arXiv:2603.09733v1 [cs.CV] 10 Mar 2026

FetalAgents：用于胎儿超声图像与视频分析的多智能体系统

XiaotianHu[1, *], Junwei Huang[1, *], Mingxuan Liu[1, *], Kasidit Anmahapong[1],Yifei Chen[1], YitongLuo[1], Yiming Huang[2], Xuguang Bai[1], Zihan Li[1],Yi Liao[3], Haibo Qu[3],和 Qiyuan Tian[1, ✉]

[1] 清华大学，北京，中国[2] 加州大学圣迭戈分校，美国加利福尼亚州拉霍亚[3] 四川大学华西第二医院，四川大学，中国成都

qiyuantian@tsinghua.edu.cn

摘要。 胎儿超声（US）是产前筛查的主要成像方式，但其判读高度依赖临床医生的专业知识。尽管深度学习和基础模型不断取得进展，现有胎儿超声分析自动化工具仍难以兼顾特定任务的准确性与支持端到端临床工作流所需的全流程通用性。为解决这些局限，我们提出 FetalAgents，这是首个用于全面胎儿超声分析的多智能体系统。借助轻量级的智能体协同框架，FetalAgents 可动态调度专门的视觉专家，以提升诊断、测量和分割任务的性能。此外，FetalAgents 突破了静态图像分析的局限，支持端到端视频流摘要：系统可自动识别多个解剖切面中的关键帧，由协同工作的专家进行分析，并结合患者元数据生成结构化临床报告。针对八项临床任务开展的大规模多中心外部评估表明，与专用模型和多模态大语言模型（MLLM）相比，FetalAgents 始终表现出最稳健、最准确的性能，最终为胎儿超声分析与报告提供了一种可审计、契合工作流的解决方案。源代码见：https://github.com/huang-jw22/FetalAgents

关键词： 胎儿超声 · 医疗智能体 · 视频摘要。

1 引言

胎儿超声（US）是常规产前筛查和结构异常评估的基石成像手段 [32]。然而，其判读高度依赖检查者的专业能力，而经验丰富的超声医师十分稀缺

* 这些作者贡献相同。✉ 通讯作者。

这进一步加剧了医疗服务质量方面的差异 [36]。为克服这些局限，研究人员已广泛探索深度学习，以实现胎儿超声分析的自动化和标准化 [10]。具体而言，针对特定任务设计的架构——从用于切面分类和孕龄（GA）回归的残差网络 [14] 和视觉变换器 [8] ，到用于解剖结构分割和生物测量的 U-Net [31] 变体 [18] 及基于 SAM 的改进模型[21,24,45] ——已为各类单项胎儿超声子任务建立了准确度很高的自动化基线。近年来，FetalCLIP [28] 和 USFM [19] 等基础模型利用大规模预训练，提取可推广至多种下游临床任务的通用表征。然而，大多数自动化模型仍然专精于狭窄任务范围，且工作流程彼此割裂，通常针对单一子任务进行优化，并作为独立预测器部署，导致临床医生仍需手动协调工具选择，并将预测结果转化为结构化文档 [10,42]。

智能体系统的最新进展为克服这些局限提供了一种颇具前景的方法。通过为大语言模型（LLM）赋予规划、记忆和工具使用能力 [17], ，智能体系统能够协调专用模型执行多步骤工作流程，正如多学科会诊智能体 [20,5], 、纵向疾病追踪系统 [23,44], 和多模态医学影像框架 [9,27,37]所展示的那样。然而，现有智能体系统均未能满足胎儿超声的独特需求：其临床推理涉及多个解剖切面，紧密结合分类、分割和生物测量，并且必须处理时间连续的视频流，而非仅分析静态图像。

为弥补这一空白，我们提出了 FetalAgents，这是首个用于胎儿超声分析的多智能体系统。FetalAgents 通过一个轻量级、由语言驱动的协调框架，统筹多种领域专用模型；该框架能够动态解读临床意图、识别解剖学背景，并将任务分派给专门的专家模块，从而模拟人类超声医师的整体推理过程。此外，考虑到真实世界中的胎儿超声检查是一个连续的时间过程，FetalAgents 不仅支持静态图像分析，还能实现端到端的视频流摘要。具体而言，自动关键帧提取模块将连续扫描提炼为具有诊断意义的帧，再由协同工作的专家进行分析，并综合生成结构化临床报告。在多个中心的外部数据集上开展的大量实验表明，与专用视觉模型和多模态大语言模型相比，FetalAgents 在八项临床任务中始终表现最佳，验证了其标准化并自动化真实世界超声工作流程的潜力。

