TRABALHO DE GRADUAÇÃO

DESENVOLVIMENTO DE TÉCNICAS PARA SEGMENTAÇÃO E DETECÇÃO DE GLOMERULOPATIAS UTILIZANDO APRENDIZAGEM DE MÁQUINA

Rodrigo Naves Rios

João Viktor de Carvalho Mota

Brasília, Novembro de 2021

ENGENHARIA MECATRÔNICA
UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASILIA
Faculdade de Tecnologia
Curso de Graduação em Engenharia de Controle e Automação

TRABALHO DE GRADUAÇÃO

DESENVOLVIMENTO DE TÉCNICAS PARA SEGMENTAÇÃO E DETECÇÃO DE GLOMERULOPATIAS UTILIZANDO APRENDIZAGEM DE MÁQUINA

Rodrigo Naves Rios

João Viktor de Carvalho Mota

Relatório submetido como requisito parcial de obtenção de grau de Engenheiro de Controle e Automação

Banca Examinadora

Prof. Dr. Flávio de Barros Vidal, CIC/UnB _____
Orientador

Prof. Dr. Marcus Chaffim - FGA/UnB _____
Examinador Externo

Prof. Dr. Luis Paulo Faina Garcia - CIC/UnB _____
Examinador Interno

Brasília, Novembro de 2021

FICHA CATALOGRÁFICA

RIOS, RODRIGO NAVES; MOTA, JOÃO VIKTOR DE CARVALHO
Desenvolvimento de técnicas para segmentação e detecção de glomerulopatias utilizando Aprendizagem de Máquina

[Distrito Federal] 2021.

x, 97p., 297 mm (FT/UnB, Engenheiro, Controle e Automação, 2021). Trabalho de Graduação – Universidade de Brasília. Faculdade de Tecnologia.

1. Redes Neurais Convolucionais 2. Glomerulopatias
3. Podócitos

I. Mecatrônica/FT/UnB II. Título (Série)

REFERÊNCIA BIBLIOGRÁFICA

RIOS, R. N.; MOTA, J. V. C., (2021). Desenvolvimento de técnicas para segmentação e detecção de glomerulopatias utilizando Aprendizagem de Máquina. Trabalho de Graduação em Engenharia de Controle e Automação, Publicação FT.TG-$n°$01, Faculdade de Tecnologia, Universidade de Brasília, Brasília, DF, 97p.

CESSÃO DE DIREITOS

AUTORES: Rodrigo Naves Rios e João Viktor de Carvalho de Mota

TÍTULO DO TRABALHO DE GRADUAÇÃO: Desenvolvimento de técnicas para segmentação e detecção de glomerulopatias utilizando Aprendizagem de Máquina.

GRAU: Engenheiro ANO: 2021

Rodrigo Naves Rios.
SHCGN 704 Bloco K, Casa 29, Asa Norte.
70730-741 Brasília – DF – Brasil.

João Viktor de Carvalho Mota.
Cond. Mirante das Paineiras Conj. 3, Lt. 18, JB.
71680-367 Brasília – DF – Brasil.

Dedicatórias

Dedico este trabalho aos meus pais, meus avós, meu irmão e meus amigos.

João Viktor de Carvalho Mota

Aos trabalhadores de serviços essenciais e à memória das vítimas da pandemia de Covid-19.

Rodrigo Naves Rios

Agradecimentos

Em primeiro lugar, agradeço a meus pais pelo apoio irrestrito e por não cobrar de mim nada mais do que minha própria felicidade. Por extensão, agradeço à minha irmã, que tanta felicidade me deu com a vinda ao mundo de meu sobrinho Luís, e a meus avós. Aos meus amigos, a quem não menciono nominalmente, menos por receio de cometer alguma injustiça do que pelo exercício de concisão a que me submeteria. Ao meu amigo de todas as horas e coautor deste texto, João. Tenho a sorte de ter ao meu lado muitos amigos e amigas, sem os quais o isolamento a que aderi convictamente nesta pandemia teria sido ainda mais desafiador. Ao meu orientador, Flávio Vidal, pela solicitude e atenção a nós dispensada ao longo da redação deste manuscrito e por todos os ensinamentos que nos passou. Por fim, ao colega de universidade e doutorando George, sem cujo suporte não seria possível a realização deste trabalho.

Rodrigo Naves Rios

Gostaria de agradecer meus pais, Wagner e Kerre Anne, por sempre me apoiarem e por sempre estarem comigo. Também agradecer meu irmão, Pietro, que sempre me ajudou em todas as minhas dificuldades. Agradecer meus avós que sempre acreditaram em mim. Agradecer meus amigos que também sempre estiveram comigo por todas as dificuldades. Agradecer todos os meus professores da UnB que me ajudaram a formar a pessoa que sou hoje, entre eles meu orientador, Flávio Vidal, que ajudou muito neste processo. Também agradecer o doutorando, George, que ajudou muito na realização deste trabalho. E por fim, agradecer a minha dupla, Rodrigo, que foi um dos amigos que mais me ajudou. Obrigado de verdade a todos vocês.

João Viktor de Carvalho Mota

RESUMO

O emprego de técnicas em Aprendizado de Máquina (*Machine Learning*) de forma combinada às de Visão Computacional tem melhorado a performance de sistemas de visão. A análise automática de imagens é um campo em que essa interseção se aplica. O campo da Patologia Digital se vale de técnicas em Processamento Digital de Imagens e Visão Computacional para prover análises de tecidos biológicos. Na última década, métodos em *Deep Learning*, em especial as redes neurais convolucionais (CNNs), têm sido amplamente utilizados como ferramenta para análise de imagens histológicas. O trabalho desenvolvido apresenta a elaboração de técnicas automáticas para a segmentação de imagens dentro do escopo do problema de identificação de glomerulopatia em imagens histológicas renais. A segmentação proposta visa a extração de estruturas de células em imagens, para que se possa posteriormente identificar e classificar lesões com um grau significativo de precisão na detecção de núcleos e podócitos em imagens histológicas de glomérulos usando redes neurais convolucionais profundas. O modelo utilizado neste trabalho foi capaz de detectar núcleos com precisão média (AP) de 0,92. Com respeito às detecções de podócitos, o modelo alcançou AP de 0,70. Em ambos os casos, o resultado foi atingido por meio de emprego de pré-treino e da expansão do conjunto de dados.

Palavras Chave: Redes Neurais Convolucionais, Glomerulopatias, podócitos.

ABSTRACT

The use of techniques in Machine Learning (*Machine Learning*) combined with Computer Vision has improved the performance of vision systems. Automatic image analysis is a field where this intersection applies. The area of Digital Pathology uses techniques in Digital Image Processing and Computer Vision to provide analysis of biological tissues. In the last decade, methods in *Deep Learning*, especially convolutional neural networks (CNNs), have been widely used as a tool for analyzing histological images. The work developed presents the development of automatic techniques for image segmentation within the scope of the problem of identifying glomerulopathy in renal histological images. The proposed segmentation aims at extracting cell structures in images. It can later identify and classify lesions with a significant degree of precision in detecting nuclei and podocytes in histological images of glomeruli using deep convolutional neural networks. The model used in this work detected nuclei with an average precision (AP) of 0.92. For podocyte detections, the model achieved an AP of 0.70. In both cases, the result was achieved by employing pre-training and expanding the dataset.

Keywords: Convolutional Neural Networks, Glomerulopathies, podocytes.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SÍMBOLOS

Siglas

UnB	Universidade de Brasília	
CIC	Ciência da Computação	
FGA	Faculdade do Gama	
ML	*Machine Learning*	Aprendizagem de Máquina
DL	*Deep Learning*	Aprendizagem Profunda
PDI	Processamento Digital de Imagens	
CNN	*Convolutional Neural Network*	Redes Neurais Convolucionais
DNN	*Deep Neural Network*	Rede Neural Profunda
AP	*Average Precision*	Precisão Média
RNN	*Recurrent Neural Network*	Redes Neurais Recursivas
CV	*Cross Validation*	Validação Cruzada
SVM	*Support Vector Machine*	Máquina de Vetores de Suporte
kNN	*k-nearest neighbors*	K-ésimo Vizinho mais Próximo
P	Precisão	
R	*Recall*	Revocação
F1	F1-*score*	
H&E	Hematoxilina-Eosina	
PAS	*Periodic acid-reactive Schiff*	Ácido Periódico de Schiff
WSI	*Whole Slide Images*	
IoU	*Intersection over Union*	Interseção sobre União
MLP	*Multi-Layer Perceptron*	Perceptron Multicamadas
SGD	*Stochastic Gradient Descent*	Gradiente Descendente Estocástico
DA	*Data Augmentation*	Aumento de Dados

Capítulo 1

Introdução

1.1 Contextualização

O campo de Processamento Digital de Imagens (PDI) pode ser definido pelo processamento de imagens digitais por meio de um computador [1]. Grosso modo, ele se refere a operações em que tanto a entrada como a saída do processo são imagens[1]. Entre seus objetivos primários está a extração de rudimentos da imagem, como remoção de ruídos, realce de contraste e filtragem.

A Visão Computacional, por outro lado, se refere ao uso de métodos estatísticos para depreender dados [2], tendo em vista produzir uma descrição semântica de objetos físicos de imagens [3]. Suas aplicações incluem reconhecimento de gestos, interpretação de conteúdo multimídia e auxílio à movimentação de robôs. Além disso, tarefas em Visão Computacional são comumente divididas em três categorias, que refletem o nível de abstração da informação: baixo nível, médio nível e alto nível. Por exemplo, a tarefa de buscar correspondentes de um conjunto de pontos em uma imagem pertence à primeira categoria, enquanto que a tarefa de descrever semanticamente um filme - identificar o gênero a que ele pertence, a título de exemplo - uma tarefa de um nível de abstração mais alto, pertence à terceira categoria [4].

Ademais, o emprego de técnicas em Aprendizagem de Máquina (*Machine Learning*) de forma combinada às de Visão Computacional tem melhorado a performance de sistemas de visão. A análise automática de imagens é um campo em que essa interseção se aplica. A Figura 1.1 mostra algumas tarefas.

O campo da Patologia Digital se vale de técnicas em Processamento Digital de Imagens e Visão Computacional para prover análises de tecidos biológicos [6]. Na última década, métodos em *Deep Learning*, em especial as redes neurais convolucionais (CNNs), têm sido amplamente utilizados como ferramenta para análise de imagens histológicas [7], embora estas tenham sido desenvolvidas há muito mais tempo.

Uma visão geral da aplicação de *Deep Learning* ao campo de Imagens Médicas, mostra que há

[1] Alguns autores adotam uma definição mais ampla sobre PDI. Por exemplo, há quem considere um *continuum* entre PDI e Visão Computacional [1]

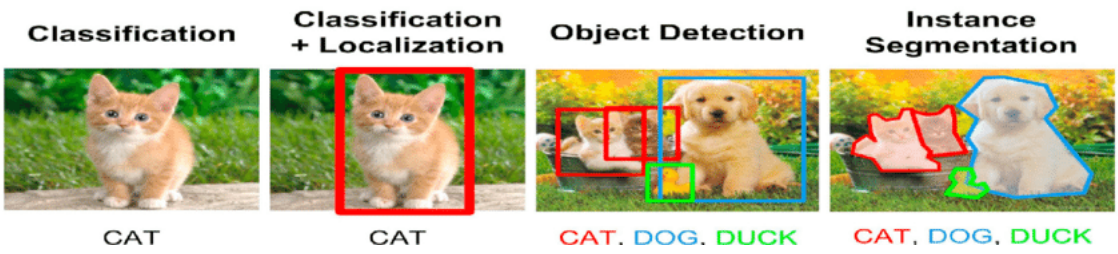


Figura 1.1: Exemplo das tarefas de classificação, localização, detecção e segmentação [5]

grande diversidade de publicações desde 2012 em termos dos objetos de análise (membranas, núcleos, citoplasma, mitose), das tarefas (segmentação semântica, detecção de objetos) e dos órgãos a que pertencem os objetos (cérebro, mamas, coluna cervical, rins) [8]. Neste trabalho, serão utilizadas redes neurais convolucionais, uma ferramenta de Aprendizagem de Máquina, para identificar e localizar estruturas biológicas em imagens de glomérulos renais. Os glomérulos são unidades dos rins responsáveis pela filtragem do sangue, direcionando as substâncias do seu filtrado - conhecido como filtrado glomerular - para os túbulos renais. As duas categorias em que se encaixam a maioria das doenças glomerulares são a glomerulonefrite e a glomeruloesclerose [9]. Com respeito a este trabalho, serão detectadas duas estruturas: núcleos, de forma geral, e podócitos. Alguns exemplos desse tipo de imagem histológica são encontrados ao longo do manuscrito, como na Seção 3.3. Futuramente, pode-se refinar o processo, classificando as células em lesionadas ou não lesionadas, ou ainda identificando o tipo específico de lesão. Além disso, partindo da localização dessas estruturas nos glomérulos, podem ser extraídas características que permitam associar uma dada amostra a um diagnóstico. Em razão da enumeração de termos acima, cabe fazer uma breve conceituação.

Aprendizagem de Máquina (*Machine Learning*) se refere a algoritmos capazes de melhorar automaticamente por meio da experiência [10]. Nesse sentido, *Deep Learning* é uma subárea de *Machine Learning* que faz uso do aumento de capacidade computacional alcançado nos últimos anos para realizar a aprendizagem. Há duas vantagens fundamentais neste campo: a escalabilidade, ou seja, a possibilidade de lidar com grandes conjuntos de dados, em contraste do que ocorre com técnicas clássicas de *Machine Learning*, e a transferência de domínio, isto é, a utilização dos dados apreendidos em um conjunto de dados para realizar uma tarefa em um outro conjunto, ainda que não haja relação entre os dois tipos de dados. Esta última característica pode ser especialmente relevante em um cenário de escassez de dados.

Devido à sua vasta aplicabilidade, entre outras razões, as redes neurais convolucionais (CNNs) têm sido empregadas na pesquisa científica. Nas redes de detecção de objetos em especial, diversos trabalhos puderam ser desenvolvidos a partir de arquiteturas desenvolvidas nos últimos anos [11] [12] [13]. As CNNs são um entre os vários tipos de redes neurais, como as recursivas (RNNs) e as *Long Short Term Memory* (LSTMs). As CNNs são uma classe aplicável tanto ao aprendizado supervisionado, que se refere ao mecanismo em que os modelos recebem dados de entrada e rótulos descritores da entrada, e devem mapear o primeiro ao segundo por meio de predições, quanto ao aprendizado não-supervisionado, em que se pretende estabelecer relações mais profundas sobre dados de entrada não rotulados. Um ponto importante é que as CNNs realizam

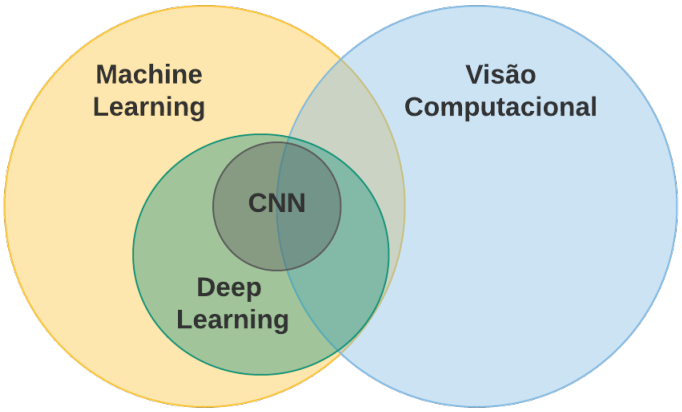


Figura 1.2: Diagrama de Venn das redes neurais convolucionais de forma simplificada.

de forma automática a hierarquização das representações de *features*, além de promover uma integração entre a etapa de extração delas e a classificação. Maiores detalhes, serão apresentados na Subseção 3.2.1.

1.2 Justificativa

A utilização de técnicas automáticas pode colaborar com a prática médica e assim contribuir com a sociedade. Em primeiro lugar, no futuro, o refinamento das técnicas em análise automática poderão auxiliar laboratórios em que haja indisponibilidade de quadros técnicos qualificados ou recursos. Além disso, a construção de grandes bases de dados anotadas poderão ser úteis no ensino de jovens patologistas. Existe ainda a possibilidade de expansão das fronteiras do conhecimento por meio do avanço da Patologia Digital: o reconhecimento de padrões em imagens histológicas pode ser empregado para melhorar os sistemas de classificação já existentes.

1.3 Descrição do problema e objetivos

Técnicas de *Machine Learning* têm sido largamente utilizadas nos últimos anos como ferramento de auxílio à Patologia. De forma especial, as redes neurais convolucionais têm chamado atenção de pesquisadores do mundo todo. Tendo em vista a sua recente incorporação ao campo de imagens histológicas, há ainda muito espaço para pesquisa com CNNs em doenças menos conhecidas ou cujo interesse de pesquisa é menor. Some-se a isto o fato de que há fatores intrínsecos como a complexidade de algumas estruturas biológicas, a indisponibilidade de grandes conjuntos de dados anotados, o dispêndio de tempo necessário para que se possa fazer anotações nas imagens e a discordância entre patologistas na análise de tecidos. Em que pese haver grande quantidade de estudos envolvendo CNNs em outras áreas, poucos estudos se dedicam, ainda que lateralmente, à detecção de podócitos no contexto da identificação de podocitopatias.

Tendo em vista o que foi acima exposto, estabelecemos como objetivo primário a verificação da viabilidade de se utilizar redes neurais convolucionais na detecção de núcleos, de forma geral, e de podócitos, de forma específica, em imagens histológicas de glomérulos, em especial do segundo, em face da escassez de trabalhos na área. Além disso, busca-se extrair métricas que permitam comparar: a) versões diferentes de uma mesma arquitetura; b) efeito do emprego de transferência de aprendizado (*transfer learning*); c) efeito do emprego de técnicas de *data augmentation*.

1.4 Apresentação do manuscrito

Este manuscrito consiste dos seguintes capítulos, a saber: No Capítulo 2, são apresentados os principais trabalhos relacionados sobre ao tema de análise de imagens histológicas, passando sobre segmentação de núcleo de células, e finalmente falando sobre segmentação automática de podócitos. Já no Capítulo 3 é apresentada a metodologia empregada neste trabalho para se alcançar os objetivos propostos, tento os resultados desta metodologia proposta apresentadas no Capítulo 4. Por último são discutidos os principais avanços e conclusões, incluindo os trabalhos futuros no Capítulo 5.

Capítulo 2

Trabalhos Relacionados

Nas duas últimas décadas, diversos trabalhos têm sido publicados relatando o desenvolvimento de técnicas automáticas na área de Patologia Digital. Em especial, os métodos de digitalização de imagens tornaram técnicas de Aprendizagem de Máquina (*Machine Learning* - ML) e Aprendizagem Profunda (*Deep Learning* - DL) adequadas aos propósitos da área. Tendo em vista os métodos utilizados neste trabalho, decidimos apresentar, neste capítulo, os trabalhos relacionados a ele divididos em três eixos: Análise de imagens histológicas, subdividido em três sub-eixos, Técnicas automáticas de segmentação de núcleos de células e Técnicas automáticas de segmentação de podócitos.

