

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

João Pedro de Oliveira Marques

**Segmentação Geométrica de Núcleos em
Imagens Microscópicas do Desenvolvimento
Embrionário da *Drosophila melanogaster***

Uberlândia, Brasil

2026

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

João Pedro de Oliveira Marques

**Segmentação Geométrica de Núcleos em Imagens
Microscópicas do Desenvolvimento Embrionário da
*Drosophila melanogaster***

Trabalho de Conclusão de Curso.

Orientadora: Dra. Daniela Justiniano de Sousa Pereira

Universidade Federal de Uberlândia – UFU

Faculdade de Computação

Bacharelado em Ciência da Computação

Uberlândia, Brasil

2026

Resumo

A *Drosophila melanogaster* é amplamente utilizada como organismo modelo em estudos de biologia do desenvolvimento, sendo a análise de imagens microscópicas um instrumento essencial para a quantificação de padrões celulares. Nesse contexto, este trabalho apresenta uma abordagem computacional baseada em visão computacional para a segmentação de instâncias nucleares em imagens microscópicas de embriões, com foco em cenários de alta densidade celular e contato entre estruturas adjacentes. O método proposto utiliza um *pipeline* geométrico que modela contornos em representação polar $r(\theta)$ e detecta concavidades a partir de variações radiais em relação a uma referência baseada no envoltório convexo (*convex hull*). A partir dessas concavidades, regras de alinhamento direcional, critérios de coerência geométrica e mecanismos de recuo (*fallback*) são empregados para determinar cortes de separação entre possíveis instâncias nucleares.

Para apoiar a aplicação e a validação da abordagem, foi desenvolvido um software *open-source* em Python com interface gráfica, integrando parametrização visual, processamento em lote, inspeção de etapas intermediárias e extração automática de métricas. A avaliação quantitativa foi conduzida sobre um subconjunto de imagens contendo amostras selvagens e mutantes, utilizando anotações manuais de referência (*ground truth*) normalizadas espacialmente via interpolação B-Spline (BOOR, 2001). Os resultados indicaram desempenho promissor na delimitação da massa nuclear, com coeficiente Dice de até 85,6% no modelo base. Em nível de instâncias, o método apresentou desempenho variável entre as amostras, com *F1-Score* de até 77,0%, evidenciando que a abordagem é promissora para a individualização de núcleos em parte dos cenários avaliados, embora ainda apresente limitações em casos de maior complexidade morfológica. Além disso, observou-se que a calibração dos parâmetros pode contribuir para a redução de falhas geométricas de segmentação.

Palavras-chave: Processamento de Imagens; Visão Computacional; Segmentação; OpenCV; *Drosophila melanogaster*.

Abstract

Drosophila melanogaster is a widely used model organism in developmental biology, and microscopy image analysis is an essential tool for quantifying cellular patterns. In this context, this work presents a computer vision-based approach for nuclear instance segmentation in microscopy images of embryos, focusing on high-density scenarios with touching structures. The proposed method uses a geometric *pipeline* that models contours in polar representation $r(\theta)$ and detects concavities from radial variations with respect to a reference derived from the convex hull. Based on these concavities, directional alignment rules, geometric consistency criteria, and *fallback* mechanisms are used to determine separation cuts between possible nuclear instances.

To support the application and validation of the approach, an open-source Python software with a graphical user interface was developed, integrating visual parameter tuning, batch processing, inspection of intermediate steps, and automatic metric extraction. The quantitative evaluation was conducted on a subset of images containing wild-type and mutant samples, using manually annotated reference masks (*ground truth*) spatially normalized through B-Spline interpolation (BOOR, 2001). The results indicated promising performance in nuclear mass delineation, reaching a Dice coefficient of up to 85.6% in the *baseline* model. At the instance level, the method showed variable performance across samples, with *F1-Score* values of up to 77.0%, indicating that the approach is promising for nucleus individualization in part of the evaluated scenarios, although it still presents limitations in cases of higher morphological complexity. Parameter calibration also showed potential for reducing geometric segmentation errors.

Keywords: Image Processing; Computer Vision; Segmentation; OpenCV; *Drosophila melanogaster*.

Lista de ilustrações

Figura 1 – Exemplo de imagem microscópica de embrião de <i>Drosophila melanogaster</i> utilizada no conjunto experimental.	21
Figura 2 – Interface do software AnyLabeling utilizada para anotação manual das máscaras de referência.	22
Figura 3 – Visão geral do <i>pipeline</i> de segmentação proposto.	22
Figura 4 – Representação esquemática das fases do pré-processamento para geração da máscara binária refinada.	23
Figura 5 – Resultado qualitativo da aplicação do filtro por elipse antes e após o refinamento geométrico.	26
Figura 6 – Comparativo do sinal polar utilizado na análise geométrica do contorno.	29
Figura 7 – Resultado da segmentação após resolução de conflitos e aplicação do mecanismo de <i>fallback</i>	31
Figura 8 – Tela do aplicador responsável pela execução automatizada das receitas de segmentação.	33
Figura 9 – Sobreposição entre a segmentação automática e a anotação de referência.	34
Figura 10 – Etapas iniciais e referência manual da amostra 3-1.tif.	38
Figura 11 – Visualizações da segmentação da amostra 3-1.tif.	39
Figura 12 – Sobreposição entre a predição do algoritmo e a referência manual da amostra 3-1.tif.	40

Sumário

1	INTRODUÇÃO	9
1.1	Motivação e problema	10
1.2	Objetivos	11
1.2.1	Objetivo geral	11
1.2.2	Objetivos específicos	12
1.3	Contribuições	12
1.4	Organização do trabalho	13
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Segmentação de imagens	14
2.2	Pré-processamento e binarização	14
2.3	Contornos e representação hierárquica	15
2.4	Envoltório convexo e detecção de concavidades	16
2.5	Parametrização polar do contorno	16
2.6	Estratégias de separação de instâncias	16
2.7	Métricas de avaliação	18
3	MATERIAIS E MÉTODOS	20
3.1	Materiais	20
3.2	Visão geral do <i>pipeline</i> de segmentação	22
3.3	Pré-processamento	22
3.3.1	Filtro por elipse	25
3.4	Limiarização adaptativa local	26
3.5	Representação espacial e modelagem do sinal	28
3.6	Heurística de pontuação global e avaliação vetorial	29
3.7	Resolução de conflitos espaciais	30
4	DESENVOLVIMENTO DO SOFTWARE EM PYTHON	32
4.1	Arquitetura e módulos do projeto	32
4.2	Interface gráfica e gerador de receitas	33
4.3	Módulo de processamento em lote e espelhamento de diretórios	33
5	RESULTADOS E DISCUSSÕES	35
5.1	Análise do desempenho de base (<i>baseline</i>)	35
5.2	Análise visual detalhada da amostra 3-1.tif	37
5.3	Precisão da máscara de pixels e validação do pré-processamento	40

5.4	Calibração geométrica de estabilização	41
6	CONCLUSÃO E TRABALHOS FUTUROS	42
	REFERÊNCIAS	44

Lista de Abreviaturas e Siglas

GUI Interface Gráfica do Usuário, do inglês *Graphical User Interface*.

ROI Região de Interesse, do inglês *Region of Interest*.

GT Verdade de referência, do inglês *ground truth*.

TP Verdadeiro Positivo, do inglês *True Positive*.

FP Falso Positivo, do inglês *False Positive*.

FN Falso Negativo, do inglês *False Negative*.

IoU Interseção sobre União, do inglês *Intersection over Union*.

WT Amostra selvagem, do inglês *Wild Type*.

GPU Unidade de Processamento Gráfico, do inglês *Graphics Processing Unit*.

Otsu Método automático de limiarização global proposto por Otsu ([OTSU, 1979](#)).

Lista de Símbolos

Ao longo do texto, $r(\theta)$ representa o raio do contorno em função do ângulo; θ representa o ângulo polar; e Δr representa a diferença radial associada ao sinal de concavidade.

1 Introdução

A análise de imagens microscópicas constitui uma etapa fundamental em estudos de biologia do desenvolvimento, pois permite observar, quantificar e comparar padrões celulares ao longo de diferentes estágios embrionários (SOUSA; LOPES; TRAVENÇOLO, 2014; SOUSA et al., 2020). Em especial, a segmentação de estruturas nucleares é uma tarefa essencial para a extração de informações morfológicas e espaciais, como quantidade de núcleos, distribuição celular, densidade local e organização tecidual (MEIJERING, 2012). Esses dados apoiam a investigação de processos associados à formação e à diferenciação dos tecidos, bem como a comparação entre condições experimentais distintas (SOUSA; LOPES; TRAVENÇOLO, 2014; SOUSA et al., 2020).

Nesse contexto, a *Drosophila melanogaster* é amplamente utilizada como organismo modelo em pesquisas biológicas, em razão do ciclo de vida curto, da facilidade de manipulação genética e de sua relevância histórica nos estudos do desenvolvimento embrionário (SOUSA; LOPES; TRAVENÇOLO, 2014; SOUSA et al., 2020). As imagens microscópicas de embriões desse organismo fornecem uma base relevante para análises quantitativas, mas também apresentam desafios computacionais. A elevada densidade celular, o contato entre núcleos adjacentes, a variação de contraste, a presença de ruído e as irregularidades morfológicas dificultam a separação automática de cada instância nuclear (SOUSA et al., 2020; MEIJERING, 2012).

A segmentação automática dessas imagens pode ser analisada em dois níveis complementares: a segmentação semântica da massa nuclear total, que separa as regiões biológicas de interesse do fundo, e a segmentação de instâncias, que individualiza núcleos próximos ou encostados (MEIJERING, 2012; SCHMIDT et al., 2018). Embora a limiarização e as operações morfológicas possam produzir máscaras globais adequadas em determinados cenários, múltiplos núcleos em contato podem permanecer unidos em uma única componente conexa (MEIJERING, 2012; SOUSA et al., 2020). Nesses casos, torna-se necessário explorar evidências adicionais, como linhas de separação por *watershed* (VINCENT; SOILLE, 1991), concavidades, contornos internos, relações direcionais e modelos explícitos de forma (SOUSA; LOPES; TRAVENÇOLO, 2014; SOUSA et al., 2020; SCHMIDT et al., 2018).

A literatura reúne diferentes famílias de técnicas para esse problema. Entre as abordagens clássicas, o algoritmo de *watershed* separa regiões a partir de uma interpretação topográfica da imagem (VINCENT; SOILLE, 1991); entre as abordagens geométricas, a análise de concavidades e o pareamento vetorial procuram localizar regiões de contato e definir cortes entre objetos adjacentes (SOUSA; LOPES; TRAVENÇOLO, 2014; SOUSA et al., 2020); e, entre os métodos baseados em aprendizado profundo, destacam-se arquiteturas

como U-Net ([RONNEBERGER; FISCHER; BROX, 2015](#)), Cellpose ([STRINGER et al., 2021](#)) e StarDist ([SCHMIDT et al., 2018](#)). Neste trabalho, adota-se como referência principal a linha metodológica de Sousa et al. ([SOUSA; LOPES; TRAVENÇOLO, 2014](#); [SOUSA et al., 2020](#)), porque ela permite formular decisões de corte por regras geométricas explícitas, compatíveis com a inspeção visual das etapas intermediárias e com o conjunto reduzido de imagens anotadas disponível para avaliação. A escolha não pressupõe superioridade geral em relação aos métodos de aprendizado profundo; ela corresponde ao recorte metodológico deste TCC, orientado à interpretabilidade, à parametrização e à execução em hardware convencional.

A proposta busca, portanto, oferecer uma solução interpretável, parametrizável e de baixo custo computacional para a segmentação de núcleos em imagens microscópicas de embriões de *Drosophila melanogaster*. Diferentemente de métodos supervisionados que dependem de treinamento ou de modelos previamente ajustados ([RONNEBERGER; FISCHER; BROX, 2015](#); [STRINGER et al., 2021](#); [SCHMIDT et al., 2018](#)), a implementação desenvolvida emprega regras geométricas que podem ser inspecionadas visualmente e calibradas pelo usuário. A documentação das etapas, dos parâmetros e dos artefatos intermediários é coerente com recomendações de rastreabilidade e reprodutibilidade em pesquisa computacional ([SANDVE et al., 2013](#); [WILSON et al., 2017](#)).

Além da formulação metodológica, foi desenvolvido um software em Python com interface gráfica, capaz de aplicar receitas de pré-processamento, executar o *pipeline* de segmentação, visualizar etapas intermediárias, processar imagens em lote e calcular métricas quantitativas. O código-fonte é disponibilizado em repositório público ([MARQUES, 2024](#)). A avaliação foi conduzida sobre amostras selvagens e mutantes, comparando as predições do algoritmo com máscaras de referência produzidas manualmente no software AnyLabeling ([CVHub520 Contributors, 2026](#)). Dessa forma, o trabalho articula o desenvolvimento computacional da ferramenta, a análise geométrica das imagens e uma avaliação experimental no contexto da biologia do desenvolvimento.

1.1 Motivação e problema

A principal motivação deste trabalho está na necessidade de automatizar, de forma confiável e interpretável, a segmentação de núcleos em imagens microscópicas de embriões de *Drosophila melanogaster*. A identificação e a separação manuais demandam tempo, atenção especializada e podem variar entre anotadores, razão pela qual a própria definição da máscara de referência deve ser tratada como uma etapa metodológica sujeita a incerteza ([WARFIELD; ZOU; WELLS, 2004](#)). Em conjuntos maiores de imagens, esse procedimento torna-se pouco escalável e dificulta a obtenção sistemática de métricas quantitativas ([MEIJERING, 2012](#)).

