MIS-FM: 基于大规模自监督预训练模型的 3D 医学图像分割

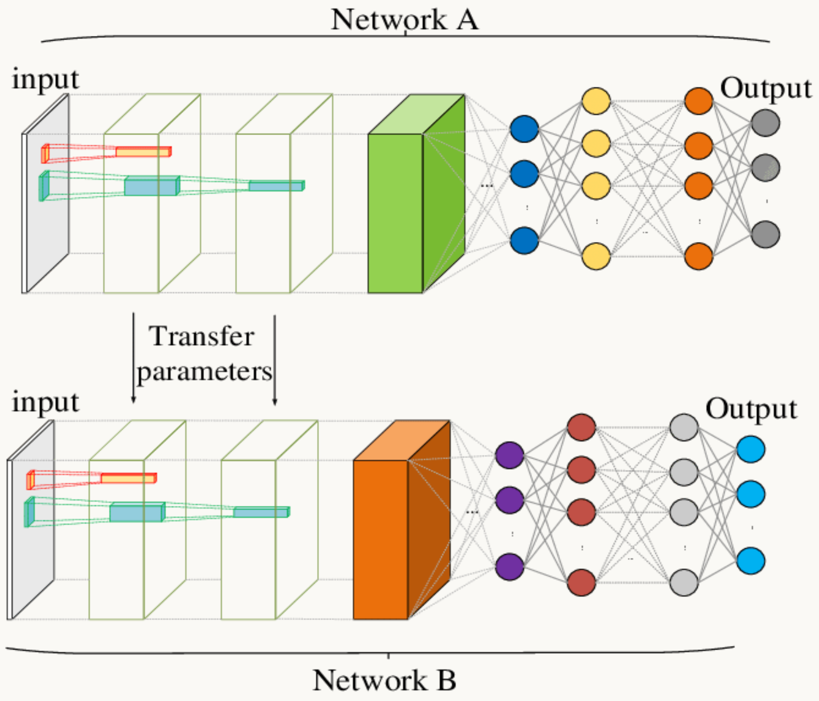
报告人：王国泰
电子科技大学
2023 年 10 月11 日

为什么使用预训练模型？

预训练模型

- 通常在一个大的数据集上进行预训练，学习到较通用的特征
- 从而可以迁移到一个较小的目标数据集，提高模型性能，降低过拟合

预训练数据（如ImageNet）

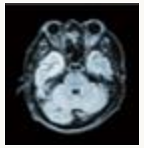


迁移学习

目标数据集

预训练模型的三大要素

数据集

	大规模获取成本	与医学图像匹配度
自然图像	低	低
医学图像	高	高

大规模三维医学图像数据集?

模型结构

- CNN
  - 模型较小，训练较快，表达能力稍弱
- Transformer
  - 模型更大，计算量大，表达能力较强

三维医学图像分割的模型?

预训练方法

- 全监督训练
  - 对特征的引导性强，标注成本高、甚至无法获取
- 自监督训练
  - 无需人工标注，需要设计合理的方式学习特征表达

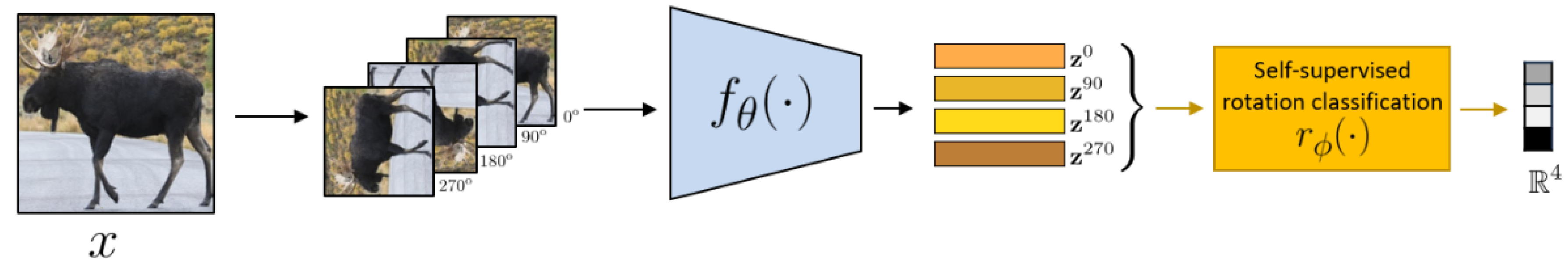
适合三维分割的自监督方法?

三维医学图像分割预训练模型

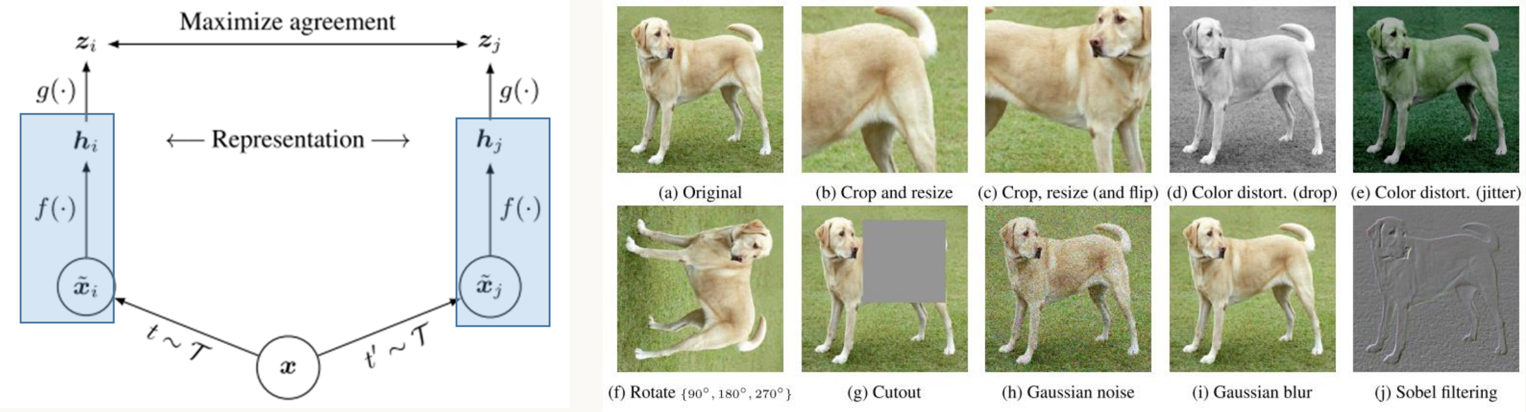
自监督预训练策略：1）训练特征提取器

可通过自监督图像分类、对比学习等训练特征提取器

旋转角度预测[1]



对比学习[2]



在不需要人工标注的情况下，增强模型对图像内容的理解能力

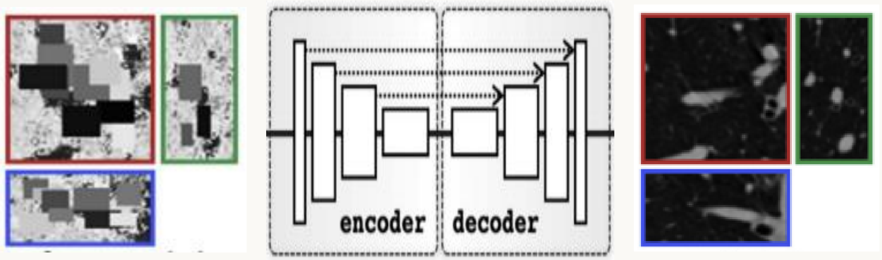
缺点：图像分割通常还需要一个解码器，这些方法只训练了编码器

[1] S. Gidaris et al., Unsupervised Representation Learning by Predicting Image Rotations, ICLR 2018
[2] T. Chen et al. A simple framework for contrastive learning of visual representations, 2020

