

第九届全国大学生生物医学工程创新设计竞赛
预赛作品报告

基于人工智能的乳腺癌免疫组化图像生成



作品 ID 号：7110

参赛学生类型：本科生

参加赛道：病理、检验与体外诊断相关

组别：命题项目组

2024 年 6 月

内容完整性自查表

完整性类别	任务或技术指标名称	完成效果	呈现方式（报告中的章节、页码，或者测试报告，或实物、视频等）
设计报告	1. 报告的完整度与美观度	完成	报告全文
	2. 模型介绍的详细程度	较好	P6
	3. 测试结果与分析的严谨性	较好	P13
任务完成情况	1. 生成模型的构建：模型架构的新颖度及精确性	较好	P6
	2. 验证与评估：图像质量评价指标表现、细节呈现的可靠性	较好	P12
	3. 模型的应用价值：利用病理组织成像特性的合理性及临床应用的可能性说明	较好	P4
	4. 数据使用：配对图像使用的巧妙性	较好	P12

摘 要

本项目旨在利用人工智能技术改进乳腺癌的诊断和治疗过程。由于临床上对肿瘤的精准诊疗离不开免疫组化染色，而相比于临床常见的苏木精-伊红染色，免疫组化染色成本高昂、技术要求高、时间消耗相对较多。因此，本项目设计并实现了一个生成式的深度学习模型，用以从苏木精-伊红染色的乳腺癌组织病理图像中生成可靠的免疫组化染色乳腺癌组织病理图像。

我们以 GAN 模型为基础，添加了注意力机制，采用处理过的的 BCI 数据进行了模型的训练，最终能够较为准确地从苏木精-伊红染色的乳腺癌组织病理图像中生成可靠的免疫组化染色乳腺癌组织病理图像。经过大量实验验证，我们的模型不仅展现出卓越的性能，其中 PSNR 和 SSIM 分别达到了 21.9 和 0.47，而且在视觉效果上接近真实的 IHC 染色图像，结构完整性和细节保留方面也达到了较高标准。此外，模型表现出极好的鲁棒性，适应多种染色差异的苏木精-伊红染色图像，辅助病理学家更准确地识别癌细胞特征。

本项目实现的生成式深度学习模型，不仅提高了工作效率，减少了人为误差，还有助于实现诊断流程的标准化。随着技术的不断进步和医疗数据积累，该项目未来有望在个性化治疗选择和预后评估中发挥关键作用，为乳腺癌患者提供更精准的医疗服务，改善患者生活质量。

关键词：人工智能，乳腺癌，免疫组化图像，图像生成，深度学习

目 录

1	选题背景与意义	3
1.1	选题背景	3
1.1.1	乳腺癌病理与诊断	3
1.1.2	乳腺癌传统诊断方式面临的问题	3
1.2	当前研究现状	3
1.3	研究目标与意义	4
2	模型原理及实验方案	4
2.1	模型原理	4
2.1.1	GAN	4
2.1.2	U-Net	5
2.1.3	注意力机制	5
2.2	实验方案	6
2.3	数据准备	9
2.4	实验步骤	10
3	结果呈现与讨论	11
3.1	技术指标	11
3.1.1	定量评估指标	11
3.1.2	定性评估指标	11
3.2	结果呈现	12
3.3	结果讨论	13
4	总结及展望	14
5	参考文献	15

1 选题背景与意义

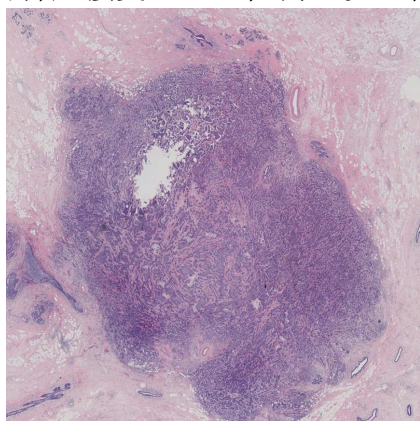
1.1 选题背景

1.1.1 乳腺癌病理与诊断

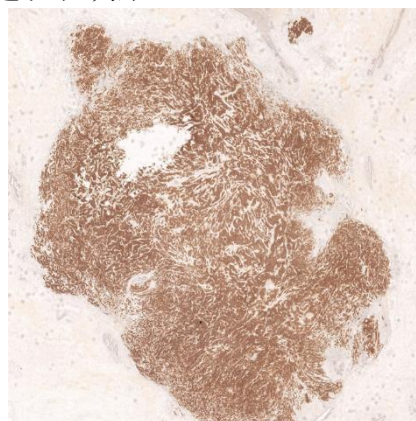
根据国际癌症研究机构（IARC）发布的数据^[1]，2020 年女性乳腺癌已超过肺癌成为最常见的癌症，估计有 230 万例新病例。乳腺癌类型和阶段的早期确定对于制定治疗计划和患者预后至关重要。

目前乳腺癌的诊断是基于用苏木精和伊红（H&E）染色的病理组织作为黄金标准。外科医生从患者的病变区域取出一块组织，经过切片制备和染色等一系列程序，最终制作病理载玻片以供观察。然后，病理学家在显微镜下观察切片并提供诊断。H&E 染色切片如图 1（a）所示。

对于被诊断患有乳腺癌的患者，通常需要进行特异性蛋白质检测以进一步评估肿瘤。例如，人表皮生长因子受体-2 的状态（阳性或阴性）（HER2）需要鉴定乳腺癌，因为 HER2 状态是治疗决策的有用标志物^[2]。如果患者的 HER2 检测呈阳性，医生将进行靶向药物治疗。及时的靶向治疗可以将 HER2 阳性患者的生存机会提高到与 HER2 阴性患者相似的水平。专家建议所有诊断为浸润性乳腺癌的患者都应接受 HER2 检测，以显著改善治疗建议和决策^[3]。



