

# Fine-Tuning do Segment Anything Model

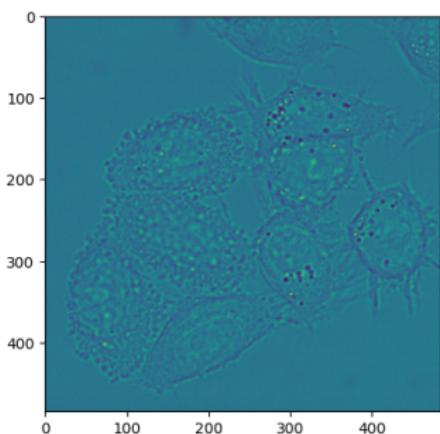
## Segmentação de Núcleos Celulares

Antonio J. Brych, Gabriel R. Abad, Tomás P. Lira

November 23, 2025

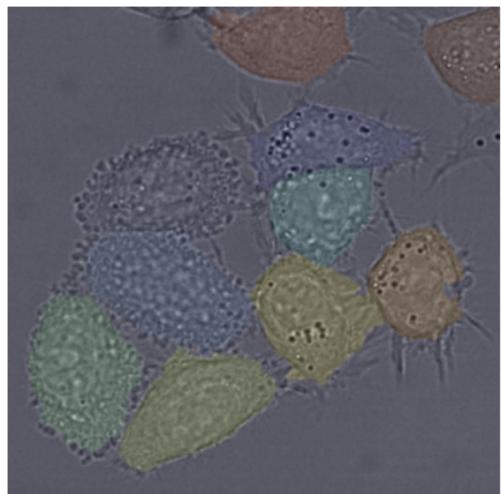
# Panorama e Motivação

- **Dataset-base:** ISBI Cell Tracking Challenge serve como referência histórica da área.
- **Bioinformática:** segmentação em microbiologia já consolidou métricas, protocolos e expectativas.
- **Pergunta-chave:** observar como o SAM, mesmo generalista, responde a um escopo super restrito.
- **Comparativo:** posicionar o SAM fine-tuned frente a modelos clássicos e ao benchmark ISBI.



# Tarefa Selecionada e Dataset

- **Base:** ISBI Cell Tracking Challenge (Fluo-N2DL-HeLa, 92 frames).
- **Amostra:** células HeLa aderidas, DNA marcado por fluorescência.
- **Aquisição:** Dr. G. van Cappellen (Erasmus MC), 512×512 px, microscopia widefield.
- **Interesse:** benchmark nuclear para contagem mitótica/triagem de drogas.
- **Rótulos:** máscaras manuais do ISBI divididas em treino/validação.



# Augmentation e Pré-Processamento

*“Data augmentation is essential to teach the network the desired invariance and robustness properties, when only few training samples are available. In case of microscopical images we primarily need shift and rotation invariance as well as robustness to deformations and gray value variations. Especially random elastic deformations of the training samples seem to be the key concept to train a segmentation network with very few annotated images.”*

— Ronneberger et al., *U-Net: Convolutional Networks for Biomedical Image Segmentation* (2015)

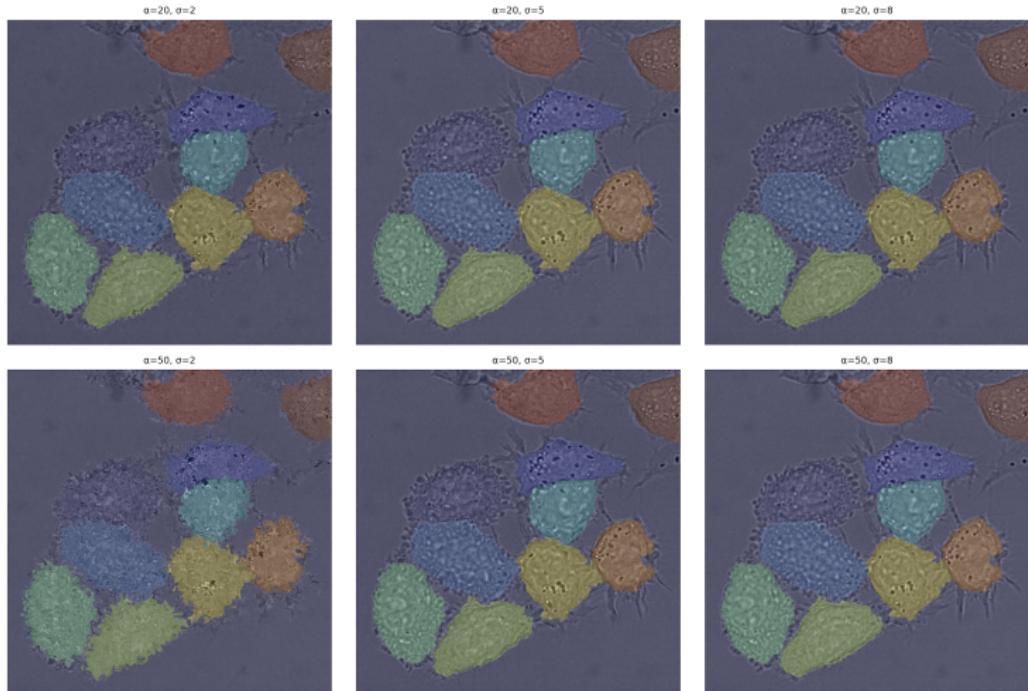
# Augmentation e Pré-Processamento

- **Invariâncias geométricas:** rotações  $\pm 45^\circ$ , flips e shifts subpixel.
- **Elasticidade:** deformações suaves ( $\sigma = 10$ ,  $\alpha \approx 30$ ) sincronizadas com as máscaras.
- **Fotometria:** jitter de intensidade, ruído Poisson/Gauss e normalize/local equalize.