2 方法

2.1 FetalAgents 概述

FetalAgents 是一个协作式多智能体系统，专为全面的胎儿超声（US）分析而设计。该系统基于 AutoGen [40] 框架构建，将专用视觉工具与推理型大语言模型智能体相结合，弥合孤立的点预测模型与全面的

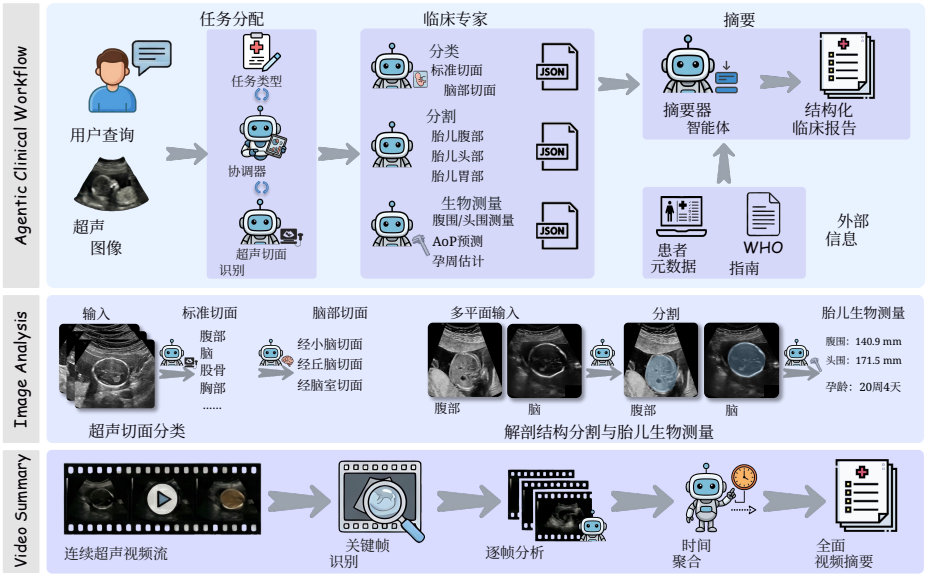

图1. FetalAgents 系统示意图。该框架由三个不同的模块组成：（1）智能体临床工作流，协调器在其中解读查询并分派任务；（2）图像分析，专业专家在其中执行特定视觉任务（分类、分割、生物测量）；（3）视频摘要，将多平面关键帧和时序发现汇总为一份全面的报告。

真实临床工作流之间的差距。为提供稳定可靠的胎儿超声分析，同时最大限度地降低幻觉风险，FetalAgents 将不同职责分配给三种专门的智能体类型：协调器、专家和摘要器：

（1）协调代理。基于 GPT-5-mini [30]，，协调器充当整个系统的智能中央编排者。给定输入的超声图像 I（或视频流 V）和用户查询 Q，协调器首先将查询分类为一种任务类型 τ（特定临床任务或报告生成），并识别标准超声切面 c。根据综合信息（Q、τ、c），它会将查询改写为结构化提示 P，并动态分派（I、P）给一组相关的专业专家。

(2) 专家代理。模块化专家代理专门用于处理特定的临床任务。每个专家 $\mathcal{E}_k$ 都将一组特定任务的视觉模型$\mathcal{M}_k = \{m_{k,1}, m_{k,2}, \ldots, m_{k,N}\}$ 封装为可调用工具，并应用确定性的融合规则 $\mathcal{F}_k$，以便在多个输出中生成最稳健的结果ˆ $o_k = \mathcal{F}(m_{k,\ 1}(I)\ldots\ldots,\ m_{k,\ N}(I))$。尤为重要的是，专家仅通过结构化 JSON 格式向上游传递信息，从而消除中间步骤中的主观幻觉。

