超声图像解释和扫描指导的语义场景图

Xuesong Li^{1,2}, Dianye Huang^{1,2}, Yameng Zhang^{1,3}, Nassir Navab¹, and Zhongliang Jiang¹

¹ Computer Aided Medical Procedures (CAMP), TU Munich, Germany

zl.jiang@tum.com

² Munich Center for Machine Learning (MCML), Munich, Germany ³ The Chinese University of Hong Kong, Hong Kong, China

摘要理解医学超声成像仍然是一个长期的挑战,这是由于成像和采集参数差异导致显著的视觉变异性。大型语言模型(LLMs)最近的进步已被用于自动生成富含术语且面向具有足够生理学知识的临床医生的摘要。然而,在非专业用户中对提高超声可解释性和基本扫描指导日益增长的需求,例如在即时护理环境中,并未得到充分探索。在这项研究中,我们首先引入了超声图像中的场景图(SG),以向普通用户提供图像内容的解释并提供超声扫描指导。该超声 SG 是通过基于变压器的一阶段方法计算得出的,无需明确的对象检测。为了生成面向普通用户的可理解的图像解释,然后使用用户查询进一步通过 LLMs 精炼抽象的 SG 表示。此外,探索了预测出的 SG 在引导超声扫描向当前成像视图中缺失解剖结构方面的作用,帮助普通用户实现更标准化和完整的解剖学探索。这种基于 SG 的图像解释和扫描指导的有效性已在五名志愿者的左、右颈部区域(包括颈动脉和甲状腺)的图像上进行了验证。结果表明该方法具有通过增强其可解释性和可用性来最大限度地普及超声技术的巨大潜力。项目页面:https://noseefood.github.io/us-scene-graph/

Keywords: 超声图像分析 · 场景图。

1 介绍

医学超声(US)在现代临床实践中被广泛用于检查如颈动脉、甲状腺和 肝脏等内部器官。由于其易用性和便携性,超声成像具有广泛部署的潜力, 使其更加普遍可用。然而,解释超声图像需要丰富的经验,因为来自成像和 获取参数差异的显著视觉变异性很大。与 CT 和 MRI 不同,超声解释不那 么直观,因为它视野有限且缺乏三维结构信息。因此,在更广泛的采用中, 2 Xuesong Li et al.

特别是在即时护理超声(POCUS)环境中 [24,2,16,5] 和非专业人士的解剖 学及生理学自我学习方面,有效的超声图像解释与扫描指导至关重要。

从经典计算机视觉中的场景图(SG)技术 [8,14,12] 和新兴的手术数据 科学 [21,25,23,31,13] 中汲取灵感,这种方法能够有效地总结图像中的关键 对象及其关系。因此,可以通过利用预测的 SG 中定义的关系生成直观的图 像解释。与最近使用大型语言模型(LLMs)[11,17] 专注于综合报告生成的 努力不同,为单个图像计算概念化的 SG 表示提供了更大的灵活性。虽然临 床医生可能需要包含医学术语的全面超声报告,但对于非专业人士来说这可 能会不太直观。最近的一项努力在引入 SG 来描述 CT 图像中的对象及其关 系方面取得了开创性成果,在 [26] 中有所报道。与全面总结不同的是,SG 表示提供了一个高度概念化的总结,仅强调关键信息。这种中间表示可以无 缝地适应各种下游任务,如超声图像摘要或用于医学学生培训的探头运动引 导,通过进一步集成全层次的解剖学知识来实现。

超声图像是一个二维横截面图像,不像自然图像那样具有明显的前景和 背景。这一特性以及相对稳定的解剖结构简化了与自然图像相比的对象关系 提取。由于图像对比度低,在超声图像中进行目标检测相对较难。在这项研 究中,我们采用最先进的 RelTR [7] 用于超声 SG 生成。这种基于单阶段变 换的方法消除了对显式对象检测的需求,实现了高效且直接的关系提取。

