

# 基于半监督学习和条件概率的膝关节软骨 MRI 图像分割<sup>①</sup>



马春帅, 程远志

(青岛科技大学 信息科学技术学院, 青岛 266061)

通信作者: 程远志, E-mail: [yzchengqust2007@163.com](mailto:yzchengqust2007@163.com)

**摘要:** 本文提出了一种基于半监督学习和条件概率的膝关节软骨分割方法, 旨在解决医学图像分割中标注样本数量不足和质量差的问题. 现有的标签树嵌入深度学习模型难以对网络输出间的层次关系进行有效建模, 而本文提出了一种条件到无条件的混合训练与任务级一致性结合的方法, 有效地利用了标签之间的层次关系和相似性, 提高了分割精度. 具体来说, 本文使用一个联合预测像素级分割图和目标的几何感知水平集表示的双任务深度网络. 通过可微分的任务变换层, 将水平集表示转换为近似的分割映射. 同时, 本文在标签和未标记数据上引入了基于水平线的分割映射与直接预测的分割映射之间的任务级一致性正则化. 在两个公共数据集上的大量实验表明, 本文的方法可以通过包含未标记的数据来显著提高性能.

**关键词:** 膝关节软骨图像分割; 多标签分类; 半监督学习; 标签相关性

引用格式: 马春帅,程远志.基于半监督学习和条件概率的膝关节软骨 MRI 图像分割.计算机系统应用,2025,34(1):100-109. <http://www.c-s-a.org.cn/1003-3254/9759.html>

## MRI Image Segmentation of Knee Cartilage Based on Semi-supervised Learning and Conditional Probability

MA Chun-Shuai, CHENG Yuan-Zhi

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China)

**Abstract:** This study introduces a knee cartilage segmentation method based on semi-supervised learning and conditional probability, to address the scarcity and quality issues of annotated samples in medical image segmentation. As it is difficult for existing embedded deep learning models to effectively model the hierarchical relationships among network outputs, the study proposes an approach combining conditional-to-unconditional mixed training and task-level consistency. In this way, the hierarchical relationships and relevance among labels are efficiently utilized, and the segmentation accuracy is enhanced. Specifically, the study employs a dual-task deep network predicting both pixel-level segmentation images and geometric perception level set representations of the target. The level set is shifted into an approximate segmentation map through a differentiable task transformation layer. Meanwhile, the study also introduces task-level consistency regularization between level line-based and directly predicted segmentation maps on labeled and unlabeled data. Extensive experiments on two public datasets demonstrate that this approach can significantly improve performance through the incorporation of unlabeled data.

**Key words:** image segmentation of knee cartilage; multi-label classification; semi-supervised learning; label relevance

① 基金项目: 国家自然科学基金 (61806107, 61702135)

收稿时间: 2024-06-08; 修改时间: 2024-07-03, 2024-08-20; 采用时间: 2024-08-27; csa 在线出版时间: 2024-11-15

CNKI 网络首发时间: 2024-11-18

膝关节关节炎是最常见的骨骼肌肉疾病,也是中老年人致残的主要原因,85%全膝关节置换是由于膝关节关节炎。因此,基于磁共振成像(MRI)的膝关节软骨分割对于早期诊断,患处定位和针对性治疗具有重要意义,因为它可以提供高精度、可重复和定量的评估。目前膝关节分割存在的主要问题还是标注样本过少,医学影像的标注需要高度专业的医学知识,不仅成本高昂,而且极其费时费力。

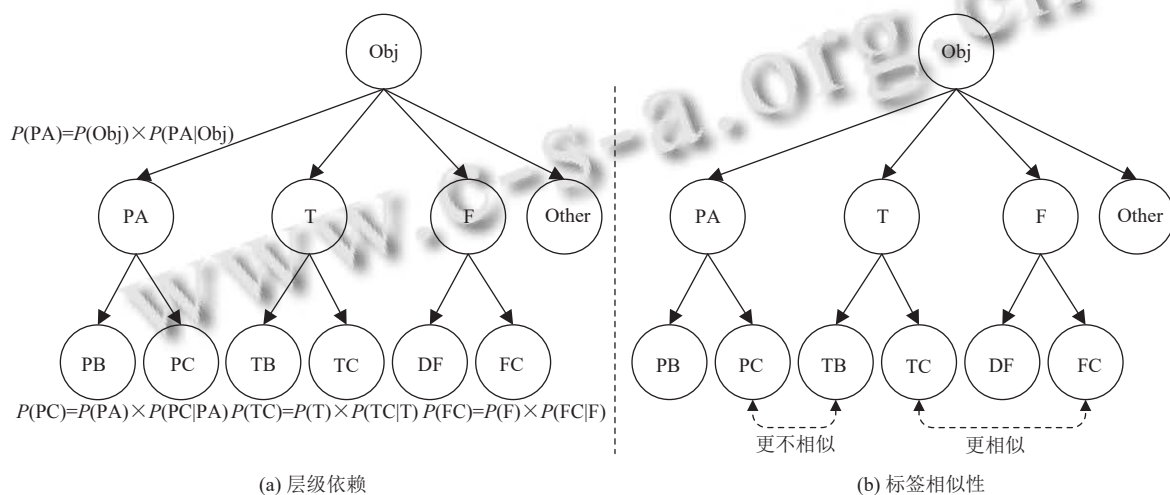
目前虽然已经出现了一些公开的骨关节炎(osteoarthritis, OA)以及肌肉骨骼影像数据库,但数据库的容量较为有限,如SKI10数据库仅收录了150个膝关节MR图像。第2届医学影像AI大会上的演讲中<sup>[1]</sup>提到,我国数据库建设仍然缺乏,目前可公开的数据库不多,数据的标注标准也不统一。同时,公共数据库样本绝大多数来自欧美人群,目前尚未出现以中国或亚洲人群为主的大型OA或肌肉骨骼影像数据库。

针对上述问题,我们希望通过将标签树嵌入半监督网络中,以实现在使用少量标注数据的情况下实现膝关节软骨精确分割。在现有的研究中,标签树嵌入深度学习模型通常面临一些挑战,特别是在处理大规模多标签问题时。现有模型的普遍问题在于,无法为网络模型输出间的层次关系进行合适的建模。在多标签分类中,标签之间往往存在一定的相关性。然而,传统的标签树嵌入模型可能无法充分捕捉这些相关性,因为它们通常将每个标签视为独立的实体。在大规模的标