(a) H&E 切片示例



(b) IHC 染色切片示例

图 1 H&E 染色切片和相应的免疫组织化学（IHC）染色切片的可视化

1.1.2 乳腺癌传统诊断方式面临的问题

HER2 表达的评估对于制定乳腺癌后续治疗计划至关重要。然而，通过制备 IHC 染色切片进行 HER2 评估的成本很高。此外，单个 IHC 染色切片可能无法全面评估肿瘤组织中 HER2 的表达水平。例如，如果 HER2 表达水平被诊断为 2+，则可能需要获取用于 IHC 染色的新组织样本。IHC 染色切片的这种额外制备要求进一步增加了与人工和材料相关的成本。

1.2 当前研究现状

随着深度学习的进步，许多智能化应用在医学领域应运而生，如肿瘤细胞分类、肿瘤分割、病理图像染色归一化等。深度学习在上述领域所展现出的强大能力，这也使得其应具备从 H&E 染色图像生成 IHC 染色病理图像的能力。

现有的方法有通过 Diffusion^{[4][5]}、Pix2Pix^[6]等深度学习模型进行从 H&E 染色图像到 IHC 染色病理图像的生成，并且能够取得较好的效果。对这项技术的成功探索有助于节省与 IHC 染色切片制备相关的大量人力和财力成本，让乳腺癌患者能够以较低价格及时获悉癌症阶段，减轻高风险潜在患者能够早发现、早治疗。

1.3 研究目标与意义

本项目的研究目标是构建一个生成式的深度学习模型，经过一定的创新和优化，实现由成本较低的 H&E 染色图像合成较昂贵的 IHC 染色风格图像，并达到较高相似度。

基于上述现有乳腺癌诊断所面临的问题，要在较低成本下能有较好的诊断效果，可借助人工智能的方法，方便对患者进行快速初步的机器诊断。

同时，为了减轻因肿瘤异质性而导致的潜在不准确的 HER2 检测^[7]，通过此生成式的深度学习模型，可以很容易地在浸润性乳腺癌患者的多个 H&E 染色肿瘤组织切片上进行 IHC 染色图像生成。此外，这项研究还将帮助理解和解释 H&E 染色图像中的哪些信息与未来的 HER2 评估相关。

2 模型原理及实验方案

2.1 模型原理

我们的模型采用了生成对抗网络（GAN）、U-Net 结构和注意力机制，以此训练从 H&E 染色图像生成 IHC 染色图像的深度学习模型。

2.1.1 GAN

GAN 模型，即生成对抗网络（Generative Adversarial Networks），是一种深度学习模型，由 Ian Goodfellow 等人^[7]在 2014 年提出。其核心理念是通过两个网络——生成器（Generator）和判别器（Discriminator）的相互对抗，不断优化生成数据的质量，使生成的数据尽可能接近真实数据的分布。

其中，生成器负责生成新的数据样本。它从一个随机噪声向量开始，通过一系列的变换和映射，最终生成与真实数据相似的样本。而判别器负责判断输入的数据是真实数据还是生成器生成的假数据。它的目标是尽量准确地分辨出真实数据和假数据。

GAN 模型自提出以来,已经衍生出了多种变种,如 CGAN、LAPGAN、DCGAN、InfoGan、LSGAN、WGAN 和 CycleGAN 等。这些变种在生成器或判别器的结构上进行了一系列创新,提高了 GAN 的性能和应用范围。

GAN 模型在多个领域都有广泛的应用,包括图像生成、图像到图像的转换、文本生成等。其生成的数据在质量上已经达到了很高的水平,如可以生成逼真的图像、连贯的文本等。

2.1.2 U-Net

U-Net 作为一种深度网络架构,其结构呈对称的 U 型。它由一个“编码器”(收缩路径)和一个“解码器”(扩张路径)组成,中间通过跳跃连接相连。

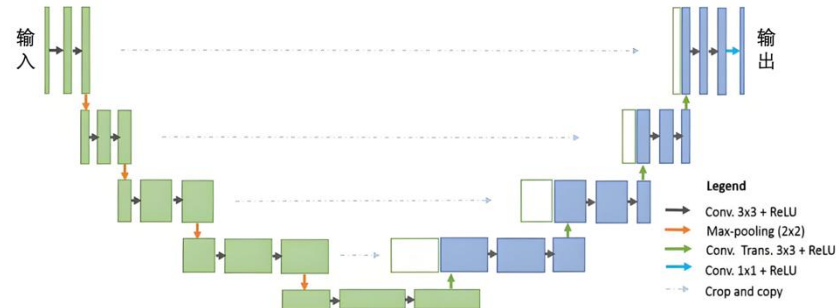


图 2 U-Net 结构示意图

其中的编码器部分是由多个卷积层和最大池化层组成,用于逐渐降低图像的空间分辨率并增加特征通道的数量。每次下采样后,输出特征图的尺寸减半,通道数翻倍。解码器部分则通过上采样过程逐步恢复图像的空间分辨率和细节。上采样层通常由转置卷积层实现,同时与编码器对应层的特征图通过跳跃连接进行合并,以恢复丢失的空间信息。

另外,U-Net 的关键特性之一是跳跃连接,它有助于在上采样过程中恢复细节信息,并允许网络学习更加精确的输出。U-Net 通过其独特的 U 型结构和跳跃连接,可以有效结合低层次和高层次的特征。

