FAMSeg：使用特征感知注意和曼巴增强技术进行胎儿股骨和颅骨超声分割

Jie He[1], Minglang Chen[2,1,*], Minying Lu[3], Bocheng Liang[4], Junming Wei[6], Guiyan Peng[4], Jiaxi Chen[5], and Ying Tan[4]

[1] Guangxi Key Laboratory of Machine Vision and Intelligent Control, Wuzhou University, Wuzhou, 543002 China.
[2] Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, 999078 China.
2240030650@student.must.edu.mo.
[3] School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, 541004 China.
[4] Shenzhen Maternity and Child Healthcare Hospital, Southern Medical University, Shenzhen, 518100 China.
[5] College of Big Data and Software Engineering, Wuzhou University, Wuzhou, 543002 China.
[6] College of Electronical and Information Engineering Wuzhou University, Wuzhou,543002 China.

Abstract. 准确的超声图像分割是精确生物测量和准确评估的前提。依赖人工描绘不仅引入了显著的误差，而且费时。然而，现有的分割模型是基于自然场景中的物体设计的，因此难以适应具有高噪声和高相似性的超声物体。在小物体分割中特别明显，常出现明显的锯齿效应。因此，本文提出了一种基于特征感知和 Mamba 增强的胎儿股骨和颅骨超声图像分割模型，以应对这些挑战。具体来说，设计了一种纵横独立视角扫描卷积块和特征感知模块，以增强捕捉局部详细信息的能力并改善上下文信息的融合。结合 Mamba 优化的残差结构，该设计抑制了原始噪声的干扰并增强了局部多维扫描。系统构建了全局信息和局部特征依赖关系，并通过不同优化器的组合进行训练以达到最优解。经过广泛的实验验证，FAMSeg 网络在各种大小和方向的图像中实现了最快的损失减少和最佳的分割性能。

1 介绍

准确描绘关键解剖结构提供了详细的空间定位和结构生物测定的适用测量方法。胎儿超声图像的精确生物测量对于准确评估胎儿健康和发育至关重要 [1] 。然而，在临床实践中，关键解剖结构的描绘常常依赖于超声技师的手动勾画。这不仅增加了工作负担，而且 [2] ，由于依赖视觉检查和生理限制，往往导致边界描绘不准确或不规则 [3] 。因此，通过结合深度学习方法，可以在像素级准确描绘关键解剖结构，有效协助超声技师。这减少了他们的工作量，提高了边界描绘的准确性，减少了测量误差，从而提升了预测和评估的准确性。

胎儿超声成像受到多种因素的影响 [4] ，包括羊水、采集参数 [5] 和扫描技术 [6] ，导致一些问题，如解剖结构小、噪声干扰显著以及对比度低。现有的深度学习方法在捕捉胎儿股骨和颅区等结构的细微变化上存在困难。然而，大多数旨在实现高效准确分割的方法采用了微调编码器、引入注意力机制以及加深解码器等技术，以获取更丰富的特征图来补充编码器的纹理提取。随着特征图

维度的增加，感受野变得更大，这在改善全局上下文的同时，降低了准确分割小目标和定义边界的能力 [7] ，从而导致分割结果中出现更明显的锯齿边缘。

因此，本研究提出了一种基于 Mamba 机制和自适应特征感知的端到端分割框架。通过设计各种卷积方法更准确地捕捉轮廓信息，并整合自适应特征感知增强局部依赖性，该方法解决了胎儿股骨和颅骨超声分割中的关键挑战。我们的方法提供了以下贡献。

- Design a multi-branch deep strip convolution module to independently scan feature maps from both horizontal and vertical perspectives, enhancing the description of target boundary features.
- Design a Mamba residual model to reduce the propagation of original features in the network through Mamba' s multi-view scanning approach, minimizing interference with segmentation accuracy.
- Design an adaptive perception module to capture the weight information of different features in the feature map, applying weighted adjustment to the original features to enhance the model's nonlinear mapping capability.
- Design a multi-optimizer alternating scheme to address issues of the network missing the optimal solution and failing to reach convergence.

