Análise do Impacto de Dados Sintéticos para Modelos Segmentadores de Pólipos Adenomatosos em Colonoscopia

Lucas Lima Neves[1], **Adalberto Ferreira Barbosa Junior**[1], **Ricardo Augusto Pereira Franco**[2]

[1]Centro de Excelência em Inteligência Artificial (CEIA)
Goiânia – GO – Brasil

[2]Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

{lucas.neves,adalbertojunior}@egresso.ufg.br, ricardofranco@ufg.br

Abstract. *Deep learning-based computer-aided diagnosis for colorectal cancer is often constrained by annotated data scarcity. This study evaluates the impact of synthetic data on training ten distinct encoder architectures for polyp segmentation. Results demonstrate consistent performance gains that vary by model scale and complexity: Global IoU increased by 1.5%–4.8%, and the Dice Coefficient rose by 2.7%–5.6%. These findings reinforce the efficacy of synthetic data for training set expansion and reveal how architectural differences influence the integration of generated data to improve diagnostic support systems.*

Resumo. *O auxílio a diagnóstico precoce do câncer colorretal, com sistemas de aprendizado profundo em colonoscopia, é comumente limitado pela escassez de dados anotados. Este estudo analisa o impacto da inclusão de dados sintéticos no treinamento de dez encoders de diferentes arquiteturas para a segmentação de pólipos. Os resultados demonstram melhorias consistentes, porém variando conforme o tamanho e a complexidade da rede: o incremento no IoU Global situou-se entre 1,5% e 4,8%, enquanto o Coeficiente Dice cresceu entre 2,7% e 5,6%. Os achados reforçam a eficácia dos dados sintéticos na ampliação de conjuntos de treinamento e evidenciam como diferentes arquiteturas interagem com dados gerados para aprimorar sistemas de apoio ao diagnóstico.*

1. Introdução

O câncer colorretal destaca-se como a terceira neoplasia mais diagnosticada e letal globalmente, sendo responsável por cerca de 10% dos óbitos oncológicos [Sung et al. 2021], com uma taxa de sobrevivência que cai drasticamente de 90% para menos de 10% se não detectado em estágios iniciais [Alberti et al. 2012]. A doença origina-se predominantemente de pólipos adenomatosos, cuja identificação e remoção via colonoscopia representam o padrão ouro para a prevenção [Corley et al. 2014]. Contudo, a eficácia desse procedimento é limitada pela subjetividade da análise humana e fatores como fadiga visual, resultando em taxas de não detecção de até 27% para lesões pequenas (inferiores a 5mm), o que compromete o diagnóstico precoce [Heresbach et al. 2008].

Nesse cenário, estratégias baseadas em aprendizado de máquina (*machine learning*) têm sido amplamente adotadas em aplicações biomédicas. O objetivo principal dessas soluções é oferecer auxílio ao diagnóstico, por meio da identificação automática de

pólipos, reduzindo o erro humano [Marques et al. 2023]. Nesse contexto, a segmentação semântica destaca-se, permitindo delimitar com precisão a área da lesão. Para performá-la, algoritmos de aprendizado profundo (*Deep Learning*), como as Redes Neurais Convolucionais (CNNs), aprendem a processar essas imagens a partir da observação de dados rotulados, dependendo geralmente de grandes volumes de imagens para entregar resultados satisfatórios e generalizáveis [de Melo et al. 2022].

Entretanto, garantir a qualidade e a quantidade de dados em medicina é um desafio complexo. A escassez de *datasets* públicos de colonoscopia, com anotações, limita o treinamento de modelos robustos, dificultando avanços que possibilitem a aplicação destas tecnologias em casos clínicos [Picard et al. 2020]. Para mitigar a escassez de dados reais e aumentar a variabilidade dos conjuntos de treinamento, a geração de dados sintéticos surge como uma fronteira promissora para a possibilitar o melhoramento de modelos de *Deep Learning* (DL) nesta área [Neves et al. 2025].

Apesar do potencial dos dados sintéticos, a literatura atual ainda carece de estudos comparativos aprofundados que avaliem como essa adição de dados impacta diferentes arquiteturas de redes neurais, uma vez que testes são realizados, na grande maioria dos casos, em um pequeno grupo de redes.

