0.827	0.875 $\pm$ 0.182

HD95 of 7.994 when segmenting the DGM fetal brain tissue, and the worst HD95 of 19.161 for segmenting the eCSF fetal brain tissue. This indicates that there are much variations in the segmentation results of our nnU-Net in terms of HD95 in segmenting the seven types of fetal brain tissues in MR images.

It can be seen from the results in Table 7 that there is little difference in the segmentation results of the nnU-Net adapted in this paper on seven types of fetal brain tissues in terms of VS metric. The best result is obtained for segmenting the WM fetal brain tissue with the VS of 0.979 while the worst is for segmenting the fetal brain tissue of eCSF, with the VS of 0.922.

The results in Table 7 on the FeTA 2021 training data also show that the nnU-Net adapted in this paper can obtain good segmentation results in terms of Dice, HD95, and VS. The average 10-fold cross-validation experimental segmentation results of seven types of brain tissues are 0.846, 12.164 and 0.951, respectively, in terms of Dice, HD95, and VS.

The average segmentation results in Table 8 of the 10 models of 10-fold cross-validation experiments on FeTA 2021 test data show that the average segmentation results of our nnU-Net on seven types of fetal brain tissues are 0.774, 14.699, and 0.875 in terms of Dice, HD95, and VS, respectively. The segmentation

results of our nnU-Net on the FeTA 2021 test data are more stable in terms of Dice and VS than that in terms of HD95.

Comparing the results in Tables 7 and 8, it can be found that in segmenting CB, DGM, and BS fetal brain tissues, the prediction results of our nnU-Net on FeTA 2021 test data are significantly different from the average results of 10-fold cross-validation experiments on FeTA 2021 training data. The segmentation results for CB, DGM, and BS fetal brain tissues on FeTA 2021 test data is much worse than that on FeTA 2021 training data.

In addition, The average segmentation results of the 10-fold cross-validation experiments of our nnU-Net on FeTA 2021 training data are better than that on FeTA 2021 test data when segmenting the six types of fetal brain tissues including eCSF, GM, WM, CB, DGM and BS, but it is not true when segmenting the VT fetal brain tissue. Our nnU-Net obtained better results on FeTA 2021 test data than that on FeTA 2021 training data when segmenting VT tissue of fetal brain MR images in terms of Dice metric.

Furthermore, the average results of our nnU-Net for segmenting seven types of fetal brain tissues of MR images in FeTA 2021 test data are not as good as that of it for the images in FeTA 2021 training data. The variance of our nnU-Net on FeTA 2021 training data is also less than that of it on FeTA 2021 test data. This fact demonstrates that the MR images of the fetal brains in the FeTA 2021 test set are quite different from that in the FeTA 2021 training set, and also demonstrates that the generalization capability of our nnU-Net model needs improving further.

The segmentation results of the Top 12 teams participating in FeTA 2021 Challenge in Table 9 show that there is a great difference between the results of each team's method in terms of HD95, with the best of 14.012 and the worst of 21.329, while there is little difference between the results of different teams' in terms of Dice and VS. In terms of Dice and HD95, NVAUTO team's model is superior to that of other teams' for segmenting seven types of fetal brain tissues of FeTA 2021 test data, with the Dice of 0.786 and HD95 of 14.012, respectively. Although ichilove-ax team achieved the best result in terms of VS of 0.888 for segmenting seven types of fetal brain tissues

**Table 9 Comparison of segmentation results between our nnU-Net and other teams on FeTA 2021 test data**

Teams	Dice	HD95	VS
NVAUTO	<b>0.786</b>	<b>14.012</b>	0.885
SJTU_EIEE_2-426Lab	0.775	14.671	0.883
our nnU-Net	0.774	14.699	0.875
Hilab	0.774	14.569	0.873
Neurophet	0.775	15.375	0.877
davoodkarimi	0.771	16.755	0.882
2Ai	0.767	14.625	0.867
xlab	0.771	15.262	0.873
ichilove-ax	0.766	21.329	<b>0.888</b>
TRABIT	0.769	14.901	0.866
ichilove-combi	0.762	16.039	0.873
muw_dsobotka	0.765	17.159	0.874

Bold values indicate the best results of the specic metric on the specific tissue among all models

of FeTA 2021 test data, its performance is the worst in terms of HD95, and is also not good in terms of Dice.

