

中图法分类号: TP391.4; TP18 文献标识码: A 文章编号: 1006-8961(2026)07-2553-16

论文引用格式: Chen Y, Zhu D X, Su L, Li K and Pecht M. 2026. Multitask learning model for fetal heart standard plane recognition with high-recall classification. Journal of Image and Graphics, 31(7):2553-2568(陈宇, 朱栋晓, 宿磊, 李可, Michael Pecht. 2026. 融合高召回分类策略的胎心标准切面多任务识别模型. 中国图象图形学报, 31(7):2553-2568)[DOI:10.11834/jig.250334]

融合高召回分类策略的胎心标准切面 多任务识别模型

陈宇¹, 朱栋晓^{2*}, 宿磊¹, 李可¹, Michael Pecht³

1. 江南大学智能制造学院, 无锡 214122; 2. 江南大学附属医院, 无锡 214122;
3. 马里兰大学CALCE中心, 马里兰州 20742, 美国

摘要: 目的 胎儿心脏超声检查是诊断先天性心脏病的重要手段, 但由于胎儿心脏体积小、搏动速率高且位置受体位变化影响, 5个标准切面的捕捉时效要求高且获取过程重复性差, 严重制约了筛查效率。为实现准确、全面的标准切面自动识别, 提出一种基于“先检测后分类”策略的胎儿心脏标准切面识别多任务模型 FHSP-Net(fetal heart standard planes net)。方法 在检测任务中, 设计路径交织特征金字塔(path weaved network, PWN)、目标感知增强模块(target-aware enhancer, TAE)并引入均衡损失函数(equalization loss version 2, EQLv2), 解决胎心超声检测数据集在多尺度与小目标检测、噪声与伪影干扰、样本数量不均等固有难题, 提供精确的第一阶段结果; 针对基于解剖结构包含条件的切面识别算法精确率高但召回率低的问题, 设计候选视图评分识别算法, 在保证高精确率的同时显著提高召回率。结果 在测试集上的实验结果表明, FHSP-Net多任务模型在解剖结构检测与标准切面识别任务中均展现出优异性能: 检测任务平均精度均值达到0.962, 相较基线提升0.027; 候选视图评分识别算法较基于解剖结构包含条件的切面识别算法准确率提升0.10。最终, FHSP-Net切面识别准确率达0.959, 较基线与解剖结构包含条件切面识别算法的组合模型准确率提高0.194, 验证了其有效性与优越性。结论 所提FHSP-Net多任务模型在胎心结构检测和标准切面识别任务上均表现优异, 路径交织特征金字塔、目标感知增强模块及均衡损失函数显著提高了结构检测模型的鲁棒性, 候选视图评分算法在提升分类准确性与召回率方面展现出显著优势, 能为超声智能辅助诊断提供有力支持。

关键词: 胎儿心脏超声检测; 标准切面识别; 关键解剖结构识别; 深度学习; 目标检测

Multitask learning model for fetal heart standard plane recognition with high-recall classification

Chen Yu¹, Zhu Dongxiao^{2*}, Su Lei¹, Li Ke¹, Pecht Michael³

1. College of Intelligent Manufacturing, Jiangnan University, Wuxi 214122, China; 2. Affiliated Hospital of Jiangnan University, Wuxi 214122, China; 3. Center for Advanced Life Cycle Engineering, University of Maryland, Maryland 20742, USA

Abstract: Objective Fetal cardiac ultrasound examination plays a pivotal role in the prenatal diagnosis of congenital heart disease, which is one of the most prevalent and serious birth defects worldwide. However, clinical acquisition of fetal heart

收稿日期: 2025-07-21; 修回日期: 2025-11-21; 预印本日期: 2025-11-28

* 通信作者: 朱栋晓 617785712@qq.com

基金项目: 国家自然科学基金项目(U23B2044, 52375099); “太湖之光”科技攻关计划(医疗卫生技术攻关)项目(Y20232028)

Supported by: National Natural Science Foundation of China (U23B2044, 52375099); “Taihu Light” Scientific and Technological Key Project (Medical and Health Technology Focus) (Y20232028)

standard planes is highly challenging because of several physiological and technical constraints. The fetal heart is small in size, has a high beating rate, and its position in the maternal abdomen is easily influenced by fetal posture and movement. As a result, capturing the five clinically required standard planes, such as the four-chamber and outflow tract views, demands high temporal precision and operator experience. Moreover, the acquisition process suffers from poor repeatability, further hindering consistent and efficient screening outcomes. These challenges limit the overall effectiveness of fetal cardiac ultrasound screening in routine clinical workflows. To address these limitations and promote the automation of fetal cardiac ultrasound analysis, this study proposes FHSP-Net, a novel multitask deep learning model designed for accurate and comprehensive recognition of fetal heart standard planes. The model adopts a detection-then-classification strategy, which explicitly separates the anatomical structure detection task from the subsequent plane classification task, enabling effective feature extraction and decision-making for each subtask. This design reflects the clinical reasoning process in which a sonographer locates key anatomical structures and interprets them in the context of a specific standard plane.

Method In the detection task, FHSP-Net incorporates several carefully designed components to handle the inherent complexities of fetal ultrasound imaging. A path-weaved network (PWN) is developed to improve the multiscale representation of features, particularly enhancing the detection of small and morphologically variable cardiac structures. In addition, a target aware enhancer (TAE) module is introduced to strengthen object-level context modeling, making the detection branch highly sensitive to anatomical structures that are often obscured by noise, artifacts, or low contrast. Equalization loss function version 2 (EQLv2) is adopted to further mitigate the effect of data imbalance, which is common in clinical datasets where some structures or views are underrepresented. This loss formulation ensures that the model maintains stable learning dynamics across different categories and structure sizes, yielding robust and generalizable detection results. While the detection task ensures a strong foundation by locating relevant anatomical regions, the plane recognition task focuses on accurately classifying ultrasound views on the basis of the detected structures. Traditional algorithms that rely on whether certain anatomical structures are present in an image (i. e., anatomical inclusion conditions) tend to have high precision but suffer from low recall. In other words, although they make only a few mistakes in their predictions, they often fail to identify all valid instances of each standard plane, resulting in missed diagnoses. To overcome this issue, we propose a candidate view scoring algorithm that evaluates and ranks multiple candidate views generated during the detection stage. This algorithm integrates structural cues and contextual features to refine classification decisions, substantially improving the model's ability to capture valid plane instances across varied imaging conditions. **Result** Experiments are conducted on a self-built fetal cardiac ultrasound dataset to evaluate the effectiveness of FHSP-Net. The dataset comprises a wide range of clinical scenarios and includes comprehensive annotations for anatomical structures and standard plane labels. Quantitative results demonstrate the superior performance of FHSP-Net across both subtasks. In the detection task, the model achieves a mean average precision of 0.962, representing a 0.027 improvement over a strong baseline model. This finding indicates that the proposed enhancements in feature fusion, object sensitivity, and loss balancing substantially benefit structure localization performance. In the plane recognition task, the candidate view scoring algorithm boosts classification accuracy by 0.10 compared with conventional structure-inclusion-based approaches. Ultimately, FHSP-Net attains an overall standard plane recognition accuracy of 0.959, representing a 0.194 improvement over the baseline model and the anatomical-inclusion-condition-based classification method. These results strongly validate the effectiveness and advantage of the proposed multitask design and its components. **Conclusion** The FHSP-Net multi-task model demonstrates strong capabilities in anatomical structure detection and fetal heart standard plane recognition. The integration of PWN, TAE, and EQLv2 enhances robustness and precision at the detection stage even under challenging imaging conditions. The candidate view scoring algorithm effectively addresses the recall bottleneck of previous classification methods, ensuring that the model captures a large portion of valid standard plane instances without sacrificing accuracy. Together, these innovations make FHSP-Net a reliable and efficient solution for intelligent ultrasound analysis. FHSP-Net has potential for advancing computer-aided diagnosis in prenatal cardiology.

