

Universidade Federal do Rio Grande do Sul
Instituto de Biociências
Curso de Biotecnologia – Ênfase em Biotecnologia Molecular

Angelo Luiz Angonezi

APRENDIZADO DE MÁQUINA APLICADO À DETECÇÃO, CLASSIFICAÇÃO E
ANÁLISE DE CÉLULAS TUMORAIS EM IMAGENS DE MICROSCOPIA DE LUZ
VISÍVEL

Porto Alegre

2022

Angelo Luiz Angonezi

APRENDIZADO DE MÁQUINA APLICADO À DETECÇÃO, CLASSIFICAÇÃO E
ANÁLISE DE CÉLULAS TUMORAIS EM IMAGENS DE MICROSCOPIA DE LUZ
VISÍVEL

Trabalho de Conclusão de Curso apresentado como
requisito parcial à obtenção do título de bacharel em
Biotecnologia – Ênfase em Biotecnologia
Molecular do Instituto de Biociências da
Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Guido Lenz

Coorientador: BSc. Juliano Luiz Faccioni

Porto Alegre

2022

CIP - Catalogação na Publicação

Angonezi, Angelo Luiz
APRENDIZADO DE MÁQUINA APLICADO À DETECÇÃO,
CLASSIFICAÇÃO E ANÁLISE DE CÉLULAS TUMORAIS EM IMAGENS
DE MICROSCÓPIA DE LUZ VISÍVEL / Angelo Luiz Angonezi.
-- 2022.
47 f.
Orientador: Guido Lenz.

Coorientador: Juliano Luiz Faccioni.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Instituto
de Biociências, Curso de Biotecnologia: Biotecnologia
Molecular, Porto Alegre, BR-RS, 2022.

1. Dinâmica tumoral. 2. Aprendizado de máquina. 3.
Análise de imagens. 4. Microscopia de luz visível. 5.
Automatização. I. Lenz, Guido, orient. II. Faccioni,
Juliano Luiz, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Angelo Luiz Angonezi

APRENDIZADO DE MÁQUINA APLICADO À DETECÇÃO, CLASSIFICAÇÃO E
ANÁLISE DE CÉLULAS TUMORAIS EM IMAGENS DE MICROSCOPIA DE LUZ
VISÍVEL

Trabalho de Conclusão de Curso apresentado como
requisito parcial à obtenção do título de bacharel em
Biotecnologia – Ênfase em Biotecnologia
Molecular do Instituto de Biociências da
Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Guido Lenz

Coorientador: Bsc. Juliano Luiz Faccioni

Avaliado em: Porto Alegre, 27 de abril de 2022

BANCA EXAMINADORA:

Prof. Dr. Guido Lenz

Universidade Federal do Rio Grande do Sul – Departamento de Biofísica

Prof. Dr. Manuel Menezes de Oliveira Neto

Universidade Federal do Rio Grande do Sul – Instituto de Informática

Prof. Dr. Marcio Dorn

Universidade Federal do Rio Grande do Sul – Instituto de Informática

Porto Alegre

2022

AGRADECIMENTOS

À Universidade Federal do Rio Grande do Sul, pelo apoio institucional e pela oportunidade de desenvolvimento científico e acadêmico promovido pelo excelente corpo docente com quem tive contato durante os anos de graduação;

Ao professor e orientador Dr. Guido Lenz, pela oportunidade de poder trabalhar com prazer, pelo desenvolvimento de visão crítica e pelo constante incentivo em ‘fazer ciência de fronteira’;

Ao coorientador BSc. Juliano Luiz Faccioni, pelas incontáveis horas de discussão que contribuíram imensamente para minha formação, sempre disponível e paciente para me apoiar ao longo destes anos;

Aos colegas do LabSinal, pelo companheirismo e troca científica diária. Em especial aos colaboradores citados no trabalho, sem os quais sua realização não teria sido possível;

Ao professor Claudio Jung e Lucas Kirsten, pelo apoio de longa data no projeto e amparo técnico computacional necessário para a realização deste trabalho;

Aos amigos de longa data de Carlos Barbosa, e a todos amigos que surgiram durante a graduação em Biotecnologia, pelo estímulo, reconhecimento e acolhimento em todos momentos.

À minha família, pelo apoio em minhas decisões, e pelo eterno carinho e interesse em me ouvir compartilhar entusiasmadamente cada passo dessa jornada.

RESUMO

A automatização da análise de células únicas em imagens de microscopia de luz visível ainda é um desafio. A variação entre configurações experimentais, microscópios e linhagens celulares impossibilitam a aplicação de uma solução única para a correta detecção e identificação de características celulares, como área, orientação e etapa do ciclo celular. Além disso, o baixo contraste entre célula e fundo da imagem impede a aplicação de *pipelines* frequentemente usados em imagens de fluorescência. O presente trabalho buscou desenvolver e validar uma ferramenta capaz de detectar e classificar células tumorais em imagens de microscopia de luz visível de maneira automatizada, empregando aprendizado de máquina. Ademais, foram propostas novas métricas para avaliação de desempenho de diferentes modelos de aprendizado de máquina, tendo em vista que a aplicação destas ferramentas em contexto biológico não necessariamente corresponde ao modo em que são aplicadas para outras abordagens. Oito modelos foram treinados e avaliados quanto à sua capacidade de identificar corretamente o número, posição, orientação, área e classe de células nas imagens. A partir dos resultados obtidos, foi possível determinar o nível de acerto dos modelos para cada aplicação, validando a ferramenta. Adicionalmente, foi desenvolvida uma versão do programa com interface, além da versão por linha de comando, facilitando sua utilização por não programadores. Espera-se que a ferramenta desenvolvida neste trabalho acelere as análises de células únicas em imagens de microscopia de luz visível, a partir da quantificação simultânea de múltiplas características celulares.

Palavras-chave: Dinâmica tumoral; aprendizado de máquina; análise de imagens.

ABSTRACT

Automating the analysis of single cells in visible light microscopy images is still a challenging task. The disparity between experimental setups, microscopes and cell lines hinders the creation of a universal solution for cell detection and identification of cell characteristics, such as area, orientation and cell cycle stage. Furthermore, the low contrast between cell and image background prevents the application of pipelines frequently used in the analyses of fluorescence imaging. This work sought to develop and validate a machine learning tool capable of detecting and classifying tumor cells in visible light microscopy images. Novel metrics were proposed to evaluate the performance of different machine learning models, taking into account that the application of these tools in a biological and non-biological context may vary considerably. Eight models were trained and evaluated for their ability to correctly identify cell number, position, orientation, area and class. The performance of the models for each application was assessed, validating the proposed tool. Furthermore, an interface version of the program was developed, in addition to the command line version, facilitating its use by non-programmers. We expect that the tool developed in this work will contribute to single-cell analysis in visible light microscopy, by simultaneously quantifying multiple cell features.

Keywords: Tumor dynamics; machine learning; image analysis.

LISTA DE FIGURAS

Figura 1: Possíveis classes de células de acordo com o estágio do ciclo celular.....	14
Figura 2: Comparação entre HBBs e OBBs como formas de descrever células em imagens de microscopia.....	15
Figura 3: Exemplo da dinâmica de determinação de melhor DT.....	17
Figura 4: Exemplo de associações entre OBBs através do algoritmo Húngaro.....	23
Figura 5: Algoritmo Húngaro e filtro de distâncias.....	26
Figura 6: FPNR médio para cada modelo, conforme diferentes DTs.....	28
Figura 7: Curvas de <i>precision-recall</i> para cada modelo.....	30
Figura 8: Distribuição de erros de contagem para cada modelo no melhor DT.....	31
Figura 9: Comparação entre imagens de diferentes níveis de ruído.....	31
Figura 10: Distribuição de erros de contagem para cada modelo no melhor DT, filtrado a partir de <i>Tukey Fence</i> para identificação de outliers.....	32
Figura 11: Distribuição de RMSE para cada modelo no melhor DT.....	33
Figura 12: Distribuição de AREs para cada modelo no melhor DT.....	34
Figura 13: Distribuição de PIs para cada modelo no melhor DT.....	35
Figura 14: Distribuição de CERs para cada modelo no melhor DT.....	36
Figura 15: Interface do programa em execução.....	39
Figura 16: Exemplo de execução do programa na versão por linha de comando.....	39
Figura 17: Exemplo de detecção com IoU baixo, mas distância entre centroides nula.....	40

LISTA DE TABELAS

Tabela 1: Resumo do conjunto de imagens usadas para geração de dados anotados.....	20
Tabela 2: Descrição dos modelos utilizados no treinamento.....	22
Tabela 3: Valor de DT ideal para cada modelo.....	29
Tabela 4: Valores médios de erro para cada modelo de acordo com as cinco diferentes etapas de validação.....	37

LISTA DE ABREVIATURAS E SIGLAS

AP: Precisão Média (*Average Precision*)

ARE: Erro Relativo de Área (*Area Relative Error*)

AUC: Área sob a curva (*Area Under the Curve*)

CNN: Rede Neural Convolutacional (*Convolutional Neural Network*)

DT: Limiar de Detecção (*Detection Threshold*)

FP: Falso Positivo

FPNR: Taxa de Falsos Positivos e Negativos (*False Positives and Negatives Rate*)

FN: Falso Negativo

HBB: Retângulo de Delimitação Horizontal (*Horizontal Bounding Box*)

IoU: Intersecção sobre União (*Intersection over Union*)

mAP: Média de AP (*mean Average Precision*)

MAPE: Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error*)

ML: Aprendizado de Máquina (*Machine Learning*)

OBB: Retângulo de Delimitação Orientado (*Oriented Bounding Box*)

PI: Índice de Perpendicularidade (*Perpendicularity Index*)

TP: Verdadeiro Positivo (*True Positive*)

SUMÁRIO

1. INTRODUÇÃO.....	13
1.1 Desafios da análise de imagens de microscopia de luz visível.....	13
1.2 Identificação de características celulares e suas aplicações.....	14
1.3 Representação de objetos com orientação em Aprendizado de Máquina.....	15
1.4 Limiar de detecção em modelos de Aprendizado de Máquina.....	16
1.5 Desenvolvimento de rede de Aprendizado de Máquina para detecção e análise de células...17	
2. OBJETIVOS E JUSTIFICATIVA.....	18
3. METODOLOGIA.....	19
3.1. Obtenção do conjunto de treino/teste.....	19
3.2. Treinamento e teste de modelos de aprendizado de máquina.....	19
3.3. Validação da ferramenta.....	21
3.3.1. Estabelecimento de associação entre OBBs.....	23
3.3.2. Determinação do melhor limiar de detecção.....	24
3.3.3. Contagem de células.....	24
3.3.4. Identificação da posição celular.....	25
3.3.5. Medida da área de OBBs.....	25
3.3.6. Medida de direcionalidade celular.....	26
3.3.7. Classificação das células.....	27
3.4. Desenvolvimento da interface.....	27
4. RESULTADOS.....	28
4.1. Determinação do melhor limiar de detecção para cada modelo.....	28
4.2. Curvas de <i>precision-recall</i>	29
4.3. Análise de erros de contagem.....	30
4.4. Análise de erros de centroide (RMSE).....	33
4.5. Análise de erros de área (ARE).....	34
4.6. Análise de erros de ângulo (PI).....	35
4.7. Análise de erros de classe (CER).....	36
4.8. Média de erro por modelo para cada aplicação.....	37
4.9. Interface da ferramenta.....	38

5. DISCUSSÃO.....	40
6. CONCLUSÃO.....	44
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	45

1. INTRODUÇÃO

1.1 Desafios da análise de imagens de microscopia de luz visível.

A heterogeneidade tumoral é um dos grandes fatores que limita o sucesso de terapias contra o câncer. Em um tumor, subpopulações de células podem resistir à terapia e causar recidiva da doença, mesmo quando compõem uma pequena parcela da massa tumoral inicial (HIRSCH et al., 1983), (DAGOGO-JACK; SHAW, 2018). Esta heterogeneidade pode ser estudada em culturas celulares *in vitro* ao se acompanhar a dinâmica de formação de colônias a partir de células únicas. Esta metodologia apresenta vantagens ao estudo populacional, como a identificação de subpopulações celulares que, embora pequenas, possuem relevância biológica (BALABAN et al., 2004). Nestes experimentos, um grande volume de imagens é gerado, dificultando - ou até mesmo impedindo - a viabilidade de análise manual destes dados.

