

Intrapartum Ultrasound Image Segmentation of Pubic Symphysis and Fetal Head Using Dual Student-Teacher Framework with CNN-ViT Collaborative Learning

Jianmei Jiang¹, Huijin Wang¹, Jieyun Bai^{1,2,3} *, Shun Long¹, Shuangping Chen¹, Victor M. Campello^{4,5}, and Karim Lekadir^{4,5}

¹ College of Information Science and Technology, Jinan University, Guangzhou 510632, China

² Guangdong Provincial Key Laboratory of Traditional Chinese Medicine Information Technology, Jinan University, Guangzhou 510632, China

³ Auckland Bioengineering Institute, University of Auckland, Auckland 1010, New Zealand

⁴ Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona 08007, Spain

⁵ Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08007, Spain

Abstract. The segmentation of the pubic symphysis and fetal head (PSFH) constitutes a pivotal step in monitoring labor progression and identifying potential delivery complications. Despite the advances in deep learning, the lack of annotated medical images hinders the training of segmentation. Traditional semi-supervised learning approaches primarily utilize a unified network model based on Convolutional Neural Networks (CNNs) and apply consistency regularization to mitigate the reliance on extensive annotated data. However, these methods often fall short in capturing the discriminative features of unlabeled data and in delineating the long-range dependencies inherent in the ambiguous boundaries of PSFH within ultrasound images. To address these limitations, we introduce a novel framework, the Dual-Student and Teacher Combining CNN and Transformer (DSTCT), which synergistically integrates the capabilities of CNNs and Transformers. Our framework comprises a Vision Transformer (ViT) as the 'teacher' and two 'student' models — one ViT and one CNN. This dual-student setup enables mutual supervision through the generation of both hard and soft pseudo-labels, with the consistency in their predictions being refined by minimizing the classifier determinacy discrepancy. The teacher model further reinforces learning within this architecture through the imposition of consistency regularization constraints. To augment the generalization abilities of our approach, we employ a blend of data and model perturbation techniques. Comprehensive evaluations on the benchmark dataset of the PSFH Segmentation Grand Challenge at MICCAI 2023 demonstrate our DSTCT framework

* Jieyun Bai: jbai996@aucklanduni.ac.nz

outperformed 10 contemporary semi-supervised segmentation methods.
Code available at <https://github.com/jjm1589/DSTCT>.

Keywords: Intrapartum Ultrasound · Semi-supervised learning · Image segmentation · Transformer.

1 Introduction

The segmentation of the pubic symphysis and fetal head (PSFH) from intrapartum ultrasound (US) images is a critical step in developing an automated diagnostic system. This process is essential for generating quantitative descriptors, such as shape and size of PSFH. These descriptors are crucial in assessing labor progression and identifying potential delivery complications [7, 1, 12, 2, 15, 4, 6, 23]. Recent advancements in deep learning, particularly in Convolutional Neural Networks (CNNs) and Transformers, have significantly improved medical image segmentation. However, these models face challenges in clinical applications due to the scarcity of large-scale annotated training datasets. Deep learning approaches generally require extensive, labeled datasets to ensure model generalization. The collection of densely annotated US images is a complex task. It demands considerable time, medical expertise, and clinical experience for accurate pixel-wise labeling [3, 24]. In clinical practice, there is often a larger amount of unlabeled data than labeled data, emphasizing the importance of semi-supervised learning techniques. These techniques aim to improve the segmentation performance of US images by leveraging unlabeled data, a practice that is garnering increasing research interest.

Despite the progress in semi-supervised learning for US image segmentation, challenges persist due to common US imaging issues like shadow artifacts and unclear boundary lines. Most semi-supervised approaches rely on CNN architectures, which may lead to under-segmentation or over-segmentation due to their localized processing nature. In contrast, Transformer-based models offer a promising alternative. They excel in capturing wide-ranging, non-local interactions, potentially resolving CNNs’ inherent limitations. These models are adept at identifying and assimilating features from distant but visually similar regions, thus improving network feature discrimination. However, their effectiveness largely hinges on having access to extensive annotated datasets—a condition seldom met in medical imaging due to the limited availability of labeled data. Thus, developing methods to efficiently train Transformer-based models with a limited set of annotated data remains a formidable challenge in the area of US image segmentation.

