

基于改进 VMamba-DFF-YOLOv10 的乳腺癌图像识别

惠雨晨¹ 戴道成¹ 王富强¹ 胡少青¹

HUI Yuchen DAI Daocheng WANG Fuqiang HU Shaoqing

摘要

多年来,利用深度学习进行早期乳腺癌诊断可以提高诊断效果。然而,传统的诊断方法有其局限性,如复杂的特征提取和多阶段分类过程。这些过程不仅增加了计算的复杂性,也影响了分类的效率和准确性,依赖单一成像模式有可能漏诊或误诊。考虑到 VMamba 在乳腺肿瘤图像识别的效率-性能权衡中经常出现的不足,以及 Transformers 固有的二次计算复杂性,文章提出了 VMamba-DFF-YOLOv10 模型。该模型集成了 VMamba 和 YOLO 技术的原理,并用空间通道解耦下采样取代了 VMamba 中的下采样组件,以促进乳腺癌图像识别和 FLOPS 的高效协同。此外,引入双相特征融合(dual-phase feature fusion)技术,结合两种成像模式(乳腺 X 射线照相术和核磁共振成像)进行分类,通过整合来自不同模式的信息,显著提高了分类准确性。实验结果表明,与其他模型相比,VMamba-DFF-YOLOv10 模型取得了更好的结果。例如,该方法的准确率、AP、AUC 和 Kappa 分别达到了 83.34%、93.79%、0.95 和 86.1%,与其他模型相比,FLOP 更少,准确率更高。

关键词

乳腺癌检测; VMamba; YOLOv10; 多模态成像融合

doi: 10.3969/j.issn.1672-9528.2025.04.009

0 引言

乳腺癌是全球女性人群中最常见的肿瘤疾病,也是对女性打击最大的癌症之一^[1]。乳房 X 线照相术是检测任何异常的重要手段之一。正确判读乳房 X 光照片对诊断非常重要,但同时也非常耗时和耗费技术。经验不足的放射科医生会因为处理的病例数量过多而无法维持许多诊断^[2]。目前在识别方面存在的主要问题是假阳性和假阴性,而减少这些错误的方法对于提高诊断准确性和病人护理非常重要^[3]。计算机辅助诊断系统可以从图像中识别和定位恶性,并将其分为恶性、良性或正常^[4]。研究人员在开发用于乳腺癌检测的 CAD 模型时面临着一些挑战。Faster R-CNN、RetinaNet、ViT 和 YOLO 鼓励开发用于乳腺癌检测的系统^[5-8]。这些框架无疑带来了三大难题:

- (1) 高参数和计算负荷:用于乳腺癌检测的人工智能和深度学习模型的部署往往涉及大量参数和高计算需求。
- (2) 精度和速度:目前的模型往往不够精确,因此如何实现乳腺癌检测的高精度仍然是一大挑战。
- (3) 实时检测和鲁棒性:实时检测能力,如识别微小物体(如微小肿瘤)和处理多个物体类别的能力,对于临床应用至关重要。

基于此,学者们提出以 YOLOv10 为基础来提高 YOLO 系列性能,用创新的框架来应对图像检测领域的各种挑战,该模型在减少延迟、提高整体效率和准确性方面表现出色。采用多任务学习和复杂的注意力机制来提高性能和效率,其可扩展性可与 YOLO 框架无缝集成。然而,Mamba 对大型数据集的依赖可能会在少数数据条件下受到限制。尽管存在这些限制,Mamba 在医疗成像和监控等领域仍大有可为,因为这些领域依赖于强大的特征提取和多模态集成。令人振奋的是,研究人员已成功地将 YOLOv10 架构引入视觉领域,并在图像分类方面取得了成功^[9]。受此启发,提出以下问题:

- (1) 能否将 Mamba 结构引入物体检测领域,并与当前的实时检测器相结合,从而受益于 Mamba 的优势,并为 YOLO 系列带来新的性能提升。
- (2) 能否将 DFF 与 Mamba 结构相结合,从而为 YOLO 系列带来新的性能提升?将 DFF 模型集成到上述结构中是否可行?