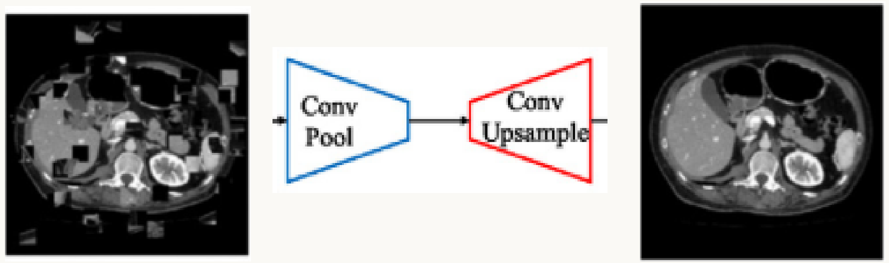
自监督预训练策略：2）训练编码器-解码器

基于图像重建任务，训练编码器-解码器结构

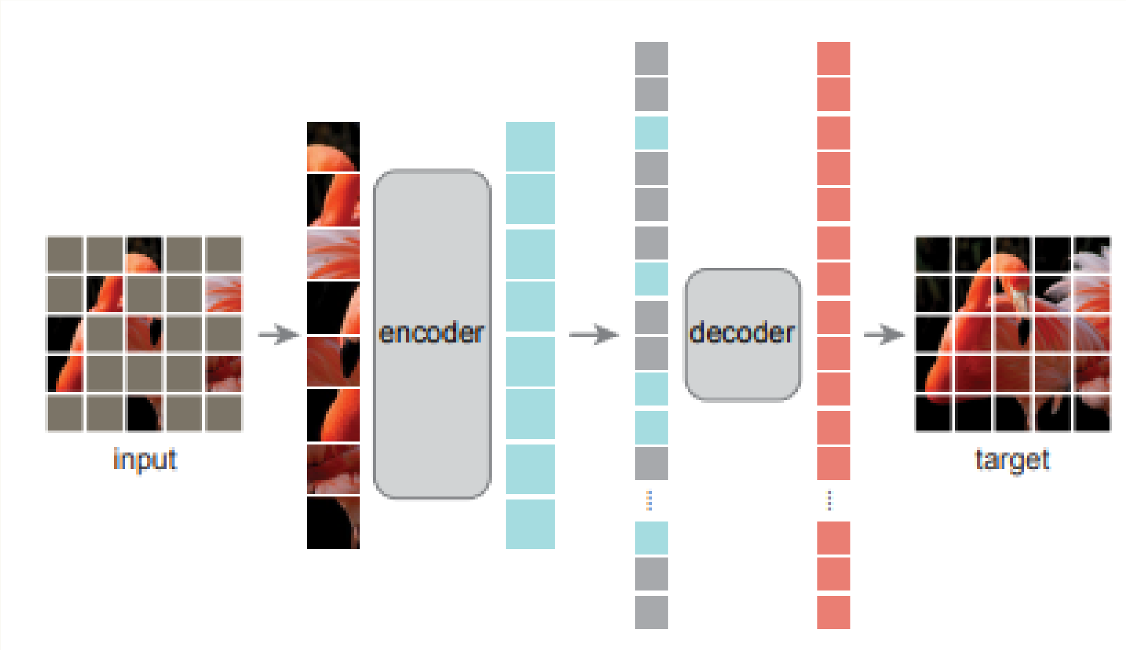
Model Genesis[1]



Patch Swapping[2]



Masked Auto Encoder[3]



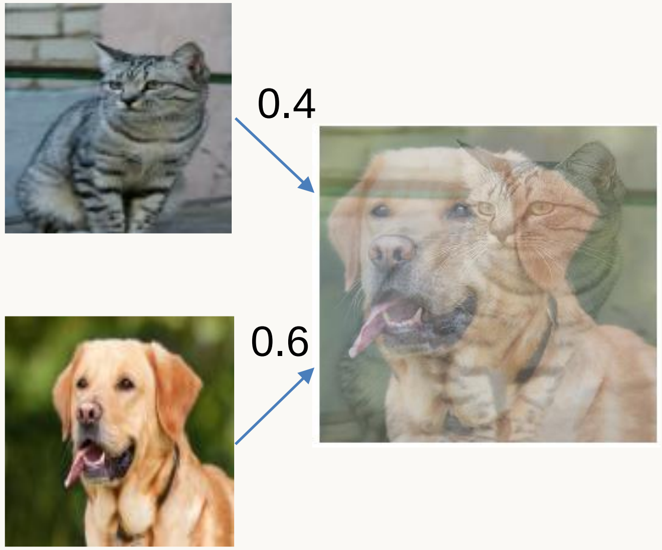
缺点：图像重建和分割是不同的任务，二者之间需要的特征可能不匹配

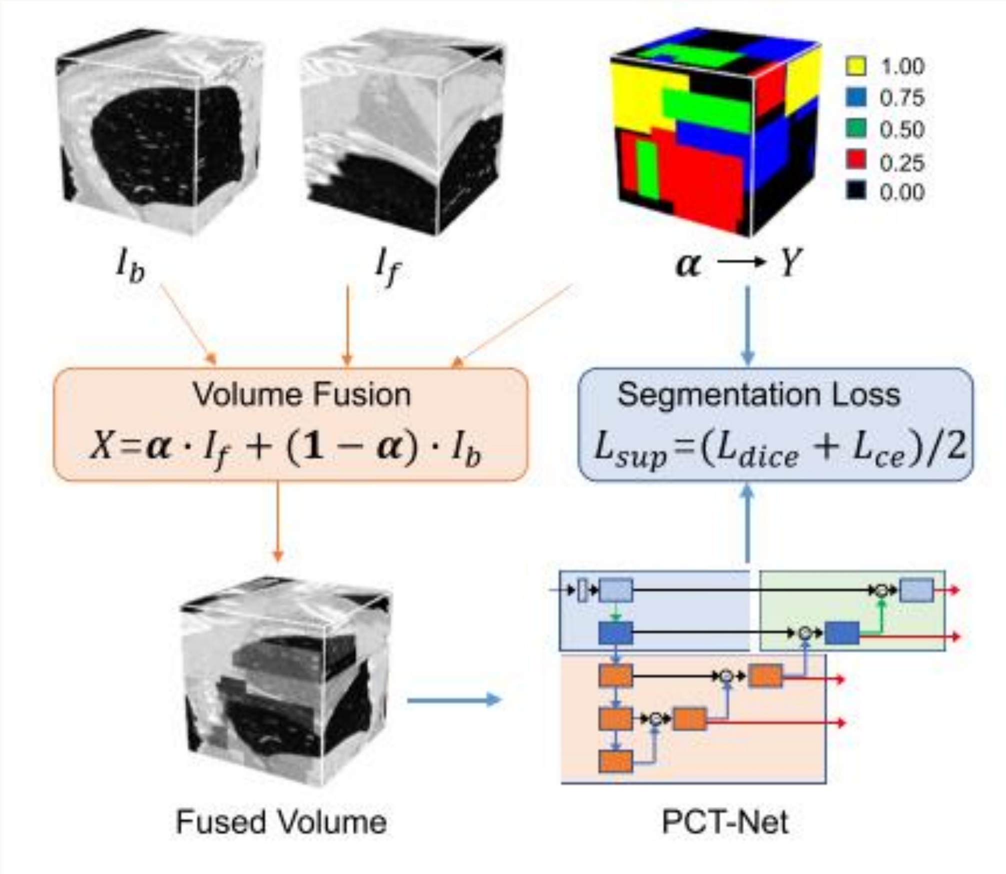
[1] Z. Zhou et al., Model Genesis, MedIA 2021
[2] L. Chen et al., Self-supervised learning for medical image analysis using image context restoration, MedIA 2019
[3] K. He et al., Masked autoencoders are scalable vision learners, CVPR 2022

