Multi-Scale Cross Contrastive Learning for Semi-Supervised Medical Image Segmentation

Qianying Liu[1], Xiao Gu[2], Paul Hendersons[1], Fani Deligianni[1]

1. School of Computing Science, University of Glasgow 2. Department of Computing, Imperial College London

Code is on GitHub

Background and Purpose

- Semi-supervised learning has demonstrated potential in medical image segmentation by utilizing unlabelled data.
- However, they do not explicitly capture high-level semantic relations between distant regions.
- Jointly train CNN and Transformer
- Regularising their features to be semantically consistent across different scales based on crossed labels.
- Code is available on GitHub (QR Code).

Supervision loss functions

- cross pseudo supervision loss (unlabelled data)

$$\mathcal{L}_{cps(cnn)} = \mathcal{L}_{dice}(P^u_{cnn}, Y^u_{tra}), \quad \mathcal{L}_{cps(tra)} = \mathcal{L}_{dice}(P^u_{tra}, Y^u_{cnn}).$$

- Multi-Scale Contrastive loss(whole data): $\mathcal{L}_{cl} = (\mathcal{L}_{cl\_1} + \ldots + \mathcal{L}_{cl\_n})$
, each scale $\mathcal{L}_{bcl}$ as $\mathcal{L}_{cl\_i}$

Balanced contrastive loss:

$$\mathcal{L}_{bcl} = -\frac{1}{|A|}\sum_{a_i \in A}\frac{1}{|A_y|-1}\sum_{p \in A_y \setminus \{i\}} \log \frac{\exp(a_i \cdot a_p/\tau)}{\sum_{j \in Y_A}\frac{1}{|A_j|}\sum_{a_k \in A_j}\exp(a_i \cdot a_k/\tau)},$$

- Total loss function:

$$\mathcal{L}_{cnn} = \mathcal{L}_{sup(cnn)} + w_{cps}\mathcal{L}_{cps(cnn)} + w_{cl}\mathcal{L}_{cl} \qquad \mathcal{L}_{tra} = \mathcal{L}_{sup(tra)} + w_{cps}\mathcal{L}_{cps(tra)} + w_{cl}\mathcal{L}_{cl}$$

MCSC framework

- On the output level, two losses:
  1 supervision loss $\mathcal{L}_{sup}$ (yellow dashed lines in Figure 1) between the segmentation predictions and the limited labelled data.
  2. cross pseudo supervision loss $\mathcal{L}_{cps}$ (green dashed lines) between the predictions and the pseudo labels in a cross teaching manner.
- On the feature level: multi-scale cross contrastive loss $\mathcal{L}_{cl}$ (black dashed lines) to enhance feature consistency/distinguishability of feature of the same /different categories across the whole data (labelled and unlabelled).

