超声图像分析基础模型挑战赛基线方法

Bo Deng[1], *Yitong Tang*[1], *Jiake Li*[1] , *Yuxin Huang*[2],*Li Wang*[3,*], *Yu Zhang*[1],*Yufei Zhan*[1],*Hua Lu*[1], *Xiaoshen Zhang* [1], *J*

[1]暨南大学附属第一医院，暨南大学，中国广州[2]南方医科大学珠江医院，中国广州[3]广州市妇女儿童医疗中心，中国广州

*通讯作者：wangli-1227@163.com；jbai996@aucklanduni.ac.nz

arXiv:2602.01055v1 [cs.CV] 1 Feb 2026

摘要

超声（US）成像在不同解剖结构和采集方案之间表现出显著的异质性，为开发具有良好泛化能力的分析模型带来了重大挑战。现有大多数方法均针对特定任务设计，因此难以作为可在临床部署的基础模型。为解决这一局限，超声图像分析基础模型挑战赛（FM UIA 2026）推出了一个大规模多任务基准，涵盖分割、分类、检测和回归四类任务，共包含 27 个子任务。本文介绍了 FM UIA 2026 的官方基线，该基线基于统一的多头多任务学习（MH-MTL）框架，可在单一共享网络中支持所有任务。该模型采用经 ImageNet 预训练的 EfficientNet–B4 主干网络，以实现稳健的特征提取，并结合特征金字塔网络（FPN）捕获多尺度上下文信息。任务特定路由策略使全局任务能够利用高级语义特征，同时使密集预测任务能够充分利用具有空间细节的 FPN 表征。训练过程中采用结合任务自适应学习率缩放的复合损失和余弦退火调度。验证结果证明了这一统一设计的可行性与稳健性，为超声基础模型研究建立了一个强大且可扩展的基线。代码和数据集已在 GitHub 上公开。

索引词— 超声分析，基础模型，多任务学习，EfficientNet，分割，分类，检测，回归

1. 引言

超声（US）成像因散斑噪声和操作者差异而面临独特的计算挑战，限制了传统任务特定“狭义人工智能”解决方案的可扩展性 [1], 这些方案通常仅限于胎头和耻骨联合分割等特定基准[2, 3]。近年来，计算机视觉领域基础模型的成功，例如 Segment Anything Model（SAM）[4], 启发了通用型医疗人工智能的发展[5]。真正的超声基础模型旨在超越器官边界，在单一统一框架内处理分割、分类、检测和回归。

然而，为此类异构任务训练单一网络并非易事。尽管多任务学习（MTL）理论上可以将相关任务作为归纳偏置来提升泛化能力 [6], ，但深度网络在实践中常遭遇“负迁移”问题，即针对某一任务的优化会损害另一任务所需的特征 [7]。如何平衡这些相互冲突的目标，仍是一个有待研究的问题。

为解决这一问题，2026 年超声图像分析基础模型挑战赛（FMC UIA）推出了一项大规模基准，涵盖 27 个不同的子任务。本文介绍官方基线方法：多头多任务学习框架。我们采用 EfficientNet-B4 编码器，并结合任务特定路由机制，以平衡语义抽象与空间分辨率之间的权衡，为竞赛提供稳健的参考方案。

2. 方法

本节将阐述所提出的多头多任务学习（MH-MTL）框架。

2.1. 统一网络架构

该框架记为 Φ，输入为一张超声图像X 和一个任务标识符 t。为兼顾计算效率与多尺度特征提取的需求，所有输入图像均调整为固定分辨率 $H \times W$ =256 × 256。该架构由三个核心组件构成：共享编码器、特征金字塔解码器和任务路由头模块。整体架构如图1所示。

2.1.1. 共享编码器

我们采用以 ImageNet 权重初始化的 EfficientNet-B4 [8] 作为共享主干网络。该编码器提取出一系列特征图，

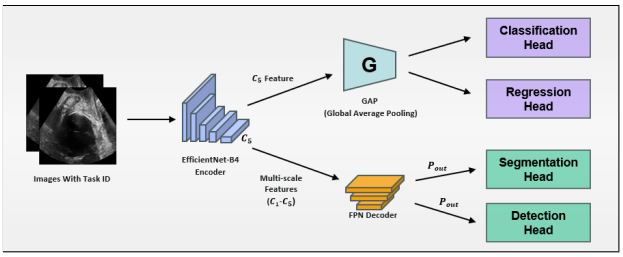


图1。所提出的基线框架概览。模型采用EfficientNet-B4 主干网络提取特征。根据输入的任务标识符，任务特定的路由机制会将高级语义特征（C_5）导向全局任务（分类、回归）的全局平均池化（GAP）模块，或通过特征金字塔网络（FPN）解码器聚合多尺度特征，为密集任务（分割、检测）生成高分辨率特征图 P_{out} 。

