噪声图像分割中的边缘阈值化

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

arXiv:2304.04116v3 [cs.CV] 8 Jul 2023

摘要

本文通过考虑一种基于高斯场形变的噪声模型，研究了医学图像分割中的标签噪声。该噪声之所以受到关注，一方面是因为它能够产生逼真的分割结果，另一方面是因为它是无偏的，即其期望形变为恒等映射。本文提供了高效的采样方法以及边缘概率的闭式解。此外，研究了交叉熵和软 *Dice* 损失函数的理论最优解，并展示了随着噪声水平增加，这两者如何逐渐分化。基于近期关于损失函数特性的研究，本文表明，可以通过对交叉熵的解采用某个特定的、先验未知但可高效计算的阈值进行阈值化，恢复软 *Dice* 的最优解。这引出了一个问题：与软 *Dice* 相比，使用交叉熵时观察到的性能下降，是否是由采用了错误的阈值所导致的。研究在来自 *TotalSegmentor* 数据集的三个器官分割问题上开展了 *5* 折实验，并使用了 *4* 种不同强度的噪声，从而验证了这一假设。结果表明，改变阈值可以使交叉熵的性能从系统性地劣于软 *Dice*，转变为与软 *Dice* 相当甚至更优。

1. 引言

在医学图像分析中，人们通常希望将图像划分为前景和背景，同时到某个感兴趣区域。这个问题被称为医学图像分割问题。自从 U-net [26], 问世以来，用于解决该问题的最先进方法主要由按照监督学习范式训练的神经网络方法构成 [6,28]。这些方法在数据丰富且干净的场景中极为有效，有些人甚至认为，在这种情况下，分割问题已经得到解决。然而，准确且可靠地手动勾画感兴趣区域可能是一个复杂且耗时的过程，通常需要医学专业知识。这导致获取用于训练的高质量数据集需要付出高昂成本，因此，实践中使用的数据往往并不完美，因为人工提供的标签受到了噪声污染 [1,5,20,24]。

大量研究致力于使用不完美数据进行医学图像分割，其中探讨的一个重要问题是各种分割方法对标签噪声的鲁棒性。通常，实验先获取一些数据集，再用各种类型的噪声污染训练数据，最后在生成的数据集上训练不同模型。随后，使用某种评估指标评估不同模型的结果，并与未添加任何噪声训练的基线模型进行比较。已研究的简单噪声模型包括随机切除部分结构 [33] ，以及以相互独立的方式翻转标签或标签簇 [9]。一种已研究的更真实的噪声模型基于随机形变，其中形变由低维高斯数组插值得到。这些高斯形变模型是无偏的，也就是说，期望形变为恒等形变。实验上已证实，机器学习模型对具有这一特性的噪声比对不具备这一特性的噪声更具鲁棒性

*作者同时隶属于 RaySearch Laboratories。

形变模型是无偏的，也就是说，期望形变为恒等形变。实验上已证实，机器学习模型对具有这一特性的噪声比对不具备这一特性的噪声更具鲁棒性 [31]。

标签噪声如何影响分割模型的性能，取决于用于训练模型的损失函数。可以说，最经典的损失函数是交叉熵。这种损失函数具有许多良好的理论性质，并且在实践中表现稳定。由于医学图像分割存在类别不平衡问题，通常前景区域明显小于背景区域，同时 Dice 又被广泛用作评估度量指标，因此后来提出了一种基于平滑版 Dice 度量指标的替代损失函数。这种损失被称为软 Dice 或 Dice 损失，并且已有研究表明，在多种实验设置下，它的性能优于交叉熵，但有时会引入一些稳定性问题 [2, 21]。因此，为了在不引入过多潜在稳定性问题的情况下取得一定的性能提升，通常会将多个损失函数进行线性组合。例如，广受欢迎的 NNUnet 所使用的就是这一类型的损失函数 [11]。

本文深入研究了一种基于高斯形变的噪声模型。具体而言，本文对该模型进行了形式化，并推导出了用于模拟的高效方法以及边际概率的闭式表达式。针对该噪声模型，本文进一步研究了交叉熵和软 Dice 的最优解，并表明：当标签无噪时，软 Dice 与交叉熵得到相同类别的最优解；但随着噪声增加，二者会出现分歧。本文还表明，软 Dice 的最优解可以通过对交叉熵的解使用某个先验未知但可高效计算的阈值进行阈值化来恢复。这引出了一个问题：采用替代阈值的交叉熵能否达到软 Dice 的性能？实验结果表明，确实如此。

贡献： 本文对一系列基于高斯场形变的噪声模型进行了形式化。推导出了一种用于采样噪声分割结果的高效方法，以及边际概率的闭式表达式。利用该噪声模型可以表明，随着噪声增加，交叉熵和软 Dice 的最优解会出现分歧；但软 Dice 的最优解可以通过使用可计算的阈值对交叉熵估计进行阈值化来恢复。随后，本文在 TotalSegmentor 数据集中的三种器官上进行了实验验证 [32]。

2. 相关工作

医学图像分割数据中的标签噪声可能是标注者绘制分割结果时受到时间限制所致，也可能源于对边界位置的不同意见。标签边界应该是.....或其他东西。众所周知，用于模型训练的数据中通常存在标签噪声，这一点已在例如 [1, 5, 20, 24]中被指出。已经提出了许多处理标签噪声的方法。示例包括主动学习方法，在这些方法中，专家会被反复询问以获取更多信息 [8]，以及重新加权方案，在这些方案中，方法会根据确定性动态学习为不同区域赋予不同权重 [19]。关于较新的综述，请参阅 [12, 30]。

标签噪声对各种损失函数的影响已在多篇论文中得到研究。这包括其对校准和使用交叉熵训练的模型所生成预测的体积偏差的影响，以及用于提升软Dice [3], 性能的重新校准方法和校准与体积偏差 [25]之间的关系。其他实验工作表明，交叉熵的期望校准误差低于软Dice，其中包括 [18]。在上述研究中，噪声是通过考察相对于某种含噪分割的期望得分来分析的。纳入噪声的另一种方式是使用软标签。使用软标签与软 Dice的工作示例包括 [7, 13]。使用软标签与其他损失函数的工作示例包括 [15, 16, 29]。

受 [17, 34], 启发， [22] 分析了标签噪声下最优分割的准确率和 Dice。文中给出了最优分割的刻画，展示了最优分割在体积偏差方面的尖锐界限，并比较了与各度量指标相关的优化器体积。随后，针对损失函数软 Dice [23]，也进行了类似的分析。具体而言，文中给出了最优解及其对应体积的刻画。此外，校准结果表明，收敛到最优软 Dice 的序列，在经过适当阈值化后，会收敛到最优 Dice。

医学图像分割对噪声的鲁棒性已在多项研究中得到探讨。在 [33], 中，研究者通过随机去除训练数据中的子结构来研究鲁棒性。高斯场形变产生的分割噪声对本研究尤为重要。已有两项研究在实验环境下研究了此类噪声，包括 [9] 和 [31]。在这些研究中，训练数据中的分割结果根据插值到像素空间的低维高斯场进行扰动。随后，研究者针对不同的噪声水平训练模型，并报告所得结果。最后， [14], 讨论了放射治疗中靶区勾画的噪声模型。该模型基于对高斯场取水平集，并提供了高效的采样方法。