Key words: fetal cardiac ultrasound detection; standard plane recognition; key anatomical structure recognition; deep learning; object detection

0 引言

先天性心脏病(congenital heart disease, CHD)是一种常见的先天性畸形,发病率约为6%至9%(Li等,2019)。CHD所致的残疾和长期医疗负担,不仅影响家庭福祉,也对社会资源造成持续压力。实现CHD的早诊断与早干预,是减少出生缺陷、提升人口素质的关键举措,对我国妇幼健康保障具有重要意义。

超声检查是评估胎心结构与功能的主要影像手段。根据国际妇产科超声学会发布的胎儿心脏超声指南,如图1所示,以腹围大血管切面(abdominal view, ABDO)为扫查起点,向上采集四腔心切面(four-chamber view, 4CH)、左室流出道切面(left ventricular outflow-tract view, LVOT)、右室流出道切面(right ventricular outflow-tract view, RVOT)、三血管切面(three-vessel view, 3VV)和三血管气管切面(three-vessel-and-trachea view, 3VT),可有效提升CHD的检出率(Carvalho等,2023)。然而,由于胎心体积小、搏动速率快以及空间位置受胎动和体位变化影响,标准切面的获取高度依赖操作者经验;且因捕捉时机稍纵即逝、获取过程重复性差,扫查过程中的延迟或暂停过早都会导致错失关键帧(徐光柱等,2024),影响切面获取的质量。因此,实现对胎儿心脏标准切面的快速、准确识别,已成为提升超声产筛效率与质量的关键路径。

近年来,深度学习在图像识别与目标检测等图像处理任务中表现出卓越性能,为胎心超声切面识别提供新思路(徐光柱等,2023)。该任务通常被建模为图像分类问题,通过改进特征融合策略、引入时间序列建模或迁移学习等方法提升模型鲁棒性。例如,Chen等人(2017)基于长短期记忆网络挖掘了超声心动图相邻帧的时间关系,将4CH切面的准确率提升至0.846。Pu等人(2021)在使用YOLOv3(you only look once version 3)实现分类功能的基础上,加入了时间依赖模块和域转移模块,进一步识别了4CH切面上的收缩末期和舒张末期帧,检测误差控制在1.25帧以内。He等人(2024)提出了以YOLOv5为骨干结构的胎儿心脏超声标准切面检测模型,引入残差结构融合不同尺度的特征图,实现对4CH、LVOT及RVOT切面的分类。

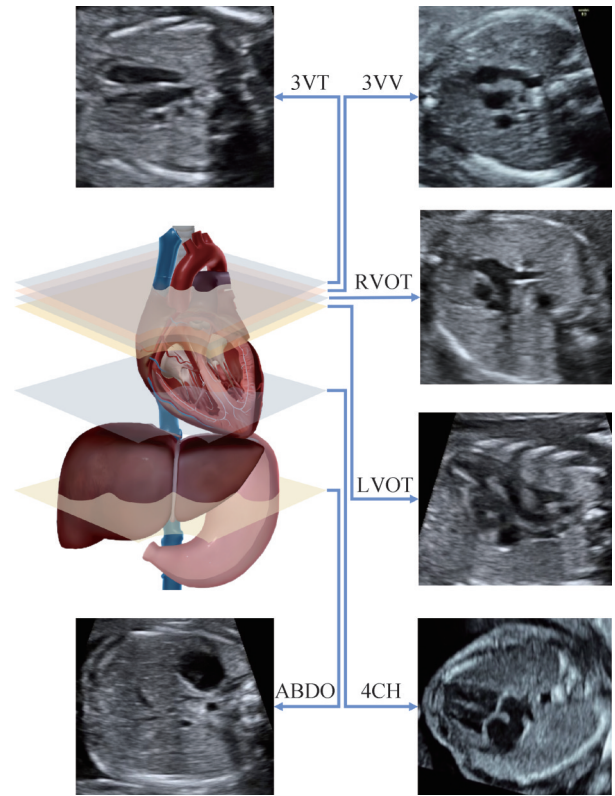


图1 胎儿心脏超声检查的重要切面

Fig. 1 Key views in fetal cardiac ultrasound examination

然而,传统分类模型由于无法精确表征图像中各解剖结构的可见程度,大多聚焦于4CH切面,无法满足临床的多切面诊断需求。为此,Zhang等人(2021)提出基于结构检测结果的切面识别策略,即先检测解剖结构的位置,再依据各切面的结构可见性信息,同时实现对切面类别的识别。这种“先检测后分类”策略在多切面识别任务中展现显著优势:目标检测模型定量分析能力的引入使模型分类结果具备结构级的可解释性,在图像质量较差或背景干扰明显时,分类结果更具鲁棒性;同时由于引入严格的结构约束,切面识别的精确率极高(Li等,2024)。但该策略高度依赖检测结果的准确性,对局部误检敏感,当存在结构误检或交叉检测时,易被误判为非标准类,导致召回率明显偏低。因此,在“先检测后分类”框架下,进一步提升目标检测性能,并优化分类逻辑与评分策略,是提升系统整体性能的关键。

为了在精准识别5个标准切面的同时准确识别13种解剖结构,如图2所示,本文构建“先检测后分类”多任务胎心超声标准切面识别模型FHSP-Net(fetal heart standard planes net):1)针对基线目标检测模型在多尺度与小目标检测、噪声伪影干扰、样本

数量不均等方面的挑战,提出3项关键改进:(1)设计路径交织特征金字塔网络(path-weaved network, PWN),增强多尺度融合能力;(2)提出目标感知增强模块(target aware enhancer, TAE),强化结构细节与上下文信息表达;(3)引入均衡损失函数(equalization loss function version 2, EQLv2),缓解轻度长尾分布下样本类别不均衡导致的性能退化。2)针对基于解剖结构包含规则的识别算法召回率显著偏低的问题,设计候选视图评分识别算法,结合置信度与全连接层分类逻辑实现对5种标准切面的精确识别与量化评估。

1 方 法

1.1 胎心解剖结构检测

针对胎儿心脏在图像中位置与朝向变化显著的

问题,本文采用旋转边界框(oriented bounding boxes, OBB)回归方式替代传统水平框,以更准确地表征解剖结构的分布,缓解因心轴偏转引起的定位误差,如图3所示。因此,构建基于YOLOv8-OBB的解剖结构检测网络PTE-YOLO,整体结构包括骨干网络、特征融合颈部网络和旋转检测头部网络,用于识别13类胎心解剖结构。输入图像首先经Mosaic数据增强策略处理以提高模型泛化能力,随后送入骨干网络提取多尺度特征;在颈部网络中,设计路径交织特征金字塔网络增强多尺度融合能力,显著提升对气管等小尺度结构的检测性能;设计目标感知增强模块强化结构细节与上下文信息表达,有效抑制噪声干扰与过拟合风险。最终,多尺度融合特征由旋转检测头输出目标的类别、位置及旋转角度,完成第1阶段的解剖结构定位与提取。

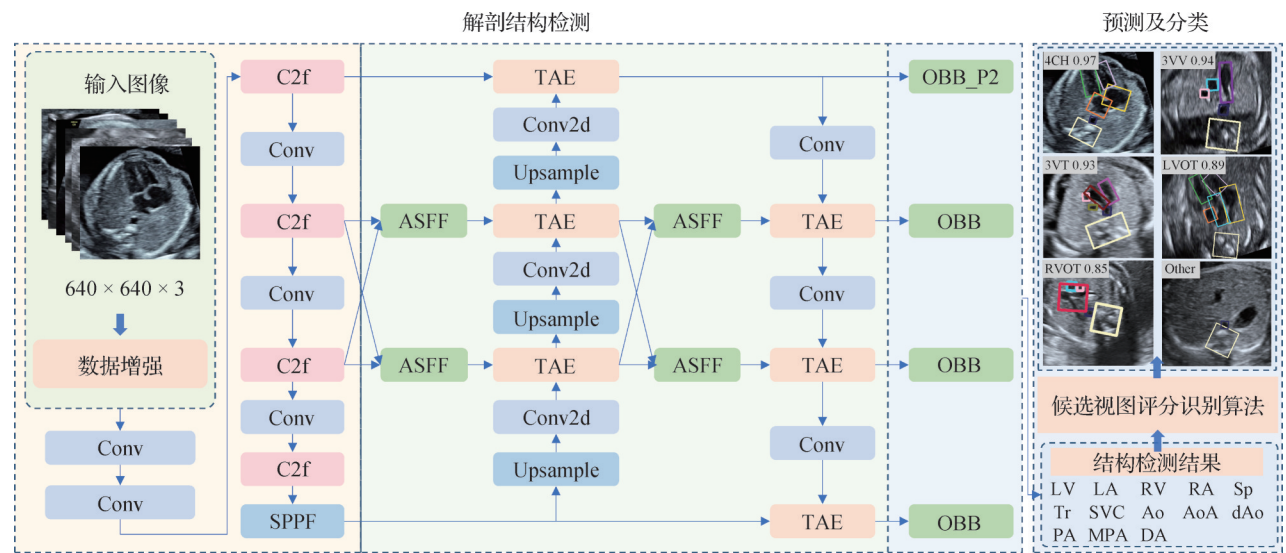


图2 FHSP-Net多任务模型结构
Fig. 2 Architecture of the FHSP-Net multitask model

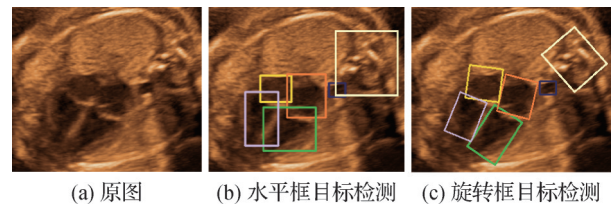


图3 水平框目标检测与旋转框目标检测效果对比
Fig. 3 Comparison of detection performance between horizontal and rotated bounding boxes ((a) original image; (b) horizontal box detection; (c) rotated box detection)

