

Crossmark

PAPER

RECEIVED
dd Month yyyyREVISED
dd Month yyyy

虚拟 DES 图像是否是真实图像的有效替代品？

Ana C. Perre¹, Luís A. Alexandre² and Luís C. Freire³

¹Faculdade Ciências da Saúde, Universidade da Beira Interior and Unidade Local de Saúde do Oeste, Av. Infante D. Henrique, 6200-506, Covilhã, Portugal

²NOVA LINCS, Universidade da Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001, Covilhã, Portugal

³Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa, Av. D. João II, lote 4.69.01, Parque das Nações, 1990-096, Lisboa, Portugal

E-mail: ana.perre@ipcb.pt

Keywords: 对比增强光谱乳腺摄影, CESM, 虚拟 DES, 图像转换

Abstract

对比增强光谱乳腺摄影 (CESM) 是一种成像方式, 提供两种类型的图像, 通常称为低能量 (LE) 和双能量减影 (DES) 图像。在许多领域, 特别是在医学领域, 图像到图像的翻译技术的出现使得能够使用其他图像作为输入来人工生成图像。在 CESM 中, 应用此类技术从 LE 图像生成 DES 图像是非常有益的, 可能减少患者因高能图像采集而受到的辐射暴露。在这项研究中, 我们调查了三种用于人工生成 DES 图像 (虚拟 DES) 的模型: 一个预训练的 U-Net 模型、一个端到端训练的 U-Net 模型和一个 CycleGAN 模型。我们也进行了一系列实验来评估使用虚拟 DES 图像对 CESM 检查分类为恶性或非恶性的影响。据我们所知, 这是首次研究评估虚拟 DES 图像对 CESM 病灶分类的影响。结果显示, 预训练的 U-Net 模型表现最佳, 在使用虚拟 DES 图像时 F1 得分为 85.59%, 而使用真实 DES 图像时得分为 90.35%。这种差异可能源于真实 DES 图像中包含的额外诊断信息, 这有助于提高分类准确性。尽管如此, 虚拟 DES 图像生成的潜力依然巨大, 未来的发展可能会缩小这一性能差距, 使仅依赖于虚拟 DES 图像成为临床上可行的选择。

1 介绍

数字化乳房摄影 (DM) 通常被认为是乳腺癌诊断中的金标准成像方式; 然而, 它有一定的局限性, 特别是在致密的乳房中, 可能会出现组织重叠的情况, 导致大量难以确定性质的病灶。[8, 18]。2011 年, 美国食品药品监督管理局 (FDA) 批准了对比增强光谱乳房摄影 (CESM) 作为一种补充成像方式, 用于与 DM 和超声检查一起使用。这种成像方式有效地克服了组织重叠相关的局限性, 并提供了类似于磁共振成像 (MRI) 获得的信息, 有助于定位和描述隐匿或难以确定性质的病灶。[8, 18]。

一般而言, 在进行 CESM 检查时, 会在静脉注射碘化对比剂 (1.5 毫升/千克) 后 2 分钟内依次获取低能和高能图像, 每个乳房视图 (头尾位-CC-和侧斜位-MLO) 都如此 [12, 8]- 参见图 1。因此输出是对比增强的低能 (LE) 图像和双能量减影 (DES) 图像 [3]。这些图像通常可以识别出血管增多或渗漏的乳腺区域, 这是新生物的常见特征 [11, 17]。

由于高能图像仅用于生成 DES 图像 [8], 如果可以从相应的低能量图像人工生成 DES 图像, 则可以避免获取这些高能图像。此过程不仅在工作流程方面具有优势, 还可以减少患者接受的辐射剂量, 从而可能降低 CESM 检查中的辐射暴露。本研究旨在评估虚拟 DES 图像是否

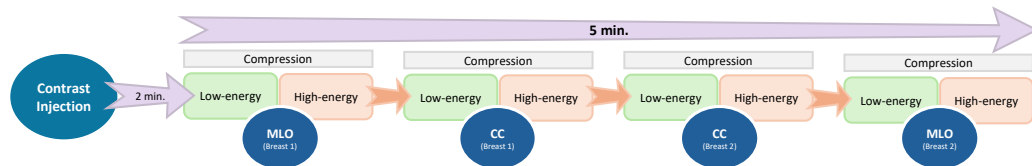


图 1. 示例的 CESM 检查：从乳房 1 开始，可疑的一个，接着是乳房 2，没有病理嫌疑。

可以在训练分类架构以区分恶性与非恶性乳腺图像时作为真实 DES 图像的有效替代品。

若干篇论文已经证明，图像到图像的翻译技术在不同的成像方式中得到了有效的应用，例如磁共振成像（MRI）、计算机断层扫描（CT）和乳房 X 线摄影，允许高效跨模式转换以提高诊断准确性和分析能力 [9, 14]。在 CESM 领域，这种技术开始被探索。自 2024 年以来，已经使用了多种模型来人工生成 DES 图像，包括自动编码器、Pix2Pix、U-net、Resnet-18 和 CycleGAN [14, 9, 6]。直到现在，人们达成共识认为，在评估峰值信噪比（PSNR）、结构相似性指数（SSIM）和均方误差（MSE）测量时，以及在呈现给专家放射科医生时的定性结果方面，CycleGAN 是提供最佳定量结果的模型 [14, 9]。Khorshidifar 等。[9] 指出，CycleGAN 成功保留了结构细节。

