杨晶东, 李皓秋, 姜泉,等. 基于协同训练的半监督学习 3D 医学图像分割模型[J]. 智能计算机与应用,2024,14(8):174-183.
DOI:10.20169/j.issn.2095-2163.240829

基于协同训练的半监督学习 3D 医学图像分割模型

杨晶东[1], 李皓秋[1], 姜 泉[2], 韩 曼[2], 宋梦歌[2]
(1 上海理工大学 光电信息与计算机工程学院自主机器人实验室, 上海 200093;
2 中国中医科学院广安门医院, 风湿病科, 北京 100053)

摘 要: 近年来人工智能应用于 COVID-19 医学影像诊断,降低了检测成本和漏检率,但临床医学图像样本数量较少和标签质量较低,影响了 3D 模型的分割性能。本文提出基于协同训练的半监督学习 3D 医学图像分割模型,使用空间翻转和窗口技术生成多视角、多模态图像,增强 3D 图像样本的空间差异性;采用一种基于加权不确定度的虚拟标签生成模块,为无标签数据生成可靠的虚拟标签,减少过拟合;采用基于三阶段的三维度六模型协同训练,增强分割精度和泛化性能。此外,本文可视化协同训练各阶段的特征关注度热力图,为临床诊断提供有效参考。针对 661 位新冠患者的 771 例 NIFTI 格式 3D COVID-19 的 CT 图像展开实验 ,5 折交叉验证结果表明,本文模型 Dice 系数为 73.30%,平均表面距离(ASD)为 10.633, 灵敏度(Sensitivity)为 0.630, 特异度(Specificity)为 0.996。与各种典型半监督学习 3D 分割模型相比,具有更好的分割精度和泛化性能。
关键词: 半监督学习; 协同训练; 3D 医学图像分割; 虚拟标签
中图分类号: TP301.6 **文献标志码**: A **文章编号**: 2095-2163(2024)08-0174-10

A semi-supervised learning 3D medical image segmentation model based on collaborative training

YANG Jingdong[1], LI Haoqiu[1], JIANG Quan[2], HAN Man[2], SONG Mengge[2]
(1 Autonomous Robot Laboratory, School of Optical-Electrial and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2 Rheumatology Department, Guang′anmen Hospital, Chinese Academy of Traditional Chinese Medicine, Beijing 100053, China)

Abstract: In recent years, the application of artificial intelligence in COVID-19 medical image diagnosis has reduced detection costs and missed detection rates. However, the small number of clinical medical image samples and low label quality have affected the segmentation performance of 3D models. This article proposes a semi supervised learning 3D medical image segmentation model based on collaborative training, which uses spatial flipping and window techniques to generate multi view and multimodal images, enhancing the spatial differences of 3D image samples; Adopting a virtual label generation module based on weighted uncertainty to generate reliable virtual labels for unlabeled data and reduce overfitting; Adopting a three-stage three-dimensional six model collaborative training to enhance segmentation accuracy and generalization performance. In addition, this article visualizes the feature attention heatmaps of each stage of collaborative training, providing effective references for clinical diagnosis. Experiments were carried out on 771 NIFTI format 3D COVID-19 CT images of 661 COVID-19 patients. The five fold cross validation results showed that the Dice coefficient of this model was 73.30%, the average surface distance (ASD) was 10.633, the sensitivity was 0.630, and the specificity was 0.996. Compared with various typical semi supervised learning 3D segmentation models, this model has better segmentation accuracy and generalization performance.
Key words: semi-supervised learning; collaborative training; 3D medical image segmentation; virtual labels

0 引 言

近年来,深度学习已经广泛应用于医学图像分割,具有较高的分割精度和诊断效率。Çiçek 等[1]提出一个可以从 2D 注释切片生成密集的体积分割的 3D U-Net 模型;Ma 等[2]使用两个放射影像数据集训

基金项目: 国家自然科学基金(81973749); 中国中医科学院科技创新工程重大攻关项目(CI2021A01503)。
作者简介: 李皓秋(1998-), 男,硕士,主要研究方向: 人工智能,机器学习在医学应用研究; 姜 泉(1961-) ,女,博士,教授,主任医师,主要研究方向: 风湿免疫病的中医、中西医结合临床及基础研究; 韩 曼(1984-) ,女,博士,副主任医师,主要研究方向: 风湿免疫病的中医、中西医结合临床及基础研究; 宋梦歌(1993-) ,女,博士,主要研究方向: 风湿免疫疾病的临床与基础研究。
通讯作者: 杨晶东(1973-),男,博士,副教授,主要研究方向:人工智能,机器学习与大数据分析,机器视觉等。Email:eerfriend@ yeah.net
收稿日期: 2023-05-07

练和评估不同的分割模型，实验结果表明 3D U-Net 模型具有最高分割精度，证明了 3D 分割模型的有效性；Milletari 等[3]在 3D U-Net 基础上增加残差链接，采用卷积代替池化，提出 V-Net 模型用于 3D 医学图像体素分割，并在 PROMISE12 数据集上验证有效性；Yu 等[4]使用全卷积残差网络（FCRN）在 Melanoma Detection Challenge 数据集上取得了较好的分割效果。为了减少对有标签数据的需求，降低标记成本，近年来人们提出了许多基于半监督学习的医学图像分割模型。Li 等[5]给输入数据增加扰动后对模型进行正则化，一次迭代模型前向传播两次，输入包括未变化图像和变化后图像，对变化后图像预测结果进行反变换，构建这两个预测结果的一致性损失，可以有效地利用未标注样本，提高模型分割性能；Li 等[6]在 Mean Teacher 模型的基础上使用旋转、翻转、尺寸变换和增加噪声等方法加入数据扰动，采用 Dropout 方法增加模型扰动，构建同一输入在不同扰动下的一致性，在 Dermoscopy image 和 Liver segmentation 数据集上取得了较好的分割效果；Yu 等[7]使用 Mean Teacher 结构和网络不确定度构建半监督学习框架，通过不确定度评估方式使 Mean Teacher 模型从未标注的数据上学习特征，但这种基于蒙特卡洛 Dropout 计算不确定度视图方法增加了许多额外的计算开销；Liu 等[8]使用多视角的联合训练构建半监督学习框架，但是并未引入一致性正则，训练过程容易出现过拟合；Luo 等[9]从多任务层面（task-level）构建基于一致性约束的半监督学习框架（DTC），使用多任务网络结构，同时进行分割和水平集回归两种任务，利用两个任务之间的差异性构建一致性正则化损失，DTC 不需要多次前向传播，减少了计算成本。本文针对小样本 3D COVID-19 CT 图像，提出一种多视角协同训练的半监督 3D COVID-19 分割模型（CTHS），采用 3D 图像翻转技术和 CT 图像窗口技术构建多视角、多模态样本集；提出一种基于加权不确定度的虚拟标签生成方法，为无标签数据生成虚拟标签，有效减少小样本过拟合问题，增加模型分类精度；采用 3 个阶段训练方式，为无标签数据生成虚拟标签数据，并采用多模型协同训练方式提高分类精度和泛化性能。