2.1 Eixo 1: Análise de imagens histológicas

2.1.1 Análises quantitativas em imagens cerebrais

Por meio do emprego de anticorpos com auxílio de técnicas imunohistoquímicas, é possível se fazer a detecção de proteínas expressas em células nervosas. Nesse sentido, pode-se relacionar uma dada função neuronal ao nível de expressão de uma dada proteína, que por sua vez impacta o número de células coradas durante a detecção. O trabalho apresentado por [14] descreve um método de contagem automática de células, ACCM, que permite a quantificação das células e, portanto, estabelecer o efeito sobre a função neuronal de estímulos induzidos. O estudo utiliza seções de córtex de ratos - separados nos grupos de controle e de teste de estímulos - com diversos marcadores: parvalbumina, GABA, c-Fos, Nissl e Neun. Após o tratamento químico do tecido, as seções foram reagidas com os anticorpos primários para cada marcador específico. Em seguida, as seções foram divididas em áreas, das quais se tiraram fotografias. A técnica utilizada para o *clustering*, isto é, a formação dos aglomerados de *pixels* que representam o objeto de interesse se dá por meio da comparação de cada *pixel* da imagem com a 8-vizinhança adjacente. O algoritmo consiste em duas passadas: de frente para trás e de trás para frente, ao cabo da qual os *pixels* de um mesmo aglomerado são rotulados com um índice, que é único para cada aglomerado. Todos os outros *pixels* que não pertencem a algum aglomerado são rotulados com -1. Previamente ao algoritmo, utiliza-se uma limiarização para identificar as regiões de interesse. Isto é feito definindo-

se um intervalo de intensidade de brilho na qual se espera que as regiões coradas se encontrem. Posteriormente ao algoritmo, definem-se limites inferior e superior para o tamanho do aglomerado, de modo a filtrar resultados como as bordas do tecido e perturbações. Por fim, o estudo compara os resultados obtidos pelo método proposto com os obtidos por duas outras formas: contagem manual e um programa comercial de análise de imagens.

Ainda em [14], os resultados identificaram um coeficiente de regressão de 0.96 entre a contagem manual e a automática. Os dados incluem diferentes marcadores, diferentes magnitudes de imagem e contagem feita por profissionais diferentes (N=18). Em relação à contagem com o programa comercial, observou-se coeficiente de regressão de 0.92. Nesse sentido, método proposto se mostrou superior para imagens com baixas densidades de células. O estudo revela que o método foi capaz de gerar resultados em apenas dois dias, em contraste com os quatro meses que levariam um histologista a contar manualmente as células. Tendo em vista a possibilidade de ajustar os limiares de nível de brilho, os autores propõem que o ACCM possa ser estendido a outros tipos de detecção, como a de vasos sanguíneos. As possibilidades de exploração da aplicabilidade do método proposto incluem a identificação de subpopulação neuronal e a diferenciação entre células apoptóticas e necróticas.

O trabalho de [15] apresenta uma análise quantitativa de microcolunas em cérebros de macaco. Os autores propõem um método semi-automático para localização dos neurônios na imagem digital. A localização dos neurônios se dá por uma série de processamentos de imagem: transformada *Watershed*, uma marcação intermediária de possíveis candidatos a neurônio, seguido de uma seleção final. Em seguida, utilizam-se do método de mapa de densidade proposto por [16] e extraem os parâmetros microcolunares (comprimento, largura, distância entre duas microcolunas, por exemplo) a partir de uma dada região de interesse. A validação do método se deu por meio da comparação com as marcações de neurônios em seis imagens. Enquanto as marcações manuais indicaram, na média por imagem, 124 neurônios, e o método proposto, 136, dos quais 106 corretamente identificados como neurônios. Apesar do esforço de caracterização quantitativa das microcolunas, o método é de díficil validação, tendo em vista que parte do processo é feito de forma não-automática, o que se reflete no número de imagens utilizadas na validação (n=6). O trabalho de [17] apresenta um método automático de localização de neurônios, denominado ANRA, que se baseia em uma abordagem diferenta da de microcolunaridade. O método proposto utiliza técnicas em aprendizado de máquina. Os autores apresentam um método de segmentação, OSM, que recebe como entradas as imagens de treino e suas respectivas anotações por um especialista, e retorna dois parâmetros primários, então aplicados ao conjunto de teste definido. O ANRA contém ainda um passo para a remoção de elementos não-neuronais, que leva à segmentação final e por último à localização dos neurônios. Classificadores como kNN, SVM, Bayes e árvores de decisão são comparados ao avaliar um dado vetor de propriedades como pertencendo à classe neurônio/não neurônio e o melhor deles (*Multilayer Perceptron*) foi escolhido como o método principal de treinamento para o ANRA. Para fins de comparação, os autores utilizam o método semi-automático proposto por [15]. Em média, o ANRA obteve maior taxa de verdadeiros positivos (86 contra 80) e menor taxa de falsos positivos (15 contra 17). Os resultados atestam a possibilidade do emprego do ANRA na localização de neurônios a partir de imagens de tecido

cerebral.

2.1.2 Classificação de câncer

Para além de análises quantitativas, estudos em classificação de patologias com base em imagens histológicas têm sido realizados ao longo dos últimos anos. Em 2021, segundo o Instituto Nacional do Câncer dos Estados Unidos, o câncer de próstata é o mais prevalente entre os homens. Trabalhos recentes têm se debruçado sobre a tarefa de classificar o nível de agressividade desse tipo de tumor por meio de técnicas automáticas. Classificação histológica é utlizada para a quantificação do nível de agressividade de câncer. Para a próstata, o método mais difundido de classificação histológica é a graduação de Gleason (*Gleason grading system*). O sistema de pontuação tem cinco escalas, de 1 a 5, ordenadas de forma crescente em nível de agressividade. O trabalho de [18], visa a automatização da graduação de Gleason. Os autores propõem um classificador kNN a partir de atributos de energia e entropia extraídos dos coeficientes *multiwavelets* das imagens. A estimação da taxa de erro se dá por meio do método de validação *leave one out* (LOO). A escolha por *multiwavelets* é justificada pelo emprego com sucesso na análise de textura, já feita à época. Foram utilizados dez diferentes tipos de de *multiwavelets*, além de dois pacotes de *wavelets* (*wavelet packets*) e matriz de co-ocorrência como formas de extração de atributos. Os resultados observados indicam que, na comparação entre as dez formas de *multiwavelets*, a SA4 se mostrou a superior em termos do percentual de classificação correta (CCP). A análise foi refinada com a introdução de duas alterações: uso de vetores de peso e ruído adicional. Ao cabo, atingiu-se 0.97 de CCP utilizando a *wavelet* SA4, 0.90 utilizando pacotes de *wavelet* e 0.84 utilizando matriz de co-ocorrêcia.

O trabalho desenvolvido por [19] apresenta uma análise de atributos de imagem utilizados para o diagnóstico de câncer e classificação em graduação de Gleason. Tanto a nível de imagem como a nível histológico, são utilizados três tipos de atributo, quais sejam: cor, textura e atributos morfométricos. Além disso, são testados três tipos de classificadores: gaussiano, kNN e *Support Vector Machine* (SVM). Com respeito aos atributos derivados dos histogramas do canais de cor, os autores encontraram um ganho de performance na classificação com a remoção de *pixels* brancos, resultado de uma transformação de espaço de cores seguida de uma limiarização. O impacto desses *pixels* sobre a classificação se deve ao fato de que eles tanto podem representar lumens, como elementos de fundo não relevantes ao processo. Ao todo, foram selecionados 48 atributos relacionados aos histogramas. Os tipos de atributos de textura são três: dimensão de fractais, código de fractais (estatísticas dos parâmetros deles extraídos), e transformada *wavelets*. Ao todo, 157 atributos. Por último, 424 atributos são extraídos a partir do sistema automático de análise MAGIC, entre os quais: a razão de cada canal de cor, seu desvio padrão e o raio de uma dada vizinhança. Para a seleção de atributos, os autores utilizaram dois métodos: validação cruzada para avaliar a acurácia de um subconjunto arbitrário de atributos e o algoritmo *sequential forward search* (SFS). Ainda em [19], primeira avaliação de resultados envolveu a classificação entre tumor e não tumor. Nela, obteve-se a mais alta acurácia com o emprego dos atributos MAGIC em dois classificadores: o gaussiano e kNN, ambos com 0.967 para um IC de 95% e com validação cruzada

(CV) *five-fold*. Para o primeiro classificador, a mediana do número de atributos foi sete, enquanto que a do segundo classificador foi oito. Este valor foi obtido tomando-se a mediana do número de atributos utilizados em cada um das cinco iterações da CV. Com respeito à segunda tarefa, isto é, a classificação na graduação de Gleason, obteve-se acurácia de 0.81 com IC de 95%. O conjunto de atributos proveio do MAGIC e a mediana encontrada foi de 10 atributos. Com vistas de comparar os resultados obtidos com o método de extração de atributos que apresentara a maior acurácia até então, *multiwavelet feature*, os autores utilizaram quatro classificadores a partir dos atributos selecionados por esse método. Nestas condições, atingiu-se acurácia de 0.72, bastante inferior à relatada pela literatura. Os autores ressalvam que a acurácia é menor que a obtida por [18] devido ao fato de utilizarem um conjunto independente do conjunto de treino para realizar a classificação de erro.

O trabalho apresentado por [20] propõe duas formas de extração de atributos de imagens histológicas com base em dimensão de fractais (FD). Os métodos propostos são utilizados em três tipos de classificadores: bayesiano, kNN e SVM. Os autores propõem ainda comparações com outros métodos de extração de atributos consagrados na literatura de então. Tendo em vista a característica irregular do crescimento do tumor, propõe-se utilizar a dimensão de fractal como medida morfométrica desse tipo de estrutura. O primeiro método de extração utilizado foi o *differential box counting* (DBC) e o segundo, proposto pelos próprios autores, *entropy-based fractal dimension estimation* (EBFDE). Desse modo, extraem-se oito diferentes atributos, metade deles obtida por DBC e a outra metade, por EBFDE. A partir de um conjunto de grades (cópias de uma dada imagem em uma escala menor), os autores fazem uma estimações do valor de FD naquela dada configuração. Para a seleção de atributos, os autores procederam com SFFS e *five-fold* CV para cada um dos três classificadores. Feita a seleção, procedeu-se com aplicação de validação cruzada *five-fold* e LOO, de modo a avaliar cada classficador. Ainda em [20], os resultados obtidos indicaram que o maior valor de *correct classification rate* (CCR) obtido pelo método de validação cruzada LOO foi de 0,932 com SVM; para *five-fold*, obteve-se 0.937 com kNN (k=1). Estes resultados se referem ao conjunto de oito atributos baseados em FD, ou seja, sem seleção de atributos (*feature selection*), tarefa feita adiante. Em seguida, os autores fazem comparações dos resultados dos classificadores basedos em FD (com e sem seleção de atributos, para os dois métodos de validação cruzada), em relação a outros conjuntos de atributos, como os filtros de Gabor, *Multiwavelet* e GLCM. Sem seleção de atributos, o método proposto atingiu os maiores valores de CCR quando avaliado o classificador bayesiano: 0.912, tanto para LOO quanto para *five-fold*. Com kNN, o método proposto seguiu com o maior CCR, quando avaliado sem seleção de atributos. Com o uso do algoritmo SFFS para a seleção de atributos, obteve-se CCR de 0.946 com o classificador SVM avaliado por LOO; avaliado por *five-fold*, nas mesmas condições, obteve-se o mesmo valor para o classificador bayesiano. Destaca-se o fato, para além do valores encontrados, de que o número de atributos utilizados (8) é bastante inferior ao de outros métodos comumente utilizados, como *multiwavelets* (56) e Gabor Filters (20).

Um classificador de câncer de próstata que pode ser entendido como um estágio anterior à graduação de Gleason foi desenvolvido por [21]. Sobre o trabalho de [20], os autores argumentam as regiões das imagens foram pré-selecionadas, o que impacta o resultado obtido. Nesse sentido,

propõe-se o desenvolvimento de um classificador com base em imagens de microscopia digital (*whole-slide images*), a partir das quais constroi-se uma pirâmide de imagens contendo diversos níveis de resolução. Os autores escolheram o algoritmo AdaBoost para a extração de atributos, tendo em conta que seja pronunciada a diferença entre as classes benigna e cancerosa em cada nível de resolução. Este algoritmo permite distinguir os atributos relevantes para a classificação, funcionando a partir da combinação de classificadores fracos (*weak classifiers*), isto é, classificadores formados a partir de um único atributo. Dessa forma, extraem-se 927 atributos, cada qual correspondente a uma função densidade de probabilidade (PDF) própria, por meio da qual se utiliza a Regra de Bayes na classificação. O algoritmo proposto é denominado BBMR (*Boosted Bayesian MultiResolution*) e utiliza três níveis de resolução (j=1,2 e 3) para realizar a classificação. Os autores explicam que em resoluções mais baixas há pouco ganho na detecção. Assim sendo, o estudo faz três experimentos: a) a avaliação do classificador BBMR; b) a comparação deste com outros cinco classificadores; c) análise dos parâmetros envolvidos. Ainda em [21], duas métricas foram avaliadas utilizando validação cruzada: acurácia (ACC) e área abaixo da curva ROC (AUC). Duas formas de avaliação foram realizadas: "por imagem", em que cada conjunto da validação é independente e "por paciente", em que se restringe que os grupos da validação cruzada contenham imagens de pacientes diferentes. Os resultados foram comparados a outras combinações de seleção de atributos (*random forests* e melhor atributo) e de estimação da PDF. Em suma, o BBMR obteve resultados superiores. Na base "por imagem", alcançou-se 0.84 de AUC e 0.70 de ACC. Na base "por paciente", 0.85 de AUC e 0.74 de ACC. De modo geral, os classificadores tiveram resultados superiores em níveis de resolução mais baixa. Além disso, constatou-se que atributos com janelas maiores extraídos pela algoritmo AdaBoost apresentaram resultados superiores em todos os níveis de resolução.

O trabalho de [22] faz uso de um sistema de *Deep Learning* (DLS) para classificação seguindo a graduação de Gleason. O estudo busca o desenvolvimento de um sistema que permita lidar com a variabilidade da graduação e melhorar o diagnóstico a partir de imagens de microscopia digital (*whole-slide sections*). Além disso, os autores encaminham uma possibilidade de refinamento da escala, o que permitiria, potencialmente, um prognóstico mais preciso. O estudo consiste em utilizar uma rede neural convolucional em cascata com um classificador kNN de grupos de graduação de Gleason (*Gleason Grade Group*). Foram utilizadas 912 seções (contendo milhões de fragmentos de imagens) para o treinamento do DLS. O DLS foi avaliado em um conjunto de 331 seções independentes, cada qual revisada por três patologistas sob o acompanhamento de um especialista genitourinário. A acurácia média obtida no conjunto de validação pelo DLS foi comparada com a obtida por um grupo de 29 patologistas. Não obstante, dez patologistas revisaram todo o conjunto de validação e, a partir disso, extraiu-se a acurácia individual de cada um deles, também comparada à do DLS.O estudo indicou que, com IC de 95%, os resultados de acurácia do DLS (0.70) foram superiores aos obtidos pelos 29 patologistas. Além disso, o sistema proposto obteve resultados superiores a 8 dos dez patologistas do subgrupo de revisão, que em média obtiveram 0.64 de acurácia, portanto, resultado inferior ao do DLS. Uma segunda avaliação feita envolveu a classificação em nível de região. Nessas condições, foram feitas anotações em 79 seções por três patologistas independentes. O intuito foi avaliar a concordância ou não do sistema proposto em classificar as regiões com as anotações feitas pelos especialistas. Desse modo,

constatou-se concordância de 97%. No subconjunto em que houve classificação em padrão de Gleason, o DLS concordou em 88% das vezes sobre qual seria o padrão. O estudo faz ainda uma comparação entre a capacidade do sistema proposto e de especialistas de realizar estratificação de risco dos pacientes. Com a adição de um padrão de Gleason (GP3.5), atingiu-se um *c-index* de 0.704 na avaliação prognóstica por meio dos modelos de Cox.

2.1.3 Segmentação de glomérulos

Ao primeiro eixo, também se relacionam tarefas como a segmentação de glomérulos, esta ainda forma mais direta com o trabalho desenvolvido. A partir da inspeção de lâminas histológicas, patologistas podem empregar técnicas de análise para prover um diagnóstico. Um exemplo disso é a identificação de glomérulos e a quantificação de quantos deles são normais ou anormais. Nesse sentido, técnicas de aprendizado de máquina têm sido desenvolvidas e testadas para estabelecer essas relações e, em última instância, contribuir para um diagnóstico de qualidade. O trabalho de [23] realiza a segmentação de glomérulos utilizando técnicas modernas de ML. Especificamente, empregam-se técnicas de aprendizado profundo: as redes neurais convolucionais (CNNs) são utilizadas para a identificação das estruturas em questão nas seções de tecido renal. O estudo inclui uma avaliação a respeito das vantagens e desavantagens do emprego dessa técnica em específico, bem como um levantamento dos resultados presentes na literatura para diversas tarefas, como segmentação de núcleos e tumores cerebrais, detecção de mitose e linfócitos e classificação de linfomas. Os autores propõem métodos para cumprir duas tarefas: classificação glomérulo/ não glomérulo e identificação de glomérulos em *whole slide images* (WSI).

Ainda em [23], O treinamento das redes consistiu em quatro combinações envolvendo duas arquiteturas de CNN distintas e já descritas anteriormente (AlexNet e GoogleNet) e a presença ou não de pré-treino. Além disso, utilizou-se *data augmentation*, em que se criou cópias das imagens por meio de rotações de 0, 90, 180 e 270 graus, *flip* vertical e modificações de cor. A validação do experimento se deu por meio de 10-*fold* CV. Além disso, tendo em conta a variabilidade dos corantes presentes nas amostras, aplica-se normalização de cor ao conjunto de imagens por meio do método de Reinhard, o que é feito a partir de dez blocos de referência. A escolha do número de blocos foi justificada pela variabilidade de cor do conjunto de dados. Os resultados de classificação indicaram que as duas configurações com redes pré-treinadas foram superiores. Em ambas arquiteturas, atingiu-se F1-score de 0.999. Houve cinco amostras de glomérulo classificadas como não glomérulo, embora não tenha havido a ocorrência contrária. Uma análise posterior de patologistas sobre as classificações incorretas revelou que eram amostras de difícil avaliação. Além disso, quatro das cinco eram relacionadas entre si, no sentido de terem sido fruto de data augmentation. Em relação à segunda tarefa, observou-se F1, precisão e *recall* de, respectivamente, 0,94, 0,88 e 1. O conjunto de teste incluia dez WSI, com 275 glomérulos ao total. Outrossim, os resultados obtidos se mostraram superiores em comparação aos que se obteve sem modificação de cor (F1=0.885), o que se pode atribuir à variabilidade dos corantes. Em suma, mostrou-se que a utilização de redes pré-treinadas e normalização de cor pode levar a resultados superiores na detecção de glomérulos com CNNs em WSI. No entanto, a diminuição dos falsos positivos deve

ser perseguida, tendo em vista a disparidade entre os valores de *recall*, mais alto, e de precisão, mais baixo. Outra possibilidade de melhora é que a divisão da imagens em blocos possa ser feito de maneira adaptativa, isto é, sem escolha de um valor fixo para o tamanho do bloco.

