Medical Image Segmentation Model Training Approaches with Low Annotation Costs
低アノテーションコストの医用画像分割モデルの訓練アプローチ

張　路暘[1]

[1)]名古屋大学大学院情報学研究科　森研究室

Overview

- **Background and Introduction**
- **Topic 1**

Towards better laparoscopic video segmentation: A class-wise contrastive learning approach with multi-scale feature extraction.

- **Topic2**

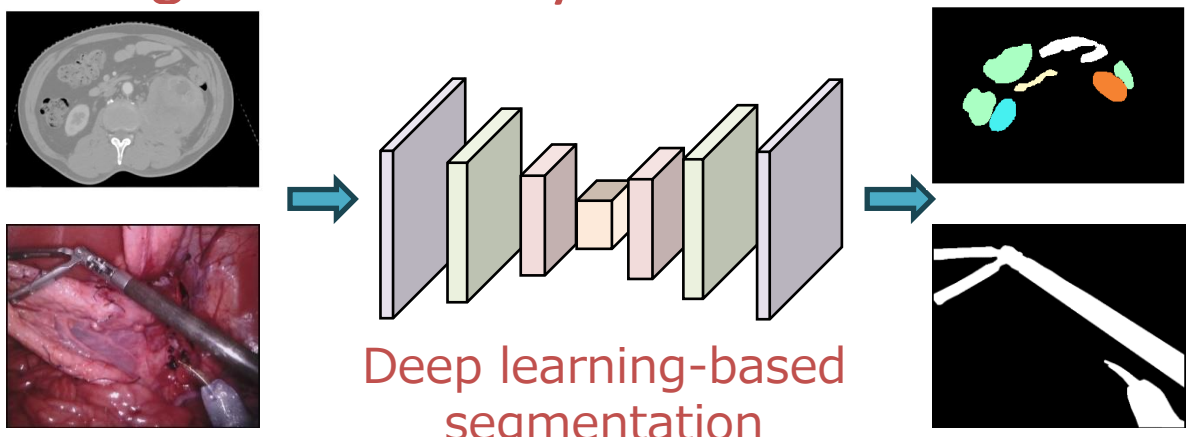
Double-Mix Pseudo-Label Framework: Enhancing Semi-Supervised Segmentation on Category-Imbalanced CT Volumes

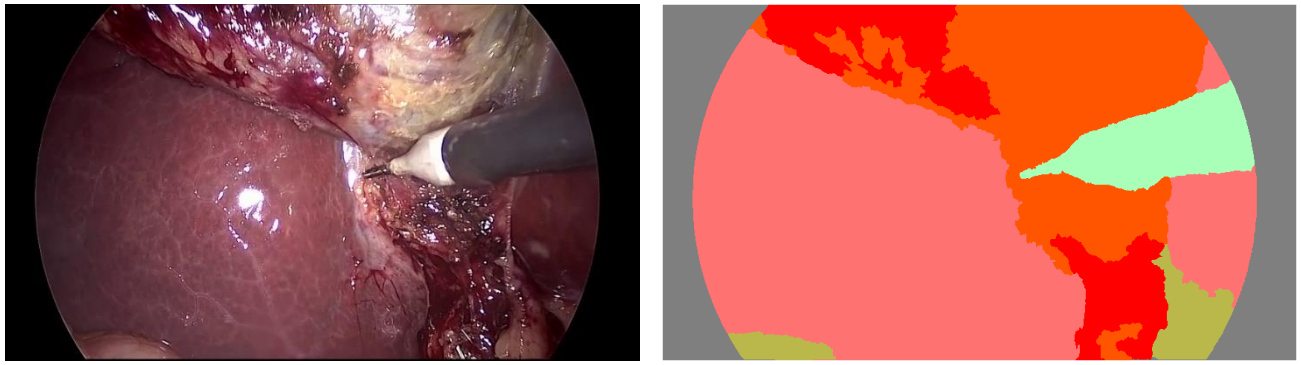
- **Conclusions and Foreseeing**

Background and Introduction

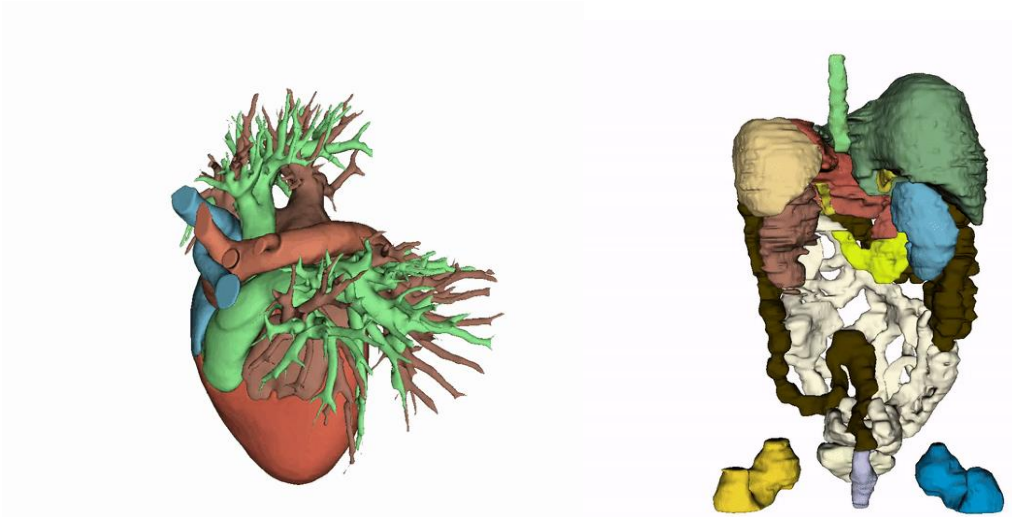
Medical image segmentation

- Enables diagnosis using high-precision CT images
- Supports surgery using high-precision laparoscopic images
- Development of deep learning-based segmentation systems is required





Segmentation of Organs and Tools from Endoscopic Video Images



Segmentation from CT scans,Heart (Left), Abdominal Organs (Right)

Challenge in medical image segmentation

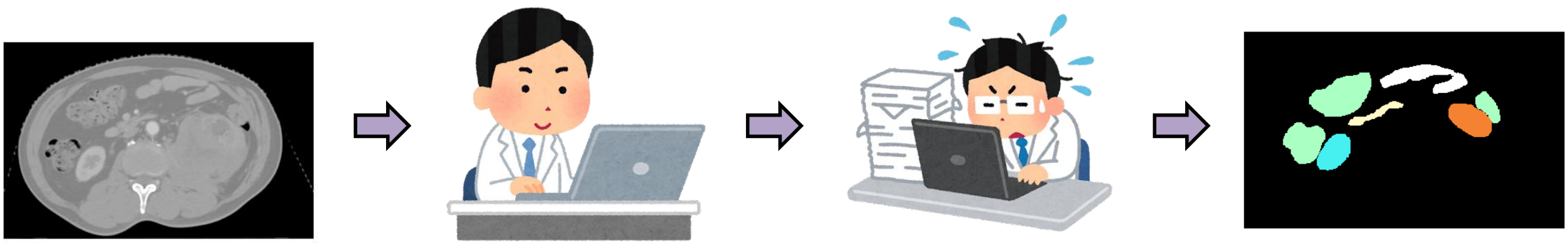
Time costs for annotation in 1 slice

Annotation type	CT	laparoscopic
Pixel-level	5-10 min	3-10 min
Category-level	5-15 s	10-30 s
No label	0	0

Wanying Shi, et al.

Problem:

- The need to create a large amount of annotated data
- Increasing annotation costs due to the growing number of medical images
- The need to **train high-accuracy segmentation models** with **low-cost annotations** in laparoscopic image and CT volume segmentation task



Creating labels requires expert knowledge and is time-consuming

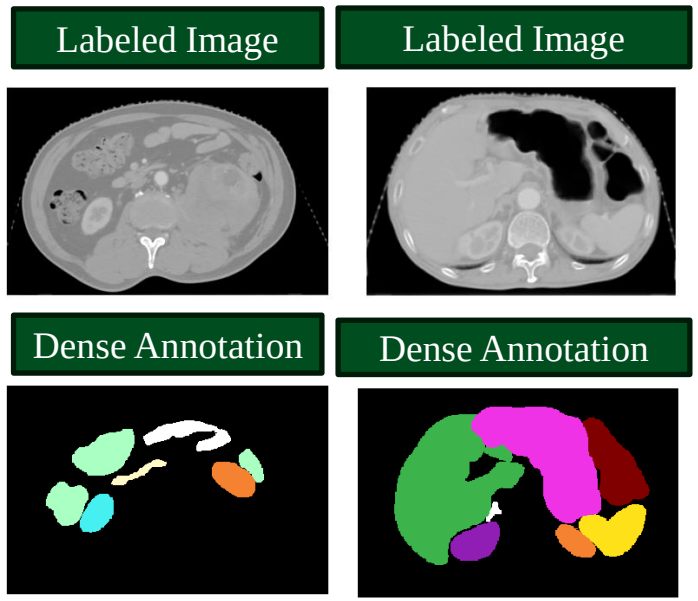
Related Works

[1] Li, Zihan, et al. “Scribblevc: Scribble-supervised medical image segmentation with vision-class embedding.” ACM, 2023.
[2] Wei, Hongbin, et al. "Only Classification Head Is Sufficient for Medical Image Segmentation." PRCV, 2023.
[3] Zeng, Dewen, et al. "Positional contrastive learning for volumetric medical image segmentation." MICCAI, 2021.

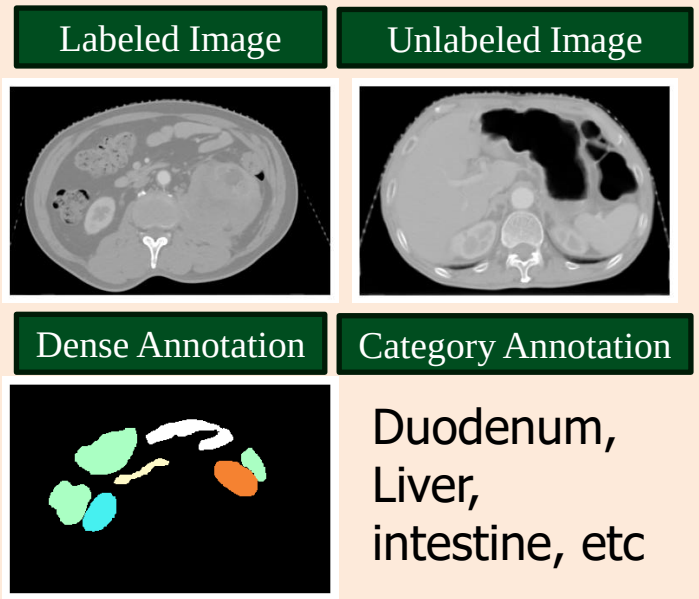
Utilizing category annotations.

- Using multi-task training for feature optimization [1,2].
- Using contrastive learning task for model pre-training [3].

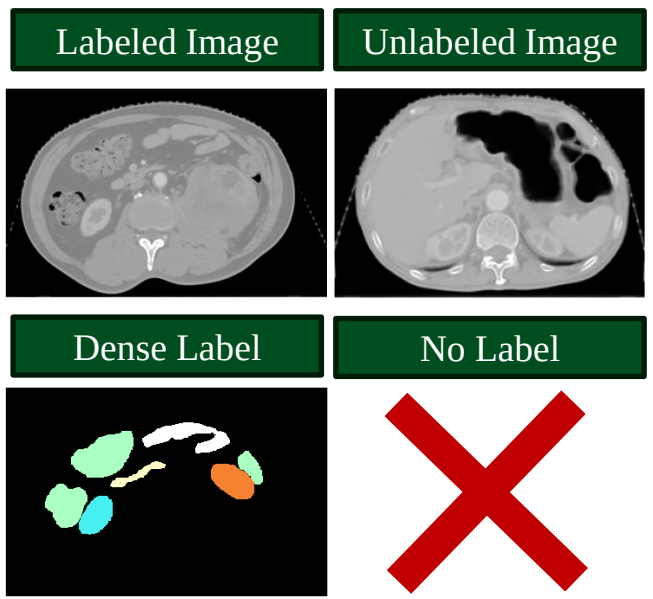
Scribble-Supervised Segmentation



Semi-Supervised Segmentation



Semi-Supervised Segmentation



High annotation cost — **Low annotation cost**

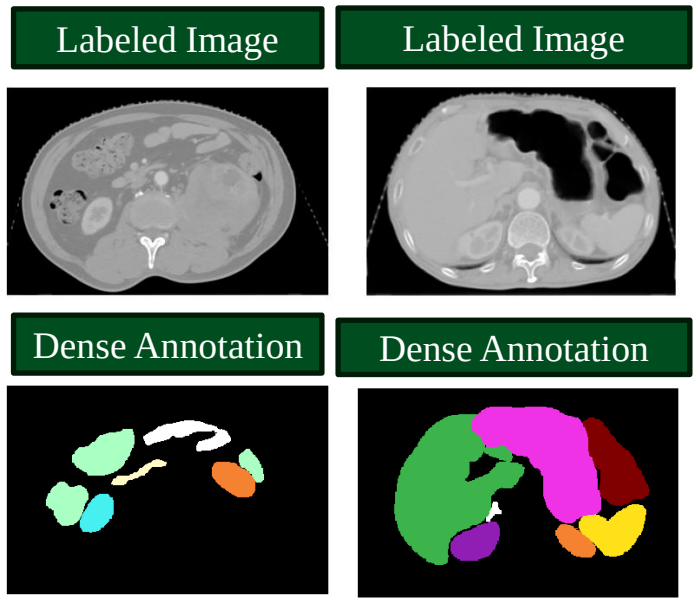
Related Works

[1] Yu, Lequan, et al. "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation." MICCAI, 2019.
[2] Kimhi, Moshe, et al. "Semi-Supervised Semantic Segmentation via Marginal Contextual Information." TMLR, 2024
[3] Chen, Xiaokang, et al. "Semi-supervised semantic segmentation with cross pseudo supervision." CVPR. 2021.

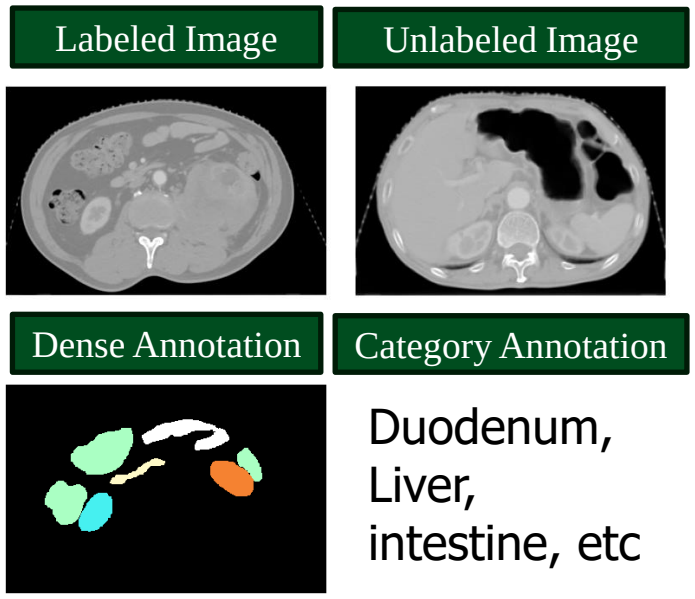
Utilizing few labeled data and lots of unlabeled data.

- Employing EMA models for pseudo-label learning [1,2].
- Applying ensemble learning for cross-supervision [3].

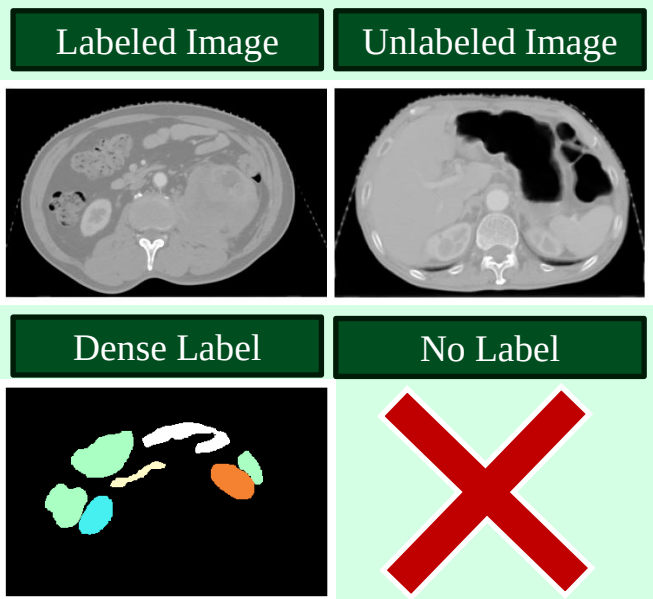
Scribble-Supervised Segmentation



Semi-Supervised Segmentation



Semi-Supervised Segmentation



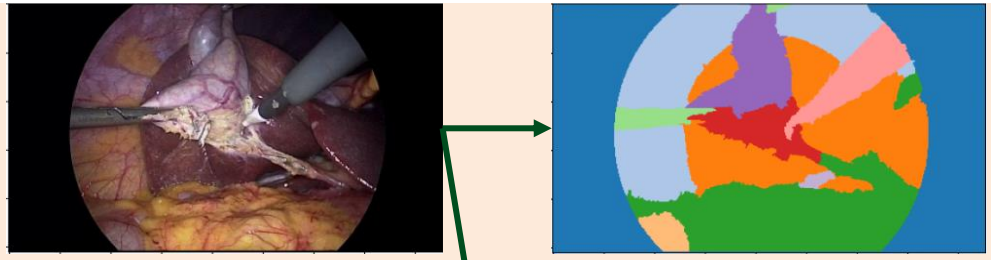
High annotation cost — **Low annotation cost**

Aim of our study

Propose two different solutions for laparoscopic and CT data.

- **Topic 1**

Train laparoscopic video segmentation model with limited pixel-Level annotated data and abundant category-Level annotated data



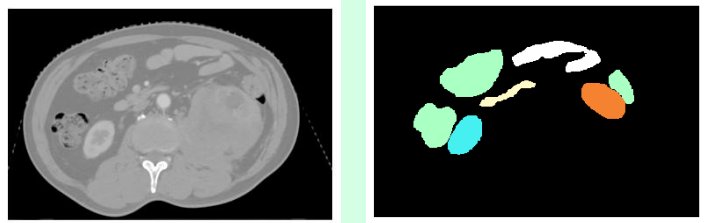
Pixel-Level annotation

"背景"、"腹壁"、"肝臓"、"脂肪"、"鉗子"、"胆嚢"、"フック"

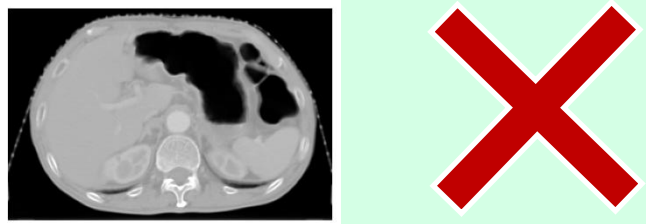
Category-Level annotation

- **Topic 2**

Train CT segmentation model with limited pixel-Level annotated data and abundant data without annotation



Volumes with annotation



Volumes with no annotation

- **Above all**

Training medical image segmentation models with low annotation costs

Topic 1

Towards better laparoscopic video segmentation: A class-wise contrastive learning approach with multi-scale feature extraction.

Zhang, Luyang, et al. "Towards better laparoscopic video segmentation: A class-wise contrastive learning approach with multi-scale feature extraction." *Healthcare Technology Letters* 11.2-3 (2024): 126-136.

Background

CAS requires segmentation

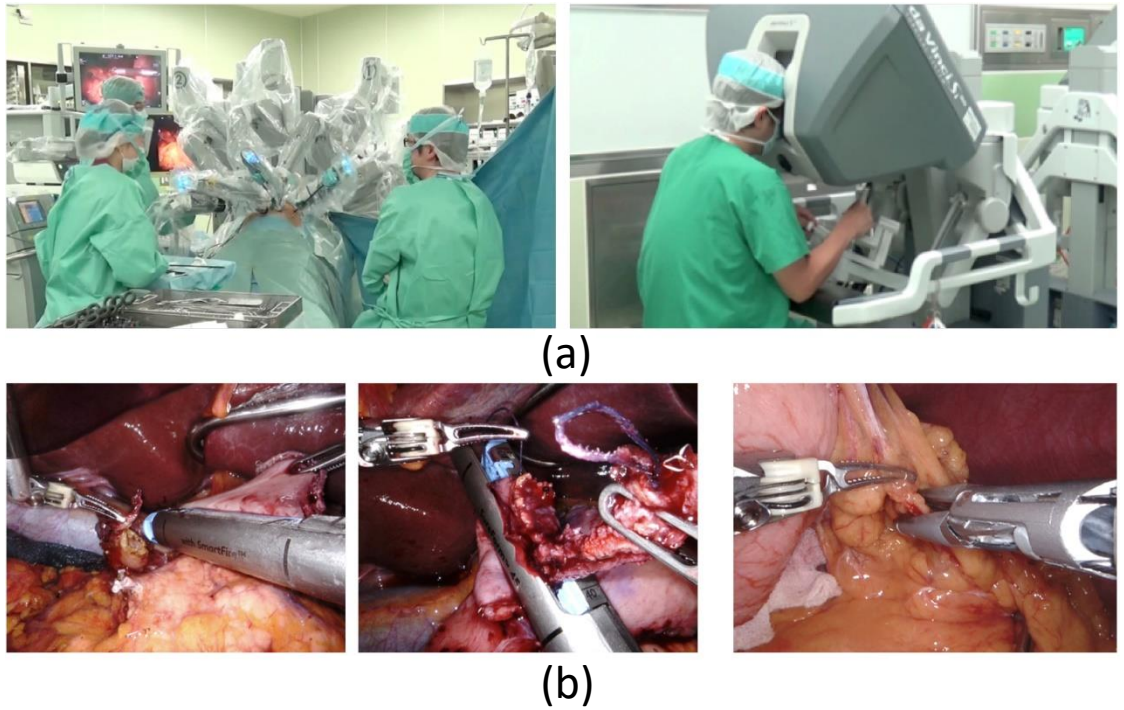
- Evaluate the condition of organs and tissues
- Identify the position and orientation of surgical tools

Deep learning-based CAS system

- To train high-precision models, extensive data annotation is necessary

Problem:

High annotation cost



(a) CAS
(b) Laparoscopic images including tools and organs [1]

[1] Hyung, Woo Jin. "Robotic surgery in gastrointestinal surgery." The Korean journal of gastroenterology= Taehan Sohwagi Hakhoe chi 50.4 (2007): 256-259.

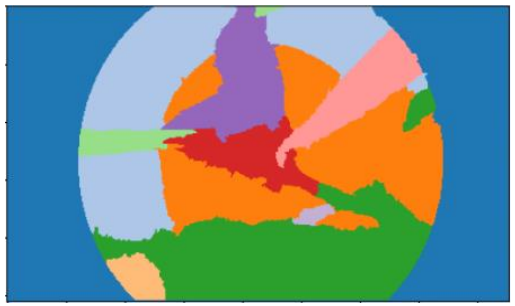
Low-cost annotation

Solution:

Utilize data with low-cost annotations.

Motivation:

Train a segmentation model using a small amount of pixel-level annotated data and a large amount of class-level annotated data.



Pixel-Level annotation

ツールと臓器：“背景”、“腹壁”、“肝臓”、“脂肪”、“鉗子”、“胆嚢”、“フック”
手術段階：“胆嚢の分離”

Category-Level annotation

Annotation cost	High ☹	**Low** ☺
Annotation information	**Complete** ☺	Incomplete ☹

Reduce annotation costs.

Multi-task training

Optimize the features extracted by the same segmentation model using multitasking.

Main Task

Segmentation Task

- Input: Data with segmentation annotations
- Label: Pixel-level annotations

Subtasks

Classification Task

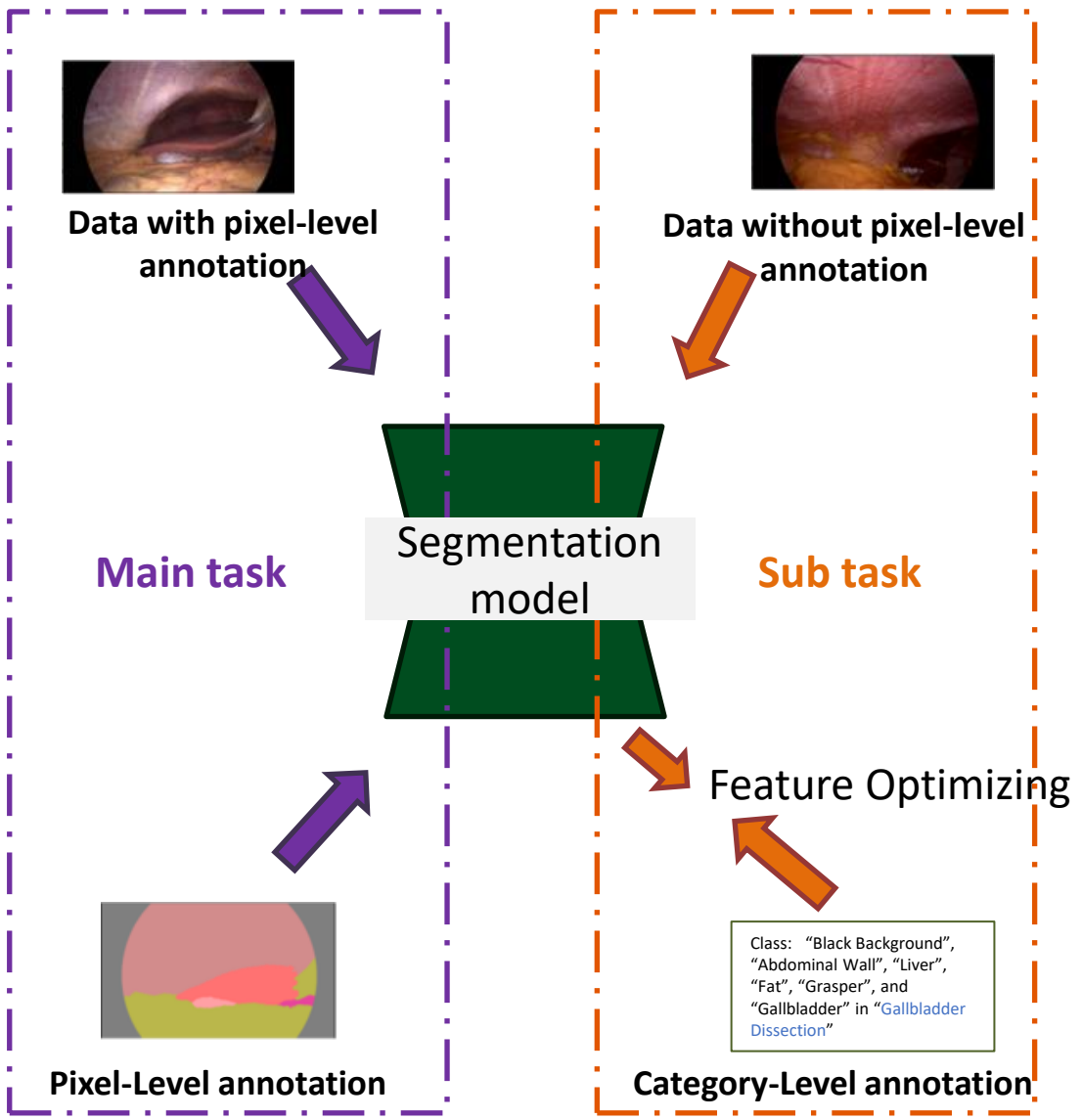
- Input: Data without segmentation annotations
- Label: Pixel-level annotations

Contrastive Learning Task

- Input: Positive pairs

Objective of subtasks:

Optimize the extracted features



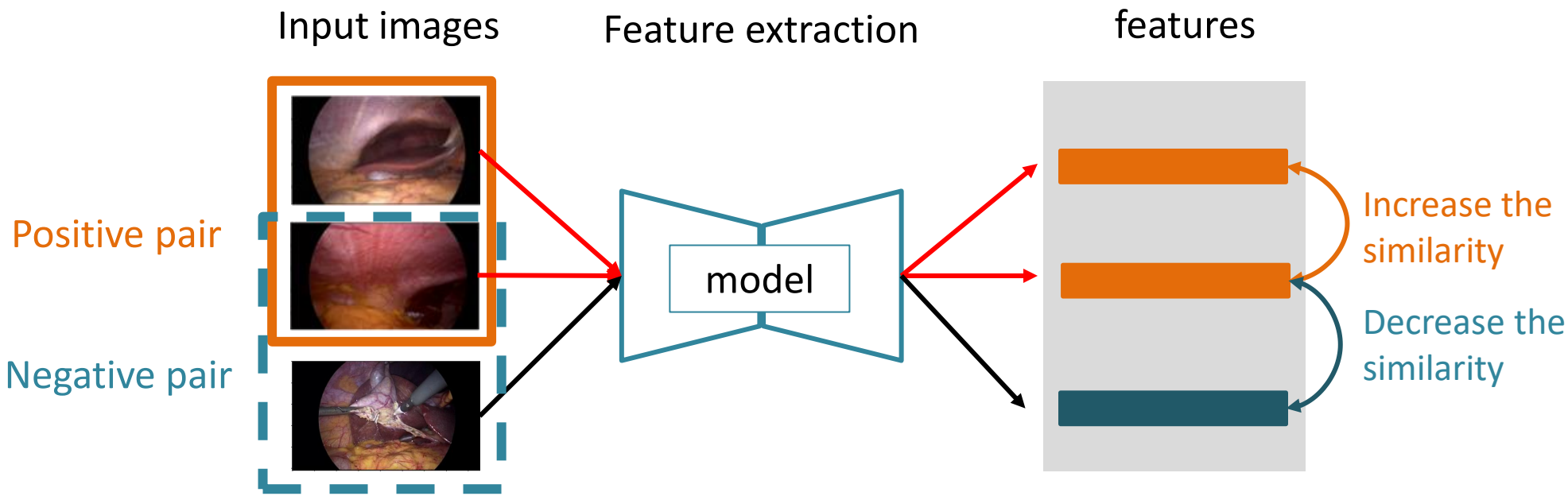
Contrastive learning

Positive Pair（正例）: Image pairs with similar information

Negative Pair: Image pairs with different information

Objective: Increase the similarity between features extracted from positive pairs

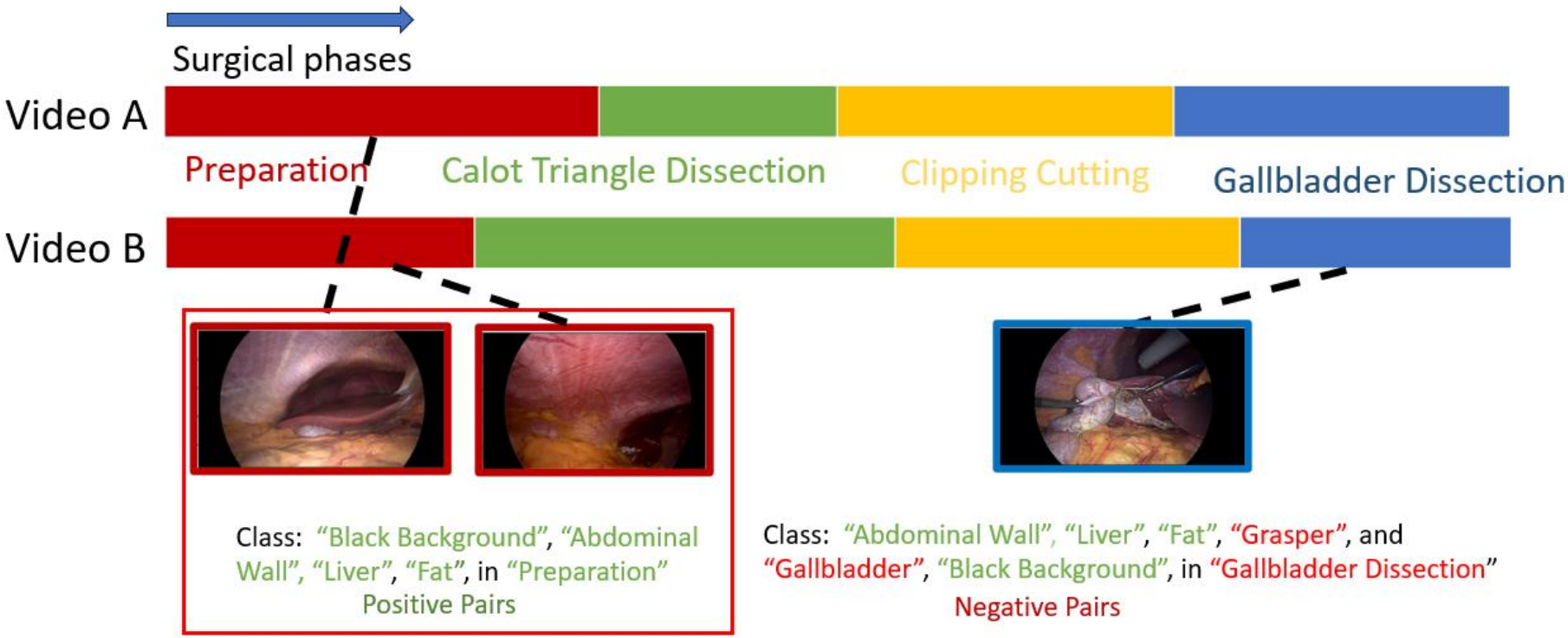
Decrease the similarity between features extracted from negative pairs



提案：Positive pairs definition

A novel positive pairs definition method in Laparoscopic image segmentation task

- Images containing the same category are similar.
- Images captured at the same surgical stage, with the same tools and organs, are set as positive pairs.



Feature Extraction in segmentation model

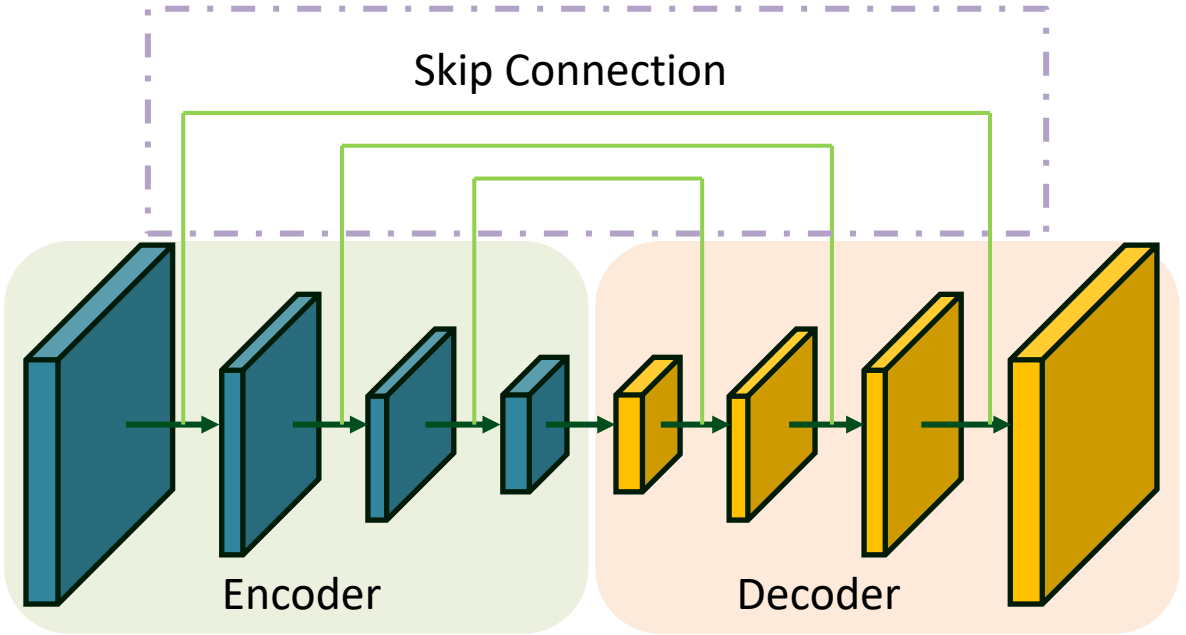
U-Net[1]

- A segmentation model for medical images
- Structure of Encoder and Decoder
- Skip connections are applied to retain information at each scale

Skip Connection

- Features extracted from each scale are transferred to the Decoder

Optimization of features in each scale is necessary



[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in MICCAI 2015, LNIP 9351, 234–241, Springer (2015).

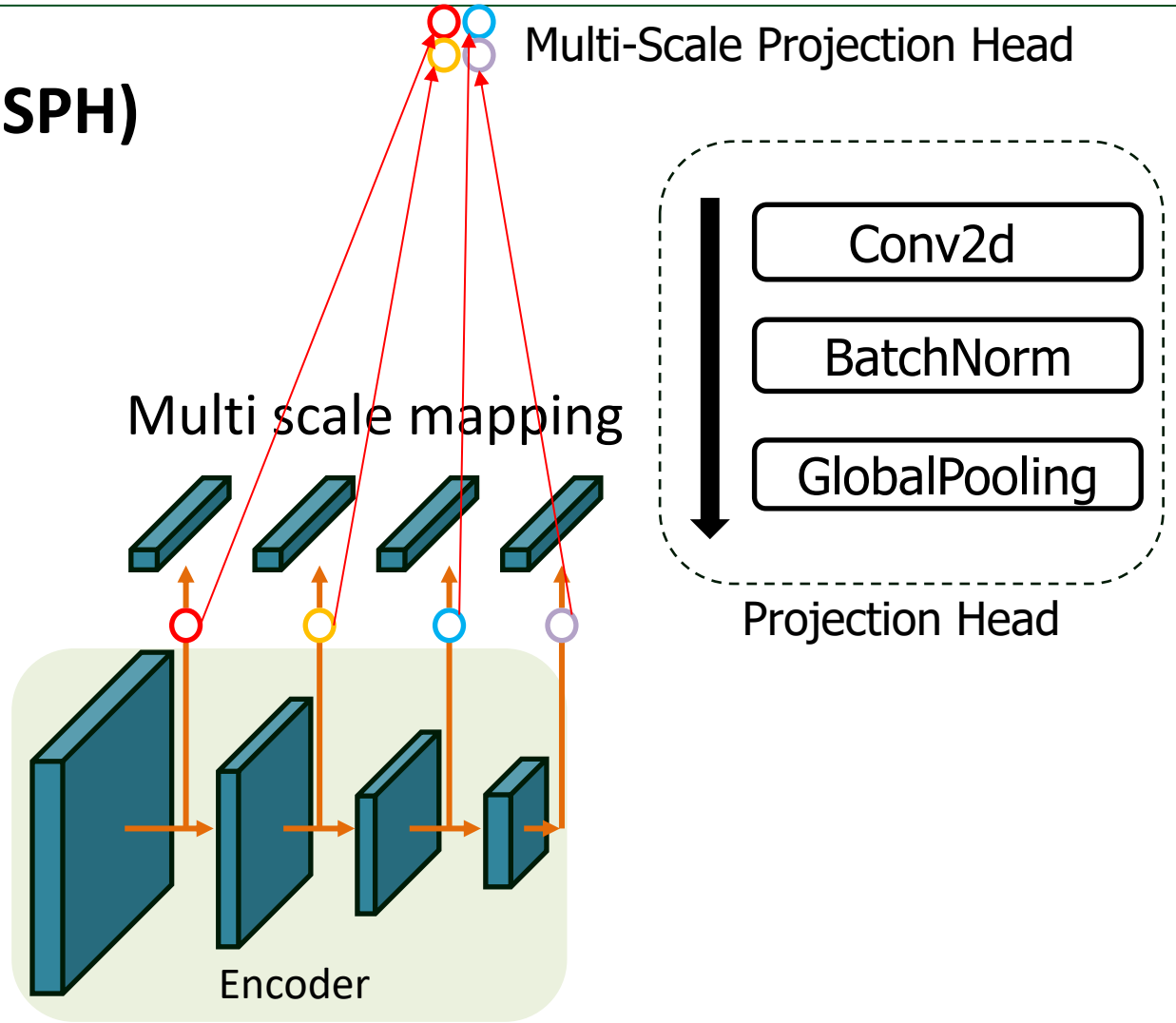
提案：Multi-Scale Projection Head (MSPH)

Projection Head [1,2,3]

- Maps high-dimensional features to a lower-dimensional space.
- Used to calculate contrastive loss, bringing features from positive pairs closer in the projection space.

Multi-Scale Projection Head (MSPH)

- In the proposed MSPH, features from each scale are mapped to a lower-dimensional space.
- Enables optimization of features from multiple scales.

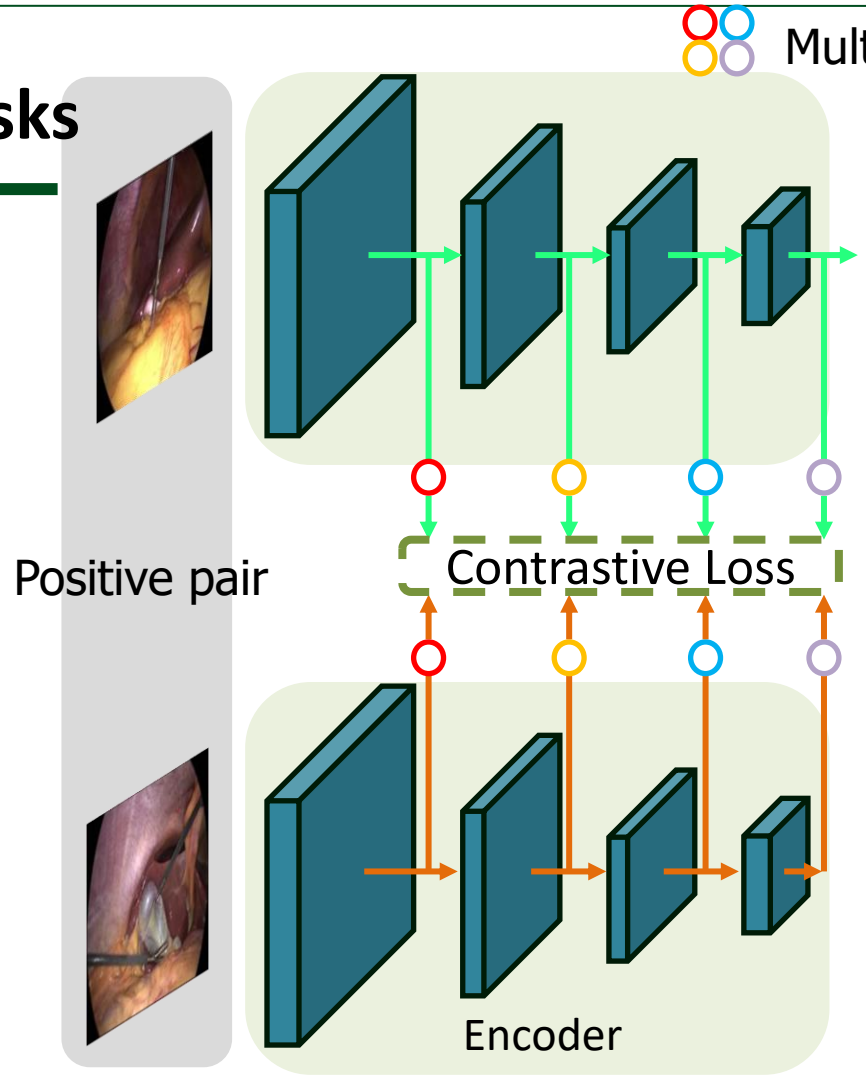


[1] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.
[2] He, Kaiming, et al. "Momentum contrast for unsupervised visual representation learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
[3] Caron, Mathilde, et al. "Emerging properties in self-supervised vision transformers." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

Optimize features using sub-tasks

Contrastive Learning

- Create image pairs using class-level labels.
- Input the image pairs into the model and extract features from each scale.
- Perform contrastive learning between features at each scale for positive pairs.



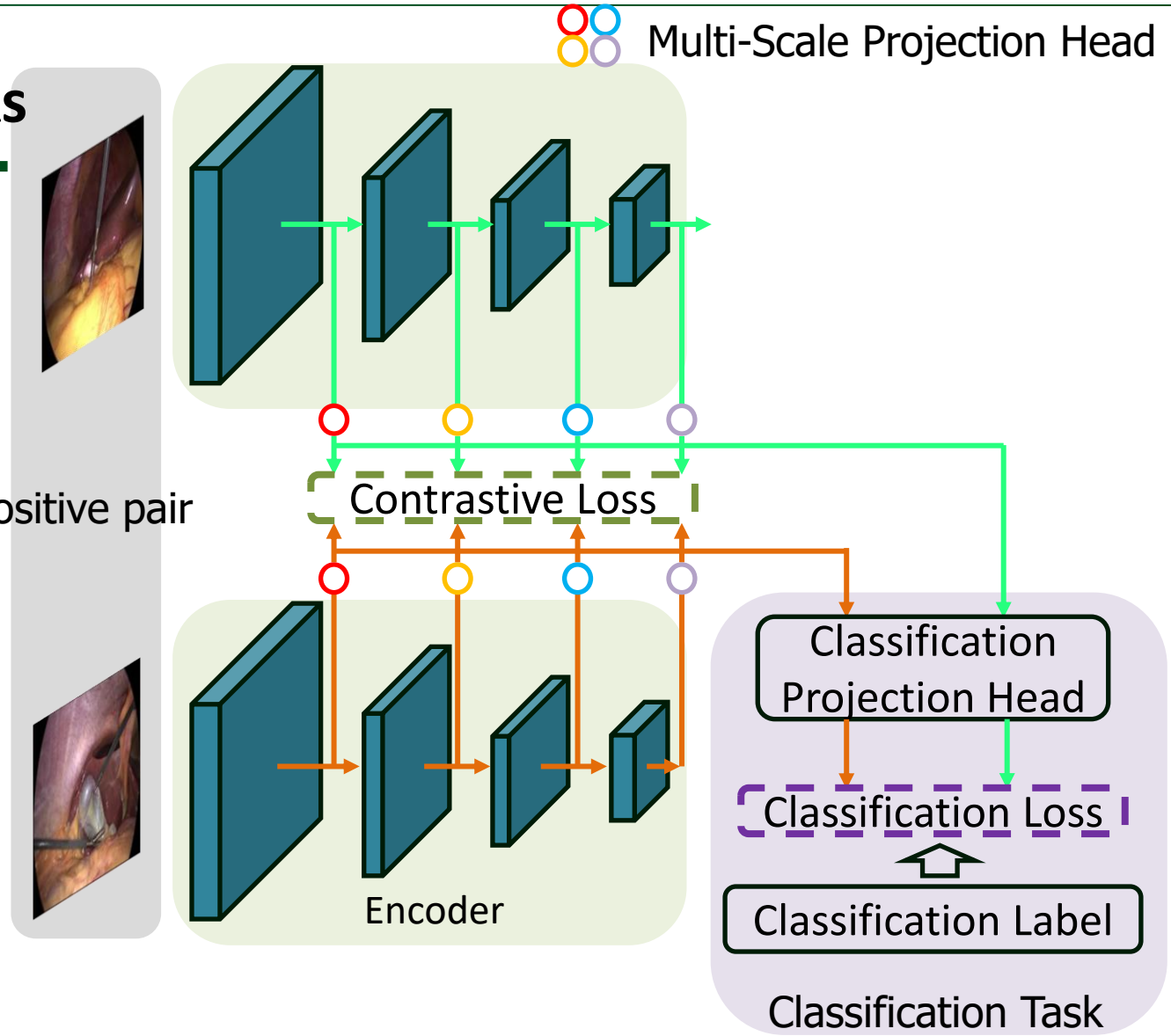
Optimize features using sub-tasks

Contrastive Learning

- Create image pairs using class-level labels.
- Input the image pairs into the model and extract features from each scale.
- Perform contrastive learning between features at each scale for positive pairs.

Classification Task

- Pass features through a classification projection head for classification learning.



提案：Proposed method

Segmentation task

Calculate segmentation loss on images with pixel-level annotations

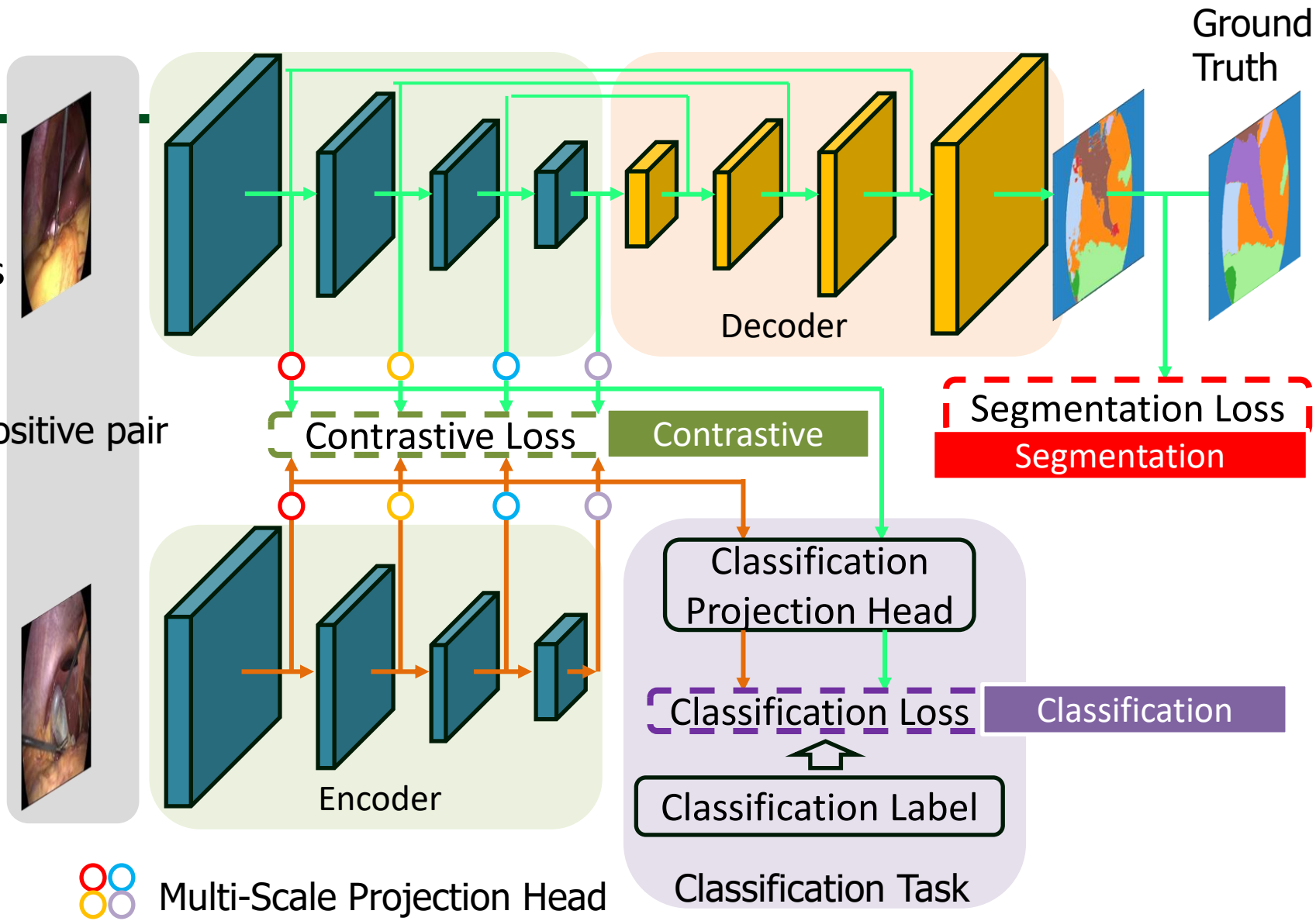
Classification task

Calculate classification loss on images without pixel-level annotations.

Contrastive learning task

Calculate contrastive loss on features extracted between positive pairs using MSPH.

The model's loss is set as the sum of segmentation loss, classification loss, and contrastive loss.



Loss function

Parameters

- $\mathbf{P}$: decoder output;
- $\mathbf{Z}$: the features extracted by MSPH from positive pairs;
- N: the count of positive pairs
- $\mathbf{c}$: the feature extracted by the classification projection head
- $\mathbf{g}$, $\mathbf{y}$: ground truth of classification and segmentation.

- L^{CE}: Cross Entropy;
- L^{Focal}: Focal Loss [1];
- L^{GDL} : Generalized Dice Loss [2];
- L^{CL} Contrastive Supervision Loss [3] ;
- α, β: hyperparameter

Segmentation task

$$L^{Seg} = L^{GDL}(\mathbf{P}, \mathbf{g}) + L^{Focal}(\mathbf{P}, \mathbf{g})$$

Classification task

$$L^{cls} = L^{CE}(\mathbf{c}, \mathbf{y})$$

Contrastive learning task

$$L^{DCL} = \sum_{i=1}^{N} L^{CL}(\mathbf{Z}_i)$$

Proposed method

$$L^{all} = L^{Seg} + \alpha L^{DCL} + \beta L^{cls}$$

[1] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988.
[2] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (pp. 240-248). Springer, Cham.
[3] Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: Supervised contrastive learning. Advances in Neural Information Processing Systems 33 (2020)

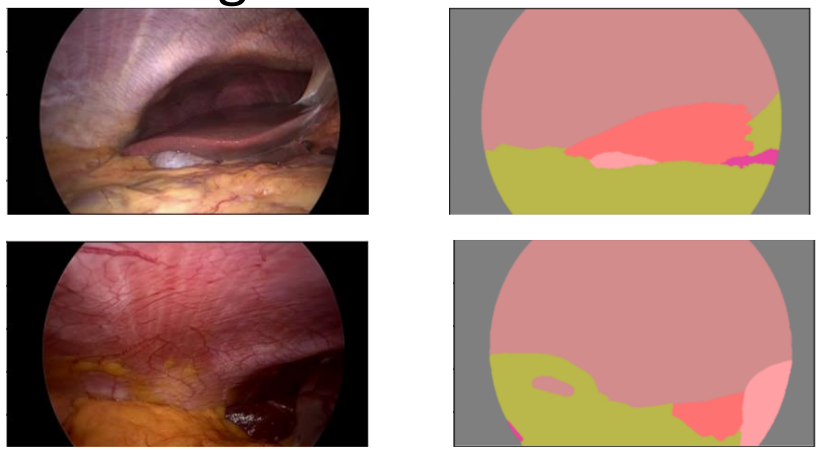
Experiment setting

[1] Hong, W-Y., et al., CholecSeg8k: A Semantic Segmentation Dataset for Laparoscopic Cholecystectomy Based on Cholec80, *arXiv preprint arXiv:2012.12453*, 2020.

- **Datasets**
  - **CholecSeg8k** [1]

    Laparoscopic images of cholecystectomy from 17 videos, totaling 8,080 frames.

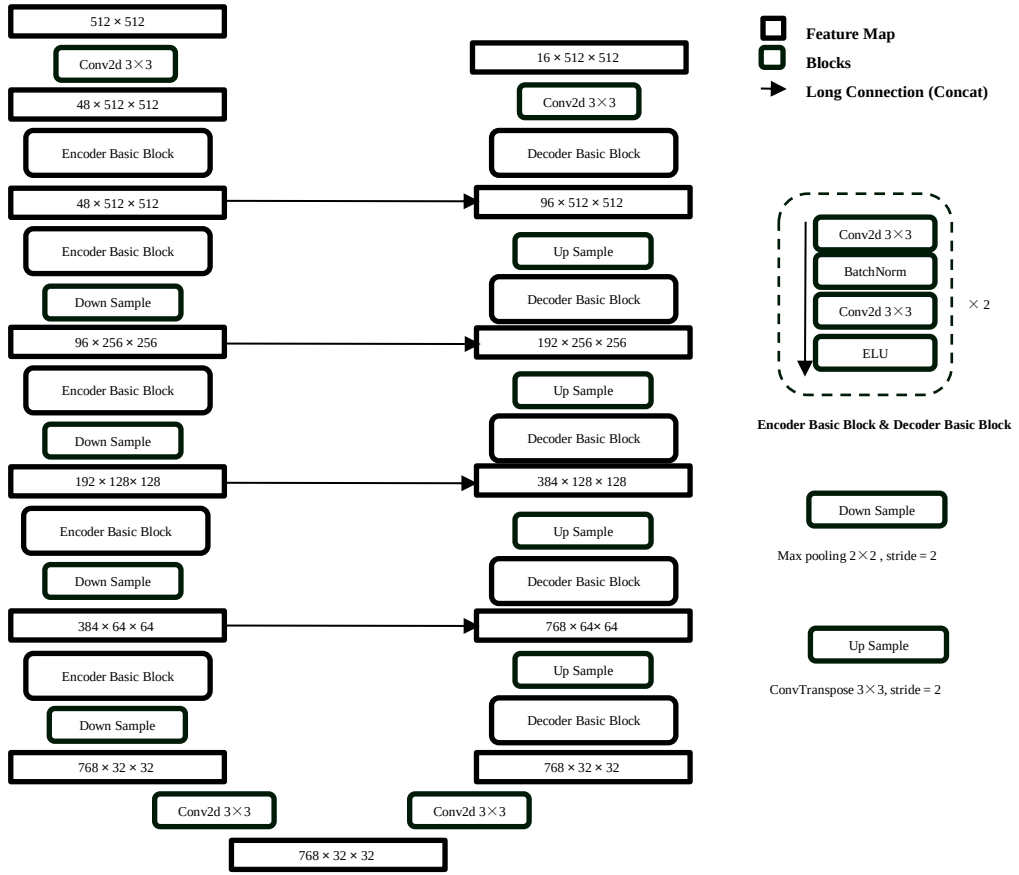
    8 categories



CholecSeg8k　　Ground-truth

- **Segmentation model**

5 layers U-Net



Experiment setting

臓器とツール（カテゴリー）の英和対照

- Background（背景）
- Abdominal Wall（腹壁）
- Liver （肝臓）
- Fat（脂肪）
- Grasper（把持鉗子）
- Connective Tissue（結合組織）
- L-hook Electrocautery（L字型電気メス）
- Gallbladder（胆嚢）

Experiment setting

- **Mainstream Approaches**

U-Net [1]	Baseline
SimCLR [2]	Positive pairs are formed by using an image and its augmented version
Ours DCL	Contrastive task + Segmentation task
Ours cls	Classification task+ Segmentation task
Ours DCL+cls	Contrastive task + classification task+ Segmentation task

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in MICCAI 2015, LNIP 9351, 234–241, Springer (2015).
[2] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.