1　基于协同训练的半监督学习模型

1.1　多视角、多模态图像生成

多视角（Multi-view）指对同一研究对象的不同表示方式，如三维图像不同角度下的成像结果[10]。多模态（Multi-modality）指同一对象不同类型的特征，彼此之间具有一定的独立性，如 CT 图像不同的窗口。有些图像自身具有不同模态，例如核磁图像的 T1、T2、T1ce、Flair 序列，每个序列可以反应不同组织特征[11]。对于 2D 医学图像，可以采用旋转、平移、滤波、增加噪声等方式对图像进行数据扩充。在三维 CT 图像分割任务中，由于医学图像本身就具有很强噪声，卷积神经网络采用卷积和池化提取的特征具有一定的平移不变性，数据增强可以使网络具有一定的旋转不变性[12]。因此通过旋转或增加噪声的方式难以产生真正的具有差异性的多视角图像。考虑到医生在临床上观察 CT 图像时会设置不同的窗宽与窗位，例如肺窗，纵膈窗等[13]，本文采用三维翻转和窗口技术产生更有利于网络训练的多视角、多模态图像。

图像在三维空间中绕某一轴旋转（以 Z 轴为例）的坐标转换如式（1）所示，在三维空间中按某一平面（以 yoz 平面为例）进行翻转的坐标变换如式（2）所示。本文将原始图像按照 3 个方向分别翻转，提取每个方向多视角图像的两个不同模态即肺窗和横隔窗，分别得到 6 个不同的多视角、多模态图像。

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{1}$$

$$P'(x',y',z') = P(-x,y,z) \tag{2}$$

其中，$P(x,y,z)$ 为三维空间图像；$P'(x',y',z')$ 为变换后图像；β 为旋转角度。

本文将 3D CT 影像的多模态特征应用于半监督学习，并利用 3D 医学影像的多视角一致性特点，从多视角和多模态角度提取 3D 医学影像的深层特征，增强临床 3D 医学样本特征差异性。

1.2　基于加权不确定度的虚拟标签生成

网络不确定度可以表示为网络预测的置信度。本文使用 MC Dropout 法确定网络的不确定度，即在网络训练和预测时都启用 Dropout，同时在预测阶段对同一批次样本进行 T 次前向传播，然后将 T 次预测结果的方差作为对网络不确定度的度量。

在监督学习中，所有样本均有对应的标签，因此可以在模型中增加正则项，减少偶然不确定度，增加样本数量，减少认知不确定性。在半监督学习中，由于临床样本数量不足，仅使用无标签数据训练模型，会增加认知不确定性，为此本文提出一种基于加权不确定度的虚拟标签生成方法（UA-CT），减少训练模型认知不确定性，采用式（3）近似表示模型的认知不确定度。

$$U(f,x)=\frac{1}{T}\sum_{t=1}^{T}f(x;w_t)^2-\left(\frac{1}{T}\sum_{t=1}^{T}f(x;w_t)\right)^2 \quad (3)$$

其中，$f(x;w_t)$ 为网络输出；x 为网络输入；T 表示 MC Dropout 方法中预测时的前向传播次数；w_t 表示打开 Dropout 时第 t 次预测时的网络权重。

当输入样本为无标签数据时，打开网络的随机 Dropout，并对同一输入进行 T 次预测（本文取 $T=8$），由于随机 Dropout 的存在，T 次的预测输出并不完全一致，通过式（3）计算当前网络的不确定度，然后将 T 次输出取平均得到当前网络对无标签数据的预测值。将第 j 独立子网的不确定度转换为置信度（Confidence Score），并作为每个独立子网的权重，通过式（4）将所有子网的预测值和不确定度转换为当前图像的虚拟标签 $\hat{y}_i$，这种方式属于决策级融合机制中的加权融合策略。

$$\hat{y}_i=\frac{\sum_{j\neq i}^{N}c(U_j)\,P_j}{\sum_{j\neq i}^{N}c(U_j)} \quad (4)$$

其中，$p_j(x)$ 表示第 j 个独立子网输出，$c(U_j)=\mathrm{Sigmoid}(1/U_j)$。

1.3 CTHS 模型

由于无标签数据不需要专业医生人工标注，与有标签数据相比，无标签数据集数量更多。因此使用半监督学习技术如协同训练，可以有效利用未标记的数据，一定程度上解决临床医学数据不足或者标注不准确问题。传统协同训练方法多数采用两阶段训练方式，而本文创新性提出三阶段协同训练策略，第一阶段使用有标签数据单独训练每个子网络，在第二阶段加入无标签数据，针对多模态数据进行双网络联合训练，生成虚拟标签，并在第三阶段针对多视角、多模态图像数据进行六模型协同训练，逐步提取病灶区域的深层特征，增强模型泛化性能。本文提出的基于协同训练的半监督分割体系结构如图 1 所示，通过三维空间翻转和窗口技术将原始图像扩充为 6 个独立的多视角图像采用 MIG（Multiple Image Generativation）表示。此框架采用三阶段训练方法，每个阶段训练 100 个批次。第一阶段采用有标签数据集单独训练每个独立子网络，损失函数为 Dice & Cross-Entropy；第二阶段采用有标签数据集和无标签数据集并行训练双模态（肺窗和横隔窗）网络，如 DCNN1 和 DCNN1s，针对无标签样本，采用 UA-CT 模块生成虚拟标签并计算模型损失，损失函数为 Dice & Cross-Entropy + Consistency Regularization；第三阶段采用所有独立子网络并行训练所有标签数据，训练方式与第二阶段相同。在 3 个阶段训练过程中，使用 Sigmoid_Rampup 函数调节一致性正则损失权重 λ_{con}，如式（5）所示：

$$\lambda_{con}(e)=\begin{cases}0, & e<e_{ini}\\ \lambda_{max}\cdot\exp\left(-5\cdot\left(1-\dfrac{e-e_{ini}}{e_{end}-e_{ini}}\right)^2\right), & e_{ini}\leqslant e<e_{end}\\ \lambda_{max}, & e\geqslant e_{end}\end{cases} \quad (5)$$