3. 预备知识

3.1. 符号说明

令 $\Omega = [0,1]^n \subset \mathbb{R}^n$ 为维数 $n \geq 1$ 的单位立方体，λ 为与 $\mathbb{R}^n$ 相关联的标准勒贝格测度，使得 $\lambda(\Omega) = 1$。令 $\mathcal{S}$ 为从 Ω 到 $\{0,1\}$ 的可测函数空间，$\mathcal{M}$ 为从 Ω 到 $[0,1]$ 的可测函数空间，$\mathcal{F}$ 为从 Ω 到 $\mathbb{R}$ 的有界可测函数空间，其中所有空间 $\mathcal{S}$、$\mathcal{M}$ 和 $\mathcal{F}$ 均配备相应的博雷尔 σ-域 $\bar{}$ 及 L_p-范数 $\|\cdot\|_p$。简言之，对于任意 $f \in \mathcal{F}$，记号 $f(\omega) = 1 - f(\omega)$ $\omega \in \Omega$ 将用于表示相应对象。字母 $s \in \mathcal{S}$ 通常表示分割，其中 $s(\omega) = 1$ 表示目标结构的前景，$s(\omega) = 0$ 表示背景。字母 $m \in \mathcal{M}$ 通常表示边缘函数，即 $m(\omega) \in [0,1]$ 表示噪声标签占据位置 $\omega \in \Omega$ 的概率。字母 $c \in \mathcal{M}$ 通常表示通过优化所选损失函数、尚未进行任何阈值处理而得到的软分割。字母 $f \in \mathcal{F}$ 通常表示与软分割相关联的 logit 函数，使得对于 $\omega \in \Omega$ 上几乎处处成立 $\sigma \circ f(\omega) = c(\omega)$，其中 $\sigma(x) = 1/(1+e^{-x})$ 为标准 sigmoid 函数。函数的复合记作 $\circ$ 和 $I_A(x)$ 用于表示集合 A 上的指示函数，即 $I_A(x) = 1$ 当 $x \in A$，且 $I_A(x) = 0$当 $x \notin A$。指示函数与某个元素 $c \in \mathcal{M}$ 结合使用，表示对软分割进行阈值化。也就是说，$I_{[t,1]} \circ c$ 对所有满足$c(\omega) \geq t$ 的 $\omega \in \Omega$ 赋值为 1，并对所有满足 $c(\omega) < t$ 的 $\omega \in \Omega$ 赋值为 0。

3.2. 损失函数和评估指标

医学图像分割领域中最常用的两种损失函数是交叉熵和软 Dice。存在噪声时，标签会成为一个随机变量，因此需要将损失扩展为定义在一个确定性软分割（预测）和一个随机分割（标签）上的泛函。本文在理论分析中采用了这种扩展的软标签化约定 [7, 13, 15, 16, 29]。具体而言，损失函数是相对于边缘函数 $m \in \mathcal{M}$ 定义的，该函数对应于取值在 $\mathcal{S}$ 中的含噪分割 L 的均值，即 $m(\omega) = L(\omega)$ $\omega \in \Omega$。

定义 1。对于任意 $m \in \mathcal{M}$，（二元）交叉熵定义为

$$\mathrm{CE}_m(c) \doteq -\int_\Omega [m(\omega)\log(c(\omega)) + \bar{m}(\omega)\log(\bar{c}(\omega))]\lambda(d\omega), \quad c \in \mathcal{M}. \tag{1}$$

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

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

可以说，最经典的评估指标是准确率。该指标只衡量被正确标注的区域占整个域的比例。另一方面，最常用的评估指标是Sørensen-Dice系数，简称Dice。为处理标签噪声，这些度量指标采用与损失函数相同的方式进行推广，即使用软标签。

定义 3. 对于任意 $m \in \mathcal{M}$，其准确率为

$$\mathrm{A}_m(s) \doteq \int_\Omega [s(\omega)m(\omega) + \bar{s}(\omega)\bar{m}(\omega)]\lambda(d\omega), s \in \mathcal{S} \tag{3}$$

定义 4. 对于任意 $m \in \mathcal{M}$，*Dice* 定义为

$$\mathrm{D}_m(s) \doteq \frac{2\int_\Omega s(\omega)m(\omega)\lambda(d\omega)}{\|s\|_1 + \|m\|_1}, \quad s \in \mathcal{S}. \tag{4}$$

进行实验时，每张图像仅有一个标签，因此更准确地说，应将该设置视为期望得分的经验近似。对于交叉熵和准确率，使用期望得分的定义与使用软标签的定义相同，即$\mathbb{E}[\mathrm{A}_L(s)] = \mathrm{A}_m(s)$以及 $\mathbb{E}[\mathrm{CE}_L(s)] = \mathrm{CE}_m(s)$。对于软-Dice和Dice，只要噪声标签的体积保持不变，这些定义就相同，即 $\mathrm{Var}[\|L\|_1] = 0$。如果这一条件近似成立——这在医学图像分割中通常是这种情况，因为噪声往往会改变边界的细节，而不会显著改变分割的体积——那么可得 $\mathbb{E}[\mathrm{D}_L(s)] \approx \mathrm{D}_m(s)$，同样可得 $\mathbb{E}[\mathrm{SD}_L(s)] \approx \mathrm{SD}_m(s)$。

3.3. 高斯场

在概率论和统计学中，最核心的对象包括正态分布随机变量，也称为高斯随机变量。这类随机变量广泛存在于科学和工程领域；由于其丰富多样且具有有趣的性质，人们对它们的各种推广形式进行了大量研究。经典高斯变量的一种简单推广是将其扩展到多维情形，此时称之为多元高斯变量。对于这种中心化且各向同性的变量，其概率密度函数如下

$$p_{\sigma^2}(\omega) \doteq \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{\|\omega\|_2^2}{2\sigma^2}\right\}, \quad \omega \in \mathbb{R}^n, \tag{5}$$

其中 σ^2 表示每个维度上的方差。一个常见的进一步推广方法是将指标集替换为实数轴的一个子集，此时得到的对象称为高斯过程。类似地，但较少见的是将指标集替换为实欧几里得空间的一个子集。此时得到的对象称为高斯场。形式上，高斯场 $Y(\omega)$ $\omega \in \mathbb{R}^n$ 是一个随机变量，使得对于任意

有限个元素 $\{\omega_1, \ldots, \omega_M\} \subset \Omega$，都有

$$(Y(\omega_1), \ldots, Y(\omega_m)) \sim \text{normal}(0, \Sigma). \tag{6}$$

协方差矩阵通常用核函数表示

$$\Sigma_{m,m'} = k(\omega_m, \omega_{m'}), \quad m, m' \in \{1, \ldots, M\}. \tag{7}$$

文献中研究了许多具有不同性质的核函数。最常见的包括

$$k_{a,b}(\omega, \omega') = a^2 \exp\left(-\frac{\|\omega - \omega'\|_2^2}{2b^2}\right), \omega, \omega' \in \mathbb{R}^n. \tag{8}$$

这种协方差函数在本文中也将发挥核心作用。具有此类协方差函数的高斯过程有一项重要性质：其样本几乎必然连续。

4. 理论

4.1. 最优分割

当标签不含噪声时，任何合理的损失函数或度量指标都以目标标签为唯一最优解。然而，当存在噪声时，解决方案便不再明确，而损失函数和度量指标会隐含地指定一组解。因此，探究针对选定损失函数或度量指标的最优解是很自然的。对于交叉熵损失，众所周知，其优化器由边缘函数唯一确定，即

$$\mathrm{CE}_m(c) = \inf_{c' \in \mathcal{M}} \mathrm{CE}_m(c') \iff c(\omega) = m(\omega). \tag{9}$$