为了展示 SG 表示在 US 成像中的影响,本研究提出了一种新颖的方法, 该方法利用高度概念化的 SG 来实现推进便携式和可访问 US 成像民主化的 两个关键任务:(1)为普通人生成可理解的 US 解释,以自学解剖学和生理 学知识,这对于 POCUS 场景也特别有用;(2)提供扫描指导,以揭示当 前成像视图中缺失的解剖结构,确保显示的内容与用户的偏好一致。US 总 结 [10,17,33] 和探头指导 [32,22,30,15] 任务都很重要,并且已经被研究过。 然而,据我们所知,这是第一项引入 SG 和 LLM 来提升直观 US 解释和扫描 指导的工作。

2 方法论

在本节中,我们首先介绍了从颈动脉扫描获得的 US 图像的 SG 预测过程。然后,我们提供了关于如何将预测的 SG 集成到 LLM 中以促进 US 图像 理解任务的见解,如图 1 所示。

目标检测与场景图预测为了预测 US 图像的场景图,应定义格式为<entity1predicate – entity2>的三元组来捕捉扫描视野内关键解剖结构之间的相互

超声图像的语义场景图 3



图1. 所提框架的概述。

作用。对于横断面颈动脉扫描,我们选择五个代表性的解剖结构作为实体: "颈总动脉"(CCA)、"颈内静脉"(IJV)、气管上方的"软骨环"(CR)、"甲 状腺"(Th)和"椎体"(VB)。此外,我们将三种相互作用模式定义为谓词: "相邻","部分包裹",和"位于上方"。这些谓词可以有效地描述选定的五 个实体在 US 图像中的基于解剖结构的关系。

与传统的两阶段 SG 预测方法不同,本研究采用最先进的 RelTR [7]。 RelTR 遵循单阶段方法,可以同时检测实体并预测 SG,而不是将其视为独 立的顺序步骤。这种设计通过直接预测解剖结构之间的关系而无需依赖中间 程序来提高效率。如图 1(a) 所示,一旦 US 图像中的 SG 被预测出来,它可 以随后解析成多个三元文本(为了简单起见,称为三元组)。基于目标检测结 果和解剖学知识先验,我们可以确定扫描是在颈部的左侧还是右侧侧边。此 外,通过比较两个连续的目标解剖结构检测结果,我们还可以识别探头的横 向移动。在下游任务中,允许大语言模型在回应用户查询之前考虑当前成像 结果的基础提示将部分或完全由提取的三元组、侧边和横向移动信息组成。 美国图像摘要此任务旨在生成强调指定目标兴趣点的美国图像摘要。在此背 景下,给定用户查询中指明的关注实体,本地部署的大语言模型被要求提供 一个连贯的美国图像概述。该概述包括对图像的一般描述、关注区域及其与 相邻实体的关系。这些需求被封装到一个固定的任务指令提示中。为了将大 语言模型置于扫描循环中,如图 1(b-任务 I)所示,一个包含三元组和侧边 的接地提示被输入到大语言模型中以引导任务。这种方法允许大语言模型理 解用户的意图,并隐含地修剪三元组,仅保留与焦点实体直接相关的实体。 通过这样做,大语言模型可以生成连贯的句子,为用户提供个性化、直观的 解释,符合用户感兴趣的区域。该方法使即使是非专家用户也能更清晰地理 解 US 图像,并学习自己的解剖学知识。

美国扫描指南在之前的 SG 预测和超声图像摘要任务的基础上,预测的 SG 也可以用于提供扫描指导,帮助非专业用户操作便携式超声探头,在自我扫描过程中揭示当前成像视野中缺失的解剖结构。类似于超声图像摘要任务,用户关于希望扫描的解剖结构或超出成像视野的解剖结构的查询,以及用于超声扫描指导任务的新任务指令提示都将通过 LLM 进行处理。然而,为了完成该任务,除了三元组和侧面之外,基础提示还整合了横向移动来指示超声探头的相对运动方向。LLM 然后分析场景图以识别缺失的实体 [参见图 1(b-任务 II) 中的情况,其中 CR 不出现在当前 US 图像中]。通过利用 SG 提供的结构信息和横向移动指示的扫描运动数据,LLM 生成自然语言的运动指导,帮助用户找到与其偏好相符的成像视图。