（3）摘要代理。摘要器充当综合报告生成器。它汇总结构化输出和工具决策说明，这些内容来自

专家，以形成连贯、人类可读的报告。通过交叉核对相关任务的生物测量结果，摘要器在格式化最终回复之前充当最终的临床保障。

2.2 胎儿超声分析

为支持多样化的临床场景，FetalAgents 配备了一套强大的工具集，可处理三类主要的胎儿超声任务：

切面分类。准确识别标准胎儿切面是进行形态学诊断的关键前提。为此，FetalAgents 集成了四种互补模型的集成方案 $\mathcal{M}_{plane}$，用于识别标准胎儿视图和三个脑部子切面：(1) FetalCLIP [28]„ 一种能够生成胎儿超声图像通用表征的视觉语言基础模型；(2) FU-LoRA [38], ，一种通过低秩适配（LoRA） [16] 微调的潜在扩散模型；(3) 使用 RadImageNet [29] 权重初始化的 ResNet-50 主干网络；以及 (4) 使用 ImageNet-1K [7] 权重初始化的视觉变换器（ ViT-B/16） [8] 。

解剖结构分割。对胎儿头部、腹部和胃等解剖结构进行精细分割，可为下游生物测量提供关键的几何先验。对于胎头分割，系统集成了 nnU-Net [18], 、超声基础模型（USFM） [19], 和轻量级卷积神经网络 [43]。对于腹部的分割，我们使用 FetalCLIP 生成粗略的初始掩膜，随后提取其几何中心和边界框，作为 SAMUS [24], 的输入提示。SAMUS 是一种基于 SAM 的交互式超声分割模型。最后，对于胎儿胃部，我们采用由 nnU-Net、FetalCLIP 和 SAMUS 组成的稳健集成方案，以确保最高的分割保真度。

胎儿生物测量。胎儿生物测量对于追踪胎儿发育和指导产时护理至关重要。为自动测量胎头下降角（AoP），系统采用了 AoP-SAM [45] （一种专为耻骨联合和胎头分割定制的 SAM）、USFM 和改进后的 UperNet[41] 模型的集成。为精准估计孕周，我们将 FetalCLIP 与专用回归头相结合，并使用 ResNet-50 和 ConvNeXt-Tiny [25] 骨干网络对其进行微调。常规生物测量指标，包括头围（HC）和腹围（AC），均根据解剖结构分割专家生成的高保真掩码，通过椭圆拟合确定性地计算得出。

2.3 全面的临床工作流 ow

超声扫描具有多切面和时间序列的固有特性。传统模型通常仅处理单个静态帧和单项任务，导致工作流中的许多关键步骤仍需人工完成，并依赖操作者判断 [13]。FetalAgents 模拟真实临床工作流，将能力扩展至孤立任务之外，以支持全面的图像描述和视频流摘要。

图像描述生成。面对宽泛的请求，FetalAgents 会利用其内在的智能体智能来综合所见。协调器

首先识别解剖平面，并调用所有适用于该平面的专家。随后，摘要器将这些多方面的输出整合为一份结构化、全面的报告模板，确保所有可用所见都得到一致汇总。此外，在生成图像描述时，FetalAgents 会根据估计孕龄，使用世界卫生组织胎儿生长图表计算标准生长百分位数，以评估生物测量（腹围、头围）。重要的是，标准百分位数还可作为自动一致性保障措施。若测量值偏离合理的生长范围，摘要器便会对专家输出执行反思步骤，并动态地以更可靠的集成预测取代异常值，从而确保报告的准确性。

视频摘要。从超声视频流中手动提取关键帧并记录逐帧所见，劳动强度极大，且依赖操作者。$V = \{I_1, I_2, \ldots, I_T\}$ 为实现这一流程的自动化，我们利用 FetalCLIP 编码器骨干网络对六类关键帧识别模型进行了微调。随后，FetalAgents 部署该模型，自动提取一组具有诊断价值的关键帧 $\mathcal{K} = \{I_{k_1}, \ldots, I_{k_M}\} \subset V$，为逐帧分析分配合适的专家，最后将时序发现汇总为全面的视频级临床报告。借助这一端到端视频摘要流程，FetalAgents 有效模拟了真实世界中的超声医师工作流程；与现有僵化的视频摘要流程相比，它具有显著更强的适应性，且无需进行大量特定于工作流程的微调，即可顺畅适应多种临床场景。