1.1.1 路径交织特征金字塔

胎心各解剖结构在尺寸上差异显著,且同一结

构在不同切面中形态多变,导致尺度跨度较大。部分结构目标较小且特征有限,进一步增加了多尺度特征图的构建难度。YOLOv8OBB采用的路径聚合特征金字塔(path aggregation network, PAN)(Liu等, 2018),通过自顶向下和自底向上的方法构建多尺度特征图,虽然实现了相邻层级之间的信息传递,但难以充分捕捉复杂结构中的跨尺度语义依赖,限制检测性能。为提升模型对多尺度结构,尤其是小目标的检测能力,本文提出路径交织特征金字塔网络,如图4所示,其中 P_2 、 P_3 、 P_4 、 P_5 分别对应4、8、16和32倍下采样尺度的特征图,通过重构特征融合路径并引入

高分辨率辅助分支增强特征表达能力。

深层特征图虽然空间分辨率低,但具备更大感受野和更强语义表达能力,适用于大目标检测;浅层的大尺寸特征图通道数较少、感受野有限,但包含更多空间信息,对于检测可用特征少的小尺寸解剖结构更为关键。因此,PWN通过增添一组上采样操作构建更高分辨率的特征图 P_2 ,并直接连接至小目标检测头OBB_P2。考虑到小目标的细节特征在深层传递中易被稀释(Lin等,2017),该分支不进行复杂融合操作,尽可能保留原始空间细节信息。

为缓解原始PAN结构中不同层级特征间信息

交互不足的问题,PWN在同级的中间层中引入跨尺度交织路径,使浅层细节与深层语义在跨级融合前实现初步耦合,提升融合特征的判别能力。在跨尺度的同级特征融合中,为避免简单相加引发的特征冲突,交织路径中引入自适应空间特征融合模块(adaptive spatial feature fusion, ASFF)(Yang等,2023)。该模块通过线性插值上采样或卷积下采样实现空间对齐和通道压缩,以统一特征图维度。随后利用卷积层和softmax激活函数生成动态权重,引导多尺度特征在空间位置上自适应融合。融合结果进一步引入残差结构,增强特征的非线性表达能力和稳定性。

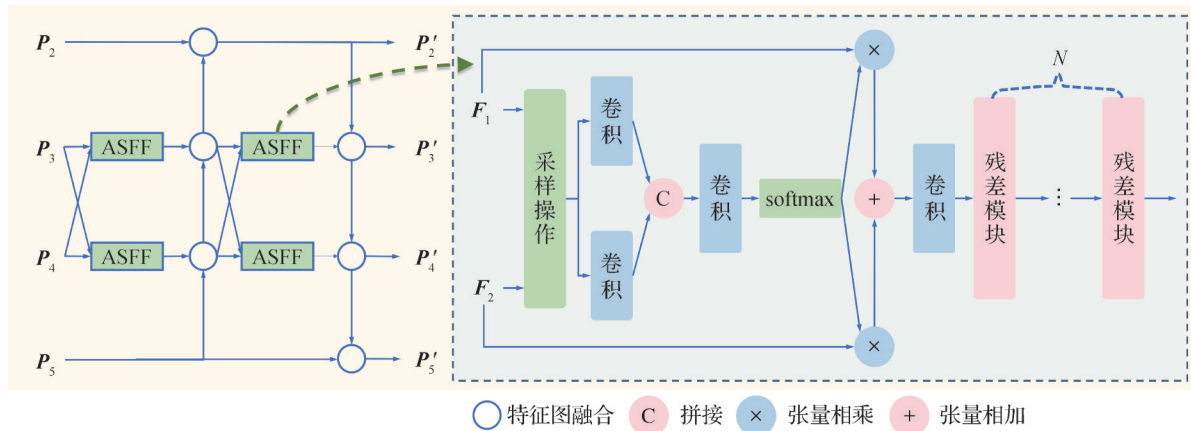


图4 PWN结构

Fig. 4 Architecture of the PWN

1.1.2 目标感知增强模块

由于胎儿心脏超声影像常伴有对比度低、边缘模糊、信号衰减、斑点噪声及多种伪影,削弱了解剖结构的可辨识性,增加了检测任务的难度。YOLOv8-OBB在解剖结构检测中采用的特征金字塔(feature pyramid network, FPN)仅通过Concat操作拼接不同尺度特征图,缺乏对各尺度特征信息质量的差异性建模,可能导致模型对噪声区域或无关背景特征过拟合,影响检测结果的鲁棒性。为增强模型对目标区域的感知能力并抑制干扰特征,本文提出目标感知增强模块TAE,如图5所示,从通道与空间两个维度对特征图进行逐级增强,强化目标表达与判别能力。

TAE模块由通道引导机制与特征聚焦机制组成,分别从通道维度和空间维度实现目标特征的感知增强。通道引导机制旨在评估各通道对目标结构的贡献度。该机制将输入特征图视为多通道集合,首先通过全局平均池化和最大池化操作聚合每个通

道的空间上下文信息,用以描述每个通道的最大特征和平均特征表现。随后,两种特征描述经由一个包含两个全连接层和ReLU(rectified linear unit)激活函数的共享前馈网络,生成通道权重,并与输入的拼接特征图 f 融合,输出通道引导特征图 M_c 。

通道引导特征图 M_c 被送至特征聚焦机制,进一步在空间维度增强目标区域响应。该机制首先通过卷积核为1的卷积对 M_c 进行通道调整,然后通过两条并行支路建模空间特征分布:一条路径通过下采样及两个卷积核为1的卷积提取全局信息,以增强语义一致性;另一条路径通过两个卷积核为1的卷积进行局部建模,挖掘细节显著区域。两路径输出结果相乘后,经sigmoid激活生成空间注意权重图,表征每个空间位置对目标的响应程度。该权重图与通道引导特征图 M_c 逐元素相乘,得到最终的融合特征图 M_r 。在该过程中,只有当某一区域同时具备显著细节与合理语义时,其特征响应才被增强,从而有

效突出解剖结构特征区域、抑制超声噪声伪影干扰,实现空间维度上的感知强化。最终,TAE模块实现了从通道层面到空间维度的逐级目标感知增强,其计算过程可表示为

$$\mathbf{M}_T = \mathbf{M}_F(\mathbf{M}_C) \quad (1)$$

$$\mathbf{M}_C(f) = f \times \sigma \left(W_1 \left(\text{ReLU} \left(W_0 \left(\text{Avg}(f) \right) \right) \right) + W_1 \left(W_0 \left(\text{ReLU} \left(W_0 \left(\text{Max}(f) \right) \right) \right) \right) \right) \quad (2)$$

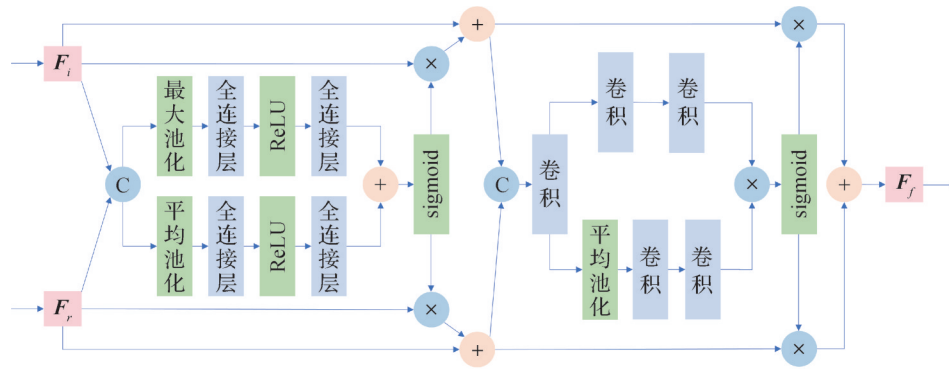


图5 TAE模块结构图

Fig. 5 Architecture of the TAE module

1.1.3 均衡损失

由于胎心超声图像各标准切面中包含的解剖结构存在差异,胎儿心脏解剖结构检测数据集样本类别分布不均,呈现出轻度的长尾分布特征,如图6所示。部分关键结构(如Sp、dAo等)在多个标准切面中可观测,成为头部类,而某些结构仅可于特定切面观测(如Tr、DA、AoA等),为尾部类。若仅通过样本数量均衡或重采样策略缓解不平衡问题,往往因切面结构间存在重叠,反而会同步增加头部类样本数量,难以从根本上解决类别不均。该分布特性不仅影响模型训练的稳定性,还会导致预测结果倾向样本数量丰富的头部类,限制了模型的实用性。为此,本文基于均衡损失(Tan等,2021)重构了YOLOv8-OBB的损失函数,提高模型对尾部类的泛化能力。