A identificação automatizada de células em imagens de microscopia de fluorescência conta com diversos programas - como ImageJ (ABRÀMOFF, 2004), CellProfiler (KAMENTSKY et al., 2011) capazes de auxiliar neste processo. A possibilidade de automatização nestes casos se deve ao elevado contraste entre as células e o plano de fundo da imagem, em virtude da presença de proteínas fluorescentes em algum compartimento celular. Esta abordagem, no entanto, requer microscópios onerosos, além da necessidade de desenvolvimento de uma linhagem celular que expresse a proteína fluorescente, o que envolve técnicas trabalhosas e pode levar a estresse celular indesejado (GANINI et al., 2017). Na prática de microscopia de *time-lapse* - em que múltiplas imagens são capturadas ao longo de um período de tempo - problemas como fototoxicidade e *photobleaching* devem ser considerados (ICHA et al., 2017; WANG et al., 2014) além da limitação de visualização da célula ao único compartimento no qual a proteína fluorescente se encontra.

Por outro lado, microscópios de luz visível são equipamentos mais acessíveis e de baixa fototoxicidade, evitando os problemas supracitados envolvendo a produção de proteínas fluorescentes. Entretanto, devido ao baixo contraste entre as células e o fundo da imagem, o processo de detecção e classificação de células em imagens de microscopia de luz visível, de forma automatizada, ainda se prova um desafio.

Embora existam múltiplas abordagens para a automatização deste processo - incluindo aprendizado de máquina (KOMURA; ISHIKAWA, 2018; ULMAN et al., 2017) - a diversidade de microscópios, linhagens celulares e magnificação impossibilita a aplicabilidade de uma mesma abordagem a todos os modelos experimentais de maneira uniforme.

1.2 Identificação de características celulares e suas aplicações

A análise de células únicas é especialmente relevante para estudos da dinâmica tumoral, ao passo que fornece informações sobre cada indivíduo da população, ao invés de uma simples média populacional. Neste contexto, além da contagem de células - relevante para ensaios clonogênicos e de proliferação celular (LENZ et al., 2021) - características como a posição, área e orientação da célula são relevantes para análises diversas de células únicas (HILSENBECK et al., 2016). Ademais, metodologias que permitam a classificação de células de acordo com o estágio do ciclo celular (Figura 1) de maneira automatizada são de interesse para correlacionar a fase do ciclo a outras medidas de interesse (WANG et al., 2014).

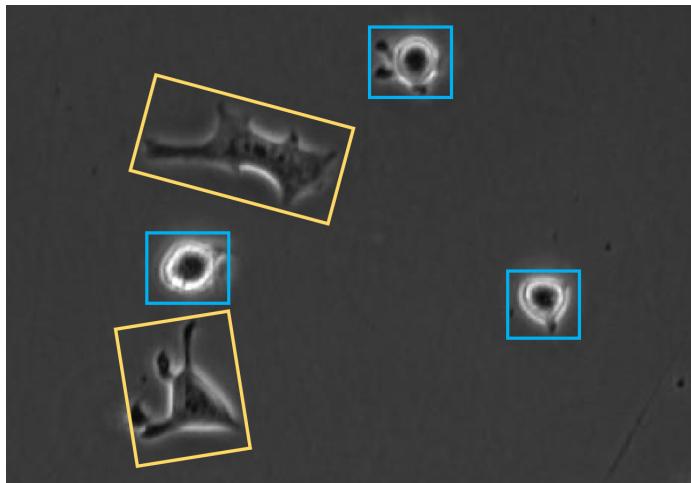


Figura 1: Comparação entre possíveis classes de células de acordo com o estágio do ciclo celular (do autor). Contornos em azul descrevem células arredondadas, supostamente próximas à fase M (mitótica), ao passo que as células marcadas em laranja refletem células mais alongadas, provavelmente em interfase.

A automatização do processo de obtenção de tais características é ainda de maior interesse quando aplicada a dados de microscopia de *time-lapse*, permitindo a análise da dinâmica de tais fenótipos, relevante no contexto de compreensão da heterogeneidade tumoral (LENZ et al., 2021). Além disso, a identificação simultânea de múltiplas características celulares facilita o processo de rastreamento das células em imagens temporalmente espaçadas (JAQAMAN et al., 2008), o que quando aplicado em conjunto à construção de árvores de parentesco entre células, se torna uma poderosa ferramenta de análise da dinâmica tumoral.

1.3 Representação de objetos com orientação em Aprendizado de Máquina.

Embora o uso de *Horizontal Bounding Boxes* (HBBs) seja o padrão em métodos de identificação de objetos, novas alternativas já foram propostas para distintas aplicações, como quadriláteros livres na detecção de placas de automóveis (SILVA; JUNG, 2018), e retângulos orientados na identificação de texto (GUPTA; VEDALDI; ZISSERMAN, 2016) e de objetos em imagens de satélite (MA et al., 2019). HBBs apresentam limitações quando aplicadas à detecção de objetos com formato alongado e com orientação, como navios em imagens aéreas (LI et al., 2018) ou detecção de células em imagens de microscopia, foco do presente trabalho. Isso ocorre de forma especial em imagens de microscopia celular com maior densidade, onde a utilização de HBBs para descrever células próximas culmina em sobreposição dos contornos (Figura 2).

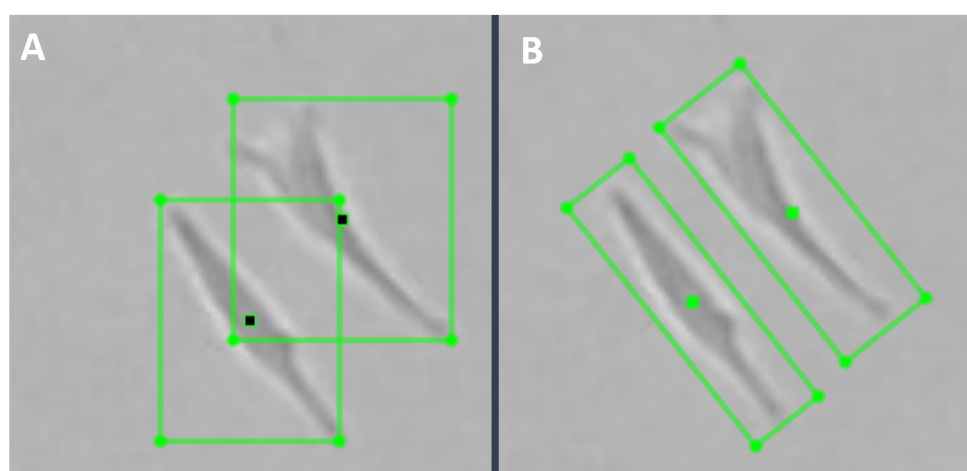


Figura 2: Comparação entre HBBs (A) e OBBs (B) como formas de descrever células em imagens de microscopia.

Abordagens como segmentação fornecem uma descrição completa do formato e características do objeto, sendo comumente empregada para detecção e rastreamento de células (RONNEBERGER; FISCHER; BROX, 2015). No entanto, para algumas modalidades de imagem de microscopia, a proximidade entre células dificulta a aplicação destas abordagens devido à sobreposição de células (ULMAN et al., 2017). Este fenômeno resulta em máscaras de segmentação conectadas e, conseqüentemente, em erros de segmentação celular. Imagens de microscopia de campo claro são particularmente suscetíveis a esse fator devido ao baixo contraste entre célula e fundo. Ademais, para o treinamento de modelos de ML com este tipo de metodologia, o anotador deve manualmente desenhar todo contorno da célula (ao invés de simplesmente sobrepor um retângulo de anotação), em um processo consideravelmente mais laborioso que Oriented Bounding Boxes (OBBs) (VICAR et al., 2019).

Dessa forma, a utilização de OBBs é uma alternativa eficaz para a identificação de células em imagens de microscopia de luz visível, já que fornece consideravelmente mais informações que HBBs tradicionais, sem requerer anotação tão complexa quanto máscaras de segmentação. Ademais, as informações extraídas a partir de OBBs, como área, orientação e formato, podem ser exploradas para facilitar a associação entre células de maneira temporal em imagens de microscopia *time-lapse*.

1.4 Limiar de detecção em modelos de Aprendizado de Máquina

Em modelos de aprendizado de máquina voltados à identificação de objetos (ZHAO et al., 2019), um valor numérico é atribuído para cada detecção, correspondendo a um intervalo de probabilidades de naquela posição haver um objeto. Este número (geralmente representado em uma escala de 0.0 a 1.0) pode ser interpretado como um nível de certeza do modelo para cada detecção. O limiar de detecção é um ponto limítrofe neste intervalo que permite a conversão do valor contínuo em binário (se aquela detecção é ou não uma célula). A definição do melhor limiar de detecção (DT) para cada modelo é extremamente importante para que detecções com nível de certeza muito baixo não entrem nas análises (evitando considerar falsos positivos, como fundo da imagem ou *debris*, Figura 3A). Do mesmo modo, não se pode elevar indiscriminadamente o DT, para que não haja perda de células verdadeiras da imagem (evitando

falsos negativos, Figura 1C). Um exemplo da dinâmica de filtro aplicado pelo DT é mostrado na Figura 3.

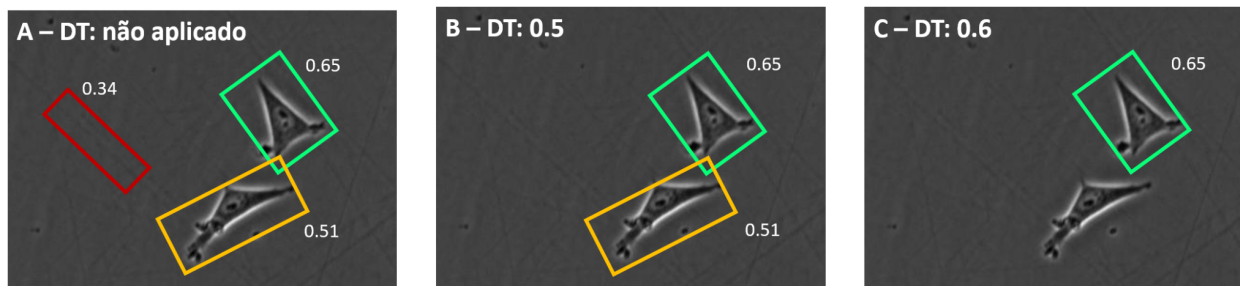


Figura 3: Exemplo da dinâmica de determinação de melhor DT. Em A, a não aplicação de um filtro resulta em consideração de um falso positivo nas análises (caixa vermelha corresponde a uma detecção incorreta, contornando o fundo da imagem). Em B, com limiar de detecção em 0.5, ambas as células na seção da imagem são incluídas na detecção, ao passo que o falso positivo em vermelho foi removido. Em C, observa-se que a elevação do DT para 0.6 culmina na remoção de uma célula verdadeira das detecções.

1.5 Desenvolvimento de rede de Aprendizado de Máquina para detecção e análise de células

Tendo em vista a necessidade de ferramentas para acelerar o processo de análise de bioimagens, o presente projeto visou desenvolver e validar uma rede de Aprendizado de Máquina para a detecção e classificação de células tumorais em imagens de microscopia de luz visível.

O trabalho, em desenvolvimento no Laboratório de Sinalização e Plasticidade Celular (LabSinal) da Universidade Federal do Rio Grande do Sul, contou com a colaboração de outros membros do laboratório nas etapas de aquisição e anotação de imagens. Ademais, uma parceria com o Instituto de Informática da Universidade foi estabelecida, para o treinamento dos modelos de aprendizado de máquina. Uma edição do modelo treinado, assim como uma versão inicial da interface já se encontra disponível para uso interno no laboratório, como fruto do trabalho desenvolvido nos últimos meses.

2. OBJETIVOS E JUSTIFICATIVA

A detecção e rastreamento de células em imagens de microscopia é uma tarefa com vastas aplicações na área médica e de pesquisa, incluindo o estudo do desenvolvimento de tumores (LENZ et al., 2021). A grande quantidade de dados produzidos por imagens de microscopia de alto *throughput*, no entanto, impõe um desafio técnico para os pesquisadores. Dessa forma, existe uma necessidade de ferramentas computacionais que auxiliem na automatização da análise dos dados, incluindo aprendizado de máquina. Assim, o presente projeto tem por objetivo desenvolver e validar uma ferramenta computacional empregando aprendizado de máquina capaz de detectar e classificar células tumorais em imagens de microscopia de luz visível de maneira automatizada.