Neste contexto, o objetivo deste estudo é avaliar o impacto da inclusão de dados sintéticos de colonoscopia no treinamento de modelos de segmentação variados. Para isso, foi adotada uma arquitetura U-Net [Ronneberger et al. 2015] como base, combinada com *encoders* de imagem que variaram entre dez arquiteturas distintas (incluindo diferentes tamanhos e famílias, como ResNet e *Transformer*), com o objetivo em analisar como se dá o benefício do uso de dados sintéticos através de diferentes capacidades computacionais e complexidades de modelo. As principais contribuições deste trabalho podem ser resumidas da seguinte forma:

- **Avaliação sistemática de *encoders*:** Condução de experimentos comparativos estruturados para mensurar o impacto da suplementação de dados sintéticos no treinamento de modelos de segmentação, variando o *backbone* entre dez arquiteturas distintas.
- **Análise de desempenho por arquitetura e escala:** Análise extensiva dos resultados, estabelecendo correlações entre o ganho de desempenho obtido com dados sintéticos e as características intrínsecas dos modelos, como a família da arquitetura e a complexidade paramétrica.

A organização do trabalho segue da seguinte forma: na Seção 2 serão apresentados os trabalhos relacionados; a metodologia utilizada para geração de dados, treinamento e avaliação, bem como breve descrição dos modelos e métricas utilizadas, são apresentadas na Seção 3; a Seção 4 contém os resultados obtidos e a análise entre os resultados de cada modelo de acordo com seu tamanho e arquitetura; por fim, serão apresentadas as conclusões e trabalhos futuros na Seção 5.

2. Trabalhos Relacionados

Trabalhos recentes têm demonstrado que o desempenho de modelos de DL em tarefas de visão computacional médica, incluindo a detecção de pólipos, está intrinsecamente ligado à disponibilidade e variabilidade dos dados de treinamento

[Castro et al. 2025]. No entanto, o avanço destas tecnologias enfrenta barreiras significativas devido à escassez de *datasets* públicos robustos, principalmente por conta de questões ligadas às preocupações com a privacidade dos dados dos pacientes [Marques et al. 2023].

2.1. Estratégias de Dados e Segmentação de Pólipos

Para mitigar a limitação de dados, De Araujo Jr. *et al.* [de Araujo Jr et al. 2024] investigaram o impacto da combinação de múltiplos *datasets* reais. O estudo demonstrou que o treinamento de modelos como o YOLOv8 com combinações de bases de dados heterogêneas (Kvasir-SEG, CVC-ClinicDB, etc.) melhora a generalização do modelo, evidenciando que a diversidade dos dados de entrada é preponderante para a eficácia da segmentação. Contudo, a simples agregação de dados reais nem sempre é viável ou suficiente para cobrir todos os cenários clínicos, motivando a exploração de dados sintéticos.

Modelos de segmentação clássicos e arquiteturas de detecção em tempo real estabeleceram-se inicialmente como o estado da arte para esta aplicação [Aguiar et al. 2024]. Entretanto, a avaliação de como dados sintéticos impactam o treinamento dessas arquiteturas específicas, variando desde *encoders* leves até *Transformers* complexos, ainda carece de uma análise sistemática e aprofundada na literatura.

2.2. Modelos Generativos em Imagens Médicas

A geração de imagens sintéticas emergiu como uma solução promissora para o aumento de dados. Inicialmente, as Redes Adversárias Generativas (GANs) dominaram este campo. O *framework* PolypConnect [Fagereng et al. 2022], por exemplo, utiliza técnicas de *inpainting* baseadas em GANs para inserir pólipos em imagens saudáveis. Posteriormente, foram apresentadas abordagens mistas, como o RePolyp [Pishva et al. 2023], que combina difusão com a inserção de partes sintéticas com uso de GANs. Apesar dos avanços, abordagens baseadas em GANs muitas vezes enfrentam desafios no equilíbrio entre fidelidade visual e diversidade, por vezes apresentando altos índices de similaridade estrutural (SSIM) que não se traduzem necessariamente em variabilidade útil para o treinamento.

Mais recentemente, métodos puramente baseados em Modelos de Difusão se apresentaram como estado da arte, oferecendo maior estabilidade e qualidade de textura. Neves *et al.* [Neves et al. 2025] propuseram uma metodologia comparativa avaliando modelos como Guided Diffusion [Dhariwal and Nichol 2021], Poisson Flow Generative Models [Xu et al. 2023] e Slicing Adversarial Networks (SAN) [Takida et al. 2024] para a geração de pólipos. Os resultados indicaram que o Improved Diffusion [Nichol and Dhariwal 2021] alcançou desempenho superior, obtendo a melhor pontuação na métrica Fréchet Inception Distance (33.89), denotando uma distribuição estatística próxima à dos dados reais, preservando características anatômicas críticas como coloração e textura.