方法：1）基于Volume Fusion的自监督训练策略

Volume Fusion: 将预训练过程表示为一个自监督图像分割任务

来自Mix-up[1]的启发

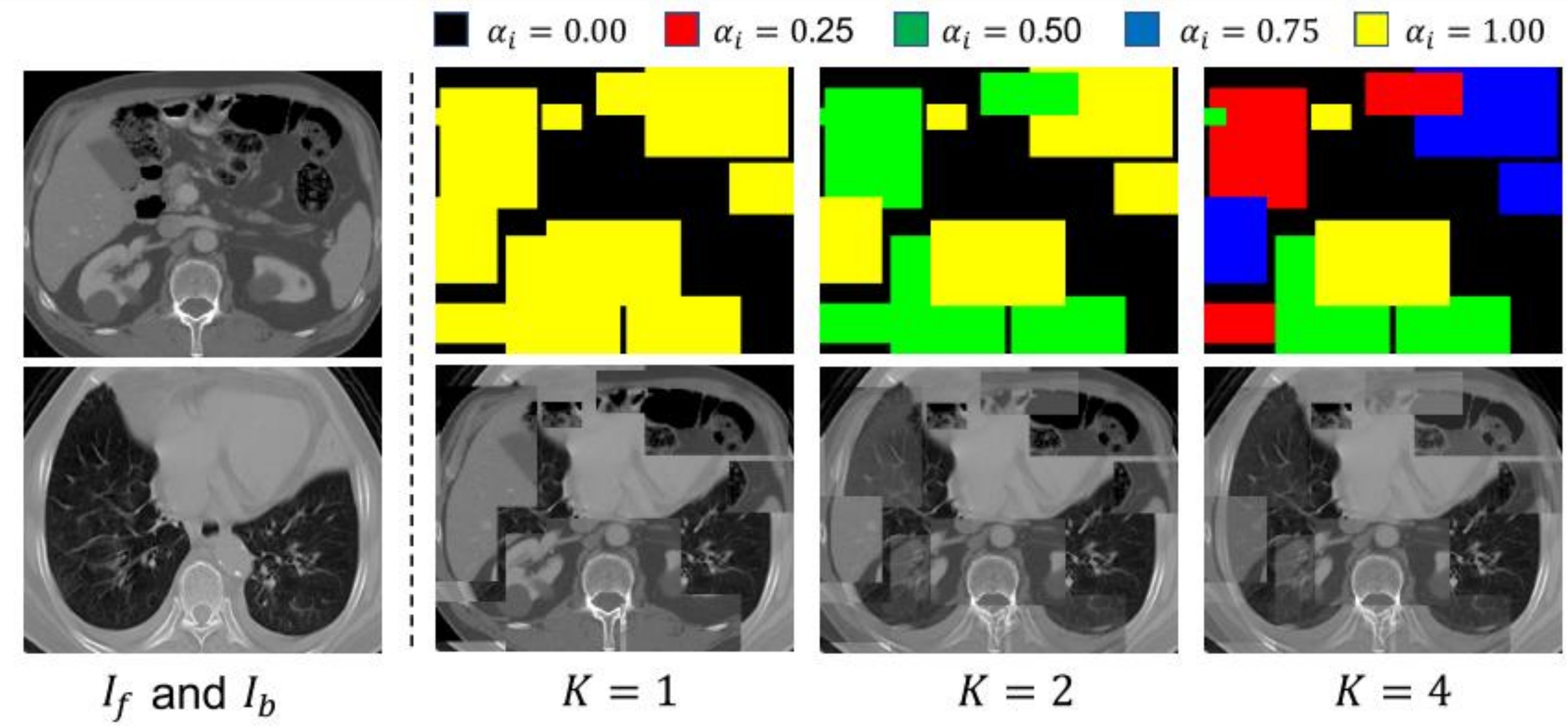




[1] H. Zhang et al, Mixup: Beyond Empirical Risk Minimization, ICLR 2018

方法：1）基于Volume Fusion的自监督训练策略

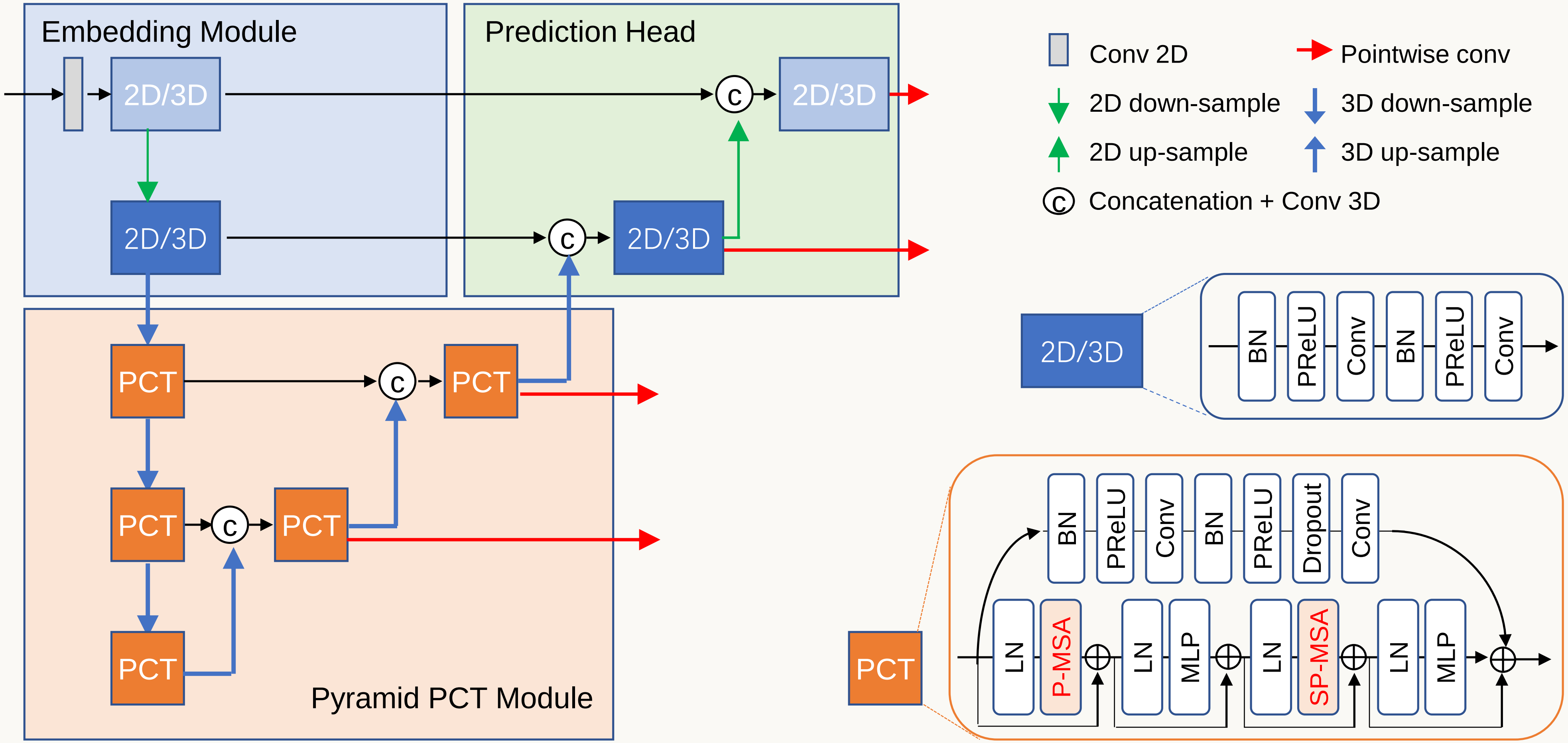
Volume Fusion可有效提升模型对图像中上下文（结构信息）的感知和识别能力



$$X = \alpha I_f + (1 - \alpha) I_b$$

G. Wang et al., MIS-FM: 3D Medical Image Segmentation using Foundation Models Pretrained on a Large-Scale Unannotated Dataset, arxiv 2023 https://arxiv.org/pdf/2306.16925.pdf