其中 $\mathcal{F}_{enc} = \{C_1, C_2, C_3, C_4, C_5\}$ 表示步幅为 2^i 的特征图。最深层 C_5 蕴含着对诊断任务至关重要的抽象语义信息，而较浅层则保留了边界勾画所需的纹理细节。

2.1.2. 特征金字塔解码器

为应对胎儿器官和病灶固有的尺度变化，我们引入了特征金字塔网络（FPN） [9]。FPN 通过自顶向下路径和横向连接聚合特征，生成融合的多尺度表示。对于密集预测任务，我们使用融合后的输出特征图 $P_{out} \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times D}$，其中 $D = 128$ 表示通道维度。

2.2. 任务特定路由与头部

与在推理时推断路径的动态路由机制不同，我们实现了一种显式的任务特定路由策略。根据任务类型 t，信息流被导向全局分支或密集分支。具体而言，这通过确定性控制流（即逻辑分支）实现，其中 t充当开关，仅激活相关的计算路径。

2.2.1. 全局任务（分类与回归）

对于需要图像级推理（例如标准切面分类）或坐标估计（例如胎儿生物测量）的任务，我们使用全局分支。为尽量减少空间冗余，该分支仅对最高层特征 C_5进行运算。

特征经过全局平均池化（GAP）处理，随后通过一个Dropout层（比率=0.2）以减轻过拟合，最后通过全连接投影。预测 Y_{global} 定义如下：

$$\hat{Y}_{global} = W_{head} \cdot \text{Dropout}(\text{GAP}(C_5)) + b_{head} \quad (1)$$

分类：输出维度对应于类别数量 K，逻辑值经Softmax处理以获得概率。

回归：头部输出大小为 $2M$ 的向量（其中 M为关键点数量），表示归一化的（x, y）坐标。

2.2.2. 密集任务（分割与检测）

需要像素级定位的任务利用特征金字塔网络的输出 P_{out}。

分割：我们采用卷积头将P_{out} 投影为 K 类别掩码，随后上采样至原始分辨率（H， W）。

检测：为了将检测统一到这一全卷积框架中，我们采用简化的无锚框网格方法。图像被划分为大小为 $h' \times w'$ 的网格（与 P_{out}的分辨率相匹配）。对于归一化真实中心为（x_{gt}， y_{gt}）的目标，我们将其映射到网格单元（i， j）：

$$i = \lfloor y_{gt} \cdot h' \rfloor, \quad j = \lfloor x_{gt} \cdot w' \rfloor \quad (2)$$

网络在每个网格单元预测一个向量$\hat{v}_{i,\ j} \in \mathbb{R}^5$ ，由边界框坐标和目标性得分组成。

2.3. 目标函数

模型使用由当前批次中激活的任务决定的复合损失函数进行训练。

2.3.1. 分割损失

为处理小病灶与背景之间的类别不平衡，我们采用 Dice损失：

$$\mathcal{L}_{seg} = 1 - \frac{2\sum_p y_p \hat{y}_p + \epsilon}{\sum_p y_p + \sum_p \hat{y}_p + \epsilon} \quad (3)$$

其中 y_p 和 $\hat{y}_p$ 分别是像素 p的真实值和预测概率。

2.3.2. 分类损失

对于诊断分类，我们采用标准的交叉熵损失：

$$\mathcal{L}_{cls} = -\sum_{c=1}^{K} y_c \log(\hat{p}_c) \quad (4)$$

2.3.3. 回归损失

对于生物测量，我们计算预测的归一化坐标与真实关键点之间的均方误差（MSE）：

$$\mathcal{L}_{reg} = \frac{1}{M}\sum_{k=1}^{M} \|\hat{y}_k - y_k\|_2^2 \tag{5}$$

2.3.4. 检测损失

鉴于超声检测目标的稀疏性（每张图像通常只有一个病灶），我们采用聚焦监督策略。我们仅在与目标中心对应的网格单元（$i,\ j$）处应用损失，定义如下：

$$\mathcal{L}_{det} = \mathbb{I}_{i,j}^{\text{obj}} \left[\mathcal{L}_{bce}(\hat{s}_{i,j}, 1) + \lambda \mathcal{L}_1(\hat{b}_{i,j}, b_{gt}) \right] \tag{6}$$