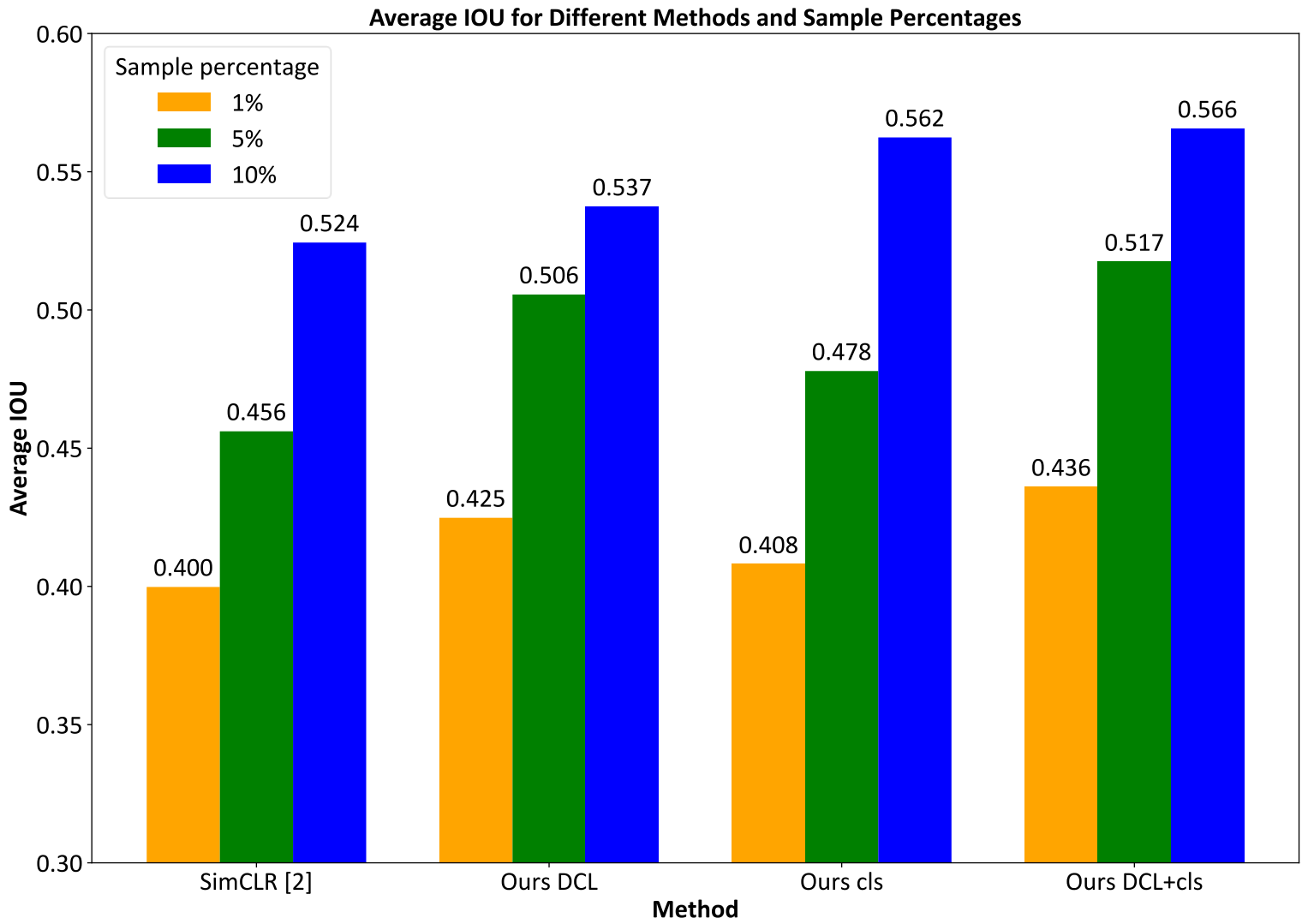
Experiment result (IOU)

The performance of the proposed methods outperforms related methods, especially in improving segmentation accuracy for small targets.

	Pixel labeled Samples	Background	Abdominal Wall	Liver	Fat	Grasper	Connective Tissue	L-hook Electrocautery	Gallbladder
U-Net [1]	1%	0.953 ± 0.004	0.519 ± 0.009	0.442 ± 0.036	0.761 ± 0.005	0.157 ± 0.006	0.007 ± 0.009	0.132 ± 0.005	0.241 ± 0.021
	5%	0.926 ± 0.034	0.664 ± 0.052	0.501 ± 0.045	0.796 ± 0.024	0.326 ± 0.112	0.265 ± 0.063	0.310 ± 0.123	0.360 ± 0.078
	10%	0.941 ± 0.008	0.612 ± 0.034	0.461 ± 0.023	0.767 ± 0.004	0.226 ± 0.042	0.091 ± 0.120	0.187 ± 0.132	0.335 ± 0.038
SimCLR [2]	1%	0.905 ± 0.012	0.482 ± 0.007	0.431 ± 0.023	0.742 ± 0.004	0.188 ± 0.030	0.040 ± 0.027	0.200 ± 0.057	0.210 ± 0.016
	5%	0.901 ± 0.008	0.558 ± 0.018	0.464 ± 0.023	0.762 ± 0.013	0.303 ± 0.014	0.009 ± 0.006	0.328 ± 0.017	0.324 ± 0.012
	10%	0.932 ± 0.002	0.633 ± 0.015	0.403 ± 0.008	0.780 ± 0.005	0.311 ± 0.030	0.342 ± 0.061	0.401 ± 0.008	0.393 ± 0.027
Ours DCL	1%	0.952 ± 0.005	0.542 ± 0.048	0.519 ± 0.017	0.755 ± 0.000	0.154 ± 0.020	0.096 ± 0.071	0.128 ± 0.148	0.252 ± 0.019
	5%	0.953 ± 0.003	0.606 ± 0.066	0.505 ± 0.038	0.797 ± 0.018	0.300 ± 0.092	0.246 ± 0.139	0.264 ± 0.188	0.374 ± 0.111
	10%	0.936 ± 0.004	0.631 ± 0.024	0.484 ± 0.007	0.792 ± 0.008	0.310 ± 0.059	0.305 ± 0.098	**0.420 ± 0.059**	0.421 ± 0.033
Ours cls	1%	0.955 ± 0.000	0.534 ± 0.034	0.494 ± 0.019	0.773 ± 0.016	0.192 ± 0.004	0.053 ± 0.047	0.001 ± 0.001	0.262 ± 0.028
	5%	0.925 ± 0.013	0.576 ± 0.057	0.444 ± 0.014	0.775 ± 0.022	0.245 ± 0.089	0.182 ± 0.079	0.356 ± 0.092	0.320 ± 0.076
	10%	0.945 ± 0.005	**0.657 ± 0.025**	0.485 ± 0.016	**0.796 ± 0.019**	0.341 ± 0.031	0.400 ± 0.133	0.417 ± 0.064	**0.458 ± 0.023**
Ours DCL+cls	1%	0.952 ± 0.006	0.614 ± 0.034	0.498 ± 0.019	0.762 ± 0.005	0.178 ± 0.005	0.060 ± 0.035	0.137 ± 0.019	0.287 ± 0.017
	5%	0.944 ± 0.012	0.624 ± 0.023	0.438 ± 0.035	0.798 ± 0.004	0.296 ± 0.070	0.336 ± 0.031	0.338 ± 0.046	0.365 ± 0.027
	10%	**0.948 ± 0.013**	0.638 ± 0.019	**0.485 ± 0.008**	0.791 ± 0.009	**0.407 ± 0.016**	**0.404 ± 0.042**	0.405 ± 0.032	0.447 ± 0.030

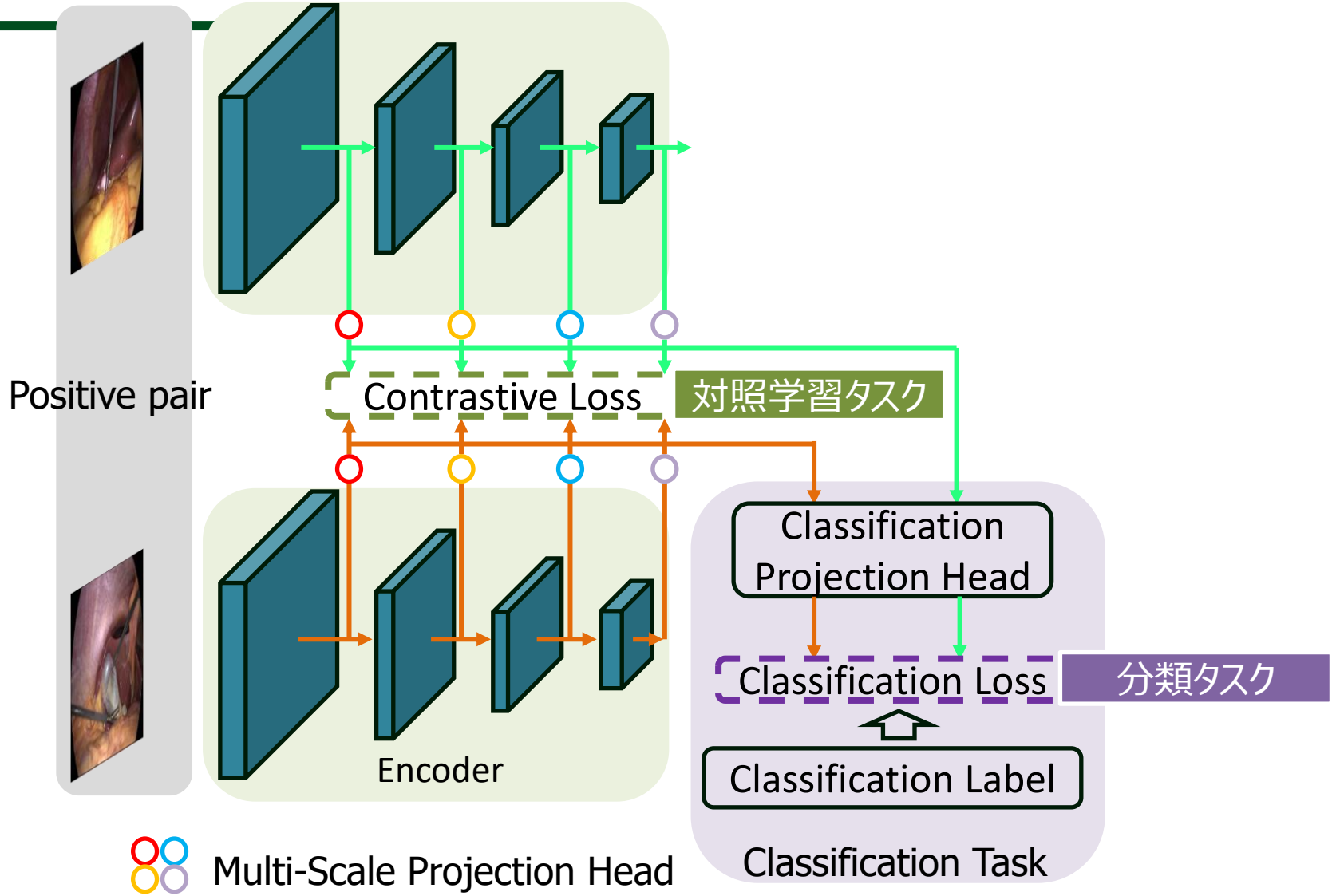
Experiment result (IOU)

The proposed method demonstrates high accuracy with few pixel-level labeled data (1% or 5%), confirming its effectiveness in situations with limited data.



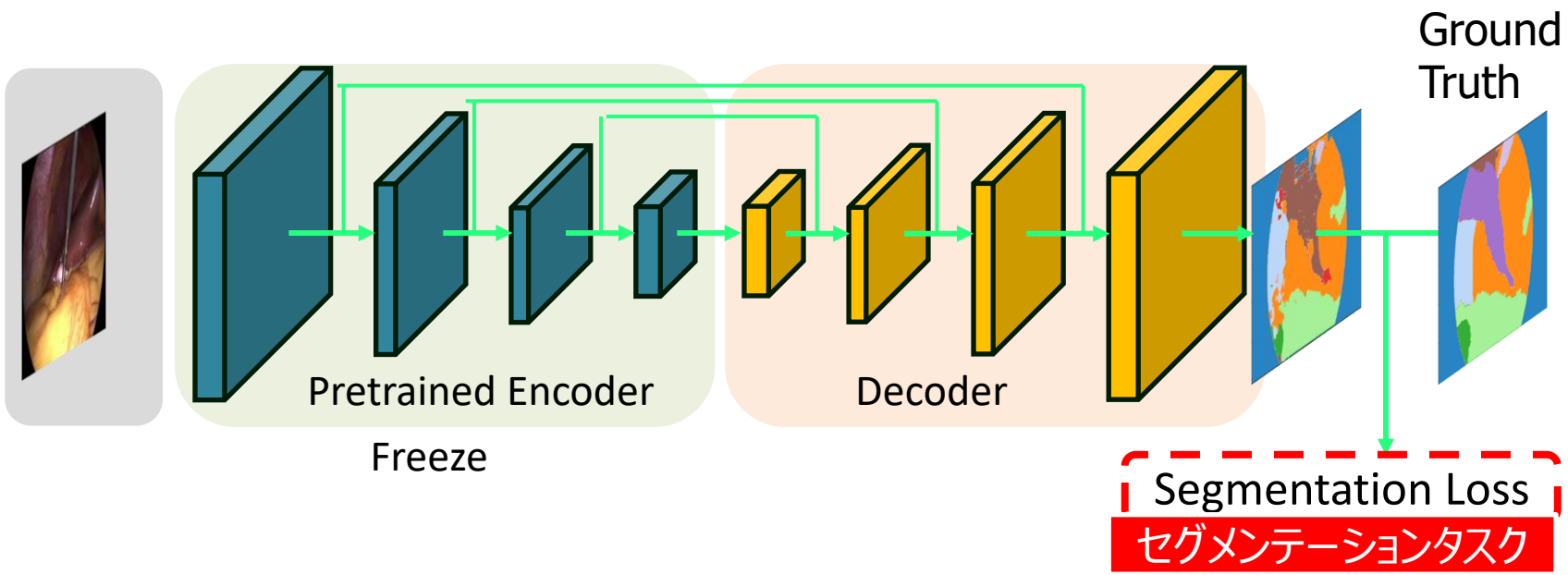
2-Step Training

Step 1 Pretrain the encoder by sub tasks



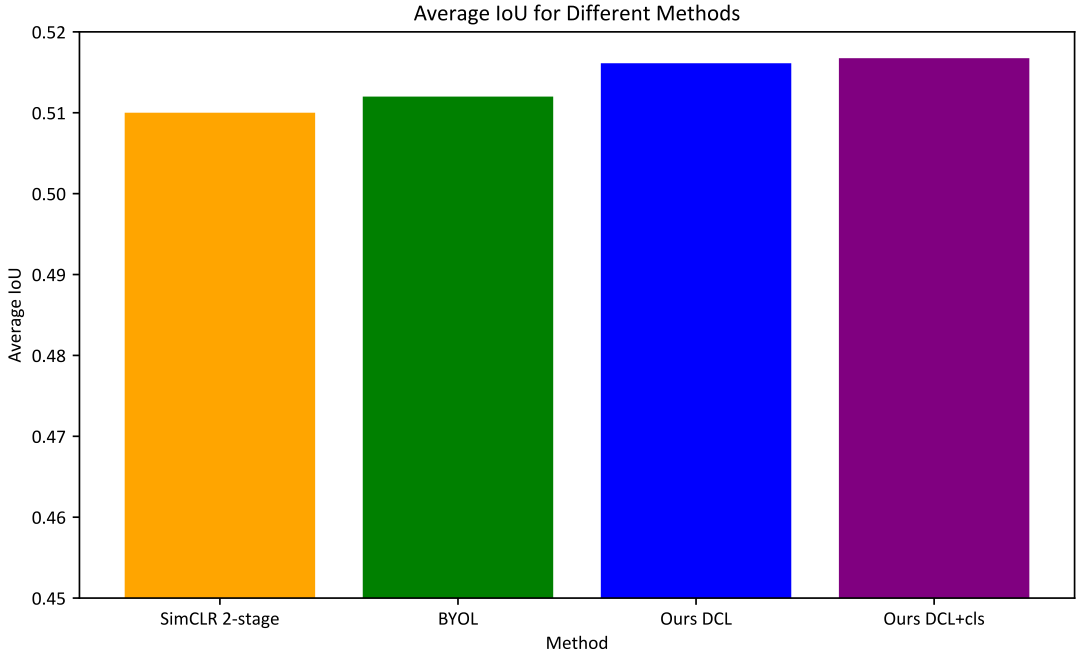
2-Step Training

Step 2 Freeze the Pretrained encoder and train the Decoder by segmentation task



2-Step Training Result

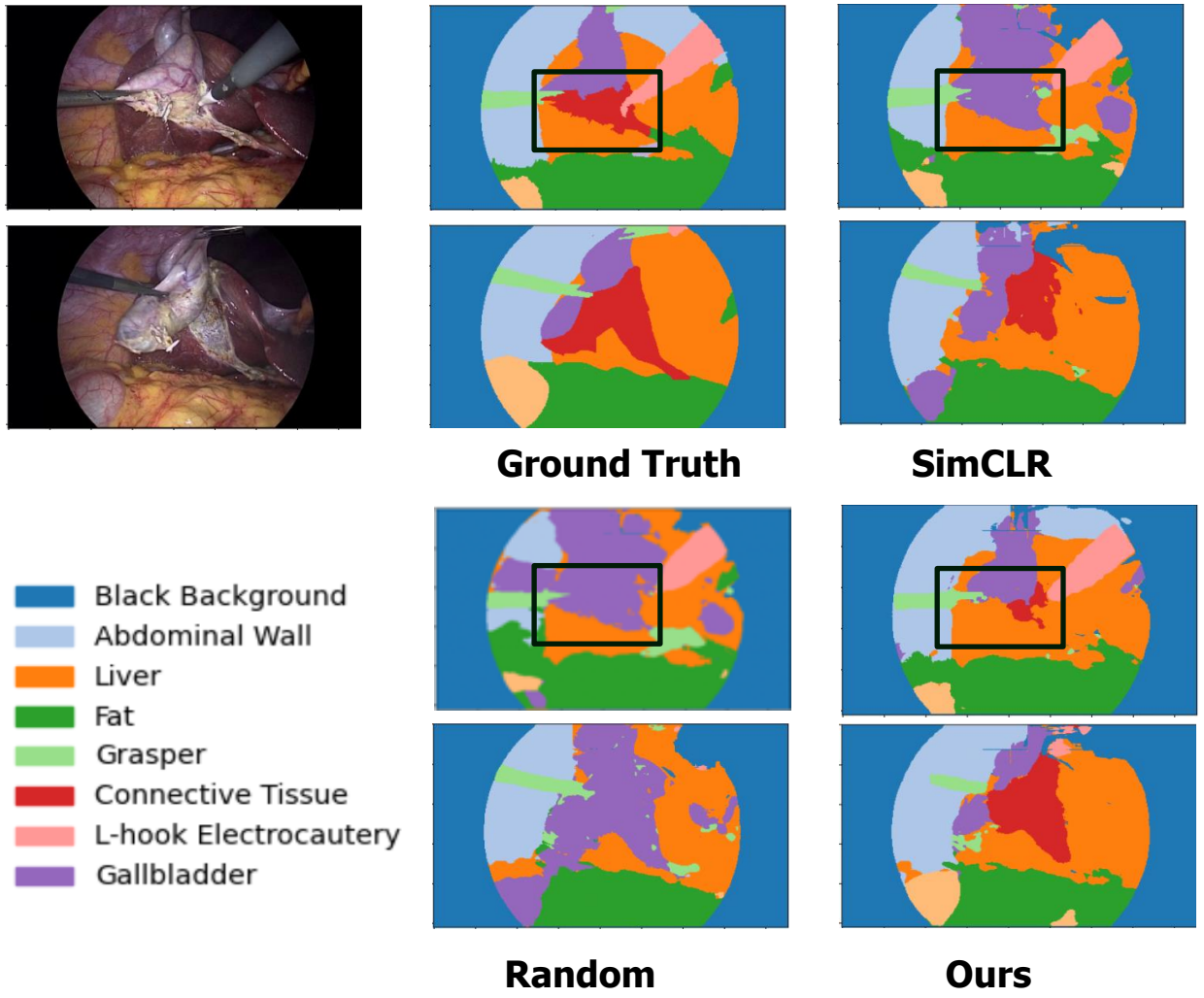
The performance of the proposed methods outperforms related 2-step methods, especially in improving segmentation accuracy for smaller targets.



	Background	Abdominal Wall	Liver	Fat	Grasper	Connective Tissue	L-hook Electrocautery	Gallbladder
SimCLR [1] 2-stage	0.921 ± 0.012	0.653 ± 0.005	**0.528 ± 0.021**	0.773 ± 0.013	0.250 ± 0.020	0.206 ± 0.038	0.360 ± 0.021	0.389 ± 0.012
BYOL [2]	0.916 ± 0.015	0.651 ± 0.027	0.520 ± 0.025	0.776 ± 0.011	0.250 ± 0.014	0.205 ± 0.037	0.359 ± 0.040	**0.419 ± 0.018**
Ours DCL	**0.933 ± 0.006**	0.657 ± 0.034	0.522 ± 0.010	0.779 ± 0.001	0.252 ± 0.015	**0.219 ± 0.037**	0.354 ± 0.026	0.413 ± 0.050
Ours DCL+cls	0.924 ± 0.005	**0.659 ± 0.022**	0.519 ± 0.015	**0.779 ± 0.014**	**0.266 ± 0.008**	0.212 ± 0.026	**0.368 ± 0.032**	0.407 ± 0.024

[1] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.
[2] Grill, J. B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., Piot, B., Kavukcuoglu, K., Munos, R., & Valko, M. (2020). Bootstrap Your Own Latent: A New Approach to Self-Supervised Learning. In Advances in Neural Information Processing Systems (NeurIPS) 2020.

Discussion



mDice using 1% and 10% data

Method	1%	10%
U-Net	0.389	0.453
SimCLR	0.400	0.513
Proposed	0.424	0.565

Ablation Study

Cls	MSPH	1%	10%
X	X	0.389	0.453
X	O	0.425	0.537
O	X	0.407	0.532
O	O	0.424	0.565

Discussion

- The method we propose outperforms related works in the field.
- In cases involving categories that are closely related or similar, our method demonstrates superior capability in distinguishing and identifying each category accurately.
- The ablation studies validate the effectiveness of each introduced improvement.

Discussion

- What issues are we dealing with in this study?
  - ☹ ➢ Creating pixel-level annotation is expensive
  - ☹ ➢ Segmentation performance with limited annotations is not good

Discussion

- What issues are we dealing with in this study?
  - ☹ Creating pixel-level annotation is expensive
  - ☹ Segmentation performance with limited annotations is not good
- How did we reduce the annotation cost?
  - Employed a sub-task to enhance model training with low-cost annotations (category-wise annotation)
  - Proposed a novel positive pair definition method for contrastive learning in Laparoscopic image segmentation task.
  - Proposed a novel MSPH for multi-scale feature optimizing.

Discussion

- What issues are we dealing with in this study?
  - ☹ Creating pixel-level annotation is expensive
  - ☹ Segmentation performance with limited annotations is not good
- How did we reduce the annotation cost?
  - Employed a sub-task to enhance model training with low-cost annotations (category-wise annotation)
  - Proposed a novel positive pair definition method for contrastive learning in Laparoscopic image segmentation task.
  - Proposed a novel MSPH for multi-scale feature optimizing.
- What are the limitations of our method?
  - The results have not been evaluated by doctors, making it impossible to estimate their actual significance in clinical practice.
  - Category–level annotation still needs annotation costs, even it is low.

In the next topic, we will discuss another method to improve segmentation performance **without category-level annotations**, cased lower annotation cost

Topic 2

Double-Mix Pseudo-Label Framework: Enhancing Semi-Supervised Segmentation on Category-Imbalanced CT Volumes

Zhang, Luyang, et al. "Double-mix pseudo-label framework: enhancing semi-supervised segmentation on category-imbalanced CT volumes." *International Journal of Computer Assisted Radiology and Surgery*, doi:10.1007/s11548-024-03281-1

Background

Automatic Abdominal Organ Segmentation from CT Images

- Enables accurate diagnosis using high-quality CT images.
- Increasing the number of CT images places a greater burden on radiologists.
- Developing a segmentation system using deep learning is essential to reduce this burden.

Semi-supervised Multi-organ Segmentation

- It is challenging to obtain a large amount of annotated data.
- Semi-supervised learning using unlabeled image data is effective.
- Semi-supervised learning methods such as Mean Teacher [1], Model Mix [2], and CPS [3] have been proposed.

[1] Tarvainen, Antti, and Harri Valpola. "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results." Advances in neural information processing systems 30 (2017).
[2] Zhang, Ke, and Vishal M. Patel. "Modelmix: A new model-mixup strategy to minimize vicinal risk across tasks for few-scribble based cardiac segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2024.
[3] Chen, Xiaokang, et al. "Semi-supervised semantic segmentation with cross pseudo supervision." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.

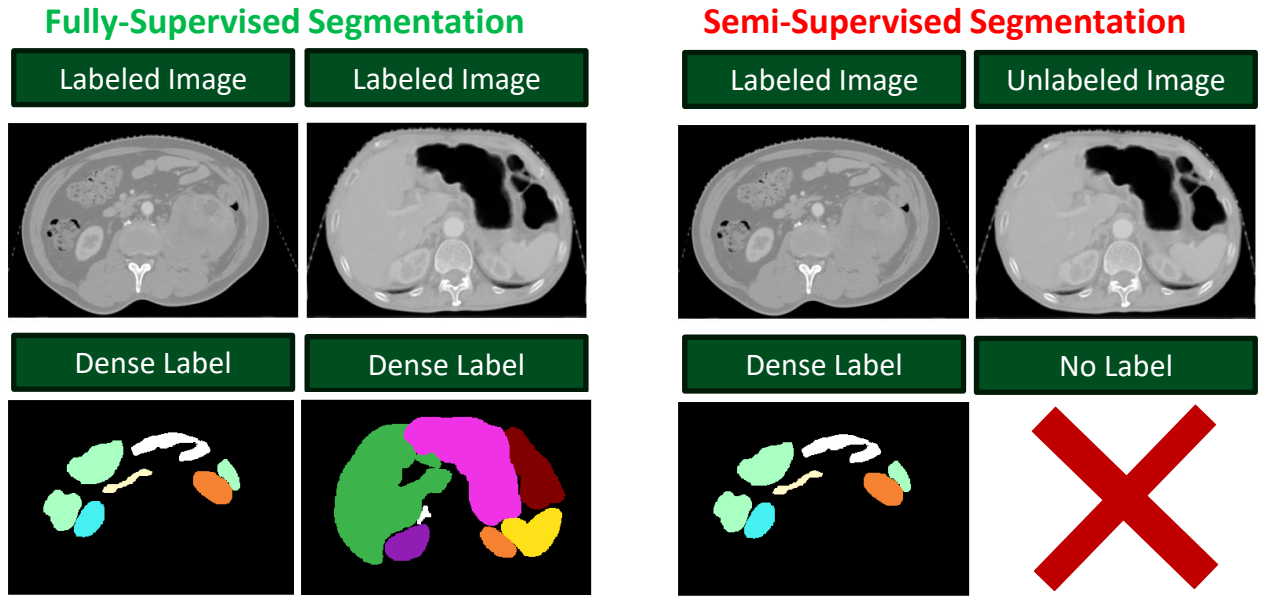
Low-cost annotation

Solution:

Utilize data without annotations.