基于以上的结构特性,U-Net 可以端到端地进行训练。由于其出色的性能和有效的利用数据的能力,U-Net 在医学图像领域被广泛采用,并也被用于其他任务中。

2.1.3 注意力机制

注意力机制(Attention Mechanism)源于对人类视觉的研究,它模拟了人类在处理信息时,能够选择性地关注某些重要部分而忽视其他部分的能力。在深度学

习中，注意力机制通过计算输入序列中不同位置的重要性权重，使得模型能够动态地聚焦于对当前任务更为关键的信息。

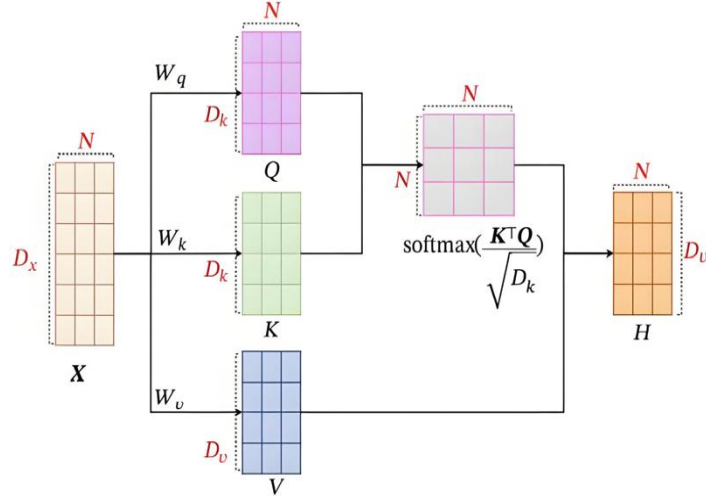


图3 自注意力机制

具体来说，注意力机制主要有两个方面：一方面决定需要关注输入的哪部分，通过计算输入序列中每个位置的重要性权重，模型可以确定哪些部分的信息对于当前任务更为重要；另一方面分配有限的信息处理资源给重要的部分，在计算能力有限的情况下，注意力机制将计算资源分配给更重要的任务，即那些权重更高的部分。

其中，自注意力机制是一种特殊的注意力机制，它关注输入序列内部的依赖关系。自注意力机制的计算过程可以分为三个步骤：首先是查询（Query）、键（Key）和值（Value）的计算，通过线性变换将输入序列转换为查询矩阵（ Q ）、键矩阵（ K ）和值矩阵（ V ）；其次是相关性得分的计算，计算查询矩阵（ Q ）和键矩阵（ K ）之间的相关性得分，通常使用点积运算得到相关性得分矩阵（ S ）；最后加权求和，使用 softmax 函数对得分矩阵（ S ）进行归一化处理，得到权重矩阵（ A ）。然后，将权重矩阵（ A ）与值矩阵（ V ）进行加权求和，得到输出序列。自注意力机制使得每个位置的输出都包含了整个序列的全局信息，从而实现了长距离依赖关系的建模。

2.2 实验方案

本实验方案的核心是开发一个基于生成式对抗网络（GAN）的模型，该模型通过引入注意力机制来改进乳腺癌免疫组化图像的生成质量。GAN 框架包括一个生成器（ G ）和一个鉴别器（ D ），其中生成器负责生成图像，而鉴别器负责区分生成的图像与真实图像。通过这种方式，模型能够生成越来越接近真实图像的合成图像。

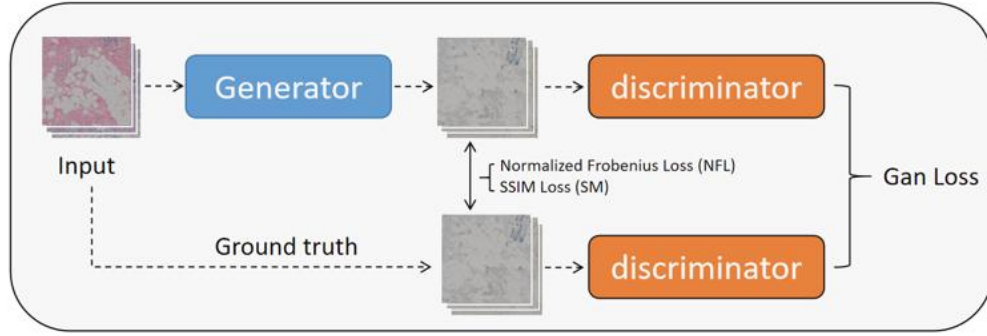


图 4 GAN 模型生成器与判别器

我们使用生成对抗网络（GAN）为训练范式，生成器使用了加入线性自注意力机制（Performer）的 U-Net，并在其中融入了表达量信息，判别器在常规的判断图片真伪的同时，也加入了表达量信息。

