

MobileNetV3 Modificado para Classificação de Tipos de Câncer: Reprodução de Experimentos

Gabriel Razzolini Pires De Paula
gabrielrazzolini@ufpr.br
Departamento de Informática
Universidade Federal do Paraná
Curitiba, Paraná, Brasil

Ludwig Aumann
ludwig@ufpr.br
Departamento de Informática
Universidade Federal do Paraná
Curitiba, Paraná, Brasil

Abstract

Este trabalho explora e reproduz a abordagem descrita no artigo “A Modified MobileNetV3 Model Using an Attention Mechanism for Eight-Class Classification of Breast Cancer Pathological Images” [1]. O estudo apresenta uma arquitetura de rede neural convolucional que integra múltiplos estágios de fusão e um kernel com mecanismo de atenção seletiva para a classificação de imagens histológicas de câncer de mama, utilizando o dataset BreakHis [15]. Detalhamos o processo de treinamento, exploramos as técnicas implementadas e discutimos os mecanismos internos do modelo, realizando uma análise comparativa dos resultados alcançados com os reportados no estudo original. Embora não tenhamos alcançado as métricas excepcionais do estudo original, nossos resultados mostraram desempenho razoável, sugerindo que diferenças nos métodos de normalização das imagens podem ter influenciado a discrepância.

Keywords

Câncer de Mama, MobileNetV3, Mecanismos de Atenção, Inteligência Artificial em Diagnóstico, Redes Neurais Convolucionais, Modelagem Computacional em Saúde

1 Introdução

A incidência crescente de câncer de mama destaca a importância do diagnóstico precoce, que é fundamental para melhorar as taxas de sobrevivência dos pacientes [3–6]. Embora o exame histopatológico seja considerado o padrão ouro para diagnósticos pré-operatórios, ele depende muito da expertise do médico e é um processo que consome tempo [2]. A escassez de patologistas e as limitações dos diagnósticos manuais têm impulsionado o uso de inteligência artificial, especialmente aprendizado de máquina, para análise automática de imagens histopatológicas de câncer de mama, aumentando a precisão do diagnóstico e ajudando os médicos na tomada de decisões clínicas [3, 4].

O artigo do título [1] apresenta uma adaptação do estudo de Zhang [8], aplicando a arquitetura MobileNetV3 com mecanismos de atenção para aprimorar a classificação de subtipos de câncer de mama. Este estudo busca integrar fusão de informações em múltiplos níveis e técnicas de atenção para manter a eficiência do modelo enquanto melhora sua acurácia, abordando desafios como a alta demanda por dados rotulados e recursos computacionais intensivos.

2 Design Geral do Modelo

O artigo de referência [1] introduz um modelo leve para a classificação de subtipos de imagens histológicas do câncer de mama, incorporando fusão de informações em múltiplos estágios e um

mecanismo de atenção. A arquitetura do modelo, ilustrada na Figura 1, utiliza o MobileNetV3 como rede principal, aproveitando de sua eficiência para garantir rapidez de processamento e facilidade de implementação. A estratégia de fusão de informações em múltiplos estágios visa explorar ao máximo as informações contidas nas imagens histopatológicas, extraindo e integrando características detalhadas e semânticas profundas ao longo das etapas de processamento.

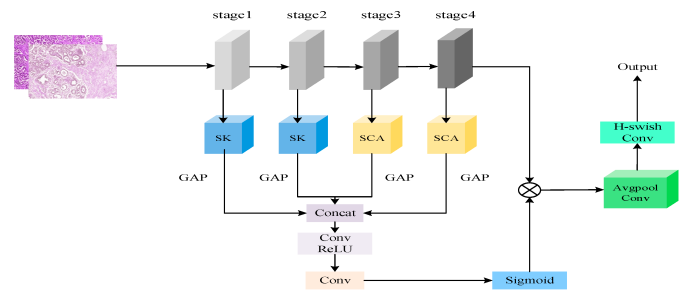


Figure 1: Design Geral da Rede [1]

2.1 Modelo MobileNetV3

Imagens histológicas muitas vezes apresentam ruído significativo e informações complexas, apresentando um desafio considerável para algoritmos de classificação. MobileNetV3, um modelo de aprendizado profundo leve, é amplamente adotado em dispositivos móveis e em computação de bordo devido ao seu tamanho compacto e alta eficiência computacional. Uma combinação de “Hardware-aware” “Network Architecture Search (NAS)” e o algoritmo NetAdapt [10], otimiza o número de kernels de convolução e canais, aumentando a eficiência do modelo, que manteve a convolução “depthwise” separável da versão V1 [11], reduzindo efetivamente o número de parâmetros e operações, tornando-o adequado para aplicações móveis. Adicionalmente, manteve a principal novidade da versão V2: o residual invertido com gargalo linear [9], permitindo ao modelo aprender estruturas de rede mais profundas com baixo custo computacional. A inclusão do mecanismo de atenção de canal squeeze-and-excitation (SE) [12] aprimora a capacidade da rede de focar nos canais de características mais relevantes, melhorando a sua capacidade de representação.