O trabalho apresentado por [24] apresenta um algoritmo de segmentação de glomérulos e de classificação de glomerulopatias utilizando técnicas modernas de ML. O localizador e o classificador utilizados se baseiam em arquiteturas de CNNs já conhecidas (AlexNet e Faster RCNN). Para além do treinamento das redes neurais, procedimentos como a extração dos tecidos de rim de ratos, aquisição de imagens, deconvolução de cor para a separação de corantes e anotações de imagens (*ground-truth boxes*.O conjunto de dados completo inclui mais de 28000 glomérulos manualmente anotados a partir de dezenas de amostras de tecido renal. O algoritmo foi testado tanto em seções de rato (n=13) quanto de origem humana (n=6), em que cada seção continha diversos glomérulos. O trabalho não envolveu somente a classificação entre glomérulo e imagem de fundo, mas também a localização das estruturas em seções renais inteiras. Na primeira etapa, utilizou-se a RCNN para atribuir as duas classes (glomérulo ou imagem de fundo) a possíveis candidatos com base em suas respectivas probabilidades, além de fazer a marcação dos contornos de cada um deles. Em seguida, a rede foi novamente treinada para que se pudesse selecionar candidatos das duas classes e então compor um segundo conjunto de dados, contendo 22 mil amostras de cada classe. Ainda em [24], o treinamento inicial forneceu acurácia de treino de 0.92 em média. Com o retreino, tendo em vista a seleção de objetos com maior classificação equivocada, atingiu-se 0.99 de acurácia de treino. No conjunto de teste, atingiu-se precisão de 0.97 e *recall* de 0.97 nas amostra de rato; nas de origem humana, 0.80 e 0.81, respectivamente. A análise contou ainda com uma avaliação do nível de enviesamento das amostras de teste em termos da graduação de cada lesão glomerular. Realizou-se a contagem de falsos negativos de um subconjunto de teste agrupado pela graduação das lesões - cujo valor vai de 0 a 4. O teste de Fisher realizado (P=046) não indicou viés de erro em nenhum valor da escala. Em suma, o trabalho apresenta bons resultados a partir de uma quantidade de amostras expressiva. Com efeito, os autores relatam que o conjunto de glomérulos utilizado havia sido o maior até então. Além disto, destaca-se o fato de que os bons resultados atingidos nas amostras humanas provieram de uma rede treinada exclusivamente com glomérulos de ratos. Uma possível extensão ao trabalho desenvolvido seria a classificação das lesões em conjunto com a localização dos glomérulos, algo que provavelmente requeria um conjunto de dados ainda mais rico.

No trabalho desenvolvido por [25], os autores propõem um método de segmentação de estruturas de tecidos renal a partir que permitam uma análise histopatológica automática. Ele se vale do uso de CNNs e se dedica a segmentar nove estruturas, entre as quais: glomérulos, túbulos, arteríolas, capilares e tecido fibrótico. Nota-se que a estruturas referidas englobam tanto tecidos saudáveis quanto patológicos. Três arquiteturas distintas são utilizadas e a validação cruzada é feita por 4-*fold*. Os autores afirmam que, até então, o estudo apresentado foi o primeiro sobre segmentação em histopatologia renal. Para cada lâmina, foram anotados glomérulos (n=875), túbulos proximais (n=853), túbulos distais (n=1118), arteríolas (n=155), capilares (69), túbulos atróficos (n=469), infiltrado inflamatório (n=67) e tecido fibrótico (n=244). Duas arquiteturas foram criadas, inicialmente, consistindo somente em redes totalmente convolucionais (*fully convo-*

lutional networks: uma *single scale output* (FCN) e outra *multiple scale output* (M-FCN). Desse modo, formou-se as três arquiteturas utilizadas: FCN + U-net (associação 1), M-FCN + U-net (associação 2) and FCN + M-FCN + U-net (associação 3). Utilizou-se diversas técnicas de *data augmentation*, tais como: rotação, espelhamento e aplicação de filtros Gaussianos. Ainda em [25], entre as três associações construídas, a terceira atingiu a maior acurácia de *pixel* para segmentar oito entre as nove estruturas. Além disso, com ela obteve-se acurácia variando de 0,84 a 0,97 nas classes patológicas e de 0,62 a 0,94 nas saudáveis. Deve-se notar que houve menor acurácia nas arteríolas (0,71) e capilares (0,62). Observa-se que existem poucos exemplares da primeira amostra em comparação com as outras classes, o que não ocorre com a segunda. De fato, os capilares são a estrutura com o quarto maior número de anotações. Os autores justificam a queda de desempenho por possíveis variâncias morfológicas e pelo tamanho destas estruturas. De todo modo, o estudo seria enriquecido com um número maior de instâncias de cada estrutura, tendo em vista que somente a classe dos glomérulos escleróticos foi anotada em todas as ocorrências de cada lâmina. Portanto, uma expansão do conjunto de dados poderia trazer maior segurança às conclusões a respeito da capacidade de generalização da rede.

O trabalho publicado por [26] propõe um método de segmentação e classificação de glomérulos utilizando CNNs. À diferença de outros trabalhos, os autores utilizam WSI de seções congeladas (sem fixação). O tecido é corado com hematoxilina-eosina (H&E). Dois modelos de rede são empregados na identificação e classificação entre glomérulo esclerótico/não esclerótico: a primeira, baseada em seção, no sentido de que as imagens de entrada são recortes de WSI com glomérulos isolados; a segunda utiliza uma rede totalmente conectada e recebe como entrada as WSI inteiras. Um dos objetivos do estudo é trazer uma comparação entre os dois métodos.No modelo baseado em seção, adotou-se uma estratégia de aumento do conjunto de treino a partir de inversões aleatórias, rotações em 90º, e translações de até 5%. A arquitetura utilizada foi a VGG16 pré-treinada, com pesos congelados nas camadas inferiores às três últimas. A estratégia de validação cruzada utilizada foi a *6-fold*. As últimas camadas da arquitetura original foram alteradas para duas camadas totalmente conectadas de 32 nós com ativação ReLu seguidas de uma única camada *softmax* de 3 nós.O modelo totalmente convolucional contou com a mesma estrutura de camadas inferiores: também se utilizou uma rede VGG16 pré-treinada. As camadas superiores contaram com uma sequência de camadas convolucionais (1x1, 3x3 e 5x5) e uma camada convolucional dilatada, em substituição às camadas finais totalmente convolucionais da arquitetura original. A estratégia de validação cruzada foi a mesma do experimento anterior.

Ainda em [26], para a rede baseada em seção obteve-se precisão, *recall* e F1 na classificação glomérulo esclerótico/não esclerótico de, respectivamente: 0,893/0,932, 0,865/0,962 e 0,879/0,947. Em termos de matriz de confusão, houve maior taxa de falsos positivos na identificação de glomérulos escleróticos. Os resultados de classificação das duas redes - entre glomérulo esclerosado, não esclerosado e região túbulo-intersticial - foram comparadas *pixel* a *pixel* com as anotações de um patologista. Houve uma tendência maior de sobrestimar a área de regiões de glomérulos não esclerosados pela rede baseada em seções, seis vezes maior que a área anotada pelo especialista; em contraste, a rede baseada em WSIs obteve sobrestimação de 1,7 vez. Houve, ainda, maior concordância entre a rede totalmente convolucional e as anotações do patologista, medida que se

fez com base no índice de Jaccard. Além disso, observou-se maior correlação entre a taxa de glomérulos esclerosados para a rede completamente convolucional ($R^2 = 0,828$) em relação ao obtido para a rede baseada em seção ($R^2 = -0,491$), ambos medidos com respeito às anotações do patologista. Para além dos procedimentos a nível de *pixel*, utilizou-se um algoritmo para identificar as regiões como um todo (*Laplacian-of-Gaussian (LoG) blob-detection algorithm*) a partir do mapa de probabilidades das predições. Com isto, obteve-se precisão e *recall* de 0,813/0,607 e 0,885/0,698 para glomérulos não esclersoados/esclerosados respectivamente. O trabalho têm o mérito de não depender do processo de fixação das amostras do tecido e, além disso, de trazer o comparativo da performance de CNN em WSIs e em seções destas. Observa-se também que a transferência de aprendizado se deu com sucesso, o que permitiu que a rede não precisasse de um grande conjunto de treino para obter bons resultados. Apesar disso, poderiam ser feitas alterações nos parâmetros de detecção a partir do mapa de probabilidades, uma vez que foram arbitrariamente definidos, de modo que as detecções de glomérulos adjacentes feitas pela rede possam ser diferenciadas pelo algoritmo LoG.

O fluxo de trabalho proposto por [27] contribuiu para a análise quantitativa de glomérulos e glomeruloscleroses a partir de lâminas de biópsia de rim. Os autores propõem um método de identificação e caracterização de glomérulo que, em conjunto com outros dados, venham a permitir o diagnóstico de doença glomerular. Para tanto, o estudo avalia o uso de técnicas de aprendizado profundo, as redes neurais convolucionais (CNN), para identificar glomérulos, a partir de imagens de biópsia do rim, além de segmentá-los entre globalmente esclerosados (GS) ou glomérulos normais/parcialmente esclerosados (NPS). A partir de biópsias de 171 pacientes, o estudo fez uso de 275 imagens, a partir das quais foram geradas outras: 745 com glomérulos e 751 sem. Utilizou-se uma arquitetura de rede já existente, a Inception v3, já pré-treinada. A rede foi treinada com as imagens tricrômicas recortadas. Alterou-se a última camada da rede, de modo que ele pudesse classificar cada imagem em uma entre três possibilidades: sem glomérulos, glomérulos NPS ou glomérulos GS. A divisão treino/teste seguiu uma razão 7:3 e avaliação foi realizada quatro vezes em conjuntos definidos aleatoriamente, para que se pudesse ter confiabilidade nos resultados. Empregou-se ainda a técnica de *data augmentation*, por meio da qual criou-se cinco cópias de cada imagem de treino - resultando em seis imagens para cada imagem recortada. O processo de expansão do conjunto de treino consistiu em adicionar *pixels* brancos aleatoriamente à imagem. A partir das identificações da rede, gerou-se um mapa de calor para a identificação de glomérulos GS, passo sucedido por uma limiarização, segmentação textitwatershed para identificação glomerular.Na tarefa de classificação nas três categorias, os resultados mostraram que o emprego das técnicas de branqueamento e de *data augmentation* alcançaram resultados superiores. Com as quatro combinações de conjuntos de treino/teste, atingiu-se valores de acurácia variando entre 0.897 e 0.951. Em relação à segmentação de glomérulos GS, que envolveu a classificação prévia da rede seguida de operações de processamento de imagens, atingiu-se um valor de F1-score de 0.623. O estudo têm importância ao constatar a efetividade do uso de redes pré-treinadas e de *data augmentation*, mas prescinde de aprimoramento para poder segmentar glomérulos normais ou parcialmente esclerosados, o que não foi feito no presente trabalho.

O trabalho desenvolvido por [28] apresenta um método de segmentação no contexto do desen-

volvimento de um *pipeline* de classificação de imagens histológicas de pacientes com nefropatia diabética (DN). A sequência de operações proposta inclui a identificação dos glomérulos, a identificação e quantização de estruturas a eles internas, a quantização destas e classificação de *features*. O diagnostico é feito por meio de uma rede neural recorrente (RNN), cuja saída é um valor contínuo, que é então discretizado para se adequar à escala utilizada (de 1 a 5).As amostras das WSI utilizadas incluem tecido de origem humana (n=54) e de rato (n=25). A detecção e posterior segmentação dos contornos glomerulares é feita com a rede ResNet. De modo a simplificar a análise, adotou-se uma divisão interna aos glomérulos em três componentes: núcleos; lúmens capilares e espaço de Bowman; componentes PAS(+), sendo este último um grupo que engloba outras estruturas. O conjunto de imagens utilizado na detecção de núcleos continha tanto glomérulos anotados automaticamente (n=400), anotados automaticamente (n=216) e imagens com grandes seções inflamadas, de modo a fornecer exemplos mais variados a aprendizagem. Do segundo conjunto, 10% das imagens foram separadas para validação. A detecção de glomérulos é avaliada por meio das métricas de sensibilidade e especificade, as quais são fruto da concordância ou não da marcação da rede com as anotações. Ainda em [28], como se trata de um conjunto encadeado de operações (*pipeline*), os resultados são apresentados de forma sequencial. Para a identificação dos conteornos dos glomérulos, atingiu-se 0,93 de acurácia. A avaliação da detecção de núcleos pôde ser feita por meia da *receiver operator curve* (ROC). Otimizando-se a perfomance pelo ajuste da ponderação das probabilidades de saída da rede, com o qual se obteve sensibilidade/especificidade 0.94/0.93 no conjunto de validação com glomérulos (n=22). A detecção de núcleos em conjunto com a detecção dos outros componentes (componentes PAS+ e capilares e espcaço de Bowman) gerou mapas glomerulares. Assim, atingiu-se sens/spec 0.98/0.99 (PAS+) e 0.98/0.99 (luminal) para 123 glomérulos anotados. A classificação de nefropatia diabética foi realizada por meio de uma RNN; atingiu-se um valor de k de Cohen (0.55), intermediário aos obtidos pelos dois outros patologistas. Computou-se estatísticas de concordância por classe, proporção de concordância por classe e probabilidade condicional de atribuir uma classe dado um certo *ground-truth*. Os dois patologistas tenderam a subestimar os contornos da região de interesse em relação ao *ground-truth*, o que não ocorreu com a rede neural. Os autores explicam que isto se deve a uma tendência de os especialistas preferirem errar "para baixo"em caso de dúvida. Por fim, uma analise de quais features seriam mais importantes para a classificação indicou as seguintes: desvio padrão dos valores vermelhos nas regiões PAS+, valor médio dos valores nucleares azuis e o desvio padrão dos valores azuis em regiões PAS+. Em que pese a longa cadeia de procedimentos descritos, há ainda possibilidades de refinamento, como estender a distinção dos componentes glomerulares a mais de três. Além disso, deve-se notar que algumas estatísticas por classe foram obtidos a partir de poucas amostras (na classe IIa, por exemplo, tem-se n=2), o que comprometeu os resultados.

O trabalho apresentado por [29] apresenta uma proposta de segmentação multiclasse em amostras de nefrectomia e biópsias de transplante utilizando CNNs. O corante utilizado é o PAS. O estudo tem como objetivo prover uma análise histológica em tecidos renais corados com PAS a partir de uma CNN. Para tanto, propõe-se que seja possível a segmentação do tecido em classes (glomérulo, túbulo, interstício), independentemente de a amostra ser saudável ou não. Para checar a robustez do modelo desenvolvido, testam-se amostras de um laboratório diferente do que se originaram as amostras de treino, de modo que se possa fazer uma avaliação sobre o impacto

da variabilidade da coloração nos resultados obtidos. Do primeiro centro médico (Radboudumc), extraíram-se, de 101 pacientes, 132 amostras, das quais 40 foram utilizadas para treinamento e 10 para teste. As 82 lâminas restantes foram utilizadas para validação pelo sistema Banff de classificação. Do segundo centro (Mayo), dez amostras foram utilizadas como validação de CNN treinada em Radboudumc. A arquitetura escolhida para a tarefa de segmentação foi a U-net. Cinco conjuntos foram criados para validação cruzada, consistindo cada qual em 37 WSI de treino e 3 de teste. Expandiu-se o conjunto de dados (*data augmentation*) por meio do emprego de técnicas espaciais e de cor. Para a otimização da taxa de aprendizado, foi utilizado o algoritmo Adam. A métrica de avaliação usada é o coeficiente de Dice (DC), que mensura a interseção espacial entre o resultado da segmentação e o *ground-truth*. Ainda em [29], no conjunto de teste Radboudumc, observou-se os menores valores de DC nas classes túbulo atrófico (0,48) e túbulo indefinido (0,32). Não obstante, no geral, a estrutura túbulo como um todo - englobando as classes subjacentes, atingiu DC de 0,93. Os glomérulos foram as estruturas com melhor nível de segmentação, DC=0,95. Já no conjunto de teste Mayo, duas estruturas estavam representadas uma única vez (glomérulo GS e cápsula de Bowman vazia). Além disso, não havia cápsulas neste conjunto. Para os glomérulos, atingiu-se DC de 0,94. Os autores fizeram uma análise da capacidade de segmentação e detecção de glomérulos em seções de nefrectomia, às quais correspondiam a 15 WSI adicionais. Estas, por sua vez, continham tanto glomérulos saudáveis (n=1747), quanto globalmente esclerosados (n=72). Nos primeiros, foram classificados corretamente 1632 (92,7%) e houve 149 falsos positivos; nos segundos, respectivamente , 55 (76,4%) e 46. O trabalho ainda apresenta os resultados obtidos com a CNN a partir dos componentes do sistema Banff de classificação. Para a contagem de glomérulos, a CNN atingiu coeficiente de correlação intraclasse (ICC) de 0,93, 0,94 e 0,96 em comparação com três patologistas. Apesar da boa performance relatada, deve-se ressaltar a possibilidade de interpretações dúbias: os túbulos, de forma combinada, foram bem segmentados, entretanto, houve grande queda na performance ao se avaliar os subtipos desta estrutura. Uma perspectiva de encaminhamento futuro é a de detecção de outras estruturas além dos glomérulos.

Um sistema de avaliação de glomeruloscleroses foi apresentado por [30]. Um dos objetivos da pesquisa é que o sistema de diagnóstico assistido por computador (CAD) possa reduzir o descarte equivocado de órgãos para doação causado pela variabilidade na observação do tecido por patologistas. O sistema proposto se baseia em segmentação semântica e detecção dos glomérulos por meio de redes neurais convolucionais. A saber, utilizam-se duas: SegNet e DeepLab V3+. O conjunto de dados consiste em glomérulos escleróticos (n=428) e não escleróticos (n=2344). Em termos da divisão dos conjuntos, são utilizadas 19 WSIs para treinamento e sete para teste. O conjunto de dados é então aumentado no processo de treinamento, sendo aplicadas operações como transformações morfológicas (rotação, redimensionamento) e nos canais de cores (deslocamento HSV). Para que a detecção pudesse ser feita em seguida da segmentação gerada pela rede - que identifica *pixels* de forma individual, sem se ater aos objetos em si - alguns pós-processamentos foram necessários. Em resumo, adaptou-se a segmentação semântica à tarefa de detecção por meio de operações de morfologia matemática seguidas do método *K-means clustering* para isolar as estruturas. Os resultados foram apresentados em duas categorias: os referentes à segmentação semântica, a nível de *pixel*,e os referentes à detecção, a nível de objeto. Com respeito à segmen-

tação, a rede SegNet apresentou acurácia superior nas classes esclerótico e não esclerótico (0,686 e 0,919), e a DeepLab, na classe fundo, 0,997. A nível de objeto, atingiu-se *F-score* de 0,924 na detecção de glomérulos não escleróticos (DeepLab) e 0,859 na de escleróticos (SegNet). Uma possibilidade de trabalho futuro é a adaptação do CAD a uma rede neural especificamente de detecção, como Faster R-CNN ou YOLO, dispensando a etapa de pós-processamento no fluxo de trabalho.

2.2 Eixo 2: Técnicas automáticas de segmentação de núcleo de células

Tendo em vista que a segmentação de núcleos de células em imagens tridimensionais de microscopia confocal é essencial para inúmeros estudos de morfologia de núcleo de célula e análise funcional , neste quesito o trabalho de [31] traz uma combinação de duas novas abordagens para realizar a segmentação de núcleo das células de um pedaço do hipocampo de ratos. Essas duas abordagens são, uma transformação de distância que considera tanto atributos geométricos de distância quanto de intensidade de gradientes chamada de transformada de distâncias ponderada por gradiente(*gradient-weighted distance transform*) e um método baseado em modelo matemático de características de um núcleo de célula obtidas automaticamente pelos dados, com o objetivo de realizar a fusão ou quebra de objetos após a super-segmentação gerada pelos processos anteriores ao pós-processamento. Primeiramente, ocorre um pré-processamento nas imagens para ajustar a intensidade de luz do primeiro plano das imagens e retirar ruídos utilizando processamento de imagem morfológica, depois se usa a transformada de distâncias ponderada por gradiente para que os objetos e o fundo da imagem sejam marcados por uma região mínima juntamente com os contornos dos objetos. Após a transformada, se utiliza a segmentação tridimensional *watershed* aprimorada com um pós-processamento para realizar o contorno final nos objetos. O pós-processamento se refere ao método baseado em modelo para a fusão e quebra de objetos. Este método serve para eliminar a super-segmentação no resultado final. Por causa da variedade do formato e intensidade dos núcleos das células nas imagens tridimensionais e também a grande quantidade de ruídos, o pré-processamento foi extremamente importante para a redução da super-segmentação. A combinação dos dois métodos também elimina quase todos os núcleos super-segmentados durante o pós processamento e assim se conseguiu uma média de 97% de concordância com o observador humano em identificar os núcleos. Há um esforço em melhorar ainda mais a acurácia na segmentação de núcleos, com o objetivo da análise FISH (*Fluorescence in situ hybridization*), que é muito importante para a pesquisa na área da neurociência.