3 实验与结果分析

3.1 实现细节

模型选择我们在实验中使用了与 RelTR [7] 类似的超参数。SG 预测网络在 配备 RTX 4080 Super GPU 的工作站上,以批处理大小为 16 的条件下训练 了 800 个周期。变压器的初始学习率为 10⁻⁴,在 200 个周期后减少至原来的 0.1 倍,以确保稳定收敛。由于这项工作旨在应用于便携式超声设备,我们优

先考虑如 LLaMa [29]、Qwen [3]、Gemma [28]、Mathstral⁴ 以及使用 Qwen 模型提炼的 DeepSeek R1 (DS-R1-Qwen) [9] 这类"轻量级"大语言模型。 这些轻量级模型是经过量化优化的版本,旨在提高效率,减少计算需求同时 保持性能,使它们更适合在资源受限环境中进行实时处理。此外,我们还使 用高容量的大语言模型作为参考模型,如 Gemini 2.0 Flash [27]和 Grok 3。 数据采集颈动脉超声图像使用配备有 12L3 线性探头的西门子柳兰超声机 (ACUSON Juniper, SIEMENS AG, 德国) 采集。成像和聚焦深度分别设 置为 45 毫米和 20 毫米。总共收集并标注了 289 张超声图像,训练集包含来 自五名志愿者的 262 张图像(分辨率: 829×770 像素)。此外,从**不同的志** 愿者处收集的另外 27 张图用于测试。由于缺乏成熟的情景图标注工具,我 们开发了一种专门针对我们的二维超声数据集的轻量级标注工具。值得注 意的是,除了目标检测标注外,情景图标注还需要标记三元组(<subject predicate - object>),这使得标注过程显著更加劳动密集型。此外,在SG 预测任务中,每个三元组中的谓词与图像中的实体之间的空间关系密切相 关。因此,传统计算机视觉任务中使用的标准数据增强技术大多不适用,水 平翻转是少数例外之一。这些挑战导致了相对较小的数据集规模。

3.2 目标检测和场景图预测

评估指标 为了评估目标检测性能,我们采用广泛使用的指标平均精度 (mAP)。具体来说,我们使用两种类型的平均精度 (AP): AP@50,在交并 比 (IoU) 阈值为 50%的情况下计算,以及 AP@[50:95],该指标在从 50%到 95%的多个 IoU 阈值上平均精度,提供更为严格的评估,如 COCO 评估协议 [19] 中所述。对于关系预测,我们采用召回率@K (R@K) [20] 和平均召 回率@K (mR@K) [6]。鉴于颈动脉超声数据集包含最多 7 种关系,我们将 K 分别设置为 5 和 20 进行评估。

鉴于所提出的方法旨在部署在资源受限的便携式 US 设备上,平衡模型 大小和预测性能至关重要,确保网络保持紧凑的同时维持最优准确率。为了 实现这一目标,我们进行了实验来评估 RelTR 变换器架构中编码器和解码 器层数量变化 [7] 对对象检测和 SG 预测的影响。表 1 概述了不同变换器编 码器和解码器层配置下的对象和关系检测性能。值得注意的是,四层模型实 现了最佳的整体性能,在目标检测中产生了最高的 AP@[50:95] (34.1%对第 二好的 32.4%)和 AP@50 (77.1%对比 70.3%)。它还在关系检测上展示了持

⁴ https://mistral.ai/

续优越的性能,除了与五层模型相比在 mR@5 上的微小 0.4%下降。虽然将 模型大小增加到五层稍微增强了关系检测的 mR@5 表现,但这导致了所有 其他评估指标中对象和关系检测性能的明显下降。鉴于有限的数据集,四层 编码器-解码器变换器在对象与关系检测之间取得了最佳平衡,使其成为该 设置下的最适选择。因此,我们采用此配置作为后续所有实验的默认设置。

| 层数 | 参数 | 目标检测 | | 关系检测 | | | |
|----------|-----|-----------------------|--------------------------------------|---------------|-------------------------|-------------------|-----------------|
| | | AP _{50:95} ? | $\uparrow \mathrm{AP}_{50} \uparrow$ | $R@5\uparrow$ | $\mathrm{mR@5}\uparrow$ | $ m R@20\uparrow$ | $mR@20\uparrow$ |
| 3 layers | 44M | 31.3 | 65.7 | 55.9 | 55.3 | 63.4 | 61.3 |
| 4 layers | 50M | 34.1 | 77.1 | 59.9 | 62.3 | 69.2 | 74.5 |
| 5 layers | 57M | 32.4 | 70.3 | 52.2 | 62.5 | 61.4 | 68.5 |