Buscando avaliar o real impacto da adição de dados sintéticos no treinamento de segmentadores, Castro *et al.* [Castro et al. 2025], combinaram as capacidades generativas do Improved Diffusion com uma ferramenta de anotação semi-automática, com validação humana, para gerar um conjunto de pares imagem-máscara de segmentação, que foi adicionado ao treinamento de modelos segmentadores, apresentando uma melhora de desempenho em todos os casos. Embora represente um avanço para a concretização da

importância dos dados sintéticos para suprir a escassez de dados na área médica, a análise se limitou a apenas duas variações de *encoders* U-Net e três versões de segmentadores YOLOv11, fazendo com que seja necessária uma análise mais ampla e detalhada para melhor compreensão do real efeito destes dados aplicados a diferentes modelos.

3. Metodologia

Esta seção descreve o *pipeline* adotado para avaliar o impacto da adição de dados sintéticos no treinamento de diferentes arquiteturas de segmentação. O processo iniciou-se com a geração e curadoria de um conjunto de dados sintéticos de alta fidelidade, seguido pela definição dos modelos de aprendizagem profunda e, finalmente, pelo protocolo de treinamento e métricas de avaliação.

3.1. Seleção de Dados e Geração Sintética

Para compor o conjunto de dados reais, foram utilizadas as bases públicas **Kvasir-SEG** [Jha et al. 2020], **CVC-ClinicDB** [Bernal et al. 2015] e **ETIS-Larib** [Silva et al. 2014]. A união destas bases totaliza 1.808 imagens (1.000 do Kvasir-SEG, 612 do CVC-ClinicDB e 196 do ETIS-Larib).

Para garantir a reprodutibilidade e a integridade da avaliação, foi replicada a estratégia de particionamento de dados utilizada em [Castro et al. 2025], utilizando-se a metodologia de repartição via *Gaussian Mixture Models*, com o objetivo de prevenir vazamento de dados. A configuração final dos dados reais resultou em 1.446 imagens para treinamento e 362 imagens para teste. Para a geração dos dados sintéticos, adotou-se, também, o *pipeline* em estado da arte proposto em [Castro et al. 2025], como ilustrado pela Figura 1. Este método utiliza o modelo generativo *Improved Diffusion* [Nichol and Dhariwal 2021] combinado a um processo de anotação semiautomática (YOLO + SAM) com validação manual. Fixou-se a proporção de 75% de adição de dados sintéticos ao conjunto de treino real (aproximadamente 1.084 imagens sintéticas adicionadas às 1.446 reais), proporção esta identificada no estudo de referência como o ponto de estabilização do ganho de desempenho.

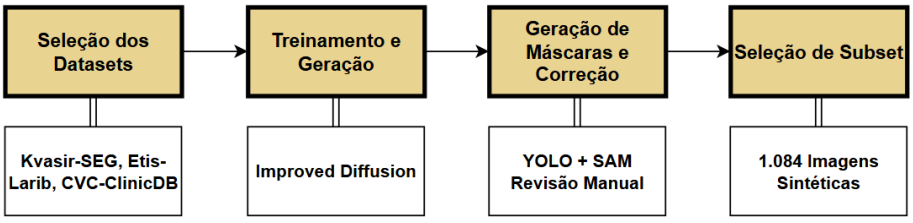

Figura 1. *Pipeline* para geração de imagens sintéticas. O procedimento compreende o treinamento do modelo Improved Diffusion em conjuntos de dados médicos consolidados, seguido por etapas de geração semi-automática de anotações com curadoria manual. Adaptado de [Castro et al. 2025].

3.2. Arquiteturas de Segmentação

Para investigar a generalização do impacto dos dados sintéticos, fixou-se a arquitetura U-Net [Ronneberger et al. 2015] como decodificador padrão, variando-se os *encoders* entre seis famílias distintas. Inicialmente, avaliou-se a família **ResNet** (ResNet34

e ResNet152) [He et al. 2016], baseada em conexões residuais, onde a versão 34 serve como um *baseline* leve e a 152 oferece alta capacidade para testar a propensão ao *overfitting*. Sua evolução, a **ResNeXt** (50_32x4d e 101_32x8d) [Xie et al. 2017], foi incluída por introduzir a cardinalidade como dimensão, equilibrando eficiência na versão 50 e maximizando a extração de *features* na versão 101. O grupo de CNNs clássicas completa-se com a **DenseNet121** [Huang et al. 2017], caracterizada pelo forte reaproveitamento de características (*feature reuse*), ideal para capturar padrões sutis.

O segundo grupo foca em eficiência e novos paradigmas. A família **EfficientNet** (B0 e B7) [Tan and Le 2019] utiliza escalonamento composto, permitindo comparar desde a versão B0 (extremamente leve) até a B7 (estado da arte em precisão). Para cenários de recursos restritos, adotou-se a **MobileNetV2** [Sandler et al. 2018], que emprega convoluções separáveis e blocos rasiduais invertidos. Por fim, incluíram-se os *encoders* **MiT** (SegFormer B0 e B4) [Xie et al. 2021], baseados em *Vision Transformers* hierárquicos, para contrastar o mecanismo de autoatenção global com as convoluções locais das arquiteturas anteriores.