其中，e 表示当前 epoch；e_{ini} 表示加入一致性正则损失的起始 epoch（本文设置为 100）；e_{end} 表示 λ_{con} 增长为 1 的分界点（本文设置为 250）。

本文基于协同训练的半监督学习模型训练伪码见表 1。

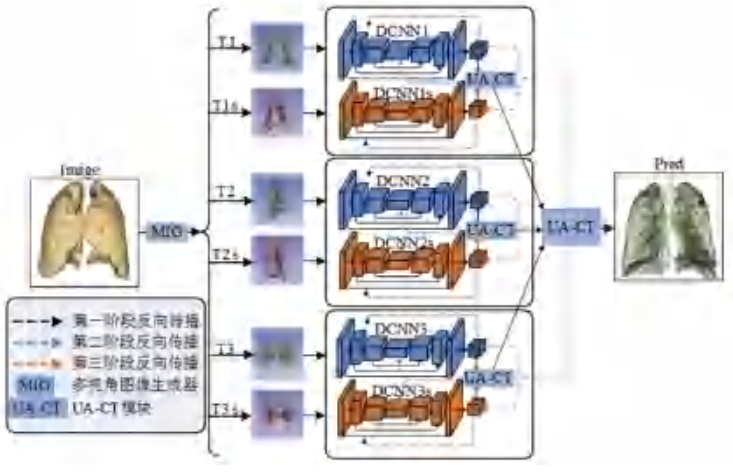


图 1 基于协同训练的半监督学习模型 CTHS 体系结构

Fig. 1 Architecture of CTHS model based on semi-supervised learning with collaborative training

表 1　基于协同训练的半监督学习模型 CTHS 训练伪码

Table 1　Pseudo-code of CTHS model based on semi-supervised learning with collaborative training

半监督学习流程

输入　有标签数据集 D_l 和无标签数据集 D_u

输出　网络预测 P

使用 3 个方向的三维翻转和两个不同的窗宽窗位组合得到 6 个图像转换器 $T_i, i \in \{1, \cdots, 6\}$

使用转换器将每个图像转换为 6 个独立的多视角图像 $D_i = T_i(D) D \in D_l \cup D_u$

for i in 所有的视角：

将图像输入独立的网络并获得输出结果 $P_i(D_i) = T_i^{-1}(f_i(D_i; w_i))$

计算无标签数据的亚标签 $\hat{y} = F(P_1(D_1), P_2(D_2), P_3(D_3), P_4(D_4), P_5(D_5), P_6(D_6))$

计算有标签数据的损失 $L_l = \frac{1}{N_l} \sum_1^{N_l} l(P_i(D_i), y_i)$

计算无标签数据的损失 $L_u = \frac{1}{N_u} \sum_1^{N_u} l(P_i(D_i), \hat{y}_i)$

计算一致性正则损失 L_{con}

总损失 $Loss = E_{(x,y) \in D_l} L_{\text{sup}}(x, y) + \lambda_{\text{sem}} E_{x \in D_u} L_{\text{sem}}(x) + \lambda_{\text{con}} E_{x \in D} L_{\text{con}}(x)$

进行反向传播，更新网络参数

2　实验与分析

2.1　实验环境与样本数据集

实验模型基于 Pytorch 1.12.0 平台构建，采用有 4 块 NVIDIA RTX3090Ti GPU 的服务器，操作系统为 Ubuntu 20.04，优化器为 SGD，初始学习率为 0.002。随机划分训练集和测试集比例为 8∶2，并进行 5-Fold 交叉验证。实验样本包含有标签数据集和无标签数据集，其中有标签数据样本采用 COVID-19-CT-Seg 和 COVID-19-20 数据集的部分样本，其中 20 例 COVID-19-CT-Seg 数据为 Corona-cases Initiative 样本，COVID19-20 数据来自 TCIA，无标签数据样本来自 TCIA，包含来自 661 个新冠患者的 771 例 NIFTI 格式的 CT 图像，部分病人具有不同时期的 CT 图像。为了保证样本的独立性，本文从此数据集中随机抽取来自 50 个患者的 50 例样本，即每个病人只选取一个 CT 样本，然后与有标签样本共同组成数据集。采用 MIG 方法生成的多视角的三维渲染图像如图 2 所示，绿色代表肺窗，红色代表纵膈窗，每个图像右下角的坐标轴代表所有的图像均按照特定的观察方向。

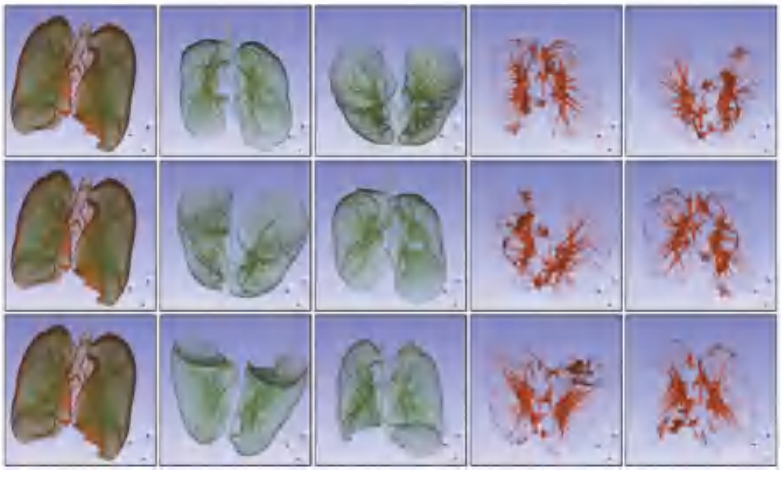
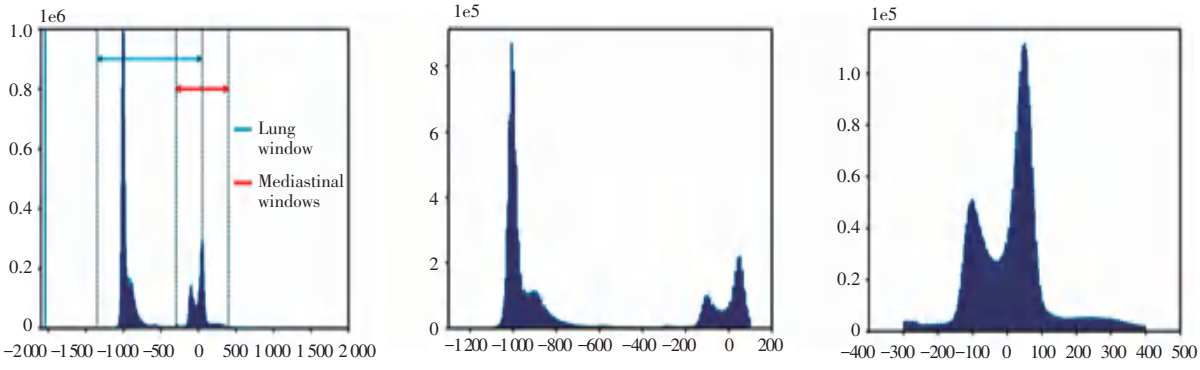


