❖专论

Intelligent recognition of prenatal ultrasound standard sections based on VMamba model

XU Caixu[1,2], XIE Longyang[3], HUANG Zhixi[4], WEN Huaxuan[5], AN Liting[5], WEI Junming[2], CHEN Jiaxi[2], HE Jie[1,2], HE Guannan[6*]

(1. Guangxi Key Laboratory of Machine Vision and Intelligent Control, 2. School of Electronic Information and Artificial Intelligence, Wuzhou University, Wuzhou 543002, China; 3. School of Information, Yunnan University, Kunming 650091, China; 4. Department of Neonatology, 5. Department of Ultrasound, Shenzhen Maternity & Child Healthcare Hospital, Southern Medical University, Shenzhen 518028, China; 6. Department of Ultrasound, Sichuan Provincial Maternal and Child Health Care Hospital, Chengdu 646199, China)

[**Abstract**] **Objective** To observe the value of intelligent recognition of 21 prenatal ultrasound standard sections based on VMamba model. **Methods** Totally 9 099 standard sections met requirements of prenatal ultrasound screening guidelines were retrospectively enrolled and divided into training set, test set and validation set at a ratio of 7∶1∶2. VMamba model was constructed to identify 21 standard sections of prenatal ultrasound, and its recognition performance was compared with other advanced deep learning models including ResNet-50, Swin-Transformer, ConvNeXtV2, RepViT and MSPANet. Then gradient-weighted class activation mapping (Grad-CAM) feature heat map and t-SNE classification effect diagram were combined to screen the best classification model. **Results** VMamba model had the highest sensitivity, specificity, accuracy, precision, area under the receiver operating characteristic curve (AUC) and F1 score for automatically recognizing standard sections of prenatal ultrasound, which was 94.54%, 99.72%, 94.42%, 94.57%, 0.997 and 0.944, respectively. Except for the transventricular view (TV), left ventricular outflow tract view (LVOT), ulna and radius coronal view (UR) and tibia and fibula coronal view (TF), high-response regions of the features showed in the other standard sections were highly coincident with the key anatomical structures. VMamba model exhibited strong intra-class aggregation and inter-class separability in two-dimensional space of features. **Conclusion** VMamba model performed outstandingly in intelligent recognition of 21 standard sections of prenatal ultrasound, which was potential for clinical application and transformation.

[**Keywords**] fetus; deep learning; ultrasonography, prenatal

DOI:10.13929/j.issn.1003-3289.2025.12.002

基于 VMamba 模型智能识别产前超声标准切面

许才项[1,2],谢龙洋[3],黄稚熙[4],文华轩[5],安丽婷[5],韦俊名[2],陈家希[2],贺　杰[1,2],何冠南[6*]

(1.梧州学院广西机器视觉与智能控制重点实验室,2.电子信息与人工智能学院,广西 梧州　543002;3.云南大学信息学院,云南 昆明　650091;4.南方医科大学深圳妇幼保健院新生儿科,5.超声科,广东 深圳　518028;6.四川省妇幼保健院超声科,四川 成都　646199)

[摘　要] **目的**　观察基于 VMamba 模型智能识别 21 种产前超声标准切面的价值。**方法**　回顾性纳入 9 099 幅符合产前超声筛查指南要求的标准切面,按 7∶1∶2 比例将其分为训练集、测试集与验证集。构建 VMamba 模型以识别产前超声

[基金项目] 成都医学院联合科研基金(25LHSFY3-01)、广西高校中青年教师科研基础能力提升项目(2024KY0694)。

[第一作者] 许才项(1991—),男,湖北孝感人,硕士,讲师、工程师。研究方向:超声人工智能应用及医学图像处理。E-mail: xucaixu@gxuwz.edu.cn

[通信作者] 何冠南,四川省妇幼保健院超声科,646199。E-mail: heguannan811003@126.com

[收稿日期] 2025-05-14　[修回日期] 2025-07-19

标准切面，比较其与其他先进深度学习模型，包括 ResNet-50、Swin-Transformer、ConvNeXtV2、RepViT 及 MSPANet 的效能，并结合梯度加权类别映射(Grad-CAM)特征热力图、t-SNE 分类效果图筛选效能最优模型。**结果**　VMamba 模型自动识别产前超声标准切面的敏感度、特异度、准确率、精确度、受试者工作特征曲线下面积(AUC)及 F1 分数均最高，分别为 94.54%、99.72%、94.42%、94.57%、0.997 及 0.944。除经侧脑室水平横切面(TV)、左心室流出道切面(LVOT)、尺桡骨冠状切面(UR)及胫腓骨冠状切面(TF)外，其他标准切面所示特征高响应区域均与关键解剖结构高度重合。VMamba 模型在特征二维空间中具有较强类内聚集性和类间可分性。**结论**　基于 VMamba 模型用于智能识别 21 种产前超声标准切面表现优异，具备临床转化潜力。