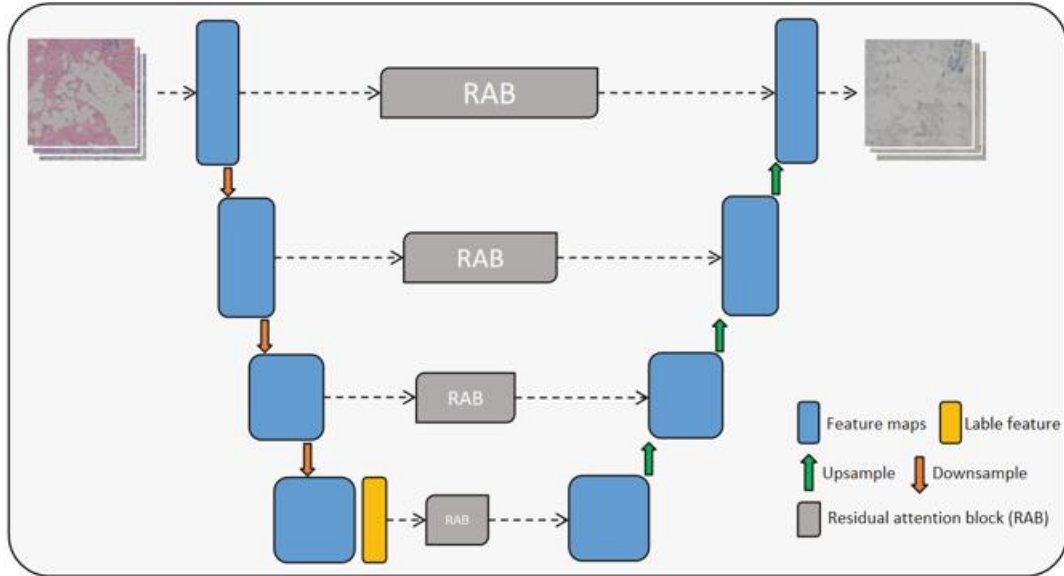


图 5 我们所使用的加入线性自注意力机制的 U-Net

生成器网络在 Unet 的基础上，在跳跃连接处增加了自注意力机制，用以帮助网络更好地融合下采样和上采样得到的特征信息。具体来说，对于得到的特征图，其维度为 $C \times H \times W$ ，其中 C 为通道数， H 和 W 分别为特征图的高和宽，维持通道数不变，将 $H \times W$ 平面展平到一维，将这个 $C \times (H \times W)$ 的张量通过线性变换生成 Q 、 K 、 V 。考虑到特征图的维度很高，为了降低计算复杂度，我们采用了 Performer 论文中提出的 Fast Attention Via positive Orthogonal Random features (FAVOR+) 技术，通过基于随机特征映射简化计算 softmax 的计算量，其将计算注意力的复杂度由 $O(n^2 \times d)$ 将至 $O(n \times d)$ ，其中 n 是序列长度， d 是特征维度。将该技术和残

差链接结合，用于处理跳跃连接后的特征图，我们将其称为 Residual attention block (RAB)。

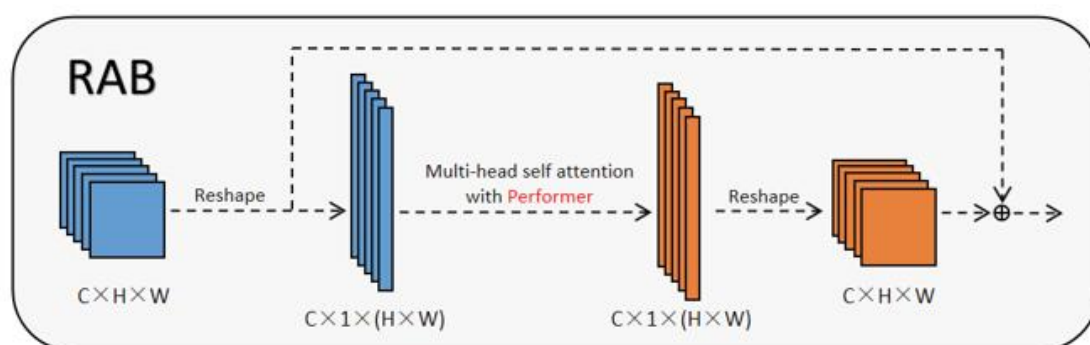


图 6 我们所使用的 Residual attention block (RAB)

为了帮助生成器生成高质量图片，我们将表达量信息融入其中，帮助其训练。具体来说，我们预先训练了一个输入图像判断表达量信息的神经网络，将其最后一层输出的特征图作为融合进生成器的特征向量。

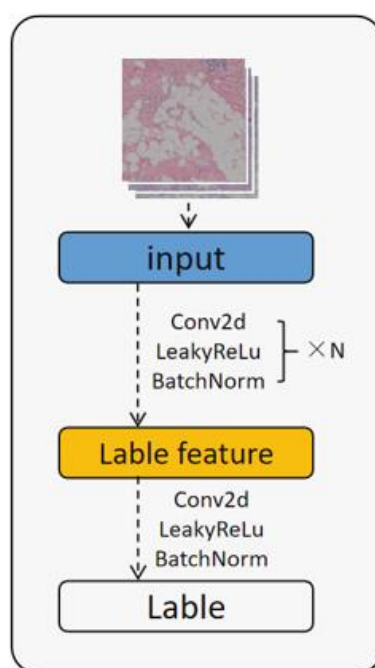


图 7 Lable feature 示意

判别器需要判断生成器生成图像的真伪，以促进生成器生成图像的质量。为了使判别器获得更多图像信息，我们同样融合了图像的表达量信息，即使判别器输出图像真伪和表达量，二者共享网络的底层，进行多任务学习，可以更好地提取图像特征。