This paper proposes a novel framework, the Dual-Student and Teacher Combining CNN and Transformer (DSTCT). Our approach uses a dual-student configuration, employing cross-supervision with hard pseudo labels to expand the training dataset. Additionally, our Consistency Learning with Soft Pseudo Labels (CLS) strategy aims to mitigate label noise and foster entropy minimization. Given the stark differences between CNN and Transformer models, we introduce

the Classifier Deterministic Difference (CDD) [8] to harmonize predictions between the two model types. The teacher model applies consistency regularization (CR) constraints, guiding the dual-student networks. This involves using data and model perturbations to align the dual-student networks' prediction maps with those of the teacher model, thus bolstering the overall model generalization. The main contributions of our work include: (1) we present the Dual-Student and Teacher Combining CNN and Transformer (DSTCT) framework. This new approach significantly enhances the utility of unlabeled data in semi-supervised segmentation of ultrasound (US) images; (2) our proposed framework effectively integrates hard and soft pseudo-label learning, entropy minimization, and consistency regularization; (3) through rigorous quantitative analysis, we establish that our DSTCT framework surpasses the performance of ten existing semi-supervised techniques in accurately segmenting the pubic symphysis and fetal head (PSFH) in intrapartum ultrasound images.

2 Method

For general semi-supervised learning, the training set is composed of a labeled dataset D_N^l with N labeled images and an unlabeled dataset D_M^u with M ($M \gg N$) raw images, and the full training dataset is denoted as $D_{N+M} = D_N^l \cup D_M^u$. For an image $X \in D_N^l$, its ground truth Y is available. But the ground truth is not provided for $X \in D_M^u$. P_{s1} and P_{s2} are the probability outputs derived from student1 and student2, respectively. P_{s1}^* is student1's soft pseudo labels.

An overview of the proposed DSTCT architecture is shown in Fig.1. It comprises a dual-student model and a single-teacher model. The dual-student model consists of a CNN model UNet (namely student1) and a Transformer model Swin-UNet (SUNet, namely student2), while the single-teacher model uses the same Transformer model SUNet. In particular, our DSTCT tackles the semi-supervised image segmentation from five various aspects: supervised learning (L_{sup}), cross-supervision with hard pseudo labels (L_h), consistency learning with soft pseudo labels (L_s), minimization of classifier determinacy discrepancy (L_{cdd}) and consistency regularization constraints from the teacher model (L_{cr}). Therefore, the overall training loss function for student1 or student2 can be defined as:

$$L_{total} = L_{sup} + \alpha L_h + \beta L_s + \gamma L_{cdd} + \mu L_{cr} \quad (1)$$

where α , β , γ and μ are trade-off weights, which were set $\alpha = 0.5$, $\beta = 1.0$, $\gamma = 3.0$, and $\mu = 0.1$ [18], respectively, in the proposed cooperative training process. Note: L_{cdd} is a common part of both students.

Supervised learning (L_{sup}). We use the labeled data to train the student models. Cross-entropy loss L_{ce} and Dice loss L_{dice} are used as follows:

$$L_{sup} = \frac{1}{2} \sum_{X \in D_N^l} (L_{ce}(P_{s1}, Y) + L_{dice}(P_{s1}, Y)) \quad (2)$$

Cross-supervision with hard pseudo labels (L_h). The predictions between the CNN and ViT have different properties, essentially in the output level. Based

