Active Semi-supervised Learning based on Global Uncertainty Variation with Noise Resistance

Yufei WEN
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
ywen196@connect.hkust-gz.edu.cn

***Abstract*—Automatic and accurate fetal brain segmentation is essential for congenital disease diagnosis and treatment. However, voxel-wise manual annotation is laborious, and the annotation quality strongly depends on the annotator's professional knowledge and clinical experience. This problem contradicts the data-hungry nature of deep learning, especially for medical image segmentation. To reduce the consumption of annotation, in this paper, we propose a novel active semi-supervised algorithm for fetal brain tissue segmentation that incorporates the active learning techniques into semi-supervised methods to minimize labeling costs. Specifically, we present a new active learning selection strategy that leverages the global uncertainty variation of a sample to measure its informativeness and adaptively adjust the time to perform active learning according to the learning state of the network. Furthermore, we design a non-parameter pool attention (PA) module to refine the prediction of the model and resist noise effectively. In addition, we introduce symmetric soft cross entropy (SSCE) loss as an unsupervised loss function to resist noise further. Extensive experiments on two fetal brain tissue segmentation datasets demonstrate the effectiveness of our model, outperforming state-of-the-art approaches. Associated codes can be accessed at: https://github.com/Dreamer1209/ASL.**

***Index Terms*—active learning, semi-supervised learning, global uncertainty, noise resistance, medical image segmentation**

I. Introduction

Congenital diseases are one of the leading causes of neonatal death worldwide [1], [12], [13], [40]. In order to detect and treat congenital disorders, antenatal maternal and fetal health care is paramount. Fetal ultrasound imaging has been an essential tool in prenatal care for many years [2]. And in recent years, fetal magnetic resonance imaging (MRI), especially of the brain, has emerged as an essential complementary tool to provide information on fetal development [3]–[5]. This advanced imaging technique enables more precise segmentation and analysis of fetal brain structures, enhancing our capability to detect, comprehend, and precisely diagnose congenital anomalies [6]. Therefore, the development of automatic segmentation methods for infant brain tissue, leveraging MRI data, is vital for the accurate diagnosis and effective treatment of congenital diseases.

Recently, deep learning methods have achieved remarkable success in medical image segmentation tasks [14]–[19], [27]–[29], but most of them rely on a large number of pixel (voxel)-wise annotations [7]. However, due to the rapid changes in fetal brain anatomy and the lack of growth data for many pathological and congenital diseases, the training data is limited. On the other hand, the corresponding annotations are extremely time-consuming and labor-intensive, and labeling accuracy depends heavily on annotators' clinical experience and expertise, especially for pixel (voxel)-wise images.

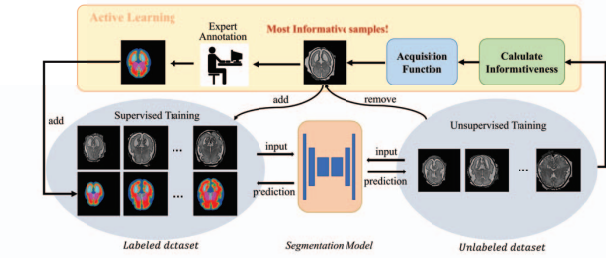

Fig. 1. Depicting the process of Semi-Supervised Learning (SSL) integrated with an active learning strategy. Duing active learning phase, active learning component efficiently identifies and selects the most informative samples and strategically moved from the unlabeled dataset into the labeled dataset.

To alleviate the heavy reliance on annotations, semi-supervised learning (SSL) is arguably one of the most feasible solutions [8], which leverages a few labeled and a large number of unlabeled samples to improve segmentation performance that approximates or exceeds fully supervised training. To connect labeled samples with the unlabeled ones in the feature space, existing SSL methods use three strategies for learning the representations of unlabeled samples, i.e., consistent learning, adversarial learning, and self-supervised learning. Nevertheless, each strategy has its own drawbacks. Consistent learning encourages the network to produce consistent predictions for different perturbations of the same sample [9], but its performance is limited by the adaptation of the perturbation methods. Adversarial learning leverages a discriminator to distinguish the authenticity of segmentation results [10]. However, it is difficult to train and sensitive to hyperparameters settings. Self-supervised learning converts model predictions into pseudo labels and treats them as guidance to train the model [11]. Nevertheless, it suffers from pixel (voxel) indecipherable, class imbalance, and insufficient label information problems, resulting in unrobust predication.

Instead of designing a sophisticated representation learning strategy, we aim to assist human experts in annotating the most effective samples for facilitating the training of SSL. Motivated by active learning (AL) [24], [25], which involves

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00025

human in the loop to select the most informative samples to maximize model performance with minimal labeling cost, we integrate active learning into the SSL process and resolve its shortcomings. The procedure is illustrated in Figure 1.

In this paper, we present a novel semi-supervised algorithm with active learning and pseudo-label noise resistance technology. Our method focuses on selecting samples for human annotation by utilizing global uncertainty variation as informativeness measurement. Specifically, we discover that the samples with lower uncertainty variation contain richer information and design an uncertainty memory bank to store the global uncertainty of unlabeled samples during the past training process. We select the sample with the lowest uncertainty fluctuation to annotate and migrate it from the unlabeled pool to the labeled pool. Furthermore, to prevent the network from overfitting the labeled data, we utilize the mean intersection of union (mIoU) between the network's predictions and the corresponding labels (pseudo labels) to decide when to perform active learning. In addition, we propose a new pool attention (PA) which is a parameter-free module to improve the model prediction provided by the teacher model. To further resist noise, we introduce the soft symmetric cross entropy (SSCE) as unsupervised loss function and leverage the predicted probabilities as soft pseudo labels instead of using pseudo hard labels.