3 实验

3.1 数据集和实验设置

训练集与外部验证数据集。FetalAgents 使用多个公开胎儿超声数据集进行训练，以满足不同任务的特定需求，包括 Fe- tal_PLANES_DB [3]（切面分类）、ACOUSLIC-AI [33] 和 [6] 的数据集（腹部和胃部分割）、PSFHS [4]（AoP预测）、HC18 [15]（头围测量）以及 PBF-US1 [11]（关键帧识别）。为评估分布外性能，FetalAgents 和对比方法在多个独立外部验证数据集上进行了评估，包括多中心非洲数据 [35,39]（标准切面分类、头围测量）、来自 [2] 的数据集（脑部平面分类、胃部分割）、[1]（腹部|||分割）、[26]（AoP预测），以及 HC18[15] 中一个包含 511 个病例的子集（孕周估计）。

基线与评估指标。FetalAgents 与专用视觉模型 [38,14,8,18,45,41,43,25], 、超声专用基础模型 [28,19], 、通用多模态大语言模型[30,12], 以及医学多模态大语言模型[34]进行了比较。对于切面分类，我们报告准确率、精确率、召回率、F1分数、Cohen's κ和 AUROC。对于解剖结构分割，我们采用 Dice相似系数（DSC）、交并比（IoU）、第 95 百分位豪斯多夫距离（HD95）、平均对称表面距离（ASSD）、阳性预测值（PPV）和敏感度

表1. 对分为三类的八项临床任务进行外部数据集定量评估。**粗体**：最佳。下划线：次佳。

方法	准确率	精确率	召回率	F1分数	Cohen's κ	AUROC
I. 切面分类						
标准切面分类（外部验证：*233* 病例）						
FetalCLIP[28]	0.914	0.942	0.915	0.927	0.887	99.7%
FU-LoRA[38]	0.764	**0.977**	0.759	0.847	0.717	98.0%
GPT-5-mini[30]	0.605	0.533	0.548	0.523	0.464	0.816
Gemini-3-flash[12]	0.880	0.894	0.843	0.856	0.857	0.961
MedGemma[34]	0.708	0.831	0.649	0.617	0.597	N/A
FetalAgents	**0.927**	0.937	**0.927**	**0.931**	**0.903**	**99.7%**
胎儿脑平面分类（外部验证：*569* 个病例）						
FetalCLIP[28]	0.893	0.880	0.911	0.883	0.830	**98.6%**
ResNet[14,29]	0.824	0.798	0.853	0.810	0.727	94.7%
ViT-16[8,7]	0.881	0.864	0.896	0.874	0.810	96.5%
GPT-5-mini[30]	0.689	0.644	0.632	0.638	0.518	0.789
Gemini-3-flash[12]	0.636	0.657	0.661	0.606	0.473	0.755
MedGemma[34]	0.691	0.723	0.693	0.684	0.509	N/A
FetalAgents	**0.909**	**0.884**	**0.925**	**0.896**	**0.856**	96.3%
II. 解剖结构分割						
方法	**DSC**	**IoU**	**HD95**	**ASSD**	**PPV**	**Sens**
腹部分割（外部验证：*187* 病例）						
FetalCLIP[28]	0.902	0.834	4.20	0.447	0.891	0.940
nnUNet[18]	0.848	0.748	129.2	15.20	0.809	0.912
FetalAgents	**0.937**	**0.937**	**2.32**	**0.202**	**0.922**	**0.955**
胃部分割（外部验证：*253* 个病例）						
FetalCLIP[28]	0.771	0.653	52.56	8.62	0.829	0.764
SAMUS[24]	0.826	0.730	14.07	5.68	0.797	0.893
nnUNet[18]	0.847	0.754	11.30	**1.95**	0.841	0.891
FetalAgents	**0.859**	**0.767**	**10.64**	3.62	**0.844**	**0.896**
III. 胎儿生物测量						
方法	**DSC**	**HD95**	**MAE**	**MdAE**	**中位数相对绝对误差**	**5%准确率**
耻骨弓角估计（外部验证：*64* 个病例）						
AoP-SAM[45]	0.927	38.0	5.59°	4.86°	3.54%	65.6%
USFM[19]	0.870	68.3	10.03°	8.16°	6.46%	43.8%
UperNet[41]	0.913	42.9	6.74°	6.31°	4.77%	53.1%
FetalAgents	**0.928**	**37.2**	**5.44°**	**4.68°**	**3.50%**	**67.2%**
腹围测量（外部验证：*187* 例）						
FetalCLIP[28]	0.902	4.20	9.94 毫米	6.59 毫米	3.77%	65.6%
nnUNet[18]	0.848	129.2	11.7 毫米	8.80 毫米	5.28%	47.1%
FetalAgents	**0.937**	**2.32**	**5.53毫米**	**4.29毫米**	**2.34%**	**84.0%**
方法	**DSC**	**HD95**	**MRAE**	**中位数相对绝对误差**	**第95百分位相对绝对误差**	**5%准确率**
头围测量（外部验证：*75* 例）						
CSM[43]	0.956	1.92	2.05%	1.48%	5.75%	89.3%
USFM[19]	0.924	11.54	3.21%	2.06%	13.0%	86.7%
nnUNet[18]	**0.964**	**1.72**	1.57%	1.21%	3.96%	**98.7%**
FetalAgents	0.961	1.93	**1.40%**	**1.12%**	**3.75%**	96.0%
方法	**有效性。**	**MAE**	**MdAE**	**MRAE**	**第95百分位绝对误差**	**5%准确率**
孕周预测（外部验证：*511*例）						
ResNet[14,29]	71.62%	1.73万	1.08万	6.74%	5.41万	52.1%
FetalCLIP[28]	82.97%	1.32w	0.72w	5.13%	4.69w	**68.9%**
ConvNext[25]	74.17%	1.53万	0.96万	5.95%	5.19万	56.75%
FetalAgents	**84.93%**	**1.23w**	**0.66w**	**4.70%**	**4.51w**	68.7%

