Geometric Visual Similarity Learning in 3D Medical Image Self-supervised Pre-training

Yuting He[1], Guanyu Yang[1*], Rongjun Ge[2], Yang Chen[1], Jean-Louis Coatrieux[3], Boyu Wang[4], Shuo Li[5]

Yuting He Paper

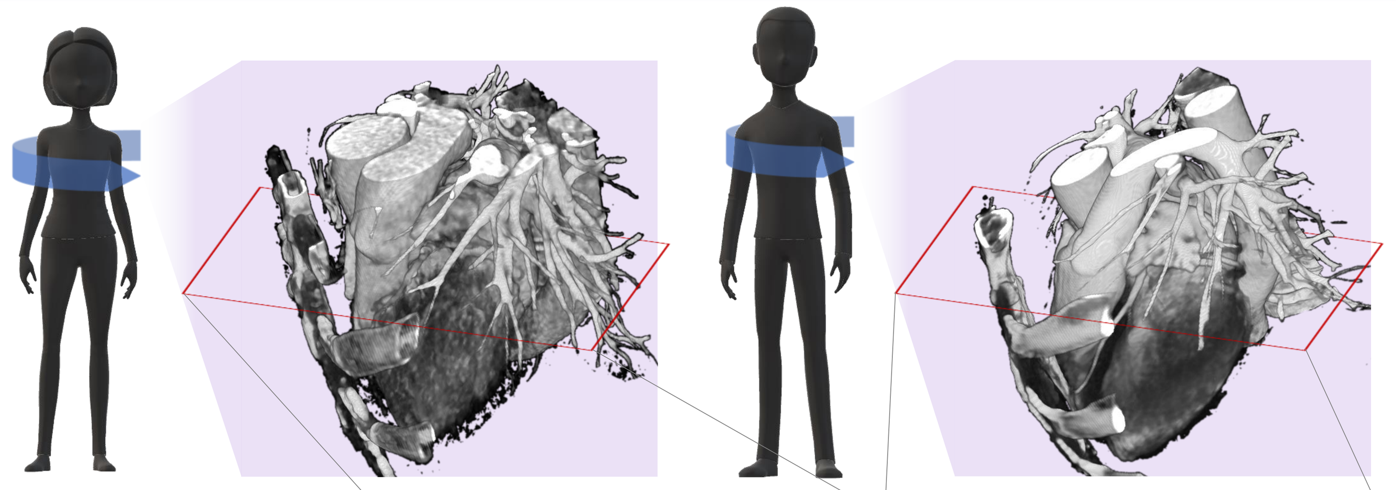
ytytyt_hhh@seu.edu.cn (Yuting He) yang.list@seu.edu.cn (Guanyu Yang)

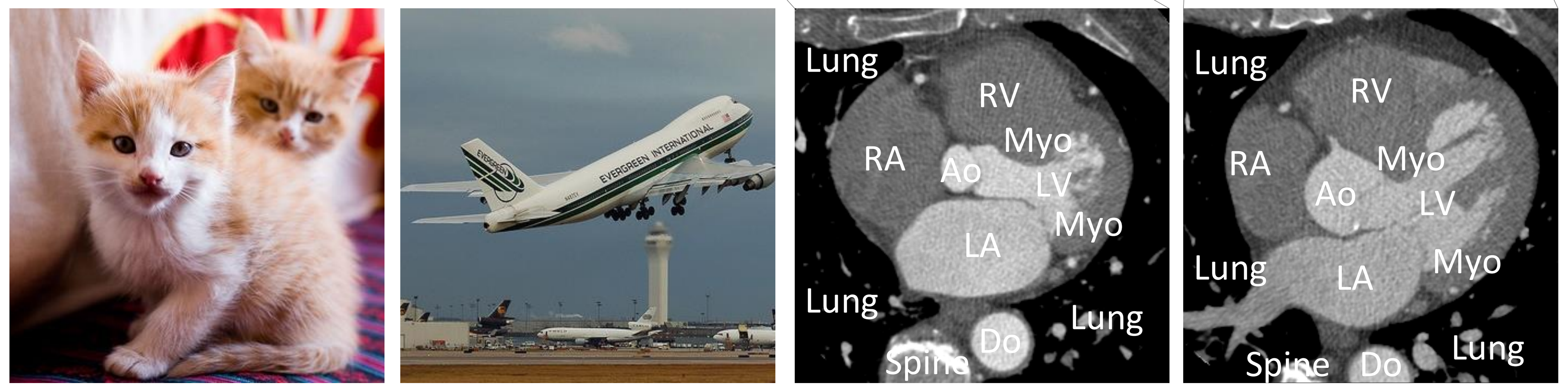
[1]Southeast University [2]Nanjing University of Aeronautics and Astronautics
[3]University of Rennes 1 [4]Western University [5]Case Western Reserve University