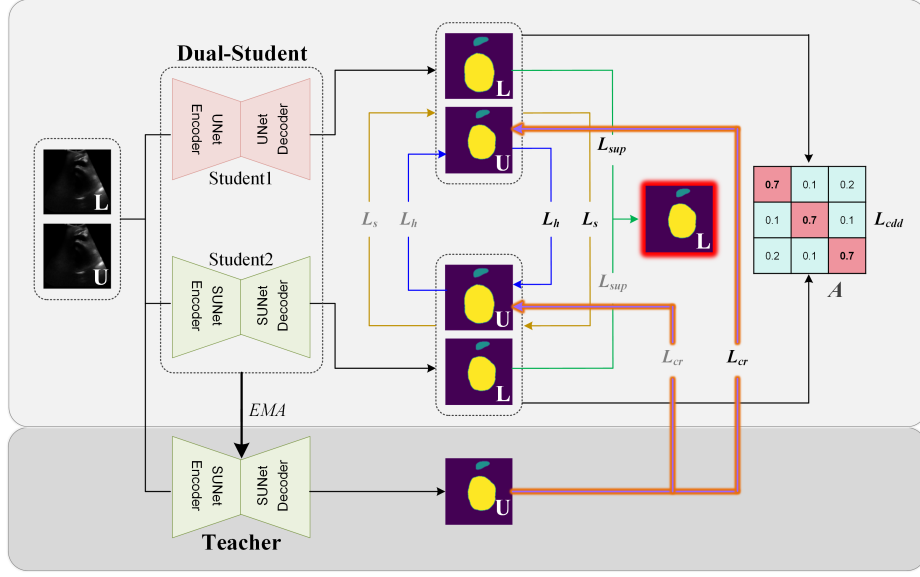


Fig. 1. An overview of the DSTCT architecture, where the black and grey parts are the loss functions for student1 and student2, respectively

on their predictions, we employ $\text{argmax}(\cdot)$ function to yield the hard pseudo labels for the cross-supervision between the peer networks. The cross-supervision loss of unlabeled data is defined as:

$$L_h = \sum_{X \in D_M^u} L_{dice}(P_{s1}, \text{argmax}(P_{s2})) \quad (3)$$

Consistency learning with soft pseudo labels (L_s). However, the hard pseudo labels generated by the maximum confidence are inevitably noisy, which may cause confusion bias during the segmentation training. To further reduce the noise of the hard pseudo labels and focus on unlabeled challenging regions, a sharpening function [9, 20] is utilized to generate soft pseudo labels, which can decrease the prediction uncertainty of the models. Soft pseudo labels can be obtained as follows:

$$P_{s1}^* = \frac{(P_{s2})^{1/\tau}}{(P_{s2})^{1/\tau} + (1 - P_{s2})^{1/\tau}} \quad (4)$$

where τ is a hyper-parameter to control the temperature of sharpening, and is set to 0.1 in our experiment [19]. Consistency learning is performed between the probability output of one model and the soft pseudo label of the other. The final loss function of consistency learning with soft pseudo labels is:

$$L_s = \sum_{X \in D_{N+M}} E[P_{s1}, P_{s1}^*] \quad (5)$$

where E is the Mean Squared Error (MSE) loss.

Classifier Determinacy Disparity (L_{cdd}). The diversity caused by structural differences between CNN and ViT may compromise the model’s performance. Directly aligning the discrepancy between the predictions is somewhat not confident. Therefore, we investigate classifier discrepancy by Bi-classifier Prediction Relevance Matrix: $A = P_{s1}P_{s2}^T$ [8]. The sum of the diagonal elements in matrix A represents the consistency of the classifier’s predictions, while the off-diagonal elements reflect the uncertainty of the predictions. We aim to maximize the former and minimize the latter. The minimizing classifier determinacy discrepancy loss is defined as follows:

$$L_{cdd} = \sum_{X \in D_{N+M}} \left[\sum_{m,n=1}^C A^{m,n} - \sum_{m=1}^C A^{m,m} \right] \quad (6)$$

where $A^{m,n}$ denotes the element in the m -th row and n -th column, and C is the number of categories.

Consistency Regularization (L_{cr}). The teacher architecture aims to minimize discrepancies between the predictions of the dual-student networks and the teacher network under both data and network perturbations. The consistency loss between the output probabilities of them is defined as follows:

$$L_{cr} = \sum_{X \in D_M} E[f_t(X; \bar{\theta}; \sigma'), f_{s1}(X; \varphi; \sigma)] \quad (7)$$

where σ and σ' represent different data perturbations and random dropout operations at the network layer, and φ and $\bar{\theta}$ represent the network parameter of student1 and teacher. $\bar{\theta}$ is updated via exponential moving average (EMA) from student2’s network. E is the MSE loss.

3 Experiments and Results

3.1 Dataset and Implementations

Dataset. This study leverages a dataset sourced from the MICCAI 2023 Grand Challenge [10], focused on the segmentation of the pubic symphysis (PS) and fetal head (FH). The comprehensive dataset includes 5,101 images, which have been methodically partitioned into training (70%), validation (10%), and testing (20%) subsets. Each image is paired with a precise segmentation mask delineating the FH and the PS, enabling the effective development and evaluation of pertinent segmentation models.