Do ponto de vista computacional, o problema torna-se especialmente desafiador quando os núcleos aparecem muito próximos ou parcialmente encostados (SOUSA et al., 2020; SCHMIDT et al., 2018). Nessas condições, a binarização pode representar diversos núcleos como uma única região contínua, e a contagem de componentes conexas deixa de corresponder à quantidade real de instâncias (MEIJERING, 2012). Esse tipo de erro afeta análises posteriores, como estimativas de densidade celular, comparações entre amostras e avaliação de padrões espaciais (SOUSA; LOPES; TRAVENÇOLO, 2014; SOUSA et al., 2020).

Além disso, imagens biológicas reais apresentam variações de iluminação, baixo contraste, ruído, bordas pouco definidas e heterogeneidade morfológica, fatores que afetam diretamente a qualidade da segmentação (MEIJERING, 2012; SOUSA et al., 2020). Métodos baseados apenas em limiarização global, limiarização adaptativa ou operações morfológicas podem ser sensíveis a essas condições, sobretudo quando a separação entre objetos depende de detalhes sutis do contorno (GONZALEZ; WOODS, 2002; SAUVOLA; PIETIKÄINEN, 2000; MEIJERING, 2012).

Uma alternativa consiste em explorar propriedades geométricas da forma segmentada. Quando dois ou mais núcleos estão em contato, o contorno externo do agrupamento pode apresentar reentrâncias ou concavidades nas regiões de encontro entre as estruturas; essas concavidades funcionam como indícios de possíveis pontos de separação (SOUSA; LOPES; TRAVENÇOLO, 2014; SOUSA et al., 2020). Entretanto, sua detecção e a definição dos cortes não são triviais, pois ruído, artefatos de aquisição e imperfeições do pré-processamento também podem produzir irregularidades no contorno (SOUSA et al., 2020; MEIJERING, 2012).

Nesse cenário, este trabalho propõe uma abordagem baseada na análise geométrica dos contornos, na modelagem polar do sinal $r(\theta)$ e na detecção de concavidades associadas a possíveis regiões de separação, em continuidade à estratégia apresentada por Sousa et al. (SOUSA; LOPES; TRAVENÇOLO, 2014; SOUSA et al., 2020). A abordagem define regras de pareamento vetorial entre concavidades externas e pontos candidatos internos. Quando esse pareamento não satisfaz os critérios de aceitação, aplica-se um mecanismo de *fallback*, no qual o corte é direcionado a um centro de referência da estrutura. Com isso, busca-se aumentar a robustez da individualização em regiões de alta densidade celular.

1.2 Objetivos

1.2.1 Objetivo geral

O objetivo geral deste trabalho é propor, desenvolver e avaliar uma abordagem computacional para segmentação de instâncias nucleares em imagens microscópicas de

embriões de *Drosophila melanogaster*. A abordagem fundamenta-se na análise geométrica de contornos, na detecção de concavidades e na aplicação de cortes de separação entre estruturas adjacentes. Além disso, incorpora-se um mecanismo de recuo (*fallback*) para lidar com situações em que a geometria da imagem não fornece um pareamento interno suficientemente confiável para a definição direta dos cortes.

1.2.2 Objetivos específicos

Para alcançar o objetivo geral, definem-se os seguintes objetivos específicos:

- desenvolver um *pipeline* de pré-processamento capaz de converter as imagens microscópicas originais em máscaras binárias adequadas à análise geométrica;
- implementar uma estratégia de detecção de concavidades baseada na representação polar do contorno e em referências geométricas, como o envoltório convexo e o ajuste de elipse;
- definir regras de pareamento vetorial para propor cortes de separação entre núcleos adjacentes, considerando proximidade, direção, alinhamento e coerência geométrica;
- incorporar um mecanismo de recuo (*fallback*) para os casos em que o pareamento geométrico direto não satisfaz os critérios mínimos de aceitação;
- implementar a abordagem em um software desenvolvido em Python, com interface gráfica, parametrização por receitas e processamento em lote; e
- avaliar quantitativa e qualitativamente o método, tanto em nível de massa segmentada quanto em nível de instâncias nucleares.

1.3 Contribuições

As contribuições deste trabalho concentram-se na construção de uma abordagem geométrica interpretável para a segmentação de núcleos em imagens microscópicas de embriões de *Drosophila melanogaster*. Diferentemente de soluções que funcionam como modelos fechados ou altamente dependentes de grandes bases anotadas, o método proposto explicita suas decisões por meio da análise de contornos, concavidades, vetores de direção e cortes geométricos. Essa característica favorece a inspeção dos resultados e permite compreender com maior clareza por que determinados cortes são aplicados ou rejeitados.

Uma contribuição central é a incorporação de um mecanismo autoral de *fallback* geométrico ao processo de segmentação, proposto e implementado por João Pedro de Oliveira Marques. Esse recurso atua em situações nas quais o pareamento direto entre concavidades externas e contornos internos não é suficientemente confiável. Nesses casos,

o método utiliza o centro de referência obtido pelo ajuste elíptico e orienta um corte alternativo segundo o padrão geométrico da elipse, com a finalidade de melhorar a definição da separação em agrupamentos celulares densos e regiões de ambiguidade morfológica.

O *pipeline* parametrizável de pré-processamento também constitui uma contribuição. Ele reúne filtragem, binarização adaptativa, remoção de ruídos, operações morfológicas e refinamento por elipse. As receitas em formato JSON permitem repetir a mesma sequência em diferentes imagens e ajustar os parâmetros de forma controlada.

Além da formulação do método, este trabalho contribui com o desenvolvimento de uma ferramenta em Python com interface gráfica. O software permite gerar receitas de processamento, executar a segmentação em lote, visualizar etapas intermediárias, inspecionar resultados e apoiar a calibração dos parâmetros. Essa implementação aproxima o método de um fluxo de uso prático, especialmente em cenários nos quais o usuário precisa testar diferentes configurações antes de consolidar uma análise.

Por fim, o trabalho também contribui com um módulo de avaliação quantitativa, capaz de comparar as máscaras produzidas automaticamente com anotações manuais de referência. A avaliação considera tanto a correspondência da massa segmentada quanto o desempenho em nível de instâncias, permitindo uma leitura mais completa do comportamento do algoritmo. A análise experimental com amostras selvagens e mutantes evidencia os pontos fortes da abordagem, suas limitações e os caminhos mais promissores para calibração e aprimoramento futuro.

1.4 Organização do trabalho

Este trabalho está organizado em seis capítulos. O Capítulo 1 apresenta a introdução, a motivação, o problema investigado, os objetivos e as contribuições da pesquisa. O Capítulo 2 descreve os fundamentos teóricos relacionados à segmentação de imagens, pré-processamento, contornos, envoltório convexo, concavidades, representação polar e métricas de avaliação. O Capítulo 3 apresenta os materiais utilizados e detalha os métodos desenvolvidos, incluindo o *pipeline* de pré-processamento, a filtragem por elipse, a limiarização adaptativa local, a modelagem do sinal e as estratégias de segmentação. O Capítulo 4 descreve o software desenvolvido em Python, sua arquitetura, seus módulos e suas funcionalidades de interface gráfica e processamento em lote. O Capítulo 5 apresenta os resultados quantitativos e qualitativos, incluindo a análise visual da amostra com melhor desempenho. Por fim, o Capítulo 6 apresenta as conclusões do trabalho e discute possibilidades de continuidade em pesquisas futuras.

2 Fundamentação Teórica

A fundamentação teórica deste trabalho aborda os principais conceitos e técnicas relacionados à segmentação de núcleos contidos em imagens biológicas, com ênfase em métodos baseados em análise geométrica de contornos.

2.1 Segmentação de imagens

A segmentação pode ser entendida como o processo de particionar uma imagem em regiões semanticamente relevantes, atribuindo rótulos a pixels a fim de separar objetos de interesse do fundo e, quando necessário, separar instâncias distintas de um mesmo tipo de objeto ([GONZALEZ; WOODS, 2002](#); [SHAPIRO; STOCKMAN, 2001](#)).

No contexto de imagens biológicas obtidas por microscopia, a segmentação de estruturas celulares constitui uma etapa crítica para análises quantitativas, permitindo a extração de características morfológicas e a investigação de padrões espaciais. No entanto, esse processo é particularmente desafiador devido a fatores como variações de iluminação e contraste, ruídos de aquisição, presença de texturas do fundo, bordas incompletas e, principalmente, o contato e a sobreposição parcial entre núcleos em cenários de alta densidade ([SOUZA et al., 2020](#)). Em tal caso, métodos baseados exclusivamente em limiarização tendem a produzir uma única componente conexa representando múltiplos objetos. Diante disso, torna-se necessário explorar evidências adicionais de separação, como concavidades no contorno, indícios topológicos e regularidades geométricas associadas à forma dos objetos.

2.2 Pré-processamento e binarização

Para que técnicas baseadas em contornos operem de forma estável, é necessário transformar a imagem original em uma máscara binária, na qual os objetos de interesse estejam representados em primeiro plano e o fundo seja suprimido. Em imagens microscópicas, essa etapa é particularmente importante, pois ruídos de aquisição, variações locais de iluminação, baixo contraste e irregularidades no fundo podem comprometer diretamente a qualidade dos contornos extraídos. Assim, o pré-processamento atua como uma etapa preparatória para reduzir interferências e tornar a binarização mais robusta ([GONZALEZ; WOODS, 2002](#)).

Entre as operações clássicas de suavização, destacam-se o filtro Gaussiano e o filtro de mediana. O filtro Gaussiano realiza a suavização da imagem por meio da convolução

com uma função Gaussiana bidimensional, atenuando variações abruptas de intensidade e reduzindo componentes de alta frequência. Essa operação é útil para diminuir pequenas oscilações locais antes da limiarização, embora possa provocar certo borramento das bordas quando aplicada com *kernels* muito grandes. O filtro de mediana, por sua vez, substitui cada pixel pelo valor mediano de sua vizinhança, sendo particularmente eficaz na remoção de ruídos impulsivos e na suavização de irregularidades locais sem comprometer de forma tão intensa as bordas dos objetos (GONZALEZ; WOODS, 2002).

Após a etapa de suavização, aplica-se um método de limiarização para separar as regiões de interesse do fundo da imagem. Dois métodos frequentemente empregados nesse contexto são o método de Otsu (OTSU, 1979) e o método de Sauvola (SAUVOLA; PIETIKÄINEN, 2000). O método de Otsu define automaticamente um limiar global a partir do histograma de intensidades da imagem, buscando maximizar a separabilidade entre as classes de fundo e objeto. Essa abordagem tende a apresentar bons resultados quando há uma distinção razoável entre as distribuições de intensidade das regiões de interesse e do fundo.

Entretanto, em imagens biológicas reais, a iluminação e o contraste podem variar significativamente ao longo da amostra. Nesses casos, métodos globais podem apresentar desempenho limitado, pois um único limiar pode não ser adequado para todas as regiões da imagem. O método de Sauvola, por outro lado, calcula limiares locais com base em estatísticas da vizinhança, como média e desvio padrão dos níveis de cinza. Essa característica torna a limiarização adaptativa especialmente relevante para imagens de embriões de *Drosophila melanogaster*, nas quais podem ocorrer variações espaciais de iluminação, contraste e densidade celular. Dessa forma, a combinação entre filtragem e limiarização adaptativa constitui uma etapa fundamental para gerar máscaras binárias adequadas à posterior análise geométrica dos contornos.

2.3 Contornos e representação hierárquica

Após a obtenção da imagem binária, uma representação eficiente dos objetos pode ser construída por meio da extração de seus contornos. Esse processo permite identificar as fronteiras das regiões segmentadas, fornecendo uma descrição geométrica adequada para análises posteriores. Em bibliotecas computacionais como o OpenCV recomenda-se o uso de imagens binárias para maior acurácia (OPENCV, 2026b). Além disso, é possível recuperar hierarquia entre contornos, o que é fundamental para lidar com buracos e estruturas internas. O modo `RETR_TREE` constrói uma hierarquia completa de contornos, permitindo identificar contornos internos como filhos do contorno externo (OPENCV, 2026a; OPENCV, 2026c). Esse aspecto é relevante neste trabalho pois, quando existe um contorno interno confiável, ele pode ser usado como alvo geométrico para cortes

de separação no mecanismo de pareamento, em contraposição ao mecanismo de recuo (*fallback*), que recorre ao centro de referência.

2.4 Envoltório convexo e detecção de concavidades

O envoltório convexo de um conjunto de pontos é definido como o menor polígono convexo capaz de conter todos os pontos considerados. No contexto de contornos, ele fornece uma aproximação convexa da forma do objeto, sem reentrâncias. Em ferramentas como o OpenCV, a função `convexHull` computa o envoltório convexo de um conjunto de pontos bidimensionais e pode retornar índices correspondentes aos pontos do contorno original (OPENCV, 2026c). As concavidades no contorno podem ser interpretadas como regiões onde o contorno real se afasta para dentro em relação ao *hull*. Essas reentrâncias são particularmente relevantes na segmentação de objetos em contato, uma vez que frequentemente indicam regiões onde duas ou mais instâncias se encontram adjacentes. Dessa forma, a identificação de concavidades constitui uma evidência geométrica importante para a separação de objetos.

