A Semantic Knowledge Complementarity based Decoupling Framework for Semi-supervised Class-imbalanced Medical Image Segmentation

Zheng Zhang[†], Guanchun Yin[†], Bo Zhang[‡], Wu Liu[↕], Xiuzhuang Zhou[†⇳] and Wendong Wang[‡]

Abstract

The limited data annotations have made semi-supervised learning (SSL) increasingly popular in medical image analysis. However, the use of pseudo labels in SSL degrades the performance of decoders that heavily rely on high-accuracy annotations. This issue is particularly pronounced in class-imbalanced multi-organ segmentation tasks, where small organs may be under-segmented or even ignored. In this paper, we propose SKCDF, a semantic knowledge complementarity based decoupling framework for multi-organ segmentation in class-imbalanced medical images. SKCDF decouples the data flow based on the responsibilities of the encoder and decoder during model training to make the model effectively learn semantic features, while mitigating the negative impact of unlabeled data on the semantic segmentation task. We also design a semantic knowledge complementarity module that adopts labeled data to guide the generation of pseudo labels and enriches the semantic features of labeled data with unlabeled data, which improves the quality of generated pseudo labels and the robustness of the overall model. Furthermore, we design an auxiliary balanced segmentation head based training strategy to further enhance the segmentation performance of small organs. Experimental results on the Synapse and AMOS datasets show that our method significantly outperforms existing methods.

1. Introduction

Computer-assisted diagnosis has been playing an increasingly significant role in the medical field, such as surgical planning and radiation therapy. Due to the high efficiency of deep learning models, researchers pay more and more attention to their application in medical data processing, such as image segmentation [5, 9, 20, 22, 24, 25, 29, 50] and etc. However, deep learning models often require a large amount of annotated data for training, which is a labor-intensive and time-consuming process, especially in medical image segmentation, where annotators need to master certain professional knowledge. As a result, semi-supervised learning (SSL) has become increasingly popular in medical image segmentation due to its ability to reduce the time and effort required for image annotation.

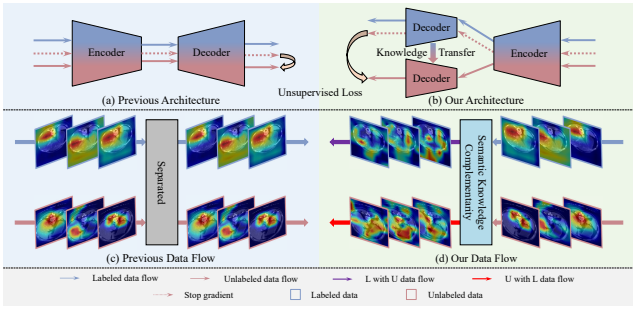

Figure 1. (a-b) Comparison between the architecture: previous methods used labeled and unlabeled data to train all parts of the model indiscriminately. Our method uses different data flows to train different parts of the model, suppressing the negative effects of unlabeled data on the model. (c-d) Comparison between the training data flows: our method adopts unlabeled data to enrich the semantic features of labeled data, while using labeled data to guide the generation of pseudo-labels.

Current SSL based medical image segmentation methods [2, 4, 15, 34, 38, 41, 43, 48] mainly focus on how to make better use of unlabeled data and generate high quality pseudo labels. However, as shown in Fig. 1 (a), these methods train all parts of themselves using both labeled and unlabeled data indiscriminately, ignoring that unlabeled data may have negative effects due to the lack of accurate annotation. In semantic segmentation tasks, the encoder is responsible for extracting the shape, texture, color and other feature information from the input images, which are subsequently used by the decoder to generate segmentation results. Since the encoder itself does not directly participate in the segmentation decision-making process, the encoder can

† School of Intelligent Engineering and Automation, Beijing University of Posts and Telecommunications (e-mail: zhangzheng, yinguanchun, xiuzhuang.zhou@bupt.edu.cn).

‡ State Key Laboratory of Networking and Switching Technology, School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications (e-mail: zbo, wdwang@bupt.edu.cn).

↕ School of Information Science and Technology, University of Science and Technology of China (e-mail: liuwu@ustc.edu.cn).

⇳ Beijing Ketai Industrial Intelligence (BKII).

Corresponding author: Bo Zhang. Our code is publicly available at: https://github.com/yinguanchun/SKCDF.

still learn rich feature representations even without precise semantic segmentation labels. In contrast to the encoder, decoder is directly involved in segmentation decisions, so the training process of the decoder relies heavily on precise semantic segmentation labels. If the labels are inaccurate, the performance of the decoder will be severely affected. GenericSSL [32] proposed an aggregating & decoupling framework, which adopted a diffusion decoder to guide the diffusion encoder to capture the distribution-invariant features, and they first decouple the labeled and unlabeled data training flows to solve the overfitting issues. However, such processing would be quite redundant in pure semi-supervised learning, and there is no detailed analysis of why decoupling data flows can improve the model's performance. Chen et al. [11] proposed an effective pseudo-labeling approach to segment unlabeled images in a prior-guided manner, but the complete independence between two decoders may result in insufficient learning from the unlabeled data. Additionally, their practice of directly averaging the outputs from various decoders not only lacks interpretability but also reintroduces the negative effects of unlabeled data on semantic segmentation tasks. To address above issues, as depicted in Fig. 1 (b), we propose a data flow decoupling framework, which adopts different data flows to train the different parts of the model so that the model can fully learn semantic features while suppressing the negative effects of unlabeled data on semantic segmentation tasks.

In addition, as shown in Fig. 1 (c), most existing methods ignore the intrinsic relationship between labeled and unlabeled data, which are completely independent of each other during the training process. In this way, deep learning models are likely to overfit labeled data and fail to adequately learn from large amounts of unlabeled data. In fact, there is an intrinsic relationship between labeled data and unlabeled data, especially in medical image data, where the same organ of different individuals has relatively fixed locations and similar semantic information. Therefore, through the semantic knowledge complementarity between labeled data and unlabeled data, on the one hand, the features of labeled data can be enriched, and a more complex environment can be provided for the learning of labeled data to prevent its overfitting. On the other hand, because the labeled data has accurate manual labels, we can use its more refined semantic knowledge to guide the generation of pseudo labels. Some studies [1, 7, 12, 33, 47] have noted this problem. AllSpark [33] adopted a channel-wise cross-attention mechanism to reborn the labeled features from unlabeled ones. Unfortunately, this method generates pseudo labels through unlabeled data flow, and does not make full use of existing high-accuracy annotations to guide the generation of pseudo labels. GuidedNet [47] proposed a 3D consistent Gaussian mixture model to leverage the knowledge from labeled data to guide the training of unlabeled data. However, this method cannot solve the problem of overfitting to labeled data, resulting in suboptimal performance. In this paper, we design a semantic knowledge complementarity module, which can adopt unlabeled data to enrich the features of labeled data, and adopt labeled data to guide the generation of high-quality pseudo labels for unlabeled data. As shown in Fig. 1 (d), the features of both labeled and unlabeled data possess richer and more organ-distribution-related semantic knowledge.