签空间中,很多标签可能只有很少的正样本与之相关联,这导致学习标签之间的联系变得困难。尽管标签树提供了一种组织标签的层次结构,但现有的模型可能无法为网络模型输出间的层次关系进行合适的建模,特别是在深度学习模型中,如何将这种层次结构有效地整合到模型中仍然是一个挑战。

如图1(a),建立分割法后,必须建立分级分割方法。因为这是一个多标签设置,所以图1中的所有标签都可以为正,也可以不为正。唯一的限制是,如果一个子类是积极的,那么他的父类也必须是积极的。而兄弟标签之间是互斥的。另外,标签之间固有的相似性也是提升预测精度的关键,如图1(b),胫骨软骨和股骨软骨的语义相似度远高于它们与其他标签的相似度,因为胫骨软骨和股骨软骨都属于软骨组织且在实际情况中距离较近。其中,PA表示髌骨和髌骨软骨,T表示胫骨和胫骨软骨,F表示股骨远端和股骨软骨,PB表示髌骨,TB表示胫骨,DF表示股骨远端,PC表示髌骨软骨,TC表示胫骨软骨,FC表示股骨软骨, $P(\text{PA})$ 表示当前区域为髌骨和髌骨软骨的预测概率。

综上所述,膝关节软骨的精准分割对于膝关节关节炎的早期诊断和治疗具有重要意义。尽管当前研究面临标注样本稀缺、多标签问题处理困难等挑战,但通过创新的分级分割方法和标签树嵌入技术,有望在少量标注数据的情况下实现更精确的分割。



PA: 髌骨和髌骨软骨, T: 胫骨和胫骨软骨, F: 股骨远端和股骨软骨, PB: 髌骨, TB: 胫骨, DF: 股骨远端, PC: 髌骨软骨, TC: 胫骨软骨, FC: 股骨软骨

图1 标签树结构

## 1 相关工作

### 1.1 半监督医学图像分割

对于医疗影像处理,最大的问题通常是标注样本

的数量过少和标注的质量差,半监督方法的目的是从标记和未标记数据提取知识,以获得一个比只使用标记数据训练得到的模型性能更好的模型。

在半监督医学图像分割领域,常见方法包括基于先验的模型<sup>[2]</sup>和基于聚类的模型<sup>[3]</sup>,但是,基于先验的模型需要为不同的器官设计特定的先验信息,这很难推广到其他器官.基于聚类的模型通常对参数敏感且不够鲁棒,这导致对形状变化较大的对象的预测较差.

深度学习由于其卓越的性能已被广泛地运用于医学图像分割<sup>[4]</sup>和半监督医学图像分割<sup>[5]</sup>,近年来发表的半监督学习医学图像分割框架几乎都是基于深度学习的. Li 等<sup>[6]</sup>在 2021 提出的 TCSM-V2 模型,在 Mean Teacher<sup>[7]</sup>的基础上引入了更多的数据扰动和模型扰动,用以构建同一输入在不同扰动下的一致性,取得了很不错的效果. Yu 等<sup>[8]</sup>提出的 UA-MT 模型整合了 Mean Teacher 和不确定性估计来进行半监督学习,通过不确定性图来指导 Mean Teacher 模型从未标注的数据上逐步学习,这样会带来一些额外的计算开销,但带来了性能的提升. Xia 等<sup>[9]</sup>将多视角的联合训练引入到医学图像的半监督和域适应的问题中,实现了比较不错的效果提升,只是随着视角的不断增多计算开销会越来越大.不同之处在于,我们的方法利用了条件概率和任务级一致性,这对于半监督医学图像分割来说既简单又有效.

## 1.2 深度度量学习

在数学中,度量函数是用来衡量集合中元素之间距离的函数.度量空间是指一个具有度量的集合.度量学习 (metric learning) 是一种用来衡量样本之间相似性,并利用最优距离度量进行学习任务的方法. Eric 等<sup>[10]</sup>在 2002 年的 NIPS 会议上提出了度量学习的概念,可以分为两种:通过线性变换的度量学习和通过非线性变化的度量.传统的度量学习方法通过投影矩阵将原始数据的特征空间转化为具有距离信息的新的变换空间.目前的研究主要涉及马氏距离<sup>[11,12]</sup>,其目的是通过协方差矩阵分解和对称正定矩阵将马氏距离转化为欧几里得距离.

2016 年,深度度量学习<sup>[13]</sup>被提出,将深度学习和度量学习结合在一起.其核心目标是最小化相似样本之间的某种距离度量,并最大化不相似样本之间的某种距离度量.

## 1.3 细粒度视觉分类

细粒度视觉分类是一种特定类型的视觉分类任务,其目标是将物体或图像分为相似的子类别.这种任务要求模型能够识别并区分具有相似外观或特征的细小

类别.

与传统的物体分类任务相比,细粒度视觉分类更关注于类别内的差异.这种任务对于模型的细节感知和特征提取能力提出了更高的要求,通常需要结合使用局部特征、关键点定位等技术来实现精细分类.

Du 等<sup>[14]</sup>通过逐步训练策略和多尺度信息融合来解决细粒度视觉分类中的挑战性问题,使用一个简单的拼图生成器来形成包含不同层次信息的图像. Dubey 等<sup>[15]</sup>提出了 pairwise confusion (PC) 方法,通过引入一个额外的距离度量来增加模型的泛化能力,用于解决在小样本数据集上训练深度神经网络时出现的过拟合问题. Hu 等<sup>[16]</sup>提出了弱监督局部注意力网络,解决了细粒度视觉分类和定位问题,即在给定一张图像的情况下,识别出其中属于某个类别的对象并确定其位置. Zhuang 等<sup>[17]</sup>通过比较图像对之间的差异来识别对比线索,并使用交互方式逐步识别一对细粒度图像.但是,他们通常忽略了对其他标签关系(如父子关系)进行编码,从而使得不同级别下的样本在级别之间传递层次知识变得比较困难.

