arXiv:2303.04873v1 [cs.CV] 8 Mar 2023

MOREA：一种用于三维医学图像多目标可变形配准的GPU加速进化算法

Georgios Andreadis 莱顿大学医学中心 莱顿，荷兰 G.Andreadis@lumc.nl

Peter A.N. Bosman 数学与计算机科学中心 阿姆斯特丹，荷兰 Peter.Bosman@cwi.nl

Tanja Alderliesten 莱顿大学医学中心 莱顿，荷兰 T.Alderliesten@lumc.nl

摘要

在需要大变形的情况下，找到一种能够将一幅图像变换为另一幅图像的逼真形变，被认为是医学图像分析中的一项关键挑战。采用恰当的图像配准方法来实现这一目标，可以催生许多应用，使信息能够在图像之间传递。目前，临床应用受到阻碍，因为许多现有方法在每次使用前都需要进行大量配置，或者无法（逼真地）捕捉大变形。近期一种多目标方法采用多目标实值基因池最优混合进化算法（MO-RV-GOMEA）和双动态网格变换模型，展现出良好前景：它揭示了图像配准问题中固有的权衡，并对二维大变形进行建模。本研究在此基础上进一步提出 MOREA：首个能够处理大变形、基于进化算法的三维图像可变形配准多目标方法。MOREA 纳入了三维生物力学网格模型，以确保物理合理性，并实现了完全 GPU 加速。我们在 4 名宫颈癌患者的腹部 CT 扫描图像上，将 MOREA 与两种最先进的方法进行比较；后两种方法均针对每位患者进行了配置，以取得最佳结果。MOREA 无需针对每位患者单独配置，在代表最困难病例的 4 名患者中，有 3 名的结果显著优于上述方法。

关键词

可变形图像配准，多目标优化，智能网格初始化，修复方法，GOMEA

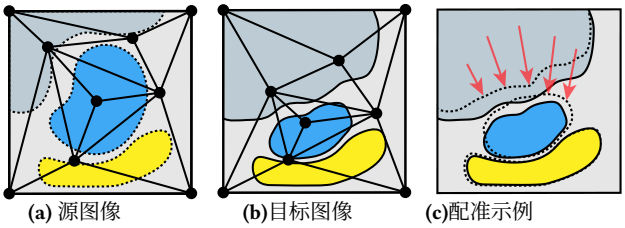
1 引言

近几十年来，放射肿瘤学领域取得了迅速发展。治疗前、治疗期间和治疗后获取的医学图像是现代实践的关键。尽管这些图像已在许多方面为临床决策提供指导，但如何在存在大变形或内容不匹配的多幅图像之间传递信息，已被证明是一项艰巨的挑战，也一直难以在临床上广泛应用。总体而言，可变形图像配准（DIR）的挑战在于找到一种真实的变换，使两个或多个图像空间彼此匹配，如图1所示。借助这一变换，还可以在图像之间传递其他元数据，例如标注的轮廓 [30] 或三维放射剂量分布 [33]，，从而为提高放射治疗的精准度创造机会 [16]。

可变形图像配准问题包含三个主要目标：基于图像的目标（用于视觉比较）、基于轮廓的目标（用于评估对象轮廓重叠情况），以及基于真实性的目标（用于衡量执行形变所需的能量）。这些目标彼此冲突，尤其是在存在大变形和内容不匹配的情况下 [1]。因此，可变形图像配准本质上是一个多目标问题，这使得进化算法（EA）非常适合用于优化该问题 [19]。

目前已有多种可变形图像配准方法 [5, 17, 45]。这些方法均采用单目标方法，要求用户针对每次使用预先选择与优化目标相关的权重。先验地然而，这可能妨碍其临床应用，因为研究表明，通常很难为特定患者选取合适的权重（及其他参数），而且这些参数会显著影响配准质量 [36]。即使经过配置以获得最佳结果，许多现有方法仍难以应对图像间的大变形和内容不匹配，这是由于其底层变换模型以及（通常基于梯度下降的）优化技术存在局限。这一缺陷构成了其应用于临床工作流程的第二个障碍。因此，仍然需要一种可变形图像配准方法，既不需要先验地配置目标权重并且 又能处理大变形。

此前已有研究通过采用多目标方法 [2]，解决了需要先验配置目标权重的问题。这样一来，在先验进行标量化问题表述时，便无需为优化目标选择权重；因为该方法可以生成一组解，恰当地体现不同冲突目标之间的权衡，使用户能够事后从中选择一个解。为克服第二个障碍，研究人员提出了一种灵活的双动态三角网格变换模型，支持逆一致的生物力学配准 [3]。该模型能够匹配两幅图像中的结构，以捕捉大变形。多目标实值基因池最优混合进化算法（MO-RV-GOMEA）已被证明，借助将问题分解为局部部分评估，该算法能够有效地使用此模型对二维图像进行可变形图像配准 [10]。图形处理器（GPU）极其适合并行执行这些部分评估，从而显著提高



(a) 源图像 (b)目标图像 (c)配准示例

图1：两幅存在大幅形变的图像，以及使用 MOREA 的双动态网格变换模型进行可变形图像配准的示例。

速度 [12]。近期，研究人员已初步将这种 GPU 加速方法扩展到三维图像 [4], 。由于图像信息量增加（从二维图像中的 6.5 万个像素增至三维图像中的 200 多万个体素），部分评估的优势可能会更加显著，但评估数量也会增加，且每次评估的成本更高。研究表明，这种扩展后的方法能够解决单个物体的简单配准问题，但它缺少应对涉及多个相互作用物体的临床问题所需的若干关键组件。

因此，在这项工作中，我们提出了 MOREA，这是首个能够借助生物力学模型配准具有大变形的三维图像、且无需先验配置目标权重的基于进化算法的多目标配准方法。在 MOREA 中，我们采用一种新颖的定制网格生成方法，在感兴趣的结构上初始化三维四面体网格，并内置了针对折叠网格的修复机制。此外，借助 MOREA，我们改进了以往针对所有目标的建模策略 [4] ，以确保实现理想的形变。

2 大变形可变形图像配准

本节将定义可变形图像配准优化问题（第 2.1 节），并考察现有方法（第 2.2 节）。

2.1 问题定义

对于一对图像，可变形图像配准问题是寻找一个非刚性变换 T，使源图像 I_s 形变后尽可能贴合目标图像 I_t [40]。我们区分单向配准与对称配准：在单向配准中，仅优化 T（I_s）$\approx I_t$；而在对称配准中，还会优化T，（I_t）$\approx I_s$ [40]。这可以提高配准的物理可行性。配准的另一个理想特性是逆一致性 [40],，它确保源图像中的任意点与其在目标图像中的对应点之间存在一一对应关系。

配准通常可以根据三类质量指标进行评估。图像强度指标使用互相关或互信息等指标 [26], 比较 $T(I_s)$ 预测的体素强度值与 I_t 的体素强度值。轮廓指标通过将 T 应用于代表轮廓（C_s 和 C_t）的点集对，并计算这些点集之间的距离，以评估配准精度。Chamfer 距离 [22] 就是一个例子：对于每一对 C_s, C_t，计算 $T(C_s)$ 中的点与 C_t 中任意一点之间的最小距离，并取其中的最大值。可变形图像配准方法也可以在初始化时使用这些轮廓，构建优化过程中使用的变换模型。最后，变形幅度指标通过使用图像空间的物理模型 [23] 测量施加形变所需的力，来表征配准真实性。这可以作为一种正则化机制，防止配准过度拟合。

2.2 相关工作

这三种质量指标是相互冲突的目标，形成了一种权衡 [1]。近年来出现了许多单目标配准方法，通常试图应对这一权衡，方法是探索不同的目标标量化方式。然而，这样做的缺点是必须<a priori>预先设定目标权重。我们根据上述定义的质量指标类别，对现有方法进行粗略分类，分为主要优化以下目标的方法：(1) 强度匹配；(2) 轮廓匹配；(3) 同时匹配两者。表1对若干具有代表性的方法及其他特征进行了比较。

第一类方法的一个示例是针对强度匹配进行优化的 Elastix 工具箱 [28]。它采用基于 B样条 的变换模型，使用贝塞尔曲线对物理空间进行建模。在该模型中，Elastix 针对图像强度进行优化，并通过变形幅度指标进行正则化。尽管这对许多应用而言是很好的选择，但我们观察到，对于具有局部不连续性（如本研究所考察的情况）的更复杂大变形，配准可能较为困难。ANTs SyN 配准方法 [5] 旨在对这类大变形进行建模，其特点是利用时变速度场执行对称、逆一致且基于强度的配准。第三种基于强度的方法是 Demons算法 [42], , 它使用光流和麦克斯韦妖的原理进行逆一致配准。该方法的较新版本还具备处理图像之间内容不匹配的机制 [34]。从理论上讲，ANTs 和 Demons 方法都能灵活地对大变形进行建模，但缺乏生物力学建模能力，且仅基于图像强度运行。这可能妨碍可靠地产生解剖学上有效的配准 [30]。

这是考虑第二类方法的原因之一。其中一种方法是薄板样条鲁棒点匹配方法（TPS-RPM），它使用薄板样条模型对轮廓进行形变 [18]。后续工作在一个腹部测试案例中验证了该方法，将膀胱和两个周围器官的形变进行了配准 [44]。此外，还有一种对称版 TPS-RPM，可提高其处理大变形时的鲁棒性 [8]。同期开展的研究也将类似模型应用于轮廓，以解决腹部配准问题 [39]。尽管可以对大变形进行建模，但无法保证变换具有生物力学合理性，而且仍需配置目标权重。另一种基于轮廓的方法是 MORFEUS [17], , 它使用有限元法（FEM）求解器对成像对象的网格表示进行配准。该方法在腹部近距离放射治疗应用中已显示出良好的前景 [37]。虽然 MORFEUS 使用生物力学模型，提升了逼真度，但它没有考虑图像强度，因此会丢失对象表面之间的细节，并过度依赖（用户提供的）轮廓。

近期研究通过提出一种结合基于轮廓和基于图像的方法，来弥补这一不足：ANAtomically CONstrained DeformationAlgorithm（ANACONDA）[45]使用拟牛顿算法优化图像项和轮廓项的固定标量化形式。然而，这种方法缺少生物力学建模，还引入了另一个需要配置的参数。其他混合方法也已出现，例如将 Demons 方法与局部有限元网格相结合 [48],，或使用基于图像的配准步骤推导组织弹性系数，之后将其用于基于有限元的配准方法 [29]。

总体而言，我们发现了一个空缺：缺少一种能在同一模型中涵盖配准各个方面的方法。如表1所示，我们通过 MOREA 来填补这一空缺：它兼具基于图像和基于轮廓的特点，采用生物力学建模，并利用可变形图像配准问题的多目标

特征	Elastix [28]	ANTs SyN [5]	Demons [42]	TPS-RPM [18]	ANACONDA [45]	MORFEUS [17]	MOREA （本研究）
基于图像的	✓	✓	✓	✗	✓	✗	✓
基于轮廓的	✗	✗	✗	✓	✓	✓	✓
生物力学的	✗	✗	✗	✗	✗	✓	✓
多目标	✗	✗	✗	✗	✗	✗	✓

表1：根据支持的配准特征，对部分现有的突出可变形图像配准方法进行比较。

特性。这些创新得益于进化算法的灵活性和稳健性，使其非常适合优化不可微的多目标问题。此外，目标函数涉及数百万个图像体素值，因此计算开销很大，需要借助硬件加速。现代基于模型的进化算法（如 MO-RV-GOMEA）与图形处理器具有出色的兼容性，因此非常适合用于优化可变形图像配准问题。

3 MO-RV-GOMEA

黑箱优化（BBO）问题的结构只能通过反复进行函数评估来揭示。另一方面，灰箱优化（GBO）问题的结构（部分）已知，可在优化过程中加以利用。GOMEA 系列进化算法已被证明特别适合高效求解基准问题和现实世界中的 GBO 问题 [41]。其向多目标实值问题的扩展 MO-RV-GOMEA [11], , 甚至已应用于临床实践中的前列腺近距离放射治疗计划 [7, 13]。本文概述 MO-RV-GOMEA 的关键工作原理。详细说明见相关文献 [14]。

非支配解会在各代中保留于容量预先设定的精英档案中 [31]。每一代开始时，先从当前种群中选出一个非支配解子集。然后将该子集划分为 k 个大小相等的簇。对于每个簇，MO-RV-GOMEA 都会采用一个连锁模型，该模型使用一组相关变量集（称为子集族（FOS）元素）来描述变量之间的依赖关系。在 BBO 设置下，可在优化过程中学习这一连锁模型；而在 MOREA 中，我们采用基于变量拓扑邻近性的静态连锁模型（见第 4.2.1 节）。随后，通过一种称为 最优混合 的过程，对 FOS 元素中的变量进行联合处理，以进行变异。在此步骤中，为每个簇中的每个 FOS 元素估计分布，并从这些分布中采样新的部分解。对新采样的部分解进行评估；若将其插入父解后得到的解支配父解，或该解在当前精英档案中属于非支配解，则接受该部分解。

4 方法

本文概述的方法建立在近期提出的适用于三维图像的多目标方法 [4]之上。本节将介绍我们新增的技术，包括问题建模（第4.1节）、解种群的初始化（第4.2节）以及形变优化（第4.3节）。

4.1 建模

4.1.1 利用组织特异性弹性提升真实感。 相关研究表明，与假设整个图像区域具有均匀弹性相比，使用组织特异性弹性能够提高所得形变的真实感 [37, 46]。基于这一发现，我们扩展了现有工作 [4] 中使用的变形幅度目标函数：根据每个四面体所在的图像区域，为其计算弹性因子。附录 A 提供了这一计算的实现细节。我们在探索性实验中观察到，这种做法能带来更好的配准结果（见附录 C.3.1 节）。

为计算形变幅度目标，我们考虑每个四面体 $\delta \in \Delta$中所有对应的边 e_s 和 e_t，它们分别属于源图像上的网格和目标图像上的网格。其中包括 4 条辐条边，能够更好地捕捉扁平化运动，因此每个四面体共有 10 条边 [4]。给定四面体特异性弹性常数 c_δ，目标函数按如下方式计算：

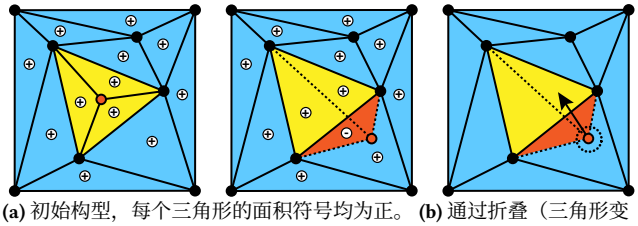
$$f_{magnitude} = \frac{1}{10|\Delta|} \sum_{\delta \in \Delta} \left[\sum_{(e_s, e_t) \in E_\delta} c_\delta (\|e_s\| - \|e_t\|)^2 \right]$$

4.1.2 稳健地估计图像相似性。我们使用的强度目标定义为体素间比较，即求和计算强度差平方，并对前景（即强度非零）体素和背景（即强度为零）体素的比较进行特殊处理。我们使用一种适合 GPU 加速的随机采样技术（详见附录 A）。利用两幅图像中所有采样图像点的集合 P_s 和P_t，以及源图像 I_s和目标图像 I_t的图像强度，目标函数定义如下：

$$f_{intensity} = \frac{1}{|P_s| + |P_t|} \left[\sum_{p_s \in P_s} h(p_s, T(p_s)) + \sum_{p_t \in P_t} h(p_t, T'(p_t)) \right]$$

$$h(p_s, p_t) = \begin{cases} (p_s - p_t)^2 & p_s > 0 \wedge p_t > 0 \\ 0 & p_s = 0 \wedge p_t = 0 \\ 1 & \text{otherwise} \end{cases}$$

4.1.3 引导误差近似。与以往将精确的引导度量用作目标之一的工作不同 [4],，本工作选择引入一种近似度量，该度量可借助我们已用于计算图像相似性目标值的 GPU 加速采样方法更高效地计算。初步实验表明，两者结果非常相似（从体素位移场来看），这也与一个完美的



(a) 初始构型，每个三角形的面积符号均为正。 **(b)** 通过折叠（三角形变为红色）时符号改变检测到折叠。 **(c)** 修复方法：移动红色点以消除折叠。

图 2：存在和不存在约束违反（折叠）的网格配置二维示意图。由于红点移出了中央三角形（以黄色标示），其中一个三角形发生了折叠。折叠区域以红色标示。

引导误差不一定是最佳解。附录 A 中提供了有关实现的详细信息。

MOREA 的引导目标在位置 P_s 和P_t处计算，使用由所有点集对C_s, C_{ti}组成的集合 G，以及最小点到点集距离 $d(p, C)$。引导点总数表示为 $|G_s|$和 $|G_t|$，截断半径表示为 r。引导目标现定义如下：

$$f_{guidance} = \frac{1}{|P_s| + |P_t|} \sum_{\langle C_s, C_t \rangle \in G} \left[\frac{|C_s|}{|G_s|} g(P_s, T, C_s, C_t) + \frac{|C_t|}{|G_t|} g(P_t, T', C_t, C_s) \right]$$

$$g(P, \Phi, C, C') = \sum_{\substack{p \in P \\ d(p,C) < r}} \left[\frac{r - d(p, C)}{r} (d(p, C) - d(\Phi(p), C'))^2 \right]$$

4.1.4 快速计算约束。MOREA 的解以包含数百个点的网格表示，这些网格很容易缠绕并形成折叠配置。为保证逆一致性，应避免此类约束违反。先前的工作 [4]采用了一种策略，但在更复杂的网格中容易出错。MOREA 引入了一种新颖的折叠检测方法，该方法基于一个观察到的现象：网格发生折叠时，至少一个四面体的体积符号会改变，如图 2 所示（图中为二维情形，但这一现象在三维中同样成立）。我们利用这一现象检测折叠并衡量其严重程度，从而为修复提供机会（参见第 4.3.1 节）。本方法的实现细节见附录 A。

4.2 配准解的初始化

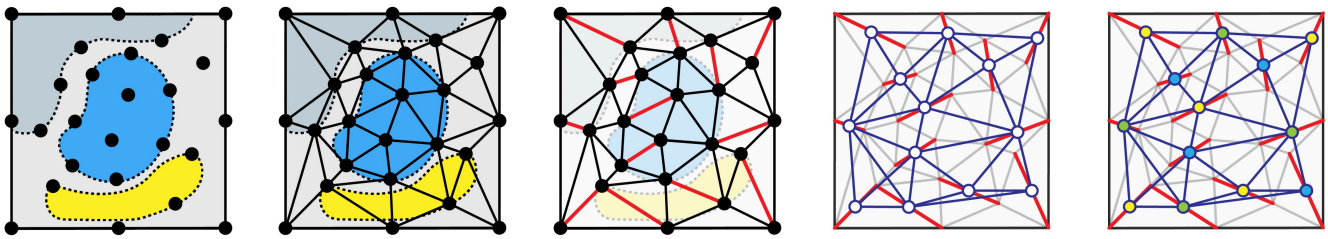
如果提供给优化器的初始猜测比随机猜测或网格状初始化 [9]更接近目标空间中理想的区域，就能显著提升性能。我们提出两种可提供此类初始猜测的技术。

