用于 MR-US 匹配与配准的 3D 跨模态关键点描述符

Daniil Morozov, Reuben Dorent, and Nazim Haouchine

Abstract—术中实时超声 (iUS) 和术前磁共振成像 (MRI) 的配准仍然是一个未解决的问题,这是由于外观、分辨率及视野 上的严重差异所致。为了解决这一问题,我们提出了一种新颖的 用于 MRI-iUS 匹配和配准的三维跨模态关键点描述符。我们 的方法采用了一种特定于患者的合成匹配方法,从术前 MRI 生 成合成的 iUS 体积, 从而实现监督对比训练, 以学习一个共享的 描述符空间。然后采用概率性关键点检测策略来识别解剖学上显 著且模态一致的位置。在训练过程中,使用基于课程的三元组损 失和动态困难负样本挖掘来学习描述符,使其对 iUS 伪影如散 斑噪声和覆盖范围有限具有鲁棒性,并且具有旋转不变性。在推 理时,该方法在 MR 和真实的 iUS 图像中检测关键点,并识别 稀疏匹配,随后用于进行刚性配准。我们的方法使用 ReMIND 数据集中的 3D MRI-iUS 对进行评估。实验表明,我们的方 法在 11 名患者中优于最先进的关键点匹配方法,平均精确度 为 69.8% 。在图像配准方面,我们的方法在 ReMIND2Reg 基准上实现了竞争力的平均目标配准误差 2.39 毫米。与现有的 iUS-MR 配准方法相比,我们的框架具有可解释性,不需要手 动初始化,并且对 iUS 视野变化表现出稳健性。代码可在此获 取: https://github.com/morozovdd/CrossKEY。

Index Terms—Cross-modality, 3D Keypoint Descriptor, MRI, Ultrasound, Matching and Registration.

在多种模式下的 配对 图像的对齐仍然是医学成像中的 一个基本挑战,因为模式之间在外观上有显著差异 [1]。 这一挑战支持了多种临床和研究应用,包括基于内容的图 像检索 [1],[2]、切片到体积重建 [3]以及可变形图像配 准 [4]-[8]。

在图像引导手术的背景下,配准互补模态的图像能够融 合不同的解剖和功能信息,提高术中关键结构的识别,最 终改善手术结果。一个有代表性的例子是神经外科,其中 实时术中超声(iUS)通常与术前磁共振成像(MRI)配准, 以补偿术中脑移位并优化肿瘤定位 [9]。

一种弥合模态特定外观差距的策略是通过关键点、分割 或潜在表示的方法来抽象每个图像。当适当地利用这些抽 象时,可以促进多模态医学图像配准。在这项工作中,我们 专注于基于关键点的多模态方法,其中对应关系是通过从 每个图像中检测出的一组稀疏的显著关键点自动建立的。 关键点特别适用于涉及部分观测或切除解剖结构的场景, 例如在术中成像中,并且在非刚性变形下仍然有效。它们 还提供了一个关键的可解释性优势,因为匹配的关键点可 以直接可视化并进行解剖合理性的评估。多模态关键点匹 配领域已经进行了大量研究,并在医学成像中取得了成功 应用。然而,现有的方法通常限于 2D 应用或者图像之间 模态的强度分布相对一致的情况。相比之下,将术前 MR 与术中超声图像对齐则是一种特别具有挑战性的场景,因

Reuben Dorent is with MIND Team, Inria Saclay, Université Paris-Saclay, Palaiseau, France and Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

Daniil Morozov is also with Technical University of Munich, Germany

为存在很大的模态差异。MRI 和术中超声不仅在它们捕获的信息(形态与回波为基础)方面不同,而且在分辨率、获取方式和噪声特性上也存在差异。MRI 通过脉冲序列参数产生高分辨率的 3D 体积,具有强烈的软组织对比,而术中超声则通过声波反射形成部分且噪声较大的视图,通常具有有限的视野(FoV)。弥合这一差距不仅需要处理大的外观差距,还要解决 3D 特有的挑战,包括旋转和视野(FoV)的变化。

贡献。我们提出了一种新颖的跨模态 3D 关键点描述符, 特别用于匹配术前 MR 和 iUS 体积。我们的主要贡献如 下:

- 一种匹配合成策略,其中从患者自身的 MR 体积生成 合成 iUS 图像,并用于训练跨模态描述符网络。
- 一种以显著性热图形式的跨模态关键点检测器,通过 在合成的 iUS 和 MR 体积中积累关键点存在性构建, 然后通过概率聚合来估计跨模态的关键点显著性和一 致性。
- 一种带有课程学习的监督对比框架,增强对 iUS 外观 变化、斑点噪声、旋转和视野变化的鲁棒性。

这项工作是我们会议论文 [10] 的实质性扩展。改进包括: 1) 扩展到 3D; 2) 一种新的对比学习方法; 3) 新实验, 包 括消融研究和图像配准的验证。

I. 相关工作

跨模态关键点匹配可以采取多种形式 [1]。一些方法是 为多光谱环境设计,比如将近红外图像与可见光图像进行 匹配 [11]。另外一些则专注于可见光到红外匹配 [12]或 者是 RGB 到深度图或者卫星图像的匹配 [1]可以说,时 间变化 [13],即在一天或一年不同时间观察同一场景,也 可以在处理极端情况时视为一种多模态性。由于组织外观 的本质上的动态和异质性,医学成像模态可以进一步区分 开来 [7],[14]。解剖结构(如实质、肿瘤、骨骼、体液和 血管)以及功能相关组织(如灰质、白质和纤维束)的可 见性在不同成像模态(如 MRI、CT、iUS、PET、fMRI 或 SPECT)之间显著变化,给用于匹配的描述符设计带来了 独特挑战,这些描述符需要能够超越外观,朝向共享的结 构、功能或语义信息进行广义化 [15],[16]。

2D 医疗图像的多模态匹配:最初,用于医疗图像多模态匹 配的方法主要是为视网膜成像开发的,例如视网膜-FA 或 视网膜-OCT 配准。起初的手工设计描述符如 PIIFD [17] 已被基于学习的方法 [18]-[20] 超越,这些方法通过端到 端关键点学习在应对强度变化、旋转和稀疏标注方面提供 了更强的鲁棒性。在 [21] 中,提出了一种端到端自监督 的 Siamese CNN,用于检测和匹配 2D 下腹部 CT 切片的 解剖标志。该网络共同学习关键点位置和描述符,在强度、 仿射和弹性变换下实现了高密度的对应关系。在 [22] 中,

Daniil Morozov and Nazim Haouchine are with Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA.