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

UWO SEU

- *Consistent human anatomies*
- *Complete spatial information in 3D vision*

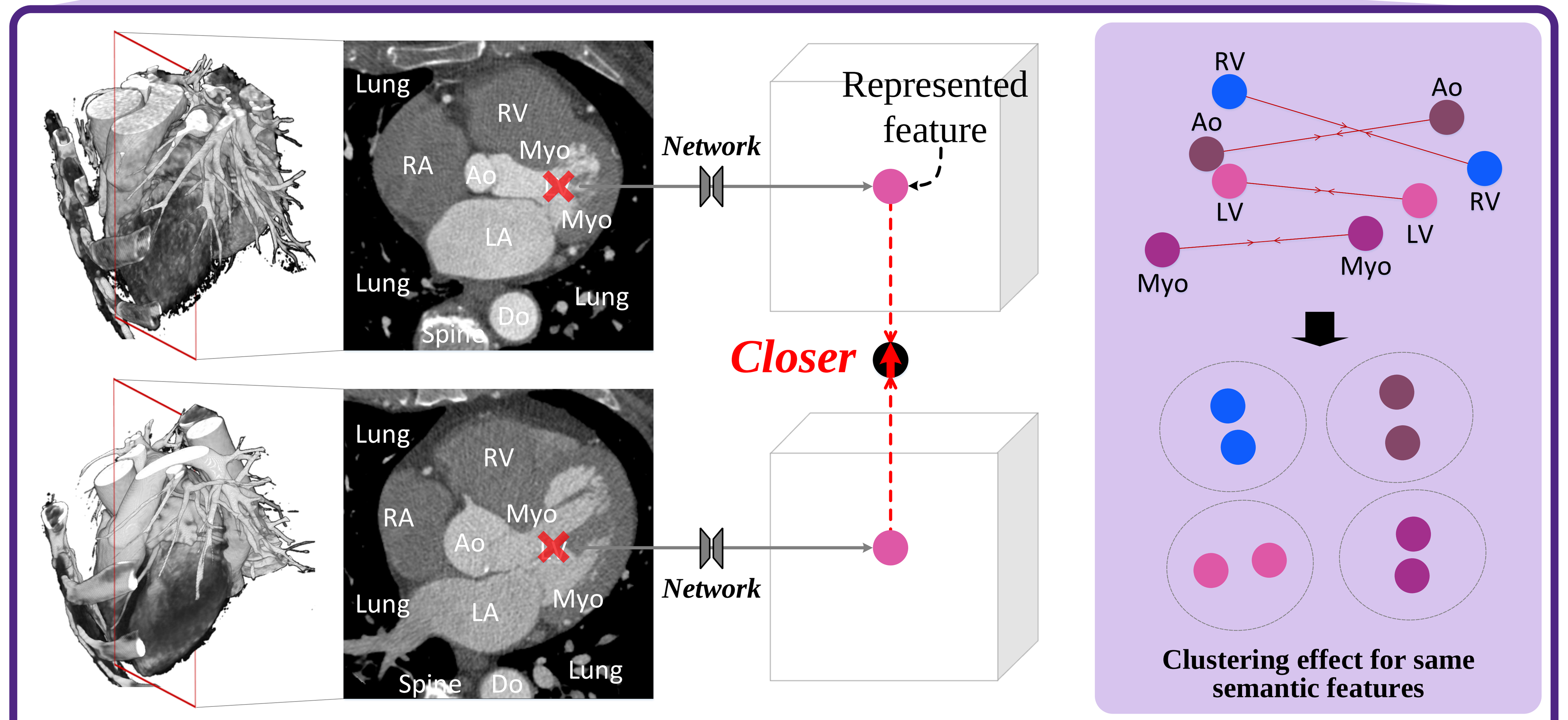




Natural images ✗
- ✓Scan from **large** scopes
- ✓**Nonlimited** range and pose
- ➤Large inter-image **difference**

