



Estudo de modelos de *deep learning* para segmentação de células em imagens microscópicas

Áureo Henrique e Silva Marques

Relatório Técnico - IC-PFG-24-42

Projeto Final de Graduação

2024 - Dezembro

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

The contents of this report are the sole responsibility of the authors.
O conteúdo deste relatório é de única responsabilidade dos autores.

Estudo de modelos de *deep learning* para segmentação de células em imagens microscópicas

Áureo Henrique e Silva Marques* Esther Luna Colombini*

Resumo

A segmentação de imagens microscópicas é uma tarefa essencial na área biomédica, sendo amplamente utilizada na análise de células. Este Projeto Final de Graduação apresenta um estudo sobre o uso de modelos de aprendizado profundo (*deep learning*) para segmentação celular, com foco nos modelos Stardist e Cellpose, ambos baseados na arquitetura U-Net. Inicialmente, foi realizada uma revisão de conceitos de microscopia e técnicas de segmentação de células, abrangendo métodos tradicionais e modernos. Em seguida, foi conduzido um experimento utilizando os modelos Cellpose e Stardist em dois *datasets* públicos do Kaggle: um mais diverso e outro com um contexto mais específico. A métrica de avaliação escolhida foi a *mean average precision* (mAP). Os resultados mostraram que, para o primeiro *dataset*, o modelo Cellpose apresentou a melhor performance, enquanto o Stardist se destacou por ser mais eficiente em tempo de execução. Para o segundo *dataset*, a performance de todos os modelos foi baixa devido à segmentação excessiva, indicando a necessidade de adaptações específicas para esse tipo de dado.

1 Introdução

A segmentação de imagens microscópicas de tecidos celulares é uma tarefa central na área biomédica, com aplicações relevantes na identificação de grupos celulares específicos, análise da geometria celular e contagem de células. Essa tarefa, no entanto, apresenta desafios significativos devido à complexidade e diversidade das imagens, que incluem variações em forma, textura, densidade e condições de aquisição. Recentemente, modelos de aprendizado profundo (*deep learning*) têm se destacado como ferramentas promissoras para abordar esses desafios, oferecendo alta precisão e robustez em tarefas de segmentação semântica e por instância.

*Instituto de Computação, Universidade Estadual de Campinas, 13081-970 Campinas, SP.

Neste contexto, este projeto final de graduação tem dois objetivos principais. O primeiro é realizar um estudo aprofundado sobre os métodos utilizados na segmentação de imagens microscópicas, com ênfase em modelos de *deep learning* recentes e de alta performance. O segundo objetivo consiste na realização de experimentos práticos utilizando os métodos estudados, aplicados a *datasets* públicos disponíveis no Kaggle. Esses experimentos visam testar e analisar a performance dos modelos em termos de métricas de segmentação e tempo de execução, comparando os resultados obtidos com referências existentes para os *datasets* escolhidos.

2 Estudo sobre segmentação de imagens microscópicas

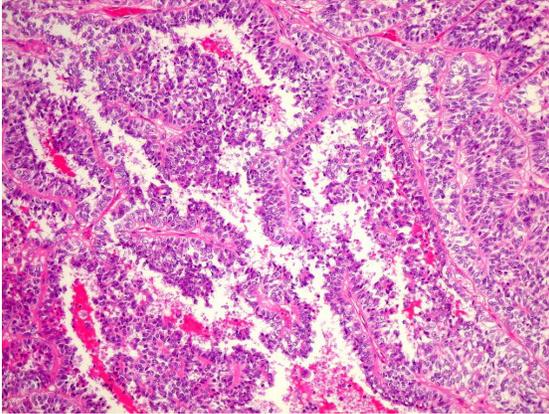
2.1 Microscopia na biomedicina

A microscopia é uma das ferramentas mais fundamentais e transformadoras da biomedicina, desempenhando um papel crucial no avanço do conhecimento sobre a estrutura e a função dos sistemas biológicos [1][2]. Sua importância reside na capacidade de revelar detalhes em escalas nano e micrométricas, permitindo o estudo aprofundado de células, tecidos e biomoléculas que não podem ser observados a olho nu. Desde a primeira observação de células por Robert Hooke no século XVII [3], a microscopia tem sido uma aliada indispensável na compreensão dos princípios básicos da vida, fornecendo as bases para a Teoria Celular e, mais recentemente, contribuindo para campos emergentes como a biologia molecular, a bioengenharia e a medicina de precisão.

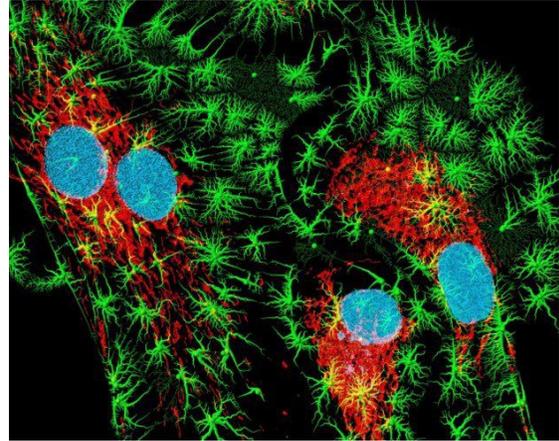
Na pesquisa biomédica, a microscopia não apenas fornece imagens detalhadas das estruturas celulares, mas também permite a análise dinâmica de processos biológicos em tempo real. Isso é essencial para a investigação de patologias humanas, como câncer, doenças neurodegenerativas e distúrbios metabólicos, ajudando a identificar alterações estruturais e funcionais em células e tecidos. Além disso, técnicas avançadas de microscopia possibilitam explorar a interação de biomoléculas e estudar mecanismos celulares em contextos específicos, como o impacto de medicamentos em células vivas ou a progressão de lesões pré-cancerosas. Essa capacidade de vincular observações estruturais e funcionais tem sido fundamental para avanços no diagnóstico, no desenvolvimento de terapias e no acompanhamento clínico de diversas condições.

Outro aspecto que destaca a relevância da microscopia é sua versatilidade. Ao longo dos últimos séculos, essa técnica evoluiu de ferramentas rudimentares para dispositivos altamente especializados, com aplicações que vão desde a caracterização básica de culturas celulares até a exploração de complexas redes de interação molecular. Combinada a avanços em áreas como engenharia óptica, química e computação, a microscopia moderna é capaz de atender às demandas de um campo biomédico em constante evolução, oferecendo uma gama de técnicas adaptadas a diferentes necessidades experimentais e diagnósticas.