MobileNetV3 está disponível em duas versões: grande e pequena. Para a tarefa de categorização de subtipos de câncer de mama, é essencial capturar sutis diferenças de características entre as imagens. Embora a versão MobileNetV3-Large tenha um número um

pouco maior de parâmetros e demanda computacional em comparação com a versão Small, ela consegue aprender características abstratas de alto nível e padrões complexos, facilitando a diferenciação dos subtipos de câncer de mama. Portanto, a versão MobileNetV3-Large foi selecionada como a rede principal para este estudo.

2.2 Estratégia de Fusão de Informações em Múltiplos Estágios

A MobileNetV3 serve como a rede backbone, extraindo informações de características abrangentes em seu quarto estágio. Devido à complexidade e diversidade das imagens histológicas, é essencial extrair características de textura e cor nos estágios iniciais e características semânticas mais profundas nos estágios posteriores. Para isso, uma estratégia de fusão de informações em múltiplos estágios foi proposta. A estrutura da MobileNetV3 divide a extração de características em quatro estágios.

Nos dois primeiros estágios de extração de características, é utilizado o mecanismo de atenção seletiva Selective Kernel (SK), que captura características fundamentais das imagens, adaptando o tamanho do kernel de convolução para otimizar o processo. Nos estágios seguintes, o modelo integra o módulo Spatial Coordinate Attention (SCA) aprimorado, que foca nas principais localizações espaciais e informações dentro da imagem, capturando características semânticas de alto nível cruciais para o reconhecimento de padrões das imagens. Esse design assegura que o modelo capture as características fundamentais nos primeiros estágios e foque nos atributos semânticos relevantes nas etapas finais, refinando a capacidade de associação das características através do cálculo dos pesos dos mapas de atenção em cada estágio [1].

A fusão de informações em múltiplos estágios processa os mapas de características dos quatro estágios e calcula um mapa de atenção. Este mapa é multiplicado pelo mapa de características do quarto estágio para produzir a saída final, destacando a importância de cada pixel e combinando as características originais com informações relevantes. Dessa forma, o modelo aprimora a interpretação das imagens e aumenta a eficiência na classificação dos subtipos de câncer de mama.

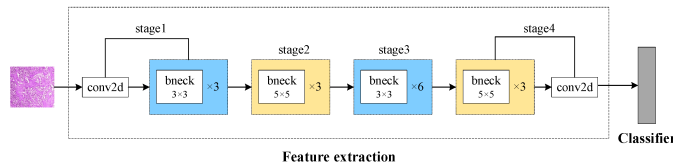


Figure 2: Diagrama Esquemático da Partição em Múltiplos Estágios [1]

2.3 Mecanismo de Atenção SK

O mecanismo de atenção Selective Kernel (SK) ajusta dinamicamente o tamanho do campo receptivo dos neurônios em CNNs [13], processo essencial para processamento de imagens onde diferentes características requerem tamanhos variados de campo receptivo. No artigo de referência, o SK é integrado após a última camada de bottleneck nos dois primeiros estágios de extração de características, melhorando a adaptabilidade do modelo a imagens histopatológicas

de câncer de mama e permitindo a captura de detalhes específicos locais e texturas em diferentes escalas.

O módulo SK utiliza convoluções com kernels de tamanhos 3×3 e 5×5 , proporcionando diferentes campos receptivos. Seu funcionamento envolve três etapas principais:

1. Divisão (Split): O mapa de características de entrada é dividido e processado por convoluções com diferentes tamanhos de kernel, extraindo características em múltiplas escalas.
2. Fusão (Fuse): Os mapas resultantes são somados, combinando informações de diferentes escalas.
3. Seleção (Select): Aplica-se o Global Average Pooling (GAP) para reduzir as dimensões espaciais. Uma camada totalmente conectada extrai informações de atenção dos canais, e uma camada Softmax gera pesos de atenção específicos para cada kernel. Esses pesos recalibram os mapas de características correspondentes.

