

AVALIANDO O IMPACTO DA QUANTIDADE DE DADOS NO DESEMPENHO DE REDES DE SEGMENTAÇÃO

MARCELO CLASEN RIBEIRO¹, MARILTON SANCHOTENE DE AGUIAR²

¹Universidade Federal de Pelotas – mcribeiro@inf.ufpel.edu.br

²Universidade Federal de Pelotas – marilton@inf.ufpel.edu.br

1. INTRODUÇÃO

Sistemas de diagnóstico auxiliado por computador são ferramentas valiosas que oferecem suporte aos médicos na detecção e diagnóstico de doenças, desempenhando um papel significativo na melhoria do atendimento ao paciente. Além disso, esses sistemas ajudam a aliviar a carga de trabalho dos profissionais de saúde, especialmente em atividades repetitivas relacionadas à análise clínica (YU et al., 2019).

A segmentação de imagens é uma das principais tarefas na análise de imagens em sistemas de diagnóstico auxiliado por computador, sendo amplamente aplicada na identificação de órgãos, anormalidades e ossos, entre outros (MAZUROWSKI et al., 2023). O resultado desse processo pode auxiliar o médico na tomada de decisões e na caracterização de achados clínicos, destacando estruturas de interesse e facilitando a análise.

Sistemas baseados em aprendizado profundo tem apresentado resultados no estado da arte em tarefas de detecção, classificação e segmentação de imagens médicas (WANG et al., 2021). Contudo, essas redes necessitam de uma quantidade significativa de dados para o treinamento de modelos robustos, que consigam inferir resultados precisos para dados fora do conjunto de treinamento. Esse requisito se torna desafiador no contexto de imagens médicas, devido à disponibilidade limitada de dados, resultante de questões relacionadas à confidencialidade e à complexidade de aquisição de imagens, especialmente as rotuladas, que exigem o acompanhamento de profissionais especializados para sua realização (GOCERI, 2023).

O objetivo desse trabalho é avaliar o impacto da quantidade de dados de treinamento no desempenho de redes de segmentação aplicadas a imagens médicas. A pesquisa visa entender como a variação na quantidade de dados influencia a precisão das segmentações e identificar a quantidade mínima necessária para garantir um desempenho robusto. Esse estudo visa contribuir para o desenvolvimento de abordagens mais eficientes em redes de segmentação no contexto de imagens médicas, fornecendo um direcionamento para pesquisadores e profissionais da área.

2. METODOLOGIA

Nesta seção serão descritos os passos realizados para o desenvolvimento da pesquisa, bem como conjunto de dados, rede de segmentação e métricas de avaliação utilizadas.

Para a realização dos experimentos, foi selecionado um conjunto de dados com quantidade significativa de dados, de tal forma que ao realizar a separação do conjunto de treinamento ainda se teriam quantidades de dados similares à de *datasets* comuns da área. Para isso, o conjunto de dados de treinamento original foi dividido nas proporções 10%, 25%, 50%, 75% e 100%, sendo o último correspondendo o conjunto original. A Tabela 1 apresenta a quantidade de dados de cada

divisão realizada, bem como a quantidade de dados de treinamento e validação, que não foram alteradas entre os conjuntos visando uma comparação justa entre os experimentos.

Tabela 1: Quantidade de dados nos conjuntos de treino, validação e teste para os experimentos realizados.

Experimento	Treino	Validação	Teste
10%	2206	7743	480
25%	5516	7743	480
50%	11032	7743	480
75%	16548	7743	480
100%	22064	7743	480

O *dataset* selecionado para o desenvolvimento foi uma versão do BRaTS (*Brain Tumor Segmentation*) *Challenge*, disponibilizado em uma competição anual do uso de aprendizado de máquina em tarefas relacionadas a tumores cerebrais (MENZE et al., 2014). A versão selecionada foi um corte do conjunto de dados do desafio de 2020, na modalidade de imagens T1, sendo imagens de exame de ressonância magnética cerebral. Além disso, para os experimentos realizados, a máscara considerada foi a união das diferentes máscaras de tumor presentes, unidas em uma classe Tumor. O número de imagens resultante desse corte está apresentado no experimento 100% na Tabela 1.

Por sua praticidade de uso e bom nível de desempenho de segmentação, a rede escolhida para desenvolvimento desse trabalho foi a YOLO (*You Only Look Once*), mais especificamente a YOLOv8¹ (JOCHER; CHAURASIA; QIU, 2023). A YOLO é uma abordagem que apresenta resultados no estado da arte em tarefas de detecção, segmentação e classificação de imagens, além de entregar resultados em tempo real, mesmo em sistemas mais modestos. Os modelos de segmentação disponibilizados pelos autores são pré-treinados no conjunto de dados COCO-Seg, variação do COCO (LIN et al., 2015) que possui 330 mil imagens distribuídas entre 80 categorias, porém com a adição de anotações detalhadas para segmentação.

Foram utilizadas as configurações de rede padrões para o treinamento do modelo, diferenciando apenas o número de épocas (definido em 200), o tamanho de *batch* (32 imagens) e o valor de *patience* (10 épocas), relacionado ao *early-stopping*. Além disso, o modelo utilizado foi o *YOLOv8n-seg*, que possui 3.4M parâmetros, sendo o menor disponibilizado pelos autores, escolhido devido à limitação do hardware disponível.