A segmentação automática permite o estudo individual do núcleo da célula dentro do seu ambiente natural no tecido. O trabalho publicado por [32] segmenta os núcleos celulares em seções de tecido em imagens feitas por microscopia de fluorescência. Nessas imagens há três problemas na segmentação: a intensidade do plano de fundo da imagem em relação ao primeiro plano, grande variação de intensidade dos núcleos que gera super-segmentação e os núcleos costumam estar aglomerados dificultando a separação individual do núcleo. Realizou-se um pequeno pré-processamento, uma segmentação *watershed* semeada (*seeded watershed segmentation*) e então

um pós-processamento para quebrar ou unir objetos. Há dois tipos de imagens sendo segmentadas , bidimensional e tridimensional , porém a metodologia para a segmentação é a mesma para ambas. O pré-processamento realizado foi um filtro gaussiano 3x3, porém em imagens tridimensionais foi realizado uma compensação para a atenuação da luz antes do filtro. Depois é feita a semeadura(*seeding*) para marcar o primeiro plano da imagem usando a transformação h-máxima estendida na imagem original e a transformação h-minima estendida na imagem gradiente magnitude para o plano de fundo. Após a semeadura, é feita a segmentação *watershed* semeada porém é necessária a união realizada baseada em força da borda dos objetos como também a quebra usando a transformação à distância, assim reduzindo a super-segmentação. Foi obtido um resultado de 90% de segmentações corretas comparando com a contagem manual nas mesmas imagens bidimensionais e tridimensionais. Percebeu-se que o pré-processamento foi pequeno e necessário. Por outro lado, processamento junto com o pós processamento podem melhorar o resultado se forem adicionados: métodos automáticos para a aproximação de parâmetros, um método para retirar núcleos danificados ou núcleos sub-segmentados ou métodos estatísticos mais avançados. Foi observado que realizar uma segmentação manual após a segmentação automática garante um resultado de 100% em um pequeno período de tempo.

O objetivo do trabalho publicado por [33] foi o de desenvolver um método totalmente automático de segmentação de núcleos em imagens microscópicas bidimensionais, consertando principalmente o problema de segmentação em núcleos que estão se tocando ou se sobrepondo. Por conta disso, foi necessário um desenvolvimento em técnicas automáticas de mais confiança em análise de imagens celulares em biologia molecular computacional. Foi utilizado um novo método para achar marcadores de forma e uma nova função de marcação para usar na segmentação tipo *watershed* e utilizada em imagens celulares de microscopia de fluorescência de células neuronais de ratos e de drosófila para segmentar seus núcleos. O método começa com uma segmentação de núcleos inicial usando contornos ativos sem pontas. Posteriormente, a segmentação foi refinada com operações morfológicas como a abertura com o elemento estruturante de disco para retirar pequenos objetos que são improváveis de ser partes reais de núcleos. Depois, é feita a transformação de distância interna e a transformação h-minima para se ter os marcadores que são utilizados para o algoritmo *watershed* controlada por marcadores(*marker-controlled watersheds*). Além dos marcadores é necessária uma função de marcação para realizar a segmentação por *watershed*, logo foi criada um novo modelo de função de marcação chamada de transformação de distância externa(*outer distance transform*) onde cada pixel do plano de fundo tem um valor correspondente a distância mínima dos marcadores.

O método de segmentação presente no trabalho de [33] foi comparado com outras técnicas mais antigas de segmentação como o clássico *watershed* e erosão condicional. Com isso foi obtido um valor de 6% a 7% de melhora na acurácia de segmentação obtendo 97.39% em células neurais de ratos e 96.30% em células neurais de drosophila. Foi observado um alto número de super-segmentação pelo *watershed,* enquanto que na erosão condicional o problema existe no tamanho da estrutura de erosão e na procura de um limiar (*threshold*) que evita tanto a super-segmentação quanto a sub-segmentação. O método utilizado obteve nos resultados núcleos mais robustos a ruídos e linhas de *watershed* mais suaves. Um problema do método é que ele funciona com

o pressuposto que, existe uma correspondência um por um dos marcadores e dos objetos, isso significa que se o tamanho e formato dos núcleos se diferirem dentro de um aglomerado, o algoritmo pode falhar em segmentar.

No trabalho desenvolvido por [34], foi apresentado um novo método para segmentação de núcleos celulares usando uma combinação de ideias para melhorar a acurácia, velocidade, nível de automação e adaptabilidade. O grande problema em segmentação de núcleos na Histologia é o fato de que a imagem é uma seção bidimensional de um tecido tridimensional resultando em núcleos parcialmente danificados por seccionar em ângulos diferentes e dano por causa do processo de seccionamento. Outros tipos de problemas são os núcleos densamente aglomerados, ruídos nas regiões do plano de fundo, principalmente em dados de fluorescência e a presença de erros espectrais de separação. Logo, foi apresentado um método para superar esse problemas ditos anteriormente como também problemas nos processos clássicos de detecção dos marcadores (*markers*) ou sementes (*seeds*). Primeiramente, é necessária a extração do primeiro plano da imagem de forma automática e para isso foi usada uma abordagem hibrida de um binarização inicial com um refinamento usando algoritmo *graph-cut*. Depois, foi usado o novo método que é a combinação do filtro Laplaciano do Gaussiano (*Laplacian-of-Gaussian (LoG)*) com uma seleção de escala automática e adaptativa. Algumas vantagens deste método é a eficiência computacional, habilidade de explorar tamanho e intensidade , fácil implementação, habilidade para especificar aproximadamente os tamanhos dos núcleos esperados e a robustez às variações. Então foi feita a segmentação por um método baseado em aglomerado de tamanho restrito (*size-constrained clustering*) que, ao contrário do método *watershed*, é capaz de evitar pequenas aglomerações e funciona apenas nos pontos do primeiro plano, sendo assim mais rápida. Por fim, realizou-se um refinamento usando um segundo algoritmo baseado em *graph-cuts* incorporando os métodos expansão α (α - *Expansions*) e gráfico de coloração (*Graph Coloring*). Também podem ser necessárias algumas interações humanas para consertar erros de segmentação para se ter uma acurácia maior. Considerando quatro tipos diferentes de erros de segmentação, o algoritmo proposto excede em 86% de acurácia e se se considerar apenas dois tipos, super-segmentação e sub-segmentação, se consegue um valor acima de 94% de acurácia. Os principais problemas continuam sendo a alta densidade de núcleos e aglomerados de núcleos, contraste pobre de imagem, ruídos no plano de fundo, núcleos danificados e informação pobre sobre a borda. Porém os resultados mostram que o algoritmo é extremamente robusto e preciso e que quando ocorre erros, o método de edição das sementes (*seeds*) seguido pelo refinamento da segmentação melhoram o resultado mesmo com o mínimo esforço de um observador humano.

Segmentação de células e núcleos é um primeiro passo importante para a análise automática de imagens microscópicas digitalizadas. Com a introdução dos *scanners* digitais de lâminas rápidos, ocorreu um renascimento no interesse de aplicações de análise de imagens na Patologia. Com isso em mente, no trabalho publicado por [35], desenvolveu-se uma técnica *watershed* controlada por marcadores para a segmentação automática de núcleos de células cancerígenas em imagens histopatológicas de câncer de mama em lâminas com coloração Hematoxilina-Eosina (*H&E stain*). A análise automática em imagens com coloração $H\&E$ é algo bem complicado por causa da complexidade e diversidade da aparência do tecido. A segmentação manual de todos os núcleos é algo inviável, logo foi feita uma abordagem de amostragem aleatória para auxiliar na segmentação

manual do valor de referência (*ground-truth*). O processo de segmentação começa com o pré-processamento para remover conteúdos irrelevantes preservando a borda dos núcleos. Utilizou-se separação de cor (*color unmixing*) e diversas operações morfológicas. A segunda parte do processo é a segmentação *watershed* controlada por marcadores extraídos usando transformação de simetria radial rápida (*fast radial symmetry transform*) e região mínima da imagem pré-processada. A ultima parte do processo é o pós-processamento para retirar regiões que provavelmente não são de núcleos e arrumar o contorno dos núcleos parametrizando com elipses. Ao se mudar o elemento estruturante do pré-processamento, a segmentação ocorre em escalas diferentes permitindo uma análise multi-escalar. A examinação visual mostrou uma boa performance com um limitado número de sub-segmentação, super-segmentação e segmentação de objetos de núcleos não epiteliais. Obteve-se uma média estimada da sensibilidade acima de 85% e uma média estimada de predição positiva próxima de 89%. A distribuição dos coeficientes de Sørensen-Dice teve um valor máximo próximo de 0.9, com a maioria da segmentação tendo valor maior que 0.8. O principal motivo da baixa sensibilidade é o grade número de núcleos bem pequenos que não foram segmentados por causa do tamanho. Os casos fora da curva com baixa predição positiva são os casos de câncer de alto nível e/ou casos com um fibroblasto largo. Uma melhoria que poderia ser inclusa seria uma pré-segmentação que divide o tecido em regiões epiteliais e estromais. Outra melhoria seria um método para segmentar linfócitos.

Um grande número de diferentes estruturas podem ser encontradas em imagens histológicas com coloração *H&E*. Além disso, diferentes imagens são caracterizadas por diferentes tipos de tecido, o que mostra a grande variabilidade de aparência das estruturas. No trabalho de [36],foi desenvolvido uma rede neural profunda (*Deep Neural Network*, cuja abreviatura é DNN) para identificar núcleos em processo de mitose já que o número de objetos fazendo mitose em seções histológicas é um grande indicador para rastreio e avaliação de câncer. A automatização do processo pode reduzir o tempo,custo e erros como também melhorar a comparabilidade de resultados obtidos em diferentes laboratórios. A *DNN* utilizada consiste em uma rede neural convolucional com *max-pooling*. O valor de referência (*ground truth*) pode ser classificado em duas classes diferentes, com mitose ou sem mitose. A classe com mitose se refere a todos os *pixels* presentes (ou próximos) no centroide de todas as mitoses vistas, enquanto que a classe sem mitose são todos os *pixels* restantes. Pelo pequeno número de exemplos de mitoses presentes no conjunto de treinamento, foi utilizado um *data-augmentation* de rotações arbitrárias e/ou espelhamento pois a detecção de mitose é invariante rotacionalmente. Os núcleos em processo de mitose são normalmente raros e separados , logo não costuma ocorrer o problema de aglomeração ou de núcleos se tocando. Porém é extremamente difícil diferenciar núcleos em processo de mitose de núcleos que não estão em processo de mitose. Foi obtido um *F-score* de 0.782 (ressaltando-se que o limiar de detecção pode mudar este resultado). Este resultado foi o melhor se comparado com todos os outros competidores em imagens histológicas de câncer de mama da época. Uma futura melhoria seria aumentar o conjunto de dados com o intuito de trazer a detecção automatizada de mitose dentro da prática clínica.

Uma análise qualitativa e quantitativa de diferentes tipos de tumores no nível celular pode ajudar a se ter um melhor entendimento do tumor e então explorar diferentes tratamentos para

câncer. A detecção e a classificação de núcleos celulares em imagens histopatológicas de tecido canceroso com coloração *H&E* é um grande desafio por causa heterogeneidade celular. No trabalho desenvolvido por [37], apresentou-se uma abordagem de aprendizagem profunda sensitiva ao vizinho local(*local neighborhood*) para a detecção e classificação de núcleos em imagens histopatológicas com coloração *H&E* de adenocarcinomas em câncer colorretal, baseadas em rede neural convolucional. Para a detecção de núcleos foi proposta a rede neural convolucional espacialmente restrita (*spatially constrained convolutional neural networks*, cuja abreviatura é SC-CNN) que inclui uma camada para estimar parâmetros e outra camada para regressão espacial. Esta rede consegue prever a probabilidade do pixel ser o centro do núcleo, isto é, *pixels* mais perto do núcleo há um valor de probabilidade maior do que aqueles longe do núcleo. Então, os núcleos são detectados quando os valores são maiores que o limiar proposto do mapa de predição. Para a classificação de núcleos foi usada o preditor de conjunto de vizinhos(*neighboring ensemble predictor*, cuja abreviatura é NEP) em conjunto com uma rede neural convolucional padrão com *softmax*. Realizou-se um aumento de dados(*data augmentation*) de rotação para ambas as redes e uma pertubação na distribuição de cor na rede neural *softmax* de classificação. Há um problema na entrada de características da rede, por exemplo, a detecção de núcleos costuma ser melhor com intensidade hematoxilina que a intensidade padrão RGB e a classificação de núcleos costuma ser melhor com a intensidade padrão RGB. O melhor resultado de detecção foi de F1-score de 0.802, enquanto que a de classificação foi de F1-score de 0.784. Porém se se usar a detecção em seguida da classificação se obtém um valor de F1-score de 0.692. Uma abordagem automática da combinação de detecção e classificação pode potencialmente oferecer uma análise sistemática quantitativa da morfologia e dos constituintes do tecido, sendo assim um benefício a patologia prática para um melhor entendimento do microambiente do tumor.

Com o grande aumento no interesse em Aprendizado Profundo (*Deep Learning*) em análises de conjunto de dados de imagens, a Histopatologia se tornou um ótimo caso de aplicação de Aprendizado de Máquina por sua alta complexidade e tamanho. A estratégia de *Stacked Sparse Autoencoder* (SSAE) foi utilizada no trabalho publicado por [38] para a detecção de núcleos em imagens histopatológicas de câncer de mama de alta resolução. A detecção de núcleos automatizada é um desafio por causa do grande número de núcleos e a grande variabilidade em tamanho, aparência,forma e textura dos núcleos individualmente. A SSAE é uma rede neural com múltiplas camadas do básico *Sparse Autoencoder* (SAE), um modelo de conexão completa para aprendizado de características de alto nível com uma matriz de pesos global para representação das características. Isto é, SSAE é uma arquitetura em que a rede *encoder* representa as intensidades do pixel modelado por atributos de baixa dimensão, enquanto uma rede *decoder* reconstrói as intensidades do pixel original usando características de baixa dimensão. Usando um *slide scanner* foi obtido o conjunto de dados de imagens histopatológicas de mama com coloração *H&E* de alta resolução. O modelo consiste na rede SSAE seguida de um classificador *softmax* que supervisiona o aprendizado para o treinamento da camada superior. O aprendizado de características do SSAE pode ser considerado não supervisionado. A estratégia de SSAE foi comparada com outras estratégias de aprendizado de máquina, algumas mais rasas como *autoencoder* comum e SAE, outras como redes neurais convolucionais (CNNs). O resultado da estratégia proposta pelo artigo foi superior a outras e obteve um F1-score de 84.49% e uma precisão média (AveP) de 78.83% . Como o

esperado, a estratégia SSAE foi melhor que os métodos baseados em características elaboradas manualmente (*handed-crafted features*) por ter uma captura melhor de informação estrutural de alto nível e melhor discriminabilidade entre núcleos e não núcleos. A detecção de núcleos apresentada pode fornecer *seed points* com melhor acurácia ou gráficos de características que ajudam na caracterização topológica das células em tumores histológicos.

Diagnósticos com a ajuda de computadores estão ficando cada vez mais importantes para auxiliar patologistas na detecção e diagnóstico de câncer de mama. Segmentação de núcleos é o primeiro passo para o auxílio na análise automática da histopatologia de células mamárias em que há um alto grau de complexidade. Esta análise é importante para uma detecção precoce de câncer de mama, que é o principal tumor maligno observado em mulheres. No trabalho desenvolvido por [39], fez-se a segmentação e classificação de imagens histopatológicas de mama com coloração $H\&E$ adquiridas por uma câmera digital adaptada em um microscópio. O método de segmentação começa com a transformação *top-bottom hat* para melhorar a imagem em nível de cinza, depois uma decomposição *Wavelet* combinada com uma *multi-scale region-growing* para extrair as regiões de interesse (ROIs) que são escolhidas por um mecanismo de votação. A segmentação termina com a separação de células sobrepostas com um modelo de estratégia de divisão dupla (*double strategy splitting model*, cuja abreviatura é DSSM) que é a combinação de operações morfológicas matemáticas adaptativas com o espaço de escala de curvatura (*curvature scale space*). O método de classificação é a seleção das características na textura e formato dos núcleos por um algoritmo genético de agente em cadeia (*chain-like agent genetic algorithm*, cuja abreviação é CAGA) aplicado em um classificador SVM (*support vector machine*). A performance da segmentação de núcleos foi melhor que outros métodos em termos de sensibilidade (91.53%) e especificidade (91.64%) com valores médios maiores que 91% e desvio de 4% , mostrando ser um método estável e com os requerimentos clínicos. A performance da classificação entre células normais e malignas atingiu uma acurácia de 96.19% , sensibilidade de 99.05% e especificidade de 93.33%, mostrando ser um método que pode ajudar no diagnóstico de câncer de mama com seu potencial de detectar câncer em células histopatológicas aparentemente normais. O método de classificação precisa ser testado em um conjunto de dados maior para confirmar se o método é robusto.

Um sistema automático de análise de imagem de alto rendimento costuma precisar de um método de segmentação de núcleos robusto e preciso. Observou-se que a ajuda de computadores para análises de imagem pode melhorar a objetividade e reprodutibilidade principalmente da natureza subjetiva da análise de imagens patológicas digitalizadas. O *framework* proposto no trabalho de [40] é de uma rede neural convolucional profunda para gerar um mapa de probabilidades e inicializar o formato dos núcleos e um algoritmo baseado em seleção de forma esparsa (*selection-based sparse shap*) para fazer o contorno dos núcleos. Como é comum núcleos estarem sobrepostos ou aglomerados em imagens histopatológicas digitalizadas, foi proposto um modelo para facilitar o contorno e obter inicializações mais robustas. Primeiramente, ocorre o treinamento da rede neural convolucional profunda para gerar o mapa de probabilidades que representa as regiões dos núcleos. Realizou-se um aumento no conjunto de dados (*data augmentation*) por meio de rotação para adquirir invariância a esta operação. No mapa de probabilidades, se utilizou dois limiares para retirar os não núcleos da imagem e também ruídos de tamanho específico. Depois, é feita a inicialização

de formato com um mapa de distância, transformação H-minima e operações morfológicas para formar marcadores do formato(*shape markers*). Então é realizada a combinação das informações ascendente e descendente (*bottom-up and top-down*) para adquirir a delineação dos núcleos. Isto é, foi realizada alternadamente uma interferência do formato com um modelo de formato esparso (*Sparse shape model*) e uma deformação do formato com um modelo de contorno ativo repulsivo até ficar em um estado estável e então obter a segmentação finalizada. O *framework* proposto foi testado em 3 conjunto de dados diferentes: câncer de mama, tumor de cérebro e tumor neuroendócrino pancreático (NET). Na detecção com inicialização de formato, obteve-se F1-score de 78%, 77% e 88% respectivamente com os três conjuntos de dados. Na segmentação, obteve-se em média 80%, 85% e 92% no coeficiente de similaridade Dice(*Dice similarity coefficient*). A inicialização de formato proposta é robusta em ruídos e intensidade não homogênea. Observou-se que o método proposto foi superior a outros métodos mais clássicos(como super-pixel,*watershed* baseada em marcadores e *graph-cut and coloring*) e que pode ser estendido para outras aplicações.

