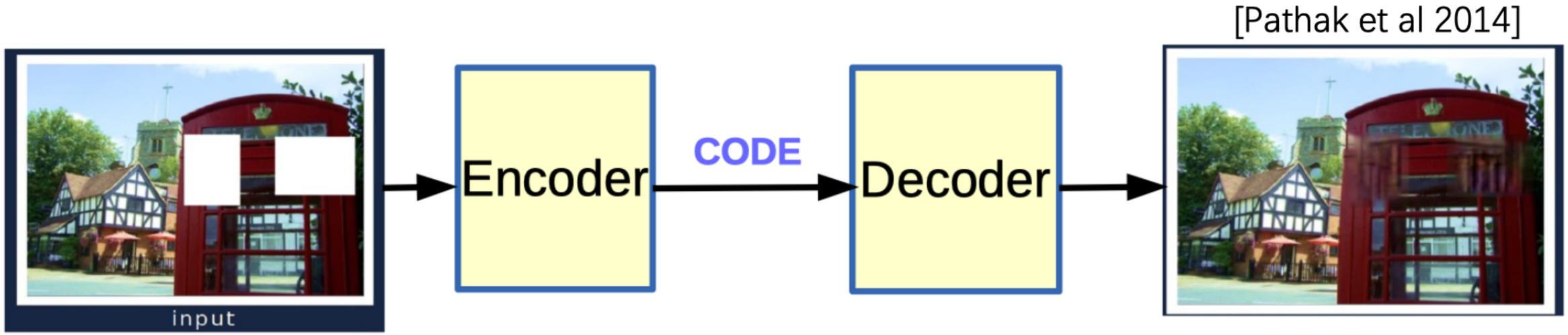
Self-supervised Transformer in Medical

Chongjie Ye

背景

自监督学习 - eg. 补全 in 2D



- 输入: 2D图片, 无label
- 目标: 让输出尽可能与输入相近
- Metrics: 原图片和补全图片的l2 loss
- 局限: 只作为单一的任务

背景

表征学习 Representation Learning

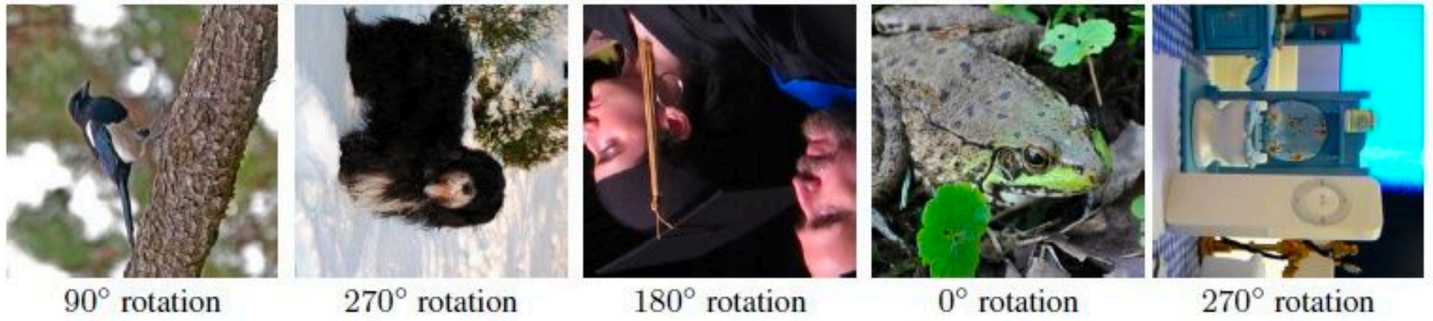


Figure 1: Images rotated by random multiples of 90 degrees (e.g., 0, 90, 180, or 270 degrees). The core intuition of our self-supervised feature learning approach is that if someone is not aware of the concepts of the objects depicted in the images, he cannot recognize the rotation that was applied to them.