# Elasticidade — Implementação

- ① **Gerar ruído:** sorteio de  $n(x) \sim \mathcal{N}(0, 1)^2$  com a mesma semente para imagem e máscara.
- ② **Suavizar:** convoluir  $n$  com  $G_\sigma$  ( $\sigma = 10$  px) para obter  $v = G_\sigma * n$  e remover dobras.
- ③ **Escalar:** amostrar  $\alpha \sim \mathcal{U}[25, 35]$  px e calcular o deslocamento  $u = \alpha v$ .
- ④ **Warp final:** usar `grid_sample(img, u)` e `grid_sample(mask, u)` com interpolação bilinear e borda refletida.

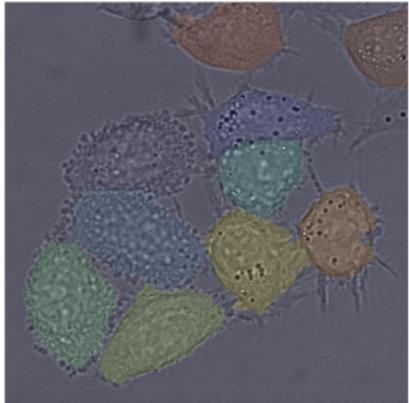
# Augmentation: Elasticidade e Deformação



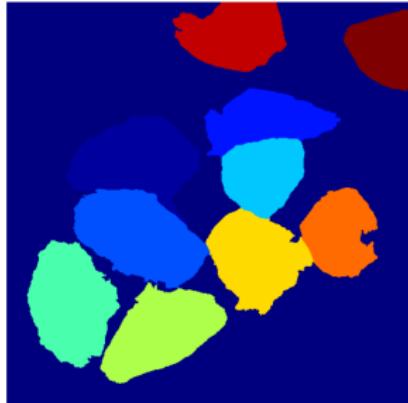
Efeitos dos parâmetros de deformação.