## 2 提出的方法

### 2.1 网络结构

图 2 展示了我们提出的半监督层次条件概率分割 (semi-supervised hierarchical conditional probability segmentation, SSHCPS) 框架.如图 2 所示,将提取出的特征分为两支.一方面,这些特征被映射到条件输出,通过我们设计的条件到无条件的训练方案进行训练,构成我们网络的全监督学习部分,即 HCPS (hierarchical conditional probability segmentation).另一方面,这些特征被投影到水平集特征空间,辅助进行半监督学习.在网络主干方面,我们选用精确度和特征提取速度都位于前列的 MedT 框架.

从网络主干提取的特征有两个主要应用.一方面,这些嵌入特征被映射到条件输出,通过我们设计的条件到无条件的训练方案进行训练.另一方面,这些嵌入特征被投影到新的特征空间,以进行半监督学习.

在网络训练中,最常见的方法是 minimize 无条件概率,但这会带来一些优化上的困难.首先,乘积可能使梯度值过小,导致梯度消失,使模型难以从局部最小值中逃脱.其次,乘积运算易引起算术下溢.为了有效训练网络,我们提出了一种混合训练方法,在第 1 阶段,

使用条件概率进行网络训练和初始化,第2阶段,我们引入了数值稳定的无条件概率损失函数,采用无条件

概率进行微调.在另一分支,我们使用水平集回归辅助进行半监督学习.

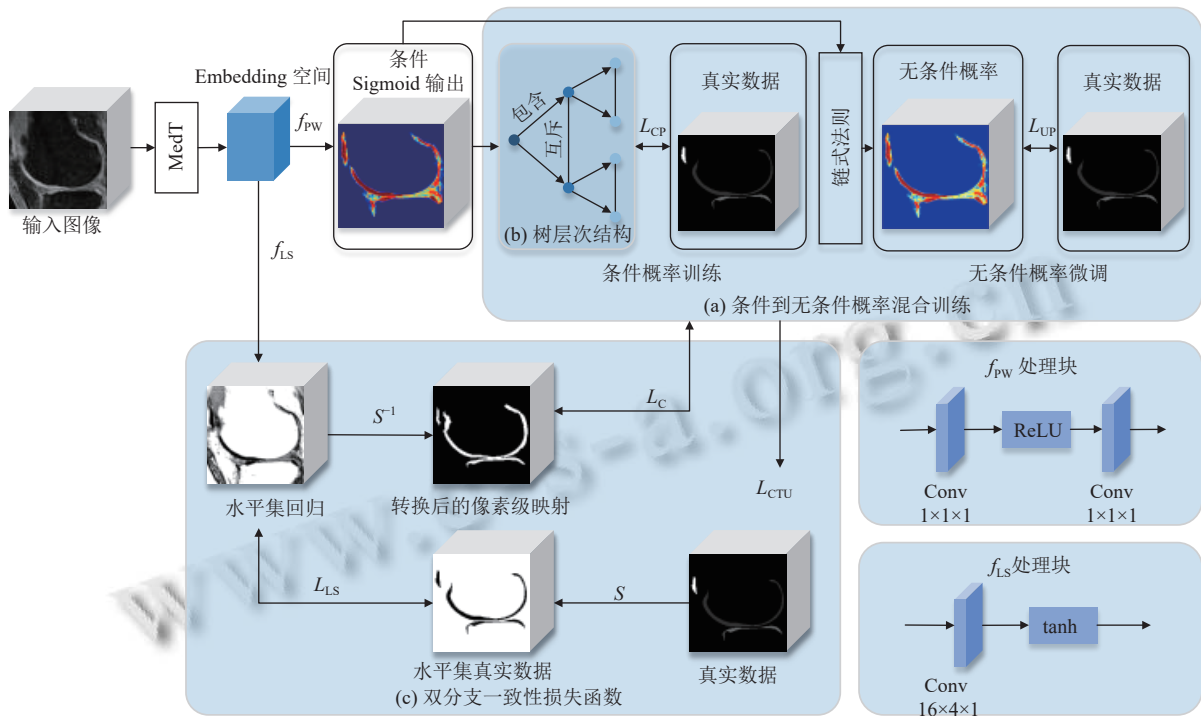


图2 SSHCPS 框架

具体地,向网络输入一个批图像特征  $C$ ,让网络输出胫骨软骨、股骨软骨和髌骨软骨的条件概率图:

$$P = Sigmoid(f_{PW}(C)) \in [0, 1]^{H \times W \times D \times 2} \quad (1)$$

其中, PW 表示像素级分支 (pixel wise map),  $P(\cdot)$  表示当前阶段的概率.  $f_{PW}$  包括中间有一个 ReLU 激活函数的两个  $1 \times 1 \times 1$  卷积. LS 表示水平集分支,  $f_{LS}$  是水平集转换和处理函数.

## 2.2 损失函数

为了应对模型标签关系利用不足的问题,我们设计了一个网络,该网络直接输出胫骨软骨、股骨软骨和髌骨软骨区域的条件概率.在该部分,每个分类器仅在其父标签为正的数据上训练.因此,训练直接对条件概率建模.分类器的共享部分,即来自主干网络的特征层,由所有任务联合训练.具体来说,对于每个图像,损失仅在其父标签也是正的标签上计算.针对该领域标注样本数量少、质量差的问题,我们提出双分支的半监督网络,在水平集回归分支和像素级分割分支上分别进行工作.

### 2.2.1 树层次结构形式化

如图 2(b),我们形式化树层次结构 (TH),以达到对

层次标签之间的语义关系进行编码的目的,进而实现带条件概率的训练.我们进一步对叶子类别施加多类交叉熵损失,这符合树层次结构中定义的细粒度类别之间的互斥约束.一个树层次结构  $T=(N, E_d, E_u)$  包括:节点集合  $N=\{n_1, n_2, \dots, n_k\}$ 、无向边  $E_u \subseteq N \times N$  和有向边  $E_d \subseteq N \times N$ .  $(v_i, v_j) \in E_u$  表示标签  $i$  和标签  $j$  互相排斥,  $(v_i, v_j) \in E_d$  表示标签  $i$  包含标签  $j$ .每个节点表示一个类标签,采用二进制值表示,即  $v_i \in \{0, 1\}$ .每条边定义了其关联节点的两个标签可以采取的二进制值的约束,所有合法全局分配的集合构成了树  $T$  的状态空间  $S_T \subseteq \{0, 1\}^k$ .