［**关键词**］　胎儿；深度学习；超声检查，产前

［**中图分类号**］　R714.5；R445.1　［**文献标识码**］　A　　［**文章编号**］　1003-3289(2025)12-1944-06

产前超声对于监测胎儿生长发育及筛查结构异常具有重要作用。孕中晚期是全面评估胎儿生长发育的最佳时期[1]，指南[2-3]要求此时超声筛查需至少对胎儿全身 21 个标准切面进行评估(图 1)；但完整获取 21 个切面高度依赖超声医师经验且耗时，并可能判断不准确或出现遗漏。近年人工智能在产前超声智能筛查领域取得一系列成果[4-7]，但多数研究仅针对单一或少数切面，不足以满足实际工作需求。本研究观察基于 VMamba 模型[8]智能识别 21 种产前超声标准切面的价值。

1　资料与方法

1.1　研究对象　回顾性收集 2020 年 12 月—2023 年 12 月深圳市妇幼保健院 9 099 幅符合产前超声筛查指南要求的标准切面图像，均于Ⅲ级产前超声检查时获取，受检时孕妇孕周 20～24^{+6}、平均体质量指数(body mass index，BMI)(21.5±3.0)kg/m^2；由具有 5 年以上产前超声筛查经验的超声科医师采集图像、具有 20 年以上产前超声筛查经验的超声科医师进行质控。纳入标准：图像质量良好，关键解剖结构完整、清晰可辨认，符合超声产前筛查指南要求；产后随访均为正常新生儿。按 7∶1∶2 比例划分训练集、测试集与验证集(表 1)。本研究获院医学伦理委员会批准(LLYJ2024-202-089)，孕妇及家属均知情同意。