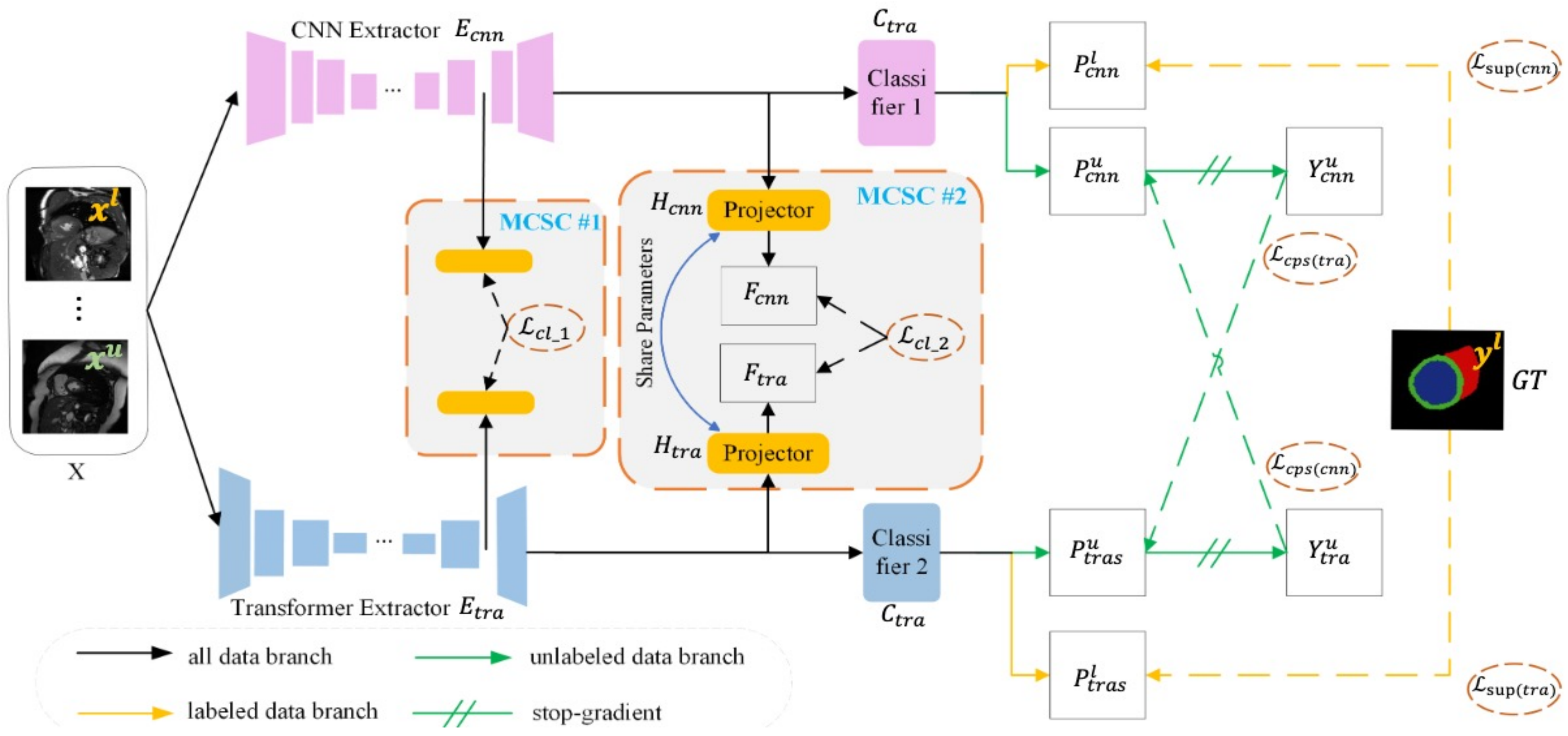


Fig. The overall architecture of our MCSC framework.

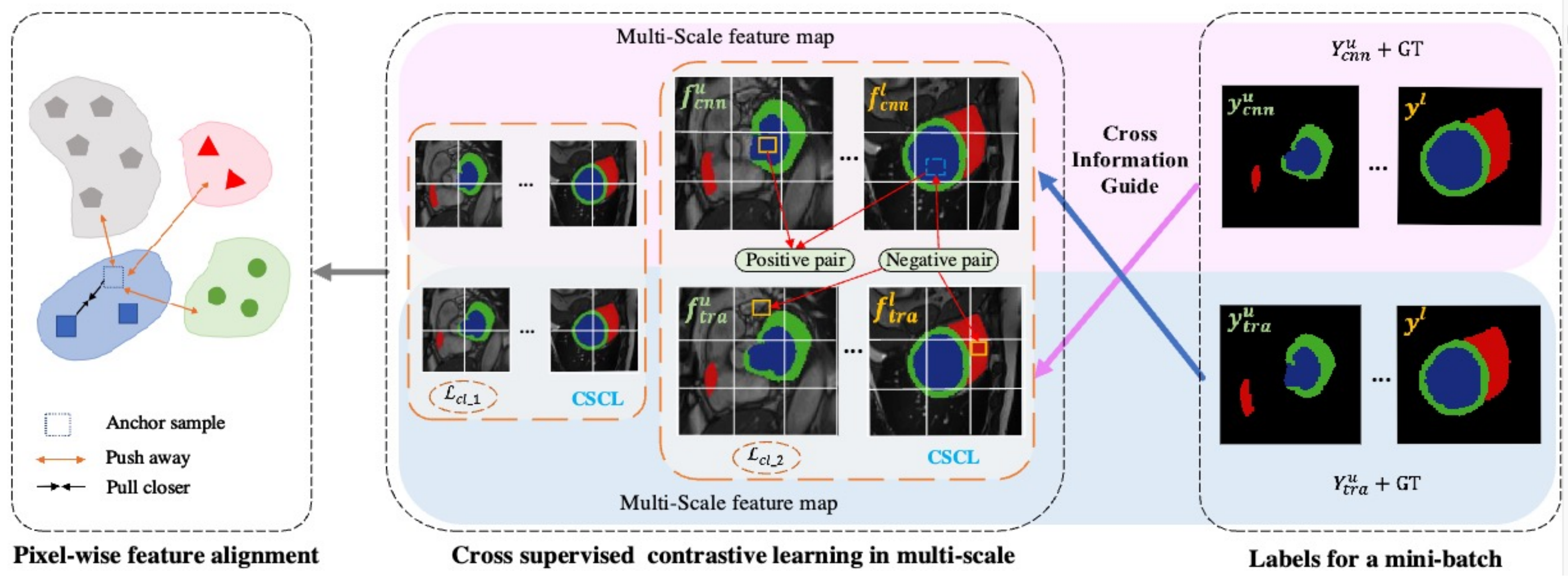


Fig. CST Multi-scale cross supervised contrastive learning. Pseudo labels from cross-teaching (right) and ground-truth, and used to guide contrastive loss.