Our contributions can be summarized as follows:

- We propose an advanced semi-supervised algorithm for fetal brain tissue segmentation and leverage active learning and pseudo-label noise resistance on mean-teacher framework, which offers easy implementation and architecture versatility.
- We present a new active selection strategy that leverages global uncertainty variation of samples to measure its informativeness and propose to utilize the learning state of the network to adjust when to engage the active learning dynamically.
- To resist noise and improve pseudo label quality, we design the pool attention module to refine the prediction and employ optimized symmetric cross entropy to resist noise.
- Extensive experiments on two public fetal brain tissue segmentation tasks demonstrate the effectiveness of the proposed method.

II. Related Works

A. Semi-supervised Medical Image Segmentation

Recent years have seen a surge in interest towards semi-supervised learning (SSL) as a means to reduce annotation costs, enhancing model performance through the combined use of labeled and unlabeled data in network training. Broadly, SSL algorithms in this domain fall into three categories: adversarial learning-based [14]–[17], [26], consistency-based [18]–[23], and self-supervised learning-based approaches [27]–[31].

Consistent-based and self-supervised learning based approaches are domain SSL methods in medical image processing. Li [22], [23] proposed TCSM (v2) algorithms for medical image segmentation tasks which introduced transformation-consistent strategy to enhance the regularization effect for pixel-level predictions. MCNet [20] encourages mutual consistency with cycled pseudo label scheme over two decoders. Another line of consistency-based methods are to enforce task-consistency [14], [21]. Self-supervised learning based approaches estimate the pseudo labels for unlabeled samples training and the key of this method is ensuring the quality of the pseudo labels. In medical scenarios, MPCT [30] estimates reliable pseudo labels by performing consistency among multiple planes for multi-organ segmentation. Sajjadi [31] presented an SSL algorithm that utilized proportional labels as weakly supervised information and generated pseudo labels for unlabeled data by negative label learning in training stage, which sift noisy labels with auxiliary information. However, existing self-supervised learning approaches still suffer from the challenges of noisy pseudo-labeling, low data utilization and class bias (class imbalance task). In our paper, we delve into a self-supervised learning method based on the mean teacher framework. To combat noise, we introduce a novel component, pool attention (PA) to enhance pseudo label quality, and employ soft symmetric cross entropy (SSCE) to address key issues in self-supervised learning.

B. Uncertainty Estimation

Uncertainty estimation evaluates how uncertain an AI system is in its predictions and is introduced into semi-supervised image segmentation tasks. For example, Yu [18] filtered out the unreliable prediction with the guidance of the estimated uncertainty of the teacher model. Xia [29] further proposed uncertainty-aware multi-view co-training framework, which integrated each view's uncertainty estimation to achieve accurate labeling. In this paper, we leverage the global uncertainty variation of a sample as the sample's uncertainty estimation to measure its informativeness.

C. Active Learning

Active learning (AL) aims to work with minimal labels while maximizing the model's performance, in other words, seeking the smallest subset of data that responds to the entire data distribution. In recent years, many researchers have incorporated active learning into semi-supervised algorithms for selecting information-rich samples from unlabeled data (samples near clustering boundaries, representative samples in the space of unlabeled distributions). TOD [24] estimates the sample loss by evaluating the discrepancy of outputs given by models at different optimization steps to select informative unlabeled samples. BoostMIS [25] found informative samples as annotation candidates using virtual adversarial perturbations and the density-aware entropy of the model. In our work, we utilize global uncertainty variation of the unlabeled samples to select informative samples. To insert newly annotated data at an appropriate time, we leverage the average intersection (mIoU) between the network's predictions and the corresponding labels (pseudo labels) to adjust when to perform active learning automatically.