2.5 Parametrização polar do contorno

Seja um centro de referência $\mathbf{c} = (c_x, c_y)$ associado a um objeto. Para cada ponto do contorno $\mathbf{p}_i = (x_i, y_i)$, pode-se definir uma representação polar baseada na distância radial ao centro e no ângulo correspondente. Ao ordenar os pontos do contorno pelo ângulo polar, obtém-se um sinal unidimensional que descreve o contorno como variação do raio. Essa transformação permite tratar o problema geométrico de detecção de concavidades como um problema de análise de sinais. Nesse contexto, as concavidades aparecem como variações positivas em um sinal de diferença entre uma tendência (*baseline* baseada no *convex hull*) e o sinal original. Após truncar valores negativos, esse procedimento reduz o problema geométrico de localizar concavidades a um problema algorítmico de localizar picos locais.

2.6 Estratégias de separação de instâncias

A partir da identificação de concavidades no contorno externo, podem ser definidas estratégias para separar objetos que aparecem em contato ou parcialmente fundidos na máscara binária. Em problemas de segmentação de instâncias, essas concavidades são relevantes porque frequentemente indicam regiões de encontro entre estruturas adjacentes. Assim, a detecção de reentrâncias no contorno pode fornecer indícios geométricos para a definição de cortes de separação.

Neste trabalho, a estratégia de separação considera dois cenários principais. No primeiro, quando há contornos internos confiáveis, busca-se estabelecer correspondências entre concavidades externas e pontos candidatos presentes nesses contornos. Para isso, são avaliados critérios como proximidade espacial, direção do vetor de corte, alinhamento entre normais e coerência geométrica. A escolha do ponto candidato é orientada por medidas vetoriais, incluindo o produto escalar, de modo que cortes mais bem alinhados com a morfologia local sejam priorizados.

No segundo cenário, quando não há contornos internos adequados ou quando o pareamento geométrico não satisfaz os critérios mínimos de aceitação, aplica-se uma estratégia alternativa baseada em um centro de referência. Esse mecanismo, denominado recuo ou *fallback*, define o ponto final do corte em direção ao centroide ou centro geométrico estimado da estrutura segmentada. A ação é formalizada no Algoritmo 1, que explicita a condição de acionamento e a construção do segmento alternativo.

Algoritmo 1 Geração de corte por pareamento direto ou por *fallback*.

Entrada: concavidade externa p , conjunto de candidatos internos Q , centro de referência c , critérios de aceitação T

Saída: segmento de corte s ou indicação de ausência de corte

- 1: avaliar cada candidato $q \in Q$ segundo distância, direção, ortogonalidade e coerência geométrica
 - 2: **if** existe candidato q^* que satisfaz T **then**
 - 3: $s \leftarrow \overline{pq^*}$ ▷ pareamento geométrico direto
 - 4: **else**
 - 5: $s \leftarrow \overline{pc}$ ▷ mecanismo de recuo
 - 6: **end if**
 - 7: recortar s aos limites válidos da componente segmentada
 - 8: **return** s
-

Dessa forma, o método mantém a capacidade de propor separações mesmo em regiões de elevada ambiguidade geométrica, nas quais os indícios internos não são suficientemente informativos.

Essas duas estratégias são complementares. O pareamento com contornos internos busca produzir cortes mais específicos e aderentes à geometria local, enquanto o mecanismo de *fallback* atua como uma alternativa para evitar que regiões densas permaneçam sem tentativa de separação. No entanto, como qualquer heurística geométrica, a qualidade dos cortes depende da estabilidade da máscara binária, da definição das concavidades e da calibração dos parâmetros utilizados no processo.

2.7 Métricas de avaliação

A avaliação de métodos de segmentação requer métricas quantitativas capazes de comparar as predições automáticas com anotações de referência previamente definidas, geralmente denominadas *ground truth*. Neste trabalho, a avaliação foi conduzida em dois níveis complementares: o nível de massa segmentada, voltado à concordância espacial entre máscaras binárias, e o nível de instâncias, voltado à verificação da individualização correta dos núcleos.

Para avaliar a correspondência espacial entre a máscara predita e a máscara de referência, utiliza-se o coeficiente Dice ([DICE, 1945](#); [TAHA; HANBURY, 2015](#)). Dadas duas regiões A e B , em que A representa a máscara predita e B representa a máscara de referência, o coeficiente Dice é definido por:

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (2.1)$$

Essa métrica assume valores entre 0 e 1, sendo que valores próximos de 1 indicam maior sobreposição entre a segmentação automática e a referência manual. O Dice é adequado para avaliar a qualidade da delimitação da massa nuclear total, pois considera a interseção entre os pixels segmentados e os pixels anotados como referência.

Outra métrica frequentemente utilizada em segmentação é a Interseção sobre União, ou *Intersection over Union* (IoU), também conhecida como índice de Jaccard ([JACCARD, 1912](#)). Essa métrica é definida por:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (2.2)$$

Embora o IoU seja uma métrica amplamente empregada, neste trabalho prioriza-se o coeficiente Dice na análise da massa segmentada, por sua interpretação direta em termos de sobreposição espacial entre a predição e a referência.

Para avaliar o desempenho em nível de instâncias, utiliza-se uma análise baseada em verdadeiros positivos (TP), falsos positivos (FP) e falsos negativos (FN). Um verdadeiro positivo corresponde a uma instância nuclear predita que possui correspondência válida com uma instância anotada na referência. Um falso positivo corresponde a uma instância predita sem correspondência adequada na referência, enquanto um falso negativo corresponde a uma instância anotada manualmente que não foi recuperada pelo algoritmo.

A partir desses valores, calculam-se a precisão, a revocação e o *F1-Score*:

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (2.3)$$

$$\text{Revocação} = \frac{TP}{TP + FN} \quad (2.4)$$

$$F1 = 2 \cdot \frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (2.5)$$

O *F1-Score* combina precisão e revocação em uma única medida, sendo útil quando se deseja avaliar simultaneamente a ocorrência de detecções indevidas e a perda de instâncias reais. Dessa forma, enquanto o coeficiente Dice avalia a fidelidade espacial da massa nuclear segmentada, o *F1-Score* por instância permite analisar a capacidade do método de separar corretamente núcleos individuais. Essa distinção é importante porque uma segmentação pode apresentar boa sobreposição global com a referência e, ainda assim, falhar na individualização de núcleos adjacentes.

3 Materiais e Métodos

Este capítulo descreve os materiais e os procedimentos metodológicos empregados no desenvolvimento e na avaliação da abordagem proposta. Inicialmente, apresentam-se os dados e o ambiente computacional. Em seguida, detalha-se o *pipeline* de segmentação e cada uma de suas etapas.

3.1 Materiais

As imagens analisadas neste trabalho correspondem a embriões de *Drosophila melanogaster* e são cedidas pelo grupo de pesquisa coordenado pela Profa. Helena Araujo, da Universidade Federal do Rio de Janeiro (UFRJ). Trata-se de uma base de imagens de acesso restrito da UFRJ, cedida para a realização desta pesquisa. O conjunto de dados contém 51 imagens, sendo 26 amostras selvagens (*Wild Type* – WT) e 25 amostras mutantes. As amostras selvagens representam embriões com fenótipo considerado padrão e funcionam como referência para a organização celular esperada. As amostras mutantes apresentam alterações genéticas induzidas e podem exibir modificações morfológicas e desorganizações teciduais observáveis em microscopia. Por esse motivo, constituem cenários de segmentação mais variáveis e desafiadores.

A Figura 1 apresenta uma das imagens microscópicas utilizadas e fornece uma referência visual para a densidade e a proximidade das estruturas nucleares descritas ao longo do capítulo.

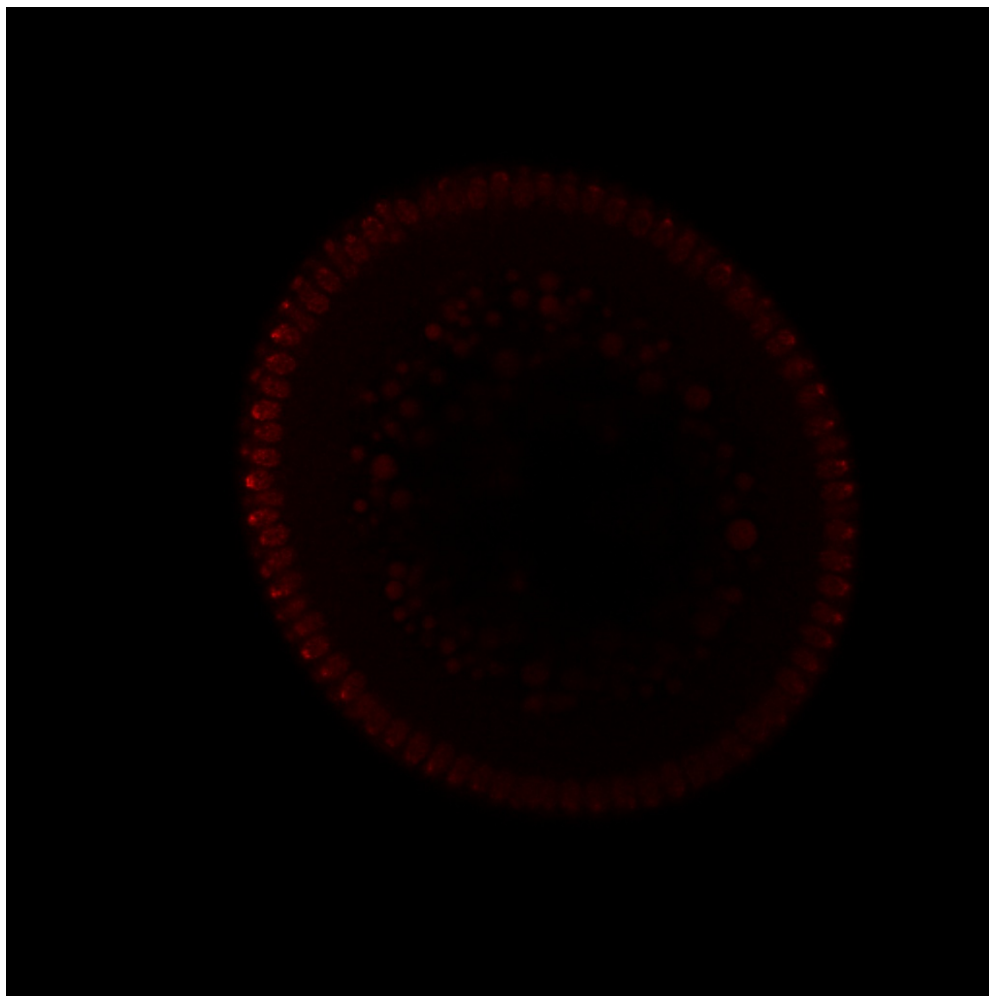


Figura 1 – Exemplo de imagem microscópica de embrião de *Drosophila melanogaster* utilizada no conjunto experimental.

O *pipeline* computacional é desenvolvido em linguagem Python e utiliza bibliotecas de processamento matricial e visual, como NumPy, OpenCV e SciPy (HARRIS et al., 2020; BRADSKI, 2000; VIRTANEN et al., 2020). Para construir a base de referência (*ground truth*), utiliza-se o software de anotação AnyLabeling (CVHub520 Contributors, 2026), no qual as bordas de referência são contornadas manualmente por meio de polígonos fechados, conforme ilustrado na Figura 2.

Embora o banco completo contenha 51 imagens, a avaliação quantitativa detalhada utiliza um subconjunto de 5 imagens selecionadas para contemplar cenários biológicos variados, sendo 3 amostras selvagens e 2 amostras mutantes. Essa escolha decorre do custo da anotação manual em nível de instância, uma vez que cada imagem pode conter dezenas ou centenas de núcleos próximos ou parcialmente sobrepostos. Assim, priorizam-se a qualidade das anotações manuais e a confiabilidade da comparação quantitativa em detrimento do volume amostral. Consequentemente, os resultados devem ser interpretados como uma avaliação experimental inicial da viabilidade da abordagem, e não como uma generalização estatística para todo o conjunto disponível.

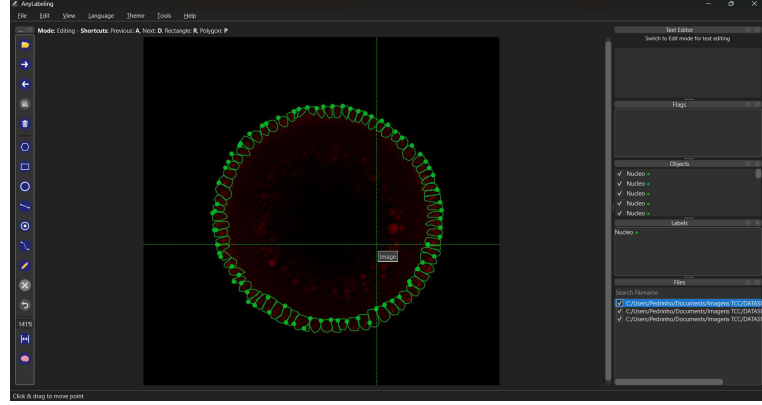


Figura 2 – Interface do software AnyLabeling utilizada para anotação manual das máscaras de referência.

3.2 Visão geral do *pipeline* de segmentação

De modo geral, o *pipeline* de segmentação proposto é composto por etapas sequenciais que visam transformar imagens de embriões da *Drosophila melanogaster* em uma representação segmentada (máscara binária) das instâncias nucleares. Inicialmente, aplica-se uma etapa de pré-processamento para redução de ruído e binarização. Em seguida, são extraídos os contornos das regiões segmentadas e estimado um centro de referência robusto. Posteriormente, realiza-se a parametrização polar do contorno e a construção do *baseline* para detecção de concavidades. A partir dessas informações, o método aplica heurísticas geométricas para geração e seleção de cortes de separação. Todas as etapas do *pipeline* operam sob um fator de escala dinâmico, de modo que os parâmetros espaciais permaneçam proporcionais à resolução da imagem. A Figura 3 sintetiza o fluxo metodológico completo, desde a entrada da imagem até a avaliação das máscaras e instâncias produzidas.