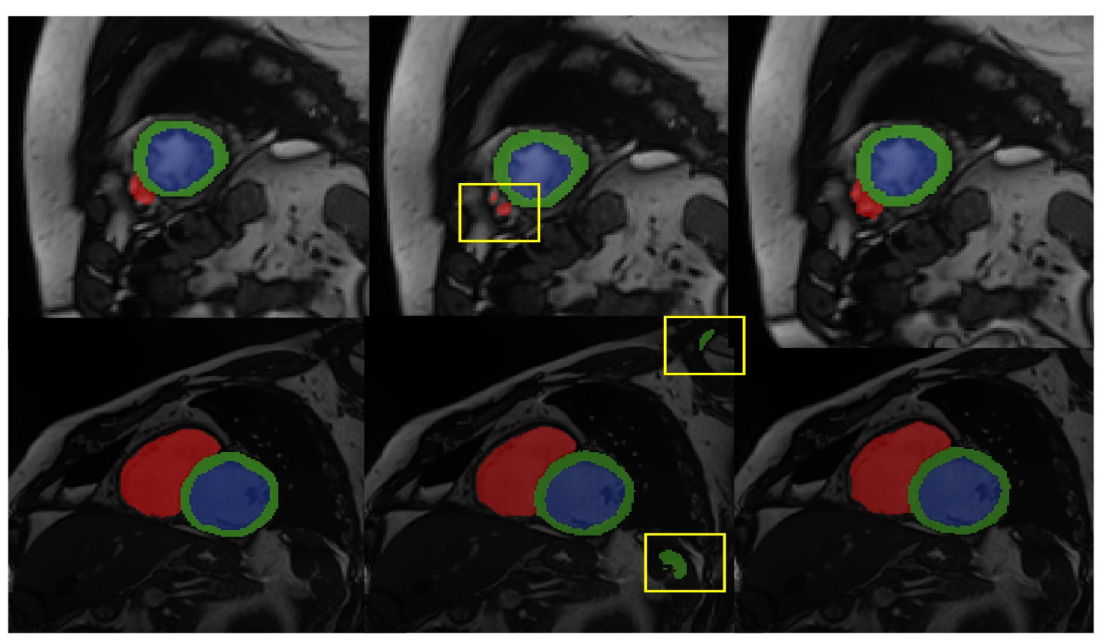
Results

- MCSC outperforms SOTA by more than 3.0% in Dice on two benchmarks.

ACDC 200 short-axis cardiac MRI, left ventricle (LV), myocardium (Myo), and right ventricle (RV).
Synapse, abdominal CT, aorta, gallbladder, spleen, left/right kidney, liver, pancreas and stomach.

Labelled	Methods	Mean DSC↑	Mean HD↓
70 cases (100%)	UNet-FS	91.7	4.0
	BATFormer [16]	92.8	8.0
7 cases (10%)	UNet-LS	75.9	10.8
	CCT [19]	84.0	6.6
	CPS [8]	85.0	6.6
	CTS [17]	86.4	8.6
	MCSC (Ours)	**89.4**	**2.3**
3 cases (5%)	UNet-LS	51.2	31.2
	CCT [19]	58.6	27.9
	CPS [8]	60.3	25.5
	CTS [17]	65.6	16.2
	MCSC (Ours)	**73.6**	**10.5**
1 case	UNet-LS	26.4	60.1
	CTS [17]	46.8	36.3
	MCSC (Ours)	**58.6**	**31.2**

Tab. Segmentation results on the ACDC dataset.