Finalmente, as características ponderadas são combinadas para produzir o mapa de características final, permitindo que o modelo selecione as informações mais relevantes em diferentes escalas e melhorando a representação das características sem aumentar significativamente a carga computacional.

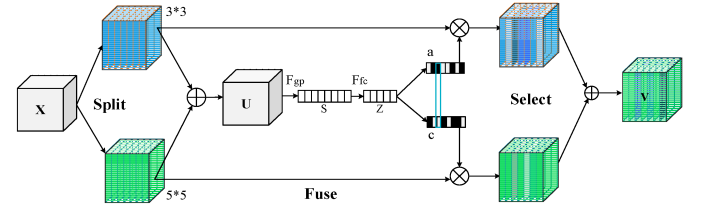


Figure 3: Mecanismo de Atenção SK [1]

2.4 Mecanismo de Atenção Coordenada

O mecanismo Spatial Coordinate Attention (SCA) é uma melhoria do mecanismo de Atenção Coordenada (CA) proposto no artigo referenciado [1]. O CA é um mecanismo de atenção projetado para modelos de deep learning com o objetivo de aprimorar a sensibilidade e o processamento de informações espaciais nos dados de entrada. Ele permite que o modelo foque em diferentes localizações espaciais dentro de uma imagem sem aumentar significativamente a carga computacional, sendo adequado para ambientes com recursos limitados [14].

O mecanismo SCA envolve duas etapas principais:

1. Incorporação de informações de coordenadas: Os mapas de características de entrada passam por um Pooling Médio Global (GAP) nas direções horizontal e vertical para capturar informações posicionais ao longo dos eixos x e y. Esses mapas reduzidos são concatenados e processados por uma convolução 1×1 para criar mapas de atenção de características.
2. Geração de atenção coordenada: Após a aplicação de uma função de ativação, a extração de características é realizada nos mapas ao longo das direções horizontal e vertical. Em seguida, esses mapas de características são multiplicados por pesos, também processados nas respectivas direções, para produzir os mapas de características finais.

Com o mecanismo SCA, o modelo aprimora sua capacidade de processar informações espaciais de forma eficaz, melhorando o reconhecimento de características sutis em diferentes tipos de imagens histológicas, demonstrando grande sinergia com o problema

da classificação de subtipos de câncer de mama, onde lesões pequenas e difíceis de identificar podem impactar o desempenho do modelo [7].

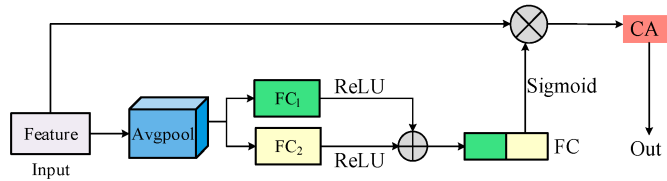


Figure 4: Mecanismo de Atenção Coordenada [1]

3 Configuração Experimental e Análise dos Resultados

3.1 Dataset

Os conjuntos de dados utilizados neste estudo foram obtidos do dataset open-source BreakHis, que é organizado em quatro grupos com base em diferentes ampliações: 40×, 100×, 200× e 400×. Todos os conjuntos de dados passaram por um pré-processamento de normalização de coloração antes de serem usados nos experimentos. Para cada grupo de ampliação, 80% das imagens foram selecionadas aleatoriamente como conjunto de treinamento, enquanto as restantes foram usadas como conjunto de teste. O tamanho das imagens de entrada foi definido em 224 × 224 pixels. Para mitigar o overfitting, as imagens do conjunto de treinamento foram aumentadas, sendo espelhadas horizontalmente e verticalmente com uma probabilidade de 25%.

3.2 Detalhes de Implementação

Todos os experimentos foram realizados em uma máquina com a seguinte configuração: taxa de aprendizado inicial de 0,0001, tamanho de lote (batch size) de 32 e 50 iterações de treinamento. O otimizador Adam foi utilizado para atualizar os pesos da rede, com a convergência controlada pela função de perda de entropia cruzada (cross-entropy loss). Para a classificação das oito classes de patologias, foi utilizada a função Softmax como função de saída.

O ambiente experimental foi reproduzido no sistema operacional ArchLinux 6.11.8, equipado com uma GPU NVIDIA GeForce RTX 1660 Ti 6GB de GPU e 16GB de RAM, fornecendo poder de processamento e espaço de armazenamento não tão próximos do adequado mas suficientes para a reprodução dos experimentos. Todo o desenvolvimento foi feito utilizando Python 3.12.7, sendo todas as tarefas de aprendizado profundo realizadas na estrutura do Pytorch 2.5.1-4. A aceleração por GPU foi possibilitada pelo uso do CUDA 12.7, garantindo eficiência tanto no treinamento do modelo quanto na inferência dos resultados.