Os métodos de microscopia são diversos, mas podem ser classificados amplamente de acordo com a fonte de iluminação utilizada [4]. No contexto desse projeto, duas técnicas de microscopia estarão presentes entre os dados analisados: a de campo claro e a de fluorescência. A Figura 1 apresenta um exemplo de imagem para cada técnica.



(a) Técnica de campo claro.



(b) Técnica de fluorescência.

Figura 1: Exemplos de imagens para as técnicas de microscopia de campo claro e de fluorescência. Imagens extraídas de [4].

A microscopia de campo claro (*bright-field*) é a técnica mais básica e amplamente utilizada. Nela, a luz visível atravessa a amostra, gerando imagens diretas que, embora simples, são limitadas pela necessidade de corantes para distinguir estruturas celulares. Por sua simplicidade, essa abordagem é frequentemente utilizada em laboratórios para análises preliminares de culturas celulares e tecidos.

Já a microscopia de fluorescência representa um marco na biomedicina devido à sua capacidade de localizar moléculas específicas em células e tecidos. Usando sondas fluorescentes, como anticorpos marcados ou proteínas fluorescentes (como GFP) [5], essa técnica permite a visualização precisa de biomarcadores, interações moleculares e processos dinâmicos dentro de células vivas.

Apesar dos avanços notáveis, a análise de imagens microscópicas apresenta desafios consideráveis. A diversidade de modalidades de microscopia, associada à variedade de amostras biológicas e condições experimentais, resulta em dados ricos, mas muitas vezes complexos. A tarefa de segmentação enfrenta dificuldades devido à heterogeneidade de células, formas e artefatos gerados durante o processo de aquisição. Sendo assim, a necessidade de métodos de análise de imagem escaláveis e robustos para interpretar esses grandes volumes de dados destaca a importância de integrar técnicas computacionais para transformar imagens em informações biológicas significativas [6].

2.2 Métodos para segmentação de células

Como exposto anteriormente, a segmentação de células é uma etapa essencial na análise de imagens microscópicas. Seja a segmentação do núcleo celular ou do citoplasma (ou seja, da célula inteira), essa tarefa é particularmente desafiadora devido à complexidade e existência de diversos tipos de células e de imagens, que exigem métodos adaptáveis para extrair contornos precisos.

Na segmentação de imagens, é importante diferenciar a segmentação semântica da segmentação por instância. Na segmentação semântica, o método atribui rótulos a cada pixel com base em categorias definidas. Por exemplo, *pixel de célula* e *pixel de fundo* para diferenciar onde na imagem tem e não tem uma célula. Já na segmentação por instância, que vai além da semântica, existe uma individualização de cada célula, atribuindo identificadores (*labels*) a cada região. Essa última é indispensável em amostras densamente populadas, onde a análise individual de células é fundamental para entender seu comportamento ou sua morfologia.

Sendo assim, métodos tradicionais de segmentação, como *thresholding*, *watershed* e operações morfológicas, têm sido amplamente utilizados pelos cientistas tanto para segmentação semântica como para por instância, mas apresentam limitações significativas. Eles exigem ajustes manuais de parâmetros que variam conforme os dados, tornando-os ineficientes e pouco generalizáveis para diferentes condições experimentais. Isso cria gargalos nos fluxos de trabalho, retardando a análise e comprometendo a consistência dos resultados.

Com isso, os avanços na área de *machine learning*, mais especificamente em *deep learning*, surgiram como uma solução promissora para superar esses desafios. Modelos baseados em redes neurais profundas conseguem aprender diretamente das imagens as características relevantes para a segmentação, reduzindo a necessidade de ajustes manuais e aumentando a adaptabilidade a diferentes conjuntos de dados [6][7]. Além disso, esses modelos demonstram melhor desempenho e precisão em relação aos métodos tradicionais, especialmente na segmentação por instância, destacando-se como uma ferramenta essencial para análises avançadas na biomedicina. As próximas sessões apresentam uma descrição sobre os principais métodos de segmentação tradicionais e os de *deep learning*, presentes na literatura.

2.2.1 Métodos tradicionais para segmentação

Thresholding O *thresholding* é uma técnica simples e amplamente empregada na segmentação de células. Ele converte imagens em escala de cinza em binárias, separando objetos do fundo com base em um valor limite. Este método é eficiente e de baixo custo computacional, sendo particularmente útil para análises de grande escala. Métodos como o *thresholding* de Otsu [8] são amplamente citados na literatura. No entanto, essa abordagem apresenta limitações significativas, como a sensibilidade a variações de intensidade na imagem, o que pode levar à segmentação incompleta de

células mal coradas. Melhorias incluem o uso de múltiplos *thresholds* ou a aplicação de janelas deslizantes para segmentação local, permitindo maior adaptabilidade às variações da imagem.

Operações Morfológicas As operações morfológicas, como erosão e dilatação, são ferramentas poderosas para refinar a segmentação de células [9][10], especialmente em amostras onde células estão muito próximas. A erosão reduz o tamanho dos objetos na imagem, separando células que estão em contato. Em seguida, a dilatação restaura o tamanho original das células, mantendo-as separadas. Essas operações também são amplamente utilizadas como etapas de pré-processamento para regularizar características estruturais antes da segmentação [11]. Abordagens avançadas, como operadores morfológicos multiescala, têm sido desenvolvidas para melhorar a precisão da segmentação ao suavizar contornos celulares e reduzir ruídos [12].

Transformada *Watershed* A transformada *watershed* [13] é um método clássico e eficaz para segmentação por instância, particularmente útil antes do surgimento de abordagens baseadas em aprendizado profundo. Ela interpreta a intensidade da imagem como um relevo topográfico, onde os mínimos regionais correspondem às áreas iniciais de "alagamento". As linhas de *watershed* definem as fronteiras entre objetos, permitindo a separação de células individuais. Embora poderosa, essa técnica frequentemente sofre com o problema de supersegmentação, onde variações de intensidade levam à divisão excessiva de objetos. Para mitigar esse problema, marcadores controlados podem ser usados para iniciar o alagamento em pontos pré-definidos, aumentando a precisão da segmentação [14]. Métodos mais recentes combinam a transformada *watershed* com técnicas de aprendizado profundo, onde mapas de segmentação gerados por redes neurais são refinados para melhorar a definição de fronteiras entre células em contato [6][7].

