胎儿脑无监督异常检测的简单基线方法

Hongjia Yang[1,‡], *Mingxuan Liu*[1,‡,*], *Yingqi Hao*[1,‡], *Xiaotian Hu*[1], *Siyao Li*[1], *Xuguang Bai*[1], *Fenglin Jia*[2], *Yijin Li*[1], *Yifei Chen*[1], *Zihan Li*[1], *Juncheng Zhu*[2], *Jing Liu*[2], *Yi Liao*[2], *Ziyu Li*[3], *Haibo Qu*[2], *Qiyuan Tian*[1,*]1清华大学清华大学医学部生物医学工程学院，中国北京2四川大学华西第二医院放射科，中国成都3 牛津大学纳菲尔德临床神经科学系，英国牛津

arXiv:2611.28505v1 [cs.CV] 21 Nov 2026

摘要—**无监督异常检测（UAD）为胎儿脑磁共振成像（MRI）分析提供了一种很有前景的方法，尤其适用于病理数据稀缺的情况。现有方法主要依赖脑龄预测或图像重建，但受到胎龄（GA）标签噪声、缺乏病灶定位能力或跨域泛化能力较差等问题的限制。我们提出 FetalBUAD，这是一种基于体素级脑龄估计的 UAD 框架。联合多任务网络预测全局脑龄、体素级脑龄图和脑分割结果，以建立内部生物学脑龄参考，从而减少对临床记录的胎龄的依赖。随后，利用不确定性感知的体素级脑龄差图进行异常定位。在 1,317 个内部胎儿脑磁共振成像体积和一个外部基准数据集上的实验表明，该方法实现了先进的异常检测性能，并提升了跨域泛化能力。FetalBUAD 还可生成解剖学可解释图，用于定位多种胎儿脑异常。我们的代码已公开发布于：https://github.com/birthlab/FetalBUAD。**

索引词—**胎儿脑磁共振成像、脑龄预测、无监督异常检测。**

I. 引言

产前脑发育是一个关键阶段，为终身神经发育健康奠定基础[1]。偏离正常发育轨迹与长期神经系统疾病相关，包括癫痫 [2], 、智力障碍 [3], 和自闭症谱系障碍 [4]。准确识别胎儿脑部发育异常对于及时开展产前干预和咨询至关重要。超声检查是主要的筛查方式，而胎儿磁共振成像（MRI）具有更优的软组织对比度和更大的视野 [5], [6],，能够全面评估复杂的颅内结构。

深度学习在胎儿脑异常诊断方面展现出潜力 [7]。然而，开发稳健的监督式诊断模型仍然面临挑战。由于许多先天性脑疾病的患病率较低、数据收集受到伦理约束，且专家标注成本高昂，病理性胎儿磁共振成像数据十分稀缺。此外，胎儿脑异常在解剖学表现和严重程度上存在显著差异，许多罕见疾病仅有少数病例记录。这些挑战妨碍了监督式模型对胎儿脑异常全貌的刻画，并限制了其临床适用性。

‡ 共同第一作者。*通讯作者：qiyuantian@tsinghua.edu.cn, arktisx@foxmail.com。

无监督异常检测（UAD）提供了另一种方案，无需病理标注即可学习正常的胎儿脑发育。一种方法将异常检测建模为脑龄预测，并将预测脑龄与临床记录的胎龄（GA）之间的差异作为异常分数 [8]–[11]。然而，这种策略存在两个根本局限。首先，由于它只估计整体年龄差，因此无法进行体素级定位。其次，其可靠性受到临床胎龄标签不准确的制约。尽管胎龄通常根据孕早期超声检查的头臀长测量值进行估算 [12],，测量不确定性和胎儿生长的个体间变异可能导致误差高达 ±5–7 天 [13]。因此，脑龄差可能反映的是临床定龄不准确，而非真正的发育异常，从而降低基于脑龄差的异常检测的可靠性。

另一大类 UAD 方法通过检测输入图像与学习到的正常表征之间的差异 [14]–[18]来发现异常，这些差异存在于图像或特征空间中。这些方法假设，仅使用正常数据训练的模型无法忠实地重现异常结构，因此在病理区域会出现较大的重建差异或特征差异。虽然这类方法对明显的强度异常效果显著，但当结构异常仅导致细微的外观变化时，其表现可能较差。此外，学习到的正常表征取决于成像特征和

容易受到不同扫描仪、采集方案和机构之间域偏移的影响，从而限制了泛化能力。

为解决这些局限性，我们提出了**FetalBUAD**，一种基于体素级生物学脑龄估计的胎儿脑异常检测框架。**FetalBUAD** 并不直接将预测结果与可能不准确的临床孕龄标签进行比较，而是通过联合多任务网络建立内部生物学年龄参考值，该网络可预测全局脑龄、体素级脑龄图以及脑分割 [19], [20]。在推理过程中，蒙特卡洛（MC）随机失活 [21]用于估计认知不确定性，以此对体素级脑龄差图相对于生物学参考值进行加权。在大规模内部和外部数据集上的实验表明，**FetalBUAD**实现了先进的异常检测性能、出色的跨域泛化能力，以及具有解剖学可解释性的病灶定位。