- 输入: 2D图片, 无label
- 输出: 物体的旋转角度
- 动机: 如果模型需要理解图片的概念信息，如目标的位置，类别及姿势，才能识别出旋转角度
- 目的: 让网络学习提取有用的特征

背景

迁移学习 - Pretrained on ImageNet

Pretraining on ImageNet Classification

Finetuning

Semantic Segmentation

Object Detection

Fine-grained Classification

- 局限: 1. 预训练网络都是2D卷积; 2. 自然图像的pattern和医疗图像很不一致

背景

Transformer

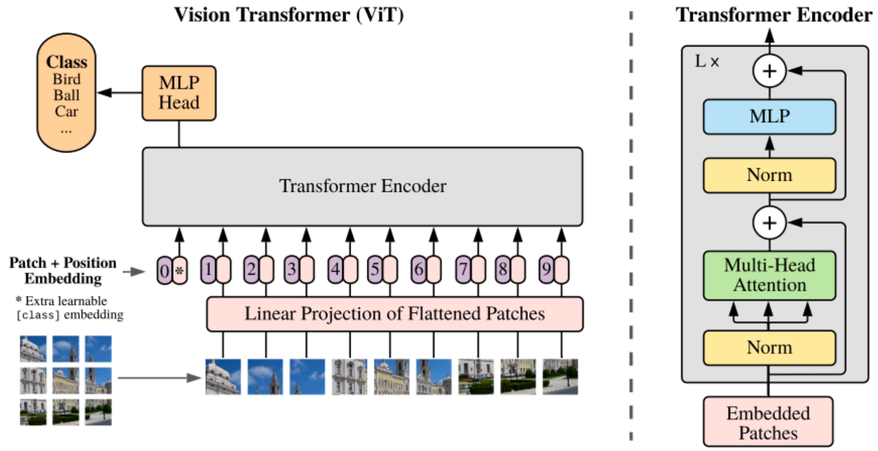


Figure 1: Model overview. We split an image into fixed-size patches, linearly embed each of them, add position embeddings, and feed the resulting sequence of vectors to a standard Transformer encoder. In order to perform classification, we use the standard approach of adding an extra learnable "classification token" to the sequence. The illustration of the Transformer encoder was inspired by Vaswani et al. (2017).

Category	Sub-category	Method	Highlights	Publication
Bakcbone	Image classification	iGPT [21]	Pixel prediction self-supervised learning, GPT model	ICML 2020
		ViT [36]	Image patches, standard transformer	ICLR 2021
High/Mid-level vision	Object detection	DETR [15]	Set-based prediction, bipartite matching, transformer	ECCV 2020
		Deformable DETR [193]	DETR, deformable attention module	ICLR 2021
		ACT [189]	Adaptive clustering transformer	arXiv 2020
		UP-DETR [33]	Unsupervised pre-training, random query patch detection	arXiv 2020
		TSP [143]	New bipartite matching, encoder-only transformer	arXiv 2020
	Segmentation	Max-DeepLab [155]	PQ-style bipartite matching, dual-path transformer	arXiv 2020
		VisTR [159]	Instance sequence matching and segmentation	arXiv 2020
		SETR [190]	sequence-to-sequence prediction, standard transformer	arXiv 2020
	Pose Estimation	Hand-Transformer [67]	Non-autoregressive transformer, 3D point set	ECCV 2020
		HOT-Net [68]	Structured-reference extractor	MM 2020
		METRO [92]	Progressive dimensionality reduction	arXiv 2020

截止2020年10月

Medical Transformer: Universal Brain Encoderfor 3D MRI Analysis

Eunji Jun, Student Member, IEEE, Seungwoo Jeong, Da-Woon Heo, and Heung-Il Suk, Member, IEEE