Furthermore, the class-imbalanced problem is very serious in medical image segmentation tasks. Most of the previous methods [3, 8, 13, 18, 31, 39] were based on reweighting or resampling, where reweighting requires additional calculations to determine new weights iteratively, and both of these methods weaken the model's learning of the majority classes. Thus, inspired by ABC [17], we develop an auxiliary balanced segmentation head training strategy, which improves the segmentation accuracy of small organs without losing the segmentation accuracy of other organs by capturing the features of small organs with auxiliary balanced segmentation heads. The main contributions can be summarized as follows:

- We propose a data flow decoupling framework for accurate multi-organ segmentation in class-imbalanced medical images, which can fully learn semantic features while suppressing the negative effects of unlabeled data on semantic segmentation tasks.
- We design a semantic knowledge complementarity module, which can adopt unlabeled data to enrich the features of labeled data to improve the robustness of the model, and adopt labeled data to guide high-quality pseudo-label generation for unlabeled data.
- We develop an auxiliary balanced segmentation head training strategy, which can improve the segmentation accuracy of small organs without losing the segmentation accuracy of other organs.
- We verify the effectiveness of our method on the Synapse and AMOS datasets. Experimental results show that our method achieves superior performance compared to several existing methods.

2. Related Work

2.1. Medical Image Segmentation

Ronneberger et al. [24] proposed the U-Net network, which significantly improves the segmentation performance of medical images through encoder-decoder structures and skip connections for fusing low resolution and high resolution semantic features, redefining the development direction of deep learning in the field of medical image segmentation. V-Net [23] proposed a volume-based, full-convolutional neural network 3D image segmentation method. Chen

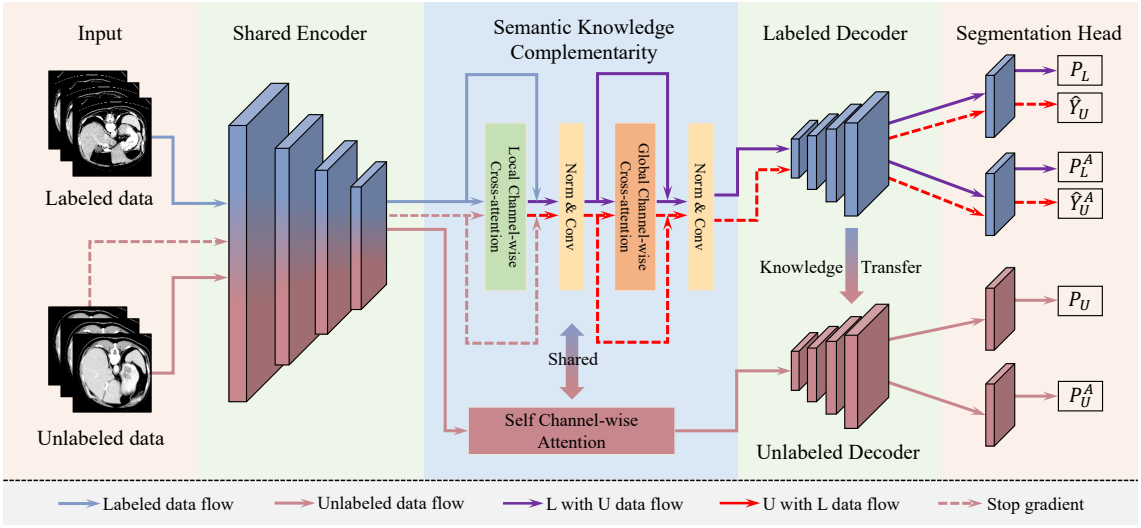

Figure 2. Overview of the proposed semantic knowledge complementarity based decoupling framework.

et al. [9] proposed TransUNet as a robust fusion network of Transformer and U-Net, which is the first network to introduce transformer into the field of medical image segmentation. Subsequently, more and more models [5, 14, 20, 25, 44, 50] have been proposed for medical image segmentation, which mainly focus on designing the network structure to adapt to the medical image segmentation task. Recently, several methods [26, 29, 30, 36, 45, 46, 49] have begun to train networks using medical prior knowledge. For example, Wang et al. [29] proposed a novel network for abdominal multi-organ segmentation, which incorporates radiologists' gaze information to boost high-precision segmentation. However, these methods often require a large amount of annotated data to train the model.

2.2. Semi-supervised Medical Image Segmentation

Semi-supervised medical image segmentation methods can be roughly classified into two categories: pseudo-labeling methods [6, 16, 19, 28, 41], and consistency regularization based methods [4, 15, 27, 34, 38, 40]. The basic pseudo-labeling method [16] first trained the model normally using labeled data, then pseudo labels are generated for the unlabeled data in each iteration, and the model is fine-tuned using this unlabeled data. Mean teacher [27] is a widely adopted consistency regularization based method for semi-supervised medical image segmentation, which learned unlabeled data through consistency training between the teacher model and the student model. UAMT [43] proposed a novel uncertainty-aware scheme to enable the student model to gradually learn from the meaningful and reliable targets by exploiting the uncertainty information. BCP [1] solved the empirical mismatch between labeled and unlabeled data distributions by copying and pasting labeled and unlabeled data bidirectionally. In fact, most of the leading methods combine pseudo-labeling and consistency regularization. UMCT [42] proposed an uncertainty-aware multi-view co-training to efficiently utilize unlabeled data. DHC [31] proposed a novel Dual-debiased Heterogeneous Co-training framework to solve the imbalanced class distribution problem. MCF [37] proposed a mutual correction framework to explore network bias correction in semi-supervised medical image segmentation.

3. Method

The overall framework is illustrated in Fig. 2. Our method adopts the basic pseudo-labeling framework as the baseline, where the unlabeled data is assigned with pseudo labels that are generated from current model predictions. Instead, we adopt a **data flow decoupling framework**, which allows the encoder to learn all the data features, and the two decoders learn the features of the labeled data and the unlabeled data, separately. To achieve the mutual guidance of labeled and unlabeled data, we incorporate a **semantic knowledge complementarity module**. Then, we propose an **auxiliary balanced segmentation head training strategy** to alleviate the class imbalance problem in semi-supervised medical image segmentation.

3.1. Data Flow Decoupling Framework

In previous semi-supervised learning, both labeled and unlabeled data are processed with the same structure, and the corresponding losses are calculated, then a weight coefficient is given to balance the loss of labeled and unlabeled data. In general, the weight coefficients are set very small, which can make the model not learn well from unlabeled data, and the model has little improvement compared to training with only labeled data. However, if it sets the weight coefficient too large, it is possible to make the model's performance worse. This is because, the unlabeled

data can still make the model learn more data features, but it is difficult for the unlabeled data to make the model perform well on the semantic segmentation task due to the lack of accurate ground truth. Therefore, we propose an end-to-end framework to make full use of unlabeled data to train the model to learn data features, while reducing the negative impact of unlabeled data on the model on semantic segmentation tasks.