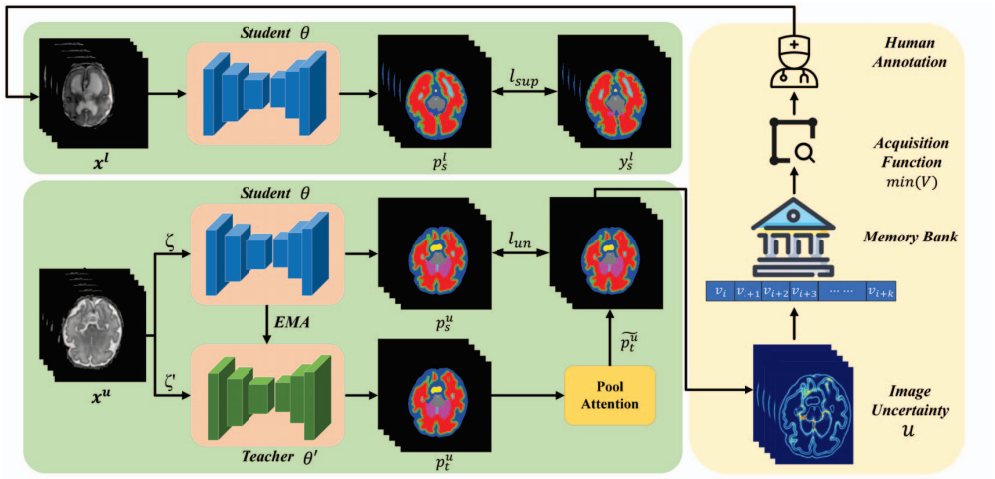

Fig. 2. Overview of our proposed model. ζ and ζ' are two data augmentation methods, θ and $\theta^{'}$ correspond to the parameters of two structurally identical models: the student model and the teacher model, where θ is updated by backpropagation algorithm and $\theta^{'}$ is updated by the exponential average movement (EMA) of θ, x, p, $\widetilde{p}$ and y represents input data, model prediction, refined model prediction and segmentation ground truth respectively. The model is initially trained with a small amount of labeled data and leverages mean-teacher structure to perform semi-supervised learning. During the training process, model will actively select the most informative samples for labeling based on unsupervised prediction. These newly labeled samples are utilized to update the model, and the model iteratively continue this process to improve model performance.

III. Methodology

A. Preliminary Definitions

In the medical image segmentation task, we define the training set includes L labeled samples and U unlabeled samples ($L \ll U$). $\mathcal{D}_l = \{(x_i, y_i)\}_{i=1}^{L}$ and $\mathcal{D}_u = \{(x_i)\}_{i=1}^{U}$ indicates labeled data set and unlabeled data set respectively, where $x_i \in \mathbb{R}^{H\times W\times D\times 1}$ denotes the input volume and $y_i \in \{0,1\}^{H\times W\times D\times C}$ is the corresponding voxel-wise ground truth.

B. Network Overview

We propose a novel semi-supervised algorithm by introducing active learning and noise resistance technologies based on mean teacher framework, as illustrated in Figure 2. For a labeled input $\mathbf{x}^l$, the student model is trained by utilizing a supervised loss between the prediction $\mathbf{p}^l$ and its corresponding ground truth $\mathbf{y}^l$. For an unlabeled input $\mathbf{x}^u$, we feed the input into student and teacher model with different augmentations and employ the teacher's prediction to guide student model learning. To resist noise, we present a pool attention (PA) module to refine the prediction of teacher model and introduce symmetric soft cross-entropy (SSCE) loss to calculate unsupervised loss between student and teacher prediction. In addition, we present a new active learning strategy, which records the variation of the global uncertainty to measure the informativeness of the sample. Moreover, we utilize the prediction mean intersection of union (mIoU) value to adaptively adjust when to engage in active learning until the workforce is depleted.

C. Active Learning via Global Uncertainty Variation

To improve data utilization efficiency, we integrates Active Learning (AL) into semi-supervised fetal brain tissue segmentation by assessing the informativeness of unlabeled samples via global uncertainty variation. Our approach tracks the training process from uncertainty to certainty and utilize the training variations as a function to select volumes for annotation. We calculate this variation by storing each sample's global uncertainty in a memory bank, then computing the prediction of model p, the volume i voxel-wise entropy. Next, we simply utilize the mean value of the entropy as the global uncertainty, which can be formulated as:

$$\mathcal{H}_t^{(i,j)} = -\sum_{c=1}^{C} p_t^{(i,j,c)} \log p_t^{(i,j,c)}, \quad \mathcal{U}_t^i = \frac{1}{N}\sum_{j=1}^{N} \mathcal{H}_t^{(i,j)} \tag{1}$$

Here, $\mathcal{H}t^{(i,j)}$ represents the entropy of the j-th voxel in volume i at step t, with C being the number of classes and N the total voxel count per sample. The calculated global uncertainty $\mathcal{U}_t^i$ is then stored in the memory bank based on a first-in-first-out principle. Then we assess the variation in global uncertainty for each sample using the uncertain variance, defined as:

$$\mathcal{V}^i = \frac{1}{K} \times \sum_{k=1}^{K} (\mathcal{U}_t^i - \bar{\mathcal{U}^i})^2 \tag{2}$$

where K is the memory bank size and $\bar{\mathcal{U}^i}$ is the mean of $\mathcal{U}_k^i$ Lastly, the acquisition function $\mathcal{A}^*$ is that:

$$\mathcal{A}^* = min_\lambda \left\{\mathcal{V}^1, \mathcal{V}^2, ..., \mathcal{V}^U\right\} \tag{3}$$

where λ is the number of actively annotated samples and U is the size of the unlabeled pool. The samples with minimum

uncertain variance is selected due to the consideration that the smaller the uncertainty variation of the sample, the richer the information it contains because the training process of the network is the process of classifying the sample from uncertainty to certainty. The selected samples are transferred from the unlabeled to the labeled pool for annotation.

