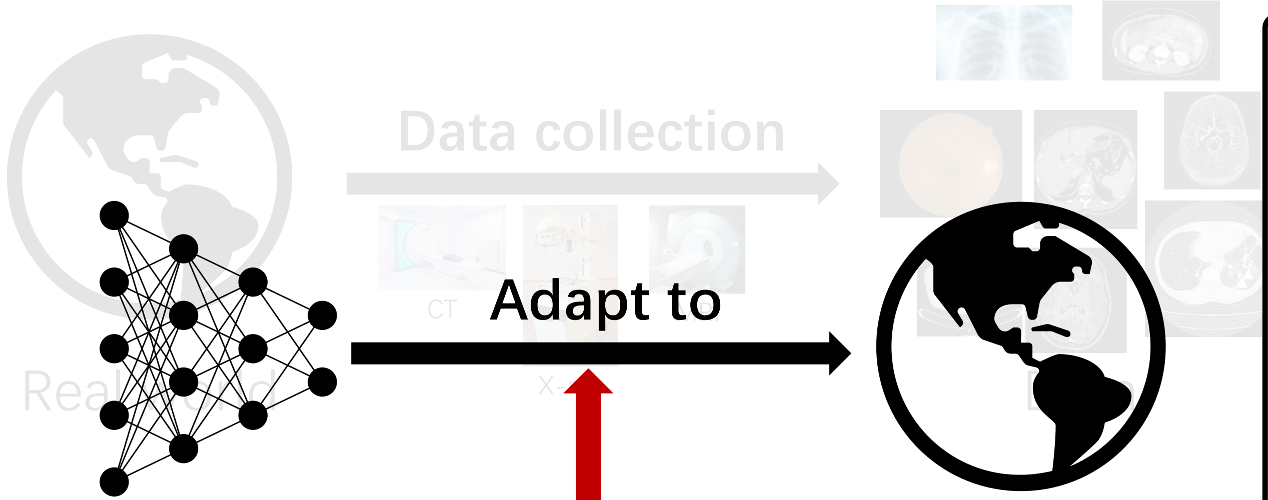
GEOMETRIC VISUAL SIMILARITY LEARNING

IN 3D MEDICAL IMAGE SELF-SUPERVISED PRE-TRAINING

He Yuting (何宇霆)
Southeast University

Background:

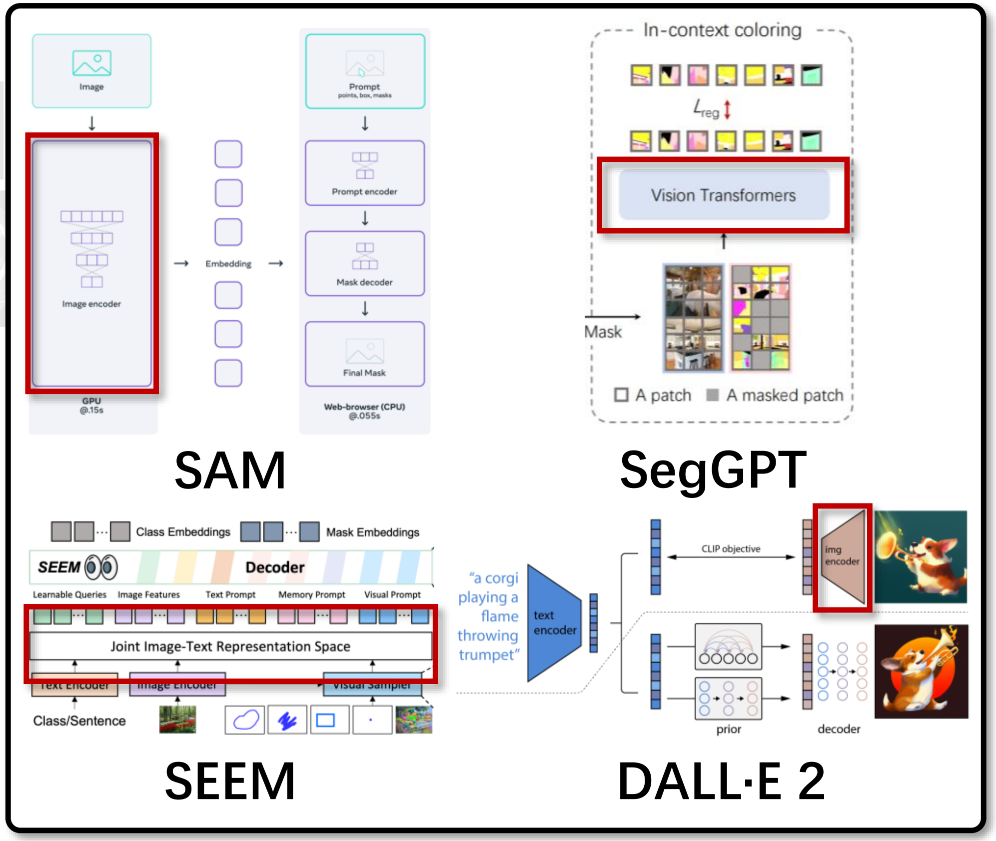
Self-supervised pre-training



Pre-trained model

Real world

Lower cost, faster adapt

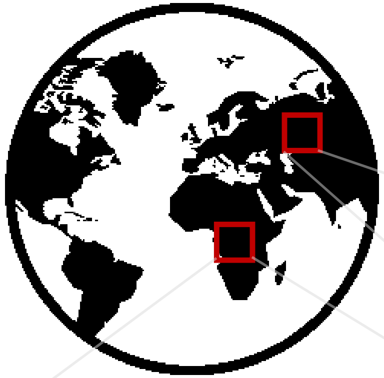


Basis of AGI···

He, Y., et al. (2023). Geometric Visual Similarity Learning in 3D Medical Image Self-supervised Pre-training. *IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023*

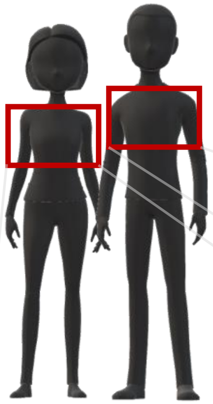
Background:

Medical images v.s. natural images



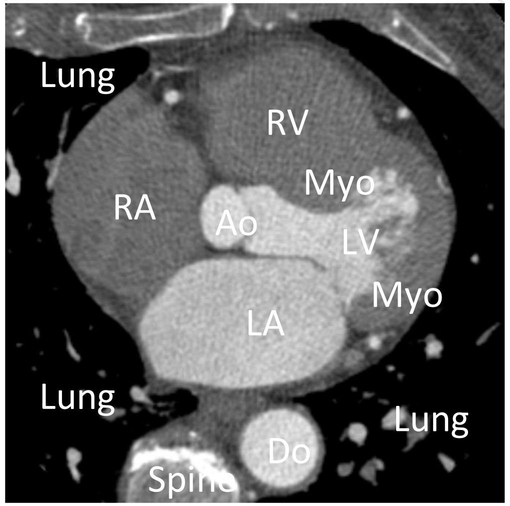
Natural images

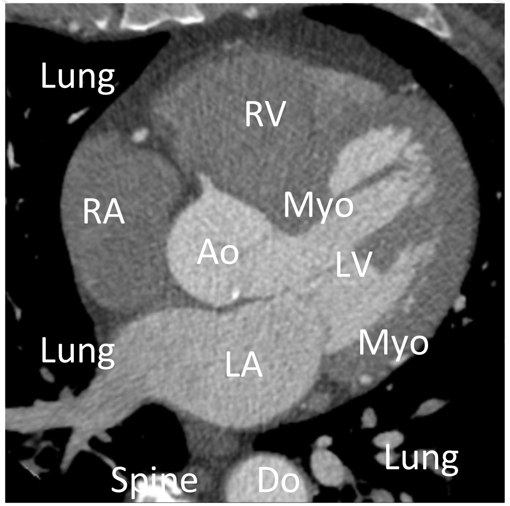
- ✓ Scan from **large** scopes
- ✓ **Nonlimited** range and pose
- ➢ *Large inter-image **difference***



Medical images

- ✓ Scan from **small** scopes
- ✓ **Limited** range and pose
- ➢ *Large inter-image **similarity***

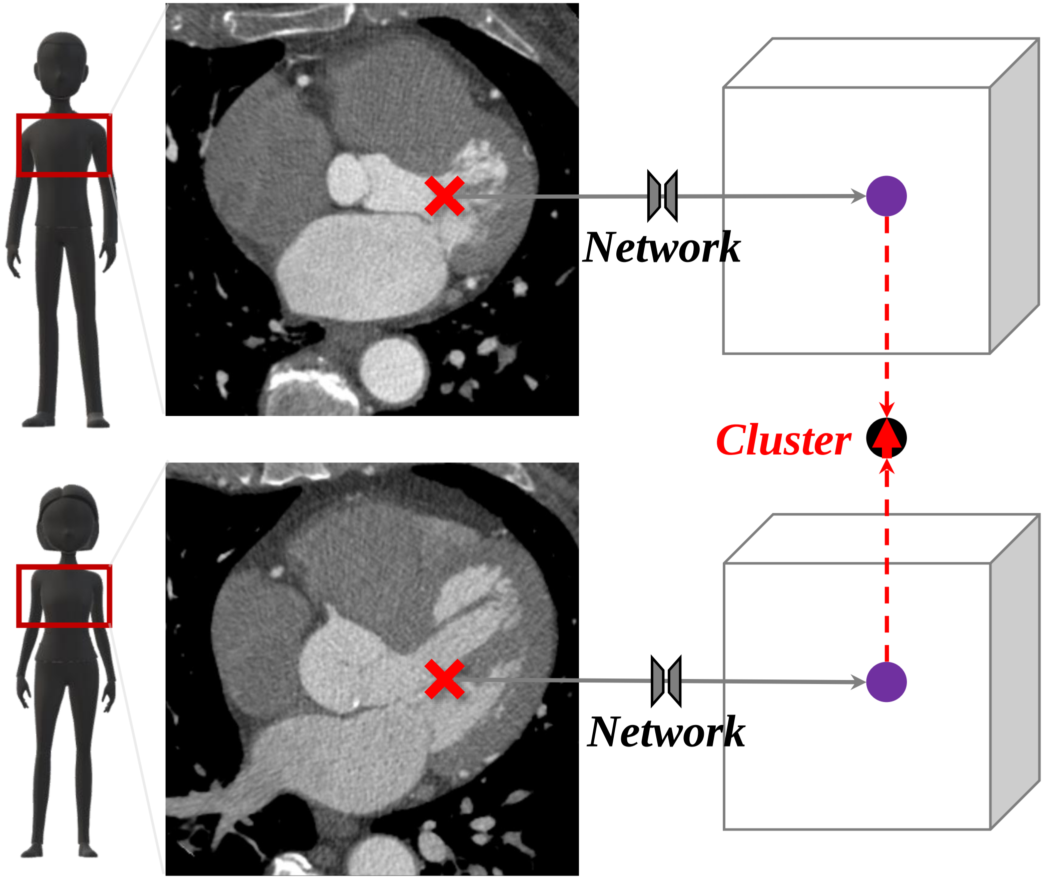




Opportunity: Learning <u>inter-image similarity</u> *for the clustering of the same semantic regions*

