Positional Contrastive Learning for Volumetric Medical Image Segmentation

Dewen Zeng[1], Yawen Wu[2], Xinrong Hu[1], Xiaowei Xu[3], Haiyun Yuan[3], Meiping Huang[3], Jian Zhuang[3], Jingtong Hu[2], Yiyu Shi[1]

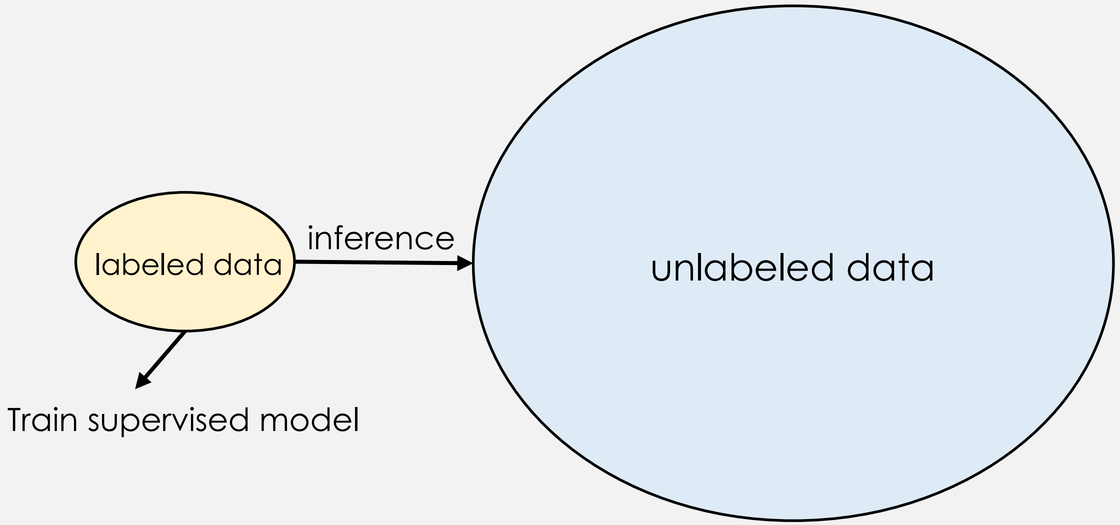
[1] University of Notre Dame

[2] University of Pittsburgh

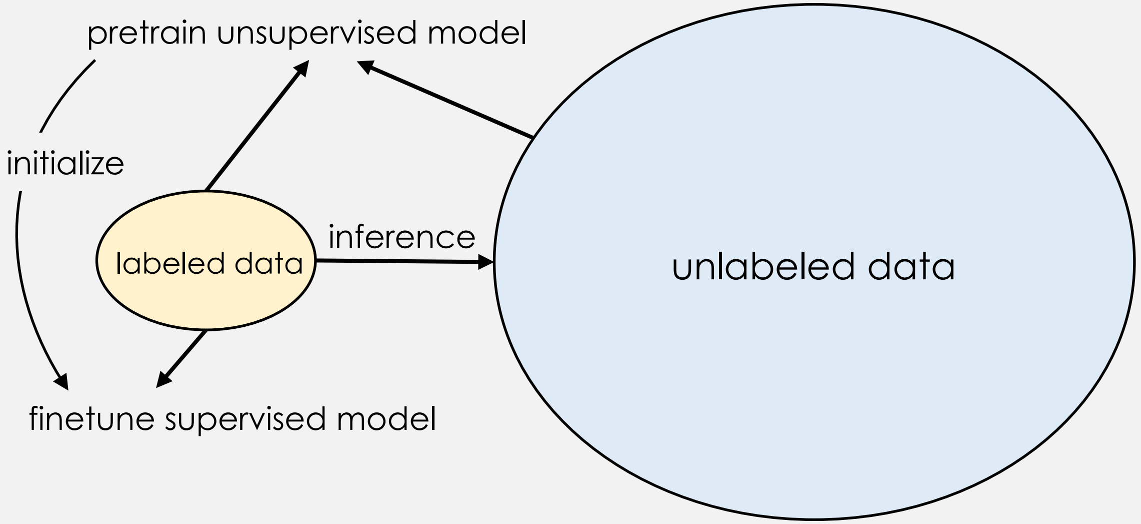
[3] Guangdong Provincial People's Hospital

Learning from unlabeled medical data is important

- SOTA models in medial image domain highly relies on labeled data.
- Medical field has a large amount of unlabeled data (CXR, CT, MRI, Ultrasound, etc. generated every day).
- Labeling costs lots of time and needs domain expertise.