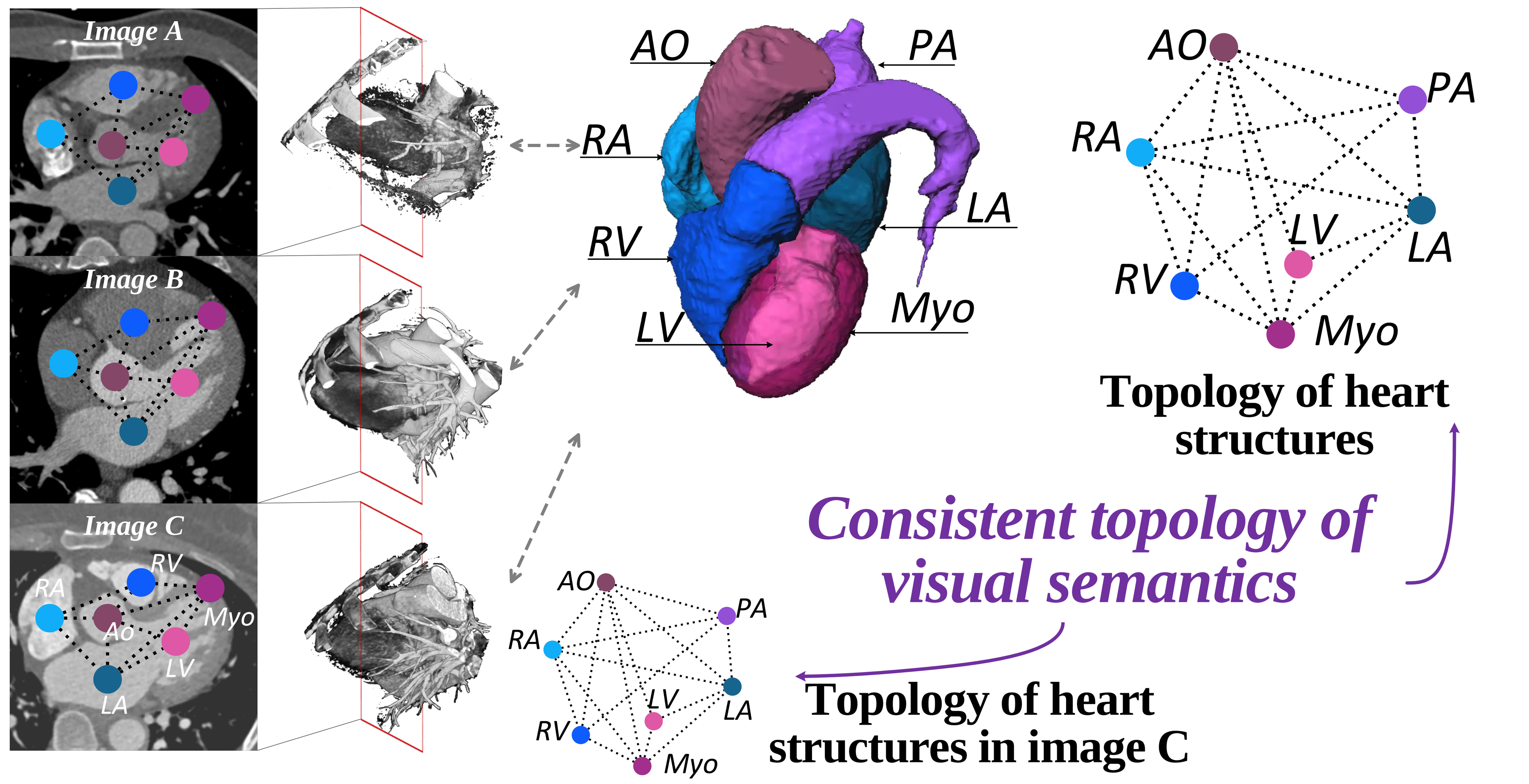
Medical images ✓
- ✓Scan from **small** scopes
- ✓**Limited** range and pose
- ➤Large inter-image **similarity**

Opportunity: Learning ***inter-image similarity*** **is crucial for 3D medical image (e.g., CT, MR) self-supervised pre-training (SSP).**