# Augmentation: Resultado do Pipeline

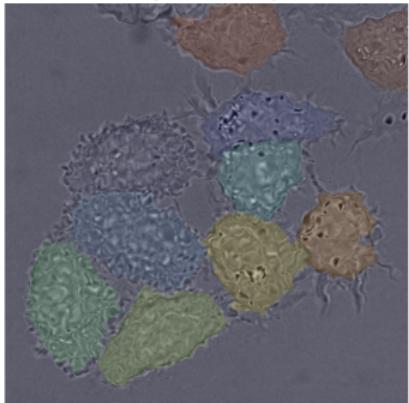
Original Overlay



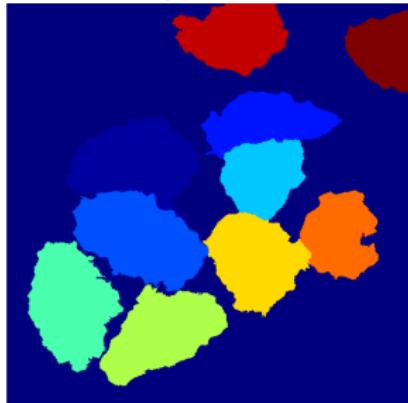
Original Mask



Augmented Overlay

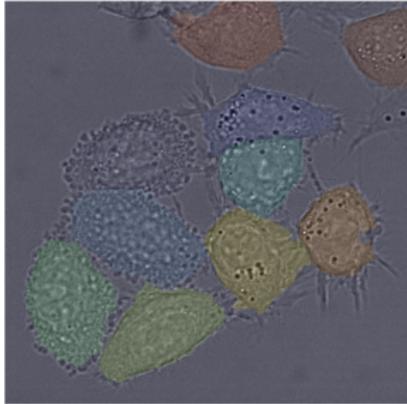


Augmented Mask

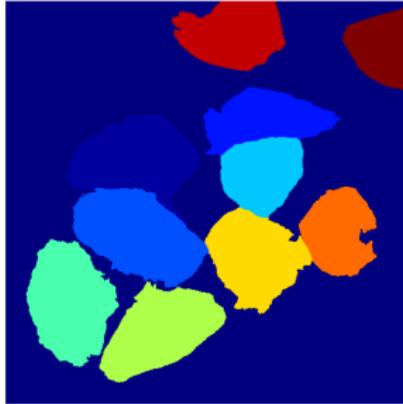


# Augmentation: Resultado do Pipeline

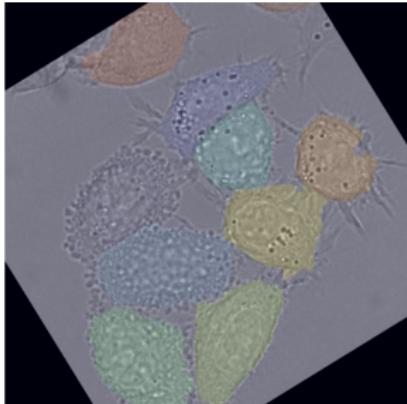
Original Overlay



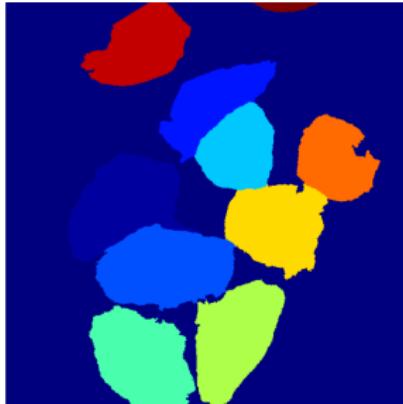
Original Mask



Augmented Overlay



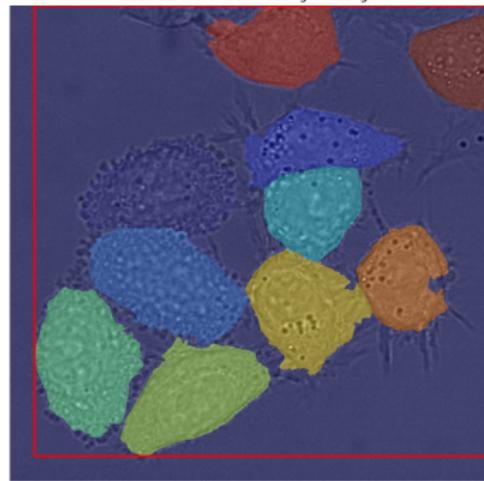
Augmented Mask



# Segment Anything Model (SAM)

- **Promptable vision transformer:**  
Necessitamos de um prompt!
- **Como pode ser feito?:** Via bounding boxes, pontos e máscaras.
- **Fluxo:** patch → prompt box → máscara refinada com foco nas células.
- **Objetivo:** Um modelo generalista retém performance num domínio microscópico?

HeLa Cells: ISBI Cell Tracking Challenge



# Tipos de Prompt Abordados

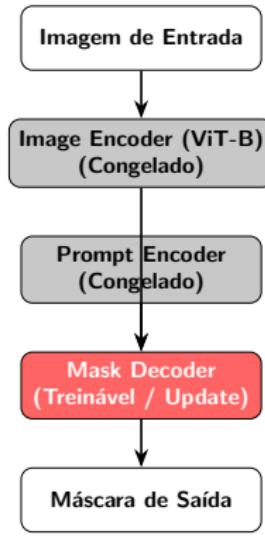
Neste estudo, avaliamos a capacidade do SAM de segmentar núcleos sob três condições de entrada distintas:

- **Box Prompts** (Bounding Boxes): Extração das coordenadas  $(x_{min}, y_{min}, x_{max}, y_{max})$  a partir do ground truth. É o método mais robusto, fornecendo escala e localização precisa.
- **Point Prompts**: Simulação de interação humana (cliques). Utiliza-se o centróide ou pontos aleatórios dentro da máscara da célula. Apresenta maior desafio devido à ambiguidade espacial.
- **Mask Prompts** (Dense): Fornecimento de uma máscara de baixa resolução (ou logits da iteração anterior) como dica densa para refinar contornos complexos.

# Metodologia: Estratégia de Fine-Tuning

Para viabilizar o treinamento, utilizamos o modelo **ViT-B (Base)**.

- **Congelamento de Pesos (Frozen):** O *Image Encoder* (ViT-B) e o *Prompt Encoder* permaneceram inalterados para preservar o conhecimento generalista prévio.
- **Pesos Treináveis:** Apenas o **Mask Decoder** teve seus gradientes calculados e pesos atualizados.
- **Hiperparâmetros:**
  - **Otimizador:** Adam ( $lr = 1e^{-5}$ ).
  - **Batch:** 1 imagem por iteração (devido ao tamanho dos embeddings).

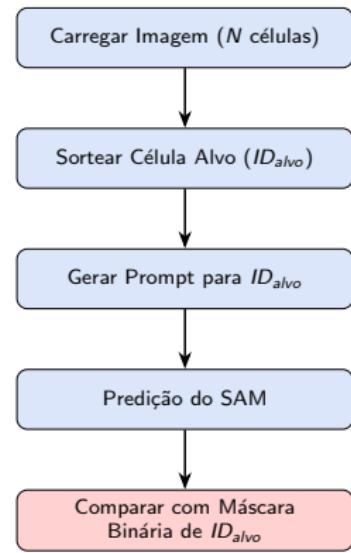


# Metodologia: Loop de Treinamento

O treinamento não utilizou todas as células simultaneamente. Para cada imagem no batch, o modelo foi treinado para focar em um único objeto por vez.

## Fluxo por Iteração:

- ① **Seleção Aleatória:** Identificam-se todos os IDs de células na imagem e sorteia-se **uma** célula alvo.
- ② **Prompting:** Gera-se o prompt (Box/Point/Mask) correspondente a esta célula sorteada.
- ③ **Segmentação Binária:** O SAM deve gerar uma máscara binária onde o alvo é **1** (Foreground) e todo o resto (outras células + fundo) é **0** (Background).
- ④ **Loss:** Calculada apenas sobre esta predição focalizada.



# Métricas de Avaliação

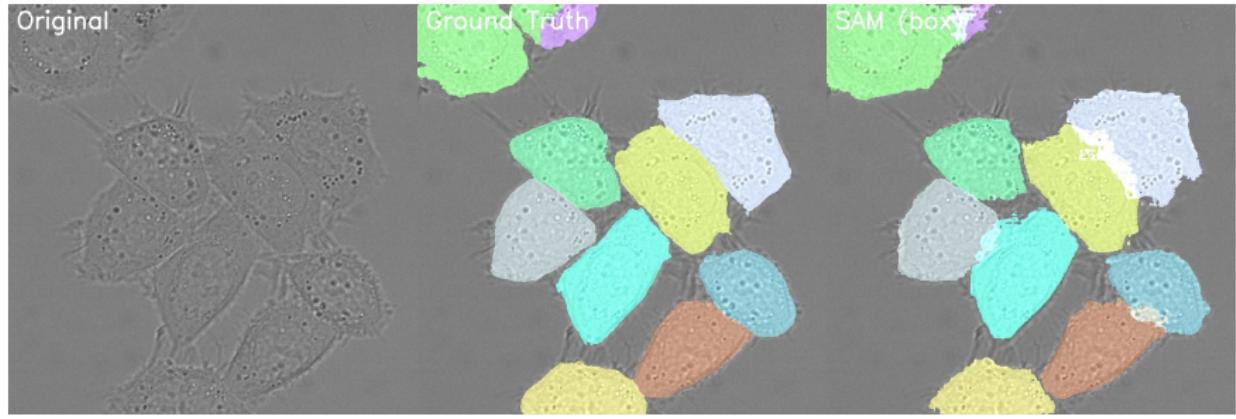
- 1. Intersection over Union (IoU)** Mede a sobreposição percentual estrita entre as áreas predita e real.