Fig. 1: 我们框架的概述。(a) 从术前 MRI 使用 MMHVAE 生成合成 iUS 体积。(b) 通过聚合合成 iUS 和 MRI 的关键 点统计构建跨模态显著图 *P*_{res}, 然后由空间先验 *M*_w 进行调制。(c) 通过三元组损失在多模态补丁对上训练孪生网络, 以生成模态不变描述符。(d) 使用最近邻搜索进行描述符匹配, 然后在 MRI 和 iUS 中采样的关键点之间执行部分分 配。关键点从 MRI 中学到的显著性分布中采样,并在真实 iUS 中均匀采样。

提出了一种利用现成扩散模型特征的扩散引导图像配准方 法。该模型在自然 RGB 图像上进行了预训练,作为可变 形匹配的语义相似性度量。应用于多模态的 2D 双能 X 光 与 X 光以及单模态 MRI (2D 切片) 匹配,他们的方法能 够在解剖结构缺失的情况下实现解剖上有意义的对齐。最 近,我们提出了一种基于对比学习策略的 Siamese 架构 [10],以学习匹配术前 MRI 和术中超声 (iUS) 图像之间的 2D 关键点。该 2D 方法表现出对斑点外观变化的鲁棒性。 与关键点描述符并行,最近的多模态匹配基础模型在医学 影像方面展示了令人印象深刻的成果。基于 Transformer 架构的方法,如 MINIMA [23]和 MatchAnything [24], 利用大规模合成数据集来实现跨多任务的模态不变对应, 包括同模态 MRI (PD-T1, PD-T2, T1-T2)、跨模态结构 和功能影像(MRI-PET, CT-SPECT),以及多种模态对, 如 CT-MRI、PET-MRI、眼底-OCT 或眼底-FA, 而无需 特定任务调优。然而,由于大多数医学 2D 图像表示的是 3D 体积的切片而非 3D 场景的 2D 投影,这些 2D 关键点 方法本质上受到了限制,因为它们假设两幅图像位于相同 的解剖平面上,并且没有考虑 3D 变形。

3D 图像和体积的关键点匹配:对三维数据的关注较少。 之前将传统的二维描述子扩展到三维的努力,如 Harris [25], SIFT [26], SIFT-rank [27]和 SURF [28],已经 强调了三维中的独特挑战,其中方向、采样和视点变化变 得更加复杂,增加了计算成本。因此,提出了更先进的方 法,例如,将 Förstner 算子与归一化梯度场 (NGF) 结合, 检测和描述 CT 图像中的三维关键点 [29],并在配准中 获得了成功使用。在 CT 匹配中, 最近的深度学习方法表 明,对合成三维补丁变形 [30]或仿射变换 [31] 进行自监 督训练显著提高了匹配和配准性能。首批多模态三维描述 子之一, MIND [7], 利用基于补丁的自相似性实现可变 形多模态配准, 最初为 MRI-CT, 最近为 MRI-iUS。最近, KevMorph [4] 及其基础变体 BrainMorph [32], 通过无 监督学习解剖学语义的关键点并以闭合形式计算变换,解 决了不同 MRI 对比度的问题,从而实现了稳健且可解释 的对齐。该方法在具有不同对比度(T1、T2和 PD 加权) 的脑部 MRI 上进行了评估,改善了大错位情况下和不同 对比度对之间的性能。还提出了基于 2D/3D 关键点的方 法,用于使用学习所得且无检测器的特征进行多模态体积 到图像的配准。值得注意的例子包括 X 光到 CT 图像的 自动配准 [33] 和自由手术 iUS 到术前 MRI 体积的配准 [34]。这些方法消除了基于优化的方法通常所需的手动初 始化,并对有限或噪声较大的训练数据具有鲁棒性,使其 特别适合于术中环境。

II. 方法

A. 问题表述,挑战和策略

问题表述:我们的方法检测、描述并预测从 3D 术前 MR 体积 $I_{MR} \in \mathbb{R}^{\Omega}$ 和 3D iUS 体积 $I_{US} \in \mathbb{R}^{\Omega}$ 中提取的两组 关键点之间的部分分配,其中 Ω 表示空间域。每个关键点 i 由一个 3D 点位置 $p_i = (x, y, z) \in [0, 1]^3$ (归一化为图 像大小)和一个描述符 $d_i \in \mathbb{R}^d$ (表征局部 3D 信息)组 成。图像 I_{MR} 和 I_{US} 有独立检测到的 N 和 M 个关键点。 给定这两组关键点,我们寻找关键点 I_{MR} 和 I_{US} 之间的 部分分配矩阵 $P \in \{0,1\}^{N \times M}$ 。由于每个关键点源自一 个唯一的 3D 位置,因此最多只能匹配一次,并且由于遮 挡或不可重复性,某些关键点可能不存在有效的对应关系。 分配矩阵 P 因此是稀疏的。

挑战:当前的方法通常需要大量配对的训练数据并且具 有已知对应关系,以学习鲁棒且具有泛化能力的关键点。 然而,将这种通用模型直接应用于新的模式或临床场景 (如 MR 和 iUS 体积)仍然具有挑战性,因为它们的性能 在很大程度上依赖于在训练期间接触到足够的代表性数 据。特别是,获取 MR 和 iUS 体积之间精确的三维对应关 系需要罕见的临床专业知识并且耗时,这限制了构建如此 大规模配对数据集的可行性。另一种选择是训练深度关键 点检测器和描述符在共配准的 MR-iUS 体积对上。然而, 即便是当大规模数据集可用时,它们通常缺乏必要的精确 MR-iUS 共配准以进行准确建模,使得这一问题成为持续 的瓶颈。 一种基于患者特异性的匹配合成策略:我们提出设计一种患者特异性的方法来检测、描述和匹配 iUS 与术前 MR 体积之间的关键点,而不是设计一个用于构建 *P* 的通用框架。在术前图像信息丰富并且术中图像可以可靠合成的情况下,患者特异性训练表现出优于患者无关方法的性能 [35]-[37]。在我们的情境中,我们提议利用从术前 MRI 生成的合成 iUS 体积来:1) 识别在不同模态中显著且共有的关键点;2) 使用对比学习以跨模态的方式描述局部信息;3) 匹配最具辨别力的对应点。在推理时,我们训练的方法用于识别术前 MRI 和真实 iUS 图像之间的一组对应关系。图 1 展示了我们的方法。

由于术中超声 (iUS) 图像无法在脑外科手术前获取,我 们提出通过从术前 MR 图像合成 iUS 来构建术前 MR 和 iUS 体积的配对训练集。为此,我们利用了最近提出的 MMHVAE 框架,这是一种为不完整多模态数据设计的分 层变分自编码器,用于从术前 MRI 生成逼真的 iUS。该框 架特别适合我们的任务,因为 1) 它能够处理不完整的多 参数 MR 输入,这在临床环境中很常见,以及 2) 通过调整 其分层潜在结构中的标准差的比例参数 γ ,能够生成具有 多样外观(例如,变化的纹理和散斑模式)的合成 iUS 图 像。为了构建一个能推广到真实超声的大型多样化数据集, 我们利用了这两个属性。具体来说,假设我们在 T1、T2 和 T2-FLAIR 中可以访问 K 个术前 MR 序列,我们为 MR 序列的每个 2^{K} – 1 种组合及比例参数 $\gamma \in \{0.3, 0.5, 0.7, 1\}$ 的多个值合成 iUS。这导致了具有不同外观的合成 iUS 图 像数据集的创建,记为 $\mathcal{D}_{synUS} := \{I_{synUS}^{(k)}\}_{k=1}^{4\times(2^{K}-1)}$ 。