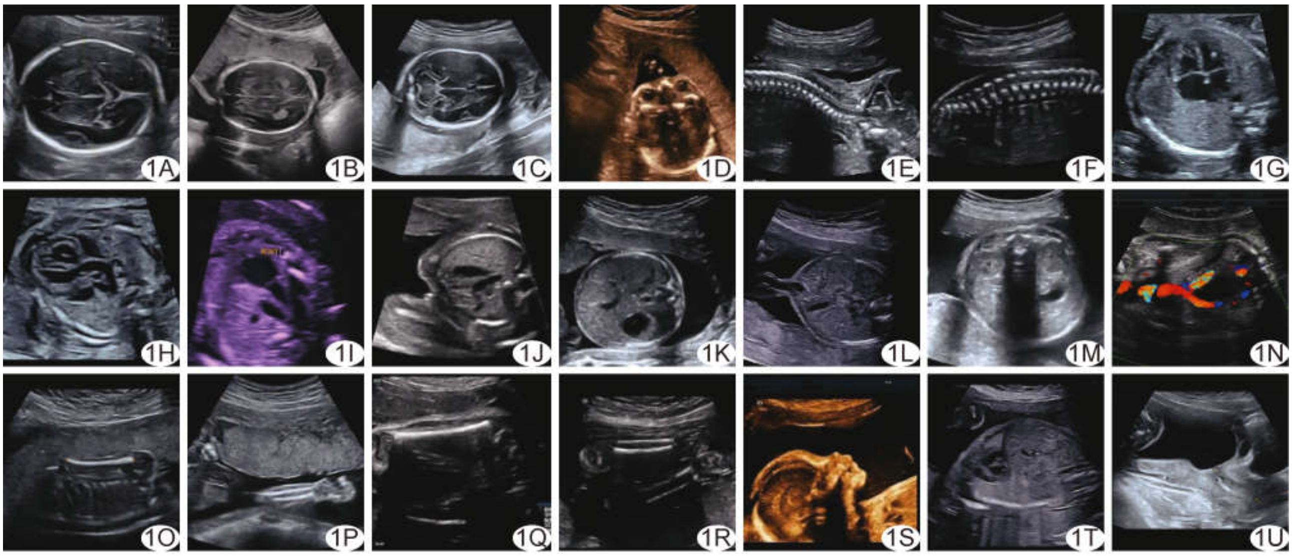


图 1　21 种产前超声标准切面　A.经丘脑水平横切面(tranthalamic view，TTV)；B.经侧脑室水平横切面(transventricular view，TV)；C.经小脑水平横切面(transcerebellar view，TCV)；D.双眼球水平横切面(eye axial view，Eyes)；E.脊柱颈胸段矢状切面(cervicothoracic spine sagittal view，CTSP)；F.脊柱腰骶尾段矢状切面(lumbosacral spine sagittal view，LSSP)；G.四腔心切面(four-chamber view，4C)；H.左心室流出道切面(left ventricular outflow tract view，LVOT)；I.右心室流出道切面(right ventricular outflow tract view，RVOT)；J.三血管气管切面(three vessels and trachea view，3VT)；K.上腹部水平横切面(upper abdomen axial view，Ab)；L.脐带腹壁插入口横切面(cord insertion abdominal axial view，CI)；M.双肾水平横切面(kidneys axial view，Kidneys)；N.膀胱水平横切面(bladder axial view，BL)；O.肱骨长轴切面(humerus long axis view，HL)；P.尺桡骨冠状切面(ulna and radius coronal view，UR)；Q.股骨长轴切面(femur long axis view，FL)；R.胫腓骨冠状切面(tibia and fibula coronal view，TF)；S.颜面部正中矢状切面(median sagittal facial profile view，MFP)；T.膈肌矢状/冠状切面(diaphragm sagittal/coronal view，DI)；U.宫颈内口矢状切面(internal cervical os sagittal view，ICO)

表 1　训练集、验证集与测试集中的标准超声切面分布(幅)

集别	TTV	TV	TCV	Eyes	CTSP	LSSP	4C	LVOT	RVOT	3VT	Ab
训练集(n=6 357)	320	319	305	298	320	303	269	279	304	300	319
测试集(n=932)	47	47	45	44	47	44	39	41	44	44	47
验证集(n=1 810)	91	91	87	85	91	86	77	79	87	85	91

集别	CI	Kidneys	BL	HL	UR	FL	TF	MFP	DI	ICO
训练集(n=6 357)	321	299	320	319	319	319	319	306	179	320
测试集(n=932)	47	44	47	47	46	47	46	45	27	47
验证集(n=1 810)	91	85	91	91	91	91	91	87	51	91

1.2　仪器与方法

1.2.1　图像预处理　对原始图像中的隐私信息进行脱敏处理。通过等比例缩放、随机裁剪、边缘填充等空间增强方法,将原始图像调至像素 224×224 并行归一化处理;利用 RandAugment 第三方库中的不同随机增强策略(色彩扰动、对比度调整、高斯模糊及随机遮挡等)处理图像。

1.2.2　构建模型　基于 Vmamba 构建自动识别产前超声标准切面的深度学习(deep learning, DL)模型。利用 Stem 模块将输入图像划分为若干图像块,生成空间维度 56×56×96 的二维特征图;以下采样和多个视觉状态空间(visual state space, VSS)模块堆叠构建 4 个维度特征提取层,其中的 VSS 模块在 Mamba 结构基础上引入 2D 选择性扫描(2D selective scan, SS2D)以实现一维序列扫描和二维平面扫描之间的衔接与统一建模;最后利用线性变换操作将高维特征向量投射至 21 种标准超声切面的概率输出空间,以实现自动识别切面。见图 2。