霍斯塞尼普尔等人。[6] 应用了一些 U-net 变体，随后意识到这些通常会减少许多明亮细节，并提出了 IAMNet 模型。该模型通过有效突出 DES 图像中的明亮细节，从而取得更好的性能，其亮度和位置与地面真实情况非常接近，可能足以引起对潜在异常的怀疑。

应用其他模型，包括 CycleGAN、Rofena 等。[15]，还提出了 Seg-CycleGAN，该方法从低能量图像合成高保真双能减影图像，利用病灶分割图指导生成过程并改进病灶重建。虽然其他研究开发了生成虚拟 DES 图像的方法，但没有一项关注这些图像对病灶分类的影响。相反，这些研究依赖于定性的人类评估或仅使用 PSNR、MSE 和 SSIM [14, 9, 6] 等指标来评估生成的图像质量，并未考虑对病灶分类系统的影响。在本文中，为了生成这种类型的图像，我们测试了几种模型，即预训练的 U-Net 模型、端到端训练的 U-Net 模型以及 CycleGAN 模型。这一选择源于这些模型是最常用且迄今提供最佳结果的模型。据我们所知，这是首次对虚拟 DES 图像在 CESM 中的病灶分类性能影响进行系统性评估的研究。目标不是获得尽可能好的分类效果，而是要评估使用虚拟 DES 图像与真实图像时相对性能差异，并推断这一差异是否可以成为放弃高能图像采集的理由，从而可能减少辐射暴露。

2 方法

2.1 图像数据库

在这项工作中，我们使用了开罗大学国家癌症研究所放射科提供的分类数字数据库（CDD-CESM） [7]，用于低能量和减影对比增强光谱乳腺摄影图像。该数据库包含 2006 张低能图像以及相应的减影 CESM 图像，在 CC 和 MLO 视图中均有。这些图像在 2019 年 1 月至 2021 年 2 月期间从 326 名女性患者（年龄 18 至 90 岁）处收集，使用了两种不同的设备型号：GE Healthcare Senographe DS 和 Hologic Selenia Dimensions 乳腺摄影系统。图像是以 DICOM 格式采集的。专家放射科医生根据 2013 年美国放射学会乳房影像报告和数据系统（ACR BIRADS）词汇表对图像进行了手动标注，包括但不限于乳房组成、肿块形状、结构扭曲、钙化类型、钙化分布、肿块增强模式以及总体 BIRADS 评估（1 到 6 分） [8]。随后，数据库的作者手动裁剪了大约 30% 的图像以“移除所有无关和未使用的边界”。之后，将所有图像导出为 JPEG 格式，并且对所有病例进行了异常分割标注、验证过的医疗报告以及病理诊断 [8]。

2.2 图像配准

由于某些未经验证的原因，提供的 JPEG 格式的相应 LE 和 DES 图像之间的视野并不相同。例如，图像 P1_L_DM_MLO (LE) 的矩阵大小为 1393×2129 像素，而相应的 CM (DES) 图像的矩阵大小为 1390×2341 像素。通过视觉检查，我们还验证了在同一病变在两张图像中可见的情况下，其坐标并不匹配。我们也得出结论，定位差异不受旋转、比例变化或变形的影响，并且它们在 x 轴或 y 轴上（包括正负两个方向）不超过 10 个像素。因此，我们决定实施一个注册程序，在 Python 中实现，旨在找到必要的注册参数（用于 x 和 y 的平移），以将每对图像带入空间对应关系。

寻找最优的 t_x 和 t_y 值是通过两级搜索方法进行的：在第一级，最佳注册参数在 x 和 y 方向上的搜索以 5 像素为步长，在每个轴上从 -10 到 +10 像素范围内进行，总共进行了 25 个搜索点；在第二级，在第一级找到的最佳平移参数周围执行了一个 5×5 像素邻域的全面搜索。在这两个级别中，平移参数（在 x 和 y 中）始终是整数值（因为搜索是以像素为单位进行的，并非以毫米表示的像素大小）。这样做是为了避免创建浮动图像对齐版本时出现插值错误。所使用的相似性度量是互信息 [10, 2, 20, 19]，其中对于每一对图像，LE 图像是参考图像，DES 图像则是浮动图像，这是考虑到其对 LE 和 DES 图像之间强度差异的鲁棒性。注册结果通过视觉检查手动验证。

2.3 模型

由于目的是评估在病变分类场景中使用虚拟 DES 图像的影响，我们开发了一种模块化分类架构，该架构允许试验不同的虚拟 DES 图像生成器。

如图 2 所示，该架构接收低能量图像 (LE) 作为输入，并将其传递给分类器，同时还带有另外两张图像。第二张图像是通过应用 OpenCV 非局部均值去噪方法获得的 LE 图像的降噪版本（我们将其指定为“LED”）。分类器接收到的第三张图像是真实的 DES 图像（作为基线）或我们希望用作真实 DES 可能替代品的生成图像（指定为“虚拟 DES”）。所使用的方法在子节 3.2.4 中进行了更详细的解释。

正如我们已经提到的，为了生成虚拟 DES 图像，我们测试了几种生成器模型，即预训练或与分类器一起端到端训练的 U-Net 模型以及 CycleGAN 模型，这些将在以下小节中讨论。

