CauSSL: Causality-inspired Semi-supervised Learning for Medical Image Segmentation

Juzheng Miao[1], Cheng Chen[2], Furui Liu[3]*, Hao Wei[4], Pheng-Ann Heng[1,5]
[1]Department of Computer Science and Engineering, The Chinese University of Hong Kong
[2]Center for Advanced Medical Computing and Analysis,
Harvard Medical School and Massachusetts General Hospital
[3]Zhejiang Lab, [4]Department of Biomedical Engineering, The Chinese University of Hong Kong
[5]Institute of Medical Intelligence and XR, The Chinese University of Hong Kong

1. Implementation Details of MCCauSSL

We use the setting of two networks/branches to illustrate our proposed method in our manuscript, but it can be easily extended to more networks or branches in the co-training framework. MC-Net+ [14] extends MC-Net [15] with 3 decoders using the same convolutional architecture but different upsampling strategies on top of a shared encoder. In this section, we'll show how to extend the network dependence for two networks to the MC-Net+ with 3 decoder branches and explain the details of MCCauSSL.

1.1. Network Dependence of MC-Net+

The dependence between two convolutional layers A and B is defined as Equation 1 in our manuscript:

$$L_{in}(A, B; G_B) = \frac{1}{C_{out}} \sum_{i=1}^{C_{out}} \left(\frac{\boldsymbol{v}_{A,i} \cdot \boldsymbol{q}_{B,i}}{|\boldsymbol{v}_{A,i}| \times |\boldsymbol{q}_{B,i}|} \right)^2 \quad (1)$$
$$\boldsymbol{q}_{B,i} = (G_B \times B)_i$$

where $\boldsymbol{v}_{A,i}$ is the i-th row vector in matrix A, and $\boldsymbol{q}_{B,i}$ is the optimal linear combination vector using the vector group of B that can approximate $\boldsymbol{v}_{A,i}$ as close as possible. G_B is the optimal coefficient matrix whose elements are the optimal linear combination coefficients, with a size of $C_{out} \times C_{out}$.

For the setting of 3 convolutional layers A, B, and C with the same convolutional architecture from the 3 decoders in the MC-Net+ method, the layer dependence is defined as the linear dependence between each layer and the matrix concatenating the other two layers:

$$L_{in}(A, [B, C]; G_{BC}) = \frac{1}{C_{out}} \sum_{i=1}^{C_{out}} \left(\frac{\boldsymbol{v}_{A,i} \cdot \boldsymbol{q}_{[B,C],i}}{|\boldsymbol{v}_{A,i}| \times |\boldsymbol{q}_{[B,C],i}|} \right)^2$$
$$\boldsymbol{q}_{[B,C],i} = (G_{BC} \times [B, C])_i \quad (2)$$

where $[B, C]$ is the extended matrix by concatenating B and C along the row dimension. Thus, the size of the corresponding optimal coefficient matrix G_{BC} is $C_{out,A} \times (C_{out,B} + C_{out,C}) = C_{out} \times (2C_{out})$, considering A, B, and C have the same size of $C_{out} \times d$.

We then define the dependence on the network level among three branches with the same convolutional architecture by taking the average over all the convolutional layers:

$$L_{in}\left(\theta_1, [\theta_2, \theta_3]; G_{23}\right) = \frac{1}{\#\text{ layers}} \sum_{i=1}^{\#\text{ layers}} L_{in}\left(\theta_{1,i}, [\theta_{2,i}, \theta_{3,i}]; G_{23,i}\right), \quad (3)$$

where $\theta_{1,i}$, $\theta_{2,i}$, $\theta_{3,i}$, and $G_{23,i}$ are the weight parameters in the format of matrices of each branch and optimal coefficient matrix of the i-th convolutional layer, respectively. Only convolutional layers are considered.

1.2. MCCauSSL Algorithm

Based on the network dependence definition, the training procedure of MCCauSSL is shown in Algorithm 1, where L_{total} is defined as:

$$\begin{aligned} L_{total,1} &= L_{s,1} + \lambda_1 L_{u,1} + \lambda_2 L_{in}(\theta_1, [\bar{\theta_2}, \bar{\theta_3}]; G_{23}) \\ L_{total,2} &= L_{s,2} + \lambda_1 L_{u,2} + \lambda_2 L_{in}(\theta_2, [\bar{\theta_1}, \bar{\theta_3}]; G_{13}) \\ L_{total,3} &= L_{s,3} + \lambda_1 L_{u,3} + \lambda_2 L_{in}(\theta_3, [\bar{\theta_1}, \bar{\theta_2}]; G_{12}) \end{aligned} \quad (4)$$

$L_{s,i}$ and $L_{u,i}$ indicate the supervised loss and unsupervised loss, respectively and λ_1, λ_2 are balancing coefficients. $\bar{\theta_1}$, $\bar{\theta_2}$, and $\bar{\theta_3}$ represent weights copy without gradient flows.