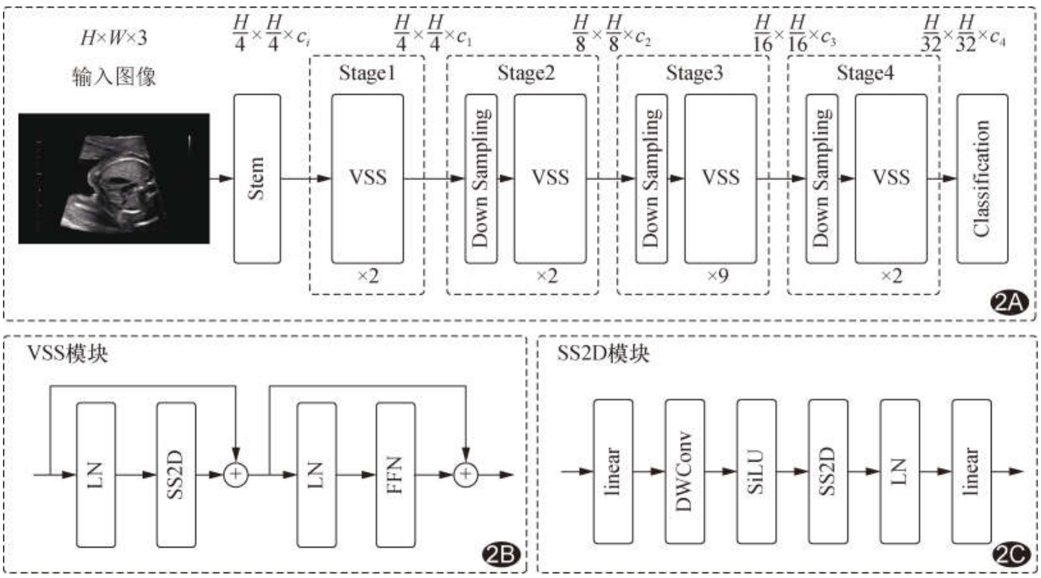


图 2　Vmamba 模型整体架构图　A. VMamba 模型架构; B. VSS 模块; C. SS2D 模块

1.2.3　训练模型及实验参数　同期建立 ResNet-50[9]、Swin-Transformer[10]、ConvNeXtV2[11]、RepViT[12] 及 MSPANet[13] 模型。上述模型及 Vmamba 模型均在 Linux 服务器[配置为 Intel(R) Xeon(R) Platinum 8375C CPU@2.90 GHz,378 GB RAM,8×NVIDIA A6000,46 068 MB]上进行训练和测试;采用交叉熵损失函数和余弦退火学习率下降调整策略,见表 2。

表 2　模型训练策略及参数

模型	优化器	初始学习率	Epoch	其他配置
VMamba	Sgd	0.1	200	DropPath+数据增强
ResNet-50	AdamW	0.000 5	200	Warmup(前 20 个 Epoch 逐步提高学习率)
Swin-Transformer	Adam	0.001	200	前 50 轮冻结主干网络训练
ConvNeXtV2	AdamW	0.000 8	600	权重衰减系数 0.05
RepViT	AdamW	0.001	300	Warmup(前 20 个 Epoch 逐步提高学习率)
MSPANet	Adam	0.001	200	前 50 轮冻结主干网络训练

1.3　统计学分析　采用 Python(3.8.20)统计分析软件,利用 scikit-learn 库评估各模型自动识别产前超声标准切面的敏感度、特异度、准确率、精确度、受试者工作特征(receiver operating characteristic, ROC)曲线下面积(area under the curve, AUC)及 F1 分数。基于 matplotlib 库绘制梯度加权类别映射(gradient-weighted class activation mapping, Grad-CAM)特征热力图、t-SNE 分类效果图,筛选性能最优的分类模型为最佳模型。

2 结果

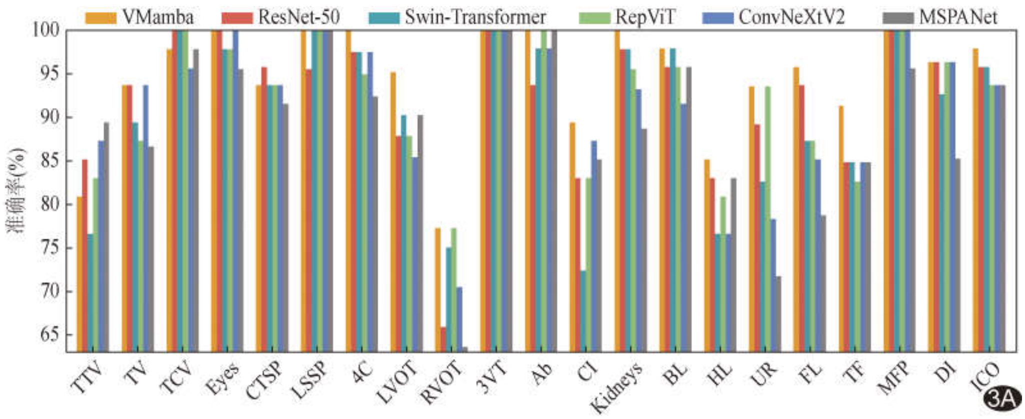
VMamba 模型自动识别产前超声标准切面模型的敏感度、特异度、准确率、精确度、AUC 及 F1 分数均最高,分别为 94.54%、99.72%、94.42%、94.57%、0.997 及 0.944(表 3);其识别 Eyes、LSSP、4C、3VT、Ab、Kidneys 及 MFP 切面的准确率均为 100%,识别 TCV、LOVT、BL、FL、DI 及 ICO 切面的准确率为 95%~99%,识别 TV、CTSP、UR、TF 切面的准确率为 90%~95%,识别 TTV、ROVT、CI 及 HL 切面识别的准确率为 70%~89%(图 3A);识别 TCV、Eyes、3VT、Ab 及 MFP 切面的 AUC 均达 1.000,识别其他 16 个切面的 AUC 为 0.996~0.999(图 3B)。