Implementation details. In this study, a compact version of ViT, pre-initialized with ImageNet weights is used for Swin-UNet (SUNet) architecture. The implementation was implemented on an Ubuntu 20.04 operating system, utilizing Python 3.8, PyTorch 1.10, and CUDA 11.3. A Tesla T4 GPU facilitated the computations. The network training regimen encompassed 30,000 iterations, employing the Stochastic Gradient Descent (SGD) optimizer, configured with

a momentum of 0.9 and a weight decay of 0.0001. We adopted a batch size 16, comprising an equal split of eight labeled and eight unlabeled images to support semi-supervised learning paradigms. The training commenced with an initial learning rate of 0.01, which was dynamically adjusted using a clustering-based learning rate strategy. Additionally, we introduced data noise perturbation within the range of $[-0.2, 0.2]$ to enhance model robustness against input variability.

Comparison Strategy. To verify the effectiveness of our DSTCT, we compare it with full/limited supervision (FS/LS) baselines and several state-of-the-art approaches for semi-supervised segmentation, including mean teacher (MT) [16], deep adversarial network (DAN) [22], uncertainty-aware mean teacher (UAMT) [21], cross consistency training (CCT) [13], cross pseudo-supervision (CPS) [5], cross teaching between CNN and transformer (CTCT) [11], interpolation consistency (ICT) [17], deep co-training (DCT) [14], self-integration method based on consent-aware pseudo-labels (S4CVnet) [18], collaborative Transformer-CNN learning (CTCL) [9].

Evaluation Metrics. The performance of these models was quantitatively evaluated using three established metrics: the Dice Similarity Coefficient (DSC), the 95% Hausdorff Distance (HD_{95}), and the Average Surface Distance (ASD). These metrics facilitate a thorough assessment of the segmentation models' accuracy and consistency, critical for validating their clinical applicability.

3.2 Comparison with Other Methods

Table 1 presents the quantitative outcomes on the PSFH dataset when trained with 20% of the total labeled training data. Notably, our proposed method significantly outperforms existing state-of-the-art approaches across most evaluation metrics. Remarkably, with only 20% labeled data for training, our DSTCT achieves 89.3% DSC performance, only 6% inferior to the upper bound performance.

Table 1. Quantitative comparison. The best result is in bold.

Labeled Method		PS			FH			PSFH		
		DSC $\uparrow$	ASD $\downarrow$	$HD_{95}\downarrow$	DSC $\uparrow$	ASD $\downarrow$	$HD_{95}\downarrow$	DSC $\uparrow$	ASD $\downarrow$	$HD_{95}\downarrow$
20%	LS(UNet)	0.835	0.793	3.875	0.897	1.901	8.868	0.866	1.347	6.372
	MT	0.803	0.760	5.143	0.872	2.912	13.608	0.837	1.836	9.376
	ICT	<u>0.847</u>	0.813	3.806	0.896	1.694	7.191	0.871	1.254	5.498
	UAMT	0.830	0.801	4.375	0.887	1.661	7.296	0.859	1.231	5.836
	DAN	0.824	1.048	4.583	0.886	1.443	6.871	0.855	1.246	5.727
	DCT	0.801	1.009	4.960	0.879	2.081	9.044	0.840	1.545	7.002
	CCT	0.824	<u>0.596</u>	4.388	0.886	2.693	11.604	0.855	1.645	7.996
	CPS	0.829	0.701	<u>3.788</u>	0.883	1.385	5.952	0.856	1.043	4.870
	CTCT	0.836	0.980	4.265	<u>0.908</u>	1.340	6.126	0.872	1.160	5.196
	CTCL	0.826	1.197	5.509	0.900	<u>0.747</u>	5.107	0.863	0.972	5.308
	S4CVnet	0.838	0.823	3.962	0.906	0.975	<u>4.658</u>	<u>0.872</u>	<u>0.899</u>	<u>4.310</u>
	OURS	0.852	0.581	3.331	0.935	0.351	2.150	0.893	0.466	2.740
100%	FL(UNet)	0.943	0.113	1.103	0.964	0.086	0.605	0.953	0.099	0.854
	FL(SUNet)	0.931	0.143	1.248	0.958	0.189	0.789	0.944	0.166	1.018