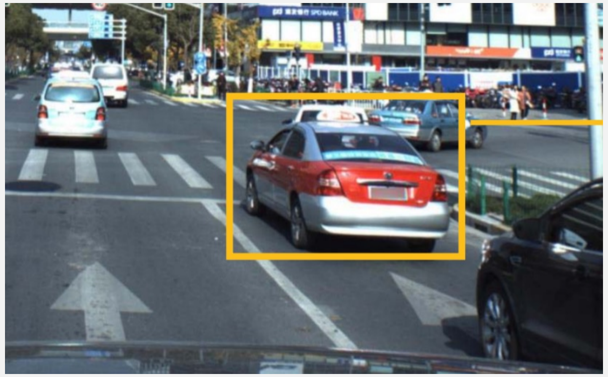
Supervised learning



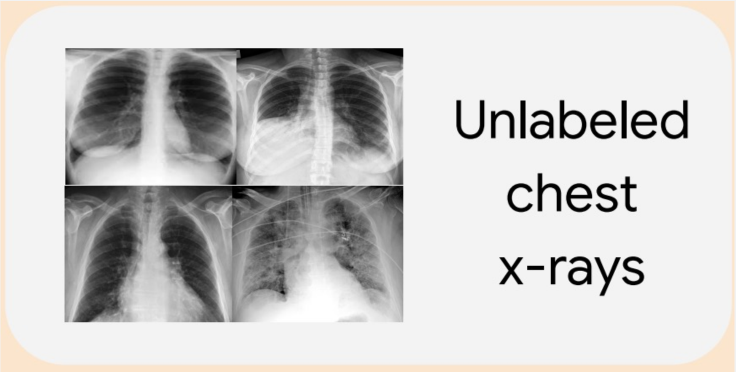
Learning from unlabeled data

Self-supervised visual representation learning

- *Pretext tasks based*
  - Generation-based methods: image inpainting, image colorization
  - Context-based methods: image rotation, patch ordering, patch distance, jigsaw puzzle
- *Contrastive learning*
  - Learning by comparing among different samples, pushing “similar” inputs together and pulling “dissimilar” inputs apart.
  - MoCo
  - SimCLR
  - SWaV

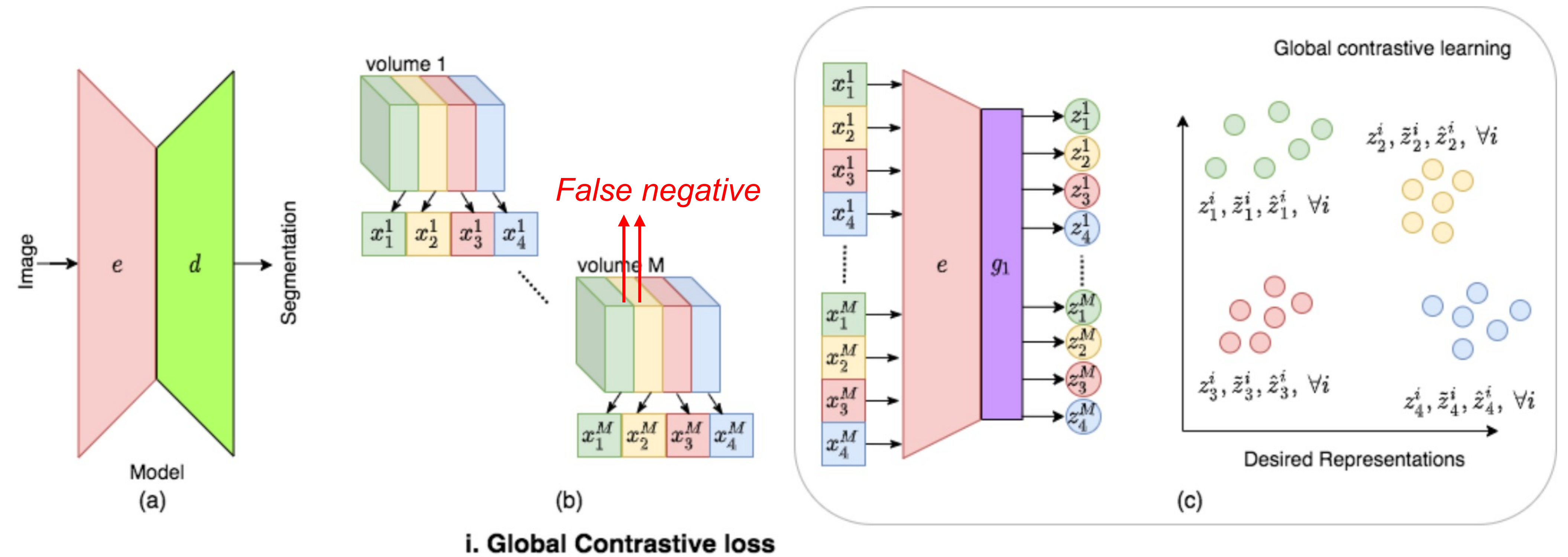
natural image classification and object detection



medical image classification[1]

[1] Azizi, Shekoofeh, et al. "**Big self-supervised models advance medical image classification.**" arXiv preprint arXiv:2101. (2021).

