Avaliação da Generalização e do Fine-Tuning de Modelos Fundacionais e Redes Neurais Convolucionais na Segmentação de Células Cervicais

Lucas Monteiro Henriques[1], Bruna Luiza Martins Santos[2], Franciele Alves Barbosa[4], Mayara Kalita Moura Gomides[5], Mateus Amaral da Silva[2], Pedro Henrique Gonçalves Pires[2], Cauan Carvalho Marotta[3], Leonardo Augusto Ferreira[4], Marcelo Antonio Pascoal Xavier[7], Walmir Matos Caminhas[8], Frederico Guimarães[6], Andrea G. Campos[9]

[1] Engenharia de Controle e Automação
Universidade Federal de Minas Gerais(UFMG)
[2]Ciencia da computação – Universidade Federal de Minas Gerais (UFMG)
[3] Medicina – Universidade Federal de Minas Gerais (UFMG)
[4]Programa de Pós Graduação em Ciencia da computação
Universidade Federal de Minas Gerais (UFMG)
[5]Programa de Pós Graduação Física
Universidade Federal de Minas Gerais (UFMG)
[6]Departamento de Ciencia da computação
Universidade Federal de Minas Gerais (UFMG)
[7]Faculdade de Medicina Universidade Federal de Minas Gerais (UFMG)
[8]Departamento de Engenharia Eletrônica
Universidade Federal de Minas Gerais (UFMG)
[9]Departamento de Computação – Universidade Federal de Ouro Preto (UFOP)

lucasmonteirohenriques@yahoo.com.br, brustsmts@gmail.com

francieleab@ufmg.br,mgomides@ufmg.br

silvaamaralmateus@gmail.com, pedro.pires.edu@gmail.com

cauanmarotta@gmail.com, leauferreira@ufmg.br, mpascoal@ufmg.br

caminhas@ufmg.br, andrea@ufop.edu.br, fredericoguimaraes@ufmg.br

Abstract. *The ability to universally identify and segment objects is a cornerstone of contemporary Artificial Intelligence, driving sectors such as robotics and biological sciences. To streamline medical image analysis and reduce the need for specialized intervention, this study implemented computational workflows using the Cervical Cancer Dataset (CCD) and the Center for Recognition and Inspection of Cells (CRIC) pathological databases. The methodology involved state-of-the-art models, such as the Medical Segment Anything Model (MedSAM) and Cellpose integrated with the Segment Anything Model (SAM), alongside spatial clustering techniques for automated image cropping. These techniques enable extraction at both the single-cell scale and in multi-cell clusters, depending on the spatial distance between nuclei. The results demonstrate that integrating foundation models with specific fine-tuning allows for a*

rigorous evaluation of various architectures and the identification of the most efficient models for cervical cancer screening.

Resumo. *A capacidade de identificar e segmentar objetos de forma universal é um pilar da Inteligência Artificial contemporânea, impulsionando setores como a robótica e as ciências biológicas. Com o intuito de simplificar a análise de imagens médicas e reduzir a necessidade de intervenção especializada, este estudo implementou fluxos de trabalho nas bases de dados patológicas Cervical Cancer Dataset (CCD) e Center for Recognition and Inspection of Cells (CRIC). A metodologia incluiu o uso de modelos em estado-da-arte, como o Medical Segment Anything Model (MedSAM) e o Cellpose integrado ao Segment Anything Model (SAM), além de técnicas de agrupamento espacial para o recorte das imagens em escala de célula individual (single-cell) ou em grupos, dependendo da distância entre os núcleos. Os resultados demonstram que a integração de modelos fundacionais com ajustes específicos permite uma avaliação criteriosa de diferentes arquiteturas e a identificação dos modelos mais eficientes para o rastreio do câncer cervical.*

1. Introdução

O exame citopatológico do colo do útero é o principal método utilizado para o rastreamento e diagnóstico precoce do câncer do colo do útero, o terceiro tumor maligno mais incidente entre as mulheres no Brasil [Instituto Nacional de Câncer 2023]. Nesse contexto, a automação na interpretação desses exames tem se tornado uma realidade promissora para reduzir a carga de trabalho dos especialistas e mitigar a subjetividade humana. Entretanto, o sucesso de qualquer sistema de classificação automatizada depende criticamente de uma etapa inicial complexa: a segmentação precisa e o isolamento de células cervicais individuais [Lu et al. 2017].