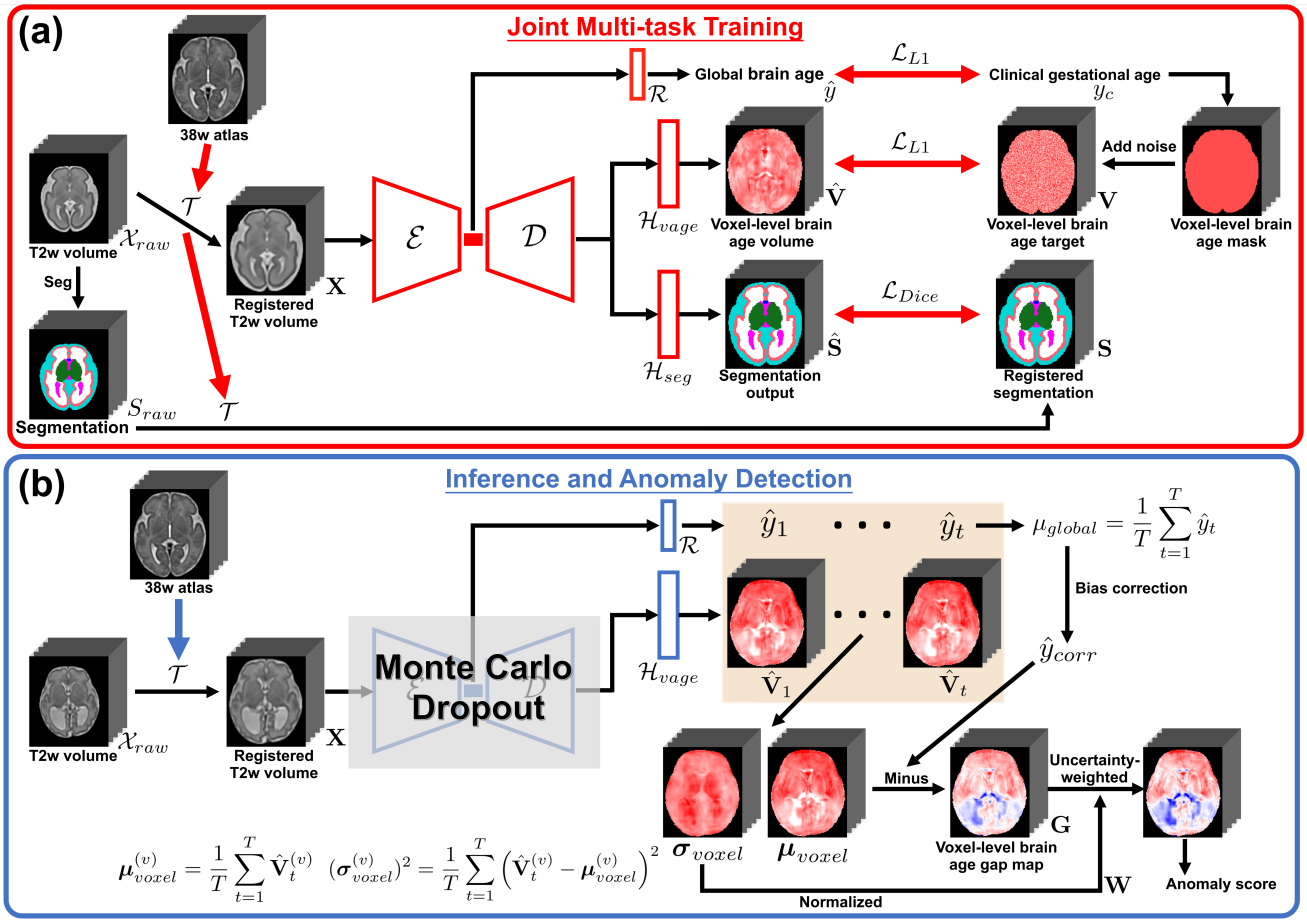


图 1. FetalBUAD 概览。(a) 联合多任务训练：共享编码器-解码器网络从正常胎儿脑部磁共振图像中预测全局脑龄、体素级脑龄和脑分割结果。对体素级目标施加噪声扰动可促进局部形态学习，而分割则提供解剖约束。(b) 推理与异常检测：蒙特卡洛Dropout估计全局脑龄和体素级脑龄预测的预测均值与认知不确定性。偏差校正后的全局脑龄作为生物学参考值，并使用归一化不确定性对体素级脑龄差进行加权，以生成最终的异常图。

II. 相关工作

现有的胎儿脑部无监督异常检测方法大致可分为两种范式：基于脑龄预测的方法[8]–[11] 和基于重建的方法 [14]–[18]。然而，这两类方法都未能充分满足可解释的发育评估和空间精确病灶定位的需求，尤其是对于细微且解剖学异质性较大的异常。

A. 基于脑龄预测的方法

基于脑龄预测的方法对正常神经发育进行建模，并通过预测脑龄与临床记录的胎龄之间的差异识别异常病例。Shi 等人提出了一种基于注意力机制的深度集成模型，用于根据常规 T2加权磁共振成像进行胎儿脑龄估计，并将预测不确定性和估计置信度作为异常标志 [8]。后续研究将二维单通道卷积神经网络应用于多平面胎儿磁共振图像切片，通过预测年龄差异检测脑室扩大及相关中枢神经系统异常 [9]。一种轻量级三维网络实现了全脑龄估计，并凸显了偏差校正在脑龄差估计中的重要性 [10]。最近，PANDA 将年龄预测扩展到局部图像块，并将各图像块中最大的绝对年龄差作为异常生物标志物，从而实现粗略的病灶定位[11]。然而，现有方法主要生成病例级评分，无法直接在体素级定位异常。其异常分数也会受到临床记录的胎龄标签中噪声的影响。

B. 基于重建的方法

基于重建的方法通过图像重建或特征表示差异来识别异常。标准三维自编码器和去噪自编码器利用受损输入的逐像素重建误差检测异常区域 [15], [16]。Iter Mask3D利用重建误差和高频结构引导，迭代细化指向异常的空间掩膜，以减少正常区域中的假阳性 [18]。特征级教师-学生方法则通过教师与学生表示之间的差异来检测异常，

而Skip-TS采用带有跳跃连接和多尺度异常一致性的异构编码器-解码器反向蒸馏 [14],；三维学生-教师特征金字塔匹配则将自监督图块学习、ACS卷积和多尺度特征匹配应用于体积MRI [17]。

这些方法会生成空间异常图，但其依据主要反映图像恢复或特征层面的差异，而非发育语义。高容量重建模型可能会复现异常结构，而更强的破坏或掩蔽则可能增加正常组织中的假阳性。此外，这些依赖外观的表征容易受到胎儿运动、对比度变化、成像伪影和跨站点差异所导致的域偏移影响。因此，这些方法能够定位外观层面的偏差，却不会明确评估局部胎儿脑结构在发育方面是否适当。

III. 方法

A. 数据预处理

令 $\mathcal{X}_{raw}$ 表示T2加权胎儿脑MRI体积。采用公开可用的胎儿脑分割模型 [22]生成相应的脑分割结果$\mathcal{S}_{raw}$。随后，通过仿射变换 $\mathcal{T}$ 将 $\mathcal{X}_{raw}$ 配准至38周胎儿脑图谱体积 [23]。具体而言，该变换使用ANTsPy[24]实现，仅涉及刚性旋转和各向同性缩放，不包含任何非线性形变。这种对齐将全局脑大小归一化至统一且一致的尺度，从而促使网络在预测脑龄时优先考虑内在形态模式，而非绝对脑体积。相应的脑分割结果使用相同的变换 $\mathcal{T}$ 进行配准，并采用最近邻插值作为训练目标$\mathbf{S} = \mathcal{T}(S_{raw})$。最后，将 $\mathcal{X}_{raw}$ 进一步归一化为零均值和单位方差，作为最终的网络输入 $\mathbf{X} = \mathcal{N}(\mathcal{T}(\mathcal{X}_{raw}))$。

B. 联合多任务网络训练

1) 网络架构： 本文提出了一种多任务框架，用于同时预测全局脑龄、体素级脑龄体积以及脑分割结果。具体而言，网络架构由共享编码器 $\mathcal{E}(\cdot)$、全局回归头 $\mathcal{R}(\cdot)$、双用途解码器 $\mathcal{D}(\cdot)$、体素级估计头 $\mathcal{H}_{vage}(\cdot)$ 和分割头 $\mathcal{H}_{seg}(\cdot)$ 组成。