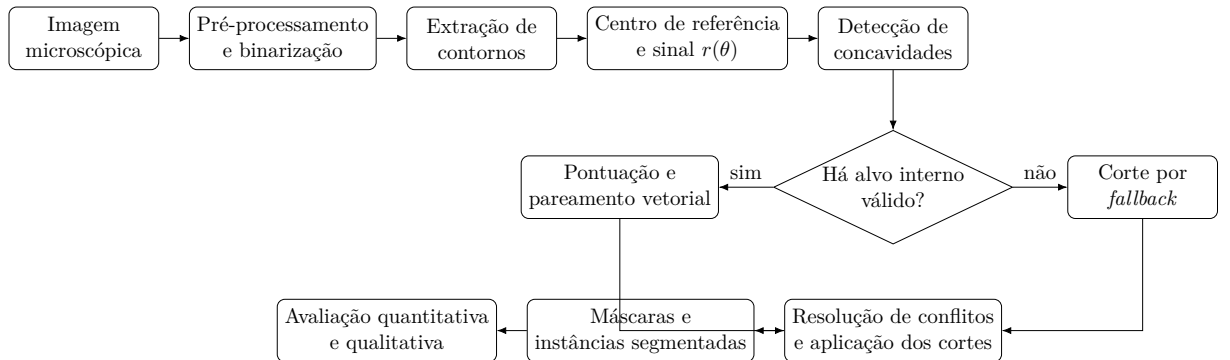


Figura 3 – Visão geral do *pipeline* de segmentação proposto.

3.3 Pré-processamento

A etapa de pré-processamento converte a imagem microscópica original em uma máscara binária adequada à análise geométrica subsequente e reduz interferências decorren-

tes de ruído, variações de iluminação e irregularidades do fundo. Utiliza-se uma receita de processamento parametrizada em arquivo JSON, denominada `Receita_Selvagens_mod2`, executada na ferramenta computacional desenvolvida neste trabalho. Essa estratégia padroniza a aplicação sequencial dos filtros e favorece a reprodutibilidade entre diferentes execuções do *pipeline*. A Figura 4 é apresentada no início da seção para antecipar a sequência de operações detalhada nos parágrafos seguintes.

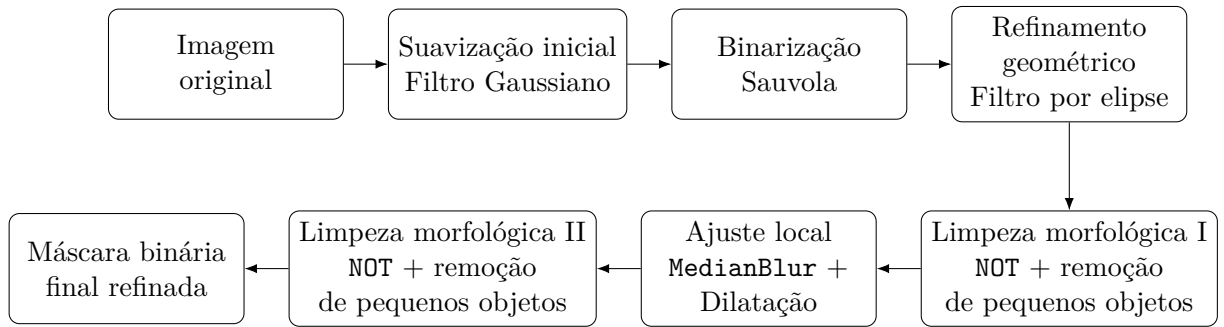


Figura 4 – Representação esquemática das fases do pré-processamento para geração da máscara binária refinada.

O pré-processamento é concebido como uma etapa estruturada, composta por uma sequência de operações complementares. Cada filtro aplicado possui uma função específica dentro do *pipeline*, e a ordem de aplicação influencia diretamente a qualidade da máscara resultante. O objetivo principal dessa etapa é gerar uma representação binária que preserve a massa nuclear real, ao mesmo tempo em que reduz ruídos e interferências externas. Além disso, busca-se obter contornos suficientemente estáveis para a etapa posterior de análise geométrica e segmentação baseada em concavidades.

Os parâmetros espaciais definidos na receita são interpretados em conjunto com um mecanismo de escala matricial implementado no sistema. Esse mecanismo ajusta automaticamente a atuação dos filtros de acordo com a resolução da imagem, garantindo que seus efeitos sejam proporcionais independentemente do tamanho da entrada. Assim, tamanhos de *kernel*, limiares de área e distâncias geométricas não são tratados como valores fixos, mas como medidas adaptadas dinamicamente com base nas dimensões da imagem processada.

Inicialmente, aplica-se um filtro Gaussiano (Seção 2.2) com *kernel* tamanho 11×11 e σ estimado automaticamente a partir do tamanho do kernel (implementado como $\sigma = 0$). Na prática, a aplicação desse filtro contribui para reduzir pequenos pontos brilhantes e oscilações locais, que poderiam comprometer o resultado da etapa subsequente de limiarização adaptativa. Em seguida, a imagem é submetida à binarização adaptativa local por meio do método de Sauvola (Seção 2.2), utilizando *kernel* de tamanho 59×59 e parâmetro de limiar 0,3. Essa combinação produz uma separação mais robusta entre regiões de interesse e fundo, mesmo na presença de heterogeneidades internas.

Após a binarização, a máscara resultante é submetida a uma etapa de filtragem geométrica baseada em ajuste elíptico, cujo objetivo é restringir e refinar a estrutura segmentada a partir de critérios de compatibilidade morfológica. Essa etapa atua como uma regularização geométrica da máscara, removendo regiões que não se ajustam ao padrão esperado da massa nuclear e preservando a componente principal associada à massa nuclear.

Na sequência, a máscara resultante passa por ciclos sucessivos de operações lógicas e morfológicas. Inicialmente, realiza-se uma inversão binária (**Bitwise NOT**), seguida da remoção de pequenos objetos com área máxima de 200 pixels. A inversão binária permite alternar a interpretação entre objeto e fundo, tornando possível remover tanto pequenos ruídos externos quanto de artefatos internos indesejados. A remoção de objetos pequenos elimina componentes desconectadas que não representam núcleos de interesse e que poderiam afetar a contagem final ou gerar falsos contornos.

Em seguida, aplica-se um filtro de mediana (Seção 2.2) com *kernel* de tamanho 9×9 . Essa operação contribui para suavizar serrilhados e oscilações locais na máscara binária, reduzindo a probabilidade de interpretá-las como concavidades na etapa subsequente de aplicação do módulo geométrico. Posteriormente, realiza-se uma dilatação morfológica com *kernel* de tamanho 3×3 e uma iteração. Essa operação tem como objetivo reforçar conexões locais frágeis e recuperar pequenas continuidades espaciais que possam ter sido comprometidas durante as etapas de binarização ou remoção de ruído. Tal procedimento é especialmente relevante em regiões de baixo contraste, onde separações muito estreitas podem levar a fragmentações artificiais da massa nuclear.

Por fim, executa-se uma nova sequência de inversão binária seguida da remoção de pequenos objetos, novamente limitada a componentes com no máximo 200 pixels. Essa etapa final atua como um refinamento adicional, removendo resíduos remanescentes e entregando uma máscara binária mais limpa, homogênea e apropriada para as fases subsequentes de extração de contornos, detecção de concavidades e aplicação dos cortes geométricos.

A relevância dessa sequência reside no fato de que as etapas geométricas subsequentes dependem diretamente da qualidade da máscara produzida. Caso a máscara apresente ruídos, falhas de preenchimento ou reentrâncias artificiais, o processo de detecção de concavidades pode interpretar tais imperfeições como indícios falsos de contato entre núcleos, induzindo falsos cortes. Por outro lado, quando o pré-processamento é bem ajustado, o contorno resultante tende a refletir mais fielmente a morfologia biológica dos agrupamentos celulares, favorecendo a aplicação consistente das heurísticas geométricas de segmentação.

3.3.1 Filtro por elipse

Entre as etapas de refinamento empregadas no pré-processamento, destaca-se o filtro geométrico baseado em ajuste elíptico, desenvolvido por João Pedro de Oliveira Marques no âmbito deste trabalho. O procedimento utiliza uma elipse de referência e limiares parametrizáveis para excluir ruídos externos à região perimetral, preservando componentes compatíveis com a estrutura nuclear principal. Essa etapa é necessária porque, após a binarização adaptativa, a máscara pode conter componentes desconectadas, fragmentos periféricos e regiões espúrias. A remoção exclusivamente por limiares de área seria inadequada, pois núcleos legítimos também podem aparecer como componentes menores ou temporariamente desconectadas da massa principal.

Para contornar essa limitação, a estratégia adotada utiliza a geometria da maior componente conexa como referência inicial. A máscara binária é inicialmente decomposta em suas componentes, sendo a maior delas assumida como a região nuclear principal ou como o núcleo inicial do agrupamento embrionário segmentado. A partir dos pontos dessa componente, ajusta-se uma elipse de referência. Essa elipse funciona como um modelo geométrico aproximado da distribuição espacial esperada da massa nuclear.

As demais componentes são então avaliadas com base em sua relação geométrica com essa elipse. Componentes próximas da maior região e com alinhamento espacial compatível com a banda elíptica são consideradas candidatas a pertencer ao agrupamento nuclear. Em contrapartida, componentes distantes ou geometricamente inconsistentes são interpretadas como ruídos ou artefatos de binarização e, portanto, descartadas.

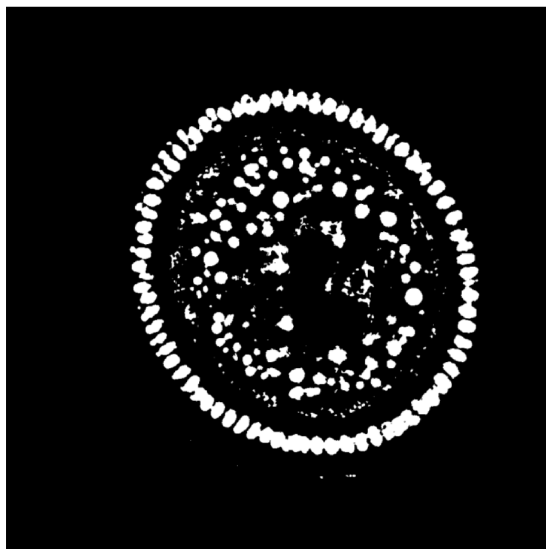
Esse processo é realizado de forma iterativa. A cada iteração, as componentes aceitas são incorporadas ao conjunto principal e uma nova elipse é ajustada com base no conjunto expandido. Dessa forma, a elipse inicial, ajustada apenas sobre a maior componente, passa a representar progressivamente uma região mais ampla e mais fiel à organização global da amostra. O processo é encerrado quando não há novas incorporações ou quando se atinge o limite máximo de iterações.

Os principais parâmetros utilizados nessa etapa incluem: espessura de banda elíptica igual a 255, fator de dilatação adicional igual a 10, distância-base igual a 18, limiar mínimo de compatibilidade geométrica igual a 0,2, área mínima relativa igual a 0,01, número máximo de iterações igual a 20, amostragem do contorno a cada 3 pontos, uso de *gap* dinâmico, uso de **MAD** dinâmico e manutenção da maior componente conexa. Esses parâmetros controlam, respectivamente, a tolerância espacial em torno da elipse, a flexibilidade do modelo, os critérios de aceitação e os limites de crescimento do processo iterativo.

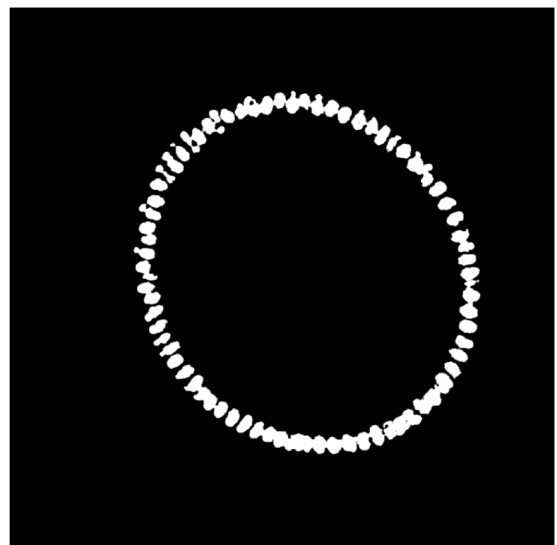
Essa abordagem difere de uma simples filtragem por tamanho porque combina critérios de conectividade, distância e forma. Assim, componentes pequenas podem ser

preservadas caso estejam próximas da massa principal e apresentem compatibilidade com o modelo elíptico em construção. Por outro lado, componentes maiores podem ser descartadas quando apresentam comportamento incompatível com a distribuição espacial esperada da estrutura nuclear.

A Figura 5 ilustra qualitativamente essa etapa em uma imagem da base. Na Figura 5a, observa-se a máscara binária antes da aplicação do filtro por elipse, ainda contendo ruídos, componentes periféricas e fragmentos desconectados. Na Figura 5b, apresenta-se a máscara após o processo iterativo de ajuste e seleção, na qual a estrutura principal é preservada e os objetos incompatíveis com a elipse são removidos.



