

DEEP LEARNING EM ENSAIOS CLÍNICOS VIRTUAIS: A MELHOR ESCOLHA DO MODELO PARA PREDIÇÃO DO ESPALHAMENTO DE FÓTONS EM MAMOGRAFIA

Palavras-Chave: DEEP LEARNING, ENSAIOS CLÍNICOS VIRTUAIS, ESPALHAMENTO, MAMOGRAFIA

Autores(as):

ALINE YOSHIDA MACHADO, UNICAMP – IFGW

Dr. RODRIGO MASSERA, KU LEUVEN

Prof^(a). Dr^(a). ALESSANDRA TOMAL (orientadora), UNICAMP - IFGW

INTRODUÇÃO:

Ensaio clínico virtual (VCT- *Virtual Clinical Trials* em inglês) em imagens médicas de mamografia são usados para avaliar e otimizar sistemas de imageamento, para prototipar ensaios clínicos reais e para regulamentação. São considerados uma alternativa rápida e de menor custo em relação aos métodos tradicionais, especialmente quando se deseja avaliar alterações em parâmetros de aquisição ou reconstrução de imagem sem a necessidade de envolver pacientes reais [1]. Para a realização de VCTs, é necessário o uso de imagens de mamografia sintetizadas com alta fidelidade às imagens obtidas em exames clínicos, o que garante representatividade e validade dos resultados.

Um método tradicionalmente usado para geração de imagens da mama é a simulação Monte Carlo (MC), um método computacional que utiliza amostragens estatísticas para modelar o comportamento de sistemas físicos com comportamento estocástico, como a interação da radiação com a mama [2]. A simulação MC permite gerar imagens de mamografias sintetizadas realísticas e estimar a dose na mama durante o exame utilizando *phantoms* computacionais. Entretanto este método é computacionalmente custoso, exigindo tempos de execução elevados para alcançar fidelidade nas imagens geradas e simulação detalhada de processos como espalhamento, atenuação e absorção depende de um conhecimento preciso dos parâmetros físicos envolvidos, o que pode limitar a escalabilidade do método para estudos de larga escala.

Nesse contexto, técnicas baseadas em deep learning (DL) têm se destacado como ferramentas promissoras para a geração de imagens sintetizadas [3,4]. Modelos de DL são capazes de aprender padrões complexos, como a estimativa do espalhamento de fótons, e vêm sendo estudados por apresentarem custo computacional e tempo de execução significativamente menores quando comparados à simulação MC. No entanto, ao aplicar DL em tarefas envolvendo imagens médicas, é fundamental preservar a fidelidade do padrão estrutural e a qualidade da imagem, evitando a perda de informações estruturais relevantes. A escolha da arquitetura de rede tem impacto direto nesse resultado [5,6]. Assim, quando o objetivo é gerar imagens para VCTs, é essencial avaliar a preservação dos princípios físicos que regem a formação da imagem.

Este trabalho propõe o desenvolvimento de modelos de redes neurais para estimativa da contribuição do espalhamento de fótons em imagens de mamografia a partir da imagem gerada por fótons primários. A partir

dessas estimativas, buscamos criar imagens sintetizadas com potencial de aplicação em ensaios clínicos virtuais. Duas arquiteturas foram avaliadas: uma baseada em ResNet e outra baseada em U-Net++, permitindo a comparação de seus desempenhos na predição do espalhamento e seu impacto na fidelidade das imagens geradas.

METODOLOGIA:

Este trabalho foi dividido em três etapas principais: a preparação dos dados, o treinamento dos modelos e a comparação dos resultados obtidos.

CRIAÇÃO DAS IMAGENS DE TREINAMENTO

As imagens utilizadas neste estudo foram geradas por simulação MC utilizando o código MC-GPU [7], a partir de *phantoms* de mama antropomórficos [8] com espessuras comprimidas entre 1,5 cm e 14 cm. As simulações consideraram duas projeções padrão em mamografia: crânio-caudal (CC) e médio-obliqua lateral (MLO). Foram simulados dez espectros monoenergéticos, com energias no intervalo utilizado em mamografia entre 12,5 keV e 35 keV. Nestas simulações, foi considerada a geometria detalhada do exame mamográfico: um detector de imagem composto por a-Se com área 24 cm x 30 cm, uma distância foco-detector de 70 cm e bandejas de compressão e suporte da mama compostas por PMMA. Para cada configuração, foram geradas duas imagens: a imagem primária (*Primary Image*), obtida considerando apenas os fótons transmitidos sem interação pela mama, e a imagem total (*Target Image*), considerada como *ground truth*, que inclui a contribuição do espalhamento na *Primary Image*. As *features* de entrada para os modelos consistem na *Primary Image*, a energia do feixe monoenergético incidente e a espessura da mama.