λ-a.e. 换言之，最小化交叉熵会得到与该标签相关的边缘概率分布的估计。对于软 Dice，已在 [23] 中证明了

$$\mathrm{SD}(c) = \inf_{c' \in \mathcal{M}} \mathrm{SD}(c') \iff c(\omega) \in \begin{cases} \{1\} & \text{if } m(\omega) > \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \\ [0,1] & \text{if } m(\omega) = \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \\ \{0\} & \text{if } m(\omega) < \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2, \end{cases} \tag{10}$$

λ-a.e. 使用本文采用的设置， [22]对准确率和 Dice 的最优分割进行了刻画。对于准确率，其解为

$$\mathrm{A}_m(s) = \sup_{s' \in \mathcal{S}} \mathrm{A}_m(s') \iff s(\omega) \in \begin{cases} \{1\} & \text{if } m(\omega) > 1/2, \\ \{0,1\} & \text{if } m(\omega) = 1/2, \\ \{0\} & \text{if } m(\omega) < 1/2, \end{cases} \tag{11}$$

λ-a.e. 类似地，对于 Dice，最优分割为

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

λ-几乎处处。理想情况下，最小化损失函数会实现所需目标指标的最大化。也就是说，软分割在使用 $1/2$-阈值进行阈值化后，能够得到使目标指标最大化的分割结果。利用这些特征，可以轻松地将经 $1/2$-阈值化的损失函数优化器与度量指标的优化器联系起来。对于交叉熵，可以推出

$$\begin{aligned} &\mathrm{CE}_m(c) = \inf_{c' \in \mathcal{M}} \mathrm{CE}(c') \implies \\ &\mathrm{A}_m(I_{[1/2,1]} \circ c) = \sup_{s' \in \mathcal{S}} \mathrm{A}(s'). \end{aligned} \tag{13}$$

同样，对于软 Dice，可以得出

$$\begin{aligned} &\mathrm{SD}_m(c) = \inf_{c' \in \mathcal{M}} \mathrm{SD}(c') \implies \\ &\mathrm{D}_m(I_{[1/2,1]} \circ c) = \sup_{s' \in \mathcal{S}} \mathrm{D}(s'). \end{aligned} \tag{14}$$

基于对软 Dice 和交叉熵优化器的刻画，可以清楚地看出，软 Dice 的优化器可以通过先优化交叉熵，然后选择一个事先未知的特定阈值来获得

$$t_m = \sup_{s' \in \mathcal{S}} \mathrm{D}_m(s')/2\,. \tag{15}$$

于是可得

$$\begin{aligned} &\mathrm{CE}_m(c) = \inf_{c' \in \mathcal{M}} \mathrm{CE}(c') \implies \\ &\mathrm{SD}_m(I_{[t_c,1]} \circ c) = \inf_{c' \in \mathcal{M}} \mathrm{SD}(c'). \end{aligned} \tag{16}$$

这种方法是统计学习文献中 [17] 所介绍方法的一种变体。然而，据我们所知，医学图像分割领域尚未对此进行过研究。

在得出这些观察结果后，一个自然的问题是，与交叉熵和软Dice 相关的性能差异是否与交叉熵所采用的阈值有关。两篇探讨最优阈值问题的论文进一步增强了这一假设 [4,10]。这些论文的实验结果表明，对于各种损失函数而言，使用 $1/2$ 作为共同阈值会导致相对于 Dice的次优分割结果。然而，两者都不是一种能够针对每幅图像唯一计算阈值的方法

既未予以考虑，也未建立与理论最优阈值的联系（15）。

在 [21], 中，有人提出，使用 sigmoid 函数施加边界时，软 Dice 有时会表现出不稳定性，这可能与其不具备凸性（也不具备拟凸性）有关，即 $f \mapsto \mathrm{SD}_m(\sigma \circ f)$ $f \in \mathcal{F}$。与此相反，交叉熵对应的 $f \mapsto \mathrm{CE}_m(\sigma \circ f)$ $f \in \mathcal{F}$ 已知是凸的。如果使用替代阈值的交叉熵能取得与软 Dice 相同的性能，那么使用交叉熵或许既能获得软 Dice 的性能，又不牺牲交叉熵的稳定性。

4.2. 标签噪声

与二元分类不同，在二元分类中，大量标签噪声可以通过对二元变量进行简单的独立翻转来建模，而分割中的标签噪声通常可能具有非常复杂的依赖结构。受噪声影响的标签常常看起来像是由一个<\styleid='1'>手在发抖的人绘制的。对于此类噪声，边界点之间的相关性将取决于它们之间的距离。也就是说，如果两个点彼此接近，那么相关性就会很高，反之亦然。

一种模拟看起来仿佛由颤抖的手绘制的分割结果的方法，是通过随机高斯场形变来实现。具体而言，令

$$X = (X_1, \ldots, X_n)^T \tag{17}$$

使得 $\{X_i\}_{i=1}^n$ 是具有核函数（8）的相互独立的高斯场。这意味着， X 是一个取值于 $\mathbb{R}^n \mapsto \mathbb{R}^n$的随机向量场。利用这样的随机向量场，通过如下方式对无噪声分割 $l \in \mathcal{S}$ 进行变形，形成随机分割

$$L(\omega) = \begin{cases} l(\omega + X(\omega)) & \text{if } \omega + X(\omega) \in \Omega, \\ 0 & \text{if } \omega + X(\omega) \notin \Omega. \end{cases}, \omega \in \Omega. \tag{18}$$

随机分割结果的样本呈现出的样子取决于指定核函数时所选择的参数。参数 a 控制形变的幅度，其中更大的 a 会导致更大的形变，反之亦然。参数 b 控制形变的平滑程度，其中更大的 b 会导致更平滑的形变，反之亦然。使用（18）对噪声分割结果进行建模十分方便，因为它可以推导出若干理论性质。在本文中，证明了其中三个重要性质。

在定理 1中，证明了与含噪分割相关的边缘概率可以通过将相应的无噪分割标签与适当的高斯密度进行卷积来计算。

定理 1. 令 l 和 L 如(18) 所定义，并令 p_{σ^2} 如(5)所定义，则可得

$$\mathbb{E}[L(\omega)] = \int_\Omega l(\omega') p_{a^2}(\omega - \omega') \lambda(d\omega'), \quad \omega \in \Omega. \tag{19}$$

在定理 2中，使用边际表示来说明，含噪分割的体积在近似意义下是无偏的，即含噪分割的期望体积近似等于相应无噪分割的体积。之所以不能完全成立，是因为目标结构有一定概率发生变形，从而使结构的某些部分离开域 Ω。对于大多数合理的参数选择，这种情况发生的概率可以忽略不计。

定理 2。 设 l 和 L 如（18）中所定义，且 p_{σ^2} 如 *5* 中所定义，则可得

$$\mathbb{E}[\|L\|_1] = \|l\|_1 - \xi, \tag{20}$$

其中

$$\xi = \int_\Omega \int_{\mathbb{R}^n \setminus \Omega} l(\omega') p_{a^2}(\omega - \omega') \lambda(d\omega) \lambda(d\omega'). \tag{21}$$

定理 3 表明， X （17）在分布上等于某个特定的随机积分向量。在 $n = 1$ 的情况下，可以使用常规的伊藤积分完成这一点；但对于 $n > 1$ 的情况，则需要使用多元对应形式。为此，使用了散布高斯测度。