When to perform active learning is critical to the training of the model. If it performs too early, the model has not yet learned enough about the labeled data. If it performs too late, there is a risk of overfitting. Inspired by [39] which demonstrated that the training mIoU increases rapidly during early learning stage and rises slowly during memorization stage. We propose to utilize the mIoU between predictions and its corresponding labels or pseudo labels as guidance to estimate the turning point. Specifically, we firstly fit the following mIoU function with exponential function using least square error:

$$f(t) = a\left(1 - e^{-b\cdot(t-d)^c}\right) \tag{4}$$

where t represents training period and $0 < a \leq 1$, $b, c, d \geq 0$ are fitting parameters. Then, we leverage the derivation $f'(t)$ relative variation with respect to start training time to measure the turning point where mIoU function starts to ascent slowly:

$$|f'(t) - f'(s)| > r|f'(s)| \tag{5}$$

where r represents a predetermined threshold, and s is the training commencement point. Through experiments, we observe that when active learning is performed, there is a drop in the mIoU function and then it continues to rise. In response to this phenomenon, we adapt our approach by refitting the mIoU function after each round of image annotation, utilizing the newly obtained mIoU values. This process will iteratively fit the mIoU function and estimating an optimal point for active sample addition, until reaching the maximum active annotation samples.

D. Pool Attention

In our semi-supervised learning method, the prediction quality of teacher model is crucial since it guides the student model's unsupervised learning process. However, the prediction of teacher model may be biased and inaccurate. Since pooling operator can make each token averagely aggregate its nearby token features and achieve excellent performance [38], we propose a non-parameter pool attention (PA) module to refine the prediction of teacher model. It conducts an average pooling operation based on the model class-wise prediction to mix the prediction of each class in adjacent domains. In medical image, tissue appears in groups, indicating that each voxel has at least one of its neighbors belonging to the same class. So average pooling operation in each class channel prediction can effectively extract domain attention and resist noise. The pipeline of PA is shown in Figure 3. Specifically, given the prediction of teacher model, we first conduct voxel-wise matrix multiplication with its average pooling output. And the final refined prediction is the matrix multiplication between normalized average pooling attention and the initial prediction.

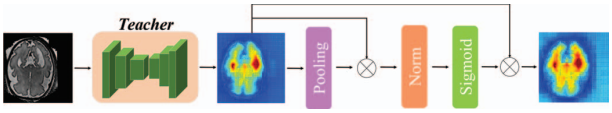

Fig. 3. Overview of our proposed Pool Attention. Norm is z-score normalization method on every class channel and dot-product is applied to calculate voxel-wise attention.

E. Objective Functions

The objective function $\mathcal{L}_{total}$ of optimization consists of two components: supervised loss $\mathcal{L}_{sup}$ and unsupervised loss $\mathcal{L}_{un}$, where $\mathcal{L}_{sup}$ is composed of cross entropy (CE) and dice loss (dice) and $\mathcal{L}_{un}$ is symmetric soft cross-entropy (SSCE) loss.

$$\mathcal{L}_{total} = \mathcal{L}_{sup} + \tau\mathcal{L}_{un} \tag{6}$$

where τ is the weighting factor, in our experiments, the τ is a gaussian ramp-up curve [44], *i.e.,* $\tau(t) = s * e^{-5(1-t)^2}$, where t is the training epoch as s scales the maximum value of the τ which is set to 0.5. $\mathcal{L}_{sup}$ is composed of a linear combination of cross entropy $\mathcal{L}_{ce}$ loss and dice loss $\mathcal{L}_{dice}$ [41] and it can be formulated as:

$$\mathcal{L}_{ce} = -\frac{1}{C} \times \sum_{c=1}^{C} y_s^l(c) \log\left(p_s^l(c)\right) \tag{7}$$

$$\mathcal{L}_{dice} = \frac{1}{C} \times \sum_{c=1}^{C} (1 - \frac{2\sum_{i=1}^{N} y_s^l(c) p_s^l(c)}{\sum_{i=1}^{N} y_s^l(c) + p_s^l(c)}) \tag{8}$$

where N is the total number of voxels in a sample and C is the number of class. And the supervised loss $\mathcal{L}_{sup}$ can be calculated by:

$$\mathcal{L}_{sup} = \mu\mathcal{L}_{ce} + \eta\mathcal{L}_{dice} \tag{9}$$

where μ and η is the weight factor which are both set to 0.5 in our experiments.

Considering that for noisy pseudo labels, the class-wise test accuracy varies significantly across different classes, and it is difficult for the network to learn difficult classes when using cross-entropy loss [43]. Due to the high similarities in representations between certain classes, the predictions for hard class examples are likely to assign a relatively large probability to those similar classes. Therefore, a simple cross-entropy loss is not sufficient for learning our task. In this paper, we propose a symmetric soft cross-entropy (SSCE) loss as unsupervised loss to exploit soft pseudo-labels to resist noise, which can be expressed as:

$$\mathcal{L}_{un} = \alpha\mathcal{L}_{sce}(p_s^u \mid \widetilde{p_t^u}) + \beta\mathcal{L}_{sce}(\widetilde{p_t^u} \mid p_s^u) \tag{10}$$

where α and β are two adjustable hyper-parameter and set to 0.1, 1, respectively. $\widetilde{p_t^u}$ and p_s^u represents refined teacher prediction and student prediction with soft label. The soft cross entropy loss $\mathcal{L}_{sce}$ can be defined as follows:

$$\mathcal{L}_{sce}(p_s^u \mid \widetilde{p_t^u}) = -\frac{1}{C} \times \sum_{c=1}^{C} p_s^u(c) \log \widetilde{p_t^u}(c) \tag{11}$$

TABLE I
DICE SCORE COMPARISON AMONG OUR METHOD AND OTHER RIVALS ON THE FETA1 DATASET WITH UNET-3D BACKBONE. 'MDC' INDICATES THE AVERAGE OF THE CLASS DICE SCORE.