$$\text{IoU}_i = \frac{|G_i \cap P_i|}{|G_i \cup P_i|}$$

- 2. Dice Coefficient (F1-Score)** Média harmônica entre precisão e recall, frequentemente utilizada em segmentação médica.

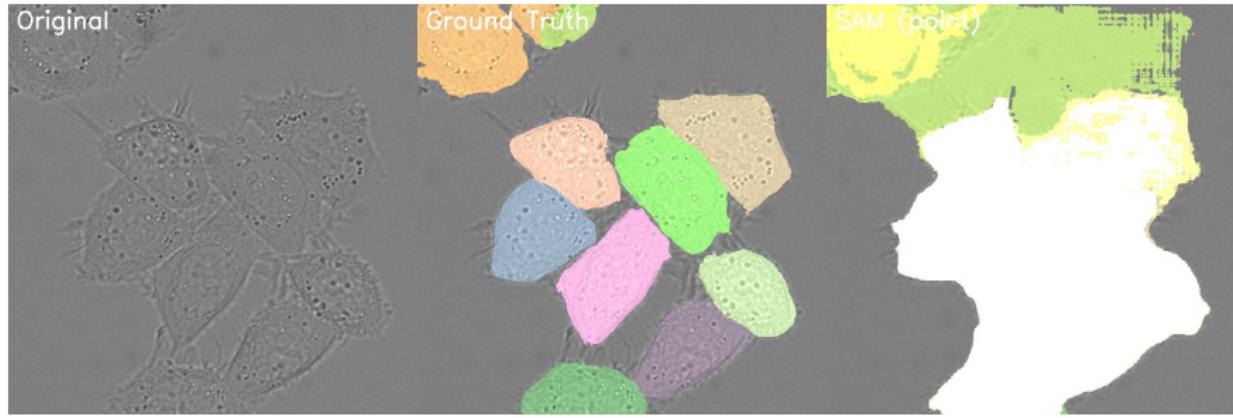
$$\text{Dice}_i = \frac{2 \cdot |G_i \cap P_i|}{|G_i| + |P_i|}$$

# SAM (box)



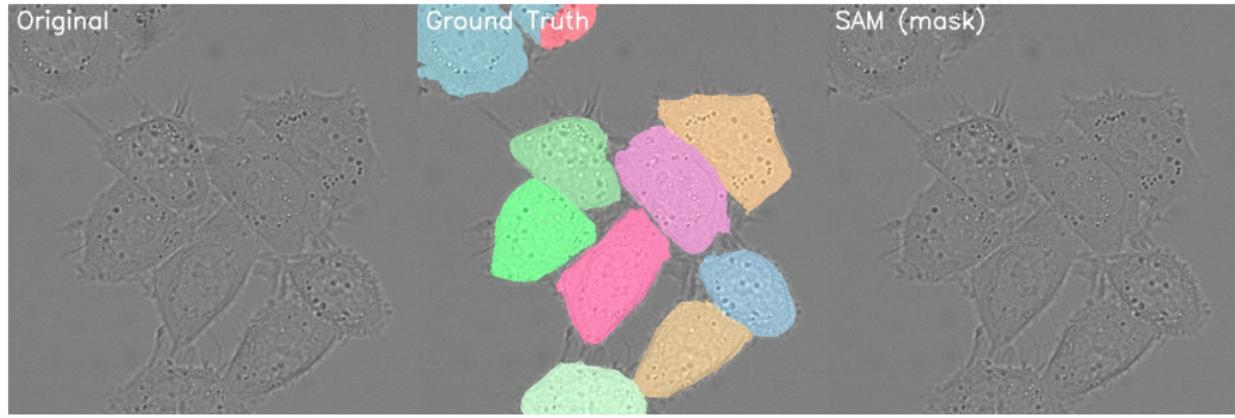
- Segmentação precisa.
- Máscaras parecidas com Ground Truth.

# SAM (point)



- Super-segmentação: células fundidas.
- Grandes regiões cobrindo múltiplos objetos.

# SAM (mask)



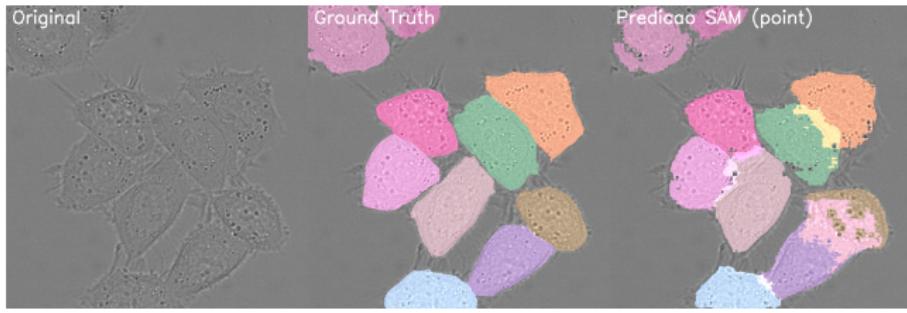
- Saída nula ou máscara vazia.
- Falha vista quando a máscara inicial não contém informação útil.