PREPARAÇÃO DAS IMAGENS DE TREINAMENTO

Na preparação das imagens de treinamento, adotou-se uma estratégia de divisão dos dados para assegurar que todas as imagens simuladas para o mesmo phantom, com diferentes espectros e projeções, fossem mantidas dentro do mesmo conjunto (treinamento, validação ou teste), a fim de evitar vazamento de informação. Cada imagem passou por uma correção da divergência do feixe, já que a intensidade dos raios X tende a diminuir em regiões mais afastadas do centro do detector. Com isso, a intensidade de cada pixel reflete unicamente as interações do feixe com o tecido mamário, removendo os efeitos puramente geométricos da aquisição. Em seguida, aplicou-se uma normalização logarítmica para reduzir a variação de intensidade entre as imagens, além de um processo de rotação padronizado.

TREINAMENTO DO MODELO

Foram avaliadas duas arquiteturas de redes neurais amplamente utilizadas em tarefas de imagens médicas. A primeira é uma ResNet personalizada, composta por 15 camadas com blocos contendo duas convoluções, *batch normalization*, função de ativação ReLU e conexões de atalho. A segunda arquitetura é a U-Net++, que possui uma estrutura em forma de *encoder-decoder* com etapas de *max pooling* na codificação e *upsampling* na decodificação, utilizando conexões entre diferentes resoluções para preservar características espaciais. Ambas as redes foram treinadas com o mesmo conjunto de dados, composto por 4.980 imagens para treinamento e validação, e 1.240 imagens para teste.

O treinamento foi realizado por 150 épocas, utilizando uma taxa de aprendizado de 0,0001 e função de custo baseada no erro quadrático médio (MSE). A avaliação dos modelos foi realizada com base em duas métricas: o erro percentual absoluto médio (MAPE) e o índice *Quality Index based on Local Variance* (QILV), que quantifica a preservação da variância local nas imagens. Além das métricas quantitativas, foi realizada uma análise visual das imagens preditas em comparação com as imagens *ground truth*.

RESULTADOS E DISCUSSÃO:

A Figura 1 apresenta um exemplo dos dados de treinamento gerados por simulação MC, correspondentes a um *phantom* de mama comprimido na projeção MLO com espessura de 8,1 cm, usando com espectro monoenergético de 20 keV. São exibidas a *Primary Image*, a contribuição de espalhamento (*Scatter*) e a *Target Image* que resulta da soma das duas componentes e é o *ground truth* do modelo de DL.

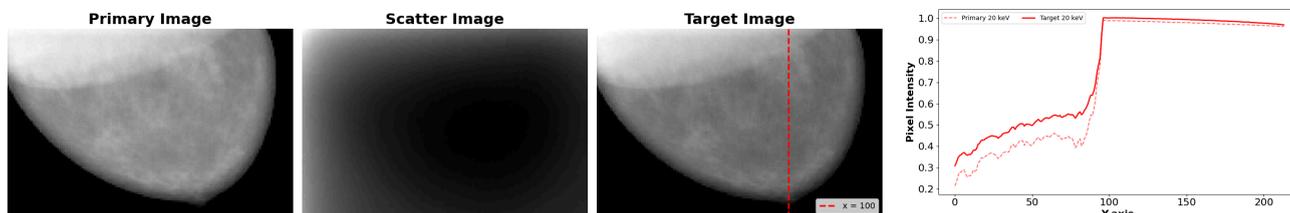


Figura 1: Dados de treinamento do modelo de DL, no qual as features são as Imagens Primárias, espectro monoenergético e a espessura, e o ground truth é o Imagem Target.