### 2.2.2 无条件概率辅助条件概率进行训练

在这部分,我们借助树层次结构来训练条件概率,设标签数为  $k$ ,  $y \in \{0, 1\}^k$  为标签向量,  $x$  是一个给定的输入图像,这里我们先对树  $T$  状态空间中的所有合法赋值  $\bar{y} \in S_T$  进行求和:

$$SUM(x) = \sum_{\bar{y} \in \{0, 1\}^k} \prod_{i=1}^k \phi_i(\bar{x}_i, \bar{y}_i) \prod_{i,j \in \{1, \dots, k\}} \psi_{i,j}(\bar{y}_i, \bar{y}_j) \quad (2)$$

其中,  $\phi_i(\bar{x}_i, \bar{y}_i) = e^{\bar{x}_i \delta_{\bar{y}_i=1}}$ ,  $\psi$  是标签层次带来的约束关系,违反约束其值为 0,符合约束则为 1.

然后,我们计算图像  $x$  关于向量  $y$  的所有  $S$  形节点的联合概率:

$$\tilde{P}(y|x) = \prod_{i=1}^k \phi_i(\bar{x}_i, y_i) \prod_{i,j \in \{1, \dots, k\}} \psi_{i,j}(y_i, y_j) \quad (3)$$

这样定义的联合概率有一个不合理的地方,由于  $\bar{x}_i \in (0, 1)$ , 有  $\phi_i(\bar{x}_i, 1) > 1$ , 因此样本属于粗粒度类别的联合概率一定大于样本属于细粒度类别的联合概率. 因此,我们使用  $SUM(x)$  对联合概率进行归一化:

$$P(y|x) = \frac{\tilde{P}(y|x)}{SUM(x)} \quad (4)$$

由此,我们可以得到  $x$  在某处被观察到的边际概率:

$$P(y_i = 1|x) = \frac{1}{SUM(x)} \sum_{\bar{y}_i=1} \prod_i \phi_i(\bar{x}_i, \bar{y}_i) \prod_{i,j} \psi_{i,j}(\bar{y}_i, \bar{y}_j) \quad (5)$$

细粒度标签对应的边际概率依赖于其父标签对应的 Sigmoid 输出,粗粒度标签对应的边际概率依赖于其所有子标签对应的 Sigmoid 输出,因此边际概率同时融合了不同粒度分类任务的特征信息. 当图像  $x$  在不同标签层次上被观察到时,我们的目标就是最大化对应的边际概率,也就是最小化如下损失函数:

$$L_{CP} = -\frac{1}{N} \sum_n \ln(P(y^n_m = 1|x^n)) \quad (6)$$

其中,  $N$  是总样本数,  $y^n$  是样本  $n$  的真值标签向量,  $m$  是被观测标签的索引.

带条件概率的训练是一个非常有效的初始化步骤. 因为它将建模能力仅集中于在同一父标签下区分兄弟姐妹,而不是必须在所有标签之间区分,这简化了收敛并减少了混淆因素.

在第 2 阶段,我们使用无条件概率 (UP)  $CE$  损耗微调模型. 当使用无条件概率进行训练时,对每个数据实例的每个分类器输出计算损失,可以写成:

$$L_{UP} = \sum_{d \in D} CE(y_d, p_d) \quad (7)$$

对于  $p_d$ , 使用链式规则计算标签  $d$  的预测无条件概率,同时,为了防止概率值乘积可能引起的算数下溢,我们使用 Logistic 函数计算该步,表示为:

$$p_d = \prod_{d'} \frac{1}{1 + \exp(-z_{d'})} \quad (8)$$

其中,  $d' \in U(d)$ ,  $U(d)$  表示标签  $d$  及其祖先的并集,  $z_{d'}$  表示 logit 输入,将式 (7) 展开:

$$\begin{aligned} L_{UP} &= \sum_{d'} (-y_d \log(p_d) - (1 - y_d) \log(1 - p_d)) \\ &= -y_d \log \left( \prod_{d'} p_d \right) - (1 - y_d) \log \left( \prod_{d'} (1 - p_d) \right) \\ &= -y_d \log \left( \prod_{d'} p_d \right) - (1 - y_d) \log \left( 1 - \prod_{d'} (p_d) \right) + \gamma \quad (9) \end{aligned}$$

对于单个  $CE$  项:

$$\begin{aligned} L_{UP} &= \sum_{d'} \left( -y_d \log \left( \frac{1}{1 + \exp(-z_{d'})} \right) \right. \\ &\quad \left. - (1 - y_d) \log \left( 1 - \frac{1}{1 + \exp(-z_{d'})} \right) \right) + \gamma \\ &= \sum_{d'} H + \gamma \quad (10) \end{aligned}$$

其中,  $H$  是交叉熵项,  $y_d$  是基准真值标签,  $\gamma$  须满足:

$$\begin{aligned} \gamma &= (1 - y_d) \log \left( \prod_{d'} \left( 1 - \frac{1}{1 + \exp(-z_{d'})} \right) \right) \\ &\quad - (1 - y_d) \log \left( 1 - \prod_{d'} \frac{1}{1 + \exp(-z_{d'})} \right) \\ &= (1 - y_d) \log \left( \frac{\prod_{d'} \exp(-z_{d'})}{\prod_{d'} (1 + \exp(-z_{d'}))} \right) \\ &\quad - (1 - y_d) \log \left( \frac{\prod_{d'} (1 + \exp(-z_{d'})) - 1}{\prod_{d'} (1 + \exp(-z_{d'}))} \right) \\ &= (1 - y_d) \log \left( \frac{\exp \left( \sum_{d'} -z_{d'} \right)}{\prod_{d'} (1 + \exp(-z_{d'})) - 1} \right) \\ &= (1 - y_d) \left( \sum_{d'} -z_{d'} - \log \left( \prod_{d'} (1 + \exp(-z_{d'})) - 1 \right) \right) \\ &= (1 - y_d) \left( \sum_{d'} -z_{d'} - LSE \left( \left\{ \sum_{i \in S} -z_i \cdot \forall S \in P(U(d)) \setminus \{\emptyset\} \right\} \right) \right) \quad (11) \end{aligned}$$

其中,  $LSE$  是 LogSumExp 函数,  $p(x)$  表示  $x$  的幂集.