As técnicas e dispositivos de *hardware* para Visão Computacional têm melhorado muito durante os últimos anos e com isso os problemas de análise de imagens histológicas normalmente manuais, como a variabilidade do observador, características visuais sutis e o alto tempo para examinar as imagens estão sendo amenizados pela patologia computacional. O objetivo principal do trabalho publicado por [41] é ajudar na patologia computacional a segmentar núcleos em imagens histológicas com coloração $H\&E$ diversas. Para isso, foi liberado um conjunto de dados grande e diverso de imagens anotadas com as bordas de núcleos que são difíceis de segmentar. Também foi proposta uma nova métrica para avaliar as técnicas de segmentação que tratam os erros de detecção e segmentação de forma unificada. Além disso, propôs-se uma nova técnica de segmentação com aprendizado profundo. O conjunto de dados liberado foi elaborado para ser o mais diverso possível. Cada imagem é de um paciente diferente, as imagens são de 18 hospitais diferentes e também há tecidos de sete órgãos diferentes. O critério de avaliação proposto possui critérios de detecção sem parâmetros, que funcionam independentemente do tamanho do núcleo e da ampliação da imagem e critério de segmentação que penaliza tanto os erros no nível de objeto e pixel. Ambos os critérios usam a equação de Jaccard, razão pela qual a combinação dos critérios foi chamada de índice agregado de Jaccard (*aggregated Jaccard index* (AJI)). O método de segmentação proposto começa no pré-processamento com a normalização de cor que considera a variação de coloração e do processo de digitalização da imagem. Depois, é realizada a segmentação com uma arquitetura de rede neural convolucional de três classes que enfatiza a detecção dos *pixels* da borda dos núcleos. Então a segmentação é finalizada com o pós-processamento, em que é feita a detecção de núcleos com um limiar no mapa de probabilidades de classe e então a técnica de crescimento dentro da borda (*grow inside into boundary*).

Com o intuito de melhorar o desenvolvimento da morfometria de núcleos e softwares de patologia computacional para uso clínico e de pesquisa, o trabalho de [41] propôs um novo conjunto de dados, um novo critério de avaliação e um novo método de segmentação. O novo conjunto de dados demonstrou ser bem diverso e largo e com isso é esperado um avanço nas técnicas de segmentação generalizada de núcleos. O novo critério de avaliação mostrou ser superior às outras métricas por ter unificado o tratamento de detecção e segmentação. As outras métricas de qualidade de formato

consideram apenas as detecções verdadeiras diferentemente do critério AJI proposto. A técnica de segmentação proposta no artigo obteve F1-score de 82.67%, média do coeficiente Dice de 76.23% e 50.83% no critério AJI proposto e então foi superior a outras técnicas no mesmo conjunto de dados proposto.

A análise histopatológica vem ganhando muita atenção nos últimos anos com o advento da Patologia Digital, que torna possível construir ferramentas para a análise automática das complexas imagens histológicas. No trabalho desenvolvido por [42] foi apresentado um fluxo de trabalho completo para a segmentação de núcleos em imagens histopatológicas usando rede neural profunda treinada com imagens anotadas manualmente e processando os mapas de probabilidades para separar núcleos segmentados unidos. A primeira contribuição deste artigo é um novo conjunto de dados para a detecção de núcleos em imagens histopatológicas com coloração *H&E*. O fluxo de trabalho começa com um aumento de dados (*data augmentation*) de rotações, ruídos, borrões e deformações elásticas aleatórias no conjunto de dados inicial. Depois, são testados diversos hiper-parâmetros da configuração da rede neural, porém com a conclusão de que a taxa de aprendizagem (*learning rate*) é a mais significante. Também foram usadas camadas já treinadas que tornaram o treinamento mais eficiente e o resultado mais robusto. Após o treinamento da rede neural profunda, é utilizada no pós-processamento uma dinâmica morfológica de um parâmetro livre para separar ou não objetos em mais objetos, isto é, separar núcleos sobrepostos ou aglomerados. Finalmente, é aplicado o algoritmo *watershed* combinado com a dinâmica morfológica para finalizar a segmentação. Este fluxo de trabalho foi testado em quatro arquiteturas diferentes de rede neural: PangNet , *Fully Convolutional Net* (FCN) , DeconvNet e uma que é o conjunto das duas arquiteturas anteriores chamada de *Ensemble*. A rede neural mais rasa PangNet obteve o pior resultado, ela aprendeu as informações das cores por exemplo porém quando o núcleo não é homogêneo a rede falha. Ela obteve F1-score de 67.6% e *Intersection over Union* (IoU) de 72.2%. As redes mais profundas FCN e DeconvNet foram mais capazes de reconhecer os núcleos inteiros e obtiveram F1-score de 76.3% e 80.5% e IoU de 78.2% e 81.4% respectivamente. Porém em ambas as redes houve super-segmentação vista no pós-processamento. A rede de conjunto *Ensemble* obteve o resultado de 80.2% de F1-score e 80.4 % de IoU, portanto não foi superior à rede DeconvNet.

A informação extraída de lâminas de tecido para se ter perfis quantitativos úteis e preditivos ainda é uma questão em aberto. Existem duas estratégias para extrair essas informações, aprender características de imagens com poder discriminativo em respeito a variável clínica e detectar elementos importantes em imagens histopatológicas e, com isso, descrever a abundância e a organização espacial destes elementos. O elemento mais importante para se detectar é o núcleo por ser um indicativo de fenótipos celulares. No trabalho publicado por [43] foi proposto um novo método de segmentação de núcleos em imagens histopatológicas com rede neural convolucional com uma tarefa de regressão em dois conjuntos de dados diferentes. O primeiro conjunto consiste em imagens histopatológicas com coloração *H&E* de pacientes com câncer de mama. O segundo conjunto consiste no conjunto de dados do mesmo tipo de imagem do conjunto anterior, porém de diversos tecidos como proposto pelo trabalho de [41]. Então realizou-se o aumento de dados (*data augmentation*) de rotação ,espelhamento, borramento e deconvolução de cor. O *workflow*

de segmentação proposta neste artigo foi de prever o mapa de distância não normalizada com a arquitetura de rede neural U-Net e uma perda de regressão (*regression loss*) e então usar a saída da rede em um pós-processamento com uma dinâmica morfológica proposta em [42]. O *workflow* proposto foi chamado de DIST. Outras abordagens neste campo focam no contorno do objeto ou na interseção dos objetos e este artigo mostrou ser vantajoso focar no interior do objeto principalmente em casos em que na imagem, o contorno está borrado ou com menos significância. Foram testadas três arquiteturas diferentes (U-Net, FCN, Mask R-CNN), com e sem pós processamento e também o *workflow* proposto no artigo. O melhor modelo em relação ao AJI foi o DIST proposto no artigo com o valor de 56%. As arquiteturas com o pós-processamento também obtiveram um bom resultado, em torno de 51%. Em questão do F1-score, tanto o DIST quanto o FCN e U-Net obtiveram bons resultados, em torno de 78%. Também é visto que o modelo proposto tem uma formulação diferente da usual: a tarefa de segmentação foi tratada como um problema de regressão ao invés de um problema de classificação, como normalmente é tratada.

2.3 Eixo 3: Técnicas automáticas de segmentação de podócitos

O trabalho publicado por [44] apresenta uma ferramenta com interface visual para assistir na anotação de dados e na exibição dos resultados de predição de redes neurais em WSIs. A interação humana ocorre ao fim do processo de segmentação, feita a partir das anotações automáticas da rede. Nesse sentido, o estudo tem o objetivo de mostrar que o método proposto é capaz de fornecer um melhor resultado na segmentação de compartimentos renais humanos e de rato, entre os quais figuram os podócitos. Além disso, a dependência de um grande conjunto de dados anotados para o uso de redes neurais convolucionais favorece o argumento dos autores de que a diminuição do peso de anotação das imagens deve ser perseguida. Outro ponto ressaltado pelos autores é a escassez de dados biológicos em larga escala, o que postergou a aplicação das CNNs nesse campo. Os autores propõem uma interface entre uma rede neural de segmentação semântica (*DeepLabV2*) e um programa de visualização de WSIs (ImageScope), denominado *Human AI Loop* (H-AI-L). O algoritmo proposto pode ser resumido pela sequência de passos: as anotações feitas são armazenadas e a partir delas geram-se máscaras de regiões; estas são utilizadas para treinar as redes que retornam as predições no formato utilizado pelo anotador de imagens; por fim, as anotações automáticas da rede podem ser visualizadas. Um ponto importante é que a visualização da WSI pode ser feita de forma integral (não somente de pedaços dela), em que se pode utilizar o recurso de *zoom*. A avaliação do algoritmo H-AI-L se dá em duas etapas: primeiro, localizam-se os glomérulos em WSIs do tecido renal de ratos; segundo, a partir destes, realiza-se a segmentação semântica. Tendo em vista o escopo desta seção, focaremos nos resultados apresentados relativos à segmentação das estruturas glomerulares, em detrimento dos resultados de performance de velocidade do algoritmo, um dos focos do estudo. Ainda em [44], um dos fenômenos observados foi que a rede foi capaz de gerar resultados superiores aos das anotações, quando havia imperfeições nestas. Em especial, esse erro ocorreu com mais frequência nas épocas iniciais de treinamento. A detecção multiclasse envolveu a alteração da rede de detecção de glomérulos para identificar núcleos de podócitos/ não podócitos. Nesse sentido, a rede de baixa resolução utilizada

para localizar os glomérulos permaneceu inalterada. A estratégia de segmentação, chamada pelos autores de multipasso *multi-pass segmentation*), envolveu identificação de candidatos em resolução mais baixa (1/16) e posterior segmentação em alta resolução. A justificativa por essa escolha de segmentação em duas etapas se deveu a redução no tempo e aumento de performance em termos de F1. Os resultados de sensitividade, especificidade, precisão e acurácia obtidos foram de, respectivamente, 0,92, 0,99, 0,93 e 0,99. Esta rede foi treinada a partir de 143 glomérulos anotados. Embora não seja o foco desta seção, pode-se destacar a possibilidade de aumentar o conjunto de treino a partir do método desenvolvido, uma vez que as anotações geradas pela rede são compatíveis com a ferramenta de anotação. Com isso, as anotações da rede podem ser editadas de maneira rápida e incluídas ao conjunto de treino. No futuro, pode-se explorar a capacidade da ferramenta em ser utilizada para criar grandes bases de dados de imagens médicas, uma vez que a redução do tempo de anotação é considerável. Além disso, poder-se-ia explorar a segmentação entre podócitos e outros núcleos para um número maior de glomérulos do que o utilizado (n=143).

O trabalho apresentado por [45] visa o estabelecimento de fatores incidentes sobre doença renal diabética (DKD), como a perda e a lesão de podócitos. Dois fatores que contribuem para a falência dos rins são a diabetes mellitus (DM) e a progressão da DKD, que pode se dar por razão do avanço de lesões glomerulares. Nesse sentido, a quantificação de podócitos é importante para os métodos em patologia digital que buscam a avaliação tanto das lesões quanto da perda deles. Uma das contribuições dos autores é a definição de um novo atributo de imagem envolvendo a quantificação da distribuição intra-glomerular dos podócitos. Ela consiste na distância euclidiana entre os podócitos e o polo urinário, definida tendo em conta a maior tensão a que essa região está submetida. Os autores já haviam desenvolvido um rede neural convolucional, H-AI-L, em trabalho anterior, a qual foi aproveitada neste. As amostras de tecido tiveram origem em 14 ratos, sete dos quais de controle e o restante acometido por DM (destes, três de forma moderada).O conjunto de treino consistiu na anotação de 3 WSIs com corante PAS, e o de teste, em 11 WSIs com o mesmo corante. Além disso, foram anotados todos os polos urinários em imagens com glomérulos, quando presentes. Ao todo, contabilizam-se 883 glomérulos e 248 polos urinários. A segmentação dos podócitos foi feita a partir da limiarização da imagem em níveis de cinza do canal de cor vermelho. Desse modo, combinou-se de forma lógica esta imagem binária com a resultante da segmentação glomerular proveniente da H-AI-L, de modo a se gerar uma máscara lógica apenas com sinais intraglomerulares e WT1 positivos (corante utilizado para visualização e anotação dos núcleos de podócitos). A imagem resultante foi então submetida a operações de morfologia matemática, de modo a refinar a máscara de podócitos. Em seguida, houve contagem dos podócitos e extração de características relevantes, como área, excentricidade e localização de centroide. A segmentação de núcleos se deu por meio de uma sequência de limiarização, operações de morfologia matemática e segmentação *watershed* para separar os núcleos que se intersectam. Então, houve contagem dos núcleos e extração de características de forma semelhante à feita para podócitos. Ainda em [45], os resultados do *pipeline* proposto foram obtidos por meio de comparação com as marcações de especialistas (*ground-truth boxes*). Desse modo, obteve-se sensitividade, especificidade e acurácia de, respectivamente, 0,727, 0,999 e 0,959 para os podócitos; para os núcleos, 0,771, 0,997 e 0,977. Com relação às propriedades morfométricas extraídas, uma importante conclusão é que houve diminuição no número de podócitos e aumento da distância para o polo urinário com o aumento

da severidade da doença. Além disso, encontrou-se algumas propriedades com comportamento não monotônico, isto é, cujos valores atingem o pico na condição DM moderada e tornam a cair na condição DM. Exemplos desse fenômeno são a área e o perímetro glomerular. Por fim, cabe destacar a importância da segmentação *watershed* realizada, tendo em vista que as estruturas intra-glomerulares que se intersectam, caso houvesse, comprometeriam resultado com um todo, uma vez que seriam tratadas como um objeto único. Os autores sugerem que trabalhos futuros possam incluir adição de novos atributos, como a distância para o polo vascular. Além disso, a construção de classificadores a partir dos principais atributos analisados - distância para UP e quantidade de podócitos - poderia indicar a validade deles para o diagnóstico da doença.

O trabalho desenvolvido por [46] apresenta um fluxo de trabalho para a caraterização de biópsias de rim humano. Entre os diversos parâmetros quantitativos analisados, estão os relacionados a podócitos. Um dos eixos do trabalho é a avaliação da perda de podócito em pacientes com glomerulonefrite associada a anticorpo citoplasmático antineutrófilo (ANCA-GN). Este objetivo em específico vem a complementar mudanças celulares conhecidas da ANCA-GN, como lesões de podócitos. Para tanto, utiliza-se imagiologia imunofluorescente para a visualização dos podócitos e, assim, a quantificação da perda deles. O conjunto de análise consistiu em mais de 27000 podócitos de 1095 glomérulos de 110 pacientes. Dos pacientes, 62 possuíam ANCA-GN e o restante pertencia ao grupo de controle. Ainda em [46], o traçamento do perfil das amostras foi realizado por meio de segmentação dual - no sentido de a saída retornar a segmentação de glomérulos e de podócitos - com a rede neural convolucional U-Net. Os autores utilizaram uma abordagem de monitoramento dos melhores pesos e hiperparâmetros ao longo do treinamento, de modo que aqueles que gerassem o melhor resultado pudessem ser armazenados. O conjunto de validação foi utilizado somente para avaliar a rede de melhor resultado durante os treinamento. A validação consistiu em métricas a nível de *pixel* e a nível de objeto. Um dos pontos focais é o estabelecimento de correlações morfométricas a partir dos resultados obtidos. Nesse sentido, a partir de uma validação LOO, os autores geraram uma pontuação podométrica englobando diversos fatores, tais como densidade, tamanho e número de podócitos. Com 65 imagens de treino e adotando uma estratégia de validação cruzada 10-*fold*, atingiu-se *Dice score* superior a 0,9. Em comparação a uma rede ImageJ, a U-Net proposta apresentou resultados similares na segmentação de glomérulos, mas superiores na de podócitos, tanto a nível de *pixel* quanto a nível de objeto. A avaliação dos resultados dos mesmos pacientes obtidos de forma diferente (diferentes locais, operadores e aparelhos) mostrou que não houve diferença significativa na densidade de podócitos observada. No entanto, os outros fatores sofreram de variância significativa nessas condições. Tendo em vista a correção deste problema, os autores propuseram uma rede adversarial generativa (GAN) baseada em U-Net, de modo aproximar imagens tiradas em condições diversas. Com a U-Net GAN, conseguiu-se melhora em termos dos *Dice-score* (de 0,65 para 0,81) nas imagens de referência, no entanto, isso não se refletiu quando as redes foram treinadas em todo o conjunto de imagens. A nível de objeto, os podócitos foram segmentados com *Dice Score* de 0,86 no grupo de controle e 0,87 no grupo ANCA-GN; a nível de *pixel*, 0,95 e 0,91, respectivamente. Um importante achado do estudo é o aumento da distância média de podócitos. Além disso, foi possível estabelecer a relação entre os números médios de podócitos e os de tamanho glomerular. Por meio de validação cruzada LOO, os autores constataram que a assinatura de perda de podócitos atingiu AUC de

0,88 na curva *precision-recall* e acurácia de 0,92 na distinção entre pacientes controle/ANCA-GN. Com efeito, a relação observada entre a perda de podócitos e a presença de doenças glomerulares é substantiva, no entanto, resta a estudos futuros avaliar a possibilidade de utilizar esse parâmetro como meio para diagnóstico de pacientes. Além disso, estudos futuros poderiam atestar a capacidade de generalização do modelo (para identificar podócitos e, a partir das características morfométricas prover um diagnóstico) utilizando um conjunto de dados maior.

No trabalho apresentado por [47], os autores utilizam redes neurais convolucionais e técnicas em processamento de imagens para realizar diversas tarefas: localização de glomérulos, identificação de lesões glomerulares e identificação de células internas a eles em amostras de pacientes com IgAN (Nefropatia da Imunoglobulina A). Tendo em vista que a segmentação semântica e análise quantitativa de podócitos é feita no artigo, optamos por incluí-lo nesta seção. O sistema desenvolvido, ARPS, engloba diversas redes neurais, uma para cada tarefa: segmentação dos glomérulos nas WSIs (U-net), seleção de glomérulos (DenseNet), classificação entre esclerose segmental (SS), crescente (C) e nenhum acima (NOA), e segmentação das células glomerulares intrínsecas (V-Net). A localização de glomérulos incluiu 360 WSIs de treino e validação, e 40 WSIs de teste (todas referentes a pacientes com IgAN), totalizando 12418 glomérulos das classses SS, C, GS (globalmente esclerosado) e NOA. O valor médio de precisão/*recall* para a localização de todos os tipos de glomérulo foi de 0.931/0,949. A detecção das células intraglomerulares foi feita a partir de 460 glomérulos, valor correspondente a cerca de 70 mil células. Neste quesito, atingiu-se precisão de 0,882 e *recall* de 0,879. A partir do reconhecimento das células, foi possível extrair diversas informações, tais como área, diâmetro e quantidade. Com respeito à distribuição das células internas, utilizou-se o ARPS para contar células mesangiais (M), endoteliais (E) e podócitos (P), tendo em vista a constatação de hipercelularidade. Esta contagem foi feita a partir de 3592 amostras NOA. Desse modo, encontrou-se uma proporção de M,E,P de 0,41:0:36:0,23 respectivamente. Em comparação aos estudos presentes na literatura, este tem um grande conjunto de dados, compreendendo mais de 400 WSIs. Além disso, o ARPS foi capaz de identificar as células intraglomerulares e estabelecer a relação entre a proporção destas e o quadro de IgAN. Em trabalhos futuros, o ARPS poderia ser adaptado para receber como entrada outros tipos de corante, uma vez que ele se baseou no corante PAS.

Capítulo 3

Metodologia Proposta

A metodologia elaborada para a realização deste trabalho é descrito pelo fluxograma da Figura 3.1. Todas as etapas serão descritas nas Seções e Subseções subsequentes do capítulo. Vale ressaltar que o desenvolvimento da metodologia proposta é definido de forma sequencial e incremental, sendo que para uma etapa ocorrer, sua anterior deverá estar concluída.