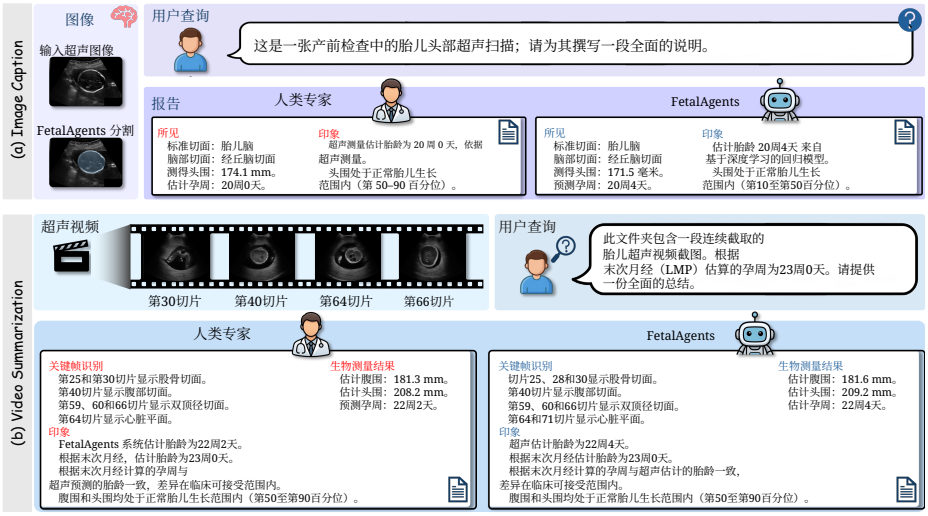

图2. 临床报告的定性评估。(a) 图像描述生成：FetalAgents 能正确识别标准切面并进行生物测量，生成与人类专家报告一致的报告。(b) 视频摘要：系统可自主从连续视频流中提取多平面关键帧，并将超声检查结果与患者元数据（末次月经，LMP）进行交叉核对，以确保临床一致性。