本文提出了一种名为 VMamba-DFF-YOLOv10 的检测器模型,将 VMamba、DFF(双相位特征融合)和 YOLOv10 的优势协同结合起来的新方法,可实现面部表情图像的高效检测和分类。这项工作的主要贡献概述如下:

- (1) 提出基于 SSM 的 Mamba-YOLOv10,为物体检测

1. 西安明德理工学院 陕西西安 710124

中的 YOLO 建立了新的基准, 并为未来基于 SSM 开发更高效的检测器奠定了坚实的基础。

(2) 设计的模型集成 VMamba 和 YOLO 技术的原理, 并用空间通道解耦下采样取代 VMamba 中的下采样组件, 促进了乳腺癌图像识别和 FLOPS 的高效协同。

(3) 引入双相特征融合技术 (dual-phase feature fusion, DFF), 结合了两种成像模式 (乳腺放射摄影和核磁共振成像) 进行分类, 并整合了来自不同模式的信息, 显著提高了分类准确性。

1 方法

二维选择性扫描机制。Mamba 最初是为为一维语言序列设计的, 由于二维视觉数据与一维序列之间固有的不兼容性, 将其直接应用于视觉任务是不切实际的。为克服这一难题, 引入了二维选择性扫描 (SS2D) 机制。如图 1 所示, SS2D 数据处理包括 3 个主要步骤: 交叉扫描、使用 S6 块进行选择性扫描和交叉合并。起初, 输入数据被分割成若干块, 然后沿着 4 条不同的遍历路径展开成序列 (交叉扫描)。使用不同的 S6 块并行处理每个片段序列。处理完成后, 对序列进行重塑和合并 (Cross-Merge), 形成最终输出地图。通过利用互补的一维遍历路径, SS2D 可确保图像中的每个像素都能有效地从不同方向的所有其他像素收集信息。这种方法有利于在二维空间内建立全局感受野。最后, 汇总的特征序列被用于重建二维特征图。

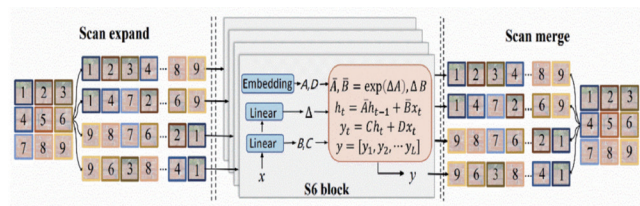


图 1 二维选择性扫描 (SS2D) 示意图

输入斑块沿着 4 条不同的扫描路径 (交叉扫描) 遍历, 每个序列由不同的 S6 块独立处理。随后, 将结果进行合并, 构建一个二维特征图作为最终输出 (扫描合并)。

双相特征融合 (DFF) 双相特征融合体系由浅层与深层融合模型共同构建, 浅层模型聚焦细粒度特征的捕获, 保留原始数据的细节信息, 深层模型则致力于抽象语义与复杂关系的挖掘, 二者有机结合, 实现特征从细节到高层语义的跨层次融合, 强化整体特征表征的丰富性与有效性。

浅层融合模型 (shallow fuse module, SFM)。受先前工作的启发, 对来自不同模态的特征 F_1 和 F_2 采用了通道交换方法。这一过程会重复多次, 使特定模态的特征逐步融合来自其他模态的信息。如图 2 所示, 一种模式的特征被整合到另一种模式的特征中。最后, 这些处理过的特征经过融合

(FUSE) 运算生成浅层融合特征。这种浅层融合方法可确保不同模态的综合特征得到有效结合, 从而在不产生大量计算开销的情况下增强整体特征表示。

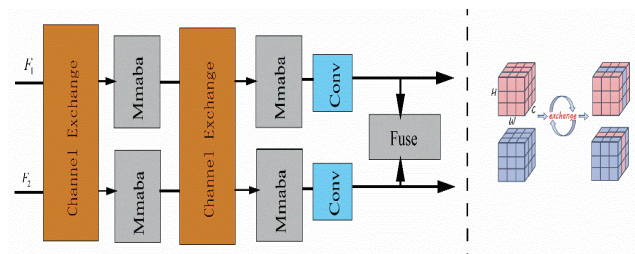


图 2 浅层融合模块的实施细节

深层融合模型 (deep fuse module, DFM)。目前 Mamba 架构由于缺乏交叉关注机制, 无法直接处理多模态图像数据。为了解决这个问题, 引入了多模态 Mamba (M3) 模块。该模块使用特定模态特征来指导融合特征的创建, 整合不同模态的详细特征。每个分支通过一系列操作处理来自不同模态的特征: 层归一化、卷积、SiLU 激活和参数离散化。然后, 这些分支通过选择性扫描机制 (SSM) 产生输出, 表示为 F_t 。使用一个门控因子来调节该网络的输出, 然后将其添加到原始分支的输出中, 以产生两个分支的最终特征组合。