图 2　三维渲染图像

Fig. 2　3D rendering image

2.2 数据预处理

首先将所有图像的空间分辨率归一化为(1.0×1.0×1.0)。为了有效利用 3D 医学图像数据的多视角、多模态特征,并能生成可靠的虚拟标签,本文采用中心裁剪将训练集样本按照肺部区域中心裁剪并保留 16 体素边缘,将所有图像进行 Z-Score 标准化处理。为了生成多视角图像,本文将原始图像分为独立的肺窗和横隔窗,原始图像与两个窗口图像的 CT 强度直方图如图 3 所示。本文的 Batch Size 设置为 8,25%无标签数据占比表示一个 Batch 中包含 6 个有标签样本和 2 个无标签样本,50%无标签数据占比表示一个 Batch 中包含 4 个有标签样本和 4 个无标签样本。



(a) 原始图像 (b) 肺窗 (c) 纵膈窗

图 3 CT 图像强度分布直方图

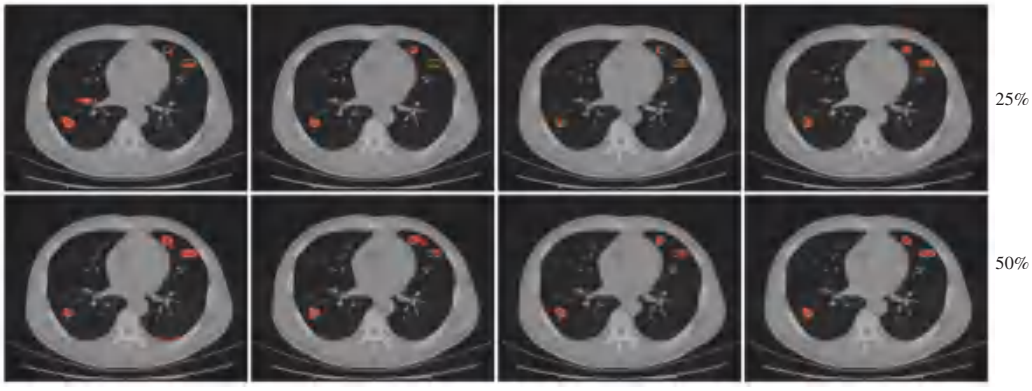
Fig. 3 Intensity distribution histogram of CT image

2.3 消融实验

本文采用准确率(Accuracy)、Dice 系数、平均表面距离(ASD)、灵敏度(Sensitivity)和特异度(Specificity)作为定量分析的评价指标。

1) 多视角、多模态数据生成方式对分割性能的影响

本文分别使用旋转和添加高斯噪声方式(R&N)以及空间翻转和窗口技术(F&W)方式生成多视角、多模态图像数据,分别采用无标签数据占比为 25%和 50%的两种训练集,针对两种多态数据生成方式,在 3D U-Net 和 V-Net 两个基础模型上的分割结果见表 2,对应的分割结果如图 4 所示。分析表 2 可知,基于 25%无标签数据占比情况下,使用 F&W 方式生成多模态数据,以 V-Net 为基础模型分割结果具有最高的 Dice 系数与最小的 ASD;与使用 R&N 方式生成的多模态数据进行训练相比,基于 F&W 方式生成的多模态数据进行训练时,3D U-Net 模型 Dice 系数增加了 6.13%,ASD 减少了9.65%,Sensitivity 增加了 6.29%,在使用 V-Net 模型时 Dice 系数增加了 13.99%,ASD 减少了 34.19%,Sensitivity 增加了 6.06%。说明 25%无标签数据占比时基于 F&W 数据生成方式 V-Net 模型具有最好分割性能。



(a) 3D U-Net/R&N (b) 3D U-Net/F&W (c) V-Net/ R&N (d) V-Net/ F&W

图 4 不同标签生成方式分割结果

Fig. 4 Comparison of performance for different label generation methods

表 2　不同标签生成方式评估参数

Table 2　Comparison of evaluation parameters for different label generation methods

模型	方法	无标签占比	Accuracy ↑	Dice ↑	ASD/mm ↓	Sensitivity ↑	Specificity ↑
3D U-Net	R&N	25%	0.995±0.002	0.636 9±0.10	20.551 3±10.94	0.572 2±0.06	0.996±0.002
		50%	0.998±0.002	0.645 4±0.12	21.835 6±15.81	0.573 0±0.09	0.995±0.002
	F&W	25%	0.998±0.002	0.674 8±0.08	18.568 1±10.08	0.608 2±0.11	**0.997±0.002**
		50%	**0.998±0.002**	0.660 1±0.13	18.363 1±11.02	0.595 2±0.05	0.991±0.002
V-Net	R&N	25%	0.988±0.010	0.642 8±0.13	16.141 6±10.41	0.593 7±0.07	0.996±0.002
		50%	0.996±0.002	0.687 2±0.10	14.193 1±5.39	0.620 1±0.10	0.998±0.001
	F&W	25%	**0.998±0.002**	**0.733 1±0.07**	**10.633 2±5.88**	**0.630 0±0.07**	0.996±0.016
		50%	0.996±0.002	**0.751 3±0.08**	**9.283 7±4.16**	**0.692 2±0.09**	**0.998±0.002**

