arXiv:2304.00801v3 [cs.CV] 4 May 2023

使用软 Dice 进行噪声图像分割

Marcus Nordström *瑞典皇家理工学院数学系，瑞典斯德哥尔摩marcno@kth.se

Henrik Hult 瑞典皇家理工学院数学系，瑞典斯德哥尔摩 hult@kth.se

Atsuto Maki 瑞典皇家理工学院计算机科学系，瑞典斯德哥尔摩 atsuto@kth.se

Fredrik Löfman 机器学习系 RaySearch Laboratories 瑞典斯德哥尔摩 fredrik.lofman@raysearchlabs.com

摘要

本文研究了软 *Dice* 损失——医学图像分割中最常用的损失函数之一——在目标标签存在噪声时的表现。具体而言，本文刻画了最优解的集合，并给出了这些解的体积偏差的紧确界。此外，研究表明，一系列收敛到最优软 *Dice* 的软分割，在通过阈值化转换为硬分割后，也会收敛到最优 *Dice*。这是一项重要结果，因为软 *Dice* 常被用作最大化 *Dice* 指标的替代目标。最后，本文通过实验验证了理论结果。

1. 引言

如今，图像分析任务中的许多先进方法都基于监督学习。对于医学图像分割问题，最常用的一类模型是采用类似 U-Net 架构的深度神经网络 [26]。这类模型有大量文献支持，并且经常在医学图像分割竞赛中名列前茅。相关综述参见 [29] 和 [8]。

这些机器学习方法的共同点是，它们都是使用某些数据集和某种损失函数进行训练的。损失函数用于量化模型对图像的预测结果与同一图像对应的真实标签之间的差异。此外，这种量化方式经过设计，可以通过连续优化方法进行优化，例如该领域广泛使用的随机梯度下降方案。对于仅包含一个目标结构的分割问题，也就是说，对于二分类问题，二元交叉熵长期以来一直被视为标准选择。这种损失函数具有若干实用的理论性质，例如凸性，并且已知在实践中表现可靠。

当数据极不平衡时，也就是说，当前景比背景常见得多，或反之亦然时，交叉熵往往无法体现相关定义域中医学从业者的优先考量。这是因为该损失函数对前景体素和背景体素的改进一视同仁，而医学从业者通常认为，相比背景，更重要的是准确处理前景及其周围区域的细节。由于类别不平衡在医学图像分割中十分常见，人们提出并评估了被认为更适合这种情况的替代损失函数。

其中一种损失函数基于 Sörensen-Dice 系数的 Sigmoid 平滑版本，简称 Dice。除了一些文献报告其表现不稳定外，这种损失函数通常称为软 Dice，或简称 Dice 损失，已在许多实验研究中被证明能够取得良好表现 [13,22] [4,7,9,13,31]。因此，它与交叉熵一同成为医学图像分割领域最常用的损失函数，且

*作者还隶属于 RaySearch Laboratories。

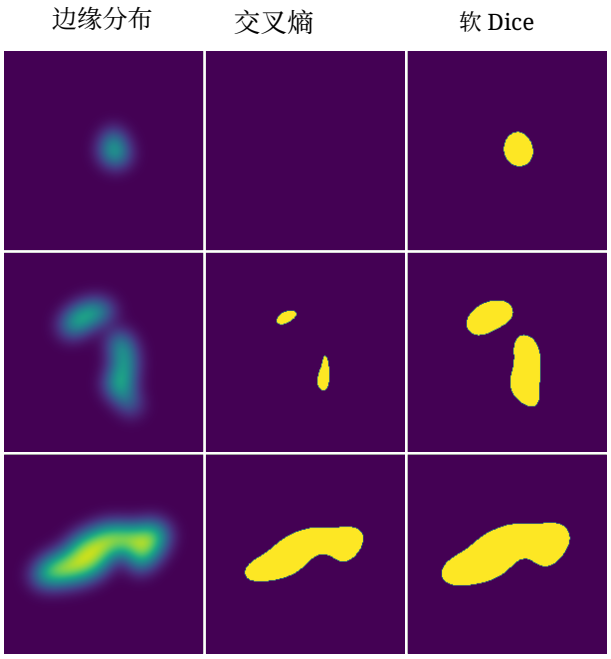

图 1。最左侧为含噪分割的逐像素边缘概率（软标签）。中间为通过最小化交叉熵并采用 $1/2$ 阈值这一标准流程得到的分割。最右侧为通过最小化软 Dice 并采用 $1/2$ 阈值这一标准流程得到的分割。根据所述理论，通过对软 Dice 优化器进行 $1/2$ 阈值化得到的分割体积，始终大于或等于通过对交叉熵优化器进行 $1/2$ 阈值化得到的分割体积。

应用于 nnUNet 等热门框架中 [12]。除了软 Dice，文献中还提出并研究了其他几种损失函数，例如 Jaccard 损失、Tversky 损失、加权交叉熵损失、TopK 损失和边界损失 [20]。然而，这些损失在实践中远不如 Dice 损失常见，这至少部分可能是因为该领域长期以来习惯使用 Dice 指标进行性能评估。

除了类别不平衡问题外，影响分割模型性能的另一个因素是标签噪声的存在 [1, 6, 21, 24]。标签噪声源于图像分割（尤其是医学图像分割）中极其繁琐的标注过程；医学图像分割通常涉及三维图像，边界可能不清晰，且需要医学专业知识。由于用于训练分割模型的数据中普遍存在标签噪声，因此，了解其如何影响模型性能，尤其是其与所选损失函数之间的相互作用，至关重要。

已有多项研究致力于理解标签噪声对使用软 Dice 损失训练的分割模型所产生的影响

然而，这些研究几乎全是实验性的，鲜少得出一般性原理。在现有的理论研究中，已有结果表明，与交叉熵不同，最小化软 Dice 不会得到前景的逐像素边缘分布概率，而且相关体积可能存在偏差 [5]。此外，有人推测，软 Dice 对 Dice [22], 是校准的，这意味着一系列软分割在软 Dice 意义下收敛至最优解后，经过阈值化转换为硬分割时，也会收敛至最优 Dice。这些进展十分重要，因为它们表明，我们对标签噪声下软 Dice 的行为仍存在巨大的认知空白；这也意味着，尽管众多论文致力于训练日益复杂的图像分割模型，但我们并不清楚这些模型实际受训要完成什么任务。

本文全面阐述了标签噪声下损失函数的理论性质。具体而言，本文研究了软 Dice 的最优解性质、这些解的体积偏差，以及软 Dice 对 Dice 的校准性质。结果证实了实践中观察到的一些实验现象，也带来了此前未曾发现的新见解。

贡献： 本文从理论上研究了存在标签噪声时软 Dice 的性质。本文刻画了最优解，给出了相关解的体积偏差的紧确界，并证明软 Dice 对 Dice 具有校准性。最后，本文报告了与 Gold Atlas 项目 [24] 相关的九个不同含噪器官分割问题，以及通过合成数据生成的九个含噪分割问题的实验结果。

2. 相关工作

医学图像分割中存在标签噪声，这一点已广为人知，并在许多研究中有所提及。相关示例包括 [1, 6, 21, 24]。

有几篇论文探讨了噪声对软 Dice 损失的影响。[5], 研究表明，在标签含噪的条件下优化软 Dice，通常无法得到前景的像素级概率估计。此外，研究还表明，由此得到的估计量的体积通常也不等于含噪目标的平均体积。 [27],进一步研究了第一个问题，并提出了一种对软 Dice 得到的估计结果进行重新校准的方法。[25]则进一步阐述了体积偏差这一问题。另有研究基于最优传输理论采用了不同的方法来处理这一问题 [19]。在噪声标签与软 Dice 的研究中，另一种处理噪声的方式是将噪声作为软标签纳入其中 [10, 15]。这一点对本研究尤为重要，因为它与所研究的问题密切相关。

在二元分类中，[18, 34] 讨论了针对 F_1 指标的最优阈值分类器，其他研究者随后对此作了进一步阐述。由于分割中的 Dice 指标与二元分类中的 F_1 指标密切相关，这一思路后来被引入分割领域，并用于刻画相对于 Dice 的最优分割 [23]。此外，研究还给出了这些解的体积的紧界。受二元分类相关研究 [2, 3] 的启发，研究者提出，校准可能是解释软 Dice 在采用 Dice 进行评估时表现良好的一个重要特性 [22]。

关于含噪分割的相关实验工作包括 [11]，其中作者考察了三种不同的标签噪声模型，并观察到噪声水平的增加会导致性能下降。[33]随后描述了一项类似的实验研究，其中也包括了对偏倚噪声的实验。最后，关于采用软 Dice 以外的其他损失函数处理软标签的实验工作包括 [16, 17, 30]。关于含噪分割和不完美数据研究的近期综合综述，参见 [32] 和 [14]。

3. 预备知识

本文将在连续图像域上进行分析，而不是在离散化区域上进行分析。这是因为部分结果在这一设置下能够得到更清晰的表述。一般而言，如果考虑适当的阶跃函数，则离散化设置可以看作连续设置的特例。

3.1. 符号

令 $\Omega = [0,1]^n \subset \mathbb{R}^n$ 为维度为$n \geq 1$ 的单位立方体，λ 为其对应的标准归一化勒贝格测度，使得 $\lambda(\Omega) = 1$。令 $\mathcal{S}$ 为从 Ω 到二进制数 $\{0,1\}$的可测函数空间，$\mathcal{M}$为从 Ω 到闭区间 $[0,1]$ 的可测函数空间，$\mathcal{F}$ 为从 Ω 到 $(-\infty, \infty)$ 的有界可测函数空间。所有函数空间 $\mathcal{S}$、$\mathcal{M}$ 和 $\mathcal{F}$ 均配备相应的 Borel σ-域和标准 L_1-范数 $\|\cdot\|_1$。字母 s、m、c 和 f 用于表示特定类型的对象。字母 $s \in \mathcal{S}$ 表示一种分割，其中 $s(\omega) = 1$表示感兴趣目标结构的前景，$s(\omega) = 0$ 表示背景。字母 $m \in \mathcal{M}$ 表示边缘概率函数，即 $m(\omega) \in [0,1]$ 表示噪声标签占据 Ω 中位置 $\omega \in$ 的概率。字母 $c \in \mathcal{M}$表示通过优化所选损失函数、在任何阈值化之前得到的软分割。字母 $f \in \mathcal{F}$ 表示通常与软分割相关联的 logit 函数，使得对 Ω 中几乎处处的 $\omega \in$，$\sigma \circ f(\omega) = c(\omega)$，其中 $\sigma(x) = 1/(1+e^{-x})$为标准 Sigmoid 函数。记号 $\circ$ 表示函数的复合，$I_A(x)$表示集合上的指示函数集合 A上的指示函数，即若 $x \in A$，则 $I_A(x) = 1$；若 $x \notin A$，则 $I_A(x) = 0$。特别地，该记号将与用于对软分割进行阈值化的某个元素 $c \in \mathcal{M}$结合使用，例如$I_{[t,1]} \circ c$：当 $c(\omega) \geq t$ 时，它对 Ω 中所有 $\omega \in$ 赋值为 1；当 $c(\omega) < t$时，则对 Ω 中所有 $\omega \in$ 赋值为 0。