方法：2）PCT-Net 基于并行的卷积-Transformer结构的分割网络



- Embedding Module: 通过2D或3D卷积实现，保留高分辨率，形成局部特征表达
- Pyramid PCT Module: 卷积与Transformer相结合，融合局部与全局特征
- Prediction Head: 产生高分辨率分割结果，并得到多尺度预测

G. Wang et al., MIS-FM: 3D Medical Image Segmentation using Foundation Models Pretrained on a Large-Scale Unannotated Dataset, arxiv 2023 https://arxiv.org/pdf/2306.16925.pdf

方法：3）3D预训练数据集

以往工作的3D医学图像预训练数据集大小为几百到5k左右

方法	预训练数据集	大小	标注情况
Model Genesis[1]	LUNA16	623	无标注
Swin UNETR[2]	LUNA16, TCIA Covid19 LiDC, HNSCC, TCIA Colon	5050	无标注
CLIP-Driven Universal Model[3]	Pancreas CT, LiTS, KiTS, WORD等16个公开数据集	3410	部分标注
STU-Net[4]	TotalSegmentor	1204	全标注

[1] Z. Zhou et al., Model Genesis, MedIA 2021
[2] Y. Tang et al., Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis, CVPR 2022
[3] J. Liu et al., CLIP-driven universal model for organ segmentation and tumor detection, ICCV 2023
[4] Z. Huang et al., STU-Net : Scalable and transferable medical image segmentation models empowered by large-scale supervised pre-training, arxiv 2023

方法：3）3D预训练数据集

我们将预训练3D图像的规模扩大到1万和10万量级

Body part	Name	Pathology	Cases
Head-Neck	Head-Neck-PET-CT	Head and neck cancer	298
	OPC-Radiomics	Oropharynx cancer	606
	HNSCC	Head and neck cancer	627
	QIN-HeadNeck	Head and neck cancer	279
	TCGA-HNSC	Head and neck cancer	227
Chest	LUNA	Lung nodule	888
	TCIA Covid19	Covid19	753
	LIDC-IDRI	Lung nodule / cancer	1,018
Abdomen	FLARE22	Abdominal lesions	2,300
	AbdomenCT-1K	Abdominal lesions	1,112
	ACRIN 6664	Colon cancer	825
Whole body	TotalSegmentator	Various pathologies	1,204
Chest	Private	Lung diseases	103k

PData-1k
(选取一部分)

PData-10k

PData-110k

GitHub: https://github.com/openmedlab/MIS-FM

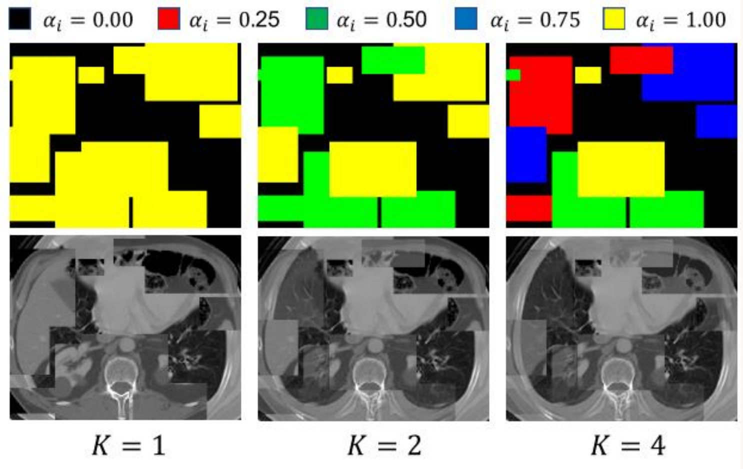
G. Wang et al., MIS-FM: 3D Medical Image Segmentation using Foundation Models Pretrained on a Large-Scale Unannotated Dataset, arxiv 2023 https://arxiv.org/pdf/2306.16925.pdf

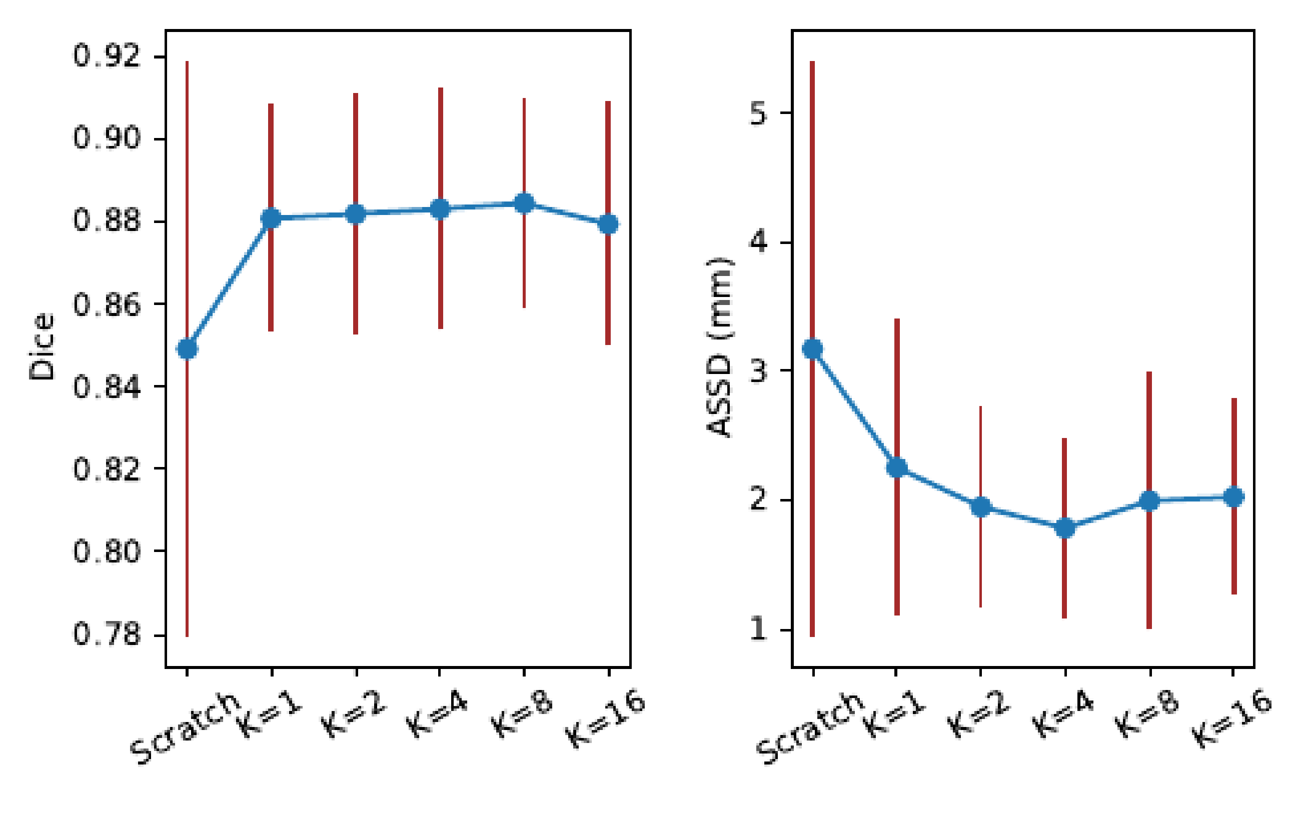
实验设置：下游分割任务数据集

基于大规模3D CT图像的自监督预训练

- CT图像中器官的分割
  - 头颈部：MICCAI 2015 Head-Neck
  - 胸部：SegTHOR
  - 腹部：Synapse
- MRI中器官的分割
  - 腹部：CHAOS
  - 前列腺：NCI_ISBI13
  - 左心室：Left Atrial
- 病灶的分割
  - 肝脏肿瘤：LiTS