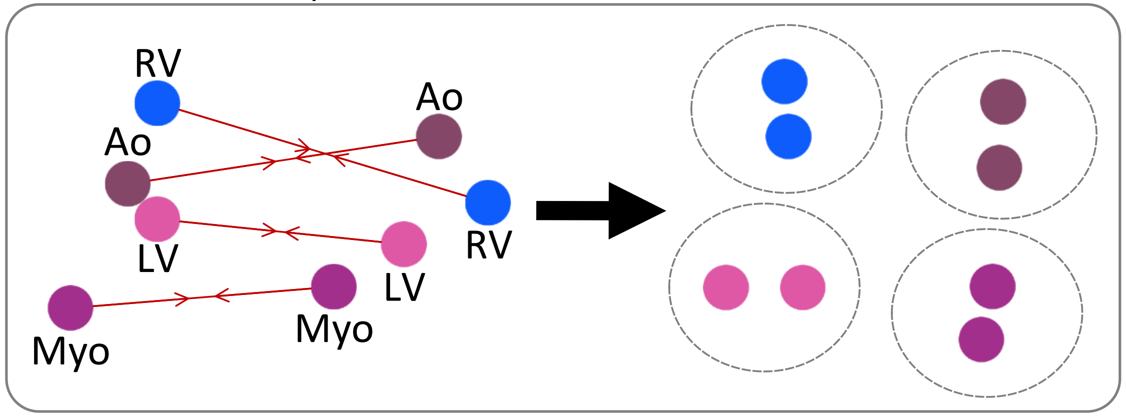
BACKGROUND:

LIMITATION

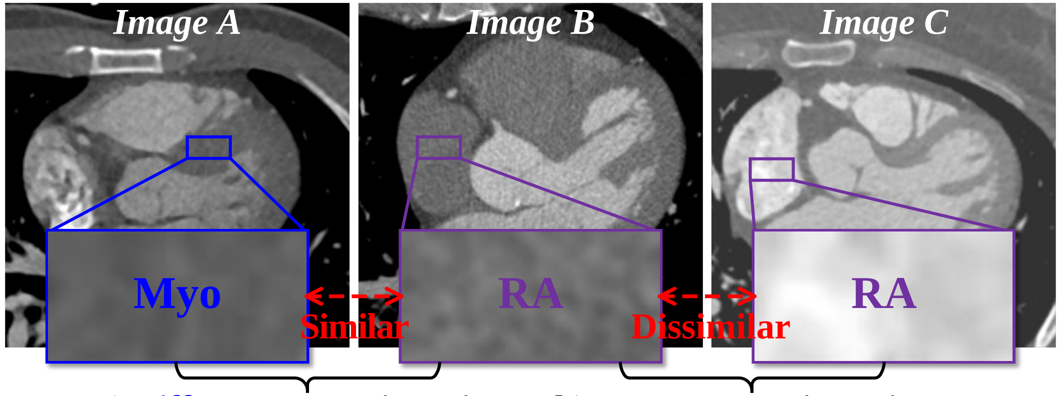


Wang, X., et al. (2021). Dense contrastive learning for self-supervised visual pre-training. CVPR (pp. 3024-3033).

DenseCL, DeepCluster, etc.



Limitation: ***unreliable inter-image correspondence***



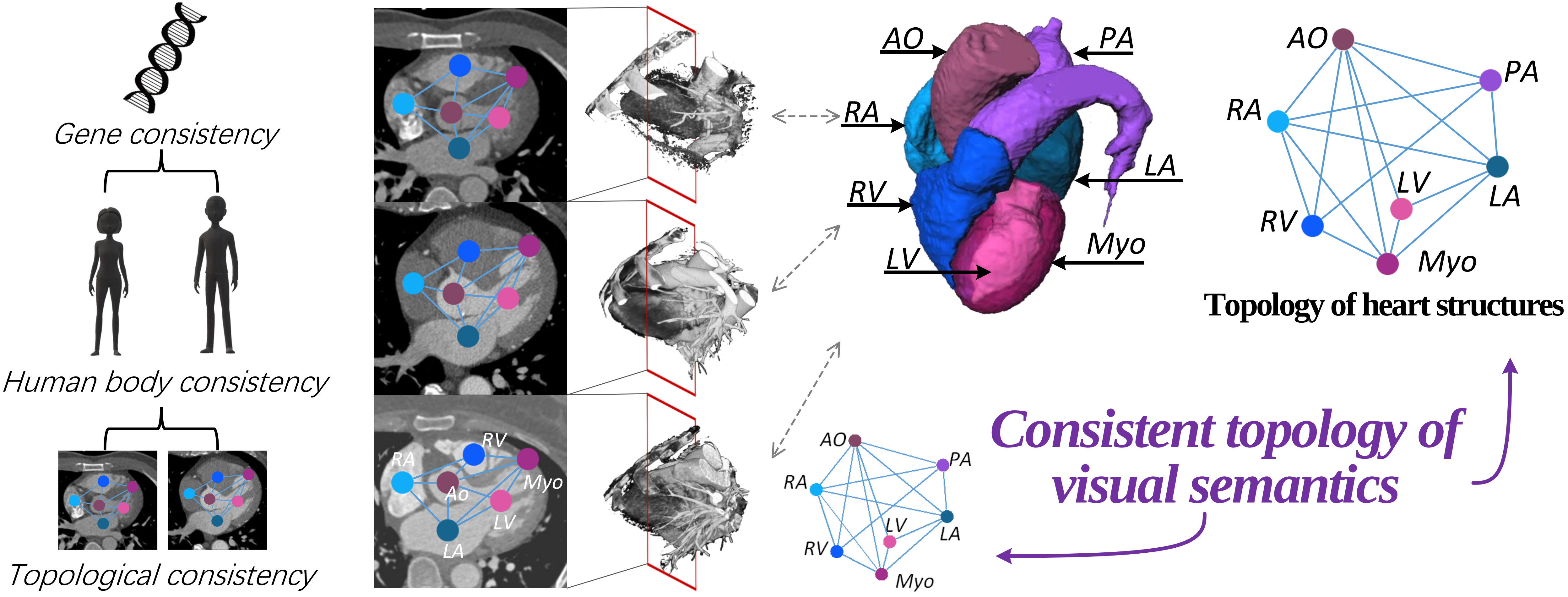
a) Different semantic regions with similar appearance

b) Same semantic regions with dissimilar appearance

Motivation:

Topological invariance

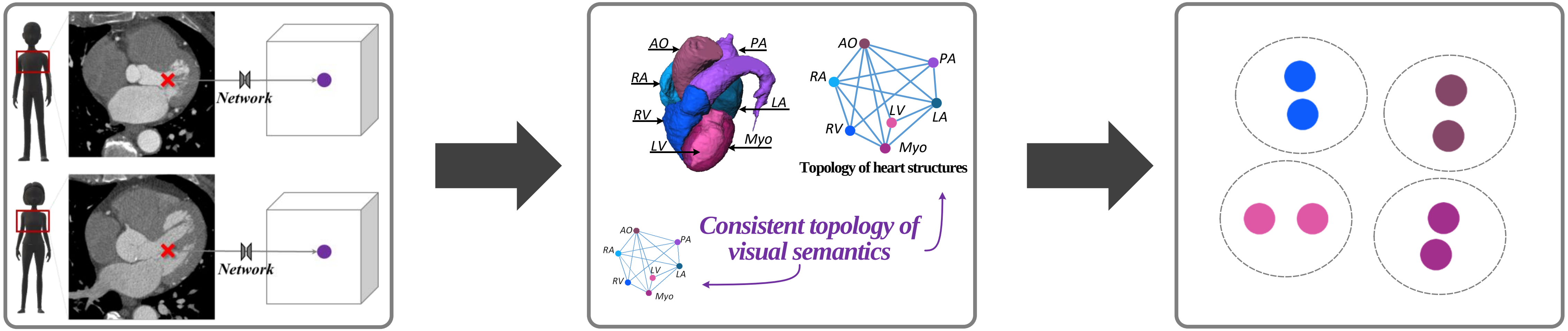
伊宇霆



Hypothesis: ***Keeping the topology of 3D medical images will enhance the correspondence discovery***

Contribution:

Geometric visual similarity learning



Representation *Correspondence discovery based on topology* *Clustering effect*

- Advances the **learning of inter-image similarity in 3D medical image SSP** pushing the representability of pre-trained models;
- Propose **the Geometric Visual Similarity Learning** (GVSL) that embeds the prior of topological invariance into the correspondence learning;
- Present a novel SSP head, **Z-Matching head**, for simultaneously powerful global and local representation via GVSL.