3.2. Dice 和软 Dice

存在标签噪声时，文献中有两种不同的流行方式来定义 Dice。第一种方式是相对于边缘概率或其估计值来计算 Dice，有时也称为软标注 [10, 15–17, 30]。定义 Dice 的第二种方式是取期望 Dice 得分。这种方式自然出现在经验风险最小化中，也是例如 [5]所研究的内容。本文遵循第一种约定。

定义 1. 对于任意 $m \in \mathcal{M}$，*Dice* 定义为

$$\mathrm{D}_m(s) \doteq \frac{2\int_\Omega s(\omega)m(\omega)\lambda(d\omega)}{\|s\|_1 + \|m\|_1}, \quad s \in \mathcal{S}. \tag{1}$$

软 Dice 定义为 1 减去 Dice，且其定义域扩展为软分割空间，而非硬分割。有时会省略常数项，但由于常数不会影响优化，因此这不会影响任何相关行为。

定义 2. 对于任意 $m \in \mathcal{M}$，软 *Dice* 定义为

$$\mathrm{SD}_m(c) \doteq 1 - \frac{2\int_\Omega c(\omega)m(\omega)\lambda(d\omega)}{\|c\|_1 + \|m\|_1}, \quad c \in \mathcal{M}. \tag{2}$$

对于取值于 $\mathcal{S}$ 且满足$\mathbb{E}[L(\omega)] = m(\omega)$ $\omega \in \Omega$ 的含噪分割 L，使用期望 Dice 得到的相应 $\mathrm{D}_m(s)$ $s \in \mathcal{S}$ 定义可表示为 $\mathbb{E}[\mathrm{D}_L(s)]$，$s \in \mathcal{S}$。关于这些定义之间关系的简要讨论见 [23]。简言之，当噪声标签体积的方差为零时，这两个定义等价；当方差较小时，它们也非常相似。在医学图像分割中，噪声通常会影响边界细节，但不会显著改变体积，因此这种情况很常见，这两个定义往往会彼此近似。相同的推理也适用于软 Dice：也就是说，当噪声没有显著改变噪声标签的体积时，有 $\mathbb{E}[\mathrm{SD}_L(c)] \approx \mathrm{SD}_m(c), c \in \mathcal{M}$。

3.3. 问题描述

当 Dice 被用作目标评估指标时，图像分割问题可以看作是寻找一种能够使 Dice 最大化的分割，即

$$\sup_{s \in \mathcal{S}} \mathrm{D}_m(s), \quad m \in \mathcal{M}. \tag{3}$$

由于该问题具有离散性，通常无法直接通过优化来求解。相反，我们要在软分割空间中最小化一个损失函数，在此例中即软 Dice

$$\inf_{c\in\mathcal{M}} \mathrm{SD}_m(c), \quad m \in \mathcal{M}, \tag{4}$$

为避免不等式约束，使用 Sigmoid 函数 $\sigma(x) = 1/(1+\exp(-x))$

$$\inf_{f\in\mathcal{F}} \mathrm{SD}_m(\sigma \circ f), \quad m \in \mathcal{M}. \tag{5}$$

这是文献中最常见的软 Dice 形式。

找到解 $f \in \mathcal{F}$ 后，需要将其转换为二值分割，换言之，即一个元素 $s \in \mathcal{S}$。最常见的做法是对解应用 $1/2$ 阈值，也就是生成

$$s = I_{[1/2,1]} \circ \sigma \circ f. \tag{6}$$

对于交叉熵损失，优化器由边缘分布估计唯一确定，且该损失针对准确率经过校准，这些都是众所周知的事实 [23]。据我们所知，文献尚未研究将软 Dice 的最小化与 Dice 的最大化联系起来的相应结果，而推导此类结果正是本文的主要目标。

4. 主要结果

本节介绍主要结果。相关证明见补充材料。本研究的理论探讨有两个目标。首先，深入分析软 Dice 的最优解。其次，探讨将软 Dice 用作 Dice 最大化的代理指标是否可靠。为此，本文给出了若干理论结果。

定理 1. 对于任意 $m \in \mathcal{M}$，达到下确界 $\inf_{c\in\mathcal{M}} \mathrm{SD}_m(c)$ 的元素类 $\mathcal{M}_m^* \subset \mathcal{M}$ 由下式给出

$$c(\omega) \in \begin{cases} \{0\} & \text{if } m(\omega) < \sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2, \\ [0,1] & \text{if } m(\omega) = \sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2, \\ \{1\} & \text{if } m(\omega) > \sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2, \end{cases} \tag{7}$$

对于 $\omega \in \Omega$，几乎处处（λ-*a.e.*）

定理 1 刻画了使软 Dice 最小化的解。请注意，除了一些潜在的边界情况外，最优解集合中的元素在定义域几乎所有位置上都取值为 0 或 1。这意味着，使用式 (4) 中的 Sigmoid 函数时，理论上永远无法达到精确解。不过，可以任意逼近精确解。

定理 2。 对于任意 $m \in \mathcal{M}$，达到下确界 下确界 $c\in\mathcal{M}\mathrm{SD}_m(c)$ 的元素类 $\mathcal{M}_m^* \subset \mathcal{M}$ 满足以下界限

$$[\inf_{c\in\mathcal{M}_m^*} \|c\|_1, \sup_{c\in\mathcal{M}_m^*} \|c\|_1] \subseteq [\|m\|_1^2, 1]. \tag{8}$$

此外，这些界是紧的，具体而言，对于任意$v \in (0,1]$，存在 $m_0, m_1 \in \mathcal{M}$，使得 $\|m_0\|_1 = \|m_1\|_1 = v$ 并且

$$\inf_{c\in\mathcal{M}_{m_0}^*} \|c\|_1 = \|m_0\|_1^2, \quad \sup_{c\in\mathcal{M}_{m_1}^*} \|c\|_1 = 1. \tag{9}$$

定理 2 给出了软 Dice 所关联体积偏差的紧界。首先，这意味着给出了体积偏差的界。其次，这意味着在某些情况下可以达到极端情形，因此不存在更优的界。

定理 3. 对于任意 $m \in \mathcal{M}$、阈值 $a \in (0,1)$ 以及相对紧序列 $\{c_l\} \subset \mathcal{M}$，若

$$\lim_{l\to\infty} \mathrm{SD}_m(c_l) = \inf_{c\in\mathcal{M}} \mathrm{SD}_m(c) \tag{10}$$

then

$$\lim_{l\to\infty} \mathrm{D}_m(I_{[a,1]} \circ c_l) = \sup_{s\in\mathcal{S}} \mathrm{D}_m(s). \tag{11}$$

定理 3 给出了一个校准结果，建立了软 Dice 与 Dice 之间的联系。通俗地说，该结果表明，若软分割序列 $\{c_l\}_{l\geq 1}$ 收敛到最优软 Dice，那么当使用某个常数 $a \in (0,1)$ 作为阈值进行适当阈值化时，该序列也会收敛到最优 Dice。这一性质很重要，因为如果目标是最大化 Dice，那么它表明基于软 Dice 训练模型是有依据的。请注意，使用 Sigmoid 函数将区间限制在指定范围内时，结果同样成立。此外，只有在处理连续域时才需要相对紧性条件，因为与有限体素化相关联的任意函数子集始终是相对紧的。

5. 影响

这些证明部分基于 [23], 中提出的工作，该工作研究了在空间 $\mathcal{S}$ 上 Dice 的最优解分析，即 $\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)$。作者表明，该问题 $\mathcal{S}_m^* \subset \mathcal{S}$ 的最优分割类别由满足以下条件的 $s \in \mathcal{S}$ 给出：

$$s(\omega) \in \begin{cases} \{0\} & \text{if } m(\omega) < \sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2, \\ \{0,1\} & \text{if } m(\omega) = \sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2, \\ \{1\} & \text{if } m(\omega) > \sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2, \end{cases} \tag{12}$$

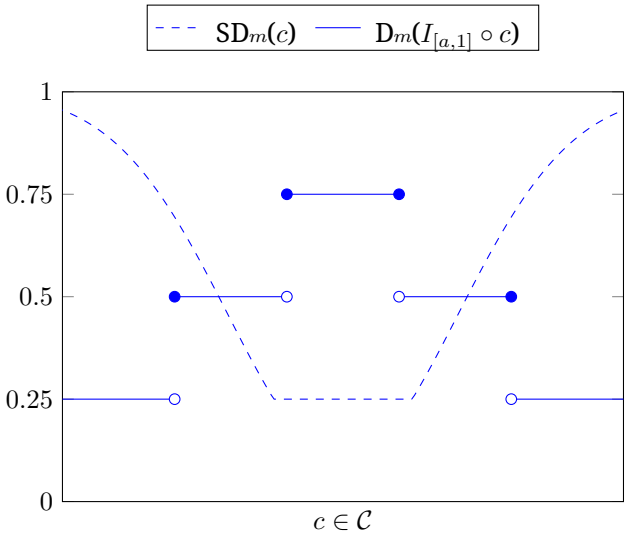

图 2。软 Dice（虚线）相对于某个软预测 $c \in \mathcal{M}$ 的示意图，以及 Dice（实线）相对于相应阈值化解 $I_{[a,1]} \circ c$的示意图，其中 $m \in \mathcal{M}$ 和 $a \in$ 满足 (0,1)。x 轴示意函数空间，并非一维空间。