2 VMamba-YOLOv10

2.1 架构

完整架构如图 3 所示, 以多模态图像作为输入。在这项工作中, 假设输入图像是乳房 X 射线照相术和核磁共振成像。

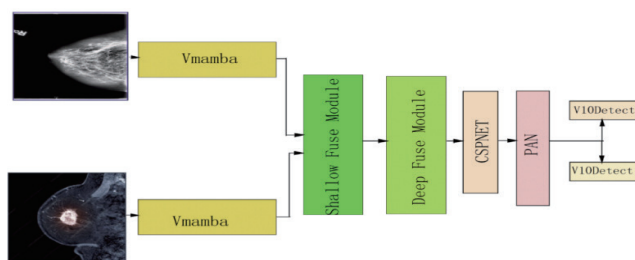


图 3 VMamba-DFF-YOLOv10 的整体架构。

VMamba 模型采用四向扫描算法, 从图像的角度挑选所有特定细节, 以获取最佳关键信息。如图 3 所示, 通道交换方法被应用于 F_1 和 F_2 , 无需额外的参数或计算, 从而实现跨多种模式的轻量级特征交换。然后通过各自的 Mamba 模块处理这些交换的特征。通过重复这些步骤, 特定模态特征 F_1 和 F_2 将整合来自其他模态的特征。最后, 执行 FUSE 操作, 生成浅层融合特征 F 。随后, 深度融合使用特定模态特征来指导融合特征的生成, 目的是整合来自不同模态的局部细节特征。初始输入是从浅层融合模块获得的融合特征。此外, 还引入了两个额外的分支, 分别接收来自不同模态的特征。

这些分支经过层归一化、卷积、SiLU 激活和参数离散化后，再通过 SSM 产生输出。在经过门控因子调制后，该输出与原始分支输出相结合，产生最终的融合特征。

2.2 Mamba-YOLOv10

为保持计算效率的同时提高模型整体性能，在传统的 VMamba 中，使用 SCDwn（空间信道去耦下采样）替代传统下采样方法。传统下采样方法，如最大池化、平均池化以及卷积下采样等，往往会造成空间细节与信息的大量丢失。而 SCDwn 通过对空间和信道维度进行解耦，有效规避了这一问题，能够切实保留关键特征信息。具体而言，SCDwn 将空间数据与信道数据分开处理，可更为精细地还原特征，进而显著降低信息损失。这一特性对提升模型检测的准确性与鲁棒性大有裨益。SCDwn 不仅能最大程度减少数据丢失，还可留存更多上下文信息，最终增强模型在物体检测任务中的有效性。从增强型 Fusion - VMamba 中提取的融合特征会被输入至 YOLOv10。然后模型的骨干、PAN（路径聚合网络）以及双标签分配头会输出边界框、类别概率和置信度分数。

3 实验

在实验中，用两个数据集来验证融合 -VMamba-YOLO 模型。使用多模态图像作为输入数据，特别是用于乳房 X 线照片的 CBIS-DDSM 和用于核磁共振成像的 Duke-Breast-Cancer-MRI。测试时，如图 4 所示，使用了由 307 对数据和 307 对来自杜克大学乳腺癌 -MRI 的数据组成的 MSRS 集，以全面验证融合性能。此外，还采用了超采样策略来明确解决类不平衡问题。这包括创建良性细胞图像的合成样本，以匹配数据集中恶性细胞图像的数量。通过平衡两类样本的数量，旨在为模型提供更均衡的训练数据集，确保模型在训练阶段能充分接触良性细胞图像。