## Resultados Experimentais (Conjunto de Teste)

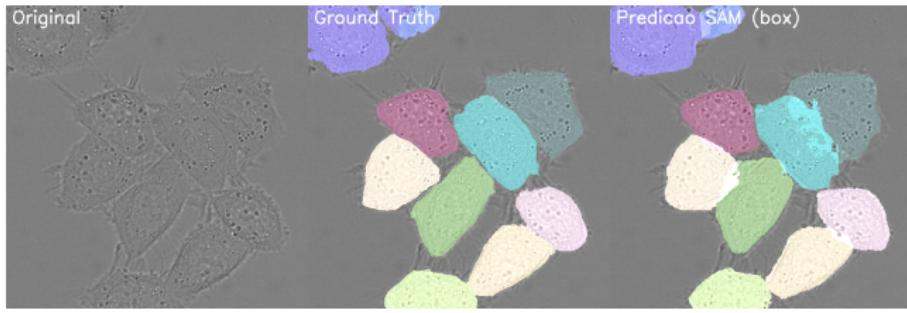
Prompt	Mean IoU	Mean Dice
Box	<b>0.8366</b>	<b>0.9081</b>
Point	0.7298	0.8356
Mask	0.7209	0.7852

- **Box:** Apresentou a melhor performance global, indicando que a restrição espacial rígida auxilia significativamente o modelo.
- **Point/Mask:** Resultados consistentes, porém inferiores, refletindo a dificuldade em delimitar bordas precisas com prompts mais ambíguos.