给定输入 $\mathbf{X}$，编码器首先通过经过随机失活正则化的层提取紧凑的潜在瓶颈表示 $\mathbf{z} = \mathcal{E}(\mathbf{X})$，随后将其输入两个并行路径进行处理：(1) 全局年龄预测路径：全局回归头 $\mathcal{R}$ 将 $\mathbf{z}$ 映射为标量全局脑龄预测，记为 $\hat{}y = \mathcal{R}(\mathbf{z})$。(2) 密集体素级路径：解码器 $\mathcal{D}$ 对 $\mathbf{z}$ 进行上采样，以重建高分辨率特征图，这些特征图再分为两个卷积分支：体素级估计头预测体素级脑龄体积 $\hat{}\mathbf{V} = \mathcal{H}_{vage}(\mathcal{D}(\mathbf{z}))$，而分割头 $\hat{}$ 则预测脑分割结果 $\mathbf{S} = \mathcal{H}_{seg}(\mathcal{D}(\mathbf{z}))$。

2) 训练目标： 对于输入 $\mathbf{X}$ 及其临床记录的胎龄 GA y_c，体素级脑龄回归目标$\mathbf{V}$ 定义如下：

$$\mathbf{V}^{(v)} = (y_c + \epsilon^{(v)}) \cdot \mathbb{M}_{\mathbf{S}}^{(v)}, \quad \text{where } \epsilon^{(v)} \sim \mathcal{U}(-\delta, \delta), \tag{1}$$

$\epsilon^{(v)}$ 是从小范围内的均匀分布中采样得到的随机噪声项，对每个体素 v 独立采样。 $\mathbb{M}_{\mathbf{S}}$表示由配准分割 $\mathbf{S}$ 得到的二值脑掩膜。这种扰动策略引入细微变化，促使网络捕获精细的局部纹理变化，而非收敛到平凡解 [20]。将噪声幅度 δ 限制在较小范围内，可确保扰动不会破坏整体发育趋势，同时有效避免梯度流完全相同。

通过最小化复合损失函数来优化网络：

$$\mathcal{L}_{total} = \mathcal{L}_{L1}(\hat{y}, y_c) + \lambda_{vage}\mathcal{L}_{L1}(\hat{\mathbf{V}}, \mathbf{V}) + \lambda_{seg}\mathcal{L}_{Dice}(\hat{\mathbf{S}}, \mathbf{S}), \tag{2}$$

其中， $\mathcal{L}_{L1}$ 表示 L1 损失， $\mathcal{L}_{Dice}$ 表示 Dice 损失，λ_{vage} 和 λ_{seg} 是用于将各项损失缩放到相同数量级并平衡其贡献的超参数。引入分割任务可作为一种解剖约束，促使共享编码器学习空间一致的表征。

3) 对临床记录噪声 *GA* 的鲁棒性：临床孕龄标签 y_c 可能包含测量噪声和系统误差。令 y^* 表示真实的生物学 GA，其中

$$y_c = y^* + \eta, \quad \mathbb{E}[\eta \mid \mathbf{X}] \approx 0, \tag{3}$$

其中 η 表示测量噪声。全局脑龄预测器 $\hat{}y = \mathcal{R}(\mathcal{E}(\mathbf{X}))$ 使用 L1 损失进行训练，其最优预测器估计临床孕龄的条件中位数：

$$\hat{y}^* = \arg\min_{\mathcal{R}\circ\mathcal{E}} \mathbb{E}_{\mathbf{X},y_c}\left[|\mathcal{R}(\mathcal{E}(\mathbf{X})) - y_c|\right] \approx \text{Median}(y_c \mid \mathbf{X}). \tag{4}$$

在测量噪声近似对称的情况下，该估计可以捕捉潜在的生物学 GA，同时对临床标签中的变异具有鲁棒性。

C. 推理与异常检测

1) 通过蒙特卡洛 *Dropout* 进行不确定性估计：为实现稳健的异常检测，我们在推理阶段使用 MC Dropout 近似贝叶斯神经网络。通过保持随机失活处于激活状态，针对给定输入 $\mathbf{X}$ 执行 T 次随机前向传播。随后，分别计算全局脑龄估计和体素级脑龄估计的预测均值与认知不确定性。对于全局脑龄，聚合均值 μ_{global} 及其对应的方差 σ^2_{global} 如下：

$$\mu_{global} = \frac{1}{T}\sum_{t=1}^{T}\hat{y}_t, \quad \sigma^2_{global} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}_t - \mu_{global}\right)^2. \tag{5}$$

同样，对于体素级脑龄，平均年龄图$\boldsymbol{\mu}_{voxel}$和不确定性图$\boldsymbol{\sigma}_{voxel}$的计算方式如下：

$$\boldsymbol{\mu}_{voxel}^{(v)} = \frac{1}{T}\sum_{t=1}^{T}\hat{\mathbf{V}}_t^{(v)}, \quad (\boldsymbol{\sigma}_{voxel}^{(v)})^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\mathbf{V}}_t^{(v)} - \boldsymbol{\mu}_{voxel}^{(v)}\right)^2, \tag{6}$$

$\hat{}y_t$和$\mathbf{V}_t$分别表示在主动丢弃下第 t 次随机前向传播输出的标量全局脑龄和体素级脑龄体积。

2) 偏差校正：由于模型正则化以及数据集 [10], 中孕龄的非高斯分布，网络预测呈现出预期的回归稀释，表现为胎儿脑龄差与生物学孕龄之间出现虚假相关。

使用线性偏差校正策略 [19], [25] 对预测的全局脑龄均值进行调整。具体而言，在内部验证集上拟合线性模型$\mu_{global} = \alpha y_c + \beta$。在对未见数据进行推理时，使用固定参数 α 和 β 计算校正后的全局脑龄$\hat{}y_{corr}$：$\hat{}y_{corr} = (\mu_{global} - \beta)/\alpha$。

偏差校正仅应用于全局脑龄预测，不调整体素级年龄图。所提出的目标随机化策略（Eq.1）从根本上抑制体素级回归器收敛到均值，从而隐式缓解体素级回归稀释。

3) 异常检测指标：体素级脑龄差图 $\mathbf{G}$ 定义为均值体素级脑龄体积与校正后的全局脑龄之间逐体素的差值：$\mathbf{G}^{(v)} = \boldsymbol{\mu}_{voxel}^{(v)} - \hat{}y_{corr}$。该策略确保$\mathbf{G}$仅反映结构偏差。$\mathbf{G}$中的正值表示区域成熟度较高，而负值则表示发育迟缓。