2.2.2 Métodos de *deep learning*

Rede U-Net A U-Net é uma arquitetura de rede neural convolucional (CNN) desenvolvida especificamente para a segmentação de imagens biomédicas, apresentada por Olaf Ronneberger, Philipp Fischer e Thomas Brox no trabalho intitulado *U-Net: Convolutional Networks for Biomedical Image Segmentation* [15]. Desde sua introdução, a U-Net se tornou uma ferramenta indispensável na área, sendo amplamente aplicado em tarefas como detecção de células, segmentação de tumores e delimitação de órgãos.

A arquitetura da U-Net é caracterizada por uma estrutura simétrica de codificador-decodificador, conectada por *skip connections* (Figura 2). Na etapa de codificação, a rede reduz progressivamente as dimensões espaciais da imagem usando camadas de

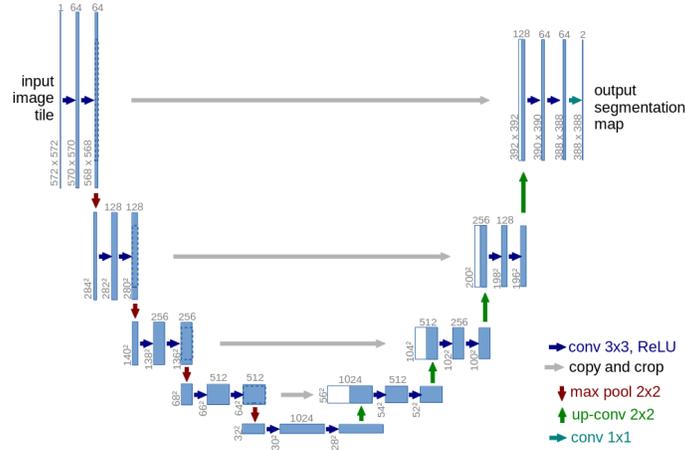


Figura 2: Exemplo de arquitetura U-Net. Figura extraída de [15].

convolução e *pooling*, extraíndo características de alto nível. Já a etapa de decodificação restaura essas dimensões por meio de *upsampling* e convoluções transpostas, gerando o mapa segmentado final. As *skip connections* conectam as camadas correspondentes do codificador e decodificador, permitindo a recuperação de informações espaciais detalhadas. Essa combinação é essencial para segmentar estruturas pequenas e complexas, comuns em imagens biomédicas.

Em relação as vantagens comparado com os métodos tradicionais, a U-Net supera esses métodos ao lidar com estruturas complexas, ruído e variações de qualidade nas imagens biomédicas. Enquanto abordagens clássicas frequentemente falham em separar células próximas ou segmentar regiões com baixa intensidade, a U-Net utiliza aprendizado profundo para capturar tanto o contexto global quanto os detalhes locais, proporcionando segmentações mais precisas.

Comparado a outras arquiteturas de CNN, a U-Net é otimizado para segmentação, destacando-se por sua eficiência e preservação de informações de baixo nível, graças às *skip connections*. Isso torna o modelo capaz de segmentar estruturas tanto grandes quanto pequenas. Além disso, sua adaptabilidade a diferentes modalidades de imagem, como microscopia de fluorescência, ressonância magnética (MRI) e tomografia computadorizada (CT), amplia seu impacto em aplicações biomédicas.

Já no quesito impacto na análise de imagens biomédicas, o modelo transformou a segmentação de imagens nessa área, permitindo avanços significativos em áreas como rastreamento de células, diagnóstico de doenças e planejamento de tratamentos. Estudos como o de Falk et al. [16], que utilizou a U-Net para contagem e detecção de células, mostram sua eficácia em tarefas desafiadoras, como segmentar células sobrepostas, com formas irregulares e em populações densas.

Além disso, a capacidade da U-Net de fornecer segmentações precisas em nível de pixel o tornou indispensável para modelos integrados, que combinam segmentação

com detecção de objetos, classificação e extração de características, abrindo novas possibilidades para a pesquisa biomédica.

Stardist Stardist é uma abordagem de aprendizado profundo para detecção e segmentação de objetos, desenvolvido por Uwe Schmidt e Martin Weigert, apresentado inicialmente no trabalho *Cell Detection with Star-Convex Polygons* [17] e expandido em *Star-Convex Polyhedra for 3D Object Detection and Segmentation in Microscopy* [18]. Baseado na arquitetura U-Net, o modelo introduz um método inovador para segmentar objetos em imagens utilizando polígonos "estrela-convexos", que aproximam as formas dos objetos.

Um polígono "estrela-convexo" é definido como uma forma onde todos os pontos podem ser conectados a um ponto central dentro do polígono por uma linha reta que não sai da forma. Essa propriedade geométrica é particularmente útil para descrever os contornos de células e outras estruturas biológicas, que frequentemente apresentam formas complexas e irregulares.

Sendo assim, o modelo Stardist utiliza uma rede neural convolucional para prever dois elementos principais: probabilidades de objeto e distâncias estrela-convexas. As probabilidades de objeto são determinadas como a distância euclidiana normalizada de cada pixel até o pixel de fundo mais próximo, auxiliando na identificação dos polígonos centrados em objetos. Já as distâncias estrela-convexas envolvem o cálculo das distâncias euclidianas entre cada pixel e a borda do objeto, traçando direções radiais até encontrar pixels de identidade diferente. A Figura 3 demonstra o funcionamento da arquitetura.

A arquitetura destaca-se por sua capacidade de lidar com objetos de formas altamente irregulares e variáveis. O modelo é particularmente eficaz em situações onde os objetos estão densamente agrupados, como em imagens de células adjacentes em amostras de microscopia. O uso de polígonos estrela-convexos permite a segmentação precisa de células que estão tocando ou até se sobrepondo, algo desafiador para métodos tradicionais. Os estudos de Schmidt e Weigert demonstraram sua eficácia em detectar e segmentar células em imagens de microscopia de fluorescência, mesmo em cenários com formas celulares variadas e densidades elevadas.

Outro ponto forte do Stardist é sua eficiência computacional. A arquitetura foi projetada para ser rápida e precisa, tornando-a ideal para análise de grandes volumes de dados e aplicações de alto rendimento, comuns em pesquisas biomédicas.