Method	L	ECF[%]	GM[%]	WM[%]	Ven[%]	Cer[%]	DGM[%]	Bra[%]	mDC[%]
UNet-3D	60	80.12	72.64	90.83	83.74	87.23	84.26	81.80	82.95
SupOnly	6	75.29	66.21	85.38	70.07	73.80	71.20	70.22	73.17
UA-MT		74.39	65.45	87.08	66.42	76.01	71.45	71.92	73.24
TCSM		74.31	64.54	86.57	68.16	77.62	72.15	**72.49**	73.69
DTC		**75.68**	66.64	**87.33**	69.55	**81.04**	74.27	70.05	74.97
CPS		75.62	66.88	85.45	70.33	78.50	**74.58**	71.13	74.64
Ours	6	75.30	**67.15**	86.62	**77.73**	79.41	74.13	70.37	**76.34**
SupOnly	12	75.28	65.38	87.09	69.86	76.09	73.68	72.58	74.28
UA-MT		75.26	71.05	87.58	75.33	81.06	78.63	74.48	77.63
TCSM		75.64	68.95	88.65	75.59	83.02	77.73	73.73	77.62
DTC		75.32	70.76	89.05	79.28	83.58	**80.15**	73.26	78.77
CPS		75.64	71.39	**89.17**	79.92	84.25	79.96	75.17	79.36
Ours	12	**79.27**	**72.26**	89.02	**82.26**	**85.06**	80.04	**80.19**	**81.16**

TABLE II
DICE SCORE COMPARISON AMONG OUR METHOD AND OTHER RIVALS ON THE FETA2 DATASET WITH UNET-3D BACKBONE. 'MDC' INDICATES THE AVERAGE OF THE CLASS DICE SCORE.

Method	L	ECF[%]	GM[%]	WM[%]	Ven[%]	Cer[%]	DGM[%]	Bra[%]	mDc[%]
UNet-3D	30	78.74	68.56	84.05	73.34	78.16	75.44	68.53	75.26
SupOnly	3	34.51	26.54	37.18	27.85	2.71	35.24	8.41	24.63
UA-MT		39.37	29.24	24.39	19.32	5.71	34.89	7.31	26.92
TCSM		40.04	30.14	44.55	31.79	6.54	32.41	9.01	27.78
DTC		40.17	31.62	49.53	37.99	2.01	**35.99**	6.22	29.08
CPS		42.56	36.43	42.83	32.08	**11.92**	31.25	13.79	30.12
Ours	3	**48.26**	**37.24**	**55.93**	**45.69**	6.95	33.36	**14.11**	**34.51**
SupOnly	6	53.91	43.93	60.99	33.26	19.47	54.11	39.53	43.60
UA-MT		63.43	53.61	71.56	54.63	33.70	54.40	46.03	53.62
TCSM		67.52	54.04	68.44	46.25	53.82	56.38	48.37	56.40
DTC		65.20	58.31	74.48	60.15	38.60	66.72	49.85	59.05
CPS		67.62	59.01	73.67	64.23	44.38	63.23	51.47	60.52
Ours	6	**73.48**	**65.23**	**80.54**	**70.42**	**56.55**	**69.92**	**62.01**	**68.31**

IV. EXPERIMENT

A. Datasets

Extensive experiments were conducted to evaluate our proposed algorithm on two public fetal brain tissue MRI datasets using UNet-3D and VoxResNet as backbone.

FeTA1 dataset. The first dataset was Fetal Brain Tissue Annotation and Segmentation Challenge released in 2021 [35] [1], which included 80 T2-weighted fetal brain reconstructions with the corresponding label map that was manually segmented into 7 different tissues, with an in-plane resolution of $0.5mm \times 0.5mm$, and a slice thickness of 3 to 5 mm. In our experiments, we used 60 scans for training and 20 scans for validation. In active learning experiments, we random select 3 training samples as initial labeled samples in our work.

FeTA2 dataset. The second was Fetal Tissue Annotation Challenge released in 2022 [35] [2], which collects 40 T2-weighted fetal brain reconstructions with the corresponding ground truth from four different sites. For FeTA2, we used 30 scans for training and 10 scans for validation. In active learning experiments, we randomly select 1 training sample as our work's initial labeled sample.

[1]https://feta.grand-challenge.org/

[2]https://feta.grand-challenge.org/feta/

TABLE III
DICE SCORE COMPARISON AMONG OUR METHOD AND OTHER RIVALS ON TWO DATASETS WITH VOXRESNET BACKBONE. 'MDC' INDICATES THE AVERAGE OF THE DICE SCORE. 'MHD' DENOTES THE AVERAGE OF THE 95HD.

Dataset	Method	L	mDC [%]	mHD[%]	L	mDC[%]	mHD[%]
FeTA1	SupOnly	60	83.80	3.91	-	-	-
	SupOnly	6	74.03	15.75	12	74.49	10.83
	CPS		75.47	10.59		79.83	6.83
	Ours		**77.33**	**6.98**		**81.80**	**4.60**
FeTA2	SupOnly	30	78.64	2.45	-	-	-
	SupOnly	3	24.03	17.85	6	52.58	8.60
	CPS		34.76	16.22		62.52	6.67
	Ours		**40.10**	**13.94**		**69.75**	**4.88**

B. Implementation Details

Environment. All experiments in our work are implemented in Pytorch 1.9.0 and conducted under python 3.7 running on a NVIDIA GeForce RTX 2080Ti.