由于偏离正常训练流形，病理区域通常会表现出较高的认知不确定性。为利用这一特性，我们定义了一种不确定性加权偏差指数，根据局部不确定性放大$\mathbf{G}$的幅度，最终异常分数定义为放大后正偏差的离散程度：

$$\text{Score} = \text{Std}\left(\{\mathbf{G}^{(v)} \cdot \mathbf{W}^{(v)} \mid \mathbf{G}^{(v)} \cdot \mathbf{W}^{(v)} > 0\}\right), \tag{7}$$

其中，$\mathbf{W}^{(v)} = 1 + (\boldsymbol{\sigma}_{voxel}^{(v)} - \min(\boldsymbol{\sigma}_{voxel}))/(\max(\boldsymbol{\sigma}_{voxel}) - \min(\boldsymbol{\sigma}_{voxel}))$ 是映射到范围 [1, 2]的归一化权重因子，可在推理过程中自适应地强调高不确定性体素。仅保留正偏差，因为结构异常往往会导致全局脑龄被低估，使病理病例相较于正常病例呈现显著的正脑龄差。

IV. 数据集与实现

A. 数据集

1) 训练集与内部测试集：在机构审查委员会批准下，回顾性收集了来自单一中心的 1,400 名孕妇（孕周：20–39 周）的 T2加权（T2w）厚层胎儿脑部MRI扫描，使用了三台扫描仪（1.5T 和 3T）。使用 NiftyMIC [26],[27] 重建了三维脑体积。经过临床医生质量评估后，排除了存在严重运动伪影、脑部覆盖不完整或其他影响质量的伪影的病例，最终纳入 1,317 例：1,207 例正常病例和 110 例病理病例，包括胼胝体发育不全（ACC, $n = 23$）、Dandy-Walker畸形（DWM, $n = 22$）以及生发基质和脑室内出血（GMH- IVH, $n = 65$）。在正常病例中，487 例用于训练，114 例用于验证，其余 606 例与全部 110 例病理病例一同用于内部测试。

2) 外部测试数据集：来自2024年胎儿组织标注与分割挑战赛的 FeTA 数据集 [28], [29],被用于外部测试，以评估跨域泛化能力。该数据集包含31例正常胎儿和49例病理性胎儿的三维胎儿脑磁共振成像体数据（孕龄：20–35周）。

B. 实现细节与对比方法

1) 实现细节：临床孕龄标签根据孕早期超声测量的头臀长推导得出，并以天为单位表示。该网络采用具有四个层级的三维 U-Net，各层通道数为 [32, 64, 128, 256]。全局年龄分支从编码器瓶颈层分出，由自适应平均池化、随机失活（0.1）和全连接层组成。两个独立的卷积分支对解码器输出进行处理，分别预测体素级脑龄和九类分割结果。在瓶颈层额外采用了随机失活（0.3）。超参数 $\delta = 2$ 天、$T = 36$、$\lambda_{vage} = 1$和 $\lambda_{seg} = 1000$ 在早期训练期间确定，用于将3个损失缩放至相同数量级。训练采用 Adam优化器和余弦退火调度器（初始学习率：0.001），在一块 NVIDIA A800 GPU 上进行，共训练150个周期，批量大小为8。

2）对比方法：基于脑龄预测的方法利用预测胎龄与临床记录的胎龄之间的绝对年龄差（AAD），包括单切片 AAD [8], 多切片 AAD [9], 全脑 AAD [10], 和基于图像块的方法PANDA [11]。基于重建的方法包括三维自编码器 [15], 3D DAE [16], 3D STFPM[17],和 Iter Mask3D [18]。所有对比方法均遵循其原始训练和推理协议实现，并在完全匹配的设置下进行评估。

V. 结果与分析

A. 病灶定位性能

如图 2 所示，**FetalBUAD** 生成的体素级脑龄差图在正常脑部中接近于零，这表明生物学脑龄参考能够有效抑制正常发育变异，并生成稳定的定位图。对于病理性受试者，该方法能够突出显示不同病灶类型中具有解剖学意义的异常，包括 ACC 中缺失的胼胝体、DWM 中扩大的后颅窝和小脑发育不全，以及 GMH-IVH 中的局灶性出血区域。与基于重建的方法相比，后者通常会产生弥散性响应或散布的假阳性，

表 I 病例级-别异常检测性能：内部测试集与外部测试集。曲线下面积 和精确率-召回率曲线下面积，以及特异度固定为0.80 时的敏感度和敏感度固定为 0.80 时的特异度。符号表示方法类别： † 脑龄预测-类方法； ‡ 重建-类方法。最佳结果为 蓝色，次-佳结果为 绿色。