（敏感度）。对于胎儿生物测量，我们再次使用 DSC 和 HD95 评估分割质量，并通过平均绝对误差（MAE）、中位绝对误差（MdAE）、平均/中位相对绝对误差（MRAE/MdRAE）、第95百分位绝对/相对误差（P95 AE/P95 RAE），以及预测值与真实值相差不超过 5% 的比例（5%准确率，Acc@5%）量化测量准确度。特别对于孕周（GA）预测，根据 [28], ，我们引入了额外的有效率指标，用于检验真实头围（HC）是否落在世界卫生组织胎儿生长图表 [22], 中预测孕周对应的第2.5至第97.5百分位范围内，从而根据临床依据对估计孕周的合理性进行验证。

3.2 定量结果

在外部验证中，FetalAgents 在 8 项临床任务上始终取得最准确、最稳定的表现（表1），证实了智能体编排和动态模型集成的有效性。具体而言，(1) 切面分类：通用及医学多模态大语言模型难以应对噪声较大的超声模态，GPT-5-mini 在标准切面上的准确率仅为 0.605。通过将图像动态路由至专用视觉主干网络，FetalAgents 的表现优于最强的独立模型 FetalCLIP：标准切面分类准确率从 0.914 提高至 0.927，Cohen's κ 从 0.887 提高至 0.903，

在标准切面分类任务中。在更具挑战性的脑部亚平面任务中，FetalAgents 的 F1分数达到 0.896， κ 0达到 0.856，分别比 FetalCLIP 提高 1.3% 和 2.6%。(2) 解剖结构分割：我们的融合规则有效纠正了单体模型常见的边界误差。在腹部|||分割任务中，FetalAgents 的 Dice相似系数达到 0.937，95%豪斯多夫距离为 2.32；与排名第二的 FetalCLIP 相比，Dice相似系数提高了 3.5%，95%豪斯多夫距离改善了 44.8%。在胃分割任务中，FetalAgents 的 Dice相似系数达到 0.859，95%豪斯多夫距离为 10.64，分别优于 nnUNet 1.2% 和 5.8%。(3) 胎儿生物测量：FetalAgents 在所有生物测量任务中均取得最低误差。在腹围测量任务中，其平均绝对误差从 FetalCLIP 的 9.94 mm 降至 5.53 mm，5%准确率则从 65.6% 提高至 84.0%。在头围测量任务中，其平均相对绝对误差为 1.40%，P95RAE 为 3.75%；相较 nnUNet，相对误差分别降低 10.8% 和 5.3%。在孕周预测任务中，FetalAgents 的有效率最高，达到 84.93%，平均绝对误差最低，为 1.23 周；与 FetalCLIP 相比，有效率提高 2.0 个百分点，平均绝对误差降低 6.8%。在耻骨弓角估计任务中，其平均绝对误差为 5.44°，5%准确率为 67.2%，始终优于所有独立模型。

3.3 图像描述生成与视频摘要结果

为展示 FetalAgents 超越孤立指标的端到端临床实用性，我们对其报告生成能力进行了定性评估。(1) 图像描述生成：面对一个宽泛的问题和一张超声图像，FetalAgents 正确识别出经丘脑平面，协调相关专家进行生物测量和孕周估计，并将所有输出整合为结构化临床报告（图 2 (a)）。(2) 视频摘要：面对包含患者元数据的连续多平面视频流，FetalAgents 自动从不同解剖结构中提取关键帧，汇总逐帧生物测量数据，并将推算出的孕龄与基于末次月经推算的孕龄进行交叉比对，以确认临床一致性（图 2 (b)），这与真实世界中的超声医师工作流程十分相似。

4 结论

本文介绍了 FetalAgents，这是首个弥合孤立的点预测模型与胎儿超声工作流程复杂现实之间差距的多智能体系统。通过在智能体框架中利用大语言模型固有的智能和工具使用能力，我们的系统可动态协调多种专用视觉骨干网络，开展准确的胎儿超声分析。FetalAgents 自动完成连续超声视频摘要这一耗时的流程，成功突破了静态图像解读的局限，并自动生成结构化的专家级临床报告。广泛的外部评估证实，在多种临床任务中，FetalAgents 的表现始终优于最先进的独立视觉模型和通用多模态大语言模型。未来的工作将重点整合更广泛的患者纵向病史，并在前瞻性临床环境中评估该系统，以进一步验证其影响。