Cellpose Cellpose é um algoritmo robusto e amplamente aplicável, projetado para superar as limitações de métodos existentes, que muitas vezes têm dificuldade em generalizar para diferentes tipos de células e condições de imagem. Desenvolvido por Stringer C. e colaboradores [19], o modelo destaca-se por sua adaptabilidade, apresentando bom desempenho em diversos contextos de imagem biológica sem a

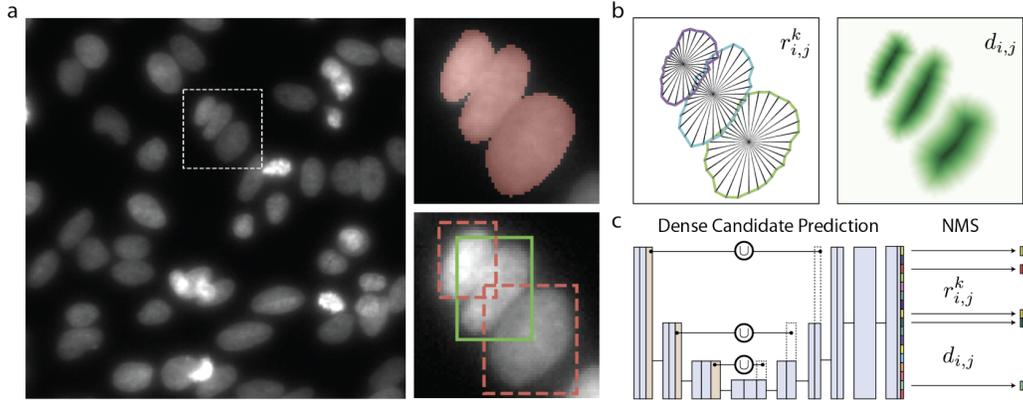


Figura 3: Arquitetura do modelo Stardist. ”(a) Potential segmentation errors for images with crowded nuclei: Merging of touching cells (upper right) or suppression of valid cell instances due to large overlap of bounding box localization (lower right). (b) The proposed StarDist method predicts object probabilities $d_{i,j}$ and star-convex polygons parameterized by the radial distances $r_{i,j}^k$. (c) We densely predict $r_{i,j}^k$ and $d_{i,j}$ using a simple U-Net architecture and then select the final instances via non-maximum suppression (NMS). Figura extraída de Uwe S. et al [17]”

necessidade de ajustes extensivos de parâmetros ou re-treinamento para cada conjunto de dados específico.

Baseado em uma variação da arquitetura U-Net, Cellpose utiliza CNNs para prever não apenas os contornos das células, mas também os gradientes espaciais que descrevem o fluxo vetorial em direção ao centro de cada célula. Essa abordagem inovadora permite que o algoritmo segmente células ao modelar características geométricas das estruturas celulares, em vez de depender exclusivamente da classificação de intensidades de pixels. Como resultado, o modelo é capaz de lidar com uma ampla variedade de morfologias celulares, desde células arredondadas até formas mais alongadas ou irregulares. A Figura 4 apresenta o funcionamento da arquitetura.

Em relação ao seu treinamento, os autores compilaram um conjunto de dados diverso, abrangendo diferentes tipos celulares e modalidades de imagem, como microscopia de fluorescência e contraste de fase. Os dados incluíam anotações das máscaras celulares e dos campos de fluxo direcional para o centro da célula, permitindo que o modelo aprendesse a delinear células individuais de maneira eficaz, mesmo em populações densas ou onde as células se tocam. O uso da representação de campo de fluxo foi crucial para resolver casos de células sobrepostas, um desafio comum na segmentação de imagens biomédicas, como já mencionado.

Com isso, o modelo foi treinado em mais de 70.000 imagens (com mais de 200.000 células segmentadas), cobrindo uma ampla gama de contextos biológicos. Além disso, imagens não microscópicas de objetos repetidos também foram utilizadas para melho-

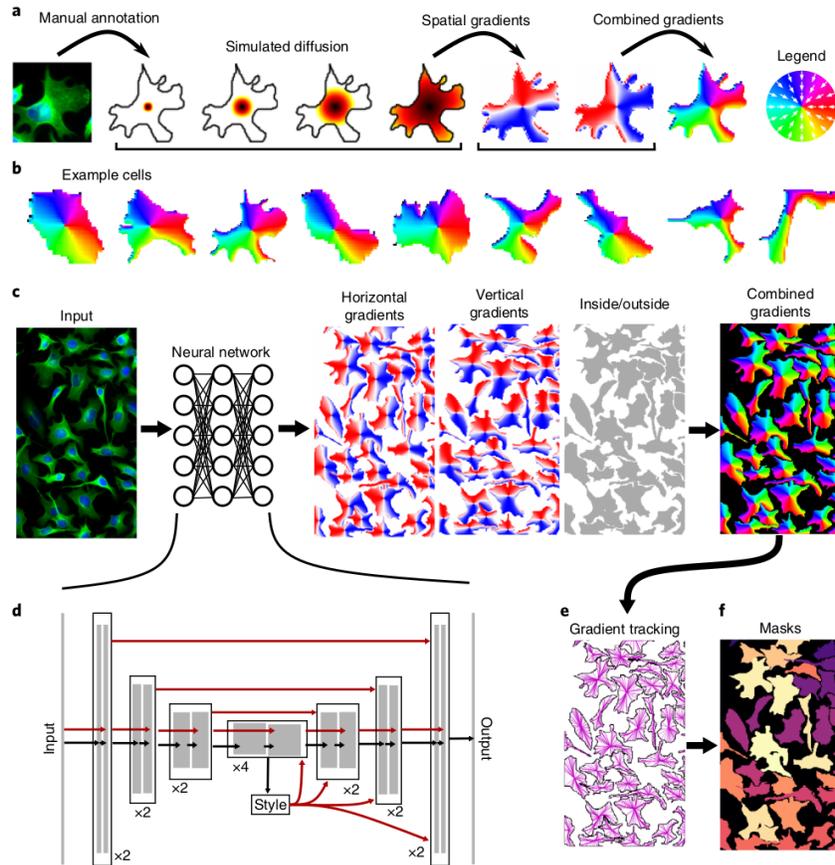


Figura 4: Arquitetura do modelo Cellpose. "a, Procedure for transforming manually annotated masks into a vector flow representation that can be predicted by a neural network. A simulated diffusion process starting from the center of the mask is used to derive spatial gradients that point toward the center of the cell, potentially indirectly around corners. The x and y gradients are combined in a single direction from 0° to 360° . b, Example spatial gradients for cells from the training dataset. c, A neural network is trained to predict the horizontal and vertical gradients, as well as whether a pixel belongs to any cell. The three predicted maps are combined into a gradient vector field. d, The details of the neural network that contains a standard backbone U-Net3 to downsample and then upsample the feature maps, with skip connections between layers of the same size and global skip connections from the image styles, computed at the lowest resolution, to all successive computations. e, At test time, the predicted gradient vector fields are used to construct a dynamical system with fixed points whose basins of attraction represent the predicted masks. Informally, every pixel 'tracks the gradients' toward their eventual fixed point. f, All the pixels that converge to the same fixed point are assigned to the same mask." Figura extraída de Stringer C. et al. [19]

rar a capacidade de generalização do modelo. Essa abordagem geral contrasta com outros métodos de *deep learning*, que frequentemente exigem re-treinamento significativo ou ajustes finos para funcionar em novos tipos de dados.