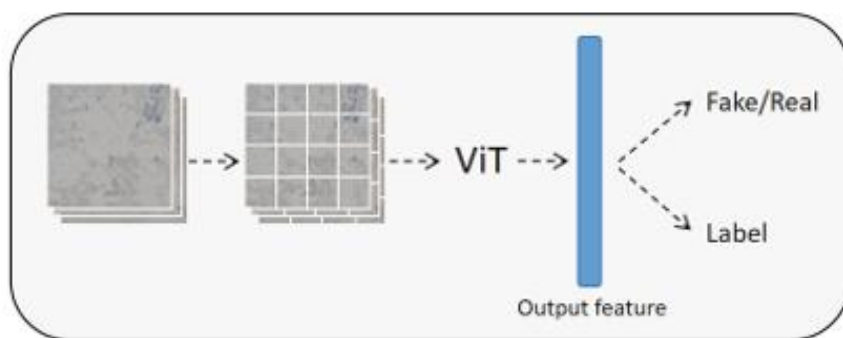


图 8 我们所使用的能够输出图像真伪和表达量的判别器

2.3 数据准备

使用 BCI 数据集^[6]作为训练和测试的基础。该数据集包含了大量的乳腺癌免疫组化图像，每张图像都有相应的标注信息。如图 9 所示，每列表示一组 HE-IHC 图像对，自上到下依次为 4 种 HER2 表达水平（0、1+、2+、3+），共 3 列。每个图像对左侧为 H&E 染色斑块，右侧为免疫组化染色的斑块。

在病理切片对的制备过程中，需要从同一肿瘤组织中连续切出两层切片，分别进行 H&E 染色和 IHC 染色。因此，H&E 染色的切片形状与 IHC 染色的相应切片相似。然后将准备好的病理载玻片扫描到全视野数字切片 WSI 中。具体来说，数据集创作者使用 Hamamatsu NanoZommer S60 扫描出 H&E 染色的 WSI 和相应的 IHC 染色的 WSI。由于计算能力和内存的限制，他们在随后的处理步骤中通过将原始图像的宽度和高度减半来对原始图像进行下采样得到 H&E-IHC WSI 对。然后，用投影变换和弹性配准的方式进行图像配准，分别在整体轮廓和内部细节方面实现了 H&E 染色 WSI 和相应的 IHC 染色 WSI 之间的对齐。

最终提供给我们使用的 BCI 数据包含 9746 张图像（4873 对），其中 3896 对用于训练，977 对用于测试，涵盖多种 HER2 表达水平。

数据集中的图像已经进行了标准化处理，包括调整图像大小、归一化像素值等，从而确保模型训练的有效性和稳定性。该数据构建过程主要包括以下步骤：切片准备、扫描、投影转换、弹性配准、图像细化和 patch 选择。

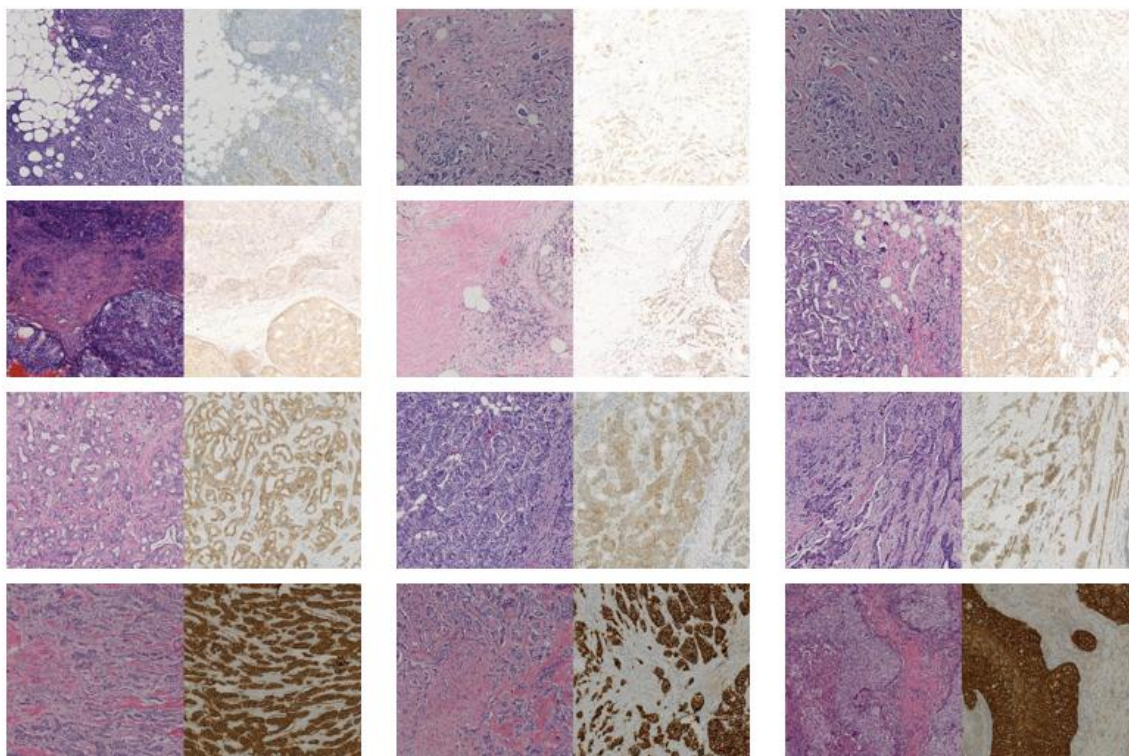


图9 BCI数据集H&E-IHC图像示例

2.4 实验步骤

(a) 数据预处理：对BCI数据集进行必要的预处理操作。

(b) 模型搭建：构建一个GAN模型，设计并实现生成器和鉴别器。具体如下：

生成器（G）：设计一个基于深度卷积神经网络的生成器，用于生成乳腺癌免疫组化图像。在生成器中引入注意力机制，以便模型能够更加关注图像中的关键特征，如细胞结构、组织特征等。