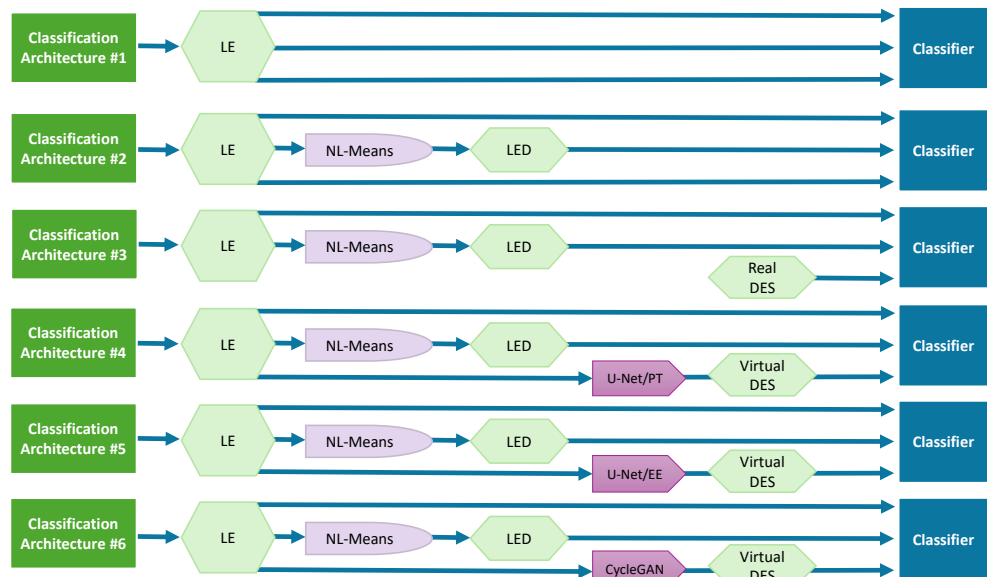


图 2. 本工作中研究的分类架构方案，编号从 1 到 6。在前三种分类架构中，给定分类器的第三张图像是 LE、LED 和真实的 DES 图像。在剩余的分类架构中，第三张图像是由生成模型 U-Net/PT、U-Net/EE 和 CycleGAN 产生的虚拟 DES 图像。

2.3.1 U-Net 创建虚拟 DES 图像的一种方法是使用 U-Net 模型 [16]。尽管它最初是为图像分割而设计的，但 U-Net 适用于图像到图像的转换，因为它能够从输入图像学习映射到输出图像，并通过跳过连接保留空间细节并提取多尺度特征并在全分辨率下重构它们。在初次尝试中，我们预先训练了一个 U-Net 来学习从 LE 到 DES 图像的翻译，仅使用训练集数据。然后我们使用训练好的模型为训练集、验证集和测试集生成虚拟 DES 图像。

第二种方法是通过 U-Net 模型与分类器一起以端到端的方式进行训练。

2.3.2 循环生成对抗网络 在这种情况下，我们研究了虚拟 DES 图像生成的最新技术，并训练了发现性能最佳的模型 [14]。CycleGAN 模型旨在学习两个视觉领域 X 和 Y 之间的映射，而无需配对示例 (x_i, y_i) 。其关键创新是将对抗训练与循环一致性约束相结合。它使用两个生成器网络 $G: X \rightarrow Y$ 和 $F: Y \rightarrow X$ ，每个都配有一个相应的判别器 D_Y 和 D_X 。判别器被训练以区分其领域中的真实图像与来自相反领域生成器的“假”输出。

每对 (G, D_Y) 和 (F, D_X) 都使用标准的 GAN 损失进行训练，由

$$L_{GAN}(G, D_Y) = E_y[\log D_Y(y)] + E_x[\log(1 - D_Y(G(x)))], \quad (1)$$

$$L_{GAN}(F, D_X) = E_x[\log D_X(x)] + E_y[\log(1 - D_X(F(y)))]. \quad (2)$$

给出

这鼓励 $G(x)$ 与 Y 中的真实样本无法区分，同样地 $F(y)$ 要模仿领域 X 。由于不需要直接的 $x \leftrightarrow y$ 对，它通过

$$L_{cyc}(G, F) = E_x[\|F(G(x)) - x\|_1] + E_y[\|G(F(y)) - y\|_1]. \quad (3)$$

强制近似可逆性最小化 L_{cyc} 确保将图像正反翻译可以重建原始内容，防止生成器任意地将所有输入映射到目标域中的单一点。还使用了一个身份映射损失来确保当在生成器中使用目标域的真实样本时，它们被同样地映射到输出。该损失由以下给出：

$$L_{id}(G, F) = E_x[\|F(x) - x\|_1] + E_y[\|G(y) - y\|_1]. \quad (4)$$

完整的训练损失为：

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y) + L_{GAN}(F, D_X) + \lambda_1 L_{cyc}(G, F) + \lambda_2 L_{id}(G, F), \quad (5)$$

其中， λ_s 平衡了视觉真实性和内容保真之间的权衡。

CycleGAN 模型仅使用训练数据集进行训练，之后用于生成训练、验证和测试集的虚拟 DES 图像。在训练和评估分类架构时按需加载这些图像。有关所用训练超参数的详细信息如下。