Motivation:

Train a segmentation model using a small amount of annotated data and a large amount of unannotated data.



Annotation cost High ☹ **Low** ☺

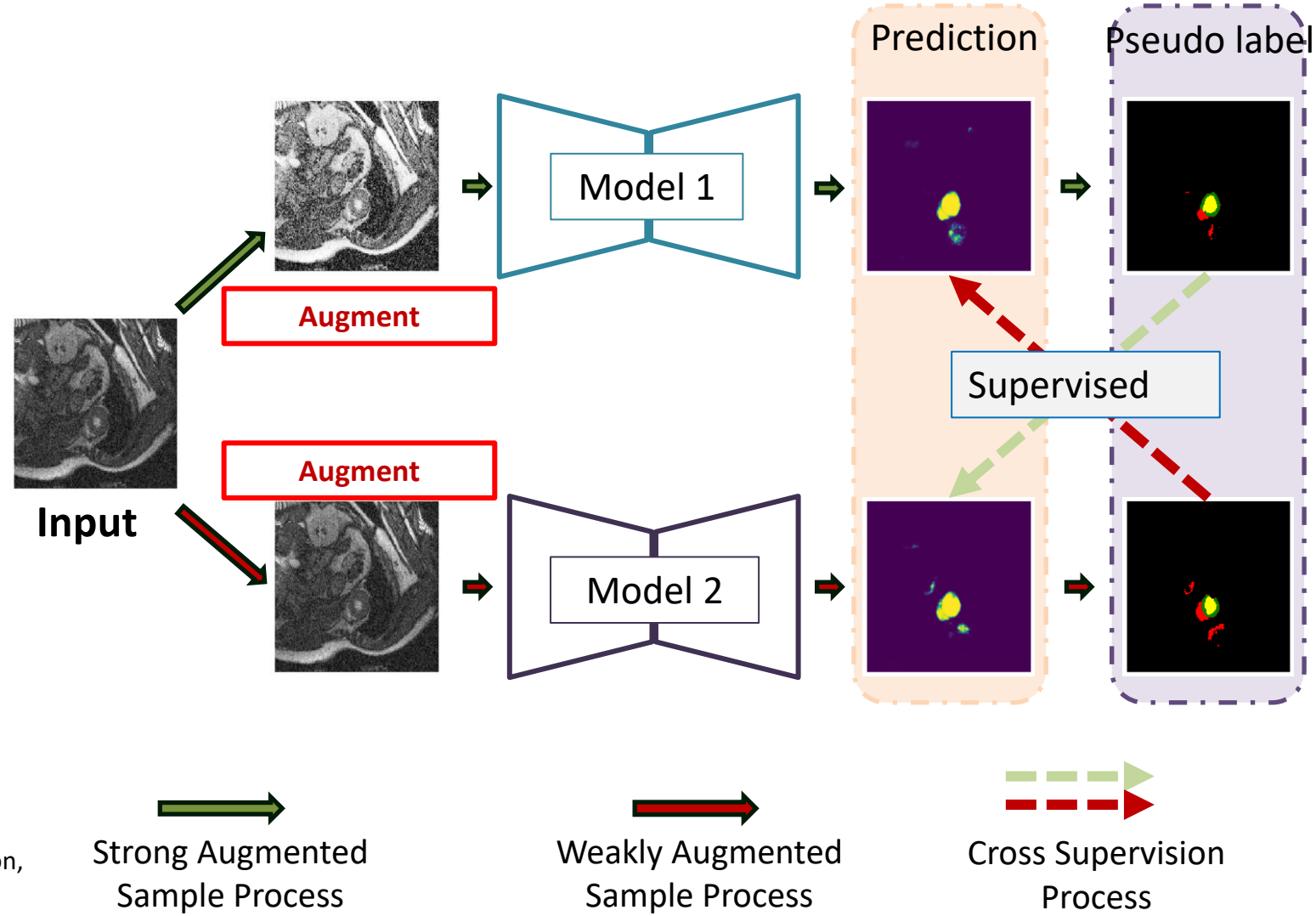
Annotation information **Complete** ☺ Incomplete ☹

Reduce annotation costs.

Utilize **Cross Pseudo Supervision (CPS)** with pseudo-labels from different models.

Cross Pseudo Supervision (CPS)

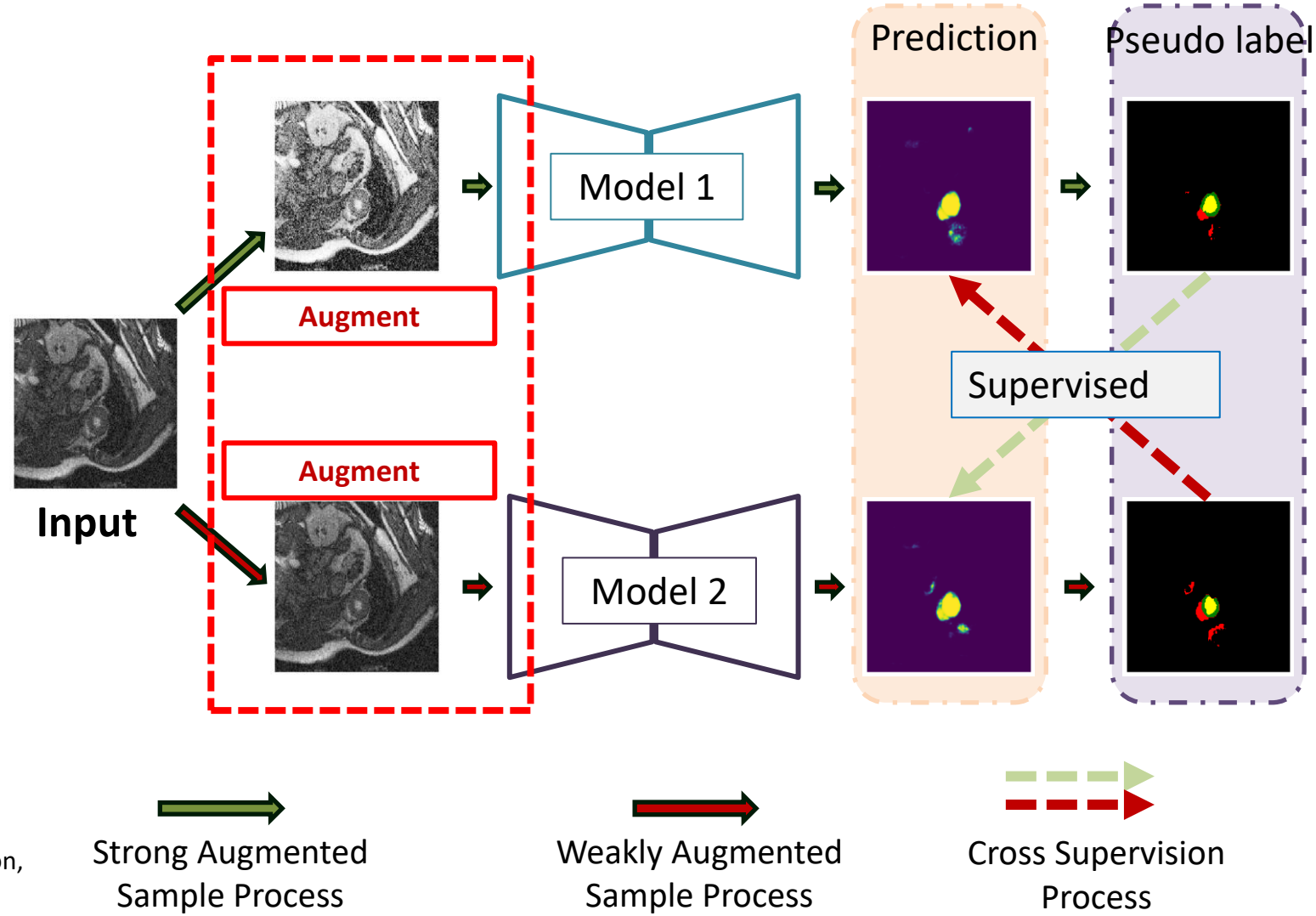
- Two different networks make predictions on different augmentations of the same image to generate pseudo-labels.
- One pseudo-label serves as training data for the other model.
- Cross Pseudo allows use in cases without labels, and accuracy improves as heterogeneity between models increases [1].



[1] Krogh A, and Jesper V. Neural network ensembles, cross validation, and active learning. NIPS 1994:7

Cross Pseudo Supervision (CPS)

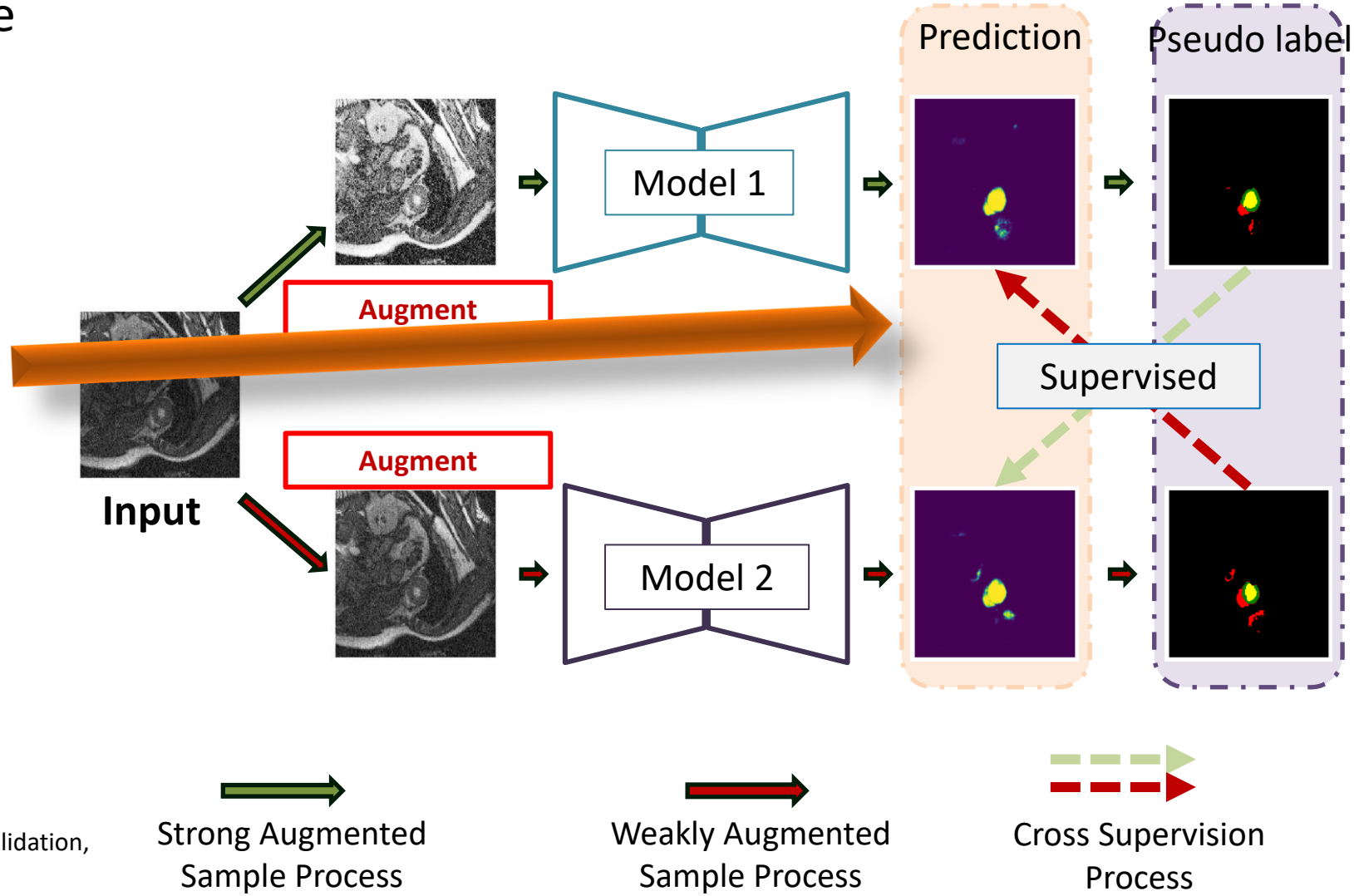
- Two different networks make predictions on different augmentations of the same image to generate **pseudo-labels.**
- One pseudo-label serves as training data for the other model.
- Cross Pseudo allows use in cases without labels, and accuracy improves as heterogeneity between models increases [1].



[1] Krogh A, and Jesper V. Neural network ensembles, cross validation, and active learning. NIPS 1994:7

Cross Pseudo Supervision (CPS)

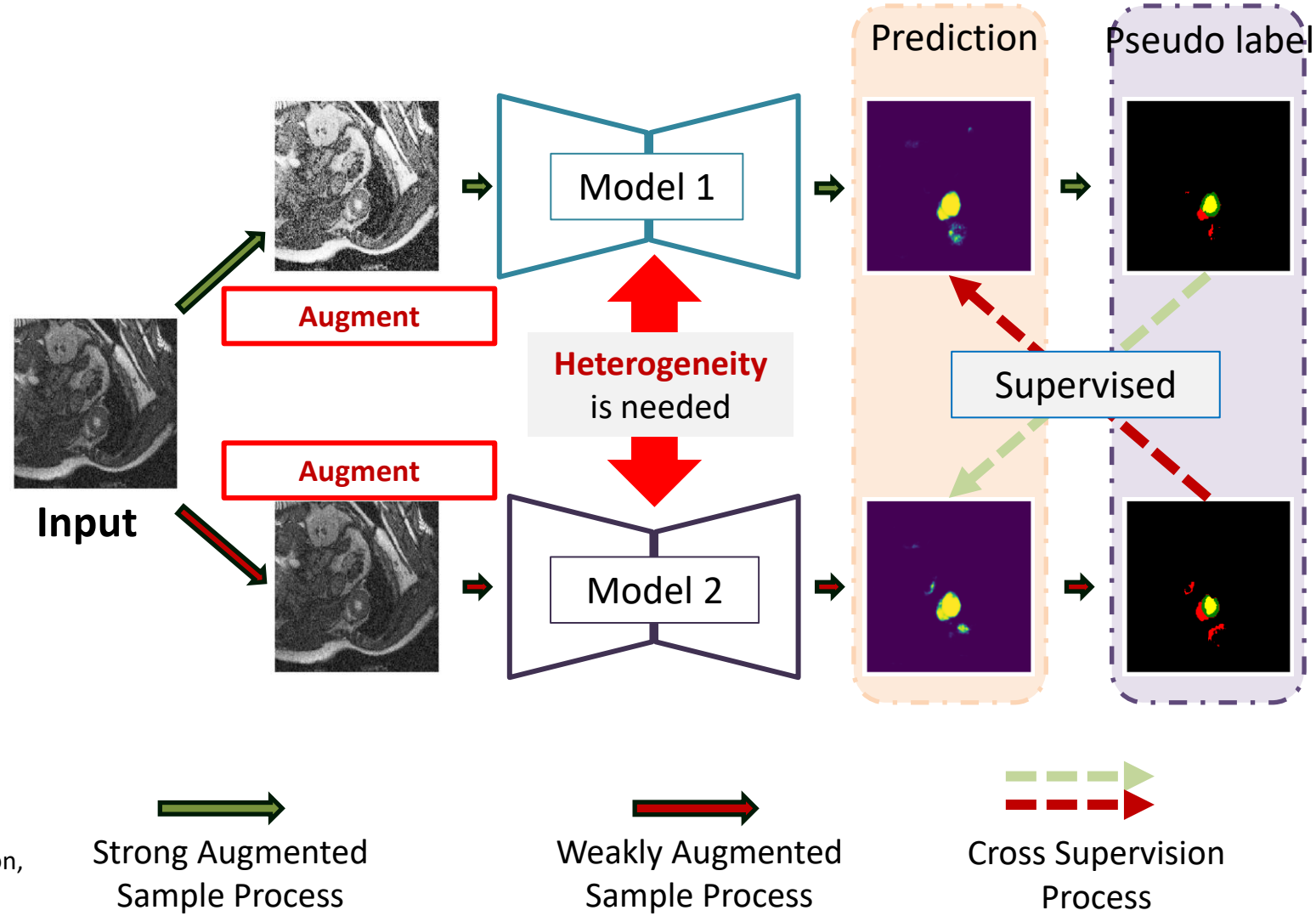
- Two different networks make predictions on different augmentations of the same image to generate **pseudo-labels.**
- One pseudo-label serves as training data for the other model.
- Cross Pseudo allows use in cases without labels, and accuracy improves as heterogeneity between models increases [1].



[1] Krogh A, and Jesper V. Neural network ensembles, cross validation, and active learning. NIPS 1994:7

Cross Pseudo Supervision (CPS)

- Two different networks make predictions on different augmentations of the same image to generate **pseudo-labels.**
- One pseudo-label serves as training data for the other model.
- Cross Pseudo allows use in cases without labels, and accuracy improves as heterogeneity (異質性) between models increases [1].



[1] Krogh A, and Jesper V. Neural network ensembles, cross validation, and active learning. NIPS 1994:7

Heterogeneity (異質性)

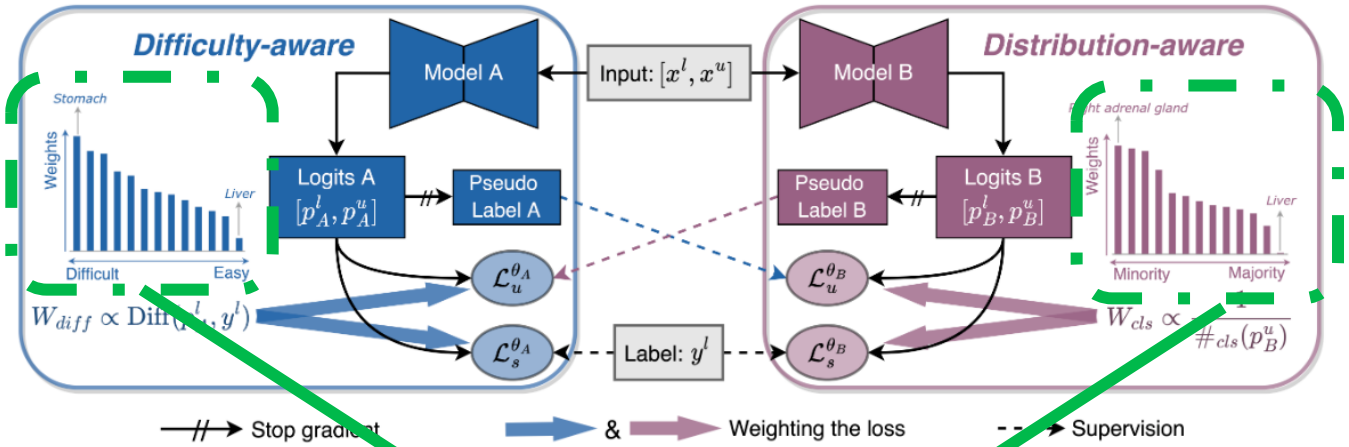
Heterogeneity

The diversity of features extracted by different models from the images [1].

Method to enhance heterogeneity

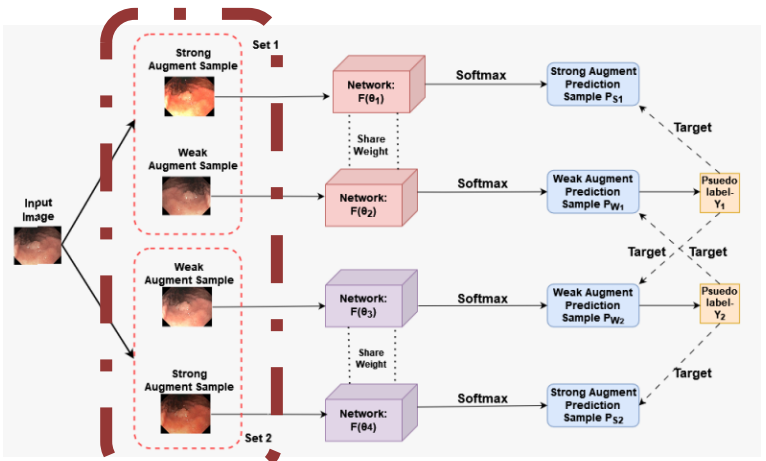
- Train different models with category-specific weights.
- Amplify the differences between input images using different augmentation methods.

[1] Wang, Haonan, and Xiaomeng Li. "DHC: Dual-debiased heterogeneous co-training framework for class-imbalanced semi-supervised medical image segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2023.
[2] Chen, Yifei, et al. "Semi-supervised Medical Image Segmentation Method Based on Cross-pseudo Labeling Leveraging Strong and Weak Data Augmentation Strategies." *arXiv preprint arXiv:2402.11273* (2024).



DHC [2]

Train different models with category-specific weights.



DFCPS [3]

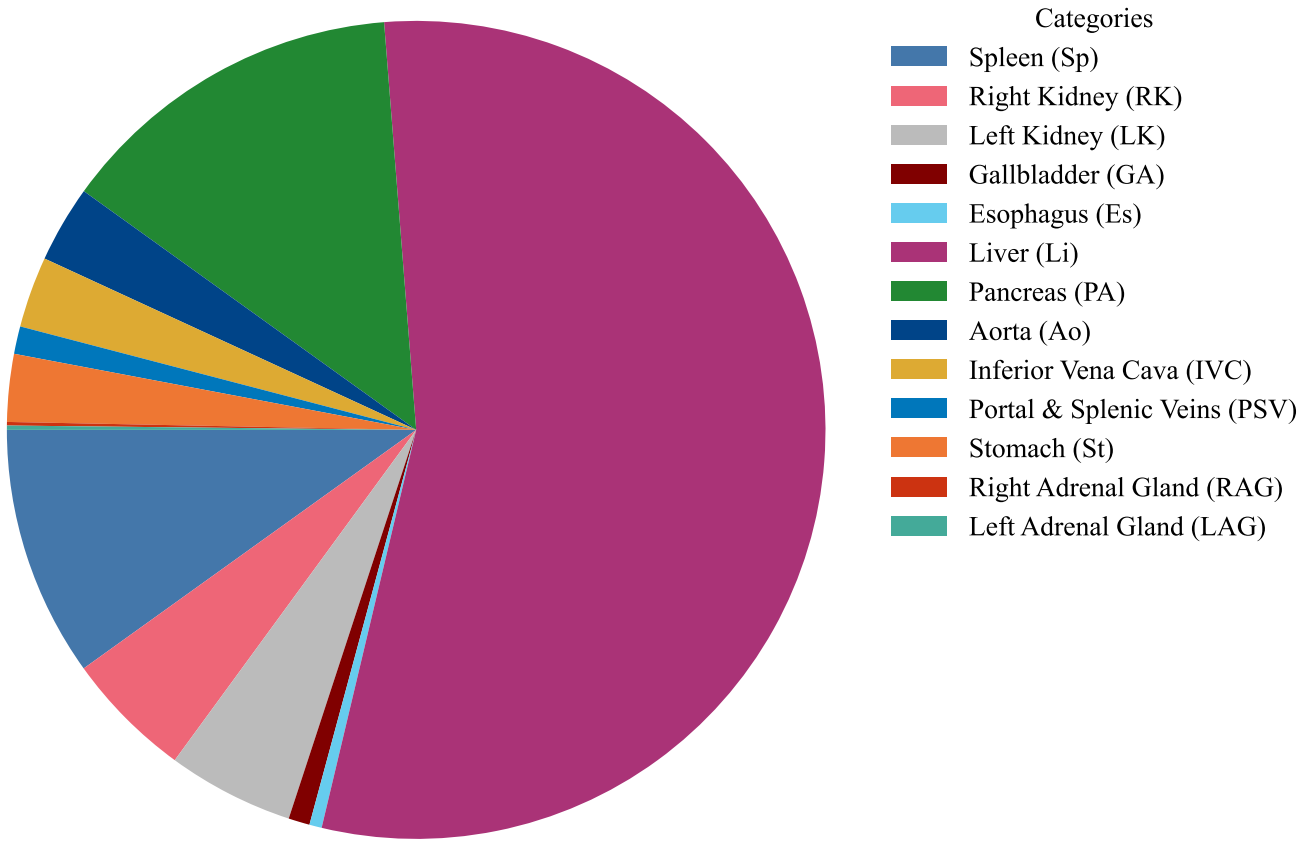
Using different data augmentations

Train different models with category-specific weights

Category imbalance

- The imbalance in **category-wise voxel counts**

The category-wise voxel-count on Organ-Segmentation dataset BTCV [1]



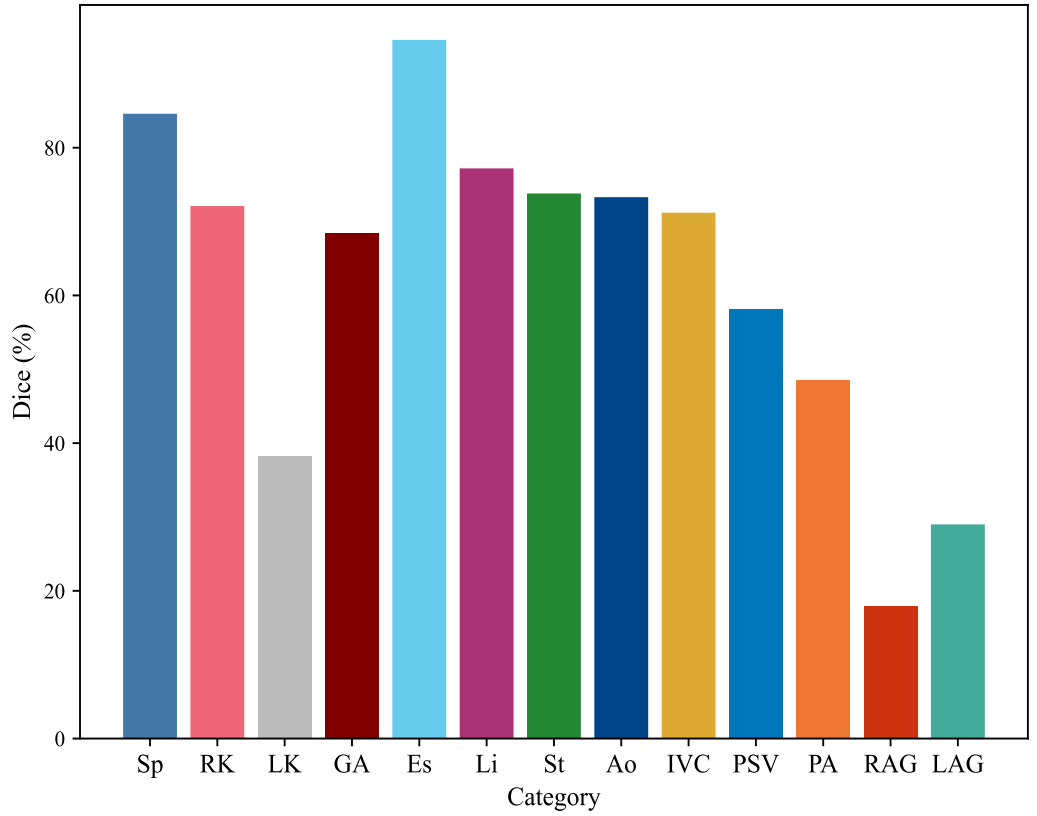
[1] Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: 2015 MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge (2015)

Train different models with category-specific weights

Category imbalance

- The imbalance in **category-wise voxel counts**
- The imbalance in **category-wise difficulty**

The category-wise Dice of full-supervision on organ segmentation dataset BTCV using U-Net[1]



[1] Wang, Haonan, and Xiaomeng Li. "DHC: Dual-debiased heterogeneous co-training framework for class-imbalanced semi-supervised medical image segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2023.