As illustrated in Fig. 2, we adopt a VNet consisting of encoder, labeled decoder and unlabeled decoder, denoted by $E(\cdot)$, $D_l(\cdot)$, and $D_u(\cdot)$, separately. A batch of input data X contains equal labeled data (X_L, Y_L) and unlabeled data X_U, and we first generate a prediction with labeled data as:

$$P_L = D_l(E(X_L)) \tag{1}$$

So, we can get the supervised loss as:

$$L_{sup} = \frac{1}{|B_l|} \sum_{i=1}^{B_l} L_{DiceCE}(P_L, Y_L) \tag{2}$$

where B_l denotes the number of labeled data in a batch, and $L_{DiceCE}(x, y) = \frac{1}{2}[L_{Dice}(x, y) + L_{CE}(x, y)]$ is the combined Dice and cross entropy loss.

Then, we can generate the pseudo label $\hat{Y}_U$ for unlabeled data and the prediction of unlabeled data P_U as:

$$\hat{Y}_U = D_l(E(X_U)), P_U = D_u(E(X_U)) \tag{3}$$

Since, we can get the unsupervised loss as:

$$L_{unsup} = \frac{1}{|B_u|} \sum_{i=1}^{B_l} L_{DiceCE}(P_U, \hat{Y}_U) \tag{4}$$

where B_u denotes the number of unlabeled data in a batch. The optimization target is to minimize the overall loss, which can be formulated as:

$$L = L_{sup} + \lambda_u L_{unsup} \tag{5}$$

where λ_u is a hyper-parameter that can adjust the influence of unsupervised loss on the model. Since the encoder is trained on both labeled and unlabeled data, it can learn rich semantic features. We decouple the labeled and unlabeled data flows into the two decoders, the unlabeled data does not affect the labeled decoder, and therefore does not affect its effect on the semantic segmentation task. It is important to note that the semantic knowledge contained in unlabeled data may improve the decoder's performance in some other ways. An intuitive idea is to use unlabeled data to fine-tune the labeled decoder after pre-training, but this approach increases the cost of training on the one hand and inevitably weakens the semantic segmentation capability of the labeled decoder on the other hand. So we use a knowledge transfer strategy for the decoder as:

$$\xi_u = \omega_{ema} \times \xi_u + (1 - \omega_{ema}) \times \xi_l \tag{6}$$

where ξ_l and ξ_u denote the parameters of labeled and unlabeled decoders, respectively, and the ω_{ema} is a hyper-parameter. In this way, we can obtain an encoder and a decoder with rich semantic information, which will be discussed in detail in the Sec. 4.4.

3.2. Semantic Knowledge Complementarity Module

In order to make full use of unlabeled data to enrich the semantic features of labeled data, and use the labeled data to guide the unlabeled data to generate pseudo labels, we propose a semantic knowledge complementarity module, which improves the model's ability to capture long- and short-range relationships, enabling efficient mutual learning of semantic knowledge between labeled and unlabeled data. As shown in Fig. 2, we decouple our semantic knowledge complementarity module into local channel-wise cross-attention and global channel-wise cross-attention, and their detailed architecture is shown in Fig. 3.

Local Channel-wise Cross-attention. Given the hidden features $h^l, h^u \in \mathbb{R}^{C \times D \times H \times W}$ of labeled and unlabeled data, we reshape and map them into $e^l, e^u \in \mathbb{R}^{\frac{DHW}{P^3} \times P^3 \times C}$, where (P, P, P) is the resolution of each channel patch. Then, we can calculate the $query$, key and $value$ as follows:

$$q^l = e^l w_q, k^u = e^u w_k, v^u = e^u w_v \tag{7}$$

$$q^u = e^u w_q, k^l = e^l w_k, v^l = e^l w_v \tag{8}$$

where $w_q, w_k, w_v \in \mathbb{R}^{C \times NC}$ are transformation weights, N is the number of heads. The local channel-wise cross-attention is defined as:

$$\hat{e}^l = \sigma[\psi(q_l)^\mathsf{T} k^u](v^u)^\mathsf{T} w_{out} \tag{9}$$

$$\hat{e}^u = \sigma[\psi(q^u)^\mathsf{T} k^l](v^l)^\mathsf{T} w_{out} \tag{10}$$

where $\psi(\cdot)$ and $\sigma(\cdot)$ denote the instance normalization and the softmax function, $w_{out} \in \mathbb{R}^{NC \times C}$.

Global Channel-wise Cross-attention. Given the hidden features h^l, h^u of labeled and unlabeled data, we calculate the $query$, key, $value$ and the global channel-wise cross-attention similar to Eqs. (7) to (10). The only difference is that we reshape and map the hidden features into $e^l, e^u \in \mathbb{R}^{DHW \times C}$. Since, the whole process of our semantic knowledge complementarity module can be expressed as follows:

$$\hat{h} = Conv_L(LN(Attention_L(h) + h)) \tag{11}$$

$$\hat{\hat{h}} = Conv_G(LN(Attention_G(\hat{h}) + \hat{h})) \tag{12}$$

where h denotes the hidden features of labeled and unlabeled data after the encoder, $LN(\cdot)$ denotes the layer normalization, $Attention_L(\cdot)$ denotes the local channel-wise

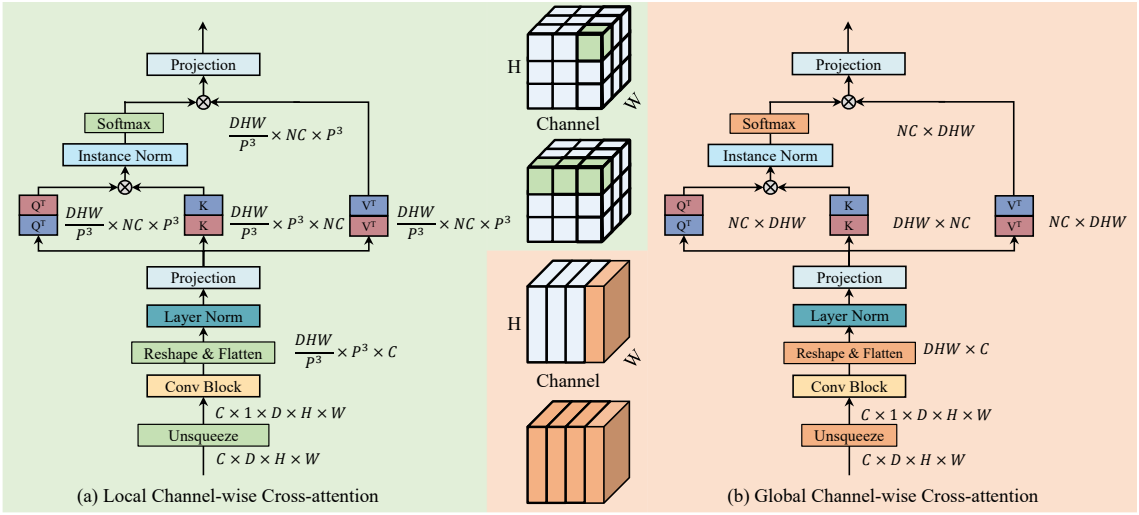

Figure 3. The detailed architecture of our proposed semantic knowledge complementarity module, which contains local channel-wise cross-attention and global channel-wise cross-attention. The different colors between Q, K, and V indicate that they come from different data flows. In order to facilitate visualization, in the middle is a schematic of the 2D image. The (a) is to perform attention operations in different channels of the two groups of data flows, and (b) is to perform attention operations in the same patch of different channels of the two groups of data flows.

cross-attention, $Attention_G(\cdot)$ denote the global channel-wise cross-attention, $Conv_L(\cdot)$ and $Conv_G(\cdot)$ denote the 1×1 convolution, respectively. Then, we input the features into the labeled decoder to generate predictions of the labeled data and the pseudo labels of unlabeled data. When we generate predictions of the unlabeled data, we replace the above operations with the corresponding self-attention mechanism.