	方法	曲线下面积	AUPR	敏感度	特异度
内部测试数据集：异常 ($n = 110$) 和 正常 ($n = 606$)	单切片 AAD† [8]	0.656 [0.595, 0.713]	0.340 [0.254, 0.432]	0.436 [0.333, 0.532]	0.340 [0.257, 0.412]
	多切片 AAD† [9]	0.596 [0.528, 0.657]	0.280 [0.208, 0.364]	0.355 [0.272, 0.450]	0.188 [0.102, 0.398]
	全脑 AAD† [10]	0.584 [0.522, 0.643]	0.269 [0.193, 0.350]	0.318 [0.213, 0.420]	0.252 [0.209, 0.342]
	PANDA† [11]	0.726 [0.674, 0.781]	0.441 [0.355, 0.533]	0.564 [0.467, 0.656]	0.465 [0.346, 0.611]
	三维自编码器‡ [15]	0.740 [0.688, 0.789]	0.392 [0.297、0.482]	0.509 [0.409, 0.621]	0.502 [0.367, 0.666]
	三维去噪自编码器‡ [16]	0.630 [0.571, 0.690]	0.285 [0.210, 0.366]	0.400 [0.287, 0.496]	0.304 [0.213, 0.443]
	3D STFPM‡ [17]	0.665 [0.608, 0.717]	0.291 [0.216, 0.375]	0.418 [0.303, 0.510]	0.422 [0.354, 0.492]
	IterMask3D‡ [18]	0.905 [0.872, 0.937]	0.779 [0.715, 0.840]	0.800 [0.714, 0.876]	0.787 [0.731, 0.931]
	FetalBUAD	0.905 [0.870, 0.935]	0.764 [0.694, 0.828]	0.845 [0.750, 0.907]	0.868 [0.715, 0.919]
亚组：ACC（$n = 23$）与正常（$n = 606$）	单切片 AAD† [8]	0.685 [0.559, 0.806]	0.230 [0.074, 0.430]	0.435 [0.210, 0.647]	0.401 [0.075, 0.622]
	多切片 AAD† [9]	0.653 [0.534, 0.762]	0.112 [0.039, 0.298]	0.391 [0.192, 0.609]	0.404 [0.138, 0.608]
	全脑 AAD† [10]	0.591 [0.451, 0.731]	0.119 [0.033, 0.258]	0.304 [0.118, 0.546]	0.251 [0.059, 0.443]
	PANDA† [11]	0.754 [0.639, 0.857]	0.255 [0.095, 0.440]	0.565 [0.350, 0.765]	0.528 [0.285, 0.752]
	三维自编码器‡ [15]	0.751 [0.645, 0.839]	0.101 [0.050, 0.212]	0.478 [0.269, 0.720]	0.502 [0.368, 0.768]
	三维去噪自编码器‡ [16]	0.613 [0.467, 0.757]	0.088 [0.037, 0.218]	0.435 [0.222, 0.650]	0.224 [0.046, 0.600]
	3D STFPM‡ [17]	0.688 [0.589, 0.777]	0.086 [0.039, 0.222]	0.348 [0.160, 0.571]	0.465 [0.268, 0.647]
	IterMask3D‡ [18]	0.920 [0.842, 0.982]	0.763 [0.588, 0.907]	0.826 [0.667, 0.963]	0.949 [0.479, 1.000]
	FetalBUAD	0.949 [0.881, 0.991]	0.753 [0.575, 0.896]	0.913 [0.769, 1.000]	0.939 [0.629, 0.997]
亚组：DWM ($n = 22$) & 正常 ($n = 606$)	单切片 AAD† [8]	0.701 [0.577, 0.823]	0.110 [0.045, 0.286]	0.500 [0.273, 0.714]	0.378 [0.280, 0.679]
	多切片 AAD† [9]	0.576 [0.420, 0.717]	0.161 [0.038, 0.327]	0.364 [0.187, 0.591]	0.188 [0.066, 0.484]
	全脑 AAD† [10]	0.638 [0.507, 0.766]	0.237 [0.081, 0.416]	0.364 [0.174, 0.588]	0.403 [0.181, 0.538]
	PANDA† [11]	0.765 [0.658, 0.867]	0.305 [0.115, 0.500]	0.636 [0.428, 0.833]	0.474 [0.312, 0.851]
	三维自编码器‡ [15]	0.886 [0.813, 0.940]	0.412 [0.205, 0.608]	0.773 [0.600, 0.950]	0.787 [0.650, 0.903]
	三维去噪自编码器‡ [16]	0.643 [0.527, 0.748]	0.096 [0.029, 0.213]	0.318 [0.136, 0.533]	0.449 [0.185, 0.619]
	3D STFPM‡ [17]	0.729 [0.615, 0.828]	0.203 [0.060, 0.375]	0.455 [0.250, 0.667]	0.531 [0.258, 0.638]
	IterMask3D‡ [18]	0.920 [0.867, 0.971]	0.610 [0.408, 0.792]	0.818 [0.625, 0.963]	0.835 [0.658, 0.977]
	FetalBUAD	0.946 [0.898, 0.983]	0.687 [0.487, 0.850]	0.909 [0.769, 1.000]	0.913 [0.740, 0.997]
亚组：生发基质出血-脑室内出血（$n = 65$）和正常（$n = 606$）	单切片 AAD† [8]	0.630 [0.557, 0.704]	0.214 [0.135, 0.316]	0.415 [0.279, 0.529]	0.272 [0.221, 0.395]
	多切片 AAD† [9]	0.583 [0.499, 0.663]	0.172 [0.106, 0.258]	0.338 [0.213, 0.449]	0.114 [0.064, 0.423]
	全脑 AAD† [10]	0.563 [0.491, 0.638]	0.144 [0.091, 0.212]	0.308 [0.169, 0.419]	0.252 [0.210, 0.333]
	PANDA† [11]	0.703 [0.623, 0.774]	0.318 [0.202, 0.431]	0.538 [0.394、0.649]	0.391 [0.231, 0.616]
	三维自编码器‡ [15]	0.686 [0.622, 0.752]	0.211 [0.141, 0.306]	0.431 [0.319, 0.574]	0.378 [0.311, 0.635]
	三维去噪自编码器‡ [16]	0.632 [0.565, 0.705]	0.212 [0.134, 0.305]	0.415 [0.262, 0.533]	0.305 [0.218, 0.453]
	3D STFPM‡ [17]	0.636 [0.570, 0.704]	0.159 [0.112, 0.235]	0.431 [0.279、0.542]	0.380 [0.317, 0.461]
	IterMask3D‡ [18]	0.895 [0.854, 0.936]	0.695 [0.594, 0.793]	0.785 [0.662, 0.897]	0.779 [0.656, 0.914]
	FetalBUAD	0.876 [0.830, 0.921]	0.618 [0.493, 0.716]	0.800 [0.667, 0.885]	0.741 [0.641, 0.894]
数据集：外部测试异常（$n = 49$）& 正常（$n = 31$）	单切片 AAD† [8]	0.512 [0.380, 0.639]	0.680 [0.540, 0.808]	0.286 [0.157, 0.449]	0.226 [0.000, 0.444]
	多切片 AAD† [9]	0.609 [0.477, 0.723]	0.704 [0.558, 0.843]	0.327 [0.130, 0.660]	0.226 [0.083, 0.539]
	全脑 AAD† [10]	0.474 [0.348, 0.616]	0.588 [0.465, 0.762]	0.224 [0.000, 0.531]	0.097 [0.000, 0.281]
	PANDA† [11]	0.461 [0.340、0.601]	0.572 [0.443, 0.732]	0.224 [0.000, 0.422]	0.129 [0.000, 0.353]
	三维自编码器‡ [15]	0.708 [0.582, 0.826]	0.749 [0.600, 0.893]	0.531 [0.127, 0.769]	0.516 [0.189, 0.778]
	三维去噪自编码器‡ [16]	0.554 [0.433, 0.676]	0.702 [0.568, 0.825]	0.347 [0.159, 0.500]	0.161 [0.032, 0.434]
	3D STFPM‡ [17]	0.766 [0.653, 0.873]	0.833 [0.720, 0.922]	0.571 [0.326, 0.813]	0.548 [0.364, 0.800]
	IterMask3D‡ [18]	0.711 [0.597, 0.818]	0.815 [0.698, 0.906]	0.551 [0.280, 0.741]	0.387 [0.192, 0.714]
	FetalBUAD	0.800 [0.697, 0.885]	0.874 [0.786, 0.937]	0.673 [0.470, 0.824]	0.613 [0.371, 0.844]

在正常组织中，**FetalBUAD** 生成的异常图更清晰，且在解剖学上的范围更集中。其响应与临床医生标注的病灶位置高度吻合，这表明体素级脑龄估计捕捉到的是局部发育偏差，而非重建残差。这些结果表明，该方法在不同类型的胎儿脑异常中具有更高的空间特异性，并能更精确地描绘病灶边界。