3.3. Protocolo Experimental

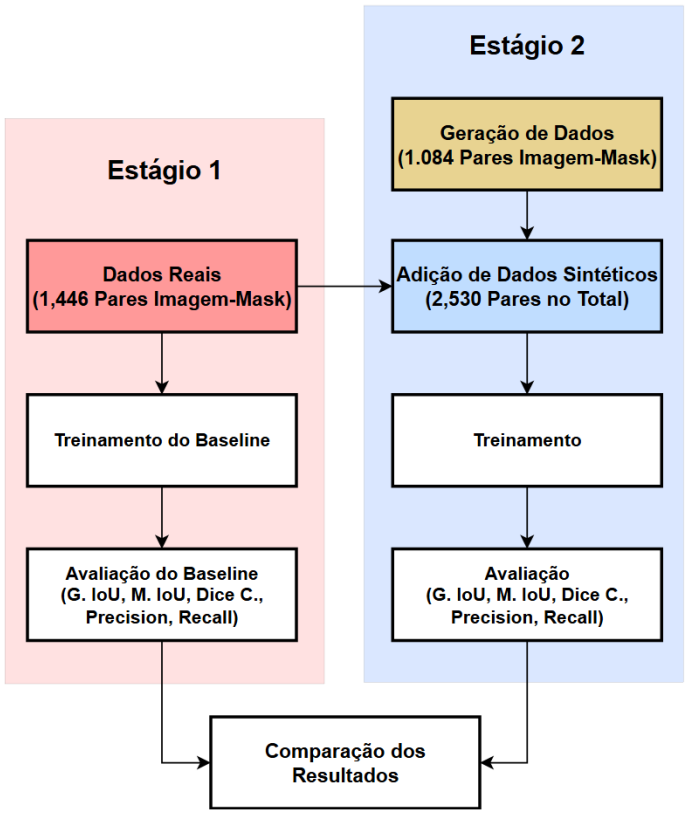

Figura 2. Fluxo experimental dividido em dois eságios. O Estágio 1 estabelece o desempenho de referência *(baseline)* utilizando dados reais, enquanto o Estágio 2 avalia o impacto da ampliação do conjunto de treinamento com imagens sintéticas geradas. Ambos os estágios seguem um protocolo padronizado de treinamento e avaliação para a síntese do desempenho.

O treinamento foi conduzido em duas etapas para cada *encoder*, conforme ilustrado pela Figura 2: (1) utilizando apenas dados reais e (2) utilizando dados reais acresci-

dos de 75% de dados sintéticos. Para garantir a confiabilidade estatística dos resultados, cada configuração foi submetida a 10 iterações de treinamento, partindo de sementes aleatórias distintas para a inicialização dos pesos. Cada iteração consistiu em um total de 50 épocas. Foi implementado um mecanismo de *early stopping* para evitar *overfitting* e economizar recursos. Para a avaliação, selecionou-se a época, de cada iteração, que apresentou o melhor valor de **Global IoU** no conjunto de validação. Foram registrados os resultados ao longo das 10 execuções para cada modelo.

3.4. Métricas de Avaliação

A avaliação seguiu as métricas padrão adotadas na literatura e no trabalho de referência [Castro et al. 2025].

- **Global IoU:** Calculado agregando as interseções e uniões de todo o conjunto de dados antes da divisão, oferecendo uma visão robusta do desempenho geral.
- **Mean IoU:** A média aritmética dos IoUs calculados individualmente para cada imagem, tratando cada amostra com igual peso.
- **Coeficiente Dice:** Mede a sobreposição harmônica entre a máscara predita e a máscara real, sendo sensível à precisão das bordas.
- **Precision (Precisão):** Indica a proporção de pixels classificados como pólipo que são realmente pólipos em relação aos que foram erroneamente classificados como pólipos.
- **Recall (Sensibilidade):** Indica a proporção de pixels de pólipos reais que foram corretamente identificados pelo modelo em relação aos que são pólipos, mas foram classificados como normais.

4. Resultados e Discussão

Para a condução dos experimentos, utilizou-se uma unidade de processamento gráfico (GPU) do modelo NVIDIA Tesla V100, integrada a um cluster de computação de alto desempenho. A implementação e o carregamento das arquiteturas de redes neurais, bem como a rotina de treinamento, foram realizados utilizando o framework *PyTorch*, garantindo a reprodutibilidade e a padronização dos hiperparâmetros experimentais.