定理 3。 设 X 如（17）中所定义， p_{a^2} 如（5）中所定义，并且 W_1、.....、 W_n 是定义在 $\mathbb{R}^n$ 上、控制测度为 λ 的相互独立散布高斯测度的独立副本。则可得

$$X(\omega) \stackrel{d}{=} \begin{pmatrix} a(2\pi b^2)^{n/4} \int_{\mathbb{R}^n} p_{b^2/2}(\omega - \omega') W_1(dw') \\ \vdots \\ a(2\pi b^2)^{n/4} \int_{\mathbb{R}^n} p_{b^2/2}(\omega - \omega') W_n(dw') \end{pmatrix}. \tag{22}$$

关于散布高斯测度的详细信息及其证明，请参见补充文档。该定理的重要推论是，当定义域离散化后，含噪分割可以通过以下方式进行采样：为所有体素生成相互独立的高斯分布变量，然后使用高斯滤波器对相关的随机数组进行平滑。这两种操作都可以得到非常高效的实现。因此，即使图像由数百万个体素组成，也能够高效地对含噪分割进行采样。

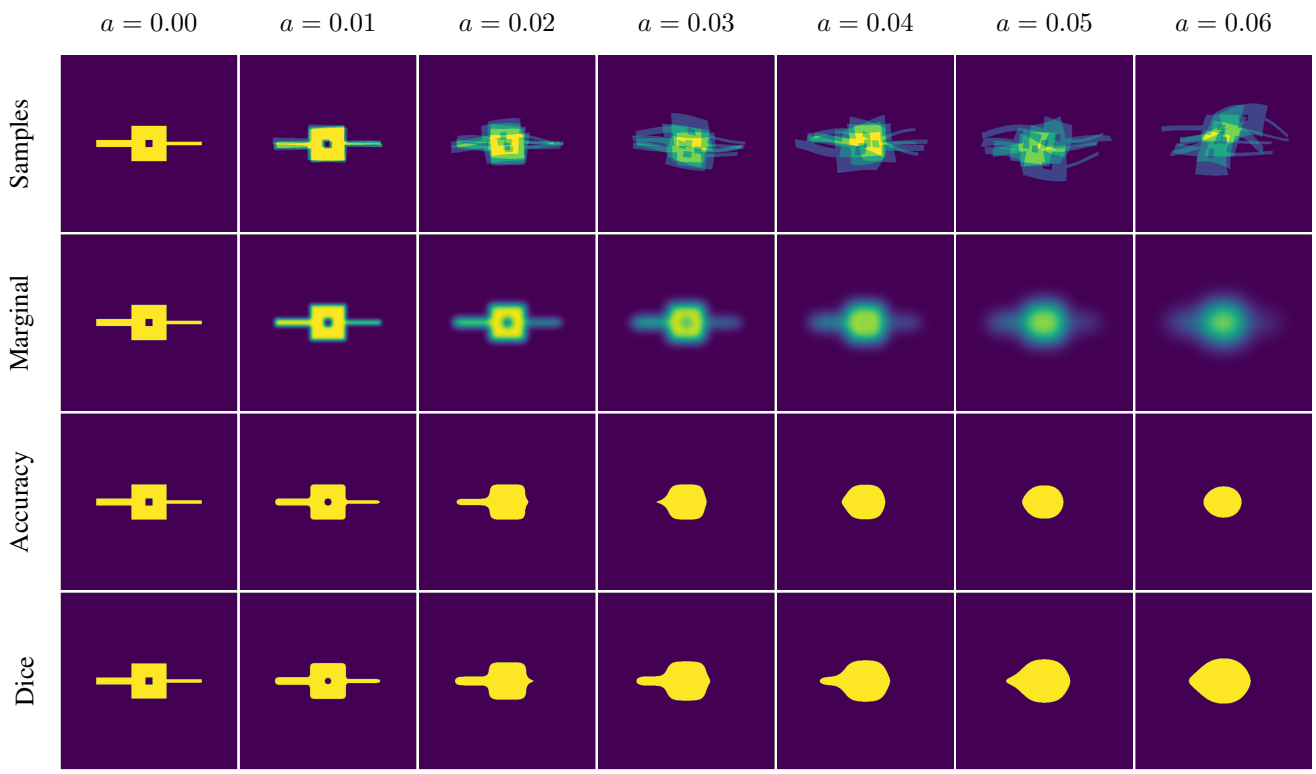


图1。噪声模型及其对二维分割问题影响的示意图，其中各列展示了不同噪声强度的影响。第一行展示了从五个样本中获得的逐像素平均值。第二行展示了边缘函数。第三行展示了与准确率（11）相关联的最优分割，在此类以及大多数自然病例中，该分割是唯一的。第四行展示了与 Dice（12）相关联的最优分割，在此类以及大多数自然病例中，该分割是唯一的。在所有报告的病例中$b = 0.15/\sqrt{2}$。

4.3. 标签噪声下的最优分割

一个自然的问题是，噪声模型的参数会以何种方式影响由损失函数或度量指标得到的最优解。由于本文研究的损失函数和度量指标仅取决于边缘函数，因此可以分两步回答这一问题：首先，考察分割噪声会以何种方式影响边缘函数；其次，考察边缘函数会以何种方式影响损失函数或度量指标的最优解。

在图1中，针对具有不同噪声强度的二维分割问题，展示了这两个步骤。第一行展示了5个样本的逐像素平均值。第二行展示了边缘函数。第三行展示了准确率的理论最优解，在这些病例以及大多数病例中，该解是唯一确定的。注意，根据(13)，这也可由以 $1/2$为阈值的交叉熵最小化器给出。第四行展示了Dice的理论最优解，在这些病例以及大多数病例中，该解是唯一确定的。根据(14)和(16)，这也可由以 $1/2$ 为阈值的软Dice最小化器，以及以 t_m 为阈值的交叉熵最小化器(15)给出。

首先可以观察到，噪声会影响根据这两种度量指标得到的最优分割，使其形状相较于无噪声结构发生变化。这一重要观察与这样一种观点相矛盾：无偏噪声（在我们的模型意义上无偏）会产生无偏的最优分割。具体而言，两种度量指标得到的分割结果都会发生变化，形状的棱角因此变得平滑，而细薄结构和孔洞等细节则会消失。笼统地说，噪声越强，最优分割就越趋近于球形。

另一个观察是，在噪声较低时，这两个度量指标产生的分割结果看起来非常相似，但随着噪声增加，它们产生的分割结果差异越来越大。由于采用 $1/2$ 阈值的交叉熵从理论上讲会得到关于准确率的最优分割，而采用 $1/2$ 阈值的软 Dice 从理论上讲会得到关于 Dice 的最优分割，因此，在低噪声情况下，以 Dice 衡量的两种损失函数的性能应该非常相似，但随着噪声增加，软 Dice 的性能会越来越好。