由于多标签任务中尾部类正样本稀缺、负样本数量庞大,EQLv2以交叉熵损失函数为基础,通过动态计算每批次的正负梯度贡献比例,自适应调整各类别梯度权重。具体而言,EQLv2设计了梯度比例权重 w_c ,根据尾部类正负样本的梯度贡献动态赋权。以 y_i^n 表示批次中第 n 个样本在第 i 类的one-hot标签, p_i^n 表示批次中第 n 个样本被识别为类别 i 的概率,

$$\mathbf{M}_F(f) = f \times \sigma \left(\left(W_4 \left(W_3 \left(\text{Avg} \left(W_7(f) \right) \right) \right) \right) \times \left(W_6 \left(W_5 \left(W_7(f) \right) \right) \right) \right) \quad (3)$$

式中, σ 为sigmoid激活函数, W_1 和 W_2 分别代表通道引导机制中两个全连接层的权重, W_3 、 W_4 、 W_5 和 W_6 表示特征聚焦机制并行路径中各卷积层的权重参数, Avg 代表平均池化操作, Max 代表最大池化操作, W_7 为 1×1 卷积操作的权重参数。

N 表示类别总数,则EQLv2可以表示为

$$L_E = w_c \times \frac{1}{N} \sum_{n \in N} - \left[y_i^n \log(p_i^n) + (1 - y_i^n) \log(1 - p_i^n) \right] \quad (4)$$

式中, w_c 为计算权重,其计算依赖于正负梯度比。为了实现这一点,将 g_i^t 定义为任务 i 在迭代 t 之前的累积正梯度与负梯度的比例,通过下式计算该迭代中正梯度 p_i^t 和负梯度 q_i^t 的权重,进而求出EQLv2的权重 w_c ,具体为

$$p_i^t = 1 + \alpha(1 - f(g_i^t)) \quad (5)$$

$$q_i^t = f(g_i^t) \quad (6)$$

$$f(x) = \frac{1}{1 + e^{-\gamma(x - \mu)}} \quad (7)$$

$$w_c = p_i^t \times y_i^n + q_i^t \times (1 - y_i^n) \quad (8)$$

为持续动态调整正负梯度比,以在不同批次中更新权重,需要在每个批次中计算本批次分类器的输出 z_i 中累计的正梯度 $\nabla_{z_i}^p(L)$ 和负梯度 $\nabla_{z_i}^n(L)$,并为下一次迭代更新梯度比 g_i^{t+1} ,具体为

$$g_i^{t+1} = \frac{\sum_0^t |p_i^t \times \nabla_{z_i}^p(L)|}{\sum_0^t |q_i^t \times \nabla_{z_i}^n(L)|} \quad (9)$$

$$\nabla_{z_i}^p(L) = \frac{1}{N} \sum_{n \in N} y_i^n (p_i^n - 1) \quad (10)$$

$$\nabla_{z_i}^n(L) = \frac{1}{N} \sum_{n \in N} (1 - y_i^n) p_i^n \quad (11)$$

因此,改进后的YOLOv8-OB模型所采用的损失函数 L 由均衡损失 L_E 、交并比损失 L_{IoU} 和分布焦点损失 L_{df} 组成,其表达式为

$$L = \lambda_1 L_E + \lambda_2 L_{IoU} + \lambda_3 L_{df} \quad (12)$$

式中, λ_1 表示EQLv2对应的权重系数; λ_2 表示交并比损失对应的权重系数; λ_3 表示分布焦点对应的权重系数。

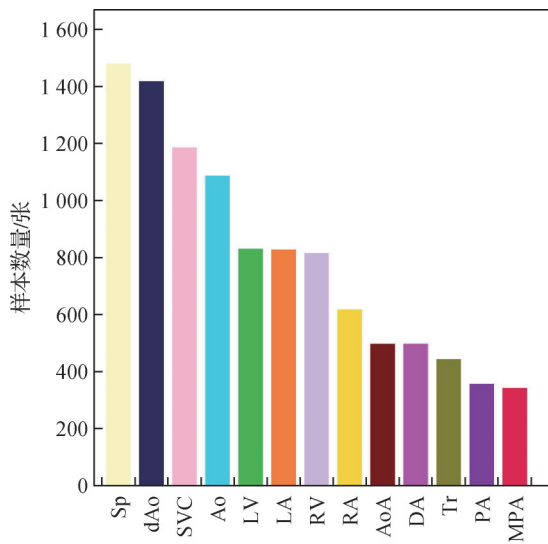


图6 数据分布图

Fig. 6 Data distribution diagram

1.2 候选视图评分识别算法

临床中超声医师通常根据多个解剖结构的组合关系判断超声图像所属切面。ABDO切面在超声扫查中通常作为扫查的起点,通过降主动脉(descending aorta, dAo)、脊椎(spine, Sp)及腹部的胃泡进行识别,且在探头移动过程中, dAo和Sp仍可在4个标准心切面中观察到。在4CH切面中,左心室(left ventricle, LV)、左心房(left atrium, LA)、右心室(right ventricle, RV)、右心房(right atrium, RA)4个心腔应清晰可见;对于LVOT切面, LV、主动脉(aorta, Ao)是观察的重点,同时切面也可能观察到LA、RV、RA; RVOT切面需包含主肺动脉(main pulmonary artery, MPA)、Ao、上腔静脉(superior vena cava, SVC); 3VV切面与RVOT切面类似,但观察的重点是肺动脉(pulmonary artery, PA)、Ao及SVC等血管的位置和大小关系;在3VT切面中,必须能观察到动脉导管

(ductus arteriosus, DA)、主动脉弓(aortic arch, AoA)和气管(trachea, Tr)等解剖结构。基于上述结构包含条件进行分类,是“先检测后分类”策略中最常用的方式,具有极高的识别精确率。然而,该方法依赖于“是否包含某结构”的刚性规则,导致对局部误检高度敏感,识别过程中的容错能力较弱,尤其在图像质量不佳时容易产生结构重复检测或结构间混淆,显著影响识别的召回率。例如,当检测结果同时包含PA与AoA,或Ao和AoA均被检测出时,仅凭结构包含条件难以准确识别图像切面。

为解决上述问题,本文提出了候选视图评分识别算法,对分类逻辑进行改进,显著提升模型召回率。该算法首先将标准切面划分为两个大类:一类包括4CH与LVOT;另一类包括RVOT、3VV与3VT。对于检测结果中包含LV、LA、RV和RA中任一结构的图像,归入第一大类;若同时检测到Ao,则判定为LVOT,否则为4CH。若未满足上述条件,但图像中包含MPA、PA、DA、Ao、AoA、SVC或Tr等结构,则归入第二大类,并分别计算其属于RVOT、3VV和3VT三类切面的得分,最终将得分最高的切面类别作为预测结果。当前两大类条件均不满足时,判定为其他切面,其中包含ABDO切面。ABDO切面位于胎儿腹部,主要用于超声筛查中的起始定位,其解剖特征与心脏结构无直接关联,因此归入“其他切面”类别。

各类别的得分通过结构置信度加权计算获得,权重值如表1所示。权重的设计遵循两个原则:1)权重值的大小需反映医师实际进行切面识别时的注意力区域,如在LVOT切面中, Ao和LV的权重高达0.5和0.25。2)在满足上述条件的前提下,同一大类下不同切面中相同或对应结构的权重比例保持一致。如在第二大类RVOT、3VV和3VT三类切面的判别中, Ao与AoA比例一致。

根据上述加权得分,算法在判别切面类型的同时,还会保留分数,用于后续的质量排序与筛选。此外,考虑到PTE-YOLO模型在处理低质量图像时可能对同一解剖结构产生多个冗余检测框,算法中对每类结构仅保留置信度最高的一个候选框用于评分,避免冗余框导致的得分偏移。该候选视图评分识别流程如图7所示,并支持后续的图像质量排序与最佳切面的选择。

表1 切面所含解剖结构类别及计分权重

Table 1 Anatomical structure categories and scoring weights for each standard plane

切面	可观测解剖结构	计分权重
4CH	LV, LA, RV, RA, dAo, Sp	0.25:0.25:0.25:0.25:0.05:0
LVOT	LV, Ao, RV, LA, RA, dAo, Sp	0.25:0.5:0.1:0.05:0.05:0.05:0
RVOT	MPA, Ao, SVC, dAo, Sp	0.35:0.35:0.2:0.05:0
3VV	PA, Ao, SVC, dAo, Sp	0.35:0.35:0.2:0.05:0
3VT	DA, AoA, SVC, Tr, dAo, Sp	0.35:0.35:0.2:0.2:0.05:0

2 实验和结果

2.1 实验设置

2.1.1 环境配置与数据集

实验在搭载 NVIDIA RTX A6000(48 GB 显存) 显卡的高性能服务器上进行,采用 PyTorch 深度学习 框架实现所有模型训练与评估。训练过程中使用 Adam 优化器,并引入 Mosaic 数据增强策略提升模型的 泛化能力与复杂场景下的鲁棒性。模型训练共迭代 180 个 epoch, batch size 设置为 16, 初始学习率为 0.001, 并采用余弦学习率衰减策略, 减少振荡。