Additionally, Fig.2 offers visual comparisons between the segmentation results. These visualizations highlight that our method achieves more accurate segmentation with reduced false positives and fewer missed segmentation areas. A detailed examination of the comparative experiment results reveals that our proposed framework effectively integrates classical semi-supervised learning strategies, including pseudo-label learning, entropy minimization, and consistency regularization. This integration underscores the efficacy of our approach in leveraging semi-supervised learning techniques for improved segmentation performance.

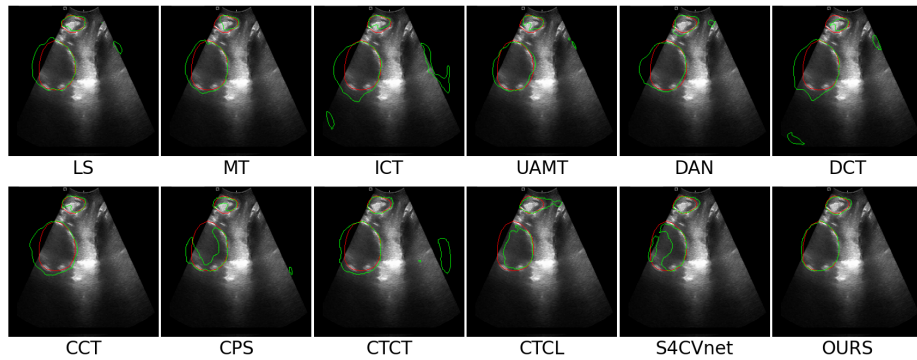


Fig. 2. Visual comparison of different methods when using 20% labeled data for testing. The figure shows the ground truth in red and the predicted results in green.

3.3 Ablation Study

Training strategy analysis. Each component of our DSTCT contributes differently to the enhancement of semi-supervised learning. From Table 2, it can be seen that each of the components and the combination of them can improve the PSFH segmentation performance, thus demonstrating the effectiveness of our method. Specifically, CDD and CR improve the DSC performance by 0.8% and 0.6%, respectively. By combining CLS, CDD and CR, the DSC performance is improved by 2.1%, the ASD is decreased by 0.626 mm, and the HD_{95} is decreased by 3.767 mm.

Different combination of Transformer and CNN. As illustrated in Table 2, The first row represents the combination of models proposed by the DSTCT framework, and its result is the best. While collaboratively training the SUNet-SUNet-SUNet or UNet-UNet-UNet models, the performance is relatively inferior to the UNet-SUNet-SUNet models, indicating the effectiveness of the complementary of these two models.

Table 2. Different strategies. Where s1, s2, and t are student1, student2, and teacher, respectively. The initial values of the trade-off weights of the loss function are $\alpha = 0.1$, $\beta = 0.1$, $\gamma = 0.01$, and $\mu = 0.1$, respectively. The best result is in bold.

Designs			PS			FH			PSFH		
CLS	CDD	CR	DSC $\uparrow$	ASD $\downarrow$	$HD_{95}\downarrow$	DSC $\uparrow$	ASD $\downarrow$	$HD_{95}\downarrow$	DSC $\uparrow$	ASD $\downarrow$	$HD_{95}\downarrow$
✓	✓	✓	0.836	0.688	3.821	0.890	2.133	9.741	0.863	1.410	6.781
			0.827	1.048	4.413	0.892	1.542	6.607	0.860	1.295	5.510
			0.830	1.002	4.323	0.912	0.815	3.775	0.871	0.909	4.049
✓	✓	✓	0.838	0.844	3.892	0.900	1.274	5.546	0.869	1.059	4.719
			0.848	0.731	3.819	0.902	1.555	2.493	0.875	1.143	5.294
			0.835	0.881	3.985	0.914	0.775	6.769	0.874	0.828	3.748
✓	✓	✓	0.842	0.606	3.594	0.891	1.742	7.664	0.866	1.174	5.829
			0.837	0.598	3.734	0.932	0.429	2.295	0.884	0.514	3.014
s1	s2	t									
UNet	SUNet	SUNet	0.837	0.598	3.734	0.932	0.429	2.295	0.884	0.514	3.014
SUNet	UNet	UNet	0.845	0.902	4.205	0.890	1.515	6.305	0.867	1.208	5.255
UNet	UNet	UNet	0.851	0.781	3.957	0.900	1.401	6.394	0.876	1.091	5.175
SUNet	SUNet	SUNet	0.820	0.966	4.336	0.920	0.433	2.810	0.870	0.699	3.573