4.2.1 利用网格初始化发挥问题结构优势。我们初始化网格，使其与图像中的物体对齐，并对一种适用于二维图像的现有方法进行调整 [9]，将其扩展以便在图形处理器上实现并行化。首先，我们在源图像中物体的轮廓上放置点，以捕捉其形状（见图3a）。我们从用于引导目标的轮廓标注中贪心选取分布较均匀的子集作为这些点，同时在图像各处随机选取少量点。然后，我们使用 TetGen 套件 [25] 对这些点进行 Delaunay 四面体剖分（见图3b）。由此得到一个网格，我们将其复制到目标图像空间，从而完成双动态变换模型。

如第3节所述，MO-RV-GOMEA 会在变异过程中联合评估变量组（即 FOS 元素）。探索性实验表明，对于此问题，将边用作 FOS 元素（即由两个相连点组成的组，其变量编码这些点的坐标）效果良好。如果两个 FOS 元素完全独立，即它们的变量不会用于各自集合的部分评估，那么这些 FOS 元素的变异和评估便可以并行执行。为便于在图形处理器上进行并行评估和优化，我们还采取了两个步骤。首先，我们执行贪心集合覆盖算法[1]，找出一个覆盖所有点的边子集（见图3c），使每个变量（点坐标）都经历变异。我们也可以使用所有边，但这样会导致点被纳入多个 FOS 集合，因此每一代都要经历多次变异。出于并行化的考虑，选择一组（近似）最小的边会更高效。

给定集合覆盖找到的边子集，我们现在确定哪些 FOS 元素可以安全地并行优化。为此，我们基于拓扑邻近性构建交互图 [12]，当两个元素对应的相关四面体集合有重叠时，就在它们之间建立连接；也就是说，这些四面体会在$_1$被更改时重新评估

来源：https://github .com/martin - steinegger/setcover



(a) 放置在潜在感兴趣的位置上的点。 **(b)** 定制网格衍生自来自这些点。 **(c)**通过集合覆-盖选定的边进行变异。 **(d)** 选定边之间的交互图（蓝色）选定边之间。 **(e)** 计算得到的图着色在交互图上。

图3：网格初始化过程的二维示意图，该过程生成定制网格，并确定哪些边组（即 FOS 元素）可以并行优化。选中的边以红色突出显示，交互边以蓝色突出显示。

元素。（见图 3d）。给定该图，我们使用 DSATUR 图着色算法 [15] 创建并行分组（见图 3e）。每个并行组中的相关四面体都可以在图形处理器上并行评估；已有研究证明，这种方法在二维图像上可实现超过 100 倍的加速 [12]。

通过指定应纳入生成网格的表面，可以进一步提高四面体网格的质量。我们将这一原则应用于膀胱：使用移动立方体算法生成表面网格，然后将其三角形表面指定为网格生成算法的约束，从而确保膀胱表面的三角形包含在网格中。探索性实验表明，采用这一步骤可获得更优性能（见附录 B.3.1）。

4.2.2 确保初始种群的多样性。为提高初始种群的多样性，先前的工作从网格初始化解 [4] 出发，为网格中的每个点生成随机偏差。我们发现，这种方法会使生成的解中出现许多网格折叠构型，而这些构型会被丢弃，进而降低收敛速度。本文采用径向基函数方法引入不产生网格折叠的大变形。有关这些场的生成方式及其在解网格上的应用方式，详见附录 A。

4.3 修复与引导

在优化过程中，我们采用两种技术，以提高所得解的质量，并缩短获得这些解所需的时间。

4.3.1 修复不可行解。默认情况下，不可行解（即两个网格中任意一个或两个网格存在一个或多个折叠的解）会被丢弃。然而，这可能会妨碍生成高质量的子代，因为不可行解仍可能为搜索空间中质量更高的区域提供有用信息。因此，我们设计了一种修复方法，尝试逐点逆转折叠。对于折叠四面体中的每个点，该方法使用高斯分布对点进行变异，其尺度根据估算的点到周围三维多边形的距离确定。经过 64 次采样后，若存在约束改善效果最佳的变化，则选取该变化。如果所有采样结果都导致情况恶化，则中止修复。图 2c 展示了对单个点的修复过程。

4.3.2 使用自适应引导施加压力。通常，近似集应尽可能多样，同时尽可能贴近帕累托集。然而在实践中，用户对帕累托前沿的不同区域感兴趣的程度并不相同。对于存在大变形的图像，进行医学图像可变形配准的用户通常不会对变形幅度较小的解感兴趣。用户真正最关注的是引导目标值良好的解，因此我们希望算法将搜索引导至目标空间中的该区域。按照先前的工作 [1],，我们实现了一种自适应引导策略，在经过 100 代的探索后，将前沿引导至高质量的引导解。给定精英档案中任一解的最佳引导目标值 s_G，我们只保留引导目标值介于 $[s_G;\ 1.5s_G]$ 之间的解，也就是说，这成为一项硬性约束。

5 实验

我们将 MOREA 与现有最先进的配准方法进行比较。由于问题十分复杂，我们不对所有方法设定统一的时间限制，而是确保它们都已（合理地）收敛。我们对所有方法的所有配置均进行 5 次重复实验，并采用可复现的随机种子。所有 MOREA 配准实验均在配备 NVIDIA RTXA5000 GPU 的 Dell Precision 7920R 机器上运行。实验设置和结果的更多信息见附录。

5.1 配准问题

我们在4个具有大变形的临床配准问题上测试所有方法（见表 2）。我们回顾性地选取了宫颈癌患者因放射治疗计划而采集的腹部区域连续两次计算机断层扫描（CT），扫描设备为 Philips Brilliance Big Bore 扫描仪。在第一次 CT 扫描中，患者的膀胱充盈；第二次扫描时，膀胱排空，因此明显缩小。要在尊重周围器官（如直肠和肠道）及骨性解剖结构的同时，正确配准这种大变形，颇具挑战性。患者1–3代表临床实践中的常见病例：全膀胱扫描中存在大变形，且膀胱与肠道之间几乎没有或完全没有间隙。患者4的膀胱形状基本保持不变，且膀胱与肠道之间有较宽的间隙，因此配准更容易。然而，这种病例在临床实践中也较为罕见，因此代表性较低。

CT扫描的轴位切片厚度为3毫米，层内分辨率介于(0.86，0.86)毫米 和(1.07，1.07)毫米之间。为确保一致性，每次扫描均重采样至(1.5，1.5，1.5)毫米 。随后，使用放射治疗技师（RTT）勾画的骨轮廓，对每对扫描图像进行刚性配准（即线性平移、旋转或缩放），使两次扫描中的骨性解剖结构对齐。每对图像均裁剪至膀胱周围的轴对齐包围盒，边缘留出30毫米，范围取两幅图像的最大边界。这样可将配准限制在实际实施治疗的区域内，并包含周围的危及器官。

每次扫描图像中关键器官的轮廓均由一名 RTT 标注，并经放射肿瘤科医生核验。定义这些轮廓的点集用作 MOREA 引导目标的输入。我们还使用这些临床轮廓，通过填充由各切片上的轮廓形成的二维多边形估计区域，为每个器官和骨骼生成二值掩膜。与临床实践中的常见情况一样，这些轮廓可能彼此重叠，因为器官是独立勾画的，且通常会留出较小的安全边界。因此，配准方法需要具备足够的鲁棒性，以应对这种重叠。为评估之用，RTT 已在两次扫描图像上标注了若干解剖学相关对应标志点，并由放射肿瘤科医生核验（见附录 D）。

5.2 配准方法

我们考虑了若干现有且广受欢迎的配准方法，这些方法均有可执行代码。对于这些方法，我们采用两阶段配置流程。首先，我们针对一个患者扫描对（患者1）探索相关的粗粒度设置，

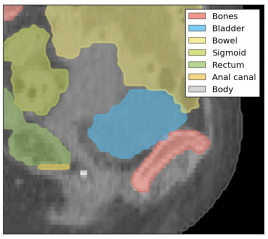
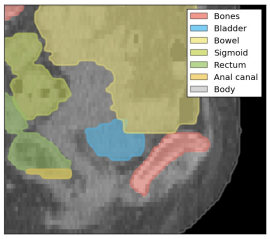
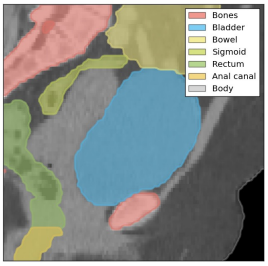
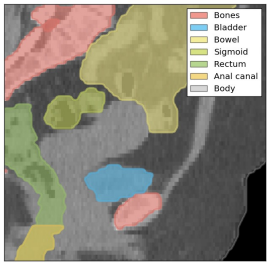
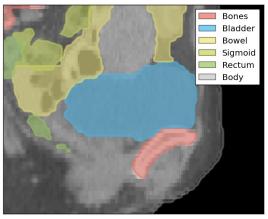
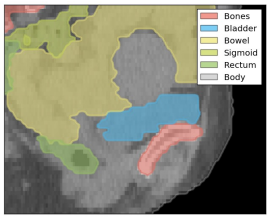
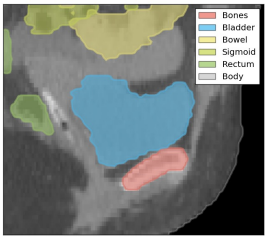
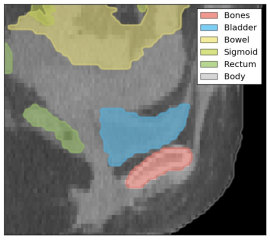
实例	来源	目标
患者1		
患者2		
患者3		
患者4		

表2：所有配准问题的矢状位切片，器官以不同颜色勾画。

以找到适合该成像模态和问题难度的配置。然后，我们针对每个患者扫描对，对其余设置（例如目标标量化权重）进行细粒度配置。下面介绍每种方法最终采用的配置，包括 MOREA 的通用粗粒度配置。关于这些配置的确定过程以及更多配置实验的详细概述，见附录 C。

5.2.1 Elastix。我们配置 Elastix，以执行正则化的多分辨率 [43] 图像配准。推荐设置[2] 在我们的扫描图像上未能取得令人满意的结果，因此我们首先针对每位患者，将复合掩膜图像相互配准。随后以此为起点，在原始图像强度上进行优化。作为针对每位患者的细粒度配置步骤，我们调整分配给形变幅度目标函数的权重，按照相关研究 [8] 的做法，在固定扫描中逐步按指数增加权重 $[0, 0.001, 0.01, \ldots, 10.0]$。

5.2.2 ANTs SyN。对于 ANTs SyN 算法，多分辨率配准的推荐设置[3] 同样不尽如人意，因此我们针对若干关键参数进行了初步配置实验，参数列于附录 C。我们还在额外的图像通道中添加了复合掩膜，并与图像一同进行配准。对于每位患者，我们通过测试与 Elastix 相同的权重，来测试整体形变的相同正则化权重。

5.2.3 本研究：*MOREA*。MOREA 使用单分辨率方法，并配置为生成包含 600 个点的网格（即问题的维数为 3600），采用第 4.2 节所述的网格生成策略。我们将精英档案容量设为 2000，并在优化过程中使用 10 个簇，运行时间预算为 500 代；在此期间，进化算法（EA）会收敛（见附录 D）。由于 MOREA 是一种返回配准近似集的多目标方法，因此无需针对每位患者进一步配置。

5.3 配准评估

对于本研究中的复杂配准问题等问题，需要采用多方面的评估方法。下面我们概述评估配准的两种主要方法：基于表面的精度分析和目视检查。附录第 B.2 节还介绍了其他方法，并在附录 C 和 D 中加以应用。

5.3.1 基于表面的配准精度。评估配准精度的一个关键部分是评估对象的表面（轮廓）对齐得有多好 [16]。我们使用豪斯多夫距离，它表示两个对象表面上任意两点之间的最小距离中的最大值。这可以理解为最差表面匹配的严重程度。为考虑图像边界区域可能存在的形变误差，在计算此度量时，我们会在每一侧舍弃 15毫米 的边缘。由于这小于之前 30毫米的裁剪边缘，因此膀胱及其周围区域不会受到这次二次裁剪的影响。

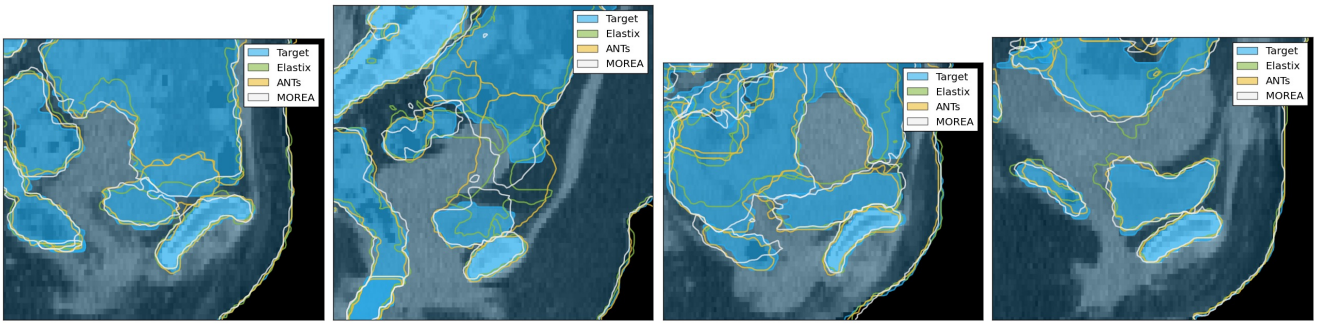
5.3.2 目视检查。除了基于表面的精度分析，我们还会进行目视检查，因为即使轮廓匹配良好的配准，在轮廓之间的区域仍可能存在不理想的形变。检查时，会查看目标图像的切片，并叠加使用配准计算出的正向形变矢量场变换后的源轮廓。为检查轮廓之间的形变，我们还会将整体形变可视化：首先，使用矢量场图呈现形变矢量场本身；其次，在切片上叠加规则网格，并用形变矢量场对其进行形变，以从另一个角度进行观察。

5.4 配准比较

所有方法得出的配准解均使用相同的评估流程进行比较，以确保比较公平。我们将每种方法配置为以正向和逆向形变矢量场的形式输出配准结果，分别定义来源图像和目标图像上的形变。现有方法要么

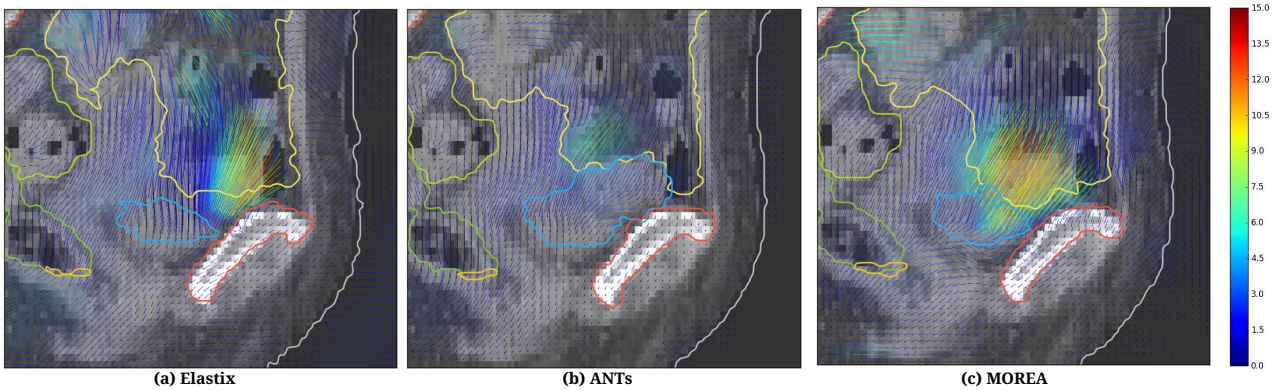
[2]基于官方参数设置数据库：https://elastix.lumc.nl/modelzoo/

[3]依据技术文档：https://github.com/ANTsX/ANTs/wiki/Anatomy-of-an-antsRegistration-call



(a) 患者1 (b)患者2 (c) 患者3 (d) 患者4

图4：每位患者预测的最佳形变示例，以变形轮廓叠加在目标图像上呈现，并显示其参考轮廓（即蓝色的目标轮廓）。表2提供了标注切片，其中显示了所有器官。



(a) Elastix (b) ANTs (c) MOREA

图 5：患者1在所有 3 种方法下选定预测形变的正向变形矢量场和变形轮廓（为便于观察，已降采样）。箭头颜色表示变形幅度，单位为体素（1 体素 = 1.5毫米）。

可以直接配置为输出此类形变矢量场，要么可以间接配置。对于 MOREA，我们使用现有的栅格化方法 [24]，将解中两个变形网格编码的形变栅格化。由于我们要将单目标方法与多目标方法（MOREA）进行比较，因此需要从 MOREA 的近似集中选择解。我们通过从引导目标值最佳的解出发，手动浏览近似前沿，寻找一个在轮廓质量和真实感之间取得良好权衡的解，以此进行事后选择。

我们还使用双侧 Mann-Whitney U 检验（一种标准的非参数检验）进行统计检验，以比较 MOREA 与 ANTs 和 Elastix。采用膀胱轮廓的豪斯多夫距离作为检验度量，因为它描述的是形变最大的器官。为校正两两比较中的多重检验，我们对 α水平应用邦费罗尼校正，并将其从 0.05 降至 0.025。

6 结果与讨论

图4展示了每位患者的细粒度配置实验中选取的结果，以及 MOREA 近似前沿中每位患者的一个解。对于 Elastix，我们分别选取患者1–4中正则化权重为 1.0、1.0、10.0 和 10.0 的运行结果；对于 ANTs，我们选取权重为 0 的所有运行结果。两种现有方法的配置实验完整结果

问题	MOREA 与 Elastix	MOREA 与 ANTs
患者1	**0.011** (+)	**0.007** (+)
患者2	**0.007** (+)	**0.007** (+)
患者3	**0.012** (+)	**0.007** (+)
患者4	**0.007** (+)	0.195 (-)

表 3：各方法间膀胱 Hausdorff 距离两两比较的p值。加号 (+) 表示 MOREA 的均值更优，减号 (-) 则相反。显著结果以高亮显示。

可在附录 B.1.2 和 B.2.2 节中查看。附录 D 中提供了患者1的收敛图，展示了所有方法如何收敛到此处呈现的结果。如第 5.1 节所述，不同扫描图像的难度存在内在差异。总体而言，我们观察到 MOREA 在较难的患者（1–3）上通常优于其他方法；这一点可从图4所示的变形轮廓，以及附录 D 中提供的其他渲染图和分析中直观地看出。