本文采用来自江南大学附属医院的胎心超声图 像数据验证所提方法的有效性。数据集采集自 GE Vivid E8 超声诊断设备, 孕周范围为 14~28 周, 经过 脱敏处理后共获得 2 949 幅二维胎心超声图像, 图像 分辨率为 1 136 × 852 像素。每幅图像由具有丰富经 验的超声医师使用 roLabelImg 旋转目标标注工具进 行人工标注, 涵盖 6 个切面(ABDO 及 5 个胎心标准 切面)中的 13 种解剖结构。图像数据随机划分为训 练集与测试集, 比例为 7:3, 具体分布如表 2 所示。

2.1.2 评价指标

为客观准确地评估模型性能, 本文从检测性能 与实用性两个维度进行实验。在解剖结构检测任务 中, 采用精确率(precision, Prec)、召回率(recall, Rec)和平均精度均值(mean average precision, mAP) 作为主要评价指标。Prec 表示预测为正例的样本中 真正例所占比例, 反映查准率; Rec 表示真正例的样 本中被预测为正例的比例, 又称查全率; mAP 表示 所有类别“Prec-Rec”曲线下的面积均值, 综合衡量 模型在不同检测阈值下的综合性能。对于标准切面

表2 数据集分布

Table 2 Dataset distribution

数据集	训练集/幅	测试集/幅
ABDO	33	15
4CH	443	191
LVOT	388	165
RVOT	342	147
3VV	359	155
3VT	497	214
总计	2 062	887

识别任务, 采用准确率(accuracy, Acc)、Prec、Rec 和 F1 值(F1-score)作为评价指标。其中, Acc 表示所有 正确分类的样本数与总样本数之比, 反映整体分类 准确性; F1 为 Prec 和 Rec 的调和平均值, 在类别分布 不均时比 Acc 能更客观地反映模型性能。此外, 为 评估模型的部署效率与资源开销, 本文还统计了模 型参数量及单幅图像推理时间, 用于衡量模型的轻 量化程度与实际应用价值。

2.2 检测性能

本文在测试集上对 PTE-YOLO 模型进行了性能 评估, 其在胎儿心脏各关键解剖结构的检测效果如 表 3 所示。整体来看, 模型在各项性能指标上均表 现优异, Prec、Rec 及 mAP 分别达到 0.925、0.930 和 0.962, 且绝大多数类别的精度较基准模型均有明显 的提升, 类别间的精度差异明显缩小, 表明模型有效 缓解了胎心超声数据集固有的多尺度目标、噪声伪 影干扰及类别不均等挑战, 整体实用性显著增强。 以小目标兼尾部类的 Tr 类为例, 其 mAP 从 0.789 提 升至 0.908, 召回率提升 0.218, 进一步验证了模型 对困难样本的识别能力增强。

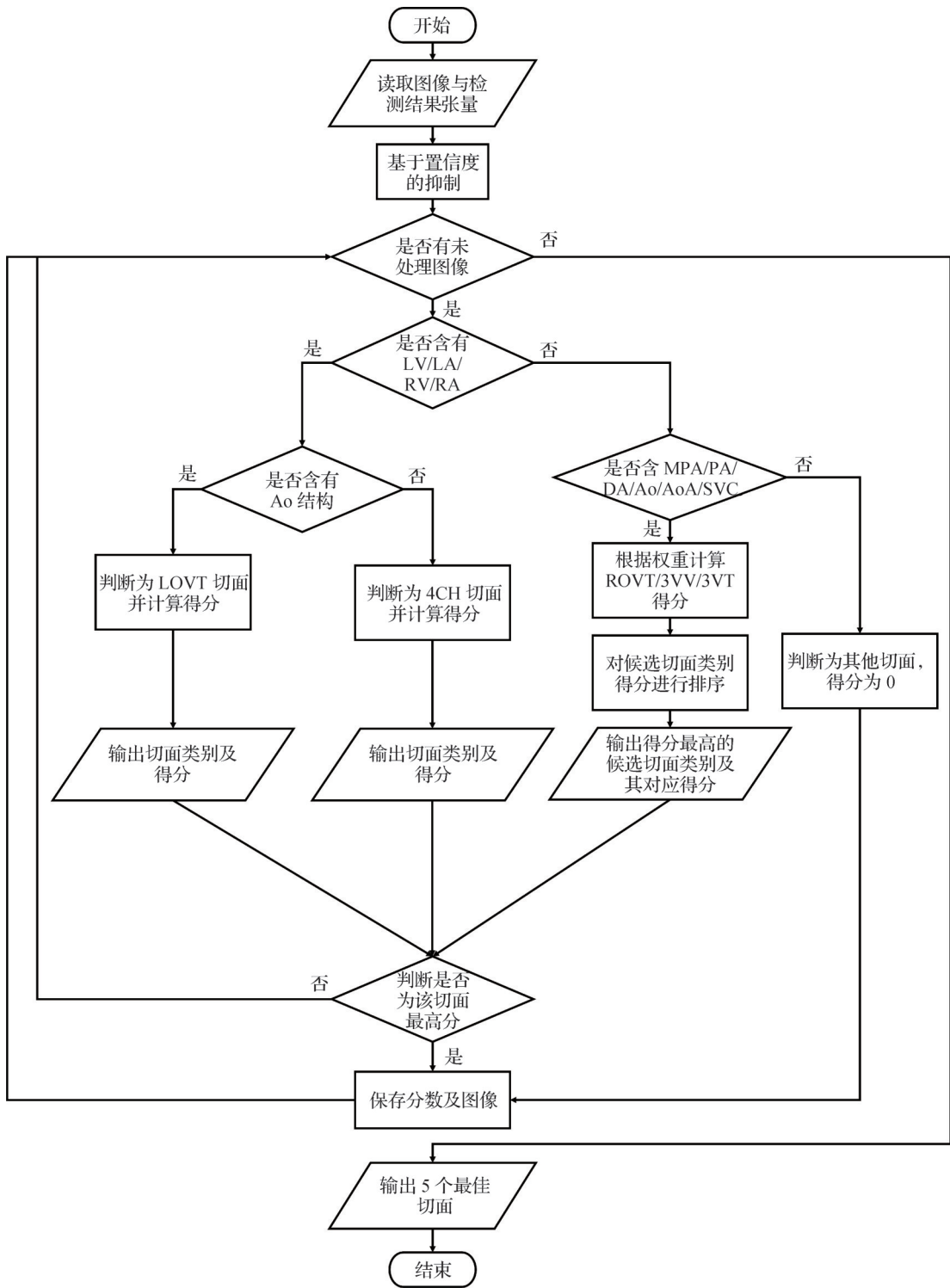


图 7 候选视图评分识别算法流程图

Fig. 7 Flowchart of the candidate view scoring recognition algorithm

2.3 消融实验

为验证 PTE-YOLO 各模块的有效性,本文设计了消融实验,如表 4 所示。

首先,采用未引入任何改进模块的 Baseline 进行检测, Prec、Rec 和 mAP 分别为 0.890、0.895 和

0.935。为验证模型在多尺度、噪声伪影干扰及类别不均等挑战下的表现,分别引入 PWN 特征金字塔、TAE 模块与 EQLv2 损失函数,评估其独立与协同性能贡献。实验结果表明,单独引入 3 种改进均可带来性能增益:引入 PWN 后, mAP 提升至 0.950,证明

表 3 PTE-YOLO 检测结果
Table 3 Detection results of PTE-YOLO

结构 名称	Prec			Rec			mAP		
	Baseline	PTE-YOLO(本文)	提升	Baseline	PTE-YOLO(本文)	提升	Baseline	PTE-YOLO(本文)	提升
LV	0.954	0.980	↑ 0.026	0.972	0.963	–	0.982	0.983	↑ 0.001
LA	0.920	0.950	↑ 0.030	0.941	0.936	–	0.962	0.958	–
RV	0.929	0.944	↑ 0.015	0.946	0.959	↑ 0.013	0.967	0.974	↑ 0.007
RA	0.878	0.923	↑ 0.045	0.920	0.941	↑ 0.021	0.955	0.978	↑ 0.023
Ao	0.916	0.952	↑ 0.036	0.942	0.965	↑ 0.023	0.968	0.984	↑ 0.016
AoA	0.907	0.980	↑ 0.073	0.883	0.934	↑ 0.051	0.948	0.991	↑ 0.043
dAo	0.891	0.890	–	0.905	0.899	–	0.947	0.943	–
Sp	0.907	0.895	–	0.957	0.929	–	0.961	0.947	–
PA	0.701	0.842	↑ 0.141	0.865	0.919	↑ 0.054	0.855	0.936	↑ 0.081
DA	0.858	0.881	↑ 0.023	0.881	0.928	↑ 0.047	0.916	0.954	↑ 0.038
MPA	0.957	0.975	↑ 0.018	0.911	0.913	↑ 0.002	0.967	0.975	↑ 0.008
SVC	0.875	0.938	↑ 0.063	0.870	0.943	↑ 0.056	0.932	0.974	↑ 0.042
Tr	0.878	0.868	–	0.636	0.854	↑ 0.218	0.789	0.908	↑ 0.119
整体	0.890	0.925	↑ 0.035	0.895	0.930	↑ 0.035	0.935	0.962	↑ 0.027