Backbone. In our experiments, we used UNet-3D [32] and VoxResNet [33] as the backbone for all experiments.

Training details. The network was trained by an SGD optimizer for 4k iterations, with an initial learning rate 0.01 and updated by $max(initial\_learning\_rate * (1 - iter\_num/total\_iter)^{0.9}, 1e^{-8})$. The batch size was 4 (2 labeled and 2 unlabeled). To avoid overfitting and mine more unseen information, we applied weak-strong augmentation strategy in our experiments. Specifically, we applied random crop, random flip and random rotate as weak augmentation (ζ') and RandAugment [34] as strong augmentation (ζ). We randomly cropped $96 \times 96 \times 96$ sub-volume as the network input and used a sliding window strategy to obtain the final results with a stride of $16 \times 16 \times 16$.

Evaluation metrics. The segmentation targets include 7 classes (except background) in our experiments, *i.e.* External Cerebrospinal Fluid (ECF), Grey Matter (GM), White Matter (WM), Ventricles (Ven), Cerebellum (Cer), Deep Grey Matter (DGM) and Brainstem (Bra). We respectively used dice (DC) and the 95% hausdorff distance (95HD) to quantitatively evaluate our method.

C. Comparison with SOTA Methods

In our study, we evaluated our semi-supervised learning (SSL) segmentation method against both fully-supervised approaches and four state-of-the-art SSL methods: the uncertainty-aware mean teacher model (UA-MT) [18], the transformation-consistent self-ensembling model (TCSM) [22], [23], the dual-task consistency method (DTC) [21], and the cross pseudo supervision strategy (CPS) [36]. All comparisons were conducted under identical experimental settings.

Results on the FeTA1 dataset, as shown in Table I, indicate that our model generally surpasses its competitors in terms of

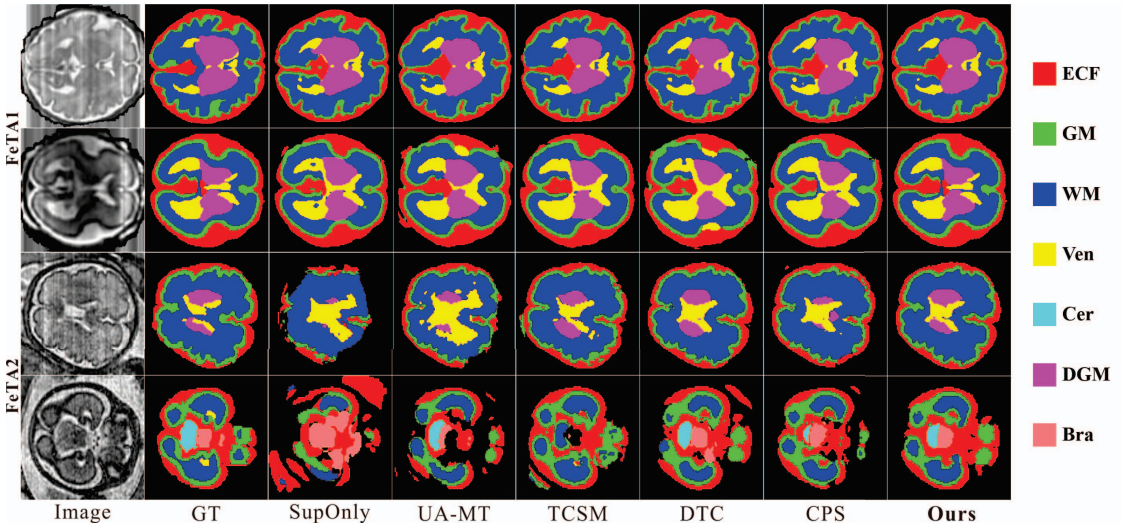

Fig. 4. Visual comparison with different state-of-the-art methods on two dataset2 with the unet_3D as backbone under 20% labeled sample. To better visualize the segmentation results, we only show one slice in one voxel.

TABLE IV
STATISTICAL COMPARISON OF OUR ABLATION STUDIES ON TWO DATASETS WITH 20% LABELED SAMPLES. 'MDC' INDICATES THE AVERAGE OF THE DICE SCORE. 'MHD' DENOTES THE AVERAGE OF THE 95HD.

Method	FeTA1			
	UNet-3D		VoxResNet	
	mDC[%]	mHD[%]	mDC[%]	mHD[%]
baseline	74.28	10.74	74.49	7.70
baseline+AL	78.58	6.15	79.31	6.54
baseline+PA	80.37	6.27	80.61	5.53
baseline+SSCE	79.92	5.05	80.22	4.87
baseline+PA+SSCE	**81.16**	**4.85**	**81.80**	**4.60**

dice score. This superiority is further illustrated in the upper two rows of Figure 4, which provide visual comparisons and highlight our method's enhanced segmentation capabilities.