Motivation

- 医疗数据集标注成本太高
- 有文献表明, 用经过pretrained后的网络参数相比随机初始化的参数进行finetune最终效果更好, 尤其是在任务训练集很小的情况下
- 之前的方法经常将将MRI或CT切割成2D volumns做pretraining, 但是会损失3D形状信息

问题定义

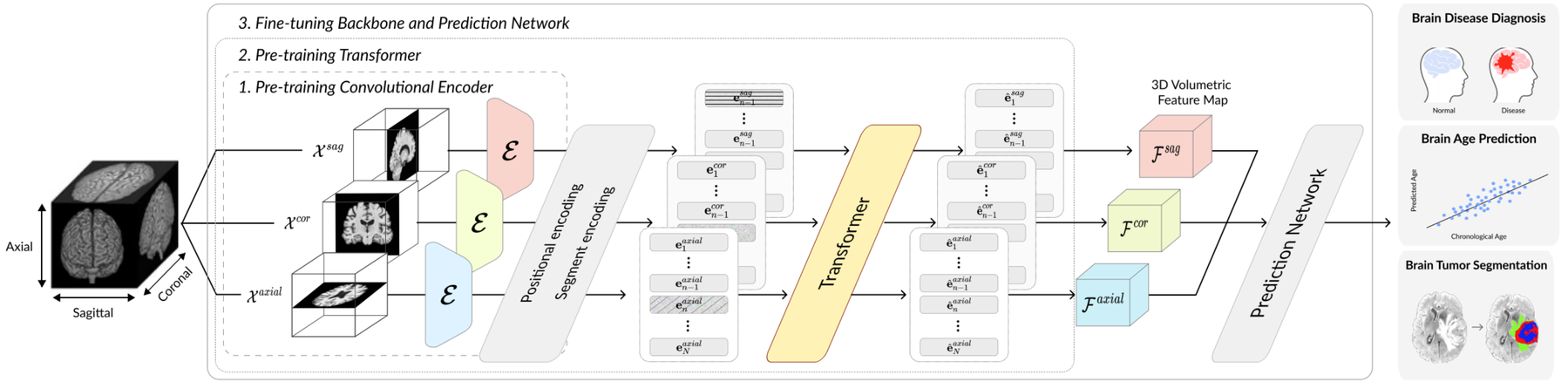
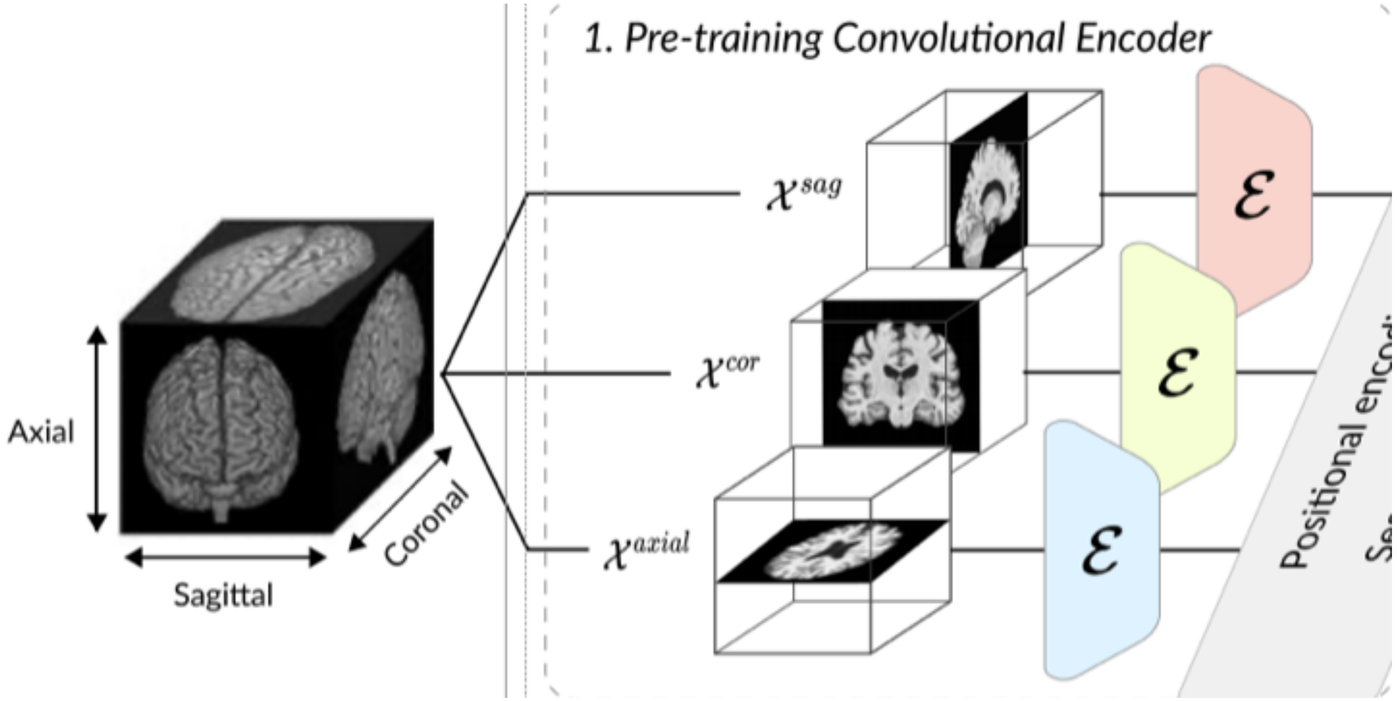