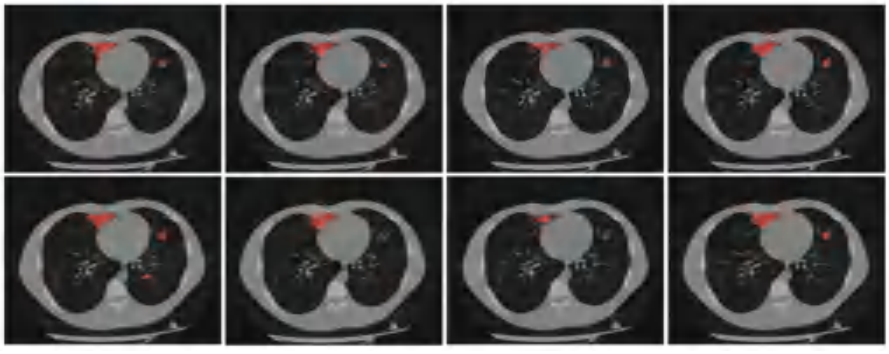
无标签数据占比 50%的情况下，使用 F&W 方式生成多模态数据训练集，以 V-Net 为基础模型的网络具有最高的 Dice 系数与最小的 ASD；与基于 R&N 方式相比，基于 F&W 方式生成的多模态数据作为训练集，3D U-Net 模型 Dice 系数增加了 2.33%，ASD 减少了 15.9%，Sensitivity 增加了 3.73%，V-Net 模型 Dice 系数增加了 9.32%，ASD 减少了 34.72%，Sensitivity 增加了 11.61%。这表明 F&W 生成方式比 R&N 方式具有更好的特征可区分性，这是因为协同训练依赖于多模态图像的一致性，每个单独模态应该包含部分独立的特征信息，而不同的子网络提取不同的特征。使用旋转方式进行数据扩充，并不能产生与原图有效的差异信息，而噪声的加入又会使子网络提取的特征不能真正的反应原始图像的特征信息。而采用窗口技术，即设置不同的窗宽和窗位，可以使图像反应不同尺度的组织特征，保证了多模态的图像彼此具有一定的独立性，而空间翻转进一步提升了多模态图像间的差异性，使多模态图像可以独立的包含不同的特征信息，因此 F&W 多模态数据生成方式更能反应原始图先后的空间体素特征。此外，与 3D U-Net 模型相比，V-Net 模型具有更好的分割性能，这是由于 Vnet 在降采样和上采样的每个阶段都加入了残差链接，相当于在 U-Net 的基础上加入了 ResBlock，同时 Vnet 在跳跃链接中加的卷积层的数量也随着网络深度逐渐增加，这种结构增加了网络的特征提取能力，降低了模型过拟合的风险。

与 25%无标签数据占比相比，当使用 50%无标签占比数据集进行训练并采用相同的 UAW 虚拟标签生成时，以 3D U-Net 作为基础模型时，R&N 数据生成方式 Dice 系数增加 1.42%，ASD 增加 6.25%，Sensitivity 增加 0.17%，F&W 数据生成方式 Dice 系数减少 2.27%，ASD 减少 1.12%，Sensitivity 减少 2.18%，以 V-Net 作为基础模型时，R&N 数据生成方式 Dice 系数增加 6.84%，ASD 减少 14.73%，Sensitivity 增加 4.38%，F&W 数据生成方式 Dice 系数增加 2.46%，ASD 减少 14.50%，Sensitivity 增加 9.84%。说明无标签数据的增加并不能在所有的基础模型上都有提升，这是因为本文采用空间翻转和窗口技术（F&W）方式进行多视角、多模态数据生成，该方法在对数据进行有效扩充的同时保证的不同视角不同模态图像间的独立性，对于 3D U-Net 模型来说，过多的无标签数据会使网络在训练过程中提取更多的冗余特征，不利于模型的分割性能。而 V-Net 在网络中加入残差链接，并且随着网络深度增加，卷积层数量也逐渐增加，因此拥有较强的特征提取能力，可以在更多的无标签数据中提取有用的差异化特征，从而增加模型的分割精度和泛化性能。

2）虚拟标签生成方式对分割性能的影响

本文采用将网络不确定度作为权重进行加权融合生成虚拟标签的方式（UAW），与将子网络预测输出平均值作为虚拟标签（AVG）。分别针对无标签数据占比 25%和 50%的两种训练集进行训练。基于 AVG 和 UAW 两种虚拟标签生成方式分割结果如图 5 所示；分别使用两种方式生成虚拟标签，在 3D-UNet 和 V-Net 两个基础模型上的分割结果见表 3。

根据表 3 可知，25%无标签数据占比的情况下，使用 UAW 方式生成虚拟标以 V-Net 为基础模型的网络具有最高的 Dice 系数与最小的 ASD。与 AVG 方式相比，基于 UAW 虚拟标签生成方式进行训练时，3D U-Net 模型 Dice 系数增加10.11%，ASD 减少 11.09%，Sensitivity 增加 0.50%；以 V-Net 作为基础模型时，Dice 系数增加 4.42%，ASD 减少 3.71%，Sensitivity 增加 0.32%。



(a) 3D U-Net/AVG (b) 3D U-Net/UAW (c) V-Net/AVG (d) V-Net/UAW

图 5 不同虚拟标签生成方式分割结果

Fig. 5 Comparison of segmentation for different virtual label generation

表 3 不同虚拟标签生成方式分割评估参数

Table 3 Comparison of evaluation indicators for different virtual label generation