随着具有强大计算能力的超级计算机的发展，具有大规模参数和复杂计算结构的深度卷积神经网络（CNNs）在图像处理和医学图像分析与诊断中取得了显著成功。2024 年，Chen 等人提出了一种多视角局部透视的特征编码模块，设计了多通道捕获来建模不同节点之间的全局依赖性，并构建了不同层次特征之间的语义关系，以增强全局和局部特征的提取和融合，从而提高模型的分割性能。2022 年，Lu 等人引入了一种新颖的多级非最大抑制（NMS）机制，以进一步增强在三个选择级别上的分割性能，并在胎儿四腔超声图像中进行 13 个解剖结构的实例分割。2022 年，Pu 等人提出构建显式 FPN 网络以增强多尺度语义信息融合，提高 MobileNet 在胎儿超声心动图的顶端四腔分割中的性能。

由于开源医学图像数据的限制，大多数语义分割模型是为工业或自然对象设计的，通常难以处理医学图像分析。因此，许多研究人员通过采用半监督网络或引入注意力机制来增强特征提取。在 2024 年，Chen 等人 [11] 提出了一个适配模块，用于在 SAM2 大模型下微调下游任务，改善复杂的分割任务，如伪装。在 2024 年，Xing 等人 [12] 集成了 Mamba 机制以捕捉在不同尺度的体积特征中的长程依赖性，增强了网络的分割性能。

本文通过设计不同的视角扫描方法，降低模型复杂度，提升轮廓和小目标信息提取。通过集成 Mamba V2 [13] 机制进行多视角扫描，它能够捕捉局部和全局特征依赖关系。采用自适应感知模块和原创特征融合方法来提高解码器的非线性映射能力。最后，使用不同优化器的交替优化来限制模型在最优解附近的波动，从而提高分割性能。

2 方法

2.1 FAMSeg 概览

图 1 ，展示了所提议的胎儿股骨和颅骨超声图像分割模型的框架。浅层特征具有较小的感受野，并在捕获边缘和纹理信息方面表现出强大的能力。然而，为

了平衡大目标和小目标的分割，许多模型通常构建一个简单的低级特征提取模块和一个复杂的高级特征提取模块。高低层特征的融合增强了对小目标的分割性能。这种方法对于自然场景中的物体分割表现良好，因为它补充了高级特征，同时忽略了低级特征提取。然而，它很难适应医学超声图像中的物体分割，这些图像通常表现出高度相似、模糊和低对比度的特征。为了应对 SegNeXt [14] 模型在小目标分割上的局限性，我们引入了 Mamba 机制并设计了一个上下文特征融合模块。这提高了浅层网络中局部特征信息的交流和多维数据的整合，从而改善了医学图像中小结构的分割精度。