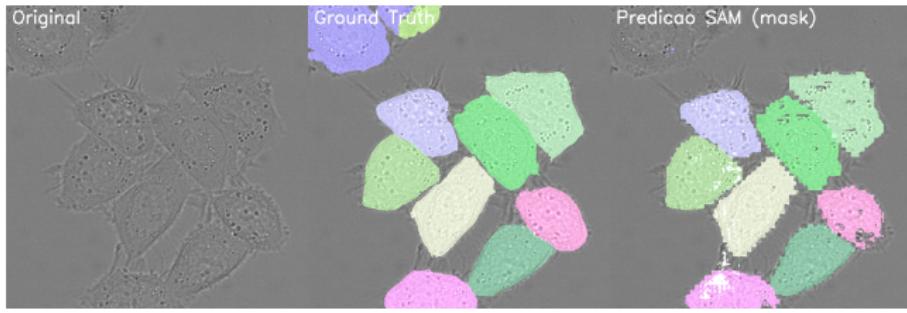
# Resultados Qualitativos: Point Prompt



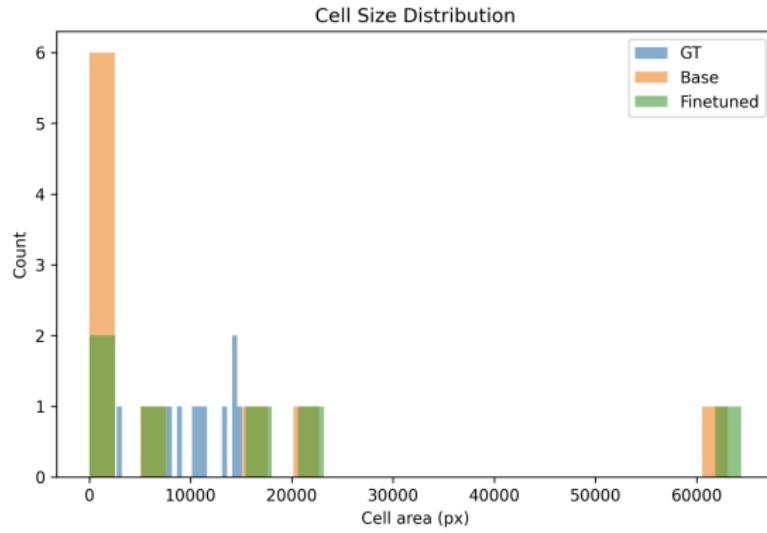
# Resultados Qualitativos: Box Prompt



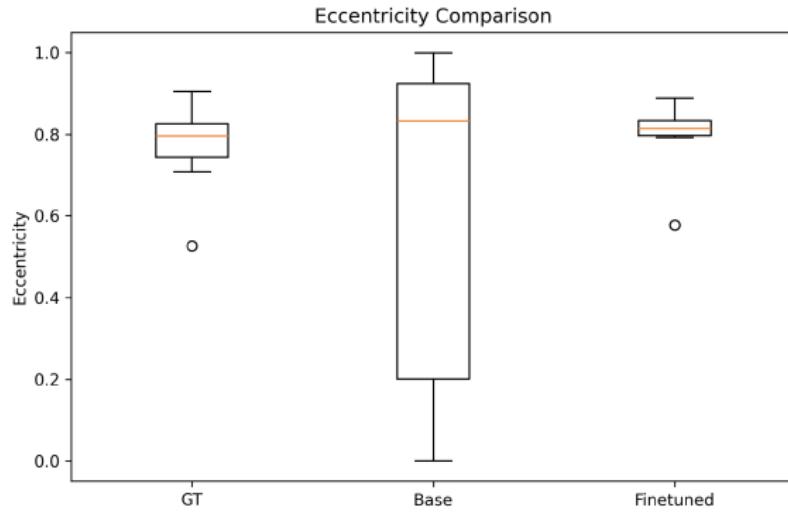
# Resultados Qualitativos: Mask Prompt



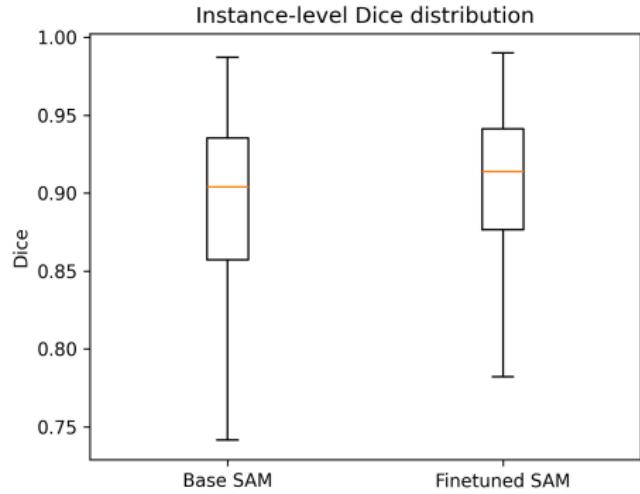
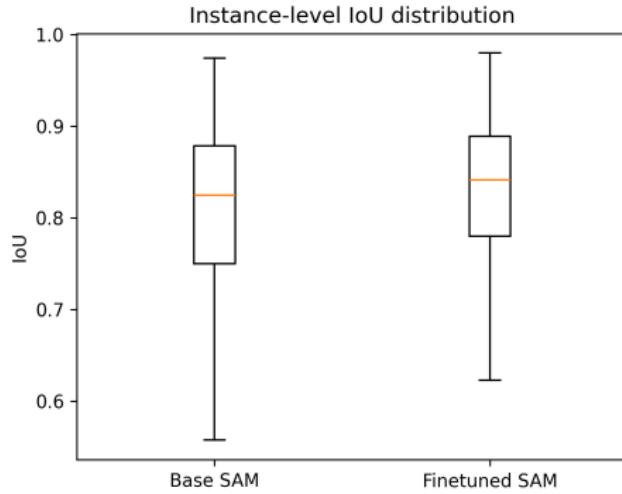
# Resultados Quantitativos: Fine Tuning



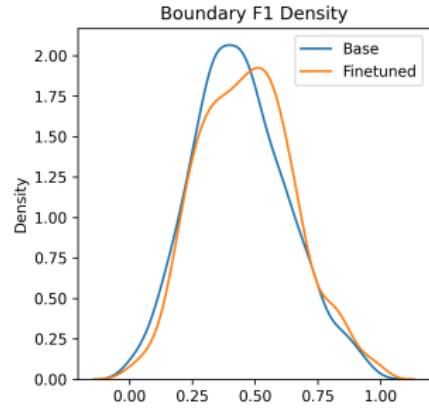
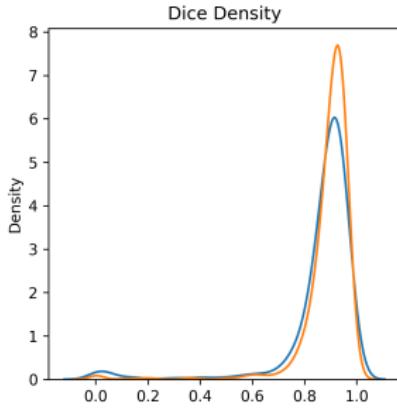
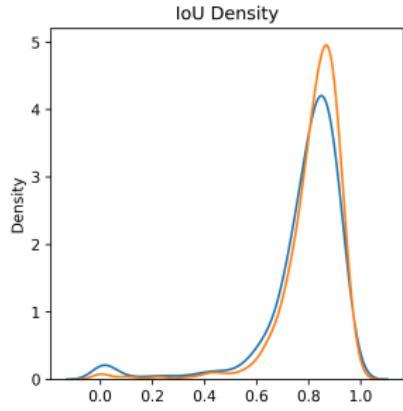
# Resultados Quantitativos: Fine Tuning