O modelo de rede neural foi implementado e treinado utilizando PyTorch v. 2.0.1 com aceleração CUDA, usando hardware GPU para otimizar o processo de treinamento, que teve duração aproximada de 8 horas para a U-Net++ e 12 horas para a ResNet. Esta diferença evidencia que a presença de operações de *downsampling* nos *feature maps* da U-Net++ contribui para uma redução no número de operações computacionais e, consequentemente, no tempo de treinamento.

A Figura 2 apresenta o resultado da predição de uma imagem de mamografia simulada com espectro de 20 keV por ambas as redes, comparando-as com a imagem Target. Nota-se que a U-Net++ tende a suavizar as estruturas, resultando em uma imagem predita com menor granularidade e possível perda de detalhes estruturais relevantes. Por outro lado, a ResNet, embora preserve melhor os contornos internos da mama, apresenta discrepâncias mais visíveis nas regiões periféricas.

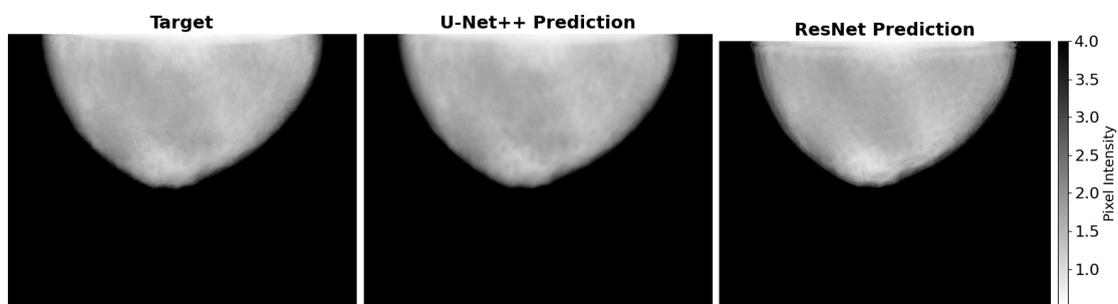


Figura 2: Comparação visual entre Imagem Target e predição da rede Unet++ e Resnet para um espectro de 35 keV

A Figura 3 apresenta a comparação pixel a pixel ao longo de um eixo central da mama, destacando a intensidade dos pixels nas regiões de fundo e de tecido mamário. De forma geral, a predição da U-Net++ aproxima-se mais do valor médio do *ground truth*, especialmente em regiões homogêneas, como o fundo da imagem. No entanto, essa arquitetura tende a suavizar variações locais dentro da região da mama e a eliminar picos de intensidade, o que pode comprometer a preservação de detalhes estruturais relevantes em aplicações como segmentação ou quantificação tecidual.

Tais picos observados na imagem Target refletem a distribuição do tecido glandular presente na imagem primária, e são suavizados durante a predição da U-Net++. Essa característica pode ser atribuída à estrutura da arquitetura, que realiza operações de *downsampling*, priorizando componentes de baixa frequência, como o espalhamento, mas perdendo parte da resolução espacial necessária para preservar a composição da mama.

Por outro lado, a ResNet apresenta maior sensibilidade às variações locais, reproduzindo com mais fidelidade as flutuações de intensidade associadas à estrutura da mama, ainda que com maior discrepância nas regiões externas ao tecido. Essa maior resposta a variações espaciais pode torná-la mais adequada em contextos onde a morfologia do tecido precisa ser mantida com precisão, mesmo que à custa de um erro médio global ligeiramente maior.

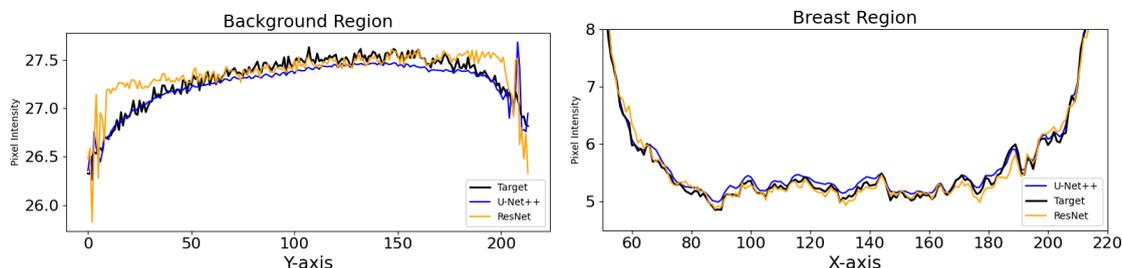


Figura 3: Comparação pixel-a-pixel das predições da rede U-Net++ e Resnet com a Imagem Target para uma região de fundo e uma região dentro da mama