Through the above methods, we adopt labeled data to guide unlabeled data to generate high quality pseudo labels, and unlabeled data to enrich the features of labeled data to improve the robustness of the model.

3.3. Auxiliary Balanced Segmentation Head Training Strategy

The class-imbalanced problem is very serious in medical image segmentation. Thus, inspired by ABC [17], we propose a method to improve the model's segmentation accuracy of the minority organs without weakening the learning of the majority organs.

As shown in Fig. 2, we first replace the last layer of the two decoders with two segmentation heads, respectively. Thus, we can get a total of 4 segmentation heads, $S_l^m(\cdot)$, $S_l^a(\cdot)$, $S_u^m(\cdot)$, and $S_u^a(\cdot)$. For the main segmentation heads $S_l^m(\cdot)$ and $S_u^m(\cdot)$, the loss function is calculated similarly to the previous formulas Eq. (2) and Eq. (4). The main difference is the loss calculation of the auxiliary balanced segmentation heads. The predictions of two auxiliary balanced segmentation heads and pseudo label are expressed as:

$$P_L^A = S_l^a(D_l(E(X_L))) \tag{13}$$

$$P_U^A = S_u^a(D_u(E(X_U))) \tag{14}$$

$$\hat{Y}_U^A = S_l^a(D_l(E(X_U))) \tag{15}$$

We first calculate the number of voxels of each organ in all labeled data before training. To train the auxiliary balanced segmentation head $S_l^a(\cdot)$ to be balanced, we generate $0/1$ mask $M(X_L)$ for each labeled data using a Bernoulli distribution $B(\cdot)$ with the parameter set to be inversely proportional to the number of voxels of each class as follows:

$$M(X_L) \sim B(\frac{N_L}{N_C}) \tag{16}$$

where N_L denotes the total number of voxels in labeled data, and N_C denotes the total number of voxels belonging to the class of the corresponding position in labeled data. This setting makes $B(\cdot)$ generate mask 1 with high probability for the voxels in the minority organs, but with low probability for those in the majority organs. Then, the supervised balanced loss is multiplied by the generated mask, which can be expressed as:

$$L_{sup}^{balance} = \frac{1}{|B_l|} \sum_{i=1}^{B_l} M(X_L) L_{DiceCE}(P_L^A, Y_L) \tag{17}$$

To train the auxiliary balanced segmentation head $S_u^a(\cdot)$ to be balanced, we generate $0/1$ mask $M(X_U)$ for each unlabeled data using a Bernoulli distribution by the pseudo label. To take full advantage of few unlabeled data voxels in the early stage, we gradually decrease the parameter of the Bernoulli distribution B$(\cdot)$ for X_U from 1 to $\frac{N_L}{N_C}$ as:

$$M(X_U) \sim B(1 - \frac{epochs}{max\_epochs}(1 - \frac{N_L}{N_C})) \tag{18}$$

where $epochs$ and max_epochs denote the current epoch and maximum epoch, respectively. Then the unsupervised balanced loss can be expressed as:

$$L_{unsup}^{balance} = \frac{1}{|B_U|} \sum_{i=1}^{B_U} M(X_U) L_{DiceCE}(P_U^A, \hat{Y}_U^A) \quad (19)$$

The final loss function is updated as:

$$L = L_{sup} + L_{sup}^{balance} + \lambda_u (L_{unsup} + L_{unsup}^{balance}) \quad (20)$$

Through the above methods, the auxiliary segmentation head can learn the semantic information of minority organs well, but the segmentation performance of the majority organs will be affected. Therefore, we adopt a knowledge transfer strategy similar to Sec. 3.1 for the two main segmentation heads.

4. Experiments

4.1. Dataset and Pre-processing

Synapse dataset. The Synapse dataset has 13 foreground classes, including spleen (Sp), right kidney (RK), left kidney (LK), gallbladder (Ga), esophagus (Es), liver (Li), stomach (St), aorta (Ao), inferior vena cava (IVC), portal & splenic veins (PSV), pancreas (Pa), right adrenal gland (RAG), left adrenal gland (LAG) with one background and 30 axial contrast-enhanced abdominal CT scans. Following DHC[31], we resample all the data to $80 \times 160 \times 160$, and randomly split them into 20, 4, and 6 scans for training, validation, and testing, respectively.

AMOS dataset. Compared with Synapse, the AMOS dataset excludes PSV but adds three new classes: duodenum(Du), bladder(Bl) and prostate/uterus(P/U). 360 scans are divided into 216, 24, and 120 scans for training, validation, and testing.

4.2. Implementation Details

We conduct our experiments using PyTorch 2.4.1, CUDA 12.6, and a single NVIDIA 3090 GPU. The network parameters are optimized with SGD with a momentum of 0.9 and an initial learning rate (lr) of 0.3 with a warming-up strategy: $lr = base\_lr \times (1 - \frac{epochs}{max\_epochs})^{0.9}$. In the training stage, we randomly crop a volume of size $64 \times 128 \times 128$, and don't use any other data augmentations. We train the networks for 1500 epochs with a batch size of 4, consisting of 2 labeled and 2 unlabeled data. We empirically set λ_u to 10, and ω_{ema} to 0.99. We run experiments on Synapse three times with different seeds to eliminate the effect of randomness due to the limited samples. In the inference stage, we use the unlabeled decoder for prediction. Final segmentation results are obtained using a sliding window strategy with a stride size of $32 \times 32 \times 16$. We choose two evaluation metrics: Dice Score (%) and Average Surface Distance (ASD) in voxel. Given two object regions, Dice computes the percentage of overlap between them, ASD computes the average distance between their boundaries.

4.3. Comparison with Existing Methods

We compare our method with several state-of-the-art semi-supervised segmentation methods [3, 6, 8, 10, 13, 18, 21, 31, 32, 35, 39, 41, 43], and some of them include class-imbalanced designs [3, 8, 13, 18, 31, 32, 39].

In Tab. 1, we summarize the results of 20% labeled Synapse dataset. We can observe that general semi-supervised methods are very poor at the segmentation performance of small organs, and the average Dice may even be 0. The methods that consider the class imbalance problem have better results on some minority classes, but they still fail to capture the features of some other small organs, such as esophagus, right adrenal gland and left adrenal gland. Our proposed method achieves the best result in terms of average Dice (64.27% ↑) and average ASD (1.45 ↓), and our method has significantly improved on some complex organs and small organs, such as the stomach (↑ 9.6%), pancreas (↑ 7.5%) and left adrenal gland (↑ 10.6%).