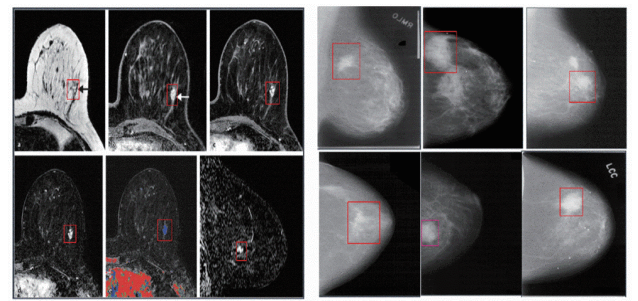


图 4 分别在 Duke-Breast-Cancer-MRI、CBIS-DDSM 数据集上预测边界框

表 1 和表 2 以融合数据集为基准，比较了文中提出的网络模型和当前最先进的方法。表中最佳性能结果以粗体标出。从表 1 和表 2 提供的数据可以得出结论，VMamba-DFF-YOLOv10 模型在 TOP-1 和 mAP50 中都取得了最高分，分

别为 83.34% 和 96.2%。这一成绩比目前领先的 YOLOv10 模型分别高出 1.3% 和 0.64%。此外，该模型还获得了最高的 KAPPA 和 AUC 分数，分别为 86.1% 和 0.95%，进一步超越了 YOLOv10 模型。

表 1 不同方法在融合数据集上的性能比较

Method	AP ^{val} /%		AUC		Acc.		KAPPA/%	
	良性	恶性	良性	恶性	良性	恶性	良性	恶性
YOLOv8	92.57	91.80	0.90	0.86	81.5	78.28	85.44	83.58
YOLOv10	95.21	93.29	0.95	0.92	83.5	80.58	86.98	84.90
Ours	95.43	92.15	0.96	0.94	84.06	82.63	86.91	85.28

表 2 所提出的方法在 GFLOPs、参数和 TOP-1 准确度之间实现了最佳平衡

Method	GFLOPs	Parameters/10 ⁶	TOP-1/%
EfficientNet-B0	0.395	5.8	78.5
Swin-T	4.6	29	80.7
GhostNet	8.1	20.1	78.8
ResNet	12.1	25.8	80.9
YOLOv8	7.3	6.6	91.1
YOLOv10	4.8	2.7	89.8
Ours	4.9	4.6	91.7

值得注意的是，VMamba-DFF-YOLOv10 模型在检测“良性”类别方面表现出色，AP 达到 95.43%。然而，其“恶性”类别的 AP 分数比“良性”类别低 3.28%。造成这种差异的主要原因是乳腺图像中恶性的形状不规则，边缘模糊，因此检测和分类更具挑战性。相比之下，良性通常表现为高对比度的小点或线状结构更加均匀、更容易检测和分类。

因此，VMamba-DFF-YOLOv10 模型有几个方面可以在未来加以改进。该模型在识别“恶性”方面的表现相对较弱，AP 分数较低，这突出表明在未来的模型迭代中需要进一步调整或特别关注这一类别。

与其他模型相比，本文模型以更少的 GFLOPs 显示出更优越的性能。与表现最好的 YOLOv10 模型相比，显著提高了 TOP-1 准确率 3.4%~2.0%。此外，与精度相似的基准模型 YOLOv8 相比，ParameterNet 减少了 45.5% 的参数和 50% 的 GFLOPs。此外，ParameterNet-600M 的 TOP-1 准确率达到 91.7%，GFLOPs 和参数均低于 ResNet 或 GhostNet。结果表明，本文提出的模型大大改进了现有的最先进方法。

这项研究的结果表明，整合不同的成像模式可显著提高性能。这种方法优于使用单一成像模式的方法，因为不同通路和单一通路之间的联系使模型能够学习到各种模式更复杂的信息，并获得准确的结果。它可以从不同角度提取特征，

获得互补信息。图 5 中的实验结果表明,使用多模态融合技术的拟议模型在 46 个 epochs 时的准确率为 93.25%,而仅使用核磁共振成像技术的准确率为 83.89%,仅使用乳腺 X 射线成像技术的准确率为 78.9%。