Como objetivos específicos, o projeto visa:

1. Obter conjunto de dados anotados de imagens de microscopia de luz visível de diferentes linhagens de glioblastoma para obtenção de conjunto de treino e teste dos modelos;
2. Treinar e testar diferentes modelos de aprendizado de máquina quanto à sua capacidade de identificação e classificação de células tumorais;
3. Validar a ferramenta quanto à sua capacidade de: i. Contagem de células; ii. Identificação da posição celular; iii. Medida da área de OBBs; iv. Identificação do ângulo da célula; v. Classificação das células (mitótica / não mitótica).
4. Desenvolver interface a fim de tornar a ferramenta de fácil uso tanto por programadores como não-programadores.

3. METODOLOGIA

3.1 Obtenção do conjunto de treino/teste

A fim de gerar dados anotados para o treinamento e validação dos modelos, imagens obtidas em diferentes microscópios e com linhagens celulares distintas foram selecionadas, de modo a gerar um conjunto de treino variado. Uma descrição do conjunto de imagens anotadas pode ser visualizada na Tabela 1. O processo de anotação contou com o apoio da graduanda em Biotecnologia Fernanda Dittrich Oliveira. As imagens utilizadas para anotação foram fornecidas por pesquisadores do LabSinal, citados junto à descrição de dados na Tabela 1.

A ferramenta roLabelimg (CGVICT, 2017) foi empregada para a anotação de objetos. Através dessa ferramenta, os anotadores são capazes de desenhar OBBs cujos valores de posição Cartesiana, altura e largura (em pixels), ângulo e classe podem ser exportados em arquivo de texto utilizado como entrada para o treinamento dos modelos, junto às imagens. Os arquivos gerados com a ferramenta de anotação foram processados por código próprio para permitir a leitura pelo programa de aprendizado de máquina. As duas possíveis classes celulares utilizadas foram “RoundCell” e “NormalCell”, cuja interpretação biológica pode variar de acordo com a linhagem celular.

Um total de 150 imagens foram anotadas, gerando 6.461 OBBs, das quais 5.810 (89,9%) foram classificadas como “não mitóticas” e 651 (10,1%) como “mitóticas”. As imagens foram subdivididas em 2 principais conjuntos: treino (120 imagens) e teste (30 imagens), empregados nas etapas descritas em 3.2 e 3.3, respectivamente. A seleção de imagens do conjunto de teste foi realizada de maneira manual, com o intuito de reduzir o viés entre linhagens celulares - 10 imagens das linhagens A172 e U251 (ambas glioblastoma humano), 10 imagens da linhagem MCF7 (câncer de mama humano) e 10 MRC5 (fibroblasto pulmonar humano).

3.2. Treinamento e teste de modelos de aprendizado de máquina

O treino e teste dos modelos de aprendizado de máquina foram realizados em parceria com o Instituto de Informática da Universidade Federal do Rio Grande do Sul, com apoio do professor Dr. Claudio Rosito Jung e o aluno de mestrado Lucas Kirsten. Oito diferentes modelos

(CSL (YANG; YAN, 2022), DCL (YANG et al., 2021), R3Det (YANG et al., 2020), RetinaNet (LIN et al., 2018), RSDet (QIAN et al., 2019), Refine RetinaNet (YANG et al., 2020), R3Det-DCL (YANG et al., 2020, 2021), R2CNN (JIANG et al., 2017)) foram testados quanto ao seu desempenho em detectar e classificar células tumorais a partir de métricas como *precision* e *recall*, permitindo indicar qual modelo apresenta maior taxa de acerto. Ademais, diferentes métricas além destas tradicionalmente empregadas são descritas na seção de validação da ferramenta.

Tabela 1: Resumo do conjunto de imagens usadas para geração de dados anotados.

Linhagens A172 e U251: glioblastoma humano; MCF7: câncer de mama humano; MRC5: fibroblasto pulmonar humano. Condição de cultivo CTR: células cultivadas sem adição de quimioterápico; TMZ: células tratadas com temozolomida em alguma etapa do cultivo.

Conjunto	Microscópio	Aumento	Pesquisador(a) responsável	Linhagem	Condição de cultivo	Número de imagens
Treino	CytoSMART	20x	Julieti Buss Karine Begnini Luana Lenz	A172	CTR	52
					TMZ	25
				U251	CTR	4
	Zeiss Axiovert 200	10x	Mauricio Tamborindeguy	MCF7	CTR	19
				MRC5	CTR	20
Teste	CytoSMART	20x	Julieti Buss Karine Begnini Luana Lenz	A172	CTR	4
				A172	TMZ	5
				U251	CTR	1
	Zeiss Axiovert 200	10x	Mauricio Tamborindeguy	MCF7	CTR	10
				MRC5	CTR	10

Etapas de *Data Augmentation* envolvendo rotação, espelhamento, mudanças de cor e resolução foram aplicadas às imagens de treino, a fim de aumentar o conjunto de dados. Além disso, é importante salientar que a utilização de OBBs permite a rotação com ângulos arbitrários na geração de dados aumentados, ao passo que HBBs ficam limitadas a ângulos múltiplos de 90°.

Todos os modelos são redes neurais convolucionais (CNNs) e usam como base o *backbone* ResNet50 (HE et al., 2016), com pesos pré-treinados no conjunto ImageNet (DENG et al., 2009). O treinamento ocorreu em 20 épocas, com 5.000 iterações, com os hiperparâmetros padrão para cada modelo (YANGXUE0827, 2021). A taxa de aprendizado iniciou em 1×10^{-3} e decresceu em uma ordem de grandeza ao passar de 12 e 16 épocas. Uma breve descrição de cada modelo pode ser encontrada na Tabela 2.

3.3. Validação da ferramenta

A fim de validar os modelos de aprendizado de máquina para aplicações biológicas, suas performances foram avaliadas nas seguintes tarefas: contagem de células, identificação correta da posição Cartesiana, área, ângulo e classe celular. Além disso, as métricas de *precision*, *recall*, e precisão média (mAP), tradicionalmente empregadas em aprendizado de máquina, foram calculadas.

Para verificar se os valores de erro para cada aplicação entre diferentes modelos foram significativamente diferentes, o teste H de Kruskal-Wallis (KRUSKAL; WALLIS, 1952) foi realizado.

Tabela 2: Descrição dos modelos utilizados no treinamento.

Modelo	Descrição
CSL (YANG; YAN, 2022)	Propõe resolver o problema dos contornos descontínuos (originados pela periodicidade angular ou ordenação dos cantos) transformando a previsão angular de uma regressão para um problema de classificação. Para isso, emprega a técnica <i>Circular Smooth Label</i> (que utiliza o método <i>Gray Coded Label</i>), juntamente com um novo modelo apropriado.
DCL (YANG et al., 2021)	Da mesma forma que o CSL, busca resolver o problema de descontinuidade de contorno em detectores OBB por meio de um novo mecanismo de codificação para classificação de ângulos em vez de regressão, mas também empregando um sistema de reponderação de perda (propondo uma distância angular e ponderação sensível à razão de aspecto). Essas mudanças foram construídas sobre a arquitetura RetinaNet.
R2CNN (JIANG et al., 2017)	Um modelo proposto para detectar textos de orientação arbitrária em imagens de cenas naturais. Ele é baseado na arquitetura Faster R-CNN (REN et al., 2017), e inclui blocos de agrupamento e concatenação de recursos até o final do módulo RPN para estimar os OBBs.
R3DET (YANG et al., 2020)	Um detector de rotação de estágio único refinado de ponta a ponta que usa uma abordagem de regressão progressiva de granularidade. Implementa um módulo de refinamento de recursos para melhorar o desempenho de detecção e também propõe uma perda aproximada de SkewIoU para regredir os OBBs.
R3DET_DCL (YANG et al., 2020, 2021)	Utiliza a mesma parametrização empregada pelo modelo DCL (YANG et al., 2021), mas, ao invés de utilizar RetinaNet como arquitetura base, utiliza R3Det (YANG et al., 2020).
Retinanet (LIN et al., 2018)	Uma rede unificada composta por dois sub módulos específicos: um para classificação de objetos e outro para regressão OBB. Propõe também o uso de <i>Focal Loss</i> como uma forma de superar o desequilíbrio de classe <i>foreground-background</i> encontrado durante o treinamento de detectores densos. Originalmente, este modelo foi desenvolvido para funcionar apenas com HBBs. Para trabalhar também com OBBs, foi necessário um parâmetro extra (angular) para ser regredido, bem como o auxílio de novas âncoras.
Rsdet (QIAN et al., 2019)	Usa a arquitetura RetinaNet com uma nova perda combinada com um método de regressão de oito parâmetros (em vez dos cinco parâmetros usuais) para resolver o problema de regressão de parâmetros inconsistentes em OBBs

3.3.1 Estabelecimento de associação entre OBBs

Para estabelecer as relações entre células (correspondência entre detecção e anotação manual) o algoritmo Húngaro (KUHN, 1955) foi empregado. O algoritmo Húngaro se baseia na construção de uma matriz de custo, constituída dos pesos de cada associação, seguida de uma etapa de identificação das associações que resultam em menor custo geral.

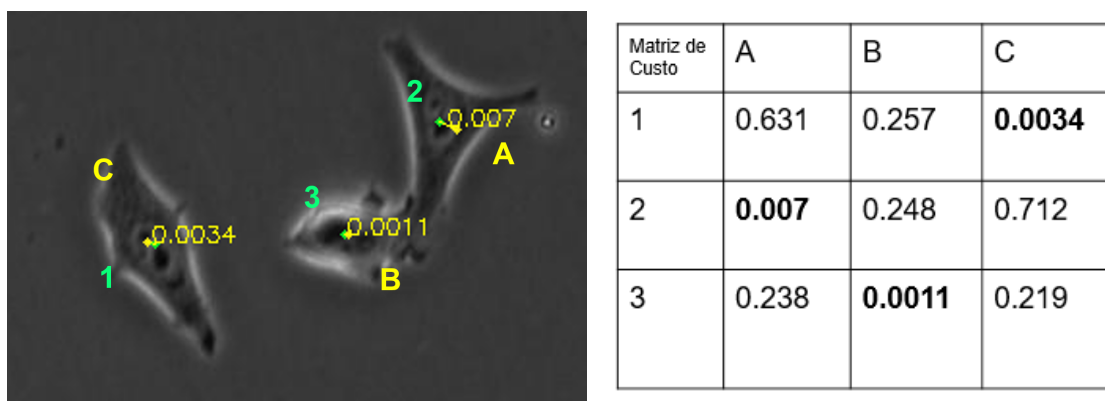


Figura 4: Exemplo de associações entre OBBs através do algoritmo Húngaro. Pontos amarelos próximo ao centro das células representam centroides de OBBs identificadas pelo modelo, e pontos verdes centroides de OBBs da anotação manual (esquerda). Matriz de custo (direita) entre as OBBs manuais (identificadas por números) e detecções do modelo (identificadas por letras), calculada a partir das distâncias entre centroides das OBBs. Em negrito, o valor de custo entre as associações finais identificadas pelo algoritmo Húngaro.

No presente trabalho, a matriz de custo foi construída com base nas distâncias entre centroides de OBBs detectadas pelo modelo e centroides de OBBs anotadas manualmente. Dessa forma, foi possível quantificar o número de Verdadeiros-Positivos (TPs), Falsos-Positivos (FPs) e Falsos-Negativos (FNs), com base no número total de células na imagem, e no número de células que tiveram associação estabelecida. As associações entre OBBs realizadas pelo algoritmo Húngaro foram necessárias para as validações descritas a seguir.

3.3.2 Determinação do melhor limiar de detecção

Devido a diferenças operacionais entre modelos, cada um poderá apresentar distintos comportamentos em um mesmo limiar de detecção. A fim de determinar o valor de DT ideal para cada modelo, as 120 imagens do conjunto de treino terão sua Taxa de Falsos Positivos e Negativos foi quantificada em 100 diferentes DTs entre 0.0 e 1.0, com base na fórmula:

$$\textit{False Positives and Negatives Ratio (FPNR)} = \frac{FP + FN}{N}$$

Onde FP corresponde ao número de falsos positivos, FN aos falsos negativos, e N equivale ao número de anotações manuais para a imagem.