Grad-CAM 自动生成的 VMamba 模型特征关注度热力图可视化结果显示,除 TV、LVOT、UR、TF 切面中模型特征关注区域受背景或其他因素干扰外,其他标准切面所示特征高响应区域与关键解剖结构高度重合,注意力分布与超声医师的实际判读焦点基本一致,表明模型决策依据具有较好的解剖学合理性与可解释性;但在 TV、LVOT 及 UR 切面对关键解剖结构的关注欠缺,TF 切面热力图高度关注区域未覆盖主要关键解剖结构,导致上述 4 种切面判别准确率较低。见图 4。

t-SNE 可视化结果显示,VMamba 模型在特征二维空间中具有较强类内聚集性和类间可分性,类别间边界区分十分明显,样本之间重叠少;而 ResNet-50、Swin-Transformer、ConvNeXtV2、RepViT 及 MSPANet 模型在多数类别特征中表现出一定可分性,但特定类别间存在显著重叠。见图 5。

表 3 各模型自动识别产前超声标准切面的效能

模型	敏感度(%)	特异度(%)	准确率(%)	精确度(%)	AUC	F1 分数
VMamba	94.54	99.72	94.42	94.57	0.997	0.944
ResNet-50	92.08	99.60	91.95	91.92	0.997	0.918
Swin-Transformer	90.73	99.53	90.56	90.79	0.997	0.906
ConvNeXtV2	91.98	99.63	90.74	92.05	0.992	0.917
RepViT	91.79	99.58	91.63	91.89	0.989	0.917
MSPANet	88.52	99.43	88.52	89.12	0.996	0.885