(a) Máscara binária antes da aplicação do filtro por elipse.



(b) Máscara binária resultante após a aplicação do filtro por elipse.

Figura 5 – Resultado qualitativo da aplicação do filtro por elipse antes e após o refinamento geométrico.

Ressalta-se que o filtro por elipse não tem como finalidade realizar a segmentação individual dos núcleos. Sua função é produzir uma máscara binária mais coerente e representativa da massa nuclear para as etapas seguintes, reduzindo a presença de ruídos e artefatos que poderiam interferir na análise geométrica. Com isso, reduz-se a probabilidade de que ruídos periféricos sejam interpretados como partes válidas do contorno ou como falsas concavidades durante a etapa de segmentação.

3.4 Limiarização adaptativa local

Embora o pré-processamento produza, na maioria dos casos, uma máscara binária adequada para a análise geométrica, verifica-se que alguns aglomerados celulares altamente densos permanecem representados como componentes conexas compactas, sem a presença de cavidades internas ou separações discerníveis entre os núcleos adjacentes. Tal

situação compromete a aplicação da estratégia de pareamento proposta por (SOUSA et al., 2020), uma vez que o algoritmo deixa de dispor de alvos confiáveis para a projeção dos vetores de corte. Para contornar essa limitação, a abordagem proposta incorpora uma etapa complementar de limiarização adaptativa local, aplicada seletivamente às regiões problemáticas. Inicialmente, a máscara binária global é decomposta em suas componentes conexas, que são analisadas individualmente quanto à presença de reentrâncias internas relevantes. Quando uma componente não exibe contornos internos suficientes para suportar a etapa de segmentação geométrica, define-se uma região de interesse (*Region of Interest* – ROI) correspondente ao seu recorte espacial. Essa estratégia permite restringir o novo processo de binarização à área estritamente necessária, reduzindo a influência de regiões externas e aumentando a sensibilidade do método às variações locais de intensidade dentro do aglomerado.

Dentro da ROI extraída, realiza-se um novo ciclo de processamento composto por suavização e limiarização adaptativa local. Inicialmente, aplica-se um desfoque Gaussiano com o objetivo de reduzir pequenas variações de intensidade e atenuar ruídos locais. Em seguida, emprega-se novamente o método de Sauvola (SAUVOLA; PIETIKÄINEN, 2000), agora em escala local, com parâmetros recalibrados em função das dimensões da própria ROI. A adoção dessa abordagem justifica-se pelo fato de que tanto a intensidade luminosa quanto a densidade celular podem variar significativamente entre diferentes aglomerados. Dessa forma, uma parametrização global fixa tende a apresentar desempenho limitado quando aplicada a regiões muito pequenas ou excessivamente compactas.

Assim como na etapa de pré-processamento global, os parâmetros espaciais dessa nova binarização não são tratados como constantes absolutas. Os tamanhos de *kernel* utilizados na suavização e na limiarização adaptativa local são redimensionados pelo módulo `utils_scaling.py`, preservando a proporcionalidade em relação à escala de referência da imagem ou da ROI. Desse modo, a atuação dos filtros permanece morfológicamente consistente, evitando que janelas muito grandes eliminem detalhes relevantes ou que janelas muito pequenas reforcem irregularidades irrelevantes do contorno.

Após a obtenção da nova máscara local, realiza-se uma operação de intersecção lógica entre essa máscara refinada e a silhueta binária original do aglomerado. Essa etapa funciona como um mecanismo de preservação topológica, uma vez que permite incorporar refinamentos internos sem alterar os limites externos previamente definidos. Em termos práticos, o refinamento local passa a atuar apenas sobre a organização interna da componente, preservando seus limites externos.

O objetivo dessa etapa é favorecer o surgimento de descontinuidades que evidenciem possíveis separações entre núcleos adjacentes. Dessa forma, a limiarização adaptativa local não substitui a segmentação global, mas funciona como um refinamento seletivo aplicado apenas quando necessário. Com isso, amplia-se a disponibilidade de estruturas internas

úteis para as etapas subsequentes de modelagem polar e pareamento vetorial, aumentando a robustez do método em cenários de alta densidade celular.

3.5 Representação espacial e modelagem do sinal

A partir da máscara binária refinada, extrai-se o contorno externo da componente conexa de interesse. Esse contorno fornece a representação geométrica da fronteira do agrupamento nuclear e constitui a base para a etapa de detecção de concavidades. Para tornar essa análise menos dependente da posição absoluta da estrutura na imagem, estima-se inicialmente um centro de referência associado ao objeto segmentado. Neste trabalho, esse centro é obtido a partir do ajuste de uma elipse ao contorno, procedimento disponibilizado pelas funções de análise de formas do OpenCV ([OPENCV, 2026c](#)).

Seja $\mathbf{c} = (c_x, c_y)$ o centro de referência estimado e seja $\mathbf{p}_i = (x_i, y_i)$ um ponto pertencente ao contorno. Cada ponto do contorno é representado em coordenadas polares por meio do raio r_i , calculado pela Equação 3.1, e do ângulo θ_i , calculado pela Equação 3.2:

$$r_i = \sqrt{(x_i - c_x)^2 + (y_i - c_y)^2} \quad (3.1)$$

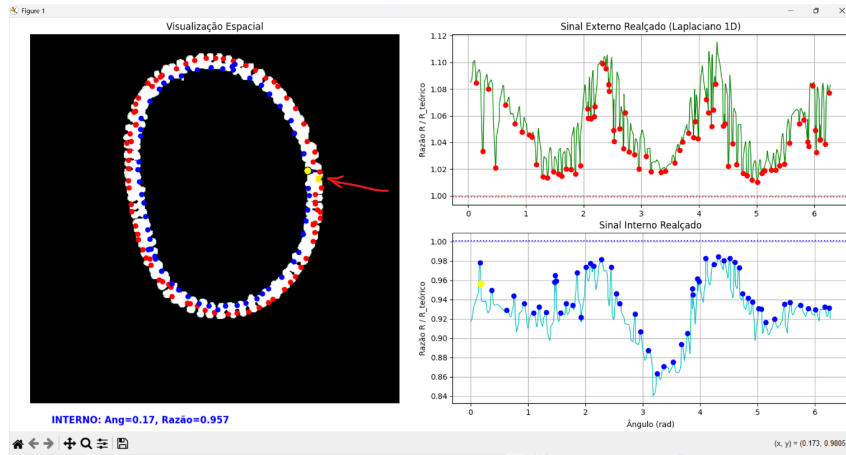
$$\theta_i = \text{atan2}(y_i - c_y, x_i - c_x) \quad (3.2)$$

Ao ordenar os pontos do contorno pelo ângulo θ_i , obtém-se um sinal unidimensional $r(\theta)$, que descreve a variação radial do contorno em torno do centro de referência. Essa transformação permite tratar a análise da forma como um problema de processamento de sinais, no qual variações abruptas no raio podem indicar regiões de curvatura acentuada ou possíveis concavidades.

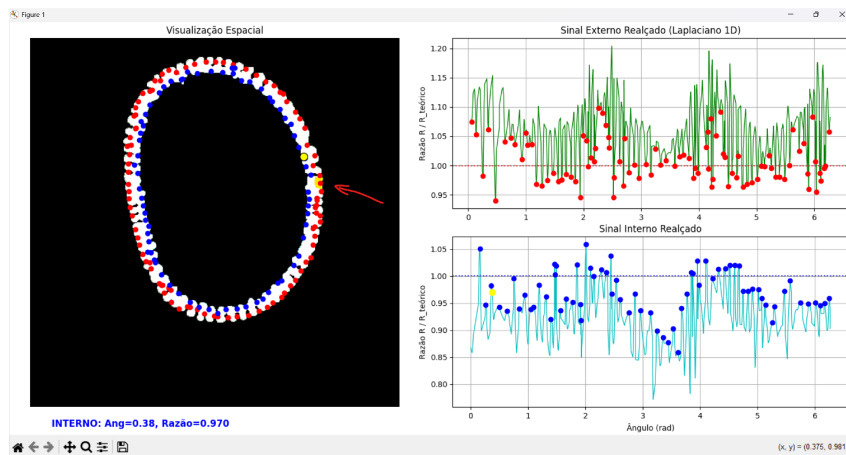
Para realçar essas variações, aplica-se o cálculo da segunda derivada sobre o sinal radial. A segunda derivada permite evidenciar mudanças na taxa de variação do raio, destacando pontos em que o contorno apresenta alterações geométricas relevantes. Em termos práticos, regiões onde o contorno sofre inflexões ou reentrâncias tendem a manifestar-se como picos no sinal processado, facilitando a identificação de candidatos a pontos de separação.

Além disso, a comparação entre o contorno real e uma referência baseada no envoltório convexo permite estimar desvios radiais associados a concavidades. O envoltório convexo representa uma aproximação sem reentrâncias da forma analisada; portanto, diferenças entre essa referência e o contorno original podem indicar regiões em que a estrutura se desloca para o interior da forma convexa. Essas regiões são interpretadas como potenciais evidências de contato entre núcleos adjacentes.

A Figura 6, gerada automaticamente pelo software desenvolvido por João Pedro de Oliveira Marques, compara o sinal polar antes e após o realce pela segunda derivada, destacando regiões candidatas à segmentação geométrica.



(a) Sinal polar 1D sem realce da segunda derivada.



(b) Sinal polar 1D realçado pela segunda derivada.

Figura 6 – Comparativo do sinal polar utilizado na análise geométrica do contorno.

3.6 Heurística de pontuação global e avaliação vetorial

Uma vez identificadas as concavidades, o sistema avalia múltiplos candidatos a cortes de separação entre os núcleos. Para cada concavidade externa, são avaliadas possíveis conexões com pontos candidatos, provenientes de contornos internos ou regiões centrais. Isto é, mapeia-se a distância geométrica em direção a todos os vértices do contorno interno, desconsiderando alvos que ultrapassem um raio limite previamente determinado.

Para avaliar a qualidade morfológica do corte, define-se uma equação de pontuação ponderada. O critério de maior relevância é a ortogonalidade de incidência, que avalia o alinhamento entre o vetor normal ao contorno externo na concavidade e o vetor de conexão até o ponto candidato. Além disso, considera-se o grau de perpendicularidade com que esse vetor intercepta o contorno interno. Em termos práticos, busca-se que o corte ocorra

de forma aproximadamente perpendicular às superfícies envolvidas, o que tende a produzir separações mais coerentes do ponto de vista morfológico.

Adicionalmente, pontuam-se o paralelismo absoluto entre a normal externa e a normal interna, o alinhamento radial do vetor de corte em relação ao centroide da elipse ajustada ao contorno e um fator de recompensa por proximidade espacial. Esses critérios complementares ajudam a refinar a escolha dos cortes mais adequados. Além da pontuação, o algoritmo impõe condições mínimas de aceitação para cada métrica considerada. Dessa forma, um candidato só é considerado válido se satisfizer simultaneamente todos os limiares estabelecidos, assegurando que os cortes resultem em separações com geometria compatível com a estrutura esperada dos núcleos.

3.7 Resolução de conflitos espaciais

A resolução dos conflitos espaciais parte da estratégia de pareamento e seleção apresentada por Sousa et al. (SOUSA; LOPES; TRAVENÇOLO, 2014; SOUSA et al., 2020), vinculada à linha de pesquisa da orientadora deste trabalho, e incorpora as adaptações implementadas neste TCC. A geração de múltiplos candidatos a cortes pode resultar em conflitos espaciais, especialmente quando diferentes pares de pontos produzem segmentos que se intersectam. Para resolver esse problema, o algoritmo ordena decrescentemente o catálogo global de escores, priorizando o processamento de ligações geometricamente ótimas, definidas aqui como aquelas que apresentam a menor distância euclidiana e a maior ortogonalidade em relação ao contorno do aglomerado. Uma vez que o vetor líder reivindica seus pontos de origem e destino, estes vértices são descartados das iterações subsequentes.

Para evitar interseções entre segmentos, instituiu-se uma malha de verificação algébrica anti-cruzamento baseada na orientação anti-horária de três pontos (CCW). Antes de registrar definitivamente uma nova linha de ruptura, os segmentos propostos são testados matematicamente contra todos os cortes de ordem superior já aprovados. Caso seja detectada interseção entre segmentos, o candidato é descartado.

Nos casos em que uma concavidade não encontra um pareamento válido, aplica-se o mecanismo de recuo (*fallback*), no qual o sistema traça um corte da concavidade ao centro de referência obtido pelo ajuste elíptico. Essa orientação segue o padrão geométrico da elipse e mantém a continuidade do processo de segmentação. A Figura 7 exibe o resultado da aplicação desse procedimento em uma amostra da base.

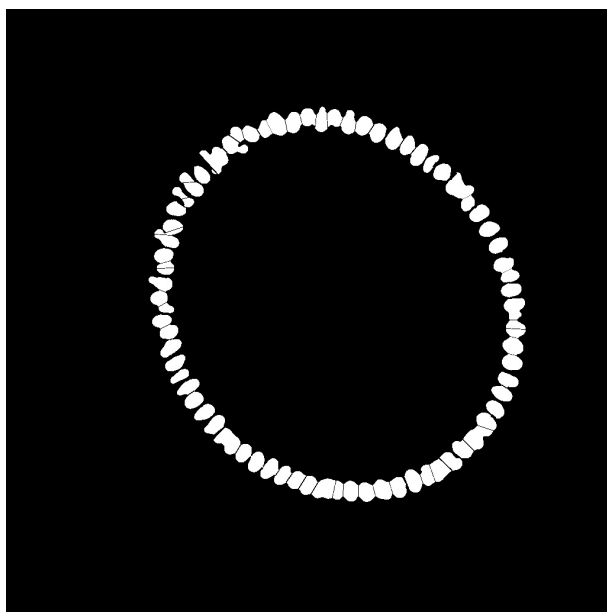


Figura 7 – Resultado da segmentação após resolução de conflitos e aplicação do mecanismo de *fallback*.