Dessa forma, o DT ideal foi calculado a partir da média de FPNR de todas as imagens para cada modelo. A utilização do conjunto de treino para a determinação de melhor DT é importante a fim de não introduzir vieses de acerto nos modelos, visto que se o DT ideal fosse calculado com base nas imagens de testes, os erros seriam equivocadamente minimizados. Para cálculo dos erros citados abaixo, as detecções serão filtradas de acordo com seu respectivo valor de DT ideal.

3.3.3 Contagem de células

Para avaliar a capacidade dos modelos em identificar corretamente o número de células em uma imagem, o erro total de contagens (diferença entre o número de detecções feitas pelo modelo em relação ao número de anotações manuais da imagem) e o valor de MAPE foram calculados conforme a fórmula abaixo:

$$\textit{Mean Absolute Percentage Error (MAPE)} = \frac{1}{n} \sum_{i=1}^n \left| \frac{N_{ml(i)} - N_{manual(i)}}{N_{manual(i)}} \right| * 100$$

onde n equivale ao número de imagens, N_{ml} se refere ao número de detecções feitas pelo modelo na imagem, e N_{manual} corresponde ao número de anotações manuais na imagem.

3.3.4 Identificação da posição celular

Para avaliar a capacidade dos modelos identificar corretamente a posição Cartesiana das células encontradas em uma imagem, as distâncias entre os centroides das detecções dos modelos e os centroides das anotações manuais foram quantificadas, e o valor de RMSE para cada imagem foi calculado a partir da fórmula:

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{\sum_{i=1}^N (\text{CellsDist}^2)_{(i)}}{N}}$$

onde N corresponde ao número de detecções cujo algoritmo Húngaro obteve associação, e *CellsDist* compreende à distância Euclidiana entre a detecção do modelo em comparação à anotação manual a que foi assinalado. A distância foi normalizada a partir da resolução da imagem, visto que imagens de diferentes resoluções foram empregadas no conjunto de treino e teste dos modelos.

Para desencorajar associações entre anotações e detecções muito distantes, pares de anotação/detecção cuja distância estava acima de um limiar tiveram o seu custo de associação aumentado na matriz de custo do algoritmo Húngaro. Apesar do alto custo, este par ainda poderia ser associado pelo algoritmo Húngaro caso não houvesse outra associação que resultasse em melhor configuração global (Figura 5). Neste caso, o par foi contabilizado como um FP e FN, e não como uma associação correta. Desta forma, apenas associações próximas entre anotações e detecções entraram nos cálculos de RMSE, erro de área, ângulo e classe.

3.3.5 Medida da área de OBBs

Após as relações entre células terem sido estabelecidas, foi avaliada a capacidade dos modelos em medir a área das células presentes em uma dada imagem, com base na área da OBB que a descreve. Para isso, foi calculado o erro de área relativo entre as áreas de OBBs detectadas pelos modelos e anotações manuais, conforme a fórmula a seguir:

$$\text{Area Relative Error (ARE)} = \frac{\text{Area}_{ml} - \text{Area}_{manual}}{\text{Area}_{manual}}$$

onde $Area_{ml}$ corresponde à área da OBB detectada pelo modelo, e $Area_{manual}$ se refere à área da OBB anotada manualmente. A partir desta fórmula, valores positivos de ARE são interpretados como detecções cuja OBB teve área maior do que a anotação manual, e valores negativos indicam detecção com área menor que a anotação.

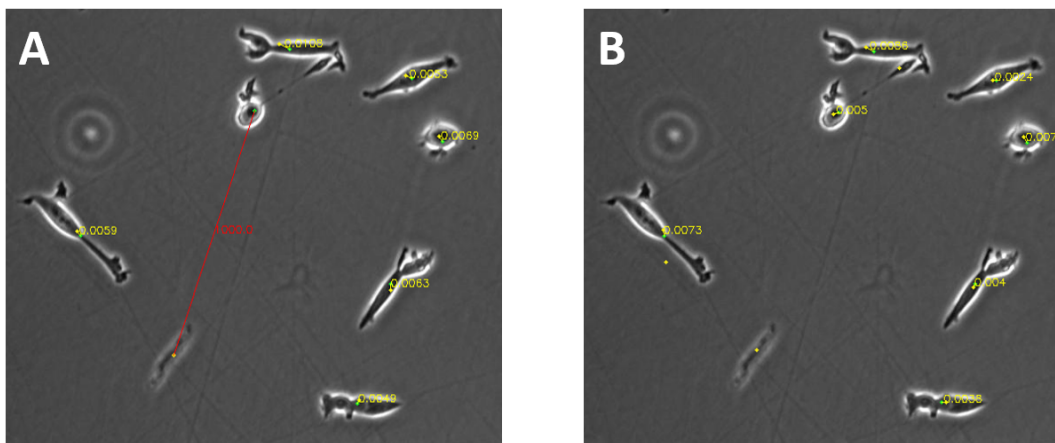


Figura 5: Algoritmo Húngaro e filtro de distâncias. Em A, uma associação que excede o limite de distâncias aceitável é marcada com a linha vermelha. Os valores em amarelo representam a distância entre células, normalizada pela resolução da imagem.

3.3.6 Medida de direcionalidade celular

Para avaliar a capacidade dos modelos em identificar corretamente a orientação das células, foi medido o índice de perpendicularidade das OBBs associadas pelo algoritmo Húngaro. O índice é calculado a partir da diferença entre ângulos, onde 0° representa paralelismo e 90° representa perpendicularidade. Para este cálculo, ângulos opostos (como 0° e 180° ou 90° e 270°) são considerados como idênticos, uma vez que determinar o sentido de uma célula depende de seus padrões de movimentação citoplasmática, os quais variam de acordo com cada linhagem (KIM et al., 2020; ZHONG; JI, 2013).

OBBs com valores de razão de eixos inferiores à 1.1 foram excluídos desta análise, visto que esta OBB representa um objeto com aspecto circular ou estrelado cuja orientação é indefinida (OBBs próximas a quadrados descrevendo um objeto com este aspecto independem de orientação).

3.3.7 Classificação das células

Para avaliar a capacidade dos modelos em identificar corretamente a classe das células (mitótica / não mitótica), a taxa de erro foi medida de acordo com a fórmula:

$$\textit{Class Error Ratio (CER)} = \frac{N_{\textit{WrongDetections}}}{N}$$

Onde $N_{\textit{WrongDetections}}$ equivale ao número de detecções com classe associada incorreta, e N compreende o total de detecções que tiveram associação estabelecida no algoritmo Húngaro.

3.4. Desenvolvimento da interface

A interface foi desenvolvida empregando as bibliotecas de construção de interfaces PyQt5 e PySide2 (QT, 2022), disponíveis na linguagem de programação Python (PYTHON SOFTWARE FOUNDATION, 2022). Esta etapa teve por objetivo fazer com que a ferramenta possa ser utilizada por pessoas com diferentes níveis de experiência em programação.

4. RESULTADOS

4.1 Determinação do melhor limiar de detecção para cada modelo

Com base nos valores de FPNR médio para cada modelo em cada DT (Figura 6), foram determinados os valores de DT de menor erro (Tabela 3).

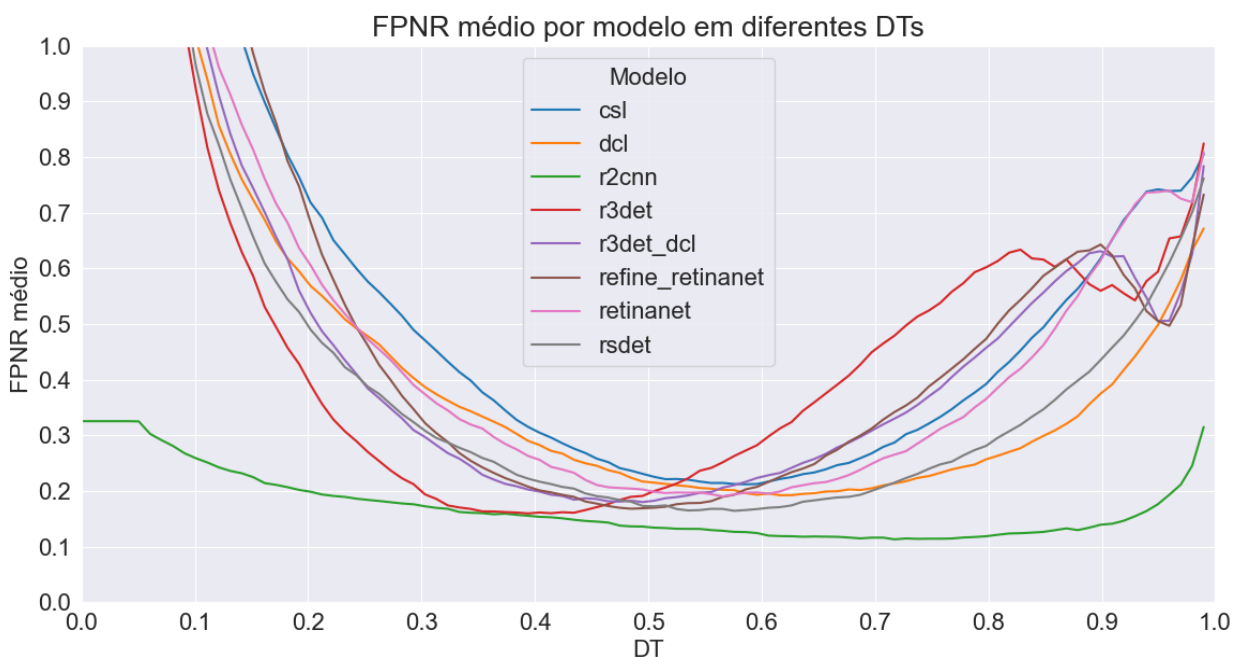


Figura 6: FPNR médio para cada modelo, conforme diferentes limiares de detecção.

Observa-se que modelo r2cnn foi capaz de atingir o menor valor médio de FPNR (Tabela 3), além de se demonstrar mais robusto entre diferentes valores de DT (Figura 6). É importante ressaltar que não é esperado que os melhores DTs, apresentados na Tabela 3, sejam ideais para todo e qualquer conjunto de dados. Fornecem, no entanto, um intervalo de DT aproximado que possa ser explorado pelo usuário, quando este estiver iniciando suas análises.

Para as análises subsequentes de erro de contagem, centroide, área, ângulo e classe, as imagens do conjunto de teste tiveram suas células detectadas pelos modelos, com cada modelo utilizando o DT de erro mais baixo encontrado com o conjunto de treino.

Tabela 3: Valor de DT ideal para cada modelo. Considera-se ideal o valor de DT que resultou em menor média de FPNR nos conjuntos de treino*.

Modelo	DT ideal*	FPNR médio do conjunto de treino no DT ideal [%]	FPNR médio do conjunto de teste no DT ideal [%]
CSL	0.58	21.0	17.4
DCL	0.62	19.1	17.1
r2cnn	0.71	11.3	11.1
r3det	0.34	15.9	14.9
r3det_dcl	0.50	18.0	16.1
refine_retinanet	0.48	16.9	16.9
retinanet	0.57	19.0	13.3
rsdet	0.57	16.3	13.7

Observa-se que alguns modelos, como r2cnn e refine_retinanet apresentam baixa variação de FPNR entre os conjuntos de treino e teste, o que sugere robustez dos modelos. Para determinar o nível de aceitação a outros conjuntos, no entanto, se faz necessária a anotação e validação com novas imagens.

4.2 Curvas de *precision-recall*

Adicionalmente ao cálculo de FNPR, proposto neste trabalho, foram avaliadas as curvas de *precision-recall* (Figura 7), usualmente empregados na análise de modelos de aprendizado de máquina.