模型	方法	无标签占比	Accuracy ↑	Dice ↑	ASD (mm) ↓	Sensitivity ↑	Specificity ↑
3D U-Net	AVG	25%	0.998±0.002	0.612 6±0.102	20.883 2±10.783	0.604 8±0.065	0.996±0.001
		50%	0.996±0.002	0.669 8±0.110	19.055 6±9.008	0.609 6±0.067	0.998±0.001
	UAW	25%	0.997±0.001	0.674 8±0.076	18.568 1±10.079	0.608 2±0.105	0.997±0.001
		50%	0.998±0.002	0.660 1±0.123	18.363 1±10.507	0.595 1±0.051	0.991±0.002
V-Net	AVG	25%	0.995±0.002	0.702 1±0.095	16.891 8±9.954	0.628 0±0.093	0.998±0.002
		50%	0.998±0.002	0.727 1±0.090	11.754 3±6.216	0.641 7±0.049	0.995±0.002
	UAW	25%	0.998±0.002	0.733 1±0.070	10.633 2±5.885	0.630 0±0.067	0.996±0.002
		50%	0.996±0.002	0.751 3±0.079	9.283 7±4.160	0.692 3±0.098	0.998±0.002

无标签数据占比 50%的情况下,基于 UAW 方式生成多模态数据 V-Net 模型具有最高的 Dice 系数与最小的 ASD。与基于取平均值方式生成的虚拟标签训练集相比,基于 UAW 多模态数据生成方式的 3D U-Net 模型 Dice 系数减少了 1.34%,ASD 减少了 3.38%,Sensitivity 减少了 2.46%;以 V-Net 为基础模型时,Dice 系数增加了 9.32%,ASD 减少了 21.01%,Sensitivity 增加了 7.79%,表明使用 UAW 方法生成的虚拟标签比 AVG 方式生成的虚拟标签具有更高的稳定性和置信度,能更好的使网络进行参数更新,提高网络的泛化性能。

无论无标签数据占比多少,在训练阶段的每个批次都需要为无标签数据赋予一个虚拟标签,而虚拟标签的质量直接影响损失函数 Loss 的计算,进而影响网络参数的更新。深度神经网络存在一定的不确定性,尤其对无标签数据,不同子网络的预测输出具有不同的置信度,将不同的子网络的输出取平均后作为虚拟标签的方式忽略了网络预测的置信度。由 MC-Dropout 可知,在预测时打开 Dropout 时,针对相同的输入,网络每次的预测结果也会有所不同,这种对同一预测的不同输出之间的变化间接反映了网络不确定度,使用 sigmoid 函数将此不确定度转换为置信度后作为每个子网络的权重,并将每个子网络的预测结果进行加权平均,可以利用这种不确定度使预测趋于稳定。即置信度高的网络拥有更高的权重,可以使虚拟标签更接近真实的标签,并具有更高的可靠性。

与训练集中无标签数据占比 25%相比,50%无标签数据占比数据集进行训练时,均使用 F&W 数据生成方式时,以 3D U-Net 作为基础模型,AVG 伪标签 Dice 系数减少 9.30%,ASD 减少了 9.59%,Sensitivity 增加 0.83%,UAW 伪标签 Dice 系数减少 2.12%,ASD 减少了 0.85%,Sensitivity 减少 2.18%;以 V-Net 作为基础模型时,AVG 伪标签 Dice 系数增加 3.56%,ASD 减少 43.71%,Sensitivity 增加 2.22%,UAW 伪标签 Dice 系数增加 2.46%,ASD 减少 14.53%,Sensitivity 增加 9.84%。这说明对于所有基础模型增加无标签数据并非均能提升精度,因为在基于协同训练的半监督学习中,并不是所有的数据都有标签,所以网络的特征提取能力对网络最后的分割结果影响较大,无标签数据的增加意味着虚拟标签的增加,网络更加容易提取到数据中冗余信

息，反而会影响分割精度。对于特征提取能力比较强的网络，如 V-Net 则可以在更多的无标签数据中提取有用的差异化特征，从而增加模型的分割精度和泛化性能。

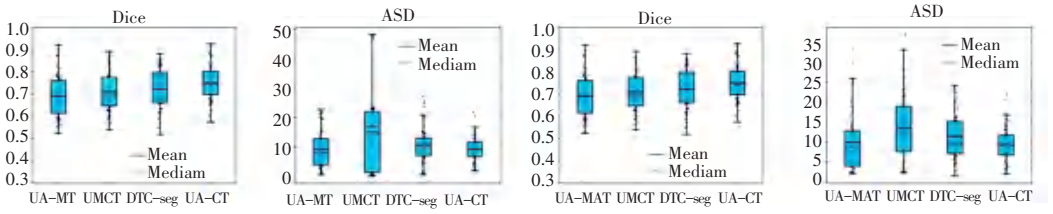
2.4　与典型模型对比实验

为了验证本文基于协同训练的半监督学习策略的有效性，将本文提出 UA-CT 模型与其他模型进行对比。本文采用 F&W 方法生成多视角、多模态图像，并使用 UAW 方法生成虚拟标签具有最佳预测性能，并与 UA-MT，UMCT 和 DCT-Seg[14] 3 种典型模型的精度对比见表 4，各模型训练集的分割性能对比如图 6 所示。

表 4　不同基础模型的分割性能

Table 4　Comparison of performance indicators for various baseline models

模型	无标签占比	Accuracy ↑	Dice ↑	ASD（mm）↓	Sensitivity ↑	Specificity ↑
UA-MT	25%	0.998±0.002	0.688 2±0.084	**9.878 5±7.334**	0.631 9±0.058	0.997±0.002
	50%	0.998±0.002	0.692 7±0.092	**9.085 8±6.093**	0.637 2±0.108	0.998±0.002
UMCT	25%	0.998±0.002	0.708 5±0.088	14.810 0±11.968	0.617 6±0.107	0.997±0.002
	50%	0.996±0.001	0.711 2±0.082	14.976 0±14.256	0.638 0±0.056	0.997±0.001
DCT-Seg	25%	**0.999±0.002**	0.719 6±0.097	11.203 4±5.897	0.712 6±0.049	0.995±0.002
	50%	0.998±0.002	0.722 6±0.090	10.633 2±5.885	0.701 1±0.110	0.996±0.002
CTHS	25%	0.998±0.002	**0.733 1±0.070**	10.633 2±5.885	0.630 0±0.067	0.996±0.002
	50%	**0.996±0.002**	**0.751 3±0.079**	9.283 7±4.160	**0.692 3±0.099**	**0.998±0.002**



（a）25%　（b）50%　（c）25%　（d）50%

图 6　各模型训练集的分割性能对比

Fig. 6　Comparison of segmentation indicators on training datasets for various typical models