因  $CE$  和  $LSE$  函数数值是稳定的,将式 (11) 代入式 (10) 后所得出的无条件概率损失函数的数值稳定性也得以保证.

### 2.2.3 双分支差异损失函数

对于水平集分支,分割图向水平集的转换通过如下函数进行:

$$S(x) = \begin{cases} -\inf_{y \in \partial\Omega} \|x-y\|_2, & x, y \in \Omega_0 - \partial\Omega_0 \\ 0, & x, y \in \partial\Omega_0 \\ +\inf_{y \in \partial\Omega} \|x-y\|_2, & x, y \in \Omega - \Omega_0 \end{cases} \quad (12)$$

其中,  $x$  和  $y$  是分割掩模中的两个不同体素,  $\|x-y\|$  表示欧氏距离,  $\Omega$  表示整个目标对象,  $\partial\Omega_0$  是零水平集, 也表示对象的轮廓,  $\Omega_0$  表示包括轮廓在内的目标对象的内部区域。

为了将水平集任务的输出映射到分割输出空间, 就需要用到  $S(x)$  的逆变换, 由于  $S(x)$  是不可微的, 不可用于反向传播, 因此, 我们需要对  $S^{-1}$  使用平滑近似, 通过该近似方法, 我们希望使得: 在转换后的输出空间映射图中, 目标物体的值为 1, 背景的为 0.  $S^{-1}$  的定义为:

$$S^{-1}(z) = \frac{1}{1 + e^{-iz}} \quad (13)$$

其中,  $z$  表示在体素  $x$  处的水平集值,  $i$  是一个输入因子, 为实现精确的逆变换, 它应当尽可能的大。

对于水平集分支, 该分支在符号距离空间上工作, 强调几何结构信息, 而另一分支则强调像素级推理, 如此, 两个分支之间必然存在预测上的差异, 我们引入一个双分支差异损失  $L_C$ , 它使用像素级预测和变换后的水平集预测得出:

$$L_C = \sum_{x_d \in D'} \|f_{PW}(x_d) - S^{-1}(f_{LS}(x_d))\|^2 \quad (14)$$

其中,  $D'$  是有标签数据和无标签数据的并集,  $f_{PW}(\cdot)$  是像素级分支输出,  $f_{LS}(\cdot)$  是水平集分支输出。

#### 2.2.4 双分支联合损失

在这部分, 我们还需要得到两个分支上各自的损失函数, 对于该部分水平集分支, 我们使用监督学习, 由水平集预测和经过  $S$  转换的水平集基准真值得到:

$$\begin{aligned} L_{LS} &= KL(f_{LS}(x_d) \| S(y_d)) \\ &= H(f_{LS}(x_d), S(y_d)) - H(f_{LS}(x_d)) \\ &= \sum_{x_d, y_d \in D} \left( f_{LS}(x_d) \cdot \log \frac{1}{S(y_d)} \right) \\ &\quad - \sum_{x_d, y_d \in D} \left( f_{LS}(x_d) \cdot \log \frac{1}{f_{LS}(x_d)} \right) \\ &= \sum_{x_d, y_d \in D} f_{LS}(x_d) \cdot \log \frac{f_{LS}(x_d)}{S(y_d)} \end{aligned} \quad (15)$$

其中,  $KL$  是相对熵,  $H$  是交叉熵项,  $D$  代表所有有标签数据,  $y_d$  是输入数据  $d$  的基准真值分割掩码. 这部分是只使用有标签数据得到的. 像素级分支损失即第 2.2.2

节中的条件到无条件概率混合训练损失:

$$\begin{aligned} L_{CTU} &= L_{CP} + L_{UP} \\ &= -\frac{1}{N} \sum_n \ln(P(y^n = 1|x^n)) + \sum_{d \in D} CE(y_d, p_d) \end{aligned} \quad (16)$$

其中,  $L_{CP}$  和  $L_{UP}$  分别是条件概率损失和无条件概率损失,  $d$  表示数据标签,  $y_d \in \{0, 1\}$  表示  $d$  的基准真值标签,  $p_d$  是标签  $d$  的预测无条件概率,  $N$  是总样本数,  $y^n$  是样本  $n$  的真值标签向量,  $I^n$  是被观测标签的索引. 如此, 我们便得到了总的损失函数:

$$L_t = L_{CTU} + \alpha L_C + \beta L_{LS} \quad (17)$$

其中,  $\alpha$  和  $\beta$  是用来控制半监督和有监督之间平衡的系数, 系数的确定取决于有标签数据和无标签数据之间的数量比例. 根据文献[7,8,18], 我们将  $\alpha$  确定为 1,  $\beta$  确定为:

$$\beta(t) = 0.1 \cdot e^{-5\left(1 - \frac{t}{t_{\max}}\right)^2} \quad (18)$$

其中,  $t$  表示当前训练步骤,  $t_{\max}$  表示最大训练步骤。

## 3 实验

### 3.1 实验数据集

本文使用的数据集是 fastMRI<sup>[19]</sup> 数据集集中的软骨图像数据和 MOST (multicenter osteoarthritis study) 数据集<sup>[20]</sup> 中的膝关节 MRI 图像. fastMRI 是 Facebook AI research (FAIR) 和 NYU Langone health 的合作研究项目. MOST 是一个公开的膝关节影像数据集. 它由 NIH (美国国立卫生研究院) 发起, 主要关注膝关节等部位的影像数据. 我们在这两个数据集中分别选取了 100 张膝关节 MRI 扫描图像, 每张 MRI 图像包含 6 个标签: 股骨、股骨软骨、胫骨、胫骨软骨、髌骨和髌骨软骨. 我们按 7:2:1 的比例将这些 MRI 图像分为训练集、测试集和验证集, 将每张 3 通道图像的大小裁剪成 288×240 像素, 并通过归一化将每张切片的像素值限制在 0-255 之间。