We then evaluate our method on 5% labeled AMOS dataset. As shown in Tab. 2, we can observe that our method achieves the best result in terms of average Dice (53.81% ↑), which achieves a 3.78% improvement compared to the state-of-the-art results. Similar to the Synapse dataset, the problem of class imbalance still exists, but our method still maintains superior performance on these complex organs and small organs, such as esophagus(↑ 5.5%), stomach(↑ 2.6%) and right adrenal gland(↑ 2.8%).

The visual segmentation results of the various methods are presented in Fig. 4. It is evident that the compared methods tend to either over-segment or under-segment, and our method performs better, especially on minority organs.

4.4. Ablation Studies

Effectiveness of each component in Method. We conduct ablation studies to show the impact of each component in our framework, the results of which are shown in Tab. 3. The first row in Tab. 3 represents the basic pseudo-labeling framework baseline. Compared to the baseline, employing the data flow decoupling framework yields improvements in the average Dice of 12.74%. When we add semantic knowledge complementarity module and auxiliary balanced segmentation head training strategy individually to the data flow decoupling framework, the average Dice increases by 1.77% and 7.17%, respectively. It is worth pointing out that our ABSH has a significant improvement on small organs, especially gallbladder, where the average Dice increases from 0.0% to 62.0%. Compared to the baseline, our complete method improves the average Dice by 19.61%.

Design choices of inference decoder and knowledge

	Method	Avg. Dice	Avg. ASD	Dice of Each Class												
				Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	Pa	RAG	LAG
	VNet (fully)	62.09 ± 1.2	10.28 ± 3.9	84.6	77.2	73.8	73.3	38.2	94.6	68.4	72.1	71.2	58.2	48.5	17.9	29.0
General	UA-MT[43]	20.26 ± 2.2	71.67 ± 7.4	48.2	31.7	22.2	0.0	0.0	81.2	29.1	23.3	27.5	0.0	0.0	0.0	0.0
General	URPC[21]	25.68 ± 5.1	72.74 ± 15.5	66.7	38.2	56.8	0.0	0.0	85.3	33.9	33.1	14.8	0.0	5.1	0.0	0.0
General	CPS[10]	33.55 ± 3.7	41.21 ± 9.1	62.8	55.2	45.4	35.9	0.0	91.1	31.3	41.9	49.2	8.8	14.5	0.0	0.0
General	SS-Net[41]	35.08 ± 2.8	50.81 ± 6.5	62.7	67.9	60.9	34.3	0.0	89.9	20.9	61.7	44.8	0.0	8.7	4.2	0.0
General	DST[6]	34.47 ± 1.6	37.69 ± 2.9	57.7	57.2	46.4	43.7	0.0	89.0	33.9	43.3	46.9	9.0	21.0	0.0	0.0
General	DePL[35]	36.27 ± 0.9	36.02 ± 0.8	62.8	61.0	48.2	54.8	0.0	90.2	36.0	42.5	48.2	10.7	17.0	0.0	0.0
Imbalance	Adsh[13]	35.29 ± 0.5	39.61 ± 4.6	55.1	59.6	45.8	52.2	0.0	89.4	32.8	47.6	53.0	8.9	14.4	0.0	0.0
Imbalance	CReST[39]	38.33 ± 3.4	22.85 ± 9.0	62.1	64.7	53.8	43.8	8.1	85.9	27.2	54.4	47.7	14.4	13.0	18.7	4.6
Imbalance	SimiS[8]	40.0 ± 0.6	32.98 ± 0.5	62.3	69.4	50.7	61.4	0.0	87.0	33.0	59.0	57.2	29.2	11.8	0.0	0.0
Imbalance	Basak et al.[3]	33.24 ± 0.6	43.78 ± 2.5	57.4	53.8	48.5	46.9	0.0	87.8	28.7	42.3	45.4	6.3	15.0	0.0	0.0
Imbalance	CLD[18]	41.07 ± 1.2	32.15 ± 3.3	62.0	66.0	59.3	61.5	0.0	89.0	31.7	62.8	49.4	28.6	18.5	0.0	0.0
Imbalance	DHC[31]	48.61 ± 0.9	10.71 ± 2.6	62.8	69.5	59.2	**66.0**	13.2	85.2	36.9	67.9	61.5	37.0	30.9	31.4	10.6
Imbalance	GenericSSL[32]	60.88 ± 0.7	2.52 ± 0.4	**85.2**	66.9	67.0	52.7	**62.9**	89.6	52.1	83.0	74.9	**41.8**	43.4	44.8	27.2
Imbalance	SKCDF (Ours)	**64.27 ± 1.36**	**1.45 ± 0.09**	79.5	**72.1**	**67.6**	59.8	60.7	**93.3**	**61.7**	**85.4**	**78.5**	**41.8**	**50.9**	**46.4**	**37.8**

Table 1. Quantitative comparison between our approach and SSL segmentation methods on **20% labeled Synapse dataset**. 'General' or 'Imbalance' indicates whether the methods consider the class imbalance issue or not. Sp: spleen, RK: right kidney, LK: left kidney, Ga: gallbladder, Es: esophagus, Li: liver, St: stomach, Ao: aorta, IVC: inferior vena cava, PSV: portal & splenic veins, Pa: pancreas, RAG: right adrenal gland, LAG: left adrenal gland. Results of 3-times repeated experiments are reported in the 'mean±std' format. Best results are boldfaced, and 2nd best results are underlined.

	Method	Avg. Dice	Avg. ASD	Dice of Each Class													
				Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	Pa	RAG	LAG	Du	Bl	P/U
	VNet (fully)	76.50	2.01	92.2	92.2	93.3	65.5	70.3	95.3	82.4	91.4	85.0	74.9	58.6	58.1	65.6	64.4	58.3
General	UA-MT[43]	42.16	15.48	59.8	64.9	64.0	35.3	34.1	77.7	37.8	61.0	46.0	33.3	26.9	12.3	18.1	29.7	31.6
General	URPC[21]	44.93	27.44	67.0	64.2	67.2	36.1	0.0	83.1	45.5	67.4	54.4	46.7	0.0	29.4	35.2	44.5	33.2
General	CPS[10]	41.08	20.37	56.1	60.3	59.4	33.3	25.4	73.8	32.4	65.7	52.1	31.1	25.5	6.2	18.4	40.7	35.8
General	SS-Net[41]	33.88	54.72	65.4	68.3	69.9	37.8	0.0	75.1	33.2	68.0	56.6	33.5	0.0	0.0	0.0	0.2	0.2
General	DST[6]	41.44	21.12	58.9	63.3	63.8	37.7	29.6	74.6	36.1	66.1	49.9	32.8	13.5	5.5	17.6	39.1	33.1
General	DePL[35]	41.97	20.42	55.7	62.4	57.7	36.6	31.3	68.4	33.9	65.6	51.9	30.2	23.3	10.2	20.9	43.9	**37.7**
Imbalance	Adsh[13]	40.33	24.53	56.0	63.6	57.3	34.7	25.7	73.9	30.7	65.7	51.9	27.1	20.2	0.0	18.6	43.5	35.9
Imbalance	CReST[39]	46.55	14.62	66.5	64.2	65.4	36.0	32.2	77.8	43.6	68.5	52.9	40.3	24.7	19.5	26.5	43.9	36.4
Imbalance	SimiS[8]	47.27	11.51	**77.4**	72.5	68.7	32.1	14.7	86.6	46.3	74.6	54.2	41.6	24.4	17.9	21.9	**47.9**	28.2
Imbalance	Basak et al.[3]	38.73	31.76	68.8	59.0	54.2	29.0	0.0	83.7	39.3	61.7	52.1	34.6	0.0	0.0	26.8	45.7	26.2
Imbalance	CLD[18]	46.10	15.86	67.2	68.5	71.4	41.0	21.0	76.1	42.4	69.8	52.1	37.9	24.7	23.4	22.7	38.1	35.2
Imbalance	DHC[31]	49.53	13.89	68.1	69.6	71.1	**42.3**	37.0	76.8	43.8	70.8	57.4	43.2	27.0	28.7	29.1	41.4	36.7
Imbalance	GenericSSL[32]	50.03	**5.21**	73.1	76.0	**76.5**	29.1	44.9	82.5	49.0	72.8	**61.7**	48.5	30.2	19.7	**36.4**	32.9	18.2
Imbalance	SKCDF (Ours)	**53.81**	5.97	77.1	**77.9**	71.2	34.1	**50.4**	**88.6**	**51.6**	**80.9**	58.9	**48.8**	**33.0**	**30.2**	32.2	45.9	26.4