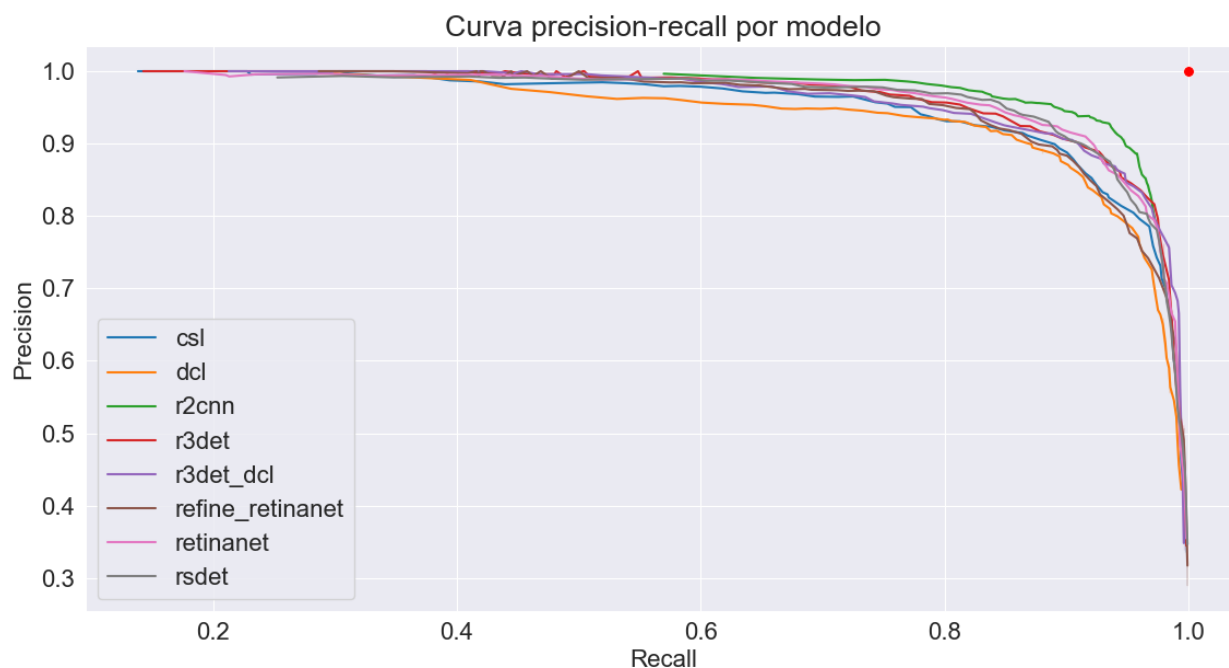


Figura 7: Curvas de *precision-recall* para cada modelo. O ponto em vermelho indica um modelo perfeito, sem FPs ou FNs em qualquer valor de DT.

De acordo com o perfil das curvas de *precision-recall* é possível inferir sobre o desempenho entre modelos - quanto mais próximo ao ponto vermelho, maior o *precision-recall*, e melhor o desempenho geral. Nota-se que r2cnn apresentou curva distinta do restante dos modelos avaliados, tendo atingido o valor mais próximo de *precision-recall* ideal, ao passo que dcl foi o que mais se distanciou.

4.3 Análise de erros de contagem

O erro de contagem se refere à diferença simples entre o número de detecções feitas pelos modelos, em comparação às anotações manuais. Dessa forma, erros de contagem positivos indicam Falsos Positivos, ao passo que erros negativos indicam falsos negativos. Nota-se tendência geral dos modelos em identificar número menor do que o esperado (Figura 8) - maior presença de falsos negativos. A dificuldade na detecção pode ser parcialmente atribuída à imagens desafiadoras - por exemplo, todos os pontos com erro de contagem abaixo de -40 se

referem a uma mesma imagem, a qual possuía maior quantidade de debris e foco distorcido (Figura 9), provocando o erro observado.

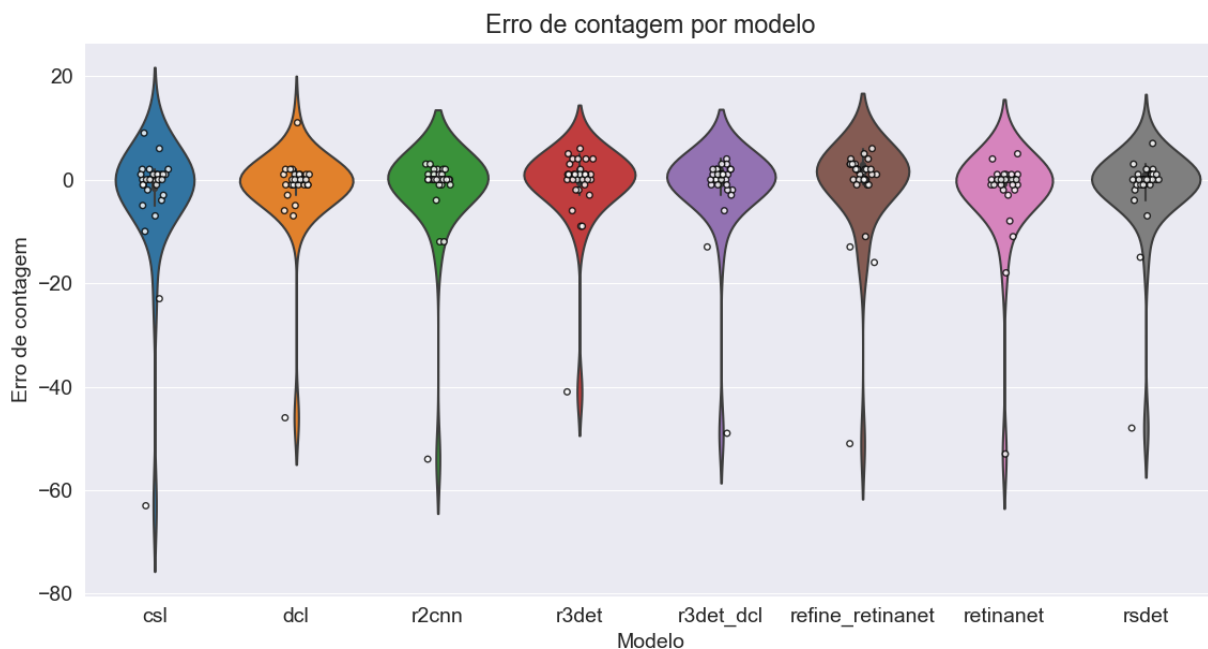


Figura 8: Distribuição de erros de contagem para cada modelo no melhor DT. Cada círculo representa o erro de contagens para uma imagem do conjunto de teste.

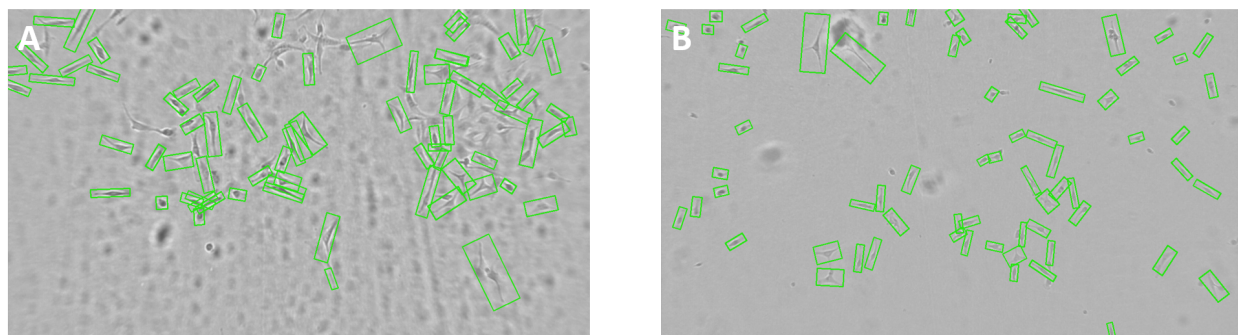


Figura 9: Comparação entre imagens de diferentes níveis de ruído. Em A, observa-se marcante presença de ruído, como células desfocadas, e *debris* presente na imagem, Em B, embora também presentes, os ruídos são pontuais, e nas posições onde há células o contraste entre fundo é maior em comparação à outra imagem. Os retângulos verdes representam OBBS detectadas pela ferramenta.

A fim de filtrar as detecções removendo os *outliers* possivelmente advindos de imagens como 8A, foi calculado o valor de corte inferior e superior com base em *Tukey Fences*, que se

baseia na distância interquartil multiplicada por um fator constante a fim de gerar limites que determinam *outliers* presentes nos dados (TUKEY, 1977). Dessa forma, imagens cujo erro de detecções excedeu os limites foram removidas (Figura 10).

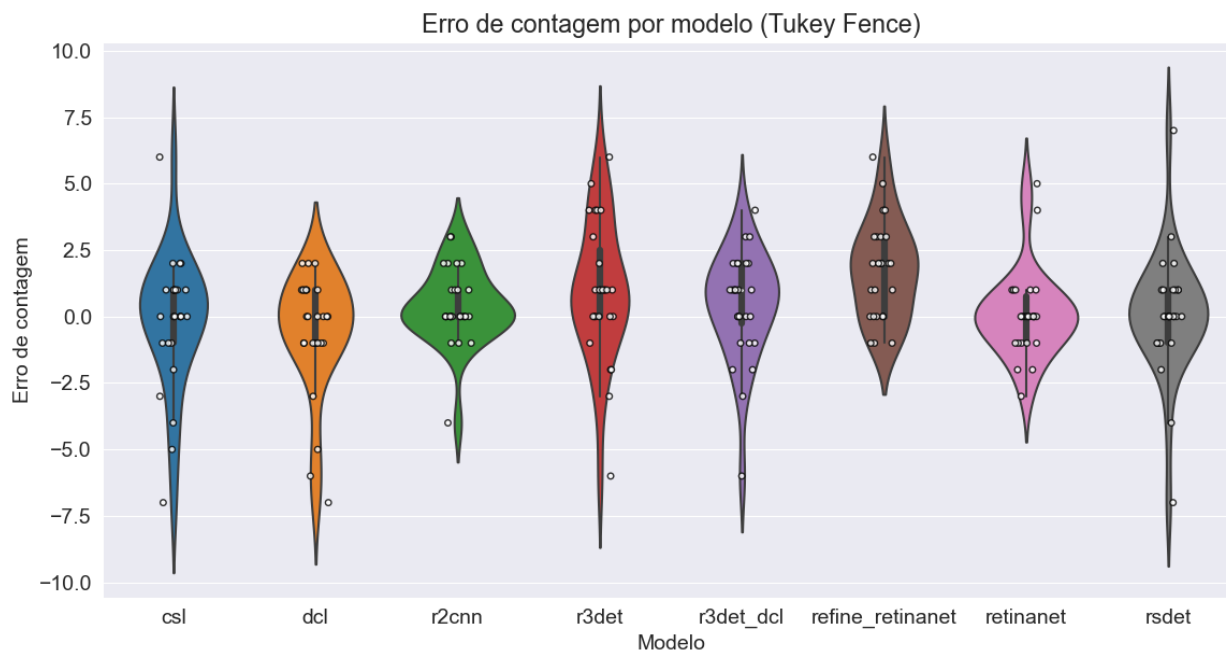


Figura 10: Distribuição de erros de contagem para cada modelo no melhor DT, filtrado a partir de *Tukey Fence* para identificação de outliers. Cada círculo representa o erro de contagens para uma imagem do conjunto de teste.

O valor de MAPE acompanha o erro de contagens em cada imagem - é nulo quando o erro de contagens na imagem é zero (número de OBBs detectadas pelo modelo se iguala ao de OBBs anotadas), e sobe ao passo que o erro aumenta. Utilizando os valores de MAPE para cada modelo (Tabela 4), foi realizado o teste H não paramétrico de Kruskal-Wallis para erros de contagem. O resultado do teste foi negativo ($p\text{-valor} > 0.05$), concluindo que não houve diferença entre modelos em relação à média de erros relativos de contagem de cada um.

4.4 Análise de erros de centroide (RMSE)

A detecção de centroides pelos modelos de aprendizado de máquina teve seu erro calculado a partir do RMSE para cada imagem.