4 Desenvolvimento do Software em Python

Este capítulo descreve a implementação computacional da abordagem proposta, incluindo a arquitetura do sistema, os principais módulos desenvolvidos e as funcionalidades voltadas à interação com o usuário.

4.1 Arquitetura e módulos do projeto

O sistema foi desenvolvido em linguagem Python, adotando o paradigma de programação orientado a objetos, com o objetivo de promover organização, reutilização de código e separação de responsabilidades lógicas e visuais em módulos independentes, cada um responsável por uma etapa específica do *pipeline* de segmentação.

- O módulo `utils_scaling.py` é responsável pelo controle de escala das imagens e ajusta automaticamente os parâmetros espaciais conforme as dimensões da entrada. Esse módulo não realiza, por si só, a calibração experimental: ele preserva a proporcionalidade dos parâmetros. A calibração ocorre pela definição dos valores das receitas JSON e pelos ajustes realizados na interface gráfica, que depois são convertidos pelo módulo para a escala efetiva de cada imagem ou ROI.
- O módulo `analisa_concavidades_anel.py` implementa os procedimentos de análise geométrica, incluindo a parametrização polar dos contornos, a detecção de concavidades e a manipulação de elipses.
- O módulo `segmentacao_aneis.py` concentra a lógica de segmentação, sendo responsável pela aplicação das heurísticas de pareamento, geração de cortes e resolução de conflitos espaciais.
- Por fim, menciona-se que a criação dos artefatos interativos e janelas ficou sob a responsabilidade do framework padrão do Python, o `Tkinter` (PYTHON SOFTWARE FOUNDATION, 2026).

O código-fonte completo da aplicação, bem como as instruções para execução e reprodução dos experimentos, está disponível no repositório público <https://github.com/markesjp/projeto_tcc> (MARQUES, 2024).

4.2 Interface gráfica e gerador de receitas

Para democratizar o uso prático da ferramenta por analistas e biólogos, desenvolveu-se uma interface gráfica interativa que permite a configuração e visualização do *pipeline* de processamento. Por meio da interface, o usuário pode carregar imagens e aplicar sequencialmente diferentes filtros a partir de um menu suspenso, como dilatadores, conversores de contraste e métodos de limiarização. O sistema atua como um laboratório visual, em que cada operação pode ser configurada por meio de parâmetros ajustáveis e seus efeitos serem visualizados em tempo real, permitindo análise imediata do impacto sobre a imagem. A Figura 8 ilustra essa interface.

Esse ambiente funciona como um espaço experimental controlado, no qual o usuário pode testar diferentes combinações de operações. Quando se alcança uma máscara binária satisfatória para a etapa de segmentação, o conjunto de operações empilhadas pode ser exportado no formato JSON. Esse modelo de “receitas” descreve formalmente o *pipeline* aplicado e pode ser reutilizado posteriormente. Isso garante a reprodutibilidade do método, visto que o arquivo salvo pode ser aplicado posteriormente de forma consistente a diferentes conjuntos de imagens, sem necessidade de intervenção manual.

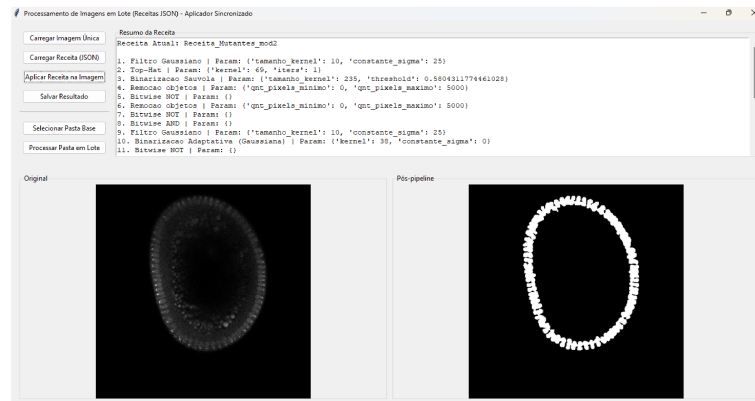


Figura 8 – Tela do aplicador responsável pela execução automatizada das receitas de segmentação.

4.3 Módulo de processamento em lote e espelhamento de diretórios

Considerando o volume de imagens geradas em experimentos de microscopia, o sistema inclui um módulo de processamento em lote, implementado no script `aplicador.py`. Esse módulo é responsável por aplicar automaticamente uma receita previamente definida a um conjunto de imagens armazenadas em diretórios. O processo é realizado por meio de uma varredura recursiva na estrutura de pastas de entrada, identificando e processando cada arquivo de imagem de forma sequencial.

Para cada imagem processada, o sistema executa o *pipeline* completo de segmentação, gera a máscara resultante, salva as imagens intermediárias relevantes e produz

visualizações auxiliares, como contornos e sobreposições entre predição e referência. Além disso, o módulo reconstrói no diretório de saída a mesma estrutura de pastas existente no diretório de entrada. Esse espelhamento facilita a organização dos resultados e permite rastrear a correspondência entre cada imagem original e os arquivos gerados pelo algoritmo.

Um desafio relevante enfrentado durante o desenvolvimento desse módulo está relacionado à diferença entre as representações das anotações manuais e das estruturas reais. As marcações oriundas do AnyLabeling consistem em polígonos definidos manualmente, formados por segmentos retilíneos, enquanto os núcleos celulares apresentam contornos naturalmente suaves, irregulares e sujeitos a variações de aquisição. Essa diferença pode influenciar a comparação entre a predição automática e a referência manual, especialmente em análises baseadas em sobreposição de pixels.

A Figura 9 ilustra uma imagem resultante da fusão entre a máscara predita e a máscara de referência relativa a uma amostra da base. Nessa visualização, regiões corretamente segmentadas e regiões associadas a erros de segmentação são destacadas por composições de cor, permitindo uma inspeção qualitativa dos acertos e das falhas do método.

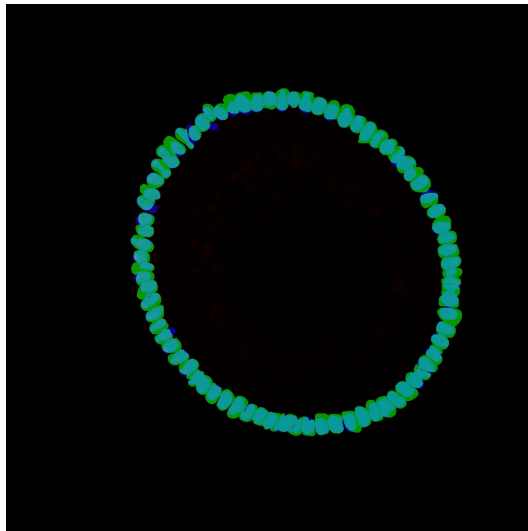


Figura 9 – Sobreposição entre a segmentação automática e a anotação de referência.

5 Resultados e Discussões

Para avaliar o desempenho do método proposto, realizou-se uma análise quantitativa baseada na comparação entre as segmentações automáticas e as anotações de referência (*ground truth*) geradas no software AnyLabeling. Os experimentos foram conduzidos em um ambiente computacional composto por processador Intel Core i7 e 16 GB de memória RAM. Sob estas condições de hardware não especializadas (sem aceleração primária via Unidade de Processamento Gráfico - GPU), o *pipeline* completo apresentou tempo médio de processamento de aproximadamente 1,2 segundos por imagem, evidenciando baixo custo computacional e viabilidade para aplicações em larga escala.

5.1 Análise do desempenho de base (*baseline*)

As coordenadas das anotações manuais foram definidas sobre as dimensões originais das imagens e posteriormente projetadas para o tamanho matricial utilizado na etapa de predição. Esse procedimento foi necessário para garantir correspondência espacial entre as máscaras de referência e as máscaras produzidas pelo algoritmo. O experimento inicial foi conduzido com parâmetros genéricos de segmentação, sem calibração individualizada para cada amostra. Os resultados quantitativos obtidos são apresentados na Tabela 1.

De modo geral, os resultados indicam comportamentos distintos entre a avaliação da massa nuclear e a avaliação em nível de instâncias. O coeficiente Dice, utilizado para mensurar a concordância espacial entre as máscaras, apresentou valores relativamente estáveis, variando entre 0,7965 e 0,8561. Esse resultado sugere que o método foi capaz de delimitar a massa nuclear global com boa aderência às anotações de referência nas amostras avaliadas.

Por outro lado, o *F1-Score* por instância apresentou maior variabilidade, oscilando entre 0,3500 e 0,7702. Essa diferença indica que a individualização dos núcleos é mais sensível às características morfológicas de cada imagem, especialmente em regiões de alta densidade celular, contato intenso entre estruturas adjacentes ou ausência de concavidades bem definidas. Portanto, embora a segmentação da massa global tenha apresentado desempenho consistente, a separação de instâncias permanece como o principal desafio do método.

Tabela 1 – Resultados quantitativos da segmentação, incluindo métricas em nível de massa (Coeficiente Dice) e em nível de instâncias (*F1-Score*), comparadas às anotações de referência (*ground truth*) para as amostras avaliadas.

Amostra	Grupo	Massa Ref. (px)	Massa Alg. (px)	Qtd. Ref.	Qtd. Alg.	Cortes Geom.	TP	FP	FN	F1 Inst.	Dice Massa
2-1.tif	Mutantes	17498	15657	73	63	64	46	17	27	0.6765	0.8340
3-1.tif	Mutantes	67181	51067	82	79	87	62	17	20	0.7702	0.8232
1-1.tif	Selvagens	15519	15203	76	44	44	21	23	55	0.3500	0.7965
11-1.tif	Selvagens	18281	16271	80	62	73	50	12	30	0.7042	0.8561
12-1.tif	Selvagens	17735	15801	79	64	70	53	11	26	0.7413	0.8482

Conforme apresentado na Tabela 1, a amostra mutante 3-1.tif apresentou o melhor desempenho em nível de instâncias, com *F1-Score* igual a 0,7702, 62 verdadeiros positivos, 17 falsos positivos e 20 falsos negativos. Esse resultado indica que, nessa amostra, o método foi capaz de individualizar os núcleos com maior estabilidade relativa quando comparado às demais imagens analisadas. Em termos de segmentação da massa global, o melhor resultado foi observado na amostra 11-1.tif, que alcançou coeficiente Dice igual a 0,8561.

A análise conjunta dos indicadores mostra que os dois níveis de avaliação utilizados neste trabalho — massa segmentada e instâncias individualizadas — capturam comportamentos complementares do algoritmo. Enquanto o coeficiente Dice reflete a aderência espacial global da máscara predita à referência, o *F1-Score* por instância revela com maior sensibilidade as dificuldades do método na separação de núcleos fortemente aglomerados.

Em particular, as amostras 3-1.tif, 11-1.tif e 12-1.tif apresentaram os melhores desempenhos em nível de instâncias, com 62, 50 e 53 verdadeiros positivos, respectivamente. Esses resultados indicam que a abordagem geométrica foi capaz de produzir separações coerentes em parte dos cenários avaliados. Em contraste, a amostra 1-1.tif apresentou o pior desempenho, com *F1-Score* igual a 0,3500 e elevado número de falsos negativos. Durante a análise qualitativa, levantou-se a hipótese de que diferenças de resolução entre as imagens também possam limitar o detalhamento do contorno e reduzir a estabilidade da detecção de concavidades. Entretanto, as dimensões exatas utilizadas na comparação mencionada durante a apresentação não foram preservadas no registro experimental disponível, o que impede quantificar esse efeito nesta versão do trabalho. Assim, essa interpretação deve ser tratada como uma hipótese explicativa, e o baixo desempenho pode estar relacionado à combinação entre resolução efetiva, aglomerações mais complexas, contatos extensos entre núcleos adjacentes, concavidades pouco definidas e limitações da parametrização inicial.

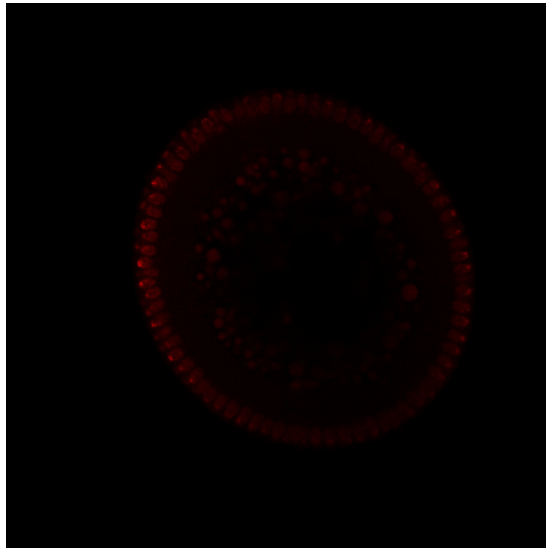
Esses resultados sugerem que a principal limitação do método não está na delimitação da massa nuclear, mas na estabilidade da individualização de instâncias em regiões fortemente aglomeradas. Em tais cenários, a ausência de concavidades bem definidas, a presença de contornos ruidosos ou a escolha inadequada de parâmetros pode comprometer a eficácia da estratégia geométrica de separação. Portanto, os resultados do modelo base devem ser interpretados como uma evidência inicial da viabilidade da abordagem, mas

também como indicação da necessidade de calibração e validação em um conjunto maior de amostras.