实验结果：1）超参数K的影响





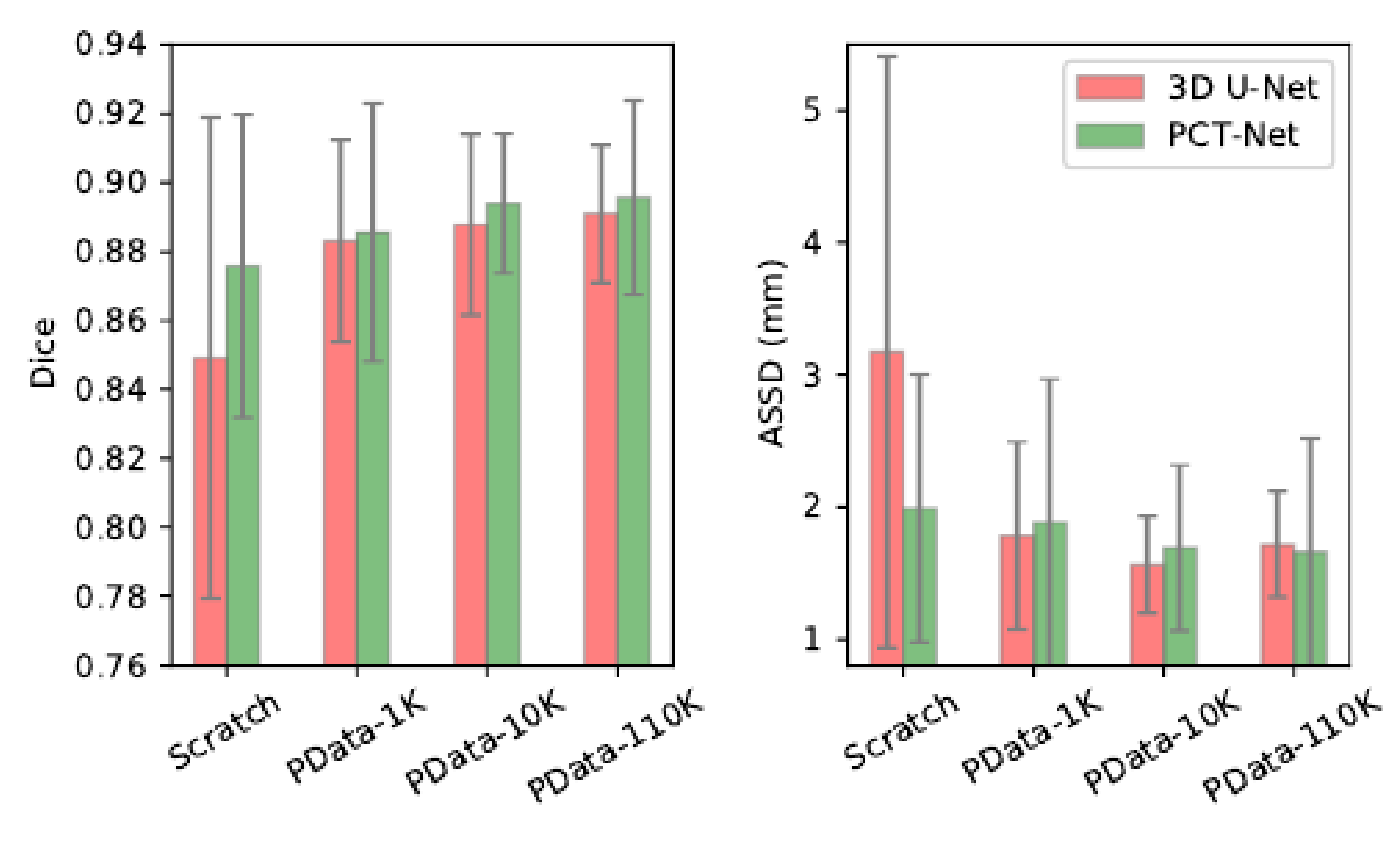
预训练：PData-1k
模　型：3D UNet
下游任务：SegTHOR

K = 1 到 8 时Dice值较平稳，显著优于不用预训练的结果

K = 4 时综合表现最好

实验结果：2）不同预训练数据规模和模型的比较

预训练：PData-1k，10k, 110k
模　型：3D UNet，PCT-Net
下游任务：SegTHOR

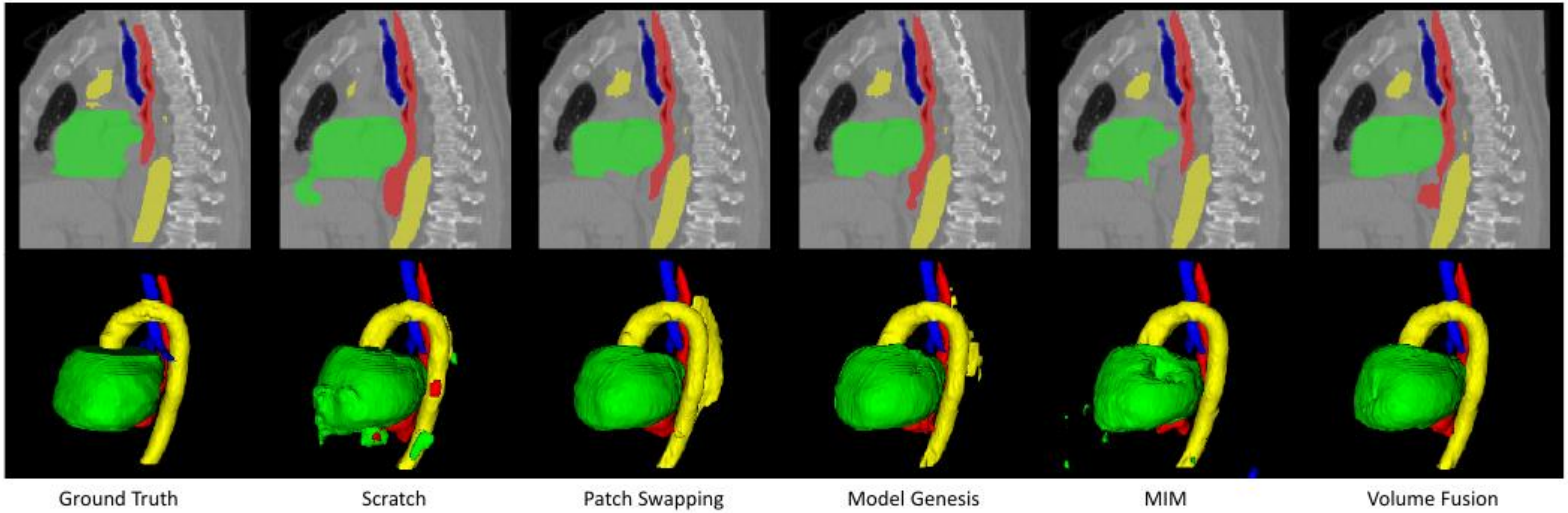


模型性能随着预训练数据规模的增加而提升
PCT-Net比3D U-Net的表现更好

实验结果：3）不同预训练策略的比较

预训练：PData-1k
模　型：3D UNet
下游任务：SegTHOR

Method	Dice (%)				
	Esophagus	Heart	Trachea	Aorta	Average
Scratch	71.87±11.27	90.82±4.72	87.63±5.77	89.34±6.19	84.92±6.47
Patch Swapping [1]	75.91±9.04	91.88±5.27	87.31±6.45	89.64±7.60	86.18±4.05
Model Genesis [2]	76.98±8.29	92.78±3.13	87.87±6.28	89.64±7.74	86.81±4.00
MIM [3]	76.29±8.91	91.85±2.82	87.52±5.43	92.19± 2.98	86.97±3.50
Volume Fusion	**77.61±7.82**	**93.72±2.28**	**88.21±4.18**	**93.67±1.68**	**88.30±2.93**

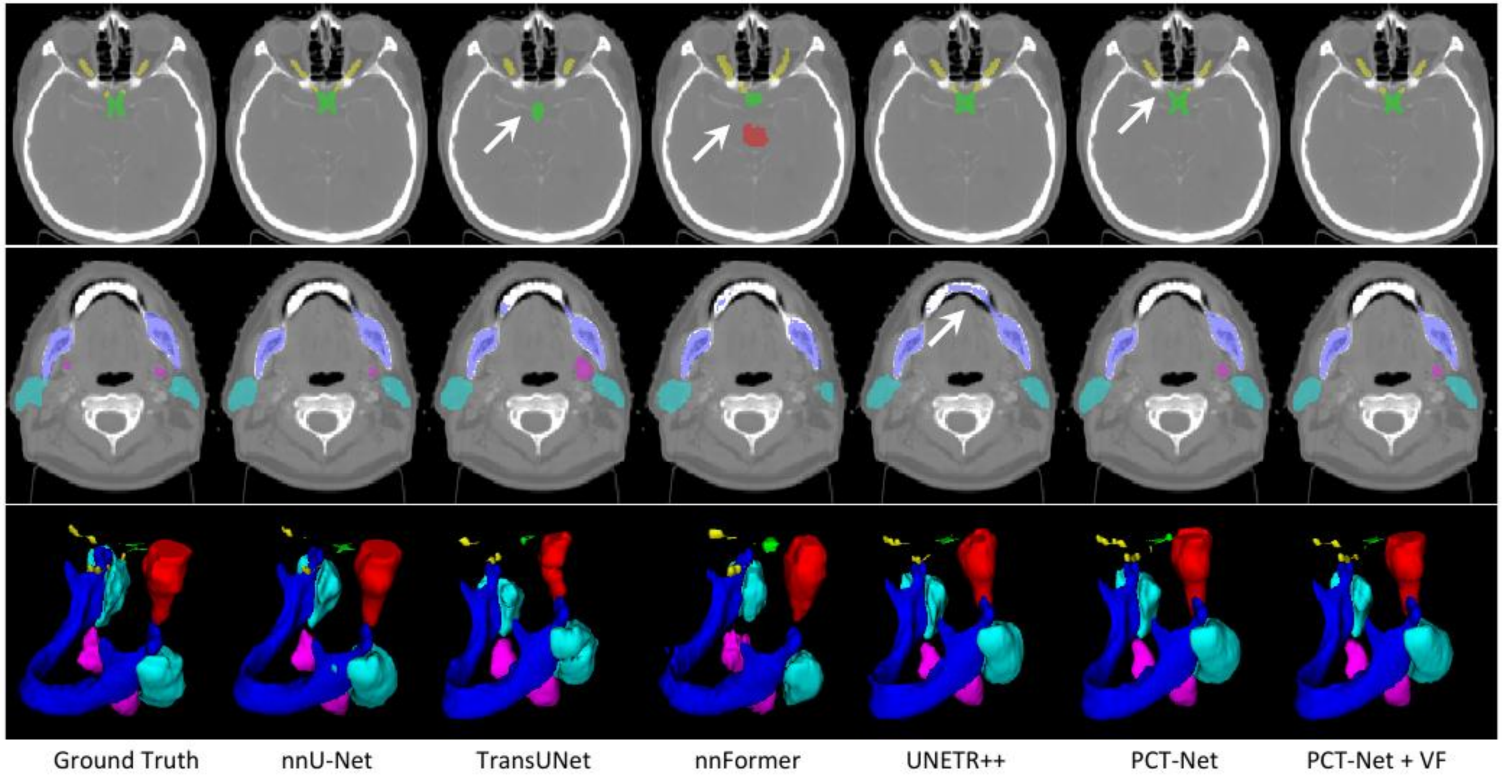


[1] L. Chen et al., Self-supervised learning for medical image analysis using image context restoration, MedIA 2019