我们寻求识别能够满足两个互补目标的关键点:1)在 MR和 iUS之间采样正负对应对,以使用有效的基于补丁 的增强来训练深度跨模态描述符,2)在推断期间提供空间 先验,以在术前 MR 图像中选择信息性关键点。为了实现 这些目标,关键点必须满足三个主要标准:(i)位于显著 区域以确保区分性,(ii)在模态间保持一致以实现高效的 跨模态匹配,(iii)空间上多样化以增强训练的多样性并避 免错误的负对应对。为此,我们引入了一种基于概率跨模 态显著图的新型随机检测策略。

1) 初始独立检测:我们首先在术前 MR 体积数据 I_{MR} 和合成的 iUS 数据集 D_{synUS} 中独立检测显著关键点,使 用的是 SIFT3D [38],这是一种经典 SIFT 算法的体积 扩展版本。虽然 SIFT3D 在检测纹理丰富的区域方面很有 效,但其对特定模态外观的敏感性,例如 MR 和 iUS 之 间的差异,导致跨模态检测之间的一致性较差。这激发了 需要一种联合关键点检测策略,该策略能够考虑跨模态的 显著性。

我们提出了一种跨模态显著性热图,通过聚合 MR 和合成 iUS 对中的 SIFT3D 关键点的空间统计来实现。具体来说,对于每个合成的 iUS 体积 $I_{synUS}^{(k)} \in \mathcal{D}_{synUS}$,我们 计算一个特征存在掩膜 $M_{synUS}^{(k)} \in \{0,1\}^{\Omega}$,如果在体素 v 上检测到 SIFT3D 关键点,则 $M_{v, synUS}^{(k)} = 1$,否则为 0。 然后我们通过在整个合成数据集上对存在掩膜求和构建一 个热图 P_{US} ,即 $P_{US} = \sum_{k=1}^{4 \times (2^{K} - 1)} M_{synUS}^{(k)}$;对 P_{US} 应 用高斯平滑滤波器 ($\sigma = 2$),然后将结果体积归一化到范 围 [0,1]。该热图突出显示了在合成 iUS 外观中稳定的显 著区域。

为了在术前 MRI 中考虑显著区域, 类似地, 通过平滑和 归一化 I_{MR} 的描述符存在掩码构建了一个热图 P_{MR} 。然 后,我们通过使用概率 OR 操作将特定模态的显著性图融 合成一个联合显著性图 $P_{comb} : P_{comb} = 1 - (1 - P_{MR})(1 - 1)$



Fig. 2: MR-iUS 补丁的例子,显示了正对(左)中特征描述符的高相似性和负对(右)中特征描述符的低相似性。*d*维特征向量根据 MR 描述符 *d*^{MR}的值进行排序,突出了 MR 和 iUS 描述符之间的相关性。

Pus),如果一个体素在至少一种模态中是显著的,那么它 被认为是辨别性的可能性就会增加,当它在两种模态中都 显著时,这种可能性会更大。

为了将采样限制在临床相关区域并避免 iUS 中的超出范围点,我们应用了一个空间加权的 FoV 掩膜 M_w 。该掩膜的构建通过计算每个体素到 iUS FoV 质心的欧几里得距离,然后进行高斯平滑以强调中心区域。最终得到的残余显著性图是 $P_{\text{res}} = P_{\text{comb}} \cdot M_w$

在训练期间,关键点通过一个连续拒绝采样程序从残差 显著性图 P_{res} 中随机采样。这个采样程序旨在执行空间和 视场 (FoV) 覆盖约束。具体而言,目标是获得一组 N 关 键点,满足以下条件: C1) 对于每个关键点 i,其对应的 以 p_i 为中心的补丁至少有 80% 位于 iUS FoV 内; C2) 选 定的任何一对关键点 p_i 和 p_j (其中 $i \neq j$)的欧几里得 距离至少为 2mm ,从而确保空间多样性并防止错误的负 对应 (见图 1)。

为了设计跨模态三维描述符,我们采用一种基于孪生深 度学习架构的对比学习方法,将 MRI 或 iUS 领域的补丁 映射到 d 维特征空间。这个共享的特征空间旨在确保 MRI 和 iUS 中对应解剖位置中心的补丁产生相似的描述符,并 且使得特征描述符具有解剖区分能力,从而能够区分不同 的解剖区域。

给定一组 N 关键点,我们在每个位置 p_i 提取大小为 s³ 的立方体补丁。由于术前 MR 和合成 iUS 体积在空间上是共同配准的,可以提取正训练对 (v_i^{MR}, v_i^{US}) ,由 锚定的 MR 补丁 v_i^{MR} 和从随机选择的合成 iUS 体积中 对应的 iUS 补丁 v_i^{US} 组成。这些补丁被输入到一个共 享的 3D ResNet-18 编码器 $\Phi : \mathbb{R}^{S^3} \to \mathbb{R}^d$ 中,该编码 器为 MR 和 iUS 模态分别生成 L2-归一化的特征描述符 $d_i^{MR} = \Phi(v_i^{MR})$ 和 $d_i^{US} = \Phi(v_i^{US})$ 。

为了构建负样本对 $(\boldsymbol{v}_i^{\text{MR}}, \boldsymbol{v}_n^{\text{US}})$,我们利用我们的关键 点采样策略,这一策略促进空间多样性并限制空间冗余。 具体来说,负的 iUS 补丁在 \boldsymbol{p}_i 和 \boldsymbol{p}_j 位置处进行采样,使 得 $j \neq i$ 。

可以使用不同的训练损失,例如二进制交叉熵 (BCE)、 噪声对比估计 (infoNCE) 和三元组损失,来学习跨模态特 征描述符。我们从经验上选择了一种三元组损失,该损失 鼓励描述符的负对 (d_i^{MR} , d_n^{US}) 与任何正对 (d_i^{MR} , d_i^{US})之 间至少保持一定的边界值 m,定义为:

$$\mathcal{L}_{\text{triplet}} = \max\left(0, \|\boldsymbol{d}_{i}^{\text{MR}} - \boldsymbol{d}_{i}^{\text{US}}\|_{2}^{2} - \|\boldsymbol{d}_{i}^{\text{MR}} - \boldsymbol{d}_{n}^{\text{US}}\|_{2}^{2} + m\right)$$
(1)