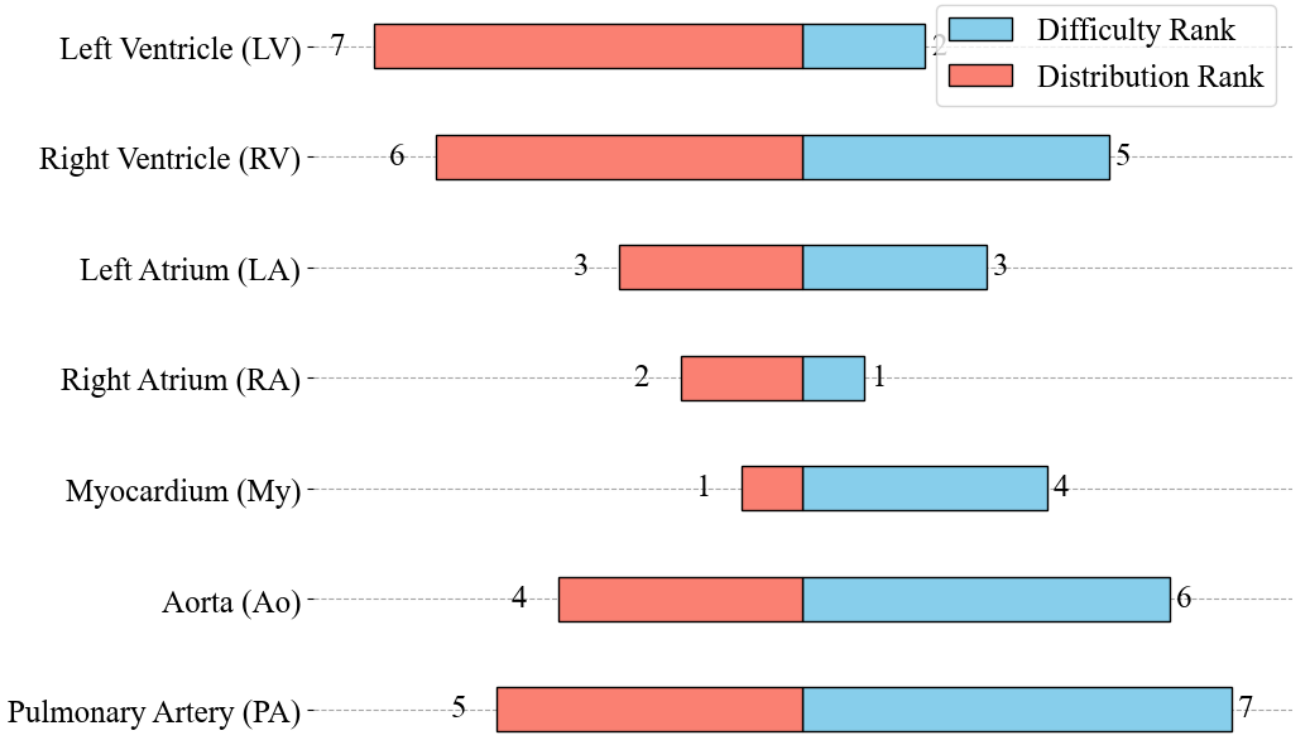
Train different models with category-specific weights

Category imbalance

- **Category-wise voxel counts** and **Category-wise difficulty** are **different**

- **Using Category-wise voxel counts based weight** and **category-wise difficulty based weight** to train different models

Comparison of category-wise difficulty ranking (inverse) and Voxel count ranking in CHD [1].



Xu, X., Wang, T., Shi, Y., Yuan, H., Jia, Q., Huang, M., Zhuang, J.: Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching. In: MICCAI, Proceedings, Part II, LNIP, vol. 11765, pp. 477–485 (2019). Springer

Problem we aiming to solve

The lack of the data → Dual-Network framework (CPS)

Dual-Network framework (CPS) → Category-wise imbalance

Dual-Network framework (CPS) → The lack of 異質性

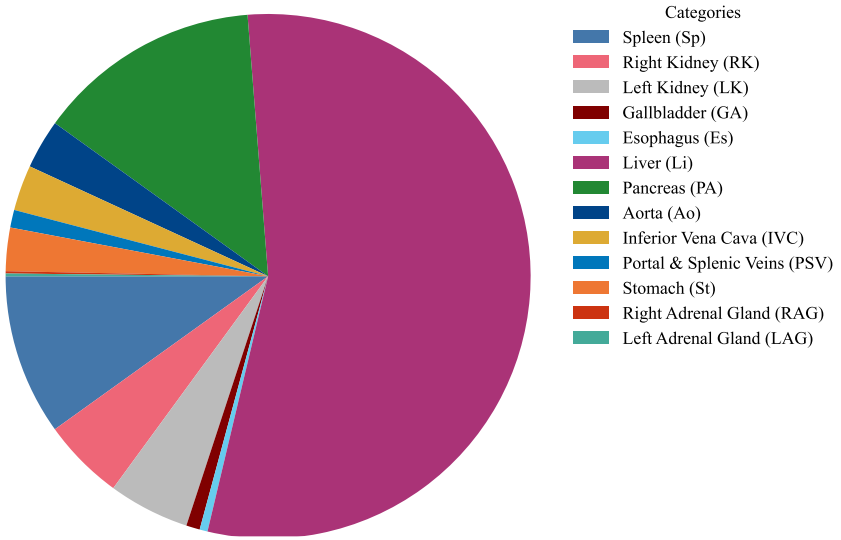
Category-wise imbalance → Train different models with category-specific weights

The lack of 異質性 → Applying different augmentation for the same image as the input

Train different models with category-specific weights

Category-wise voxel counts based weight (DisW) W^{dis}[1,2]

Category-wise voxel counts in iteration t



Counts: $A_t = \{a_{t,c} \mid c = 1,2, \dots K\}$

K: the total number of categories

Inverse voxel ratios of category c in iteration t

$$r_{t,c} = \frac{\max A_t}{a_{t,c}}$$

$a_{t,c}$ ↑ $r_{t,c}$ ↓

Voxel counts based weight of category c in iteration t

$$w_{t,c}^{dis} = \frac{\log(r_{t,c})}{max_{\rho \in \{1,2,\dots,K\}} \log(r_{t,\rho})}$$

$a_{t,c}$ ↑ $r_{t,c}$ ↓ $w_{t,c}^{dis}$ ↓

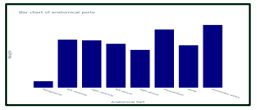
[1] Wang, Haonan, and Xiaomeng Li. "DHC: Dual-debiased heterogeneous co-training framework for class-imbalanced semi-supervised medical image segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2023.
[2] Chen, H., Fan, Y., Wang, Y., Wang, J., Schiele, B., Xie, X., Savvides, M., Raj, B.: An embarrassingly simple baseline for imbalanced semi-supervised learning. arXiv preprint arXiv:2211.11086, 2023.

Train different models with category-specific weights

Category-wise difficulty based weight (DifW) $w_{t,c}^{dif}$ [1]

Category-wise Segmentation difficulty in iteration t

- Dice score [2]



Use the **Dice score** for each category in each iteration as the difficulty evaluation criterion.

- Well-learned speed [3]

Use the **rate of change in the Dice score (Population Stability Index, PSI)** for each category as the difficulty evaluation criterion.

[1] Wang, Haonan, and Xiaomeng Li. "DHC: Dual-debiased heterogeneous co-training framework for class-imbalanced semi-supervised medical image segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2023.
[2]Sudre, Carole H., et al. "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations." *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*. Springer International Publishing, 2017.
[3] Yurdakul, B.: Statistical properties of population stability index (psi). PhD thesis, Western Michigan University, 2018

Train different models with category-specific weights

Category-wise difficulty based weight (DifW) $w_{t,c}^{dif}$

Learning Speed

Considering the Dice changes between iterations *o* and *t*

Well-learned iterations speed

$$s_{t,c}^{l} = \sum_{\mathrm{T=o}}^{t} I_{\Delta>0} \ln\left(\frac{d_{\mathrm{T},c}}{d_{\mathrm{T}-1,c}}\right)$$

$$\Delta = d_{t,c} - d_{t-1,c}$$

Not learned iterations speed

$$s_{t,c}^{u} = \sum_{\mathrm{T=o}}^{t} I_{\Delta\leq 0} \ln\left(\frac{d_{\mathrm{T},c}}{d_{\mathrm{T}-1,c}}\right)$$

Learning speed (PSI)

$$s_{t,c} = \left(\frac{s_{t,c}^{u} + \epsilon}{s_{t,c}^{l} + \epsilon}\right)^{\alpha}$$

α, ϵ: hyperparameters

$d_{\mathrm{T},c}$: Dice score of category c in iteration T

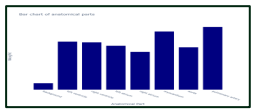
Train different models with category-specific weights

Category-wise difficulty based weight (DifW) $w_{t,c}^{dif}$

DHC [1]

Category-wise Segmentation difficulty in iter t

Dice score $d_{t,c}$



Well-learned speed $s_{t,c} = \left(\frac{s_{t,c}^{u} + \epsilon}{s_{t,c}^{l} + \epsilon}\right)^{\alpha}$

$$w_{t,c}^{dif} = (1 - d_{t,c}) s_{t,c}$$

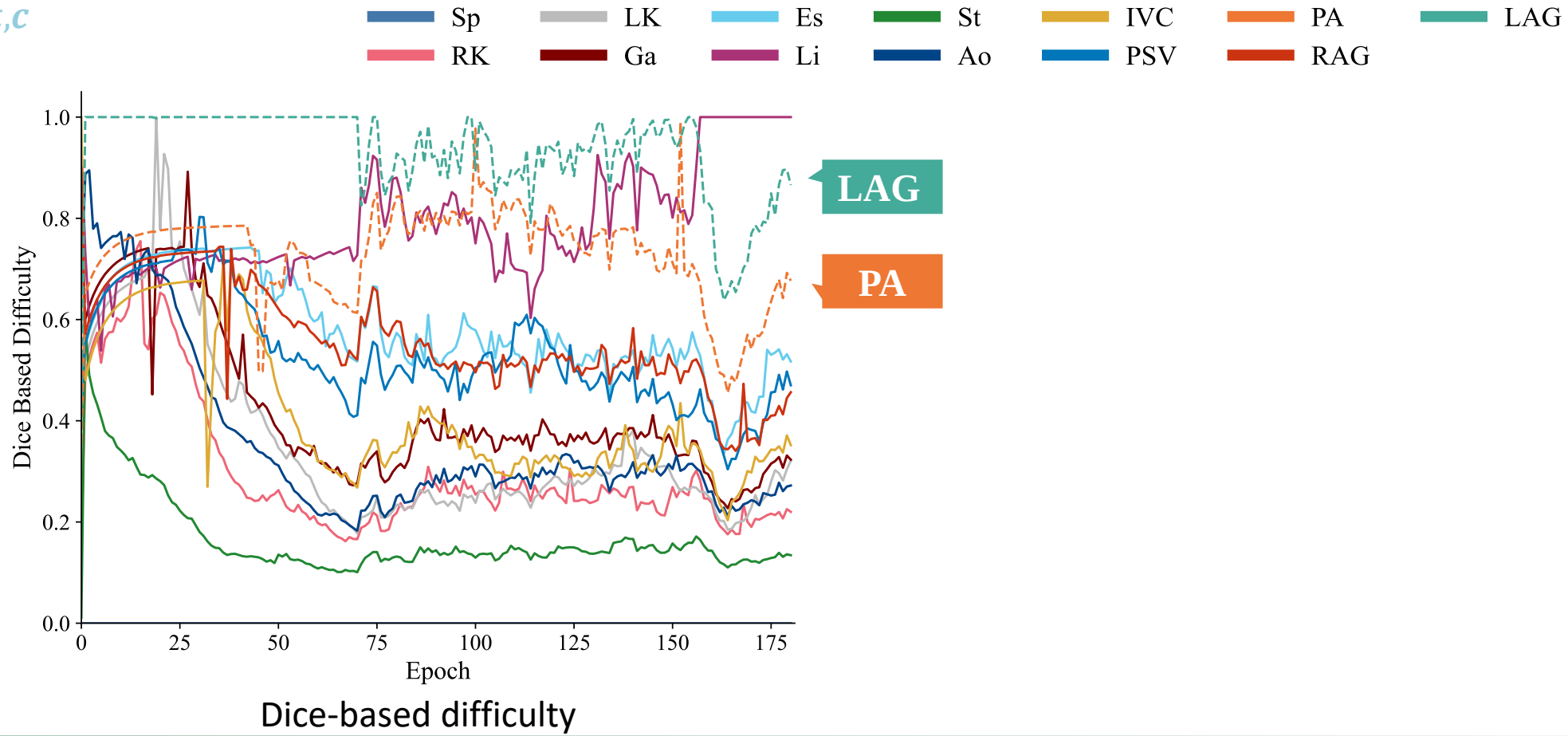
Well learned ↑ $s_{t,c}$ ↓

[1] Wang, Haonan, and Xiaomeng Li. "DHC: Dual-debiased heterogeneous co-training framework for class-imbalanced semi-supervised medical image segmentation." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2023.

Train different models with category-specific weights

Issues in $w_{t,c}^{dif}$

Dice fluctuates significantly, making the training process unstable.



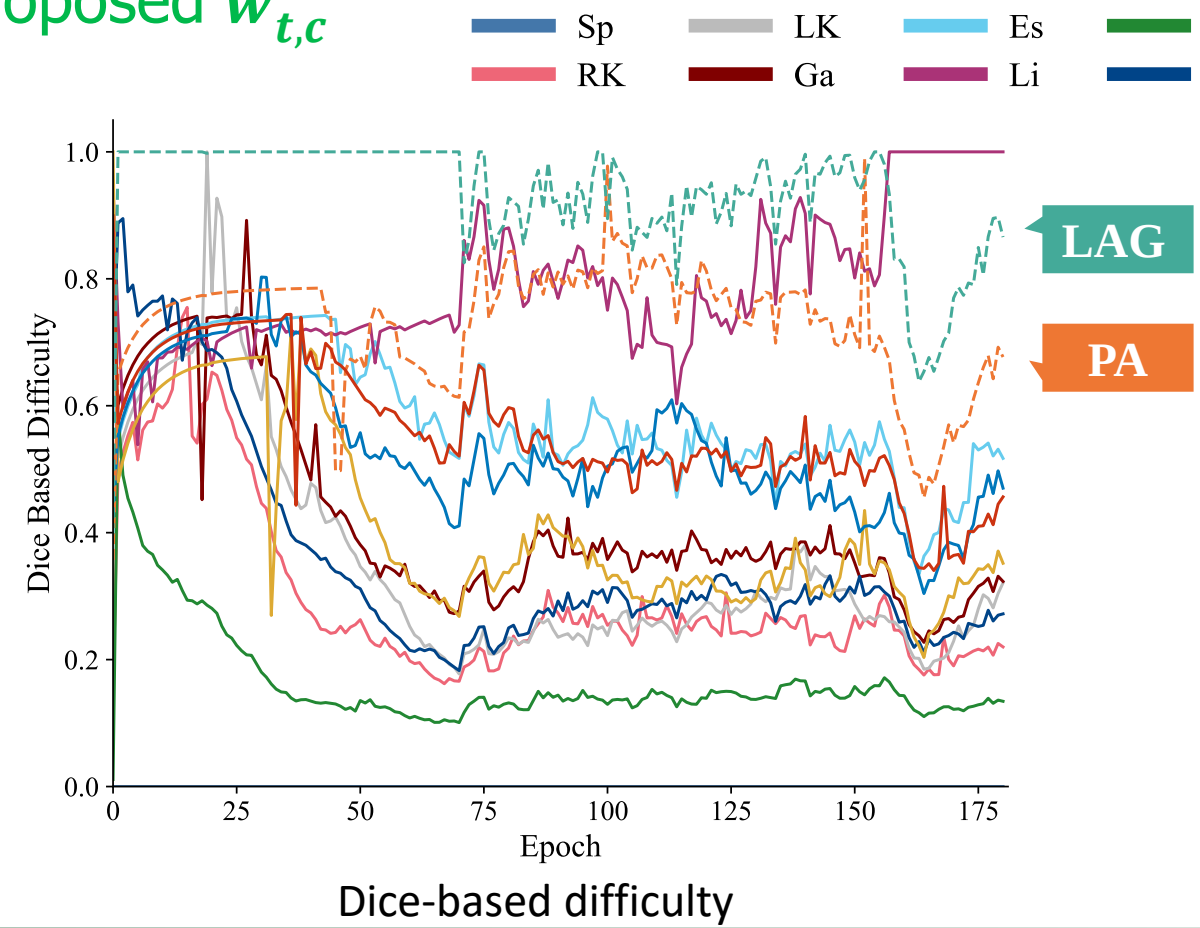
Dice-based difficulty

Train different models with category-specific weights

Soluation

Changes in Confidence can smoothly reflect the difficulty level of each category [1].

Proposed $w_{t,c}^{cdif}$



Dice-based difficulty

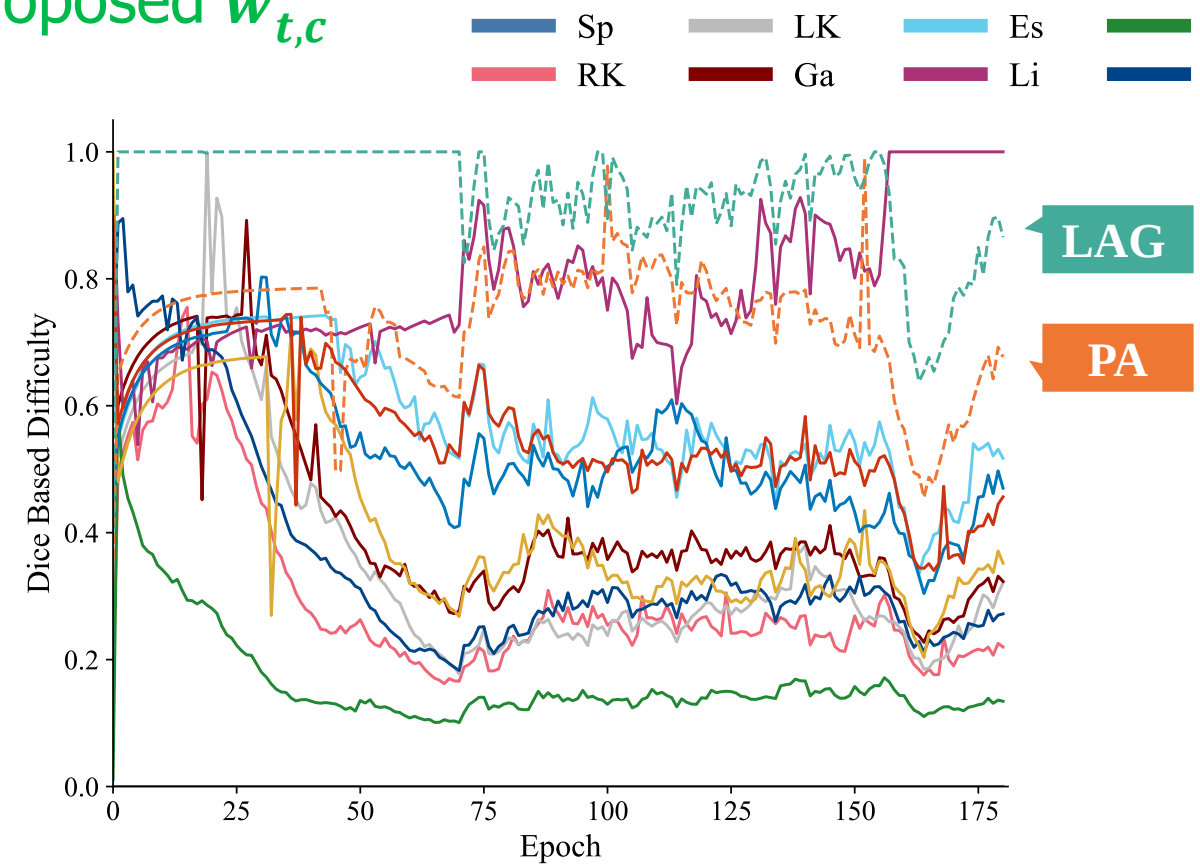
[1] Qiu, J., Hayashi, Y., Oda, M., Kitasaka, T., Mori, K.: Class-wise confidence-aware active learning for laparoscopic images segmentation. International Journalof Computer Assisted Radiology and Surgery 18(3), 473–482, 2023

Train different models with category-specific weights

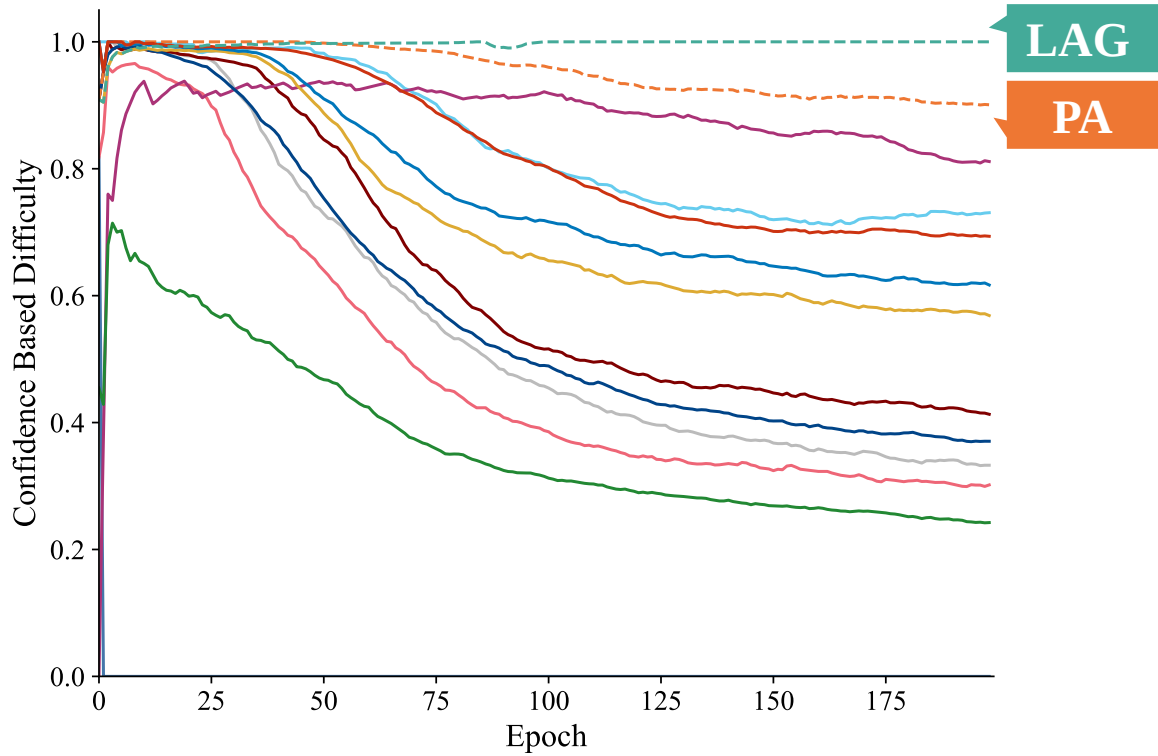
Soluation

Proposed $w_{t,c}^{cdif}$

By combining Dice and Confidence [1], fluctuations become more stable.



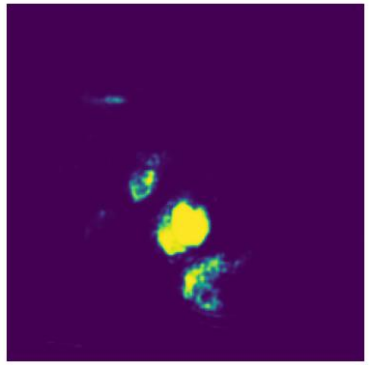
Dice-based difficulty



Confidence-based difficulty

Train different models with category-specific weights

Propose: Confidence-Difficulty Weight (CDifW) $w_{t,c}^{cdif}$



Prediction map $\mathbf{P}_t$

Category-wise position on Ground Truth

$\mathbf{J}_t = \{\mathbf{J}_{t,c} \mid \mathrm{c} = 1,2, \ldots K\}$

Category-wise voxel counts on Ground Truth

$\mathbf{Z}_t = \{z_{t,c} \mid \mathrm{c} = 1,2, \ldots K\}$

$$r_{t,c} = \frac{1}{z_{t,c}} \sum_{j \in J_{t,c}} p_{c,j}$$

Category-wise average Confidence r

$$= i_{t,c} \frac{1 - r_{t,c}}{\max\limits_{c \in \{1,2,\ldots K\}} (1 - r_{t,c})}$$

The information score i [1] of Category-wise average Confidence r

[1] Qiu, J., Hayashi, Y., Oda, M., Kitasaka, T., Mori, K.: Class-wise confidence-aware active learning for laparoscopic images segmentation. International Journalof Computer Assisted Radiology and Surgery 18(3), 473–482, 2023

Train different models with category-specific weights

Propose: Confidence-Difficulty Weight (CDifW) $w_{t,c}^{cdif}$

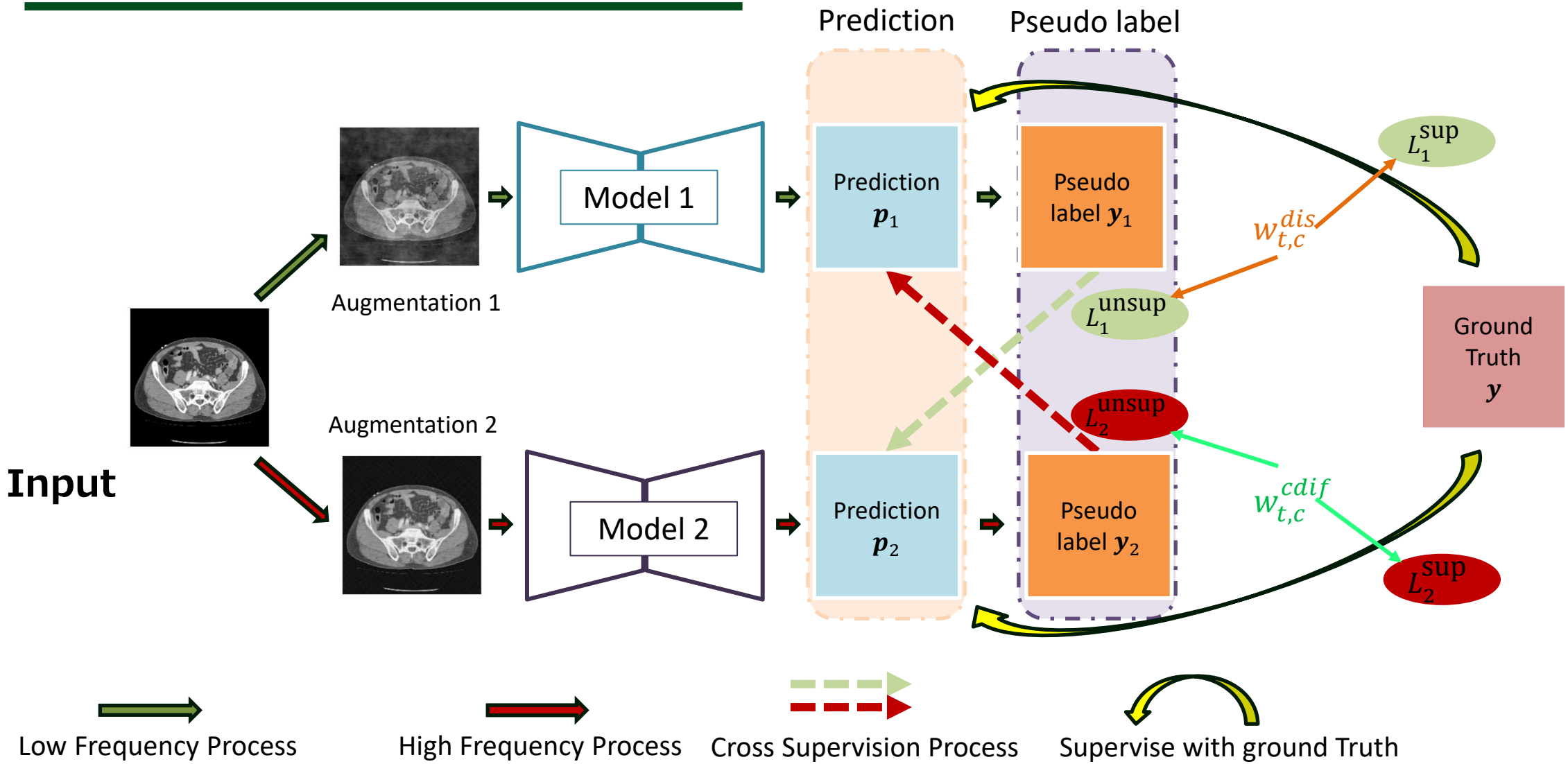
Category-wise Segmentation difficulty in iteration t

- Dice score $d_{t,c}$
- Confidence score $i_{t,c}$
- Well-learned speed $s_{t,c} = \left(\frac{s_{t,c}^{u} + \epsilon}{s_{t,c}^{l} + \epsilon}\right)^{\alpha}$

$$w_{t,c}^{cdif} = i_{t,c}^{\beta}\left(1 - d_{t,c}\right)s_{t,c}$$

β:hyperparameter

Train different models with category-specific weights



Problem we aiming to solve

The lack of the data → Dual-Network framework (CPS)

Dual-Network framework (CPS) → Category-wise imbalance

Dual-Network framework (CPS) → The lack of 異質性

Category-wise imbalance → Train different models with category-specific weights

Train different models with category-specific weights → Difficulty based DifW

Train different models with category-specific weights → Distribution based DisW

Difficulty based DifW → Difficulty-Confidence based CDifW

The lack of 異質性 → Applying different augmentation for the same image as the input

Different image augmentations

Related works

Weak Augmentations
- Gaussian Noise
- Gamma Correction
- Gaussian Blur

etc.