Table 2. Quantitative comparison between our approach and SSL segmentation methods on **5% labeled AMOS dataset**. Sp: spleen, RK: right kidney, LK: left kidney, Ga: gallbladder, Es: esophagus, Li: liver, St: stomach, Ao: aorta, IVC: inferior vena cava, Pa: pancreas, RAG: right adrenal gland, LAG: left adrenal gland, Du: duodenum, Bl: bladder, P/U: prostate/uterus.

transfer strategy. We compare the results with different inference decoder choices and different knowledge transfer strategies. As shown in Tab. 4, When we do not adopt knowledge transfer strategy, using a labeled decoder as the final output results in an average Dice improvement of 3.18% compared to using an unlabeled decoder. This is because labeled data can provide more precise semantic information for segmentation tasks. When we use the knowledge transfer strategy, there is an extremely significant improvement in both decoders. This is because the knowledge transfer strategy makes the unlabeled decoder better, so that the unlabeled data can better train the encoder, and then make the labeled decoder better, forming a virtuous cycle. The average Dice of the unlabeled decoder has a 0.19% improvement compared to the labeled decoder, demonstrating that unlabeled decoders make full use of both labeled and unlabeled data, which is highly consistent with our motivation.

The knowledge transfer sequence between decoders and segmentation heads. Since we use the knowledge transfer strategy between the segmentation heads and the decoders, we compare different sequences of knowledge transfer strategy in order to choose the best one. As shown in Tab. 5, we observe that when the knowledge transfer strategy is used for the segmentation heads, the average Dice improves by at least 5.93% compared with no use. And when the order of "$Head \rightarrow Decoder \rightarrow Head$" is adopted, the average Dice can reach the highest 63.57%. In this sequence, the main segmentation head of the labeled decoder first learns

Baseline	DF	SKCM	ABSH	Avg. Dice	Avg. ASD	Dice of Each Class												
						Sp	RK	LK	Ga	Es	Li	St	Ao	IVC	PSV	Pa	RAG	LAG
✓				44.66 ± 2.5	33.46 ± 1.99	66.5	62.2	47.7	41.8	0.0	91.6	52.5	81.7	74.0	31.9	30.8	0.0	0.0
✓	✓			57.4 ± 0.9	13.0 ± 2.08	74.1	69.3	61.8	61.8	0.0	90.0	60.3	82.7	79.1	41.3	48.4	48.3	29.1
✓	✓	✓		59.17 ± 1.95	7.78 ± 4.61	76.9	67.2	66.9	49.9	17.9	92.8	63.2	84.4	78.0	40.2	50.1	42.9	38.8
✓	✓		✓	63.57 ± 1.21	1.26 ± 0.14	79.0	70.7	64.1	59.9	62.0	93.7	64.2	84.5	78.1	39.5	47.1	46.9	36.7
✓	✓	✓	✓	64.27 ± 1.36	1.45 ± 0.09	79.5	72.1	67.6	59.8	60.7	93.3	61.7	85.4	78.5	41.8	50.9	46.4	37.8

Table 3. Ablation study for the effectiveness of each component on **20% labeled Synapse dataset**. DF: data flow decoupling framework. SKCM: semantic knowledge complementarity module. ABSH: auxiliary balanced segmentation head training strategy.

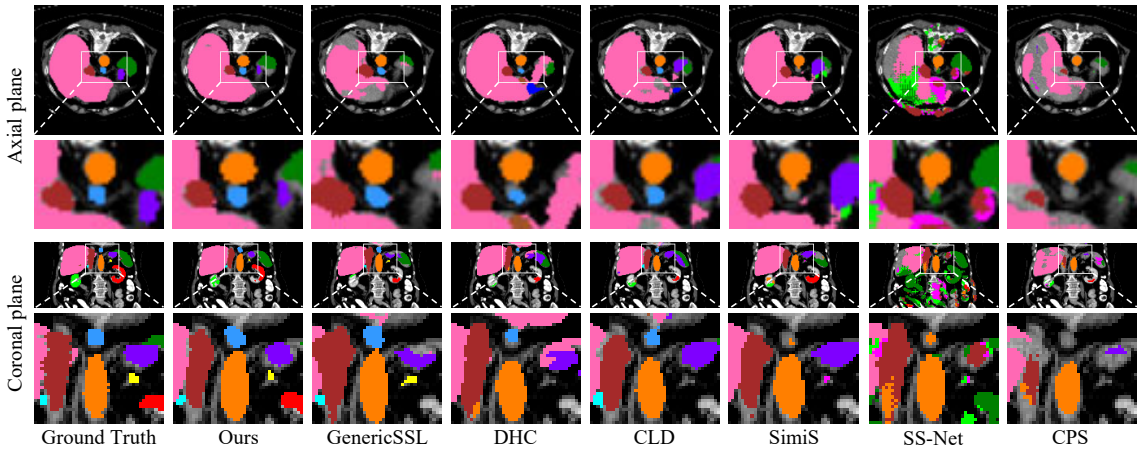

Figure 4. Visual comparison on the 20% labeled Synapse dataset: spleen, right kidney, left kidney, gallbladder, esophagus, liver, stomach, aorta, inferior vena cava, protal & splenic veins, pancreas, right adrenal gland, and left adrenal gland.