Contrastive learning for medical image segmentation

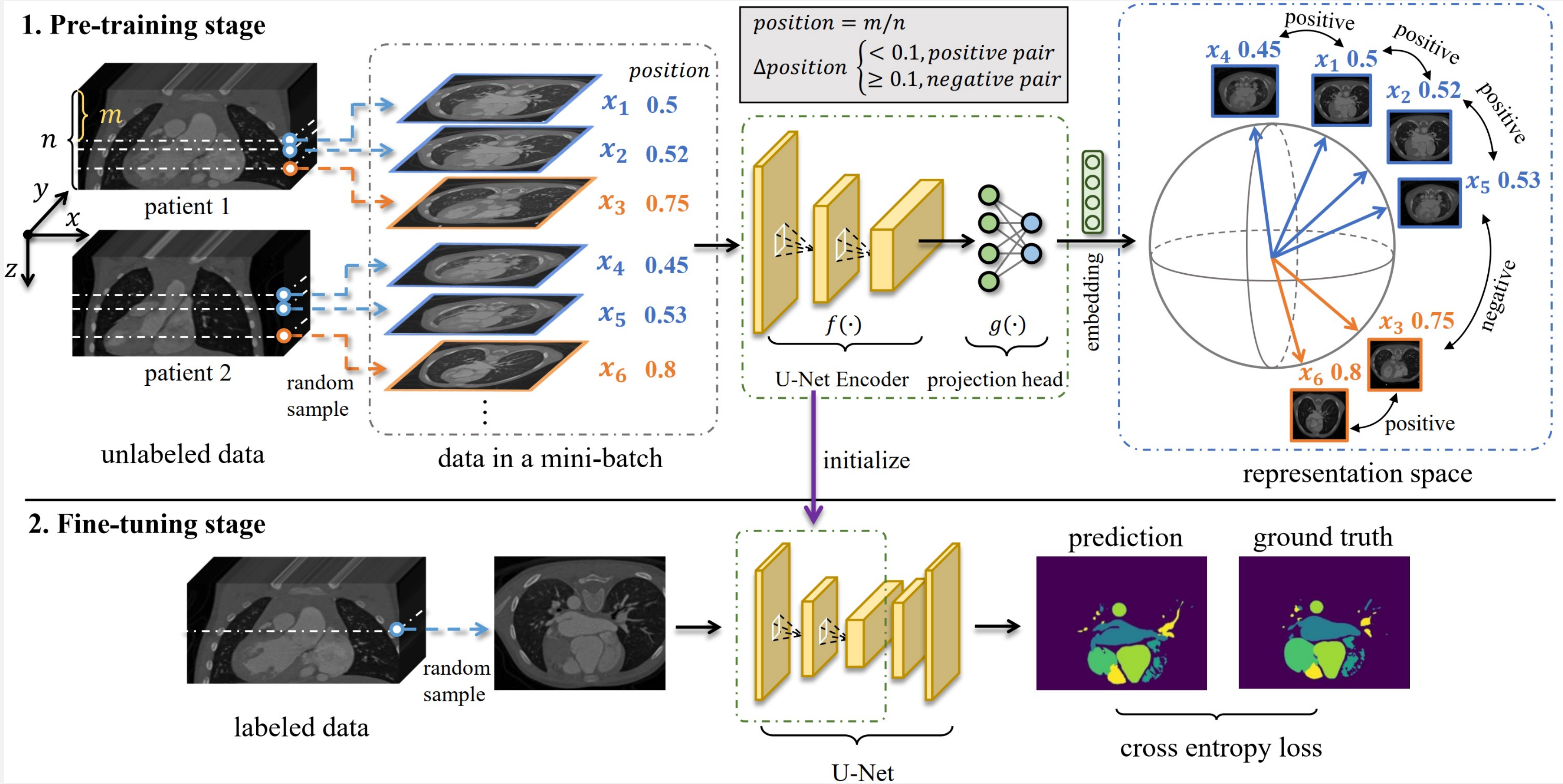


Dividing each volume into four partitions, slices from the same partition are positive pairs.