Strong Augmentations
- CutOut [1]
- CutMix [2]
- ClassMix [3]

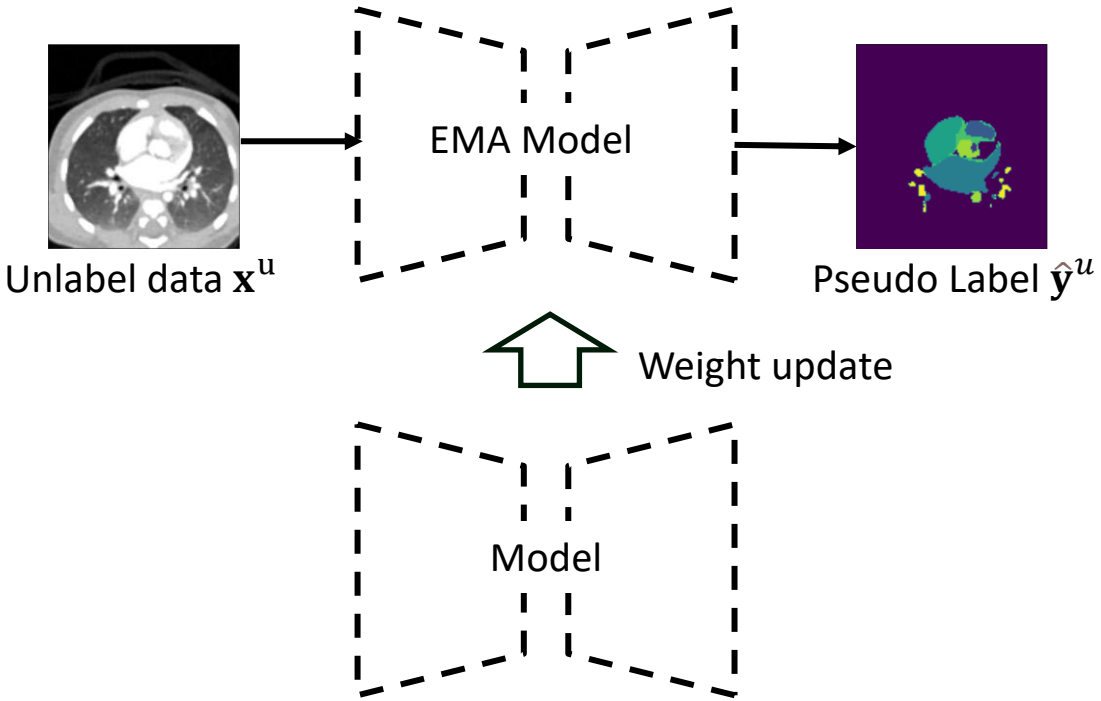
etc.

Differences	☹	☺	☺
Imbalance	☹	☹	☺

Different image augmentations

Propose ： Double-Mix Pseudo Label (DMP)

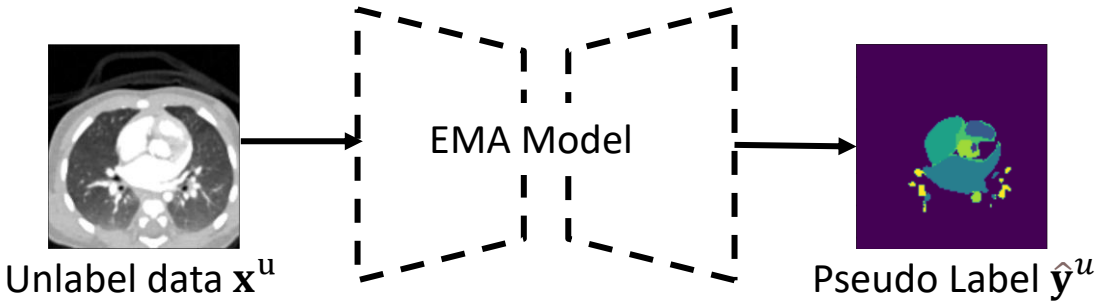
Step 1 reference Pseudo Label $\hat{\mathbf{y}}^u$ by EMA model

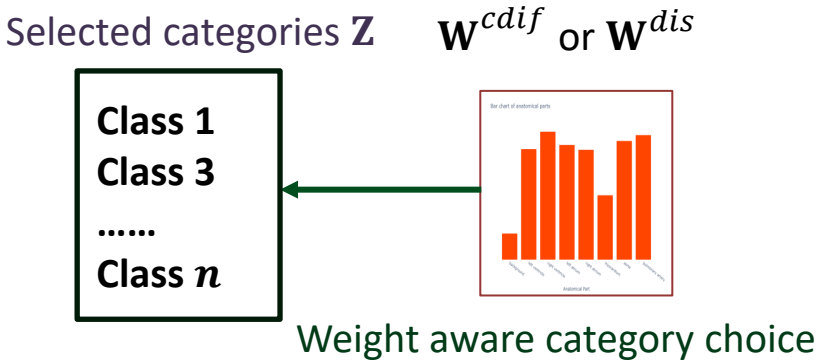


Different image augmentations

Propose ： Double-Mix Pseudo Label (DMP)

Step 2 Select categories using category-wise weight (in this experiment, $\mathbf{W}^{cdif}$ or $\mathbf{W}^{dis}$)



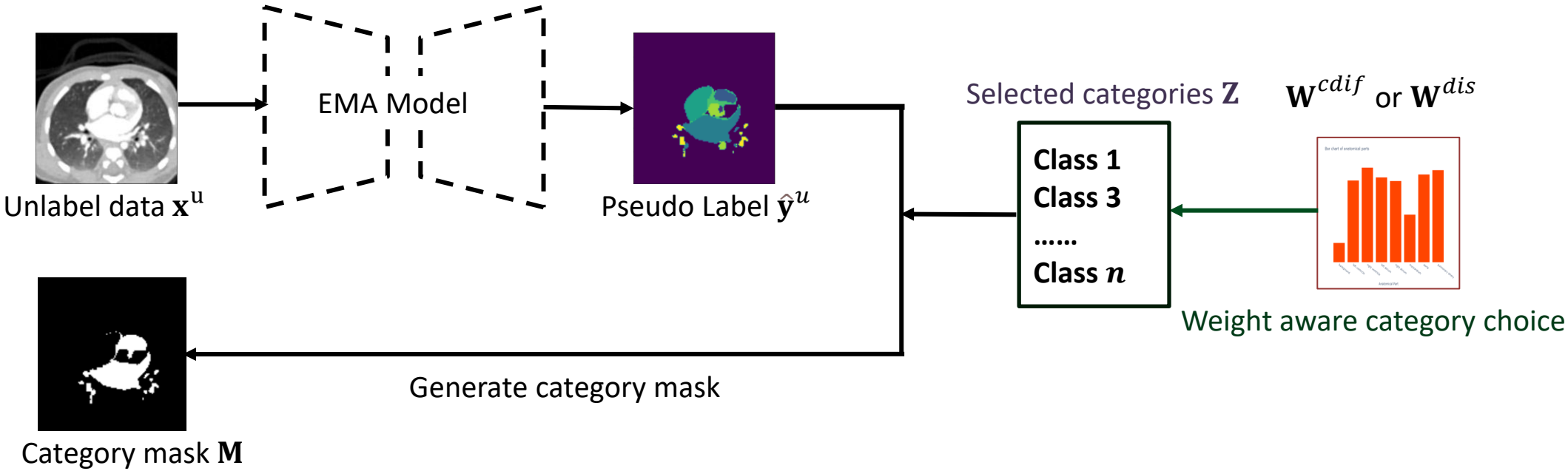


The weight for category c represents the probability of this category being sampled.

Different image augmentations

Propose ： Double-Mix Pseudo Label (DMP)

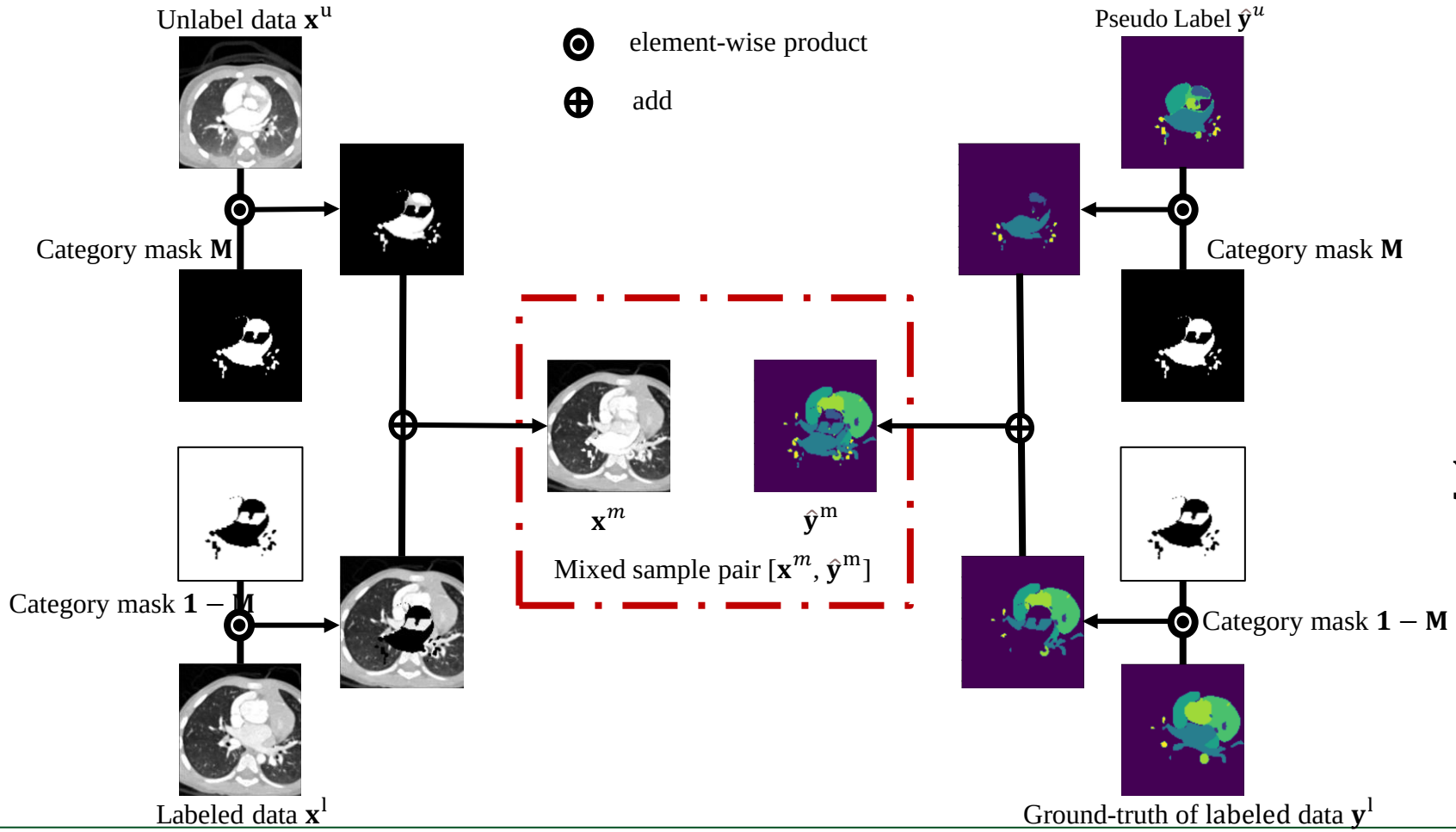
Step 3 Generate category mask $\mathbf{M}$ by selected categories $\mathbf{Z}$ and pseudo label $\hat{\mathbf{y}}^u$



Different image augmentations

Step 4 Generate mixed sample pair $[\mathbf{x}^m, \hat{\mathbf{y}}^m]$

Propose： Double-Mix Pseudo Label (DMP)



Unlabeled data $\mathbf{x}^u$

Pseudo label $\hat{\mathbf{y}}^u$

Labeled data $\mathbf{x}^l$

Ground truth $\mathbf{y}^l$

$$\mathbf{x}^m = \mathbf{x}^u \odot \mathbf{M} + \mathbf{x}^l \odot (1 - \mathbf{M})$$

$$\hat{\mathbf{y}}^m = \hat{\mathbf{y}}^u \odot \mathbf{M} + \mathbf{y}^l \odot (1 - \mathbf{M})$$

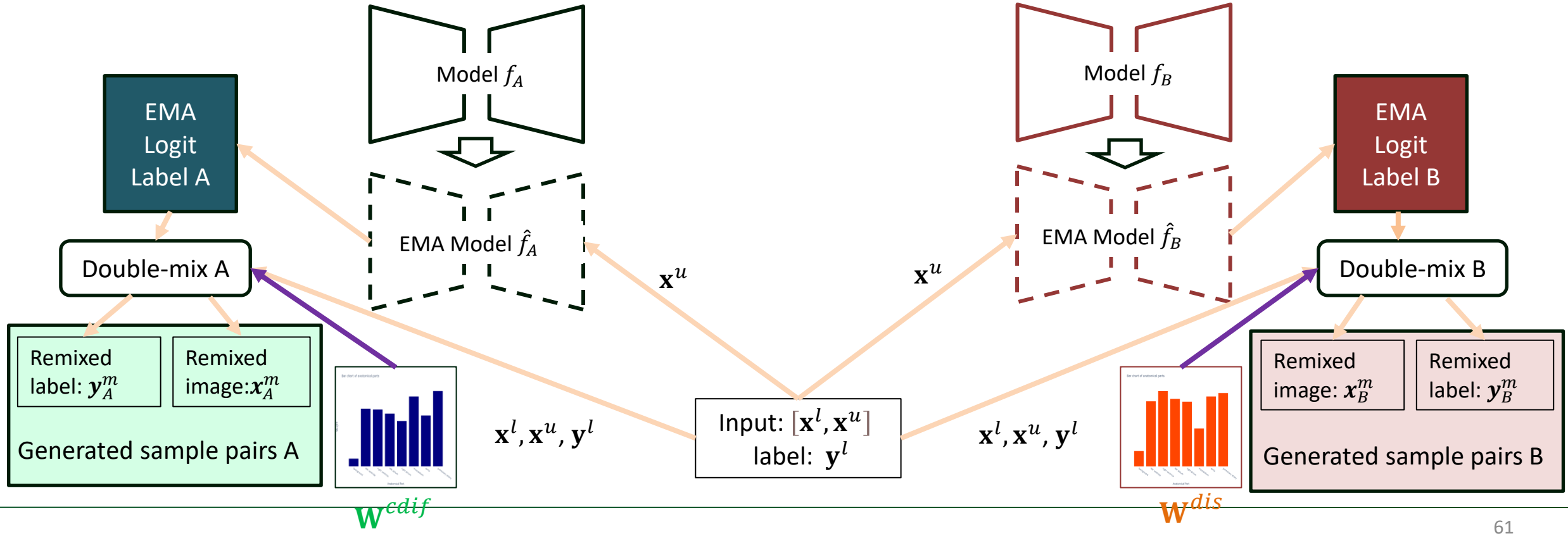
Mixed sample pair $[\mathbf{x}^m, \hat{\mathbf{y}}^m]$

Different image augmentations

f_A, f_B:Segmentation models
$\hat{f}_A$, $\hat{f}_B$: EMA models of f_A, f_B

Propose ： Double-Mix Pseudo-label (DMP)

Generate mixed sample pairs $[\boldsymbol{y}_A^m, \boldsymbol{x}_A^m]$ and $[\boldsymbol{y}_B^m, \boldsymbol{x}_B^m]$ based on $\mathbf{W}^{cdif}$ and $\mathbf{W}^{dis}$



Different image augmentations

Related works

Weak Augmentations

- Gaussian Noise
- Gamma Correction
- Gaussian Blur

etc ..

Strong Augmentations

- CutOut [1]
- CutMix [2]
- ClassMix [3]

etc ..

Proposed method

Double-Mix Pseudo Label

- Used different weights for data augmentation
- Considered category-wise imbalance

Differences	☹	☺	☺
Imbalance	☹	☹	☺

Problem we aiming to solve

The lack of the data

Dual-Network framework (CPS)

Category-wise imbalance

The lack of 異質性

Train different models with category-specific weights

Applying different augmentation for the same image as the input

Difficulty based DifW

Distribution based DisW

Double-Mix Pseudo Label

Difficulty-Confidence based CDifW

Double-Mix Pseudo Label Framework (DMPF) (in iteration t)

f_A, f_B:Segmentation models
$\hat{f}_A$, $\hat{f}_B$: EMA models of f_A, f_B

- **Step1:** Calculate $\mathbf{W}_t^{cdif}$ and $\mathbf{W}_t^{dis}$
- **Step2:** Update the EMA models and Generate pseudo label of unlabeled data $\hat{\mathbf{y}}_A^u$, $\hat{\mathbf{y}}_B^u$
- **Step3:** Generate DMP sample pairs $[\mathbf{y}_A^m, \mathbf{x}_A^m]$ and $[\mathbf{y}_B^m, \mathbf{x}_B^m]$ by DMP, using $[\mathbf{W}_t^{cdif}, \mathbf{x}^l, \mathbf{y}^l, \mathbf{x}^u, \hat{\mathbf{y}}_A^u]$and $[\mathbf{W}_t^{dis}, \mathbf{x}^l, \mathbf{y}^l, \mathbf{x}^u, \hat{\mathbf{y}}_B^u]$, respectively
- **Step4:** Calculate the unsupervision loss of sample pairs

$$L_m^{unsup} = L_{Seg}^{unsup}(\mathbf{W}_t^{cdif}, f_A(\boldsymbol{x}_A^m), \boldsymbol{y}_A^m) + L_{Seg}^{unsup}(\mathbf{W}_t^{dis}, f_B(\boldsymbol{x}_B^m), \boldsymbol{y}_B^m)$$

$$L_{Seg}^{sup}(\mathbf{W}, \mathbf{x}, \mathbf{y}) = L_{Dice}(\mathbf{W}, \mathbf{x}, \mathbf{y}) + \frac{1}{2} L_{CE}(\mathbf{W}, \mathbf{x}, \mathbf{y})$$

$$L_{Seg}^{unsup}(\mathbf{W}, \mathbf{x}, \mathbf{y}) = L_{CE}(\mathbf{W}, \mathbf{x}, \mathbf{y})$$

L_{CE}: weighted Cross Entropy loss

L_{Dice}: weighted Dice loss

Double-Mix Pseudo Label Framework (DMPF) (in iteration t)

- **Step5:** Calculate the supervision loss

$$L^{sup} = L_{Seg}^{sup}(\mathbf{W}_t^{cdif}, f_A(\boldsymbol{x}^l), \boldsymbol{y}^l) + L_{Seg}^{sup}(\mathbf{W}_t^{dis}, f_B(\boldsymbol{x}^l), \boldsymbol{y}^l)$$

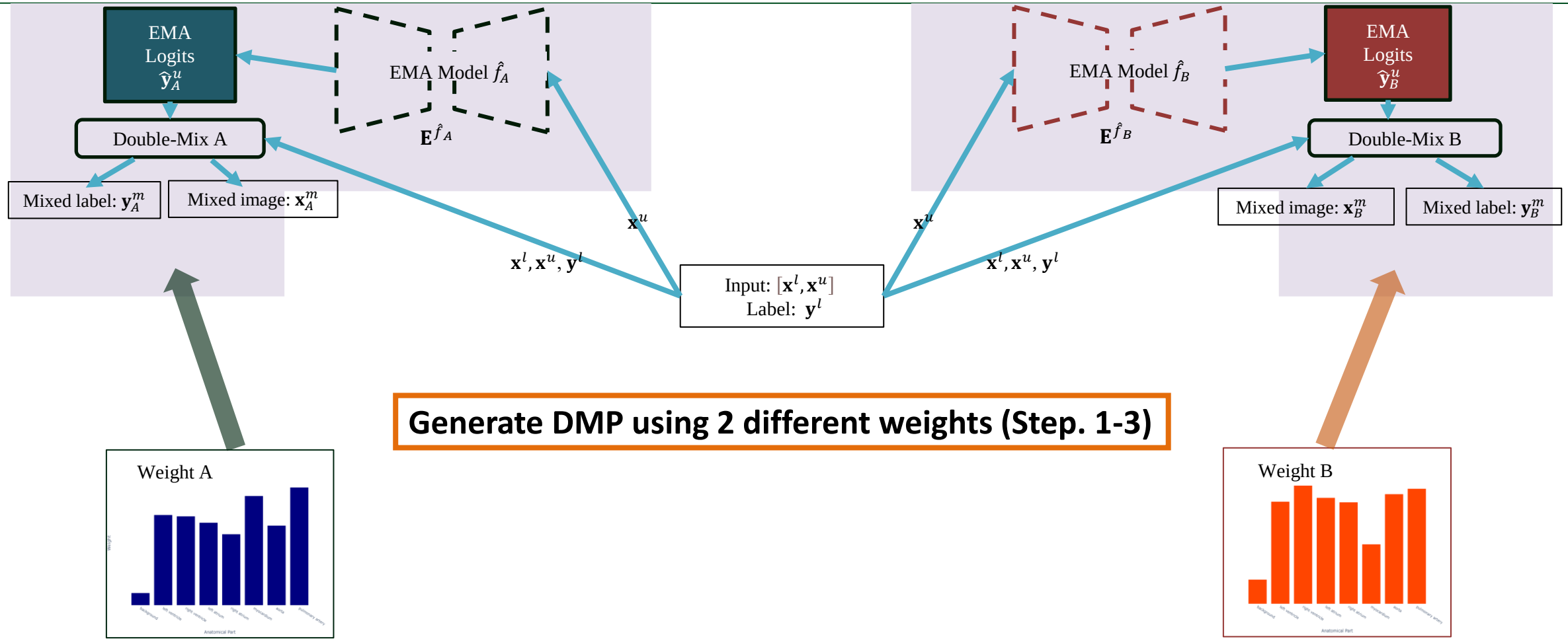
- **Step6:** Calculate the unsupervision loss of unlabeled sample pairs

$$L_u^{unsup} = L_{Seg}^{sup}(\mathbf{W}_t^{cdif}, f_A(\boldsymbol{x}^u), \hat{\mathbf{y}}_B^u) + L_{Seg}^{sup}(\mathbf{W}_t^{dis}, f_B(\boldsymbol{x}^l), \hat{\mathbf{y}}_A^u)$$

- **Step7:** Calculate the total loss and optimize the models

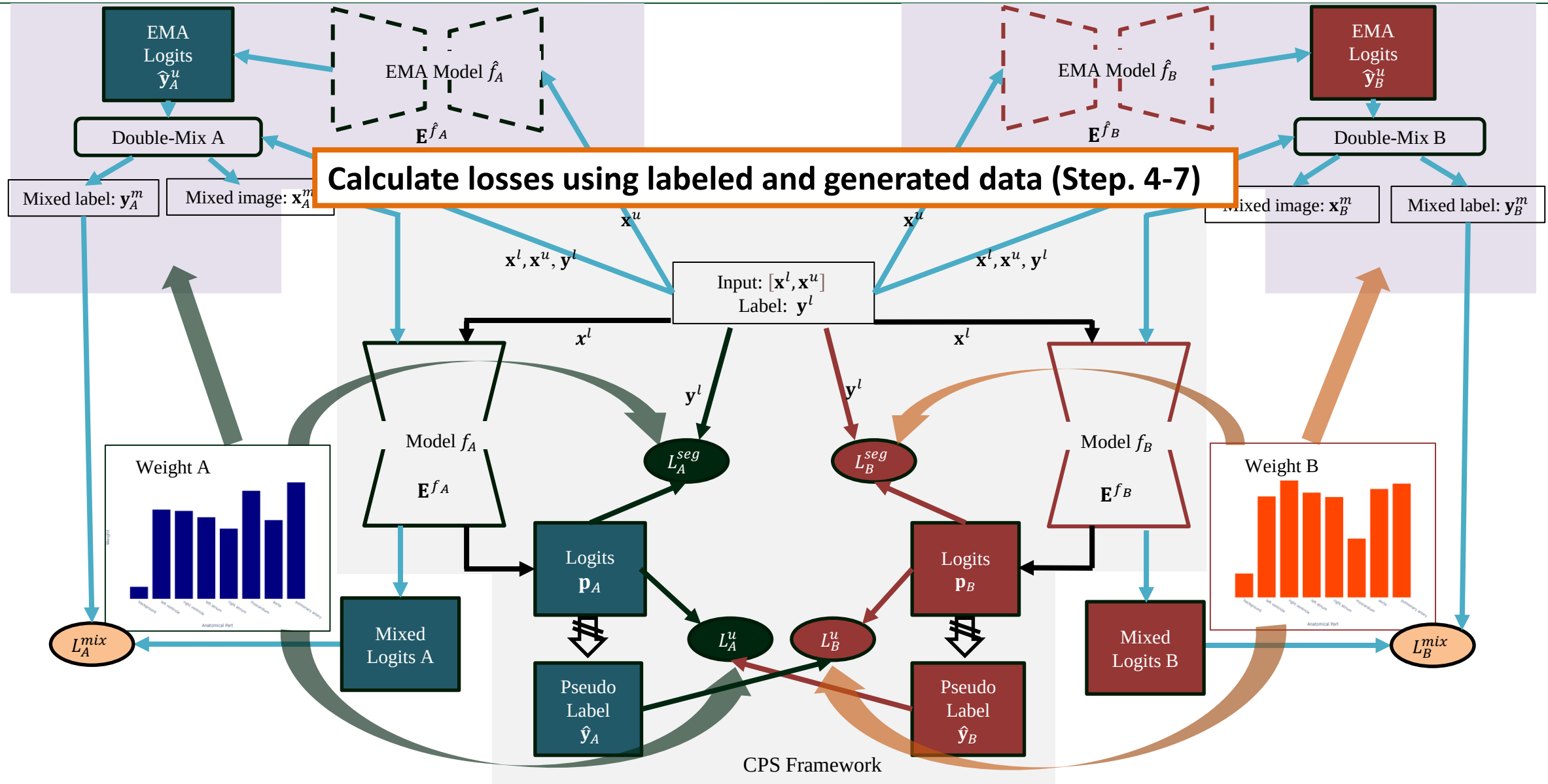
$$L^{total} = L^{sup} + L_m^{unsup} + \theta L_u^{unsup}$$

θ: Hyperparater



Generate DMP using 2 different weights (Step. 1-3)

DMP Modules | Infer without gradient | Double-mix process | CPS framework training | Apply the weight distribution



DMP Modules | Infer without gradient | Double-mix process | CPS framework training | Apply the weight distribution

Experiment setting

- **Datasets**

CHD [1]

Whole heart and great vessel segmentation

Training set: 88

Validation set: 11

Test set: 11

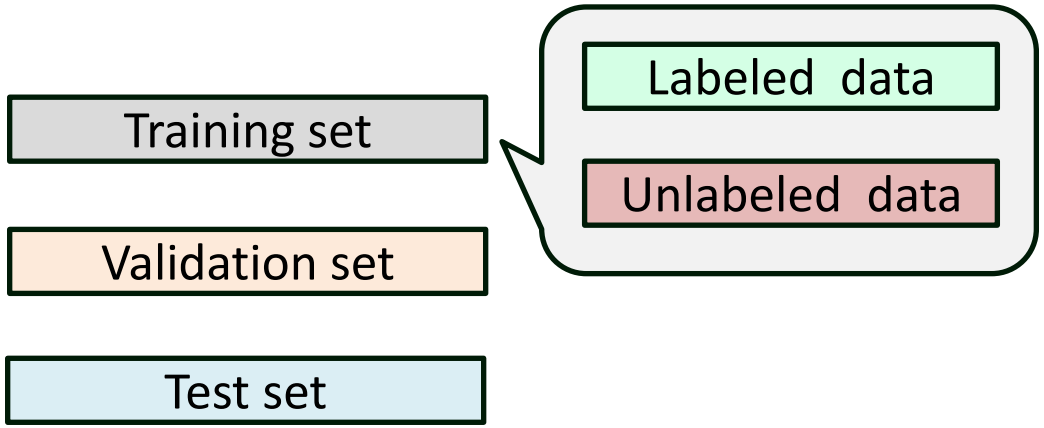
BTCV [2] (腹部臓器)

Abdominal Organ Segmentation Dataset

Training set: 20

Validation set: 4

Test set: 6



[1] Xu, X., Wang, T., Shi, Y., Yuan, H., Jia, Q., Huang, M., Zhuang, J.: Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching. In: MICCAI, Proceedings, Part II, LNIP, vol. 11765, pp. 477–485, 2019
[2] Landman, B., Xu, Z., Igelsias, J., Styner, M., Langerak, T., Klein, A.: 2015 MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge, 2015

Experiment setting

臓器の英和対照
- 脾臓 (Sp)
- 右腎臓 (RK)
- 左腎臓 (LK)
- 胆嚢 (Ga)
- 食道 (Es)
- 肝臓 (Li)
- 胃 (St)
- 大動脈 (Ao)
- 下大静脈 (IVC)
- 門脈・脾静脈 (PSV)
- 膵臓 (Pa)
- 右副腎 (RAG)
- 左副腎 (LAG)

心臓構造の英和対照
- 左心室 (LV)
- 右心室 (RV)
- 左心房 (LA)
- 右心房 (RA)
- 心筋 (Myo)
- 大動脈 (Ao)
- 肺動脈 (PA)

Experiment setting

- **Model Training**

Training Settings:

3 random seeds, trained 3 times.