对于**患者1**，我们还在图 5 中绘制了 DVF 切片，展示了针对某一切片中各个区域计算出的变换。我们观察到，Elastix 和 ANTs 返回的形变只会使

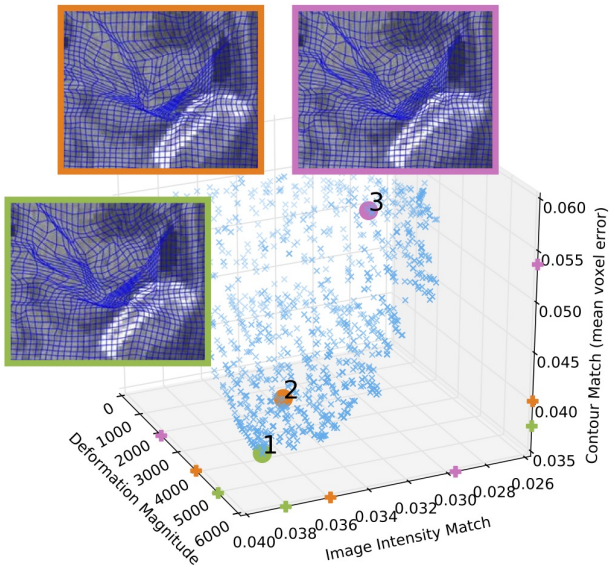


图6：MOREA 在患者1上生成的近似前沿。我们展示了3个放大的配准解。

膀胱的顶部区域发生形变。MOREA 是唯一一种将这种形变分布到整个膀胱的方法；对于这个柔性体积而言，这种形变更为真实。图 6 绘制了 MOREA 在患者1上生成的近似集，其中突显了 3 个形变略有不同的解。这展示了呈现给用户的解的范围，它们都将形变分散到了整个膀胱。

患者2的膀胱体积变化最大，似乎也是最难处理的病例：MOREA 对其形变的建模最为接近（见图4），但代价是肠道也向下移动。可能的原因是源图像中两个器官之间的空间（即间隔）很小。在此，MOREA 的结果揭示了一个影响所有方法的更根本问题：如果两个结构在一幅图像中的间隔很小甚至没有间隔，那么由单一网格构成的变换模型就无法在另一幅图像中将它们分开。**患者3**的膀胱形状变化没有患者2那么严重，但仍给 Elastix 和 ANTs 带来了挑战（见图4）。尤其是后部区域（位于图像中心左侧）与目标不匹配。**患者4**代表一个相对简单的配准问题：膀胱形状变化很小，且膀胱与肠道之间有明显的间隔（见图2）。目视检查表明，在这一问题上，ANTs 和 MOREA 都能很好地拟合膀胱轮廓，而 Elastix 对膀胱和肠道的配准效果都不理想。

对这些结果进行定量考察时，我们对表 3 所列的膀胱豪斯多夫距离进行了显著性检验。在所有患者中，MOREA 变形得到的膀胱轮廓匹配显著优于 Elastix 的轮廓匹配。在 4 个病例中的 3 个病例中，ANTs 对膀胱轮廓的建模精度显著低于 MOREA；在第 4 个病例（患者4）中，结果则无显著差异。附录 D 列出了所有器官的显著性检验结果，这些结果证实了上述趋势，但也表明 MOREA 的豪斯多夫距离有时可能显著高于 ANTs 或 Elastix 的结果。然而，这并不一定意味着配准性能更差，定性分析便说明了这一点。例如，ANTs 得到的患者2乙状结肠变形形状明显偏离（见图4）。然而，其度量值被认为显著优于 MOREA，尽管就整体形状而言，MOREA 更接近目标。

7 结论

本研究独特地汇集了可变形图像配准领域的多条研究方向。我们提出了一种名为 MOREA 的配准方法，该方法兼具基于轮廓和基于图像的特点，采用生物力学模型，并执行多目标优化。这种组合使 MOREA 能够独特地应对具有大幅形变和内容不匹配的复杂三维图像配准问题。MOREA 基于 MO-RV-GOMEA 这一基于模型的进化算法构建，并进行了若干针对问题的扩展，例如 GPU 加速、解修复和对象对齐网格初始化。我们的实验在 4 例宫颈癌患者扫描图像上取得了令人鼓舞的结果：在 4 名患者中的 3 名（代表最困难的病例）中，其轮廓配准精度高于两种最先进的方法。值得注意的是，MOREA 产生的形变似乎比现有方法产生的形变更均匀地分布于各个对象上，因此被认为更加真实。

MOREA获得的解仍存在局部不准确性，因而仍有改进空间，尤其是在器官交界的区域。事实上，本研究结果揭示了可变形图像配准中一个更根本的问题，即典型的可变形图像配准模型无法捕捉局部不连续和内容不匹配。这促使未来开展有关独立器官运动建模的研究，延续近期在该主题上的工作 [35, 38]。MOREA可扩展的生物力学模型或许非常适合进行扩展，以捕捉这些现象。鉴于这种扩展后的方法，开展一项涵盖更多患者并有领域专家参与的更大规模验证研究，可能有助于缩小与临床实践之间的差距。

致谢

作者感谢 W. Visser-Groot 和 S.M. de Boer（荷兰莱顿莱顿大学医学中心放射肿瘤科）对本研究所作的贡献。本研究属于开放技术计划研究项目，项目编号为 15586，由荷兰研究委员会（NWO）、Elekta 和 Xomnia 资助。此外，本研究还获得了荷兰经济事务部知识与创新顶级联盟（TKIs）公私合作伙伴关系补助金的共同资助。

A MOREA 方法的技术实现细节

在本附录中，我们将进一步介绍第4节提出的 MOREA 方法的技术实现细节。

A.1 建模变形幅度

MOREA 的形变幅度目标会针对不同图像区域对非均质弹性进行建模。对于每个四面体 δ，我们通过从特定对象的二值掩膜中采样，确定其下方图像区域的弹性（见图7）。这些掩膜通过填充各对象的轮廓内部（可用作引导）计算得出，从而获得离散的对象分割。我们计算每个对象掩膜与四面体 δ 的重叠部分，由此得到每个对象对应的一个占比。在图7所示的示例中，与该掩膜对应的对象占比为 0.4。随后，将这些对象占比分别乘以针对不同组织类型预先设定的弹性系数，得出 δ的整体单元特定因子。目前，只有骨骼和膀胱设有自定义因子。将 δ的幅度目标函数值乘以该因子，以更准确地模拟使该图像区域发生形变所需的实际能量。

A.2 图像相似度建模

MOREA 的图像强度目标函数定义为特定采样点处强度差平方的求和。要对单个四面体的部分目标函数值进行建模，必须确定要采样哪些图像体素。现有原型 [4] 尝试使用一种受线搜索启发的方法，找出所有中心点位于四面体内部的体素。然而，我们发现，这种将体素离散地关联到四面体的方式会在四面体表面附近产生不理想的行为：有时体素会被关联到多个相邻四面体，或未关联到任何相邻四面体。这种现象可用于提高采样值，却无法提高甚至会降低真实值。

因此，我们在方法中引入一种基于随机采样的方法，对图像空间进行连续采样，并在体素中心之间插值强度值。这种方法也更适合 GPU加速，因为需要暂停执行的决策点更少。我们使用每个四面体的重心坐标系，在其中均匀采样 N个点，

其中， N 由四面体的体积决定。对于每个点，我们采样 4 个随机实数 $r_i \in [0;\ 1]$ 并取 $-\log(r_i)$，以获得均匀分布。然后，我们用它们的总和对坐标进行归一化，确保这些坐标位于四面体内。我们没有使用传统的随机数生成器，而是采用 Sobol序列，使采样点分布得更加均匀。为确保可复现性，我们使用一个根据每个四面体的坐标推导出的种子来初始化其 Sobol序列。因此，对于相同的四面体配置，始终会采样到相同的位置。

A.3 对引导误差进行建模

MOREA 的引导误差目标用于近似解的轮廓匹配。先前的工作 [4] 通过考虑 C_s 中的每个点，并计算其在目标空间中的对应点到集合 C_t中最近点的距离，来衡量轮廓匹配的程度。这需要遍历 C_s中的所有点，

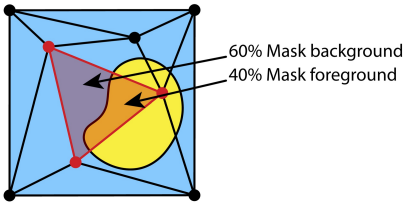
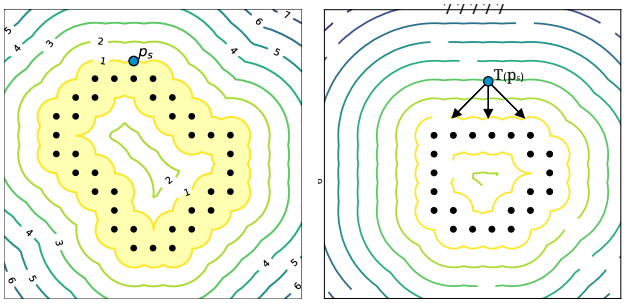


图7：一个四面体单元（此处为红色三角形）与器官掩膜重叠情况的二维示意图。计算得到的重叠比例用于确定该四面体形变幅度的弹性因子。



(a) 源轮廓点集。 **(b)**目标轮廓点集。

图 8：源图像和目标图像中的两个对象轮廓点集，其最小距离图以等值线形式可视化。随机采样的点 p_s 靠近源轮廓，但变换后的 $T(p_s)$ 距离目标轮廓更远。黄色阴影区域表示截断区域，超出该区域的采样点将被丢弃。

确定这些点位于哪个四面体中，并计算该点处的变换。我们提出了一种新的连续引导公式，用于近似逐点距离；初步实验表明，该方法速度更快、鲁棒性更强。

在源图像 I_s上用于强度目标的随机采样过程中，我们还会考虑 C_s距离图上的相同位置，该距离图给出距离源轮廓最近的点（见图 8）。从 C_s距离图中该点处的距离减去 C_t距离图中对应点处的距离，并根据其到源轮廓的距离进行反向加权。距离会截断在每个引导点周围的一个半径范围内，该半径为图像宽度的 2.5%，以确保远距离移动不会影响点集的引导误差。我们根据每个点集中的点数占引导点总数的比例，对其引导误差进行归一化，以抵消对轮廓定义更清晰或尺寸更大的点集的偏向。

A.4 准确检测网格折叠

用于检测约束违反的函数需要具有高精确率（即准确识别所有违规情况）和低延迟（即快速给出结果）。此外，该函数应当是连续定义的，以便该方法还能评估违规的严重程度。这对于修复违规情况的方法而言非常重要。

先前基于网格的三维图像配准研究 [4]采用射线相交法，通过检测一个点是否位于所谓的包围

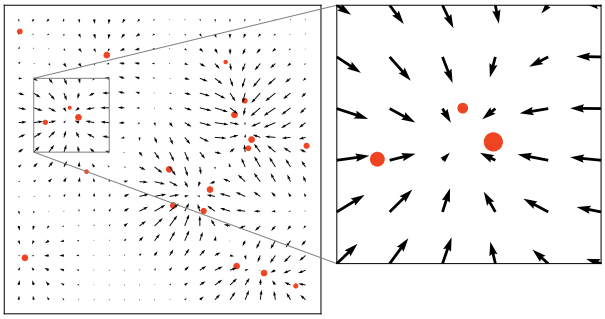


图 9：我们使用径向基函数方法生成解时产生的二维矢量场。红点表示吸引子，其大小表示其权重。

多边形内部来进行判断。初步实验表明，该方法在三维空间中容易出错，会产生假阳性和假阴性。因此，我们基于四面体网格中各四面体的有向体积，开发了一种检测折叠的新方法 [21]。我们计算初始网格配置中每个四面体的有向体积，以建立一组参考符号。当一个点移动时，我们会重新计算所有受其影响的四面体的有向体积，并将其与相应的参考符号进行比较。如果发生折叠，至少一个四面体的符号会翻转。我们利用这一现象检测网格约束违规，并通过违规有向体积的绝对值计算每项违规的严重程度。

A.5 确保初始种群的多样性

即使网格经过了精心初始化，第0代种群的多样性仍然发挥着重要作用 [32]。以往的工作使用一个参考解，并通过围绕每个网格点进行采样、逐渐增大方差来生成随机偏差 [4]。对于低分辨率网格，这种方法可能有效；但对于更高分辨率的网格，它可能导致生成的解中出现许多约束违规（即网格配置发生折叠）。受其他领域中使用径向基函数的方法启发 [47]，我们提出了一种初始化噪声生成方法，可产生不存在约束违规的大变形。该方法在源图像和目标图像上分别放置若干高斯核，并对网格点朝这些核移动的引力进行建模。只要这些力不会导致约束违规，就会以增量轮次的方式施加。图 9 展示了采用这一策略生成的变形矢量场。

B 扩展的问题规范

在本附录中，我们将补充说明本研究使用的配准问题，并详细介绍用于评估和比较配准质量的其他方法。

B.1 其他问题信息

表4列出了所用 CT扫描的层内分辨率。这是每个切片在重采样至 (1.5,1.5)毫米之前的物理分辨率。我们还为每幅医学图像提供了其他视图：对于每位患者，表5列出了每个来源图像和目标图像的两个切片。这提供了有用的补充视角，因为从不同角度观察时，某些运动会更加明显。

B.2 其他评估方法

我们采用四类方法评估每个解，分别基于：(1) 基于表面的配准精度；(2) 使用二维和三维可视化进行目视检查；(3) 基于体积的配准精度；(4) 标志点配准精度。第 5.3 节已介绍第 (1) 类和第 (2) 类方法。这里，我们为第 (1) 类方法补充一种策略，并概述第 (3) 类和第 (4) 类方法。

B.2.1 基于表面的配准精度。除了豪斯多夫距离外，我们的研究还采用豪斯多夫距离的第95百分位数作为另一项指标。它表示一个距离值，使得95%的表面点距离都小于该值。豪斯多夫距离和豪斯多夫距离的第95百分位数均使用 pymia Python 包索引包计算。

B.2.2 基于体积的配准精度。除了表面精度之外，我们还关注图像中各个体积（例如器官、骨骼）的精度。Dice系数是衡量这一点的常用指标，表示重叠体积占总体积的比例。我们使用图像中每个标注对象的二值掩膜，逐体素计算该指标。我们将目标图像对应的二值掩膜与经过计算得到的形变变换后的源图像二值掩膜进行比较。基于与表面评估相同的理由（见第5.3节），评估基于体积的指标时，我们也舍弃相同的边界区域。

B.2.3 标志点配准精度。 优化过程中未提供给算法的一组对应标志点，可用于局部评估配准精度。对于每对标志点，我们使用正向变换将来源标志点变换到目标空间，并将变换后的来源标志点与其对应的目标标志点之间的欧几里得距离作为标志点精度。这是图像配准研究中常用的精度度量 [16, 20], ，但作为整体配准质量的指标，其准确性可能较低，因为标志点被放置在可见的解剖结构上，而这些结构通常移动幅度有限，我们的扫描图像便是如此。

B.3 比较多目标指标

单独来看，各个器官的指标难以得到充分解读，因为器官运动彼此相关，因此会形成权衡。我们通过在同一张平行坐标图中绘制不同器官的得分，将这些权衡可视化，方式类似于 [27]中展示的彩色编码热图比较。这些折线图有助于为决策提供依据；此类决策需要综合考虑各个配准目标的配准质量。

患者	Scan	层内分辨率
患者1	膀胱充盈	(0.86, 0.86)毫米
	膀胱排空	(0.98, 0.98)毫米
患者2	膀胱充盈	(1.04, 1.04)毫米
	膀胱排空	(1.07, 1.07)毫米
患者3	膀胱充盈	(0.98, 0.98)毫米
	膀胱排空	(0.98, 0.98)毫米
患者4	膀胱充盈	(1.04, 1.04)毫米
	膀胱排空	(1。00，1。00)毫米

表4：每次CT扫描的切片在重采样至(1.5,1.5)毫米之前的层内分辨率。

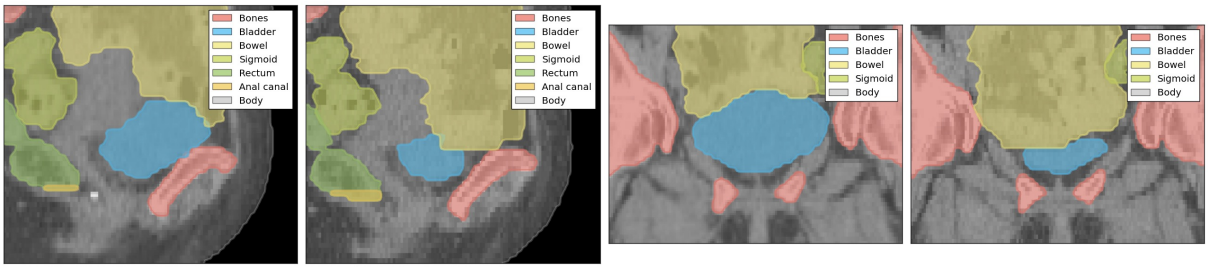
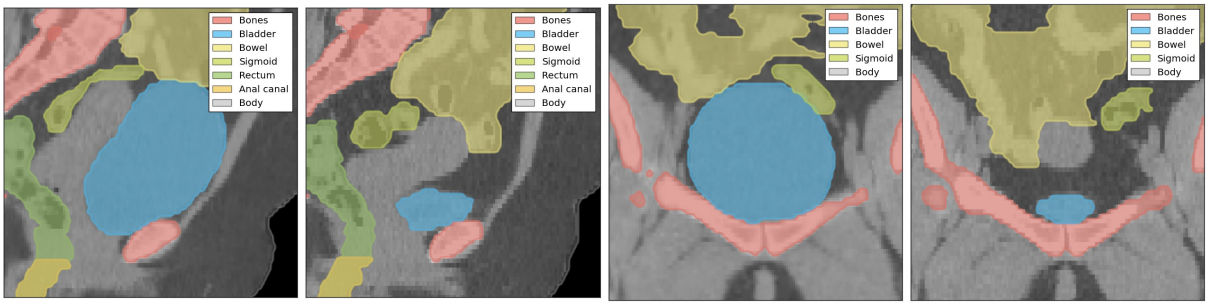
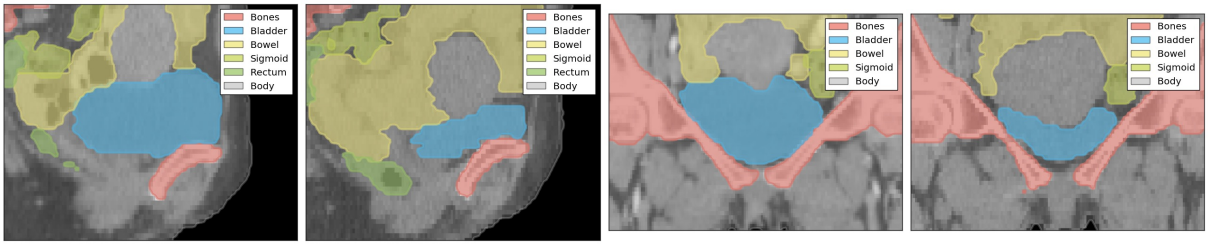
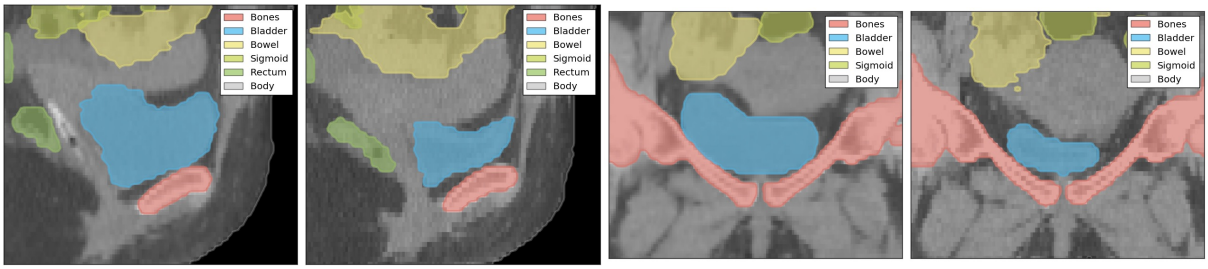
实例	源图像：矢状位	目标图像：矢状位	源图像：冠状位	目标图像：冠状位
患者1				
患者2				
患者3				
患者4				