Method		Avg. Dice	Avg. ASD
Labeled	w/o EMA	53.71 ± 1.64	11.39 ± 0.04
Unlabeled	w/o EMA	50.53 ± 1.00	31.23 ± 0.78
Labeled	EMA	57.21 ± 0.85	11.56 ± 0.75
Unlabeled	EMA	57.40 ± 0.90	13.00 ±2.08

Table 4. Comparison of different training strategies and decoders in inference stage on **20% labeled Synapse dataset**. Labeled/unlabeled: inference with labeled/unlabeled decoder, EMA: adopt knowledge transfer strategy to decoder in training stage.

Method	Avg. Dice	Avg. ASD
D	54.79 ± 1.88	14.58 ± 4.91
H → D	62.51 ± 1.01	1.89 ± 0.76
D → H → H	60.90 ± 1.46	2.67 ± 0.64
H → H → D	62.40 ± 0.67	1.34 ± 0.12
H → D → H	63.57 ± 1.21	1.26 ± 0.14

Table 5. Comparison of different knowledge transfer sequences between decoders and segmentation heads on **20% labeled Synapse dataset**. D: decoder, H: segmentation head, →: direction of sequence.

the knowledge of the auxiliary balanced segmentation head, then passes it to the unlabeled decoder, and finally the main segmentation head of the unlabeled decoder can learn the knowledge of all the segmentation heads, so the best performance can be achieved.

5. Conclusions

Current semi-supervised methods usually use both labeled and unlabeled data to perform uniform model training without distinguishing them, which will degrade model performance. To address this issue, we propose a data flow decoupling framework, which adopts both labeled and unlabeled data to train the encoder, and the decoder is trained separately. This can make full use of the unlabeled data, and reduce the adverse impact of the unlabeled data due to the lack of accurate ground truth. Meanwhile, in order to use labeled data to guide the segmentation of unlabeled data and use unlabeled data to enrich the features of labeled data, we design a semantic knowledge complementarity module, which provides multi-scale features from different 3D medical image individuals and effectively improves the results of semi-supervised segmentation. We further develop a semi-supervised learning strategy based on auxiliary balanced segmentation head to improve the model's segmentation accuracy of the minority organs without weakening the learning of the majority organs. Finally, experiments on the Synapse and AMOS datasets show that our method significantly outperforms existing methods.

Acknowledgement

This work was supported in part by the Proof of Concept Program of Zhongguancun Science City and Peking University Third Hospital (No. HDCXZHKC2022202), the Natural Science Foundation of Beijing (No. 4252046), and the National Natural Science Foundation of China (No. 61972046, 61802022 and 61802027).

References

[1] Yunhao Bai, Duowen Chen, Qingli Li, Wei Shen, and Yan Wang. Bidirectional copy-paste for semi-supervised medical image segmentation. In *CVPR'23*, pages 11514–11524, 2023. 2, 3

[2] Hritam Basak and Zhaozheng Yin. Pseudo-label guided contrastive learning for semi-supervised medical image segmentation. In *CVPR'23*, pages 19786–19797, 2023. 1

[3] Hritam Basak, Sagnik Ghosal, and Ram Sarkar. Addressing class imbalance in semi-supervised image segmentation: A study on cardiac mri. In *MICCAI'22*, pages 224–233, 2022. 2, 6, 7

[4] Gerda Bortsova, Florian Dubost, Laurens Hogeweg, Ioannis Katramados, and Marleen De Bruijne. Semi-supervised medical image segmentation via learning consistency under transformations. In *MICCAI'19*, pages 810–818, 2019. 1, 3

[5] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *ECCV'22*, pages 205–218, 2022. 1, 3

[6] Baixu Chen, Junguang Jiang, Ximei Wang, Pengfei Wan, Jianmin Wang, and Mingsheng Long. Debiased self-training for semi-supervised learning. *NeurIPS'22*, 35:32424–32437, 2022. 3, 6, 7

[7] Duowen Chen, Yunhao Bai, Wei Shen, Qingli Li, Lequan Yu, and Yan Wang. Magicnet: Semi-supervised multi-organ segmentation via magic-cube partition and recovery. In *CVPR'23*, pages 23869–23878, 2023. 2

[8] Hao Chen, Yue Fan, Yidong Wang, Jindong Wang, Bernt Schiele, Xing Xie, Marios Savvides, and Bhiksha Raj. An embarrassingly simple baseline for imbalanced semi-supervised learning. *arXiv preprint arXiv:2211.11086*, 2022. 2, 6, 7

[9] Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, 97:103280, 2024. 1, 3

[10] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *CVPR'21*, pages 2613–2622, 2021. 6, 7

[11] Yaxiong Chen, Yujie Wang, Zixuan Zheng, Jingliang Hu, Yilei Shi, Shengwu Xiong, Xiao Xiang Zhu, and Lichao Mou. Striving for simplicity: Simple yet effective prior-aware pseudo-labeling for semi-supervised ultrasound image segmentation. In *MICCAI'24*, pages 604–614, 2024. 2

[12] Shengbo Gao, Ziji Zhang, Jiechao Ma, Zihao Li, and Shu Zhang. Correlation-aware mutual learning for semi-supervised medical image segmentation. In *MICCAI'23*, pages 98–108, 2023. 2

[13] Lan-Zhe Guo and Yu-Feng Li. Class-imbalanced semi-supervised learning with adaptive thresholding. In *ICML'22*, pages 8082–8094, 2022. 2, 6, 7

[14] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *WACV'22*, pages 574–584, 2022. 3

[15] Wei Huang, Chang Chen, Zhiwei Xiong, Yueyi Zhang, Xuejin Chen, Xiaoyan Sun, and Feng Wu. Semi-supervised neuron segmentation via reinforced consistency learning. *IEEE Transactions on Medical Imaging*, 41(11):3016–3028, 2022. 1, 3

[16] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *ICML'13*, page 896, 2013. 3

[17] Hyuck Lee, Seungjae Shin, and Heeyoung Kim. Abc: Auxiliary balanced classifier for class-imbalanced semi-supervised learning. *NeurIPS'21*, 34:7082–7094, 2021. 2, 5

[18] Yiqun Lin, Huifeng Yao, Zezhong Li, Guoyan Zheng, and Xiaomeng Li. Calibrating label distribution for class-imbalanced barely-supervised knee segmentation. In *MICCAI'22*, pages 109–118, 2022. 2, 6, 7

[19] Han Liu, Zhoubing Xu, Riqiang Gao, Hao Li, Jianing Wang, Guillaume Chabin, Ipek Oguz, and Sasa Grbic. Cosst: Multi-organ segmentation with partially labeled datasets using comprehensive supervisions and self-training. *IEEE Transactions on Medical Imaging*, 2024. 3

[20] Wentao Liu, Tong Tian, Weijin Xu, Huihua Yang, Xipeng Pan, Songlin Yan, and Lemeng Wang. Phtrans: Parallelly aggregating global and local representations for medical image segmentation. In *MICCAI'22*, pages 235–244, 2022. 1, 3

[21] Xiangde Luo, Wenjun Liao, Jieneng Chen, Tao Song, Yinan Chen, Shichuan Zhang, Nianyong Chen, Guotai Wang, and Shaoting Zhang. Efficient semi-supervised gross target volume of nasopharyngeal carcinoma segmentation via uncertainty rectified pyramid consistency. In *MICCAI'21*, pages 318–329, 2021. 6, 7