Apesar dos avanços na visão computacional, a segmentação na prática clínica é severamente dificultada pela frequente sobreposição celular, presença de artefatos e variações de coloração nas lâminas [Lu et al. 2017]. Soluções tradicionais sofrem com a degradação de desempenho diante de novos dados, enquanto novas abordagens de inteligência artificial geral ainda carecem de validação empírica rigorosa neste nicho específico da citopatologia, especialmente em cenários de escassez de dados [Matta et al. 2024, Chen et al. 2026].

Para preencher essa lacuna, o objetivo principal deste trabalho é aplicar e avaliar modelos do estado da arte na segmentação automática de células em lâminas do exame citopatológico do colo do útero presentes nos *datasets Cell Recognition for Inspection of Cervix* (CRIC) [Rezende et al. 2021] e *Cervical Cell Detection* (CCD) [Liang et al. 2021]. Como desdobramentos deste objetivo, a pesquisa busca avaliar a capacidade de generalização desses modelos em cenários com múltiplas células agrupadas e variação de domínio, por meio de experimentos interdomínio. Tais experimentos utilizam o aprendizado por transferência (*transfer learning*) com e sem a aplicação de *fine-tuning*, técnica essencial para adaptar modelos pré-treinados a tarefas específicas com poucos dados [Pan and Yang 2009]. Consequentemente, o estudo visa realizar a extração de instâncias celulares individuais, incluindo a geração de *bounding boxes* (retângulos

delimitadores), que definem a localização espacial exata do objeto [Zhao et al. 2019] e o recorte automático, a fim de viabilizar etapas posteriores de classificação diagnóstica.

Para facilitar a compreensão do leitor, o restante deste artigo está organizado da seguinte forma. A Seção 2 apresenta a fundamentação teórica e discute os principais trabalhos relacionados, abordando desde arquiteturas convolucionais clássicas até modelos fundacionais e suas limitações no contexto da citopatologia. Na Seção 3, são detalhados os materiais e métodos adotados, incluindo as bases de dados utilizadas, o processo de curadoria, as estratégias de *fine-tuning*, o protocolo experimental e as métricas de avaliação. A Seção 4 expõe e analisa os resultados obtidos, com ênfase na capacidade de generalização dos modelos, no impacto do ajuste fino e na viabilidade da extração de instâncias. Por fim, a Seção 5 apresenta as conclusões do trabalho, destacando as principais contribuições, limitações e possíveis direções para pesquisas futuras.

2. Fundamentação Teórica e Trabalhos Relacionados

2.1. Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNNs) consolidaram-se como o estado da arte na detecção celular por sua capacidade de aprender padrões complexos, como bordas e texturas, transformando a segmentação em um problema de classificação *pixel* a *pixel* [Neha et al. 2025]. Um dos modelos mais influentes nessa vertente é a arquitetura *U-Net* [Ronneberger et al. 2015], que frequentemente serve como base (*backbone*) para abordagens avançadas como o *StarDist*. O *StarDist*, especificamente, otimiza a detecção em cenários de aglomeração celular ao prever polígonos estrela-convexos para cada núcleo, oferecendo uma representação morfológica superior às *bounding boxes* tradicionais [Schmidt et al. 2018].

Contudo, embora sejam altamente eficientes em distribuições controladas, essas arquiteturas puramente supervisionadas apresentam uma fragilidade fundamental diante do *domain shift*: a queda abrupta de desempenho quando aplicadas a imagens de diferentes *scanners*, laboratórios ou condições de iluminação [Matta et al. 2024]. Além disso, o treinamento de CNNs robustas exige vastos conjuntos de dados com anotações densas (máscaras *pixel* a *pixel*), um recurso escasso, de alto custo e sujeito a vieses de anotação na citopatologia [Lu et al. 2017].

2.2. Modelos Fundacionais e SAM

Os Modelos Fundacionais (*Foundation Models* – FMs) são arquiteturas de grande escala pré-treinadas em diversos conjuntos de dados por meio de estratégias de auto supervisão. Diferente das CNNs, os FMs adquirem conhecimentos prévios universais sobre a estrutura dos objetos, exibindo notável capacidade de transferência *zero-shot* e eliminando a necessidade de milhares de novos exemplos anotados [Chen et al. 2026]. O *Segment Anything Model* (SAM), desenvolvido pela equipe da FAIR (*Fundamental AI Research* / Meta AI), é um marco nesse cenário [Kirillov et al. 2023].

Utilizando uma arquitetura *Vision Transformer* (ViT) altamente escalável, o SAM realiza segmentação baseada em *prompts* (como pontos ou caixas). Seus fortes vieses indutivos conferem-lhe imunidade a vícios de anotação humana e o tornam consciente de ambiguidades na imagem [Pachitariu et al. 2025]. Modelos como o MedSAM

[Ma et al. 2024] adaptaram essa capacidade de generalização especificamente para o domínio médico.