Methodology:

Geometric visual similarity learning

Image x_B

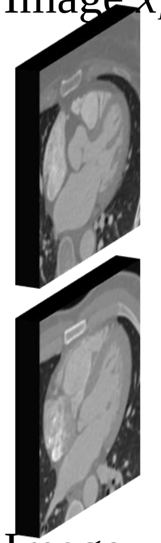
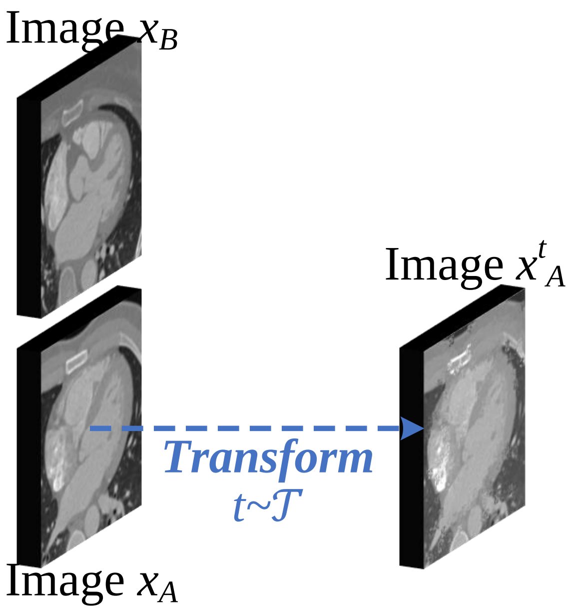


Image x_A

Two 3D images

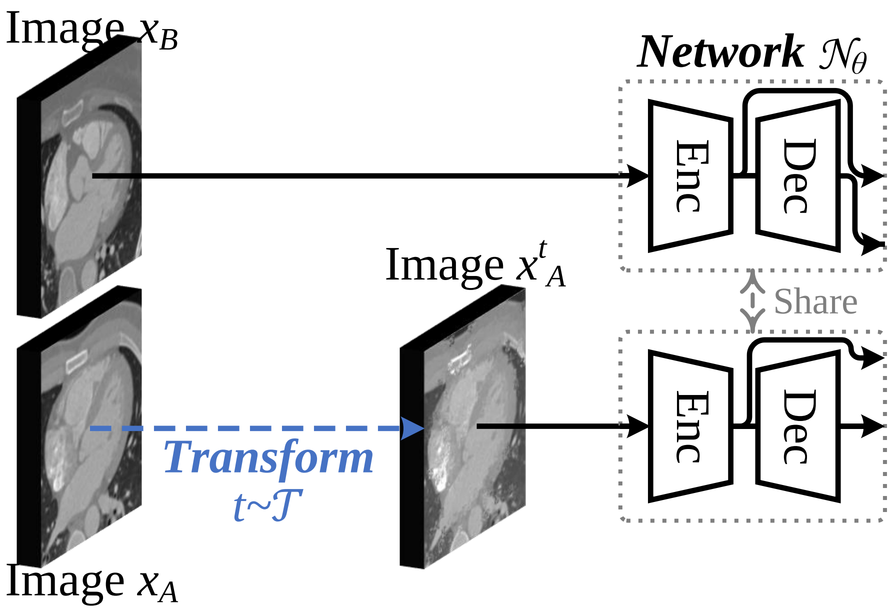
METHODOLOGY:
GEOMETRIC VISUAL SIMILARITY LEARNING



Augmentation for feature diversity

Methodology:

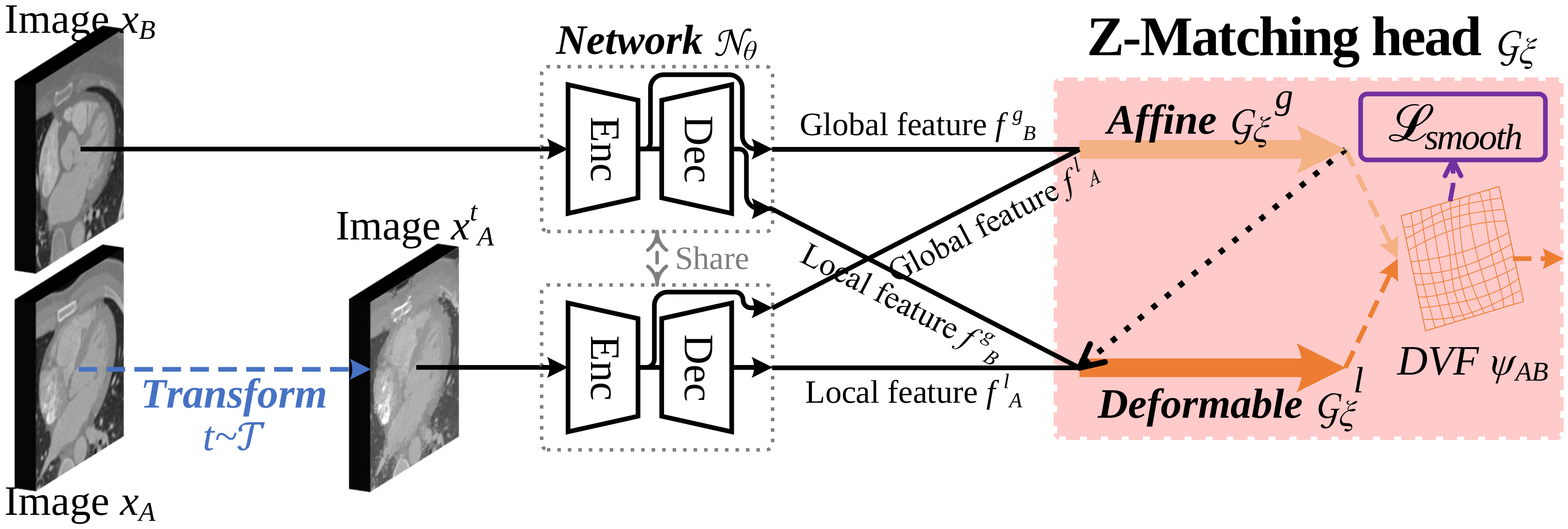
Geometric visual similarity learning



Feature extraction via two shared-weight networks

Methodology:

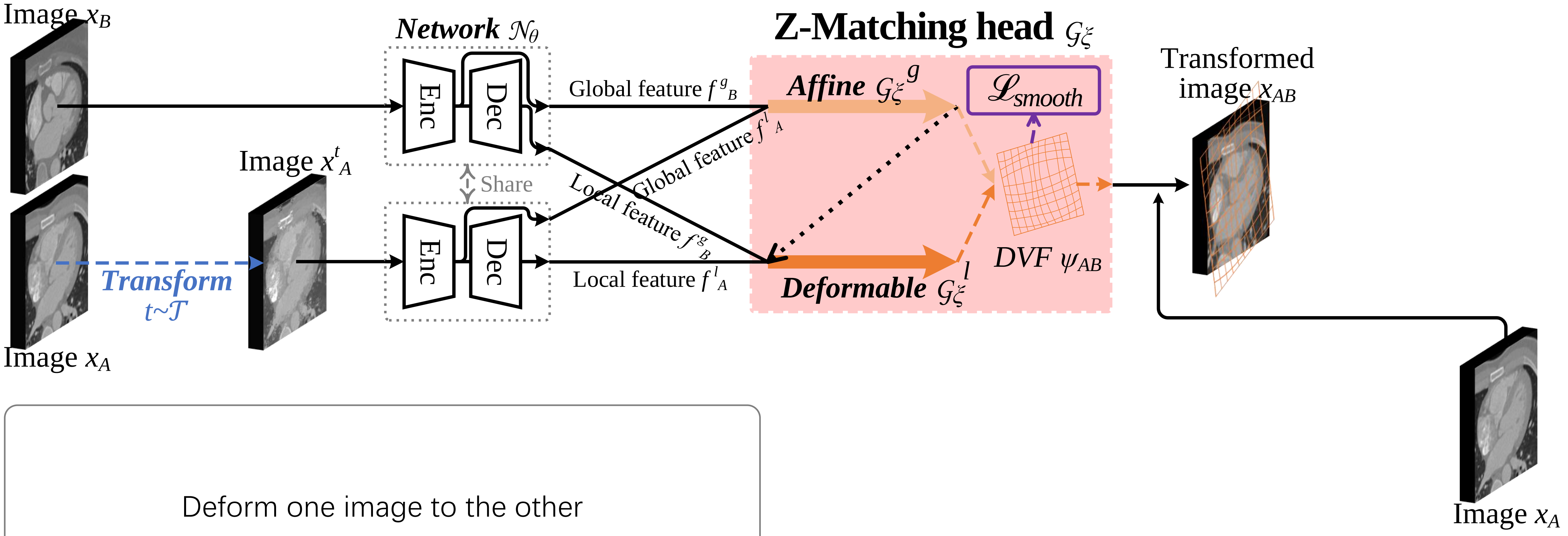
Geometric visual similarity learning



Predict correspondence

Methodology:

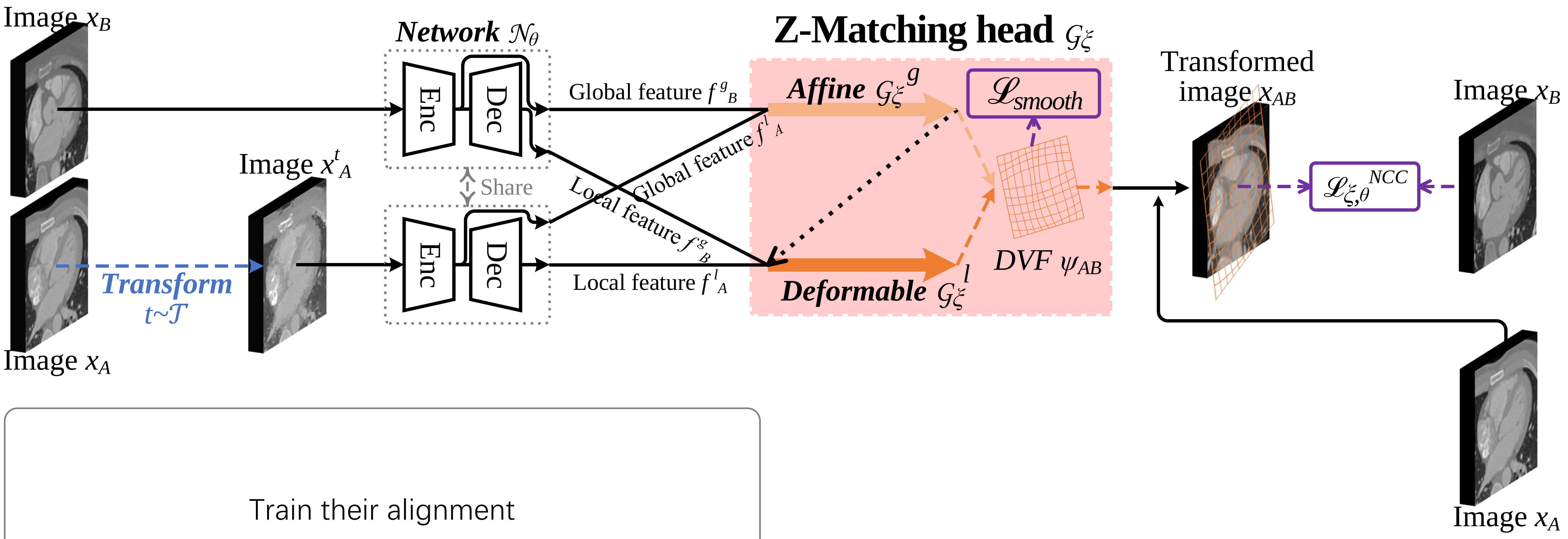
Geometric visual similarity learning



Deform one image to the other

Methodology:

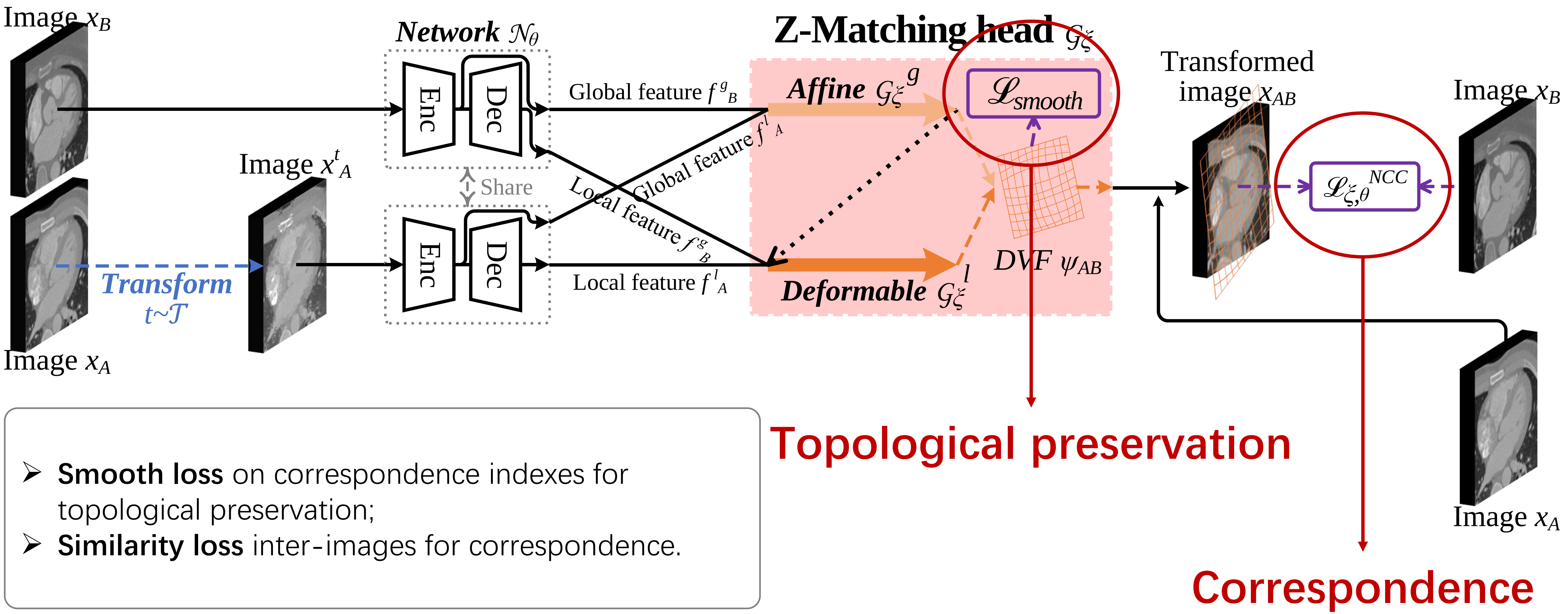
Geometric visual similarity learning



Train their alignment

Methodology:

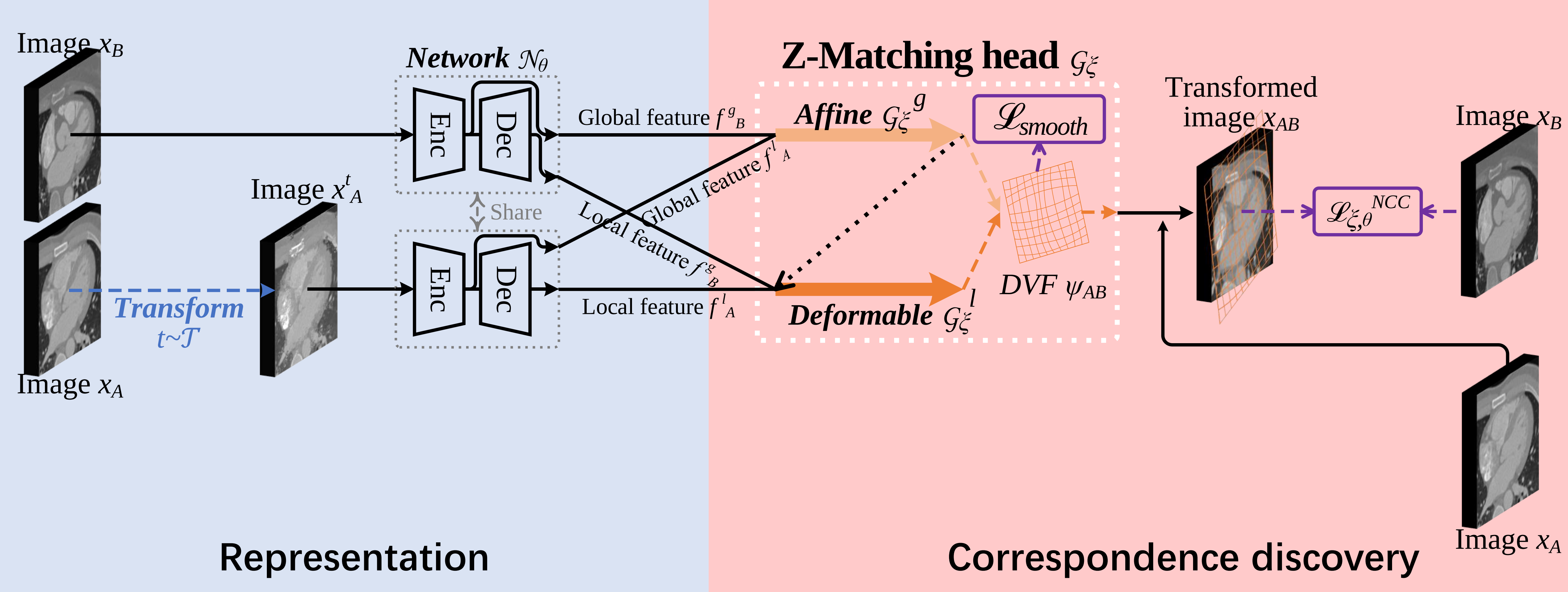
Geometric visual similarity learning



- **Smooth loss** on correspondence indexes for topological preservation;
- **Similarity loss** inter-images for correspondence.

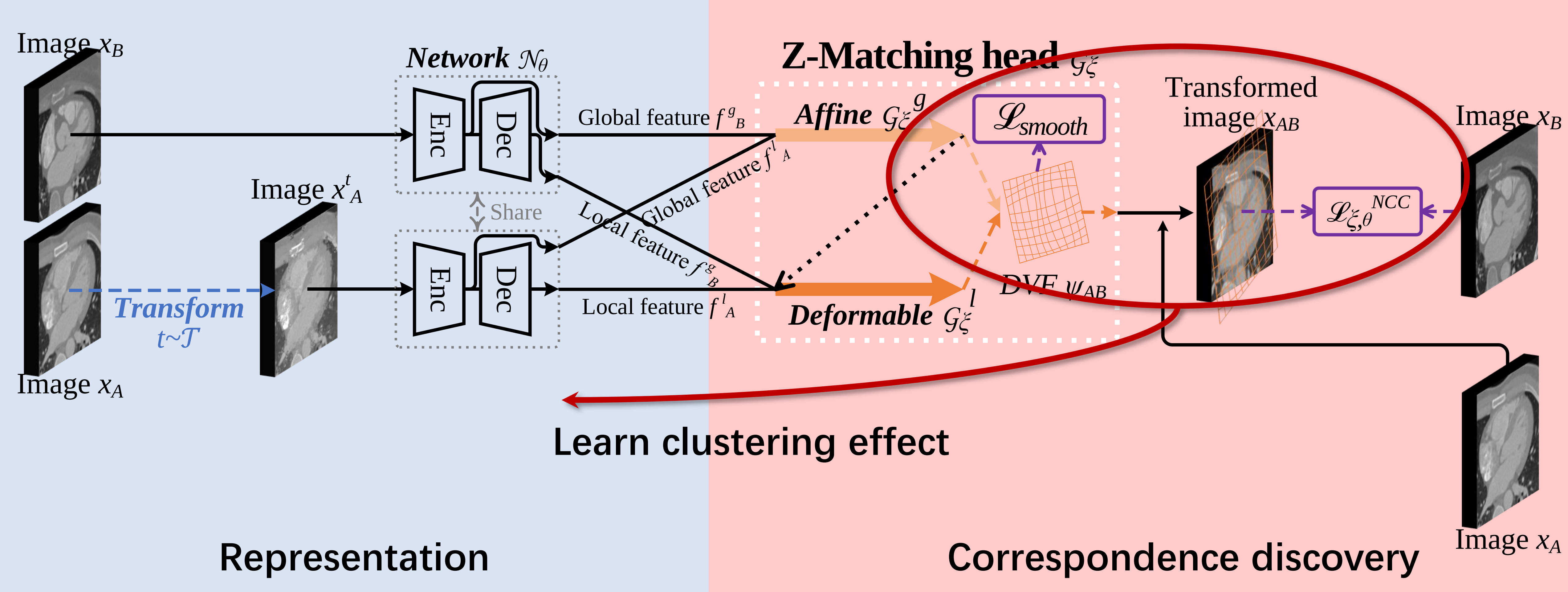
Methodology:

Geometric visual similarity learning



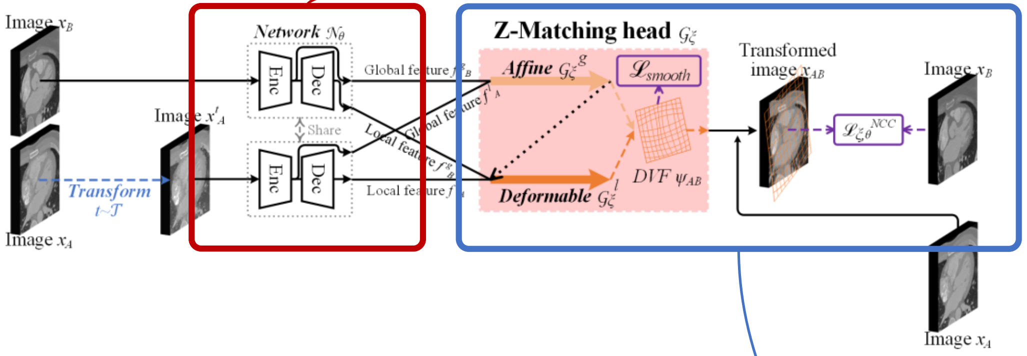
Methodology:

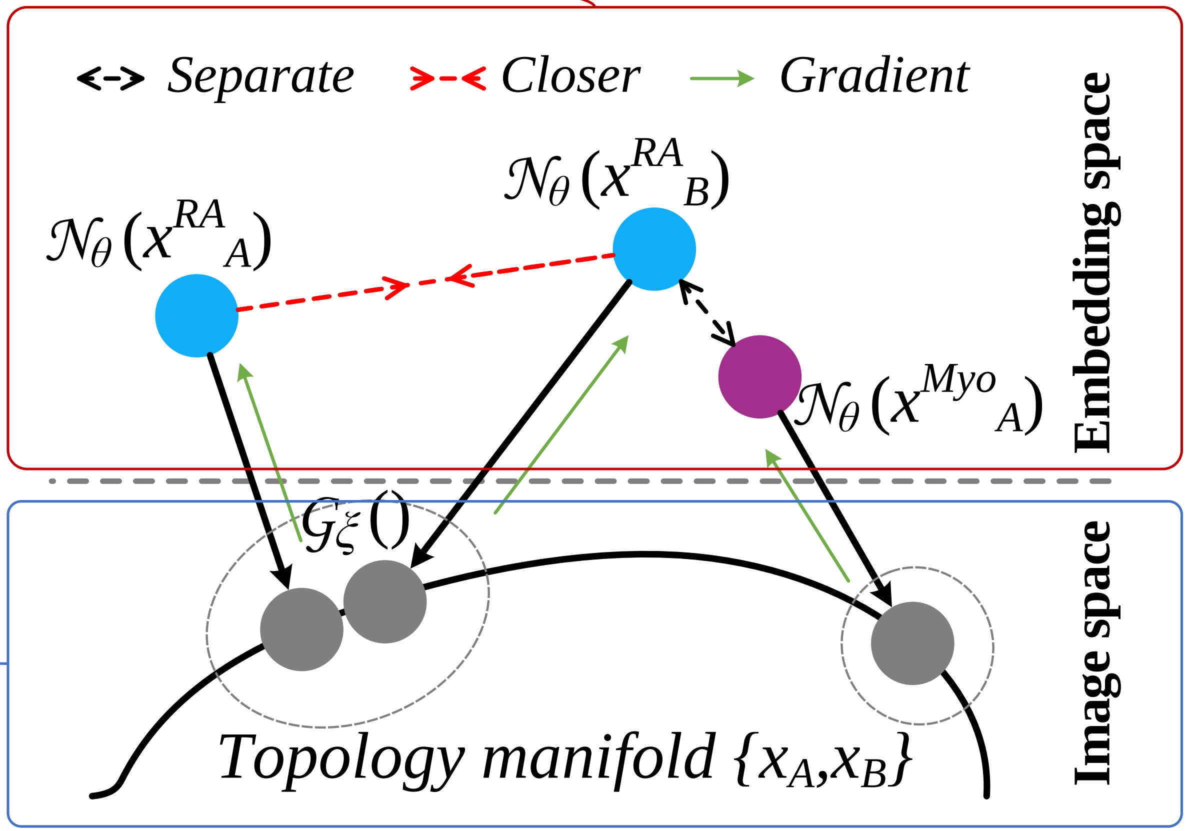
Geometric visual similarity learning



Methodology (Intuitions):

Geometric visual similarity learning

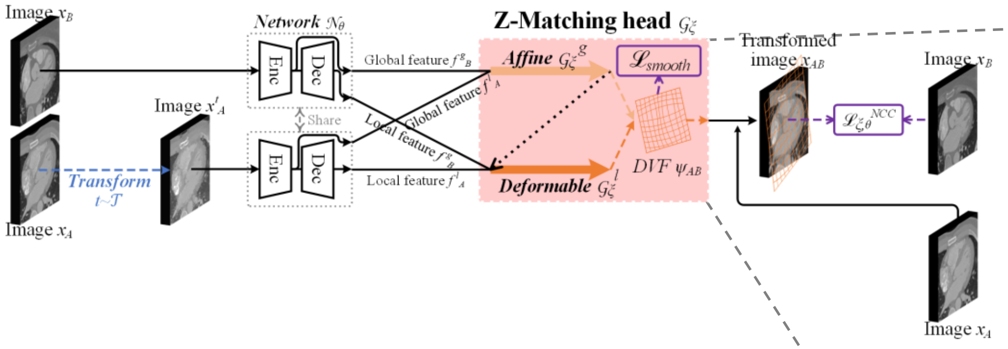


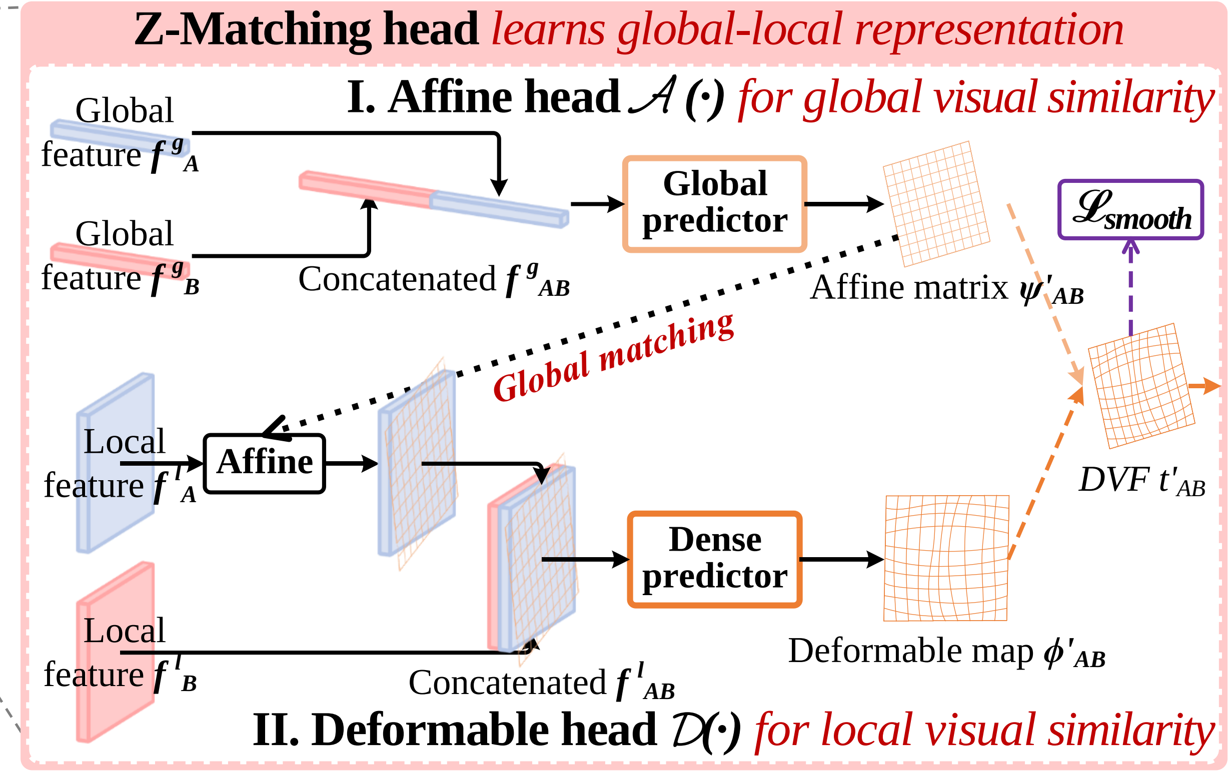


- Implicitly embed a topology manifold inner the images into the measurement process, and measure the similarity on this topology manifold.