As métricas escolhidas para avaliar o desempenho de segmentação dos experimentos realizados foram *Intersection over Union* (*IoU*) em combinação com *Average Precision* (*AP*). Essa combinação de métricas é comum na área de segmentação de imagens, ao oferecer informações importantes sobre a eficácia do sistema. Para o cálculo do *AP*, define-se um valor mínimo de *IoU* para um objeto ser considerado corretamente segmentado, isto é, quanto do objeto precisa ser corretamente identificado para a segmentação ser avaliada como verdadeira. Valores comuns para a apresentação dessas métricas são *AP*₅₀ e *AP*₅₀₋₉₅, onde o número que acompanha o *AP* se refere ao valor do *IoU*. A métrica *AP*₅₀₋₉₅ é mais

¹<<https://github.com/ultralytics/ultralytics>>

rigorosa, ao avaliar a precisão em múltiplos níveis de IoU , tornando a avaliação mais exigente.

3. RESULTADOS E DISCUSSÃO

Nesta seção serão apresentados e analisados os resultados dos experimentos realizados utilizando a YOLOv8 em diferentes taxas de dados de treinamento. É importante salientar que são apresentados tanto os resultados para caixas delimitadoras (*Bounding box*) quanto para máscaras (*Mask*), nas taxas de $mAP50$ e $mAP50 - 95$. A Tabela 2 resume os valores de $mAP50$ obtidos para cada proporção de dados utilizada.

Tabela 2: Resultados de $mAP50$ no conjunto de teste.

Proporção de Dados (%)	$mAP50$ (Bounding Box)	$mAP50$ (Mask)
10%	0.276	0.230
25%	0.365	0.314
50%	0.681	0.669
75%	0.763	0.764
100%	0.825	0.810

Com os resultados apresentados é possível observar um aumento significativo de desempenho à medida que se aumenta o número de dados, por exemplo, acentuadamente da faixa 25% para 50%. No entanto, a partir de 75% esses ganhos se tornam menos significativos, sugerindo uma saturação na quantidade de dados necessários para melhorar a precisão do modelo.

A Tabela 3 apresenta os resultados de $mAP50 - 95$. Essa métrica é mais rigorosa do que o $mAP50$, apresentando uma relação maior com a precisão da rede.

Tabela 3: Resultados de $mAP50 - 95$ no conjunto de teste.

Proporção de Dados (%)	$mAP50-95$ (Bounding Box)	$mAP50-95$ (Mask)
10%	0.118	0.071
25%	0.141	0.111
50%	0.450	0.373
75%	0.533	0.453
100%	0.619	0.528

Os resultados de $mAP50 - 95$ apresentam tendência similar aos de $mAP50$, porém com valores absolutos mais baixos, devido ao tipo mais rigoroso de avaliação. Em ambos os casos é possível observar que o desempenho de *Bounding Box* é superior à de *Mask*, decorrente da complexidade maior na segmentação de máscaras em relação à de definição de caixas delimitadoras.

Os resultados apresentados indicam forte relação entre a quantidade de dados de treinamento e o desempenho da rede de segmentação. É possível observar um crescimento consistente nas métricas de desempenho com o aumento da quantidade de dados, porém este aumento da precisão não se dá na mesma proporção da quantidade de dados utilizada, como pode ser notado principalmente na métrica $mAP50$.

4. CONCLUSÕES

Neste estudo, investigou-se o impacto da quantidade de dados de treinamento no desempenho de redes de segmentação aplicadas a imagens médicas, utilizando a rede YOLOv8. Os resultados demonstraram que, embora o desempenho da rede melhore significativamente com o aumento da quantidade de dados, esse crescimento se torna menos expressivo após atingir 75% dos dados originais. Esse comportamento sugere que um limiar pode ser estabelecido em uma base de dados de modo que seja suficiente para alcançar um desempenho razoável na tarefa de segmentação de imagens médicas. Contudo, é importante destacar que essa conclusão foi elaborada com base em um único conjunto de dados, podendo não representar a realidade em outros contextos ou modalidades.

5. AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- GOCERI, E. Medical image data augmentation: techniques, comparisons and interpretations. **Artificial Intelligence Review**, Springer, Berlin, Germany, v. 56, n. 11, p. 12561–12605, 2023.
- JOCHER, G.; CHAURASIA, A.; QIU, J. **Ultralytics YOLO**. 2023. Disponível em: <<https://github.com/ultralytics/ultralytics>>.
- LIN, T.-Y.; MAIRE, M.; BELONGIE, S.; BOURDEV, L.; GIRSHICK, R.; HAYS, J.; PERONA, P.; RAMANAN, D.; ZITNICK, C. L.; DOLLÁR, P. **Microsoft COCO: Common Objects in Context**. 2015.
- MAZUROWSKI, M. A.; DONG, H.; GU, H.; YANG, J.; KONZ, N.; ZHANG, Y. Segment anything model for medical image analysis: an experimental study. **Medical Image Analysis**, Elsevier, Amsterdam, Netherlands, v. 89, p. 102918, 2023.
- MENZE, B. H.; JAKAB, A.; BAUER, S.; KALPATHY-CRAMER, J.; FARAHANI, K.; KIRBY, J.; BURREN, Y.; PORZ, N.; SLOTBOOM, J.; WIEST, R. et al. The multimodal brain tumor image segmentation benchmark (brats). **IEEE transactions on medical imaging**, IEEE, v. 34, n. 10, p. 1993–2024, 2014.
- WANG, J.; ZHU, H.; WANG, S.-H.; ZHANG, Y.-D. A review of deep learning on medical image analysis. **Mobile Networks and Applications**, Springer, Berlin, Germany, v. 26, n. 1, p. 351–380, 2021.
- YU, Z.; XIANG, Q.; MENG, J.; KOU, C.; REN, Q.; LU, Y. Retinal image synthesis from multiple-landmarks input with generative adversarial networks. **Biomedical engineering online**, BioMed Central, London, United Kingdom, v. 18, n. 1, p. 1–15, 2019.