Segmentation Model:

5-layer V-Net [1].

Data Augmentation:

Gaussian noise, random flip, random rotation, random crop

- **Details**

V-Net

kernel numbers:

[32, 64, 128, 256, 512] in encoder and decoder

Input patch size:

(128, 128, 64)

Metrics

Dice score

Average Surface Distance (ASD)

[1] Milletari, F., Navab, N., Ahmadi, S.-A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3DV, pp. 565–571, 2016. IEEE

Experiment setting

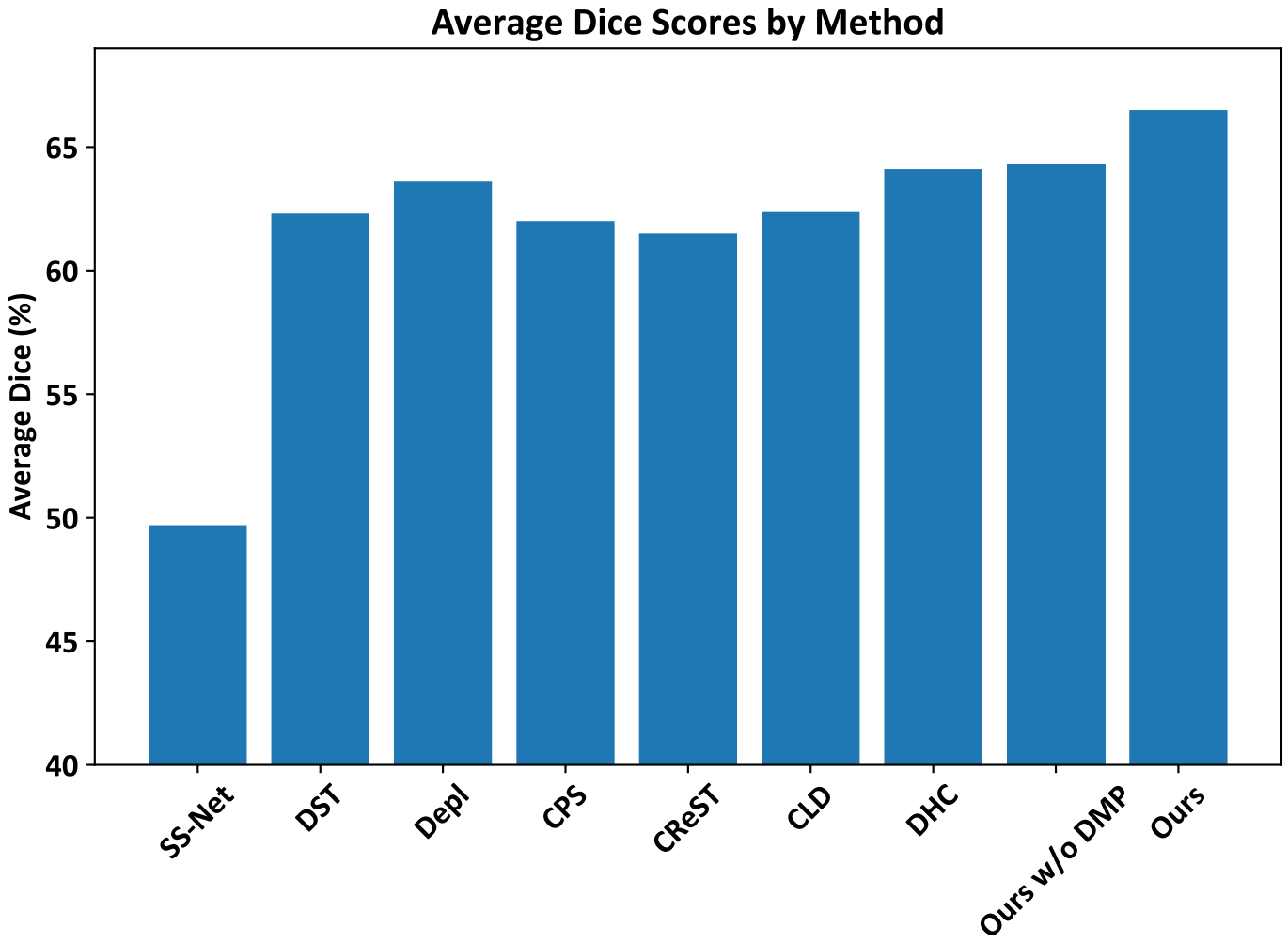
- **Mainstream Approaches**
  - SS-Net [1]
  - DST [2]
  - Depl [3]
  - CPS [4]
  - CReST [5]
  - CLD [6]
  - DHC (DisW + DifW) [7]
  - Ours w/o DMP (DisW + CDifW)
  - Ours (DisW + CDifW + DMP)

[1] Wu, Y., Wu, Z., Wu, Q., Ge, Z., Cai, J.: Exploring smoothness and classseparation for semi-supervised medical image segmentation. In: MICCAI, LNCS, vol. 13435, pp. 34–43 (2022). Springer
[2] Chen, B., Jiang, J., Wang, X., Wan, P., Wang, J., Long, M.: Debiased self-training for semi-supervised learning. In: NeurIPS, vol. 35, pp. 32424–32437 (2022)
[3] Wang, X., Wu, Z., Lian, L., Yu, S.X.: Debiased learning from naturally imbalanced pseudo-labels. In: CVPR, pp. 14647–14657 (2022)
[4] Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: CVPR, pp. 2613–2622 (2021)
[5] Wei, C., Sohn, K., Mellina, C., Yuille, A., Yang, F.: CReST: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In: CVPR, pp. 10857–10866 (2021)
[6] Lin, Y., Yao, H., Li, Z., Zheng, G., Li, X.: Calibrating label distribution for classimbalanced barely-supervised knee segmentation. In: MICCAI, LNCS, vol. 13438, pp. 109–118 (2022). Springer
[7] Wang, H., Li, X.: DHC: Dual-debiased heterogeneous co-training framework for class-imbalanced semi-supervised medical image segmentation. In: MICCAI, LNCS, vol. 14222, pp. 582–591 (2022). Springer

Result on using 5% labeled CHD dataset

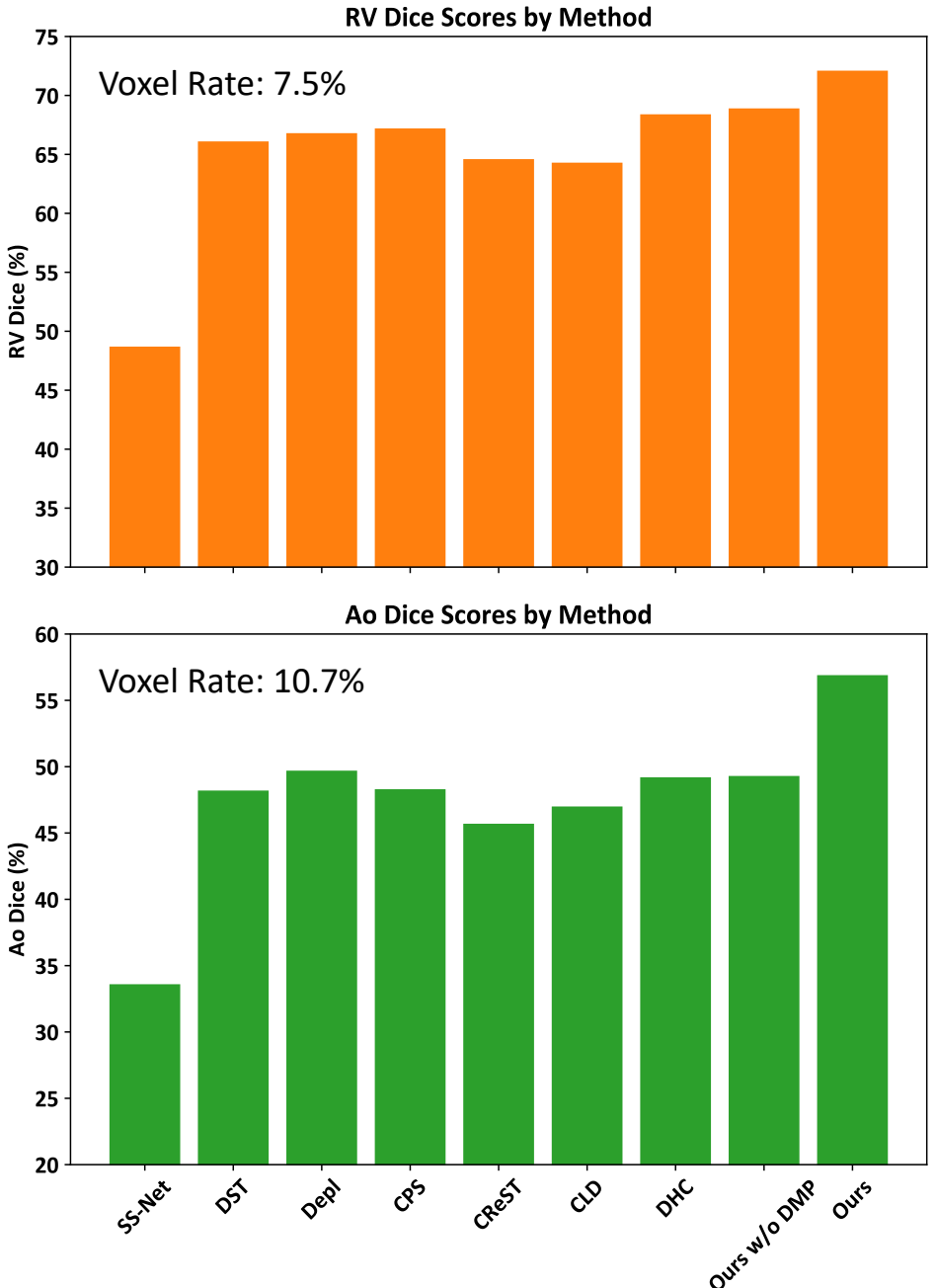
- The proposed method achieves higher average accuracy than related methods.

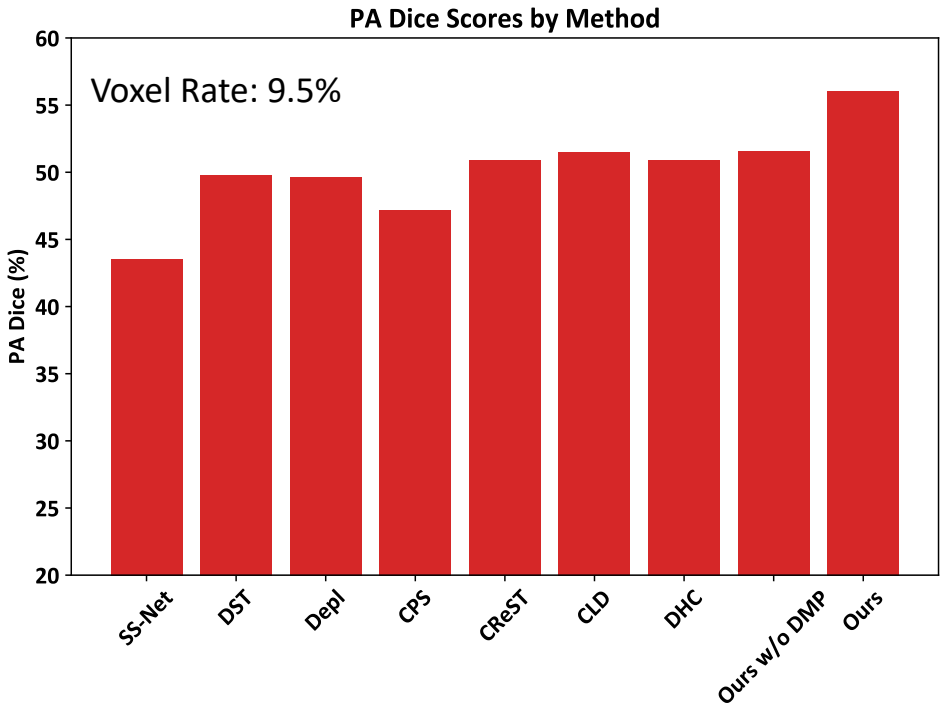
Method	Average Dice	ASD
SS-Net	49.7	7.9
DST	62.3	5.6
Depl	63.6	**5.1**
CPS	62.0	5.5
CReST	61.5	6.4
CLD	62.4	5.9
DHC	64.1	6.7
Ours w/o DMP	64.3	6.0
Ours	**66.5**	6.0



Result on using 5% labeled CHD dataset

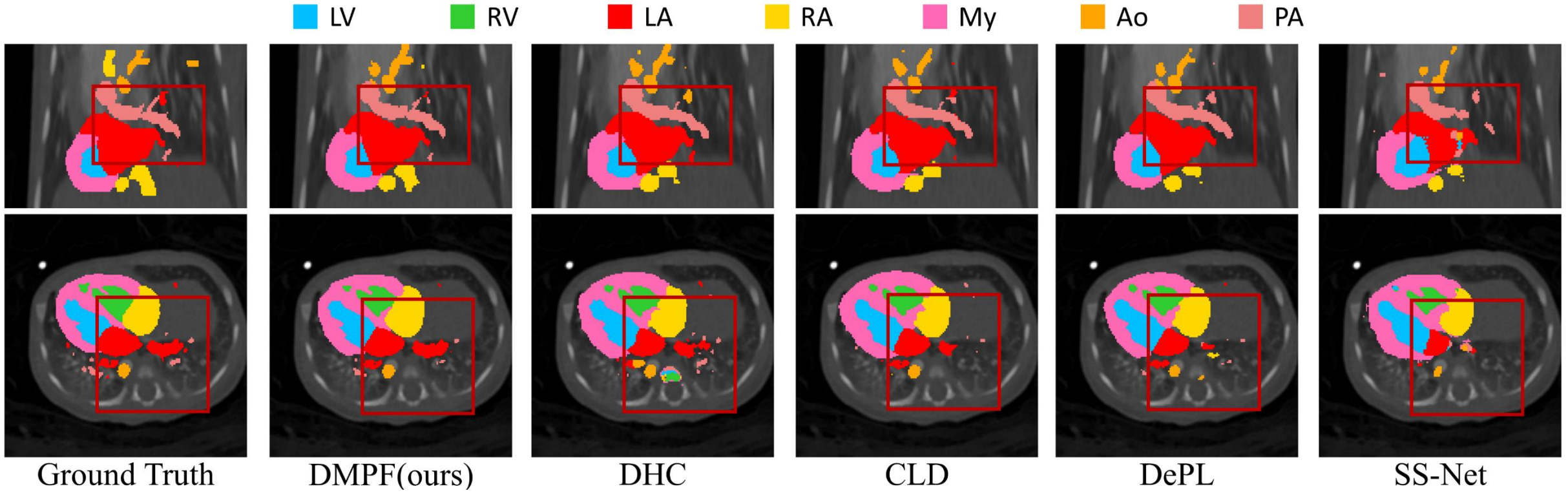
- The proposed method achieves higher accuracy for categories with fewer voxels compared to related methods.





Result on using 5% labeled CHD dataset

The proposed method achieves better segmentation accuracy, especially in challenging categories (PV, RA, Ao).



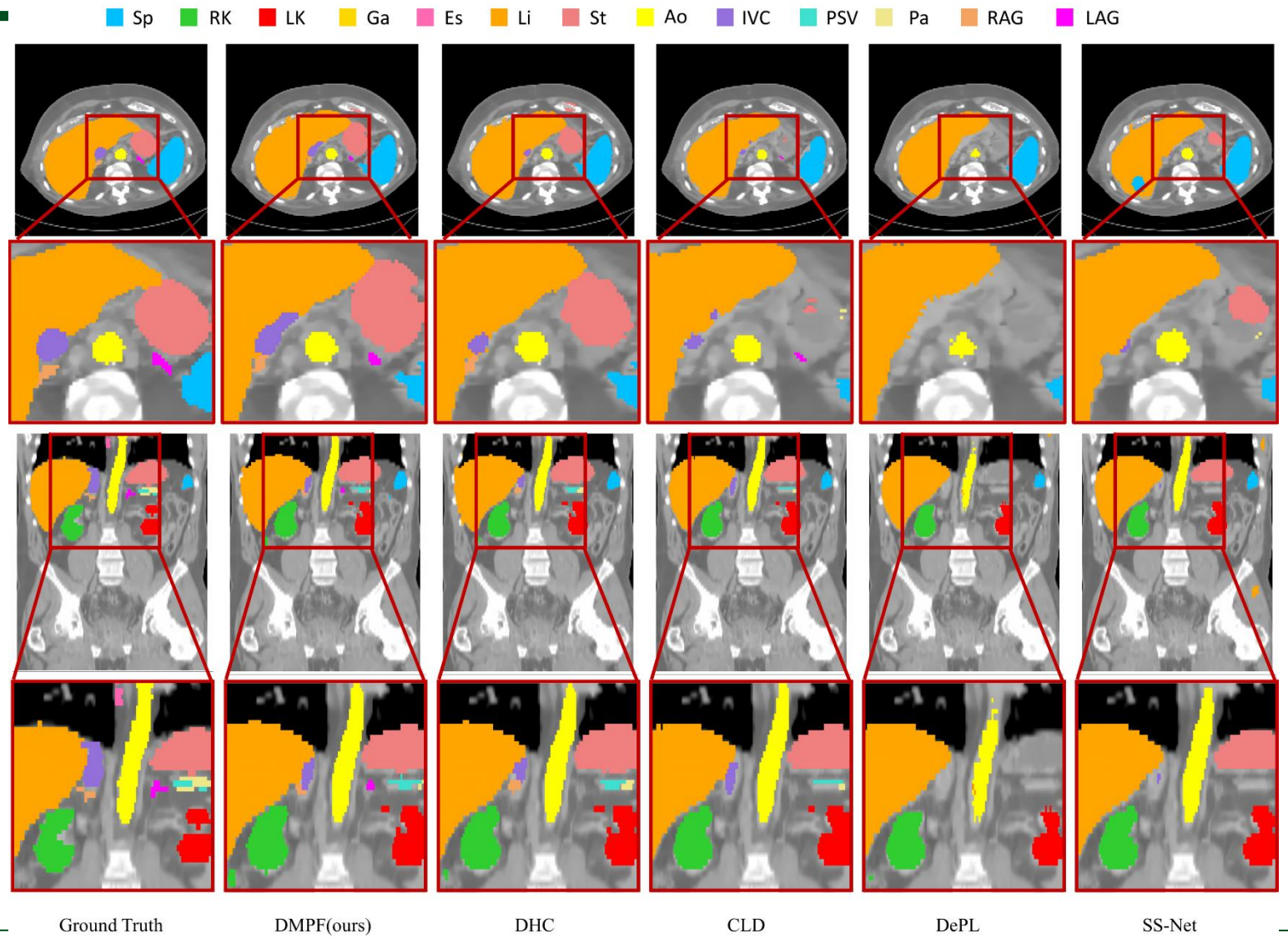
Result on using 40% labeled BTCV dataset

- The proposed method achieves higher accuracy for categories with fewer voxels compared to conventional methods.

Methods	Average Dice and ASD		Low Voxel count category		
	Dice (%)	ASD	Es (0.49%)	RAG (0.14%)	LAG (0.17%)
SS-Net	42.5±6.5	49.2±10.1	0	0	0
DST	40.1±0.9	46.8±2.2	0	0	0
Depl	41.2±0.9	48.1±0.5	0	0	0
CPS	37.5±2.1	52.5±11.1	0	0	0
CReST	38.5±3.8	22.1±8.7	21.2	18.1	9.5
CLD	54.7±1.2	7.6±0.6	28.7	25.3	27
DHC	59.6±1.2	4.5±0.6	44.8	33.1	40.9
ours w/o DMP	60.0±0.7	**3.9±0.5**	45.8	28.9	**50.5**
ours	**61.2±0.7**	4.06±0.6	**48.5**	**36.4**	48

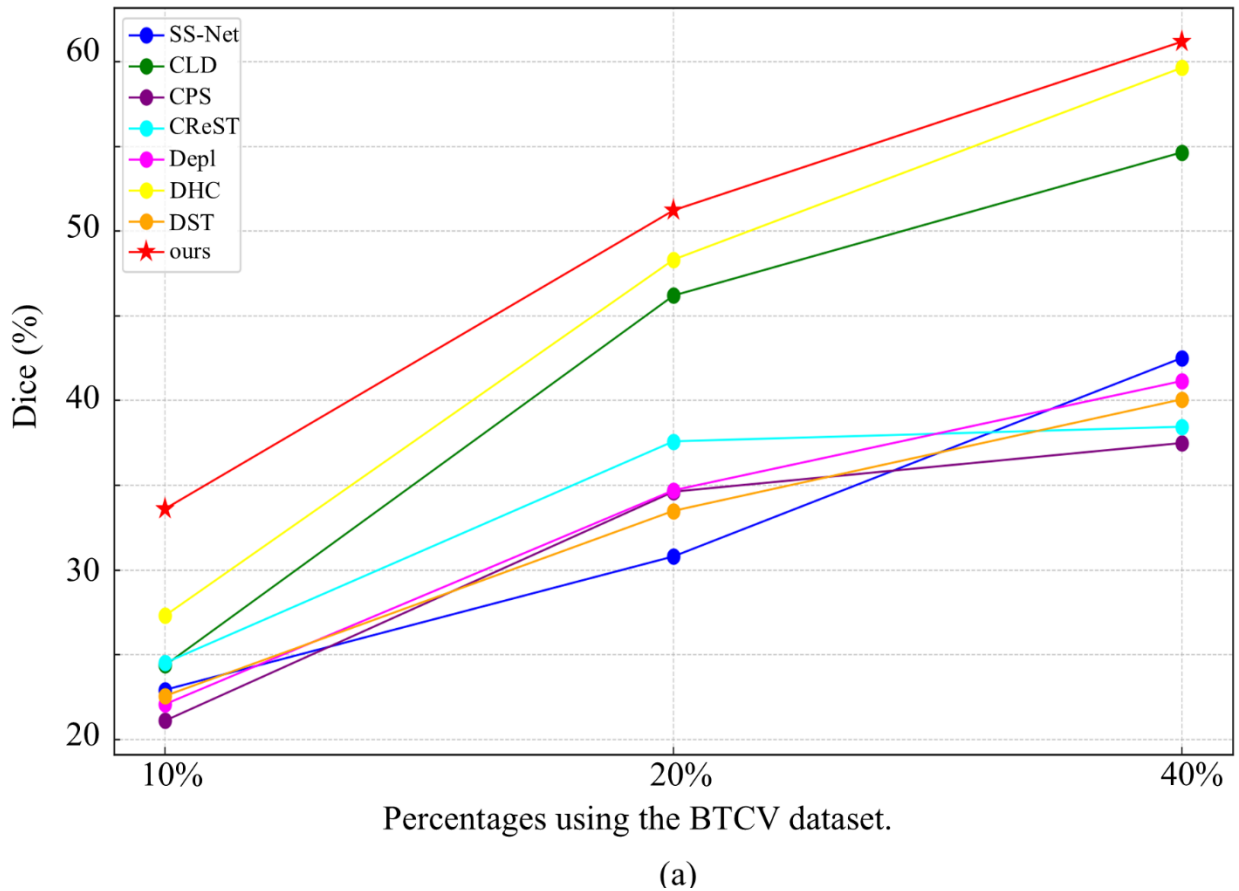
Result on using 40% labeled BTCV dataset

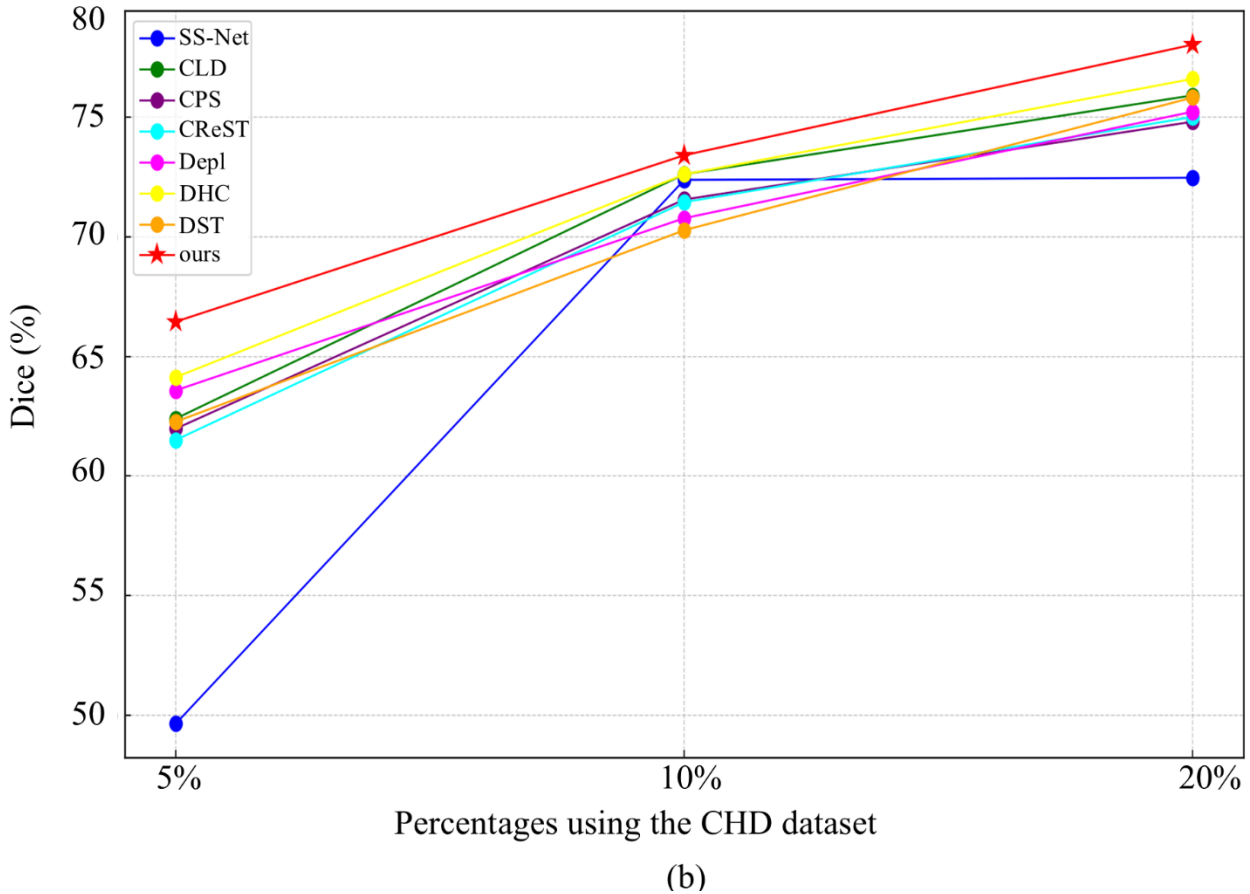
The proposed method achieves better segmentation accuracy, especially in challenging categories (Es, RAG, LAG).



Results on using different ratio labeled dataset

Our method has advantages when using a smaller amount of annotated data.