Methodology:

Z-Matching head

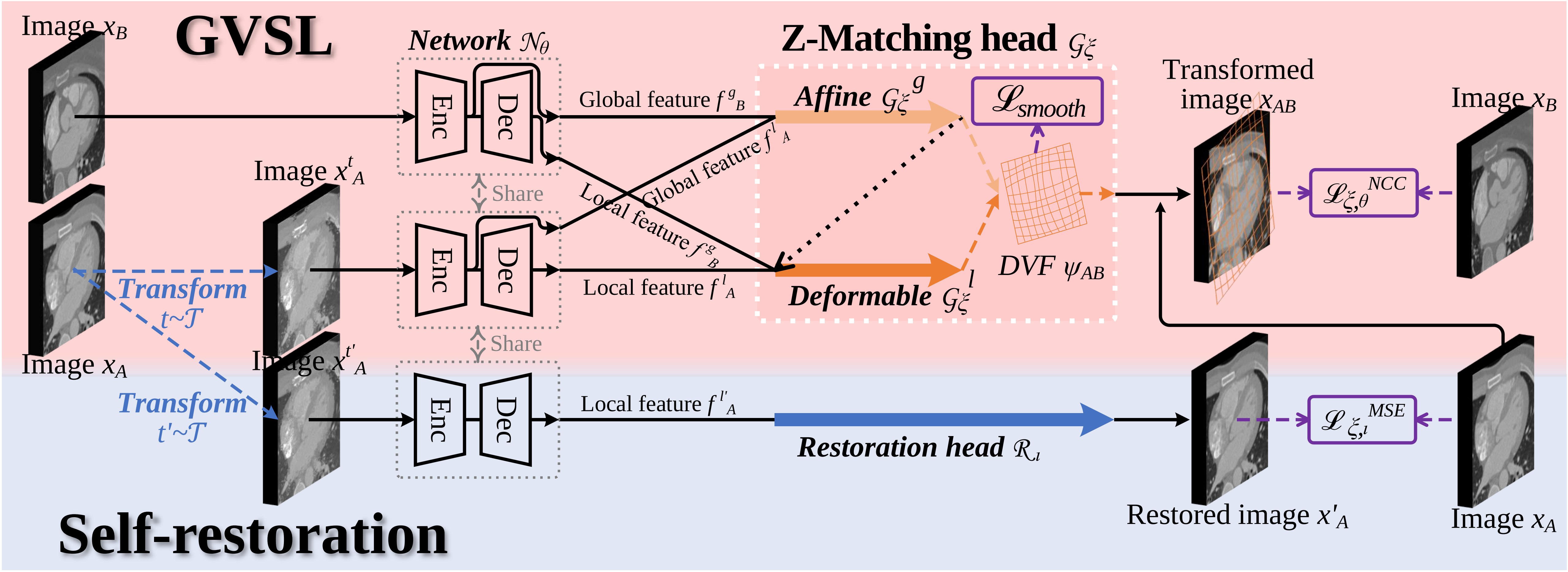




- **Affine head:** global visual similarity and alignment for global representation
- **Deformable head:** local visual similarity and alignment for dense representation

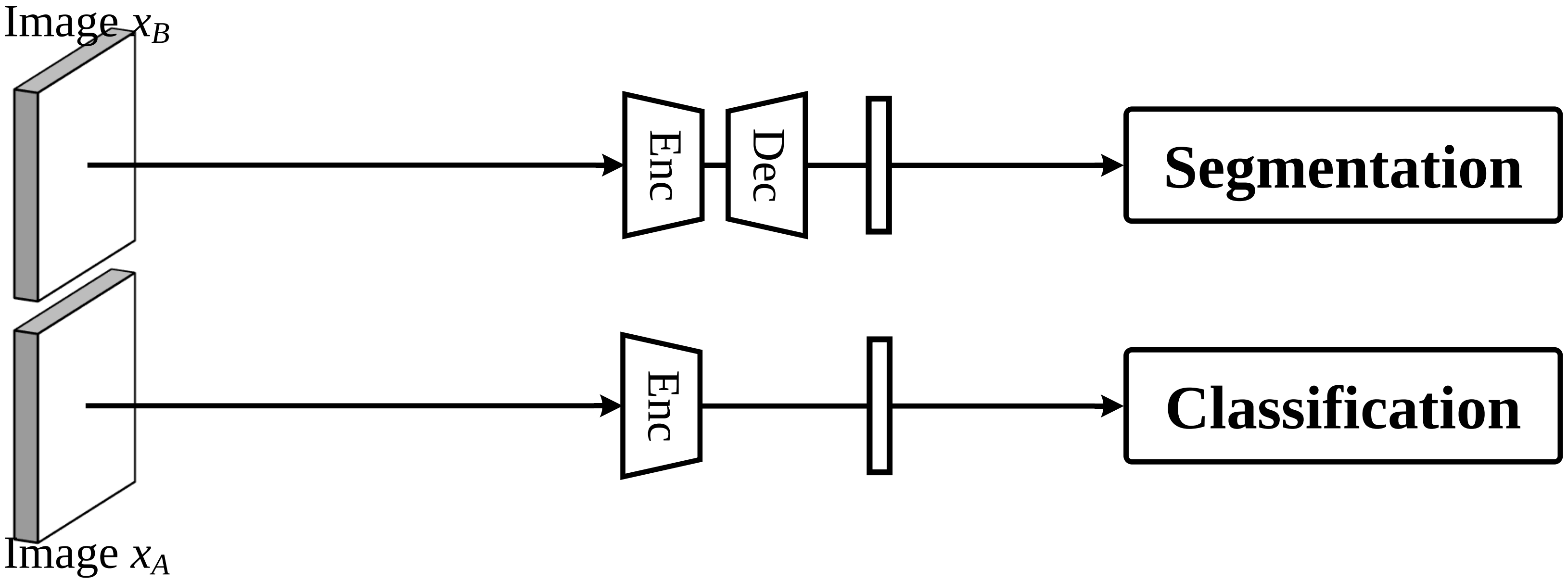
Methodology:

Self-restoration for warm-up



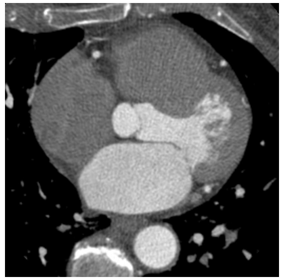
Experiment:

Evaluation tasks



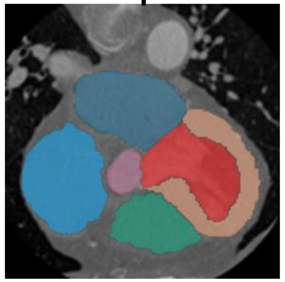
- Heart structures on CT (SHC)
- Coronary artery on CT (SAC)
- Brain tissues on MR (SBM)

- COVID-19 on CT (CCC)

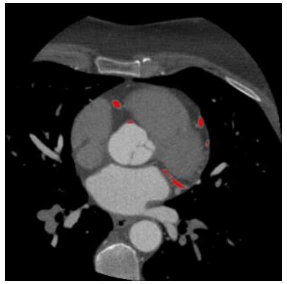


Pretrain dataset:
302 CCTA images

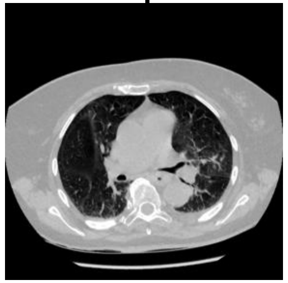
Inner-scene



SHC
-MMWHS
-15/5/40

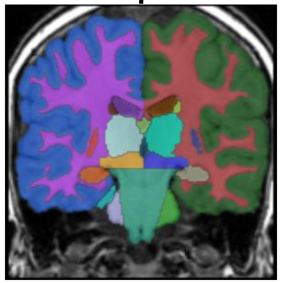


SAC
-ASOCA
-15/5/20



CCC
-STOIC
-1k/400/
600

Inter-scene



SBM
-CANDI
-40/20/43

Experiment:

Linear and fine-tuning evaluation

Pre-training	a) Linear: powerful representation				b) Fine-tuning: great transferring			
	$SHC_{DSC\%}$	$SAC_{DSC\%}$	$CCC_{AUC\%}$	$SBM_{DSC\%}$	$SHC_{DSC\%}$	$SAC_{DSC\%}$	$CCC_{AUC\%}$	$SBM_{DSC\%}$
	Inner scene			*Inter scene*	*Inner scene*			*Inter scene*
Scratch	21.9	10.0	52.7	56.4	87.8	80.4	74.4	89.7
Denosing [40]	31.4 (+9.5)	9.3 (−0.7)	57.9 (+5.2)	28.3 (−28.1)	90.3 (+2.5)	80.5 (+0.1)	75.6 (+1.2)	89.7
In-painting [30]	32.3 (+10.4)	5.9 (−4.1)	57.1 (+4.4)	25.0 (−31.4)	90.4 (+2.6)	80.3 (−0.1)	79.9 (+5.5)	89.9 (+0.2)
Models Genesis [48]	47.4 (+25.5)	22.5 (+12.5)	60.4 (+7.7)	44.9 (−11.5)	90.3 (+2.5)	79.9 (−0.5)	80.7 (+6.3)	89.4 (−0.3)
Rotation [23]	56.1 (+34.2)	21.9 (+11.9)	**62.1** (+9.4)	54.1 (−2.3)	90.6 (+2.8)	81.1 (+0.7)	77.1 (+2.7)	89.6 (−0.1)
DeepCluster [2]	55.9 (+34.0)	4.4 (−5.6)	57.9 (+5.2)	67.5 (+11.1)	85.4 (−2.4)	80.5 (+0.1)	59.9 (−14.5)	89.1 (−0.6)
SimSiam [4]	56.5 (+34.6)	9.7 (−0.3)	61.0 (+8.3)	66.2 (+9.8)	87.5 (−0.3)	80.1 (−0.3)	73.6 (−0.8)	89.8 (+0.1)
BYOL [7]	46.9 (+25.0)	8.6 (−1.4)	53.7 (+1.0)	52.7 (−3.7)	88.6 (+0.8)	80.7 (+0.3)	76.5 (+2.1)	89.5 (−0.2)
SimCLR [3]	48.7 (+26.8)	15.5 (+5.5)	61.3 (+8.6)	58.7 (+2.3)	86.9 (−0.9)	79.9 (−0.5)	74.3 (−0.1)	89.3 (−0.4)
w/o Z-Matching	49.1 (+27.2)	21.1 (+11.1)	55.8 (+3.4)	45.1 (−11.3)	88.3 (+0.5)	81.2 (+0.8)	81.3 (+6.9)	89.7
w/o Fundament	45.3 (+23.4)	0.0 (−10.0)	58.8 (+6.4)	48.5 (−7.9)	87.0 (−0.8)	79.5 (−0.9)	76.6 (+2.2)	89.0 (−0.7)
w/o Affine head	57.7 (+35.8)	17.9 (+7.9)	57.6 (+4.9)	53.4 (−3.0)	89.4 (+1.6)	**82.3** (+1.9)	79.8 (+5.4)	89.8 (+0.1)
Our GVSL (Whole)	**68.4** (+46.5)	**28.7** (+18.7)	60.8 (+8.1)	**79.9** (+23.5)	**91.2** (+3.4)	81.3 (+0.9)	**82.2** (+7.8)	**90.0** (+0.3)