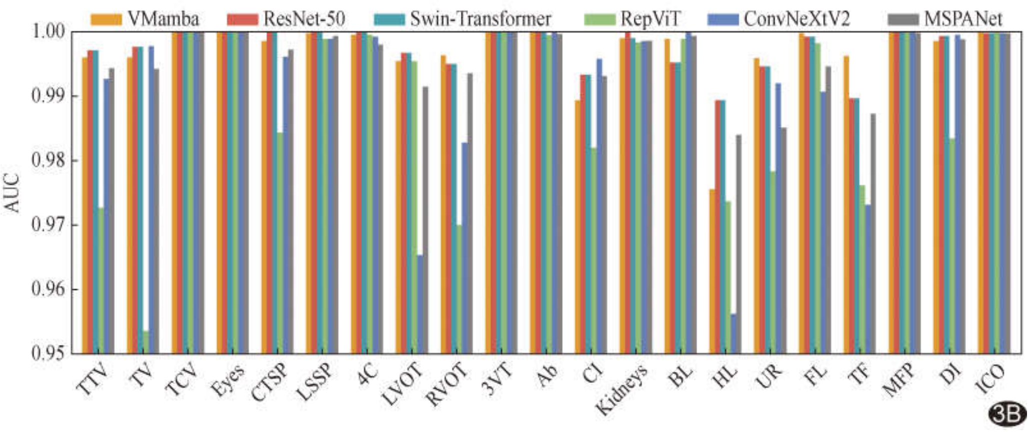


图 3 各模型自动识别 21 种产前超声标准切面的效能 A.准确率;B.AUC

3 讨论

本研究收集Ⅲ级产前超声检查中的 21 种标准超声切面,用于训练和测试 VMamba 模型自动识别 21 种产前超声标准切面的性能,结果显示该模型的敏感度、特异度、准确率、精确度、AUC 及 F1 分数均最高,分别为 94.54%、99.72%、94.42%、94.57%、0.997 及 0.944,尤其准确率及精确度相较其他模型高出 2%以上[14-15];分析原因,主要在于其在学习特征时获得了更多有效的长距离依赖信息,对于识别复杂超声切面任务表现出更强的阳性检出能力与整体判别能力,能有效减少漏检和误检风险,整体性能更为均衡、稳定,可在一定程度上缓解临床获取切面高度依赖操作者经验、识别耗时且可重复性不佳等问题。

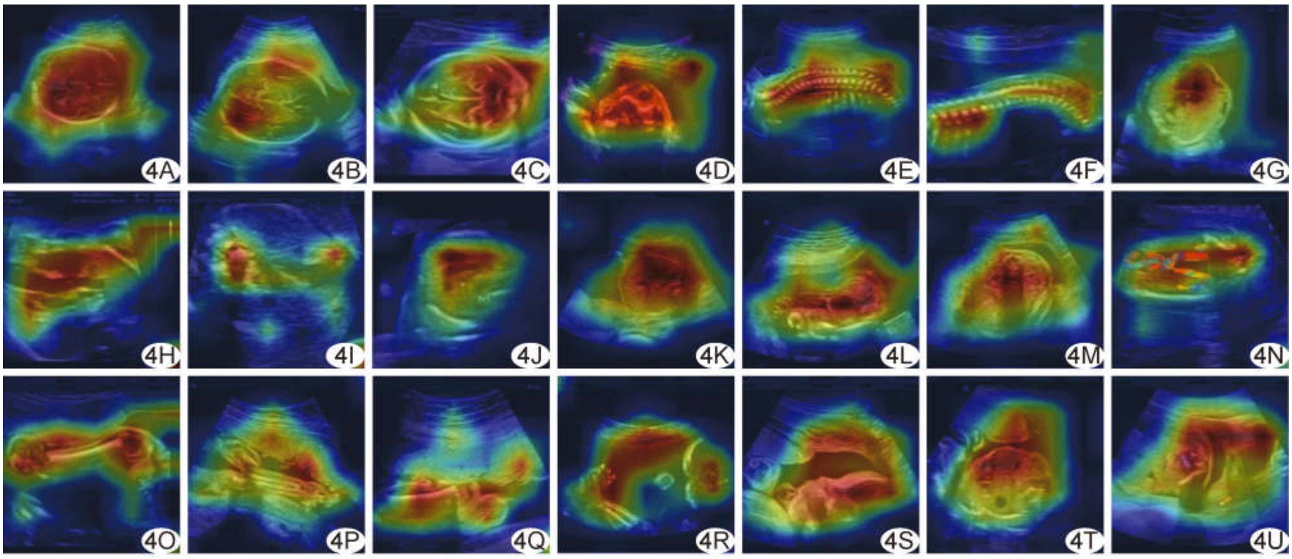


图 4　VMamba 模型学习 21 种产前超声标准切面后关注区域热力图　A. TTV；B. TV；C. TCV；D. Eyes；E. CTSP；F. LSSP；G. 4C；H. LVOT；I. RVOT；J. 3VT；K. Ab；L. CI；M. Kidneys；N. BL；O. HL；P. UR；Q. FL；R. TF；S. MFP；T. DI；U. ICO　（红色区域为最关注区域，黄色区域次之，蓝色区域为关注度最低区域）