Contudo, o SAM isolado apresenta limitações críticas na segmentação densa automatizada exigida na citopatologia, onde dezenas de células se sobrepõem simultaneamente. Para solucionar esse gargalo, abordagens híbridas passaram a integrar a robustez dos *transformers* a *frameworks* já consagrados no isolamento celular. O ecossistema do *Cellpose*, por exemplo, já havia demonstrado excepcional capacidade de segmentar morfologias complexas e aglomeradas por meio da previsão de campos de fluxo vetorial, como evidenciado em variantes como o *Omnipose* [Cutler et al. 2022].

A evolução natural dessa linha resultou no *Cellpose-SAM* [Pachitariu et al. 2025], uma arquitetura que une a imunidade ao *domain shift* do SAM com o rigor de processamento denso do *Cellpose*. Essa integração viabiliza a extração de instâncias celulares sobrepostas, configurando a solução ideal para lidar com situações inesperadas nas lâminas de triagem clínica.

2.3. Fine-tuning

Fine-tuning é uma técnica de aprendizado de máquina que adapta um modelo pré-treinado para ter melhor desempenho na sua tarefa específica [Microsoft 2023]. Em vez de treinar um modelo do zero, você começa com um modelo que já entende padrões gerais e o ajusta para funcionar com seus dados. Essa abordagem aproveita o aprendizado por transferência usando o conhecimento adquirido em uma tarefa para melhorar o desempenho em uma tarefa relacionada. O ajuste fino funciona bem quando você tem uma pequena quantidade de dados e quer melhorar o desempenho do seu modelo. Ao começar com um modelo pré-treinado, você pode usar o conhecimento que o modelo já adquiriu e ajustá-lo para melhor se ajustar aos seus dados. Essa abordagem ajuda a melhorar o desempenho do seu modelo e reduz a quantidade de dados necessária para o treinamento.

2.4. Trabalhos Relacionados

Diversos estudos têm explorado soluções computacionais para a segmentação de células cervicais. Em [Lu et al. 2017], os autores avaliaram o desempenho de diferentes algoritmos tradicionais na segmentação de células cervicais sobrepostas, destacando a dificuldade de extrair limites precisos devido ao baixo contraste e à aglomeração celular. Mais recentemente, com a ascensão do aprendizado profundo (*deep learning*), a literatura tem convergido para o uso de modelos fundacionais. No trabalho de [Ma et al. 2024], foi proposto o MedSAM, demonstrando sua eficácia em uma ampla gama de modalidades de imagens médicas; no entanto, a aplicação do modelo dependeu de marcações manuais (*prompts*) e não foi otimizada para cenários autônomos de alta densidade celular.

Para lidar especificamente com morfologias complexas e aglomeradas, o Cellpose-SAM foi desenvolvido em [Pachitariu et al. 2025], alcançando resultados próximos ao consenso humano em imagens biológicas genéricas. Apesar desses avanços contínuos, a transposição direta e a literatura indica que a transposição direta desses modelos para o domínio da citopatologia cervical ainda enfrenta desafios de generalização, causados pela alta variabilidade nas colorações do esfregaço cervicovaginal e pelo *domain shift* inerente a diferentes equipamentos de captura [Matta et al. 2024]."

3. Materiais e Métodos

A estratégia adotada neste trabalho consiste em identificar modelos com elevada capacidade de generalização, capazes de alcançar altas taxas de acerto em domínios distintos. O objetivo final do fluxo de processamento é recortar, segmentar e catalogar imagens complexas do *dataset* CCD (*Cervical Cell Detection*) sem a necessidade de um treinamento completo e exaustivo. Para viabilizar a triagem clínica autônoma nesse contexto, a importância do *fine-tuning* (detalhado na Seção 2.3) torna-se evidente: ele permite que o sistema herde capacidades de segmentação de modelos robustos e as especialize para a morfologia celular específica do CCD, garantindo precisão mesmo com amostras limitadas. Assim, a metodologia foi estruturada nas etapas de curadoria de dados, configuração experimental e avaliação de desempenho.