[2] Z. Zhou et al., Model Genesis, MedIA 2021
[3] Z. Chen et al., Masked Image Modeling Advances 3D Medical Image Analysis., IEEE WCACV 2023

实验结果：4）不同模型在CT图像中多器官分割的比较

下游任务：头颈部器官分割（MICCAI 2015 Head-Neck）

	Method	Brain stem	Optic chiasm	Mandible	Optic nerves	Parotid Glands	SM glands	Average
Dice (%)	nnU-Net [2]	89.27±2.42	57.47±24.24	90.12±5.84	72.50±7.85	**87.53±3.43**	75.06±12.58	78.66±5.42
	TransUNet [34]	75.52±5.58	41.92±15.99	92.28±1.45	58.36±6.91	76.70±6.85	69.80±8.27	69.10±3.08
	nnFormer [37]	80.02±3.53	52.72±14.70	87.96±2.27	57.34±7.72	75.31±7.24	68.25±5.53	70.27±4.19
	UNETR++ [33]	87.26±2.13	60.44±22.49	93.99±1.30	75.19± 5.85	84.61±3.89	80.74± 4.53	80.37±3.94
	PCT-Net	89.25±1.86	58.09±18.59	94.17±1.66	77.04±4.84	87.44±3.37	82.49±4.55	81.41±3.67
	PCT-Net + VF	**90.24±1.78**	**62.93±20.73**	**94.85±1.36**	**78.11±4.04**	87.07±3.69	**83.25±3.90**	**82.74±3.95**