# Resultados Quantitativos: Fine Tuning



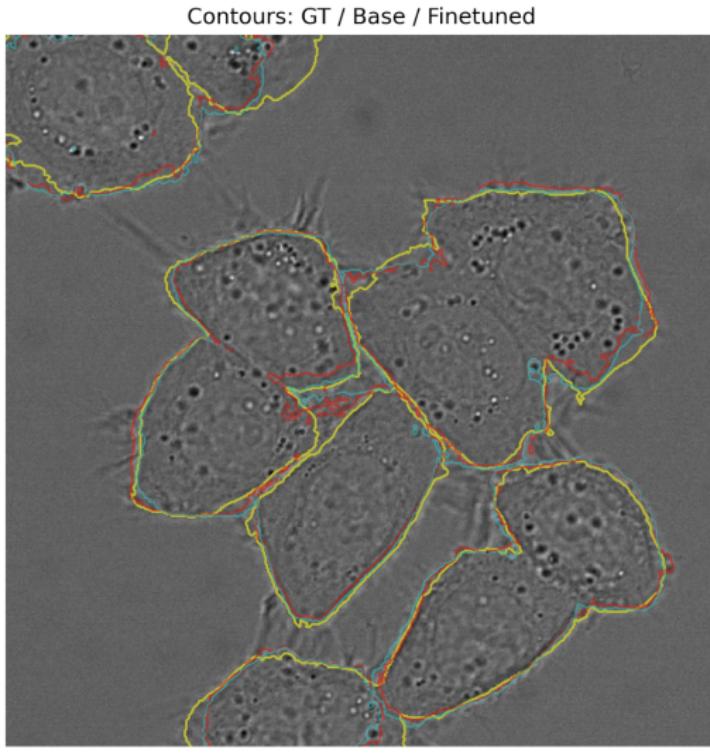
# Resultados Quantitativos: Fine Tuning



# Resultados Quantitativos: Fine Tuning

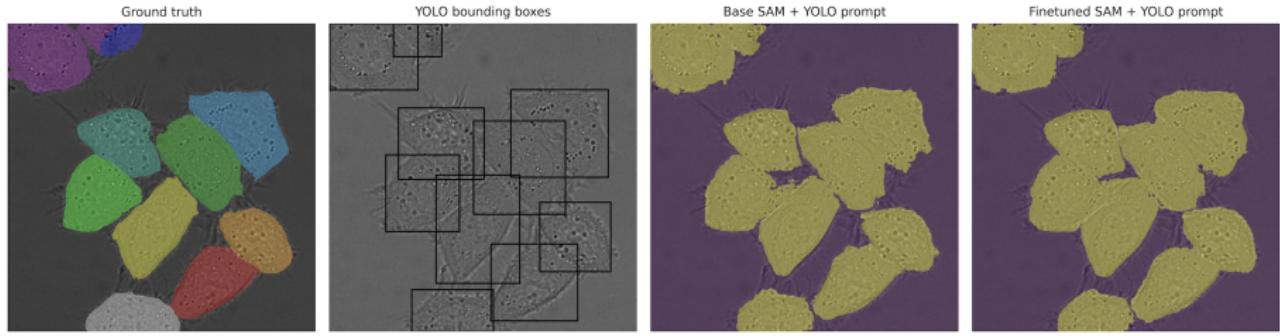


# Resultados Quantitativos: Fine Tuning



# Automação com YOLOv11

- **Problema:** Gerar prompts sem depender de máscaras!
- **Solução:** YOLOV11



Observação: Ground Truth somente para comparação, pipeline não necessita deste.

# YOLO como Guia de Detecção

**Motivação:** O SAM não opera eficientemente na imagem completa; ele requer uma região.

**Solução:** Usar o YOLOv11s para gerar *bounding boxes* que delimitam cada célula.

# Função do YOLO no Pipeline SAM

**Papel Central:** Atuar como filtro preliminar que determina “*onde*” o SAM deve atuar.

## Contribuições Principais

- **Localização Confiável**

- Geração de coordenadas precisas para cada célula pois tem alta precisão de classe.
- Reduz buscas desnecessárias do SAM.

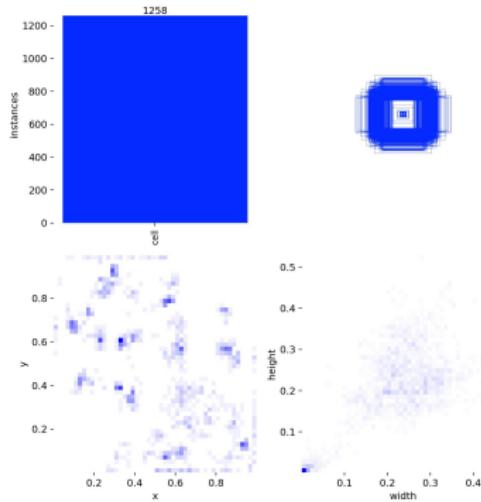
- **Seleção Biologicamente Informada**

- SAM é agnóstico de classe; segmentaria ruídos.
- YOLO filtra apenas objetos compatíveis com células HeLa.

- **Viabilidade Computacional**

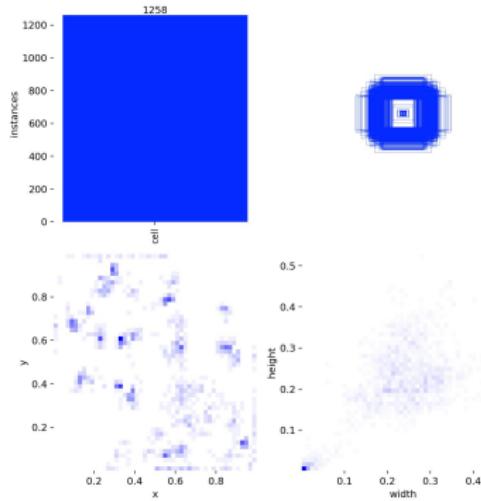
- Detecção ultrarrápida mantém o pipeline escalável.

# Correlograma — Tamanho das Células



- A maior parte das células ocupa apenas  $\approx 10\%-20\%$  da imagem.
- Objetos muito pequenos são facilmente perdidos por modelos fracos.
- Justifica o uso do **YOLOv11**, cuja **Feature Pyramid Network** mantém alta resolução e melhora a detecção de objetos muito pequenos.

# Correlograma — Distribuição Espacial

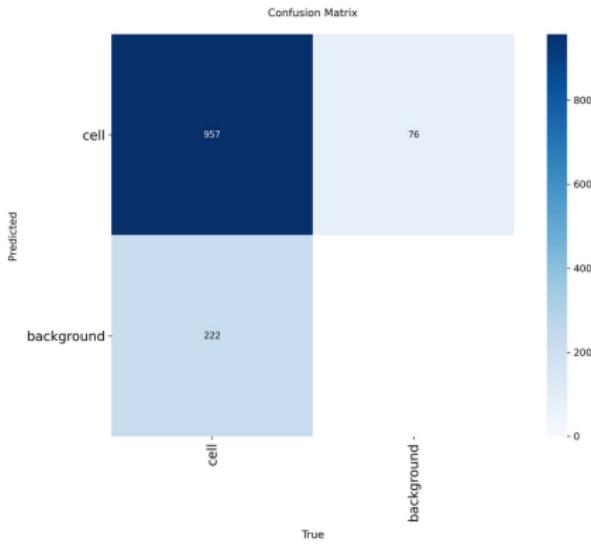


- As células aparecem em clusters, refletindo colônias HeLa.
- O YOLO lida melhor com **ocluções e objetos densos** que detectores mais antigos.