3.1. Bases de Dados

Foram utilizados dois repositórios de imagens citopatológicas para conduzir os experimentos interdomínio:

- ***Dataset* CRIC (*Cell Recognition for Inspection of Cervix*):** desenvolvido por pesquisadores de instituições brasileiras e do *Berkeley Institute for Data Science* (BIDS), este repositório atuou como o domínio de calibração do estudo [Rezende et al. 2021]. A escolha do CRIC justifica-se por ser composto por 400 imagens do exame citopatológico do colo do útero convencional e fornece anotações espaciais, representadas exclusivamente pelos centróides (pontos centrais x, y) de cada célula, contemplando as classes: ASC-US, LSIL, HSIL, ASC-H e SCC.
- ***Dataset* CCD (*Cervical Cell Detection*):** introduzido em [Liang et al. 2021], constitui o domínio alvo da aplicação. A base possui 7.410 imagens complexas, sem segmentação densa prévia.Este estudo foca no mapeamento e isolamento das cinco classes de células escamosas compatíveis com o CRIC, utilizando o CCD como o conjunto de teste independente para avaliar a robustez interdomínio.

3.2. Curadoria e *Fine-Tuning*

Como o CRIC fornece apenas centróides, realizou-se a segmentação manual densa de 45 imagens na plataforma *Roboflow*, com o auxílio da *API* do modelo SAM, a fim de permitir a validação morfológica das máscaras. A escolha de um conjunto reduzido justifica-se pela natureza das arquiteturas avaliadas. Modelos fundacionais possuem representações universais consolidadas; assim, o ajuste fino (*fine-tuning*) teve como escopo apenas a adaptação ao domínio da coloração cervical (*few-shot learning*), minimizando a demanda por mão de obra especializada.

Para a avaliação de desempenho, adotou-se uma estratégia de validação do tipo *hold-out*. O conjunto de 45 imagens com anotações densas foi utilizado exclusivamente para o ajuste fino dos modelos. As 355 imagens restantes, não utilizadas no treinamento, foram empregadas como conjunto de teste independente, permitindo avaliar a capacidade de generalização dos modelos em dados não vistos. Os resultados desse experimento podem ser observados na Tabela 1.

3.3. Protocolo Experimental e Modelos Avaliados

Foram avaliadas diferentes arquiteturas submetidas a cenários com e sem *fine-tuning*:

1. ***Cellpose-SAM*:** Realizou-se uma busca em grade (*Grid Search*) nos parâmetros *flow threshold* (limiar de erro de fluxo) e *cell probability threshold* (limiar de probabilidade).
    - ***Flow Threshold*:** Controla o rigor da máscara. Valores altos permitem máscaras mais complexas, enquanto valores baixos evitam segmentações espúrias.
    - ***Cell Probability Threshold*:** Define a sensibilidade de detecção. O valor de $-0,0$ foi adotado para maximizar o *Recall*, indicando que qualquer pixel com probabilidade neutra (após a normalização logit do modelo) é considerado candidato a núcleo, minimizando Falsos Negativos em detrimento de uma Precisão mais conservadora.
2. ***MedSAM*:** avaliado em dois cenários: (i) utilizando *prompts* das coordenadas exatas do CRIC, simulando a orientação de um especialista; e (ii) utilizando *prompts* aleatórios, para quantificar a robustez do decodificador frente a ruídos de posicionamento [Ma et al. 2024].
3. ***StarDist*:** Esta arquitetura baseia-se na predição de polígonos estrela-convexos para localizar instâncias celulares, sendo particularmente eficiente para núcleos densamente agrupados [Schmidt et al. 2018]. Diferente de modelos que utilizam máscaras de *pixels* independentes, o *StarDist* prediz, para cada *pixel*, a probabilidade de pertencer a um objeto e um conjunto de distâncias ($radii$) até a fronteira do polígono.

3.4. Métricas de Avaliação Espacial

A validação baseou-se no cruzamento espacial entre as máscaras preditas e o *Ground Truth* (verdade fundamental). O critério de acerto define:

- **Verdadeiro Positivo (TP):** o centróide anotado está contido no interior da máscara gerada.
- **Falso Negativo (FN):** o centróide anotado não é englobado por nenhuma máscara.
- **Falso Positivo (FP):** o modelo gerou uma máscara em uma região sem centróide correspondente.

A partir dessas definições, calcularam-se a Precisão, o *Recall* e o *F1-Score*. Adicionalmente, adotou-se o *Error Rate* (ER) para quantificar o erro total de detecção em relação ao número de células esperadas, conforme a Equação 1:

$$ER = \frac{FP + FN}{TP + FN} \tag{1}$$

A rejeição da fórmula de erro tradicional baseada na classificação binária de pixels em favor da métrica da Equação 1 justifica-se pelo fenômeno da diluição estatística pelo fundo (*background*). Em imagens de microscopia, o número de pixels vazios é vastamente superior ao número de pixels celulares; assim, uma métrica tradicional que inclua Verdadeiros Negativos (TN) no denominador apresentaria um erro artificialmente baixo, mesmo que o modelo falhasse em segmentar células individuais. Ao utilizar a fórmula centrada em objetos, isola-se a capacidade do sistema de discernir unidades discretas. Diferente do

erro tradicional, esta métrica penaliza erros topológicos críticos, como a subsegmentação (fusão de células), garantindo que a precisão reportada reflita a contagem real de células e não apenas a ocupação de área na imagem, o que é essencial para a validade de análises diagnósticas e quantitativas.