Parameter analysis. The parameters α , β and γ are essential to control the importance of each loss in the objective function. Table 3 lists the results obtained with different parameter settings. When $\alpha = 0.5$, $\beta = 1.0$, $\gamma = 3.0$, DSTCT achieves the best DSC performance of 89.32%.

Table 3. Initial value analysis of the trade-off weights of the loss function. The results of the PSFH segmentation are shown.

γ		1.0			2.0			3.0		
α	β	DSC $\uparrow$	ASD $\downarrow$	$HD_{95}\downarrow$	DSC $\uparrow$	ASD $\downarrow$	$HD_{95}\downarrow$	DSC $\uparrow$	ASD $\downarrow$	$HD_{95}\downarrow$
0.5	1.0	0.8701	0.8985	4.1565	0.8917	0.4741	2.7726	0.8932	0.4659	2.7405
1.0	1.0	0.8671	0.7914	4.5107	0.8923	0.4979	2.7862	0.8912	0.4867	2.7949
2.0	1.0	0.8829	0.6551	3.2419	0.8903	0.5501	2.8618	0.8931	0.5049	2.7588
0.5	2.0	0.8732	0.7037	3.8760	0.8890	0.4487	2.7636	0.8839	0.4840	2.9388

4 Conclusion

In this study, we introduce a novel framework termed Dual-Student and Teacher Combining CNN and Transformer (DSTCT), designed to leverage the distinct inherent characteristics of CNN and ViT models through a synergistic training approach. The DSTCT framework intricately combines three key components: Consistency Learning with Soft Pseudo Labels (CLS), minimizing Classifier Determinacy Discrepancy (CDD), and Consistency Regularization (CR)

through mean teacher architecture. This comprehensive integration aims to bolster the generalization capabilities of CNNs. Empirical evaluations conducted on a widely recognized benchmark indicate that the DSTCT framework significantly enhances the performance of CNN architecture, outstripping competing state-of-the-art methods by a substantial margin. Furthermore, this study catalyzes advancing the application of Transformer models within the realm of semi-supervised image segmentation tasks, encouraging further research and development in this promising area. The PSFHS Challenge of MICCAI 2023 and the IUGC Challenge of MICCAI 2024 are available at <https://ps-fh-aop-2023.grand-challenge.org/> and <https://codalab.lisn.upsaclay.fr/competitions/18413>, respectively.

Acknowledgments. Supported in part by the Natural Science Foundation of Guangdong Province under Grant 2024A1515011886 and 2023A1515012833, the Guangzhou Science and Technology Planning Project under Grant 2023B03J1297, 2024B03J1283 and 2024B03J1289, Key Research and Development Program of Guangxi Province under Project No. 2023AB22074, the National Natural Science Foundation of China under Grant 61901192, and the China Scholarship Council 202206785002.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Angeli, L., Conversano, F., Dall’Asta, A., Eggebø, T., Volpe, N., Marta, S., Pisani, P., Casciaro, S., Di Paola, M., Frusca, T., et al.: Automatic measurement of head-perineum distance during intrapartum ultrasound: description of the technique and preliminary results. *The Journal of Maternal-Fetal & Neonatal Medicine* **35**(14), 2759–2764 (2022)
2. Bai, J., Lu, Y., Liu, H., He, F., Guo, X.: New technologies improve maternal and newborn safety (2024)
3. Bai, J., Sun, Z., Yu, S., Lu, Y., Long, S., Wang, H., Qiu, R., Ou, Z., Zhou, M., Zhi, D., et al.: A framework for computing angle of progression from transperineal ultrasound images for evaluating fetal head descent using a novel double branch network. *Frontiers in Physiology* **13**, 2565 (2022)
4. Chen, G., Bai, J., Ou, Z., Lu, Y., Wang, H.: Psfhs: Intrapartum ultrasound image dataset for ai-based segmentation of pubic symphysis and fetal head. *Scientific Data* **11**(1), 436 (2024)
5. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 2613–2622 (2021)
6. Chen, Z., Lu, Y., Long, S., Campello, V.M., Bai, J., Lekadir, K.: Fetal head and pubic symphysis segmentation in intrapartum ultrasound image using a dual-path boundary-guided residual network. *IEEE Journal of Biomedical and Health Informatics* (2024)
7. Chen, Z., Ou, Z., Lu, Y., Bai, J.: Direction-guided and multi-scale feature screening for fetal head–pubic symphysis segmentation and angle of progression calculation. *Expert Systems with Applications* **245**, 123096 (2024)