表 1. 对象检测和场景图检测的结果

3.3 场景图增强的 LLM 驱动 US 理解

评估指标为了评估 LLM 生成文本在超声图像总结和扫描指导任务中的 准确性 [参见图 1(b)],我们采用主观评价(称为 Acc)与客观指标相结合的 方法。主观评价由第三方专家进行,他们评估 LLM 是否准确遵循任务指令 提示并正确执行预期操作。另一方面,客观评估依赖于广泛使用的自然语言 处理指标,包括 METEOR [4] 和 ROUGE_L [18],这些指标衡量大型语言模 型生成的输出与参考文本之间的语言相似性。用于评估所需的参考文本是使 用 GPT-40 [1] 生成的,并经过人工验证以确保准确性和可靠性。

任务 I: 美国图像概要生成结果 (见表 2)显示,大规模模型在任务完成方面显 著优于轻量级量化模型,特别是 Grok 3,在所有指标中得分最高。在轻量级 模型中,Qwen 2.5 (14B)和 Gemma 2 (27B)表现相对较强,但在 METEOR (Grok: 0.880对比第二高分的 0.709)和 ROUGE_L (Grok: 0.841对比第二 高分的 0.641)指标上仍明显落后于大容量模型。这些发现清楚地表明了一 个趋势:随着模型规模的增大,改进后的推理能力增强了指令执行效果,使 得大规模模型成为处理更复杂和要求更高的任务时的首选。为了更加直观地 理解这一趋势,可以参考图 2 中不同 LLM 的 US 图像摘要以获取更多细节。

表 2. 不同 LLM 模型在任务 I 和任务 II 上的评估结果,使用准确率 (*Acc*)、METEOR 和 ROUGE_L。括号表示模型参数规模(例如,LLaMA 3.2 (1B) = 1B 个参数)。†:高容量的 LLMs。

| 構刑 | 任务 I | | | 任务II | | |
|-------------------------------|------------------------|---------|----------------------------|----------------|---------|----------------------------|
| 快王 | $\mathrm{Acc}\uparrow$ | METEOR↑ | $\mathrm{ROUGE}_L\uparrow$ | Acc \uparrow | METEOR↑ | $\mathrm{ROUGE}_L\uparrow$ |
| LLaMA 3.2(1B) | 0.265 | 0.550 | 0.387 | 0.408 | 0.392 | 0.300 |
| LLaMA 3.2(3B) | 0.531 | 0.534 | 0.365 | 0.347 | 0.403 | 0.390 |
| LLaMA 3.1(8B) | 0.735 | 0.489 | 0.335 | 0.633 | 0.395 | 0.290 |
| Mathstral $v0.1(7B)$ | 0.612 | 0.545 | 0.384 | 0.327 | 0.447 | 0.496 |
| DS-R1-Qwen(7B) | 0.551 | 0.590 | 0.576 | 0.265 | 0.400 | 0.478 |
| Qwen $2.5(14B)$ | 0.755 | 0.709 | 0.641 | 0.429 | 0.404 | 0.423 |
| Gemma $2(27B)$ | 1.000 | 0.590 | 0.615 | 0.469 | 0.401 | 0.508 |
| [†] Gemini 2.0 Flash | 0.980 | 0.589 | 0.736 | 0.592 | 0.452 | 0.623 |
| † Grok 3 | 1.000 | 0.880 | 0.841 | 0.776 | 0.490 | 0.665 |