3.5. Extração Espacial e Geração de *Bounding Boxes*

O fluxo de processamento proposto prevê que as máscaras geradas na etapa de inferência sejam submetidas a algoritmos de *clustering* espacial, como o DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*). O objetivo desta etapa é identificar a densidade das instâncias segmentadas para determinar a estratégia de recorte: caso a distância entre os núcleos seja reduzida, o sistema agrupa múltiplas células em uma única região de interesse; caso contrário, realiza o isolamento individual (*single-cell*). A partir do mapeamento desses agrupamentos, o sistema gera automaticamente *bounding boxes* para a execução do recorte (*cropping*) das imagens originais. Este procedimento visa garantir que os modelos de classificação subsequentes processem apenas as informações morfológicas pertinentes seja de uma célula isolada ou de um pequeno grupo de células sobrepostas minimizando a influência de ruídos de fundo e garantindo a integridade dos dados para a triagem diagnóstica.

4. Resultados e Discussão

Para avaliar a capacidade de generalização e adaptação das arquiteturas propostas, os modelos foram submetidos a testes rigorosos no domínio do *dataset* CRIC, alternando entre cenários *zero-shot* (sem *fine-tuning*) e adaptados com *fine-tuning* de 45 imagens. Na Tabela 1 são sintetizadas as métricas globais obtidas para cada configuração espacial.

Tabela 1. Desempenho comparativo dos modelos na segmentação espacial.

Modelo	Cenário	Precisão	Recall	F1-Score	ER
Cellpose-SAM	Zero-shot	0,301	0,887	0,401	2,172
Cellpose-SAM	Fine-tuning	**0,857**	**0,850**	**0,843**	**0,292**
StarDist	Zero-shot	0,134	0,011	0,020	1,060
StarDist	Fine-tuning	**0,657**	**0,838**	**0,715**	**0,599**
MedSAM	Zero-shot	0,929	0,941	0,935	0,131
MedSAM	Fine-tuning	**0,973**	**0,984**	**0,978**	**0,043**
MedSAM	Autônomo	0,008	0,047	0,013	6,781

4.1. Avaliação de Generalização *Zero-Shot*

Considerando a Tabela 1, no cenário sem ajuste prévio, os resultados revelam disparidades arquiteturais significativas. O modelo *StarDist* sofreu uma degradação severa, com um índice de *Recall* de 0,011, o que indica uma dificuldade acentuada na identificação de células em domínios distintos daqueles utilizados em seu treinamento original.

Em contrapartida, o *Cellpose-SAM* demonstrou um alto *Recall* (0,887), evidenciando uma elevada capacidade de identificar células mesmo sem ajuste específico ao domínio das imagens cervicais. A baixa precisão (0,301) observada não reflete, necessariamente, uma falha de detecção; pelo contrário, ocorre porque o modelo identifica corretamente diversas células que não constam nas anotações do conjunto de dados CRIC, sendo estas contabilizadas erroneamente como Falsos Positivos.

O *MedSAM* apresentou um excelente desempenho quando auxiliado por *prompts* ideais, alcançando um *F1-Score* de 0,935. Tal resultado confirma sua robustez na delimitação de bordas quando o objeto de interesse é previamente sinalizado. Contudo, no cenário autônomo, sem o auxílio de pontos de guia, o desempenho apresenta uma queda vertiginosa para um *F1-Score* de 0,013. Esta discrepância evidencia que o *MedSAM*, embora seja um segmentador poderoso, carece de um mecanismo interno de detecção eficaz para a citopatologia cervical, dependendo estritamente de informações de orientação para localizar os alvos de interesse.

4.2. Impacto do *Fine-Tuning* e Análise Comparativa

Após o *fine-tuning*, houve um aumento expressivo na precisão do *Cellpose-SAM*, que saltou para 0,857. Esse resultado indica que o ajuste ao domínio permitiu um maior alinhamento entre as detecções realizadas e as instâncias efetivamente anotadas.