表5：所有配准　　问题的切片，其中器官已勾画轮廓。矢状位：侧视图；冠状位：前后视图。

C 对比方法的配置

C.1 Elastix

我们使用 Elastix 5.0.0版。根据 Elastix模型库4 中的参数设置，我们采用多分辨率Elastix配准，以多种配置处理配准问题，并尝试为每个问题找到最优配置（参数文件见 C.1.3 节）。受一种采用组间方法在 Elastix 中实现对称配准的方法 [6],启发，我们还测试了一种对称变体，该变体将两幅图像都配准到一个公共图像中间空间。对于所有设置，我们都放宽了收敛要求，将每个分辨率下的迭代次数增加到 10,000 次，显著高于（5 倍于）大多数参考文件中给定的计算预算。这样做是为了让 Elastix 有充分机会对其中存在的大幅变形进行建模。我们还通过将图像采样点数量从常用的 10,000 个增加到 20,000 个，来稳定优化过程。尽管这会增加计算复杂度，但应能提高优化过程中内部使用的图像强度近似的准确性，并使计算梯度更加可靠。

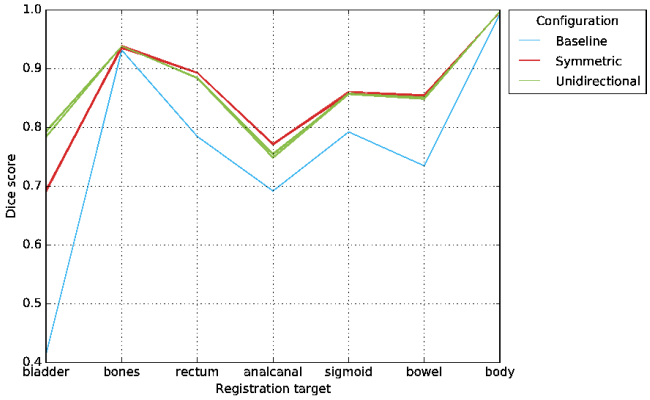
Elastix 默认计算逆向变换，即定义在固定（目标）空间并指向移动（来源）空间的形变矢量场。为计算正向变换（将标注从移动（来源）空间变换到固定（目标）空间所需的变换），我们使用给定的参数文件和已计算出的变换作为初始变换重新运行配准，同时将度量替换为*DisplacementMagnitudePenalty*度量。这实际上求得了已计算逆向变换的正向变换。从参数文件中移除初始变换指针，单独导出此正向变换，即可得到所需的形变矢量场。

Elastix 不支持优化目标轮廓匹配，而 MOREA 方法通过引导目标进行优化。为确保比较公平，我们尝试将这些信息作为一对复合掩膜图像输入，以隐式传递轮廓信息。每张掩膜图像由每次扫描可用的不同二值目标掩膜合并而成，并为每个对象分割赋予不同的均匀强度值。在启用此功能的运行中，我们会先对这些预处理后的复合掩膜进行配准，再进行 CT图像配准。

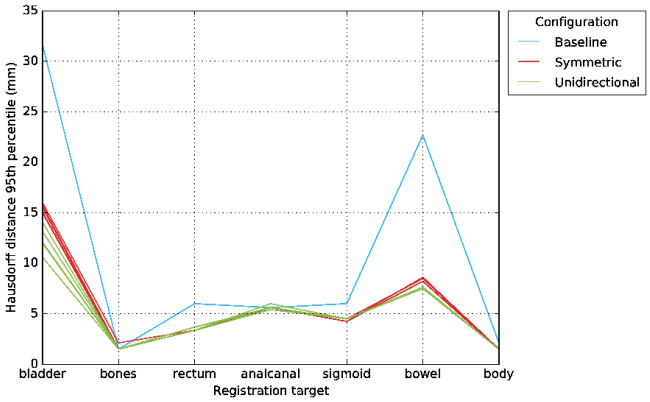
C.1.1 粗粒度配置实验。首先，我们在患者1上进行一组初步运行，以确定适合此问题类型和难度的基础配置。我们探究配准方向（单向配准与对称配准）以及是否采用复合掩膜配准步骤的影响（假设正则化权重为0.001），以便让 Elastix 能够灵活处理大变形（较大的形变幅度权重可能会阻碍大变形）。

在图 10 中，我们绘制了 Elastix 使用对称配准和单向配准时的性能，并报告两种不同的指标（Dice 分数和 Hausdorff距离的第95百分位数）。我们观察到，就 Dice 分数而言，除直肠和肛管外，单向配准的表现通常与对称配准相当或更好。由于相对

[4]https://elastix.lumc.nl/modelzoo/

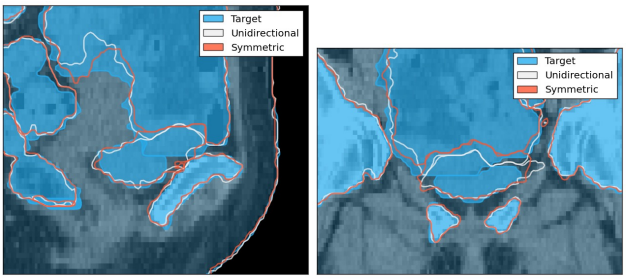


(a) Dice 分数。



(b) 豪斯多夫距离的第95百分位数。

图10：Elastix 中对称配准与单向配准在多次运行中的比较。刚性配准后的基线分数以蓝色绘制。

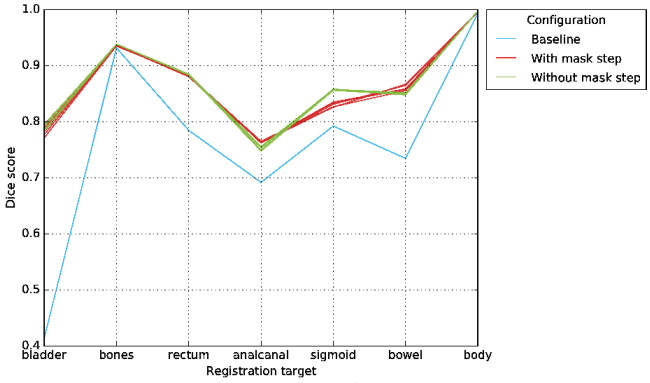


(a) 矢状切片。 **(b) 冠状切片。**

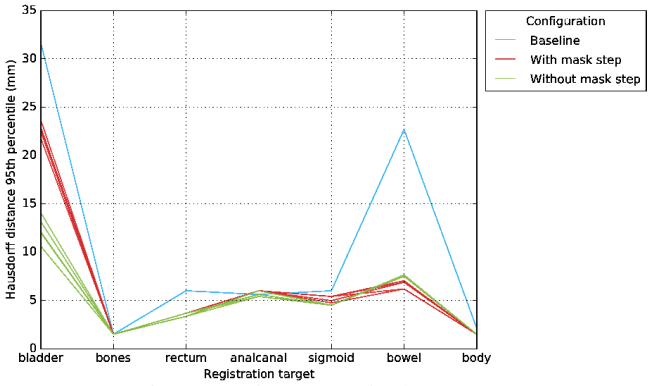
图 11：使用单向配准和对称配准的 Elastix 配置所预测形变的渲染图，未进行掩膜配准步骤。

膀胱（形变最明显的器官）的性能提升幅度很大，因此我们在此处选择单向配准。图 11 的目视检查结果支持这一选择：冠状切片中膀胱的配准效果略好。

现在，我们转而使用复合掩膜配准步骤，尝试通过简化输入给 Elastix 的信息来获得更大的形变。图 12 展示了相同的指标，但

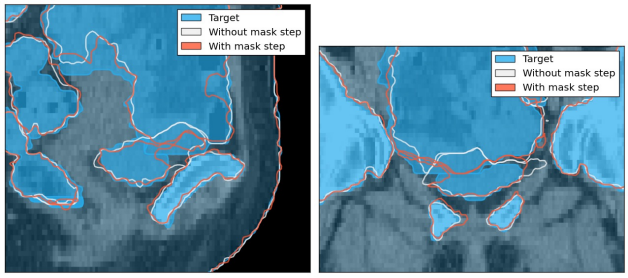


(a) Dice 分数。



(b) 豪斯多夫距离的第95百分位数。

图 12：比较 Elastix 中采用和未采用复合掩膜配准步骤的单向配准，针对多次运行。刚性配准后的基线分数以蓝色绘出。



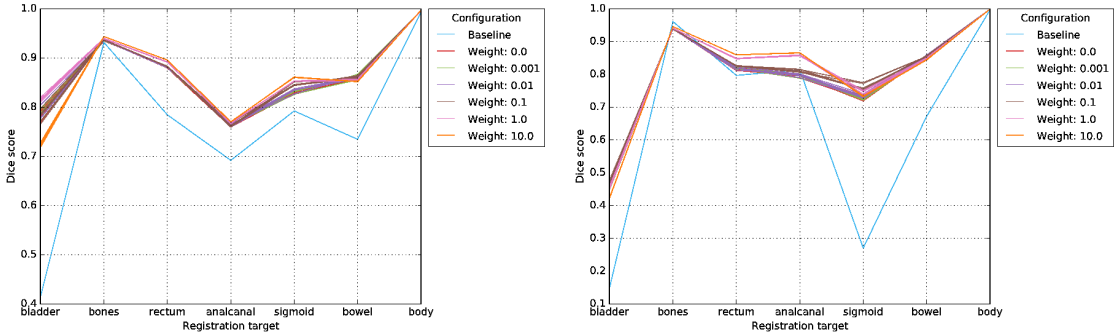
(a) 矢状切片。 **(b) 冠状切片。**

图13：使用单向配准时，采用和未采用复合掩膜配准步骤的 Elastix 配置所预测形变的可视化渲染图。

比较了使用和不使用该步骤的情况（同时采用单向配准）。结果并未显示哪种方法明显更优，因为使用掩膜的配置通常 Dice 分数更高，但不使用掩膜的配置的 Hausdorff 第95百分位数更低。图 13 表明，添加掩膜步骤改善了对膀胱基底区域的建模，但中间区域只是向侧面收缩，顶部区域并未向下移动，因此产生的形变不符合解剖学实际。尽管如此，我们还是选择这一版本，而非不使用掩膜配准步骤的版本，因为添加该步骤后，更接近于对所需大幅形变的建模。

C.1.2 患者级细粒度配置实验。对于每位患者，我们尝试以指数级递增的正则化权重；这种指数级正则化权重扫描也用于类似的研究 [8]。每位患者的 Dice 分数见图 14，豪斯多夫距离的第95百分位数见图 15。每个问题的渲染图见图 16–19。

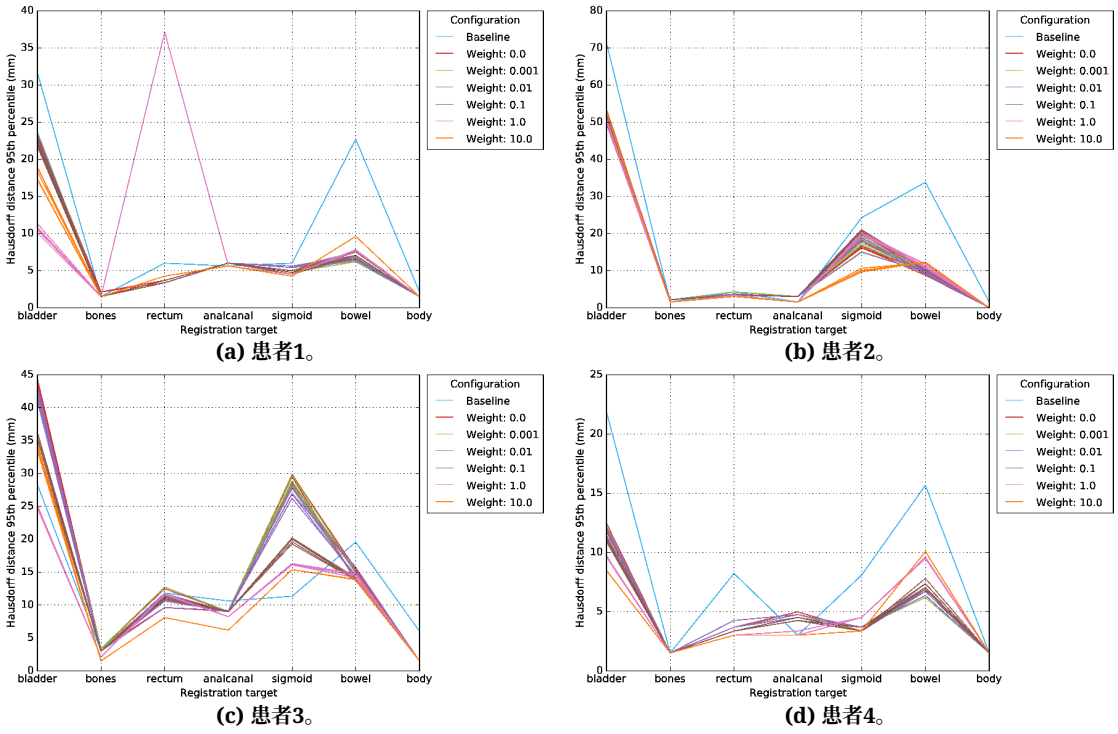
我们观察到，不同配准问题的最佳正则化权重差异很大。对于患者1的扫描图像（图 16），在尝试的设置中，权重为 1.0 时效果最佳；而患者3的扫描图像（图 18）似乎在权重为 10.0 时效果更好。



(a) 患者1。 (b) 患者2。

(c) 患者3。 (d) 患者4。

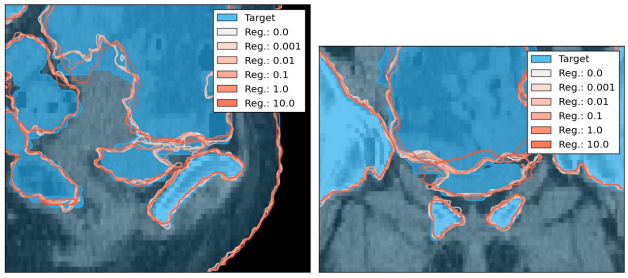
图14：Elastix 中逐患者细粒度配置运行的 Dice分数。刚性配准后的基线分数以蓝色绘制。



(a) 患者1。 (b) 患者2。

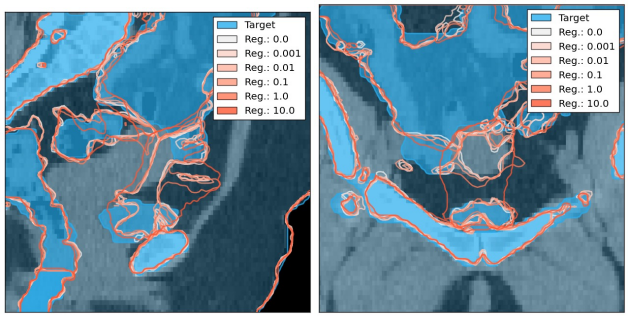
(c) 患者3。 (d) 患者4。

图15：Elastix中按患者进行的细粒度配置运行的Hausdorff第95百分位数。刚性配准后的基线分数以蓝色绘制。



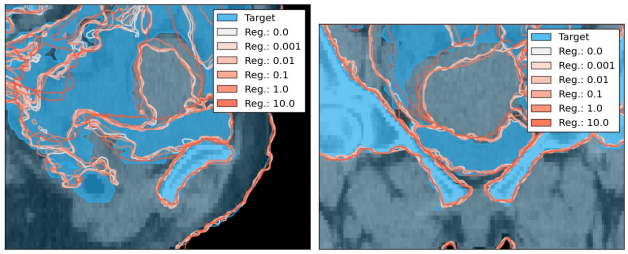
(a) 矢状切片。 **(b)** 冠状切片。

图16：Elastix 在患者1上使用不同正则化权重预测的形变渲染图。



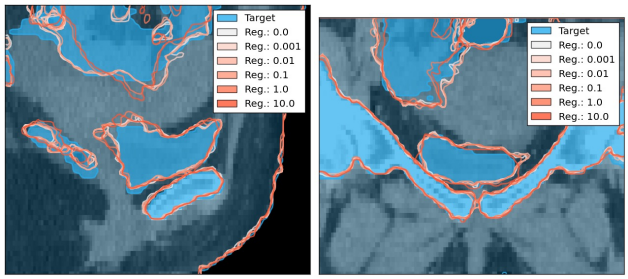
(a) 矢状切片。 **(b)** 冠状切片。

图17：Elastix 在患者2上使用不同正则化权重预测的形变渲染图。



(a) 矢状切片。 **(b)** 冠状切片。

图18：Elastix 在患者3上使用不同正则化权重预测的形变渲染图。



(a) 矢状切片。 **(b)** 冠状切片。

图19：Elastix 在患者4上使用不同正则化权重预测的形变渲染图。

C.1.3 参数文件。 下文列出了我们用于不同Elastix配准变体的参数文件。以 $ 字符开头的标记表示在我们将文件传递给Elastix之前解析的变量（例如，我们在每次重复时递增的随机种子）。