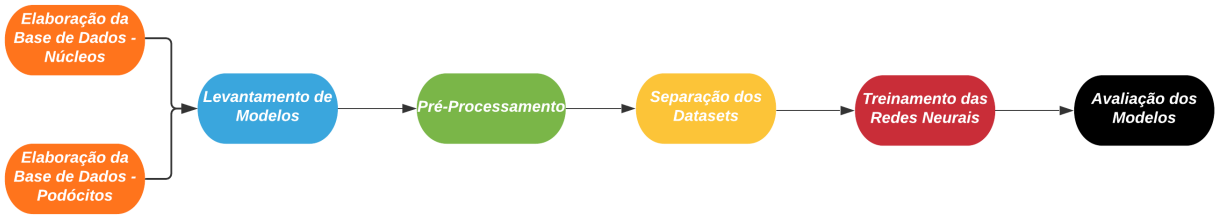


Figura 3.1: Fluxo de trabalho

3.1 Elaboração da Base de Dados de Imagens

As imagens histológicas foram obtidas a partir do Centro de Pesquisa Gonçalo Moniz da Fiocruz - Bahia. Elas foram obtidas por meio de exames de biópsia, em que se retira um pedaço microscópico de tecido do paciente, que então é cortado em seções transversais, coradas, compactadas em uma lâmina e, finalmente, ampliadas e capturadas por uma câmera acoplada no microscópio.

3.1.1 Núcleos

A base de dados utilizada na tarefa de detecção de núcleos nas imagens histológicas foi anotada pelos autores deste texto e então validada por um patologista experiente. Após a marcação de todas as regiões de núcleos com *bounding box* (caixa delimitadora) das imagens pelo software *Labelme*, o patologista validou uma amostra de algumas imagens. Todas as 99 imagens deste conjunto, conforme mostra a Tabela 3.1, são de células saudáveis com coloração H&E e resolu-

ção variando de 327x362 até 1459x1333. Por fim, utilizou-se o software *Roboflow* para mudar a marcação do formato do *Labelme* para o formato do *YOLOv5 PyTorch* e também realizar um pré-processamento para deixar todas as imagens no mesmo tamanho (416x416) com *resize* e realizar a auto-orientação de dados com a padronização de pixels, descartando a orientação *EXIF*.

Núcleos				
Condição				Total
Saudáveis		Lesionadas		
99 (100%)		0		99
Coloração				Total
H&E	PAM	PAS	Tricrômico	
99 (100%)	0	0	0	99
Número total de núcleos anotados: 21464				

Tabela 3.1: Base de dados de núcleos

3.1.2 Podócitos

A base de dados utilizada na tarefa de detecção de podócitos nas imagens histológicas foi anotada e validada por dois patologistas experientes. O primeiro patologista marcou as regiões dos podócitos com *bounding boxes* nas 122 imagens utilizando o software *Labelme* e então o segundo patologista validou algumas amostras. Todo o processo de anotação durou cerca de 1 ano. Este conjunto possui em torno de 60 imagens em comum com a base de dados de núcleos, resoluções variando de 640x330 até 1024x768 e, além disso, são divididas entre imagens com podócitos saudáveis e lesionados e com respeito à coloração utilizada, conforme mostra a Tabela 3.2. As imagens têm coloração H&E, *PAM*, *PAS* e *tricrômico*. Por fim, utilizou-se o software *Roboflow* para mudar a marcação do formato do *Labelme* para o formato do *YOLOv5 PyTorch*, realizar um pré-processamento para deixar todas as imagens no mesmo tamanho (416x416) e realizar a auto-orientação de dados com a padronização de *pixels*, descartando orientação *EXIF*.

Podócito				
Condição				Total
Saudáveis		Lesionadas		
99 (81.1%)		23 (18.9%)		122
Coloração				Total
H&E	PAM	PAS	Tricrômico	
71 (58.2%)	3 (2.5%)	42 (34.4%)	6 (4.9%)	122
Número total de podócitos anotados: 3707				

Tabela 3.2: Base de dados de podócitos

3.2 Levantamento de Modelos

Nesta etapa, faz necessário apresentar algumas informações importantes sobre Redes Convolucionais, como descrito a na Subseção 3.2.1.

3.2.1 Redes Neurais Convolucionais

Como em outras redes neurais, as CNNs são compostas de blocos básicos, chamados camadas. O grande diferencial deste tipo de rede é a presença de um tipo específico de camada: a convolucional. Esta camada consiste em um conjunto de filtros, isto é, de matrizes de números discretos, com os quais se aplica uma operação de convolução a uma matriz de entrada. A saída desse tipo de operação é chamada de *feature map*. O processo de aprendizado durante o treinamento é responsável pela alteração dos valores dos pesos de cada filtro.

A intuição matemática por trás da operação de convolução é a soma da multiplicação de termos correspondentes do filtro por uma submatriz da entrada de mesmo tamanho. O filtro então percorre ao longo das linhas e colunas da matriz de entrada, até que não se possa mais realizar a operação. Formalmente[1], a convolução de uma imagem I de duas dimensões por um filtro (também comumente chamado de *kernel*) é dada pela Equação 3.1.

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n) \tag{3.1}$$

Um importante parâmetro utilizado durante a convolução é o *stride* (passo). Ele se refere ao número de unidades de deslocamento do filtro em cada passo da operação. Quando o número de *stride* é igual ao tamanho do filtro, por exemplo, não há interseção entre os elementos de entrada utilizados na convolução. A consequência da aplicação de stride maior que um é gerar um espaçamento nas regiões em que se aplica o filtro. Um efeito óbvio deste espaçamento gerado com o *stride* é a sub-amostragem: o *feature map* de saída de resultante da operação com *stride* n=2 será menor do que o *feature map* de saída obtido com *stride* n=1. Com efeito, as dimensões de saída para um *feature map* de entrada com dimensões h x w, *stride* s e filtro f x f são dadas pela Equação 3.2.

$$h_1 = \left\lfloor \frac{h - f + s}{s} \right\rfloor, \quad w_1 = \left\lfloor \frac{w - f + s}{s} \right\rfloor \tag{3.2}$$

Nem sempre, contudo, a redução da dimensionalidade é desejável. Para além de questões como características próprias da aplicação, que pode requerer que as convoluções mantenham o tamanho espacial, é possível que ocorra um colapso da dimensão das *features* após poucas operações, o que impediria que se projetasse redes muito profundas. Tendo em vista mitigar esses efeitos, é comum que se aplique *zero-padding*, técnica que consiste em aumentar o *feature map* em ambas as dimensões por meio do acréscimo de linhas e colunas com zeros. Com isto, as dimensões

[1] A equação mencionada descreve a convolução com o *kernel* invertido. Ver Seção 6.1 sobre a diferença de operar com o *kernel* original

de saída da convolução considerando um fato de *padding* p podem ser reescritas como na Equação 3.3. A Figura 3.2 ilustra a convolução de um *feature map* 4x4 (desconsiderando-se o *zero-padding*) por um filtro 2x2, com parâmetros s=2 e p=1.

$$h_1 = \left\lfloor \frac{h - f + s + p}{s} \right\rfloor, \quad w_1 = \left\lfloor \frac{w - f + s + p}{s} \right\rfloor \tag{3.3}$$

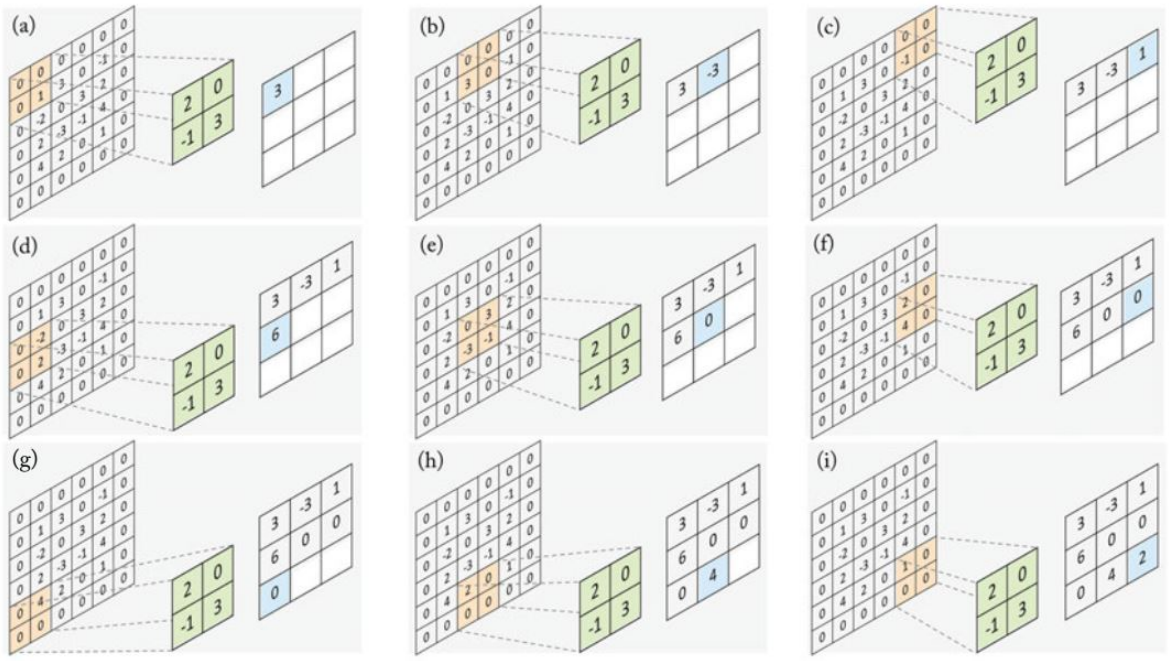


Figura 3.2: Exemplo de convolução [48]

Um tipo de camada importante nas CNNs é a de *pooling*. Em linhas gerais, o funcionamento desta camada consiste em combinar as ativações de *feature* presentes em um dado bloco do *feature map* de entrada. Esta combinação se dá por meio de uma função, como o valor máximo ou a média. Um exemplo de *max pooling*, caso em que se toma o valor máximo com função de operação sobre os blocos, é mostrado na Figura 3.3. Este processo de subamostragem é importante para que se tenha um representação compacta das *features* que não dependa da escala e posição de um objeto, tornando-as invariantes a pequenas alterações nos objetos, como translações [49].

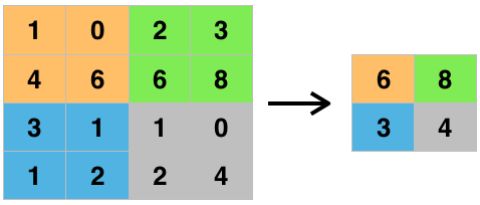


Figura 3.3: Exemplo de *max pooling*

É comum que se utilize alguma função de ativação nos pesos das camadas de CNNs, tanto convolucionais, quanto densas. Uma das razões para isto é mapear o eixo (ou semi-eixo) real

para um intervalo comprimido, como [0, 1]. Além disso, a aplicação dessas funções de ativação, dentre as quais pode-se citar as funções ReLU, sigmoide e tangente hiperbólico, é útil para que se possa identificar mapeamentos não lineares. Uma descrição rápida sobre algumas funções de ativação pode ser conferida na Seção 6.2. As camadas densas, mencionadas acima, são também conhecidas como camadas completamente conectadas (fully connected layers). Elas são idênticas as camadas das MLPs (*Multi-layer Perceptron*). Embora existam redes neurais convolucionais em que camadas intermediárias são desse tipo, o mais comum é que elas sejam encontradas na parte final da arquitetura. Em essência, elas correspondem a camadas convolucionais com filtro de tamanho 1x1. A operação a ela associada pode ser descrita por meio da aplicação de uma função não-linear, como as que vimos, à combinação linear entre os pesos de conexão das camadas e os vetores ativação de entrada. Desse modo, de posse das noções básicas apresentadas nesta seção, podemos apresentar uma arquitetura básica de rede neural convolucional, como a que é mostrada na Figura 3.4.

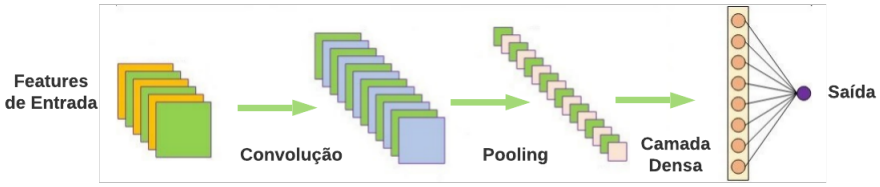


Figura 3.4: Arquitetura básica de uma rede neural convolucional

3.2.2 Segmentação de Imagens por Redes Neurais Convolucionais

Em 2015 [12], um pesquisador chamado Joseph Redmon e seus colegas introduziram um sistema de detecção de objeto em tempo real em um *framework* chamado *Darknet*. Este sistema ficou conhecido como YOLO (abreviatura do inglês *You Only Look Once*). É um sistema de detecção de objetos que considera o problema como uma simples regressão, um modelo que prediz as *bounding box* simultaneamente com as classes de probabilidades desse objeto. A motivação do autor foi criar um modelo unificado de todas as fases em uma rede neural, isto é, um modelo que computa todas as características de uma imagem e realiza as predições de objetos ao mesmo tempo. O YOLO é uma rede neural convolucional que foi evoluindo com o tempo e terminou na terceira versão em 2017 [50], porém um pesquisador chamado Alexey Bochkovskiy criou a quarta versão em 2020 [51] baseada no *Darknet framework* das primeiras três versões . Após um mês da criação da quarta versão, foi criada a quinta e atual versão pelo pesquisador Glenn Jocher e o departamento de pesquisa da *Ultralytics LLC* com *Pytorch framework*, a qual inclui algumas melhorias e diferenças. Isso significa que o YOLOv5 é um sistema em tempo real de detecção de objetos que usa o *framework Pytorch* e o modelo de arquitetura do YOLOv4.

O modelo de arquitetura do YOLOv5, mostrado na Figura 3.5, pode ser resumido em:

- *Backbone*: este bloco é um extrator de *features* das imagens de entrada. Ele é composto

de um *Focus structure* e de uma rede CSP (*Cross-stage partial connections*), que mantém as *features* por propagação e incentiva a rede em reusar *features* reduzindo o número de parâmetros da rede, permite as *features* mais refinadas a irem nas partes mais profundas da rede eficientemente.

- Bloco Adicional: o bloco SPP (*Spatial Pyramid Pooling block*) recebe o mapa de *features* da saída do *Backbone* e então aumenta o campo receptivo (*receptive field*) e depois separa as *features* mais importantes.

- *Neck*: este bloco combina as *features* formadas pelo *Backbone* para o passo de detecção da *Head*. Ele é composto da arquitetura PANet (*Path Aggregation Network*) e é uma versão avançada da arquitetura FPN (*Feature Pyramid Network*) das versões mais antigas do YOLO. A arquitetura do PANet transfere as *features* semânticas das camadas mais altas da rede para concatenar com as *features* mais refinadas das camadas mais baixas e profundas da rede. A proposta utilizada para evoluir a arquitetura de FPN para PANet consistiu em adicionar um caminho da parte mais profunda da rede para a parte mais rasa da rede e assim conectar diretamente as *features* mais refinadas com as *features* semânticas.

- *Head*: este bloco é um detector de um estágio que serve para realizar previsões densas. Ele é composto por uma detecção baseada em âncora(*anchor-based detection*) com três níveis diferentes de detecção. É usado o GIoU-loss (*Generalized Intersection over Union loss*) como função de regressão.

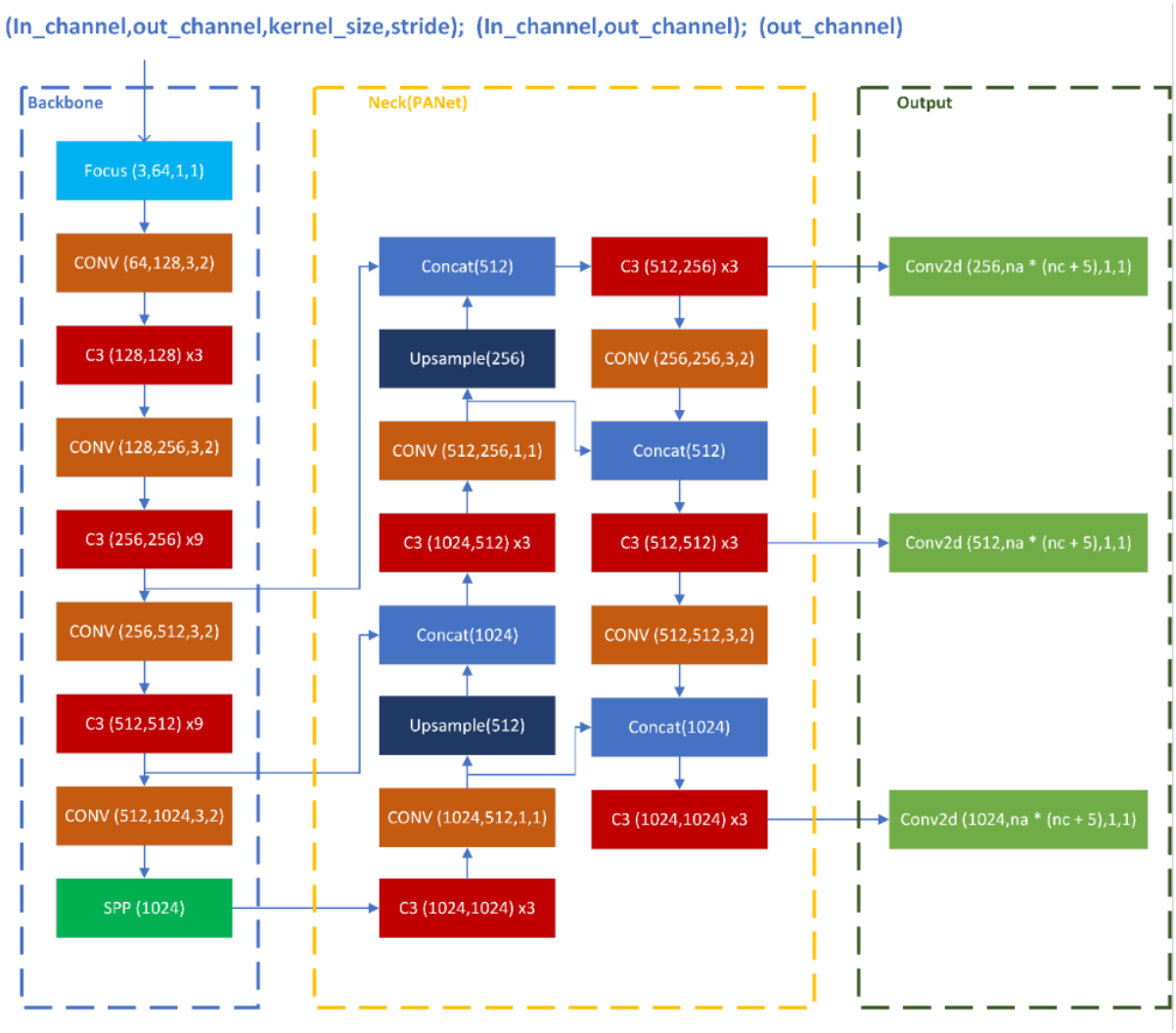


Figura 3.5: Modelo de arquitetura do YOLOv5 v4.0.[2]

A arquitetura do YOLOv5 é extremamente parecida com a do YOLOv4, porém com algumas diferenças de engenharia, além de utilizar um *framework* diferente. Um exemplo dentre essas diferenças é a integração do processo de seleção das caixas de âncora para que se possa usar qualquer conjunto de dados como entrada além do usual conjunto de dados COCO, como foi proposto nas versões mais antigas do YOLO.

Utilizou-se quatro modelos diferentes do YOLOv5 com profundidades diferentes. Quando o modelo é mais profundo, observa-se que o treinamento e a detecção são mais lentos e computacionalmente pesados. Porém o modelo costuma obter um valor de mAP superior, como mostrado na Figura 3.6 no conjunto de dados COCO.