鉴别器（D）：设计一个鉴别器，其结构也基于深度卷积神经网络。鉴别器的任务是区分生成的图像与真实图像。

(c) 模型训练：使用预处理后的数据集对模型进行训练。在训练过程中，逐步调整生成器和鉴别器的参数，直至模型收敛。其中，我们采用交叉熵损失函数作为GAN的损失函数，同时结合生成器和鉴别器的损失来优化模型。除此之外，我们使用Adam优化器进行模型的训练，为生成器和鉴别器设置不同的学习率和beta参数。

(d) 性能评估：使用一组独立的测试数据来评估生成的图像质量和模型性能，可以采用诸如PSNR、SSIM等指标进行评估。

(e) 结果分析：对比生成图像与真实图像，分析模型的生成效果，并探讨注意力机制对图像生成质量的影响。

预期通过引入注意力机制的 GAN 能够有效地生成高质量的乳腺癌免疫组化图像，这些图像在细节上与真实图像高度相似，能够为乳腺癌的研究和诊断提供有价值的视觉材料。

3 结果呈现与讨论

3.1 技术指标

评估生成的乳腺癌免疫组化图像与真实图像之间的相似度通常涉及定量和定性两个方面的指标。

3.1.1 定量评估指标

常用的定量评估指标有以下几种：

1. 峰值信噪比（PSNR）：衡量图像重建质量的指标，主要反映图像的亮度误差和对比度误差。PSNR 值越高，表示图像质量越好。

2. 结构相似性指数（SSIM）：衡量两张图像的结构信息、亮度和对比度的相似性。SSIM 值在 -1 到 1 之间，值越接近 1 表示相似度越高。

3. LPIPS（Learned Perceptual Image Patch Similarity）：使用深度学习模型来评估图像之间的感知相似度，更接近于人类的视觉感知。

4. Inception Score（IS）：通过 Inception 模型计算图像质量和多样性，通常用于评估生成图像的多样性和真实性。

5. Frechet Inception Distance（FID）：计算真实图像和生成图像特征的 Frechet 距离，值越小表示生成图像与真实图像越相似。

根据本竞赛要求，我们选择了 PSNR 与 SSIM 作为定量评估的指标。

3.1.2 定性评估指标

常用的定性评估指标有以下几种：

1. 专家评审：请相关领域的专家对生成的图像进行评估，专家可以基于他们的专业知识来判断生成图像的真实性和适用性。

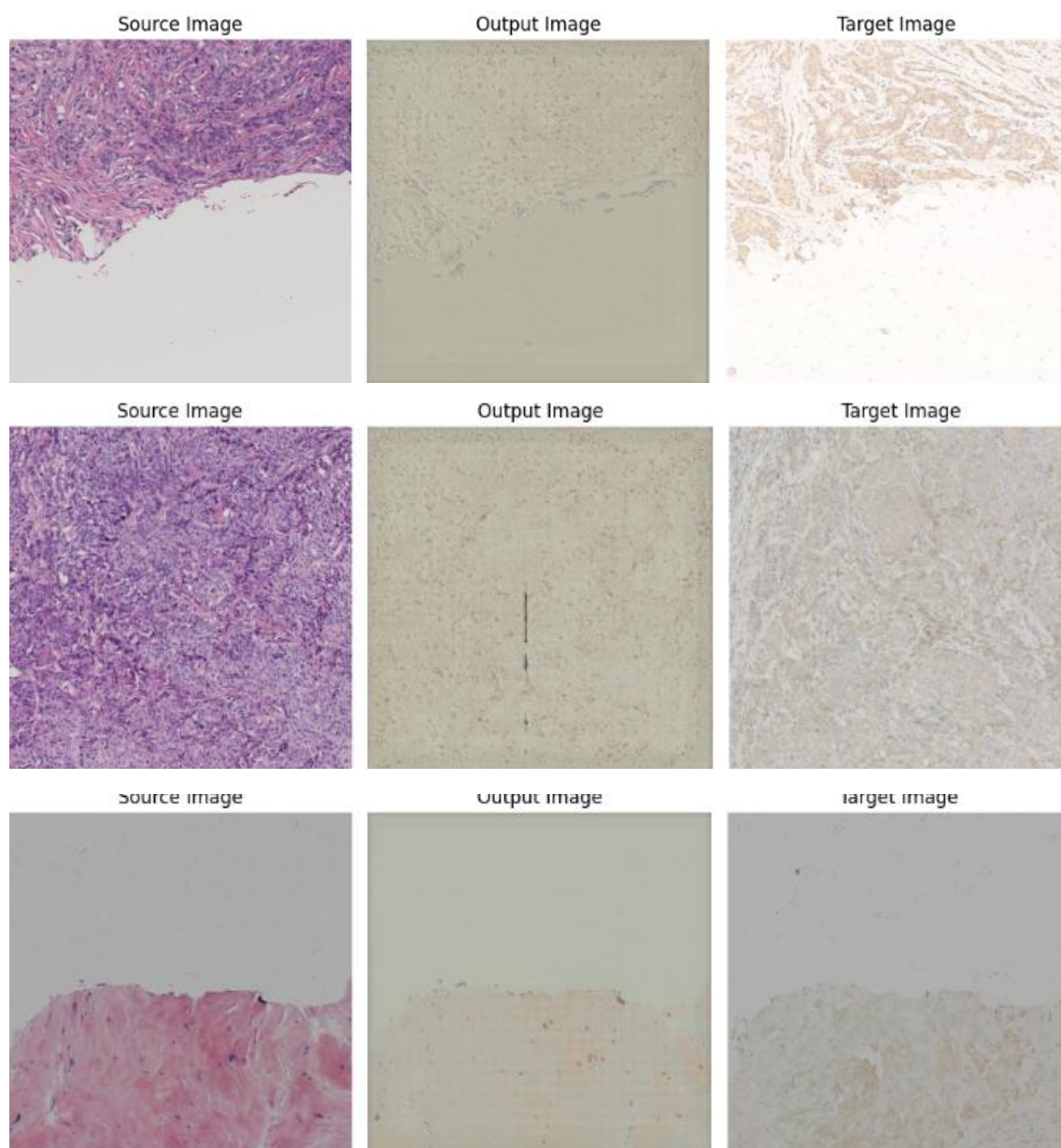
2. 用户研究：邀请一组用户对生成的图像进行评分或排序，以收集他们对图像质量的主观看法。