3 实验

3.1 设置

我们首先仅使用训练集和验证集运行初步实验。这些实验旨在评估最佳的超参数配置，然后我们在最终实验中仅使用选定的设置与测试集进行实验。

初步实验运行了固定的 30 个周期，学习率 (LR) 从集合 $\{1 \times 10^{-3}, 5 \times 10^{-4}, 1 \times 10^{-4}, 5 \times 10^{-5}\}$ 中选取；优化器为 Adam，并保留了这 30 个周期中最佳的验证指标作为特定设置的代表。此外，我们对所有初步实验重复进行了 30 次，并展示了准确率 and F1 分数的平均值与标准差。关于 U-Net 模型，我们改变了第一层中的通道数 $\{4, 8, 12, 16, 32, 64\}$ 和深度 (从 2 到 7)。结果显示，使用 32 个通道和深度 6 时性能最佳，因此在其余实验中固定了这些值。需要注意的是，使用更高的参数值可能会获得更好的结果，但

表 1. 数据集详情: 每个子集中农作物的数量、它们的类别以及每组中图像块总数的百分比。

Set	Non-malignant	Malignant	Total crops	Percentage
Train	561	397	958	70.3
Validation	104	95	199	14.6
Test	116	89	205	15.1
Total	781	581	1362	100

生成的架构将无法适应我们的 GPU 内存。另外, 并非所有这两个参数的组合都可以进行测试, 因为生成的架构对于 GPU 来说可能过大。我们在 U-Net 模型和分类器中测试了 dropout, 其值等于 $\{0.1, 0.2, 0.3\}$ 。这很重要, 因为在使用数据增强的情况下, 数据集大小仍然不大, 而 dropout 有助于对模型进行正则化。在初步实验中发现, 使用 0.2 的 dropout 值是一个好的设置, 因此在其余实验中采用了这一设定。使用的损失函数为交叉熵。

CycleGAN 生成器模型的训练参数为: epoch 数=300, 批量大小=8, 学习率= 10^{-5} , $\lambda_1=10$ 和 $\lambda_2=5$ 。这些值是最初在 [14] 上使用的数值, 我们没有对其进行修改。

我们使用了一台配备 AMD Ryzen 7 3700X 处理器、64GB 内存、1TB 固态硬盘以及 Nvidia RTX 4090 显卡 (24GB 内存) 的 PC, 运行 Pop!OS 22.04 和 Pytorch。

3.2 数据集

图像配准后, 从 LE 和 DES 图像中手动裁剪出 224×224 像素的正方形。对于良性病例和恶性病例, 这是根据病变的位置来考虑进行的。对于正常情况, 这是在乳腺组织内随机进行的。

3.2.1 类标签 根据美国放射学会 (ACR) 的乳腺影像报告和数据系统 (BI-RADS) [4], 如果报告中使用的术语对应于 BI-RADS 类别 1 或 2, 则乳房 X 光检查被认为是正常的或包含良性发现, 因此基本上没有恶性肿瘤的可能性。这些类别通常不需要进一步的影像学检查, 并且预计在随访期间不会发生变化。在这两种情况下, 建议进行常规筛查随访。

考虑到这些信息, 我们决定考虑两类: “非恶性”, 其中包含被认为是正常的 (BI-RADS 1) 和良性的 (BI-RADS 2) 图像, 以及 “恶性” (包括高度可疑和恶性病变即 $\text{BI-RADS} \geq 4$), 我们考虑了具有恶性病变的图像。考虑到本工作中使用的数据库, 这一决定得到了事实的支持, 即被分类为良性 (BI-RADS 2) 的图像数量很少。

3.2.2 数据拆分 作物被随机划分为训练集、验证集和测试集, 使得: 1) 来自同一患者的作物仅出现在其中一个集合中 (患者级别的划分防止数据泄露), 以及; 2) 非恶性与恶性作物的比例在三个集合中大致保持相似 (分层划分)。每个集合中的作物比例分别为 70%、15% 和 15%, 分别对应训练集、验证集和测试集。由于上述两个限制条件, 各部分的图像数量并不完全符合这些百分比, 但非常接近。每组中的作物总数以及非恶性与恶性作物的数量在表 1 中给出, 并附有每个集合中实际图像所占的比例。

3.2.3 数据增强 我们使用数据增强来训练分类器。所使用的变换包括水平和垂直翻转、以 90 度的倍数旋转、锐度调整和自动对比度, 这些都保持了病变的形状。这些可以累积应用, 并且在使用 LE 和 DES 图像进行训练时, 特别注意使两张图像接受相同的变换。图 3 包含了这些变换的一个示例。基于颜色修改的变换在此背景下不适用, 并且任何可能影响病变形状的变化都可能对结果产生负面影响。

3.2.4 输入图像 所有实验均使用 LE 图像作为第一个输入图像。第二个图像是 LE 图像的去噪版本 (我们将其指定为 “LED”), 通过应用 OpenCV 非局部平均去噪方法获得。有一个例外,