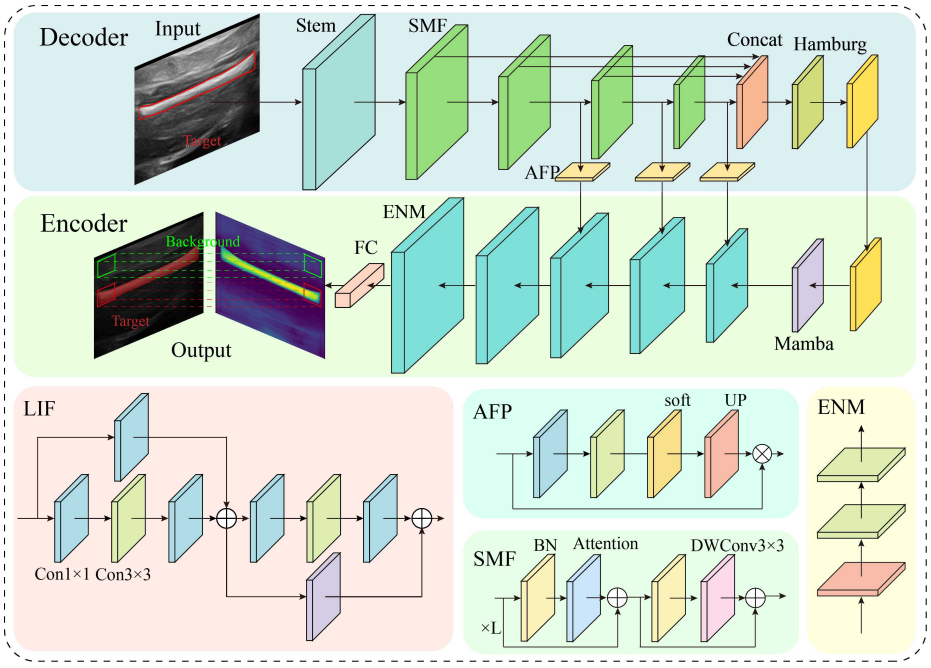


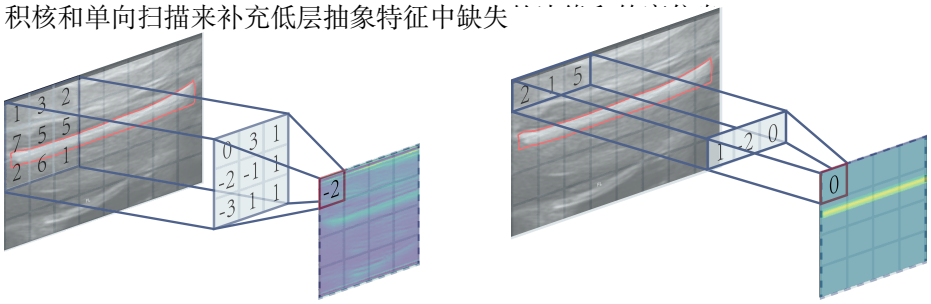
Fig. 1: FAMSeg 模型细节的示意图。

2.2 特征编码器设计

大多数编码器主要由以单向连接方式组合的卷积模块或残差结构组成。这导致局部特征流动和提取能力弱，难以深入挖掘和表达关键物体边缘轮廓信息。因此，解码器在特征恢复过程中缺乏边缘参考，导致明显的伪影效应和小物体特征的消失。为了解决伪影和小物体特征消失的问题，本工作中的编码器主要由多分支深条纹卷积模块和 Mamba 残差结构组成，有效地整合了物体的纹理信息。

2.2.1 多分支深度条带卷积 多分支深条带卷积方法以一种更高效的计算方式从水平和垂直的角度对低层特征进行重点扫描。这减少了计算周期并缓解了同时从两个方向扫描特征时的特征退化现象。如图 2 (a) 所示，标准卷积通过卷积核 K 形成的 $K \times K$ 滑动窗口水平和垂直地扫描特征。窗口以特定的步长 S 在输入特征图上滑动。对于每个覆盖的特征区域，将卷积核的元素与输入特征图的相应元素相乘，并将结果相加以完成特征提取。标准卷积通过同时进行水平

和垂直扫描来捕获抽象和丰富的特征信息。然而，双视图方法往往容易忽视较小的边缘细节，从而导致小物体信息的丢失和特征恢复期间的混叠问题。因此，如图 2 (b) 所示，我们设计了一个多分支深条带卷积模块，该模块使用多个卷积核和单向扫描来补充低层抽象特征中缺失



(a) 卷积块　　(b) 步幅卷积模块

Fig. 2: 不同卷积操作的原理。从水平和垂直的独立视角扫描特征在保持物体边缘信息方面更为有效。

首先，使用 5×5 卷积提取抽象特征。然后，多个卷积核从水平和垂直视角分别扫描特征。最后，通过特征堆叠，将单向信息集成到复杂的多视角信息中，从而缓解小物体和轮廓特征丢失的问题。此外，为了增强编码器单向特征的感受野，我们设计了一个聚合深度卷积模块，以融合来自不同编码器层的特征信息。进一步使用来自 SegNeXt 网络的汉堡模块实现特征增强，从而提高模型的分割性能和鲁棒性。此外，多分支深条带卷积模块与标准卷积相比提供了更高的计算效率，如方程（1）和方程（2）所示，