B. 异常检测性能

1）内部测试：如表 I 所示，**FetalBUAD**的曲线下面积为 0.905，与表现最佳的基于重建的方法 Iter Mask3D 持平，同时优于所有脑龄预测基线和其余基于重建的方法。除曲线下面积外，**FetalBUAD**在预设运行阈值下的敏感度（0.845）和特异度（0.868）也均为最高

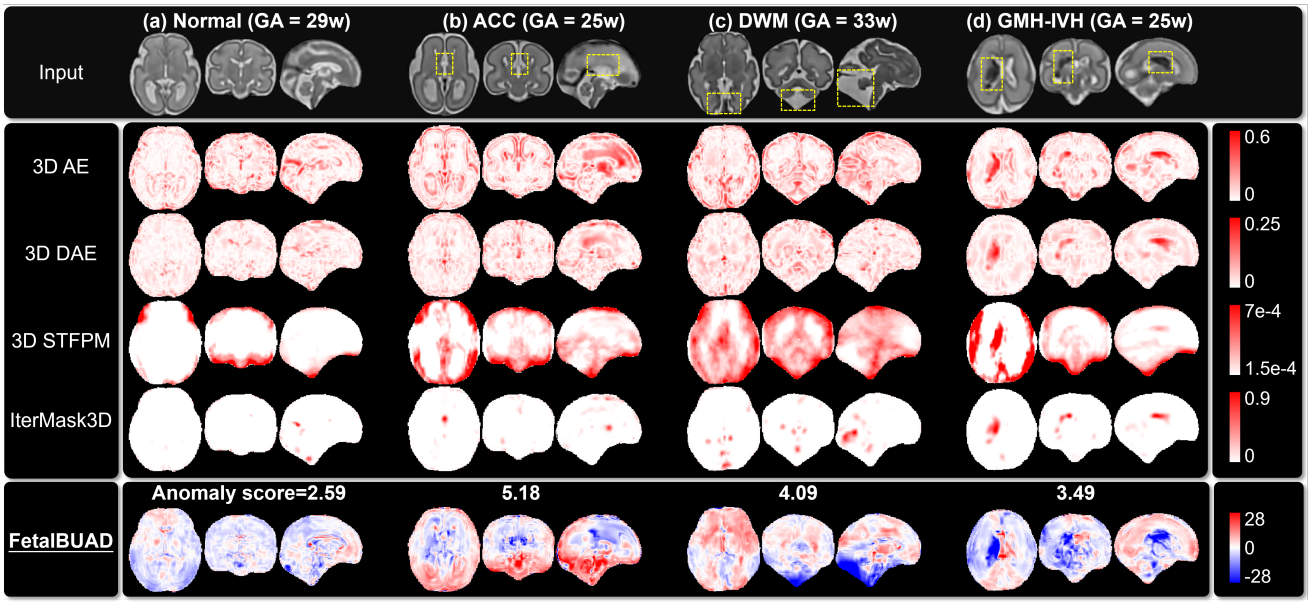


图 2. 内部测试集中具有代表性的正常和病理病例的病灶定位。黄色虚线框表示真实病灶。病理发现包括 ACC 中的胼胝体发育不全、DWM 中的后颅窝囊性扩张和小脑发育不全，以及 GMH-IVH 中的低信号出血区域。各行分别显示基于重建的基线方法（三维自编码器、三维去噪自编码器、3D STFPM 和 IterMask3D）以及所提出的 FetalBUAD 生成的异常图。

表 I内部测试集上的消融研究。曲线下面积和精确率-召回率曲线下面积同时报告了特异度固定为0.80时的敏感度，以及敏感度固定为0.80。最佳结果以蓝色标示，次佳-结果以绿色标示。

	方法	曲线下面积	AUPR	敏感度	特异度
内部测试数据集：异常（n = 110）和正常（n = 606）	未配准至38周	0.870 [0.827，0.911]	0.718 [0.638，0.794]	0.809 [0.720，0.877]	0.802 [0.662，0.878]
	不含分割任务	0.829 [0.786，0.870]	0.616 [0.529，0.697]	0.655 [0.561，0.762]	0.652 [0.503，0.779]
	不使用不确定性加权	0.887 [0.849, 0.920]	0.737 [0.665，0.804]	0.800 [0.722，0.868]	0.759 [0.625, 0.887]
	指标：均值	0.881 [0.843, 0.914]	0.711 [0.636，0.778]	0.755 [0.675，0.841]	0.748 [0.643, 0.862]
	FetalBUAD	0.905 [0.870，0.935]	0.764 [0.694, 0.828]	0.845 [0.750，0.907]	0.868 [0.715、0.919]

，表明其在病理病例检测与假阳性控制之间实现了良好的平衡。

疾病特异性评估表明，**FetalBUAD**在 ACC（曲线下面积 0.949）和 DWM（曲线下面积 0.946）上取得最佳表现，优于 IterMask3D（两者均为 0.920）。这些疾病主要涉及局部结构畸形，而非显著的强度变化，这凸显了体素级脑龄估计在检测区域性发育异常方面的优势。相反，IterMask3D 在 GMH-IVH 上的表现略好（0.895 对比 0.876），这可能是因为出血性病灶具有明显的强度变化，基于重建的方法能够轻松捕捉这些变化。

2）外部测试：如表 I 所示，**FetalBUAD**在外部 FeTA 数据集上无需针对目标域进行微调，即可在所有指标上取得最佳性能，包括曲线下面积（0.800）、精确率-召回率曲线下面积（0.874）、敏感度（0.673）和特异度（0.613）。这些结果表明，所提出的生物学脑龄表征具有更强的泛化能力，能够更有效地适应不同的采集方案和机构，相较于基于重建的异常建模方法表现更佳。该方法通过评估发育一致性而非重建保真度，可能对域偏移不那么敏感，从而提升跨数据集鲁棒性。

C. 消融研究

我们在内部测试集上进行了消融实验，以验证关键组件的贡献（表 II）。

*1)*移除与 ***38*** 周图谱的仿射配准：曲线下面积、精确率-召回率曲线下面积、敏感度和特异度分别下降了 3.9%、6.0 %、4.3% 和 7.6%。这表明，使用原始尺寸的脑体积会使模型在胎龄预测时过度依赖脑大小，而非内部结构关系，从而降低异常检测性能。

*2)*训练中排除辅助脑分割任务：这导致性能下降最为明显，其中 AUROC、精确率-召回率曲线下面积、敏感度和特异度分别下降了 8.4%、19.4%、22.5% 和 24.9%。这凸显了分割在提供解剖约束和促进空间一致的形态表征方面的作用。

3) 推理过程中弃用不确定性加权：AUROC、精确率-召回率曲线下面积、敏感度和特异度分别下降了 2.0%、3.5%、5.3% 和 12.6%。这些结果证实，不确定性加权能够突出体素级脑龄差图中的病理区域，从而提高检测性能。