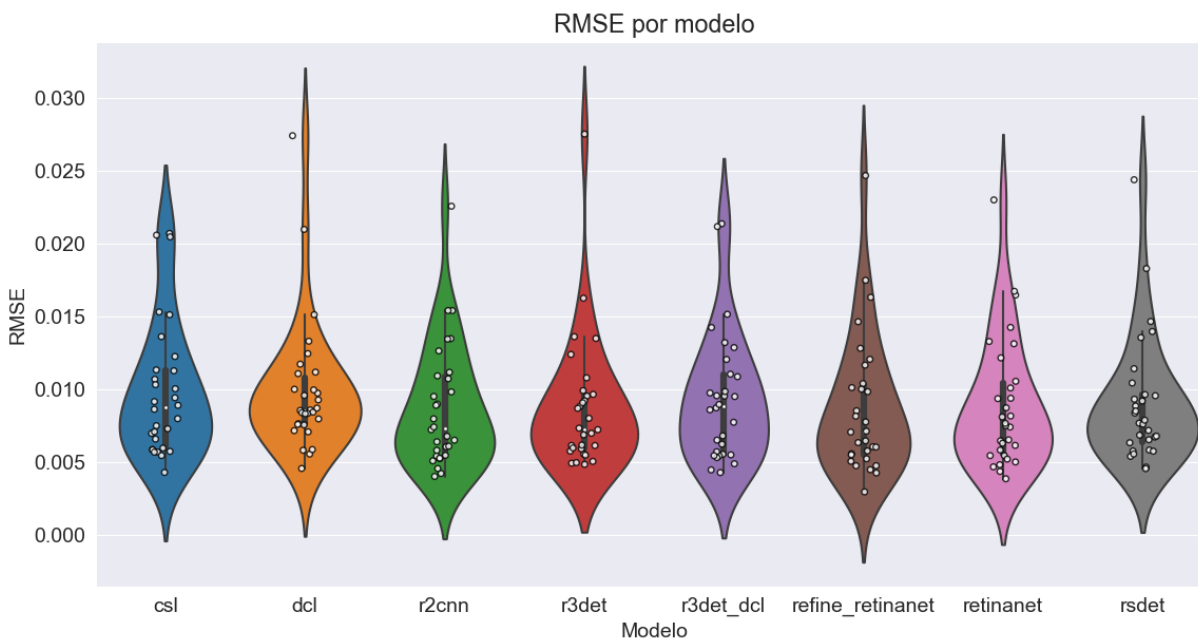


Figura 11: Distribuição de RMSE para cada modelo no melhor DT. Cada círculo representa o erro de contagens para uma imagem do conjunto de teste.

Nota-se que os diferentes modelos apresentaram comportamento bastante similar no que tange o RMSE, salvo os modelos dcl e r3det, com RMSE maior que 0.025 para uma imagem. É importante salientar que os valores de RMSE são importantes descritores para o restante das análises, tendo em vista que refletem as distâncias entre detecções quando comparadas às suas respectivas associações estabelecidas pelo algoritmo Húngaro. É ressaltado que para os cálculos de RMSE, as posições Cartesianas das OBBs foram normalizadas pela resolução da imagem, de modo que a distância utilizada para o cálculo não está em valor de pixels, mas sim relativizadas para um quadrado de 1:1.

Para o cálculo de erro desta etapa e das seguintes, não foi empregado método de remoção de outliers, tendo em vista que estes já foram removidos por conta do filtro de distâncias prévio à

aplicação do algoritmo Húngaro no momento de estabelecer as relações entre células, de modo que os valores descritos por estes erros compreendem somente às OBBs que foram associadas à alguma anotação manual.

O teste H de Kruskal-Wallis também foi realizado utilizando os valores de RMSE médio para cada modelo. No entanto, nenhum se sobressaiu significativamente.

4.5 Análise de erros de área (ARE)

Para cálculo de erro de área, os AREs de cada detecção foram calculados com base na fórmula descrita em 3.3.5, e sua distribuição é ilustrada na Figura 12. Diferentemente dos valores de erro de contagem e RMSE (onde havia um valor por imagem), tem-se expresso o valor de ARE para cada associação de OBBs na imagem, agrupado por modelo. Salienta-se que os maiores erros de área são procedentes de somente duas imagens (círculos em tons laranjas acima de ARE 6.0, Figura 12), podendo ser indicativo de uma imagem de dificuldade elevada.

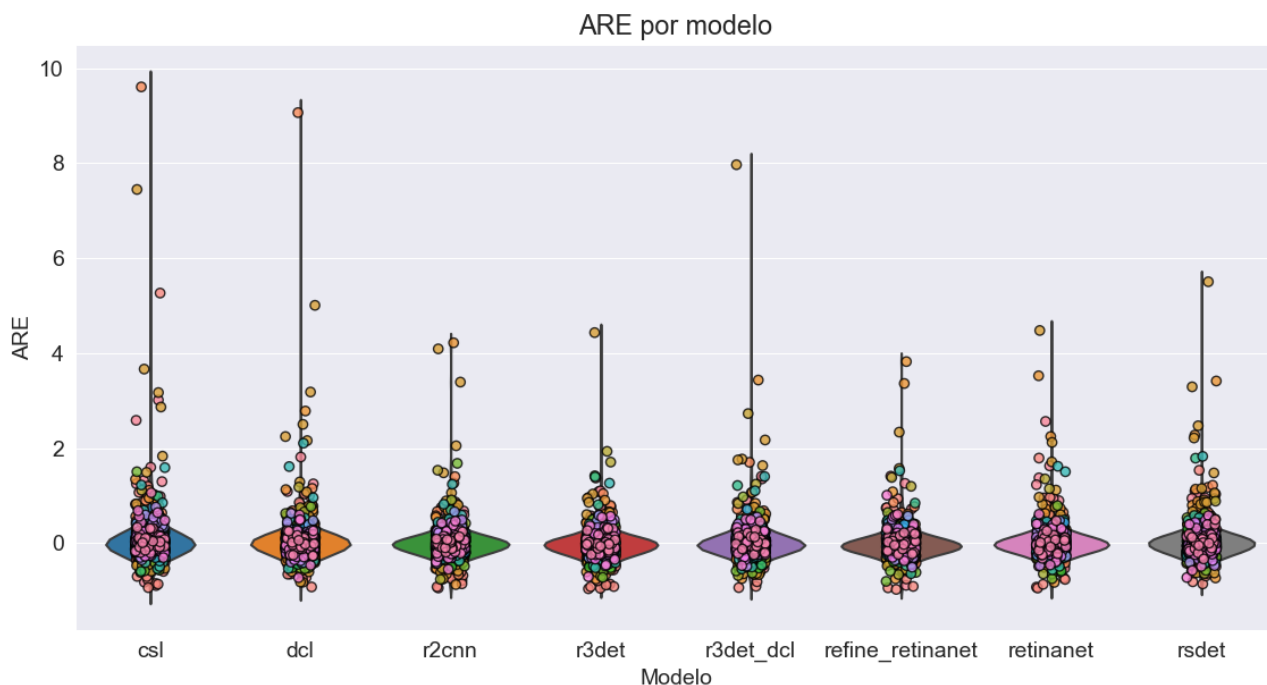


Figura 12: Distribuição de AREs para cada modelo no melhor DT. Cada círculo representa o erro entre um par de OBBs (detecção e anotação), e sua cor indica a imagem a que pertence.

O teste estatístico para erros de área também foi realizado, utilizando os valores absolutos de AREs de cada modelo, que não tiveram diferença estatística.

4.6 Análise de erros de ângulo (PI)

Os erros de ângulo foram analisados com base no seu índice de perpendicularidade (PI), onde os valores descrevem - em um intervalo de 0° a 90° - o quanto a orientação da OBB detectada pelos modelos se distancia da orientação da OBB da anotação.

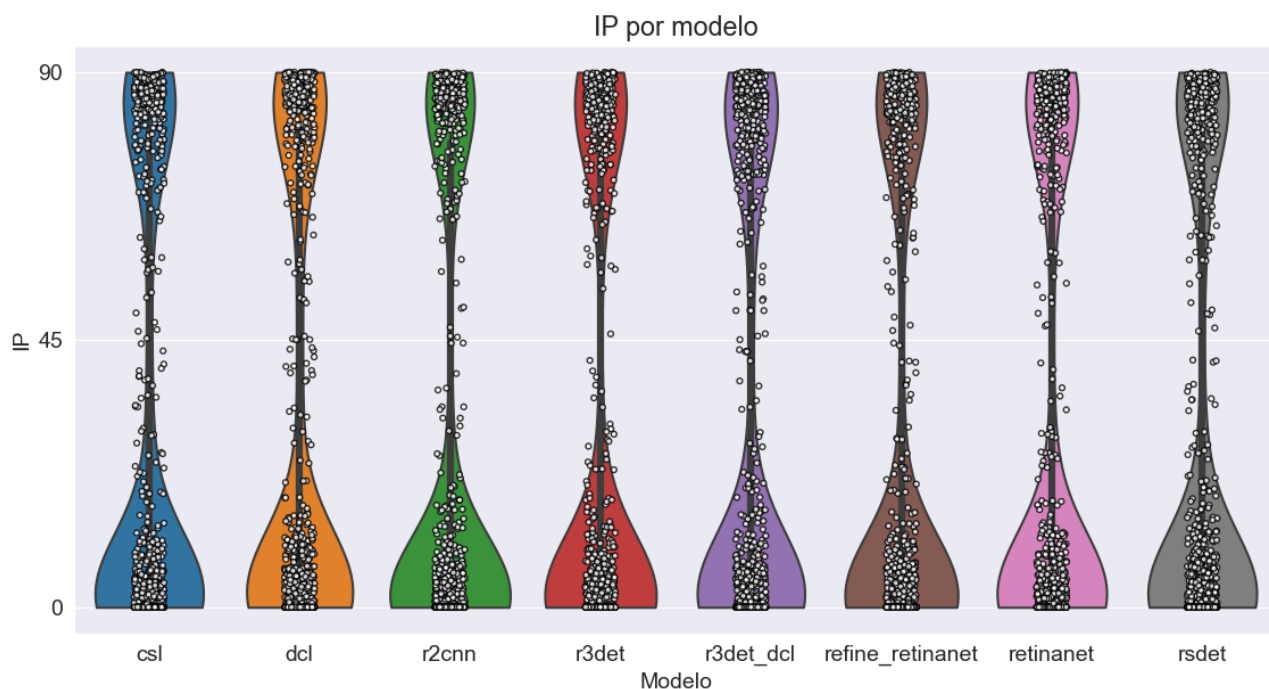


Figura 13: Distribuição de PIs para cada modelo no melhor DT. Cada círculo representa o erro de contagens para uma imagem do conjunto de teste.

Todos os modelos apresentaram PI médio menor que 35° (Tabela 4), o que pode apresentar limitações em aplicações como análise de migração com base na orientação celular, por exemplo. Não houve diferença estatística para erros de ângulo entre modelos.

4.7 Análise de erros de classe (CER)

O desempenho dos modelos em identificar corretamente a classe das células foi avaliado com base na razão de erro de classe por modelo, indicada na Figura 14.

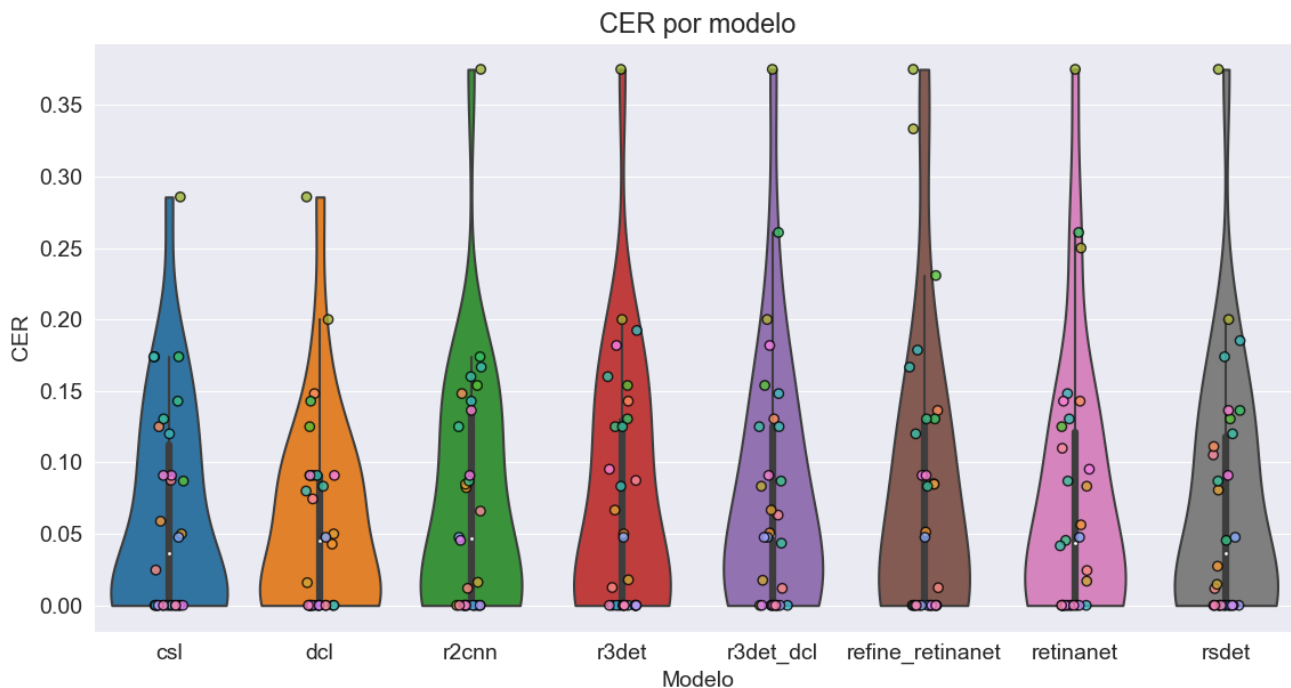


Figura 14: Distribuição de CERs para cada modelo no melhor DT. Cada círculo representa o erro entre um par de OBBs (detecção e anotação), e sua cor indica a imagem a que pertence.