Limitation: The last a few slices of a partition can be very similar to the first a few slices of the next partition as they are naturally adjacent, causing false negative pairs.

Chaitanya, Krishna, et al. "**Contrastive learning of global and local features for medical image segmentation with limited annotations**." arXiv preprint arXiv:2006.10511 (2020).

Positional contrastive learning mitigates false positive



- Based on SimCLR
- Slices with small position difference are positive pair
- There could be multiple positive pairs for each data point in the mini-batch

Experimental Results

Semi-supervised setting

U-Net encoder is pretrained on CHD (CT for whole heart segmentation) or ACDC (MRI) dataset and then finetuned with limited number of labeled data. (M is the number of finetuning patients)

CHD (68 patients in total)							
Method	M=2	M=6	M=10	M=15	M=20	M=30	M=51
Random	0.184(.06)	0.508(.06)	0.584(.05)	0.627(.05)	0.658(.04)	0.693(.04)	0.754(.02)
Rotation [6]	0.171(.06)	0.488(.07)	0.575(.04)	0.625(.04)	0.651(.04)	0.691(.04)	0.749(.03)
PIRL [14]	0.196(.07)	0.504(.08)	0.617(.05)	0.658(.03)	0.674(.04)	0.714(.04)	0.761(.03)
SimCLR [3]	0.192(.06)	0.515(.06)	0.599(.06)	0.631(.05)	0.666(.05)	0.699(.05)	0.756(.03)
GCL [2]	0.255(.10)	0.564(.04)	0.646(.03)	0.669(.04)	0.697(.04)	0.725(.04)	0.766(.03)
PCL	**0.356(.08)**	**0.600(.06)**	**0.661(.05)**	**0.686(.05)**	**0.716(.04)**	**0.735(.05)**	**0.774(.03)**

ACDC (100 patients in total)							
Method	M=2	M=6	M=10	M=15	M=20	M=30	M=80
Random	0.588(.07)	0.782(.03)	0.840(.03)	0.876(.01)	0.894(.01)	0.909(.01)	0.928(.00)
Rotation [6]	0.572(.08)	0.809(.03)	0.868(.02)	0.886(.01)	0.898(.01)	0.910(.01)	0.925(.00)
PIRL [14]	0.492(.03)	0.823(.04)	0.865(.01)	0.880(.02)	0.896(.02)	0.912(.01)	0.927(.00)
SimCLR [3]	0.352(.06)	0.725(.08)	0.824(.04)	0.869(.02)	0.894(.01)	0.913(.01)	0.927(.00)
GCL [2]	0.636(.05)	0.803(.04)	0.872(.01)	0.891(.01)	0.902(.01)	0.913(.01)	0.927(.01)
PCL	**0.671(.06)**	**0.850(.01)**	**0.885(.01)**	**0.904(.01)**	**0.909(.01)**	**0.919(.00)**	**0.929(.00)**