*4)*在异常评分中用平均偏差替代标准差（***std***）：曲线下面积、精确率-召回率曲线下面积、敏感度和特异度分别下降了 2.7%、6.9%、10.7% 和 13.8%。这表明，标准差能够捕捉偏差的空间异质性，相较于仅考虑平均偏移，它是更敏感的异常指标。

VI. 结论

我们提出了 **FetalBUAD**，一种基于体素级生物学脑龄估计、具备不确定性感知能力的无监督胎儿脑异常检测框架。通过建立内部生物学年龄参考，而非依赖噪声较大的临床孕龄标签，**FetalBUAD** 能够实现病例级异常检测和具有解剖学可解释性的病灶定位。本研究凸显了体素级生物学年龄建模作为基于图像一致性的异常检测替代方案的潜力。未来工作将把该框架拓展至更广泛的胎儿脑异常及纵向MRI分析。

VII. 致谢

本研究得到中国国家自然科学基金（项目编号 82572198、82672619）、四川省医学会科研项目（项目编号 2024HR130）、中国四川省科学技术厅（项目编号 25SYSX0255）、清华大学启动基金以及清华大学笃实计划（项目编号 20241080026、20251080056）的资助。

参考文献

[1] S. Meredith Weiss、E. Aydin、S. Lloyd-Fox 和 M. H. Johnson，“胎儿期、出生时及婴儿期的脑部与行为发育轨迹”，*NatureHuman Behaviour*，第 8 卷，第 7 期，第 1251–1262 页，2024 年。[2] S. K. Urhoj 、 J. Morris 、 M. Loane 、 E. Ballardini、L. Barrachina-Bonet、C. Cavero-Carbonell、A. Coi、M. Gissler、J. Given、A. Heino、S. Jordan、A. Neville、M. Santoro、J. Tan、D. Tucker、D. Wellesley、E. Garne 和 M. Damkjaer，“先天性异常儿童发生脑性瘫痪、癫痫发作/癫痫、视力和听力障碍、癌症、受伤及虐待的风险较高：来自 eurolinkcat研究的数据”，*Acta Paediatrica*，第 113 卷，第 5 期，第 1024–1031 页，2024 年。[3] J. B. Russ、S. Agarwal、C. Venkatesan、B. Scelsa、B. Vollmer、T. Tarui、A. C. Pardo、M. E. Lemmon、S. B. Mulkey、A. R. Hart、U. D. Nagaraj、J. A. Kuller、M. T. Whitehead、J. L. Cohen、J. S. Gebb、O. A. Glenn、M. E. Norton 和 D. Gano，“胎儿皮层发育畸形：综述与临床指导”，*Brain*，第 148 卷，第 6 期，第 1888–1903 页，2025 年 3 月。[4] A. Ortug、Y. Guo、H. A. Feldman、Y. Ou、J. L. A. Warren、H. Dieuveuil、N. T. Baumer、S. K. Faja 和 E. Takahashi，“利用磁共振成像可在宫内观察到自闭症相关脑部差异”，*CerebralCortex*，第 34 卷，第 4 期，文章编号 bhae117，2024 年 4 月。[5] P. D. Griffiths、M. Bradburn、M. J. Campbell、C. L. Cooper、R. Graham、D. Jarvis、M. D. Kilby、G. Mason、C. Mooney、S. C. Robson 和 A. Wailoo，“在宫内诊断胎儿脑异常时使用磁共振成像（MERIDIAN）：一项多中心前瞻性队列研究”，*The Lancet*，第 389 卷，第 10068 期，第 538–546 页，2017 年。

[6] M. Liu、Y. Li、J. Zhu、H. Yang、Y. Huang、H. Li、Y. Chen、X. Bai、Y. Liao、H. Qu 和 Q. Tian，“Fetalextract-llm：基于隐私保护型开放权重大语言模型从胎儿 MRI 自由文本报告中进行结构化信息抽取”，载于围产期、早产儿与儿科图像分析。Cham：Springer Nature Switzerland，2026 年，第 119–129 页。[7] A. A. Chowdhury、S. Hasan Mahmud、K. K. Shahjalal Hoque、K. Ahmed、F. M. Bui、P. Lio、M. A. Moni 和 F. A. Al-Zahrani，“Stackf-bas：利用卷积神经网络与堆叠策略从 MRI 图像中检测胎儿脑异常”，《沙特国王大学学报：计算机与信息科学》，第 35 卷，第 8 期，2023 年 9 月。[8] W. Shi、G. Yan、Y. Li、H. Li、T. Liu、C. Sun、G. Wang、Y. Zhang、Y. Zou 和 D. Wu，“利用基于注意力的深度集成与不确定性进行胎儿脑龄估计和异常检测”，*NeuroImage*，第 223 卷，文章 117316，2020 年。[9] H. J. Yun、H.-J. Lee、S. You、J. Y. Lee、J. Aguirre-Chavez、L. Vasung、H. J. Lee、T. Tarui、H. A. Feldman、P. E. Grant 和 K. Im，“基于深度学习的脑龄预测：利用 MRI 识别患有脑室扩张的胎儿”，《放射学：人工智能》，第 7 卷，第 2 期，文章 e240115，2025 年。[10] H. Peng、W. Gong、C. F. Beckmann、A. Vedaldi 和 S. M. Smith，“利用轻量级深度神经网络进行准确的脑龄预测”，《医学图像分析》，第 68 卷，文章 101871，2021 年。[11] Y. Hao、M. Liu、J. Zhu、H. Yang、H. Li、M. Kang、Y. Song、H. Lai、X. Zhou、G. Ning、Y. Liao、H. Qu 和 Q. Tian，“PANDA：通过胎儿 MRI 中的年龄预测进行基于图像块的无监督深度学习脑异常检测”，《神经影像学》，2026 年 6 月。 [在线]。可访问：https://doi.org/ 10 1162 1293 。/IMAG.a. A. T.