A Figura 4 apresenta a distribuição do erro MAPE obtido para os modelos U-Net++ e ResNet. O histograma à esquerda mostra a densidade em função do MAPE, evidenciando que a U-Net++ apresenta menor erro percentual em grande parte dos casos, com concentração mais acentuada de amostras com MAPE abaixo de 10%. Em contrapartida, a ResNet apresenta uma cauda mais longa, indicando maior variabilidade no desempenho e ocorrência de erros mais elevados em alguns casos. Essa tendência se mantém na análise por energia do feixe, representada no *boxplot* à direita, onde observa-se que o desempenho da U-Net++ é consistentemente superior em termos de mediana e dispersão, especialmente para energias mais baixas entre 12,5 a 20 keV, onde a diferença entre os modelos é mais pronunciada.

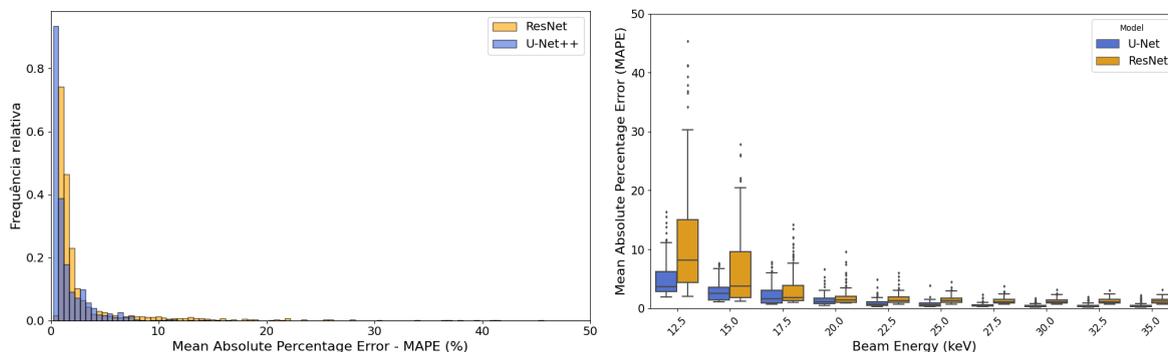


Figura 4: Histograma da distribuição do MAPE e boxplot do MAPE para cada espectro monoenergético.

As métricas de avaliação foram calculadas considerando apenas a região da mama, delimitada por uma máscara. O MAPE para a ResNet foi de $2,2 \pm 4,7\%$, enquanto para a U-Net++ foi de $1,5 \pm 1,8\%$, indicando desempenho superior da U-Net++ em termos de erro médio. Por outro lado, ao se considerar a métrica QILV, que avalia a qualidade da imagem com ênfase na preservação da estrutura espacial, a ResNet apresentou valor médio de $0,99 \pm 0,04$, ligeiramente superior ao obtido pela U-Net++ ($0,98 \pm 0,05$). Esse resultado sugere que, embora a U-Net++ produza predições mais próximas da média da imagem Target, a ResNet demonstra maior capacidade de preservar características espaciais relevantes, possivelmente por ser mais sensível a variações locais, como aquelas presentes nas bordas da mama. É importante destacar que a máscara utilizada para restringir a região de avaliação inclui áreas de transição entre fundo e mama, onde pequenas diferenças na predição podem impactar significativamente as métricas que avaliam a estrutura.

Os modelos escolhidos para a geração de imagens de mamografia sintetizadas impactam diretamente na qualidade da predição e devem ser selecionados com base na aplicação específica nos VCTs. Em aplicações como quantificação do tecido glandular, reconstrução ou processamento de imagens, é fundamental optar por

modelos que mantenham a fidelidade estrutural da imagem, mesmo que apresentem maior erro médio global. Nesse contexto, a ResNet se mostra mais adequada por capturar melhor as variações espaciais associadas à morfologia do tecido. Por outro lado, a U-Net++ é vantajosa em aplicações em que a uniformidade da predição e a redução do erro médio sejam mais importantes que a preservação de detalhes finos, como em geração de grandes volumes de dados para treinamento supervisionado em tarefas onde a estrutura local não é crítica.