有效的三元组损失训练的一个关键组成部分是选择有信息的负样本。如果没有困难负样本挖掘策略,训练会迅速饱和,因为许多三元组贡献的梯度可以忽略不计[39]。为了解决这个问题,我们提出了一种基于课程学习的渐进困难负样本挖掘方案:在训练的早期阶段,从空间上远离的区域采样负样本,这些样本在特征空间中预期更容易区分。

随着训练的进展和模型的改进,挖掘逐渐转向更困难的负 样本,这对应于在特征空间中更接近的负描述符(见图 2)。

在每个训练步骤中,对于给定的锚点描述符 d_i^{MR} ,我 们计算批次中所有 $j \neq i$ 个关键点的选择得分 $S_{i,j}$,具体 计算如下:

$$S_{i,j} = (1 - \lambda_t) \underbrace{\min\left(\frac{\|(\boldsymbol{p}_i - \boldsymbol{p}_j)\|_2}{D_{\max}}, 1\right)}_{\text{spatial distance}} - \lambda_t \underbrace{\|\boldsymbol{d}_i^{\text{MR}} - \boldsymbol{d}_i^{\text{US}}\|_2}_{\text{feature similarity}}$$

,其中 $\lambda_t \in [0,1]$ 权衡由 $D_{\text{max}} = 24$ mm 限制的归一化欧 几里得距离和特征相似度。权重 λ_t 充当课程难度调度器, 它逐渐将负采样的重点从解剖学上距离较远的对(高空间 权重)转移到语义上相似的对(高特征相似度)。在实践 中, λ_t 在第 t 轮中被定义为: $\lambda_t = \min(t/T, 1.0)$,其中 T 是预热周期的数量,在我们的实验中设为 200。对于每 个锚点 *i*,候选的负样本 *j* 是得分最低的关键点 S_{i,j}。

为了在学习的描述符中鼓励旋转不变性,我们在训练过 程中对 MR 锚定补丁应用随机 3D 旋转。遵循类似于我 们困难负样本挖掘方案的课程学习策略,旋转角度从范围 $[0^\circ, \theta_{max}]$ 中均匀采样,其中 θ_{max} 在线性增加从 0° 到 30° 的前 1000 个训练周期内。为了避免在旋转过程中信息丢 失,我们最初沿每个空间维度提取比目标大小 s大 1.5 倍 的补丁,然后进行中心裁剪以获得最终的 s^3 体素输入。

在每个轮次中,使用第 Section ?? 节定义的策略,对 每一个合成的 iUS $I_{synUS}^{(k)}$ 和 1024 关键点进行采样。我们 在术前 I_{MR} 和合成 iUS $I_{synUS}^{(k)}$ 的每个关键点周围提取大 小为 s = 32 的补丁。我们使用的批大小为 N = 256,每 个轮次有 4 次迭代。在批处理内执行在线负样本挖掘,以 动态构建硬三元组。使用 AdamW 优化器,该网络在初始 学习率为 10^{-3} 、权重衰减为 2×10^{-3} ,以及最低学习率 为 10^{-6} 的余弦退火调度下,训练 2000 个轮次。

在真实的 iUS 体积的测试阶段,我们使用最近邻匹配策 略构建一个稀疏分配矩阵 $P \in \{0,1\}^{N \times M}$,然后进行歧义 过滤步骤。具体来说,我们在 MR 体积中采样 N = 1024个关键点,使用第??节中描述的过程。由于真实 iUS 图 像没有显著性图,我们使用真实 iUS 体积视野范围内一个 步长为 4 毫米的均匀三维网格来定义 M 个关键点。这允 许对可见解剖结构进行无偏覆盖,同时避免检测到特定模 态的显著结构。对于每个 MR 关键点 i,我们通过使用 L_2 范数比较描述符 d_i^{MR} 和 d_j^{US} 来识别其最佳匹配的 iUS 关 键点 j。使用 Lowe 的比率测试来过滤掉不可靠的对应关 系。

III. 结果

为了评估我们的方法,我们在脑肿瘤患者的 MRI 和 iUS 图像匹配和配准上进行了全面的实验。我们通过以下方式 评估性能:(i)使用标准成对评估指标进行成对图像匹配, (ii)进行消融研究分析每个新组件和旋转的鲁棒性分析, 以及(iii)将术前 MRI 与术后 iUS 进行图像配准。

我们使用大脑切除多模态成像数据库(ReMIND)进行 了实验,该数据库包含来自连续 114 名患者的术前多参数 MRI 和 iUS 图像。在本研究中,我们专注于 13 名提供完 整术前 3D MRI (T1、T2 和 FLAIR)和术前硬膜 iUS 的 患者的子集。iUS 体积图像是通过跟踪的 2D 手持探针采 集重建的。所有图像被重新采样到等方分辨率为 0.5 mm 的分辨率,零填充到平面大小为 192 × 192,并将强度归 一化到 [0,1] 范围。

为了评估跨模态图像匹配,我们按照 [40] 中描述的协 议,通过共同配准硬膜前 iUS 体积和术前 MRI 构建了一 个配对数据集。在方法开发过程中,随机选择了两个病例 用于参数调整和原型设计。最终的方法在其余 11 个测试 病例中进行了评估。我们记得,每个病例(特定患者)训 练一个单独的模型。

对于下游配准任务,我们使用了来自 ReMIND2Reg 挑战赛 2024 中的 9 个验证案例, [41],这些案例提供了手动标注的成对地标作为真实值。

我们评估了我们的跨模态匹配方法识别不同模态之间解 剖对应关键点的能力。实验在 11 测试案例上进行,使用 它们的术前 T2 加权 MRI 和真实硬脑膜前 iUS 图像,并 通过 3D 共配准获得真实对应关系 [40]。

如果匹配位于转换后真实位置的预定义空间容差(2.5 mm)内,则认为匹配是正确的。特征点匹配性能通过以 下指标进行评估:1)精确度,作为正确匹配对的比例,2) 匹配得分,作为正确匹配数相对于 MR 体积中检测到的特 征点数量的比例,以及3)建立的匹配点的数量。

我们将我们提出的特征描述符(我们的)与八种竞争 方法进行对比测试。四种 2D 匹配方法: SuperPoint [42] , ALIKED [43] ,和 LoFTR [12] ,我们与多模式匹配器 MINIMA [23] 配对,以及 Rasheed 等人提出的特定位患 者的 LM2DK 方法;以及四种 3D 描述符方法: SIFT3D [26] , MIND [7] ,Förstner+NGF [29] 和 MedicalNet [44] ,这是一个用于医学成像任务的预训练 CNN (相当于 ImageNet)。

实现细节。二维方法需要为了体数据进行适配。Super-Point、ALIKED和LoFTR,与MINIMA匹配器配对,被 应用于通过其本身的检测机制对三维体积逐切片进行,并 对相应切片进行匹配以及计算聚合指标。LM2DK使用其 原有的基于SuperPoint的检测器和KNN匹配器进行了 逐切片(全部对全部)评估。