$$Parameters_{conv} = K \times H_o \times W_o \times K, \tag{1}$$

$$Parameters_{our} = 2 \times H_o \times W_o \times K. \tag{2}$$

设 H_i 和 W_i 表示输入特征的维度，H_o 和 W_o 表示输出特征的维度。步幅记为 S 。当核大小为 $K = 7$ 时，水平方向和垂直方向的计算成本为

$$Cost_{7\times7} = 2 \times H_o \times W_o \times 7, \tag{3}$$

而标准卷积的成本为

$$Cost_{7\times7(\text{standard})} = 7 \times H_o \times W_o \times 7. \tag{4}$$

如果改用 3×3 卷积，并得到相同的输出，计算成本将变为

$$Cost_{3\times3} = 2 \times H_o \times W_o \times 9. \tag{5}$$

因此，显然独立扫描会导致较低的计算成本。

为了实现更稳定的模型传播并获得精确的分割掩码，我们引入了两个残差结构：卷积块和瓶颈块。卷积块对输入特征进行下采样，减少计算数据，为瓶颈块提供更丰富的感受野。这使得对抽象特征的深入挖掘和学习成为可能，增强了局部和全局特征的依赖性。它改进了多分支深条带卷积模块的感受野，解决了单视图扫描中特征多样性有限的问题，并优化了模型在轮廓分割方面的表现。

这两个模块的主要分支都由两个 1×1 卷积和一个 3×3 卷积组成。它们的关键区别在于侧支结构：卷积模块的侧支包含一个 1×1 卷积用于特征融合后降采样，而瓶颈模块的侧支则直接捷径连接到输出特征用于堆叠和特征提取。这些残差结构通过信息分支有助于缓解梯度消失问题。

然而，快捷连接可能允许未经过滤的噪声进入网络，特别是在多个 Bottleneck 模块堆叠时。这导致输出特征中原始噪声的累积，减少后续层的过滤能力，并影响网络的稳定性和分割精度。为了解决这个问题，我们通过引入 Mamba V2 模块优化了 Bottleneck 的快捷分支，该模块通过其多角度扫描方法滤除噪声，并增强编码器中低级特征和轮廓的提取。

2.3 特征感知解码器

低级抽象特征为编码器提供了捕捉全局图像信息的特征图。然而，这些抽象特征图失去了局部信息和空间细节，导致无法直接从其生成分割掩膜。依赖单一特征上采样逐步恢复分辨率常常导致目标细节和空间信息的丢失，尤其是在超声图像分割任务中，其中关键的解剖结构不清晰且图像质量较差。然而，来自编码器的原始特征包含了高分辨率信息，这对精确定位和边缘分割至关重要。因此，解码器结合特征融合和上采样方法逐步改进抽象特征信息。通过集成一种特征感知注意机制，它增强了特征的非线性能力，恢复了原始图像的空间分辨率以生成分割掩膜。

尽管原始特征在精确定位和边缘分割方面表现出色，但它们包含显著的背景噪声。然而，传统的特征上采样方法通常仅在输入特征图上执行简单的空间扩展，往往导致图像模糊或细节丢失。因此，FAM 机制引入了一种动态内容感知方法，使用输入特征图每个位置的空间分布作为权重。这种方法结合并扩展特征图的不同部分，同时带有映射约束，使得上采样不仅仅是简单的空间扩展，而是一个整合输入特征全局信息的过程，从而生成更精细、高质量的特征图。