其中，$\mathbb{I}_{i}^{\text{obj}}{}_{,\ j}$ 是一个指示函数，仅在真实中心网格处为 1，其他位置均为 0。$\hat{s}$表示目标置信度ˆ分数，b表示边界框坐标。λ 是一个设为 8 的平衡超参数，旨在优先保证局部定位的精确性。该公式与我们的基线实现一致，优先将模型容量分配给正样本。

3. 实验与结果

3.1. 数据集

FMC UIA 2026 挑战赛推出了一个大规模基准，包含数万张来自多中心临床队列的超声图像[10, 11, 12, 3]。如此规模的数据集有助于开发稳健的通用模型。该数据集在图像质量、解剖视图和采集设备方面呈现显著异质性，涵盖 27 个不同的子任务。

在本基线研究中，模型使用大规模训练集进行训练，并在官方验证集上进行定量评估。该验证集专门由来自未见领域的数据构成，用于严格测试模型在四个主要任务类别中的泛化能力：

- **分割（12 个子任务）：** 涉及对胎儿器官（如头部、心脏、腹部）、母体结构及病灶进行像素级勾画。验证集中包含 2,674 个此类任务的样本。
- **分类（9 个子任务）：** 涵盖胎儿标准切面验证和肿瘤恶性程度评估等诊断任务。验证集包含 2,727 个样本。
- **检测（3 个子任务）：** 重点是定位甲状腺结节、子宫肌瘤和脊髓损伤，共有 725 个验证样本。
- **回归（3 个子任务）：** 目标是生物测量指标，包括宫内生长受限（IUGC）标记物和胎儿股骨长度，由 617 个验证样本表示。

表 1。基线方法在官方 FMC UIA 2026 验证集上的定量结果。评估指标为各类别内所有子任务的平均值。

任务类别	评估指标	数值
分割	平均DSC (↑)	0.7543
	平均HD (↓)	81.18
分类	平均AUC (↑)	0.9155
	平均F1分数 (↑)	0.7896
	平均MCC (↑)	0.6766
检测	平均IoU (↑)	0.2641
回归	平均平均径向误差 (像素) (↓)	67.43

3.2. 评估指标

性能采用挑战平台标准化的任务特定评估指标进行评估：

- **分割：** 我们报告 Dice相似系数（DSC）以评估体积重叠，并报告豪斯多夫距离（HD）以评估边界精度。
- **分类：** 为严格评估类别不平衡情况下的判别能力，我们采用曲线下面积（AUC）、F1分数和 Matthews 相关系数（MCC）进行评估。
- **检测：** 交并比（IoU）用于衡量预测边界框的定位准确性。
- **回归：** 我们以像素为单位报告平均径向误差（MRE）。值得注意的是，MRE 是在原始图像分辨率下计算的，以反映真实临床测量的精度，而非在训练时使用的调整尺寸后的输入上计算。

3.3. 实现细节

该框架使用 PyTorch 和 Segmentation Models PyTorch（SMP）库实现。所有输入图像均调整为 256×256。我们使用 Albumentations 库进行严格的数据增强，包括随机亮度对比度调整和高斯噪声。模型使用 AdamW 优化器和余弦退火调度器训练 50 个周期。主干网络的学习率设为 $1e-4$，而特定任务的头部采用 $1e-3$，以加快收敛。

3.4. 结果与分析

我们在官方验证集上评估了所提出的多头多任务学习基线方法。按任务类别取平均后的定量结果汇总于表 1。

实验结果验证了使用共享编码器处理异质超声任务的可行性。该模型在分类（平均AUC 0.9155）和分割（平均DSC 0.7543）任务上表现稳健。这

表明，采用预训练权重初始化的 EfficientNet 主干网络能够有效捕捉用于区分解剖平面和勾勒主要器官的高层语义表征。对于全局上下文或大型结构特征占主导的任务，该架构已足够胜任。

然而，在定位敏感任务中观察到明显的性能差距，凸显了这一统一基线方法的局限性。检测任务的平均IoU仅为0.2641，表现平平。性能较低的情况主要出现在涉及小病灶或细微异常的子任务中。尽管基于网格的检测头在架构上实现了统一，但与大型解剖结构相比，它难以精确定位小目标。此外，在完整尺寸的超声图像中，背景与小病灶之间极端的类别不平衡进一步阻碍了收敛。