A Tabela 1 apresenta a relação dos modelos utilizados como *encoders* na arquitetura U-Net, acompanhados de seus respectivos tamanhos (em milhões de parâmetros). A seleção abrange uma ampla variedade de configurações, desde modelos leves e otimizados para eficiência computacional, como o mit_b0 (5.55M) e o mobilenet_v2 (6.63M), até arquiteturas profundas e de alta capacidade paramétrica, como a resnext101_32x32d (475.49M).

Os resultados médios de desempenho obtidos ao longo das 10 iterações de treinamento utilizando exclusivamente o conjunto de dados reais são exibidos na Tabela 2 (à esquerda). O melhor resultado para cada métrica está destacado em negrito, enquanto o segundo melhor é destacado em itálico. Observa-se que modelos baseados em *Transformers* (MiT) e variações robustas de CNNs apresentaram os melhores resultados. O modelo mit_b4 destacou-se com o maior Coeficiente Dice (0.8297) e Global IoU (0.7980), sugerindo que a capacidade dos *Transformers* de capturar dependências globais é benéfica para a segmentação de pólipos. Por outro lado, arquiteturas mais leves, como o

Modelo	Parâmetros (M)
densenet121	13.61
efficientnet-b0	6.25
efficientnet-b7	67.10
mit_b0	5.55
mit_b4	64.12
mobilenet_v2	6.63
resnet152	67.16
resnet34	24.44
resnext50_32x4d	31.99
resnext101_32x32d	475.49

Tabela 1. Tamanho dos Encoders

efficientnet-b0 e o mobilenet_v2, apresentaram os desempenhos mais baixos, o que é esperado dada a menor capacidade de representação destes modelos quando treinados com uma quantidade limitada de dados reais.

	Apenas Dados Reais					**Com Adição de Dados Sintéticos**				
Modelo	Mean IoU	Global IoU	Dice	Precision	Recall	Mean IoU	Global IoU	Dice	Precision	Recall
densenet121	0.7416	0.7822	0.8115	**0.9395**	0.8237	0.7700	0.7939	0.8417	0.9318	0.8429
efficientnet-b0	0.7058	0.7714	0.7804	0.9141	0.8319	0.7522	0.7970	0.8238	0.9287	0.8491
efficientnet-b7	*0.7494*	*0.7921*	*0.8186*	0.9225	*0.8488*	**0.7837**	*0.8127*	**0.8524**	0.9309	*0.8649*
mit_b0	0.7217	0.7675	0.7971	0.9149	0.8267	0.7547	0.7889	0.8284	0.9252	0.8428
mit_b4	**0.7599**	**0.7980**	**0.8297**	*0.9284*	**0.8505**	*0.7817*	**0.8166**	*0.8521*	*0.9321*	**0.8683**
mobilenet_v2	0.7094	0.7490	0.7848	0.9157	0.8048	0.7543	0.7849	0.8261	0.9306	0.8337
resnet152	0.7173	0.7583	0.7909	0.9186	0.8129	0.7475	0.7789	0.8197	0.9241	0.8323
resnet34	0.7274	0.7674	0.8013	0.9213	0.8213	0.7559	0.7902	0.8298	0.9268	0.8428
resnext50_32x4d	0.7332	0.7764	0.8068	0.9271	0.8271	0.7643	0.7945	0.8369	0.9305	0.8447
resnext101_32x32d	0.7225	0.7631	0.7949	0.9159	0.8206	0.7598	0.7828	0.8329	**0.9331**	0.8293

Tabela 2. Comparação dos Resultados Médios: Treinamento Apenas com Dados Reais e Treinamento com Adição de Dados Sintéticos

Com a adição dos dados sintéticos ao conjunto de treinamento, observa-se um incremento consistente em quase todas as métricas para os modelos avaliados, conforme demonstrado na Tabela 2 (à direita). O modelo mit_b4 manteve sua posição de liderança, atingindo um Dice de 0.8521 e Global IoU de 0.8166. O modelo efficientnet-b7 também demonstrou desempenho notável, alcançando resultados muito próximos aos do mit_b4.

A Tabela 3 detalha a variação entre os resultados obtidos com e sem o uso de dados sintéticos. De modo geral, diversos modelos que apresentaram os menores desempenhos iniciais beneficiaram-se significativamente com a adição dos dados gerados. Especificamente, o mobilenet_v2 obteve o maior aumento no Global IoU (0.0358), seguido pelo efficientnet-b0, que alcançou o maior ganho no Coeficiente Dice (0.0434). Este fenômeno sugere que modelos mais compactos, que tipicamente sofrem mais com a falta de generalização em conjuntos de dados reduzidos, são capazes de extrair características mais robustas quando expostos à variabilidade adicional proporcionada pelos dados sintéticos.