Adicionalmente, a análise do *Error Rate* (ER) revela nuances importantes sobre a eficácia da adaptação. Embora o ER obtido para o *Cellpose-SAM* com ajuste fino (0,292) seja numericamente superior ao reportado no trabalho original de [Pachitariu et al. 2025], tal diferença não implica inferioridade técnica, mas sim variações de protocolo e volume de dados. Enquanto o estudo original utilizou 67 imagens, este trabalho operou em um regime de *few-shot learning* com apenas 45 amostras. Somado a isso, a natureza não exaustiva do *dataset* CRIC penaliza o modelo com falsos positivos ao segmentar corretamente células saudáveis não anotadas, elevando artificialmente o ER. É imperativo destacar que o desempenho obtido situa-se em um patamar de erro comparável ao de um anotador humano secundário, validando a aplicabilidade clínica da arquitetura proposta.

No que diz respeito ao *StarDist*, houve uma evolução significativa após o *fine-tuning*, com o F1-Score saltando de 0,020 para 0,715 e o ER reduzindo de 1,060 para 0,599. Esse ganho evidencia a forte dependência deste modelo em relação ao ajuste de domínio.

Por fim, o *MedSAM* destaca-se com as métricas mais elevadas da Tabela 1, atingindo um F1-Score de 0,978 e um ER de apenas 0,043. Este desempenho superior justifica-se pelo uso de *prompts* ideais (coordenadas exatas), o que elimina a necessidade de o modelo realizar a detecção autônoma das instâncias. Dessa forma, o *MedSAM* opera estritamente como um segmentador guiado de alta precisão, dependendo de informações prévias de localização para alcançar tais patamares de assertividade.

4.3. Extração de *Bounding Boxes* (*Single-Cell*)

A viabilidade de um sistema autônomo de triagem depende da capacidade do modelo em isolar instâncias de forma independente e precisa. Embora o *MedSAM* tenha apresentado estatisticamente o menor índice de erro ($ER = 0,043$ com *fine-tuning*), este resultado é estritamente condicionado ao uso dos centróides do *dataset* CRIC como *prompts* manuais, o que desqualifica sua aplicação em fluxos automatizados. Em contraste, o *Cellpose-SAM* consolidou-se como a arquitetura superior, superando os demais concorrentes em cenários de autonomia. Enquanto modelos como o *StarDist* apresentaram falhas críticas de generalização e o *MedSAM* falhou na ausência de guias, o *Cellpose-SAM* manteve uma detecção exaustiva.

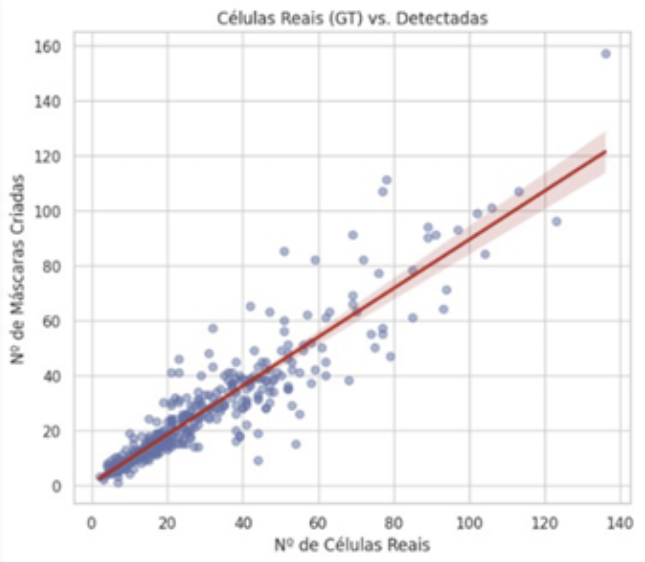

Figura 1. Distribuição de máscaras preditas versus células reais para o *Cellpose-SAM* com *fine-tuning*.

Conforme fundamentado por [Pachitariu et al. 2025], a superioridade dessa arquitetura reside na combinação dos campos de fluxo vetorial do *Cellpose* com os vieses indutivos do SAM, permitindo a localização de núcleos em lâminas complexas sem a necessidade de auxílio externo. Devido a essa característica espacial, o *Cellpose-SAM* qualifica-se como uma solução ideal para a geração automática de caixas delimitadoras (*bounding boxes*), conforme exemplificado na Figura 3, garantindo que os recortes (*crops*) para classificação diagnóstica sejam extraídos com máxima integridade.

Tamanho da imagem (pixels)	Tamanho do arquivo	VRAM (GPU)	RAM (CPU)	Batch Size	Tempo (s)
150	270KB	2,45 GB	2,18 GB	1	0,37
300	1,08MB	2,45 GB	2,18 GB	4	0,41
600	4,32MB	3,39 GB	2,23 GB	9	0,87
1.200	17,2MB	8,90 GB	3,57 GB	32	3,24
2.400	69MB	8,90 GB	3,84 GB	32	12,48
4.800	276MB	8,90 GB	4,80 GB	32	46,71
9.600	1,11GB	12,28 GB	18,40 GB	32	368,88

Tabela 2. Consumo de recursos e tempo de execução por tamanho de imagem. Fonte: adaptado de [Pachitariu et al. 2025].