注:加粗字体为每项最优结果。“↑”表示对比Baseline效果上升。“–”表示对该类效果没有提升。

表 4 消融实验结果
Table 4 Ablation study results

方法	PWN	TAE	EQLv2	Prec	Rec	mAP
Baseline	×	×	×	0.890	0.895	0.935
P-YOLO	√	×	×	0.925	0.917	0.950
T-YOLO	×	√	×	0.902	0.916	0.948
E-YOLO	×	×	√	0.915	0.894	0.943
PT-YOLO	√	√	×	0.917	0.927	0.957
PTE-YOLO(本文)	√	√	√	0.925	0.930	0.962

注:加粗字体表示各列最优结果。“√”和“×”分别表示是否包含该模块。

其在多尺度与小目标特征建模方面效果显著;引入TAE模块后,mAP提升至0.948,验证其在目标感知强化与噪声抑制方面的有效性;引入EQLv2后,mAP达到0.943,显示其对类别不均有积极作用。为进一步验证模块间的协同效果,在PWN的基础上引入TAE模块,构建PT-YOLO,mAP提升至0.957。PWN与TAE的组合进一步提升了关键结构区域的特征表达能力,展现出较强的互补性。最终,引入EQLv2构成完整的PTE-YOLO,Prec、Rec和mAP分别提升至0.925、0.930和0.962,验证了所提方法在多个关

键问题上的综合改进能力与整体检测表现。

为验证本文所提出特征金字塔效果,将PWN分别替换为多种主流特征金字塔网络,包括双向特征金字塔网络(bi-directional feature pyramid network, BiFPN)(Tan等,2020)、重参数化泛化特征金字塔(reparameterized generalized feature pyramid network, RepGFPN)(Xu等,2023)、渐进特征金字塔网络(asymptotic feature pyramid network, AFPN)(Yang等,2023)和递归特征金字塔(recursive feature pyramid, RFP)网络(Qiao等,2020),进行关于特征金字

塔的消融实验。为确保公平性,所有特征金字塔均基于 4 个尺度的特征图(P_2-P_5)构建,并统一引入 P_2 小目标检测头,实验结果如表 5 所示。从结果可知,本文改进的 PWN 在多个关键性能维度上均优于现有方法。与 AFPN 和 RFP 相比,PWN 在 Prec、Rec 与 mAP 等指标上表现更优且同时具备更低的参数量。相较于 RepGFPN 和 BiFPN,PWN 在精确率、召回率及 mAP 方面均表现更佳,尽管参数量相对更大,但推理时间较 BiFPN 基本持平,优于 RepGFPN,体现出更优的性能与效率平衡能力。

为验证 TAE 模块在应对胎心超声图像中的模

糊伪影和背景噪声等干扰因素中的效果,本文设计了关于 TAE 模块的消融实验。具体方法为,将 PTE-YOLO 中的 TAE 模块替换为 Concat 模块与主流注意力机制的组合,包括 SE(squeeze and excitation networks)(Hu 等,2019)、SA(shuffle attention)(Zhang 和 Yang, 2021)、CBAM(convolutional block attention module)(Woo 等,2018)及 EMA(efficient multi-scale attention module)(Ouyang 等,2023)等,实验结果如表 6 所示。结果表明,TAE 模块在 Prec、Rec 及 mAP 指标上较其他方法性能最优,在模型参数量与推理时间方面也保持领先水平,综合性能优异。

表 5 关于特征金字塔的消融实验
Table 5 Ablation study on the feature pyramid

方法	Prec	Rec	mAP	参数量	推理时间/(ms/幅)
TE-YOLO-BiFPN	0.909	0.918	0.953	3 584 554	10.5
TE-YOLO-RepGFPN	0.926	0.915	0.958	7 623 904	11.1
TE-YOLO-AFPN	0.916	0.925	0.957	38 885 213	12.0
TE-YOLO-RFP	0.910	0.919	0.953	7 957 752	13.4
PTE-YOLO(本文)	0.925	0.930	0.962	9 192 576	10.7

注:加粗字体表示各列最优结果。

表 6 关于 TAE 模块的消融实验
Table 6 Ablation study on the TAE module

方法	Prec	Rec	mAP	参数量	推理时间/(ms/幅)
PE-YOLO-SE	0.914	0.913	0.956	9 430 464	10.9
PE-YOLO-SA	0.699	0.748	0.770	9 435 636	10.7
PE-YOLO-CBAM	0.919	0.919	0.956	9 525 804	10.8
PE-YOLO-EMA	0.927	0.919	0.957	9 434 160	10.9
PTE-YOLO(本文)	0.925	0.930	0.962	9 192 576	10.7

注:加粗字体表示各列最优结果。

2.4 对比实验

2.4.1 检测性能对比实验

为了验证 PTE-YOLO 性能,本文在测试集上与多种主流旋转目标检测模型进行了对比实验,结果如表 7 所示。PTE-YOLO 在各项指标中表现优异,mAP 达到最高值 0.962,显著优于其他模型。与基线模型相比,Prec、Rec 和 mAP 分别提升了 0.035、0.035 和 0.027,尽管模型参数量和计算复杂度有所增加,但推理速度仍保持在 10.7 ms/幅,对应帧率约 93.45 帧/s,满足实时检测要求。综合来看,PTE-YOLO 在保持较

高运行效率的同时显著提升了检测精度,为后续切面识别任务提供了可靠的第一阶段检测结果。

图 8 展示了部分模型的可视化预测,可见其他主流网络检测普遍存在不同程度的误检和漏检,而 PTE-YOLO 检测结果更接近真实标注,表现更为稳定可靠。

2.4.2 切面识别性能对比实验

为系统评估所提多任务识别模型 FHSP-Net 的切面识别性能,本文在测试集上设计了 3 组对比实验:第 1 组实验对比单任务模型与多任务模型,验证多任

表 7 胎心解剖结构检测结果对比

Table 7 Comparison of detection results for fetal cardiac anatomical structures

方法	Prec	Rec	mAP	参数量	推理时间/(ms/幅)
Baseline	0.890	0.895	0.935	3 085 050	7.9
YOLO11OBB	0.884	0.880	0.927	2 664 042	3.3
YOLO12OBB (Tian 等, 2025)	0.894	0.872	0.930	2 642 250	3.6
ROI Transformer (Ding 等, 2019)	0.837	0.827	0.797	55 110 000	45.4
Oriented R-CNN (Xie 等, 2021)	0.712	0.707	0.751	41 140 000	42.6
ReDet (Han 等, 2021)	0.851	0.843	0.804	31 630 000	27.1
Oriented RepPoints (Li 等, 2022)	0.883	0.875	0.784	36 610 000	38.5
PTE-YOLO (本文)	0.925	0.930	0.962	9 192 576	10.7