CONCLUSÕES:

Neste trabalho, foi utilizado um modelo para estimar o espalhamento em imagens de mamografia geradas por simulações MC, com o objetivo de produzir imagens sintetizadas destinadas a VCTs. Foram treinadas e comparadas duas arquiteturas de rede neural, uma baseada em ResNet e outra em U-Net++, avaliando-se seu desempenho quantitativo e qualitativo. Em termos médios, a UNet++ apresentou métricas superiores, como menor MAPE, mas não preservou adequadamente a distribuição espacial associada à composição tecidual da mama. Por sua vez, a ResNet demonstrou maior sensibilidade às variações locais, mantendo melhor a fidelidade estrutural da imagem, apesar de apresentar erros mais elevados nas regiões periféricas.

Esses resultados reforçam que a escolha do modelo deve ser orientada pelo propósito específico do VCT. Em aplicações que exigem acurácia estrutural, como segmentação, cálculo de dose glandular média ou reconstrução de imagem, a preservação da morfologia da mama é crítica, e modelos como a ResNet se mostram mais adequados. Já em cenários onde o desempenho médio e a consistência global são prioritários, a U-Net++ pode ser uma alternativa mais eficiente. Por fim, ressalta-se a importância da escolha criteriosa das métricas de avaliação, pois diferentes métricas capturam aspectos distintos da qualidade da predição, e podem conduzir a interpretações distintas quanto à aplicabilidade do modelo em VCTs.

BIBLIOGRAFIA

- [1] Ehsan Abadi, William P. Segars, Benjamin M. W. Tsui, Paul E. Kinahan, Nick Bottenus, Alejandro F. Frangi, Andrew Maidment, Joseph Lo, Ehsan Samei, "Virtual clinical trials in medical imaging: a review" *J. Med. Imag.* 7(4) 042805 (2020). doi: [10.1117/1.JMI.7.4.042805](https://doi.org/10.1117/1.JMI.7.4.042805)
- [2] Di Franco F., Sarno A., Mettivier G., Hernandez A., Bliznakova K., Boone J., Russo P. "GEANT4 Monte Carlo simulations for virtual clinical trials in breast X-ray imaging: Proof of concept". *Phys. Med.* 2020;74:133–142 (2020). doi: [10.1016/j.ejmp.2020.05.007](https://doi.org/10.1016/j.ejmp.2020.05.007).
- [3] Oyelade, O.N., Ezugwu, A.E., Almutairi, M.S. et al." A generative adversarial network for synthetization of regions of interest based on digital mammograms". *Sci Rep* 12, 6166 (2022). doi: [10.1038/s41598-022-09929-9](https://doi.org/10.1038/s41598-022-09929-9)
- [4] Pinto MC, Mauter F, Michielsen K, Biniazan R, Kappler S, Sechopoulos I. "A deep learning approach to estimate x-ray scatter in digital breast tomosynthesis: From phantom models to clinical applications". *Med Phys.* 50(8):4744-4757 (2023). doi: [10.1002/mp.16589](https://doi.org/10.1002/mp.16589).
- [5] Ghodrati V, Shao J, Bydder M, Zhou Z, Yin W, Nguyen KL, Yang Y, Hu P. "MR image reconstruction using deep learning: evaluation of network structure and loss functions". *Quant Imaging Med Surg.*9(9):1516-1527 (2019). doi: [10.21037/qims.2019.08.10](https://doi.org/10.21037/qims.2019.08.10).
- [6] Baccouche, A., Garcia-Zapirain, B., Castillo Olea, C. *et al.* "Connected-UNets: a deep learning architecture for breast mass segmentation". *npj Breast Cancer* 7, 151 (2021). doi: [10.1038/s41523-021-00358-x](https://doi.org/10.1038/s41523-021-00358-x)
- [7] Badal A, Badano A. "Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit". *Med Phys.*36(11):4878-80 (2009). doi: [10.1118/1.3231824](https://doi.org/10.1118/1.3231824).
- [8] Graff, Christian G. "A new, open-source, multi-modality digital breast phantom." *Medical Imaging 2016: Physics of Medical Imaging*. Vol. 9783. SPIE, 2016.