参考文献

1. Ashkani Chenarlogh, V., Ghelich Oghli, M., Shabanzadeh, A., 等：用于胎儿超声图像分割的快速、准确 U-Net 模型。《超声成像》**44**(1), 25–38（2022年） 2. Belciug, S.：胎儿切面与器官（2024年）。https://doi.org/10.5281/zenodo. 14093338 3. Burgos-Artizzu, X.P., Coronado-Gutierrez, D., Valenzuela-Alcaraz, B., 等：Fe- t al_切面_db：常见母胎超声图像（2020年）。https://doi. org/10 3904280 4 5281. / zenodo. . Chen, G., Bai, J., Ou, Z., 等：PSFHS：用于耻骨联合和胎儿头部基于人工智能分割的产时超声图像数据集。《科学数据》**11**, 436（2024年） 5. Chen, K., Li, X., Yang, T., 等：Mdteamgpt：一种基于大语言模型、自我演进的多智能体框架，用于多学科团队医疗会诊。arXiv 预印本 arXiv:2503.13856（2025年） 6. Da Correggio, K.S., Noya Galluzzo, R., Santos, L.O., 等：基于超声图像的胎儿腹部结构分割数据集（2023年）。https://doi.org/ 10 1 7 17632 4 9 3 . / gcpm dsc . . Deng, J., Dong, W., Socher, R., 等：ImageNet：大规模分层图像数据库。载于：CVPR。第248–255页（2009年） 8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., 等：一张图像胜过16×16个词：用于大规模图像识别的Transformer。载于：ICLR（2021年） 9. Fallahpour, A., Ma, J., Munim, A., 等：Medrax：用于胸部X线的医学推理智能体。arXiv 预印本 arXiv:2502.02673（2025年） 10. Fiorentino, M.C., Villani, F.P., Di Cosmo, M., 等：胎儿超声图像分析深度学习算法综述。《医学图像分析》**83**, 102629（2023年） 11. González, D., 等：Natalia：PBF-US1（胎儿超声扫描体模盲扫）（2024 14193949 12 10 5281 3 2026年）。https://doi.org/ . /zenodo. . 谷歌：Gemini Flash。大语言模型（） 13. Guo, X., Men, Q., Noble, J.A.：Mmsummary：胎儿超声视频的多模态摘要生成。载于：MICCAI（2024年） 14. He, K., Zhang, X., Ren, S., 等：用于图像识别的深度残差学习。载于：CVPR。第770–778页（2016年） 15. van den Heuvel, T.L.A., de Bruijn, D., de Korte, C.L., van Ginneken, B.：使用二维超声图像自动测量胎儿头围（2018年）。https: //doi.org/10 1327317 16 5281. /zenodo. . Hu, E.J., 等：LoRA：大语言模型的低秩适配。ICLR（2022年） 17. Hu, X., Qian, Y., Yu, J., 等：医疗智能体概况：综述。TechRxiv（2026年） 18. Isensee, F., 等：nnU-Net：一种用于基于深度学习的生物医学图像分割的自配置方法。《自然·方法》**18**, 203–211（2021年） 19. Jiao, J., Zhou, J., Li, X., 等：USFM：一种通用超声基础模型，可推广至不同任务和器官，实现高标签效率的图像分析。《医学图像分析》**96**, 103202（2024年） 20. Kim, Y., Park, C., Jeong, H., 等：MDAgents：用于医疗决策的多大语言模型自适应协作。arXiv 预印本 arXiv:2404.15155（2024年） 21. Kirillov, A., 等：Segment Anything。arXiv 预印本 arXiv:2304.02643（2023年） 22. Kiserud, T., Piaggio, G., Carroli, G., 等：世界卫生组织胎儿生长曲线：一项关于超声生物测量和估计胎儿体重的多国纵向研究。《公共科学图书馆·医学》**14**(1), e1002220（2017年）