对于 $\omega \in \Omega$， λ-几乎处处成立。请注意，最优解与定理 1 中提出的 $\mathcal{M}_m^*$ 几乎完全相同。唯一的区别是，边缘分布为 $m(\omega)$、满足 $=\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2$， $\omega \in \Omega$ 的定义域可以取整个区间 $[0,1]$ 内的值，而非二元值 $\{0,1\}$。

由此显然可知，若 $c \in \mathcal{M}_m^*$，则对于任意 $a \in$ (0,1)，都有$I_{[a,1]} \circ c \in \mathcal{S}_m^*$。或者说，软 Dice 优化器的阈值化解也是 Dice 的优化解，因此具有相应的性质。图 2 示意了软 Dice 与 Dice 之间的这种关系。

由此还可得出另一个重要相关性质：软 Dice 的阈值化最小化解的体积始终大于或等于交叉熵的阈值化最小化解的体积。由于$I_{[1/2,1]} \circ m$ 就是这样一个最大化解，因此可得

$$\inf_{s\in\mathcal{S}_m^*} \|s\|_1 \geq \|I_{[1/2,1]} \circ m\|_1. \tag{13}$$

这一性质从理论上解释了图 1 中的示意图，并意味着我们不可能遇到相反的情况：对交叉熵解进行 $1/2$ 阈值化后得到的极小化解，其体积大于对软 Dice 的极大化解进行 $1/2$ 阈值化后得到的体积。然而，在通过阈值化将软分割转换为硬分割之前，交叉熵的优化器始终具有体积 $\|m\|_1$；根据定理 2，软 Dice 的优化器在某些情况下具有体积 $\|m\|_1^2$。回想一下，由于 $\lambda(\Omega)$ $= 1$，因此可得 $\|m\|_1 \leq 1$，进而 $\|m\|_1^2 \leq \|m\|_1$。

6. 实验

为验证这些理论结果在实践中的相关性，研究人员开展了一系列实验。具体而言，他们考察了通过优化软 Dice 得到的结果是否

$$\inf_{f\in\mathcal{F}} \mathrm{SD}_m(\sigma \circ f), \tag{14}$$

在各种情况下采用简单的优化方案，得到的解都接近定理 1 所描述的解。由于图像域中边缘值恰好等于 $\sup_{s\in\mathcal{S}}\mathrm{D}_m(s)/2$ 的区域不太可能占据显著比例，换言之，

$$\lambda(\{\omega \in \Omega : m(\omega) = \sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2\}) > 0, \tag{15}$$

任务便简化为只需将数值解与下述单一分割进行比较

$$s(\omega) = I_{[\sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2,1]} \circ m(\omega), \quad \omega \in \Omega. \tag{16}$$

在实验中的数值案例里，这种分割很容易高效计算。

6.1. 设置

实验针对将图像域离散为 $N \geq 1$ 个等体积体素的情况进行。也就是说，Ω 被划分为 $\{$个$\Omega_i\}_{1\leq i\leq N}$，使得 Ω $=\Omega_1 \cup \cdots \cup \Omega_N$，其中 $\Omega_i \cap \Omega_j = \emptyset$，当 $i \neq j$ 时，且对所有 $1 \leq i$， $j \leq N$，都有 $\lambda(\Omega_i)$ $= \lambda(\Omega i)$。请注意，数据集中的不同图像可能具有不同数量的体素 N。对于特定的边缘函数 m，初始化元素 $f_1 \in \mathcal{F}$ 时，将每个体素的值设为独立抽取的各向同性高斯随机变量的值。随后，采用学习率固定为 $\gamma = 10N$ 的简单归一化梯度下降方案生成一系列更新。学习率会影响优化的收敛速度。不过，由于主要关注的是该方法收敛到的解，只要该方法在所运行的步数内能够较好地收敛，这就不成问题。因此，经过少量微调后选定了学习率。

为了衡量计算所得解与理论解的接近程度，我们考虑了两个度量指标。第一个度量指标定义为

$$e_{0,l} = \|\sigma \circ f_l - I_{[\sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2,1]} \circ m\|_1 \tag{17}$$

并测量软值与理论值之间的平均体素级距离。第二个度量指标定义为

$$e_{1,l} = \|I_{[1/2,1]} \circ \sigma \circ f_l - I_{[\sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2,1]} \circ m\|_1 \tag{18}$$

ROI	样本	$\bar{e}_{0,1}$	$\bar{e}_{1,1}$	$\bar{e}_{0,10}$	$\bar{e}_{1,10}$	$\bar{e}_{0,20}$	$\bar{e}_{1,20}$	$\bar{e}_{0,100}$	$\bar{e}_{1,100}$	$\bar{e}_{0,200}$	$\bar{e}_{1,200}$
(G) 膀胱	19	0.500	0.500	0.004	0.002	0.001	0.000	0.000	0.000	0.000	0.000
(G) 直肠	19	0.500	0.500	0.040	0.011	0.003	0.002	0.000	0.000	0.000	0.000
(G) 肛管	19	0.500	0.500	0.408	0.307	0.236	0.087	0.000	0.000	0.000	0.000
(G) 阴茎球	19	0.500	0.500	0.466	0.425	0.417	0.329	0.044	0.014	0.000	0.000
(G) 神经血管束。	19	0.500	0.500	0.390	0.286	0.210	0.095	0.000	0.000	0.000	0.000
(G) 股骨头 R	19	0.500	0.500	0.028	0.005	0.001	0.000	0.000	0.000	0.000	0.000
(G) 股骨头 L	19	0.500	0.500	0.031	0.006	0.001	0.000	0.000	0.000	0.000	0.000
(G) 前列腺	19	0.500	0.500	0.166	0.060	0.013	0.004	0.000	0.000	0.000	0.000
(G) 精囊	19	0.500	0.500	0.413	0.322	0.257	0.136	0.000	0.000	0.000	0.000
(S) $\rho = 0.01$	1000	0.500	0.500	0.022	0.008	0.001	0.001	0.000	0.000	0.000	0.000
(S) $\rho = 0.02$	1000	0.500	0.500	0.026	0.011	0.002	0.002	0.000	0.000	0.000	0.000
(S) $\rho = 0.03$	1000	0.500	0.500	0.027	0.015	0.004	0.003	0.001	0.000	0.000	0.000
(S) $\rho = 0.04$	1000	0.500	0.500	0.034	0.021	0.005	0.005	0.001	0.001	0.000	0.000
(S) $\rho = 0.05$	1000	0.500	0.500	0.039	0.025	0.007	0.006	0.001	0.001	0.000	0.000
(S) $\rho = 0.06$	1000	0.500	0.500	0.051	0.034	0.010	0.009	0.001	0.001	0.001	0.001
(S) $\rho = 0.07$	1000	0.500	0.500	0.057	0.040	0.013	0.011	0.002	0.001	0.001	0.001
(S) $\rho = 0.08$	1000	0.500	0.500	0.067	0.048	0.017	0.015	0.002	0.002	0.001	0.001
(S) $\rho = 0.09$	1000	0.500	0.500	0.080	0.057	0.022	0.019	0.003	0.003	0.001	0.001

表 1。关于 Gold Atlas 项目（G）中的骨盆数据和合成数据（S）的实验结果。各列展示了感兴趣区域的名称、用于计算的样本数量，以及相对于软标签的平均绝对差 $\bar{e}_{0,\,l}\, e_{0,\,l} = \|\sigma \circ f_l - I_{[\sup_{s' \in \mathcal{S}} \mathrm{D}m(s')/2,\ 1]} \circ m\|_1$ 以及相对于硬标签的平均绝对差 $\bar{e}_{1,\,l}\, e_{1,\,l} = \|I_{[1/2,1]} \circ \sigma \circ f_l - I_{[\sup_{s' \in \mathcal{S}} \mathrm{D}m(s')/2,\ 1]} \circ m\|_1$，对应的梯度下降迭代次数为 $l = 1$，10，20，100，200。

并测量相同的平均体素级距离，但使用经过 $1/2$ 阈值处理的值。这两个度量指标分别在运行 $l = 1$、10、20、100、200 次迭代后报告。有关实验和代码的详细信息请参见补充材料。

6.2. 数据

实验在两个不同的数据集上进行：一个是真实世界数据集（G），另一个是合成生成的数据集（S）。有关如何访问和处理数据的说明见补充材料。

(G)：第一个数据集包含盆腔区域中的九种结构，是 Gold Atlas 项目 [24]的一部分。该数据集包含 19 个病例，采用三维数据，切片分辨率为 512×512 像素。每种结构均由五位临床专家勾画，并通过对不同标签进行逐像素平均来形成边缘分布。在原始分辨率下，提取以每种结构为中心、大小为 128×128 的图像块。提取数据时使用了 *Plastimatch* [28] 软件。

(S)：第二个数据集是合成数据，通过三个步骤生成。首先，在包含 200×200 个像素的二维离散化区域上构造一个半径为0.2的球。其次，将这个球与具有方差 ρ 的高斯滤波器进行卷积，方差取一组不同的值。每个值对应一个特定的噪声水平，其中 ρ 越大，噪声越强。第三，将图像通过随机高斯场进行形变，以获得不同的形状。其原理是，ρ越大，图像中的噪声越多。

6.3. 结果

表 1 列出了实验结果。其中，对于每个感兴趣区域，均给出了不同迭代次数下的平均误差，误差由 $e_{0,\,l}$ 和 $e_{1,\,l}$ 衡量；迭代次数为$l = 1$、10、20、100、200。显然，优化方案平均而言会收敛到定理 1 中描述的优化器。迹象还表明，由 ρ 衡量的噪声量对收敛速度有一定影响。也就是说，较高的 ρ 似乎会使收敛速度稍慢。附录 A 中的图 3 展示了实验（G）中的不同样本，以及不同迭代次数下的 $\sigma \circ f_l$ 和理论最优解 s。同样，附录 B 中的图 4 展示了实验（S）中的不同样本，以及不同迭代次数下的 $\sigma \circ f_l$ l 和理论最优解 s。