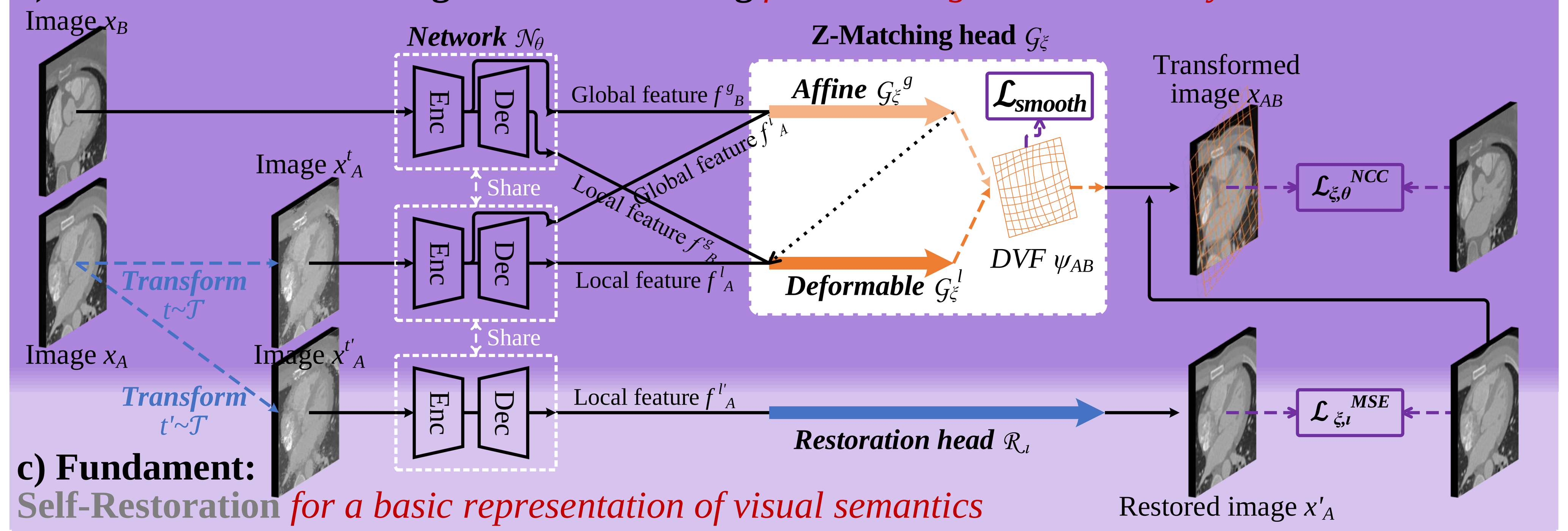
Target: Constrain pre-training network for ***consistent representation of these same semantic regions*** between images without annotations. Once successful, it will bring great clustering effect for same semantic features, powerful representability of pre-trained network, and effective transferring for potential downstream tasks.

Scientific question: How to discover the inter-image correspondence of same semantics for a reliable consistency learning?



Hypothesis: ***Topological invariance*** between 3D medical images is able to drive the discovery of inter-image correspondence, thus guiding networks to learn inter-image similarity.