[2]Declaração de direitos autorais: este artigo é o artigo original do blogger que segue o acordo de direitos autorais CC 4.0 BY-SA. Por favor coloque o link da fonte original e essa declaração para reimpressão. Link do artigo: https://blog.csdn.net/Q1u1NG/article/details/107511465

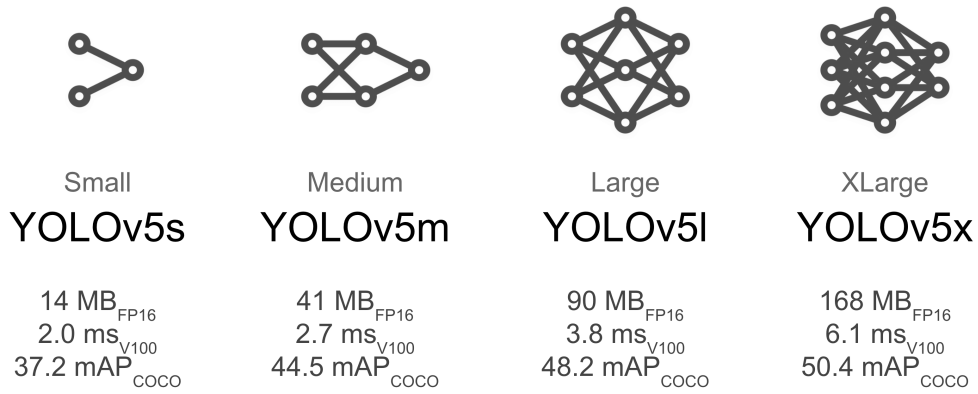


Figura 3.6: Diferenças das versões do YOLOv5. Crédito:
`https://github.com/ultralytics/yolov5/issues/475`

3.3 Pré-processamento

Normalmente, é necessária uma grande base de dados para que a rede neural profunda consiga aprender melhor. Para aumentar o tamanho do conjunto de dados do treinamento para redes neurais, pode-se utilizar a estratégia de *data augmentation*. Ela consiste em aumentar o número de dados de treinamento, podendo melhorar de forma limitada o aprendizado e os resultados. Utilizou-se estratégias para ampliar o conjunto de dados como rotação, *flip* (espelhamento), mudança no brilho e uma combinação de recorte, saturação, *blur* (borramento) e *salt-and-pepper noise* (ruído). O número de cópias geradas por imagem original pode ser vista na Tabela 3.3.

A rotação foi realizada de forma a se gerar 36 novas imagens por imagem original. Rotacionou-se a imagem original de 30°até 330°, resultando em, aproximadamente, 8°de espaçamento a cada imagem gerada, com relação à anterior. A rotação garante que o modelo treine para imagem rotacionadas e com isso fique invariante a rotações. Além disso, realizou-se três tipos de *flips*: horizontal, vertical e ambos simultaneamente. Logo, gerou-se 3 novas imagens para cada original que garantem a invariância ao espelhamento. A variação no brilho é escolhida aleatoriamente dentro do intervalo entre -30% a 30% para que o modelo fique invariante a diferenças no nível de brilho das imagens. Gerou-se uma nova imagem por imagem original. Por fim, realizou-se uma combinação randômica de recorte (entre 0% a 20%), saturação (-25% a 25%), borramento (até 3 *pixels*) e ruído (até 5% dos *pixels*). O valor de cada parâmetro da combinação é escolhido aleatoriamente dentro do intervalo destacado e foram gerada três cópias para cada original.

Algumas imagens de exemplo podem ser vistas na Figura 3.7.

Cópias Geradas por Imagem					
Original	Rotação	Flip	Brilho	Combinação	Total
1	36	3	1	3	44

Tabela 3.3: Imagens geradas pelo Data Augmentation

Figura 3.7: Exemplos de imagens do Data Augmentation

3.4 Separação dos *datasets*

Para o treinamento e posterior avaliação do modelo, deve-se estabelecer o percentual de dados usado no conjunto de treino. Tendo em conta os principais trabalhos presentes na literatura, conforme se observa na Seção 2, definiu-se a razão 0.7:0.3 para treino e teste.

Base	Procedimento	Dataset	Treino	Teste
Podócito	Modelo Final	Original	85	37
		Aumentado	3740	37
	Validação Cruzada	Original	68	17
		Aumentado	2992	748
Núcleo	Modelo Final	Original	69	30
		Aumentado	3036	30
	Validação Cruzada	Original	55	14
		Aumentado	2420	616

Tabela 3.4: Separação dos *datasets*

A Tabela 3.4 mostra a divisão das imagens para cada base de dados e para cada configuração de treinamento. Os dados de validação cruzada, método a ser explicado a diante, se referem a cada *split*. Deve se ter em conta que, no caso da base de núcleos, os valores da validação cruzada são os mais frequentes, tendo em vista que a divisão do conjunto de treino (n=69) em cinco subconjuntos não é exata. Por exemplo, no conjunto original, há um *fold* com 13 imagens e quatro com 14.

Uma consideração importante é a forma como é feita a divisão das imagens com *data augmentation* entre os conjuntos de treino e de teste. Ela é feita da seguinte maneira: primeiramente, o conjunto original é dividido entre treino e teste, nas proporções de 70% e 30%, respectivamente. Em seguida, são geradas cópias para cada imagem do conjunto de treino, resultando em um aumento de 44 vezes (a imagem original e as 43 geradas por procedimentos diversos). Como um exemplo, pode-se tomar a base de podócitos anotados, em que as 122 imagens são divididas entre teste (n=37) e treino, e estas são então ampliadas (n=3740), conforme mostra a Tabela 3.4. O mesmo procedimento vale para a base anotada de núcleos. O conjunto de teste no qual se avaliam as redes treinadas em ambos conjuntos de dados, original e ampliado, é idêntico nas duas situações.

3.5 Treinamento das Redes Neurais Convolucionais

Em uma etapa inicial, são feitos alguns testes preliminares para a definição dos hiper-parâmetros básicos a serem utilizados nas diversas configurações de treino, que serão detalhadas ainda nesta seção. Neste procedimento inicial, variam-se *batch-size* e número de épocas para cada treinamento utilizando a arquitetura YOLO. Com isto, objetiva-se escolher valores apropriados à nossa aplicação, tendo em vista bons resultados, por um lado, e as limitações de *hardware*, por outro. Nesse sentido, avalia-se a pertinência de aumentar o número de épocas quando há pouca ou nenhuma nos resultados de saída da rede, uma vez que, deste aumento, decorre maior tempo computacional. Os valores estabelecidos para os principais hiper-parâmetros utilizados em cada configuração de treinamento são mostrados na Tabela 3.5.

Nome	Valor ou Definição
Batch Size	4
Número de Épocas	200
Learning Rate	0.01
Momentum	0.937
Otimizador	SGD
Cálculo de Loss	Entropia cruzada

Tabela 3.5: Principais definições adotadas

O treinamento das redes consistirá em dois procedimentos, cada qual com seu objetivo específico: a validação cruzada e a geração dos modelos finais. No segundo, as métricas são analisadas a fim de avaliar cada configuração e compará-las entre si. Além disso, as redes são testadas no conjunto definido correspondente a 30% do total, conforme explicado na Seção 3.4. No primeiro procedimento, por outro lado, a rede é avaliada *única e exclusivamente* no conjunto de treino. Isto se deve ao fato de que o objetivo da validação cruzada é a avaliação da capacidade de generalização dos nossos modelos. Este método é utilizado também para otimização de hiper-parâmetros [52], o que não será feito neste trabalho, uma vez que eles são fixados em todas as configurações de treino.

Há diversos métodos de validação cruzada descritos na literatura. Neste trabalho, será realizada a validação *5-fold*, ilustrada na Figura 3.8. O procedimento consiste em dividir o conjunto de treino em cinco subconjuntos disjuntos de tamanho aproximadamente igual - idealmente, teriam o mesmo tamanho. Em seguida, um destes subconjuntos é escolhido para validação, enquanto que os quatro restantes são fixados no conjunto de treino. A rede é então treinada, validada e tem suas métricas extraídas. O procedimento é então repetido quatro vezes, até que todos os subconjuntos tenham sido fixados como validação, varrendo todas as possibilidades de atribuição de um subconjunto ao conjunto de validação, sem que haja repetição de algum deles. Como o conjunto de treino corresponde a 70% do total, cada subconjunto (*fold*) corresponderá a, aproximadamente, 14% do total, o que pode ser conferido na Tabela 3.4

As configurações de treino são mostradas na Tabela 3.6. Ela mostra oito configurações, as

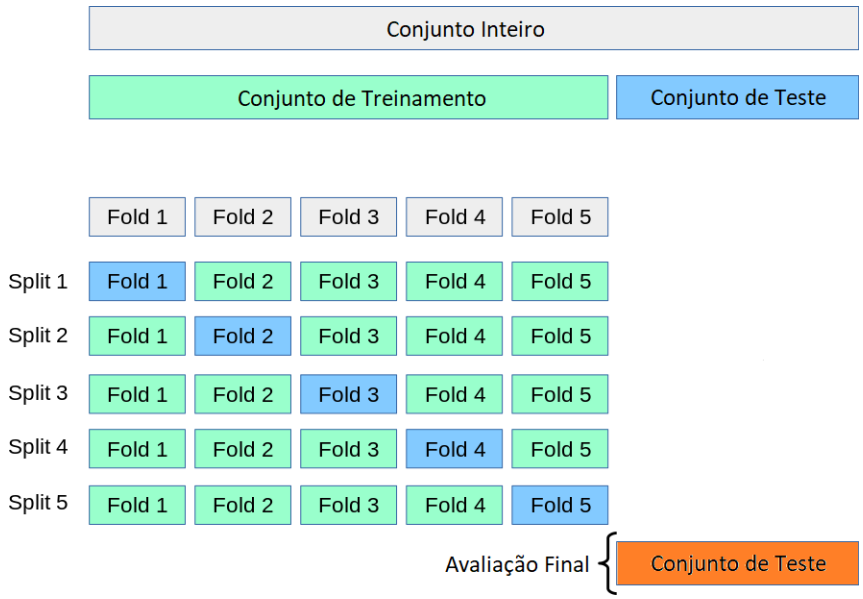


Figura 3.8: Ilustração da validação cruzada *5-fold*. Adaptado de: `https://scikit-learn.org/stable/modules/cross_validation.html`

Configuração	Arquitetura	Pré-Treino
# 1A	YOLO 5S	Nenhum
# 2A	YOLO 5M	Nenhum
# 3A	YOLO 5L	Nenhum
# 4A	YOLO 5X	Nenhum
# 1B	YOLO 5S	COCO
# 2B	YOLO 5M	COCO
# 3B	YOLO 5L	COCO
# 4B	YOLO 5X	COCO

Tabela 3.6: Configurações de treinamento

quais serão aplicadas aos dois conjuntos (original e ampliado) das duas bases de dados anotadas (núcleos e podócitos), resultando em 32 modelos gerados. Em algumas configurações, emprega-se *transfer learning*, de modo a aproveitar os pesos gerados por uma rede treinada em um conjunto de dados muito maior, que pode ou não conter a classe de objetos que se queira detectar. Neste trabalho, o pré-treino é realizado na base de dados de detecção de objetos COCO[53], que contém mais de 1,5 milhão de objetos de 80 classes em mais de 300 mil imagens.

3.6 Avaliação dos modelos

Com o treinamento encerrado, após as 200 épocas, algumas métricas são extraídas. Neste sentido, vale definir alguns conceitos principais, das quais as métricas derivam:

- Verdadeiros positivos (TP): ocorrem quando a detecção da rede coincide com o objeto;

- Falsos positivos (FP): ocorrem quando a detecção da rede não coincide com o objeto;
- Falsos Negativos (FN): ocorrem quando um objeto não é detectado pela rede;

Ressalta-se que a categoria verdadeiro negativo não se aplica à tarefa em questão, de detecção em uma única classe. Com estes conceitos, utilizamos as métricas de Precisão (P), *Recall* (R) e F1-*score* (F1), definidas nas Equações 3.4-3.6:

$$P = \frac{TP}{TP + FP} \tag{3.4}$$

$$R = \frac{TP}{TP + FN} \tag{3.5}$$

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{3.6}$$

Duas métricas adicionais derivam de uma curva comumente utilizada na detecção de objetos, a de precisão-*recall*. Um bom detector de objetos deve ser capaz de manter a precisão alta à medida que o o *recall* aumenta[54]. Desse modo, o limiar de confiança pode ser aumentado sem que haja queda na performance para essas duas métricas. Um ponto importante é que o número de falsos positivos, portanto, tende a aumentar com a diminuição do limiar de confiança, o que diminui a precisão. Quando se aumenta o limiar de confiança, por outro lado, o número de falsos negativos tende a aumentar, o que diminui o *recall*. O primeiro passo para a construção da curva de precisão-*recall* consiste em computar cada detecção como verdadeiro positivo ou falso positivo e, em seguida, calcular os valores de P e R acumulados, até que se chegue a última detecção. No entanto, esse cálculo acumulado é feito a partir da ordenação decrescente do nível de confiança, razão pela qual a curva precisão-*recall* tende a decair.

A partir dos valores de precisão e *recall* acumulados, pode-se fazer uma interpolação entre todos os pontos experimentais, ou seja, as detecções da rede para uma mesma classe. A interpolação deste pontos gera a curva precisão-*recall*. Tomando-se a área abaixo da curva, obtém-se a métrica de *Average Precision* (AP). Quando há mais de uma classe a ser detectada, é comum que se utilize como métrica a média entre os valores de AP para cada classe, denominada *mean Average Precision* (mAP). Quando há somente uma classe de na tarefa de detecção, estas métricas são redundantes. Portanto, chamaremos doravante a área sob a curva precisão-*recall* de AP.

É importante destacar que cada uma das métricas supracitadas depende, direta ou indiretamente, de como são definidos cada um dos conceitos apresentados no início desta seção. No caso do falsos negativos, a definição é natural: o objeto não foi identificado pela rede neural. Cabe perguntar, entretanto, como se atribui, dada uma detecção da rede, a condição de falso positivo, quando ela é incorreta, ou de verdadeiro positivo, quando ela é correta. Esta decisão é feita com base no índice de *Jaccard*. Ele é definido como a razão entre a a interseção e a união da detecção da rede com a anotação de referência (*ground-truth box*), conforme mostra a Figura 3.9. A Equação 3.7 mostra a definição matemática do índice de Jaccard, considerando um polígono de detecção

B_d e um polígono de anotação B_a. Quando este valor está acima de um certo limiar pré-definido, considera-se a detecção como um verdadeiro positivo. Caso contrário, como um falso positivo. No campo de detecção de objetos, o índice de *Jaccard* é também conhecido como *Intersection over Union* (IoU). O limiar de IoU definido em nossas detecções foi de 0,5. Isto significa que nosso detector identifica corretamente um objeto quando IoU é maior que 0,5.

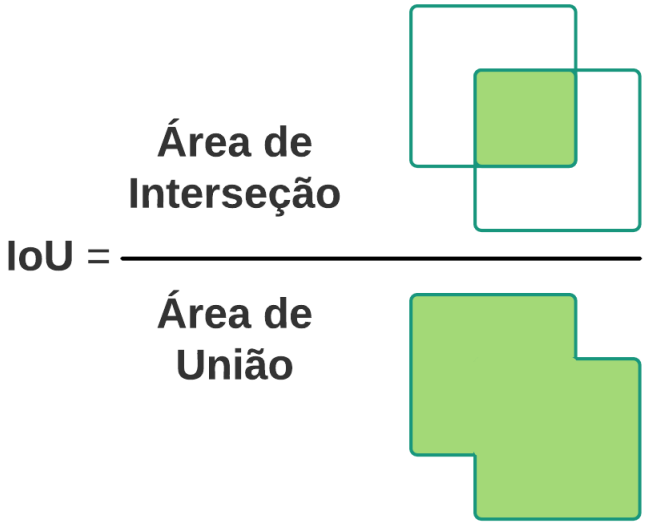


Figura 3.9: *Intersection over Union*

$$J = \frac{B_d \cap B_a}{B_d \cup B_a} \tag{3.7}$$

Definidas as métricas, passamos a discutir como definir o melhor resultado de uma época. Em primeiro lugar, deve-se esclarecer que o comportamento de aprendizado da rede pode ser não-monotônico, no sentido de nem sempre melhorar suas predições ao longo do treinamento. Portanto, não se pode garantir que os pesos gerados após a última época serão os melhores. Tendo em vista a possibilidade de exportar os melhores pesos ao fim do treinamento, armazenaremos, juntamente com os pesos da última época, aqueles que tiverem tido o melhor desempenho até então. Esta definição leva a um segundo questionamento: qual critério utilizar para escolher a melhor época. Para tanto, definimos inicialmente uma nova métrica, que leva em conta diversos limiares de IoU. Isto se deve ao fato de que é importante constatar se o modelo gerado é capaz de identificar os objetos de interesse com restrições maiores, isto é, se ele é capaz de fornecer detecções mais bem ajustadas. A Equação 3.8 descreve a média entre 10 valores de AP, com limiar de IoU entre 0.5 e 0.95. Cada elemento da sequência de valores AP_i é definido como o valor de AP para o i-ésimo limiar. Os limiares L_i são definidos na Equação 3.9.

$$AP^* = \frac{1}{10} \sum_{n=1}^{10} AP_i \tag{3.8}$$

$$L_i = 0.5 \cdot (1 + 0.1 \cdot i), i \in [0, 9] \cap \mathbb{Z} \tag{3.9}$$

Estas equações permitem definir o critério $\overline{AP}$ de escolha do melhor peso gerado durante o treinamento. Ele consiste em uma média ponderada entre AP e AP^*, conforme mostra a Equação 3.10

$$\overline{AP} = \langle\ (AP, AP^*), (0.1, 0.9)\rangle \tag{3.10}$$

Por fim, será feita uma comparação visual entre as marcações da rede e as anotações dos patologistas. Isto servirá, além de um auxílio visual para avaliar a rede, para identificar possíveis vícios, conforme discutido em um dos trabalhos referidos no Capítulo 2. Isto pode servir, por exemplo, para identificar se as marcações da rede são sempre, ou quase sempre, maiores do que os *ground-truth boxes*. Além disso, este procedimento pode ajudar a constatar se existem padrões nos falsos positivos e falsos negativos, de modo que se possa corrigir a rede futuramente.

Capítulo 4

Resultados

Este capitulo se dedica à exposição dos resultados, bem como à análise destes. Em primeiro lugar, segmentamos a análise para cada base de dados anotada, cada qual correspondendo a uma seção deste capítulo. Além disso, subdividimos cada seção por cenários, em que cada um destes corresponde ao conjunto de dados utilizado, isto é, se original ou aumentado. Tendo em vista a quantidade de modelos criados, mostraremos, em alguns casos, figuras destacando o melhor e o pior modelo de cada cenário. Contudo, informações mais detalhadas de cada modelo poderão ser encontradas nos materiais suplementares do Capítulo 6.

4.1 Base anotada de núcleos

4.1.1 Cenário 1: conjunto de imagens original

Neste cenário, são avaliados os resultados do conjunto original de núcleos, conforme nomenclatura definida na Seção 3.4. O valor de AP de cada *split* da validação cruzada é mostrado na Figura 4.1. Os valores de média dos *splits* podem ser conferidos no Capítulo 6. Conforme explicado na Seção 3.5, o método de validação cruzada foi aplicado única e exclusivamente ao conjunto de treino, de modo a se avaliar a capacidade de generalização dos modelos. Observa-se que houve apenas uma ocorrência, na versão S, em que una rede sem pré-treino obteve AP superior a algum *split* da rede pré-treinada. Além disso, observa-se que a média dos valores de AP teve o maior resultado para a configuração 4B.