7. 结论

本研究考察了在引入标签噪声并使用软标签的情况下，关于软 Dice 的最优解

。本文刻画了这些最优解，并给出了其体积偏差的紧确界。此外，研究表明软 Dice 对 Dice 是校准的，也就是说，任何收敛到最优软 Dice 的序列，在经过适当阈值化后，都会收敛到最优 Dice。最后，通过在一个合成数据集和来自 Gold Atlas 项目 [24] 的数据上进行实验，展示了这种刻画在实践中的意义。所有结果均支持理论结论。本文所呈现的结果有助于深入理解噪声如何影响医学图像分割中最常见的训练设置之一。

局限性：(1) 在数值环境中，可能无法精确构造定理 2 中用于证明紧性的必要情形。不过，数值误差的量级为 $1/N$，其中 N 是体素数量，而实际应用中的体素数量通常很大。(2) 这些实验针对的是无约束函数空间，而非受特定神经网络架构约束的函数空间。这样做是有意为之，目的是尽可能单独考察目标函数的行为。

附录 A. 实验样本 (G)

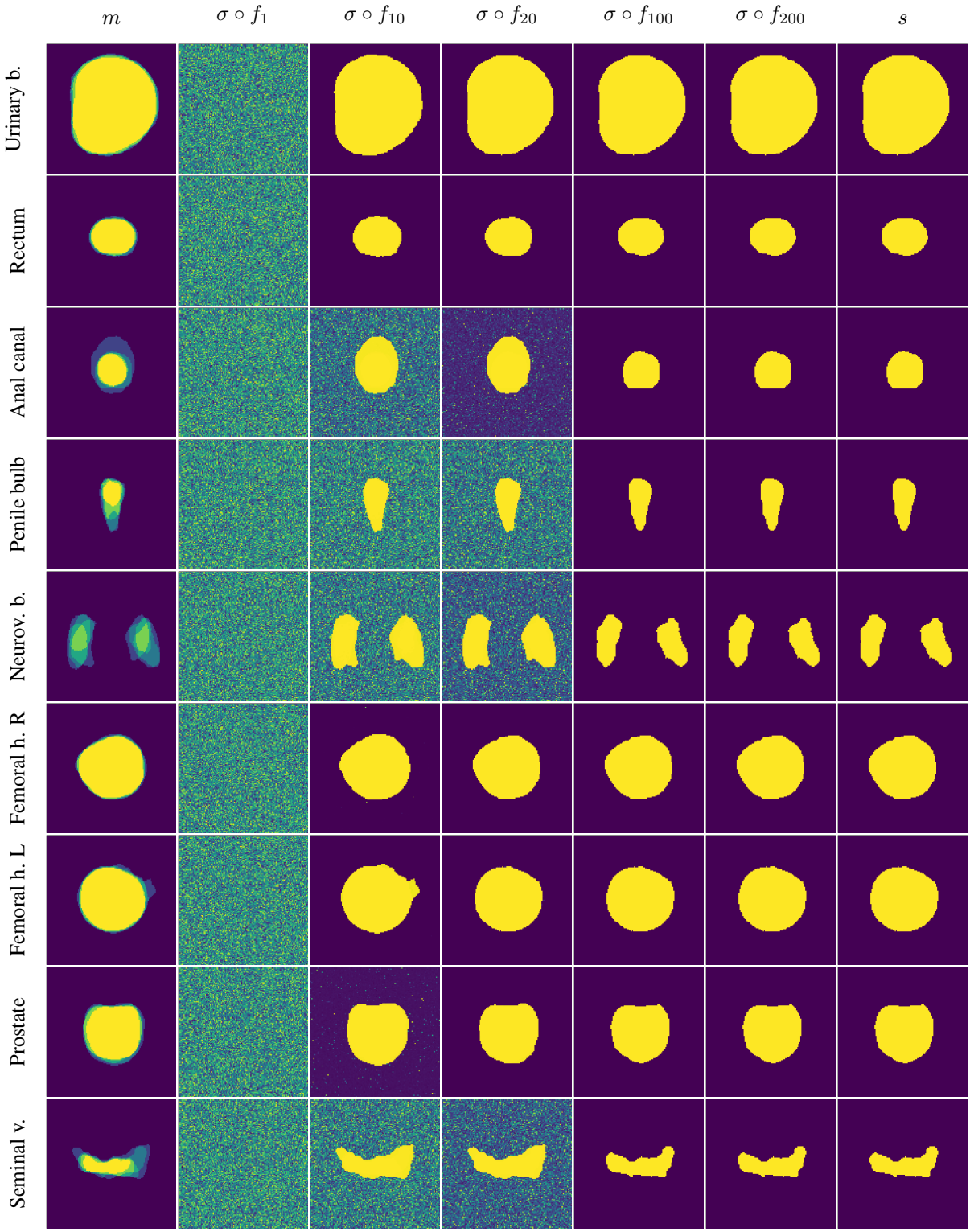

图 3. 每一行对应于 (G) 中某个实验示例所关联的一个二维切片。第一列显示一个边缘函数 m。第 2–6 列显示从软 Dice $\mathrm{SD}_m(\sigma \circ f)$ 的梯度下降最小化过程中若干次迭代的结果，以及一个随机起点。最后一列显示定理 1 中描述的理论最优解 $s = I_{[\sup_{s' \in \mathcal{S}} \mathrm{D}m(s')/2,\, 1]} \circ m$。

附录 B. 实验中的样本（S）

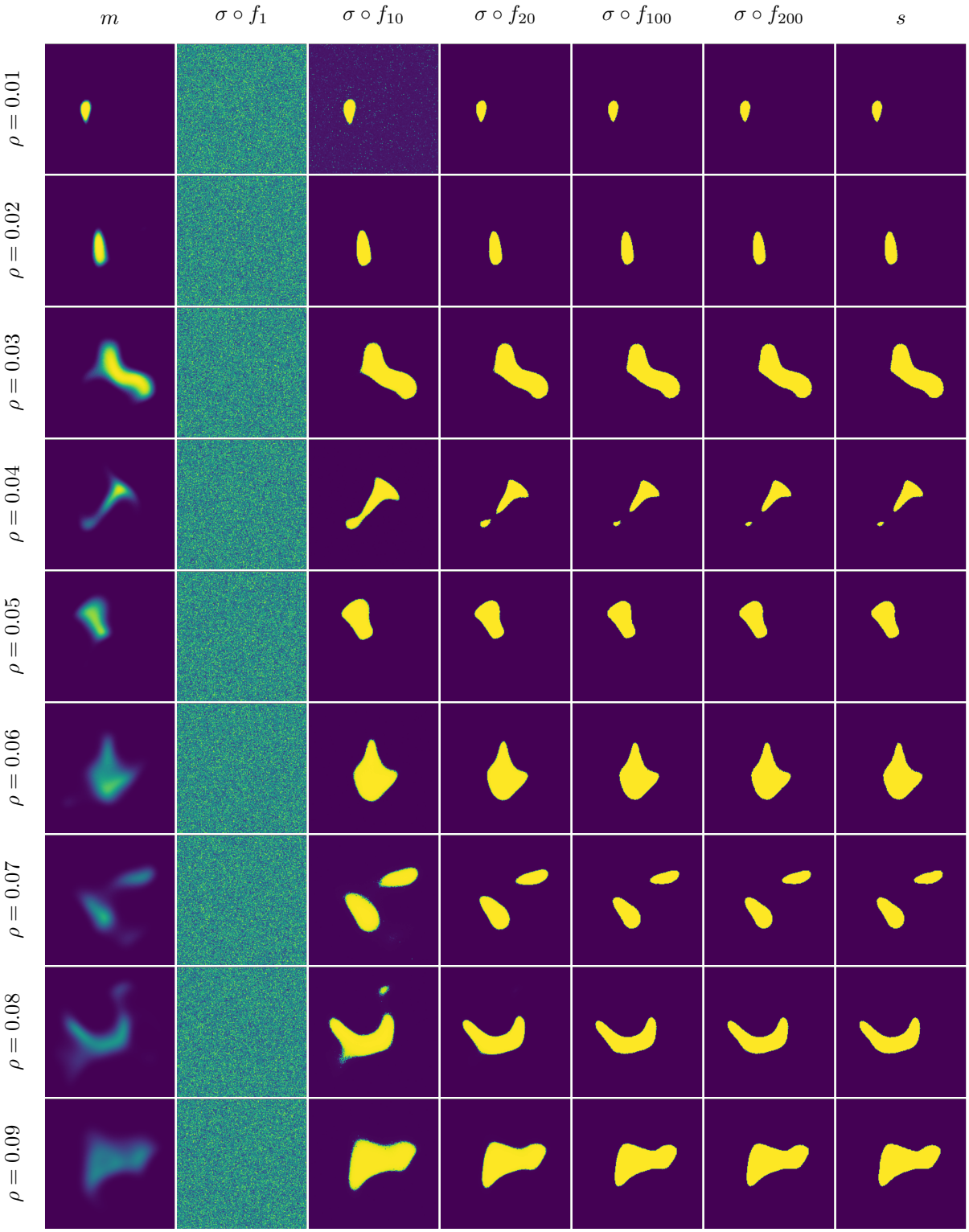

图 4。每一行对应（S）中某个实验的一个示例。第一列展示了一个边缘函数 m。第 2 至第 6 列展示了软 Dice $\mathrm{SD}_m(\sigma \circ f)$ 梯度下降最小化过程中的若干次迭代。最后一列展示了定理 1 中描述的理论最优解 $s = I_{[\sup_{s' \in \mathcal{S}} \mathrm{D}m(s')/2,\ 1]} \circ m$