a) Ours GVSL learns the geometric matching *for inter-image visual similarity*

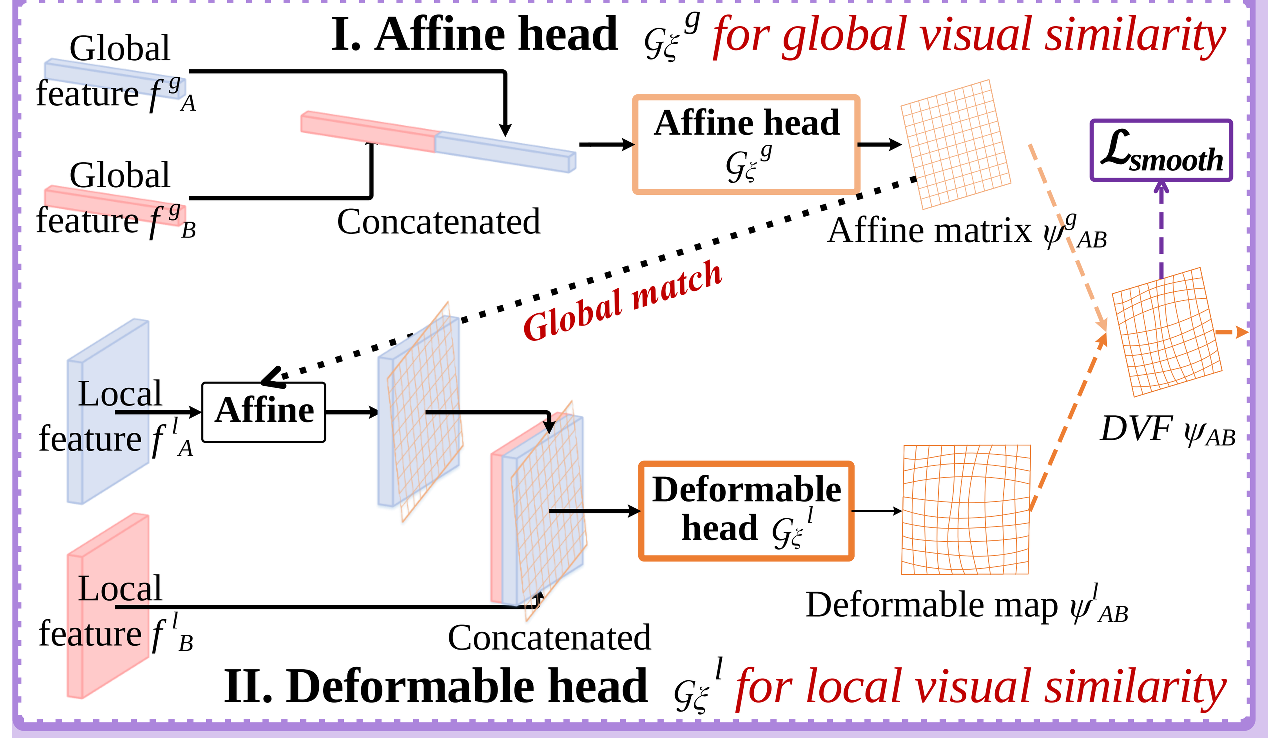


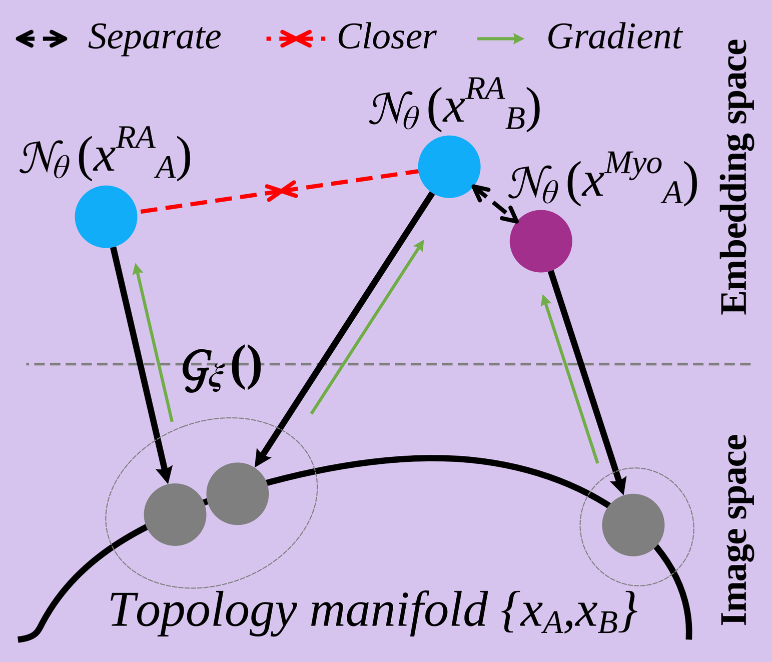
c) Fundament: Self-Restoration *for a basic representation of visual semantics*

Contribution: In this paper, we propose a novel SSP paradigm, ***Geometric Visual Similarity Learning (GVSL)***, to learn the inter-image similarity in 3D medical images. It embeds the prior of topological invariance into the measurement of the similarities, and train network to estimate semantics' correspondence from the represented features in geometric matching (GM).