Um outro ponto é a versatilidade do Cellpose, evidenciada por sua oferta de diferentes modelos, dependendo do tipo de imagem. Existem modelos voltados para imagens em campo claro, como *TissueNet* e *LiveCell*, e outros para imagens de fluorescência, como *Cyto* e *CP*. Embora esses modelos sejam projetados principalmente para segmentação citoplasmática, o Cellpose também inclui um modelo específico para segmentação nuclear (*nuclei model*).

O algoritmo integra um passo de pré-processamento que normaliza as imagens, garantindo desempenho consistente em diferentes condições. Ele também permite a entrada do usuário para refinar os resultados de segmentação, oferecendo um equilíbrio entre automação total e ajustes manuais quando necessário. Além disso, o algoritmo oferece uma interface gráfica amigável que simplifica o uso do modelo em novos conjuntos de dados, permitindo que pesquisadores se beneficiem de suas capacidades avançadas de segmentação sem necessidade de conhecimento profundo em análise de imagens.

Sendo assim, o modelo demonstrou desempenho de ponta em diversos conjuntos de dados de referência, superando ou igualando a precisão de métodos especializados projetados para tipos celulares específicos, mas com muito menos necessidade de personalização. Recentemente, uma segunda publicação foi lançada para facilitar ainda mais o treinamento do modelo em novos dados [20], tornando mais simples adaptar o algoritmo a contextos específicos e melhorar seu desempenho em cenários inéditos.

3 Metodologia

Nesta etapa do projeto, abordamos o segundo objetivo proposto, que consiste na avaliação experimental de modelos de *deep learning* utilizando *datasets* disponíveis no Kaggle, uma plataforma amplamente utilizada que oferece uma grande variedade de *datasets* para treinamento e teste de modelos de aprendizado de máquina.

Modelos Selecionados Para os experimentos, optamos por utilizar os modelos analisados na seção anterior: Cellpose e StarDist. Ambos possuem bibliotecas implementadas em Python, que se destacam pela facilidade de uso e ampla documentação. Os testes foram realizados no ambiente Google Colab, que permite a instalação e importação de bibliotecas de forma prática, além de oferecer acesso a uma unidade GPU gratuita, suficiente para os requisitos computacionais deste experimento.

O modelo Cellpose possui diferentes submodelos, cada um projetado para atender a contextos específicos. Para o nosso experimento, escolhemos o submodelo *cyto3*,

a versão mais recente, desenvolvida para segmentação de citoplasma e treinada com imagens fluorescentes provenientes de diversos tipos de tecido. Além disso, o Cellpose conta com um hiperparâmetro relacionado ao diâmetro médio das células, que auxilia na identificação das regiões com base nessa medida fornecida previamente. Esse hiperparâmetro pode ser definido de forma automática (estimado pelo próprio modelo) ou manualmente pelo usuário.

Com isso, realizamos dois testes: um utilizando a escolha automática do diâmetro e outro variando o valor do diâmetro dentro de um intervalo, para identificar a métrica de desempenho ideal com base no valor ótimo do diâmetro para cada imagem. Assim, pudemos avaliar se o diâmetro estimado automaticamente pelo modelo é de fato uma boa escolha para a predição. Dado o custo computacional elevado para testar múltiplos valores de diâmetro, restringimos a análise a, no máximo, sete valores distintos, definidos com base nas características de cada *dataset*.

Por sua vez, o modelo StarDist também conta com submodelos, e selecionamos para os experimentos o *2D-versatile_fluo*, uma versão treinada com imagens fluorescentes de diversos contextos. Diferentemente do Cellpose, o StarDist não utiliza hiperparâmetros relacionados ao diâmetro celular, o que dispensou a necessidade de testes adicionais.

Datasets Selecionados Dois *datasets* do Kaggle foram selecionados para os testes:

- **Dataset 1:** "Nuclei Segmentation in Microscope Cell Images"¹ Este *dataset* consiste em 664 imagens de células de diferentes tipos de tecido, capturadas por variadas técnicas de aquisição (como *bright-field* e fluorescência). Ele foi originalmente proposto para a competição Data Science Bowl de 2018² (DSB2018), cujo objetivo era segmentar os núcleos celulares. Como parte dos recursos disponíveis, o *dataset* inclui a lista de modelos vencedores e suas respectivas performances, medidas pela métrica oficial da competição. Esses dados oferecem uma base sólida para compararmos os resultados obtidos pelos modelos estudados com os melhores desempenhos alcançados na competição.

- **Dataset 2:** "Fluorescent Neuronal Cells"³ Este *dataset* é mais específico e consiste em 283 imagens fluorescentes de tecido cerebral de camundongos. Diferentemente do *dataset* 1, ele não foi produzido para uma competição e, portanto, não possui valores de referência para avaliação da performance dos modelos. Esse *dataset* representa um caso de uso mais específico, permitindo testar a capacidade dos modelos em lidar com um contexto mais restrito.