# Interpretação da Detecção e Impacto no SAM

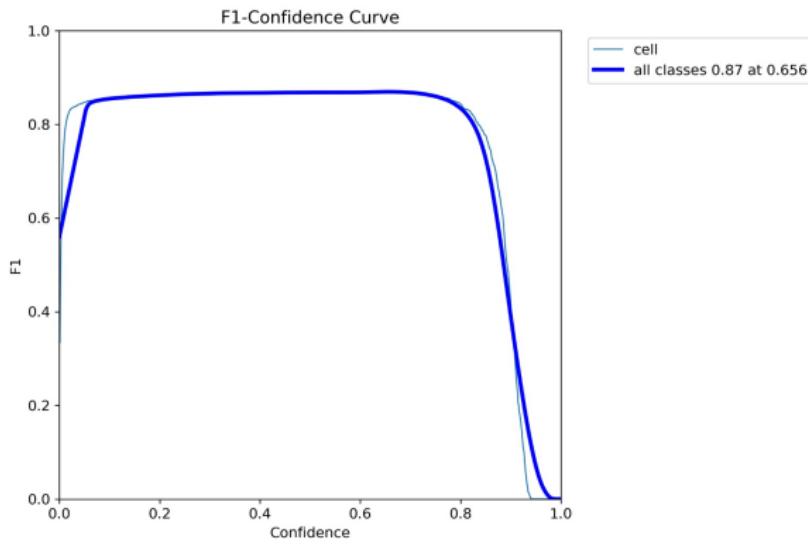
- O padrão observado é esperado para um modelo **rápido e pequeno** como o YOLOv11s, que prioriza detecções **confiáveis** em vez de cobertura total.
- As bounding boxes fornecidas ao SAM são **bem alinhadas** com células reais, reduzindo segmentações incorretas de ruído.
- Porém, células associadas a FN não recebem prompt, ou seja, o SAM **não que elas existem**.

# Matriz de Confusão — YOLOv11 (Detecção de Células)



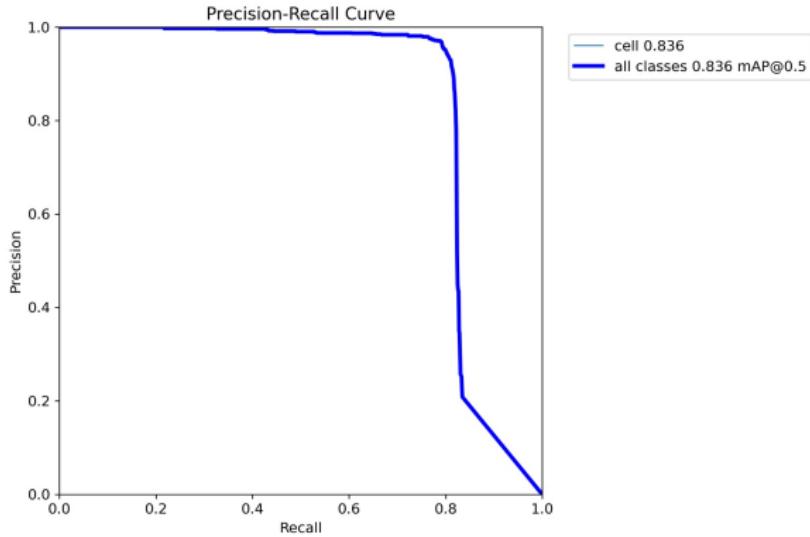
- A maioria das células verdadeiras é detectada corretamente como **cell** ( $TP = 957$ ).
- Falsos negativos moderados ( $FN = 222$ ) reduzem o **recall** ( $\approx 0.81$ ).
- Falsos positivos baixos ( $FP = 76$ ) resultam em **alta precisão** ( $\approx 0.93$ ).

# Curva F1–Confidence



- O F1 atinge seu pico em **0.86** com confiança  $\approx 0.50$ .
- A região superior é estável com precisão e recall retos.

# Curva Recall–Confidence



- O recall permanece estável até aproximadamente **0.5**, mas cai rapidamente acima de **0.6**.
- Limiares elevados aumentam os falsos negativos: células deixam de ser detectadas.

# Considerações Finais e Trabalhos Futuros

- **Adaptação Eficiente do Encoder (LoRA):**

Aplicar *Low-Rank Adaptation* (LoRA) no Image Encoder para ajustar as features ao domínio de microscopia com baixo custo computacional. [Ref3]

- **Funções de Perda Especializadas:**

Utilizar *Boundary Loss* ou *Hausdorff Loss* para melhorar a precisão nas bordas, onde os métodos mais falharam. [Ref4, Ref5]

# Referências

- Ref1. Xie, Y. et al. *MaskSAM: Auto-prompt SAM with Mask Classification for Volumetric Medical Image Segmentation*. ICCV, 2025.
- Ref2. Zhang, H. et al. *AoP-SAM: Automation of Prompts for Efficient Segmentation*. arXiv, 2025.
- Ref3. Wei, Y. et al. *Convolution Meets LoRA: Parameter-Efficient Finetuning for the Segment Anything Model*. arXiv, 2024.
- Ref4. Karimi, D.; Salcudean, S. *Reducing the Hausdorff Distance in Medical Image Segmentation with Convolutional Neural Networks*. IEEE TMI, 2019.
- Ref5. Ma, J. et al. *A Weighted Normalized Boundary Loss for Medical Image Segmentation*. 2023.