[22] Zibo Ma, Bo Zhang, Zheng Zhang, Wu Liu, Wufan Wang, Hui Gao, and Wendong Wang. Addg: An adaptive domain generalization framework for cross-plane mri segmentation. In *MM'24*, pages 5384–5392, 2024. 1

[23] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3DV'16*, pages 565–571, 2016. 2

[24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI'15*, pages 234–241, 2015. 1, 2

[25] Saikat Roy, Gregor Koehler, Constantin Ulrich, Michael Baumgartner, Jens Petersen, Fabian Isensee, Paul F Jaeger, and Klaus H Maier-Hein. Mednext: transformer-driven scaling of convnets for medical image segmentation. In *MICCAI'23*, pages 405–415, 2023. 1, 3

[26] Yucheng Tang, Dong Yang, Wenqi Li, Holger R Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-supervised pre-training of swin transformers for 3d medical image analysis. In *CVPR'22*, pages 20730–20740, 2022. 3

[27] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets im-

prove semi-supervised deep learning results. *NeurIPS'17*, 30, 2017. 3

[28] Bethany H Thompson, Gaetano Di Caterina, and Jeremy P Voisey. Pseudo-label refinement using superpixels for semi-supervised brain tumour segmentation. In *ISBI'22*, pages 1–5, 2022. 3

[29] Chong Wang, Daoqiang Zhang, and Rongjun Ge. Eye-guided dual-path network for multi-organ segmentation of abdomen. In *MICCAI'23*, pages 23–32, 2023. 1, 3

[30] Fakai Wang, Kang Zheng, Le Lu, Jing Xiao, Min Wu, and Shun Miao. Automatic vertebra localization and identification in ct by spine rectification and anatomically-constrained optimization. In *CVPR'21*, pages 5280–5288, 2021. 3

[31] Haonan Wang and Xiaomeng Li. Dhc: Dual-debiased heterogeneous co-training framework for class-imbalanced semi-supervised medical image segmentation. In *MICCAI'23*, pages 582–591, 2023. 2, 3, 6, 7

[32] Haonan Wang and Xiaomeng Li. Towards generic semi-supervised framework for volumetric medical image segmentation. *NeurIPS'24*, 36, 2024. 2, 6, 7

[33] Haonan Wang, Qixiang Zhang, Yi Li, and Xiaomeng Li. Allspark: Reborn labeled features from unlabeled in transformer for semi-supervised semantic segmentation. In *CVPR'24*, pages 3627–3636, 2024. 2

[34] Liansheng Wang, Jiacheng Wang, Lei Zhu, Huazhu Fu, Ping Li, Gary Cheng, Zhipeng Feng, Shuo Li, and Pheng-Ann Heng. Dual multiscale mean teacher network for semi-supervised infection segmentation in chest ct volume for covid-19. *IEEE Transactions on Cybernetics*, 53(10):6363–6375, 2022. 1, 3

[35] Xudong Wang, Zhirong Wu, Long Lian, and Stella X Yu. Debiased learning from naturally imbalanced pseudo-labels. In *CVPR'22*, pages 14647–14657, 2022. 6, 7

[36] Yan Wang, Xu Wei, Fengze Liu, Jieneng Chen, Yuyin Zhou, Wei Shen, Elliot K Fishman, and Alan L Yuille. Deep distance transform for tubular structure segmentation in ct scans. In *CVPR'20*, pages 3833–3842, 2020. 3

[37] Yongchao Wang, Bin Xiao, Xiuli Bi, Weisheng Li, and Xinbo Gao. Mcf: Mutual correction framework for semi-supervised medical image segmentation. In *CVPR'23*, pages 15651–15660, 2023. 3

[38] Ziyang Wang and Congying Ma. Dual-contrastive dual-consistency dual-transformer: A semi-supervised approach to medical image segmentation. In *ICCV'23*, pages 870–879, 2023. 1, 3

[39] Chen Wei, Kihyuk Sohn, Clayton Mellina, Alan Yuille, and Fan Yang. Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In *CVPR'21*, pages 10857–10866, 2021. 2, 6, 7

[40] Yicheng Wu, Minfeng Xu, Zongyuan Ge, Jianfei Cai, and Lei Zhang. Semi-supervised left atrium segmentation with mutual consistency training. In *MICCAI'21*, pages 297–306, 2021. 3

[41] Yicheng Wu, Zhonghua Wu, Qianyi Wu, Zongyuan Ge, and Jianfei Cai. Exploring smoothness and class-separation for semi-supervised medical image segmentation. In *MICCAI'22*, pages 34–43, 2022. 1, 3, 6, 7

[42] Yingda Xia, Dong Yang, Zhiding Yu, Fengze Liu, Jinzheng Cai, Lequan Yu, Zhuotun Zhu, Daguang Xu, Alan Yuille, and Holger Roth. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. *Medical Image Analysis*, 65:101766, 2020. 3

[43] Lequan Yu, Shujun Wang, Xiaomeng Li, Chi-Wing Fu, and Pheng-Ann Heng. Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In *MICCAI'19*, pages 605–613, 2019. 1, 3, 6, 7

[44] Bo Zhang, YunPeng Tan, Zheng Zhang, Wu Liu, Hui Gao, Zhijun Xi, and Wendong Wang. Factorized omnidirectional representation based vision gnn for anisotropic 3d multimodal mr image segmentation. In *MM'23*, pages 1607–1615, 2023. 3

[45] Zheng Zhang, Yushan Song, Yunpeng Tan, Shuo Yan, Bo Zhang, and Yufeng Zhuang. Segmentation assisted prostate cancer grading with multitask collaborative learning. *Pattern Recognition Letters*, 183:42–48, 2024. 3

[46] Zheng Zhang, Guanchun Yin, Zibo Ma, Yunpeng Tan, Bo Zhang, and Yufeng Zhuang. Ida-net: Individual difference aware medical image segmentation with meta-learning. *Pattern Recognition Letters*, 187:21–27, 2025. 3

[47] Haochen Zhao, Hui Meng, Deqian Yang, Xiaozheng Xie, Xiaoze Wu, Qingfeng Li, and Jianwei Niu. Guidednet: Semi-supervised multi-organ segmentation via labeled data guide unlabeled data. In *MM'24*, pages 886–895, 2024. 2

[48] Xiangyu Zhao, Zengxin Qi, Sheng Wang, Qian Wang, Xuehai Wu, Ying Mao, and Lichi Zhang. Rcps: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2023. 1

[49] Yuan Zhong, Chenhui Tang, Yumeng Yang, Ruoxi Qi, Kang Zhou, Yuqi Gong, Pheng Ann Heng, Janet H Hsiao, and Qi Dou. Weakly-supervised medical image segmentation with gaze annotations. In *MICCAI'24*, pages 530–540, 2024. 3

[50] Hong-Yu Zhou, Jiansen Guo, Yinghao Zhang, Xiaoguang Han, Lequan Yu, Liansheng Wang, and Yizhou Yu. nnformer: Volumetric medical image segmentation via a 3d transformer. *IEEE Transactions on Image Processing*, 2023. 1, 3