

# 融合 Resnet 与 Transformer 的乳腺癌超声图像分类研究

黄文剑<sup>1</sup> 赵思朝<sup>2</sup> 邱学军<sup>2\*</sup>

1. 华南理工大学附属第六医院 广东佛山 528200

2. 广东药科大学 广东广州 510000

**摘要:** 目的 乳腺癌是女性中最常见的恶性肿瘤之一, 其早期诊断对提高治愈率至关重要。本文提出一种融合 ResNet 和 Transformer 的乳腺癌超声图像分类方法, 旨在结合卷积神经网络 (CNN) 的局部特征提取能力和 Transformer 的全局特征建模能力, 提升乳腺癌超声图像的分类性能。方法 通过 ResNet 网络提取图像局部特征, 并利用 Transformer 结构对这些特征进行全局建模, 进一步增强特征的表达能力。实验在 Kaggle 乳腺癌超声图像数据集上进行。结果 所提方法在良恶性及正常三分类任务中取得了优异的性能, 分类准确率、精确率、召回率分别达到 93.2%、92.1% 及 92.6%, 显著优于单一模型方法。结论 本文研究为乳腺癌超声图像的自动化诊断提供了一种有效的解决方案。

**关键词:** 乳腺癌; 超声图像分类; ResNet; Transformer; 深度学习

乳腺癌是全球女性中最常见的恶性肿瘤之一, 其发病率和死亡率均居高不下。早期诊断对于提高乳腺癌患者的治愈率和生存率至关重要。近年来, 随着医学影像技术的发展, 超声成像和病理图像分析已成为乳腺癌诊断的重要手段。超声图像具有无创性、便捷性和低成本等优点, 而病理图像则提供了细胞和组织水平的详细信息。因此, 基于超声图像和病理图像的乳腺癌诊断方法受到了广泛关注。

随着深度学习技术的兴起, 其在医学图像分析领域展现了巨大的潜力。Hou<sup>[1]</sup> 提出了一种基于深度学习的乳腺癌病理图像分类方法, 通过卷积神经网络提取图像特征, 实现了较高的分类准确率。Chen<sup>[2]</sup> 结合深度卷积神经网络和浅层局部二进制模式特征, 提出了一种用于乳腺肿瘤分类的方法, 进一步提升了分类性能。Du<sup>[3]</sup> 利用卷积神经网络对超声图像进行分析, 成功实现了乳腺癌的诊断。Xie<sup>[4]</sup> 提出了一种双采样卷积神经网络的新方法, 用于乳腺肿瘤超声图像的分类, 取得了显著的效果。Wang<sup>[5]</sup> 基于深度迁移学习提出了一种乳腺癌图像分类方法, 展示了其在小样本数据集上的有效性。

在国内, 张建新等<sup>[6]</sup> 提出了一种混合监督学习方法, 用于乳腺癌全切片病理图像的分类, 通过结合有监督和无监督学习的优势, 提高了分类精度。庄建军等<sup>[7]</sup> 提出了一种多尺度特征融合的改进残差网络, 用于乳腺癌病理图像分类, 显著提升了模型的特征提取能力。禰浚波等<sup>[8]</sup> 结合

Swin Transformer 和卷积注意力机制, 提出了一种新的乳腺癌病理图像诊断方法, 展示了 Transformer 在医学图像分析中的潜力。杨芳等<sup>[9]</sup> 通过多层级深度特征融合, 进一步提升了乳腺癌病理图像分类的性能。

此外, GoogLeNet 作为一种经典的深度学习架构, 已被成功应用于乳腺癌超声图像分类。韩杨等<sup>[10]</sup> 基于 GoogLeNet 提出了一种高效的分类方法, 验证了其在乳腺癌超声图像分析中的有效性。方于华等<sup>[11]</sup> 提出了一种融合多尺度特征和注意力机制的 MFDC-Net, 用于乳腺癌病理图像分类, 进一步提升了模型的鲁棒性。龚安等<sup>[12]</sup> 和张庙林等<sup>[13]</sup> 分别基于卷积神经网络提出了新的分类方法, 展示了 CNN 在乳腺癌病理图像分析中的广泛应用。胡天寒等<sup>[14]</sup> 提出了一种基于 SE-DenseNet 的分类方法, 进一步优化了模型的特征提取能力。