同样，回归任务的平均平均径向误差为67.43像素。需要注意的是，该误差是在原始高分辨率图像上计算的，而模型处理的输入则被调整为 256×256。这种分辨率不匹配不可避免地限制了精细生物测量所需的空间精度。

4. 结论

总而言之，尽管所提出的框架为语义理解奠定了坚实的基线，但结果凸显了未来改进多尺度处理和专用定位模块的必要性，以处理小目标并实现精确测量。

5. 致谢

本研究获得广州市科学技术局项目资助（项目编号：2025B03J0127 和 2023A03J0565）以及广东省科学技术厅项目资助（项目编号：2025B1111030001）。

6. 参考文献

[1] Shengfeng Liu、Yi Wang、Xin Yang、Baiying Lei、Li Liu、Shawn Xiang Li、Dong Ni 和 Tianfu Wang，《医学超声分析中的深度学习：综述》，工程，第 5 卷，第 2 期，第 261–275 页，2019年。[2] Jieyun Bai、Zihao Zhou、Zhanhong Ou、Gregor Koehler、Raphael Stock、Klaus Maier-Hein、Marawan Elbatel、Robert Martí、Xiaomeng Li、Yaoyang Qiu 等，《PSFHS 挑战赛报告：基于产时超声图像的耻骨联合和胎头分割》，医学图像分析，第 99 卷，第 103353 页，2025年。

[3] Gaowen Chen、Jieyun Bai、Zhanhong Ou、Yaosheng Lu 和 Huijin Wang，《PSFHS：用于耻骨联合和胎头基于人工智能的分割的产时超声图像数据集》，*ScientificData*，第 11 卷，第 1 期，第 436 页，2024年。[4]Zihao Zhou、Yaosheng Lu、Jieyun Bai、Victor M Campello、Fan Feng 和 Karim Lekadir，《用于产时超声图像分析中胎头与耻骨联合分割的 Segment Anything 模型》，*ExpertSystems with Applications*，第 263 卷，第 125699 页，2025年。[5] Michael Moor、Oishi Banerjee、Zahra Shakeri Hossein Abad、Harlan M Krumholz、Jure Leskovec、Eric J Topol 和 Pranav Rajpurkar，《通用医学人工智能的基础模型》，*Nature*，第 616 卷，第 7956 期，第 259–265 页，2023 年。[6] Yu Zhang 和 Qiang Yang，《多任务学习概述》，*National Science Review*，第 5 卷，第 1 期，第 30–43 页，2018年。[7] Simon Vandenhende、Stamatios Georgoulis、Wouter Van Gansbeke、Marc Proesmans、Dengxin Dai 和 Luc Van Gool，《密集预测任务的多任务学习：综述》，*IEEE* 模式分析与机器智能汇刊，第 44 卷，第 7 期，第 3614–3633 页，2021年。[8] Mingxing Tan 和 Quoc Le，《EfficientNet：重新思考卷积神经网络的模型缩放》，载于国际机器学习会议。PMLR，2019年，第 6105–6114 页。[9] Tsung-Yi Lin、Piotr Doll´ar、Ross Girshick、Kaiming He、Bharath Hariharan 和 Serge Belongie，《用于目标检测的特征金字塔网络》，载于*IEEE* 计算机视觉与模式识别会议论文集，2017年，第 2117–2125 页。

[10]Yaosheng Lu、Mengqiang Zhou、Dengjiang Zhi、Minghong Zhou、Xiaosong Jiang、Ruiyu Qiu、Zhanhong Ou、Huijin Wang、Di Qiu、Mei Zhong 等，“用于分割耻骨联合与胎头的 jnu-ifm 数据集”，*Data in brief*，第 41 卷，第 107904 页，2022 年。

[11] Yu Chen、Bo Deng、Zilun Peng、Yongle Zhou、Pan Sun、Jin Wan 和 Jieyun Bai，“CoMT：用于宫颈分割的协同训练均值教师半监督训练框架”，载于 *2025* 年第 *22* 届 *IEEE* 生物医学成像国际研讨会 *(ISBI)*。IEEE，2025 年，第 1–5 页。

[12]Yifan Zhang、Jinwen Zhu、Shun Long、Jieyun Bai、Yaosheng Lu 和 Gaowen Chen，“产时超声图像中宫颈扩张的自动测量”，*Signal, Image and VideoProcessing*，第 19 卷第 3 期，第 229 页，2025年。