### 3.2 实验细节及评价指标

我们采用尺寸为 288×240 的随机采样块进行训练. 最终选定的模型展现出所有模型中最高的指数移动平均值 (EMA-VL). 在训练过程中, 我们采用了多种数据增强技术, 如随机缩放、随机翻转、高斯噪声添加以及对比度随机调整。

进行本文实验的训练与测试环境均为 Ubuntu 20.04

操作系统, 硬件为一块 NVIDIA GeForce RTX 3090 显卡, 编程语言为 Python 3.7, 使用 PyTorch 深度学习框架完成整个模型的训练与测试过程. 在这项工作中, 我们使用 MedT<sup>[21]</sup>作为 backbone, 并通过在原始 MedT 末尾添加新的回归层来实现双任务 MedT. 该框架使用 Adam 优化器做梯度下降算法, 默认参数设置为  $\beta_1=0.9, \beta_2=0.99$ , 权重衰减率为  $1 \times 10^{-5}$ . 在网络的半监督分支, 初始学习率为 0.01, 每 200 次迭代衰减 0.1, 批大小为 4, 包含 2 张带标签图像和 2 张无标签图像. 在全监督分支的两个训练阶段, 我们都采用了自适应学习率, 第 1 阶段的学习率设定为  $1 \times 10^{-3}$ , 第 2 阶段调整为  $1 \times 10^{-4}$ . 若在经过 40 个轮次后 EMA-VL 趋于稳定, 则学习率将减少一半. 最大训练轮数设置为 500 轮. 若指数移动平均值在 60 个轮次后保持不变, 训练将提前终止. 为了防止过拟合, 在训练阶段使用标准实时数据增强方法.

根据文献, 我们设置有标签数据和无标签数据的数量为 1:4, 即我们共使用 70 张 MRI 图像, 其中有标

签数据 14 张, 无标签数据 56 张.

为了公平比较, 我们不使用任何后处理或集成方法. 为了评估分割的准确性, 我们研究了每个样本的 Dice 和 95% Hausdorff distance (95HD).

### 3.3 实验结果和分析

#### 3.3.1 消融实验

为了研究不同任务对个体的影响, 我们首先只使用标记图像进行训练, 并分析此时双分支差异性如何执行. 在 fastMRI 数据集上, 根据 Yu 等<sup>[8]</sup>, 我们分别使用 14 个标记数据和全部 70 个标记数据来训练网络.

我们比较了不同的训练策略: (1) 仅使用像素级 (Seg) 分支的条件概率 (TH) 阶段; (2) 仅使用像素级 (Seg) 分支; (3) 使用像素级 (Seg) 分支和水平集任务 (LS) 分支; (4) 以及我们的半监督层次条件概率 (SSHCPs) 方法. 这些变体的性能如表 1 所示, 可以看出, 引入无条件概率 (UP) 微调和水平集 (LS) 回归对结果有着显著提高. 在表 1 中 TC 代表胫骨软骨, FC 代表股骨软骨, PC 代表髌骨软骨.

表 1 fastMRI 数据集上的消融实验结果

Methods	Scans used		Dice (%)			95HD (voxel)		
	Labeled	Unlabeled	TC	FC	PC	TC	FC	PC
Seg (TH)	14	0	77.32	78.36	77.23	11.37	10.98	10.65
Seg (TH+UP)	14	0	77.65	78.51	77.72	10.59	10.23	10.11
Seg+LS	14	0	78.16	80.03	78.19	9.75	9.08	9.59
SSHCPs	14	0	<b>78.53</b>	<b>80.74</b>	<b>78.76</b>	<b>9.17</b>	<b>8.66</b>	<b>9.15</b>
Seg (TH)	70	0	85.56	87.05	85.21	5.68	5.62	5.57
Seg (TH+UP)	70	0	85.77	87.21	85.63	5.60	5.53	5.50
Seg+LS	70	0	86.18	88.79	86.37	5.43	5.32	5.39
SSHCPs	70	0	<b>86.36</b>	<b>89.22</b>	<b>86.48</b>	<b>5.35</b>	<b>5.26</b>	<b>5.31</b>

此外, 我们在 MOST 数据集上比较了我们的半监督 SSHCPs 方法与全监督 HCPS 及全监督的 MedT (Baseline) 之间的数据利用表现. 如图 3 所示, 我们绘制了 3 块软骨测试结果的平均 Dice 得分. 可以观察到, 在不同的标签数据设置中, 半监督双任务一致性方法始终比全监督方法表现得更好, 这表明我们的方法有效地利用了未标记的数据并带来了性能提升. 当标记数据的数量很少时, 我们的方法也获得了比全监督方法更好的分割结果, 表明我们方法在临床应用中的巨大潜力. 另外, 我们的全监督 HCPS 方法的 Dice 曲线始终在 MedT 之上, 进一步证明了我们方法的有效性.

与单任务相比, 双任务一致性提高了分割精度. 在不同数据集上, 双任务一致性表现一致, 这进一步证明了多任务学习的有效性. 总之, 这项研究表明, 通过无

条件概率微调和结合双任务一致性方法与水平集函数回归, 可以提高医学影像分割的准确性和效率.

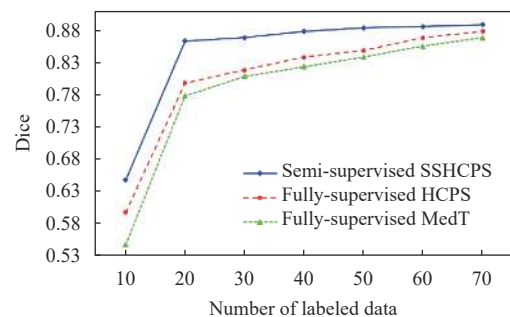


图 3 本文的半监督方法在 MOST 数据集上, 不同数量标记数据下的表现与全监督方法的对比

#### 3.3.2 对比实验

我们使用了 4 个最先进半监督方法: 校准标签分

布算法 (CLD)<sup>[22]</sup>, GenericSSL<sup>[23]</sup>, 双任务一致性算法 (DTC)<sup>[24]</sup>, 交叉一致性训练 (CCT)<sup>[25]</sup>和基线模型 medical Transformer (MedT) 与我们的方法进行对比, 3D 可视化图如图 4 所示. 如表 2 和表 3 所示, 分别在 fastMRI 和 MOST 数据集上评估了我们的框架. 可以看到, 在两个数据集上, 与仅使用 14 个标记数据的全监督 MedT 相比, 所有利用未标记图像的半监督方法都显著提高

了分割性能. 如表 2 所示, 在 fastMRI 数据集上, 本文方法的 Dice 得分均高于表中所列的其他半监督方法, 如表 3 所示, 在 MOST 数据集上, 本文方法在 3 个区域的评价标准上都取得了出色的表现, 除了在 PC 区域上的 Dice 得分略低于 CCT. 值得注意的是, CCT 方法是通过多域交叉一致性来训练网络的, 需要比较复杂的编码和解码, 且需要使用额外的图像级标签.