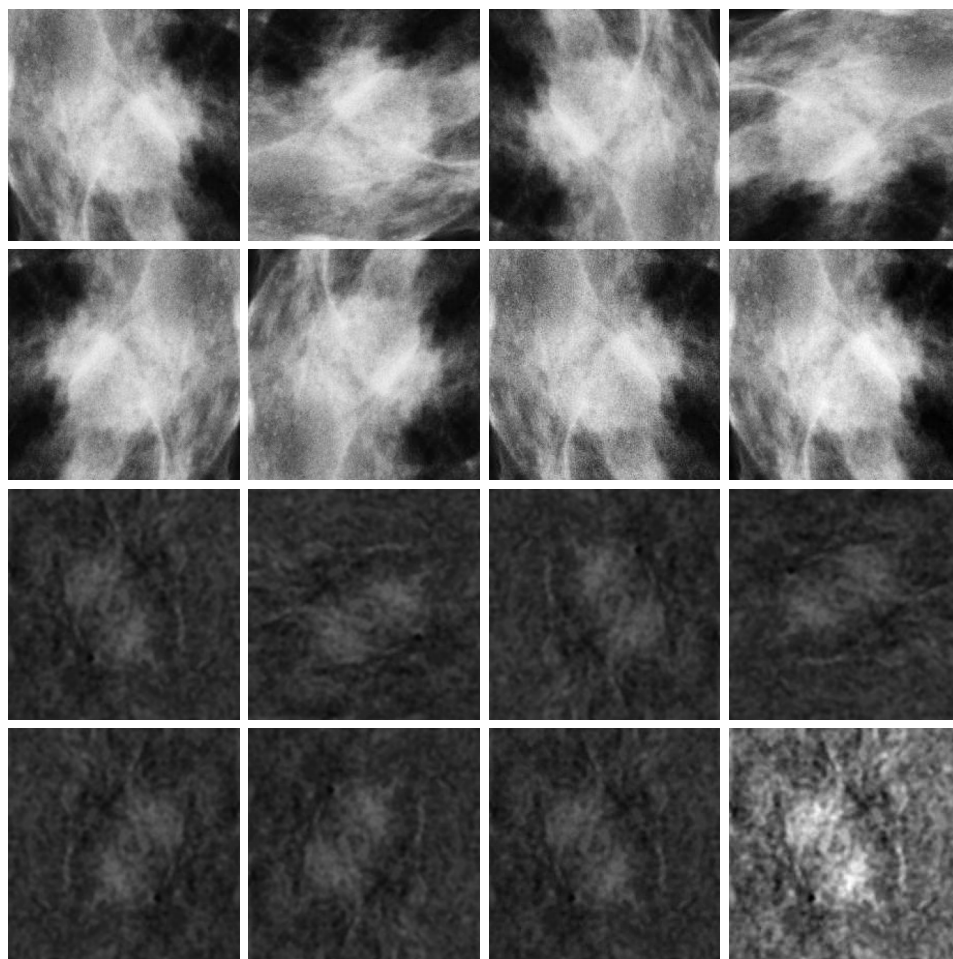


图 3. 示例的同步数据增强：LE 和 DES 图像每次都会收到相同的随机变换。这里从左到右举例说明，原始输入图像（第一行左侧为 LE，第三行左侧为 DES），然后分别是 90、180 和 270 度旋转，接着是水平和垂直翻转，最后进行锐度调整和自动对比度。

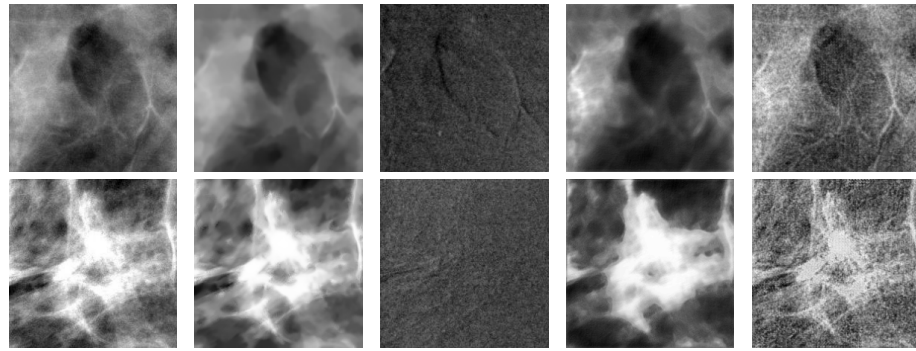


图 4. 每一行从左到右依次包含一个 LE 作物及其相应的 LED、真实的 DES 作物以及使用预训练的 U-Net 和 CycleGAN 生成器架构制作的虚拟 DES 图像。

即我们在三个输入通道中全部使用 LE，以作为没有额外预处理或生成的基线（参见图 2，分类结构#1）。创建 LED 图像的目标是去除图像中的可见噪声，包括 JPEG 压缩伪影。该方法使用三个参数： h 值、模板窗口大小和搜索窗口大小。我们使用的 OpenCV NL-Means 实现建议对于 8 位医学图像， h 值的范围为 10 到 15；我们选择了 $h\text{-value} = 10$ 因为它比较保守，在避免过度平滑精细结构的同时仍然抑制了块效应 [1]。在原始的非局部均值论文 [1] 中，显示 5×5 和 7×7 大小的补丁提供了一个很好的折衷，在捕获足够的局部结构和避免过度平滑之间。我们使用了 OpenCV 默认的 7×7 模板窗口，该窗口直接源自早期实验。我们使用了一个 21×21 像素的搜索窗口，以确保每个参考补丁可以从足够大的邻域中找到良好的“匹配”，这个邻域足以包含结构相似的组织区域（例如附近的腺体模式），同时保持足够小以排除远处无关的解剖结构。