- ✓ GVSL learns the geometric matching for the learning of inter-image similarity via the gradient;
- ✓ **Z-Matching head** learns affine and deformable matchings for global and local representations;
- ✓ Fundamental task, **self-restoration**, for a basic representation gives a warm-up for GM learning.

b) Z-Matching head $\mathcal{G}_\xi$ *learns global-local representation*

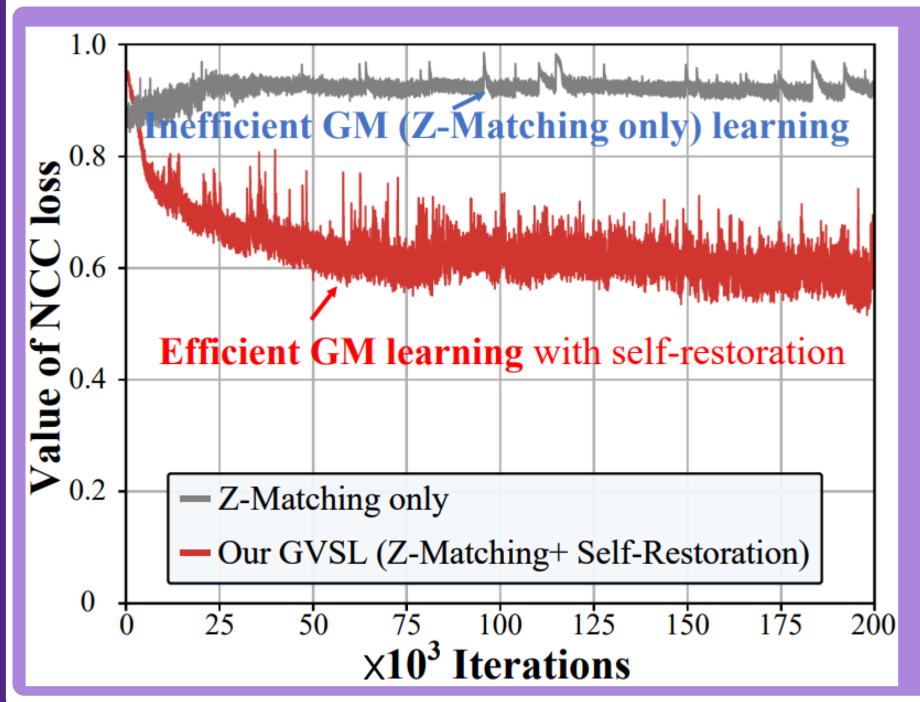




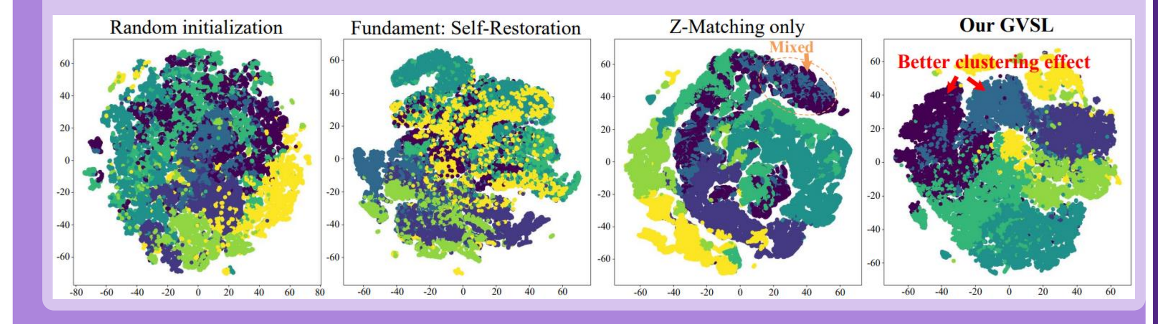
Intuition: Implicitly embed topology manifold inner the images into the measurement process, and measure the similarity on this topology manifold.

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)

Result 1:
1. Powerful ***inner-scene transferring*** for both large and small structures;
2. Effective ***inter-scene transferring***, but is not significant in fine-tuning;
3. Superiority in both ***global and dense*** prediction tasks.



Result 2: Pre-trained models in the SHC task show GVSL's promotion for the clustering effect.



Result 3: The self-restoration learns a basic representation for visual semantic regions, driving the learning of inter-image similarity in our GM.

Acknowledgements. This research was supported by the Intergovernmental Cooperation Project of the National Key Research and Development Program of China(2022YFE0116700), CAAI-Huawei MindSpore Open Fund, Natural Sciences and Engineering Research Council of Canada (NSERC). We thank the Big Data Computing Center of Southeast University for providing the facility support. We also thank the Key Laboratory of Computer Network and Information Integration, Ministry of Education, and the Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Nanjing, China.