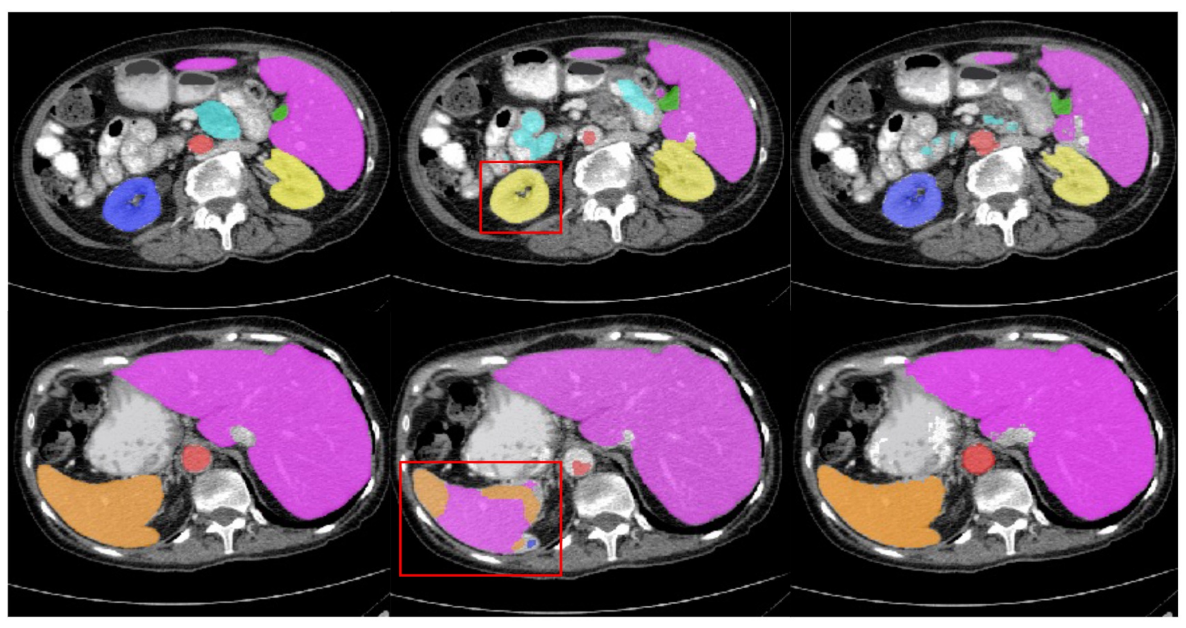


GT CTS Ours

Fig. Visualizations on the ACDC.

Labelled	Methods	DSC↑	HD↓
18 cases(100 %)	UNet-FS	75.6	42.3
	nnFormer [39]	86.6	10.6
4 cases(20 %)	UNet-LS	47.2	122.3
	CCT [19]	51.4	102.9
	CPS [8]	57.9	62.6
	CTS [17]	64.0	56.4
	MCSC (Ours)	**68.5**	**24.8**
2 cases(10 %)	UNet-LS	45.2	55.6
	CCT [19]	46.9	58.2
	CPS [8]	48.8	65.6
	CTS [17]	52.0	63.7
	MCSC (Ours)	**61.1**	**32.6**

Tab. Segmentation results on the Synapse dataset.



GT CTS Ours

Fig. Visualizations on the Synapse.

Ablations

SCL	DB	CroLab	Balanced	MulSca	Unet DSC↑	Unet HD↓	Transformer DSC↑	Transformer HD↓
					86.40	8.6	85.22	5.1
✓	✓				87.50	7.4	86.02	4.5
✓	✓	✓			88.23	3.4	86.13	3.2
✓		✓	✓		88.80	4.6	86.53	**2.4**
✓		✓	✓	✓	**89.38**	**2.3**	**87.28**	3.5

Tab. Ablation study for the primary components of our model. SCL, supervised local contrastive loss. DB, discard background pixels as anchor. CroLab, cross label information of two models to select contrastive sample. Balanced, average the instances of each class in denominator of SCL. MulSca, multi-scale feature maps.

Branches 256	Branches 56	Branches 28	Mean DSC↑	Mean HD↓
✓			88.80	4.6
	✓		88.88	4.2
		✓	88.39	4.5
✓		✓	**89.38**	**2.3**
✓	✓		88.92	2.9
✓	✓	✓	88.35	4.3

Tab. Ablation on the choice of feature maps for the multi-scale (ACDC, 7 labelled cases).

References

1. Xiangde Luo etal. Semi-supervised medical image segmentation via cross teaching between cnn and transformer. Medical Imaging with Deep Learning, 2022.
2. Wenguan Wang, etal. Exploring cross-image pixel contrast for semantic segmentation. ICCV, 2021.
3. Jianggang Zhu, etal. Balanced contrastive learning for long-tailed visual recognition. CVPR, 2022.

Acknowledgements

We acknowledge funding by China Scholarship Council, EPSRC (EP/W01212X/1) and Royal Society (RGS/R2/212199).

University of Glasgow | School of Computing Science

Contact: 2665227L@student.gla.ac.uk