任务 II: 扫描指导与任务 I 相比,任务 II 需要大语言模型具备更强的逻辑推理能力以确保准确执行。因此,所有模型在准确性、METEOR 和 ROUGE_L方面的整体得分都较低。在所有模型中,Grok 3 继续领先于其他模型(Acc: 0.776 对比第二好的 0.633),展示了其处理复杂推理任务的卓越能力。然而,轻量级量化模型如 Qwen 2.5 (14B)和 Gemma 2 (27B)保持了相对较高的准确性,类似于它们在任务 I 中的表现。因此,在资源受限的情况下,Qwen 2.5 (14B)和 Gemma 2 (27B)可以提供一种实用的解决方案;通过适当的量化进行本地部署,这两个模型分别只需 8GB 和 14GB 的 VRAM 即可执行推理。

4 讨论与结论

本研究引入了超声图像的语义场景图,为生理知识有限的人提供直观的 图像解释和有效的扫描指导。基于变压器的 RelTR 模型用于预测 US 图像 的语义 SG,捕捉关键解剖结构及其关系。这种信息丰富的 SG 进一步与最 近先进的 LLMs 结合使用,表明该系统可以帮助非专业人士更好地理解和分 析超声图像。虽然所提出的框架在颈动脉扫描中显示出有希望的结果,但当 前的方法仅在颈动脉图像上进行了验证。为了在不同解剖部位具有稳健的性 能,应收集包括不同解剖部位图像的大数据集。 8 Xuesong Li et al.



图 2. 由不同大语言模型在任务 I 中生成的美国图像摘要示例,关键医学术语被高亮显示 以便于清晰和比较。用灰色背景突出显示的文本表示包含显著错误的区域。

然而,本文展示了由 SG 引导的 LLM 框架在解释超声图像和提供扫描 指导方面具有广阔的前景。鉴于超声数据本身的稀缺性,所提出的框架为大 规模视觉语言模型(VLMs)提供了实用且创新的替代方案,这些模型需要 大量的超声数据集进行训练。此外,这些应用展示了促进解剖学和生理学知 识自我学习的巨大潜力,尤其是在年轻人中。

参考文献

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
- Andersen, C.A., Holden, S., Vela, J., Rathleff, M.S., Jensen, M.B.: Point-of-care ultrasound in general practice: a systematic review. The Annals of Family Medicine 17(1), 61–69 (2019)
- Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., et al.: Qwen technical report. arXiv preprint arXiv:2309.16609 (2023)
- Banerjee, S., Lavie, A.: Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In: Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization. pp. 65–72 (2005)
- Bi, Y., Jiang, Z., Duelmer, F., Huang, D., Navab, N.: Machine learning in robotic ultrasound imaging: Challenges and perspectives. Annual Review of Control, Robotics, and Autonomous Systems 7 (2024)
- Chen, T., Yu, W., Chen, R., Lin, L.: Knowledge-embedded routing network for scene graph generation. In: CVPR. pp. 6163–6171 (2019)
- Cong, Y., Yang, M.Y., Rosenhahn, B.: Reltr: Relation transformer for scene graph generation. IEEE Transactions on Pattern Analysis and Machine Intelligence 45(9), 11169–11183 (2023)
- Dhamo, H., Farshad, A., Laina, I., Navab, N., Hager, G.D., Tombari, F., Rupprecht, C.: Semantic image manipulation using scene graphs. In: CVPR. pp. 5213–5222 (2020)
- Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al.: Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948 (2025)
- Guo, X., Men, Q., Noble, J.A.: Mmsummary: Multimodal summary generation for fetal ultrasound video. In: MICCAI. pp. 678–688. Springer (2024)
- Huh, J., Park, H.J., Ye, J.C.: Breast ultrasound report generation using langchain. arXiv preprint arXiv:2312.03013 (2023)
- Im, J., Nam, J., Park, N., Lee, H., Park, S.: Egtr: Extracting graph from transformer for scene graph generation. In: CVPR. pp. 24229–24238 (2024)
- Islam, M., Seenivasan, L., Ming, L.C., Ren, H.: Learning and reasoning with the graph structure representation in robotic surgery. In: MICCAI. pp. 627–636. Springer (2020)