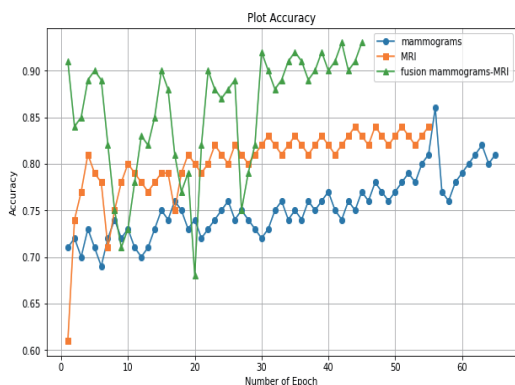


图 5 三类数据集的性能比较

4 结论

本文旨在解决传统 VMamba 模型存在的高计算复杂性和可扩展性难题,同时提高 YOLOv10 模型的性能。探讨了 VMamba 在多模态图像融合中的潜力,并提出了一种高效、有效的基于 Mamba 的双阶段模型 VMamba-DFF。一方面,采用 SCDwn(空间通道去耦下采样)的特征提取模块增强了提取特定模态特征的能力;另一方面,融合模块促进了综合和互补模态融合特征的合成。此外,本文还介绍了 VMamba-DFF-YOLOv10 网络模型,有效地结合了深度学习的高效特征提取能力和实时端到端模型的综合效率-精度驱动能力。在 CBIS-DDSM 和 Duke-Breast-Cancer-MRI 多模态数据集上的实验结果表明,所提出的 VMamba-DFF-YOLOv10 模型在乳腺癌图像诊断中达到了最先进的性能。然而,该模型可能无法达到最高准确度,尤其是在处理大尺寸、高分辨率图像时,因为追求更快的性能很可能会导致取舍。VMamba-DFF-YOLOv10 模型能有效处理各种形式的复杂纹理和噪声图像,在乳腺癌图像检测中实现了检测准确性和计算复杂性的双提升,适用于临床诊断目的,更重要的是,在未来的研究中可以整合多模态数据,将其应用于其他癌症的诊断。

参考文献:

[1] POONIA A, SHARMA V K, SINGH H K, et al. Breast cancer detection: challenges and future research developments[C/OL]. // 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS). Piscataway: IEEE, 2024[2024-10-02]. <https://ieeexplore.ieee.org/document/10498444>. DOI: 10.1109/

ICICACS60521.2024.10498444.

- [2] SATHISH K P J, SHIBU S, MOHAN M, et al. Hybrid deep learning enabled breast cancer detection using mammogram images[J]. Biomedical signal processing and control, 2024, 95:106310.
- [3] EKPO E U, VALAKHRAS M, BRENNAN P. Errors in mammography cannot be solved through technology alone[J]. Asian pacific journal of cancer prevention, 2018, 19(2):291-301.
- [4] 付荣, 丁海滨. 基于混合机器学习模型的乳腺癌预测[J]. 信息技术与信息化, 2023(1):195-198.
- [5] KULANDAIVELU G, TALUJA A, GAWAS M, et al. Automated breast cancer diagnosis optimized with higher-order attribute-enhancing heterogeneous graph neural networks using mammogram images[J]. Biomedical signal processing and control, 2024, 97:106659.
- [6] ALY G H, MAREY M, EL-SAYED S A, et al. YOLO based breast masses detection and classification in full-field digital mammograms[J]. Computer methods and programs in biomedicine, 2021, 200:105823.
- [7] BACCOUCHE A, GARCIA-ZAPIRAIN B, OLEA C C, et al. Breast lesions detection and classification via YOLO-based fusion models[J]. Computers, materials and continua, 2021, 69(1):1407-1425.
- [8] HASSAN N M, HAMAD S, MAHAR K. YOLO-based CAD framework with ViT transformer for breast mass detection and classification in CESM and FFDM images[J]. Neural computing and applications, 2024, 36:6467-6496.
- [9] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection[DB/OL]. (2024-10-30)[2024-11-04]. <https://doi.org/10.48550/arXiv.2405.14458>.

【作者简介】

惠雨晨(1992—), 男, 陕西西安人, 硕士, 助教、专任教师, 研究方向: 智能图像识别、研究各类人工智能算法和 AI 大模型。

戴道成(1995—), 男, 陕西西安人, 硕士研究生, 讲师、专任教师, 研究方向: 数据挖掘、机器学习。

王富强(1994—), 男, 甘肃庆阳人, 硕士研究生, 讲师, 研究方向: 图像识别。

胡少青(2003—), 男, 甘肃天水人, 本科, 研究方向: 图像处理。

(收稿日期: 2024-11-28)