参考文献

[1] Samuel G Armato III、Geoffrey McLennan、Luc Bidaut、Michael F McNitt-Gray、Charles R Meyer、Anthony P Reeves、Binsheng Zhao、Denise R Aberle、Claudia I Henschke、Eric A Hoffman 等。肺部图像数据库联盟（LIDC）与图像数据库资源倡议（IDRI）：一项关于CT扫描中肺结节的完整参考数据库。《医学物理》，38(2):915–931，2011年。[2] Han Bao 和 Masashi Sugiyama。二元分类中线性分式效用的校准替代最大化。载于国际人工智能与统计学会议，第2337–2347页。PMLR，2020年。[3] Peter L Bartlett、Michael I Jordan 和 Jon D McAuliffe。凸性、分类与风险界。《美国统计学会会刊》，101(473):138–156，2006年。[4] Jeroen Bertels、Tom Eelbode、Maxim Berman、Dirk Vandermeulen、Frederik Maes、Raf Bisschops 和 Matthew B Blaschko。医学图像分割中Dice系数与Jaccard指数的优化：理论与实践。载于医学图像计算与计算机辅助干预国际会议，第92–100页。施普林格，2019年。[5] Jeroen Bertels、David Robben、Dirk Vandermeulen 和 Paul Suetens。在固有不确定性背景下，对软Dice系数优化的分割图体积偏差进行理论分析与实验验证。医学图像分析，67:101833，2021年。[6] Pete Bridge、Andrew Fielding、Pamela Rowntree 和 Andrew Pullar。观察者内变异性：我们是否应该担心？医学影像与放射科学杂志，47：220，2016年。[7] Michal Drozdzal、Eugene Vorontsov、Gabriel Chartrand、Samuel Kadoury 和 Chris Pal。跳跃连接在生物医学图像分割中的重要性。载于医学应用中的深度学习与数据标注，第179–187页。施普林格，2016年。[8] Getao Du、Xu Cao、Jimin Liang、Xueli Chen 和 Yonghua Zhan。基于U-Net的医学图像分割：综述。《图像科学与技术杂志》，64:1–12，2020年。[9] Tom Eelbode、Jeroen Bertels、Maxim Berman、Dirk Vandermeulen、Frederik Maes、Raf Bisschops 和 Matthew B Blaschko。医学图像分割的优化：使用Dice系数或Jaccard指数进行评估时的理论与实践。*IEEE*医学影像汇刊，39(11):3679–3690，2020年。[10] Charley Gros、Andreanne Lemay 和 Julien Cohen-Adad。SoftSeg：图像分割中软标签训练相对于二元标签训练的优势。医学图像分析，71:102038，2021年。[11] Nicholas Heller、Joshua Dean 和 Nikolaos Papanikolopoulos。不完美分割标签：它们究竟有多重要？载于血管内成像与计算机辅助支架置入，以及生物医学数据大规模标注与专家标签合成，第112–120页。施普林格，2018年。[12] Fabian Isensee、Paul F Jaeger、Simon AA Kohl、Jens Petersen 和 Klaus H Maier-Hein。nnU-Net：一种用于基于深度学习的生物医学图像分割的自配置方法。《自然方法》，18(2):203–211，2021年。[13] Zeyu Jiang、Changxing Ding、Minfeng Liu 和 Dacheng Tao。《两阶段级联 U-Net：BraTS 挑战赛 2019 分割任务的第一名方案》。载于*MICCAI* 国际脑病变研讨会，第 231–241 页。施普林格，2019年。[14] Davood Karimi、Haoran Dou、Simon K Warfield 和 Ali Gholipour。《带噪声标签的深度学习：探索医学图像分析中的技术与补救方法》。《医 学图像分析》，65：101759，2020年。[15] Eytan Kats、Jacob Goldberger 和 Hayit Greenspan。《通过提炼解剖学知识进行软标签化，以改进多发性硬化病灶分割》。载于*2019*年*IEEE* 第*16*届国际生物医学成像研讨会（*ISBI 2019*），第 1563–1566 页。IEEE，2019年。[16] Andreanne Lemay、Charley Gros 和 Julien Cohen-Adad。《用于可靠表征评审者间不确定性的标签融合与训练方法》。《arXiv 预印本 arXiv:2202 07550》，2022年。。[17] Hang Li、Dong Wei、Shilei Cao、Kai Ma、Liansheng Wang 和 Yefeng Zheng。《超像素引导的标签软化在医学图像分割中的应用》。载于医学图像计算与计算机辅助干预国际会议，第 227–237 页。施普林格，2020年。[18] Zachary C Lipton、Charles Elkan 和 Balakrishnan Naryanaswamy。《分类器的最优阈值设定：最大化 F1 度量》。载于欧洲机器学习与数据库知识发现联合会议，第 225–239 页。施普林格，2014年。[19] Jun Liu、Xiangyue Wang 和 Xue-cheng Tai。《结合空间正则化、体积与星形先验的深度卷积神经网络用于图像分割》。《数学成像与视觉杂志》，第 1–21 页，2022年。[20] Jun Ma、Jianan Chen、Matthew Ng、Rui Huang、Yu Li、Chen Li、Xiaoping Yang 和 Anne L Martel。《医学图像分割中的损失函数探索》。《医学图像分析》，71：102035，2021年。[21] Guy Nir、Soheil Hor、Davood Karimi、Ladan Fazli、Brian F Skinnider、Peyman Tavassoli、Dmitry Turbin、Carlos F Villamil、Gang Wang、R Storey Wilson 等。《从多位专家学习：数字化组织病理图像中前列腺癌的自动分级》。《医学图像分析》，50：167–180，2018年。[22] Marcus Nordström、Han Bao、Fredrik Löfman、Henrik Hult、Atsuto Maki 和 Masashi Sugiyama。《Dice 系数的校准代理最大化》。载于医学图像计算与计算机辅助干预国际会议，第 269–278 页。施普林格，2020年。[23] Marcus Nordstrom、Henrik Hult、Fredrik Löfman 和 Jonas Söderberg。《带噪声标签的图像分割：准确率与 Dice 最优解的特征及体积属性》。《神经信息处理系统进展》，35：34321–34333，2022年。[24] Tufve Nyholm、Stina Svensson、Sebastian Andersson、Joakim Jonsson、Maja Sohlin、Christian Gustafsson、Elisabeth Kjellén、Karin Söderström、Per Albertsson、Lennart Blomqvist 等。《含盆腔区域器官多观察者勾画的磁共振成像与计算机断层扫描数据——Gold Atlas 项目的一部分》。《医学物理》，45(3)：1295–1300，2018年。

[25] Teodora Popordanoska、Jeroen Bertels、Dirk Vandermeulen、Frederik Maes 和 Matthew B Blaschko。《校准预测器与无偏体积估计之间的关系》。载于医学图像计算与计算机辅助干预国际会议，第 678–688 页。施普林格，2021年。[26] Olaf Ronneberger、Philipp Fischer 和 Thomas Brox。《U-Net：用于生物医学图像分割的卷积网络》。载于医学图像计算与计算机辅助干预国际会议，第 234–241 页。施普林格，2015年。[27] Axel-Jan Rousseau、Thijs Becker、Jeroen Bertels、Matthew B Blaschko 和 Dirk Valkenborg。《深度网络医学图像分割的训练后不确定性校准》。载于*2021*年 *IEEE* 第*18*届国际生物医学成像研讨会（*ISBI*），第 1052–1056 页。IEEE，2021年。[28] Gregory C Sharp、Rui Li、John Wolfgang、G Chen、Marta Peroni、Maria Francesca Spadea、Shinichro Mori、Junan Zhang、James Shackleford 和 Nagarajan Kandasamy。《Plastimatch：用于放射治疗图像处理的开源软件套件》。载于第*16*届放射治疗计算机应用国际会议（*ICCR*），阿姆斯特丹，荷兰，2010年。[29] Nahian Siddique、Sidike Paheding、Colin P Elkin 和 Vijay Devabhaktuni。《U-Net 及其变体在医学图像分割中的应用：理论与应用综述》。《IEEE Access》，9：82031–82057，2021年。[30] Joao Lourenço Silva 和 Arlindo L Oliveira。《使用软标签对医学图像分割中的不确定性建模》。《arXiv 预印本 arXiv:2109.12622》，2021年。[31] Carole H Sudre、Wenqi Li、Tom Vercauteren、Sebastien Ourselin 和 M Jorge Cardoso。《作为高度类别不平衡分割深度学习损失函数的广义 Dice 重叠度》。载于医学图像分析中的深度学习与临床决策支持的多模态学习，第 240–248 页。施普林格，2017年。[32] Nima Tajbakhsh、Laura Jeyaseelan、Qian Li、Jeffrey N Chiang、Zhihao Wu 和 Xiaowei Ding。《拥抱不完美数据集：医学图像分割深度学习解决方案综述》。《医学图像分析》，63：101693，2020年。[33] Eugene Vorontsov 和 Samuel Kadoury。《分割网络中的标签噪声：缓解方法必须应对偏差》。载于深度生成模型、数据增强、标注与瑕疵，第 251–258 页。施普林格，2021年。[34] Ming-Jie Zhao、Narayanan Edakunni、Adam Pocock 和 Gavin Brown。《超越 Fano 不等式：最优 F 分数、平衡错误率和代价敏感风险的界及其影响》。《机器学习研究期刊》，14(1)：1033–1090，2013年。

arXiv:2304.00801v3 [cs.CV] 4 May 2023

— 补充文件— 采用软 Dice 的噪声图像分割

Marcus Nordström *数学系 瑞典皇家理工学院 斯德哥尔摩, 瑞典marcno@kth.se

Henrik Hult 数学系 瑞典皇家理工学院 斯德哥尔摩，瑞典hult@kth.se

Atsuto Maki 计算机科学系 瑞典皇家理工学院 斯德哥尔摩, 瑞典atsuto@kth.se

Fredrik Löfman 机器学习系 RaySearch Laboratories 斯德哥尔摩，瑞典 fredrik.lofman@raysearchlabs.com

目录

1. 证明 **2**
1.1. 定理 1 的证明.2 1.2. 定理 2 的证明.4 1.3. 定理 3 的证明.5

2. 实验 **7**
2.1. 实验（G）.8 2.2. 实验（S）.8

*作者也隶属于 RaySearch Laboratories。

1. 证明

1.1. 定理 1 的证明

首先介绍 Dice 对软标签的一种推广

$$\mathrm{D}'_m(c) = \frac{2\int_\Omega c(\omega)m(\omega)\lambda(d\omega)}{\|c\|_1 + \|m\|_1}, \quad c \in \mathcal{M}, \tag{1}$$

并注意，对于任意 $c^* \in \mathcal{M}$，都有

$$\mathrm{SD}_m(c^*) = \inf_{c\in\mathcal{C}} \mathrm{SD}_m(c) \iff \mathrm{D}'_m(c^*) = \sup_{c\in\mathcal{C}} \mathrm{D}'_m(c). \tag{2}$$