对于基线 3D 方法,我们将 SIFT3D 与我们的基于采样的检测器配对,因为其原生关键点检测器无法生成对应点。 由于 MIND 仅提供描述符,我们同样使用了我们的采样 检测器。对于 MedicalNet,我们在其预训练 ResNet-18 结 构的最后一层应用了全局平均池化以获得描述符,同时也 使用了我们的采样检测器进行关键点检测。

由于我们的检测策略具有随机性,我们对每个受试者重 复进行了 10 次评估协议,并将结果在所有受试者中取平 均。对于基于检测的算法,Lowe 比率阈值参数 l 通过网 格搜索进行优化,以在确保至少 40 个匹配点的情况下最 大化精度。这导致我们选择了 l = 0.75 作为我们的参数, l = 0.8 用于 LM2DK 和 MedicalNet,以及 l = 0.9 用于 MIND、SIFT3D 和 Förstner+NGF。

结果如表 I 和图 3 所示,我们的方法在精度和匹配评 分上始终优于所有基线方法。它实现了 69.8% 的匹配精度, 4.90% 的匹配评分,和 72 的平均正确对应数——显示出较 高的辨别能力和对跨模态变化的鲁棒性。相比之下,传统 的 3D 方法如 SIFT3D、MIND 和 Förstner+NGF 即使与 我们的检测器结合,跨模态的实用性也有限。MedicalNet 表现尤差,仅有 1.3% 的精度。在 2D 方法中,LM2DK 表 现最好,精度相对较高 (56.6%),但匹配覆盖率却低得多, 而与 MINIMA 结合的 ALIKED、SuperPoint 和 LoFTR 显示出中等的表现,但并不能匹敌我们专用的 3D 方法的 鲁棒性。 TABLE I: 各个案例在最先进方法上的表现。报告的指标包括精确度 (P. %)、匹配分数 (MS %) 和总匹配点 (MP)。 由于 2D 方法无需检测器,因此不报告其匹配分数。

Case	SIFT3D		Förstner+NGF		MIND		MedicalNet		ALIKED		SuperPoint		LoFTR		LM2KD		Ours						
	Р.	MS	MP	Р.	MS	MP	Р.	MS	MP	Р.	MS	MP	Р.	MP	Р.	MP	Р.	MP	Р.	MP	Р.	MS	MP
C1	48.0	3.19	68	1.3	0.04	152	0.8	0.08	99	0.7	0.05	73	10.5	$2.5\mathrm{k}$	77.5	271	33.0	$4.2\mathrm{k}$	46.5	172	64.5	5.78	92
C2	26.5	0.88	34	0.8	0.04	133	2.7	0.56	209	4.5	0.46	104	4.3	$1.7\mathrm{k}$	57.9	394	24.1	$2.6\mathrm{k}$	35.8	81	64.2	5.00	80
C3	47.4	0.98	21	0.5	0.02	190	2.3	0.54	238	1.1	0.06	60	6.7	$2.4\mathrm{k}$	93.3	$2.7\mathrm{k}$	54.4	$5.1\mathrm{k}$	62.9	35	66.3	3.60	55
C4	43.2	1.69	41	1.3	0.06	309	0.6	0.11	188	0.5	0.03	59	13.0	$3.1\mathrm{k}$	64.8	364	41.9	$6.9\mathrm{k}$	79.8	114	73.9	5.30	74
C5	48.3	1.42	30	0.4	0.02	241	1.2	0.32	279	0.8	0.04	55	4.7	$2.8\mathrm{k}$	79.5	809	53.7	$5.5\mathrm{k}$	63.4	112	66.7	6.50	100
C6	73.0	4.67	66	0.0	0.00	148	0.3	0.07	207	1.4	0.07	48	7.7	$1.8\mathrm{k}$	3.3	30	33.3	$1.9\mathrm{k}$	90.8	76	80.2	3.90	50
C7	46.4	2.35	52	0.0	0.00	145	0.1	0.03	215	0.3	0.02	62	4.1	$1.5\mathrm{k}$	29.0	62	46.0	$4.9\mathrm{k}$	66.7	72	58.4	2.30	40
C8	57.3	1.90	34	0.0	0.00	195	2.1	0.39	188	1.0	0.06	59	7.5	$3.0\mathrm{k}$	72.3	376	57.6	$3.5\mathrm{k}$	58.8	160	73.7	5.90	81
C9	68.2	5.18	78	2.0	0.12	150	0.5	0.11	226	2.6	0.17	66	5.0	$1.2\mathrm{k}$	45.6	228	35.4	$3.1\mathrm{k}$	56.0	25	72.1	6.40	90
C10	55.2	2.40	45	0.0	0.00	184	0.5	0.10	216	1.0	0.08	76	9.3	$1.8\mathrm{k}$	55.9	152	34.3	$4.1\mathrm{k}$	15.9	63	74.9	3.50	47
C11	27.5	0.44	17	0.0	0.00	198	0.9	0.31	356	0.9	0.06	70	5.4	$2.0\mathrm{k}$	2.4	125	39.9	$3.2\mathrm{k}$	46.2	106	73.0	6.20	87
Mean	49.2	2.28	44	0.6	0.03	186	1.1	0.24	220	1.3	0.10	66	7.1	$2.2\mathrm{k}$	52.9	502	41.2	$4.1\mathrm{k}$	56.6	92	69.8	4.90	72
Std	14.4	1.52	20	0.7	0.04	52	0.9	0.19	63	1.2	0.13	15	2.9	643	30.2	764	10.6	$1.4\mathrm{k}$	20.5	46	6.3	1.40	21

1) 消融研究:我们进行了全面的消融研究,以系统地 评估我们方法中的四个关键设计选择:1) 合成训练 MR 序列,2) 关键点检测和采样策略;3) 对比优化目标;以 及4) 旋转不变性策略。所有组件均在关键点匹配任务中 进行评估,结果汇总在表 II 和图 4 中。我们的基线配置 采用了所有合成模态(T2、T1 和 FLAIR)、课程三元组 损失、课程旋转增强和随机关键点检测。

iUS 合成变化的影响。我们通过从 T2、T1 和 FLAIR 序列的各种组合进行训练,研究利用不同的 MR 衍生合成 US 图像的贡献。在所有配置中,T2 都作为基础模态,因 为它对应于我们的目标 MR 序列。表格 II 的结果显示了 在合成 US 图像时合并额外 MR 序列以增加 iUS 数据多 样性的好处。虽然单独添加 T1 或 FLAIR 相较仅使用 T2 训练能提高性能,但结合所有三种模态可获得最佳结果, 匹配的平均数量从 18.9 显著增加到 72.3 。这表明每种模 态提供了互补的解剖信息,从而增强了描述符的区分能力。