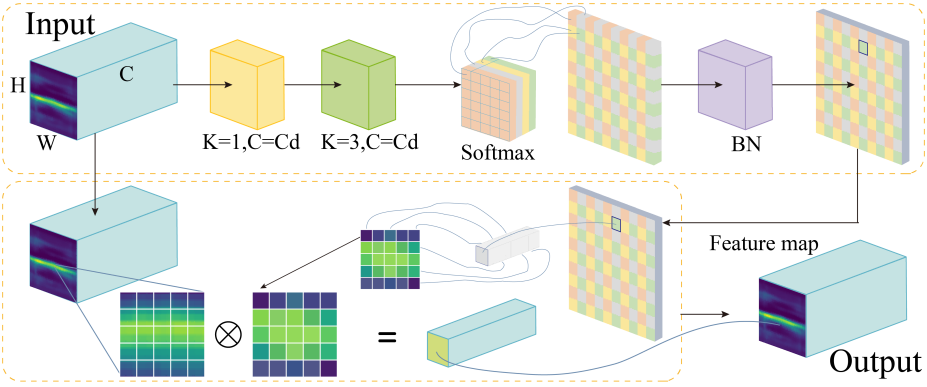


Fig. 3: 特征感知注意力机制原理的示意图。

如图 3 所示，FAM 主要由通道压缩器、内核重组器和内核归一化组成。首先，1×1 通道压缩器将输入特征通道从 C 减少到 Cd，从而通过减少特征图中的通道数来降低模块的计算复杂性。接下来，具有内核大小为 k 的内核重组器根据通道压缩后的特征图内容生成重组内核，形成大小为 $K \times K \times C$ 的特征图。这扩展了感受野，并能够使用来自更大区域的上下文信息进行特征非线性映射。

然后，softmax 函数根据空间信息对特征进行归一化，以评估其重要性。最后，特征重要性与原始特征进行加权，动态感知并重组它们以增强模型的理解能力。

为了减少原始噪声的干扰，我们仅融合编码器最后三层的特征。此外，我们对上采样后的特征进行卷积操作，以加强非线性特征映射并提取局部特征，从而提高模型捕捉更复杂特征的能力并进一步细化特征图。一旦特征图恢复到原始分辨率，使用 1×1 卷积代替全连接方法识别并分类特征图中的特征，从而完成解剖结构分割任务。

当前主流的优化器包括随机梯度下降法 (SGD) [15] 、带权重衰减的 Adam (AdamW) [16] 、以及自适应矩估计 (Adam) [17] 。然而，每种优化器都有其显著优势和不可避免的缺点。因此，为了实现更高效的学习和更高的分割准确性，本文同时使用了 SGD 和 AdamW 优化器，在它们之间交替使用以利用各自的优势。这种方法弥补了每个优化器的局限性，降低了学习成本并提升了分割性能。

SGD 优化器每次只使用一个样本来计算模型梯度，使其易于实现且所需超参数极少。与一些自适应优化器相比，它提供了更好的泛化能力。SGD 以其简单性、计算效率和强泛化能力而闻名。如方程 (6) 所示，SGD 优化器在计算时不包含梯度动量，并通常需要学习率衰减策略以实现最佳性能。因此，SGD 优化器通常较慢，需要更长的拟合周期，但它可以在最优解附近振荡，使其适合于微调模型以达到最优解。

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla_\theta L(\theta_t). \tag{6}$$

然而，Adam 和 AdamW 优化器都结合了梯度动量和方差来优化网络，与 SGD 相比显著提高了优化效率，从而加快了模型的收敛速度并减少了训练时间。尽管 Adam 优化器比 SGD 效率要高得多，但它对初始学习率高度依赖。较大的初始学习率可能导致网络错失最优解，而较小的初始学习率可能会阻碍收敛。然而，AdamW 优化器将权重衰减应用于参数更新，而不是与梯度一同更新，这不仅解决了对初始学习率的依赖，还提供了更稳定和改进的优化性能。

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla_\theta L(\theta_t), v_t = \beta_2 v_{t-1} + (1-\beta_2)\nabla_\theta L(\theta_t)^2, \tag{7}$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t}, \tag{8}$$

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}, \tag{9}$$

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} - \eta \cdot \lambda \cdot \theta_t. \tag{10}$$

总之，我们首先使用 AdamW 优化器进行 150 个周期，快速收敛到接近最优解，避免由于迭代不足而不能达到最优解。然后，我们切换到 SGD 优化器进行 50 个周期的训练，以便进行充分的优化，确保网络达到最优梯度解。