Observa-se que a maior parte de CERs para as imagens de teste ficou inferior à 0.15 (Figura 14). Novamente, os valores mais extremos de erro são provenientes de uma mesma imagem (círculos esverdeados acima de CER 0.3, Figura 14). Todos os modelos treinados apresentaram média de CER inferior a 10%, o que indica a elevada capacidade de classificação pelos modelos em identificar corretamente se uma célula é mitótica ou não mitótica. De acordo com o teste H realizado, as diferenças observadas entre os modelos não foram significativas.

4.8 Média de erro por modelo para cada aplicação

A fim de avaliar de maneira geral o desempenho entre diferentes modelos para cada aplicação, os resultados de erro médio foram calculados, culminando na Tabela 4.

Observa-se que o modelo r2cnn se sobressaiu em média aos outros modelos nos erros de contagem (MAPE), se equiparou a outros modelos nas análises de centroide, área e ângulo, mas apresentou erro mais expressivo na classificação celular. De forma geral, no entanto, todos modelos apresentaram métricas de performance similares, não refletindo em diferença estatística quando avaliados pelo teste H de Kruskal-Wallis.

Tabela 4: Valores médios* de erro para cada modelo de acordo com as cinco diferentes etapas de validação.

Modelo	MAPEm* (%)	RMSEm* (%)	AREm* (%)	PIIm* (%)	CERm* (%)
CSL	7.5	0.0098	0.27	30.6	6.2
DCL	5.7	0.0099	0.23	31.4	5.8
r2cnn	4.8	0.0087	0.21	27.2	7.0
r3det	8.1	0.0087	0.21	29.8	7.4
r3det_dcl	6.4	0.0093	0.22	31.0	7.7
refine_retinanet	9.1	0.0087	0.21	29.3	7.8
retinanet	4.9	0.0088	0.22	30.3	7.2
rsdet	5.7	0.0090	0.23	30.4	6.9

4.9 Interface da ferramenta

A fim de facilitar a utilização da ferramenta por não programadores, uma versão de interface foi desenvolvida (Figura 15), assim como uma versão em linha de comando (Figura 16). As seções essenciais para execução de ambas são descritas abaixo:

1. Seleção da pasta contendo imagens de microscopia celular (aceitando múltiplos formatos como jpg, tiff, jpeg)
2. Seleção do arquivo de pesos do modelo (gerado a partir do treinamento, armazena os parâmetros necessários para detecção de células em novas imagens)
3. Seleção da pasta de *output* (nesta pasta serão salvos os arquivos .csv contendo os dados de detecções, assim como as imagens com o contorno das OBBs, caso o *checkbox/flag* de salvamento com contorno esteja marcado)
4. Seleção do limiar de detecção (pode haver limiar ideal distinto para aplicações ou conjuntos de dados diferentes)

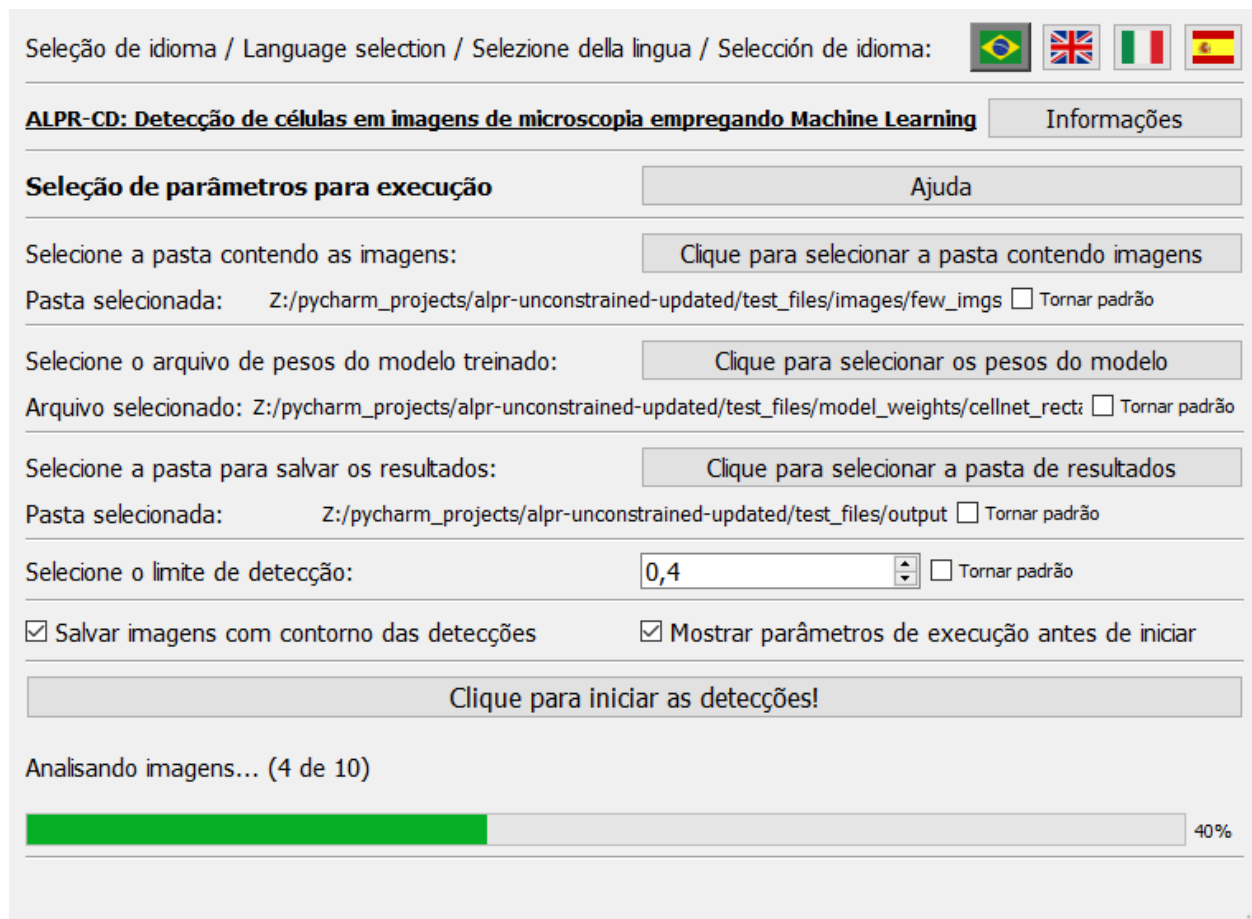


Figura 15: Interface do programa em execução. O programa permite ao usuário selecionar a pasta contendo as imagens, a pasta de resultados, o arquivo contendo os pesos do modelo, bem como o DT a ser aplicado.

```
usage: run_app_cl.py [-h] -i IMAGES_FOLDER_PATH -w MODEL_WEIGHTS_PATH -o OUTPUT_FOLDER_PATH [-t DETECTION_THRESHOLD] [-s]
alpr-CD - ML Based Cell Detection in Microscopy Images

optional arguments:
  -h, --help            show this help message and exit
  -i IMAGES_FOLDER_PATH, --imgs-folder IMAGES_FOLDER_PATH
                        defines path to folder containing cell microscopy images
  -w MODEL_WEIGHTS_PATH, --model_weights MODEL_WEIGHTS_PATH
                        defines path to trained model weights (.h5 file)
  -o OUTPUT_FOLDER_PATH, --output-folder OUTPUT_FOLDER_PATH
                        defines path to output folder (where images and files will be stored)
  -t DETECTION_THRESHOLD, --detection-threshold DETECTION_THRESHOLD
                        defines threshold or cells detection (default: 0.3)
  -s, --save-images     defines whether or not to save images outlined with detections (default: False)
```

Figura 16: Exemplo de execução do comando de ajuda do programa na versão por linha de comando.

5. DISCUSSÃO

Tradicionalmente em modelos de aprendizado de máquina, a intersecção sobre união (IoU) é utilizada como parâmetro para determinação de acerto da predição feita pela rede. No entanto, ao se fazer valer da sobreposição e união da área de células, esta métrica pode ser excessivamente restritiva para aplicações onde o interesse do pesquisador é somente o número ou a posição das células, por exemplo (Figura 17). Uma pequena variação no ângulo das OBBs pode representar uma grande oscilação no valor de IoU, ainda que, do ponto de vista de contagem ou determinação da posição do centroide, isso não seja significativo. Dessa forma, foi decidido utilizar o valor de FPNR em conjunto com o algoritmo Húngaro para associar detecções a anotações manuais, em diferentes valores de DT.

No presente trabalho, oito modelos de aprendizado de máquina previamente treinados foram avaliados quanto à sua capacidade de detecção de células tumorais em imagens de microscopia de luz visível. Inicialmente, para a determinação do melhor limiar de detecção, foi utilizado o cálculo de FPNR, como definido na seção 3.3.2. Esta métrica parte de duas premissas principais: 1), falsos positivos têm o mesmo peso que falsos negativos; 2), falsos positivos e falsos negativos são mais graves quando há menos células na imagem. Compreende-se, no entanto, que de acordo com a situação e aplicação de interesse do pesquisador, estas premissas podem não ser verdadeiras, e diferentes métricas para estabelecimento de melhor DT podem ser exploradas.

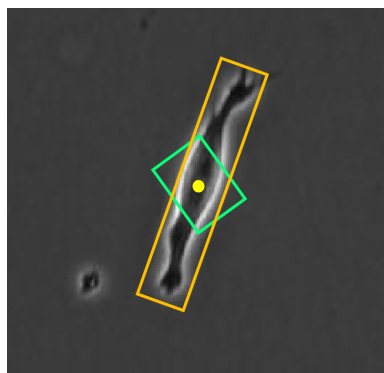


Figura 17: Exemplo de detecção com IoU baixo, mas distância entre centroides nula. Caso IoU fosse empregado como métrica para determinação de acerto, a detecção em verde seria desconsiderada, embora para fins de análise de simples contagem ou posição Cartesiana estaria adequada.

Neste projeto, optou-se por usar a distância Euclidiana entre centroides de OBBs para o estabelecimento de relações entre detecções dos modelos e anotações manuais. Esta abordagem, no entanto, sofre de um problema: associações entre OBBs distantes, ainda que sejam de alto custo, ainda podem ser feitas caso somente o seu centroide seja considerado, enquanto uma métrica como o IoU evitaria associações em que não há sobreposição da área de OBBs. Para solucionar este desafio, um filtro anterior ao estabelecimento de relação entre células foi aplicado, de modo que associações com distância maior do que um limite foram desconsideradas e somadas aos valores de falsos positivos e falsos negativos do FPNR.

Para o estabelecimento das relações entre OBBs detectadas pelos modelos às OBBs manualmente anotadas na imagem, foi empregado o algoritmo Húngaro, com base na criação de uma matriz de custo gerada a partir das distâncias de todas as associações possíveis entre OBBs. O algoritmo Húngaro, no entanto, permite com que outros valores - além do centroide - sejam considerados no momento de geração desta matriz de custo (KUHN, 1955). A área, ângulo e classe celular, por exemplo, poderiam entrar com pesos diferentes nesta etapa, e esta abordagem se provar mais eficaz do que simplesmente a utilização do IoU. Tem-se como perspectiva a realização de comparações entre diferentes métricas, como entre mAP obtida por IoU e a partir de FPNR.

Com base nos resultados apresentados, foi possível avaliar os modelos quanto a suas capacidades de identificação de células em imagens de microscopia, bem como as taxas de erro para cada aplicação.