23. Li, R., Wang, X., Berlowitz, D., 等：Care-ad：一种利用纵向临床记录预测阿尔茨海默病的多智能体大语言模型框架。npj Digit. Med. **8**(1), 541（2025年）24. Lin, X., 等：超越 SAM 适配：通过自动提示实现端到端超声图像分割。arXiv 预印本 arXiv:2309.06824（2023年）25. Liu, Z., Mao, H., Wu, C.Y., 等：面向 2020 年代的卷积网络。CVPR（2022年）26. Lu, Y., Zhou, M., Zhi, D., 等：用于耻骨联合-胎头分割的 JNU-IFM 数据集。Data Brief **41**, 107904（2022年）27. Lyu, X., Liang, Y., Chen, W., 等：Wsi-agents：用于多模态全视野切片图像分析的协作式多智能体系统。载于：MICCAI。第 680–690 页（2025年）28. Maani, F., Saeed, N., Saleem, T., 等：Fetalclip：用于胎儿超声图像分析的视觉语言基础模型。arXiv 预印本 arXiv:2502.14807（2025年）29. Mei, X., 等：Radimagenet：用于有效迁移学习的开放放射学深度学习研究数据集。Radiol. Artif. Intell. **4**(5), e210315（2022年）30. OpenAI：GPT-5 mini。大语言模型（2026年）31. Ronneberger, O., Fischer, P., Brox, T.：U-Net：用于生物医学图像分割的卷积网络。载于：MICCAI。第 234–241 页（2015年）32. Salomon, L.J., 等：常规孕中期胎儿超声扫描实施指南。Ultrasound Obstet. Gynecol. **37**(1), 116–126（2011年）33. Sappia, M., de Korte, C., van Ginneken, B., 等：ACOUSLIC-AI 挑战赛报告：来自低收入国家的盲扫超声数据中的胎儿腹围测量。Med. Image Anal. **105**, 103640（2025年）34. Sellergren, A., Kazemzadeh, S., Jaroensri, T., 等：Medgemma 技术报告。arXiv 预印本 arXiv:2507.05201（2025年）35. Sendra-Balcells, C., Campello, V.M., Torrents-Barrena, J., 等：来自五个非洲国家低资源成像环境的母胎超声切面（2023年）。https://doi.org/10 7540448 36 5281. /zenodo. . Sippel, S., Muruganandan, K., Levine, A., 等：综述文章：超声在发展中世界的应用。Int. J. Emerg. Med. **4**, 72（2011年）36. Vaidya, S., 等：MedPAO：用于结构化医疗报告的协议驱动型智能体。载于：AI for Clinical Applications。第 33–45 页（2025年）38. Wang, F., Whelan, K., Silvestre, G., 等：生成式扩散模型助力代表性不足的非洲人群胎儿超声图像的零样本分类。arXiv 预印本 arXiv:2407.20072（2024年）39. Wang, F., Silvestre, G., Curran, K.M.：在低资源环境中采用高效微调策略分割胎儿头部：一项基于 U-Net 的实证研究。arXiv 预印本 arXiv:2407.20086（2024年）40. Wu, Q., Bansal, G., Zhang, J., 等：AutoGen：通过多智能体对话赋能新一代大语言模型应用。arXiv 预印本 arXiv:2308.08155（2023年）41. Xiao, T., Liu, Y., Zhou, B., 等：用于场景理解的统一感知解析。arXiv 预印本 arXiv:1807.10221（2018年）42. Yan, L., Li, Q., Fu, K., 等：人工智能在超声辅助医学诊断中的应用进展。Bioengineering **12**(3), 288（2025年）43. Zeng, W., Luo, J., Cheng, J., 等：利用轻量级深度卷积神经网络实现高效胎儿超声图像分割，以自动测量头围。Med. Phys. **49**(8), 5081–5092（2022年）44. Zhao, W., Wu, C., Fan, Y., 等：一种具有可追溯推理能力的罕见病诊断智能体系统。Nature（2026年）45. Zhou, Z., Lu, Y., Bai, J., 等：用于产时超声图像分析中胎头-耻骨联合分割的 Segment Anything Model。Expert Syst. Appl. **263**, 125699（2025年）