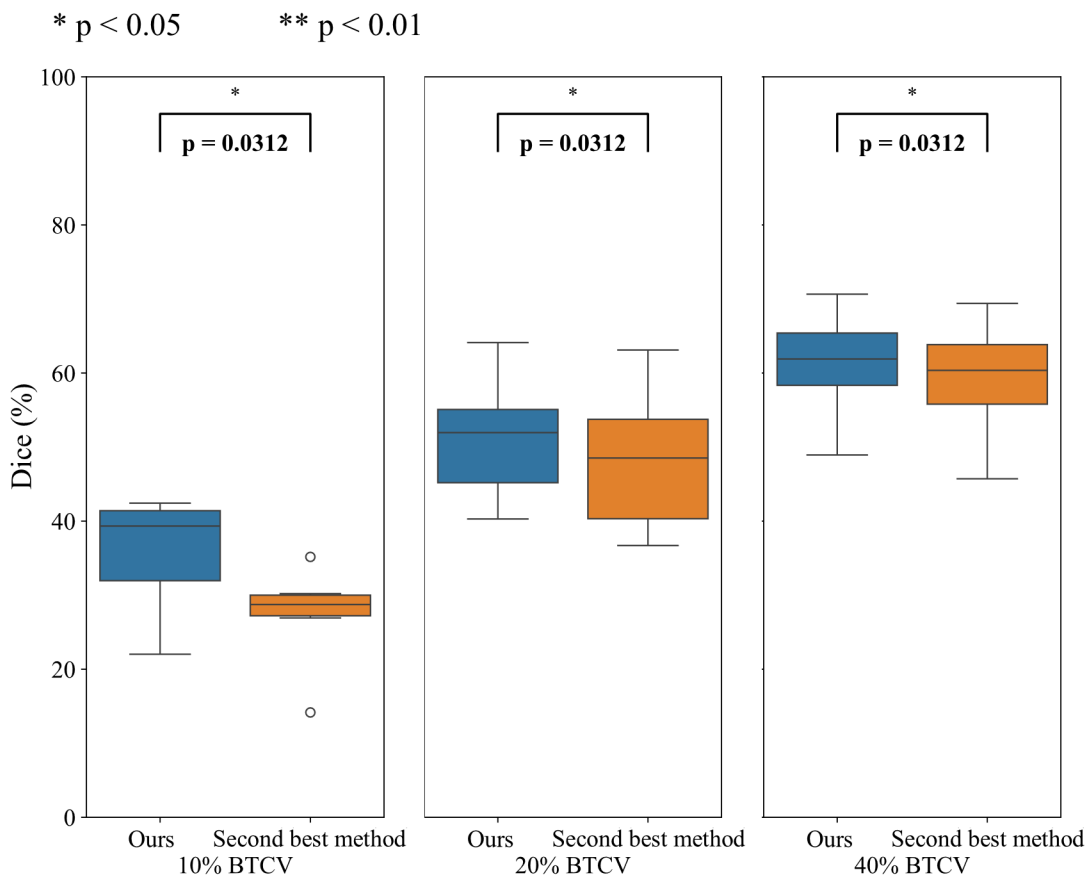




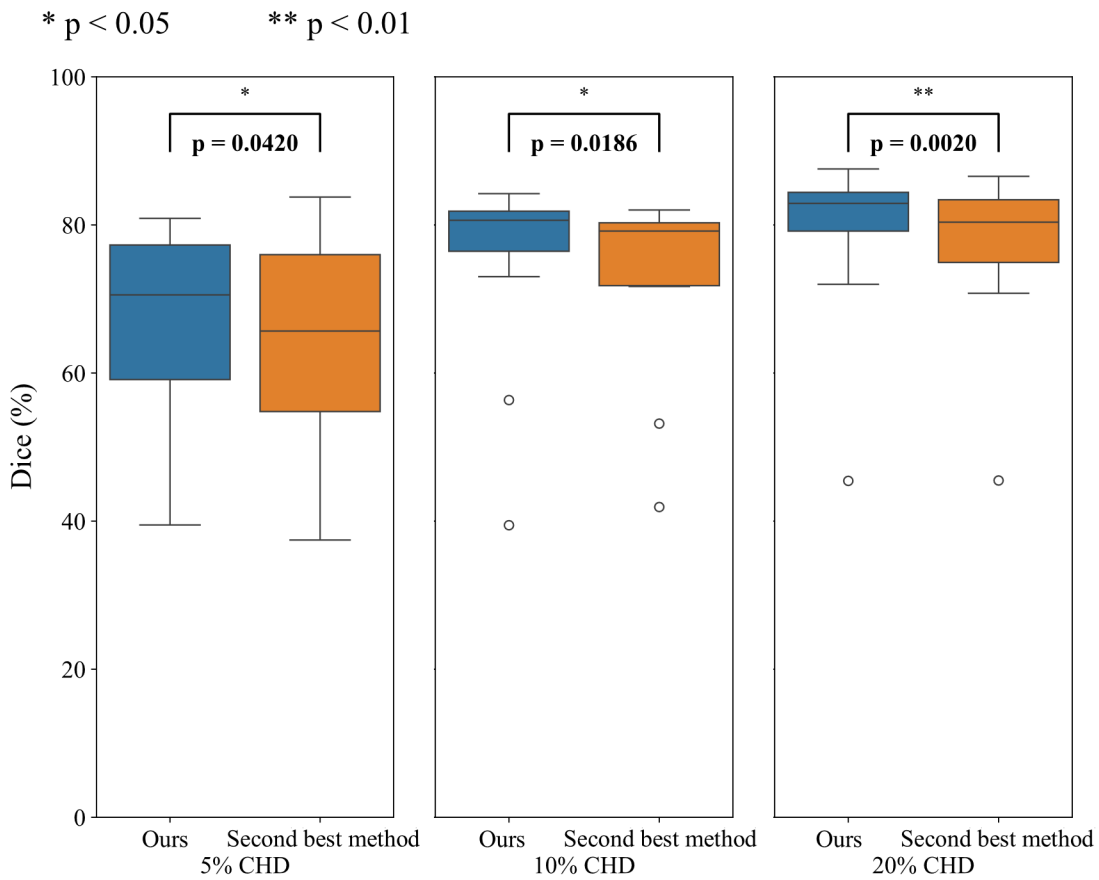
Results using different percentages of labeled-data. (a) BTCV, (b) CHD

Significance test

Our method achieved p-values less than 0.05 across multiple splits on two datasets, demonstrating its effectiveness.



(a)



(b)

The result of Wilcoxon signed-rank test. (a) The results using 10%, 20%, and 40% of the BTCV dataset as the labeled data, (b) the results using 5%, 10%, and 20% of the CHD dataset as the labeled data

The effect of using two distinct category-wise weights in model training

CDisW: $w_{t,c}^{CDisW} = i_{t,c}^{\beta} \, w_{t,c}^{disw}$

Confidence information score in $w_{t,c}^{cdifw}$

CDifW: Confidence + Difficulty score
CDifW: Confidence + Distribution score
DisW: Distribution score
DifW: Difficulty score

The results of training CPS module using different weights on 10% labeled BTCV dataset.

Methods	Dice (%)	ASD
DisW-DifW	29.8±5.4	28.1±8.5
CDifW-DisW	**29.9±2.7**	**25.3±8.3**
CDifW-CDifW	26.4±2.8	25.2±3.5
CDifW-CDisW	29.4±2.6	27.3±3.4
DisW-DisW	26.2±4.7	32.2±5.0
CDisW-CDisW	28.9±2.6	25.5±4.4

The effect of using two distinct category-wise weights in model training

CDisW: $w_{t,c}^{CDisW} = i_{t,c}^{\beta} \, w_{t,c}^{disw}$

Confidence information score in $w_{t,c}^{cdifw}$

- Using the same weights fails to account for category-specific differences, resulting in performance degradation (e.g., CDifW-CDifW, DisW-DisW).

CDifW: Confidence + Difficulty score
CDifW: Confidence + Distribution score
DisW: Distribution score
DifW: Difficulty score

The results of training CPS module using different weights on 10% labeled BTCV dataset.

Methods	Dice (%)	ASD
DisW-DifW	29.8±5.4	28.1±8.5
CDifW-DisW	**29.9±2.7**	**25.3±8.3**
CDifW-CDifW	26.4±2.8	25.2±3.5
CDifW-CDisW	29.4±2.6	27.3±3.4
DisW-DisW	26.2±4.7	32.2±5.0
CDisW-CDisW	28.9±2.6	25.5±4.4

The effect of using two distinct category-wise weights in model training

CDisW: $w_{t,c}^{CDisW} = i_{t,c}^{\beta} \, w_{t,c}^{disw}$

Confidence information score in $w_{t,c}^{cdifw}$

- Using the same weights fails to account for category-specific differences, resulting in performance degradation (e.g., CDifW-CDifW, DisW-DisW).
- CDisW improves performance by considering difficulty and distribution (CDisW-CDisW outperforms DisW-DisW).

CDifW: Confidence + Difficulty score
CDifW: Confidence + Distribution score
DisW: Distribution score
DifW: Difficulty score

The results of training CPS module using different weights on 10% labeled BTCV dataset.

Methods	Dice (%)	ASD
DisW-DifW	29.8±5.4	28.1±8.5
CDifW-DisW	**29.9±2.7**	**25.3±8.3**
CDifW-CDifW	26.4±2.8	25.2±3.5
CDifW-CDisW	29.4±2.6	27.3±3.4
DisW-DisW	26.2±4.7	32.2±5.0
CDisW-CDisW	28.9±2.6	25.5±4.4

The effect of using two distinct category-wise weights in model training

CDisW: $w_{t,c}^{CDisW} = i_{t,c}^{\beta} \, w_{t,c}^{disw}$

Confidence information score in $w_{t,c}^{cdifw}$

- Using the same weights fails to account for category-specific differences, resulting in performance degradation (e.g., CDifW-CDifW, DisW-DisW).
- CDisW improves performance by considering difficulty and distribution (CDisW-CDisW outperforms DisW-DisW).
- CDifW-CDisW has lower heterogeneity due to the introduction of confidence on both sides, resulting in lower accuracy compared to CDifW-DisW.

CDifW: Confidence + Difficulty score
CDifW: Confidence + Distribution score
DisW: Distribution score
DifW: Difficulty score

The results of training CPS module using different weights on 10% labeled BTCV dataset.

Methods	Dice (%)	ASD
DisW-DifW	29.8±5.4	28.1±8.5
CDifW-DisW	**29.9±2.7**	**25.3±8.3**
CDifW-CDifW	26.4±2.8	25.2±3.5
CDifW-CDisW	29.4±2.6	27.3±3.4
DisW-DisW	26.2±4.7	32.2±5.0
CDisW-CDisW	28.9±2.6	25.5±4.4

異質性の高い重みほど望ましい
The higher the heterogeneity of the weights, the more desirable it is.

Compasion with other strong augmentation methods

- Our method considers category-wise imbalance, making it superior to other strong data augmentation methods.
- ClassMix augmented the images without considering the category-wise weights, caused performance reduce

Comparsion with other strong data augmentations on 10% labeled BTCV dataset. All the experiments are applied CDifW-DisW

Methods	Dice (%)	ASD
CDifW-DisW	29.9±2.7	25.3±8.3
CutMix [1]	31.5±2.6	20.4±5.8
CutOut [2]	30.9±3.7	24.5±6.8
ClassMix[3]	29.3±8.3	33.1±7.0
Ours	**35.7±1.0**	**18.2±4.3**

It is crucial to perform image augmentation targeted at category imbalance.

[1] DeVries, T., & Taylor, G. W. (2017). Improved Regularization of Convolutional Neural Networks with Cutout. arXiv preprint arXiv:1708.04552.
[2] Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019). CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6023-6032.
[3] Olsson, V., Tranheden, W., Pinto, J., & Svensson, L. (2021). ClassMix: Segmentation-based Data Augmentation for Semi-Supervised Learning. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1369-1378.

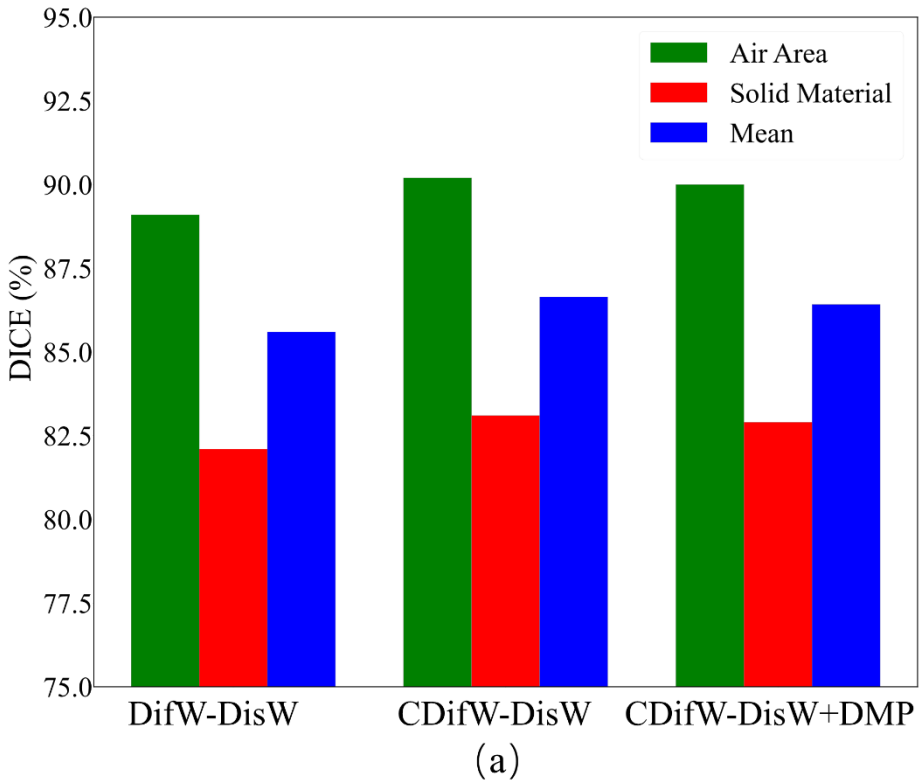
Performance on balanced dataset

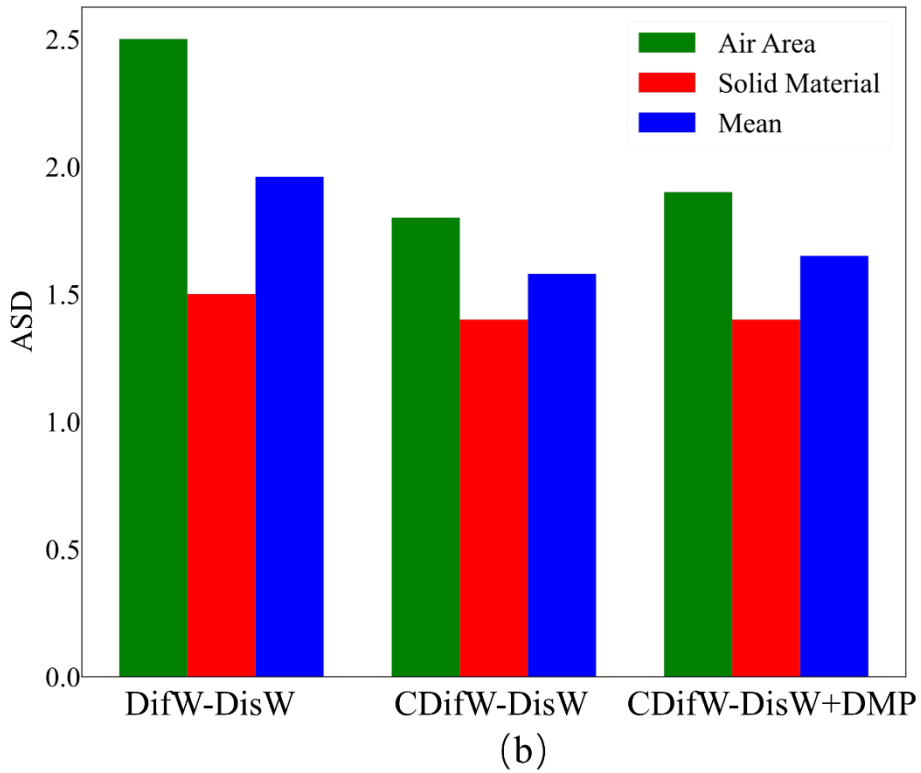
[1] Long, J.R., Frew, M.I., Brazaitis, M.P.: Virtual colonoscopy in the US army: current utilization at the Walter Reed Army Medical Center. Abdominal Imaging 36, 149–152 (2011). Springer

- Colon Segmentation Task (based on WRAMC [1])
- 2 balance categories: Air area and Solid Material
- 10 Labeled cases, 3-fold cross validation

Our CDifW-DisW is better than DifW-DisW.

After using DMP, **a decline in accuracy was observed.**

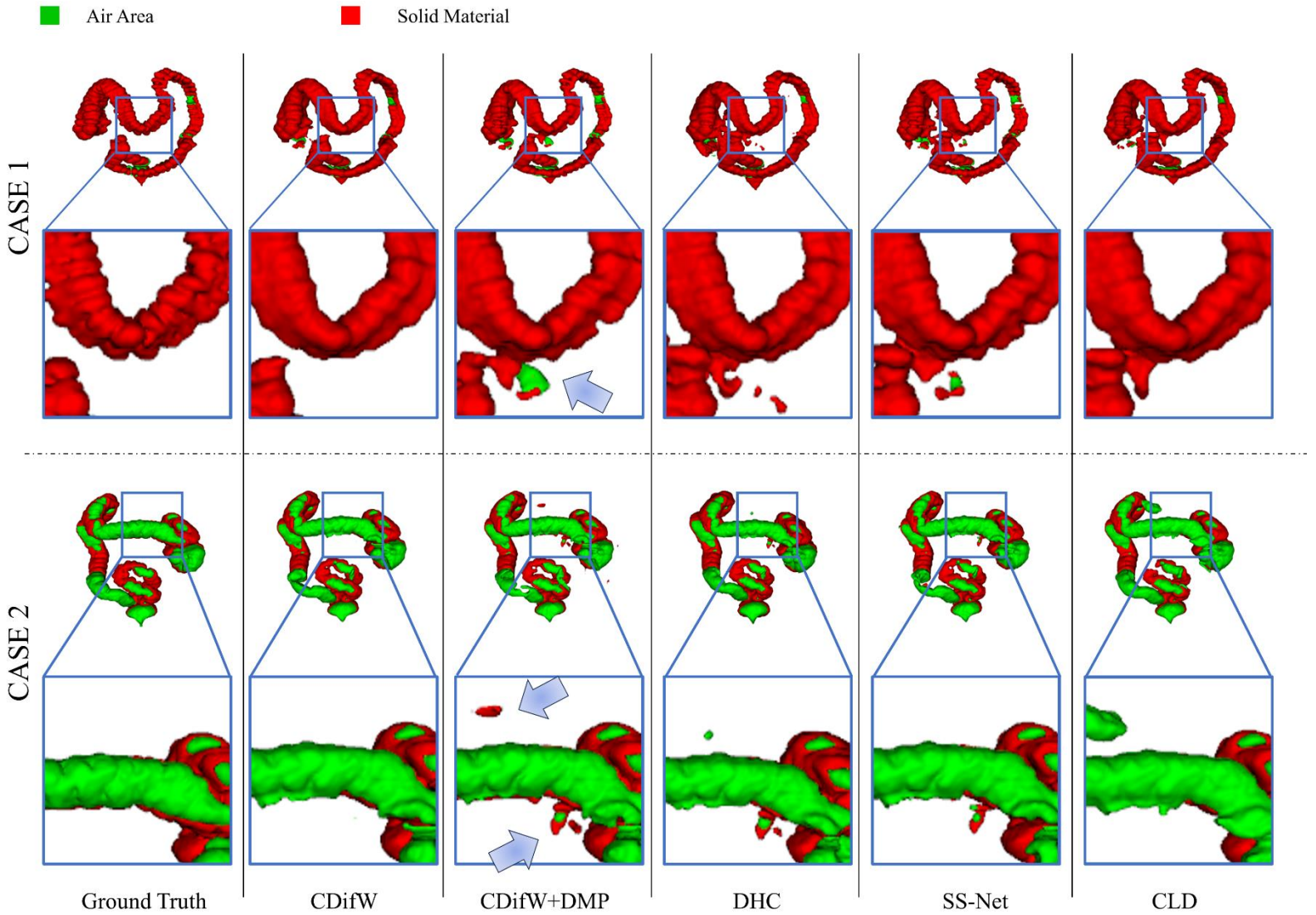




Performance on balanced dataset

Applying the DMP module (CDifW-DisW+DMP) to the balanced dataset likely compromises some spatial information, resulting in reduced performance.

On a simple and balanced dataset, overly strong data augmentation may not be necessary.



Discussion

◆ What issues did we face in this study?

- ➢ CT segmentation requires extensive pixel-level annotated data, which is high-cost.
- ➢ The category-wise weight is not stable, may cased performance increasing
- ➢ Augmentation methods don’t consider class imbalance, leading to poor performance for challenging categories.

Discussion

- What issues did we face in this study?
  - 😟 CT segmentation requires extensive pixel-level annotated data, which is high-cost.
  - 😕 The category-wise weight is not stable, may cased performance increasing
  - 😕 Augmentation methods don't consider class imbalance, leading to poor performance for challenging categories.
- What methods we proposed to address above issues?
  - Proposed **Confidence-Difficulty Weight (CDifW)** to balance training across classes based on confidence and Dice score.
  - Introduced **Double-Mix Pseudo-label Framework (DMPF)** to augment images based on class distribution and difficulty, enhancing segmentation for challenging categories.

Discussion

- What issues did we face in this study?
  - CT segmentation requires extensive pixel-level annotated data, which is high-cost.
  - The category-wise weight is not stable, may cased performance increasing
  - Augmentation methods don't consider class imbalance, leading to poor performance for challenging categories.
- What methods we proposed to address above issues?
  - Proposed **Confidence-Difficulty Weight (CDifW)** to balance training across classes based on confidence and Dice score.
  - Introduced **Double-Mix Pseudo-label Framework (DMPF)** to augment images based on class distribution and difficulty, enhancing segmentation for challenging categories.
- What are the limitations of our method?
  - DMP module introduces noise in balanced datasets by potentially disrupting spatial information.
  - The practicality of the results needs to be evaluated by clinicians.

Conclusions and Foreseeing

Summary of the topics

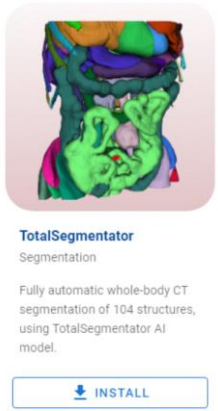
- High annotation cost in data annotation for medical image segmentation
- Since endoscopic data differs from CT data, we proposed two approaches tailored to each data type, to solve this problem
  - Train laparoscopic video segmentation model with limited Pixel-Level annotated data and abundant category-Level annotated data (Topic 1)
  - Train CT segmentation model with limited Pixel-Level annotated data and abundant non-annotated data (Topic 2)
- Provide successful solutions to two important tasks in medical image segmentation with low annotations cost

Main limitations:

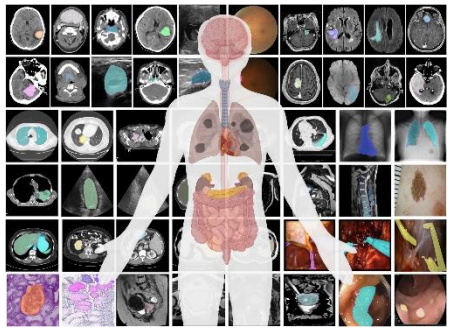
- Still need annotated data in model training (Topic 1)
- Performance reduce in simple and balanced dataset (Topic 2)

Issues to be Solved

- Can we finetuning the large pre-trained models (Totalsegmentator, MedSAM, etc.) to further reduce the cost of required annotations?



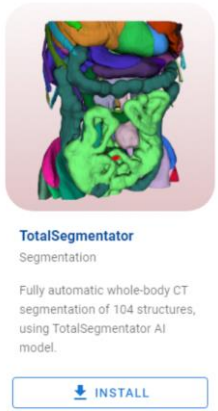
Totalsegmentator



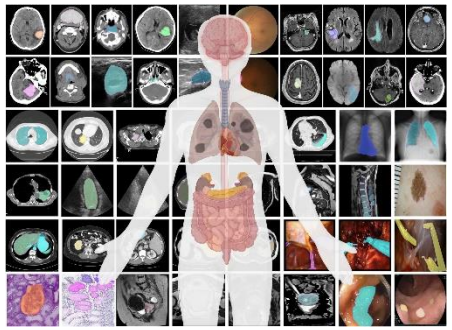
MedSAM

Issues to be Solved

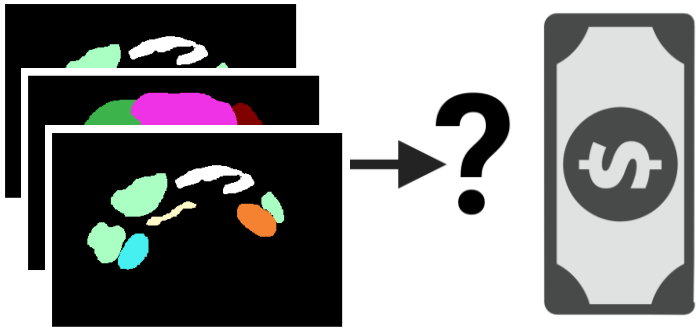
- Can we finetuning the large pre-trained models (Totalsegmentator, MedSAM, etc.) to further reduce the cost of required annotations?
- How can annotation cost be quantified and used as a standard for model evaluation?



Totalsegmentator



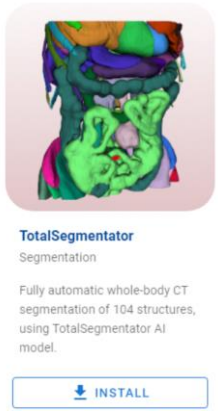
MedSAM



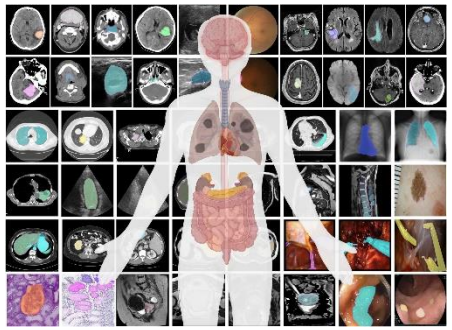
Quantification of annotation cost

Issues to be Solved

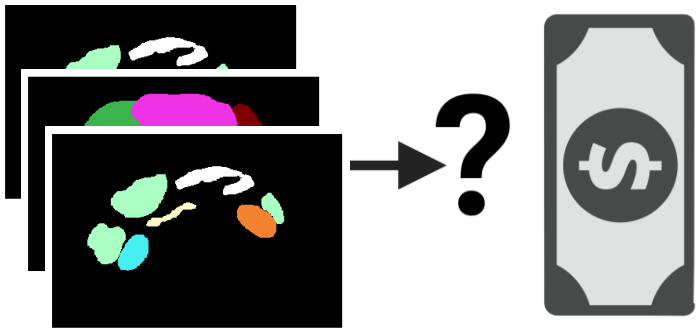
- Can we finetuning the large pre-trained models (Totalsegmentator, MedSAM, etc.) to further reduce the cost of required annotations?
- How can annotation cost be quantified and used as a standard for model evaluation?
- The results should be confirmed by clinical doctors to evaluate the clinical significance of our method.



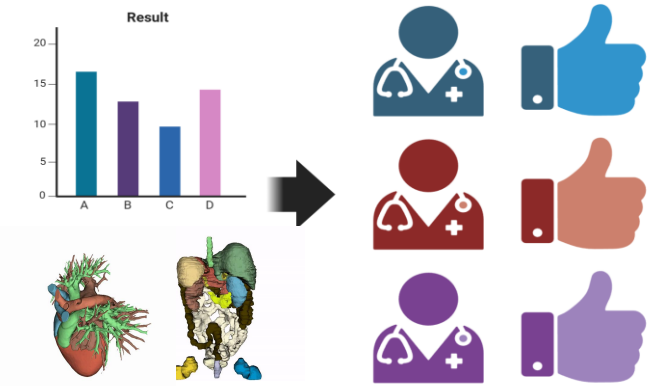
Totalsegmentator



MedSAM



Quantification of annotation cost



Evaluation of clinical doctors

Publication List

- 学術雑誌論文（査読付き）
- [1] Zhang, Luyang, Yuichiro Hayashi, Masahiro Oda, and Kensaku Mori. "Towards better laparoscopic video segmentation: A Class-Wise Contrastive Learning Approach with Multi-Scale Feature Extraction." Healthcare Technology Letters 11.2-3:126-136 (2024)

- [2] Zhang, Luyang, Yuichiro Hayashi, Masahiro Oda, and Kensaku Mori. "Double-Mix Pseudo-Label Framework: Enhancing Semi-Supervised Segmentation on Category-Imbalanced CT Volumes." International Journal of Computer Assisted Radiology and Surgery (accepted)
- その他の研究業績
- [3] Luyang Zhang, Yuichiro Hayashi, Masahiro Oda, Kensaku Mori. "Towards Better Laparoscopic Video Segmentation: A Class-Wise Contrastive Learning Approach with Multi-Scale Feature Extraction." Joint MICCAI workshop 2023, AE-CAI/CARE/OR2.0 (2023)

- [4] 張 路暘，小田 昌宏，森 健策．〝腹部臓器CTセグメンテーションのための角度許容位置ベース対比損失．〞2023年度日本生体医工学会東海支部大会（2023）

- [5] 張 路暘，小田 昌宏，森 健策．〝ランダム周波数マスキングと疑似ラベル微調整を用いたCT像からの多臓器半教師ありセグメンテーション．〞第33回日本コンピュータ外科学会大会（2024）

- [6] Zhang, Luyang, Masahiro Oda, and Kensaku Mori. "SemiOrth: A Novel Orthogonal Dual Network Architecture for Enhanced Semi-Supervised Medical Image Segmentation." SPIE Medical Imaging (2024) (accepted)

ご清聴いただきありがとうございます

If you are interested in my work, please visit