3. 可视化对比：通过并排对比生成图像和真实图像，直观地展示两者之间的差异。

由于实验过程中出现了实验结果 PSNR 与 SSIM 数值较高但肉眼观测与验证集并不相似的情况，我们队伍与具有临床经验的医生取得联系并参与定性评估。

3.2 结果呈现

图 10 是我们团队使用 GAN 模型得到的 IHC 效果图，与数据集中 IHC 图片十分相似。图 10 左侧为原始的 H&E 染色图像，中间为我们模型输出的结果图像，右侧为数据集中对应的 IHC 目标图像。我们的结果图像与目标图像间的 PSNR 平均值为 18.24，SSIM 平均值为 0.46。



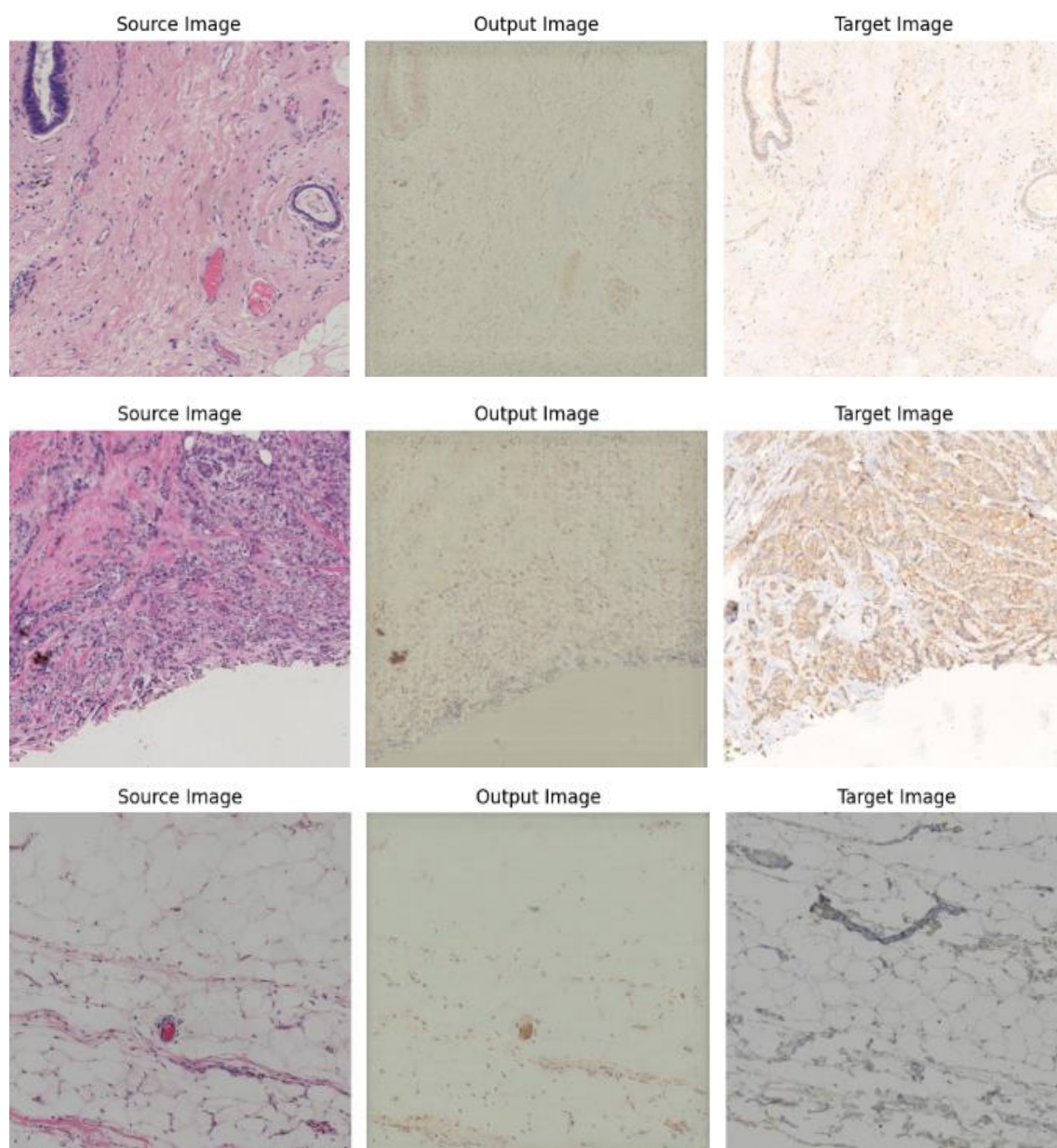


图 10 实验结果与数据集对比展示

3.3 结果讨论

我们通过对模型的创新和优化，达到了较好的定量评估指标，图像的 PSNR 和 SSIM 与此前研究工作接近。同时，我们的工作也具有比较好的视觉效果，肉眼观察可以发现所生成的图像与目标图像具有较高的相似性。根据具有临床经验的医生的意见，我们的工作对临床诊疗工作具有一定参考价值，可以作为癌症诊疗的低成本参考图像。我们认为，在实际诊疗过程中，这有助于患者根据自身的经济状况选择进一步检查方案。