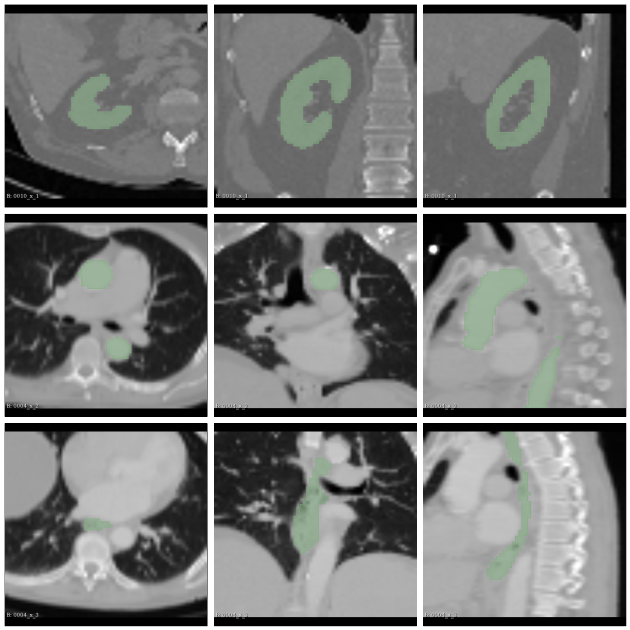


图2。用于实验的 TotalSegmentor数据集中的数据示例 [32] 。第一行展示了穿过其中一个数据点中肾脏质心的轴位、冠状位和矢状位视图。第二行展示了穿过其中一个数据点中主动脉质心的轴位、冠状位和矢状位视图。第三行展示了穿过其中一个数据点中食管质心的轴位、冠状位和矢状位视图。

器官	a	$\mathrm{CE}^{(0)}$	$\mathrm{SD}^{(0)}$	$\mathrm{CE}^{(*)}$
肾脏	0.0000	0.9611	0.9634	0.9615
	0.0100	0.8762	0.8794	0.8774
	0.0200	0.7883	0.7914	0.7909
	0.0300	0.6947	0.7080	0.7055
主动脉	0.0000	0.9525	0.9515	0.9524
	0.0100	0.8639	0.8654	0.8653
	0.0200	0.7557	0.7569	0.7600
	0.0300	0.6215	0.6513	0.6560
食管	0.0000	0.8552	0.8603	0.8602
	0.0100	0.6671	0.6722	0.6814
	0.0200	0.4168	0.4527	0.4829
	0.0300	0.1441	0.3105	0.3489

表 1. 各器官和噪声水平 a 的实验结果表。表中条目是对五折结果取平均后得到的平均 Dice分数。$\mathrm{CE}^{(0)}$ 表示使用了带有 $1/2$ 阈值的交叉熵，$\mathrm{SD}^{(0)}$ 表示使用了带有 $1/2$阈值的软 Dice，$\mathrm{CE}^{(*)}$ 表示使用了公式 (15) 中所述阈值的交叉熵。

5. 实验

在本节中，将使用推导出的噪声模型对讨论的三种方法进行实验测试。简言之，记号 $\mathrm{CE}^{(0)}$ 表示使用交叉熵训练并以 $1/2$进行阈值处理的模型，$\mathrm{SD}^{(0)}$ 表示使用交叉熵训练并以 $1/2$ 进行阈值处理的模型，而 $\mathrm{CE}^{(*)}$ 表示根据 (15) 进行阈值处理、并使用交叉熵训练的模型。

5.1. 数据

实验基于 To-talSegmentor 数据集 [32]进行。该数据集包含 1204 张 CT图像，涵盖 104 种解剖结构（27 个器官、59 块骨骼、10 块肌肉、8 条血管）。为展示噪声对不同形状器官的影响，我们选取了三种不同的器官，包括通常近似球形的右肾、管状且相对较粗的主动脉，以及管状且相对较细的食管。每种器官选取 400 个病例，并分为 5 折，每折 80 个病例。最后，将图像下采样至一半分辨率，并提取以每种结构为中心、大小为 64^3 体素的图像块。

5.2. 模型

实现了一个标准的3D U-Net，由包含两个卷积、实例归一化和 ReLU激活的模块组成。每次下采样时，所使用的特征数量都会增加，依次为 16, 32,64, 128, 256，其中16对应输入分辨率，256 对应瓶颈层。使用内核为 2^3 、步幅为2的卷积进行下采样，并使用内核为2^3 、步幅为 2 的转置卷积进行上采样。

5.3. 训练

模型采用交叉熵或 soft-Dice 进行训练，两种情况下均使用权重为10^{-5}的 l_2 正则化。优化时使用学习率为 10^{-4} 的 Adam，并配合一个调度器：连续 30 个 epoch 未见性能提升后，将学习率除以 5。训练 epoch 数固定为 100，训练数据的顺序随机打乱。批量大小设为 1，并以*0.5* 的概率对数据进行增强。增强方式包括随机平移、随机旋转，以及使用具有随机平滑程度的高斯滤波器。性能根据模型在训练数据上取得的最高度量指标分数进行评估。

5.4. 结果

实验结果如表1和图3所示。结果包括五种不同折数和四种不同噪声水平下，各种分割结果获取方法的平均 Dice分数：$\mathrm{CE}^{(0)}$、$\mathrm{SD}^{(0)}$ 和$\mathrm{CE}^{(*)}$。

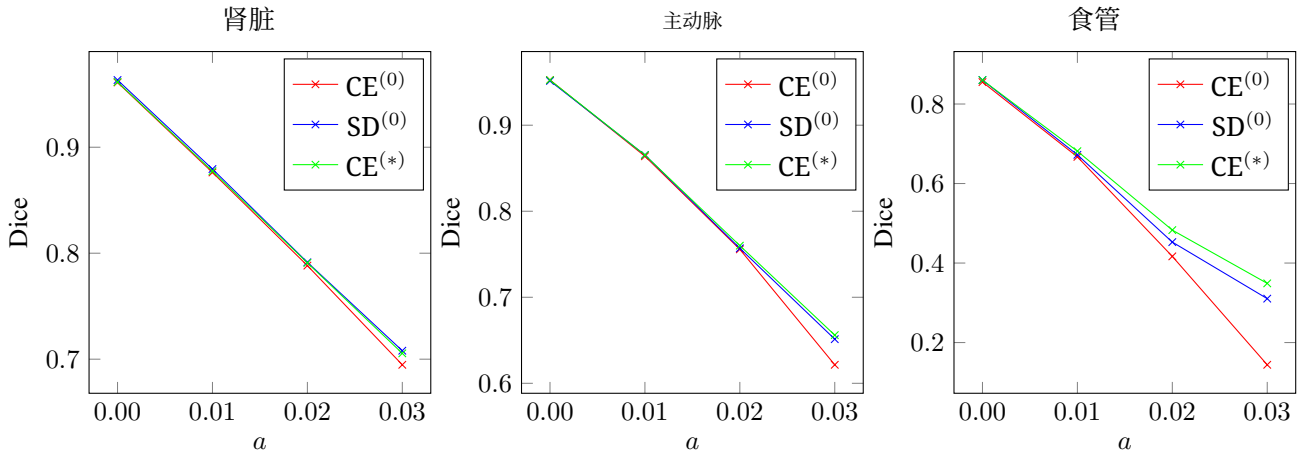


图 3. 表 1 所示实验结果示意图。左侧为肾脏分割问题的结果，中间为主动脉分割问题的结果，右侧为食管分割问题的结果。在每幅图中，绘制了五折实验得到的平均 Dice 值随噪声水平 a 变化的曲线。图例 $\mathrm{CE}^{(0)}$ 表示使用了 $1/2$ 阈值的交叉熵，$\mathrm{SD}^{(0)}$ 表示使用了 $1/2$ 阈值的软 Dice，$\mathrm{CE}^{(*)}$ 表示使用了式 (15) 中描述的阈值的交叉熵。所有报告的病例均为 $b = 0.15/\sqrt{2}$。