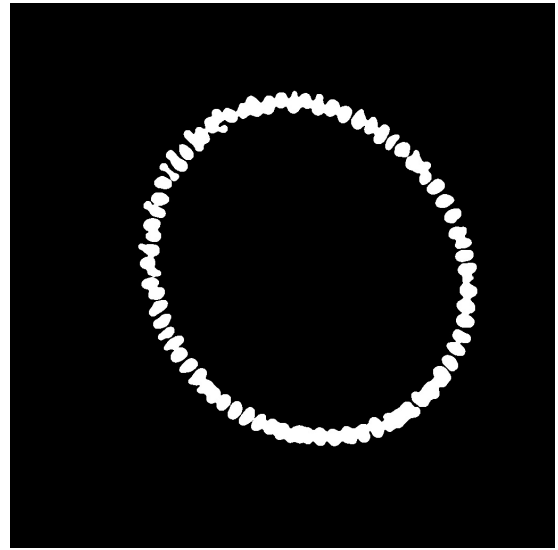
5.2 Análise visual detalhada da amostra 3-1.tif

A amostra 3-1.tif foi selecionada para análise visual detalhada por apresentar o melhor desempenho em nível de instâncias ($F1-Score=0,7702$). Além disso, essa amostra apresentou 62 verdadeiros positivos, 17 falsos positivos e 20 falsos negativos, o que a torna um caso representativo para compreender, de forma qualitativa, os fatores que contribuíram para o comportamento do método.

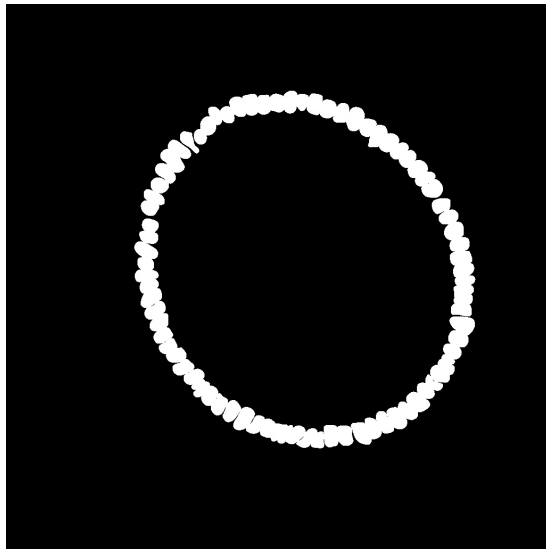
A Figura 10 apresenta as etapas iniciais do *pipeline* para uma amostra da base. Na Figura 10a observa-se a amostra original 3-1.tif do embrião de *Drosophila*, exibindo um cenário biologicamente complexo, com alta densidade celular e contato entre estruturas nucleares. Na Figura 10b tem-se o resultado da imagem pré-processada, obtido após aplicação da receita de filtragem descrita neste trabalho. Nessa etapa, nota-se a redução de ruído e a melhor definição da massa segmentada, tornando o contorno mais adequado para análise geométrica subsequente. Essa condição é fundamental, pois a qualidade do contorno influencia diretamente a etapa de detecção de concavidades. Já as Figuras 10c e 10d apresentam, respectivamente, a máscara de referência (*ground truth*) produzida manualmente e a sua representação em contorno, utilizadas como base para comparação quantitativa e inspeção qualitativa.



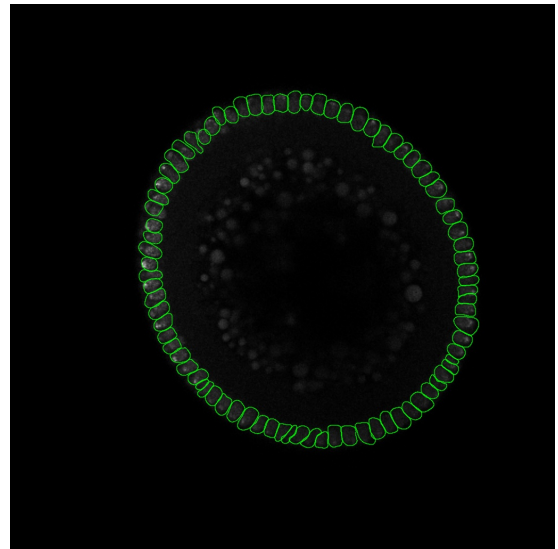
(a) Imagem original.



(b) Imagem pré-processada.



(c) *ground truth*.

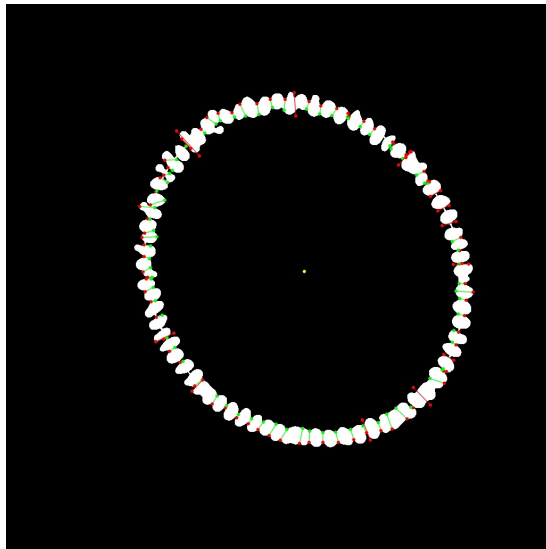


(d) *ground truth* em contorno.

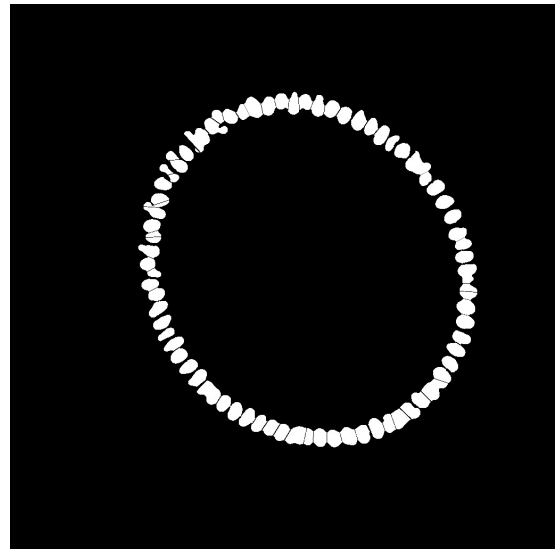
Figura 10 – Etapas iniciais e referência manual da amostra 3-1.tif.

A Figura 11 exibe as imagens intermediárias e finais relacionadas ao processo de segmentação propriamente dito. A imagem de depuração da segmentação, apresentada na Figura 11a, é particularmente importante porque permite inspecionar o comportamento interno do algoritmo, evidenciando os elementos geométricos analisados para proposição dos cortes. Esse tipo de visualização auxilia na interpretação das decisões do método, sobretudo em regiões com múltiplas concavidades ou adjacência entre núcleos. A Figura 11b demonstra a predição final do algoritmo, isto é, a máscara segmentada obtida automaticamente. Já a Figura 11c apresenta essa mesma predição destacada por contornos, o que facilita a comparação estrutural com a referência manual. Por fim, a Figura 11d apresenta a visualização de depuração em contorno, útil para compreender como o algoritmo interpreta

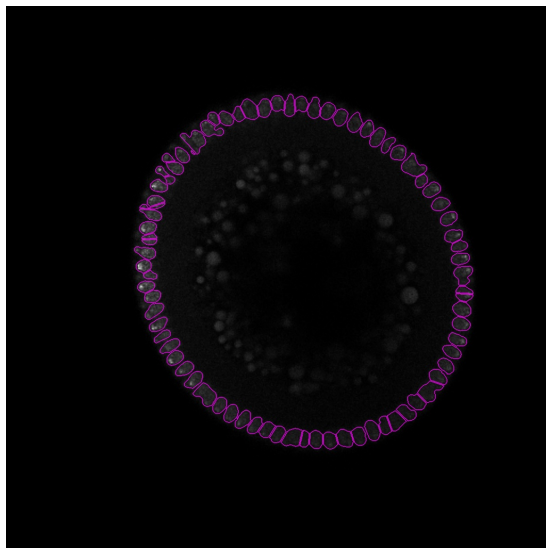
a geometria externa e interna da componente analisada.



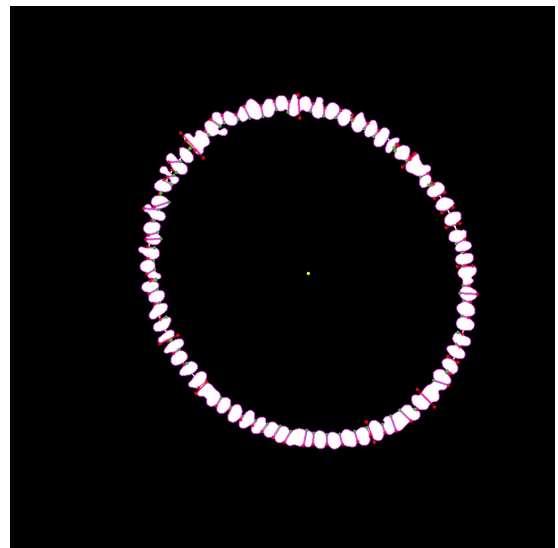
(a) Depuração da segmentação.



(b) Predição do algoritmo.



(c) Predição do algoritmo em contorno.



(d) Depuração da segmentação em contorno.

Figura 11 – Visualizações da segmentação da amostra 3-1.tif.

A Figura 12 apresenta a inspeção visual por sobreposição (*overlay*) entre os resultados do algoritmo e a referência manual. Esse tipo de visualização permite examinar, de forma qualitativa, a correspondência espacial entre a massa segmentada automaticamente e a massa anotada manualmente. No caso da amostra 3-1.tif, essa inspeção reforça o bom desempenho observado quantitativamente, indicando que a predição preserva de forma satisfatória a ocupação global da massa nuclear e apresenta capacidade relevante de individualização em boa parte das regiões analisadas.

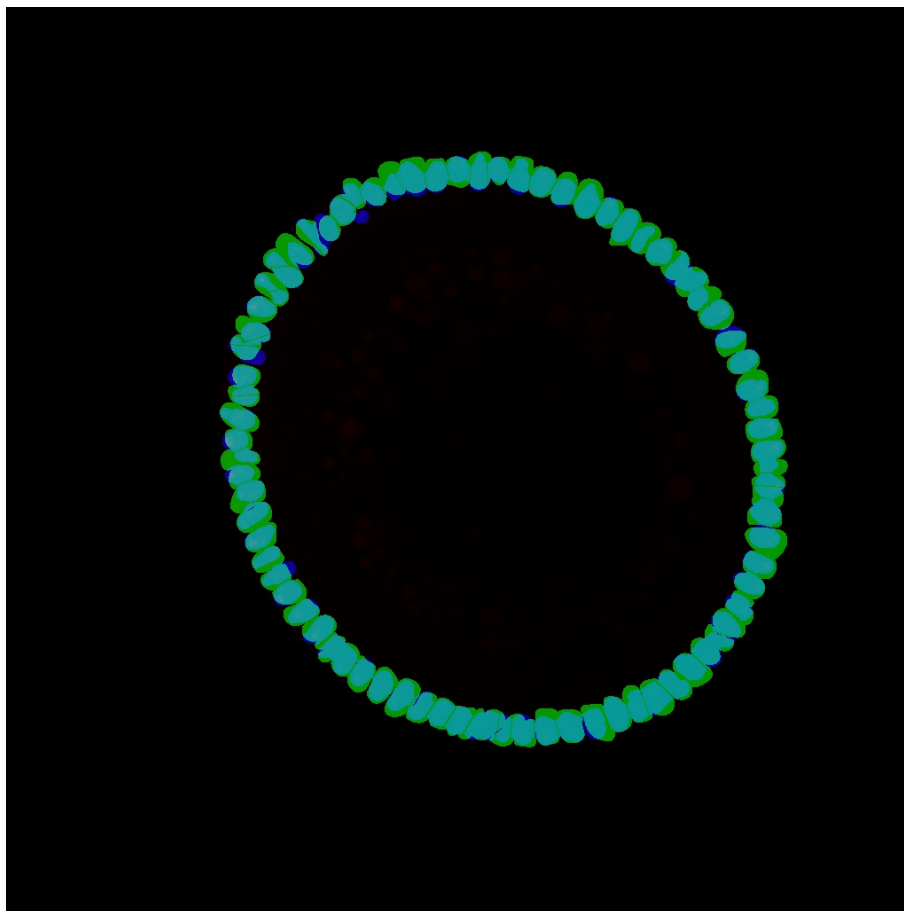


Figura 12 – Sobreposição entre a predição do algoritmo e a referência manual da amostra 3-1.tif.