注:加粗字体表示各列最优结果。

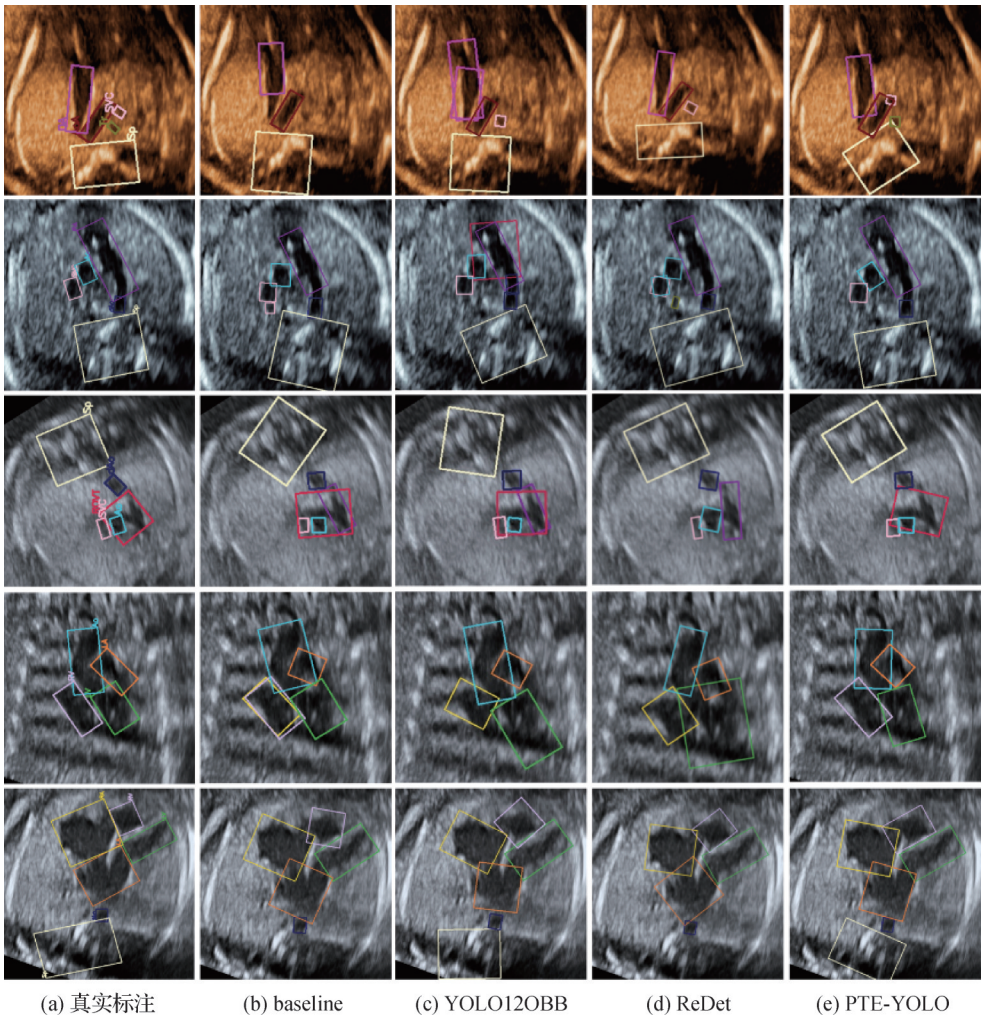


图 8 胎心解剖结构检测网络结果可视化

Fig. 8 Visualization of detection results from the fetal cardiac anatomical structure network

((a) ground truth; (b) baseline; (c) YOLO12OBB; (d) ReDet; (e) PTE-YOLO)

务策略对切面识别的影响;第2组实验设置在统一的分类算法下,在多任务模型中引入不同的结构检测

算法,评估检测性能对切面识别的影响;第3组实验在固定检测算法的基础上,对比不同的切面识别算

法,验证所提出候选视图评分识别算法的有效性。

如表8所示,本文提出的FHSP-Net(PTE-YOLO + 候选视图评分识别算法)获得最高的 Acc,为 0.959,验证了所提方法的优越性。第1组实验中,多任务模型在胎心超声数据集上的切面识别精度整体更高,精确率显著优于单任务模型,验证了“先检测后分类”策略的有效性;第2组实验显示,无论采取何种识别算法,PTE-YOLO在各项指标上均优于其他检测模型,表明检测精度对分类效果有决定

性影响;第3组实验证明,在相同的检测结果基础上,候选视图评分识别算法性能总是显著优于解剖结构包含条件,在保持高精确率的同时大幅提升了召回率。

图9展示了部分标准切面识别方法的可视化结果。对比4CH、3VV和3VT切面的分类结果可以分别看出,当出现误检、交叉检测或漏检时,候选视图评分识别算法能通过置信度抑制和分组评分,减弱刚性约束导致的误判。

表 8 胎心标准切面识别结果对比
Table 8 Comparison of fetal heart standard plane recognition results

类别	方法	Acc	Prec	Rec	F1	推理时间/(ms/幅)
单任务	VGG16(Simonyan 和 Zisserman, 2015)	0.636	0.635	0.628	0.630	12.1
	ResNet50(He 等, 2016)	0.508	0.492	0.492	0.481	7.3
	EfficientNet(Tan 和 Le, 2019)	0.500	0.474	0.478	0.461	5.1
	DenseNet(Huang 等, 2017)	0.841	0.839	0.832	0.833	8.4
多任务	Baseline+解剖结构包含条件	0.765	0.969	0.769	0.847	7.9
	YOLO12OBB+解剖结构包含条件	0.705	0.975	0.708	0.803	3.6
	PTE-YOLO+解剖结构包含条件	0.859	0.980	0.880	0.926	10.7
	Baseline+候选视图评分识别算法	0.933	0.930	0.931	0.930	7.9
	YOLO12OBB+候选视图评分识别算法	0.933	0.933	0.932	0.932	3.6
	PTE-YOLO+候选视图评分识别算法(本文)	0.959	0.962	0.957	0.959	10.7

注:加粗字体表示各列最优结果。

3 结 论

本文针对当前胎心标准切面识别中存在的可识别切面类型有限、分类召回率低等问题,提出一种“先检测后分类”的多任务识别模型FHSP-Net:在检测阶段,针对多尺度小目标、背景伪影及类别不均等挑战,构建了路径交织特征金字塔与目标感知增强模块,并结合均衡损失函数优化YOLOv8OBB模型,提供更准确的第一阶段检测结果;在分类阶段,提出候选视图评分识别算法,引入置信度抑制策略,结合医学先验知识构建结构组合得分机制,显著缓解了强约束结构包含条件带来的误判问题,进一步提升切面识别性能。

在自建胎儿心脏超声数据集上的实验结果表

明,FHSP-Net在结构检测与切面识别任务中均取得了优异表现。相较基准模型,检测阶段的Prec、Rec与mAP分别提升了0.035、0.035与0.027;切面识别阶段Acc、Prec和Rec分别达到0.959、0.962和0.957,优于现有方法。以上结果验证了所提FHSP-Net模型在胎心多标准切面识别任务中的有效性。同时,平均推理时间为10.7 ms/幅(约93.45 帧/s),满足超声实时检测需求。

尽管FHSP-Net在模型性能与实用性方面取得了积极进展,但仍存在两个主要局限:训练与评估数据仅来自同一型号设备,且尚未部署于实际超声设备。这导致模型在真实应用中可能仍面临鲁棒性方面的挑战,如低质量图像、不同仪器成像风格差异或异常心脏病例等情况,其在临床环境下的实际表现仍有待进一步验证。后续工作将重点开展用户界面