4.4. Custo computacional

A análise da eficiência computacional revela distinções significativas entre as arquiteturas avaliadas. O *StarDist* [Schmidt et al. 2018], por sua natureza baseada em regressão de polígonos convexos sobre uma *U-Net* simplificada, apresenta o menor custo operacional, sendo ideal para aplicações de alta taxa de transferência em núcleos celulares. Em contrapartida, o *Cellpose* [Pachitariu et al. 2025] demonstra uma demanda crescente de

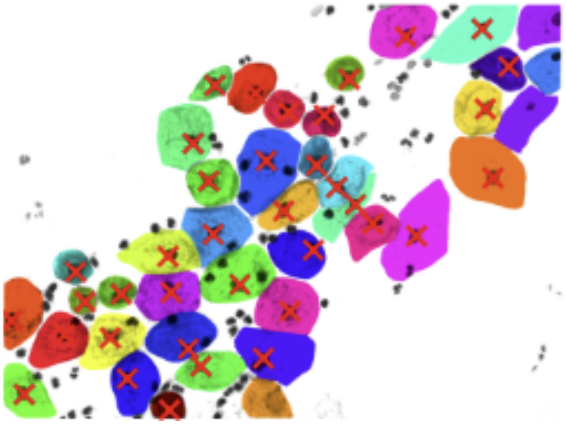

Figura 2. Previsão de máscaras de segmentação pelo modelo *Cellpose-SAM*, onde os marcadores (X) representam as anotações do *Ground Truth*.

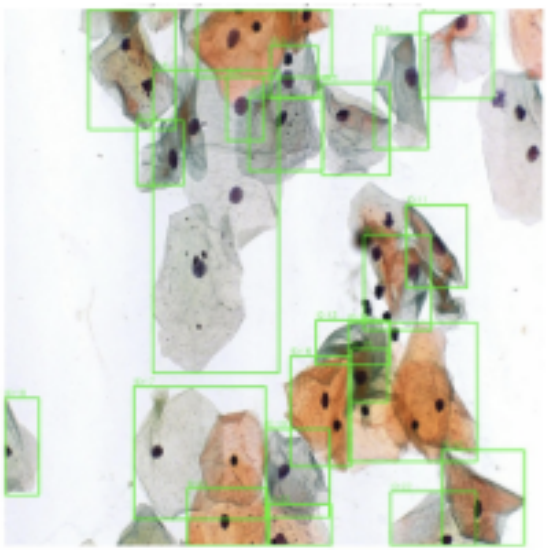

Figura 3. Processo de geração automática de bounding boxes a partir das máscaras segmentadas, permitindo o recorte individual (cropping) das células para posterior classificação em sistemas de auxílio ao diagnóstico.

recursos conforme a resolução da imagem escala; conforme observado nos dados experimentais Tabela 2, resoluções de 9.600 pixels podem elevar o tempo de inferência para $368,88$ segundos, exigindo até $18,40$ GB de RAM, o que pode inviabilizar o uso de GPUs de consumo padrão. O *MedSAM* [Ma et al. 2024] posiciona-se como uma alternativa de alto desempenho que, embora utilize um *backbone* robusto de *Vision Transformer* (ViT-Base) com cerca de 94 milhões de parâmetros, mitiga o escalonamento explosivo de memória ao padronizar a entrada em 1024×1024 pixels [Microsoft 2023]. Essa característica garante uma latência previsível e um uso de VRAM estável, fatores cruciais para a implementação de um fluxo de triagem clínica autônoma no *dataset* CCD.

5. Conclusão

Este trabalho propôs a avaliação de modelos de última geração para a segmentação autônoma de células cervicais, preenchendo uma lacuna na literatura quanto à viabili-

dade de modelos fundacionais em cenários clínicos sem intervenção humana. A principal contribuição desta pesquisa reside na demonstração de que a integração dos campos de fluxo vetorial do *Cellpose* com os vieses indutivos do *SAM* (*Cellpose-SAM*) supera arquiteturas consolidadas como o *StarDist*, especialmente em termos de generalização e economia de mão de obra especializada. Além disso, nossa abordagem mostrou-se robusta com um *fine-tuning* de apenas 45 imagens, facilitando a adoção tecnológica em laboratórios com recursos limitados de anotação médica.