2. Implementation Details on Each Dataset

2.1. ACDC Dataset

The ACDC dataset[1] [4] contains 200 labeled cine MRI scans from 100 patients for training and 100 unlabeled scans

*Corresponding author.

[1]https://acdc.creatis.insa-lyon.fr/

Algorithm 1 Pseudocode of MCCauSSL

Input: labeled data $\mathcal{L}$, unlabeled data $\mathcal{U}$ and hyperparameters λ_1 and λ_2.

Output: Three independent decoder branches parameterized by θ_1, θ_2 and θ_3, and a shared encoder parameterized by θ.

Randomly initialize the network weights θ_1, θ_2, θ_3, θ and linear coefficients G_{12}, G_{13}, G_{23}. // initialization
$i = 0$ // iteration number
while $i \leq$maximum iterations **do** // training
  for j=1:s_{max} **do** // maximize
    Fix θ_1, θ_2, θ_3, θ. Update G_{12}, G_{13}, G_{23} by
    maximizing $L_{in}(\theta_3, [\bar{\theta_1}, \bar{\theta_2}]; G_{12})$,
    $L_{in}(\theta_2, [\bar{\theta_1}, \bar{\theta_3}]; G_{13})$ & $L_{in}(\theta_1, [\bar{\theta_2}, \bar{\theta_3}]; G_{23})$.
  for j=1:s_{min} **do** // minimize
    Fix G_{12}, G_{13}, G_{23}. Update θ_1, θ_2, θ_3, θ by
    minimizing $L_{total,1}$, $L_{total,2}$, and $L_{total,3}$ using
    Equation 4.
    $i = i + 1$.
Return θ_1, θ_2, θ_3, and θ.

from 50 patients for testing. Three regions of interest are delineated for the segmentation task: the right ventricle (RV) cavity, the myocardium (Myo), and the left ventricle (LV) cavity. Following [14], only the training dataset was used in our experiments, which was randomly split at the patient level, with 70 patients for training, 10 for validation, and 20 for testing. The volume thickness ranges from 5 to 10 mm while the spatial resolution is between 1.34 and 1.68 mm^2/pixel. We directly leverage the processed data by Luo *et al*. [7], where the 3D volume data are normalized to [0,1] first and then sliced to 2D images.

On the ACDC dataset, we employed 2D U-Net [11] as the segmentation backbone by treating the 3D volume slice by slice and developed our method on the public source code of Luo *et al*. [7] with the same training setting. We used an SGD optimizer to update network weights and set the weight decay and the momentum as 0.0001 and 0.9, respectively. We initialized the learning rate to be 0.01 and updated it following $lr\,(t) = 0.01*\left(1 - \frac{t}{t_{max}}\right)^{0.9}$, where t is the iteration number and t_{max} means the total iteration number, i.e., 30000 on this dataset. The batch size was 24, containing 12 labeled images and 12 unlabeled ones. The coefficient of the algorithmic independence constraint λ_2 was set as 1.0 and 0.2 for CPSCauSSL and MCCauSSL, respectively. In training, we first resized all the sliced images to 256×256 and took the random rotation between [-20°, 20°] and random flip operations as data augmentation. The predictions were scaled back to the original size and stacked into 3D volumes in the testing phase for metric evaluation.

2.2. Pancreas-CT Dataset

The Pancreas-CT dataset[2] [6, 12, 13] collected by the National Institutes of Health Clinical Center contains 82 abdominal contrast enhanced 3D CT scans for pancreas segmentation. The slice thickness varies in the range between 1.5 and 2.5 mm. For data preprocessing, all the volumes were resampled into an isotropic resolution of 1.0 mm in all axes. The intensity values were also clipped to [-125, 275] Hounsfield units (HU) following [8, 14]. Moreover, we cropped the region including the pancreas from the original volume with an enlarged margin of 25 voxels and normalized them to zero mean and unit variance as done in [8, 14]. We adopted the same data split as [8], using 62 and 20 volume data for training and testing, respectively.