近年来, Transformer 架构在计算机视觉领域的应用逐渐受到关注。王剑等<sup>[15]</sup> 提出了一种基于 VGG-19 和 Vision Transformer 的乳腺超声图像分类方法, 展示卷积神经网络和 Transformer 融合模型在医学图像分类中的潜力。

尽管已有研究在乳腺癌超声图像分类方面取得了显著进展, 但现有方法大多基于单一的卷积神经网络架构, 难以充分利用图像的全局特征和局部特征。因此, 本研究提出了一种融合 ResNet 和 Transformer 的乳腺癌超声图像分类方法, 旨在结合 ResNet 的局部特征提取能力和 Transformer 的全局

特征建模能力, 进一步提升分类性能。通过实验验证, 所提出的方法在准确率、精确率和召回率等指标上均优于现有的主流方法, 为乳腺癌超声图像的自动化诊断提供了新的解决方案。

## 1. 资料与方法

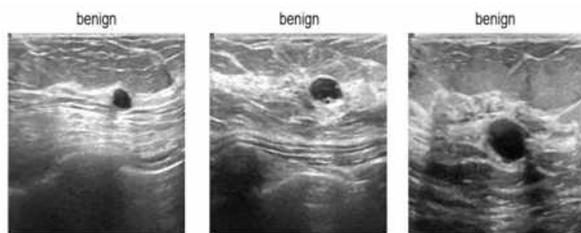
### 1.1 数据集

本研究使用的乳腺癌超声图像数据集来源于 Kaggle 平台的 Dataset of breast ultrasound images 数据集<sup>[16]</sup>和 华南理工大学附属第六医院。其中 Kaggle 平台数据集包含 780 张图像, 采集于 2018 年, 涵盖 25 至 75 岁女性的乳腺超声检查结果, 涉及 600 名女性患者, 华南理工大学附属第六医院的数据集包含 31 张图像, 涉及 31 名女性患者, 平均年龄 42, 中位数年龄 45, 涵盖 25 至 65 岁女性的乳腺超声检查结果, 本研究经医院伦理委员会审查批准, 均已获得患者知情同意。

两个数据集图像均以 PNG 格式存储, 平均大小为  $500 \times 500$  像素, 图像分为三类: 包含正常、良性与恶性, 不同数据集中每个类别中的数量如表 1 所示, 不同类别的数据样本如下图所示。

表 1 数据集不同类别的数据样本数量

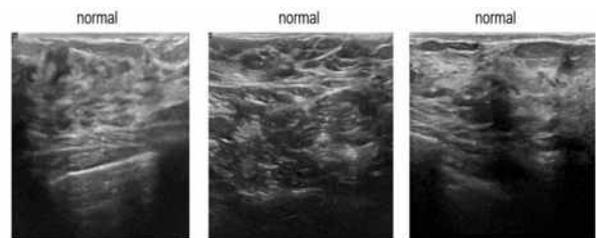
数据集	类别	数量
Kaggle	良性	487
	恶性	210
	正常	133
华南理工大学 附属第六医院	良性	15
	恶性	9
	正常	7
总数	/	811



(a) 良性图像



(b) 恶性图像



(c) 正常图像

图 1 良性 (a)、恶性 (b) 以及正常 (c) 乳腺癌超声图像

### 1.2 基于 Resnet+Transformer 的乳腺癌超声图像分类

乳腺癌超声图像分类的目标是通过深度学习模型对超声图像进行分类, 判断其属于良性、恶性还是正常组织。深度学习是一种先进的机器学习技术, 它模拟人类大脑神经网络的运作方式, 通过构建多层非线性处理单元 (即神经元), 对输入数据进行逐级抽象和特征提取, 从而实现了对复杂数据模式的自动学习和高效识别。