Fig. 1: Schematic diagram of the proposed Medical Transformer. Based on a multi-view approach, a given 3D volumetric image is split into 2D slices from three planes (sagittal, coronal, axial), and these 2D image slices are fed to the network as inputs. First of all, we pre-train a backbone network that consists of a convolutional encoder and a transformer in a self-supervised learning scheme. Then, after passing through the pre-trained backbone network, the 2D slice features are recovered by their combinations into 3D-form representations, and finally fed into the prediction network for three medical imaging tasks.

- 输入: a large-scale 3D brain MRI (from IXI, Cam-CAN, ABIDE)
- 目的: 预训练网络, 以迁移到所有基于3D brain MRI的子任务

网络结构

- convolutional encoder

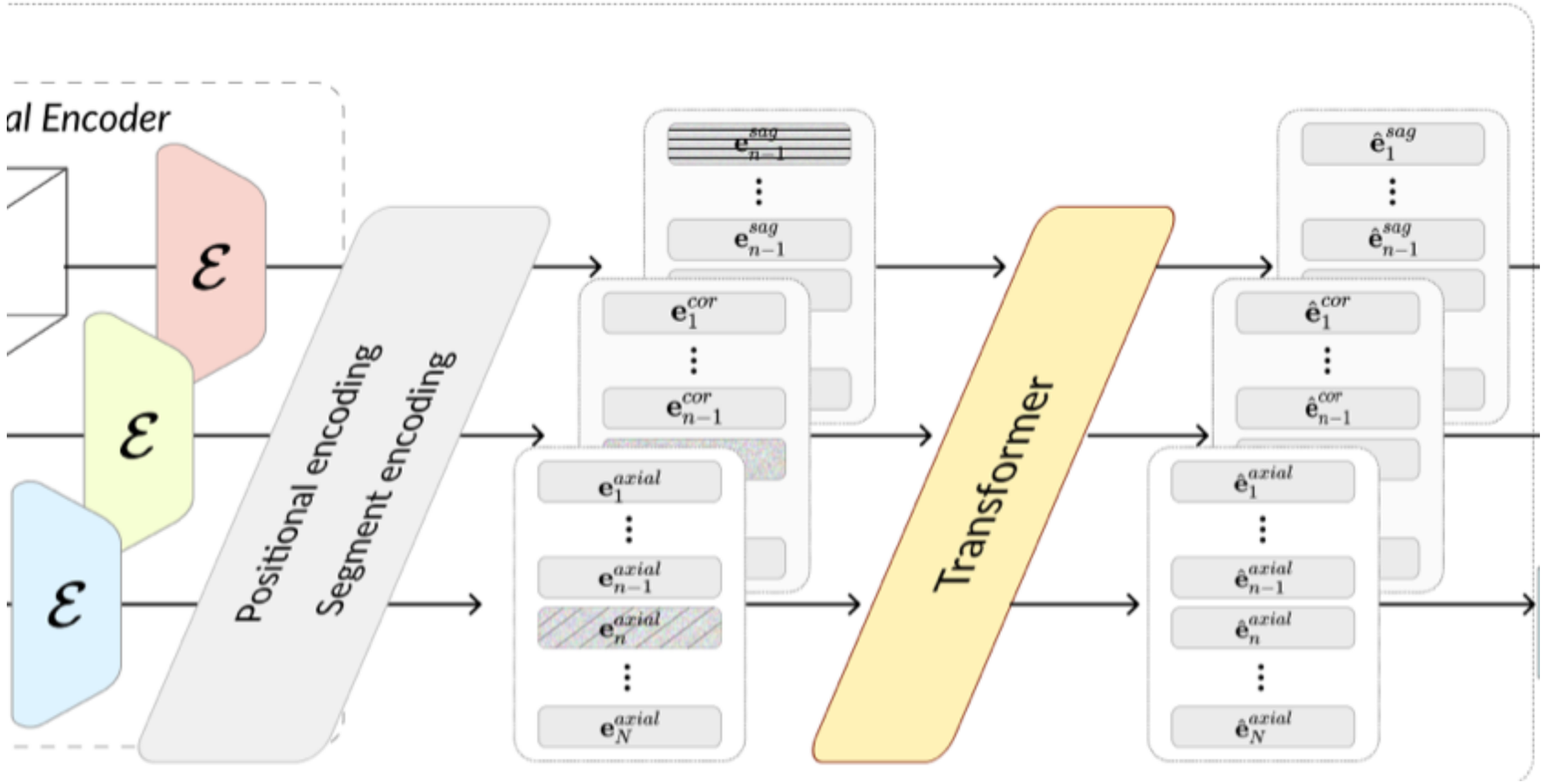


- 将MRI按照三个方向分别切割, 得到3组 2D volume
- 输入conv中提取特征

网络结构

- transformer network

动机: 通过Transformer的自注意力机制, 提取2D slicing之间的相互关系.



- 每一组slicing输入一个共享权重的transformer做特征增强
- 对于每组slicing, 先对每个2D volume做max pooling,得到一个一维的embedding vector, 然后作为一个sequence 输入transformer中