- 10 Xuesong Li et al.
- J. Y., Lu, J., Lee, S., Batra, D., Parikh, D.: Graph r-cnn for scene graph generation. In: ECCV. pp. 670–685 (2018)
- Jiang, Z., Bi, Y., Zhou, M., Hu, Y., Burke, M., Navab, N.: Intelligent robotic sonographer: Mutual information-based disentangled reward learning from few demonstrations. The International Journal of Robotics Research 43(7), 981–1002 (2024)
- Jiang, Z., Salcudean, S.E., Navab, N.: Robotic ultrasound imaging: State-of-the-art and future perspectives. Medical image analysis 89, 102878 (2023)
- Li, J., Su, T., Zhao, B., Lv, F., Wang, Q., Navab, N., Hu, Y., Jiang, Z.: Ultrasound report generation with cross-modality feature alignment via unsupervised guidance. IEEE Transactions on Medical Imaging (2024)
- Lin, C.Y.: Rouge: A package for automatic evaluation of summaries. In: Text summarization branches out. pp. 74–81 (2004)
- Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: Computer vision– ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13. pp. 740–755. Springer (2014)
- Lu, C., Krishna, R., Bernstein, M., Fei-Fei, L.: Visual relationship detection with language priors. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. pp. 852–869. Springer (2016)
- Maier-Hein, L., Vedula, S.S., Speidel, S., Navab, N., Kikinis, R., Park, A., Eisenmann, M., Feussner, H., Forestier, G., Giannarou, S., et al.: Surgical data science for next-generation interventions. Nature Biomedical Engineering 1(9), 691–696 (2017)
- Men, Q., Teng, C., Drukker, L., Papageorghiou, A.T., Noble, J.A.: Multimodalguidenet: Gaze-probe bidirectional guidance in obstetric ultrasound scanning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 94–103. Springer (2022)
- Murali, A., Alapatt, D., Mascagni, P., Vardazaryan, A., Garcia, A., Okamoto, N., Mutter, D., Padoy, N.: Latent graph representations for critical view of safety assessment. IEEE Transactions on Medical Imaging 43(3), 1247–1258 (2023)
- Ockenden, E.S., Frischer, S.R., Cheng, H., Noble, J.A., Chami, G.F.: The role of point-of-care ultrasound in the assessment of schistosomiasis-induced liver fibrosis: A systematic scoping review. PLOS Neglected Tropical Diseases 18(3), e0012033 (2024)

- Özsoy, E., Örnek, E.P., Eck, U., Czempiel, T., Tombari, F., Navab, N.: 4d-or: Semantic scene graphs for or domain modeling. In: MICCAI. pp. 475–485. Springer (2022)
- Sanner, A.P., Grauhan, N.F., Brockmann, M.A., Othman, A.E., Mukhopadhyay, A.: Voxel scene graph for intracranial hemorrhage. In: MICCAI. pp. 519–529. Springer (2024)
- Team, G., Anil, R., Borgeaud, S., Alayrac, J.B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A.M., Hauth, A., Millican, K., et al.: Gemini: a family of highly capable multimodal models. arXiv preprint arXiv:2312.11805 (2023)
- Team, G., Riviere, M., Pathak, S., Sessa, P.G., Hardin, C., Bhupatiraju, S., Hussenot, L., Mesnard, T., Shahriari, B., Ramé, A., et al.: Gemma 2: Improving open language models at a practical size. arXiv preprint arXiv:2408.00118 (2024)
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al.: Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023)
- Xu, H., Wu, J., Cao, G., Chen, Z., Lei, Z., Liu, H.: Transforming surgical interventions with embodied intelligence for ultrasound robotics. In: MICCAI. pp. 703–713. Springer (2024)
- Yuan, K., Kattel, M., Lavanchy, J.L., Navab, N., Srivastav, V., Padoy, N.: Advancing surgical vqa with scene graph knowledge. International Journal of Computer Assisted Radiology and Surgery 19(7), 1409–1417 (2024)
- Zhao, C., Droste, R., Drukker, L., Papageorghiou, A.T., Noble, J.A.: Uspoint: Selfsupervised interest point detection and description for ultrasound-probe motion estimation during fine-adjustment standard fetal plane finding. In: MICCAI. pp. 104–114. Springer (2022)
- Zhou, Y., Bi, Y., Tong, W., Wang, W., Navab, N., Jiang, Z.: Ultraad: Fine-grained ultrasound anomaly classification via few-shot clip adaptation (2025), https:// arxiv.org/abs/2506.19694