首先值得注意的是，所有方法的表现都非常相似，或者 $\mathrm{SD}^{(0)}$ 阈值优于 $\mathrm{CE}^{(0)}$。此外，每当 $\mathrm{SD}^{(0)}$ 的表现优于 $\mathrm{CE}^{(*)}$时，采用 Dice 最优阈值的交叉熵表现相近或更优。简而言之，交叉熵使用替代阈值（15）可达到软 Dice 的表现。其次值得注意的是，随着噪声增大，各方法之间的差异也随之增大。也就是说，在噪声较小时，各方法的表现几乎相同；但随着噪声增大，相应的分数开始出现差异。理论上预期会出现这类结果，如图1所示。

6. 讨论

本文中的实验系统地验证了理论观察结果。然而，在相关影响开始显现之前，以平均 Dice 值衡量的得分非常低。此前比较软 Dice 与交叉熵的研究似乎也呈现出这一规律：得分越低，两者之间的差异越大 [2]。不过，也存在 Dice分数较高但差异仍然显著的情况。这可能有很多原因，其中一种可能是存在具有这种特性的其他噪声。总体而言，结果表明，只要采用合适的阈值，使用交叉熵而非软 Dice 是有依据的。出于其他一些原因，这一选择也更受青睐，例如交叉熵具有更好的稳定性，并且能够提供有用的边缘估计。

7. 结论

本研究比较了软 Dice 和交叉熵的最优解，并说明通常可以通过对交叉熵进行阈值化来获得软 Dice 的最优解，所用阈值可计算，且一般不同于$1/2$。本文提出了一种基于高斯场的现实标签噪声模型，并推导出边缘函数的显式表达式。此外，还推导出一种高效的噪声分割结果采样方法。利用该噪声模型，本文展示了经过 $1/2$ 阈值化的交叉熵和软 Dice 最优解如何受到不同噪声水平的影响，以及噪声增大时所得分割结果会如何逐渐分化。研究还表明，可以通过对交叉熵的最优解进行阈值化来找到软 Dice 的最优解，所用阈值可计算，但先验未知，且不同于 $1/2$。最后，本文在 TotalSegmentor 数据集的真实数据上验证了理论观察结果 [32]。

局限性： 本研究有两个主要局限性。首先，理论结果仅在噪声标签的体积保持不变时严格成立。其次，假设噪声可以用所提出的高斯场模型进行建模。在实践中，噪声不会完全遵循这样的模型，因为目标结构的不同区域通常会存在不同程度的噪声。

致谢： 本研究部分经费由 RaySearch Laboratories AB 提供。

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）与图像数据库资源计划（IDRI）：一套已完成的 CT扫描肺结节参考数据库。《医学物理学》，38(2):915–931，2011。[2] Jeroen Bertels、Tom Eelbode、Maxim Berman、Dirk Van- dermeulen、Frederik Maes、Raf Bisschops 和 Matthew B Blaschko。优化医学图像分割的 Dice分数和 Jaccard指数：理论与实践。载于国际医学图像计算与计算机辅助介入会议，第92–100页。施普林格，2019。[3] Jeroen Bertels、David Robben、Dirk Vandermeulen 和 Paul Suetens。在固有不确定性背景下，对软Dice优化分割图体积偏差的理论分析与实验验证。《医学图像分析》，67:101833，2021。[4] Noah Bice、Neil Kirby、Ruiqi Li、Dan Nguyen、Tyler Bahr、Christopher Kabat、Pamela Myers、Niko Papanikolaou 和 Mohamad Fakhreddine。基于深度学习的解剖结构分割中概率图的敏感性分析。《应用临床医学物理学杂志》，22 119，2021 8 105( ): – 。[5] Pete Bridge、Andrew Fielding、Pamela Rowntree 和 Andrew Pullar。观察者内变异性：我们需要担心吗？《医学影像与放射科学杂志》，47 220，2016 3 217( ): – 。[6] Getao Du、Xu Cao、Jimin Liang、Xueli Chen 和 Yonghua Zhan。基于 U-Net 的医学图像分割：综述。《影像科学与技术杂志》，64:1–12，2020。[7] Charley Gros、Andreanne Lemay 和 Julien Cohen-Adad。SoftSeg：图像分割中软标签训练相较于二值训练的优势。《医学图像分析》，71:102038，2021。[8] Yun Gu、Mali Shen、Jie Yang 和 Guang-Zhong Yang。用于生物医学图像识别的可靠标签高效学习。《*IEEE*生物医学工程汇刊》，66(9):2423–2432，2018。[9] 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，以及生物医学数据的大规模标注与专家标签合成，第112–120页。施普林格，2018。[10] Burhan Rashid Hussein、Cédric Meurée、Malo Gaubert、Arthur Masson、Anne Kerbrat、Benoît Combès 和 Francesca Galassi。脊髓磁共振成像多发性硬化病灶分割的损失函数与决策阈值研究。*HAL* 预印本 *hal-03865212*，2022年。[11] 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。[12] 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。[13] Eytan Kats、Jacob Goldberger 和 Hayit Greenspan。通过蒸馏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。[14] Matthieu Lê、Jan Unkelbach、Nicholas Ayache 和 Hervé Delingette。对图像分割结果进行采样以实现不确定性量化。医学图像分析，34:42–51，2016年。[15] 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。。[16] Hang Li、Dong Wei、Shilei Cao、Kai Ma、Liansheng Wang 和 Yefeng Zheng。超像素引导的标签软化，用于医学图像分割。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。[17] Zachary C Lipton、Charles Elkan 和 Balakrishnan Naryanaswamy。对分类器进行最优阈值化以最大化 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。[18] Alireza Mehrtash、William M Wells、Clare M Tempany、Purang Abolmaesumi 和 Tina Kapur。深度医学图像分割的置信度校准与预测不确定性估计。*IEEE*医学影像汇刊，39(12):3868–3878，2020年。[19] Zahra Mirikharaji、Yiqi Yan 和 Ghassan Hamarneh。学习利用噪声标注分割皮肤病灶。载于标签较少且数据不完美条件下的领域自适应、表征迁移与医学图像学习，第207–215页。施普林格，2019年。[20] Guy Nir、Soheil Hor、Davood Karimi、Ladan Fazli、Brian F Skinnider、Peyman Tavassoli、Dmitry Turbin、Carlos F Villamil、Gang Wang、R Storey Wilson 等。数字化组织病理图像中的前列腺癌自动分级：从多位专家处学习。医学图像分析，50:167–180，2018年。[21] Marcus Nordström、Han Bao、Fredrik Löfman、Henrik Hult、Atsuto Maki 和 Masashi Sugiyama。Dice 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。[22] Marcus Nordstrom、Henrik Hult、Fredrik Löfman 和 Jonas Söderberg。带噪声标签的图像分割：准确率与 Dice 最优解的表征及体积属性。神经信息处理系统进展，35:34321–34333，2022年。[23] Marcus Nordstrom、Henrik Hult、Atsuto Maki 和 Fredrik Löfman。采用软 Dice 的噪声图像分割。*arXiv*预印本 *arXiv*，2023年。[24] Tufve Nyholm、Stina Svensson、Sebastian Andersson、Joakim J onsson、Maja Sohlin、Christian Gustafsson、Elisabeth Kjellén、Karin Söderström、Per Albertsson、Lennart Blomqvist 等。骨盆区域器官的磁共振成像和计算机断层扫描数据及多观察者勾画