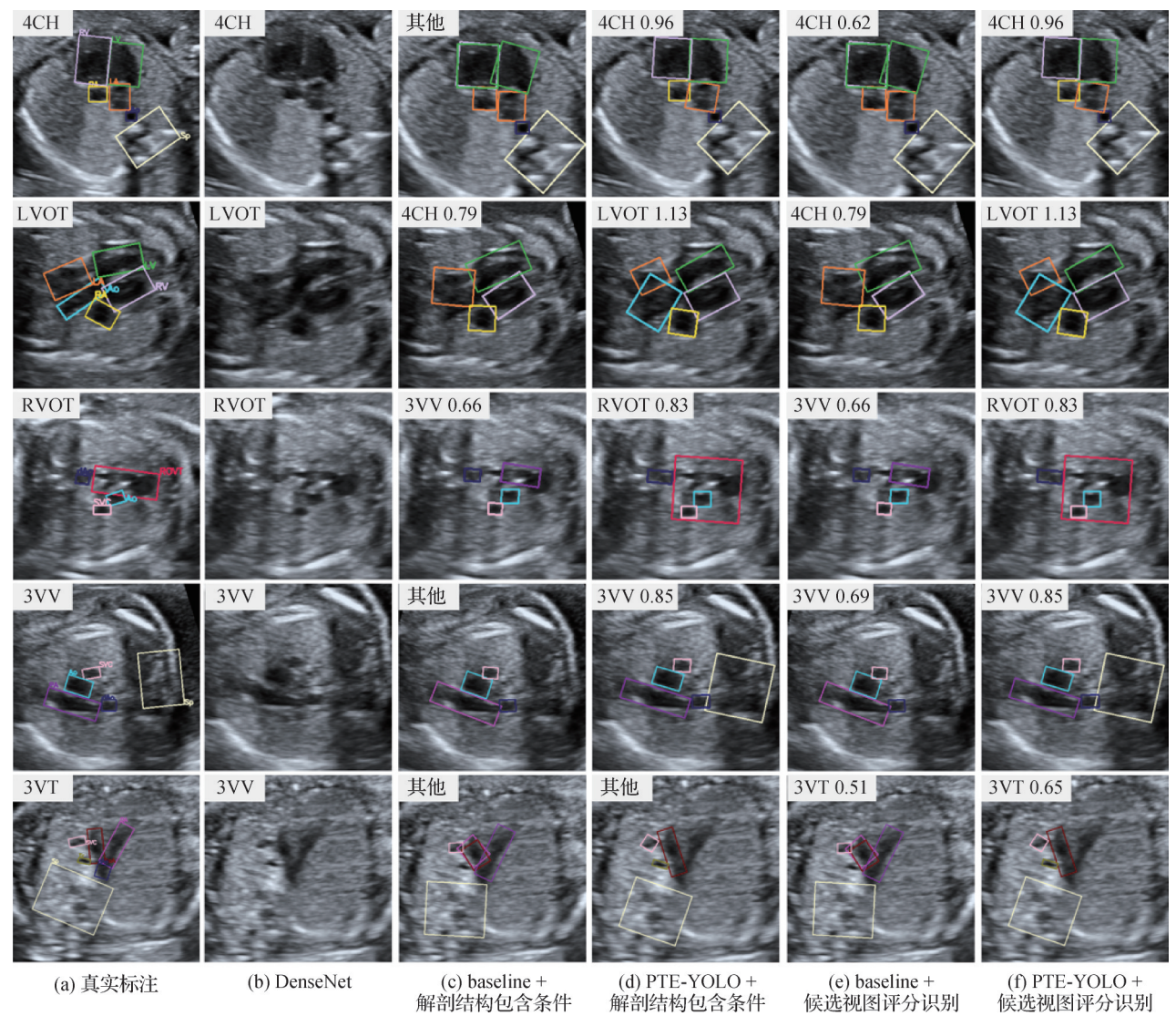


图9 分类及评分结果可视化

Fig. 9 Visualization of classification and scoring results ((a) ground truth; (b) DenseNet; (c) baseline and anatomical structure conditions; (d) PTE-YOLO and anatomical structure conditions; (e) baseline and candidate view scoring recognition algorithm; (f) PTE-YOLO and candidate view scoring recognition algorithm)

开发与系统集成测试,推动模型在产前筛查流程中的落地与临床验证。

参考文献 (References)

Carvalho J S, Axt-Flidner R, Chaoui R, Copel J A, Cuneo B F, Goff D, et al. 2023. ISUOG practice guidelines (updated): fetal cardiac screening. *Ultrasound in Obstetrics and Gynecology*, 61(6): 788-803 [DOI: 10.1002/uog.26224]

Chen H, Wu L Y, Dou Q, Qin J, Li S L, Cheng J Z, et al. 2017. Ultrasound standard plane detection using a composite neural network framework. *IEEE Transactions on Cybernetics*, 47(6): 1576-1586 [DOI: 10.1109/TCYB.2017.2685080]

Ding J, Xue N, Long Y, Xia G S and Lu Q K. 2019. Learning RoI transformer for oriented object detection in aerial images//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 2844-2853 [DOI: 10.1109/CVPR.2019.00296]

Han J M, Ding J, Xue N and Xia G S. 2021. ReDet: a rotation-equivariant detector for aerial object detection//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 2785-2794 [DOI: 10.1109/CVPR46437.2021.00281]

He J, Yang L, Liang B C, Li S L and Xu C X. 2024. Fetal cardiac ultrasound standard section detection model based on multitask learning and mixed attention mechanism. *Neurocomputing*, 579: #127443 [DOI: 10.1016/j.neucom.2024.127443]

- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hu J, Shen L, Albanie S, Sun G and Wu E. 2019. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8): 2011-2023 [DOI: 10.1109/TPAMI.2019.2913372]
- Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2261-2269 [DOI: 10.1109/CVPR.2017.243]
- Li F R, Li P, Wu X M, Zeng P, Lyu G R, Fan Y L, et al. 2024. FHUSP-NET: a multi-task model for fetal heart ultrasound standard plane recognition and key anatomical structures detection. *Computers in Biology and Medicine*, 168: #107741 [DOI: 10.1016/j.combiomed.2023.107741]
- Li W T, Chen Y J, Hu K X and Zhu J K. 2022. Oriented RepPoints for aerial object detection//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 1819-1828 [DOI: 10.1109/CVPR52688.2022.00187]
- Li Z Y, Chen Y M, Qiu L Q, Chen D Q, Hu C G, Xu J Y, et al. 2019. Prevalence, types, and malformations in congenital anomalies of the kidney and urinary tract in newborns: a retrospective hospital-based study. *Italian Journal of Pediatrics*, 45(1): #50 [DOI: 10.1186/s13052-019-0635-9]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Liu S, Qi L, Qin H F, Shi J P and Jia J Y. 2018. Path aggregation network for instance segmentation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8759-8768 [DOI: 10.1109/CVPR.2018.00913]
- Ouyang D L, He S, Zhang G Z, Luo M Z, Guo H Y, Zhan J, et al. 2023. Efficient multi-scale attention module with cross-spatial learning//Proceedings of 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island, Greece: IEEE: 1-5 [DOI: 10.1109/ICASSP49357.2023.10096516]
- Pu B, Zhu N B, Li K L and Li S L. 2021. Fetal cardiac cycle detection in multi-resource echocardiograms using hybrid classification framework. *Future Generation Computer Systems*, 115: 825-836 [DOI: 10.1016/j.future.2020.09.014]
- Qiao S Y, Chen L C and Yuille A. 2020. DetectoRS: detecting objects with recursive feature pyramid and switchable atrous convolution//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 10208-10219 [DOI: 10.1109/CVPR46437.2021.01008]
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition//Proceedings of 2015 International Conference on Learning Representations. San Diego, US: Computational and Biological Learning Society: 1-14 [DOI: 10.48550/arXiv.1409.1556]
- Tan J R, Lu X, Zhang G, Yin C Q and Li Q Q. 2021. Equalization loss v2: a new gradient balance approach for long-tailed object detection//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 1685-1694 [DOI: 10.1109/CVPR46437.2021.00173]
- Tan M X and Le Q. 2019. EfficientNet: rethinking model scaling for convolutional neural networks//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA: PMLR: 6105-6114
- Tan M X, Pang R M and Le Q V. 2020. EfficientDet: scalable and efficient object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10778-10787 [DOI: 10.1109/CVPR42600.2020.01079]
- Tian Y J, Ye Q X and Doermann D. 2025. YOLOv12: attention-centric real-time object detectors [EB/OL]. [2025-07-10]. <https://arxiv.org/pdf/2502.12524.pdf>
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]
- Xie X X, Cheng G, Wang J B, Yao X W and Han J W. 2021. Oriented R-CNN for object detection//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 3500-3509 [DOI: 10.1109/ICCV48922.2021.00350]
- Xu G Z, Qian Y F, Wang Y, Liu R, Zhou J and Lei B J. 2023. Quality assessment for fetal four-chamber ultrasound views based on two-stage segmentation. *Journal of Image and Graphics*, 28(8): 2476-2490 (徐光柱, 钱奕凡, 王阳, 刘蓉, 周军, 雷帮军. 2023. 基于两级分割的胎儿四腔心超声切面质量评测. *中国图象图形学报*, 28(8): 2476-2490) [DOI: 10.11834/jig.220347]
- Xu G Z, Wu M Q, Qian Y F, Wang Y, Liu R, Zhou J, et al. 2024. Automatic capture for standard fetal cardiac four-chamber ultrasound view by fusing frame sequential relationships. *Journal of Image and Graphics*, 29(3): 782-797 (徐光柱, 吴梦琦, 钱奕凡, 王阳, 刘蓉, 周军, 等. 2024. 融合帧间时序关系的标准胎儿四腔心超声切面自动获取. *中国图象图形学报*, 29(3): 782-797) [DOI: 10.11834/jig.230251]
- Xu X Z, Jiang Y Q, Chen W H, Huang Y L, Zhang Y and Sun X Y. 2023. DAMO-YOLO: a report on real-time object detection design [EB/OL]. [2025-07-10]. <https://arxiv.org/pdf/2211.15444.pdf>
- Yang G Y, Lei J, Zhu Z K, Cheng S Y, Feng Z L and Liang R H. 2023. AFPN: asymptotic feature pyramid network for object detection//Proceedings of 2023 IEEE International Conference on Systems,

Man, and Cybernetics. Honolulu, USA: IEEE: 2184-2189 [DOI: 10.1109/SMC53992.2023.10394415]

Zhang B, Liu H, Luo H and Li K J. 2021. Automatic quality assessment for 2D fetal sonographic standard plane based on multi-task learning. Medicine, 100 (4) : #24427 [DOI: 10.1097/MD.0000000000024427]

Zhang Q L and Yang Y B. 2021. SA-Net: shuffle attention for deep convolutional neural networks//Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). Toronto, Canada: IEEE: 2235-2239 [DOI: 10.1109/ICASSP39728.2021.9414568]

作者简介

陈宇,女,硕士研究生,主要研究方向为医学图像处理。
E-mail:6230805002@stu.jiangnan.edu.cn

朱栋晓,通信作者,男,主任医师,主要研究方向为产科超声和超声人工智能。E-mail:617785712@qq.com

宿磊,男,教授,主要研究方向为机器视觉和故障诊断。
E-mail:lei_su2015@jiangnan.edu.cn

李可,男,教授,主要研究方向为智能检测和机器视觉。
E-mail:like_jiangnan@163.com

Michael Pecht,男,教授,主要研究方向为故障诊断。
E-mail:pecht@umd.edu