Considering the Dice, HD95, and VS metrics used by FeTA 2021 challenge, our nnU-Net achieved good results and ranked the third in FeTA 2021 Challenge. The NVAUTO and SJTU\_EIEE\_2-426Lab teams ranked the first and the second, respectively.

## Conclusions

This paper adapted the nnU-Net to segment seven types of fetal brain tissues, including external cerebrospinal fluid, gray matter, white matter, ventricle, cerebellum, deep gray matter, and brainstem, based on the MR images from normal and pathological fetal brains.

The extensive experiments were conducted to test the segmentation capability of the improved nnU-Net. The experimental results of our nnU-Net were compared to that of the peers and the methods participating in FeTA 2021 challenge.

The results of each model for segmenting the eCSF and GM tissues are not as good as that for segmenting the other five types of fetal brain tissues, due to the complex shape structure of these two types of tissues and the large variation in the shapes.

In addition, the overall segmentation results of each segmentation model for the pathological fetal brains are not as good as that for the normal fetal brains, except for the segmentation results of the fetal brain VT tissue, which may be caused by the fact that the morphologies of the pathological fetal brains are always quite different to that of the normal fetal brains.

Furthermore, the average segmentation results of 10-fold cross-validation experiments of our adapted nnU-Net on FeTA 2021 training data are better than that on FeTA 2021 test data, except for that on VT tissue of fetal brains.

Compared with the typical and the latest methods, our adjusted nnU-Net in this paper can obtain accurate and robust results in segmenting the normal and pathological fetal brain tissues based on MR images.

However, the average segmentation results of our nnU-Net on FeTA2021 test data is only 0.774 in terms of Dice, which is much less than 0.8, and much less than its performance on FeTA 2021 training data with the Dice of 0.846. Therefore, we need further improve our nnU-Net by changing its structure or experiment design or propose a more appropriate loss function, so as to improve its segmentation capability and the generalization for segmenting the fetal brain tissues. This is our future work.

## Acknowledgements

This work is supported in part by the National Natural Science Foundation of China under Grant Nos. of 62076159, 12031010, and 61673251; and is also supported by the Fundamental Research Funds for the Central Universities under Grant Nos. of GK202105003 and 202207017. We acknowledge those who organized the FeTA 2021 challenge and published the training data and the related informations for us to use in this research.

## Declarations

### Conflict of interest

There is not any conflict of interest in the manuscript.

### Author details

<sup>1</sup>School of Computer Science, Shaanxi Normal University, West Chang'an Avenue, Chang'an District, Xi'an 710119, Shaanxi, People's Republic of China. <sup>2</sup>School of Medicine, Tulane University, 1430 Tulane Ave, New Orleans, LA 70112, USA.

Received: 24 December 2022 Accepted: 9 March 2023

Published: 27 March 2023

## References

- Modell B, Darlison MW, Malherbe H, Moorhith S, Blencowe H, Mahaini R, El-Adawy M. Congenital disorders: epidemiological methods for answering calls for action. *J Commun Genetics*. 2018;9(4):335–40.
- Makropoulos A, Counsell SJ, Rueckert D. A review on automatic fetal and neonatal brain MRI segmentation. *NeuroImage*. 2018;170:231–48.
- Habas PA, Kim K, Rousseau F, Glenn OA, Barkovich AJ, Studholme C. Atlas-based segmentation of the germinal matrix from in utero clinical MRI of the fetal brain. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 351–358. Springer, 2008.
- Habas PA, Kim K, Corbett-Detig JM, Rousseau F, Glenn OA, Barkovich AJ, Studholme C. A spatiotemporal atlas of MR intensity, tissue probability and shape of the fetal brain with application to segmentation. *NeuroImage*. 2010;53(2):460–70.
- Serag A, Kyriakopoulou V, Rutherford MA, Edwards AD, Hajnal JV, Aljabar P, Counsell SJ, Boardman J, Rueckert D. A multi-channel 4D probabilistic atlas of the developing brain: application to fetuses and neonates. *Ann BMVA*. 2012;2012(3):1–14.
- Sarki R, Ahmed K, Wang H, Zhang Y, Wang K. Convolutional neural network for multi-class classification of diabetic eye disease. *EAI Endorsed Trans Scalable Inf Syst*. 2022;9(4): e15.
- Sarki R, Ahmed K, Wang H, Zhang Y. Automated detection of mild and multi-class diabetic eye diseases using deep learning. *Health Inf Sci Syst*. 2020;8(32):1–9.
- Pandey D, Wang H, Yin X, Wang K, Zhang Y, Shen J. Automatic breast lesion segmentation in phase preserved DCE-MRIs. *Health Inf Sci Syst*. 2022;10(9):1–19.
- Jiahua D, Michalska S, Subramani S, Wang H, Zhang Y. Neural attention with character embeddings for hay fever detection from Twitter. *Health Inf Sci Syst*. 2019;7(21):1–7.
- Alvi AM, Siuly S, Wang H. A long short-term memory based framework for early detection of mild cognitive impairment from EEG signals. *IEEE Trans Emerg Top Comput Intell*. 2022. <https://doi.org/10.1109/TETCI.2022.3186180>.
- Supriya S, Siuly S, Wang H, Zhang Y. Automated epilepsy detection techniques from electroencephalogram signals: a review study. *Health Inf Sci Syst*. 2020;8(33):1–15.
- He J, Rong J, Sun L, Wang H, Zhang Y, Ma J. A framework for cardiac arrhythmia detection from IoT-based ECGs. *World Wide Web*. 2020;23(5):2835–50.
- Khalili N, Lessmann N, Turk E, Claessens N, de Heus R, Kolk T, Viergever MA, Benders MJ, Išgum I. Automatic brain tissue segmentation in fetal MRI using convolutional neural networks. *Magn Reson Imaging*. 2019;64:77–89.
- Payette K, Kottke R, Jakab A. Efficient multi-class fetal brain segmentation in high resolution MRI reconstructions with noisy labels. In *Medical*

- Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis, pp. 295–304. Springer, 2020.
15. Hong J, Yun HJ, Park G, Kim S, Laurentys CT, Siqueira LC, Tarui T, Rollins CK, Ortinau CM, Grant PE, Lee JM. Fetal cortical plate segmentation using fully convolutional networks with multiple plane aggregation. *Front Neurosci*. 2020;14:591683.
  16. Dou H, Karimi D, Rollins CK, Ortinau CM, Vasung L, Velasco-Annis C, Ouaalim A, Yang X, Ni D, Gholipour A. A deep attentive convolutional neural network for automatic cortical plate segmentation in fetal MRI. *IEEE Trans Med Imaging*. 2021;40(4):1123–33.
  17. de Dumast P, Kebiri H, Atat C, Dunet V, Koob M, Cuadra MB. Segmentation of the cortical plate in fetal brain MRI with a topological loss. In: Uncertainty for safe utilization of machine learning in medical imaging, and perinatal imaging, placental and preterm image analysis, pp. 200–209. Springer, 2021.
  18. Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(12):2481–95.
  19. Xie Y, Zhang J, Shen C, Xia Y. Cotr: Efficiently bridging CNN and transformer for 3D medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 171–180. Springer, 2021.
  20. Chen X, Williams BM, Vallabhaneni SR, Czanner G, Williams R, Zheng Y. Learning active contour models for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11632–11640, 2019.
  21. Zhang Z, Liu Q, Wang Y. Road extraction by deep residual U-Net. *IEEE Geosci Remote Sens Lett*. 2018;15(5):749–53.
  22. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203–11.
  23. Payette K, de Dumast P, Kebiri H, Ezhov I, Paetzold JC, Shit S, Iqbal A, Khan R, Kottke R, Grehten P, Ji H. An automatic multi-tissue human fetal brain segmentation benchmark using the fetal tissue annotation dataset. *Sci Data*. 2021;8(1):1–14.
  24. Tourbier S, Bresson X, Hagmann P, Thiran JP, Meuli R, Cuadra MB. An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization. *NeuroImage*. 2015;118:584–97.
  25. Kuklisova-Murgasova M, Quaghebeur G, Rutherford MA, Hajnal JV, Schnabel JA. Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. *Med Image Anal*. 2012;16(8):1550–64.
  26. Payette K, Li H, de Dumast P, Licandro R, Ji H, Siddiquee MM, Xu D, Myronenko A, Liu H, Pei Y, Wang L. Fetal brain tissue annotation and segmentation challenge results. 2022. [arXiv:2204.09573](https://arxiv.org/abs/2204.09573)
  27. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, Böhm A, Deubner J, Jäckel Z, Seiwald K, et al. U-net: deep learning for cell counting, detection, and morphometry. *Nat Methods*. 2019;16(1):67–70.
  28. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26(3):297–302.
  29. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell*. 1993;15(9):850–63.
  30. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell*. 2017;40(4):834–48.

---

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.