Métrica de Avaliação de Performance Para avaliar o desempenho dos modelos, utilizamos a métrica oficial da competição DSB2018, associada ao *dataset* 1. A

¹<https://www.kaggle.com/datasets/gangadhar/nuclei-segmentation-in-microscope-cell-images>

²<https://www.kaggle.com/c/data-science-bowl-2018/>

³<https://www.kaggle.com/datasets/nbroad/fluorescent-neuronal-cells>

métrica é chamada de *mean average precision* (mAP). Trata-se de uma métrica amplamente utilizada para avaliar o desempenho de modelos em tarefas de detecção e segmentação de objetos. Ela calcula a média da precisão média (*average precision*, AP) em diferentes valores de limiar de *Intersection over Union* (IoU). O IoU é definido como a razão entre a interseção e a união dos pixels de um objeto previsto e do respectivo objeto verdadeiro. Matematicamente, é expresso como:

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (1)$$

O passo-a-passo do cálculo é explicado abaixo:

1. **Limiar de IoU:** O cálculo do mAP envolve múltiplos valores de limiar de IoU, variando de 0.5 a 0.95 com passos de 0.05. Por exemplo, para um limiar de 0.5, um objeto previsto é considerado uma correspondência correta (*true positive*) se o IoU com o objeto verdadeiro for maior ou igual a 0.5.
2. **Cálculo da Precisão em Cada Limiar:** Para cada limiar t , a precisão é calculada considerando:
 - **Verdadeiros Positivos (TP):** Previsões corretas que correspondem a um objeto verdadeiro com IoU acima do limiar.
 - **Falsos Positivos (FP):** Objetos previstos que não correspondem a nenhum objeto verdadeiro.
 - **Falsos Negativos (FN):** Objetos verdadeiros que não têm correspondência com nenhuma previsão.

A fórmula da precisão para um limiar específico t é dada por

$$Precisão(t) = \frac{TP(t)}{TP(t) + FP(t) + FN(t)}. \quad (2)$$

3. **Cálculo da Average Precision (AP):** Para cada imagem, a AP é a média das precisões calculadas em todos os limiares de IoU

$$AP = \frac{1}{|T|} \sum_{t \in T} Precisão(t), \quad (3)$$

onde T é o conjunto de limiares (de 0,5 a 0,95).

4. **Cálculo do mAP:** O mAP final é a média das APs de todas as imagens no conjunto de teste. Essa média reflete o desempenho global do modelo.

4 Resultados

4.1 Dataset 1

Os resultados para o Dataset 1 estão apresentados na Tabela 1. Nessa tabela, comparamos os três modelos utilizados — Modelo 1 (Cellpose com diâmetro automático), Modelo 2 (Cellpose sem diâmetro automático) e Modelo 3 (Stardist) — quanto à métrica mAP obtida e ao tempo de execução necessário para processar o conjunto de imagens. Para o Modelo 2, foram selecionados empiricamente sete valores de diâmetro médio (10, 15, 20, 25, 30, 35 e 40 pixels), com base no diâmetro estimado pelo Modelo 1, no tamanho das imagens do dataset e na diversidade de tamanhos das células presentes. Durante os experimentos, utilizamos a unidade GPU do Google Colab para executar os modelos Cellpose, enquanto o Stardist foi executado apenas com CPU.

Tabela 1: Resultados obtidos para o Dataset 1.

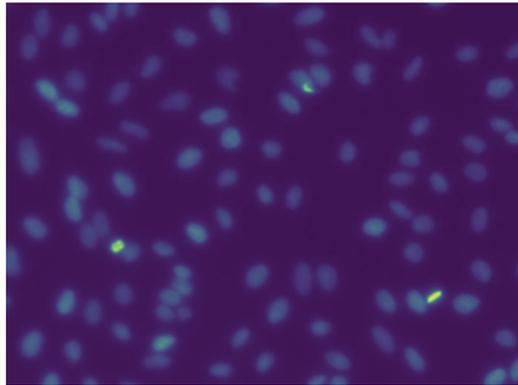
Modelo	mAP	Tempo de Execução
1 - Cellpose (com diam. automático)	0,2827	0h52min
2 - Cellpose (sem diam. automático)	0,5686	5h30min
3 - Stardist	0,5136	1h25min

A Figura 5 apresenta um exemplo visual das segmentações realizadas pelos três modelos em uma imagem do Dataset 1. A figura contém cinco subfiguras: a imagem original, a imagem de referência (ground truth) e as máscaras preditas pelos modelos 1, 2 e 3.

A partir da Figura 5, podemos realizar uma análise qualitativa dos resultados. Comparando com a imagem de referência 5a (*ground truth*), nota-se que os modelos, em geral, conseguiram identificar adequadamente as posições e os formatos das células. A máscara gerada pelo Modelo 1, mostrada na subfigura 5c, apresentou o menor valor de mAP (0,2539). Esse desempenho inferior pode ser atribuído a um diâmetro estimado maior, que não correspondeu adequadamente às dimensões reais das células no *dataset*. Por outro lado, os outros dois modelos apresentaram métricas superiores, sendo o Modelo 3 o que obteve o maior mAP (0,5826).

Passando para a análise quantitativa, conforme apresentado na Tabela 1, os resultados confirmam a tendência observada na análise qualitativa. O Modelo 1 teve a menor performance, enquanto os Modelos 2 e 3 apresentaram melhores resultados. Nesse caso, contudo, o Modelo 2 alcançou o maior valor de mAP (0,5686), superando o Modelo 3 (0,5136). Comparando esse resultado com a métrica do modelo vencedor da competição DSB2018 (0,6316), podemos considerar a segmentação obtida pelo Cellpose como satisfatória, dada a proximidade do *score*.

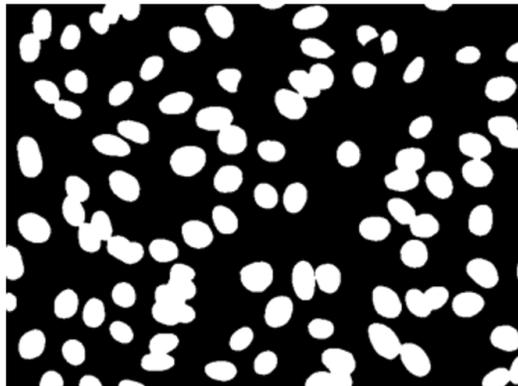
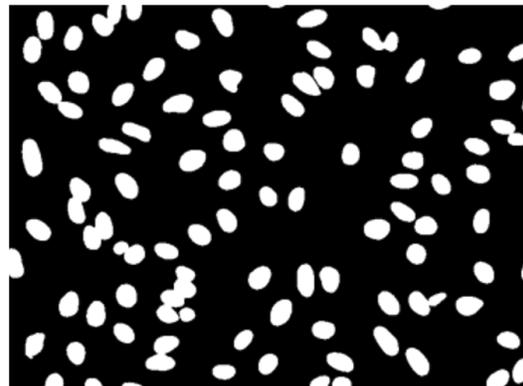
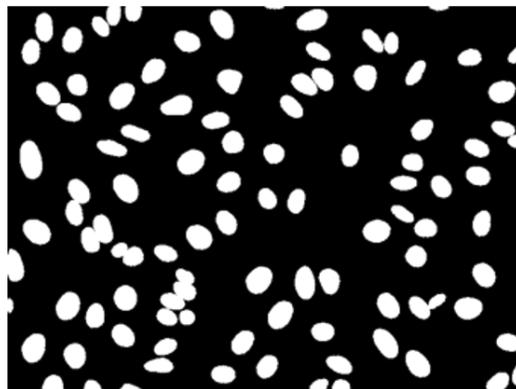
Em relação ao tempo de execução, o Modelo 1 foi o mais rápido (0h52min), seguido pelo Modelo 3 (1h25min), enquanto o Modelo 2 foi o mais lento (5h30min). Esse au-