此选择第二个输入图像的理由也得到了表 3 中第一和第二行数据的支持，针对验证集结果，在第一行我们仅有使用 LE 图像的结果，而在第二行则是将 LED 用作第二个图像时的结果：使用 LED 图像的结果相对于使用 LE 图像在平均 F1 分数方面提高了超过 0.8%。

关于第三幅图像，作为基线，我们测试了使用 LE 图像本身以及真实的 DES 图像。这些出现在表 3 的上半部分，并分别位于第二行和第三行数据中。其他版本将使用的第三幅图像要么是 U-Net 模型（预训练或与分类器一起端到端训练）的输出，要么是预训练 CycleGAN 模型的输出。这些选项将分别称为 LE+LED+U-Net/PT、LE+LED+U-Net/EE 和 LE+LED+CG。在分类器中用作输入的图像示例如图 4 所示，为了更好地理解不同的分类架构，请参见图 2。

3.3 分类器

分类器将在我们改变输入图像类型时保持固定，因此可以被视为模型中的非基础块。目标不是获得可能的最佳分类结果，而是评估使用不同输入图像时性能的相对差异，以了解使用虚拟 DES 图像是否合理。尽管如此，我们测试了三种不同的网络作为分类器：ResNet[5] 的两个变体，即形式为 ResNet18 的 ResNet V1.5 和一个 ResNet34，以及 RegNetY_1.6GF[13]。经过几次初步实验后，后者始终取得了最佳结果，因此我们将其固定为剩余实验中的分类器。这些实验显示在表 2 中，并且也用于通过改变学习率和分类器来确定最佳学习率，具体值为 0.001、0.0005 和 0.0001。选定的学习率为 0.0005。输入图像是三个相同的 LE 图像，对应于分类架构#1。这些实验仅使用训练集和验证集进行。

4 结果与讨论

我们使用的数据是不平衡的：非恶性类有 781 张图像，恶性类有 581 张图像，如上所述。虽然我们在表 3 中同时展示了准确率和 F1 分数，但我们认为 F1 分数是相关的度量标准，因为在不平衡分类场景中，准确率可能会误导人，因为分类器可能只是倾向于预测多数类并仍然达到整体较高的准确率。另一方面，F1 分数将精确率和召回率结合成一个单一的指标，从而更好

表 2. 准确率和 F1 分数在不同分类器和学习率下，30 次重复的平均值和标准差（百分比）。输入图像均为 LE，对应分类架构#1。验证集上的结果。

分类器	Learning rate	Mean Valid. Accuracy (std)	Mean Valid. F1 (std)
ResNet18	0.0001	86.77 (1.88)	85.53 (2.44)
ResNet18	0.0005	87.99 (1.90)	87.41 (2.21)
ResNet18	0.001	86.21 (1.58)	85.98 (1.72)
ResNet34	0.0001	88.21 (1.68)	87.21 (2.04)
ResNet34	0.0005	88.06 (1.33)	87.64 (1.47)
ResNet34	0.001	86.05 (1.46)	85.51 (1.84)
RegNetY_1.6GF	0.0001	87.76 (1.64)	86.47 (1.99)
RegNetY_1.6GF	0.0005	89.01 (1.78)	88.27 (2.06)
RegNetY_1.6GF	0.001	88.51 (1.99)	87.72 (2.47)

表 3. 准确率和 F1 分数的平均值及标准差（百分比形式），针对验证集和测试集上 6 种分类架构的不同输入图像和生成器模型，重复 30 次的结果。

分类模型（输入图像）	验证集		测试集	
	Mean Acc. (std)	Mean F1 (std)	Mean Acc. (std)	Mean F1 (std)
#1 (LE+LE+LE)	89.01 (1.78)	88.27 (2.06)	86.13 (2.22)	85.38 (2.07)
#2 (LE+LED+LE)	89.65 (1.57)	89.10 (1.70)	85.24 (2.00)	84.67 (1.97)
#3 (LE+LED+DES)	91.21 (1.65)	90.62 (1.77)	91.48 (1.52)	90.35 (1.72)
#4 (LE+LED+U-Net/PT)	90.13 (1.89)	89.58 (2.07)	86.26 (2.18)	85.59 (2.06)
#5 (LE+LED+U-Net/EE)	89.51 (1.82)	89.02 (2.07)	86.15 (2.47)	85.43 (2.37)
#6 (LE+LED+CG)	88.86 (2.02)	88.20 (2.25)	84.91 (2.04)	84.30 (1.86)

地理解分类器在每个类别中的表现，尤其是在少数类上的表现。这使得 F1 成为一个更具信息量且公平的度量标准，当我们关心正确检测少数类时。此外，当假阴性和假阳性的成本很高时，如在医疗诊断中，我们的案例就是如此，F1 分数尤其有用。

表 3 所示的结果展示了在 6 种分类架构中不同输入图像配置的性能。评估指标，准确率和 F1 分数，在验证集和测试集上重复计算了 30 次。

基线分类模型（前三个数据行）为评估虚拟 DES 方法提供了参考。第一个分类架构#1（使用 LE+LE+LE 图像）采用了三张低排放图像，实现了平均验证准确率 89.01%（标准差：1.78%），F1 分数 88.27%（标准差：2.06%），测试集性能略低，准确率为 86.13%（标准差：2.22%），F1 分数为 85.38%（标准差：2.07%）。