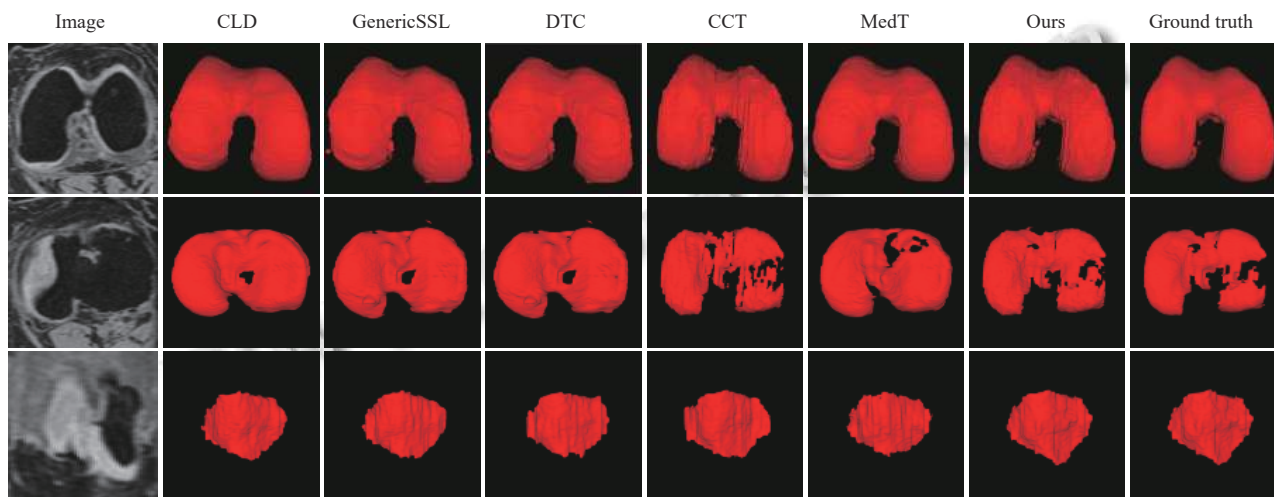


图 4 不同方法在膝关节 MRI 图像上的 3D 可视化图

表 2 不同方法在 70 张选自 fastMRI 数据集的膝关节 MRI 扫描图像上的表现

Methods	Scans used		Dice (%)			95HD (voxel)			Cost Params (M)
	Labeled	Unlabeled	TC	FC	PC	TC	FC	PC	
MedT (Baseline)	70	0	85.26	86.41	85.81	5.71	5.06	5.60	1.73
MedT (Baseline)	14	0	76.57	77.88	76.08	10.93	9.15	11.83	1.73
CCT	14	56	84.29	86.10	85.35	6.42	6.38	6.40	15.65
CLD	14	56	84.25	86.05	85.22	6.47	6.41	7.01	75.55
GenericSSL	14	56	83.63	84.92	84.13	8.11	7.81	8.97	57.89
DTC	14	56	84.11	85.76	85.29	6.49	6.46	6.43	9.44
Ours	14	56	<b>84.36</b>	<b>86.22</b>	<b>85.68</b>	<b>6.35</b>	<b>6.26</b>	<b>6.31</b>	<b>1.73</b>

表 3 不同方法在 70 张选自 MOST 数据集的膝关节 MRI 扫描图像上的表现

Methods	Scans used		Dice (%)			95HD (voxel)		
	Labeled	Unlabeled	TC	FC	PC	TC	FC	PC
MedT (Baseline)	70	0	86.16	89.62	86.94	5.13	4.86	5.36
MedT (Baseline)	14	0	78.56	79.35	78.95	8.43	7.80	8.32
CCT	14	56	85.57	87.11	<b>86.23</b>	6.32	6.18	6.29
CLD	14	56	85.32	87.06	85.88	6.37	6.24	6.30
GenericSSL	14	56	85.02	86.36	85.26	7.45	7.31	7.41
DTC	14	56	85.43	86.99	85.85	6.35	6.22	6.33
Ours	14	56	<b>85.69</b>	<b>87.16</b>	85.93	<b>6.30</b>	<b>6.16</b>	<b>6.27</b>

另外, 在表 2 和表 3 中可以看到, 我们的 95HD 指标比其他方法更低, 从图 4 中可以看到, 我们方法的分割结果与真实值重叠率更高, 产生的假阳性更少且保留了更多细节, 这进一步证明了我们提出的方法的有

效性、通用性和鲁棒性. 且与其他方法相比, 我们的方法网络结构较简单, 得益于先进的 MedT 方法, 我们的网络具有较少的参数, 并且无需在迭代过程中多次传递图像, 实验表明, 我们的方法在准确度、网络参数和



计算成本之间取得了最佳的折衷。

## 4 结论与展望

在本文中,我们介绍了双分支联合训练,这是一种基于条件概率和任务级一致性的半监督分割框架。我们使用双分支网络,同时预测像素级分类图和能够捕获全局形状和几何信息分割的水平集表示,最大化地利用图像标签信息。我们的方法在 fastMRI 和 MOST 数据集上均取得了卓越的结果。出色的性能证明了我们提出的方法的有效性、稳健性和泛化能力。在本文中,我们专注于多标签分割,同时,我们的方法在单类分割任务中也同样性能卓越。