随机与确定性检测。我们比较了确定性和随机关键点采 样策略。在确定性设置中,从术前 MRI (*I*_{MR}) 和合成 数据集 *D*_{synUS} 中预先选取了一组固定的 SIFT3D 关键点。 相比之下,随机方法在训练过程中动态采样关键点。如表 II 所示,随机采样显著提高了模型性能,尤其是在精度方 面 (69.79% vs. 30.40%)。这一结果表明,在训练过程中 让模型接触到更广泛的空间上下文和更多样化的负样本是 有益的。

难例挖掘策略。为了评估我们结合三元组损失的难例挖 掘策略的有效性,我们将其与 BCE 和 InfoNCE 进行了 比较。如表 II 所示, BCE 的表现最差,精确度仅为 8.6% ,表明其判别能力有限。InfoNCE 获得了更高的匹配得分 (6.2%),但代价是精确度降低(60.4%),这表明其在接 受匹配时过于宽松。相比之下,我们基于课程的三元组损 失在各项指标中提供了最佳的折衷,达到了最高的精确度 (60.4%)、正确的匹配得分(4.9%)以及足够的正确匹配 数量(72.3),显示出它在学习具有判别性和空间意义的 表征方面的稳健性。

课程旋转增强。最后,我们通过比较三种训练策略来评 估模型对旋转的鲁棒性:无旋转增强、完全旋转增强(即 从一开始随机旋转)以及我们的课程旋转增强。为了评估 旋转不变性,我们在10个系统增加的方向差异下测试性 能,这些差异从0°到30°,递增步长为3°,并围绕5 个随机采样的3D轴应用。正如图4所示,随着旋转的增 加,无旋转基线的性能下降,突显了其无法泛化到未见过 的方向的能力。完整旋转模型在旋转角度上表现出更一致 的性能,但总体表现不佳,这表明过早暴露于高旋转变化

TABLE II: 消融研究

Configurations	Prec. (%)	MSc. (%)	MP		
Synth. Modalities					
T2 T1 FLAIR					
• • •	68.47 ± 10.42	1.26 ± 0.26	18.94 ± 3.30		
• • •	68.86 ± 5.94	3.68 ± 0.57	54.67 ± 6.20		
• • •	69.86 ± 6.49	3.59 ± 0.57	51.94 ± 6.62		
• • •	69.79 ± 4.94	4.92 ± 0.62	72.32 ± 7.92		
Optimization Loss			(
BCE	8.64 ± 1.00	4.20 ± 0.53	503.07 ± 18.31		
InfoNCE	60.39 ± 4.67	6.18 ± 0.70	104.13 ± 8.78		
Triplet	69.79 ± 4.94	4.92 ± 0.62	72.32 ± 7.92		
Point Detection					
Deterministic	30.40	3.11	105.55		
Stochastic	69.79 ± 4.94	4.92 ± 0.62	72.32 ± 7.92		

可能限制模型提取语义上有意义特征的能力。相反,我们 基于课程的方法在中等旋转角度(< 20°)内获得了高性 能,并在此后逐渐下降,展示了鲁棒性和判别能力之间更 有效的权衡。

A. 图像配准

我们在注册术前 MR 和术后切除 iUS 体积的任务上测 试了我们的方法,使用的是公开可用的 ReMIND2Reg 数 据集,这是 Learn2Reg 2024 挑战的一部分。该数据集包含 由于肿瘤切除导致的大幅组织变形和拓扑变化的病例,并 在两种模式下都包括手动注释的真实标记,通过目标注册 误差(TRE)可以进行定量评估。

1) 基于关键点的迭代配准:我们的处理流程经过三个 刚性对齐迭代来执行配准。在每次迭代中,通过我们的描述 符计算固定的 MR 体积和当前状态移动的 iUS 体积之间 的关键点对应关系。然后使用这些稀疏匹配通过 RANSAC 估计刚性变换,配置为最多 4000 次迭代和 5.0 mm 内点 阈值。所得变换与先前的变换合成,原始移动的 iUS 图像 相应地重新采样以进行下一次迭代。最终输出是跨所有迭 代的累积刚性变换。

2) 竞争的配准方法: 我们将我们的方法与 Re-MIND2Reg 2024 挑战赛排行榜上的顶级提交进行比较: 1) Madesta 等人的 VROC (Variational Registration on Crack) 方法,使用常规迭代优化的两阶段刚性配准流程, 使用高斯平滑输入、掩模和强度阈值,并依次优化 NCC 和 NGF 指标; 2) Wang 等人的 next-gen-nn,是一种基于 MIND-SSC 特征提取器的多层次相关性平衡优化策略的 无监督方法; 3) Li 等人的 Topology Higher-Order MRF,



Fig. 3: 三个病例(列)中的定性匹配结果。第1至5行显示来自Top5表现方法的切片结果。绿线表示正确匹配;红 点表示错配。最后一行显示使用我们的描述符进行匹配的体积渲染。

这是一种基于拓扑高阶 MRF 并使用多分辨率二次伪布尔 优化策略进行多尺度优化的可变形配准框架;以及 Wang 等人的 Coarse-to-Fine Registration with Style Transfer, 采用粗到细的策略,使用 3D CycleGAN 进行无配对成像 风格迁移,将 iUS 图像转换为合成的 T1 风格 MR 图像, 以在配准前在不同模态之间创建更统一的信号分布,然后 进行 NiftyReg 注册。

验证数据集中 9 个案例的结果如表 III 所示,并按其平 均 TRE (对应于排行榜名次)排序。我们的方法实现了

2.385±0.397 毫米的平均 TRE,在挑战排行榜中排名第 三。尽管只使用了刚性变换模型,并完全依赖于稀疏的关 键点对应,我们的方法相对于几个基于学习和可变形的策 略,展示了具有竞争力的性能。

为了定性评估配准精度,我们在图 5 中展示了在应用我 们的方法之前和之后的标志对齐情况,其中有两个代表性 的病例显示了叠加的 MR 和 iUS 体积。即使在涉及显著 组织变形和特定于模态的伪影的复杂场景中,我们的方法 也能实现视觉上连贯的配准。iUS 的视野与 MRI 中对应



Fig. 4: 在不断增加的旋转角度下的旋转不变性分析。我们的基于课程的模型在旋转失配逐渐增加的情况下保持了更高的匹配质量。

TABLE III: 在 ReMIND2Reg 验证集上的配准方法结果。

Rank	Method	Mean TRE (mm)
1	VROC	1.903 ± 0.582
2	next-gen-nn	1.969 ± 0.459
*	Ours (Rigid, Keypoint-Based)	2.385 ± 0.397
3	Coarse-to-Fine w/ Style Transfer	2.419 ± 0.669
4	Topological Higher-Order MRF	3.680 ± 0.620

的解剖区域很好地对齐,而标志对应关系显示出一致的空间关系。配准质量在复杂解剖结构的案例中尤其明显,例 如脑室边界和组织界面,在变换后保持了预期的空间关系。

结果展示了我们基于关键点的配准框架的潜力,提供了 一种稳健且精确的解决方案,无需复杂的优化策略或手动 初始化,并通过匹配的关键点为配准过程提供了清晰的视 觉提示。