3.3 Critérios de Avaliação

O desempenho do modelo foi avaliado de forma abrangente utilizando as métricas de acurácia, precisão, recall, valor F1 e matrizes de confusão. Considerando que o foco está na classificação de subtipos, foi empregada a média macro para calcular os resultados finais. Esta abordagem garante que cada categoria tenha o mesmo peso no

cálculo do valor médio. Consequentemente, o resultado final foi a média aritmética das métricas correspondentes para cada categoria.

4 Resultados

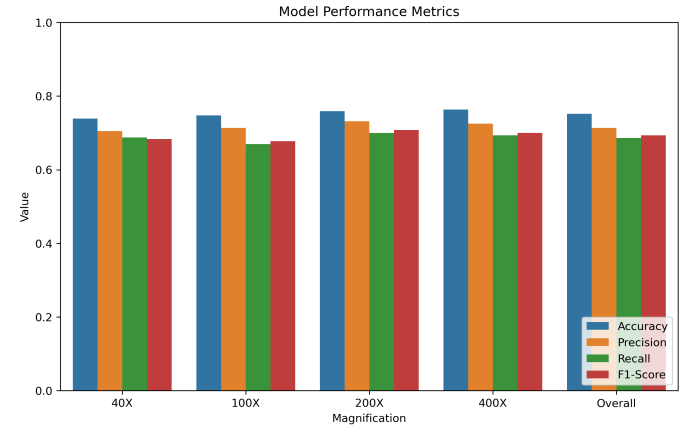


Figure 5: Métricas obtidas: por magnificação e gerais

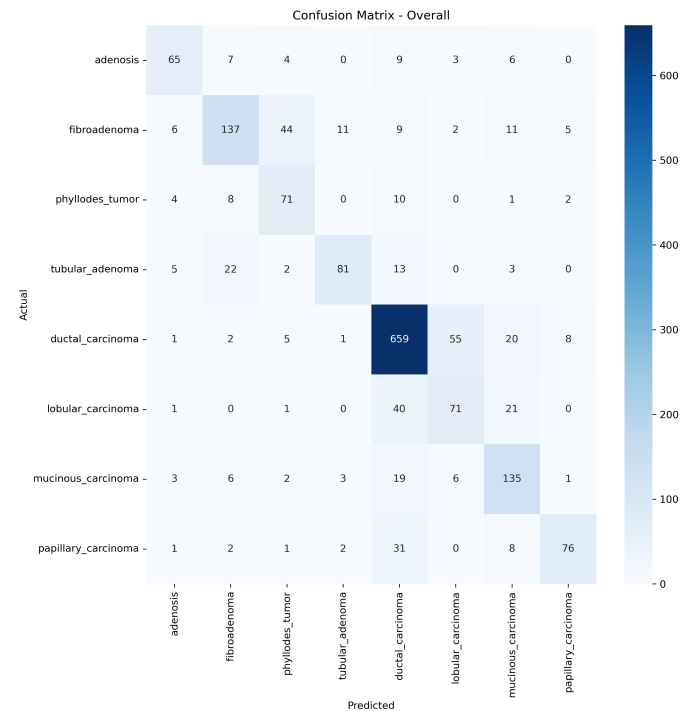


Figure 6: Matriz de confusão do modelo treinado

5 Conclusões

Apesar da insuficiência dos resultados quando contrapostos às excelentes métricas reportadas no artigo original, nossa experiência apresentou um desempenho razoável. Observamos que, enquanto o

estudo original obteve acurácias superiores a 94% em todas as magnificações, nossas métricas variaram entre 76,8% e 81,1% de acurácia. Acreditamos que essa discrepância possa ser atribuída principalmente ao método de normalização e transformação das imagens utilizadas durante o treinamento. O artigo original menciona que todas as imagens passaram por um pré-processamento de normalização de corantes, mas não especifica o método utilizado: “*All the datasets underwent dye normalization preprocessing before being used in the experiments.*” [1]. Além de que a diferença de ambiente e poder computacional é grande, o que os autores acreditam que seja um fator relevante para o desempenho da reprodução da rede e de experimentos feitos.

Diversas alternativas para o processo de normalização foram experimentadas, incluindo o cálculo dinâmico da média e do desvio padrão dos pixels das imagens no conjunto de dados, visando facilitar a normalização durante o pré-processamento. Em outro esforço, diferentes técnicas de melhoria de imagem, como ajuste de saturação, equalização adaptativa de histograma e conversão de RGB para Lab foram verificadas; no entanto, essas abordagens não melhoraram os resultados e, em alguns casos, geraram resultados nas fronteiras do teórico.