(a) Imagem Original



(b) Ground Truth

(c) Cellpose (com diam. automático) -
mAP = 0,2539(d) Cellpose (sem diam. automático) -
mAP = 0,5605

(e) Stardist - mAP = 0,5826

Figura 5: Exemplo de execução para o Dataset 1: (a) Imagem Original; (b) Ground Truth; (c) Predição do Modelo 1; (d) Predição do Modelo 2; (e) Predição do Stardist.

mento significativo no tempo de execução do Modelo 2 é justificável pelo processo de segmentação realizado com sete diferentes valores de diâmetro, o que aumentou consideravelmente o custo computacional. Assim, embora o Modelo 2 tenha apresentado a maior métrica, sua desvantagem está no tempo elevado de processamento.

4.2 Dataset 2

Os resultados para o Dataset 2 são apresentados na Tabela 2, contendo as mesmas métricas de comparação entre os três modelos. Para o Modelo 2, foram escolhidos cinco valores de diâmetro médio (40, 45, 50, 55 e 60). Estes são maiores que o experimento anterior, devido à dimensão maior das imagens, e menos medidas foram escolhidas, devido à maior uniformidade de tamanho de células do Dataset 2. Assim como no experimento anterior, utilizamos a unidade GPU do Google Colab para os modelos Cellpose, enquanto o Stardist foi executado exclusivamente com CPU.

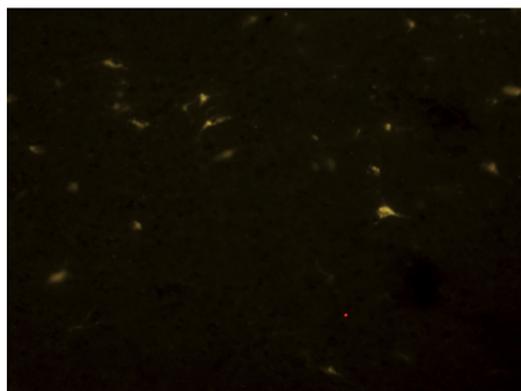
Tabela 2: Resultados obtidos para o Dataset 2.

Modelo	mAP	Tempo de Execução
1 - Cellpose (com diam. automático)	0,0221	0h28min
2 - Cellpose (sem diam. automático)	0,0478	2h30min
3 - Stardist	0,0125	2h50min

A Figura 6 apresenta um exemplo visual das segmentações realizadas pelos três modelos em uma imagem do Dataset 2. A figura inclui cinco subfiguras: a imagem original, a imagem *ground truth* e as máscaras preditas pelos Modelos 1, 2 e 3.

A análise quantitativa dos resultados, conforme apresentado na Tabela 2, revela uma performance bastante baixa para todos os modelos, com mAPs inferiores a 0,05. O Modelo 2 apresentou a melhor métrica (0,0478). Essa baixa performance pode ser explicada pela análise qualitativa apresentada na Figura 6. Comparando as máscaras preditas com a máscara de referência (Figura 6b), observa-se que todos os modelos apresentaram uma supersegmentação significativa, identificando muito mais regiões do que o esperado. Esse comportamento reduz drasticamente a métrica mAP. Essa segmentação excessiva provavelmente está relacionada à natureza específica das imagens do Dataset 2, que apresentam células neuronais com formatos distintos daqueles usados no treinamento das arquiteturas. Assim, os modelos avaliados mostram-se inadequados para esse contexto.

Em relação ao tempo de execução, o Modelo 1 novamente foi o mais rápido (0h28min), enquanto os outros dois modelos apresentaram tempos semelhantes: 2h30min para o Modelo 2 e 2h50min para o Modelo 3. No geral, os tempos de processamento foram menores em comparação com o Dataset 1 devido à menor quantidade de imagens no Dataset 2 e, no caso do Modelo 2, à redução no número de valores de diâmetro utilizados no experimento, o que diminuiu significativamente o custo computacional.



(a) Imagem Original



(b) Ground Truth

(c) Cellpose (com diam. automático) -
mAP = 0,0168(d) Cellpose (sem diam. automático) -
mAP = 0,0431

(e) Stardist - mAP = 0,0081

Figura 6: Exemplo de execução para o Dataset 2: (a) Imagem Original; (b) Ground Truth; (c) Predição do Modelo 1; (d) Predição do Modelo 2; (e) Predição do Stardist.

5 Conclusão

Esse Projeto Final de Graduação apresentou um estudo aprofundado sobre o uso de modelos de aprendizado profundo para a segmentação de imagens microscópicas na área biomédica. Inicialmente, foram discutidos os conceitos fundamentais de microscopia, destacando sua importância no contexto de análise de processos biológicos e patologias humanas. Em seguida, abordamos métodos de segmentação de células, partindo de técnicas tradicionais, como *thresholding*, operações morfológicas e transformada watershed, que, embora amplamente utilizadas, possuem limitações devido à baixa capacidade de generalização. O avanço das técnicas de aprendizado de máquina, em particular os modelos de *deep learning*, mostrou-se uma solução promissora para superar essas limitações, permitindo maior precisão e robustez na segmentação. Entre os métodos de *deep learning*, apresentamos os modelos Stardist e Cellpose, ambos baseados na arquitetura U-Net e focados na segmentação de células.

Na segunda parte do projeto, realizamos experimentos com esses dois modelos utilizando dois *datasets* do Kaggle: o primeiro, maior e mais diverso, com imagens de diferentes tipos de células, e o segundo, menor e mais específico, contendo imagens de células neuronais. Três abordagens foram exploradas: Cellpose com diâmetro automático (Modelo 1), Cellpose sem diâmetro automático (Modelo 2) e Stardist (Modelo 3). A métrica de avaliação escolhida foi a *mean average precision* (mAP), definida pela competição DSB2018.