8. Li, S., Lv, F., Xie, B., Liu, C.H., Liang, J., Qin, C.: Bi-classifier determinacy maximization for unsupervised domain adaptation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 8455–8464 (2021)
9. Li, W., Yang, H.: Collaborative transformer-cnn learning for semi-supervised medical image segmentation. In: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1058–1065 (2022)
10. Lu, Y., Zhou, M., Zhi, D., Zhou, M., Jiang, X., Qiu, R., Ou, Z., Wang, H., Qiu, D., Zhong, M., et al.: The jnu-ifm dataset for segmenting pubic symphysis-fetal head. *Data in Brief* **41**, 107904 (2022)
11. Luo, X., Hu, M., Song, T., Wang, G., Zhang, S.: Semi-supervised medical image segmentation via cross teaching between cnn and transformer. In: International Conference on Medical Imaging with Deep Learning. pp. 820–833 (2022)
12. Ou, Z., Bai, J., Chen, Z., Lu, Y., Wang, H., Long, S., Chen, G.: Rtseg-net: a lightweight network for real-time segmentation of fetal head and pubic symphysis from intrapartum ultrasound images. *Computers in biology and medicine* **175**, 108501 (2024)
13. Ouali, Y., Hudelot, C., Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12674–12684 (2020)
14. Qiao, S., Shen, W., Zhang, Z., Wang, B., Yuille, A.: Deep co-training for semi-supervised image recognition. In: Proceedings of the european conference on computer vision (eccv). pp. 135–152 (2018)
15. Qiu, R., Zhou, M., Bai, J., Lu, Y., Wang, H.: Psfhsp-net: an efficient lightweight network for identifying pubic symphysis-fetal head standard plane from intrapartum ultrasound images. *Medical & Biological Engineering & Computing* pp. 1–12 (2024)
16. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems* **30** (2017)
17. Verma, V., Kawaguchi, K., Lamb, A., Kannala, J., Solin, A., Bengio, Y., Lopez-Paz, D.: Interpolation consistency training for semi-supervised learning. *Neural Networks* **145**, 90–106 (2022)
18. Wang, Z., Li, T., Zheng, J.Q., Huang, B.: When cnn meet with vit: Towards semi-supervised learning for multi-class medical image semantic segmentation. In: European Conference on Computer Vision. pp. 424–441 (2022)
19. Wu, Y., Ge, Z., Zhang, D., Xu, M., Zhang, L., Xia, Y., Cai, J.: Mutual consistency learning for semi-supervised medical image segmentation. *Medical Image Analysis* **81**, 102530 (2022)
20. Xie, Q., Dai, Z., Hovy, E., Luong, T., Le, Q.: Unsupervised data augmentation for consistency training. *Advances in neural information processing systems* **33**, 6256–6268 (2020)
21. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22. pp. 605–613 (2019)
22. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11–13, 2017, Proceedings, Part III 20. pp. 408–416 (2017)

23. Zhou, M., Wang, C., Lu, Y., Qiu, R., Zeng, R., Zhi, D., Jiang, X., Ou, Z., Wang, H., Chen, G., et al.: The segmentation effect of style transfer on fetal head ultrasound image: a study of multi-source data. *Medical & Biological Engineering & Computing* **61**(5), 1017–1031 (2023)
24. Zhou, M., Yuan, C., Chen, Z., Wang, C., Lu, Y.: Automatic angle of progress measurement of intrapartum transperineal ultrasound image with deep learning. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI* 23. pp. 406–414 (2020)