引入去噪低发射率图像（分类架构#2，使用 LE+LED+LE 图像）略微提高了验证性能至 89.65% 的准确性（标准差：1.57%）和 89.10% 的 F1 分数（标准差：1.70%），但测试集的表现仍与 LE+LE+LE 配置相似（85.24% 的准确性，标准差：2.00%；84.67% 的 F1 分数，标准差：1.97%）。这表明去噪 LE 图像提供了边际收益，可能通过减少 JPEG 压缩噪声等伪影来实现，但并未显著提升诊断性能。

第三个分类架构#3（包含图像 LE+LED+DES）整合了真实的 DES 图像，并在所有分类架构中展现出最高的性能，其验证准确率平均值为 91.21%（标准差：1.65%），F1 得分为 90.62%（标准差：1.77%），测试集的结果为 91.48% 的准确率（标准差：1.52%）和 90.35% 的 F1 得分（标准差：1.72%）。这一卓越性能突显了真实 DES 图像的诊断价值，这些图像利用双能量成像的对比度增强来改善恶性特征的检测。

表中后半部分（最后三行）的虚拟 DES 配置旨在复制真实 DES 图像的诊断效用，同时尽量减少辐射暴露。分类架构#4 采用 LE+LED+U-Net/PT 配置，在此配置下，首先训练一个 U-Net 模型并生成虚拟 DES 图像，然后再进行分类器的训练，实现了 90.13%（标准差：1.89%）的平均验证准确率和 89.58%（标准差：2.07%）的 F1 分数，在测试集上的性能为

86.26%的准确率（标准差：2.18%）和 85.59%的 F1 分数（标准差：2.06%）。这些结果与分类架构#2（LE+LED+LE 基准线）相当，并接近分类架构#3（真实 DES）的表现，特别是在验证集上。预训练的 U-Net 生成器产生保留关键诊断信息的虚拟 DES 图像。

在分类架构#5 中，U-Net 模型与分类器进行端到端训练的 LE+LED+U-Net/EE 配置性能稍低，验证准确率为 89.51%（标准差：1.82%），F1 得分为 89.02%（标准差：2.07%），测试集结果为准确率 86.15%（标准差：2.47%）和 F1 得分 85.43%（标准差：2.37%）。端到端训练方法可能面临优化挑战，因为生成器与分类器的联合训练可能导致次优收敛，相比预训练 U-Net 模型而言。然而，测试集上的 F1 性能仍然高于 LE+LED+LE 基线#2，表明这些虚拟 DES 图像仍提供有意义的诊断信息。

LE+LED+CG 架构 #6 使用预训练的 CycleGAN 模型生成虚拟 DES 图像，在所有虚拟 DES 方法中表现最差，验证准确率为 88.86%（标准差：2.02%），F1 分数为 88.20%（标准差：2.25%），测试集结果的准确率为 84.91%（标准差：2.04%），F1 分数为 84.30%（标准差：1.86%）。CycleGAN 模型性能较低的原因可能是其对抗训练框架优先考虑视觉真实感而非诊断准确性，可能会引入伪影或丢失对恶性肿瘤检测至关重要的细微特征。测试集中 F1 分数相比 LE+LED+DES 基线下降约 6%，表明 CycleGAN 的映射未能保留探测器用于区分恶性的细微对比度特征。较高的标准差也表明，与提供真实 DES 图像时相比，CycleGAN 引入了额外的变异性。

5 结论

从低能量图像生成虚拟 DES 图像的想法颇具吸引力，并且在过去 [14, 15, 6, 9] 中一直被追求。在这种情况下，将不再获取高能量图像，这会减少患者的辐射暴露。

这项工作首次量化了 CESM 中虚拟和真实 DES 图像之间的分类性能差距，为未来的研究奠定了基础。我们研究了在训练用于区分恶性与非恶性乳腺图像的分类模型时，这些图像是否可以作为真实 DES 图像的有效替代品。为了实现这一目标，我们首先注册了一个开源 CESM 数据库 [7] 中的 LE 和 DES 图像，之后手动生成选定图像的裁剪，并创建了一种通用架构，使我们能够比较使用真实和虚拟 DES 图像之间的分类性能。最后，我们进行了几次实验测试以了解使用虚拟 DES 图像的影响。

我们发现，正如预期的那样，真实的 DES 图像包含额外的信息，有助于实现更高的 F1 分数。我们的设置性能差异表明，使用虚拟 DES 图像并不能达到与真实 DES 图像相同的分类质量。

所呈现的工作有一些局限性，具体如下：1) 可用的 CESM 图像数量较少，并且不足以对较大模型进行训练，尽管使用了数据增强；2) 我们提出了一种特定的管道，该管道使用三张输入图像到分类器中，虽然可以设计不同的方法；3) 数据集提供者使用了 JPEG 压缩，这在图像中引入了不希望看到的压缩伪影（像 DICOM 这样的格式会更好）；4) 还可以使用其他虚拟 DES 方法，并可能影响这里得出的结论。

可能的后续工作可以集中在通过整合特定领域的知识（例如，血管模式）来创建改进的虚拟 DES 图像生成方法，使用更大的数据集，或探索其他生成模型如扩散模型，在进行风险-收益评估以证明消除真实 DES 图像采集合理性之前。

Funding

此项工作得到了 UID/04516/NOVA 计算机科学与信息学实验室（NOVA LINCIS）的资助，该实验室获得了 FCT.IP 的财政支持。

Author contributions

Ana Perre: Preparation, creation and/or presentation of the published work, specifically writing the initial draft; Data curation, writing—review and editing. Luís Alexandre: Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components; writing—review and editing. Luís Freire: Programming, software development; writing—review and editing, specifically critical review, commentary or revision.