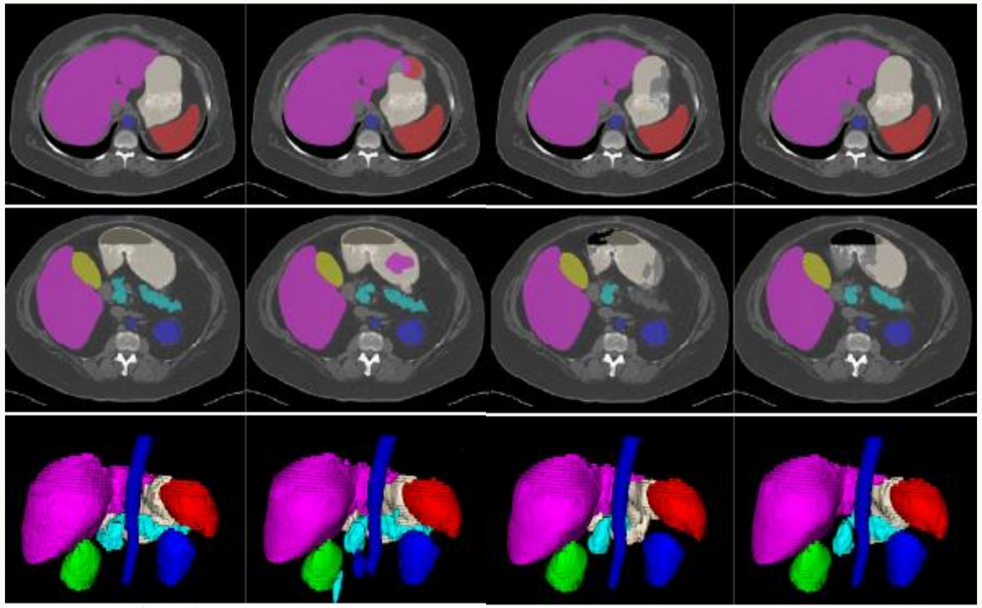


实验结果：4）不同模型在CT图像中多器官分割的比较

下游任务：胸、腹部器官分割

SegTHOR

Method	Dice (%)				
	Esophagus	Heart	Trachea	Aorta	Average
nnU-Net [2]	80.15±7.85	**93.05±2.59**	87.33±5.84	93.78±1.67	88.57±3.84
TransUNet [34]	75.66±9.36	85.27±16.18	**89.37±4.26**	91.52±2.97	85.46±7.20
nnFormer [37]	78.00±6.86	92.47±2.06	84.87±10.10	90.07±4.89	86.35±3.78
UNETR++ [33]	76.24±10.43	92.00±4.79	88.61±4.49	92.48± 2.32	87.33±3.67
PCT-Net	82.08±6.19	88.47±11.18	88.11±4.43	91.65±3.67	87.58±4.39
PCT-Net + VF	**83.45±4.78**	91.66±7.14	89.26±4.47	**93.88±1.79**	**89.56±2.81**



Ground Truth nnU-Net PCT-Net PCT-Net + VF

Synapse

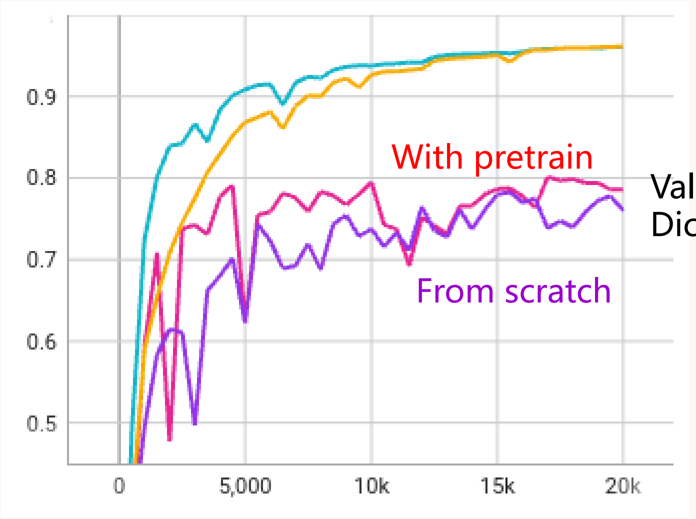
Method	Spleen	R Kidney	L Kidney	Gallbladder	Pancreas	Liver	Stomach	Aorta	Average
nnU-Net [2]	**94.00±4.26**	91.89±7.72	93.30±4.17	78.17±18.48	**83.27±3.98**	94.33±3.89	79.30±19.89	89.26±3.38	87.94±5.26
TransUNet [34]	92.00±7.15	92.48±4.11	92.30±4.77	74.21±12.31	72.18±16.12	94.73±4.11	75.72±15.54	90.67±4.16	85.54±5.40
nnFormer [37]	92.25±5.83	92.86±2.11	**93.84±1.45**	73.56±14.48	72.02±6.22	95.31±1.28	80.77±10.29	90.47±3.54	86.39±2.81
UNETR++ [33]	89.26±15.54	93.40±1.61	93.19±2.31	70.96± 28.24	74.70±12.14	95.76± 0.68	**82.79±15.22**	88.79±4..82	86.11±6.55
PCT-Net	91.36±13.77	95.21±5.46	90.78±9.79	**80.94±9.91**	79.13±9.86	96.63±7.04	79.25±23.33	90.48±5.11	87.97±5.22
PCT-Net + VF	91.38±12.97	**95.31±0.55**	92.17±8.00	80.79±13.58	83.24±3.97	**96.70±6.30**	82.46±15.99	**90.86±4.10**	**89.11±4.43**

实验结果：4）跨模态迁移能力——MRI图像分割

下游任务：MRI图像中器官的分割

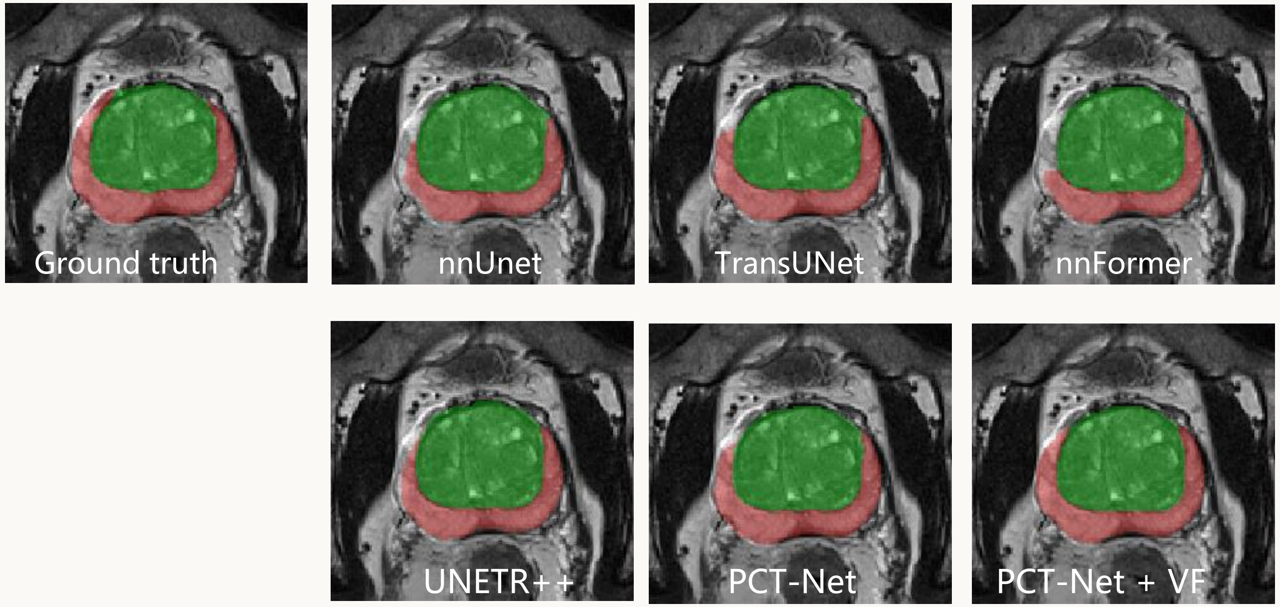
腹部器官（T2-MRI） 前列腺（T2-MRI） 左心室(LGE-MRI)