我们收集了来自深圳市妇幼保健院的 3,798 张胎儿股骨和颅骨结构的超声图像数据集，涵盖了不同的发育阶段。胎儿股骨缩写为 FL，颅骨结构缩写为 FB。该数据集是使用来自飞利浦和西门子的设备在怀孕的 14 到 28 周期间收集的。所有实验是在一台安装了 Arch Linux 系统的计算机上进行的，该计算机配备了 Intel Xeon Platinum 8360Y CPU 和四个 NVIDIA A100 GPU，使用 PyTorch 作为模型开发的框架。初始化时的最大和最小学习率分别设置为 0.01 和 0.0001，

学习率的上下界分别为 0.001 和 0.0001。为了动态调整学习率并增强模型的稳定性，我们基于批量大小自适应地调整最大和最小学习率。为了全面评估 FAMSeg 模型的分割和泛化性能，我们设计了与 BisNet V2 和 U-Net 等主流或相对较新的分割模型的比较。为了确保一致的测量，我们使用 IoU（交并比）参数来评估分割结果。实验结果在表 **??** 中展示。FAMSeg 模型不仅优于其他网络，而且在包含大量小目标的股骨数据集上也取得了更优异的结果。各模型的分割结果如图 **??** 所示。

FAMSeg 模型能够有效地对胎儿骨骼和颅骨结构进行超声图像分割，这些结构具有不同的形状、大小和方向，没有出现明显的问题，如锯齿效应、误报、漏报或误分类。虽然 Swin Transformer 的分割结果次优，但在小目标分割中表现不佳，并存在分类错误，如图 **??** 中第二和第三实验组的可视化所示。U-Net 网络从视觉角度可以正确执行分割，但在物体边缘表现出明显的锯齿效应和分割间隙。为进一步验证 FAMSeg 设计方案的可靠性和可行性，我们如表 **??** 所示，通过控制相同变量进行了一项消融研究。在相同条件下，通过去除曼巴结构和特征融合组件，独立训练分割模型，并将结果与基准模型进行比较。实验组 B 和 C 表明，集成上下文融合带来了最显著的整体性能提升。然而，小目标的分割仍然次优。因此，如实验组 A 和 D 所示，通过结合局部和全局信息，我们提高了小目标分割的性能。此外如表 **??** 所示，我们探讨了不同优化器和损失衰减方法对模型训练性能和效率的影响。虽然 AdamW 优化器与余弦损失衰减的组合结果与我们提出的方法相当，但我们的设计在整体分割和小目标分割性能之间提供了更好的平衡。

图 **??** 展示了不同优化器和衰减方法组合的训练损失曲线，说明了模型的收敛速度。本研究中使用的 AdamW 和 Adadelta 组合优化方法实现了最快和最稳定的收敛。本文提出了 FAMSeg 语义分割模型，以解决小对象分割精度不足和分割不平滑的问题。该模型探索了横向和纵向独立视角卷积块、自适应感知机制、多尺度上下文特征融合及不同优化器对分割性能的影响。所提出的横向和纵向独立视角卷积块旨在保留对象轮廓的纹理特征。此外，自适应感知机制增强了对局部特征和全局信息的依赖，加强了模型恢复特征映射的能力。尽管 FAMSeg 在胎儿股骨和脑部分割中取得了显著结果，但仍存在一些局限性。首先，它无法平衡多对象分割的精度。其次，它无法对齐多个模型的信息。最后，模型的训练过程相对耗时。在未来，我们计划通过将大模型与知识蒸馏技术结合，进一步平衡多对象分割性能，并将应用扩展到更广泛的医学领域。该研究得到广西自然科学基金（项目编号 2020JJA170007）、广西自然科学基金（项目编号 2024JJA141093）、国家自然科学基金（项目编号 62162054）、梧州大学重点研究项目（项目编号 2023C004 和 2024QN001）、梧州市科技计划项目（项目编号 202302036）部分支持。