Experimental Results

Transfer learning setting

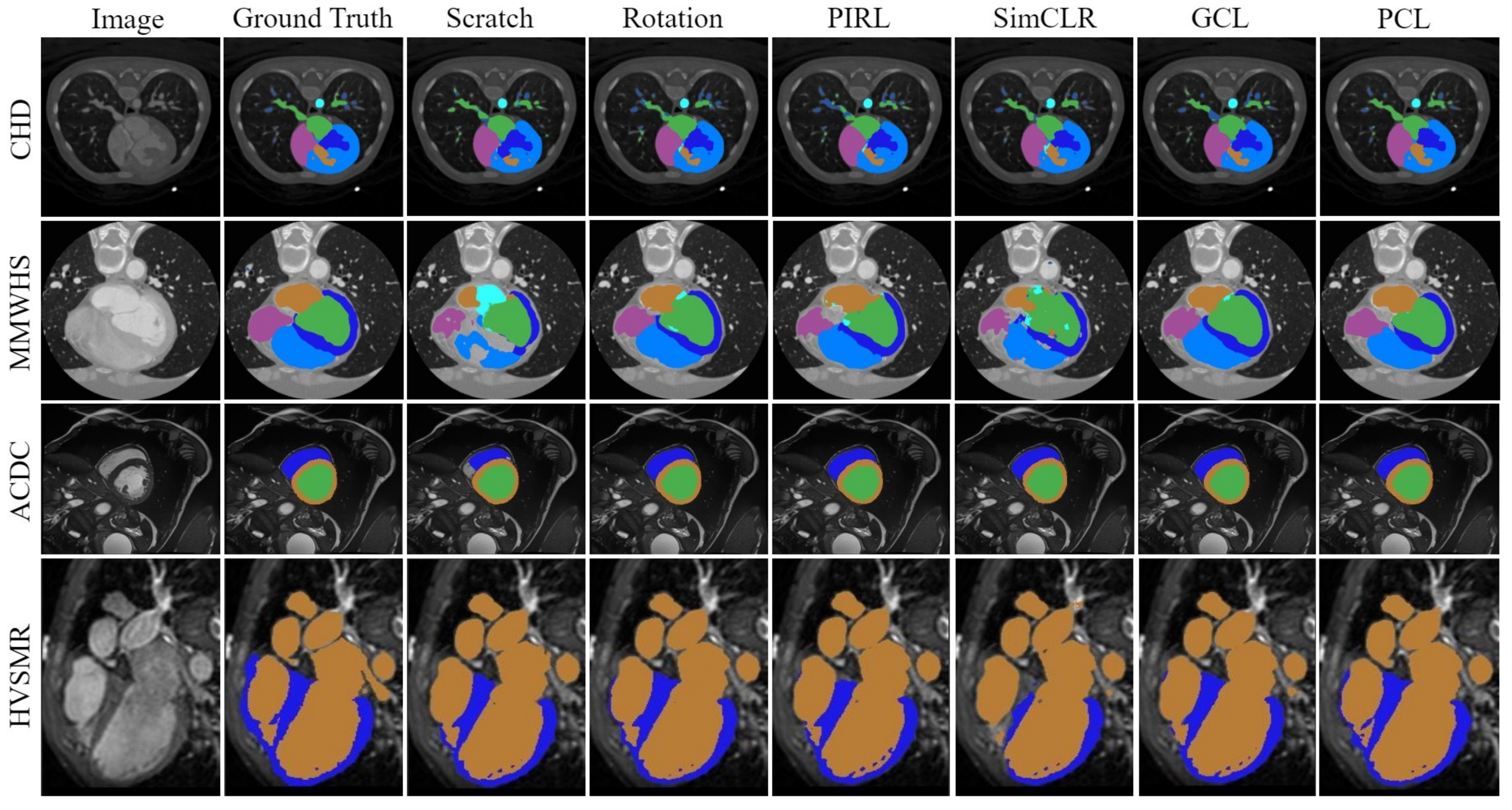
- Encoder pretrained on CHD transfers to MMWHS (CT) dataset
- Encoder pretrained on ACDC transfers to HVSMR (MRI) dataset

CHD transferring to MMWHS (20 patients in total)						
Method	M=2	M=4	M=6	M=8	M=10	M=16
Random	0.232(.14)	0.661(.10)	0.732(.07)	0.769(.06)	0.808(.05)	0.834(.05)
Rotation [6]	0.247(.16)	0.659(.13)	0.751(.07)	0.768(.07)	0.803(.06)	0.850(.04)
PIRL [14]	0.251(.10)	0.670(.11)	0.755(.07)	0.774(.06)	0.821(.05)	0.851(.04)
SimCLR [3]	0.269(.17)	0.683(.10)	0.751(.07)	0.783(.06)	0.818(.05)	0.850(.04)
GCL [2]	0.262(.11)	0.703(.07)	0.768(.05)	0.805(.04)	0.820(.04)	0.851(.03)
PCL	**0.339(.15)**	**0.748(.08)**	**0.792(.05)**	**0.820(.05)**	**0.840(.04)**	**0.869(.03)**

ACDC transferring to HVSMR (10 patients in total)				
Method	M=2	M=4	M=6	M=8
Random	0.742(.06)	0.813(.05)	0.842(.03)	0.842(.04)
Rotation [6]	0.737(.07)	0.816(.06)	0.845(.03)	0.844(.03)
PIRL [14]	0.740(.05)	0.826(.04)	0.849(.03)	0.846(.03)
SimCLR [3]	0.700(.07)	0.779(.05)	0.808(.04)	0.815(.04)
GCL [2]	0.770(.05)	0.818(.05)	0.842(.03)	0.843(.03)
PCL	**0.781(.05)**	**0.832(.05)**	**0.857(.03)**	**0.857(.03)**

M is the number of finetuned patients

Experimental Results



Conclusions

1. Positional contrastive learning can mitigate the false negative rate in the state of art global contrastive learning, hence improve the performance of segmentation downstream task.

2. In addition to semi-supervised setting, positional contrastive learning shows superior transfer learning ability over all baselines, which is not discussed in global contrastive learning.