根据表 4 可知，UA-MT 模型具有最小的 ASD 系数，本文模型具有最高 Dice 系数与 Sensitivity。与 UA-MT 模型相比，本文模型 Dice 系数增加 2.60%，ASD 增加 2.18%，Sensitivity 增加 9.15%；与 UMCT 模型相比，本文模型 Dice 系数增加 5.63%，ASD 减少 2.18%，Sensitivity 增加 8.46%；与 DCT-Seg 模型相比，本文方法 Dice 系数增加 3.87%，ASD 减少 12.69%，Sensitivity 增加 9.15%。

50%无标签数据占比的情况下，UA-MT 模型具有最小的 ASD 系数，本文方法具有最高的 Dice 系数与 Sensitivity。与 UA-MT 模型相比本文方法的 Dice 系数增加了 2.60%，ASD 增加了 2.18%，Sensitivity 增加了 9.15%；与 UMCT 模型相比本文方法 Dice 系数增加了 5.63%，ASD 减少了 2.18%，Sensitivity 增加了 8.46%；与 DCT-Seg 模型相比本文方法 Dice 系数增加了 3.87%，ASD 减少了 12.69%，Sensitivity 增加了 9.15%。分析可知本文方法具有最好的分割性能，这是因为 UA-MT 模型使用 Mean Teacher 半监督学习策略，为了网络间差异，图像分别输入 Student Model 和 Teacher Model 时加入了一定的噪声，这种噪声的加入有利于 Teacher Model 和 Student Model 产生不同的预测，从而通过一致性约束使网络从无标签数据中提取信息，但也降低了网络的特征提取能力，使网络更容易去拟合医学图像的噪声，而不是原始的语义特征。本文提出的 CTHS 分割模型使用空间翻转和 CT 图像的窗口技术生成多视角、多模态图像，在不加入额外噪声的前提下保证了不同视角的差异性，而一致性正则的加入约束网络对同一个样本的不同视角产生相似的预测，可以根据网络的不确定度生成更加可靠的虚拟标签，综合了协同训练和一致性正则的优势。

50%无标签样本的 5 折交叉验证 Dice 系数、Loss 曲线如图 7 所示。50%无标签数据占比的数据集进行训练时在第 150 个批次由于新的数据加入训练，Dice 系数有所下降。在第 250 个批次由于将所有的子网络并行训练，Dice 系数同样有所下降，但

由于第三阶段的虚拟标签融合了 6 个子网络的输出,使训练过程更加稳定,所以 Dice 系数在下降后立即增加并很快趋于稳定。在第 200 个批次附近,Dice 系数也有一定程度下降后快速增加。这是因为本文在训练过程中采用带有动量的 SGD 优化器和固定步长学习率衰减,每隔 50 个批次学习率更新为原来的 0.5 倍,而动量机制有利于网络跳出局部最优解,快速达到全局最优解,在第 200 批次时网络陷入局部最优解,随着学习率的更新,网络迅速跳出局部最优解,按照新的梯度继续更新网络参数。但是这种现象在使用 25% 无标签数据占比的数据集进行训练时并未发生,说明更多地无标签数据可以有效避免网络陷入局部最优,快速达到全局最优解,增加网络训练速度。

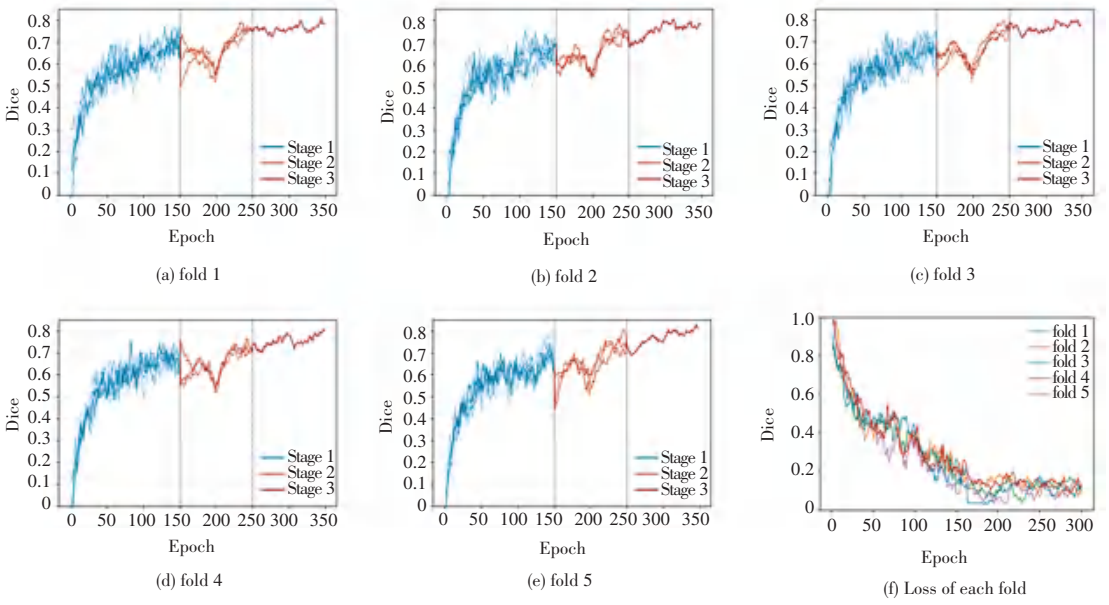


图 7 50%无标签样本的 5 折交叉验证 Dice 系数、Loss 曲线

Fig. 7 Dice and Loss on 5-Fold Cross-validation for 50% No-label generation methods

2.5 特征关注度可视化

为了进一步验证模型的有效性,本文将 3 个训练阶段的网络对不同区域的关注度进行可视化输出。3 个样本每个阶段对样本的感兴趣区域如图 8 所示。从图 8 可以看出,在第一个训练阶段,由于只使用有标签样本,网络注意力集中范围较大,并在部分集中在无病灶区域;网络训练的第二个阶段,由于无标签样本的加入,网络的注意力集中在病灶区域附近,而且范围更加集中,但仍有部分集中在无病灶区域;网络训练的第三阶段,由于并行训练所有网络,经过第二阶段训练虚拟标签已经具有较高的可信度,更多的网络关注度集中在病灶区域附近,说明本文提出的基于协同训练的半监督学习模型可以有效利用无标签数据,提升网络的分割性能。