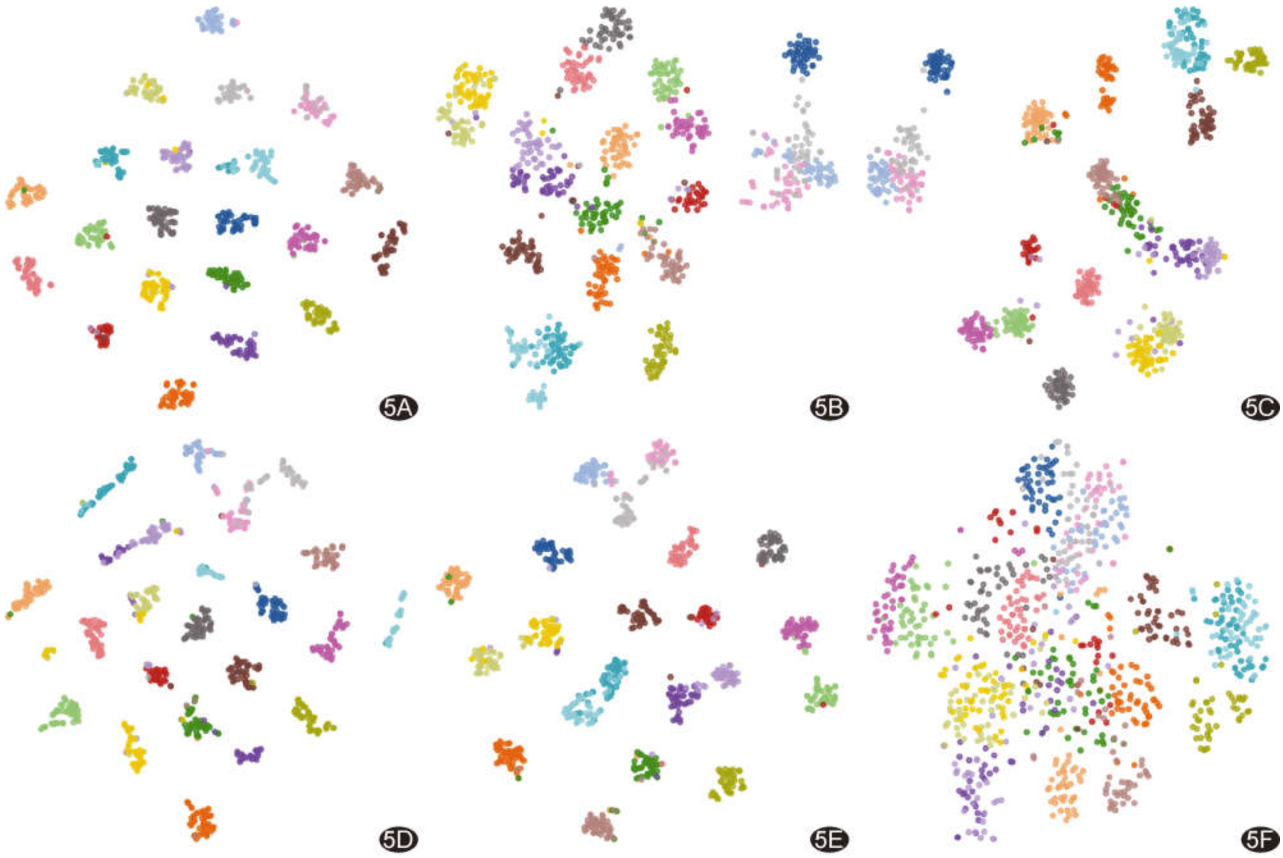


图 5　各模型学习特征的 t-SNE 可视化效果　A. VMamba 模型；B. ResNet-50 模型；C. Swin-Transformer 模型；D. ConvNeXtV2 模型；E. RepViT模型；F. MSPANet 模型

本研究所涉 6 种模型识别 HL、Kidneys、LVOT、RVOT、TF、TTV、TV 和 UR 等切面表现均存在一定波动，而识别 3VT、4C、Ab、LSSP、MPF 和 TCV 等结构特征清晰切面则具有高度一致性与稳定性，提示结构特征清晰、解剖标志稳定、图像特异性强的切面更利于 DL 模型进行高效、准确的学习。总体而言，

VMamba 模型识别全部 21 种产前超声切面性能最为均衡(AUC 为 0.996～1.000),具有强大的长距离建模和适应多类别超声切面分类任务能力。本组 DI 和 Eyes 切面数量相对较少,但 VMamba 模型预测精度仍达 96.3%和 100%,进一步证实该模型针对结构清晰、特征显著的超声切面具备良好鲁棒性与小样本学习能力。RVOT 与 LVOT 切面结构形态高度相似且图像上下文信息有限,导致各模型均难以提取有效的鉴别特征并捕捉关键空间动态特征,值得重点关注。此外,本研究所构建的模型均采用将自然图像领域学习的特征迁移至超声图像分类的预训练策略,在一定程度上提升了模型分类性能、减少了样本稀缺所致模型过拟合风险。t-SNE 结果显示,VMamba 模型在特征二维空间中具有较强类内聚集性和类间可分性,类别间边界区分十分明显,样本之间重叠少,表明 VMamba 模型对类别的判别能力与特征提取能力良好;而其他模型在多数类别特征中表现出一定可分性,但在特定类别间存在显著重叠,提示其处理结构相似或判别难度较高切面时存在局限性。