卷积神经网络 (CNN) 是经典深度学习模型, 常用于视觉图像分析。它由输入层、隐藏层和输出层组成, 可处理一维或多维数据。ResNet 是 CNN 中的经典模型, 其擅长提取图像的局部特征, 并通过残差连接有效避免了梯度消失问题;

Transformer 是一种深度学习模型架构, 最初应用于自然语言处理领域。它通过自注意力机制捕捉输入序列中各元素之间的关系, 从而提高了并行处理能力。这种机制使得 Transformer 模型能够在处理图像时, 还能够考虑到整个图像中其他区域的信息, 从而更准确地理解图像内容。

本研究通过结合 ResNet 和 Transformer 的优势, 先利用 ResNet 提取局部特征, 再通过 Transformer 捕捉全局上下文信息, 可以显著提升分类性能。能够更全面地理解图像内容, 从而提高分类的准确性和鲁棒性。整体网络结构如下图所示。

网络架构由 ResNet18 与 Transformer 双分支并行组成, 通过层次特征互补实现乳腺癌超声图像的分类。其中, ResNet18 分支专注于局部特征提取: 输入图像 ( $224 \times 224 \times 3$ ) 经过  $7 \times 7$  卷积和最大池化后, 通过四个阶段的残差块逐步下采样, 最终经全局平均池化输出 512 维特征向量, 完整保留了 CNN 对局部纹理、边缘等细节特征的强捕捉能力。Transformer 分支则采用视觉 Transformer 架构实现全局建模: 将图像分割为  $16 \times 16$  的 196 个图像块, 经线性投影生成 768 维嵌入向量并添加位置编码, 随后通过 12 层 Transformer 编码器进行跨区域关系建模, 最终通过序列平均

池化得到 768 维全局特征向量。两分支特征在拼接层进行融合，将 512 维 CNN 特征与 768 维 Transformer 特征直接拼接为 1280 维混合特征，最后通过三层全连接网络完成三分类。

此设计通过 Transformer 结合 ResNet18 的深度残差学习，使模型既能捕捉局部病灶细节，又能建模全局上下文依赖，显著提升了复杂超声图像的分类鲁棒性。

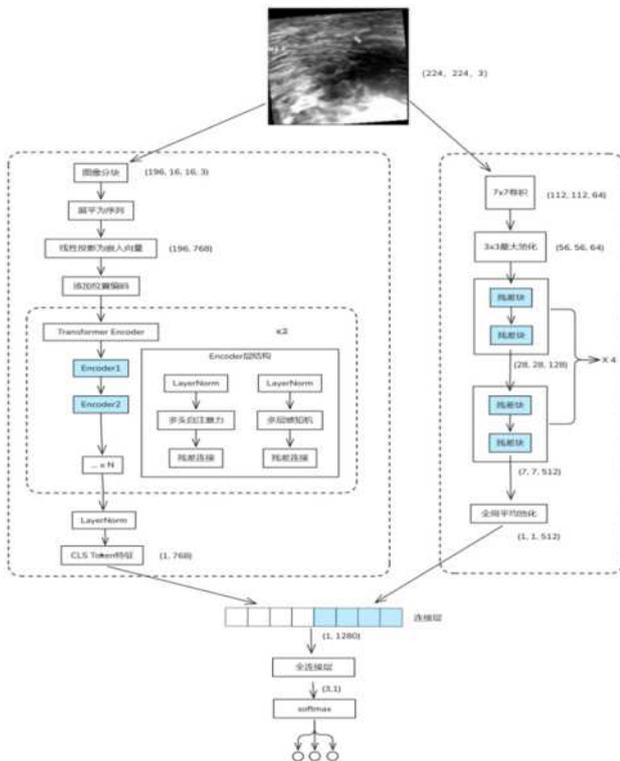


图 2 整体网络架构

### 1.3 评估指标

为了验证模型的性能，本文也采用在多分类问题中经常使用的准确率 (Accuracy)、召回率 (Recall)、精确率 (Precision)，作为衡量深度学习模型在乳腺癌超声图像诊断中的分类评价指标，这些评价指标的百分比数值越大，表示对应的深度学习模型在该数据集集中的性能越好。计算公式如下所示：