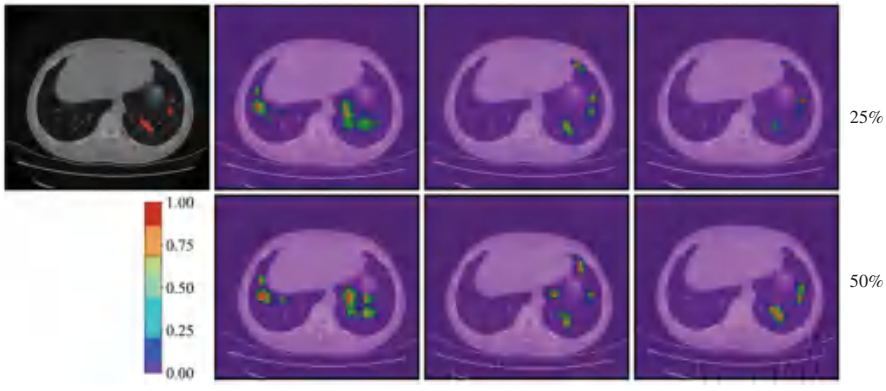


图 8 采用三阶段训练策略每个阶段网络的感兴趣区域

Fig. 8 ROI attention of each stage for 3 stages

为了更好检验本文提出的半监督学习框架以及三阶段训练策略的有效性，本文对网络的特征注意力区域进行进一步量化，将 Grad-cam 中热力值前75%为阈值进行二值化处理，计算每个连通域的中心点与距离最近的标签区域的中心点距离，最后得到网络对当前样本的注意力区域与标签区域的平均距离，使用这种距离近似反应网络注意力区域与标签区域偏差，如图 9 所示。

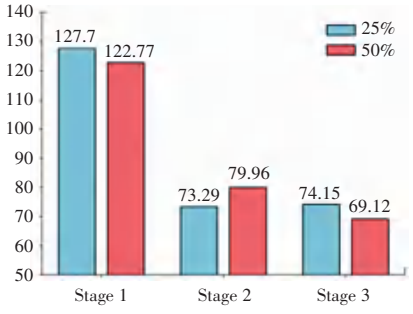


图 9　各阶段网络对当前样本注意力区域与标签区域平均距离

Fig. 9　Mean Distance of attention and labeled region for each stage of 3 stages

当 25%无标签数据占比时，与第一阶段相比，网络训练第二阶段注意力区域像素中心与标签中心距离减少 42.61%，第三阶段相比第二阶段增加1.18%；当 50%无标签数据占比时，与第一阶段相比，第二阶段注意力区域与标签的距离减少34.87%，第三阶段相比第二阶段减少了 13.56%；由此可见，随着伪标签加入，在协同训练的 3 个阶段中，分割精度逐渐增加，这是由于在第二阶段使用双模态网络进行协同训练，通过利用无标签数据，增加了网络对病灶区域的定位能力，网络注意力更集中于病灶区域附近。而训练第三阶段，随着虚拟标签数据逐渐增加，训练网络对病灶区域的重要特征提取能力逐渐增强。与 25%无标签数据占比相比，当 50%无标签数据占比时，在第一阶段时，网络注意力区域像素质心与标签质心距离减少 3.86%，在第二阶段时，增加 9.10%，在第三阶段减少 6.78%，表明当更多虚拟标签数据加入时，可以有效提升训练网络对病灶区域的定位能力。

3　结束语

本文构建了基于协同训练的半监督学习 3D 医学图像分割模型，使用空间翻转和窗口技术生成多视角、多模态图像，增强 3D 图像样本的空间差异性；采用一种基于加权不确定度的虚拟标签生成模块，为无标签数据生成可靠的虚拟标签，减少过拟合；采用基于三阶段的三维度六模型协同训练，增强 3D 医学样本分割精度。今后重点研究改进训练策略和 3D 分割模型结构，进行更多外部验证，增强模型泛化性能。

参考文献

[1] ÇIÇEK Ö, ABDULKADIR A, LIENKAMPS S, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation [C]//Proceedings of Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016: 19th International Conference, Athens, Greece: IEEE, 2016: 424-432.

[2] MA J, WANG Y, AN X, et al. Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation [J]. Medical Physics, 2021, 48(3): 1197-1210.

[3] MILLETARI F, NAVAB N, AHMADI S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation [C]//Proceedings of 2016 Fourth International Conference on 3D Vision (3DV). IEEE, 2016: 565-571.

[4] YU L, CHEN H, DOU Q, et al. Automated melanoma recognition in dermoscopy images via very deep residual networks [J]. IEEE Transactions on Medical Imaging, 2016, 36(4): 994-1004.

[5] LI X, YU L, CHEN H, et al. Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model [J]. arXiv preprint arXiv: 1808.03887, 2018.

[6] LI X, YU L, CHEN H, et al. Transformation-consistent self-ensembling model for semisupervised medical image segmentation [J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(2): 523-534.

[7] YU L, WANG S, LI X, et al. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation [C]// Proceedings of Medical Image Computing and Computer Assisted Intervention - MICCAI 2019: 22nd International Conference. Shenzhen, China: IEEE, 2019: 605-613.

[8] LIU F, YANG D. 3D semi-supervised learning with uncertainty-aware multi-view co-training [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE, 2020: 3646-3655.

[9] LUO X, CHEN J, SONG T, et al. Semi-supervised medical image segmentation through dual-task consistency [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021: 8801-8809.

[10] XU C, TAO D, XU C. A survey on multi-view learning [J]. arXiv preprint arXiv: 1304.5634, 2013.

[11] 苏庆华，张一晨，杨翼臣，等. 基于 ResNet50 的脑胶质瘤甲基转移酶生物标志检测[J]. Advances in Clinical Medicine, 2022, 12: 8756.

[12] KAYHAN O S, GEMERT J C. On translation invariance in CNNS: Convolutional layers can exploit absolute spatial location [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 14274-14285.

[13] LU H, MU W, BALAGURUNATHAN Y, et al. Multi-window CT based radiomic signatures in differentiating indolent versus aggressive lung cancers in the National Lung Screening Trial: a retrospective study [J]. Cancer Imaging, 2019, 19: 1-11.

[14] PENG J, ESTRADA G, PEDERSOLI M, et al. Deep co-training for semi-supervised image segmentation [J]. Pattern Recognition, 2020, 107: 107269.