[12] Papageorghiou、S. H. Kennedy、L. J. Salomon、E. O. Ohuma、L. Cheikh Ismail、F. C. Barros、A. Lambert、M. Carvalho、Y. A. Jaffer、E. Bertino、M. G. Gravett、D. G. Altman、M. Purwar、J. A. Noble、R. Pang、C. G. Victora、Z. A. Bhutta、J. Villar，以及国际21世纪胎儿与新生儿生长联盟（INTERGROWTH-21st），“基于孕早期超声检查头臀长测量的早期胎儿大小和妊娠孕周判定国际标准”，《妇产科超声》，第 44 卷，第 6 期，第 641–648 页，2014 年。[13] G. E. Chalouhi、J. P. Bernard、G. Benoist、B. Nasr、Y. Ville 和 L. J. Salomon，“用于预测分娩日期的孕早期测量比较”，《母胎与新生儿医学杂志》，第 24 卷，第 1 期，第 51–57 页，2011 年。[14] M. Liu、Y. Jiao、J. Lu 和 H. Chen，“利用带跳跃连接和多尺度异常一致性的师生模型进行医学图像异常检测”，《*IEEE*仪器与测量汇刊》，第 73 卷，第 1–15 页，2024 年。[15] G. Luo、W. Xie、R. Gao、T. Zheng、L. Chen 和 H. Sun，“脑部 MRI 中的无监督异常检测：从海量健康脑部中学习抽象分布”，《生物学与医学中的计算机》，第 154 卷，文章 106610，2023 年。[16] A. Kascenas、N. Pugeault 和 A. Q. O’Neil，“用于脑部 MRI 无监督异常检测的去噪自编码器”，载于《第*5*届医学成像深度学习国际会议论文集》，丛书《机器学习研究论文集》，E. Konukoglu、B. Menze、A. Venkataraman、C. Baumgartner、Q. Dou 和 S. Albarqouni 编，第 172 卷。PMLR，2022 年 7 月 6–8 日，第 653–664 页。[17] J. Schwarz、L. Will、J. Wellmer 和 A. Mosig，“三维磁共振成像中基于图像块的师生金字塔匹配异常检测方法”，载于《第*7*届医学成像深度学习国际会议论文集》，丛书《机器学习研究论文集》，N. Burgos、C. Petitjean、M. Vakalopoulou、S. Christodoulidis、P. Coupe、H. Delingette、C. Lartizien 和 D. Mateus 编，第 250 卷。PMLR，2024 年 7 月 3–5 日，第 1357–1370 页。[18] Z. Liang、X. Guo、W. Xu、Y. Ibrahim、N. Voets、P. M. Pretorius、J. A. Noble 和 K. Kamnitsas，“IterMask3D：利用测试时迭代掩码细化进行三维脑部 MRI 无监督异常检测与分割”，《医学图像分析》，第 107 卷，文章 103763，2026 年。[19] S. G. Popescu、B. Glocker、D. J. Sharp 和 J. H. Cole，“局部脑龄：一个 U-Net 模型”，《衰老神经科学前沿》，第 13 卷，2021 年。[20] N. Gianchandani、M. Dibaji、J. Ospel、F. Vega、M. Bento、M. E. MacDonald 和 R. Souza，“脑龄预测的体素级方法：评估脑区衰老的一种方法”，《生物医学成像机器学习》，第 2 卷，第 761–795 页，2024 年。[21] Y. Gal 和 Z. Ghahramani，“随机失活作为贝叶斯近似：在深度学习中表示模型不确定性”，载于《第

第*33*届国际机器学习会议，丛书《机器学习研究论文集》，M. F. Balcan 和 K. Q. Weinberger 编，第 48 卷。美国纽约州纽约：PMLR，2016 年 6 月 20–22 日，第 1050–1059 页。[22] L. Fidon、M. Aertsen、F. Kofler、A. Bink、A. L. David、T. Deprest、D. Emam、F. Guffens、A. Jakab、G. Kasprian、P. Kienast、A. Melbourne、B. Menze、N. Mufti、I. Pogledic、D. Prayer、M. Stuempflen、E. Van El-slander、S. Ourselin、J. Deprest 和 T. Vercauteren，“基于 Dempster-Shafer 方法的可信人工智能及其在胎儿脑磁共振成像分割中的应用”，*IEEE*模式分析与机器智能汇刊，第 46 卷，第 5 期，第 3784–3795 页，2024 年。[23] A. Gholipour、C. K. Rollins、C. Velasco-Annis、A. Ouaalam、A. Akhondi-Asl、O. Afacan、C. M. Ortinau、S. Clancy、C. Limperopoulos、E. Yang 等，“用于胎儿脑部自动分割、分析早期脑部生长的规范性时空磁共振成像图谱”，《科学报告》，第 7 卷，第 1 期，文章 476，2017 年。[24] N. J. Tustison、P. A. Cook、A. J. Holbrook、H. J. Johnson、J. Muschelli、G. A. Devenyi、J. T. Duda、S. R. Das、N. C. Cullen、D. L. Gillen 等，“用于定量生物医学成像的 ANTsX生态系统”，《科学报告》，第 11 卷，第 1 期，文章 9068，2021 年。[25] S. M. Smith、D. Vidaurre、F. Alfaro-Almagro、T. E. Nichols 和 K. L. Miller，“从脑成像中估计脑龄差值”，*NeuroImage*，第 200 卷，第 528–539 页，2019 年。[26] M. Ebner、G. Wang、W. Li、M. Aertsen、P. A. Patel、R. Aughwane、A. Melbourne、T. Doel、S. Dymarkowski、P. De Coppi、A. L. David、J. Deprest、S. Ourselin 和 T. Vercauteren，“胎儿脑磁共振成像定位、分割与超分辨率重建的自动化框架”，*NeuroImage*，第 206 卷，文章 116324，2020 年。[27] M. Liu、Y. Liao、H. Li、J. Zhu、H. Yang、Y. Hao、H. Qu 和 Q. Tian，“基于图像方向识别不确定性的无质量标签胎儿脑磁共振成像质量控制”，《医学图像分析》，文章 103994，2026 年。[28] K. Payette、P. de Dumast、H. Kebiri、I. Ezhov、J. C. Paetzold、S. Shit、A. Iqbal、R. Khan、R. Kottke、P. Grehten 等，“使用胎儿组织标注数据集的自动多组织人类胎儿脑部分割基准”，《科学数据》，第 8 卷，第 1 期，文章 167，2021 年。[29] K. Payette、H. B. Li、P. de Dumast、R. Licandro、H. Ji、M. M. R. Siddiquee、D. Xu、A. Myronenko、H. Liu、Y. Pei、L. Wang、Y. Peng、J. Xie、H. Zhang、G. Dong、H. Fu、G. Wang、Z. Rieu、D. Kim、H. G. Kim、D. Karimi、A. Gholipour、H. R. Torres、B. Oliveira、J. L. Vilac̣a、Y. Lin、N. Avisdris、O. Ben-Zvi、D. B. Bashat、L. Fidon、M. Aertsen、T. Vercauteren、D. Sobotka、G. Langs、G. Aleny`a、M. I. Villanueva、O. Camara、B. S. Fadida、L. Joskowicz、L. Weibin、L. Yi、L. Xuesong、M. Mazher、A. Qayyum、D. Puig、H. Kebiri、Z. Zhang、X. Xu、D. Wu、K. Liao、Y. Wu、J. Chen、Y. Xu、L. Zhao、L. Vasung、B. Menze、M. B. Cuadra 和 A. Jakab，“胎儿脑组织标注与分割挑战赛结果”，《医学图像分析》，第 88 卷，文章 102833，2023 年。