On the Pancreas-CT dataset, 3D V-Net [10] was chosen as the baseline backbone. According to [8, 14], the input size of each volume patch was 96×96×96 and the batch size was set as 4, with 2 labeled data. All the networks were trained for 5k iterations using an SGD optimizer with the same settings as above. In addition, we scaled down the learning rate by 0.1 every 2.5k iterations. During inference, a sliding window strategy with a stride of 16 in each axis was utilized following [8, 14]. We empirically set λ_2 as 3.0 for both CPSCauSSL and MCCauSSL on this dataset.

2.3. BraTS'2019 Dataset

BraTS'2019 dataset [1, 2, 3, 9] contains multi-institutional pre-operative MRI scans collected from 335 glioma patients. For each patient, there are four modalities of MRI scans, including T1, T2, Flair, and T1Ce. 3 sub-regions are delineated: 1) the "enhancing tumor" (ET), 2) the "tumor core" (TC), and 3) the "whole tumor" (WT). Following Xu *et al*. [16], we only used the Flair images for whole tumor segmentation in our experiments to evaluate the performance of our proposed semi-supervised learning method, since the WT segmentation is able to describe the complete extent of the disease and is critical to brain surgery of low-grade glioma [17]. Also, it is typically depicted by hyper-intense signals in FLAIR. We directly used the data processed by Luo *et al*. [7], where 250, 25, and 60 samples are used for training, validation, and testing, respectively.

We developed our method on the public source code of Luo *et al*. [7] on this dataset, where a 3D U-Net [5] with an input patch size of 96×96×96 was adopted as the backbone. The batch size was set as 4, containing 2 labeled data and 2 unlabeled ones. All the networks were optimized for 30000 iterations using an SGD optimizer, whose initial learning rate was 0.01. The learning rate was updated according to the same way on the ACDC dataset. The coefficient of the algorithmic independence constraint λ_2 was set as 0.05 and 1.0 for CPSCauSSL and MCCauSSL, re-

[2]https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT

spectively. During training, random rotation, flip, and crop augmentations were taken. For the inference, we utilized a sliding window strategy based on the stride of 64 in all axes.

References

[1] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13, 2017.

[2] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018.

[3] Spyridon (Spyros) Bakas. Brats miccai brain tumor dataset, 2020.

[4] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging*, 37(11):2514–2525, 2018.

[5] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.

[6] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6):1045–1057, 2013.

[7] Xiangde Luo. SSL4MIS. `https://github.com/HiLab-git/SSL4MIS`, 2020.

[8] Xiangde Luo, Jieneng Chen, Tao Song, and Guotai Wang. Semi-supervised medical image segmentation through dual-task consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8801–8809, 2021.

[9] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[10] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016.

[11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[12] Holger R Roth, Amal Farag, E Turkbey, Le Lu, Jiamin Liu, and Ronald M Summers. Data from pancreas-ct. the cancer imaging archive. *IEEE Transactions on Image Processing*, 2016.

[13] Holger R Roth, Le Lu, Amal Farag, Hoo-Chang Shin, Jiamin Liu, Evrim B Turkbey, and Ronald M Summers. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 556–564. Springer, 2015.

[14] Yicheng Wu, Zongyuan Ge, Donghao Zhang, Minfeng Xu, Lei Zhang, Yong Xia, and Jianfei Cai. Enforcing mutual consistency of hard regions for semi-supervised medical image segmentation. *arXiv preprint arXiv:2109.09960*, 2021.

[15] Yicheng Wu, Minfeng Xu, Zongyuan Ge, Jianfei Cai, and Lei Zhang. Semi-supervised left atrium segmentation with mutual consistency training. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 297–306. Springer, 2021.

[16] Zhe Xu, Yixin Wang, Donghuan Lu, Lequan Yu, Jiangpeng Yan, Jie Luo, Kai Ma, Yefeng Zheng, and Raymond Kai-yu Tong. All-around real label supervision: Cyclic prototype consistency learning for semi-supervised medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(7):3174–3184, 2022.

[17] Ramy A Zeineldin, Mohamed E Karar, Jan Coburger, Christian R Wirtz, and Oliver Burgert. Deepseg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance flair images. *International journal of computer assisted radiology and surgery*, 15:909–920, 2020.