综上,基于 VMamba 模型可智能识别 21 种产前超声标准切面,具备临床转化潜力。本研究的主要局限性:①数据来源于单中心,多样性不足,可致模型泛化能力受限;②主要纳入指南规定的 21 种标准超声切面,未来将纳入非标准切面,进一步完善标准超声切面的自动质量控制;③VMamba 模型区分 RVOT 与 LVOT 存在一定困难,未来将引入解剖学先验知识等方法,提升模型对切面的判别能力与适应性。

利益冲突:全体作者声明无利益冲突。

作者贡献:许才项研究设计和实施、查阅文献、图像处理、数据分析、撰写文章;谢龙洋研究实施、图像处理和分析、统计分析;黄稚熙数据分析、图像处理和分析;文华轩指导、图像处理、修改文章;安丽婷查阅文献、图像和统计分析;韦俊名研究设计和实施、数据分析、图像处理;陈家希研究设计、数据分析;贺杰指导、数据分析、审阅文章、经费支持;何冠南研究设计和实施、指导、修改和审阅文章、经费支持。

[参考文献]

[1] 罗丹丹,文华轩,彭桂艳,等.新型智能产前超声技术在中晚孕期产前超声检查中的应用价值[J/CD].中华医学超声杂志(电子版),2020,17(11):1061-1069.

[2] 中华医学会超声医学分会妇产超声学组,国家卫生健康委妇幼司全国产前诊断专家组医学影像组.超声产前筛查指南[J].中华超声影像学杂志,2022,31(1):1-12.

[3] SALOMON L J, ALFIREVIC Z, BERGHELLA V, et al. ISUOG practice guidelines (updated): Performance of the routine mid-trimester fetal ultrasound scan [J]. Ultrasound Obstet Gynecol, 2022,59(6):840-856.

[4] PU B, LV X, YANG J, et al. Leveraging anatomical consistency for multi-object detection in ultrasound images via source-free unsupervised domain adaptation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Pennsylvania: AAAI,2025, 39(6):6532-6540.

[5] HE J, YANG L, LIANG B, et al. Fetal cardiac ultrasound standard section detection model based on multitask learning and mixed attention mechanism [J]. Neurocomputing, 2024, 579:127443.

[6] LIANG B, PENG F, LUO D, et al. Automatic segmentation of 15 critical anatomical labels and measurements of cardiac axis and cardiothoracic ratio in fetal four chambers using nnU-NetV2[J]. BMC Med Inform Decis Mak, 2024,24(1):128.

[7] PU B, WANG L, YANG J, et al. M3-uda: A new benchmark for unsupervised domain adaptive fetal cardiac structure detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Tennessee: IEEE, 2024:11621-11630.

[8] LIU Y, TIAN Y, ZHAO Y, et al. Vmamba: Visual state space model[C]. Neural Information Processing Systems. Tennessee: IEEE, 2024:103031-103063.

[9] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Tennessee: IEEE, 2016:770-778.

[10] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Tennessee: IEEE, 2021:10012-10022.

[11] WOO S, DEBNATH S, HU R, et al. Convnext v2: Co-designing and scaling convnets with masked autoencoders[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 16133-16142.

[12] WANG A, CHEN H, LIN Z, et al. Repvit: Revisiting mobile cnn from vit perspective[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Tennessee: IEEE, 2024:15909-15920.

[13] YU Y, ZHANG Y, CHENG Z, et al. Multi-scale spatial pyramid attention mechanism for image recognition: An effective approach[J]. Engineering Applications of Artificial Intelligence, 2024,133:10826.

[14] 宋书豪,曾施.应用深度学习模型分类正常胎儿心脏超声切面[J].中国医学影像技术,2025,41(1):70-73.

[15] 何冠南,李胜利.人工智能在胎儿超声心动图应用中的研究进展[J/OL].中华医学超声杂志(电子版),2023,20(5):554-556.