➢ Powerful inner-scene transferring for both large and small structures

Experiment:

Linear and fine-tuning evaluation

何宇霆

Pre-training	a) Linear: powerful representation				b) Fine-tuning: great transferring			
	SHC$_{DSC\%}$	SAC$_{DSC\%}$	CCC$_{AUC\%}$	SBM$_{DSC\%}$	SHC$_{DSC\%}$	SAC$_{DSC\%}$	CCC$_{AUC\%}$	SBM$_{DSC\%}$
	Inner scene			*Inter scene*	*Inner scene*			*Inter scene*
Scratch	21.9	10.0	52.7	56.4	87.8	80.4	74.4	89.7
Denosing [40]	31.4$_{(+9.5)}$	9.3$_{(-0.7)}$	57.9$_{(+5.2)}$	28.3$_{(-28.1)}$	90.3$_{(+2.5)}$	80.5 $_{(+0.1)}$	75.6$_{(+1.2)}$	89.7
In-painting [30]	32.3$_{(+10.4)}$	5.9$_{(-4.1)}$	57.1$_{(+4.4)}$	25.0$_{(-31.4)}$	90.4$_{(+2.6)}$	80.3 $_{(-0.1)}$	79.9$_{(+5.5)}$	89.9$_{(+0.2)}$
Models Genesis [48]	47.4$_{(+25.5)}$	22.5$_{(+12.5)}$	60.4$_{(+7.7)}$	44.9$_{(-11.5)}$	90.3$_{(+2.5)}$	79.9 $_{(-0.5)}$	80.7$_{(+6.3)}$	89.4$_{(-0.3)}$
Rotation [23]	56.1$_{(+34.2)}$	21.9$_{(+11.9)}$	**62.1**$_{(+9.4)}$	54.1$_{(-2.3)}$	90.6$_{(+2.8)}$	81.1$_{(+0.7)}$	77.1$_{(+2.7)}$	89.6$_{(-0.1)}$
DeepCluster [2]	55.9$_{(+34.0)}$	4.4$_{(-5.6)}$	57.9$_{(+5.2)}$	67.5$_{(+11.1)}$	85.4$_{(-2.4)}$	80.5$_{(+0.1)}$	59.9$_{(-14.5)}$	89.1$_{(-0.6)}$
SimSiam [4]	56.5$_{(+34.6)}$	9.7$_{(-0.3)}$	61.0$_{(+8.3)}$	66.2$_{(+9.8)}$	87.5$_{(-0.3)}$	80.1 $_{(-0.3)}$	73.6$_{(-0.8)}$	89.8$_{(+0.1)}$
BYOL [7]	46.9$_{(+25.0)}$	8.6$_{(-1.4)}$	53.7$_{(+1.0)}$	52.7$_{(-3.7)}$	88.6$_{(+0.8)}$	80.7 $_{(+0.3)}$	76.5$_{(+2.1)}$	89.5$_{(-0.2)}$
SimCLR [3]	48.7$_{(+26.8)}$	15.5$_{(+5.5)}$	61.3$_{(+8.6)}$	58.7$_{(+2.3)}$	86.9 $_{(-0.9)}$	79.9$_{(-0.5)}$	74.3$_{(-0.1)}$	89.3$_{(-0.4)}$
w/o Z-Matching	49.1$_{(+27.2)}$	21.1$_{(+11.1)}$	55.8$_{(+3.4)}$	45.1$_{(-11.3)}$	88.3$_{(+0.5)}$	81.2$_{(+0.8)}$	81.3$_{(+6.9)}$	89.7
w/o Fundament	45.3$_{(+23.4)}$	0.0$_{(-10.0)}$	58.8$_{(+6.4)}$	48.5$_{(-7.9)}$	87.0$_{(-0.8)}$	79.5 $_{(-0.9)}$	76.6$_{(+2.2)}$	89.0$_{(-0.7)}$
w/o Affine head	57.7$_{(+35.8)}$	17.9$_{(+7.9)}$	57.6$_{(+4.9)}$	53.4$_{(-3.0)}$	89.4$_{(+1.6)}$	**82.3**$_{(+1.9)}$	79.8$_{(+5.4)}$	89.8$_{(+0.1)}$
Our GVSL (Whole)	**68.4**$_{(+46.5)}$	**28.7**$_{(+18.7)}$	60.8$_{(+8.1)}$	**79.9**$_{(+23.5)}$	**91.2**$_{(+3.4)}$	81.3$_{(+0.9)}$	**82.2**$_{(+7.8)}$	**90.0**$_{(+0.3)}$

➢ **Effective inter-scene transferring, but is not significant in fine-tuning**

Experiment:

Linear and fine-tuning evaluation

Dense

Pre-training	a) Linear: powerful representation				b) Fine-tuning: great transferring			
	SHC$_{DSC\%}$	SAC$_{DSC\%}$	CCC$_{AUC\%}$	SBM$_{DSC\%}$	SHC$_{DSC\%}$	SAC$_{DSC\%}$	CCC$_{AUC\%}$	SBM$_{DSC\%}$
	Inner scene			*Inter scene*	*Inner scene*			*Inter scene*
Scratch	21.9	10.0	52.7	56.4	87.8	80.4	74.4	89.7
Denosing [40]	31.4(+9.5)	9.3(−0.7)	57.9(+5.2)	28.3(−28.1)	90.3(+2.5)	80.5 (+0.1)	75.6(+1.2)	89.7
In-painting [30]	32.3(+10.4)	5.9(−4.1)	57.1(+4.4)	25.0(−31.4)	90.4(+2.6)	80.3 (−0.1)	79.9(+5.5)	89.9(+0.2)
Models Genesis [48]	47.4(+25.5)	22.5(+12.5)	60.4(+7.7)	44.9(−11.5)	90.3(+2.5)	79.9 (−0.5)	80.7(+6.3)	89.4(−0.3)
Rotation [23]	56.1(+34.2)	21.9(+11.9)	**62.1**(+9.4)	54.1(−2.3)	90.6(+2.8)	81.1(+0.7)	77.1(+2.7)	89.6(−0.1)
DeepCluster [2]	55.9(+34.0)	4.4(−5.6)	57.9(+5.2)	67.5(+11.1)	85.4(−2.4)	80.5(+0.1)	59.9(−14.5)	89.1(−0.6)
SimSiam [4]	56.5(+34.6)	9.7(−0.3)	61.0(+8.3)	66.2(+9.8)	87.5(−0.3)	80.1 (−0.3)	73.6(−0.8)	89.8(+0.1)
BYOL [7]	46.9(+25.0)	8.6(−1.4)	53.7(+1.0)	52.7(−3.7)	88.6(+0.8)	80.7 (+0.3)	76.5(+2.1)	89.5(−0.2)
SimCLR [3]	48.7(+26.8)	15.5(+5.5)	61.3(+8.6)	58.7(+2.3)	86.9 (−0.9)	79.9(−0.5)	74.3(−0.1)	89.3(−0.4)
w/o Z-Matching	49.1(+27.2)	21.1(+11.1)	55.8(+3.4)	45.1(−11.3)	88.3(+0.5)	81.2(+0.8)	81.3(+6.9)	89.7
w/o Fundament	45.3(+23.4)	0.0(−10.0)	58.8(+6.4)	48.5(−7.9)	87.0(−0.8)	79.5 (−0.9)	76.6(+2.2)	89.0(−0.7)
w/o Affine head	57.7(+35.8)	17.9(+7.9)	57.6(+4.9)	53.4(−3.0)	89.4(+1.6)	**82.3**(+1.9)	79.8(+5.4)	89.8(+0.1)
Our GVSL (Whole)	**68.4**(+46.5)	**28.7**(+18.7)	60.8(+8.1)	**79.9**(+23.5)	**91.2**(+3.4)	81.3(+0.9)	**82.2**(+7.8)	**90.0**(+0.3)

Global

➢ Superiority in global and dense prediction tasks

Experiment:

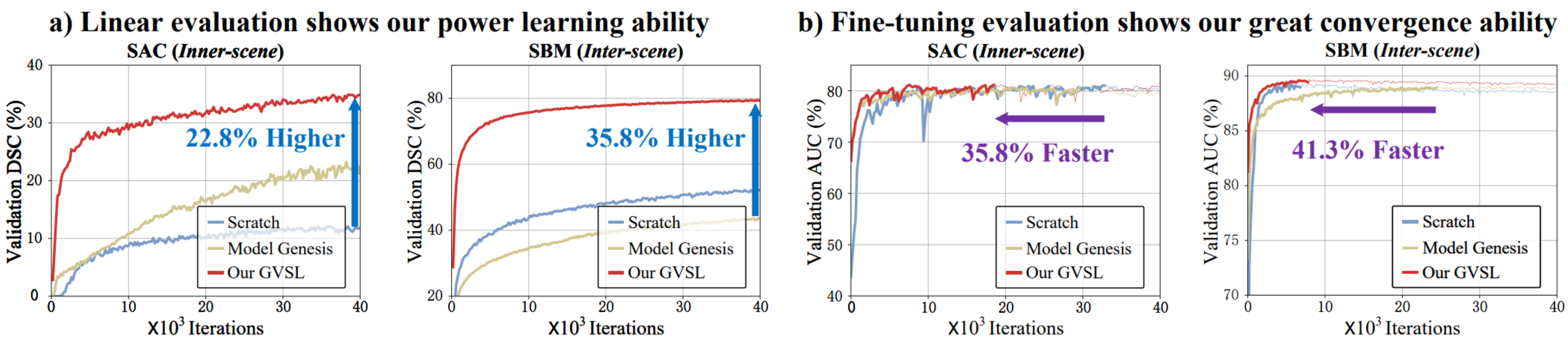
Ablation study

Pre-training	a) Linear: powerful representation				b) Fine-tuning: great transferring			
	$SHC_{DSC\%}$	$SAC_{DSC\%}$	$CCC_{AUC\%}$	$SBM_{DSC\%}$	$SHC_{DSC\%}$	$SAC_{DSC\%}$	$CCC_{AUC\%}$	$SBM_{DSC\%}$
	Inner scene			*Inter scene*	*Inner scene*			*Inter scene*
Scratch	21.9	10.0	52.7	56.4	87.8	80.4	74.4	89.7
Denosing [40]	31.4 (+9.5)	9.3 (−0.7)	57.9 (+5.2)	28.3 (−28.1)	90.3 (+2.5)	80.5 (+0.1)	75.6 (+1.2)	89.7
In-painting [30]	32.3 (+10.4)	5.9 (−4.1)	57.1 (+4.4)	25.0 (−31.4)	90.4 (+2.6)	80.3 (−0.1)	79.9 (+5.5)	89.9 (+0.2)
Models Genesis [48]	47.4 (+25.5)	22.5 (+12.5)	60.4 (+7.7)	44.9 (−11.5)	90.3 (+2.5)	79.9 (−0.5)	80.7 (+6.3)	89.4 (−0.3)
Rotation [23]	56.1 (+34.2)	21.9 (+11.9)	**62.1** (+9.4)	54.1 (−2.3)	90.6 (+2.8)	81.1 (+0.7)	77.1 (+2.7)	89.6 (−0.1)
DeepCluster [2]	55.9 (+34.0)	4.4 (−5.6)	57.9 (+5.2)	67.5 (+11.1)	85.4 (−2.4)	80.5 (+0.1)	59.9 (−14.5)	89.1 (−0.6)
SimSiam [4]	56.5 (+34.6)	9.7 (−0.3)	61.0 (+8.3)	66.2 (+9.8)	87.5 (−0.3)	80.1 (−0.3)	73.6 (−0.8)	89.8 (+0.1)
BYOL [7]	46.9 (+25.0)	8.6 (−1.4)	53.7 (+1.0)	52.7 (−3.7)	88.6 (+0.8)	80.7 (+0.3)	76.5 (+2.1)	89.5 (−0.2)
SimCLR [3]	48.7 (+26.8)	15.5 (+5.5)	61.3 (+8.6)	58.7 (+2.3)	86.9 (−0.9)	79.9 (−0.5)	74.3 (−0.1)	89.3 (−0.4)
w/o Z-Matching	49.1 (+27.2)	21.1 (+11.1)	55.8 (+3.4)	45.1 (−11.3)	88.3 (+0.5)	81.2 (+0.8)	81.3 (+6.9)	89.7
w/o Fundament	45.3 (+23.4)	0.0 (−10.0)	58.8 (+6.4)	48.5 (−7.9)	87.0 (−0.8)	79.5 (−0.9)	76.6 (+2.2)	89.0 (−0.7)
w/o Affine head	57.7 (+35.8)	17.9 (+7.9)	57.6 (+4.9)	53.4 (−3.0)	89.4 (+1.6)	**82.3** (+1.9)	79.8 (+5.4)	89.8 (+0.1)
Our GVSL (Whole)	**68.4** (+46.5)	**28.7** (+18.7)	60.8 (+8.1)	**79.9** (+23.5)	**91.2** (+3.4)	81.3 (+0.9)	**82.2** (+7.8)	**90.0** (+0.3)

- When **only learning the GM** (Z-Matching), its initial weak representability makes the pre-trained model have inefficient optimization and brings poor representation
- When **adding the fundamental task**, our GVSL has better performance than the single two sub-pretext tasks on all four downstream tasks.
- When **removing the Affine head** in the Z-Matching head, it reduces 3.2% and 2.4% AUC in the linear and fine-tuning evaluations of CCC task due to the lack of global representation learning.

Experiment:

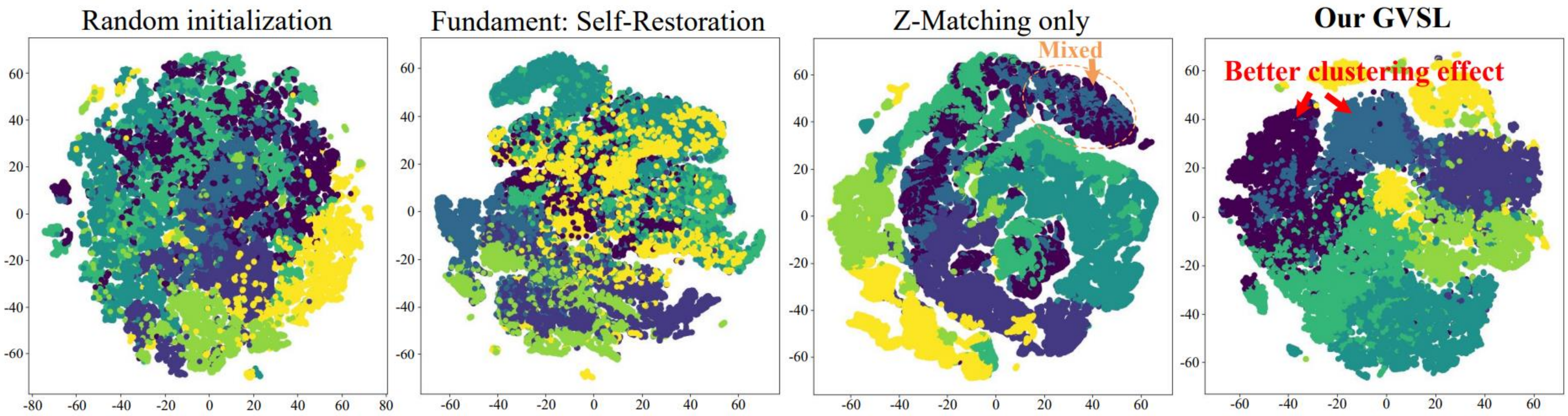
Ablation study



- ➢ Our powerful representability and much faster convergence ability.

Experiment:

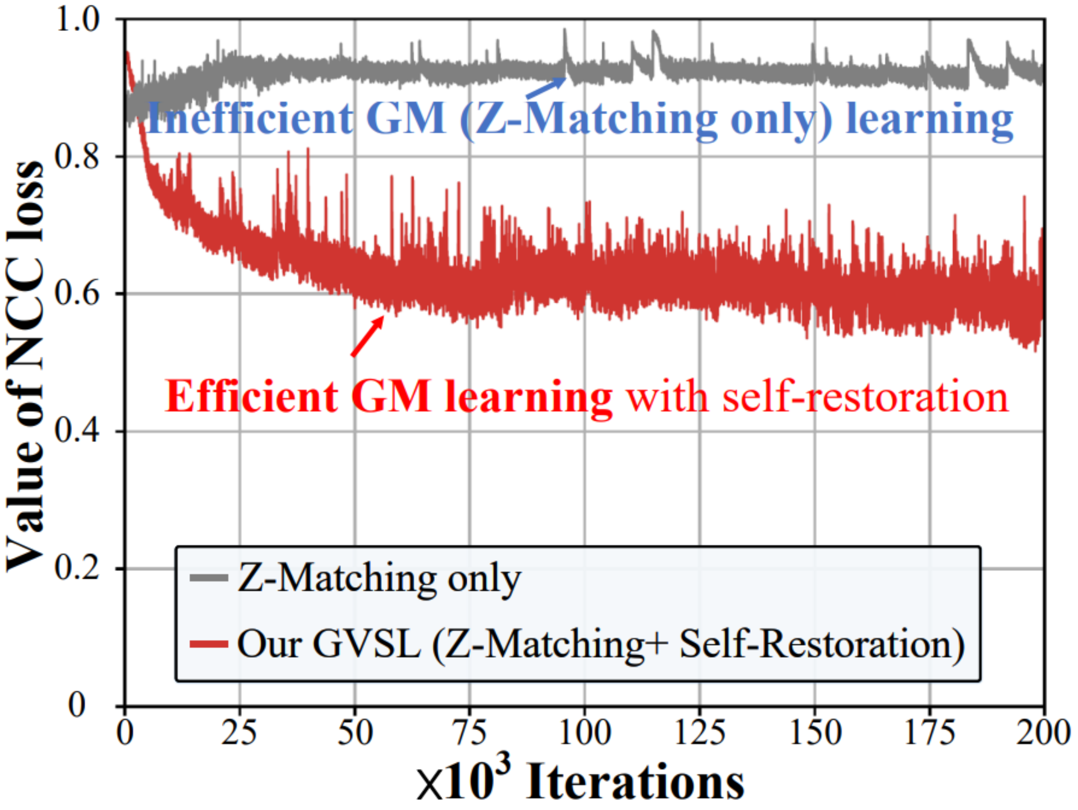
Ablation study



- Pre-trained models in the SHC task demonstrate our GVSL's promotion for the clustering effect.

Experiment:

Ablation study



- The self-restoration learns a basic representation for visual semantic regions, thus driving the learning of inter-image similarity in our GM.

Discussion and conclusion

伊宇霆

- **Conclusion of method:** Geometric Visual Similarity Learning based on the topological invariance of 3D medical images is a powerful prior for the representation pre-training of inter-image similarity;
- **Future work:** Expand the learning of inter-image similarity to some images without topological invariance, i.e., whole slide imaging.

Thanks, Q&A

He Yuting (何宇霆)
Southeast University