A eficácia desta proposta é evidenciada na Figura 1, que apresenta a análise comparativa entre a quantidade de células reais e as máscaras geradas automaticamente no *dataset* CRIC. A linha de regressão linear demonstra uma forte correlação positiva, comprovando que o modelo mantém uma alta taxa de acerto na identificação de instâncias, mesmo em lâminas com alta densidade celular. Observa-se que a delimitação precisa ocorre inclusive em *clusters* densos; contudo, os *outliers* presentes em imagens com mais de 100 células indicam que a sobreposição extrema ainda representa o limite atual da arquitetura, onde o deslocamento do centroide entre núcleo e citoplasma pode impactar as métricas de distância.

Adicionalmente, a análise das Figuras 1, 2 e 3 permite elucidar de forma clara o fluxo de processamento proposto. Inicialmente, a etapa de identificação é responsável pela localização precisa dos núcleos celulares, conforme ilustrado na 1. Em seguida, o modelo *Cellpose-SAM*, na etapa de segmentação, realiza a delimitação das células previamente identificadas, como apresentado na 2. Posteriormente, o algoritmo de clusterização espacial organiza essas detecções e gera automaticamente retângulos delimitadores (*bounding boxes*) com margens de segurança adaptativas, conforme observado na 3. Esse mecanismo assegura a extração íntegra das células de interesse, reduzindo o risco de truncamento de estruturas relevantes e preservando características morfológicas essenciais. Como resultado, obtêm-se amostras com maior consistência estrutural e qualidade informacional, tornando-as mais adequadas para etapas subsequentes de classificação em sistemas de apoio ao diagnóstico citopatológico.

Em conclusão, os resultados permitem recomendar o *Cellpose-SAM*, mesmo apresentando um custo computacional mais elevado, tanto em termos de uso de memória quanto de tempo de processamento, conforme discutido na Seção 4.4. Ainda assim, o modelo permanece como a opção mais vantajosa para fluxos de trabalho totalmente autônomos, dada sua capacidade superior de identificar ”o que é uma célula”sem auxílio externo. Ressalta-se, contudo, que, caso o cenário permita o uso de *prompts* manuais ou pontos de referência (cenário semi-autônomo), modelos como o *MedSAM* tornam-se preferíveis devido à sua precisão quando guiados. Para a triagem clínica em larga escala, onde a automação total é o objetivo, o *Cellpose-SAM* estabelece-se como o novo padrão de aplicabilidade.

Agradecimentos

Este estudo foi financiado com recursos do Centro de Inovação e Inteligência Artificial para Saúde (CI-IA Saúde), em parte com recursos da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) Processo nº 2020/09866-4, da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) Processo nº PPE-00030-21 e da UNIMED Belo Horizonte.

Referências

Chen, Y. et al. (2026). Foundation models in medical imaging: A review. *EngMedicine*, 3(2).

Cutler, K. J. et al. (2022). Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. *Nature Methods*, 19(11):1438–1448.

Instituto Nacional de Câncer (2023). Dados e números sobre câncer do colo do útero. Acesso em: 23 fev. 2026.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, C.-Y., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4015–4026.

Liang, Y. et al. (2021). Global context-aware cervical cell detection with soft scale anchor matching. *Computer Methods and Programs in Biomedicine*, 204.

Lu, Z. et al. (2017). Evaluation of three algorithms for the segmentation of overlapping cervical cells. *IEEE Journal of Biomedical and Health Informatics*, 21(2):441–450.

Ma, J. et al. (2024). Segment anything in medical images (medsam). *Nature Communications*, 15:654.

Matta, R. et al. (2024). A systematic review of generalization research in medical image classification. *Computers in Biology and Medicine*, 183.

Microsoft (2023). Fine-tuning AI models. *Microsoft Learn*. Acesso em: 29 abr. 2026. Fine-tuning helps you adapt pre-trained AI models to work better with your specific data.

Neha, F. et al. (2025). An analytics-driven review of u-net for medical image segmentation. *Healthcare Analytics*, 8.

Pachitariu, M., Stringer, C., and Rariden, M. (2025). Cellpose-sam: superhuman generalization for cellular segmentation. *BioRxiv: The preprint server for biology*.

Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

Rezende, M. T. et al. (2021). Cric searchable image database as a public platform for conventional pap smear cytology data. *Scientific Data*, 8(151).

Roboflow (2026). Roboflow: Computer vision platform. Disponível em: https://roboflow.com/. Acesso em: 27 fev. 2026.

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, Cham.

Schmidt, U. et al. (2018). Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 265–273. Springer, Cham.

Zhao, Z.-Q., Zheng, P., Xu, S.-t., and Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11):3212–3232.