现在，利用这一点，分四步证明该定理。首先，考虑 $\|m\|_1 = 0$ 的情形。其余情形均假设 $\|m\|_1 > 0$。其次，证明一个上界。第三，给出最优解的刻画。第四，将该刻画与集合 $\mathcal{M}^*_m$ 联系起来。全文中，λ_1 表示 $[0,1]$ 上的标准勒贝格测度（与 λ 类似，后者在 $\Omega = [0,1]^n$ 上以类似方式定义，其中 $n \geq 1$）。

第一部分，边界情况。对于 $\|m\|_1 = 0$ 的情形，任意 $c \in \mathcal{M}$ 都有 $\mathrm{D}'_m(c) = 0$，因此对任意 $c \in \mathcal{M}$，上确界均可取到。由于对于 $\omega \in \Omega$ λ 几乎处处都有 $m(\omega) = 0$，且 $\sup_{c\in\mathcal{M}} \mathrm{D}'_m(c) = 0$，可知对于 $\omega \in \Omega$ λ 几乎处处都有 $m(\omega) = \sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2$。因此，$\mathcal{M}^*_m = \mathcal{M}$ 当且仅当 $\|m\|_1 = 0$。

第 2 部分，存在性。首先注意，对于任意软分割 $c \in \mathcal{M}$，以下结论成立，

$$c = \int_{(0,1)} I_{[a,1]} \circ c \lambda_1(da), \quad \lambda - \text{a.e.} \tag{3}$$

此外，由于 $c(\omega) \geq 0, \ \omega \in \Omega$,

$$\|c\|_1 = \int_{(0,1)} \|I_{[a,1]} \circ c\|_1 \lambda_1(da). \tag{4}$$

结合 Dice 的定义可得

$$\begin{aligned}
\mathrm{D}'_m(c) &= \int_{(0,1)} \mathrm{D}_m(I_{[a,1]} \circ c) \frac{\|I_{[a,1]} \circ c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} \lambda_1(da) && (5)\\
&\leq \sup_{s\in\mathcal{S}} \mathrm{D}_m(s) \int_{(0,1)} \frac{\|I_{[a,1]} \circ c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} \lambda_1(da) && (6)\\
&= \sup_{s\in\mathcal{S}} \mathrm{D}_m(s) \frac{\|\int_{(0,1)} I_{[a,1]} \circ c \lambda_1(da)\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} && (7)\\
&= \sup_{s\in\mathcal{S}} \mathrm{D}_m(s) \frac{\|c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} && (8)\\
&= \sup_{s\in\mathcal{S}} \mathrm{D}_m(s)。 && (9)
\end{aligned}$$

因此,

$$\sup_{c\in\mathcal{M}} \mathrm{D}'_m(c) \leq \sup_{s\in\mathcal{S}} \mathrm{D}_m(s). \tag{10}$$

然而，由于 $\mathcal{S} \subset \mathcal{M}$,

$$\sup_{c\in\mathcal{M}} \mathrm{D}'_m(c) = \sup_{s\in\mathcal{S}} \mathrm{D}_m(s). \tag{11}$$

第 3 部分，表征。现在考虑集合

$$A_\epsilon = \{a \in (0,1) : \mathrm{D}_m(I_{[a,1]} \circ c) \geq \sup_{s\in\mathcal{S}} \mathrm{D}_m(s) - \epsilon\}. \tag{12}$$

并注意到有两种情况

$$(1)\exists \epsilon > 0 \implies \lambda_1(A_\epsilon^C) > 0, \tag{13}$$

$$(2)\forall \epsilon > 0 \implies \lambda_1(A_\epsilon^C) = 0. \tag{14}$$

对于第一种情况 (1)，则存在某个 $\epsilon > 0$

$$\mathrm{D}'_m(c) = \int_{(0,1)} \mathrm{D}_m(I_{[a,1]} \circ c) \frac{\|I_{[a,1]} \circ c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} \lambda_1(da) \tag{15}$$

$$\leq \sup_{s\in\mathcal{S}} \mathrm{D}_m(s) \int_{A_\epsilon} \frac{\|I_{[a,1]} \circ c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} \lambda_1(da) + (\sup_{s\in\mathcal{S}} \mathrm{D}_m(s) - \epsilon) \int_{A_\epsilon^C} \frac{\|I_{[a,1]} \circ c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} \lambda_1(da) \tag{16}$$

$$= \text{上确界}_{s\in\mathcal{S}} \mathrm{D}_m(s) - \epsilon \int_{A_\epsilon^C} \frac{\|I_{[a,1]} \circ c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} \lambda_1(da)。 \tag{17}$$

由于对每个 $a \in (0,\ 1)$ 和 $c \in \mathcal{M}$,

$$\frac{\|I_{[a,1]} \circ c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} \geq \frac{\|m\|_1}{1 + \|m\|_1}, \tag{18}$$

因此，前一个公式中的表达式有如下上界

$$\sup_{s\in\mathcal{S}} \mathrm{D}_m(s) - \epsilon\lambda_1(A_\epsilon^C) \frac{\|m\|_1}{1 + \|m\|_1} < \sup_{s\in\mathcal{S}} \mathrm{D}_m(s), \tag{19}$$

其中，最后一步由假设 $\|m\|_1 > 0$得出。我们得出，在情形（1）下

$$\sup_{c\in\mathcal{M}} \mathrm{D}'_m(c) < \sup_{s\in\mathcal{S}} \mathrm{D}_m(s). \tag{20}$$

对于第二种情形（2），

$$\mathrm{D}'_m(c) = \int_{(0,1)} \mathrm{D}_m(I_{[a,1]} \circ c) \frac{\|I_{[a,1]} \circ c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} \lambda_1(da) \tag{21}$$

$$= \sup_{s\in\mathcal{S}} \mathrm{D}_m(s) \int_{(0,1)} \frac{\|I_{[a,1]} \circ c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} \lambda_1(da) \tag{22}$$

$$= \sup_{s\in\mathcal{S}} \mathrm{D}_m(s) \frac{\int_{(0,1)} \|I_{[a,1]} \circ c\|_1 \lambda_1(da) + \|m\|_1}{\|c\|_1 + \|m\|_1} \tag{23}$$

$$= \sup_{s\in\mathcal{S}} \mathrm{D}_m(s) \frac{\|c\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} \tag{24}$$

$$= \text{上确界}_{s\in\mathcal{S}} \mathrm{D}_m(s)。 \tag{25}$$

第 4 部分，最终形式。 由定理 2 和定理 5 可知，在 [1], 中，使 $\mathcal{S}_m^* \subset \mathcal{S}$ 到 $\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)$ 的优化器由满足以下条件的元素给出

$$s(\omega) \in \begin{cases} \{0\} & \text{if } m(\omega) < \sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2, \\ \{0,1\} & \text{if } m(\omega) = \sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2, \\ \{1\} & \text{if } m(\omega) > \sup_{s'\in\mathcal{S}} \mathrm{D}_m(s')/2, \end{cases} \tag{26}$$

$\omega \in \Omega$， λ-几乎处处成立。若第二种情况 (2) 成立，则几乎每个 $a \in (0,\ 1)$ 都满足

$$\mathrm{D}_m(I_{[a,1]} \circ c) = \sup_{s \in \mathcal{S}} \mathrm{D}_m(s), \tag{27}$$

因此， $I_{[a,1]} \circ c \in \mathcal{S}_m^*$ 对于几乎所有的 $a \in (0,\ 1)$。现在，由于

$$c = \int_{(0,1)} I_{[a,1]} \circ c \lambda_1(da), \tag{28}$$

可知， $\mathcal{M}_m^* \subset \mathcal{M}$ 使 $\sup_{c \in \mathcal{M}} \mathrm{D}'_m(c)$ 达到最大值的优化器是满足以下条件的元素

$$c(\omega) \in \begin{cases} \{0\} & \text{if } m(\omega) < \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \\ [0,1] & \text{if } m(\omega) = \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \\ \{1\} & \text{if } m(\omega) > \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \end{cases} \tag{29}$$

$\omega \in \Omega$， λ-几乎处处成立。证明完毕。 □

1.2. 定理 2 的证明

本定理分两个简单步骤证明。首先，观察到它与先前结果的相似之处。其次，通过一个简单的观察，将其与本文给出的结果联系起来。

第 1 部分，硬分割。以下结论是定理 2 和定理 5 在 [1]中的推论。对于任意 $m \in \mathcal{M}$，达到 $\sup_{s \in \mathcal{S}} \mathrm{D}_m(s)$ 上确界的最大化元类 $\mathcal{S}_m^* \subset \mathcal{S}$ 满足以下界

$$[\inf_{s \in \mathcal{S}_m^*} \|s\|_1,\ \sup_{s \in \mathcal{S}_m^*} \|s\|_1] \subseteq [\|m\|_1^2, 1]. \tag{30}$$

此外，这些界是紧的，意即对于任意 $v \in (0,1]$， 存在 m_0， $m_1 \in \mathcal{M}$ 使得 $\|m_0\|_1 = \|m_1\|_1 = v$且

$$\inf_{s \in \mathcal{S}_{m_0}^*} \|s\|_1 = \|m_0\|_1^2, \quad \sup_{s \in \mathcal{S}_{m_1}^*} \|s\|_1 = 1. \tag{31}$$

此外， $\mathcal{S}_m^*$ 由这些元素给出。

$$s(\omega) \in \begin{cases} \{0\} & \text{if } m(\omega) < \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \\ \{0,1\} & \text{if } m(\omega) = \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \\ \{1\} & \text{if } m(\omega) > \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \end{cases} \tag{32}$$

$\omega \in \Omega$， λ-几乎处
处。

第 2 部分，软分割。令 $\mathcal{M}_m^* \subset \mathcal{M}$ 为使 $\inf_{c \in \mathcal{M}} \mathrm{SD}_m(c)$ 取得下确界的元素；如定理 1 所示，该下确界由下式给出

$$c(\omega) \in \begin{cases} \{0\} & \text{if } m(\omega) < \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \\ [0,1] & \text{if } m(\omega) = \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \\ \{1\} & \text{if } m(\omega) > \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \end{cases} \tag{33}$$