清单1：传统单向形变的正向变换参数。

```
// 图像类型
(固定图像像素类型       "short")
(固定图像维度 3)
(移动图像像素类型 "short")
(移动图像维度 3)

// 多分辨率 (配准 "多指标多分辨率配准") (变换组合方式 "复合
") (直方图箱数 32) (分辨率数 4) (最大迭代次数 10000)



// 优化器 (优化器 "自适应随机梯度下降") (自动参数
估计 "true") (使用自适应步长 "true") (检查样本数 "
true") (使用方向余弦 "true") (随机种子 $random_
seed)


// 度量 (度量 "高级 Mattes 互信息" "变换弯曲
能量惩罚") (度量0权重 1.0) (度量1权重 $
regularization_weight)


// 组件 (固定图像金字塔 "固定图像平滑金字塔") (移动图
像金字塔 "移动图像平滑金字塔") (插值器 "B样条插值器")
( 重采样插值器 " 最终 B 样条插值器 ") ( 重采样器 " 默认
重采样器") (变换 "B样条变换")


// 变换 (物理单位下的最终网格间距 2.0)

// 采样 (图像采样器 "随机坐标") (每
次迭代生成新样本 "true") (空间采样
数 20000)

// 插值与重采样 (B样条插值阶数 1) (
最终B样条插值阶数 3) (默认像素值
0)

// 输出及其他（WriteTransformParametersEachIteration "false" "f
alse" "false" "false" "false"）
(WriteTransformParametersEachResolution "true" "true" "true" "t
rue" "true") (ShowExactMetricValue "false" "false" "false" "false" "
false") (WriteResultImageAfterEachResolution "false")
(WriteResultImage "true") (ResultImagePixelType "short")
(ResultImageFormat "nii.gz")
```

列表 2：对称形变的正向变换参数。

```
// 图像类型（FixedImagePixelType "short"）
(FixedInternalImagePixelType "short") (FixedImageDimension 4)
(MovingImagePixelType "short") (MovingInternalImagePixelType "
short") (MovingImageDimension 4) // 多分辨率（Registration "
MultiResolutionRegistration"）(HowToCombineTransforms "
Compose") (NumberOfHistogramBins 32) (NumberOfResolutions 4)
(MaximumNumberOfIterations 10000)
(MaximumNumberOfSamplingAttempts 10) // 优化器（Optimizer "
AdaptiveStochasticGradientDescent"）
(AutomaticParameterEstimation "true") (UseAdaptiveStepSizes "true")
(CheckNumberOfSamples "true") (UseDirectionCosines "true")
(RandomSeed \$random_seed) // 度量（Metric "$metric"）
(NumEigenValues 1) (TemplateImage "ArithmeticAverage" "
ArithmeticAverage") (Combination "Sum" "Sum") (SubtractMean "
true") (MovingImageDerivativeScales 1.0 1.0 1.0 0.0) // 组件（
FixedImagePyramid "FixedSmoothingImagePyramid"）
(MovingImagePyramid "MovingSmoothingImagePyramid")
(ImagePyramidSchedule 8 8 8 0 4 4 4 0 2 2 2 0 1 1 1 0) (Interpolator "
ReducedDimensionBSplineInterpolator") (ResampleInterpolator "
FinalReducedDimensionBSplineInterpolator") (Resampler "
DefaultResampler") (Transform "BSplineStackTransform") // 变换（
FinalGridSpacingInPhysicalUnits 2.0）// 采样（ImageSampler "
RandomCoordinate"）(NewSamplesEveryIteration "true")
(NumberOfSpatialSamples 20000) // 插值与重采样（
BSplineTransformSplineOrder 1）(FinalBSplineInterpolationOrder
3) (DefaultPixelValue 0) // 输出及其他（
WriteTransformParametersEachIteration "false" "false" "false" "false")
(WriteTransformParametersEachResolution "true" "true" "true" "t
rue"
                                                                    )
(显示精确度量值 "false" "false" "false" "false") (每个分辨率
后写入结果图像 "false") (写入结果图像 "true") (结果图像像
素类型 "short") (结果图像格式 "nii.gz")
```

C.2 ANTs SyN

我们使用 2.4.2 版 ANTs SyN 算法。我们以 antsRegistrationSyN.sh 脚本为基础构建配准命令，并根据问题进行定制（运行命令见 C.2.3 节）。遵循官方建议[5]，我们认为以下设置在本问题中可以调整：（1）互相关度量使用的区域半径；（2）是否在配准过程中将复合掩膜作为额外的图像模态通道；（3）使用的梯度步长；（4）时间步之间局部形变的正则化权重；（5）总形变的正则化权重。我们为患者1配置前四个参数，然后针对每位患者分别配置第五个参数。

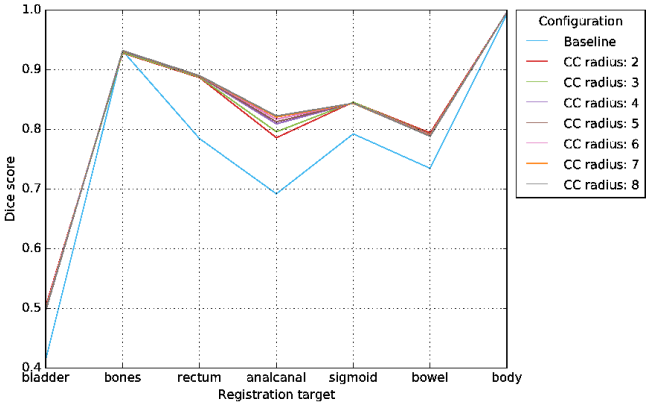
在我们的设置中，与指南相比，我们放宽了收敛限制，以允许更长时间、并希望能更准确地进行配准。在指标方面，我们没有使用手册中提到的点集配准度量，因为手册指出，该度量目前在 ANTs SyN 中尚不受支持。

我们发现，即使采用随机采样策略，ANTs SyN 的随机种子对使用交叉相关（CC）度量进行配准的结果也没有任何影响。当前版本似乎是完全确定性的，但没有将随机种子纳入考虑，因此无论种子为何，总是产生相同的输出。这是有问题的，因为我们希望获得多个输出，以揭示配准方法对略有变化的输入的反应。为了缓解配准确定性缺乏控制的问题，我们使用非常小的（确定性随机）增量，对 σ 平滑因子进行了轻微扰动（见列表 3）。 Δ_3 服从正态分布，并限制在 $[-0.1, 0.1]$ 和 Δ_2 之间； $[-0.05, 0.05]$限制在 Δ_1 和 $[-0.01, 0.01]$ 之间。

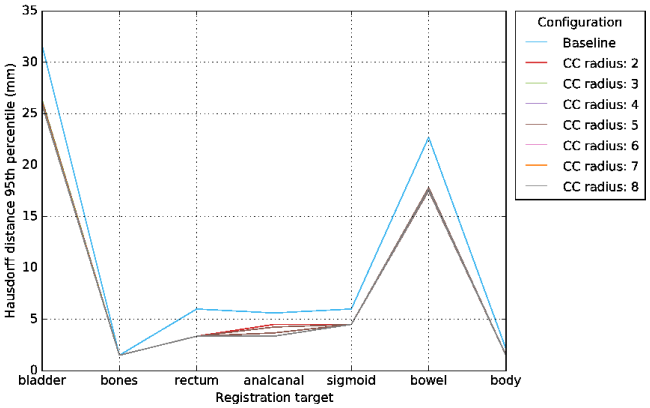
C.2.1 粗粒度配置实验。我们使用 ANTs SyN 算法，在患者1上进行了一组初步的粗粒度配置实验。官方推荐的设置作为基线：互相关半径为 4 个体素，梯度步长为 0.1，仅对图像本身进行配准（不使用额外通道），更新正则化权重为 3.0。对于每项设置，我们都尝试了相对于基线的不同偏差。

互相关半径首先，我们研究了不同互相关半径的影响。较大的值应该能提高配准精度，因为计算样本的互相关时会考虑更多上下文信息。图 20 证实了这一预期，不过总体影响不大。大多数器官的分数变化不大，但增大半径后，肛管的 Dice 分数有所提高。我们观察到此处的收益递减，例如，将半径从 7 增至 8 只能带来轻微改善。尽管如此，我们还是决定采用测试过的最大设置（8 个体素，即临床问题中为 12毫米），因为这一设置效果最好，而且本研究对配准没有时间限制。图 21 的可视化呈现展示了这一设置的视觉影响，影响可以说较为有限。

[5]https://github.com/ANTsX/ANTs/wiki/Anatomy-of-an-antsRegistration-call

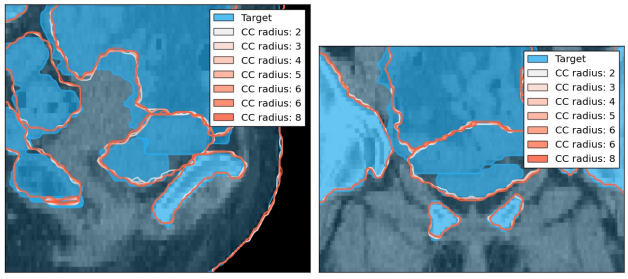


(a) Dice 分数。



(b) 豪斯多夫距离的第95百分位数。

图 20：使用不同区域半径的 ANTs 互相关度量进行配准的比较。刚性配准后的基线分数以蓝色绘出。

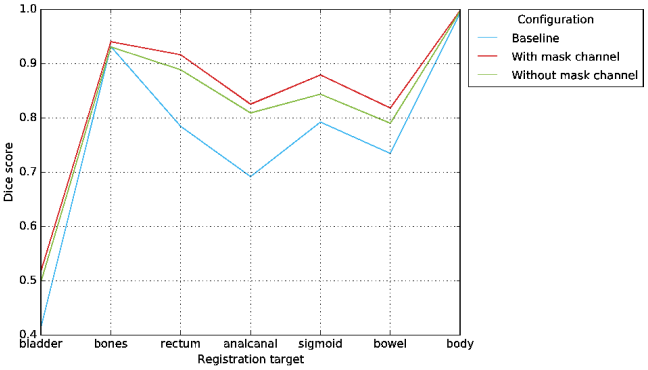


(a) 矢状切片。 **(b) 冠状切片。**

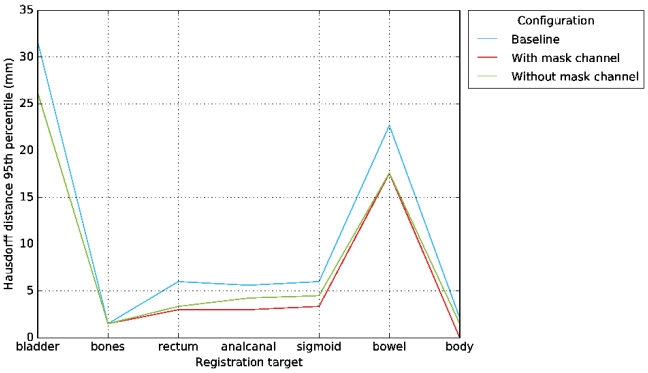
图 21：使用不同 CC 半径的 ANT 配置预测的形变渲染图。

复合掩膜通道 其次，我们探讨了在配准过程中加入复合掩膜图像通道的影响。图22的证据表明，加入掩膜通道有助于提升所有器官配准的 Dice 分数。图23中只能略微看出性能差异，但度量值的差异促使我们决定在接下来的患者特异性配置步骤中使用掩膜通道。

梯度步长第三，我们考察采用不同梯度步长对 ANTs 配准性能的影响。

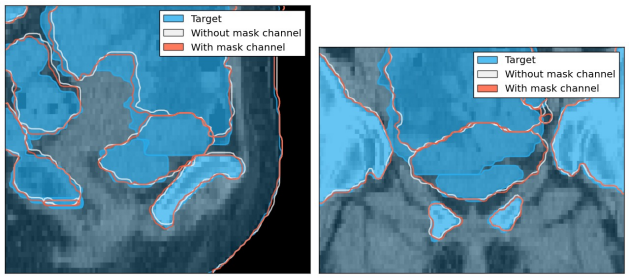


(a) Dice分数。



(b) 豪斯多夫距离的第95百分位数。

图22：ANTs 中使用和不使用复合掩膜通道时的配准比较。刚性配准后的基线分数以蓝色绘出。



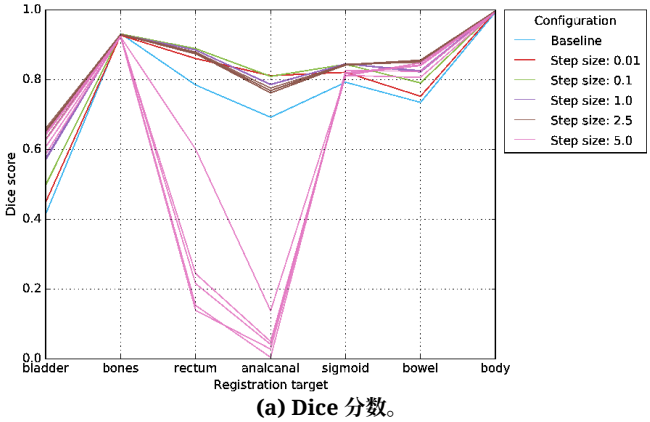
(a) 矢状切片。

(b) 冠状切片。

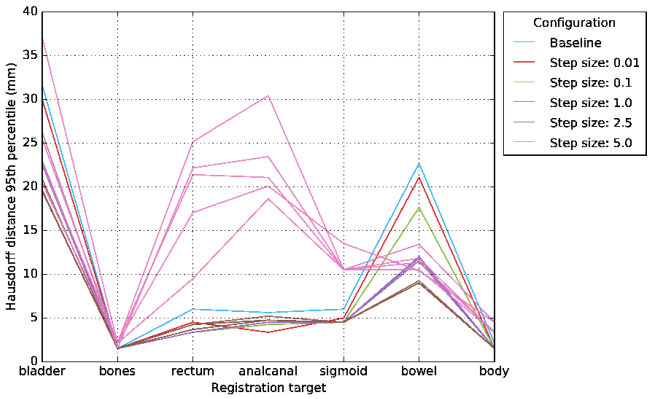
图23：使用和不使用复合掩膜通道的 ANT 配置所预测形变的可视化渲染图。

在 ANTs 中，时间点之间采用较大的步长可能会使较大的形变变得可行，因为优化较不容易陷入局部极小值。图24表明，选择大于建议值 0.1 的步长可能有益，其中 1.0 能为不同器官提供良好的权衡。较大的步长（如 5.0）会导致算法越过目标，并使多个器官发生剧烈形变，这一点可从轮廓渲染图（图25）中看出。我们选择 1.0 作为梯度步长，因为它能在各项性能目标之间实现良好的权衡。

更新正则化权重 最后，我们使用前面三轮扫描推导出的设置，测试哪种更新正则化

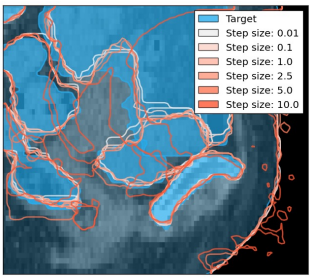


(a) Dice 分数。

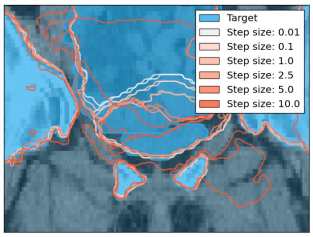


(b) 豪斯多夫距离的第95百分位数。

图24：不同时间点之间采用不同梯度步长的 ANTs 配准对比。刚性配准后的基线分数以蓝色绘出。



(a) 矢状切片。

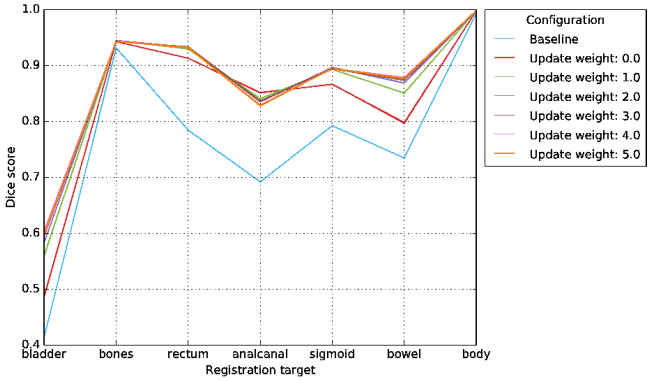


(b) 冠状切片。

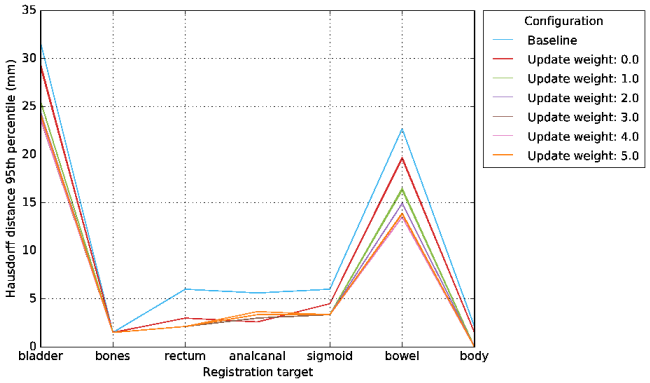
图25：不同梯度步长的 ANT配置预测的形变可视化渲染图。

权重的效果最好。图26显示，权重为4.0时两项指标的总体表现均最佳。从视觉效果来看，图27表明，权重为4.0和5.0时的配准结果最好，二者之间几乎没有明显差异。综合视觉和定量结果，我们为患者特定配置实验选择4.0作为更新正则化权重。

C.2.2 患者级细粒度配置实验。我们针对所有问题实例尝试了呈指数递增的总正则化权重。图28和图29绘出了每个问题实例的 Dice 分数和 Hausdorff第95百分位数，图30–33则展示了

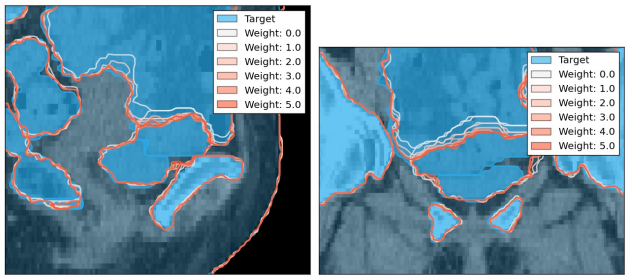


(a) Dice分数。



(b) 第95 p百分位的豪斯多夫距离。

图26：比较采用不同更新正则化权重的 ANTs 配准在各时间点之间的结果。刚性配准后的基线分数以蓝色绘出。



(a) 矢状切片。 **(b) 冠状切片。**

图27：不同更新正则化权重下，ANT配置预测的形变渲染图。

ANTs 为这些实例预测的变形轮廓渲染图。我们观察到，正则化对所有考察案例的性能都有显著影响，但通常（相对）较好的结果仍是在不使用正则化的情况下获得的。图30–32显示，无论是否进行正则化，ANTs 都无法对膀胱及其周围器官发生的大幅形变进行建模。Dice 分数和豪斯多夫指标的结果印证了这些观察。在图33中，ANTs 展示了其能够较为精确地对膀胱形变进行建模，但需要注意的是，从形态学角度来看，这也是最简单的问题。