Method	CHAOS					NCI_ISBI13			
	Liver	R Kidney	L Kidney	Spleen	Average	PZ	CG	Average	Left Atrial
nnU-Net [2]	92.48±0.46	91.56±6.67	91.76±3.67	**91.62±4.70**	91.85±3.69	78.08±11.18	86.12±7.28	82.10±7.40	89.80±6.06
TransUNet [34]	92.74±1.81	93.96±1.59	90.36±6.08	89.33±3.23	91.60±1.62	77.32±11.61	86.10±6.35	81.71±7.47	89.63±2.62
nnFormer [37]	91.47±3.59	93.20±1.79	90.83±5.21	90.49±3.83	91.50±3.60	74.64±12.50	81.88±15.76	78.26±11.35	86.62±5.31
UNETR++ [33]	91.85±2.18	93.27±1.20	90.88±4.52	91.36±3.78	91.84±1.25	76.69±13.04	84.05±12.93	80.37±9.15	89.00±4.33
PCT-Net	94.55±1.19	93.36±1.10	92.91±1.97	88.95±7.71	92.44±1.89	75.91±11.73	**87.79±4.90**	81.85±6.43	89.94±4.52
PCT-Net + VF	**95.08±1.20**	**94.38±0.94**	**94.20±1.11**	90.39±7.01	**93.51±1.96**	**80.30±9.71**	86.64±7.87	**83.47±6.74**	**90.93±3.34**



Validation Dice

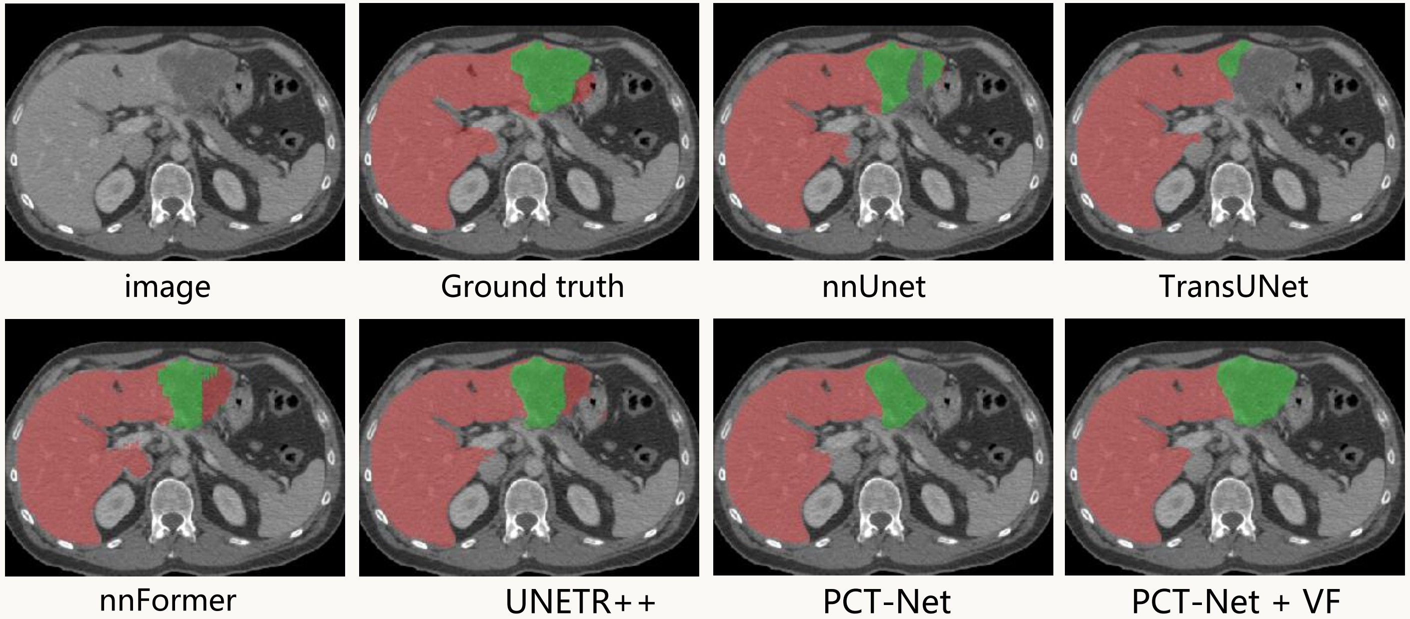
NCI_ISBI 13



实验结果：5）病灶分割

下游任务：肝脏肿瘤分割 LiTS数据集

Method	LiTS Liver	LiTS Tumor	LiTS Average
nnU-Net [2]	95.03±2.50	66.45±22.15	80.74±10.93
TransUNet [34]	94.12±3.29	65.17±20.50	79.65±10.27
nnFormer [37]	92.02±3.56	57.05±21.60	74.53±10.45
UNETR++ [33]	94.61±3.23	64.81±20.94	79.71±10.42
PCT-Net	94.00±2.70	71.42±19.62	82.71±9.72
PCT-Net + VF	**95.17±2.36**	**74.41±13.92**	**84.79±6.87**



总结与展望

针对三维医学图像分割任务：

数据：大规模未标注3D图像数据集（110k）

模型：PCT-Net 并行卷积与Transformer结构

训练：基于Volume Fusion的自监督预训练方法

- 分割任务作为预训练任务
- 有效训练模型对图像中上下文结构的理解

CT图像中器官的分割

MRI中器官的分割

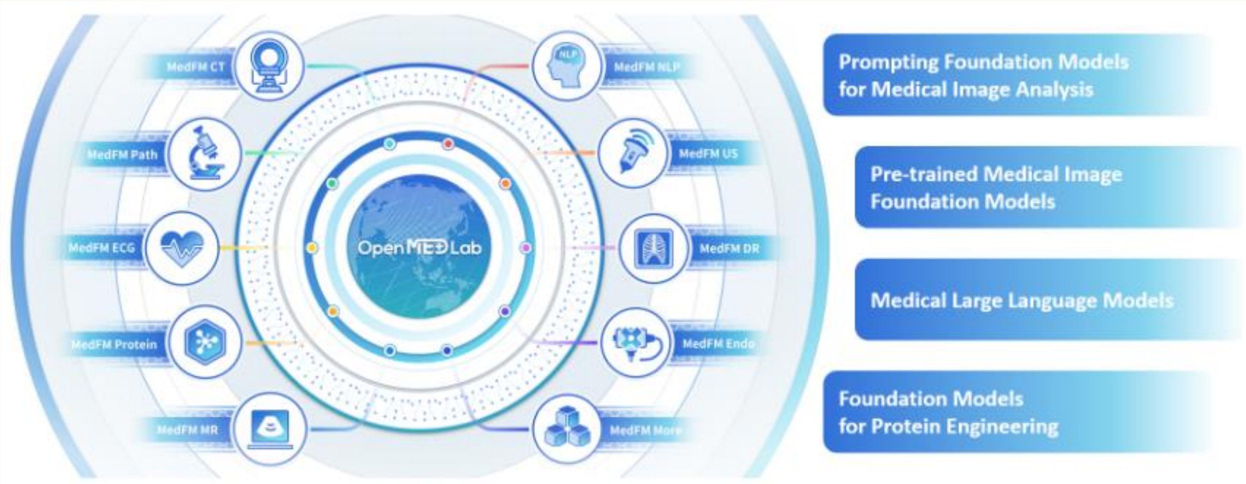
病灶的分割

预训练模型的作用：减少下游任务数据依赖、提高训练速度、泛化能力等

展望：

- 如何在下游任务中进一步减少数据和标注依赖，实现one-shot, few-shot分割?
- 更多模态、更大容量的预训练数据集(收集困难)

Thanks



GitHub: https://github.com/openmedlab/MIS-FM

联系方式：guotai.wang@uestc.edu.cn