Data availability

所使用的数据集来自 - www.cancerimagingarchive.net/collection/cdd-cesm/

参考文献

- [1] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A non-local algorithm for image denoising. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 60–65, 2005.
- [2] André Collignon, Frederik Maes, Dominique Delaere, Dirk Vandermeulen, Paul Suetens, Guy Marchal, Y Bizais, C Barillot, and R Di Paola. Automated multi-modality image registration based on information theory. *Computational Imaging and Vision*, 3, 1995-01-01.
- [3] Gopichandh Danala, Bhavika Patel, Faranak Aghaei, Morteza Heidari, Jing Li, Teresa Wu, and Bin Zheng. Classification of breast masses using a computer-aided diagnosis scheme of contrast enhanced digital mammograms. *Annals of Biomedical Engineering*, 46:1419–1431, 9 2018.
- [4] C.J. D’Orsi, E.A. Sickles, E.B. Mendelson, and E.A. Morris. *2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*. American College of Radiology, 2014.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [6] Mohammad Hosseinipour, Luca Bergamin, Harel Kotler, Gisella Gennaro, and Fabio Aioli. Addressing challenges in image translation for contrast-enhanced mammography using generative adversarial networks. *EasyChair Preprint 15473*, EasyChair, 2024.
- [7] Rana Khaled, Maha Helal, Omar Alfarghaly, Omnia Mokhtar, Abeer Elkorany, Hebatalla El Kassas, and Aly Fahmy. Categorized digital database for low energy and subtracted contrast enhanced spectral mammography images [dataset]. The Cancer Imaging Archive, 2021.
- [8] Rana Khaled, Maha Helal, Omar Alfarghaly, Omnia Mokhtar, Abeer Elkorany, Hebatalla El Kassas, and Aly Fahmy. Categorized contrast enhanced mammography dataset for diagnostic and artificial intelligence research. *Scientific Data*, 8:122, 2022.
- [9] Asma Khorshidifar, Ghazal Mostaghel, Kaveh Dastvareh, Yashar Ahmadyar, and Rezvan Samimi. Generating pseudo-subtracted image in dual-energy contrast-enhanced spectral mammography using transfer learning, 03 2025.

- [10] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.
- [11] Bhavika K. Patel, Sara Ranjbar, Teresa Wu, Barbara A. Pockaj, Jing Li, Nan Zhang, Mark Lobbes, Bin Zhang, and J. Ross Mitchell. Computer-aided diagnosis of contrast-enhanced spectral mammography: A feasibility study. *European Journal of Radiology*, 98:207–213, 1 2018.
- [12] Shaked Perek, Nahum Kiryati, Gali Zimmerman-Moreno, Miri Sklair-Levy, Eli Konen, and Arnaldo Mayer. Classification of contrast-enhanced spectral mammography (cesm) images. *International Journal of Computer Assisted Radiology and Surgery*, 14:249–257, 2 2019.
- [13] Ilija Radosavovic, Raj Prateek Kosaraju, Ross Girshick, Kaiming He, and Piotr Dollar. Designing network design spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [14] Aurora Rofena, Valerio Guarrasi, Marina Sarli, Claudia Lucia Piccolo, Matteo Sammarra, Bruno Beomonte Zobel, and Paolo Soda. A deep learning approach for virtual contrast enhancement in contrast enhanced spectral mammography. *Computerized Medical Imaging and Graphics*, 116:102398, 2024.
- [15] Aurora Rofena, Arianna Manchia, Claudia Lucia Piccolo, Bruno Beomonte Zobel, Paolo Soda, and Valerio Guarrasi. Lesion-aware generative artificial intelligence for virtual contrast-enhanced mammography in breast cancer, 2025.
- [16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
- [17] William F. Sensakovic, Molly B. Carnahan, Christopher D. Czaplicki, Samuel Fahrenholtz, Anshuman Panda, Yuxiang Zhou, William Pavlicek, and Bhavika Patel. Contrast-enhanced mammography: How does it work? *RadioGraphics*, 41(3):829–839, 2021. PMID: 33835871.
- [18] Jingqi Song, Yuanjie Zheng, Chenxi Xu, Zhenxing Zou, Guocheng Ding, and Wenhui Huang. Improving the classification ability of network utilizing fusion technique in contrast-enhanced spectral mammography. *Medical Physics*, 49:966–977, 2 2022.
- [19] Paul Viola and William M Wells III. Alignment by maximization of mutual information. *International journal of computer vision*, 24(2):137–154, 1997.
- [20] William M. Wells, Paul Viola, Hideki Atsumi, Shin Nakajima, and Ron Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.