$\omega \in \Omega$， λ-几乎处处。现在注意，对任意 $m \in \mathcal{M}$,

$$\inf_{s \in \mathcal{S}_m^*} \|s\|_1 = \inf_{s \in \mathcal{M}_m^*} \|c\|_1, \tag{34}$$

在这两种情况下，上确界和下确界均可由以下对象取得:

$$s(\omega) = I_{(\sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, 1]} \circ m, \quad \omega \in \Omega. \tag{35}$$

类似地，对于任意 $m \in \mathcal{M}$,

$$\sup_{s \in \mathcal{S}_m^*} \|s\|_1 = \sup_{s \in \mathcal{M}_m^*} \|c\|_1. \tag{36}$$

并且，在这两种情况下，上确界均可取到：

$$s(\omega) = I_{[\sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, 1]} \circ m, \quad \omega \in \Omega. \tag{37}$$

因此，对于任意 $m \in \mathcal{M}$,

$$[\inf_{c \in \mathcal{M}_m^*} \|c\|_1, \sup_{c \in \mathcal{M}_m^*} \|c\|_1] = [\inf_{s \in \mathcal{S}_m^*} \|s\|_1, \sup_{s \in \mathcal{S}_m^*} \|s\|_1] \subseteq [\|m\|_1^2, 1]. \tag{38}$$

此外，这些界是紧的，也就是说，对于任意 $v \in (0, 1]$，都存在 m_0，$m_1 \in \mathcal{M}$ 使得 $\|m_0\|_1 = \|m_1\|_1 = v$并且

$$\inf_{c \in \mathcal{M}_{m_0}^*} \|c\|_1 = \inf_{s \in \mathcal{S}_{m_0}^*} \|s\|_1 = \|m_0\|_1^2, \quad \sup_{c \in \mathcal{M}_{m_1}^*} \|c\|_1 = \sup_{s \in \mathcal{S}_{m_1}^*} \|s\|_1 = 1. \tag{39}$$

证明至此完成。 □

1.3. 定理 3 的证明

首先，将 Dice 推广到软标签，定义如下

$$\mathrm{D}'_m(c) = \frac{2 \int_\Omega c(\omega) m(\omega) \lambda(d\omega)}{\|c\|_1 + \|m\|_1}, \quad c \in \mathcal{M}, \tag{40}$$

并注意，对于任意 $c^* \in \mathcal{M}$,

$$\lim_{l \to \infty} \mathrm{SD}_m(c_l) = \inf_{c \in \mathcal{M}} \mathrm{SD}_m(c) \iff \lim_{l \to \infty} \mathrm{D}'_m(c_l) = \sup_{c \in \mathcal{M}} \mathrm{D}'_m(c). \tag{41}$$

现在，利用这一点，定理分三步证明。首先，证明 D'_m 关于 $\mathcal{M}$ 的连续性。其次，证明软标签收敛到最优软 Dice 的一个推论。最后，利用这一推论证明，对同一序列进行适当阈值化后，它会收敛到最优 Dice。

第 1 部分，连续性。首先，注意如果 $\|m\|_1 = 0$，那么对于任意 c，$\mathrm{D}'_m(c) = 0$，这意味着 D'_m 是连续的。现在，假设 $\|m\|_1 > 0$，并令 c，$c_0 \in \mathcal{M}$ 为满足以下条件的任意两个元素

$$\|c - c_0\|_1 \leq \delta, \tag{42}$$

Then

$$|\mathrm{D}'_m(c_0) - \mathrm{D}'_m(c)| = \left| \mathrm{D}'_m(c_0) - \mathrm{D}'_m(c_0) \frac{\|c_0\|_1 + \|m\|_1}{\|c\|_1 + \|m\|_1} - \frac{2 \int_\Omega (c(\omega) - c_0(\omega)) m(\omega) \lambda(d\omega)}{\|c\|_1 + \|m\|_1} \right| \tag{43}$$

$$= \left| \mathrm{D}'_m(c_0) \frac{\|c\|_1 - \|c_0\|_1}{\|c\|_1 + \|m\|_1} - \frac{2 \int_\Omega (c(\omega) - c_0(\omega)) m(\omega) \lambda(d\omega)}{\|c\|_1 + \|m\|_1} \right| \tag{44}$$

$$\leq \mathrm{D}'_m(c_0) \frac{|\|c\|_1 - \|c_0\|_1|}{\|c\|_1 + \|m\|_1} + \frac{2 |\int_\omega (c(\omega) - c_0(\omega)) m(\omega) \lambda(d\omega)|}{\|c\|_1 + \|m\|_1} \tag{45}$$

$$\leq \frac{\|c - c_0\|_1}{\|c\|_1 + \|m\|_1} + \frac{2\|c - c_0\|_1}{\|s\|_1 + \|m\|_1} \tag{46}$$

$$\leq \frac{3\delta}{\|m\|_1} \tag{47}$$

因此，对于任意 $\epsilon > 0$,

$$\|c - c_0\|_1 \leq \delta = \frac{\|m\|_1}{3} \epsilon \tag{48}$$

意味着

$$|\mathrm{D}'_m(c_0) - \mathrm{D}'_m(c)| \leq \epsilon. \tag{49}$$

第 2 部分，收敛性　　软标签。在本部分中，我们证明，对于任意相对紧序列 $\{c_l\}_{l\geq 1} \subset \mathcal{M}$, n

$$\lim_{l\to\infty} \mathrm{D}'_m(c_l) = \sup_{c\in\mathcal{M}} \mathrm{D}'_m(c) \quad \Longrightarrow \quad \lim_{l\to\infty} \inf_{c^*\in\mathcal{M}^*_m} \|c_l - c^*\|_1 = 0. \tag{50}$$

为此，引入以下 ϵ-展开式： $\mathcal{M}^*_m$

$$\mathcal{M}^{*,\epsilon}_m = \{c \in \mathcal{M} : \inf_{c^*\in\mathcal{M}^*_m} \|c - c^*\|_1 \leq \epsilon\}. \tag{51}$$

由于相对紧集的子集的闭包是紧集，因此 closure($\{c_l\}_{l\geq 1} \setminus \mathcal{M}^{*,\ }_{\ m}{}^{\epsilon}$) 是紧集。由魏尔斯特拉斯极值定理可得

$$\exists c' \in \text{closure}(\{c_l\}_{l\geq 1} \setminus \mathcal{M}^{*,\epsilon}_m) \quad \Longrightarrow \quad \mathrm{D}'_m(c') = \sup_{c\in\{c_l\}_{l\geq 1}\setminus\mathcal{M}^{*,\epsilon}_m} \mathrm{D}'_m(c). \tag{52}$$

现在，注意到如果

$$\mathrm{D}'_m(c') = \sup_{c\in\mathcal{M}} \mathrm{D}'_m(c), \tag{53}$$

那么可得 $c' \in \mathcal{M}^*_m$ ，这与构造矛盾，因此

$$\sup_{c\in\{c_l\}_{l\geq 1}\setminus\mathcal{M}^{\epsilon,*}_m} \mathrm{D}'_m(c) < \sup_{c\in\mathcal{M}} \mathrm{D}'_m(c), \tag{54}$$

从而

$$\lim_{l\to\infty} \mathrm{D}'_m(c_l) = \sup_{c\in\mathcal{M}} \mathrm{D}'_m(c) \quad \Longrightarrow \quad \lim_{l\to\infty} \inf_{c^*\in\mathcal{M}^{*,\epsilon}_m} \|c_l - c^*\|_1 = 0. \tag{55}$$

但由于 $\epsilon > 0$ 是任意的，我们必有

$$\lim_{l\to\infty} \mathrm{D}'_m(c_l) = \sup_{c\in\mathcal{M}} \mathrm{D}'_m(c) \quad \Longrightarrow \quad \lim_{l\to\infty} \inf_{c^*\in\mathcal{M}^*_m} \|c_l - c^*\|_1 = 0. \tag{56}$$

第 3 部分。阈值化软标签的收敛。在本部分中，我们证明，对于任意序列 $\{c_l\}_{l\geq 1} \subset \mathcal{M}$ 和常数$a \in (0,1)$，均有

$$\lim_{l\to\infty} \inf_{c^*\in\mathcal{M}^*_m} \|c_l - c^*\|_1 = 0 \implies \lim_{l\to\infty} \mathrm{D}_m(I_{[a,1]} \circ c_l) = \sup_{s\in\mathcal{S}} \mathrm{D}_m(s). \tag{57}$$

为此，任取 $a \in (0,1)$和 $\epsilon > 0$。那么存在某个 $L \geq 1$ ，使得对于 $l \geq L$

$$\inf_{c^*\in\mathcal{M}^*_m} \|c_l - c^*\|_1 \leq \epsilon. \tag{58}$$

此外，由于

$$\inf_{c^*\in\mathcal{M}^*_m} \|c_l - c^*\|_1 = \tag{59}$$

$$\int_\Omega I_{[0,\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2)}(m(\omega))|c_l(\omega) - 0|\lambda(d\omega) + \int_\Omega I_{(\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2,1]}(m(\omega))|c_l(\omega) - 1|\lambda(d\omega), \tag{60}$$

由此可得

$$\int_\Omega I_{[0,\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2)}(m(\omega))|c_l(\omega) - 0|\lambda(d\omega) \leq \epsilon, \tag{61}$$

$$\int_\Omega I_{(\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2,1]}(m(\omega))|c_l(\omega) - 1|\lambda(d\omega) \leq \epsilon. \tag{62}$$

这进一步意味着

$$\int_{\Omega} I_{[0,\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2)}(m(\omega))|I_{[a,1]}(c_l(\omega)) - 0|\lambda(d\omega) \leq \epsilon/\min\{a, 1-a\}, \tag{63}$$

$$\int_{\Omega} I_{(\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2,1]}(m(\omega))|I_{[a,1]}(c_l(\omega)) - 1|\lambda(d\omega) \leq \epsilon/\min\{a, 1-a\}, \tag{64}$$