Similarly, on the FeTA2 dataset, our method consistently outperforms others, as evidenced in Table II. In particular, the experimental results demonstrates our method's robust performance even with limited labeled data. The bottom two lines of Figure 4 offer additional visual comparisons, further underscoring the effectiveness of our approach in achieving superior segmentation results.

More experiments. To further verify the proposed algorithm's superiority, we conducted comparative experiments on different backbones. As shown in Table III, our method can still achieve better segmentation results than competitors under different backbones on both datasets.

D. Ablation Studies

To evaluate the effectiveness of components in the proposed method, we conducted the following ablation studies on FeTA1 datasets with two different backbones (UNet-3D and VoxResNet). The extensive ablation study results were shown in Table IV, demonstrating each component's contribution.

In our ablation studies, we defined the model trained over labeled data with only supervised loss as the 'baseline'. Except for 'baseline', all ablation experiments are conducted in active learning manner. 'baseline+AL' denoted that we select the most informative samples from the unlabeled pool for annotating during network training and add them to the labeled samples training process. For adding sample points, we estimate the growth of the mIoU curve to choose adding points. 'PA' was the proposed pool attention module leveraged to refine the prediction of the teacher model. 'SSCE' was the proposed softly-symmetrical cross entropy for resisting noisy pseudo labels. In Table IV, the experimental statistical results clearly demonstrated the effectiveness of 'PA' and 'SSCE', and the combination of the two owns better performance.

V. CONCLUSION

In this paper, we explore the potential of active annotations in semi-supervised fetal brain tissue segmentation task. We propose a novel SSL segmentation algorithm based on mean teacher framework with active learning and pseudo-label noise resistance. Specifically, we present an actively select strategy that leverages the global uncertainty variation of the teacher model to measure the information of unlabeled samples. Furthermore, we utilize the mIoU value between student model prediction and ground-truth labels/pseudo-labels to adjust the timing of performing active learning adaptively. In addition, we further present a parameter-free pooling attention (PA)

module to resist noise and improve model prediction quality and introduce a symmetric soft cross-entropy (SSCE) as unsupervised loss to resist noise. The quantitative comparison experiments and ablation studies based on two public fetal brain datasets demonstrate the superior performance of the proposed algorithm.

REFERENCES

[1] B. Boyle, M. C. Addor, L. Arriola, et al., "Estimating global burden of disease due to congenital anomaly: an analysis of European data," Archives of Disease in Childhood-Fetal and Neonatal Edition, vol. 103, no. 1, pp. 22-28, 2018.

[2] E. Buskens, D. E. Grobbee, I. M. E. Frohn-Mulder, et al., "Efficacy of routine fetal ultrasound screening for congenital heart disease in normal pregnancy," Circulation, vol. 94, no. 1, pp. 67-72, 1996.

[3] J. Espinoza, "Fetal MRI and prenatal diagnosis of congenital heart defects," The Lancet, vol. 393, no. 10181, pp. 1574-1576, 2019.

[4] C. Votino, J. Jani, N. Damry, et al., "Magnetic resonance imaging in the normal fetal heart and in congenital heart disease," Ultrasound in Obstetrics & Gynecology, vol. 39, no. 3, pp. 322-329, 2012.

[5] A. Khalil, S. Bennet, B. Thilaganathan, et al., "Prevalence of prenatal brain abnormalities in fetuses with congenital heart disease: a systematic review," Ultrasound in Obstetrics & Gynecology, vol. 48, no. 3, pp. 296-307, 2016.

[6] S. N. Saleem, "Fetal MRI: An approach to practice: A review," Journal of Advanced Research, vol. 5, no. 5, pp. 507-523, 2014.

[7] S. Asgari Taghanaki, A. Abhishek, J. P. Cohen, et al., "Deep semantic segmentation of natural and medical images: a review," Artificial Intelligence Review, vol. 54, pp. 137-178, 2021.

[8] X. J. Zhu, "Semi-supervised learning literature survey," 2005.

[9] V. Guizilini, R. Hou, J. Li, et al., "Semantically-guided representation learning for self-supervised monocular depth," arXiv preprint arXiv:2002.12319, 2020.

[10] T. Miyato, S. Maeda, M. Koyama, et al., "Virtual adversarial training: a regularization method for supervised and semi-supervised learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 8, pp. 1979-1993, 2018.

[11] X. Zhai, A. Oliver, A. Kolesnikov, et al., "S4l: Self-supervised semi-supervised learning," in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 1476-1485.

[12] J. E. Lawn, K. Wilczynska-Ketende, and S. N. Cousens, "Estimating the causes of 4 million neonatal deaths in the year 2000," International Journal of Epidemiology, vol. 35, no. 3, pp. 706-718, 2006.

[13] Lawn J E, Kerber K, Enweronu-Laryea C, et al. 3.6 million neonatal deaths—what is progressing and what is not?[C]//Seminars in perinatology. WB Saunders, 2010, 34(6): 371-386.

[14] S. Li, C. Zhang, and X. He, "Shape-aware semi-supervised 3D semantic segmentation for medical images," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2020, pp. 552–561.

[15] L. Wang, D. Guo, G. Wang, et al., "Annotation-efficient learning for medical image segmentation based on noisy pseudo labels and adversarial learning," IEEE Trans. Med. Imaging, vol. 40, no. 10, pp. 2795-2807, 2020.