Os resultados mostraram que, para o Dataset 1, o Modelo 2 apresentou a melhor performance (mAP = 0,5686), seguido pelo Modelo 3 (mAP = 0,5136). No entanto, o tempo de execução do Modelo 2 foi significativamente maior, o que torna o Stardist (Modelo 3) uma alternativa mais conveniente caso o tempo seja uma prioridade. Para o Dataset 2, a performance dos três modelos foi bastante baixa, devido a uma segmentação excessiva em todas as predições. Isso indica que tanto o Cellpose quanto o Stardist provavelmente não foram treinados com dados semelhantes aos do Dataset 2, que apresenta células neuronais com características muito específicas. Para esse caso, seria interessante explorar outros submodelos do Cellpose ou Stardist, ou mesmo, redes de *deep learning* mais adequadas a esse contexto.

Por fim, em relação ao Dataset 1, um próximo passo pertinente seria a realização de *transfer learning* com o Cellpose, ou seja, utilizar as imagens do Dataset 1 para re-treinamento do modelo. Visto que o Cellpose permite esse re-treinamento de sua rede através de inserção de dados pelo usuário, um segundo experimento poderia ser elaborado para avaliar se a rede consegue melhorar sua performance e se aproximar, ou até mesmo superar, a métrica alcançada pelo modelo vencedor da competição DSB2018. Tal investigação seria fundamental para explorar o potencial do Cellpose e sua capacidade de adaptação em contextos específicos, ampliando as aplicações do modelo na segmentação de imagens biomédicas.

Referências

- [1] Sonali Karhana et al. “Chapter 11 - Advances in microscopy and their applications in biomedical research”. Em: *Biomedical Imaging Instrumentation*. Ed. por Mrutyunjay Suar, Namrata Misra e Neel Sarovar Bhavesh. Primers in Biomedical Imaging Devices and Systems. Academic Press, 2022, pp. 185–212. ISBN: 978-0-323-85650-8. DOI: <https://doi.org/10.1016/B978-0-323-85650-8.00008-5>. URL: <https://www.sciencedirect.com/science/article/pii/B9780323856508000085>.
- [2] Ana Maria Reigoto et al. “A comparative study on the use of microscopy in pharmacology and cell biology research”. Em: *PLoS ONE* 16.1 (2021), e0245795. DOI: [10.1371/journal.pone.0245795](https://doi.org/10.1371/journal.pone.0245795). URL: <https://doi.org/10.1371/journal.pone.0245795>.
- [3] Hooke R. *Micrographia: or, Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses*. 1st. London: J. Martyn e J. Allestry, 1665.
- [4] *Microscópio – conheça as diversas técnicas de microscopia*. URL: <https://kasvi.com.br/microscopio-tecnicas-microscopia/> (acesso em 05/09/2024).
- [5] Roger Y. Tsien. “The Green Fluorescent Protein”. Em: *Annual Review of Biochemistry* 67 (1998), pp. 509–544. DOI: [10.1146/annurev.biochem.67.1.509](https://doi.org/10.1146/annurev.biochem.67.1.509).
- [6] Juan C. Caicedo et al. “Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl”. Em: *Nature Methods* 16 (2019), pp. 1247–1253. DOI: [10.1038/s41592-019-0612-7](https://doi.org/10.1038/s41592-019-0612-7).
- [7] Neeraj Kumar et al. “A Multi-Organ Nucleus Segmentation Challenge”. Em: *IEEE Transactions on Medical Imaging* 39.5 (2020), pp. 1380–1391. DOI: [10.1109/TMI.2019.2947628](https://doi.org/10.1109/TMI.2019.2947628). URL: <https://doi.org/10.1109/TMI.2019.2947628>.
- [8] Nobuyuki Otsu. “A threshold selection method from gray-level histograms”. Em: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pp. 62–66. DOI: [10.1109/TSMC.1979.4310076](https://doi.org/10.1109/TSMC.1979.4310076).
- [9] O. Schmitt e M. Hasse. “Morphological multiscale decomposition of connected regions with emphasis on cell clusters”. Em: *Computer Vision and Image Understanding* 113.2 (2009), pp. 188–201.
- [10] Q. Wang et al. “Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy”. Em: *Cytometry Part A: The Journal of the International Society for Advancement of Cytometry* 77.1 (2010), pp. 101–110.
- [11] F. Xing e L. Yang. “Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review”. Em: *IEEE reviews in biomedical engineering* 9 (2016), pp. 234–263.

- [12] L. B. Dorini, R. Minetto e N. J. Leite. “Semiautomatic white blood cell segmentation based on multiscale analysis”. Em: *IEEE journal of biomedical and health informatics* 17.1 (2012), pp. 250–256.
- [13] J. B. Roerdink e A. Meijster. “The watershed transform: Definitions, algorithms and parallelization strategies”. Em: *Fundamenta Informaticae* 41.1-2 (2000), pp. 187–228.
- [14] L. Yang et al. “A new framework for particle detection in low-snr fluorescence live-cell images and its application for improved particle tracking”. Em: *IEEE transactions on biomedical engineering* 59.7 (2012), pp. 2040–2050.
- [15] Olaf Ronneberger, Philipp Fischer e Thomas Brox. “U-Net: Convolutional networks for biomedical image segmentation”. Em: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, pp. 234–241.
- [16] Thorsten Falk e Olaf Ronneberger. “U-Net: deep learning for cell counting, detection, and morphometry”. Em: *Nature Methods* 16 (2019), pp. 67–70. DOI: 10.1038/s41592-018-0261-2.
- [17] Uwe Schmidt et al. “Cell detection with star-convex polygons”. Em: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2018, pp. 265–273.
- [18] Martin Weigert et al. “Star-convex polyhedra for 3D object detection and segmentation in microscopy”. Em: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2020, pp. 311–319.
- [19] Carsen Stringer et al. “Cellpose: a generalist algorithm for cellular segmentation”. Em: *Nature Methods* 18 (2021), pp. 100–106. DOI: 10.1038/s41592-020-01018-x.
- [20] Marius Pachitariu e Carsen Stringer. “Cellpose 2.0: how to train your own model”. Em: *Nature Methods* 19 (2022), pp. 1634–1641. DOI: 10.1038/s41592-022-01663-4.