Na aplicação de contagem de células na imagem, os modelos apresentaram elevada taxa de acerto, considerando que os valores de MAPE médio para todos ficaram abaixo de 10%. Com base nas experiências de análises manuais realizadas no laboratório, este erro é próximo do encontrado quando diversos avaliadores realizam a contagem manual de células, especialmente em imagens com maior densidade celular. Neste cenário de análises de grandes conjuntos de dados, em que mecanismos de reconhecimento de *outliers* possam ser aplicados, a taxa de erro próxima a 10% pode ser considerada aceitável.

Já quanto à identificação da posição celular, foi empregado o cálculo de RMSE. Tendo em vista que as distâncias foram normalizadas de acordo com a resolução da imagem, o valor de RMSE pode ser interpretado como a distância dentro de um quadrado de lado 1 - onde a distância máxima possível entre dois objetos seria $\sim 1,4$. Assim, como todos modelos

apresentaram distância média entre OBBs menor que 0,01, conclui-se que também foram capazes de identificar corretamente a posição das células nas imagens.

Para a análise de área, as diferenças de área relativas para cada OBB foram calculadas, gerando o ARE. A maior parte dos modelos ficou com erro relativo em torno de 0,21, representando desvio médio de ~20% (para mais ou menos) quanto à identificação de área das células. Ressalta-se que, dependendo da aplicação, esta taxa de erro seria incompatível com as análises, visto que variações na área de OBBs dentro deste intervalo não teriam como ser discriminadas entre seu contexto biológico e margem de erro da ferramenta.

Quanto ao ângulo de orientação das OBBs, observa-se que em média os modelos apresentaram erro próximo à 30° (Tabela 4). No entanto, ao se analisar a distribuição de erros na Figura 13, podem ser percebidos dois grandes grupos. O grupo próximo a zero compreende detecções cuja diferença de ângulo foi baixa, indicando elevada taxa de acerto. Já o grupo próximo a 90° pode ser explicado devido ao limite de razão de eixos aplicado. Ao filtrar as detecções por “quadrados” e “retângulos não quadrados”, é introduzida uma variável adicional. Desse modo, quadrados com diferença de ângulo alta podem estar sobrepostos, e não representam um erro, mas em virtude do baixo limite de razão de eixos, são contabilizados como OBBs não quadradas, tendo seu desvio de ângulo calculado, conseqüentemente deslocando o erro de ângulo para próximo a 90°.

Já a capacidade de classificação das OBBs foi avaliada com base na taxa de erro de classe. Para esta aplicação, todos modelos apresentaram erro inferior a 8%, validando a ferramenta como um classificador de células mitóticas / não mitóticas para diversas aplicações. Nas anotações utilizadas para treinamento dos modelos, o número de OBBs classificadas como não-mitóticas foi nove vezes maior que o de mitóticas. Este importante desbalanço pode provocar viés nas predições pelo modelo. Dessa forma, é esperado que o treinamento com conjunto de dados com maior equilíbrio entre classes resulte em ainda menor erro de classificação.

É importante salientar que, para todas as métricas de erro calculadas, é esperada variabilidade com outros conjuntos de imagens, em especial caso estas contenham linhagens celulares distintas, ou sua captura tenha sido feita em microscópios diferentes. Novas etapas de treinamento com conjuntos de imagens adicionais, advindos de configurações experimentais distintas poderão ser realizadas, a fim de ajustar a ferramenta para outras aplicações.

Ademais, o processo de anotação de imagens para geração do conjunto de dados de treinamento também é uma etapa com variabilidade inerente. Diferentes pesquisadores poderão descrever uma mesma célula com OBBs distintas, em especial nos casos mais desafiadores, como ilustrado na Figura 9. Tem-se como perspectiva analisar a variabilidade entre anotadores, de modo a avaliar se os modelos se enquadram em uma taxa de erro compatível com o esperado entre diferentes humanos.

A partir da área sob a curva (AUC) das curvas de *precision-recall*, será possível calcular o valor de mAP para cada modelo, permitindo comparação com diferentes métricas de estabelecimento de TPs, FPs e FNs, como a partir da IoU. Futuramente, espera-se explorar as comparações entre tais métricas para avaliação de desempenho de modelos de aprendizado de máquina voltados à análise de células em imagens de microscopia.

6. CONCLUSÃO

O presente trabalho teve por objetivo o desenvolvimento de uma ferramenta capaz de detectar células em imagens de microscopia de luz visível, bem como obter características celulares como área, orientação e fase do ciclo. A partir de um conjunto de imagens anotadas, oito modelos foram treinados e testados quanto à capacidade de identificação destas características, e os respectivos erros foram mensurados. O teste estatístico não paramétrico de Kruskal-Wallis foi realizado para verificar diferenças entre modelos. Os resultados não foram significativos em nenhuma das cinco tarefas, indicando que os modelos testados tiveram desempenhos similares. Todos apresentaram faixas de erro aceitáveis para múltiplas aplicações, em especial para cenários de enormes conjuntos de dados, onde a análise manual é inviabilizada. Nota-se, no entanto, limitações da ferramenta que devem ser consideradas, como a densidade de células na imagem, conforme avaliações nas imagens de teste. Em imagens com maior confluência (densidade de células na imagem), os modelos apresentaram dificuldade de identificar corretamente o número de células, resultando em elevada taxa de falsos negativos. Ressalta-se, também, que o conjunto de dados utilizado continha imagens de duas diferentes magnificações (10x e 20x), com resultados de detecção semelhantes. No entanto, ressalta-se que a aplicação da ferramenta para identificar células em imagens com ampliações diferentes das empregadas no conjunto de treino poderá resultar em elevação das taxas de erro. A partir das versões desenvolvidas da ferramenta (linha de comando e interface), espera-se que o processo de análises envolvendo células únicas em imagens de microscopia de luz visível seja acelerado - desde que respeitadas as magnificações e linhagens celulares utilizadas, bem como o DT ideal determinado para cada modelo no presente trabalho. Futuramente, tem-se como perspectiva realizar novas etapas de treinamento, com conjuntos de dados mais variados, permitindo a aplicabilidade da ferramenta em contextos diversos.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ABRÀMOFF, D. M. D. Image Processing with ImageJ. p. 7, 2004.

BALABAN, N. Q. et al. Bacterial Persistence as a Phenotypic Switch. **Science**, v. 305, n. 5690, p. 1622–1625, 10 set. 2004.

CGVICT. **roLabelImg**. Disponível em: <<https://github.com/cgvict/roLabelImg>>. Acesso em: 10 abr. 2022.

DAGOGO-JACK, I.; SHAW, A. T. Tumour heterogeneity and resistance to cancer therapies. **Nature Reviews Clinical Oncology**, v. 15, n. 2, p. 81–94, fev. 2018.

DENG, J. et al. ImageNet: A Large-Scale Hierarchical Image Database. p. 8, 1 jan. 2009.

GANINI, D. et al. Fluorescent proteins such as eGFP lead to catalytic oxidative stress in cells. **Redox Biology**, v. 12, p. 462–468, ago. 2017.

GUPTA, A.; VEDALDI, A.; ZISSERMAN, A. Synthetic Data for Text Localisation in Natural Images. **arXiv:1604.06646 [cs]**, 22 abr. 2016.

HE, K. et al. **Deep Residual Learning for Image Recognition**. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). **Anais...** Em: 2016 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR). Las Vegas, NV, USA: IEEE, jun. 2016. Disponível em: <<http://ieeexplore.ieee.org/document/7780459/>>. Acesso em: 11 abr. 2022

HILSENBECK, O. et al. Software tools for single-cell tracking and quantification of cellular and molecular properties. **Nature Biotechnology**, v. 34, n. 7, p. 703–706, jul. 2016.

HIRSCH, F. R. et al. Tumor heterogeneity in lung cancer based on light microscopic features: A retrospective study of a consecutive series of 200 patients, treated surgically. **Virchows Archiv A Pathological Anatomy and Histopathology**, v. 402, n. 2, p. 147–153, 1983.

ICHA, J. et al. Phototoxicity in live fluorescence microscopy, and how to avoid it. **BioEssays**, v. 39, n. 8, p. 1700003, ago. 2017.

JAQAMAN, K. et al. Robust single-particle tracking in live-cell time-lapse sequences. **Nature Methods**, v. 5, n. 8, p. 695–702, ago. 2008.

JIANG, Y. et al. R2CNN: Rotational Region CNN for Orientation Robust Scene Text Detection. **arXiv:1706.09579 [cs]**, 30 jun. 2017.

KAMENTSKY, L. et al. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. **Bioinformatics**, v. 27, n. 8, p. 1179–1180, 15 abr. 2011.

- KIM, J. H. et al. Unjamming and collective migration in MCF10A breast cancer cell lines. **Biochemical and Biophysical Research Communications**, v. 521, n. 3, p. 706–715, jan. 2020.
- KOMURA, D.; ISHIKAWA, S. Machine Learning Methods for Histopathological Image Analysis. **Computational and Structural Biotechnology Journal**, v. 16, p. 34–42, 2018.
- KRUSKAL, W. H.; WALLIS, W. A. Use of Ranks in One-Criterion Variance Analysis. **Journal of the American Statistical Association**, v. 47, n. 260, p. 583–621, dez. 1952.
- KUHN, H. W. The Hungarian method for the assignment problem. **Naval Research Logistics Quarterly**, v. 2, n. 1–2, p. 83–97, mar. 1955.
- LENZ, L. S. et al. Cancer Cell Fitness Is Dynamic. **Cancer Research**, v. 81, n. 4, p. 1040–1051, 15 fev. 2021.
- LI, S. et al. Multiscale Rotated Bounding Box-Based Deep Learning Method for Detecting Ship Targets in Remote Sensing Images. **Sensors**, v. 18, n. 8, p. 2702, 17 ago. 2018.
- LIN, T.-Y. et al. Focal Loss for Dense Object Detection. **arXiv:1708.02002 [cs]**, 7 fev. 2018.
- MA, J. et al. Ship Detection in Optical Satellite Images via Directional Bounding Boxes Based on Ship Center and Orientation Prediction. **Remote Sensing**, v. 11, n. 18, p. 2173, 18 set. 2019.
- PYTHON SOFTWARE FOUNDATION. **Python**. Disponível em: <<https://www.python.org/>>. Acesso em: 10 abr. 2022.
- QIAN, W. et al. Learning Modulated Loss for Rotated Object Detection. **arXiv:1911.08299 [cs]**, 20 dez. 2019.
- QT. **Qt for Python**. Disponível em: <<https://doc.qt.io/qtforpython/>>. Acesso em: 10 abr. 2022.
- REN, S. et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 39, n. 6, p. 1137–1149, 1 jun. 2017.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. **arXiv:1505.04597 [cs]**, 18 maio 2015.
- SILVA, S. M.; JUNG, C. R. License Plate Detection and Recognition in Unconstrained Scenarios. Em: FERRARI, V. et al. (Eds.). **Computer Vision – ECCV 2018**. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018. v. 11216p. 593–609.
- TUKEY, J. W. **Exploratory data analysis**. 17. ed. [s.l.: s.n.].
- ULMAN, V. et al. An objective comparison of cell-tracking algorithms. **Nature Methods**, v. 14, n. 12, p. 1141–1152, dez. 2017.

VICAR, T. et al. Cell segmentation methods for label-free contrast microscopy: review and comprehensive comparison. **BMC Bioinformatics**, v. 20, n. 1, p. 360, 28 jun. 2019.

WANG, Y. et al. Quantitative Characterization of Cell Behaviors through Cell Cycle Progression via Automated Cell Tracking. **PLoS ONE**, v. 9, n. 6, p. e98762, 9 jun. 2014.

YANG, X. et al. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. **arXiv:1908.05612 [cs, eess]**, 8 dez. 2020.

YANG, X. et al. Dense Label Encoding for Boundary Discontinuity Free Rotation Detection. **arXiv:2011.09670 [cs]**, 25 maio 2021.

YANG, X.; YAN, J. On the Arbitrary-Oriented Object Detection: Classification based Approaches Revisited. **arXiv:2003.05597 [cs]**, 23 mar. 2022.

YANGXUE0827. **RotationDetection**. Disponível em: <<https://github.com/yangxue0827/RotationDetection>>. Acesso em: 10 abr. 2022.

ZHAO, Z.-Q. et al. Object Detection With Deep Learning: A Review. **IEEE Transactions on Neural Networks and Learning Systems**, v. 30, n. 11, p. 3212–3232, nov. 2019.

ZHONG, Y.; JI, B. Impact of cell shape on cell migration behavior on elastic substrate. **Biofabrication**, v. 5, n. 1, p. 015011, 10 jan. 2013.