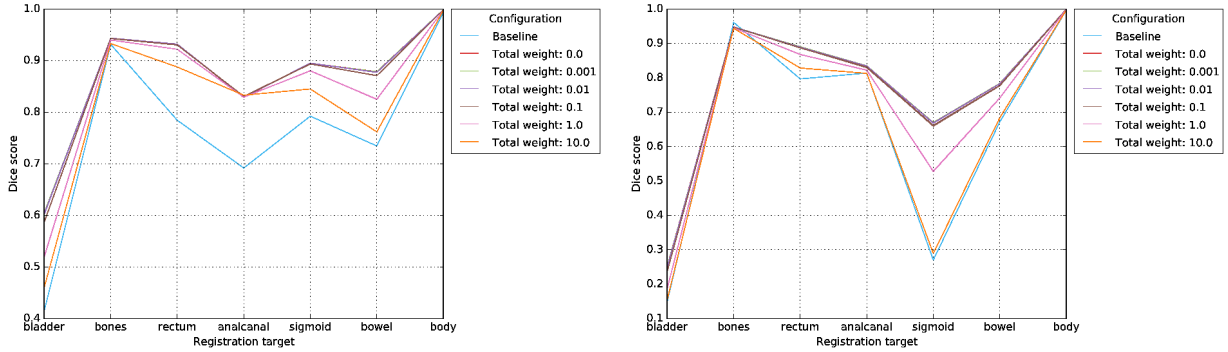
C.2.3 运行命令。 我们列出用于通过 ANTs 进行配准的两条命令。以 $ 字符开头的标记表示变量，这些变量会在执行命令前解析。请注意，尽管命令中提供了随机种子，但它并不起作用，也不会改变输出。

列表3：用于多变量配准的 ANTs 配准命令（使用 复合掩膜。

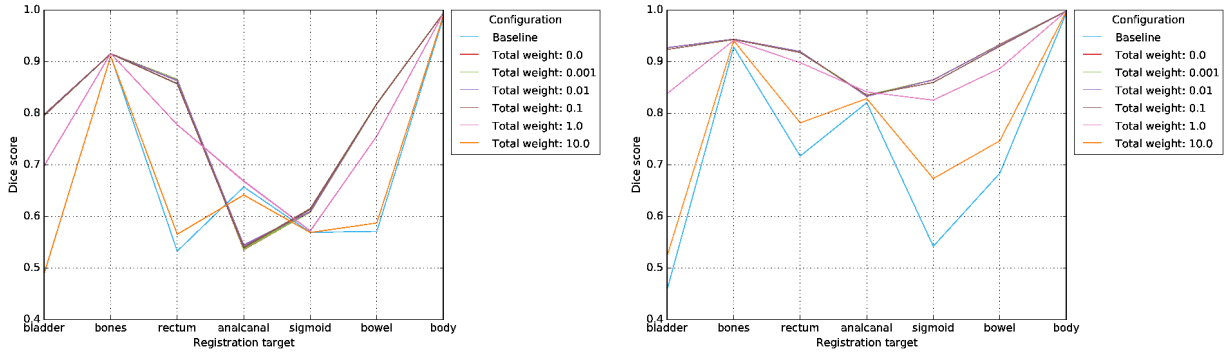
```
$ANTSPATH/antsRegistration
    --verbose 1
    --random-seed $random    _seed
    --dimensionality 3
    --float 0
    --collapse-output-transforms 1
    --output [ , Warped.nii.gz,    InverseWarped.nii.gz ]
    --interpolation 线性
    --use-histogram-matching 0
    --缩尾处理图像强度 [ 0.005,                0.995 ]
    --初始移动变换 [$fixed_composite_mask,
          $moving_复合_掩膜,  1 ]
    --transform SyN[$gradient_步长，_
          $update_正则化_权重,
          $total_正则化_权重 ]
    --metric 互相关[ $fixed_复合_掩膜,     $moving_复合_掩膜，1,
          $互相关_半径_ ]
    --度量 互相关[ $固定_图像,   $移动_图像,     1，
          $互相关_半径_ ]
    --收敛 [ 2000x1000x500x250,          1e-6,  10 ]
    --shrink-factors 8x4x2x1
    --smoothing-sigmas {3+delta_3}x{2+delta_2}x{1+delta_1}x0vox
```

列表 4：不使用复合掩膜进行多变量配准的 ANTs 配准命令。

```
$ANTSPATH/antsRegistration --verbose 1 --random-seed $rando
m_seed --dimensionality 3 --float 0 --collapse-output-transforms 1 --
output [ , Warped.nii.gz, InverseWarped.nii.gz ]--interpolation
线性 --use-histogram-matching 0 --winsorize-image-intensities [ 0.
005, 0.995 ]--初始移动变换 [$fixed_图像, $moving_图像, 1 ]--变换
SyN[$gradient_步长_大小, $update_正则化_权重, $total_正则化_权重 ]--
度量 互相关[ $fixed_图像, $moving_图像, 1, $交叉_相关_半径 ]--收敛
[ 2000x1000x500x250, 1e-6, 10 ]--shrink-factors 8x4x2x1 --
smoothing-sigmas {3+delta_3}x{2+delta_2}x{1+delta_1}x0vox
```

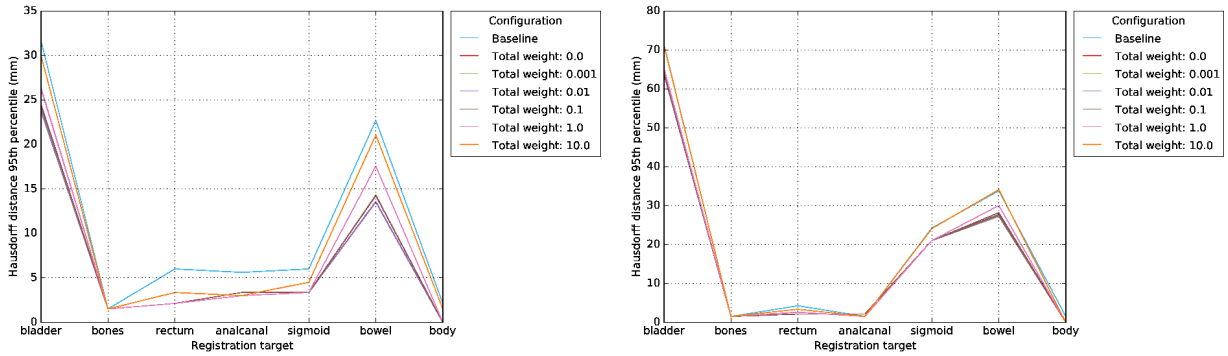


(a) 患者1。 **(b) 患者2。**

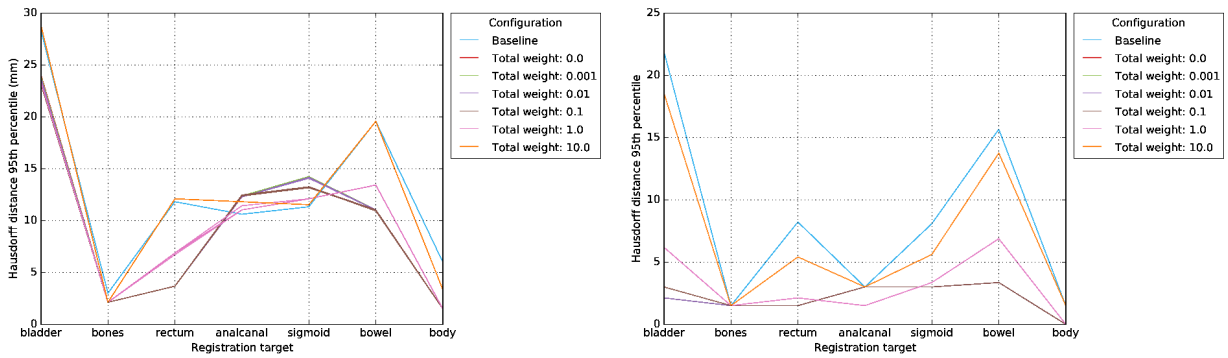


(c) 患者3。 **(d) 患者4。**

图28：ANTs 中逐患者细粒度配置运行的 Dice 分数，蓝色表示刚性配准后的基线。

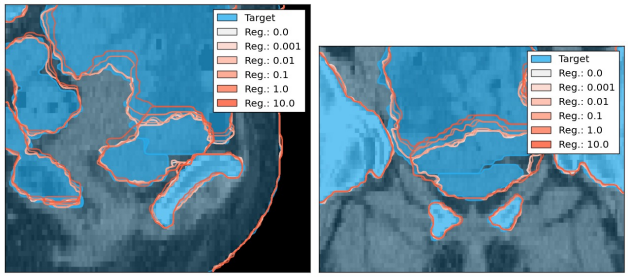


(a) 患者1。 **(b) 患者2。**



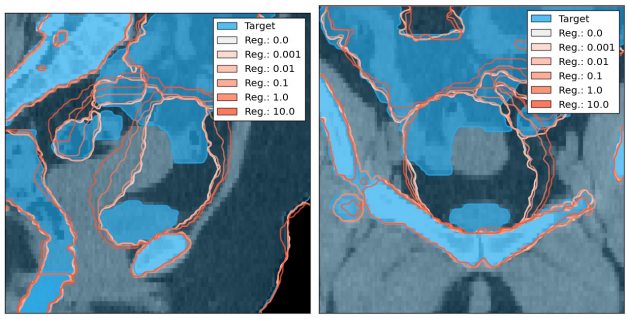
(c) 患者3。 **(d) 患者4。**

图29：ANTs 中逐患者细粒度配置运行的 Hausdorff第95百分位数，其中刚性配准后的基线以蓝色显示。



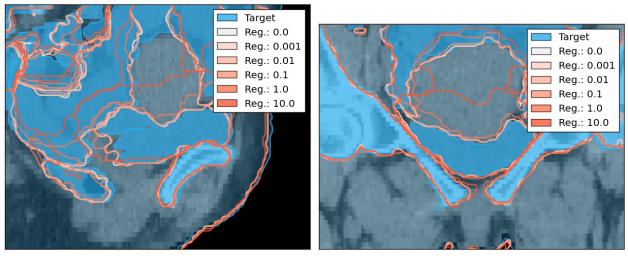
（a）矢状切片。　　（b）冠状切片。

图30：ANTs 在患者1上使用不同总正则化权重预测的形变渲染图。



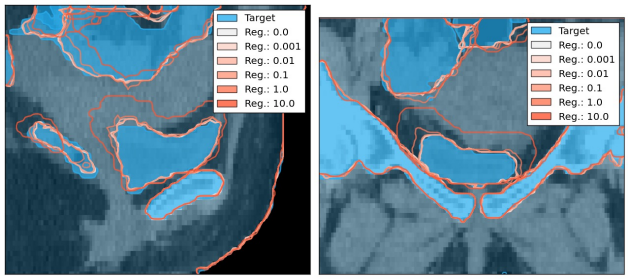
（a）矢状切片。　　（b）冠状切片。

图31：ANTs 在患者2上使用不同总正则化权重预测的形变渲染图。



(a) 矢状切片。　　**(b)** 冠状切片。

图32：在患者3上，ANTs 使用不同总正则化权重预测的形变渲染图。



(a) 矢状切片。　　**(b)** 冠状切片。

图33：在患者4上，ANTs 使用不同总正则化权重预测的形变渲染图。

C.3 本文工作：MOREA

我们介绍了使用 MOREA 在患者1上进行的若干粗粒度配置实验。根据这些实验得出的基础参数文件见 C.3.2节。由于 MOREA 是一种多目标方法，我们不进行细粒度配置步骤。

对于MOREA的引导目标，我们对每个扫描执行额外的预处理步骤，以解决不同维度之间分辨率不一致的问题。将每个扫描重采样至1.5毫米 的均匀体素分辨率的初始重采样步骤，导致层间维度被过度采样（原始切片间距为3毫米 ）。轮廓标注仅置于切片上，这意味着在原始切片之间通过重采样至1.5毫米而新增的切片不具有轮廓信息。对象轮廓中的这些切片“间隙”可能会在优化过程中被利用。我们通过一个中间步骤来解决这一问题，即构建跨切片的每个对象的三维模型，并从该模型生成边界点。

C.3.1 粗粒度配置实验。

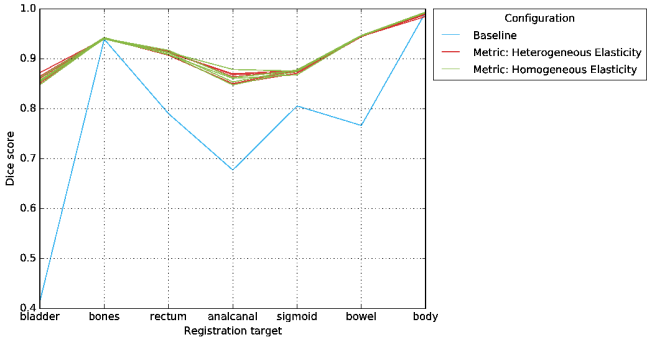
异质弹性 在第4.1节中，我们描述了一种模型，该模型能够在形变幅度目标中捕捉不同组织类型的生物力学属性。该生物力学模型的核心原则是，为图像空间中与存在的对象（例如器官和骨骼）相对应的不同区域赋予非均质弹性。在这项首次配置实验中，我们将该模型的性能与先前工作 [4], 所使用的、假设图像空间具有均匀弹性的模型的性能进行了比较。本实验未使用身体轮廓，而后续实验均使用了该轮廓。

图34中的度量结果表明，异质模型通常能获得更高的Dice分数，而Hausdorff第95百分位数相近。图35展示了使用异质模型和均质模型得到的部分解的渲染图，证实了这一趋势。在两个切片中，我们都观察到，异质弹性尤其能改善膀胱形变的性能，这可能是因为该模型赋予膀胱更高的弹性。

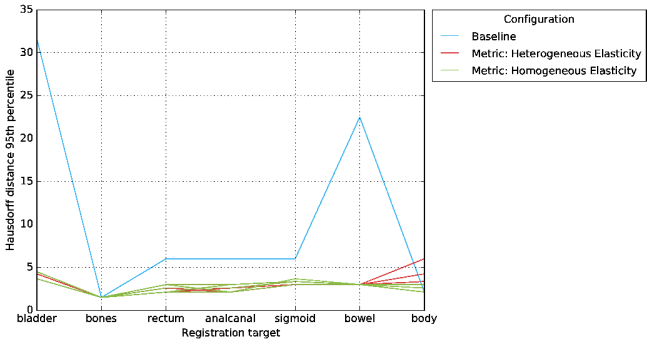
网格生成 基于上一小节实验所涵盖的生物力学模型，我们现在研究不同网格点布置策略的影响。第 4.2.1 节介绍了如何根据这些点生成网格。

在本实验中，我们比较随机（基于Sobol序列）布点方式与两种基于轮廓的策略的效果：一种是针对每条轮廓采样，另一种则对膀胱表面进行特殊处理。图36a显示，最后一种策略对膀胱的建模效果最佳；总体而言，在各个器官上，基于轮廓的策略都优于随机布点。图37中的渲染图表明，随机布点方法能够模拟整体形变，但粒度过粗，无法准确处理特定器官和骨骼部位的细节。两种基于轮廓的策略表现都不错，但在膀胱表面附近，采用特殊表面约束的策略效果尤为出色。

提供引导信息 MOREA 所延续的多目标配准方法系列，可以包含第三个

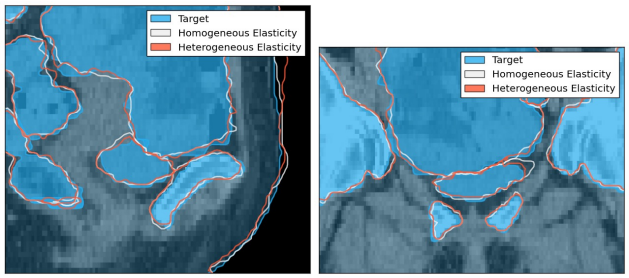


(a) Dice分数。



(b) 第95 p百分位豪斯多夫距离。

图34：比较在 MOREA 的形变幅度目标中使用异质弹性与先前使用均质弹性模型的效果，结果基于多次运行。刚性配准后的基线分数以蓝色绘出。

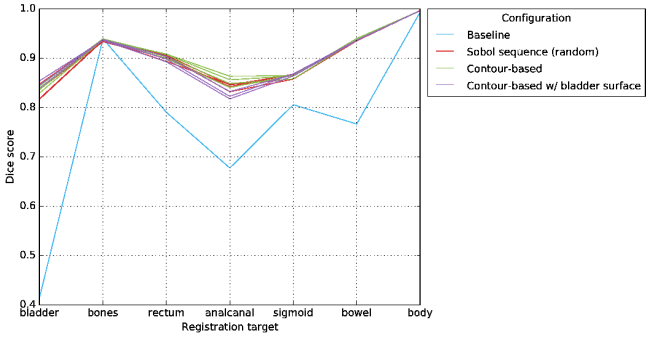


(a) 矢状切片。 **(b) 冠状切片。**

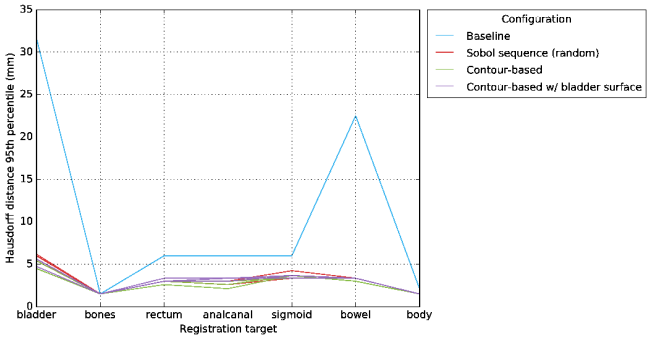
图35：MOREA 使用非均质弹性形变模型和均质模型预测的形变渲染图。

目标，用于衡量引导信息（轮廓）的匹配程度。在本实验中，我们评估这一目标对配准质量的影响。

图38中的定量结果几乎毫无疑问地表明，采用引导目标对于建模大幅形变至关重要。没有引导目标时，膀胱基本保持原位，如图39所示。在这个问题中，图像信息似乎不足以引导优化。