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

其中，TP (True Positives): 真正例，预测为正例而且实际上也是正例；FP (False Positives): 假正例，预测为正例

然而实际上却是负例；FN (false Negatives): 假负例，预测为负例然而实际上却是正例；TN (True Negatives): 真负例，预测为负例而且实际上也是负例。

### 1.4 数据增强与扩充

为提升模型的泛化能力和鲁棒性，本文对有限的乳腺癌超声图像数据进行了增强处理。具体方法包括：先将图像调整至  $256 \times 256$  像素，再随机裁剪并缩放回  $224 \times 224$  像素；以 50% 概率进行水平和垂直翻转；随机旋转  $\pm 15$  度；随机调整亮度、对比度、饱和度和色调 (范围分别为  $\pm 0.2$ 、 $\pm 0.2$ 、 $\pm 0.2$  和  $\pm 0.1$ )；随机平移图像宽高的 10%；最后转换 Tensor 格式并归一化 (均值为  $[0.485, 0.456, 0.406]$ ，标准差为  $[0.229, 0.224, 0.225]$ )。通过这些操作，原始数据集扩充至 5 倍，为模型训练提供了更丰富的样本，有效缓解了数据不足的问题。

### 1.5 实验设置

在本研究中，为了实现基于 ResNet 和 Transformer 的乳腺癌超声图像分类模型，本文设计了下述实验环境，并对数据处理和模型训练过程进行了详细配置。

(1) 硬件与软件环境 本文实验环境如下表 2 所示：

表 2 实验环境配置

项目	配置详情
硬件配置	GeForce GTX 3090 GPU, 显存 24GB
操作系统	Ubuntu 20.04
编译器	GCC 9.4.0-1ubuntu1~20.04.1
深度学习框架	PyTorch 1.9.0
CUDA 版本	CUDA 11.6
编程语言	Python 3.8.2

(2) 模型训练配置

在模型训练中，本研究采用交叉熵损失函数 (CrossEntropyLoss) 和 Adam 优化器 (初始学习率 0.001，权重衰减系数  $1e-5$ )。学习率调度器为 StepLR，每 10 个 Epoch 将学习率乘以 0.1。训练总轮数为 30 轮，每批次样本数为 64。最后将 Kaggle 数据集划分为训练集与测试集，比例为 8:2，将华南理工大学附属第六医院数据集作为验证集，进行模型性能的验证，训练集、测试集及验证集数量如下表 3 所示。这些超参数设置经过多次实验验证，可有效平衡收敛速度和分类准确率。

表 3 数据增强与扩充后训练集、测试集及验证集数量分布

数据集	数量
训练集	3120
测试集	780
验证集	155

## 2 结果

在本研究中, 本文对基于 ResNet 和 Transformer 的乳腺癌超声图像分类模型进行了详细的实验验证, 实验结果表明, 融合 ResNet 和 Transformer 的模型在性能上具有显著优势, 如表 4 所示。

为了直观展示模型在训练过程中的性能变化, 本文绘制了 ResNet18 及融合模型在测试集的准确率 (Accuracy) 和损失 (Loss) 曲线。如下图 3 所示:

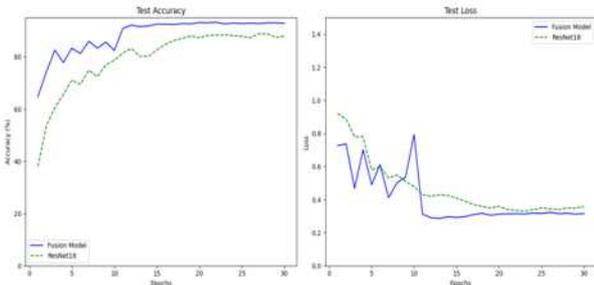


图 3 准确率 (Accuracy) 和损失 (Loss) 曲线

为了全面评估模型的性能, 本文计算了准确率 (Accuracy)、精确率 (Precision) 和召回率 (Recall) 三个指标。实验结果如表 4 所示:

表 4 不同深度学习模型的性能结果

模型	精确率	召回率	准确率
GoogleNet <sup>[10]</sup>	0.827	0.863	0.846
ResNet18	0.861	0.874	0.881
ResNet18+Transformer	0.921	0.926	0.932

为了进一步验证 ResNet+Transformer 模型的泛化性, 本文对验证集进行了预测, 并且对验证集的分类混淆矩阵结果进行了可视化。如下图 4 所示:

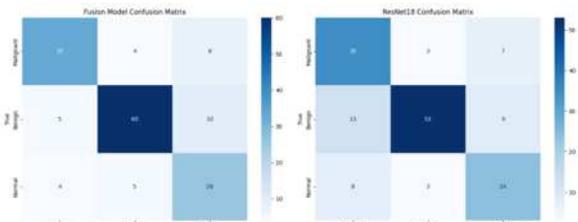


图 4 模型在验证集的混淆矩阵图

混淆矩阵直观地展示了不同模型对不同类别 (良性、恶性、正常) 的分类结果。

## 3. 讨论

从测试集的准确率 (Accuracy) 与损失 (Loss) 曲线 (图 6) 可见, ResNet18+Transformer 融合模型在训练过程中展现出显著的性能优势, 指标均优于单一 Resnet18 模型。这一现象表明, 双分支架构通过局部特征 (ResNet18) 与全局语义 (Transformer) 的互补性, 有效缓解了单一模型的过拟合问题。

如表 4 所示, ResNet18+Transformer 模型的准确率 (93.2%)、精确率 (92.1%) 与召回率 (92.6%) 均显著优于单一模型 (ResNet18: 88.6%) 及经典 GoogLeNet (84.6%)。这一结果归因于双分支的协同效应, 此外, 融合模型在召回率上的优势表明其对恶性样本的漏检率更低, 这对临床筛查具有重要意义。

混淆矩阵可视化结果 (图 7) 进一步揭示了模型的特性。在验证集上, 融合模型对良性样本的分类准确率最高 (80.6%, 75 例中 60 例正确), 恶性样本次之 (73.3%, 45 例中 33 例正确), 而正常样本存在 4 例误判为恶性 (准确率 74.2%)。这种差异可能源于数据分布的不平衡 (良性样本占比达 62.4%) 及正常样本特征的模糊性。相比之下, 单一 Transformer 模型对正常样本的误判率显著升高 (8 例误判为恶性), 凸显了局部特征建模在区分细微病变中的必要性。

虽然本研究中融合模型的各项精度较高, 但是仍存在以下局限性: 数据集中良性样本占比过高 (62.4%), 可能导致模型对正常、恶性类别的敏感性不足; Transformer 分支的 12 层编码器结构使模型参数量增至 85M, 增加了计算资源消耗; 双分支特征融合机制缺乏可解释性量化指标, 难以明确局部与全局特征的贡献权重。

综上, 本文提出的 ResNet+Transformer 模型在乳腺癌超声图像分类任务中表现出了显著的性能优势。通过结合 ResNet 的局部特征提取能力 Transformer 的全局特征建模能力, 该模型能够有效地处理复杂的超声图像, 并在准确率、精确率和召回率等指标上均优于传统的卷积神经网络 (如 GoogLeNet) 和单一模型方法。后续将探索更高效的特征提取方法, 并进一步优化模型的泛化能力, 以推动乳腺癌超声图像自动化诊断的临床应用。