[16] H. Wu, G. Chen, Z. Wen, et al., "Collaborative and adversarial learning of focused and dispersive representations for semi-supervised polyp segmentation," in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021, pp. 3489–3498.

[17] H. Peiris, Z. Chen, G. Egan, et al., "Duo-SegNet: Adversarial Dual-Views for Semi-Supervised Medical Image Segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2021, pp. 428–438.

[18] L. Yu, S. Wang, X. Li, et al., "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation," in Proc. Med. Image Comput. Comput. Assist. Intervent., Springer, 2019, pp. 605–613.

[19] Y. Wang, Y. Zhang, J. Tian, et al., "Double-uncertainty weighted method for semi-supervised learning," in Proc. Med. Image Comput. Comput. Assist. Intervent., Springer, 2020, pp. 542–551.

[20] Y. Wu, M. Xu, Z. Ge, et al., "Semi-supervised left atrium segmentation with mutual consistency training," in Proc. Med. Image Comput. Comput. Assist. Intervent.–MICCAI, Springer, 2021, pp. 297–306.

[21] X. Luo, J. Chen, T. Song, et al., "Semi-supervised Medical Image Segmentation through Dual-task Consistency," in Proc. AAAI Conf. Artif. Intell., vol. 35, no. 10, pp. 8801-8809, 2021.

[22] X. Li, L. Yu, H. Chen, et al., "Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model," arXiv preprint arXiv:1808.03887, 2018.

[23] X. Li, L. Yu, H. Chen, et al., "Transformation-Consistent Self-Ensembling Model for Semisupervised Medical Image Segmentation," IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 2, pp. 523-534, 2021.

[24] S. Huang, T. Wang, H. Xiong, et al., "Semi-supervised active learning with temporal output discrepancy," in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021, pp. 3447–3456.

[25] W. Zhang, L. Zhu, J. Hallinan, et al., "Boostmis: Boosting medical image semi-supervised learning with adaptive pseudo labeling and informative active annotation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2022, pp. 20666–20676.

[26] K. Chaitanya, N. Karani, C. F. Baumgartner, et al., "Semi-supervised and task-driven data augmentation," in Proc. Int. Conf. Inf. Process. Med. Imaging, Springer, 2019, pp. 29–41.

[27] P. Wang, J. Peng, M. Pedersoli, et al., "Self-paced and self-consistent co-training for semi-supervised image segmentation," Medical Image Analysis, vol. 73, 102146, 2021.

[28] K. Chaitanya, E. Erdil, N. Karani, et al., "Local contrastive loss with pseudo-label based self-training for semi-supervised medical image segmentation," arXiv preprint arXiv:2112.09645, 2021.

[29] Y. Xia, D. Yang, Z. Yu, et al., "Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation," Medical Image Analysis, vol. 65, 101766, 2020.

[30] Y. Zhou, Y. Wang, P. Tang, et al., "Semi-Supervised 3D Abdominal Multi-Organ Segmentation Via Deep Multi-Planar Co-Training," in 2019 IEEE Winter Conf. Appl. Comput. Vis., pp. 121–140.

[31] H. Tokunaga, B. K. Iwana, Y. Teramoto, et al., "Negative pseudo labeling using class proportion for semantic segmentation in pathology," in Computer Vision–ECCV, Springer, 2020, pp. 430–446.

[32] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, et al., "3D U-Net: learning dense volumetric segmentation from sparse annotation," in Proc. Med. Image Comput. Comput.-Assist. Intervent., Springer, 2016, pp. 424–432.

[33] H. Chen, Q. Dou, L. Yu, et al., "Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation," arXiv preprint arXiv:1608.05895, 2016.

[34] E. D. Cubuk, B. Zoph, J. Shlens, et al., "Randaugment: Practical automated data augmentation with a reduced search space," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2020, pp. 3008–3017.

[35] K. Payette, P. de Dumast, H. Kebiri, et al., "An automatic multi-tissue human fetal brain segmentation benchmark using the fetal tissue annotation dataset," Scientific Data, vol. 8, no. 1, pp. 1–14, 2021.

[36] X. Chen, Y. Yuan, G. Zeng, et al., "Semi-supervised semantic segmentation with cross pseudo supervision," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 2613–2622.

[37] A. Tarvainen, H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in Adv. Neural Inf. Process. Syst., vol. 30, 2017.

[38] W. Yu, M. Luo, P. Zhou, et al., "Metaformer is actually what you need for vision," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2022, pp. 10819-10829.

[39] S. Liu, K. Liu, W. Zhu, et al., "Adaptive early-learning correction for segmentation from noisy annotations," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2022, pp. 2606-2616.

[40] L. Liu, H. L. Johnson, S. Cousens, et al., "Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000," The Lancet, vol. 379, no. 9832, pp. 2151-2161, 2012.

[41] A. Abdollahi, B. Pradhan, A. Alamri, "VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data," IEEE Access, vol. 8, pp. 179424-179436, 2020.

[42] A. Odena, "Semi-supervised learning with generative adversarial networks," arXiv preprint arXiv:1606.01583, 2016.

[43] Y. Wang, X. Ma, Z. Chen, et al., "Symmetric cross entropy for robust learning with noisy labels," in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 322-330.

[44] S. Laine, T. Aila, "Temporal ensembling for semi-supervised learning," arXiv preprint arXiv:1610.02242, 2016.