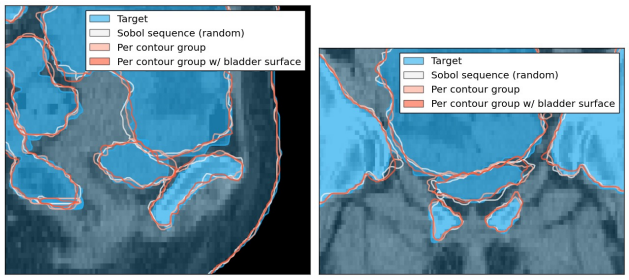


(a) Dice分数。



(b) 豪斯多夫距离的第95百分位数。

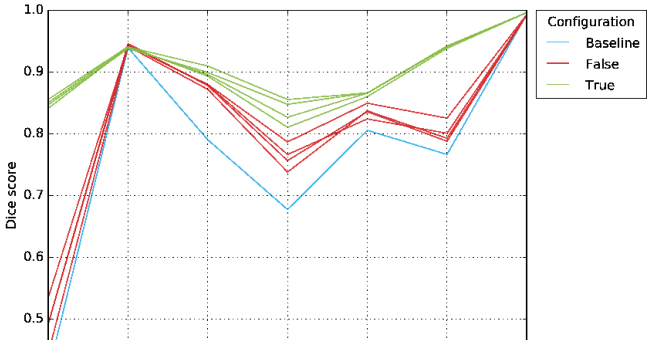
图36：多次运行中不同网格点布置策略的比较。刚性配准后的基线分数以蓝色绘制。



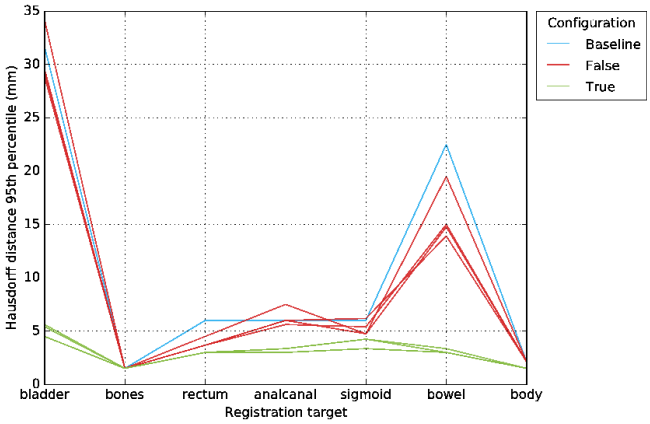
(a) 矢状切片。

(b) 冠状切片。

图37：采用不同网格点布置策略时，MOREA预测的形变渲染图。

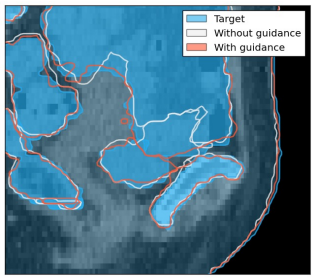


（a）Dice分数。

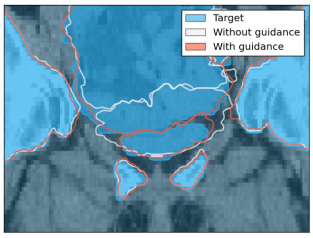


（b）豪斯多夫距离的第95百分位数。

图38：多次运行中，MOREA在有无引导信息时的配准结果比较。刚性配准后的基线分数以蓝色绘制。



（a）矢状切片。



（b）冠状切片。

图39：MOREA 在启用和未启用引导的情况下预测的形变渲染图。

C.3.2 参数文件。 我们以自行编写的参数文件格式向 MOREA 传递参数。下面列出本文所述实验所依据的参数文件。

列表 5：主要 MOREA 实验所依据的参数文件。

```
sweep_描述符 = "$实验_描述符" num_运行次
数 = 5问题_id = "$问题_id" zip = true


问题_指导_已启用 = true 问题_指导_选
择 = "-1"

cuda_计算_级别 = 80cuda_
图形处理器_ID = 0

进化算法_数量_代 = 500进化算法
_种群_大小 = 700进化算法_数量_
簇 = 10进化算法_档案_大小
 = 2000

进化算法_自适应_引导_已启用 = true 进化算法_自适应_引导_在_第_
数量_代时启动 = 100 进化算法_自适应_引导_指导_阈值 = 1.5


MOREA_初始化_噪声_方法 ="全局高斯"
MOREA _init_噪声 _因子 = 1.0

morea_网格_代_方法 = "annotation-group-random-bladder-10"
morea_网格_点_数量 = 600morea_最大_网格_层级_数量 = 1morea_每个_
层级的_代_数量 = 0

morea_幅度_度量 = "生物力学的" morea_图像_度
量 = "平方差" morea_引导_度量 = "按组连续"
morea_采样_率 = 1.0

morea_fos_类型 = "边" morea_对称性_
模式 = "双向变换" morea_双_动态_模式
 ="双" morea_修复_方法 = "高斯"
morea_ams_策略 = "无" morea_数量_扰
动_核 = 0morea_扰动_频率 = 0
```

D 完整实验结果

在本附录中，我们列出第6节所介绍实验的更详尽结果。图40和图41通过平行坐标图给出了所有患者的完整度量结果，并对三种方法进行了比较。表6列出了所有器官豪斯多夫距离的显著性检验结果。表8提供了一个直观视角，其中展示了每位患者的另一张切片，并叠加了预测的变换结果。下面，我们将分析收敛行为（D.1节）和标志点性能（D.2节）。

D.1 收敛行为

我们在图 42 中绘制了各方法在患者1上的收敛行为，以展示它们在得出本文所示结果之前的收敛情况。Elastix 和 ANTs 都采用多分辨率方法。为处理多阶段分辨率中的不连续性，我们在图中用红色竖线标出分辨率切换。我们的 Elastix 配置还包含一个掩膜配准步骤，因此总共有 8 个阶段（掩膜配准的 4 个分辨率和图像配准的 4 个分辨率）。待优化值的缩放并不总是在不同分辨率之间归一化，这解释了不同分辨率之间数值范围的跳变。请注意，ANTs 使用单独的“收敛值”来判断是否收敛，该值绘制在图42d中。对于 MOREA，我们绘制了达到的超体积和达到的最佳引导目标值。第100代时超体积突然下降，与所用的自适应引导策略有关；该策略会从精英档案中清除所有引导目标值不理想的解。

D.2 标志点精度

表7列出了全部4位患者的标志点配准精度。对于每位患者和每种方法，我们汇总各次重复试验中所有标志点的误差，并计算该样本集的均值和标准差。由于这些标志点通常位于清晰可见且解剖学稳定的位置，而不在形变强烈的区域，因此这一精度应被理解为衡量该方法保留某些解剖结构的能力。这一指标因此不太适合用于衡量配准问题的“解决”程度；对此，仍需通过目视检查（形变矢量场和渲染结果）。对于某些标志点，其精确位置的定义可能存在歧义，或在某些患者的扫描图像上不太清晰。不过，仍可借助其所在位置的视觉上下文，并对每对扫描图像采用一致的标记决策，从而在扫描图像之间准确放置这些标志点。

总体而言，我们观察到 Elastix 的表现逊于 ANTs 和 MOREA，而 MOREA 总能改善或大致维持基线标志点配准误差。由于上述原因，我们没有发现大型形变对象上的实际配准性能与目标配准误差值之间存在一致的相关性。

问题	轮廓	MOREA/ Elastix MOREA / ANTs	
患者1	膀胱	**0.011** (+)	**0.007** (+)
	骨骼	**0.009** (+)	**0.006** (+)
	直肠	**0.007** (+)	**0.007** (+)
	肛管	**0.007** (+)	**0.007** (+)
	乙状结肠	**0.007** (+)	**0.007** (+)
	肠道	**0.010** (+)	**0.011** (+)
	body	**0.006** (+)	**0.006** (-)
患者2	膀胱	**0.007** (+)	**0.007** (+)
	骨骼	**0.007** (+)	**0.007** (+)
	直肠	0.118 (+)	**0.007** (-)
	肛管	0.123 (-)	0.180 (-)
	乙状结肠	**0.007** (+)	**0.007** (-)
	肠道	0.401 (+)	**0.007** (+)
	body	0.655 (+)	1.000 (-)
患者3	膀胱	**0.012** (+)	**0.007** (+)
	骨骼	**0.007** (+)	**0.007** (+)
	直肠	0.290 (+)	**0.007** (-)
	肛管	0.118 (-)	**0.007** (+)
	乙状结肠	**0.007** (+)	**0.007** (+)
	肠道	**0.007** (+)	0.056 (+)
	body	**0.007** (+)	0.118 (+)
患者4	膀胱	**0.007** (+)	0.195 (-)
	骨骼	**0.007** (-)	**0.007** (-)
	直肠	**0.010** (-)	**0.007** (-)
	肛管	0.606 (+)	**0.007** (-)
	乙状结肠	**0.009** (+)	0.118 (+)
	肠道	0.119 (+)	0.119 (-)
	body	**0.020** (-)	**0.020** (-)

表6：采用双侧曼-惠特尼U检验计算的各方法间所有轮廓的 Hausdorff 距离两两比较的 p 值。加号（+）表示 MOREA 的均值更优，减号（-）则表示相反。根据 0.025 的 α，显著结果以高亮显示。

问题	基线	Elastix	ANTs	MOREA
患者1	4.8 ± 3.1	5.6 ± 2.8	4.2 ± 2.0	4.8 ± 2.0
患者2	7.5 ± 4.0	11.8 ± 7.3	7.7 ± 4.3	7.8 ± 3.8
患者3	9.5 ± 6.7	6.4 ± 2.0	7.7 ± 2.6	6.5 ± 1.9
患者4	14.1 ± 9.5	8.1 ± 4.3	6.3 ± 3.4	6.8 ± 4.0

表7：所示方法在每位患者上的配准结果经多次重复试验得到的目标配准误差（均值和标准差）。所有误差均以毫米为单位。

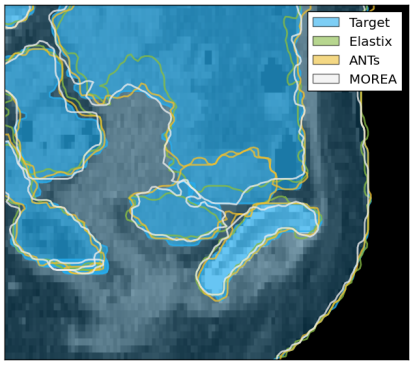
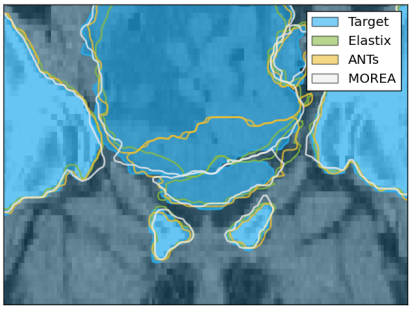
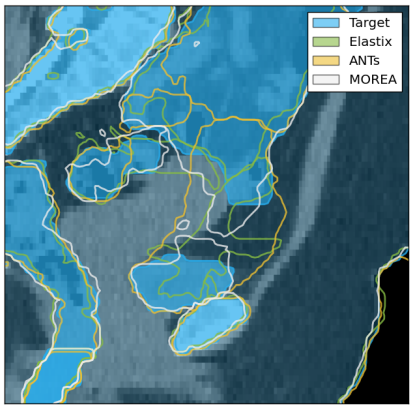
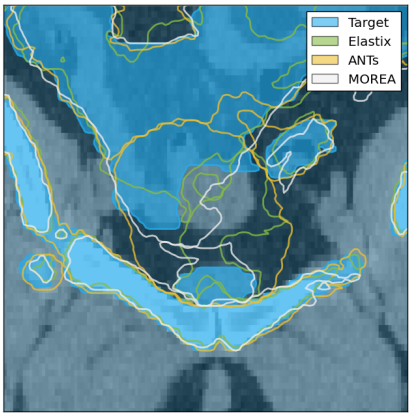
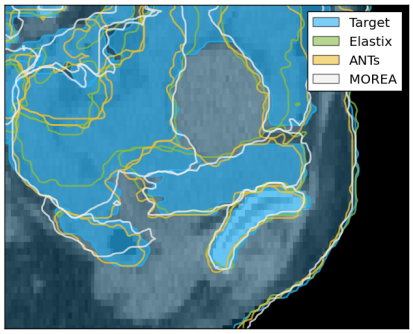
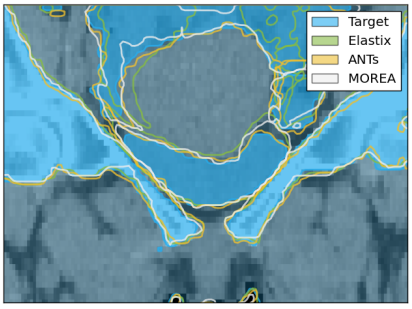
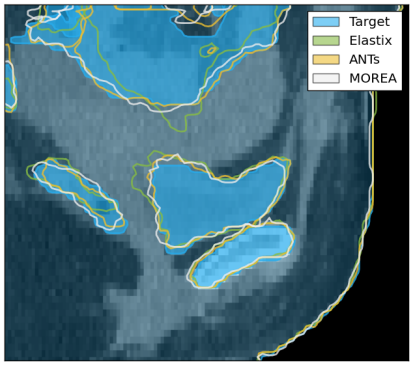
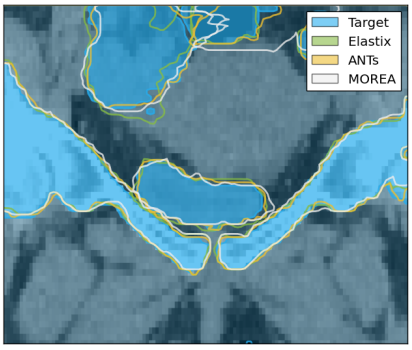
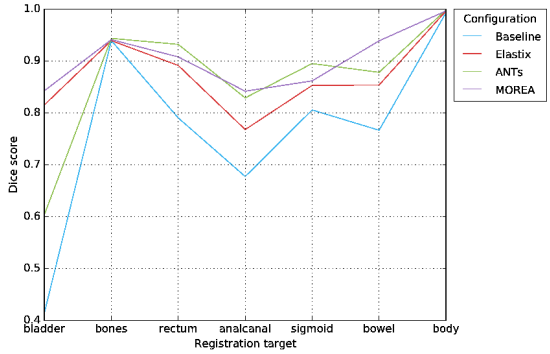
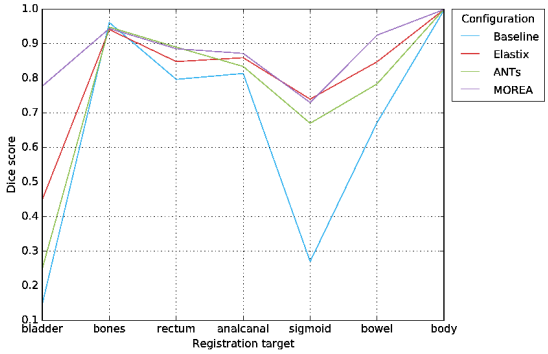
实例	变换后：矢状位	变换后：冠状位
患者1		
患者2		
患者3		
患者4		