### 参考文献:

- [1] Hou Y. Breast cancer pathological image classification based on deep learning[J]. Journal of X-ray science and technology, 2020, 28(4): 727-738.
- [2] Chen H, Ma M, Liu G, et al. Breast tumor classification in ultrasound images by fusion of deep convolutional neural network and shallow LBP feature[J]. Journal of digital imaging, 2023, 36(3): 932-946.
- [3] Du R, Chen Y, Li T, et al. Discrimination of breast cancer based on ultrasound images and convolutional neural network[J]. Journal of oncology, 2022, 2022(1): 7733583.
- [4] Xie J, Song X, Zhang W, et al. A novel approach with dual-sampling convolutional neural network for ultrasound image classification of breast tumors[J]. Physics in Medicine & Biology, 2020, 65(24): 245001.
- [5] Wang W, Li Y, Yan X, et al. Breast cancer image classification method based on deep transfer learning[C]// Proceedings of the International Conference on Image Processing, Machine Learning and Pattern Recognition. 2024: 190-197.
- [6] 张建新, 高程阳, 孙鉴, 等. 混合监督学习的乳腺癌全切片病理图像分类[J]. 中国图象图形学报, 2024, 29(09): 2825-2836.
- [7] 庄建军, 吴晓慧, 景生华, 等. 多尺度特征融合和改进残差网络乳腺癌病理图像分类[J]. 中国生物医学工程学报, 2024, 43(04): 419-428.
- [8] 襍浚波, 周立广, 梁英豪, 梁淑慧, 付志鸿, 关志广, 毕明霞. 基于 Swin Transformer 和卷积注意力的乳腺癌病理图像诊断研究[J]. 现代电子技术, 2024, 47(13): 36-42.
- [9] 杨芳, 邹迎, 丁雪妍, 张建新. 多层次深度特征融合的乳腺癌病理图像分类[J]. 光电子·激光, 2025, 36(02): 185-192.
- [10] 韩杨, 孙悦, 郭金兴, 等. 基于 GoogLeNet 的乳腺癌超声图像分类[J]. 新一代信息技术, 2023, 6(22): 24-28.
- [11] 方于华, 叶枫. MFDC-Net: 一种融合多尺度特征和注意力机制的乳腺癌病理图像分类算法[J]. 浙江大学学报(理学版), 2023, 50(04): 455-464.
- [12] 龚安, 吕秀明. 基于卷积神经网络的乳腺癌病理图像分类方法[J]. 计算机应用与软件, 2023, 40(06): 133-139+198.
- [13] 张庙林, 帅仁俊. 基于 DC-DenseNet 的乳腺癌病理图像分类[J]. 计算机应用与软件, 2023, 40(04): 116-121.
- [14] 胡天寒, 吴敏, 刘影, 等. 基于 SE-DenseNet 的乳腺癌组织病理学图像分类[J]. 韶关学院学报, 2023, 44(03): 20-27.
- [15] 王剑, 王晓锋. 基于 VGG-19 和 Vision Transformer 的乳腺超声图像分类方法[J]. 信息技术与信息化, 2022, (11): 25-28.
- [16] Pawowska A, Karwat P, oek N. Re: “[Dataset of breast ultrasound images by W. Al-Dhabyani, M. Gomaa, H. Khaled & A. Fahmy, Data in Brief, 2020, 28, 104863]” [J]. Data in Brief, 2023, 48: 109247.

**作者简介:** 黄文剑(1990-), 男, 汉族, 湖北黄石人, 研究生学历, 华南理工大学附属第六医院, 住院医师, 研究方向: 乳腺等相关疾病诊断及治疗

**通讯作者:** 邱学军(1982-), 男, 汉族, 湖南益阳人, 博士, 广东药科大学, 讲师, 研究方向: 生物医学工程

**基金项目:** 线粒体介导的内源性凋亡通路在氧化修饰蛋白诱导干眼发生中的机制研究. 广东省医学科学技术研究基金项目, 2024, B2024035