IV. 讨论、局限性和结论

我们提出了一种新颖的、完全三维、旋转不变且跨模态 的关键点描述符,专为 MRI 到 iUS 体积匹配任务设计。 我们的方法连接了两种高度不同的模态,超越了对比无关 或单模态的方法。据我们所知,这是该领域中第一个量身 定制的关键点描述符。我们的方法始终优于现有的最先进 的描述符和匹配技术,并实现了与最新基准挑战中的领先 方法相当的配准精度。除了精度之外,我们的方法在可用 性和可解释性方面提供了关键优势。匹配和不匹配的关键 点可以直接可视化,使临床医生能够评估解剖合理性和配 准信心。此外,我们的配准流程不需要手动初始化,这在 基于优化的方法中是常见的,并且是临床采用的重大障碍。

我们的方法存在一些局限性。首先,尽管患者特异性训练策略可以提高准确性,但需要大约五小时的训练。然而,由于术前 MRI 通常在手术前早已获得,这种离线训练时间在临床上是可以接受的。此外,最近的研究表明,在给定一个预训练模型的情况下,对新患者的术前图像进行快速微调可以在大大减少训练时间的同时保持高性能 [35]。 其次,我们基于合成的训练利用了固定的视场(FoV)先验来生成合成的超声图像。这样会引发一个问题,即当手术过程中 iUS 视场偏离模拟的训练视场时,性能可能会下降。为了解这一问题,我们进行了一个实验,使用了两个案例,每个案例采用了三种不同的 iUS 视场。如图 6 所示, 模型在不同视场间表现出高度的可重复性和一致性。在原始视场检测到的关键点能够可靠地在其他视场中恢复,并且大部分匹配的关键点在所有三个视图中都是一致的。这表明学习到的描述符能够很好地推广超出合成的训练条件,并在真实手术场景中 iUS 覆盖范围的实际变化下保持鲁棒性。最后,尽管我们在切除的非刚性组织上使用刚性配准管道兼容。它们可以作为稀疏约束整合到生物力学模型或 B 样条框架中 [45]。此外,它们非常适合处理诸如肿瘤切除等拓扑变化,其中匹配的关键点可以提供组织应力估计并指导 MRI 更新。未来的工作将探索这些扩展,并研究通过关键点引导的交互式编辑用于切片到体积配准的下游任务,包括自由手超声重建和分割传播。

ν.

致谢 这项工作得到了国家卫生研究院 (R01EB032387, R01EB034223,和 K25EB035166)的支持。R.D. 获得了玛 丽·居里资助 No 101154248 (项目: SafeREG)。

References

- X. Jiang, J. Ma, G. Xiao, Z. Shao, and X. Guo, "A review of multimodal image matching: Methods and applications," *Information Fusion*, vol. 73, pp. 22–71, 2021.
- [2] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Contentbased medical image retrieval: a survey of applications to multidimensional and multimodality data," *Journal of digital imaging*, vol. 26, pp. 1025–1039, 2013.
- [3] E. Ferrante and N. Paragios, "Slice-to-volume medical image registration: A survey," *Medical image analysis*, vol. 39, pp. 101– 123, 2017.
- [4] M. Y. Evan, A. Q. Wang, A. V. Dalca, and M. R. Sabuncu, "Keymorph: Robust multi-modal affine registration via unsupervised keypoint detection," in *Medical imaging with deep learning*, 2021.
- [5] I. Machado, M. Toews, J. Luo, P. Unadkat, W. Essayed, E. George, P. Teodoro, H. Carvalho, J. Martins, P. Golland, S. Pieper, S. Frisken, A. Golby, and W. III, "Non-rigid registration of 3d ultrasound for neurosurgery using automatic feature detection and matching," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, 06 2018.
- [6] J. Luo, M. Toews, I. Machado, S. Frisken, M. Zhang, F. Preiswerk, A. Sedghi, H. Ding, S. Pieper, P. Golland, A. Golby, M. Sugiyama, and W. M. Wells III, "A feature-driven active framework for ultrasound-based brain shift compensation," in *MICCAI 2018*, 2018, pp. 30–38.
- [7] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, M. Brady, and J. A. Schnabel, "Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration," *Medical image analysis*, vol. 16, no. 7, pp. 1423–1435, 2012.
- [8] S. Joutard, R. Dorent, S. Ourselin, T. Vercauteren, and M. Modat, "Driving points prediction for abdominal probabilistic registration," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2022, pp. 288–297.
- [9] J. M. Gonzalez-Darder, State of the Art of the Craniotomy in the Early Twenty-First Century and Future Development. Cham: Springer International Publishing, 2019, pp. 421–427.
- [10] H. Rasheed, R. Dorent, M. Fehrentz, T. Kapur, W. M. Wells III, A. Golby, S. Frisken, J. A. Schnabel, and N. Haouchine, "Learning to match 2d keypoints across preoperative mr and intraoperative ultrasound," in *International Workshop on Advances in Simplifying Medical Ultrasound*. Springer, 2024, pp. 78–87.
- [11] E. B. Baruch and Y. Keller, "Joint detection and matching of feature points in multimodal images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6585–6593, 2021.
- [12] Ö. Tuzcuoğlu, A. Köksal, B. Sofu, S. Kalkan, and A. A. Alatan, "Xoftr: Cross-modal feature matching transformer," in *Proceed*ings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 4275–4286.



Fig. 5: MR-iUS 配准的定性结果显示 T2 加权 MR(灰度)与叠加的 iUS(颜色覆盖,例如红色),具有表示 TREs 的 MR 标志(例如,蓝点)和 iUS 标志(例如,黄色叉)。我们的方法显示了解剖结构的良好对齐和标志之间的小 TREs。