因此，由于

$$\inf_{s^*\in\mathcal{S}_m^*} \|I_{[a,1]} \circ c_l - s^*\|_1 = \tag{65}$$

$$\int_{\Omega} I_{[0,\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2)}(m(\omega))|I_{[a,1]}(c_l(\omega)) - 0|\lambda(d\omega) + \int_{\Omega} I_{(\sup_{s\in\mathcal{S}} \mathrm{D}_m(s)/2,1]}(m(\omega))|I_{[a,1]}(c_l(\omega)) - 1|\lambda(d\omega) \tag{66}$$

也就是说

$$\inf_{s^*\in\mathcal{S}_m^*} \|I_{[a,1]} \circ c_l - s^*\|_1 \leq 2\epsilon/\min\{a, 1-a\}. \tag{67}$$

由于 $\epsilon > 0$ 是任意的，这意味着

$$\lim_{l\to\infty} \inf_{s^*\in\mathcal{S}_m^*} \|I_{[a,1]} \circ c_l - s^*\|_1 = 0. \tag{68}$$

根据 D_m 的连续性，我们最终得到

$$\lim_{l\to\infty} \mathrm{D}_m(I_{[a,1]} \circ c_l) = \sup_{s\in\mathcal{S}} \mathrm{D}_m(s). \tag{69}$$

证明至此完成。 □

2. 实验

所有实验都只是计算梯度下降方案中的各个步骤。我们关注的是所考虑的边缘函数（软标签） $m \in \mathcal{M}$、通过梯度下降方案生成的序列 $\{f_l\}_{l\geq 1} \subset \mathcal{F}$，以及理论最优分割$s = I_{[\sup_{s'\in\mathcal{S}} \mathrm{D}m(s')/2,1]} \circ m \in \mathcal{S}_{m\circ}^*$ 计算时，将连续域离散化为一组体素。设 N为体素数量，这些体素由划分 $\{\Omega_i\}_{i=1}^N$ 给出，并满足 $\Omega = \Omega_1 \cup \cdots \cup \Omega_n$，其中 $\Omega_i \cap \Omega_j = \emptyset$ 当 $i \neq j$ 时成立，且对于所有 $1 \leq i \leq N$，$(\Omega_i) = 1/N$。随后，通过将函数限制为在每个体素上保持常数，在该体素化网格上进行计算

$$m(\omega) \doteq M(i), \quad \omega \in \Omega_i, \tag{70}$$

$$f_l(\omega) \doteq F_l(i), \quad \omega \in \Omega_i, \tag{71}$$

$$s(\omega) \doteq S(i), \quad \omega \in \Omega_i, \tag{72}$$

对于 $i = 1, \ldots, N$和 $l = 1, 2, \ldots$。开始时， M以离散形式给出，而 $\{F_l\}_{l\geq 1}$和 S则通过计算得到。更确切地说，首先从标准正态分布中独立抽取样本，并将其赋给 F_1 中的元素，从而生成 $\{F_l\}_{l\geq 1}$

$$F_1(i) \sim \mathrm{normal}(0, 1), \tag{73}$$

然后依次计算其余部分

$$F_{l+1}(i) = F_l(i) - \gamma \frac{d}{dF_l(i)} \left[\frac{2\frac{1}{N}\sum_{j=1}^N \sigma(F_l(j))M(j)}{\frac{1}{N}\sum_{j=1}^N M(j) + \frac{1}{N}\sum_{j=1}^N \sigma(F_l(i))} \right] \tag{74}$$

$$= F_l(i) - \gamma \frac{2M(i) - 2\frac{1}{N}\sum_{j=1}^N \sigma(F_l(j))M(j)}{(\sum_{j=1}^N M(j) + \sum_{j=1}^N \sigma(F_l(i)))^2} \sigma(F_l(i))\sigma(-F_l(i)). \tag{75}$$

通过以下方式计算理论上的最优解

$$S(i) = I\left\{M(i) \geq \max_{t\in[0,1]} \frac{2\frac{1}{N}\sum_{j=1}^{N} I\{M(j)\geq t\}M(j)}{\frac{1}{N}\sum_{j=1}^{N} M(j) + \frac{1}{N}\sum_{j=1}^{N} I\{M(j)\geq t\}} \frac{1}{2}\right\} \tag{76}$$

$$= I\left\{M(i) \geq \max_{t\in\{M(1),\ldots,M(N)\}} \frac{\sum_{j=1}^{N} I\{M(j)\geq t\}M(j)}{\sum_{j=1}^{N} M(j) + \sum_{j=1}^{N} I\{M(j)\geq t\}}\right\}, \tag{77}$$

以及论文中报告的最终结果是

$$e_{0,l} = \frac{1}{N}\sum_{j=1}^{N} |\sigma(F_l(j)) - S(j)|, \tag{78}$$

$$e_{1,l} = \frac{1}{N}\sum_{j=1}^{N} |I\{\sigma(F_l(j)) \geq 1/2\} - S(j)|. \tag{79}$$

实验使用了三个不同的数据集，下面将对其进行介绍。

2.1. 实验（G）

实验（G）是在三维空间中进行的，基于 Gold Atlas 项目的骨盆数据 [2]。具体而言，它包含九个不同的感兴趣区域：膀胱、直肠、肛管、阴茎球、神经血管束、右侧股骨头、左侧股骨头、前列腺和精囊。每个感兴趣区域均由 5 位不同的医学从业者勾画。实验中使用的软标签，是针对每个病例和感兴趣区域，对不同从业者所做的分割进行体素级平均后得到的。

如需获取原始数据，请访问以下链接并申请访问权限。

`https://doi.org/10.5281/zenodo.583096`

获得权限后，下载所有文件并将其放入文件夹 Experiments_G/dicom 中。还要确保将 `3_03_P(1).zip` 重命名为 `3_03_P.zip`。完成后，文件夹中应包含以下文件：`1_01_P.zip`，`1_02_P.zip`，`1_03_P.zip`，`1_04_P.zip`，`1_05_P.zip`，`1_06_P.zip`，`1_07_P.zip`，`1_08_P.zip`，`2_03_P.zip`，`2_04_P.zip`，`2_05_P.zip`，`2_06_P.zip`，`2_09_P.zip`，`2_10_P.zip`，`2_11_P.zip`，`3_01_P.zip`，`3_02_P.zip`，`3_03_P.zip`，`3_04_P.zip`。

为提取数据，使用版本为 1.9.3 的*Plastimatch* [3] 软件。安装程序和许可证相关信息可通过以下链接获取。

`http://plastimatch.org/`

随附代码和 Python 依赖文件。为保护患者隐私，参与者均签署了知情同意书。据我们所知，数据并未附带明确的许可证；不过，原始文献指出，数据可免费用于非商业应用。实验基于中央处理器，在装有 Python 3.10.4 的 Windows 10 计算机上进行。

2.2. 实验（S）

实验（S）基于合成数据。生成这类合成数据的目的是构造边缘函数，使其类似于我们预期在实践中出现的真实边缘函数。请注意，实验（G）使用的是真实世界数据；不过，我们采用有限样本近似，即计算样本的体素级平均值。这种近似受样本数量限制；如果我们能够获得无限多个样本，预期得到的边缘函数会更加平滑，而不是有限样本所产生的阶梯状边缘分布。我们生成的合成数据在这个意义上是平滑的。其计算分为以下三个步骤。

步骤1。 在二维空间中生成一个球 m_0 。

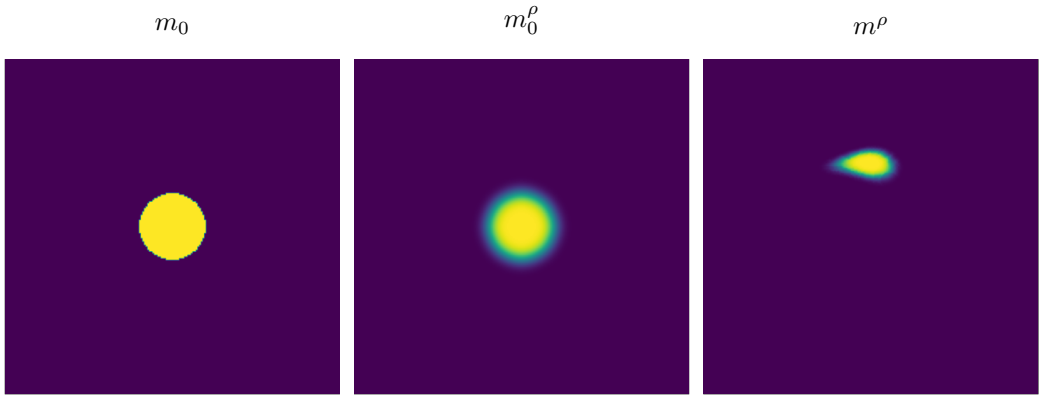

图 1. 实验（S）中用于生成数据的三个步骤示意图。

步骤2。 通过将 m_0^ρ 与协方差为零、方差为 ρ^2的各向同性高斯分布进行卷积，生成函数 m_0^ρ 。

步骤3. 最后，通过连续高斯场随机变形 m_0^ρ，生成 m^ρ 。

随附代码和 Python 依赖项文件。由于数据是合成数据，因此不存在隐私或许可方面的顾虑。该实验基于中央处理器，在一台安装了 Python 3.10.4 的 Windows 10 机器上进行。

参考文献

[1] Marcus Nordstrom、Henrik Hult、Fredrik Löfman 和 Jonas Söderberg。《噪声标签下的图像分割：准确率和 Dice 最优解的特征及体积属性》。《神经信息处理系统进展》，35 34333，2022 34321：–。[2] Tufve Nyholm、Stina Svensson、Sebastian Andersson、Joakim Jonsson、Maja Sohlin、Christian Gustafsson、Elisabeth Kjellén、Karin Söderström、Per Albertsson、Lennart Blomqvist 等。《磁共振成像和计算机断层扫描数据及盆腔区域器官的多观察者勾画——Gold Atlas 项目的一部分》。《医学物理》，45(3):1295–1300，2018年。[3] Gregory C Sharp、Rui Li、John Wolfgang、G Chen、Marta Peroni、Maria Francesca Spadea、Shinichro Mori、Junan Zhang、James Shackleford 和 Nagarajan Kandasamy。Plastimatch：用于放射治疗图像处理的开源软件套件。载于第*16*届放射治疗计算机应用国际会议（*ICCR*），阿姆斯特丹，荷兰，2010年。