Pretraining Method

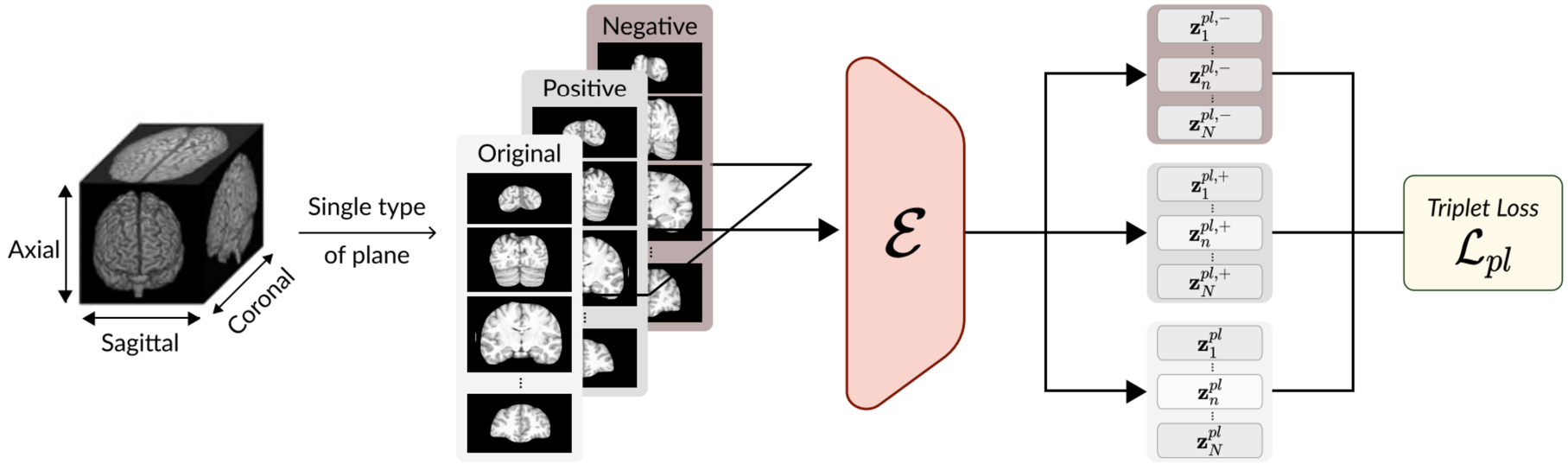


Fig. 2: Illustration of pre-training a convolutional encoder.

Results

Pre-training	Approach	Target tasks				
		Classification	Regression	Segmentation		
		mAUC	MAE (years)	Dice WT	Dice TC	Dice ET
No	Training from scratch	0.7728 ± 0.0077	4.4357 ± 0.3260	0.8649 ± 0.0094	0.6392 ± 0.0556	0.4830 ± 0.0485
(Fully) supervised	I3D [7]	0.7325 ± 0.0164	4.6561 ± 0.3209	0.6607 ± 0.0542	0.4708 ± 0.0593	0.0569 ± 0.0159
	NiftyNet [8]	0.5031 ± 0.0165	4.6580 ± 0.3161	0.8395 ± 0.0065	0.5295 ± 0.0148	0.5046 ± 0.0278
	MedicalNet [9]	0.6910 ± 0.0063	4.6443 ± 0.3626	0.7885 ± 0.0378	0.5681 ± 0.0572	0.0809 ± 0.0298
Self-supervised	3D-RPL [13]	0.4849 ± 0.0333	5.1237 ± 0.7086	0.8555 ± 0.0462	0.6595 ± 0.0322	0.3897 ± 0.0078
	3D-Rotation [13]	0.4965 ± 0.0077	4.9799 ± 0.4365	0.8672 ± 0.0344	0.6756 ± 0.0204	0.3717 ± 0.0452
	3D-Jigsaw [13]	0.4950 ± 0.0202	4.7719 ± 0.4784	0.8671 ± 0.0453	0.6739 ± 0.0218	0.3789 ± 0.0415
	3D-CPC [13]	0.4943 ± 0.0109	5.0091 ± 0.7856	0.8879 ± 0.0089	0.6844 ± 0.0086	0.3760 ± 0.0159
	3D-Exemplar [13]	0.5085 ± 0.0163	5.4434 ± 0.9623	$\mathbf{0.8975 \pm 0.0123}$	0.6912 ± 0.0120	0.3819 ± 0.0134
	Model Genesis [14]	0.4997 ± 0.0004	4.6377 ± 0.3411	0.8505 ± 0.0203	0.6201 ± 0.0289	0.0896 ± 0.0329
	Medical Transformer (Ours)	$\mathbf{0.8347 \pm 0.0072}$	$\mathbf{3.4924 \pm 0.0863}$	0.8733 ± 0.0086	$\mathbf{0.6969 \pm 0.0470}$	$\mathbf{0.5882 \pm 0.0437}$