Além das arquiteturas menores, o modelo de maior escala, resnext101_32x32d, também se destacou como um dos principais beneficiados. Este apresentou o maior ganho de precisão entre todos os modelos testados (0.0172) e uma melhoria substancial no

coeficiente Dice (0.0381).

Modelo	Δ Mean IoU	Δ Global IoU	Δ Dice	Δ Prec.	Δ Recall
densenet121	0.0283	0.0117	0.0302	-0.0077	0.0192
efficientnet-b0	**0.0464**	*0.0256*	**0.0434**	0.0146	0.0173
efficientnet-b7	0.0343	0.0206	0.0339	0.0083	0.0161
mit_b0	0.0329	0.0214	0.0313	0.0103	0.0160
mit_b4	0.0218	0.0185	0.0223	0.0037	0.0179
mobilenet_v2	*0.0449*	**0.0358**	*0.0413*	*0.0148*	**0.0289**
resnet152	0.0302	0.0206	0.0288	0.0055	0.0194
resnet34	0.0285	0.0228	0.0285	0.0055	*0.0215*
resnext50_32x4d	0.0311	0.0181	0.0300	0.0034	0.0176
resnext101_32x32d	0.0373	0.0196	0.0381	**0.0172**	0.0087

Tabela 3. Ganho de Desempenho com Adição dos Dados Sintéticos

A Figura 3 sintetiza visualmente o impacto da adição de dados sintéticos. As barras em vermelho representam o desempenho médio base (treinamento apenas com dados reais), enquanto as extensões em azul denotam o incremento absoluto proporcionado pela adição de 75% de dados sintéticos gerados pelo *pipeline* adotado.

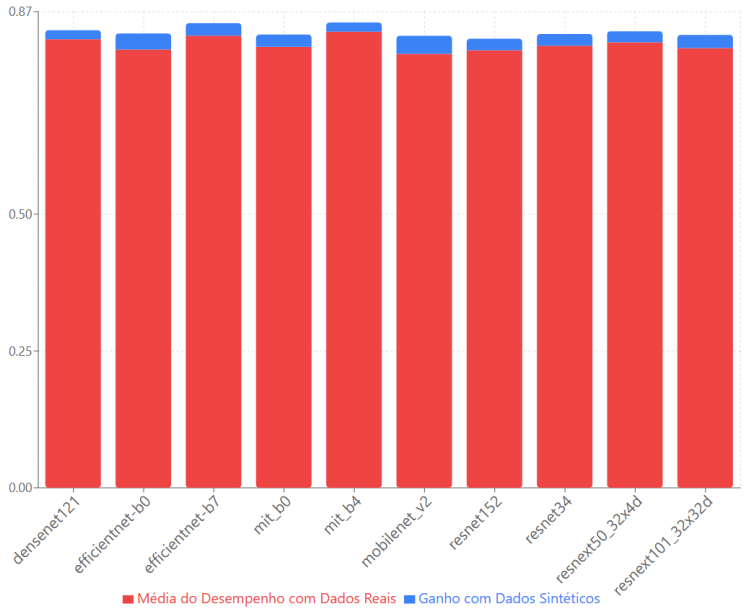

Figura 3. Desempenho médio base (vermelho) e ganho absoluto com dados sintéticos (azul) por modelo.

A análise permite identificar tendências claras relacionadas às famílias das arquiteturas. Nota-se que a família EfficientNet demonstrou uma receptividade superior à expansão do *dataset*, com ambas as versões exibindo ganhos expressivos. Em contrapartida, arquiteturas baseadas em CNNs mais tradicionais ou com forte reuso de *features*, especificamente as famílias ResNet e a DenseNet121, apresentaram incrementos comparativamente inferiores. A DenseNet121, apesar de ter um alto desempenho base, registrou o

menor ganho absoluto entre todos os modelos testados, indicando uma possível saturação na sua capacidade de extrair novas informações relevantes deste domínio.

Além das diferenças entre famílias, é possível notar discrepâncias intra-familiares significativas, corroborando a relação entre tamanho do modelo e ganho marginal. Este fenômeno é particularmente visível na família de Transformers, conforme exibido na Figura 4, que ilustra a relação entre o tamanho dos modelos (em escala logarítmica) e o incremento percentual de desempenho obtido com a adição de dados sintéticos. Enquanto a versão leve (mit_b0) obteve um aumento considerável de desempenho, a versão de alta capacidade (mit_b4) apresentou um incremento mais contido.