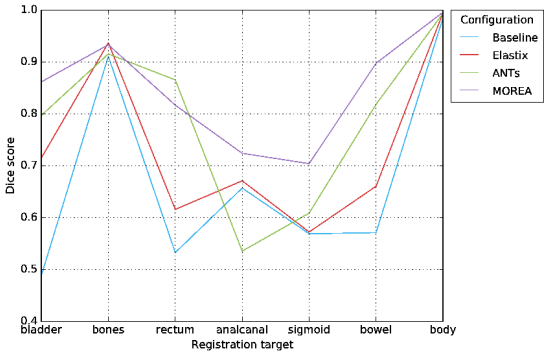
表8：所比较配准方法中部分最佳预测形变，以变形轮廓与目标轮廓和图像的对比表示。



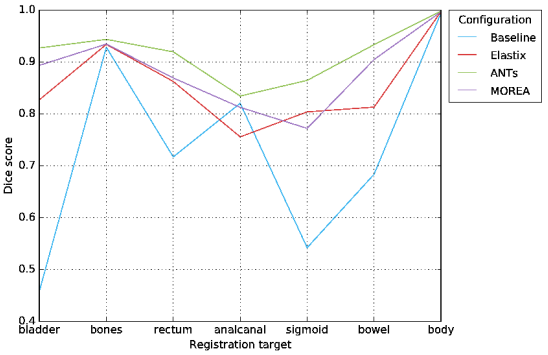
(a) 患者1。



(b) 患者2。

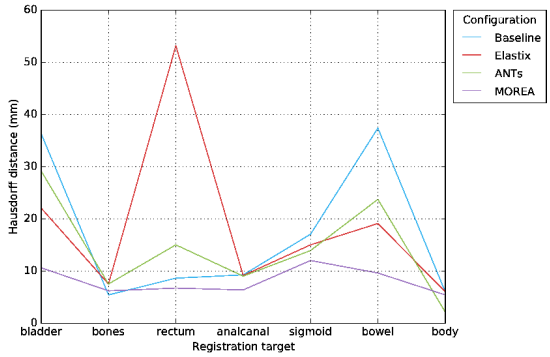


(c) 患者3。

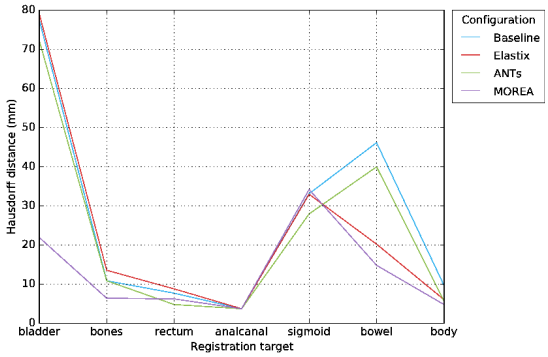


(d) 患者4。

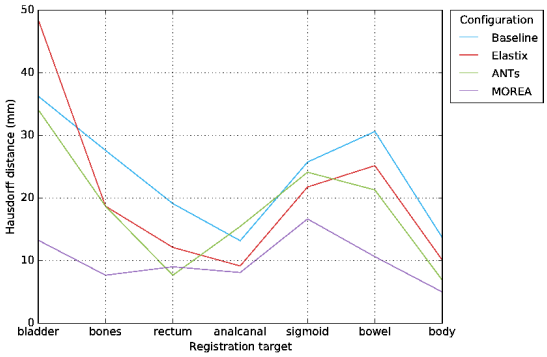
图40：所有方法在所有患者上的 Dice分数。刚性配准后的基线分数以蓝色绘制。



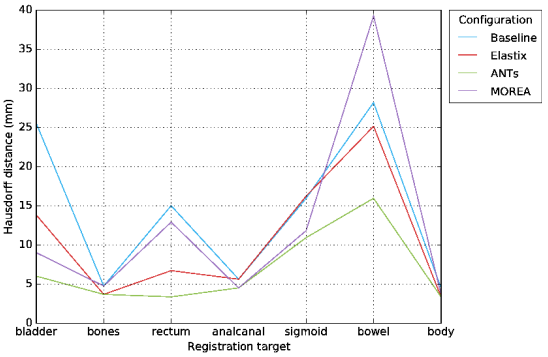
(a) 患者1。



(b) 患者2。

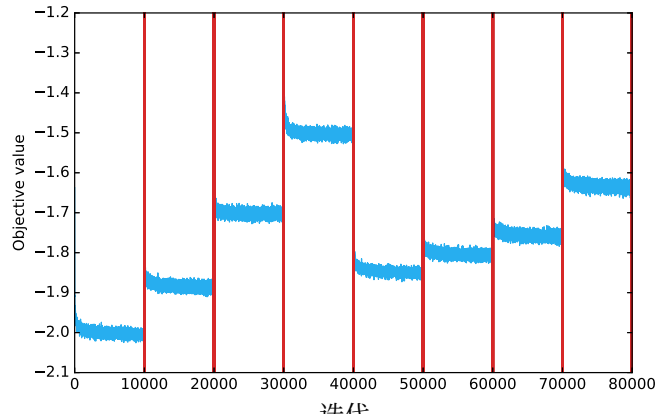


(c) 患者3。

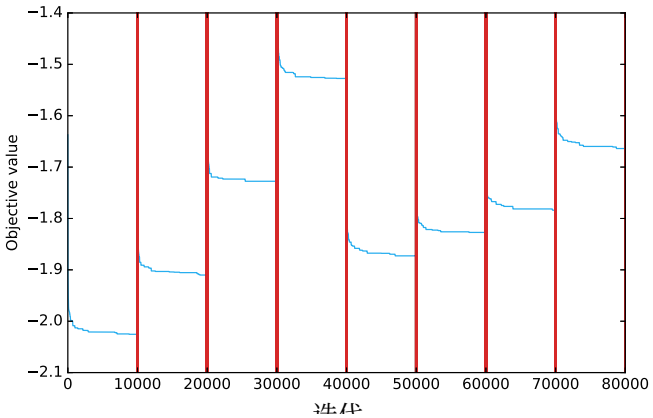


(d) 患者4。

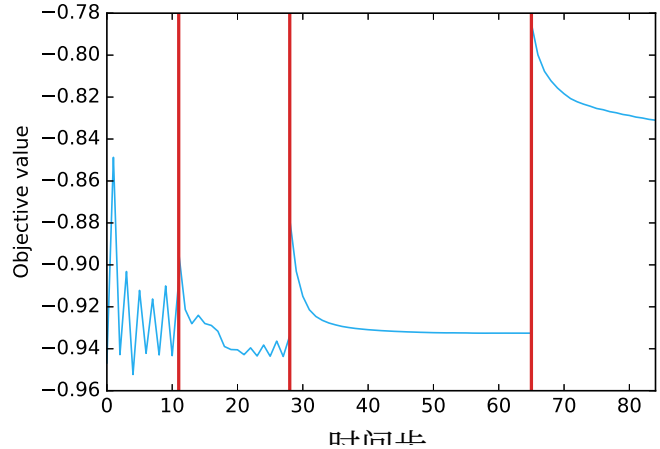
图41：所有患者中各方法的豪斯多夫距离。刚性配准后的基线分数以蓝色绘制。



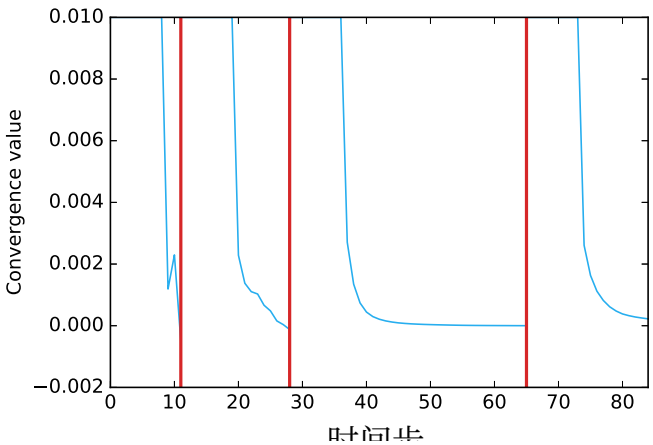
(a) Elastix：每次迭代的目标值。



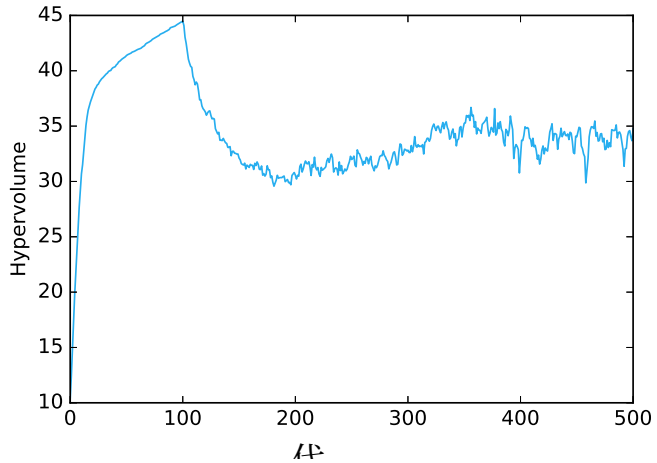
(b) Elastix：每个点达到的最佳目标值。



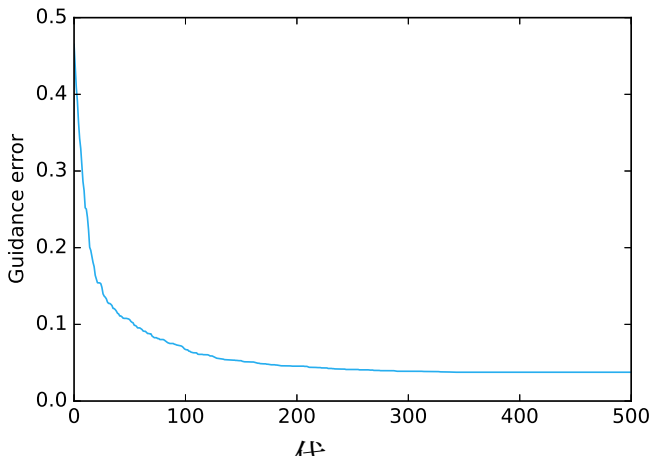
(c) ANTs：每次迭代的目标值。



(d) ANTs：每次迭代的收敛度量。



(e) MOREA：每一代的超体积。



(f) MOREA：每一代找到的最佳引导目标值。

图 42：患者1一次运行中 3 种方法的收敛图。红色竖线表示分辨率发生变化。对于 ANTs，这会产生 4 个优化阶段。对于 Elastix，我们先运行掩膜配准步骤（分为 4 个阶段），然后运行图像配准步骤（同样分为 4 个阶段）。

参考文献

[1] T. Alderliesten、P. A. N. Bosman 和 A. Bel。2015。通过利用多目标优化，充分发挥可变形图像配准中额外引导信息的作用。载于 SPIE医学影像2015：图像处理。94131R。[2] T. Alderliesten、J. J. Sonke 和 P. A. N. Bosman。2012。可变形图像配准的多目标优化：概念验证。载于 SPIE Medical Imaging 2012：图像处理，第8314卷。831420。[3] T. Alderliesten、J. J. Sonke 和 P. A. N. Bosman。2013。通过多目标优化进行可变形图像配准：采用双动态变换模型来应对较大的解剖差异。载于 SPIE医学影像2013：图像处理，第8669卷。866910。[4] G. Andreadis、P. A. N. Bosman 和 T. Alderliesten。2022。基于双单纯形网格的三维医学图像多目标可变形图像配准——概念验证。载于 SPIE医学影像2022：图像处理。744–750。[5] B. B. Avants、C. L. Epstein、M. Grossman 和 J. C. Gee。2008。采用互相关的对称微分同胚图像配准：评估老年及神经退行性脑病患者的自动标注。《医学图像分析》12，1（2008），26–41。[6] F. Bartel、M. Visser、M. de Ruiter、J. Belderbos、F. Barkhof、H. Vrenken、J. C. de Munck 和 M. van Herk。2019。非线性配准提高了检测衰老和痴呆中海马萎缩的统计效能。NeuroImage: Clinical 23（2019），101902。[7] D. L. J. Barten、B. R. Pieters、A. Bouter、M. C. van der Meer、S. C. Maree、K. A. Hinnen、H. Westerveld、P. A. N. Bosman、T. Alderliesten、N. van Wieringen 和 A. Bel。2023。迈向临床实践中的人工智能自动化治疗计划：前列腺高剂量率近距离放射治疗首次临床应用经验的前瞻性研究。近距离放射治疗，即将刊出（2023）。[8] L. Bondar、M. S. Hoogeman、E. M. Vásquez Osorio 和 B. J.M. Heijmen。2010。一种用于处理宫颈癌患者大幅器官形变的对称非刚性配准方法。Medical Physics 37，7（2010），3760–3772。[9] P. A. N. Bosman 和 T. Alderliesten。2016。智能网格初始化降低了基于双动态变换模型、用于处理较大解剖差异的多目标图像配准的计算复杂度。载于 SPIE Medical Imaging 2016：图像处理。978447。[10] A. Bouter、T. Alderliesten 和 P. A. N. Bosman。2017。一种用于内容不匹配和大变形情况下多目标可变形图像配准的新型基于模型的进化算法：效率与质量基准测试。载于 SPIE Medical Imaging 2017：图像处理，第10133卷。1013312。[11] A. Bouter、T. Alderliesten 和 P. A. N. Bosman。2021。利用部分评估实现高度可扩展的实值进化优化。Evolutionary Computation 29，1（2021），129–155。[12] A. Bouter、T. Alderliesten 和 P. A. N. Bosman。2021。应用于多目标可变形图像配准的 GPU加速并行基因池最优混合。载于 IEEE进化计算大会。2539–2548。[13] A. Bouter、T. Alderliesten、B. R. Pieters、A. Bel、Y. Niatsetski 和 P. A. N. Bosman。2019。用于前列腺高剂量率近距离放射治疗的 GPU加速双目标治疗计划。Medical Physics 46，9（2019），3776–3787。[14] A. Bouter、N. H. Luong、C. Witteveen、T. Alderliesten 和 P. A. N. Bosman。2017。多目标实值基因池最优混合进化算法。载于《2017年遗传与进化计算会议论文集》。537–544。[15] D. Brélaz。1979。图顶点着色的新方法。Commun. ACM 22，4（1979），251–256。[16] K. K. Brock、S. Mutic、T. R. McNutt、H. Li 和 M. L. Kessler。2017。图像配准与融合算法及技术在放射治疗中的应用：AAPM放射治疗委员会第132工作组报告。Medical Physics 44，7（2017），e43–e76。[17] K. K. Brock、M. B. Sharpe、L. A. Dawson、S. M. Kim 和 D. A. Jaffray。2005。基于有限元模型的多器官可变形图像配准的准确性。Medical Physics 32，6（2005），1647–1659。[18] H. Chui 和 A. Rangarajan。2000。一种新的非刚性点匹配算法。载于 IEEE计算机视觉与模式识别会议。44–51。[19] K. Deb。2001。《使用进化算法进行多目标优化》。Wiley。[20] B. Eiben、V. Vavourakis、J. H. Hipwell、S. Kabus、T. Buelow、C. Lorenz、T. Mertzanidou、S. Reis、N. R. Williams、M. Keshtgar 和 D. J. Hawkes。2016。生物力学引导的乳房俯卧位至仰卧位对称图像配准。《生物医学工程年鉴》44，1（2016），154–173。[21] C. Ericson。2004。《实时碰撞检测》（第1版）。CRC Press。[22] M. Faisal Beg、M. I. Miller、A. Trouvétrouv 和 L. Younes。2005。通过微分同胚的测地流计算大形变度量映射。《国际计算机视觉杂志》61，2（2005），139–157。[23] B. Fischer 和 J. Modersitzki。2008。病态医学问题——图像配准简介。《逆问题》24，3（2008），1–16。[24] J. Gascon、J. M. Espadero、A. G. Perez、R. Torres 和 M. A. Otaduy。2013。利用四面体网格光栅化快速形变体数据。载于《论文集——SCA 2013：第12届 ACM SIGGRAPH / Eurographics研讨会：计算机动画。181–186。[25] S. Hang。2015。TetGen，一种基于德洛内法的高质量四面体网格生成器。ACM《数学软件汇刊》41, 2（2015），1–36。[26] F. Khalifa、G. M. Beache、G. Gimel'farb、J. S. Suri 和 A. S. El-Baz。2011。《最先进的医学图像配准方法：综述》。载于《多模态最先进医学图像分割与配准方法》。Springer，235–280。[27] A. Klein、J. Andersson、B. A. Ardekani、J. Ashburner、B. Avants、M. C. Chiang、G. E. Christensen、D. L. Collins、J. Gee、P. Hellier、J. H. Song、M. Jenkinson、C. Lepage、D. Rueckert、P. Thompson、T. Vercauteren、R. P. Woods、J. J. Mann 和 R. V. Parsey。2009。应用于人脑磁共振图像配准的14种非线性形变算法评估。NeuroImage 46, 3（2009），786–802。[28] S. Klein、M. Staring、K. Murphy、M. A. Viergever 和 J. P.W. Pluim。2010。Elastix：用于基于强度的医学图像配准的工具箱。IEEE《医学影像汇刊》29, 1（2010），196–205。[29] M. Li、E. Castillo、X. L. Zheng、H. Y. Luo、R. Castillo、Y. Wu 和 T. Guerrero。2013。肺部形变建模：一种结合空间变化的杨氏模量估计的可变形图像配准方法。Medical Physics 40, 8（2013），1–10。[30] G. Loi、M. Fusella、E. Lanzi、E. Cagni、C. Garibaldi、G. Iacoviello、F. Lucio、E. Menghi、R. Miceli、L. C. Orlandini、Antonella Roggio、Federica Rosica、Michele Stasi、Lidia Strigari、Silvia Strolin 和 Christian Fiandra。2018。商用可变形图像配准平台利用基于患者的计算体模进行轮廓传播的性能：一项多机构研究。Medical Physics 45, 2（2018），748–757。[31] H. N. Luong 和 P. A. N. Bosman。2012。多目标进化算法的精英存档：是否适应？载于《第12届自然问题并行求解会议论文集》。72–81。[32] H. Maaranen、K. Miettinen 和 A. Penttinen。2007。连续优化问题遗传算法的初始种群研究。《全球优化杂志》37, 3（2007），405–436。[33] R. Mohammadi、S. R. Mahdavi、R. Jaberi、Z. Siavashpour、L. Janani、A. S. Meigooni 和 R. Reiazi。2019。评估可变形图像配准算法在宫颈癌患者近距离放射治疗中确定累积剂量的效果。《当代近距离放射治疗杂志》11, 5（2019），469–478。[34] S. Nithiananthan、S. Schafer、D. J. Mirota、J. W. Stayman、W. Zbijewski、D. D. Reh、G. L. Gallia 和 J. H. Siewerdsen。2012。超维 Demons：一种在可变形图像配准中纳入缺失组织的方法。Medical Physics 39, 9（2012），5718–5731。[35] D. F. Pace、M. Niethammer 和 S. R. Aylward。2012。可变形图像配准中的滑动几何。载于《国际 MICCAI 腹部影像计算与临床挑战研讨会》。141–148。[36] K. Pirpinia、P. A. N. Bosman、C. E. Loo、G. Winter-Warnars、N. N. Y. Janssen、A. N. Scholten、J. J. Sonke、M. van Herk 和 T. Alderliesten。2017。从多目标优化角度探讨手动参数调优用于可变形乳腺磁共振图像配准的可行性。《医学与生物学中的物理学》62, 14（2017），5723–5743。[37] B. Rigaud、A. Klopp、S. Vedam、A. Venkatesan、N. Taku、A. Simon、P. Haigron、R. De Crevoisier、K. K. Brock 和 G. Cazoulat。2019。用于宫颈癌外照射放射治疗图像与近距离放射治疗图像之间剂量映射的可变形图像配准。《医学与生物学中的物理学》64, 11（2019），115023。[38] L. Risser、F. X. Vialard、H. Y. Baluwala 和 J. A. Schnabel。2013。分段微分同胚图像配准：将其应用于具有滑动条件的三维肺部 CT 图像运动估计。《医学图像分析》17, 2（2013），182–193。[39] B. Schaly、J. A. Kempe、G. S. Bauman、J. J. Battista 和 J. van Dyk。2004。通过考虑可变解剖结构来追踪放射治疗中的剂量分布。《医学与生物学中的物理学》49, 5（2004），791–805。[40] A. Sotiras 和 N. Paragios。2012。可变形图像配准：综述。技术报告。视觉计算中心、应用数学系、巴黎中央理工学院、GALEN 团队、INRIA 萨克雷中心。[41] D. Thierens 和 P. A. N. Bosman。2011。最优混合进化算法。载于《2011年遗传与进化计算会议论文集》。617–624。[42] J.-P. Thirion。1998。图像匹配作为扩散过程：与麦克斯韦妖的类比。《医学图像分析》2, 3（1998），243–260。[43] M. Unser、A. Aldroubi 和 C. R. Gerfen。1993。使用样条金字塔的多分辨率图像配准流程。载于《SPIE 数学成像：小波在信号与图像处理中的应用》，第2034卷。160–170。[44] E. M. Vásquez Osorio、M. S. Hoogeman、L. Bondar、P. C. Levendag 和 B. J. M. Heijmen。2009。一种用于放射治疗非刚性配准的全新灵活框架，具有自动特征对应优化功能。《Medical Physics》36, 7（2009）2848–2859。[45] O. Weistrand 和 S. Svensson。2015。用于放射治疗可变形图像配准的 ANACONDA 算法。Medical Physics 42, 1（2015），40–53。[46] S. Wognum、L. Bondar、A. G. Zolnay、X. Chai、M. C. C. M. Hulshof、M. S. Hoogeman 和 A. Bel。2013。控制结构特异性灵活度可提高膀胱癌放射治疗中基于点的可变形配准的解剖精度。

Medical Physics 40, 2 (2013), 1–15. [47] W. Zhang、Y. Ma、J. Zheng 和 W. J. Allen。2020。带位置约束的四面体网格形变。计算机辅助几何设计 81 (2020), 1–16。

[48] H. Zhong、J. Kim、H. Li、T. Nurushev、B. Movsas 和 I. J. Chetty。2012。一种用于校正低对比度区域可变形图像配准误差的有限元法。《医学与生物学中的物理学》57, 11（2012），3499–3515。