在未来,我们将扩展此方法以用于更多计算机视觉应用,以减少注释工作并进一步研究融合策略,以组合所有不同任务的预测结果以获得更好的性能。

### 参考文献

- 1 翟万江. 热情拥抱人工智能共促医学影像创新发展——第二届中国医学影像 AI 大会侧记. 中国科技产业, 2020(10): 14–17. [doi: [10.16277/j.cnki.cn11-2502/n.2020.10.006](https://doi.org/10.16277/j.cnki.cn11-2502/n.2020.10.006)]
- 2 Zhao CJ, Xiang S, Wang YQ, *et al.* Context-aware network fusing Transformer and V-Net for semi-supervised segmentation of 3D left atrium. *Expert Systems with Applications*, 2023, 214: 119105. [doi: [10.1016/j.eswa.2022.119105](https://doi.org/10.1016/j.eswa.2022.119105)]
- 3 Portela NM, Cavalcanti GDC, Ren TI. Semi-supervised clustering for MR brain image segmentation. *Expert Systems with Applications*, 2014, 41(4): 1492–1497. [doi: [10.1016/j.eswa.2013.08.046](https://doi.org/10.1016/j.eswa.2013.08.046)]
- 4 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference*. Munich: Springer, 2015. 234–241.
- 5 Berthelot D, Carlini N, Goodfellow I, *et al.* MixMatch: A holistic approach to semi-supervised learning. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2019. 454.
- 6 Li XM, Yu LQ, Chen H, *et al.* Transformation-consistent self-ensembling model for semisupervised medical image segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(2): 523–534. [doi: [10.1109/TNNLS.2020.2995319](https://doi.org/10.1109/TNNLS.2020.2995319)]
- 7 Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 1195–1204.
- 8 Yu LQ, Wang SJ, Li XM, *et al.* Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. *Proceedings of the 22nd International Conference on Medical Image Computing and Computer Assisted Intervention*. Shenzhen: Springer, 2019. 605–613.
- 9 Xia YD, Liu FZ, Yang D, *et al.* 3D semi-supervised learning with uncertainty-aware multi-view co-training. *Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision*. Snowmass: IEEE, 2020. 3635–3644. [doi: [10.1109/WACV45572.2020.9093608](https://doi.org/10.1109/WACV45572.2020.9093608)]
- 10 Eric PX, Andrew YN, Jordan MI, *et al.* Distance metric learning, with application to clustering with side-information. *Proceedings of the 15th International Conference on Neural Information Processing Systems*. Vancouver: MIT Press, 2003. 521–528.
- 11 Zhuang ZF, Kong XN, Rundensteiner E, *et al.* MLAS: Metric learning on attributed sequences. *Proceedings of the 2020 IEEE International Conference on Big Data*. Atlanta: IEEE, 2020. 1449–1454. [doi: [10.1109/BigData50022.2020.9378490](https://doi.org/10.1109/BigData50022.2020.9378490)]
- 12 Thaler S, Menkovski V, Petkovic M. Deep metric learning for sequential data using approximate information. *Proceedings of the 14th International Conference on Machine Learning and Data Mining in Pattern Recognition*. New York: Springer, 2018. 269–282. [doi: [10.1007/978-3-319-96136-1\\_22](https://doi.org/10.1007/978-3-319-96136-1_22)]
- 13 Song HO, Xiang Y, Jegelka S, *et al.* Deep metric learning via lifted structured feature embedding. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 4004–4012. [doi: [10.1109/CVPR.2016.434](https://doi.org/10.1109/CVPR.2016.434)]
- 14 Du RY, Chang DL, Bhunia AK, *et al.* Fine-grained visual classification via progressive multi-granularity training of jigsaw patches. *Proceedings of the 16th European Conference Computer Vision*. Glasgow: Springer, 2020. 153–168. [doi: [10.1007/978-3-030-58565-5\\_10](https://doi.org/10.1007/978-3-030-58565-5_10)]
- 15 Dubey A, Gupta O, Guo P, *et al.* Improving fine-grained visual classification using pairwise confusion. *arXiv:1705.08016*, 2018.
- 16 Hu T, Xu JZ, Huang C, *et al.* Weakly supervised bilinear attention network for fine-grained visual classification.

- arXiv:1808.02152, 2019.
- 17 Zhuang PQ, Wang YL, Qiao Y. Learning attentive pairwise interaction for fine-grained classification. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020. 13130–13137. [doi: [10.1609/aaai.v34i07.7016](https://doi.org/10.1609/aaai.v34i07.7016)]
- 18 Laine S, Aila T. Temporal ensembling for semi-supervised learning. Proceedings of the 5th International Conference on Learning Representations. arXiv:1610.02242v3, 2017.
- 19 Knoll F, Zbontar J, Sriram A, *et al.* fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. Radiology: Artificial Intelligence, 2020, 2(1): e190007. [doi: [10.1148/ryai.2020190007](https://doi.org/10.1148/ryai.2020190007)]
- 20 Segal NA, Nevitt MC, Gross KD, *et al.* The multicenter osteoarthritis study: Opportunities for rehabilitation research. PM& R, 2013, 5(8): 647–654. [doi: [10.1016/j.pmrj.2013.04.014](https://doi.org/10.1016/j.pmrj.2013.04.014)]
- 21 Valanarasu JM, Oza P, Hacihaliloglu I, *et al.* Medical transformer: Gated axial-attention for medical image segmentation. Proceedings of the 24th International Conference on Medical Image Computing and Computer-assisted Intervention. Strasbourg: Springer, 2021. 36–46.
- 22 Lin YQ, Yao HF, Li ZZ, *et al.* Calibrating label distribution for class-imbalanced barely-supervised knee segmentation. Proceedings of the 25th International Conference on Medical Image Computing and Computer Assisted Intervention. Singapore: Springer, 2022. 109–118.
- 23 Wang HN, Li XM. Towards generic semi-supervised framework for volumetric medical image segmentation. Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2024. 88.
- 24 Luo XD, Chen JN, Song T, *et al.* Semi-supervised medical image segmentation through dual-task consistency. Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI Press, 2021. 8801–8809.
- 25 Ouali Y, Hudelot C, Tami M. Semi-supervised semantic segmentation with cross-consistency training. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 12674–12684.

(校对责编: 王欣欣)