Ao analisar a distribuição geral dos pontos, nota-se um comportamento não linear, onde os maiores beneficiários da suplementação de dados tendem a se concentrar nas extremidades do espectro de complexidade.

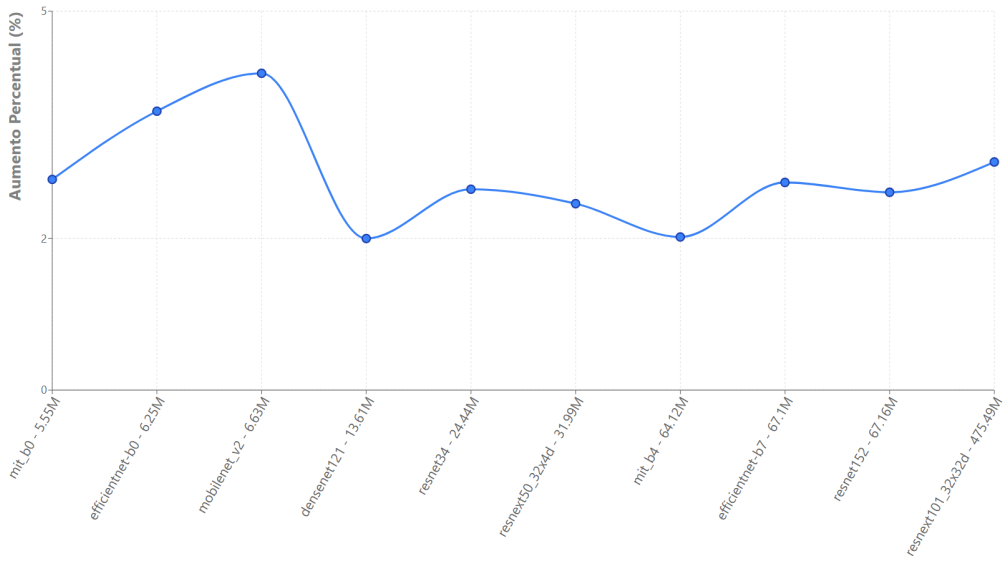

Figura 4. Relação entre o tamanho do modelo (milhões de parâmetros) e o incremento percentual de desempenho.

É possível notar que modelos compactos (MobileNetV2 e EfficientNet-B0) e modelos massivos (ResNeXt101) apresentam ganhos superiores em comparação aos modelos de porte intermediário (DenseNet121, ResNeXt50 e ResNet34). Este comportamento sugere que a adição de dados sintéticos atua por mecanismos distintos dependendo da capacidade do modelo.

Para modelos compactos, a escassez de dados reais pode levar a um aprendizado de características insuficientemente generalizáveis. A adição de dados sintéticos atua preenchendo lacunas no espaço de características, fornecendo variações que permitem a esses modelos convergir para soluções mais robustas, maximizando sua eficiência.

Já para os modelos massivos, por outro lado, sua capacidade paramétrica tende a exceder a complexidade necessária para se ajustar ao conjunto de dados reais limitado (1.446 imagens), tornando-os propensos ao *overfitting*. Neste caso, os dados sintéticos atuam como uma regularização, fornecendo o volume de dados necessário para justificar a profundidade da rede e forçando o modelo a aprender padrões mais invariantes.

Por fim, modelos intermediários, como a DenseNet121 e a ResNet34, demonstram ocupar um ponto de equilibrio entre complexidade do modelo e tamanho do dataset original, com capacidade suficiente para aprender bem sem subajuste, mas não tão excessiva que leve ao *overfitting*. Consequentemente, o ganho marginal proporcionado pelos dados sintéticos, embora positivo, é menor, pois o desempenho base já se encontra próximo do limiar de saturação para a tarefa.

5. Conclusões e Trabalhos Futuros

Este estudo demonstrou que a suplementação do treinamento com imagens sintéticas, geradas por modelos de difusão com anotação semiautomática, promoveu ganhos consistentes de desempenho em dez arquiteturas de segmentação de pólipos colorretais, validando esta abordagem como alternativa robusta à escassez de dados anotados no domínio médico [Castro et al. 2025, de Araujo Jr et al. 2024]. A análise granular revelou, contudo, que o impacto não é uniforme: modelos compactos e de grande porte foram os maiores beneficiários, os primeiros pela diversificação do aprendizado de características em redes com capacidade restrita, e os segundos pela prevenção do *overfitting*. Por outro lado, modelos intermediários apresentaram incrementos mais modestos, sugerindo saturação para o volume de dados atual.