A análise conjunta dessas figuras permite compreender com maior profundidade por que a amostra 3-1.tif obteve o melhor *F1-Score* por instância entre as imagens avaliadas. Em primeiro lugar, o pré-processamento produziu uma máscara estável para preservar a massa biológica principal e reduzir artefatos. Em segundo lugar, a geometria do contorno resultante mostrou-se favorável à detecção de concavidades relevantes, o que ampliou a capacidade do algoritmo de propor cortes úteis para a separação de núcleos em contato. Por fim, a comparação visual com a referência manual indica que, embora ainda existam divergências localizadas, o método foi capaz de recuperar de forma consistente a estrutura global da amostra e uma parcela significativa de suas instâncias individuais.

5.3 Precisão da máscara de pixels e validação do pré-processamento

Os resultados obtidos para o coeficiente Dice indicam que o *pipeline* apresentou boa fidelidade na segmentação da massa biológica total em relação às anotações de referência. As cinco amostras avaliadas alcançaram valores entre 79,65% e 85,61%, indicando que a etapa de pré-processamento foi eficaz para separar, com boa concordância espacial, as regiões nucleares em relação ao fundo da imagem.

Esse comportamento reforça a consistência da receita de pré-processamento adotada, composta por realce de contraste, suavização e limiarização adaptativa, seguida de refinamentos morfológicos. Em particular, os resultados sugerem que a massa segmentada pelo algoritmo permaneceu próxima da massa de referência em todas as amostras, o que valida a adequação da etapa inicial do *pipeline* como suporte à análise geométrica subsequente.

Dessa forma, ainda que o desempenho em nível de instâncias apresente variabilidade entre as amostras, a análise em nível de pixels sugere que o método é consistente para delimitar a ocupação global dos núcleos nas imagens de embriões de *Drosophila melanogaster*. Isso é particularmente relevante porque confirma que eventuais erros remanescentes concentram-se mais na separação interna entre núcleos adjacentes do que na identificação da massa biológica de interesse.

5.4 Calibração geométrica de estabilização

Verificada a boa correspondência espacial entre as máscaras binárias e as anotações de referência, a etapa seguinte concentrou-se na análise do mecanismo geométrico de proposição de cortes. A inspeção qualitativa indicou que parte dos erros remanescentes em nível de instâncias estava associada à sensibilidade do detector de concavidades a pequenas irregularidades do contorno, decorrentes de ruído residual, desfoque óptico e variações locais de contraste.

Observou-se que o uso de critérios excessivamente permissivos para detecção de concavidades e aceitação de alinhamentos vetoriais pode levar à geração de cortes desnecessários ou pouco coerentes com a geometria real dos agrupamentos celulares. Em contrapartida, o ajuste mais restritivo desses parâmetros mostrou-se promissor para reduzir segmentações indevidas, principalmente em regiões onde as bordas apresentam oscilações pouco significativas do ponto de vista morfológico.

A interface gráfica desenvolvida em Python mostrou-se particularmente relevante nesse processo, pois permitiu inspecionar visualmente os efeitos da calibração sem a necessidade de modificar diretamente a infraestrutura do código. Em termos práticos, verificou-se que o aumento do limiar mínimo de concavidade, a redução da sensibilidade a flutuações locais do sinal radial e o enrijecimento dos critérios de alinhamento geométrico contribuíram para maior estabilidade da segmentação.

Assim, os experimentos qualitativos de calibração sugerem que a arquitetura proposta possui capacidade real de refinamento paramétrico. Ainda que os resultados quantitativos consolidados deste trabalho se refiram ao experimento inicial, a análise conduzida indica que ajustes geométricos adicionais podem ampliar a robustez do método e reduzir falhas de individualização em cenários mais complexos.

6 Conclusão e Trabalhos Futuros

Este trabalho apresentou uma abordagem computacional para segmentação geométrica de núcleos em imagens microscópicas de embriões de *Drosophila melanogaster*. A proposta foi baseada na análise de contornos, na representação polar do sinal $r(\theta)$, na detecção de concavidades e na aplicação de cortes geométricos para separação de instâncias nucleares. Além disso, foi incorporado um mecanismo de recuo (*fallback*) para lidar com situações em que o pareamento entre concavidades externas e contornos internos não fornece uma solução suficientemente confiável.

Os resultados obtidos permitem uma conclusão em dois níveis. Em primeiro lugar, o método apresentou bom desempenho na delimitação da massa nuclear total. No experimento inicial, o coeficiente Dice variou entre 79,65% e 85,61%, indicando que a etapa de pré-processamento e binarização foi capaz de separar, com boa fidelidade espacial, as regiões nucleares do fundo da imagem nas amostras avaliadas. Esse resultado sugere que a combinação entre filtragem, limiarização adaptativa, operações morfológicas e refinamento geométrico produziu máscaras adequadas para a etapa posterior de análise de contornos. Do ponto de vista computacional, o *pipeline* também se mostrou viável em hardware convencional, com tempo médio de processamento de aproximadamente 1,2 segundos por imagem.

Em segundo lugar, os experimentos mostraram que a separação de instâncias nucleares apresentou resultados promissores, embora com variabilidade entre as amostras. Os valores de *F1-Score* por instância variaram entre 0,3500 e 0,7702, evidenciando que o método foi capaz de individualizar núcleos com desempenho satisfatório em parte dos casos analisados, mas ainda encontra dificuldades em cenários de maior complexidade morfológica. Assim, a principal limitação observada não está na extração da massa nuclear global, mas na estabilidade da individualização de instâncias em regiões com alta densidade celular, contatos extensos entre núcleos ou concavidades pouco definidas.

A implementação do software em Python também constitui uma contribuição relevante do trabalho. A interface gráfica desenvolvida permitiu configurar receitas de pré-processamento, visualizar etapas intermediárias, executar o processamento em lote e apoiar a análise qualitativa dos resultados. Esse aspecto reforça o caráter interpretável e rastreável da abordagem, permitindo que o usuário acompanhe as decisões do algoritmo e ajuste parâmetros conforme as características das imagens analisadas.

Apesar dos resultados promissores, algumas limitações devem ser destacadas. A avaliação quantitativa foi realizada sobre um subconjunto reduzido de cinco imagens, selecionadas em razão do alto custo de anotação manual em nível de instância. Dessa

forma, os resultados obtidos não devem ser interpretados como uma generalização estatística para todo o banco de imagens, mas como uma análise inicial da viabilidade da abordagem proposta. Além disso, o desempenho do método mostrou-se sensível à calibração dos parâmetros e à qualidade da máscara binária inicial, especialmente em imagens com maior ambiguidade geométrica.

Como trabalhos futuros, recomenda-se ampliar a base de *ground truth*, incorporando mais imagens, diferentes fenótipos mutantes e maior variabilidade de estágios embrionários. Também se mostra relevante realizar uma nova rodada experimental comparando, de forma controlada, o modelo base e versões calibradas do algoritmo, utilizando as mesmas métricas de avaliação. Outra possibilidade consiste em investigar mecanismos automáticos de ajuste de parâmetros, reduzindo a dependência da calibração manual. Por fim, seria pertinente comparar a abordagem proposta com métodos clássicos, como *watershed*, e com métodos baseados em aprendizado profundo, como U-Net, Cellpose ou StarDist, a fim de posicionar melhor o desempenho do método no contexto de técnicas modernas de segmentação celular.

Conclui-se, portanto, que o objetivo geral deste trabalho foi atingido: foi proposta, implementada e avaliada uma estratégia de segmentação geométrica interpretável, de baixo custo computacional e aplicável ao contexto de imagens microscópicas de alta densidade. A principal contribuição do estudo está na formalização de um *pipeline* rastreável, apoiado por software próprio, capaz de servir como base para investigações futuras em segmentação celular assistida por computador.

Referências

- BOOR, C. de. *A Practical Guide to Splines*. Revised. New York: Springer, 2001. ISBN 9780387953663. Citado 2 vezes nas páginas 2 e 3.
- BRADSKI, G. The opencv library. *Dr. Dobbs's Journal of Software Tools*, v. 25, n. 11, p. 120–123, 2000. Citado na página 21.
- CVHub520 Contributors. *X-AnyLabeling*. 2026. GitHub repository. Ferramenta de anotação utilizada no trabalho. Acesso em: 05 abr. 2026. Disponível em: <<https://github.com/CVHub520/X-AnyLabeling>>. Acesso em: 05 abr. 2026. Citado 2 vezes nas páginas 10 e 21.
- DICE, L. R. Measures of the amount of ecologic association between species. *Ecology*, v. 26, n. 3, p. 297–302, 1945. Citado na página 18.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 2. ed. New Jersey: Prentice Hall, 2002. ISBN 9780130946504. Citado 3 vezes nas páginas 11, 14 e 15.
- HARRIS, C. R.; MILLMAN, K. J.; van der Walt, S. J. et al. Array programming with numpy. *Nature*, v. 585, n. 7825, p. 357–362, 2020. Citado na página 21.
- JACCARD, P. The distribution of the flora in the alpine zone. *New Phytologist*, v. 11, n. 2, p. 37–50, 1912. Citado na página 18.
- MARQUES, J. P. de O. *Drosophila Segmentation Tool*. 2024. Repositório GitHub. Acesso em: 09 fev. 2026. Disponível em: <https://github.com/markesjp/projeto_tcc>. Acesso em: 09 fev. 2026. Citado 2 vezes nas páginas 10 e 32.
- MEIJERING, E. Cell segmentation: 50 years down the road. *IEEE Signal Processing Magazine*, v. 29, n. 5, p. 140–145, 2012. Citado 3 vezes nas páginas 9, 10 e 11.
- OPENCV. *OpenCV Documentation: Contours Hierarchy*. [S.l.], 2026. Acesso em: 05 abr. 2026. Disponível em: <https://docs.opencv.org/4.x/d9/d8b/tutorial_py_contours_hierarchy.html>. Acesso em: 05 abr. 2026. Citado na página 15.
- OPENCV. *OpenCV Documentation: Finding contours in your image*. [S.l.], 2026. Acesso em: 05 abr. 2026. Disponível em: <https://docs.opencv.org/4.x/df/d0d/tutorial_find_contours.html>. Acesso em: 05 abr. 2026. Citado na página 15.
- OPENCV. *OpenCV Documentation: Structural Analysis and Shape Descriptors*. [S.l.], 2026. Acesso em: 05 abr. 2026. Disponível em: <https://docs.opencv.org/4.x/d3/dc0/group__imgproc__shape.html>. Acesso em: 05 abr. 2026. Citado 3 vezes nas páginas 15, 16 e 28.
- OTSU, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 9, n. 1, p. 62–66, 1979. Citado 2 vezes nas páginas 7 e 15.

PYTHON SOFTWARE FOUNDATION. *Tkinter — Python interface to Tcl/Tk*. [S.l.], 2026. Acesso em: 09 fev. 2026. Disponível em: <<https://docs.python.org/3/library/tkinter.html>>. Acesso em: 09 fev. 2026. Citado na página 32.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015. p. 234–241. Citado na página 10.

SANDVE, G. K.; NEKRUTENKO, A.; TAYLOR, J.; HOVIG, E. Ten simple rules for reproducible computational research. *PLOS Computational Biology*, v. 9, n. 10, p. e1003285, 2013. Citado na página 10.

SAUVOLA, J.; PIETIKÄINEN, M. Adaptive document image binarization. *Pattern Recognition*, v. 33, n. 2, p. 225–236, 2000. Citado 3 vezes nas páginas 11, 15 e 27.

SCHMIDT, U.; WEIGERT, M.; BROADDUS, C.; MYERS, G. Cell detection with star-convex polygons. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. [S.l.]: Springer, 2018. (Lecture Notes in Computer Science, v. 11071), p. 265–273. Citado 3 vezes nas páginas 9, 10 e 11.

SHAPIRO, L. G.; STOCKMAN, G. C. *Computer Vision*. New Jersey: Prentice Hall, 2001. ISBN 9780130307965. Citado na página 14.

SOUSA, D. J. de; CARDOSO, M. A.; BISCH, P. M.; LOPES, F. J. P.; TRAVENÇOLO, B. A. N. Automated standardization of images of drosophila embryos. *Journal of Visual Communication and Image Representation*, v. 71, p. 102758, 2020. Citado 6 vezes nas páginas 9, 10, 11, 14, 27 e 30.

SOUSA, D. J. de; LOPES, F. J. P.; TRAVENÇOLO, B. A. N. Methods for nuclear segmentation and evaluation of gene expression in sagittal images of drosophila melanogaster embryos. In: *Workshop of Theses and Dissertations (WTD) – 27th Conference on Graphics, Patterns and Images (SIBGRAPI)*. Rio de Janeiro, Brazil: [s.n.], 2014. Citado 4 vezes nas páginas 9, 10, 11 e 30.

STRINGER, C.; WANG, T.; MICHAELLOS, M.; PACHITARIU, M. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, v. 18, n. 1, p. 100–106, 2021. Citado na página 10.

TAHA, A. A.; HANBURY, A. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, v. 15, n. 1, p. 29, 2015. Citado na página 18.

VINCENT, L.; SOILLE, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 13, n. 6, p. 583–598, 1991. Citado na página 9.

VIRTANEN, P.; GOMMERS, R.; OLIPHANT, T. E. et al. Scipy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, v. 17, n. 3, p. 261–272, 2020. Citado na página 21.

WARFIELD, S. K.; ZOU, K. H.; WELLS, W. M. Simultaneous truth and performance level estimation (staple): An algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, v. 23, n. 7, p. 903–921, 2004. Citado na página 10.

WILSON, G.; BRYAN, J.; CRANSTON, K.; KITZES, J.; NEDERBRAGT, L.; TEAL, T. K. Good enough practices in scientific computing. *PLOS Computational Biology*, v. 13, n. 6, p. e1005510, 2017. Citado na página [10](#).