但是，在生成的乳腺癌免疫组化图像中，仍存在一些棘手的问题。例如，生成的图像中非 HER2 表达区域出现为淡黄色，而在真实的 IHC 数据集中这些区域

为白色，这可能会严重影响到该图像的诊疗价值。在乳腺癌早期阶段，HER2 表达水平较低，IHC 图像也为淡黄色，与我们模型生成的图像背景颜色相近，在早期临床诊断中容易出现误诊、漏诊的情况，使模型的临床价值有所降低。我们认为这可能是由以下几个原因造成的：

1. 颜色空间的处理不当：在图像生成过程中，可能由于颜色空间转换或处理不当导致颜色失真。

2. 模型训练不足：如果模型在训练过程中没有充分暴露于足够多样化的颜色分布，或者训练数据不足以捕捉到染色的具体特征，模型可能无法准确地学习和再现特定的颜色。

3. 数据集质量和一致性问题：输入数据集的质量和一致性直接影响模型的学习效果。如果数据集中染色质量不一或存在变异，可能导致模型无法正确学习到特定的染色模式。我们观察到虽然 BCI 数据集经过了挑选，但图片的风格并不完全相同，有的出现大面积的空白、缝隙，有的则几乎均匀填满整个图像。

4. 损失函数和优化问题：生成模型使用的损失函数可能没有很好地惩罚颜色误差，或者优化过程没有充分考虑颜色准确性。

之后我们会考虑采用从 RGB 到 LAB 或 HSV 等颜色空间的转换，在模型中加入专门处理颜色信息的组件，例如使用颜色相关的损失函数或包含颜色编码的网络层，或者通过更多的训练迭代，更细致的参数调整，或者采用更适合颜色生成的训练策略，来提高模型对颜色的生成能力，亦或者增加 BCI 之外的数据集。

实际上，我们观察到生成的 IHC 图像区域灰色颜色深度与真实 IHC 染色图像存在一定的联系，真实 IHC 染色图像颜色越深，生成的图像灰色越深。

4 总结及展望

本报告详细描述了我们在基于人工智能的乳腺癌免疫组化（IHC）图像生成中的实验方案及其实施过程。我们采用了生成式对抗网络（GAN）并引入了注意力机制，旨在生成高质量的乳腺癌 IHC 图像。在实验过程中，我们使用了 BCI 数据集并在 PyTorch 框架下进行了模型的训练和优化。

我们的模型在生成图像质量上取得了一定的成果，其中 PSNR 值达到了 18.24，显示了模型在亮度和对比度上的一定还原能力。然而，SSIM 值为 0.46，表明在结构相似性方面仍有提升空间。特别是，我们注意到生成图像中非 HER2 表达区域的颜色与真实数据集存在差异，这可能是由于模型对颜色信息学习的局限性、数据集多样性不足、颜色空间处理不当、模型架构选择、以及损失函数设置不足等原因造成的。

我们未来的工作将集中在以下几个方面：

1. 数据集的丰富与优化：我们计划扩充和增强数据集，引入更多样化的图像，以训练模型更准确地学习和再现 HER2 表达区域的颜色。
 2. 模型架构的改进：我们将探索和测试新的模型架构，特别是那些能够更好处理颜色信息的网络设计，以提高图像的颜色准确性。
 3. 损失函数的定制：我们将设计新的损失函数，特别强化对颜色准确性的惩罚，以促使模型在训练过程中更加注重颜色信息的恢复。
 4. 专家参与和反馈：我们将邀请领域专家参与评估流程，并提供宝贵的反馈，以指导模型训练和优化的方向。
 5. 技术的迭代与创新：我们将持续关注人工智能领域的最新研究成果，特别是在图像生成和处理方面的技术进展，以便将最新技术应用到我们的工作中。
- 通过对这些关键领域的持续研究和改进，我们相信能够进一步提高模型的性能，生成的乳腺癌 IHC 图像将在质量上更接近真实情况，为乳腺癌的研究和诊断提供更加准确和有用的视觉资料。我们期待在未来的工作中取得更多突破，并为医疗图像分析领域做出更大的贡献。

5 参考文献

- [1] Sung H, Ferlay J, Siegel R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: a cancer journal for clinicians, 2021, 71(3): 209-249.
- [2] Iqbal N, Iqbal N. Human epidermal growth factor receptor 2 (HER2) in cancers: overexpression and therapeutic implications[J]. Molecular biology international, 2014, 2014(1): 852748.
- [3] Yamauchi D H, Hayes D. HER2 and predicting response to therapy in breast cancer[J]. UpToDate, last updated Oct, 2008, 4: 20122012.
- [4] Shen Y, Ke J. StainDiff: Transfer Stain Styles of Histology Images with Denoising Diffusion Probabilistic Models and Self-ensemble[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 549-559.
- [5] Harb R, Pock T, Müller H. Diffusion-based generation of Histopathological Whole Slide Images at a Gigapixel scale[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 5131-5140.
- [6] Liu S, Zhu C, Xu F, et al. Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 1815-1824.
- [7] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.
- [8] Turashvili G, Brogi E. Tumor heterogeneity in breast cancer[J]. Frontiers in medicine, 2017, 4: 227.