Como direções para trabalhos futuros, destaca-se a necessidade de investigar o impacto da geração e segmentação em resoluções espaciais superiores. Além disso, sugere-se estender esta análise para além da arquitetura U-Net, explorando o comportamento de decodificadores alternativos e segmentadores baseados puramente em *Transformers* ou arquiteturas de detecção em tempo real. Por fim, uma análise estatística em relação aos resultados poderia ser conduzida.

Agradecimentos

Este trabalho foi parcialmente financiado pelo projeto Inteligência Artificial Aplicada na Detecção de Anomalias em Vídeos no Apoio à Tomada de Decisão, apoiado pelo Centro de Excelência em Inteligência Artificial (CEIA), Embrapii, ZSCAN e SEBRAE, com recursos financeiros do processo nº PEIA-2501.0109.

Referências

Aguiar, R. M. G., Scheeren, M. H., de Araujo Jr, S. L., Mendes, E., de Paula Filho, P. L., and Franco, R. A. P. (2024). Aplicação de modelos de aprendizado profundo para a segmentação semântica de imagens de colonoscopia. In *Anais do Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. SBC.

Alberti, L. R., Lima, D. C. A. D., Rodrigues, K. C. D. L., Taranto, M. P. L., Gonçalves, S. H. L., and Petroianu, A. (2012). The impact of colonoscopy for colorectal cancer screening. *Surgical Endoscopy*, 26:2308–2313.

Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., et al. (2015). Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43:99–111.

Castro, A. C., Naves, A. A., Neves, L. L., Fernandes, J. P., Santos, R., Oliveira, C., and Franco, R. A. P. (2025). A semi-automated pipeline for generating and annotating co-

lorectal polyp data for semantic segmentation tasks. In *2025 International Conference on Machine Learning and Applications (ICMLA)*. IEEE.

Corley, D. A., Jensen, C. D., Marks, A. R., et al. (2014). Adenoma detection rate and risk of colorectal cancer and death. *New England Journal of Medicine*, 370(14):1298–1306.

de Araujo Jr, S. L., Scheeren, M. H., Aguiar, R. M. G., Mendes, E., Franco, R. A. P., and de Paula Filho, P. L. (2024). Segmentação de pólipos em imagens de colonoscopia utilizando YOLOv8. In *Anais do Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. SBC.

de Melo, C. M., Torralba, A., Guibas, L., DiCarlo, J., Chellappa, R., and Hodgins, J. (2022). Next-generation deep learning based on simulators and synthetic data. *Trends in Cognitive Sciences*, 26:174–187.

Dhariwal, P. and Nichol, A. (2021). Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 8780–8794.

Fagereng, J. A., Thambawita, V., Storås, A. M., et al. (2022). Polypconnect: Image inpainting for generating realistic gastrointestinal tract images with polyps. In *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 66–71.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Heresbach, D., Barrioz, T., Lapalus, M. G., et al. (2008). Miss rate for colorectal neoplastic polyps: a prospective multicenter study of back-to-back video colonoscopies. *Endoscopy*, 40(4):284–290.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708.

Jha, D., Smedsrud, P. H., Riegler, M. A., et al. (2020). Kvasir-seg: A segmented polyp dataset. In *International Conference on Multimedia Modeling (MMM)*, pages 451–462. Springer.

Marques, A. F., Marques, K. F., dos Santos Beraldo, M. N. M., et al. (2023). Inteligência artificial na colonoscopia no rastreio do câncer colorretal: revisão de literatura. *Brazilian Journal of Health Review*, 6(4).

Neves, L. L., Castro, A. C., Naves, A. A., Paiva, H. S. G., Franco, R. A. P., and Cardoso, A. A. (2025). Methodology for generating medical images applied to the generation of synthetic colon polyps. In *2025 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE.

Nichol, A. and Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139, pages 8162–8171.

Picard, S., Chapdelaine, C., Cappi, C., et al. (2020). Ensuring dataset quality for machine learning certification. In *2020 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, pages 275–282.

Pishva, A. K., Thambawita, V., Torresen, J., and Hicks, S. A. (2023). Repolyp: A framework for generating realistic colon polyps with corresponding segmentation masks using diffusion models. In *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 47–52.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4510–4520.

Silva, J., Histace, A., Romain, O., Dray, X., and Granado, B. (2014). Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery*, 9(2):283–293.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249.

Takida, Y., Imaizumi, M., Shibuya, T., et al. (2024). San: Inducing metrizability of gan with discriminative normalized linear layer. In *The Twelfth International Conference on Learning Representations (ICLR)*.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 6105–6114.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 12077–12090.

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1492–1500.

Xu, Y., Liu, Z., Tian, Y., et al. (2023). Pfgm++: Unlocking the potential of physics-inspired generative models. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, volume 202, pages 38566–38591.