骨盆区域器官——金色图谱项目的一部分。医学物理学，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。医学图像分割深度网络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] 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。[29] 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。[30] 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 : 。[31] Eugene Vorontsov 和 Samuel Kadoury。分割网络中的标签噪声：缓解方法必须处理偏差。载于深度生成模型、数据增强、标注与不完美性：第一届 *DGM4MICCAI 2021* 研讨会和第一届 *DALI 2021* 研讨会，与 *MICCAI 2021* 联合举办，法国斯特拉斯堡，*2021*年*10*月*1*日，论文集 *1*，第251–258页。施普林格，2021年。[32] Jakob Wasserthal、Manfred Meyer、Hanns-Christian Breit、Joshy Cyriac、Shan Yang 和 Martin Segeroth。TotalSegmentator：对104种解剖结构进行稳健分割。*arXiv*预印本 *arXiv:2208.05868*，2022年。[33] Shaode Yu、Mingli Chen、Erlei Zhang、Junjie Wu、Hang Yu、Zi Yang、Lin Ma、Xuejun Gu 和 Weiguo Lu。基于深度学习的医学图像分割中噪声标注的鲁棒性研究。《医学与生物学中的物理学》，65(17):175007，2020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。《机器学习研究杂志》，14(1):1033–1090，2013年。

arXiv:2304.04116v3 [cs.CV] 8 Jul 2023

— 补充文档 — 边缘阈值化在噪声图像分割中的应用

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

目录

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

2. 实验 **4**
2.1. 计算边缘分布.4 2.2. 采样噪声分割结果.5 2.3. Dice 最优分割结果.5

*作者还隶属于 RaySearch Laboratories。

1. 证明

1.1. 定理 1 的证明

首先，由于 X 的各个分量是具有协方差核的独立高斯场。

$$k_{a,b}(\omega,\omega') = a^2 \exp\left(-\frac{\|\omega-\omega'\|_2^2}{2b^2}\right), \quad \omega,\omega' \in \mathbb{R}^n, \tag{1}$$

由此可得

$$\mathrm{Var}((X_1(\omega),\ldots,X_n(\omega))^T) = (k_{a,b}(\omega,\omega),\ldots,k_{a,b}(\omega,\omega))^T = (a^2,\ldots,a^2)^T, \tag{2}$$

从而，边缘概率密度 $X(\omega)$ $\omega \in \Omega$ 由下式给出

$$X(\omega) \sim p_{a^2}(x), x \in \mathbb{R}^n. \tag{3}$$

现在，令

$$\hat{l}(\omega) = \begin{cases} l(\omega) & \text{if } \omega \in \Omega, \\ 0 & \text{otherwise}, \end{cases} \tag{4}$$

这意味着

$$L(\omega) \overset{d}{=} \hat{l}(\omega + X(\omega)), \tag{5}$$

其中 $\overset{d}{=}$ 表示分布相同。这与无意识统计学家定律以及对称性 $p_{a^2}(x) = p_{a^2}(-x)$ $x \in \mathbb{R}^n$ 相结合，则对任意 $\omega \in \Omega$ 有

$$\begin{aligned} \mathbb{E}[L(\omega)] &= \mathbb{E}[\hat{l}(\omega + X(\omega))] && (6)\\ &= \int_{\mathbb{R}^n} \hat{l}(\omega+x) p_{a^2}(x) \lambda(dx) && (7)\\ &= \int_{\mathbb{R}^n} \hat{l}(\omega') p_{a^2}(\omega'-\omega) \lambda(d\omega') && (8)\\ &= \int_{\mathbb{R}^n} \hat{l}(\omega') p_{a^2}(\omega-\omega') \lambda(d\omega') && (9)\\ &= \int_{\Omega} l(\omega') p_{a^2}(\omega-\omega') \lambda(d\omega'). && (10) \end{aligned}$$

证明完毕。 □

1.2. 定理 2 的证明

根据定理 1，交换积分次序，并回想 $\int_{\mathbb{R}^n} p_{a^2}(\omega-\omega')\lambda(d\omega) = 1$ 对任意 $\omega' \in \mathbb{R}^n$，可得

$$\mathbb{E}\left[\|L\|_1\right] = \mathbb{E}\left[\int_\Omega |L(\omega)|\lambda(d\omega)\right] \tag{11}$$

$$= E\left[\int_\Omega L(\omega)\lambda(d\omega)\right] \tag{12}$$

$$= \int_\Omega \mathbb{E}[L(\omega)]\lambda(d\omega) \tag{13}$$

$$= \int_\Omega \int_\Omega l(\omega')p_{a^2}(\omega-\omega')\lambda(d\omega')\lambda(d\omega) \tag{14}$$

$$= \int_\Omega l(\omega')\left[\int_{\mathbb{R}^n} p_{a^2}(\omega-\omega')\lambda(d\omega)\right]\lambda(d\omega') - \int_\Omega l(\omega')\left[\int_{\mathbb{R}^n\setminus\Omega} p_{a^2}(\omega-\omega')\lambda(d\omega)\right]\lambda(d\omega') \tag{15}$$

$$= \int_\Omega l(\omega')\lambda(d\omega') - \xi \tag{16}$$

$$= \int_\Omega |l(\omega')|\lambda(d\omega') - \xi \tag{17}$$

$$= \|l\|_1 - \xi. \tag{18}$$

这就完成了证明。 □

1.3. 定理 3 的证明

独立散布测度的定义在 [3, 第 3 节3] 中针对 α-稳定分布作了更一般的介绍。本工作中使用的高斯情形只是 $\alpha = 2$ 时的特殊情形。令 W 为定义在 $\mathbb{R}^n$ 上、控制测度为 λ 的独立散布高斯测度。也就是说，令 $\mathcal{B}_0$ 为具有有限测度的勒贝格可测集合，对于 $\mathcal{B}_0$ 中任意有限个两两不相交的集合 A_1、.....、A_k，随机变量 $W(A_1)$、.....、$W(A_k)$ 相互独立，且 $W(A_i)$ 服从方差为 $\lambda(A_i)$ 的中心高斯分布。对于 $f \in L^2(\mathbb{R}^n)$，随机积分 $I(f) = \int_{\mathbb{R}^n} f(\omega)W(d\omega)$ 定义良好，且服从中心高斯分布，方差为 $\|f\|^2_{L^2}$。事实上，$\{I(f)\ f \in L^2(\mathbb{R}^n)\}$ 是以 $L^2(\mathbb{R}^n)$ 为指标的高斯过程。特别地，对于给定的 $f \in L^2(\mathbb{R}^n)$，过程 $Y = \{Y(\omega)\ \omega \in \mathbb{R}^n\}$ 定义为

$$Y(\omega) = \int_{\mathbb{R}^n} f(\omega-\omega')W(d\omega'), \tag{19}$$

是一个均值为零的高斯过程，其协方差核为

$$k(\omega,\omega') = \int_{\mathbb{R}^n} f(\omega-u)f(\omega'-u)\lambda(du). \tag{20}$$

对于平方指数核

$$k(\omega,\omega') = a^2\exp\left(-\frac{\|\omega-\omega'\|^2}{2b^2}\right) \tag{21}$$

我们可以将 f 识别为

$$f(\omega) = \frac{a}{(\pi b^2/2)^{n/4}}\exp\left(-\frac{\|\omega\|^2}{b^2}\right). \tag{22}$$