Os melhores resultados foram obtidos utilizando uma normalização fixa com os valores $\text{mean}=[0.485, 0.456, 0.406]$ e $\text{std}=[0.229, 0.224, 0.225]$, amplamente utilizados em modelos pré-treinados no ImageNet. Esses valores representam as médias e desvios padrão dos canais de cor no conjunto ImageNet, permitindo uma inicialização mais consistente do modelo. Apesar das diferenças nos resultados, todos os processos e detalhes da implementação do artigo foram seguidos rigorosamente. O código completo, incluindo o modelo, treinamento e testes, está disponível publicamente no GitHub para consulta e replicação: <https://github.com/gabdepa/Deep-Learning>.

Acknowledgments

Os autores gostariam de expressar sua gratidão a Chang Guo, Qingjian Zhou, Jia Jiao, Qingyang Li e Lin Zhu pelo trabalho “A Modified MobileNetV3 Model Using an Attention Mechanism for Eight-Class Classification of Breast Cancer Pathological Images”, que serviu como base integral para este projeto, sendo uma reprodução detalhada de seus experimentos. Agradecemos também ao professor Paulo R. Lisboa de Almeida, cuja orientação nos proporcionaram as ferramentas para explorar as técnicas descritas.

References

- [1] Guo, C.; Zhou, Q.; Jiao, J.; Li, Q.; Zhu, L. A Modified MobileNetV3 Model Using an Attention Mechanism for Eight-Class Classification of Breast Cancer Pathological Images. *Appl. Sci.* **2024**, *14*, 7564.
- [2] Siegel, R.L.; Miller, K.D.; Wagle, N.S.; Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **2023**, *73*, 17–48.
- [3] Ding, R.; Xiao, Y.; Mo, M.; Zheng, Y.; Jiang, Y.Z.; Shao, Z.M. Breast cancer screening and early diagnosis in Chinese women. *Cancer Biol. Med.* **2022**, *19*, 450–467.
- [4] Wang, W.; Li, Q.; Lan, B.; Ma, F. Interpretation of Quality Control Yearbook of Standardized Diagnosis and Treatment of Breast Cancer in China. *Electron. J. Compr. Tumor Ther.* **2023**, *9*, 23–29.
- [5] Williams, B.; Hanby, A.; Millican-Slater, R.; Verghese, E.; Nijhawani, A.; Wilson, I.; Besusparis, J.; Clark, D.; Snead, D.; Rakha, E.; et al. Digital pathology for primary diagnosis of screen-detected breast lesions—experimental data, validation and experience from four centres. *Histopathology* **2020**, *76*, 968–975.
- [6] Liang, Y.; Hao, M.; Guo, R.; Li, X.; Li, Y.; Yu, C.; Yang, Z. Biomarkers for early screening and diagnosis of breast cancer: A review. *Chin. J. Biotechnol.* **2023**, *39*, 1425–1444.
- [7] Radak, M.; Lafta, H.Y.; Fallahi, H. Machine learning and deep learning techniques for breast cancer diagnosis and classification: A comprehensive review of medical imaging studies. *J. Cancer Res. Clin. Oncol.* **2023**, *149*, 10473–10491.
- [8] Zhang, L.; Zhang, C.; Hao, Y.; Cheng, R.; Bai, Y. Breast Cancer Histopathological Image Classification Based on Multi-scale and Multi-gamut Feature Fusion. *Comput. Technol. Dev.* **2022**, *32*, 175–180+185.
- [9] Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for mobilenetv3. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
- [10] Yang, T.J.; Howard, A.; Chen, B.; Zhang, X.; Go, A.; Sandler, M.; Sze, V.; Adam, H. Netadapt: Platform-aware neural network adaptation for mobile applications. In *Computer Vision—ECCV 2018: 15th European Conference*, Munich, Germany, 8–14 September 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, **2018**; Volume 11214.
- [11] Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
- [12] Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
- [13] Li, X.; Wang, W.; Hu, X.; Yang, J. Selective kernel networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 510–519.
- [14] Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 19–25 June 2021; pp. 13708–13717.
- [15] Spanhol, F.; Oliveira, L. S.; Petitjean, C.; Heutte, L. A Dataset for Breast Cancer Histopathological Image Classification, IEEE Transactions on Biomedical Engineering (TBME), 63(7):1455-1462, **2016**. [pdf]