Fig. 6: 评估描述符在不同 iUS 视野范围内的稳健性。尽管 视野发生变化,关键点匹配仍保持一致并重复,这表明其 超出训练条件的泛化能力。

- [13] Y. Verdie, K. Yi, P. Fua, and V. Lepetit, "Tilde: A temporally invariant learned detector," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5279–5288.
- [14] P. Juvekar, R. Dorent, F. Kögl, E. Torio, C. Barr, L. Rigolo, C. Galvin, N. Jowkar, A. Kazi, N. Haouchine *et al.*, "Remind: The brain resection multimodal imaging database," *medRxiv*, 2023.
- [15] N. Dey, B. Billot, H. E. Wong, C. J. Wang, M. Ren, P. E. Grant, A. V. Dalca, and P. Golland, "Learning general-purpose biomedical volume representations using randomized synthesis," 2024.
- [16] R. Dorent, N. Haouchine, F. Kogl, S. Joutard, P. Juvekar, E. Torio, A. J. Golby, S. Ourselin, S. Frisken, T. Vercauteren et al., "Unified brain mr-ultrasound synthesis using multi-modal hierarchical representations," in *MICCAI*. Springer, 2023, pp. 448–458.
- [17] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, "A partial intensity invariant feature descriptor for multimodal retinal image registration," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1707–1718, 2010.
- [18] J. A. Lee, P. Liu, J. Cheng, and H. Fu, "A deep step pattern representation for multimodal retinal image registration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 5077–5086.
- [19] J. Liu, X. Li, Q. Wei, J. Xu, and D. Ding, "Semi-supervised keypoint detector and descriptor for retinal image matching," in *European Conference on Computer Vision*. Springer, 2022, pp. 593–609.
- [20] M. Santarossa, A. Kilic, C. von der Burchard, L. Schmarje, C. Zelenka, S. Reinhold, R. Koch, and J. Roider, "Medregnet: Unsupervised multimodal retinal-image registration with gans and ranking loss," in *Medical Imaging 2022: Image Processing*, vol. 12032. SPIE, 2022, pp. 321–333.
- [21] M. Grewal, T. M. Deist, J. Wiersma, P. A. Bosman, and T. Alderliesten, "An end-to-end deep learning approach for landmark detection and matching in medical images," in *Medical Imaging 2020: Image Processing*, vol. 11313. SPIE, 2020, pp. 548–557.

- [22] N. Tursynbek, H. Greer, B. Demir, and M. Niethammer, "Guiding registration with emergent similarity from pre-trained diffusion models," arXiv preprint arXiv:2506.02419, 2025.
- [23] J. Ren, X. Jiang, Z. Li, D. Liang, X. Zhou, and X. Bai, "Minima: Modality invariant image matching," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 23059–23068.
- [24] X. He, H. Yu, S. Peng, D. Tan, Z. Shen, H. Bao, and X. Zhou, "Matchanything: Universal cross-modality image matching with large-scale pre-training," 2025.
- [25] I. Sipiran and B. Bustos, "Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes," *The Visual Computer*, vol. 27, pp. 963–976, 2011.
- [26] B. Rister, M. A. Horowitz, and D. L. Rubin, "Volumetric image registration from invariant keypoints," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4900–4910, 2017.
- [27] L. Chauvin, K. Kumar, C. Wachinger, M. Vangel, J. de Guise, C. Desrosiers, W. Wells, M. Toews, A. D. N. Initiative *et al.*, "Neuroimage signature from salient keypoints is highly specific to individuals and shared by close relatives," *NeuroImage*, vol. 204, p. 116208, 2020.
- [28] R. Agier, S. Valette, L. Fanton, P. Croisille, and R. Prost, "Hubless 3d medical image bundle registration," in VISAPP 2016 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, vol. 3, 2016, pp. 265–272.
- [29] J. Rühaak, T. Polzin, S. Heldmann, I. J. Simpson, H. Handels, J. Modersitzki, and M. P. Heinrich, "Estimation of large motion in lung ct by integrating regularized keypoint correspondences into dense deformable registration," *IEEE transactions on medical imaging*, vol. 36, no. 8, pp. 1746–1757, 2017.
- [30] M. Grewal, J. Wiersma, H. Westerveld, P. A. Bosman, and T. Alderliesten, "Automatic landmark correspondence detection in medical images with an application to deformable image registration," *Journal of Medical Imaging*, vol. 10, no. 1, pp. 014007-014007, 2023.
- [31] N. Loiseau-Witon, R. Kéchichian, S. Valette, P. Bailly, and I. E. Magnin, "Learning 3d medical image keypoint descriptors with the triplet loss," *International Journal of Computer Assisted Radiology and Surgery*, vol. 17, pp. 141–146, 2022.
- [32] A. Q. Wang, R. Saluja, H. Kim, X. He, A. Dalca, and M. R. Sabuncu, "Brainmorph: A foundational keypoint model for robust and flexible brain mri registration," arXiv preprint arXiv:2405.14019, 2024.
- [33] J. Esteban, M. Grimm, M. Unberath, G. Zahnd, and N. Navab, "Towards fully automatic x-ray to ct registration," in Medical Image Computing and Computer Assisted Intervention-MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13-17, 2019, Proceedings, Part VI 22. Springer, 2019, pp. 631-639.
- [34] V. Markova, M. Ronchetti, W. Wein, O. Zettinig, and R. Prevost, "Global multi-modal 2d/3d registration via local descriptors learning," in *MICCAI*. Springer, 2022, pp. 269–279.
- [35] V. Gopalakrishnan, N. Dey, D.-D. Chlorogiannis, A. Abumoussa, A. M. Larson, D. B. Orbach, S. Frisken, and P. Golland, "Rapid patient-specific neural networks for intraoperative x-ray to volume registration," 2025.

- [36] M. Fehrentz, M. F. Azampour, R. Dorent, H. Rasheed, C. Galvin, A. Golby, W. M. Wells, S. Frisken, N. Navab, and N. Haouchine, "Intraoperative registration by cross-modal inverse neural rendering," in *MICCAI*. Springer, 2024, pp. 317– 327.
- [37] R. Dorent, E. Torio, N. Haouchine, C. Galvin, S. Frisken, A. Golby, T. Kapur, and W. M. Wells, "Patient-specific realtime segmentation in trackerless brain ultrasound," in *MICCAI*. Springer, 2024, pp. 477–487.
- [38] B. Rister, M. A. Horowitz, and D. L. Rubin, "Volumetric Image Registration From Invariant Keypoints," vol. 26, no. 10, pp. 4900–4910.
- [39] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," arXiv preprint arXiv:1703.07737, 2017.
- [40] R. Dorent, N. Haouchine, A. Golby, S. Frisken, T. Kapur, and W. Wells. Unified Cross-Modal Image Synthesis with Hierarchical Mixture of Product-of-Experts.
- [41] R. Dorent, T. Kapur, S. Wells, A. Golby, W. Heyer, J. Chen, Y. Liu, M. Heinrich, A. Walter, J. Lindblad, N. Slodaje, P. Paul-Gilloteaux, L. Hansen, M.-C. Domart, L. Collinson, and M. Jones, "Learn2reg 2024," Apr. 2024.
- [42] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018, pp. 224–236.
- [43] X. Zhao, X. Wu, W. Chen, P. C. Chen, Q. Xu, and Z. Li, "Aliked: A lighter keypoint and descriptor extraction network via deformable transformation," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–16, 2023.
- [44] L. Alzubaidi, J. Santamaría, M. Manoufali, B. Mohammed, M. A. Fadhel, J. Zhang, A. H. Al-Timemy, O. Al-Shamma, and Y. Duan, "Mednet: pre-trained convolutional neural network model for the medical imaging tasks," arXiv preprint arXiv:2110.06512, 2021.
- [45] N. Haouchine, P. Juvekar, M. Nercessian, W. M. Wells III, A. Golby, and S. Frisken, "Pose estimation and non-rigid registration for augmented reality during neurosurgery," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 4, pp. 1310–1317, 2022.