事实上，对于任意 ω，$\omega' \in \mathbb{R}^n$ 经过配方平方完成后可得

$$
\begin{aligned}
k(\omega, \omega') &= \int_{\mathbb{R}^n} f(\omega - u) f(\omega' - u) \lambda(du) && (23)\\
&= \frac{a^2}{(\pi b^2/2)^{n/2}} \int_{\mathbb{R}^n} \exp\left(-\frac{\|\omega - u\|_2^2}{b^2}\right) \exp\left(-\frac{\|\omega' - u\|_2^2}{b^2}\right) \lambda(du) && (24)\\
&= \frac{a^2}{(\pi b^2/2)^{n/2}} \exp\left(-\frac{\|\omega - \omega'\|_2^2}{2b^2}\right) \int_{\mathbb{R}^d} \exp\left(-\frac{\|u - \frac{\omega+\omega'}{2}\|_2^2}{2(b/2)^2}\right) \lambda(du) && (25)\\
&= \frac{a^2 (2\pi (b/2)^2)^{d/2}}{(\pi b^2/2)^{d/2}} \exp\left(-\frac{\|\omega - \omega'\|_2^2}{2b^2}\right) && (26)\\
&= a^2 \exp\left(-\frac{\|\omega - \omega'\|_2^2}{2b^2}\right) && (27)
\end{aligned}
$$

现在考虑维度为 d、方差为 $b^2/2$的各向同性正态密度

$$p_{b^2/2}(\omega) = \frac{1}{(\pi b^2)^{d/2}} \exp\left\{-\frac{\|\omega\|_2^2}{b^2}\right\} \tag{28}$$

并注意到它可用于将 f 改写如下

$$f(\omega) = a(2\pi b^2)^{n/4} p_{b^2/2}(\omega). \tag{29}$$

因此可得

$$Y(\omega) = a(2\pi b^2)^{d/4} \int_{\mathbb{R}^n} p_{b^2/2}(\omega - \omega') W(dw'). \tag{30}$$

现在，设 W_1，...，W_n 是 W的独立副本，则可得

$$X(\omega) \stackrel{d}{=} \begin{pmatrix} a(2\pi b^2)^{n/4} \int_{\mathbb{R}^n} p_{b^2/2}(\omega - \omega') W_1(dw') \\ \vdots \\ a(2\pi b^2)^{n/4} \int_{\mathbb{R}^n} p_{b^2/2}(\omega - \omega') W_n(dw') \end{pmatrix}, \tag{31}$$

其中 $\stackrel{d}{=}$ 表示同分布。证明至此完成。 □

2. 实验

实验代码由两部分组成。第一部分用于提取数据，第二部分用于训练模型。大部分代码都很直接，与使用交叉熵或软 Dice 训练的标准 UNet 实现中的代码类似。不过，其中有三种基于本文所述理论的方法需要进一步说明。

2.1. 计算边缘分布

```
导入 numpy as np导入 scipy.ndimage定义 get_noisy_marginals(l,a):scaled_a =
np.array(a)*np.array(l.shape)noisy_marginals =scipy.ndimage.gaus
sian_filter(m,scaled_a,mode='constant')返回 noisy_marginals
```

清单1。用于计算与含噪分割相关的边缘分布的 Python代码，该含噪分割由无噪分割 l 和噪声强度参数 a 形成。该代码是定理 1 的直接实现。

第一个方法用于计算与特定含噪分割相关的边缘分布，其依据是定理 1。此方法的代码列于清单 1。输入由以下内容组成： l 以多维 NumPy 数组表示的无噪声参考分割的离散化版本，以及 a 以浮点数表示的噪声模型参数，用于编码噪声强度。输出是以多维 NumPy 数组表示的边缘分布离散化版本。

2.2. 含噪分割结果的采样

```
导入 numpy a s np
导入 s c i p y . ndimage

定义  获取 _ 含噪 _ 样本 ( l , a ) :
      b = 0.15* np . s q r t ( 2 )
      缩放后的 a_ =np . a r r a y ( a ) *np . a r r a y ( l . s h a p e )
      缩放后的 b_ =np . a r r a y ( b ) *np . a r r a y ( l . s h a p e )
      权重 =缩放后的_a * (2 * np . p i * 缩放后的_b ** 2) * * ( len ( np . s h a p e (m)) / 4 )

      每 b =np . a r r a y ( [ w e i g h t * s c i p y . ndimage . g a u s s i a n _ f i l t e r (
          np . random . n or ma l ( s i z e =l . sha pe ) , s c a l e d _ b [ i ] / np . s q r t ( 2 ) , mode=' c o n s t a n t ' )
          for i in 范围 ( len ( l . s ha pe ) ) ] )
      网 格 _ 网 格 =np . m e s h g r i d ( * [ 范围 ( l . s h a p e [ i ] ) 对 i 在 范围 ( 长度 ( l . sh a pe ) ) ] ,
          索引 =' i j ' )
      噪声 _ 样本 =np . 四舍五入 ( s c i p y . ndimage . m a p _ c o o r d i n a t e s ( l , g r i d _ m e s h + p e r b , mode='
          n e a r e s t ' ) )

      返回噪声_样本
```

列表 2. Python 代码，用于 com 计算与含噪分割相关的一个随机样本，该含噪分割由无噪分割形成 n l
以及噪声强度参数 er a。该代码是定理 3 的直接实现。

第二种方法用于采样含噪分割，并基于定理 3。该方法的代码列于代码清单 2 中。输入由以下内容组成： l 以多维 NumPy 数组表示的无噪参考分割的离散化版本，以及 a 编码噪声强度的噪声模型参数，以浮点数表示。输出是与含噪分割相关的随机样本的离散化版本，以多维 NumPy 数组表示。该方法可以分为三个步骤。首先，计算用于重新缩放的常数。其次，生成高斯场向量；在数值设置中，通过为 l 中的每个条目抽取独立同分布的高斯变量，并使用高斯滤波器处理所得数组来近似该向量。第三，使用所得的随机形变数组对无噪分割 l 进行形变。

2.3. Dice 最优分割结果

```
导入 numpy a s np

de  g e t _ o p t _ d i c e _ s e g (m) :
f     p s i = np . f l i p ( np . s o r t (m. f l a t t e n ( ) ) )
      d = 2* np . cumsum ( p s i ) / ( np . sum (m) +np . a r a n g e ( 1 , len ( p s i ) +1 ) )
      t =np . max ( d ) / 2
      s = 1.0* (m>=t)
      返回s
```

代码清单 3。用于从边缘函数生成 Dice 最优分割的 Python 代码 m。

第三种方法用于计算 Dice 意义下的最优分割。该方法的代码见代码清单 3，摘自 [2]。这是对二元分类中提出的一种方法 [1]的高效变体。输入由以下内容组成： m 以多维 NumPy 数组表示的边缘函数离散化版本。输出是以多维 NumPy 数组表示的 Dice 意义下最优分割的离散化版本。其思路是将 m 中的体素按从大到小排序，然后计算与分割结果相对应的 Dice 分数；这些分割结果

通过选取排序列表中对应的前若干个体素形成。对每一种可能的体素数量都执行此操作。然后用最高分数确定一个阈值，并据此生成最终分割结果。

参考文献

[1] Zachary C Lipton、Charles Elkan 和 Balakrishnan Naryanaswamy。使 F1 度量最大化的分类器最优阈值设定。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。[2] 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，第 35 卷，34333，2022年，34321：–。[3] Gennady Samoradnitsky 和 Murad S. Taqqu。稳定非高斯随机过程：无限方差随机模型。*Chapman and Hall，1994*年。