References

1. Pu, B., Li, K., Li, S., et al.: Automatic fetal ultrasound standard plane recognition based on deep learning and iiot. IEEE Transactions on Industrial Informatics 17 (11), 7771–7780 (2021)
2. Gao, Z., Tan, G., Wang, C., et al.: Graph-enhanced ensembles of multi-scale structure perception deep architecture for fetal ultrasound plane recognition. Engineering Applications of Artificial Intelligence 136 , 108,885 (2024)

3. Pu, B., Lv, X., Yang, J., et al.: Unsupervised domain adaptation for anatomical structure detection in ultrasound images. In: Forty-first International Conference on Machine Learning (2024)
4. Zhao, L., Tan, G., Wu, Q., et al.: Farn: fetal anatomy reasoning network for detection with global context semantic and local topology relationship. IEEE Journal of Biomedical and Health Informatics (2024)
5. Zhao, L., Li, K., Pu, B., et al.: An ultrasound standard plane detection model of fetal head based on multi-task learning and hybrid knowledge graph. Future Generation Computer Systems 135 , 234–243 (2022)
6. Pu, B., Wang, L., Yang, J., et al.: M3-uda: a new benchmark for unsupervised domain adaptive fetal cardiac structure detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11,621–11,630 (2024)
7. Yang, J., Lin, Y., Pu, B., Li, X.: Bidirectional recurrence for cardiac motion tracking with gaussian process latent coding. Advances in Neural Information Processing Systems 37 , 34,800–34,823 (2025)
8. Chen, G., Tan, G., Duan, M., et al.: Mlmseg: A multi-view learning model for ultrasound thyroid nodule segmentation. Computers in Biology and Medicine 169 , 107,898 (2024)
9. Lu, Y., Li, K., Pu, B., et al.: A yolox-based deep instance segmentation neural network for cardiac anatomical structures in fetal ultrasound images. IEEE/ACM Transactions on Computational Biology and Bioinformatics 21 (4), 1007–1018 (2024)
10. Pu, B., Lu, Y., Chen, J., et al.: Mobileunet-fpn: A semantic segmentation model for fetal ultrasound four-chamber segmentation in edge computing environments. IEEE Journal of Biomedical and Health Informatics 26 (11), 5540–5550 (2022)
11. Chen, T., Lu, A., Zhu, L., et al.: Sam2-adapter: Evaluating & adapting segment anything 2 in downstream tasks: Camouflage, shadow, medical image segmentation, and more. arXiv preprint arXiv:2408.04579 (2024)
12. Xing, Z., Ye, T., Yang, Y., et al.: Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 578–588. Springer (2024)
13. Dao, T., Gu, A.: Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In: International Conference on Machine Learning (ICML) (2024)
14. Guo, M.H., Lu, C.Z., Hou, Q., et al.: Segnext: Rethinking convolutional attention design for semantic segmentation. Advances in neural information processing systems 35 , 1140–1156 (2022)
15. Ketkar, N.: Stochastic gradient descent. In: Deep learning with Python: A hands-on introduction, pp. 113–132. Springer (2017)
16. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
17. Loshchilov, I., Hutter, F., et al.: Fixing weight decay regularization in adam. arXiv preprint arXiv:1711.05101 5 , 5 (2017)
18. Xiao, T., Liu, Y., Zhou, B., et al.: Unified perceptual parsing for scene understanding. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 418–434 (2018)
19. Chu, X., Tian, Z., Wang, Y., et al.: Twins: Revisiting spatial attention design in vision transformers. arXiv preprint arXiv:2104.13840 (2021)
20. Liu, Z., Lin, Y., Cao, Y., et al.: Swin transformer: Hierarchical vision transformer using shifted windows. arXiv preprint arXiv:2103.14030 (2021)

21. Howard, A., Sandler, M., et al.: Searching for mobilenetv3. In: The IEEE International Conference on Computer Vision (ICCV), pp. 1314–1324 (2019). DOI 10.1109/ICCV.2019.00140
22. Liu, Z., Mao, H., Wu, C.Y., et al.: A convnet for the 2020s. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
23. Yu, C., Gao, C., Wang, J., et al.: Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. International Journal of Computer Vision pp. 1–18 (2021)