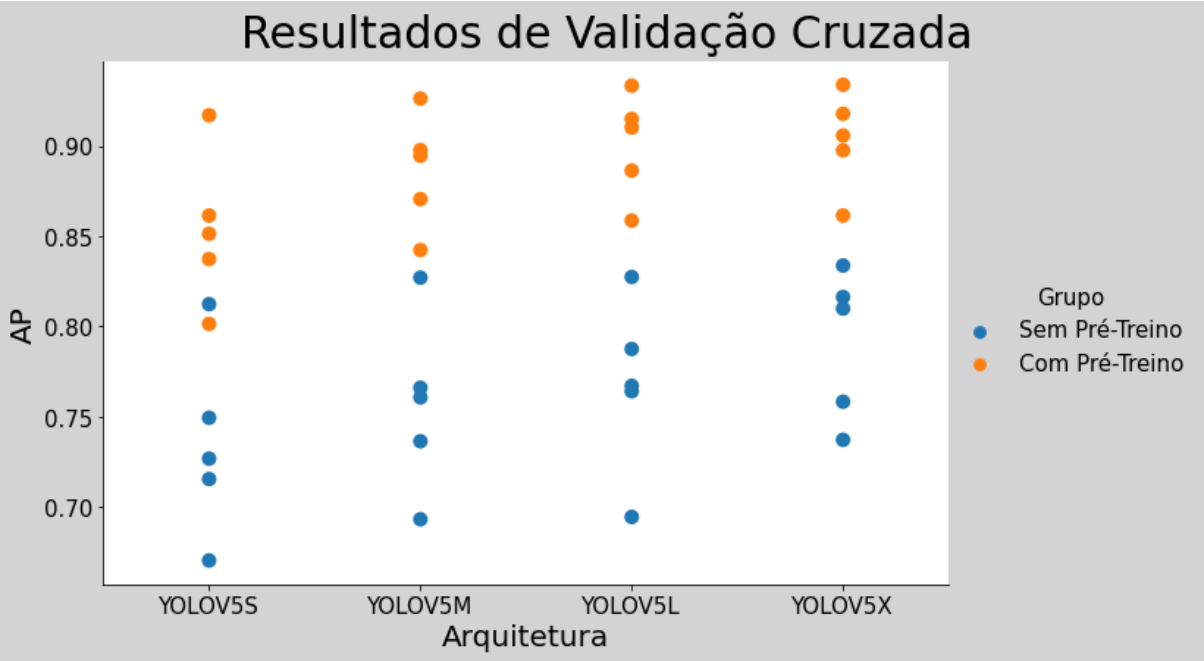


Figura 4.1: Gráfico de dispersão dos valores de AP de cada split da validação cruzada no conjunto original separados por configuração

As Figuras 4.2 e 4.3 mostram as curvas de *loss* de treino e teste para a melhor (3B) e a pior (2A) redes, respectivamente. Os critérios de definição do desempenho das redes, para posterior classificação em melhor ou pior, podem ser conferidos na Seção 3.6. Em ambas as redes, observa-se um comportamento não-monotônico, tanto para as curvas de treino, quanto para as de teste. Nas curvas 3B, observa-se que o valor de teste se mantém abaixo do de treino em todas as épocas. Nas curvas 2A, por outro lado, essa tendência se mantém, mas com seis ocorrências de valor de teste acima do de treino, em especial entre as épocas 52-56. Nas curvas 3B, a de teste atinge seu ponto de mínimo da curva de teste ocorra logo nas primeiras épocas (n=9). O mesmo ocorre com a curva de treino: o ponto de mínimo ocorre na época 12. Com relação à configuração 2A, observa-se que os pontos de mínimo das curvas de treino e de teste, respectivamente, ocorrem na mesma época, 22. Neste sentido, dadas as características de cada uma das curvas, como a não-monotonicidade e os valores de mínimo, ressalta-se a importância de armazenar os pesos da melhor época, apesar de as curvas voltarem a diminuir após atingir o valor de mínimo. Além disso, deve-se esperar que o comportamento oscilatório das curvas de teste também se reflitam nas curvas das métricas extraídas.

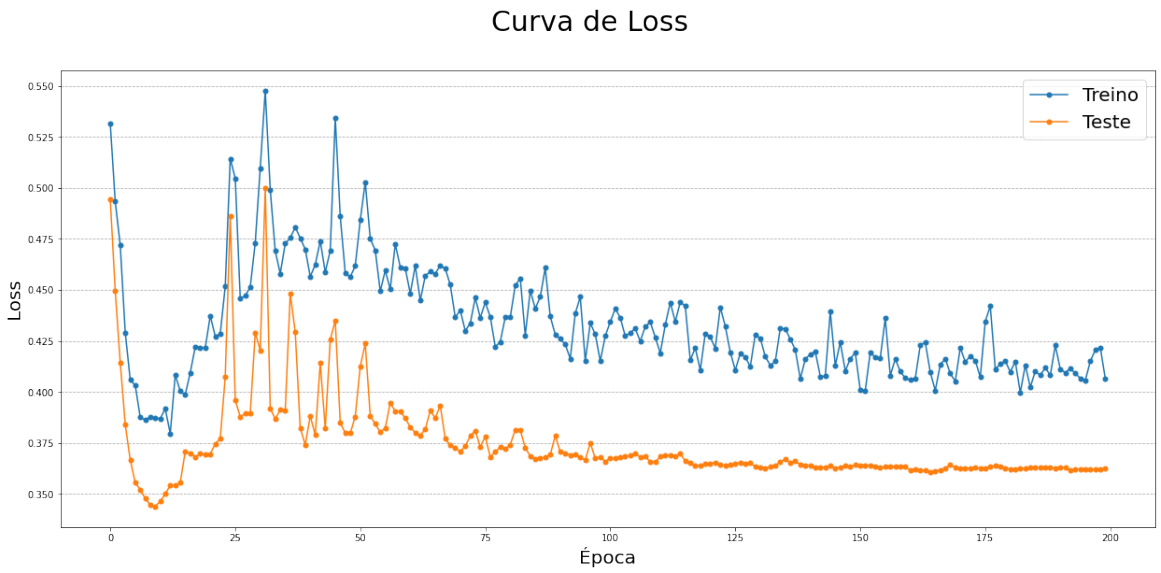


Figura 4.2: Curva de Loss para a rede de melhor desempenho

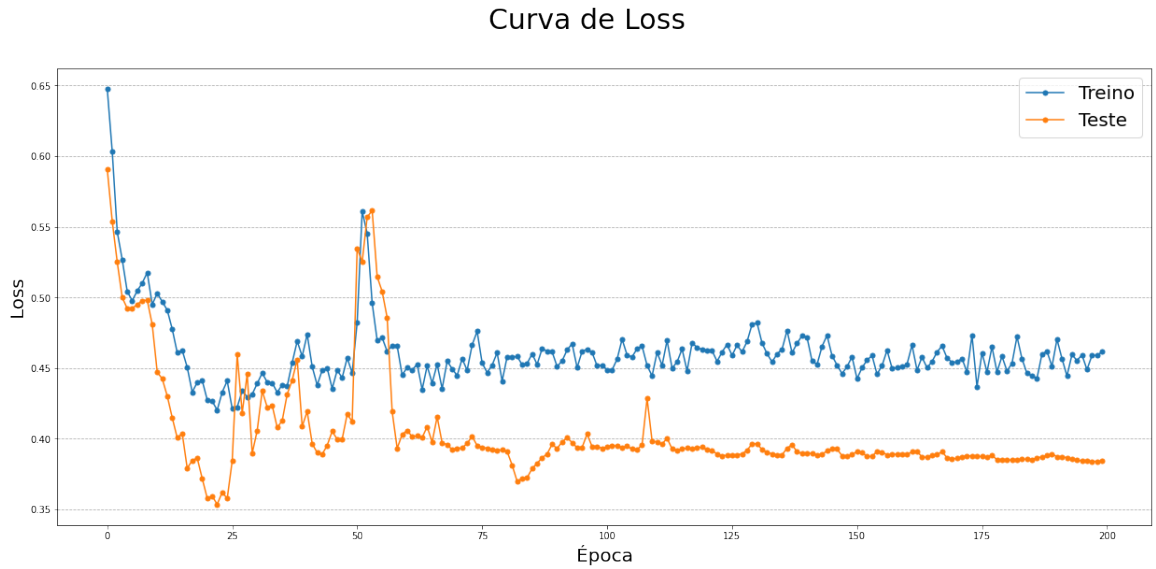


Figura 4.3: Curva de Loss para a rede de pior desempenho

A Figura 4.4 mostra a curva de AP obtida no conjunto de teste ao longo das 200 épocas de treinamento para a melhor e a pior redes. Em ambos os casos, observa-se um comportamento não-monotônico. A rede mais profunda tem desempenho superior logo nas primeiras épocas, como observamos graficamente. Além disso, observa-se que há grande oscilação no comportamento da melhor rede, em especial nas épocas anteriores a cem. De fato, na época 42, e somente nela, a rede mais rasa teve valor de AP superior. Em ambos os casos, o valor máximo é atingido no último quarto do treinamento: nas épocas 155, no caso da rede mais profunda e 199, no caso da mais rasa.

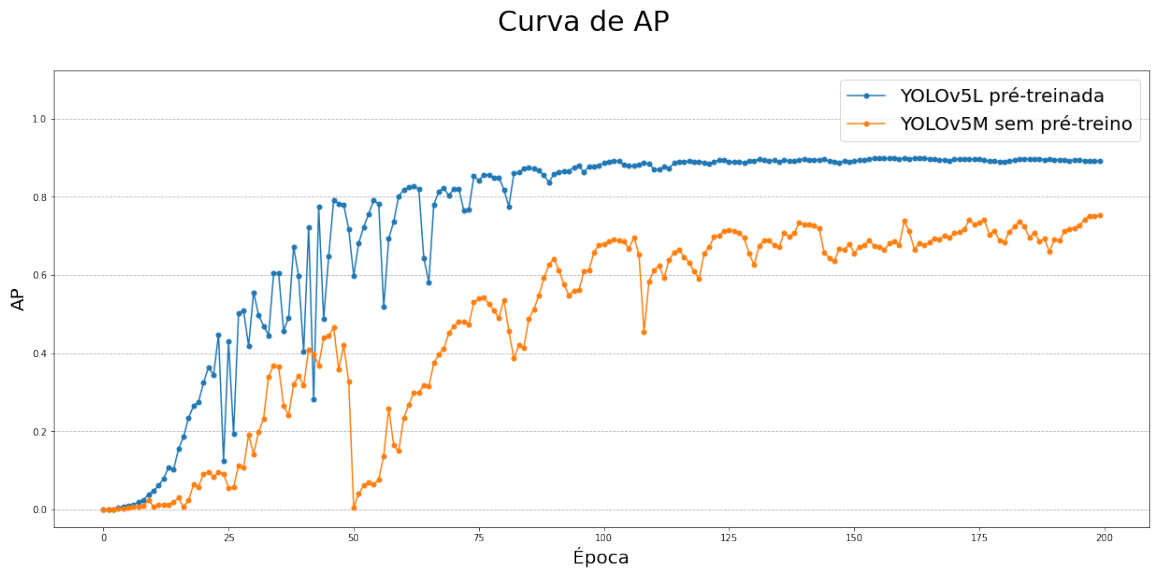


Figura 4.4: Curvas de AP da melhor e da pior redes

Uma vez extraídos os pesos da melhor (3B) e da pior (2A) redes, escolheu-se uma imagem padrão para realizar a detecção sobre ela. Esta imagem é a mesma que será apresenta no cenário 2 desta seção. Ela é mostrada na Figura 4.5. Os contornos em verde representam as anotações (*ground-truth*). Os contornos em vermelho representam as detecções da rede. Esta convenção é adotada em todas os comparativos desse tipo. O objetivo de mostrar as detecções não é fornecer métricas para rede, o que se faz na Tabela 4.1, apenas ilustrar o comportamento de detecção. Um fato que se destaca é que a pior rede parece ter gerado menos falsos positivos do que a de melhor desempenho nas regiões interna e externa do glomérulo. Nesse sentido, deve-se ter em conta a possibilidade de que o melhor desempenho global no conjunto de teste não se reflita nesta imagem em específico. De fato, ao se fazer uma inferência com as redes nesta imagem, observou-se que a 1B apresentou maior *recall* e AP, além de uma precisão próxima à obtida pela 3B (0,882 contra 0,758).

Figura 4.5: Comparação das detecções da melhor e da pior redes em uma imagem padrão

Outra inspeção visual pode ser feita a partir da Figura 4.6. Com os pesos da melhor rede, realiza-se uma inferência sobre cada imagem do conjunto de teste individualmente. As imagens mostradas se referem àquelas de melhor e pior AP. Observa-se que a imagem de pior resultado contém grande número de objetos a serem detectados, em comparação com a outra, o que por si só pode ser um fator que impacta a qualidade da detecção. O tamanho dos objetos detectados também é contrastante. Além disso, são visivelmente perceptíveis falsos negativos na região que delimita o glomérulo.

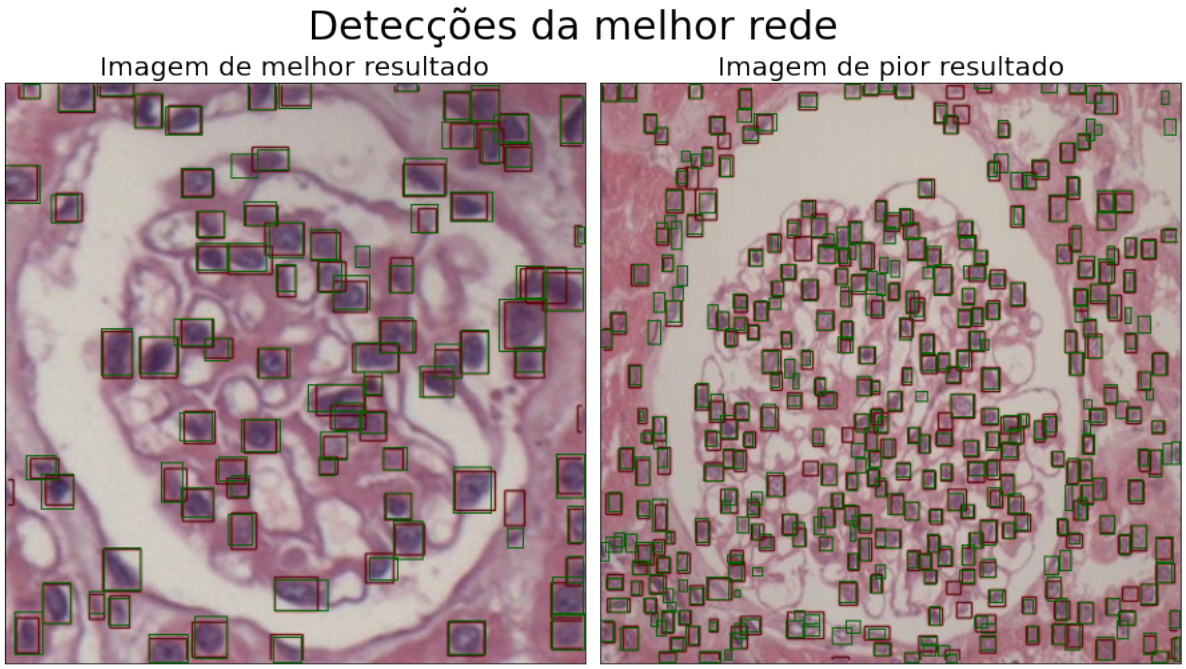


Figura 4.6: Comparação entre as imagens de melhor e de pior resultado para a rede de melhor desempenho

A Tabela 4.1 mostra as métricas obtidas no conjunto de teste para cada uma das configurações definidas na Seção 3.5. A coluna época se refere à época de melhor resultado durante o treinamento, cujo critério de definição é exposto na Seção 3.6. Deve-se notar, como ocorre neste caso, que nem sempre a melhor rede terá todas as melhores métricas. Por exemplo, a melhor rede, 3B, apresenta o segundo melhor AP. Além disso, constata-se que somente em duas configurações se atingiu o melhor resultado na última época: 1A e 2A. De um modo geral, os melhores pesos foram obtidos nas últimas épocas, visto que em todos os casos isto se deu a partir do último quarto do treinamento (n=150).

Configuração	P	R	AP	F1	$\overline{AP}$	Época
# 1A	0.821	0.755	0.768	0.787	0.333	199
# 2A	0.843	0.727	0.752	0.781	0.319	199
# 3A	0.876	0.761	0.800	0.814	0.352	174
# 4A	0.862	0.783	0.812	0.821	0.371	196
# 1B	0.904	0.818	0.864	0.859	0.416	168
# 2B	0.910	0.855	0.887	0.881	0.440	169
# 3B	0.924	0.863	0.896	0.892	0.462	170
# 4B	0.916	0.873	0.899	0.894	0.462	194

Tabela 4.1: Métricas de teste de cada configuração, obtidas na melhor época durante o treinamento

4.1.2 Cenário 2: conjunto de imagens aumentado

A Figura 4.7 mostra os resultados de AP de validação cruzada. De forma geral, as redes sem pré-treino desempenharam melhor, como se nota na Tabela 6.2. Com efeito, a diferença entre as médias observadas foi pequena: os valores de AP variaram entre 0,922 e 0,931. Observa-se um ponto de mínimo com maior desvio em relação à média para todas as configurações. Em todas elas, o ponto mínimo se refere a um *split* específico (k=3). Nesse sentido, observa-se que o conjunto de treino não está perfeitamente balanceado.

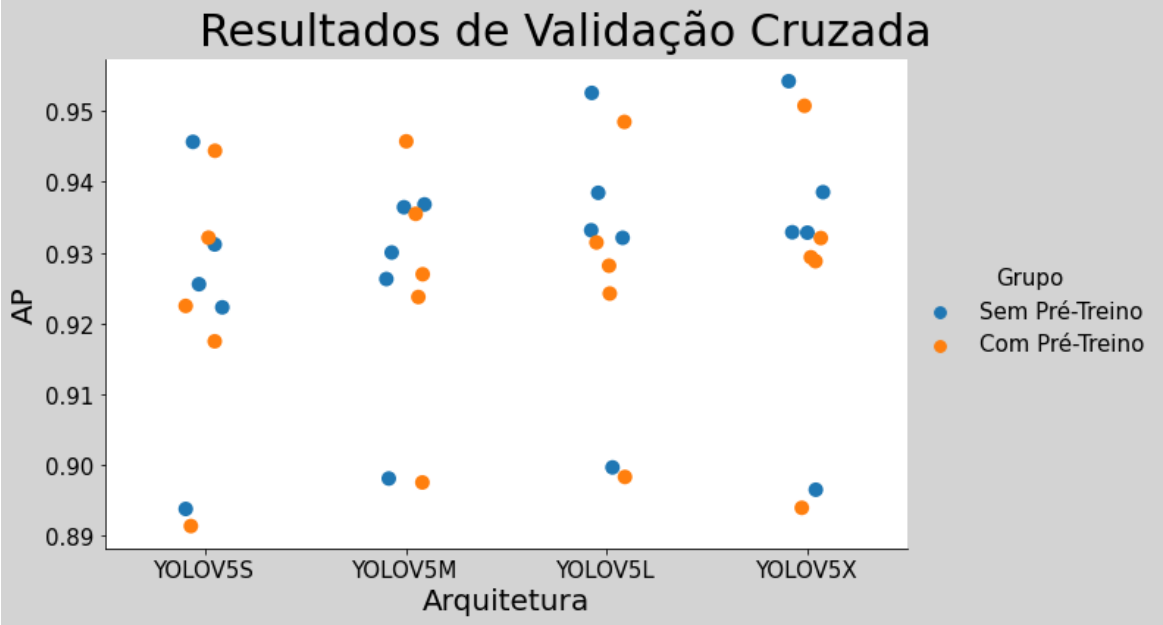


Figura 4.7: Gráfico de dispersão dos valores de AP de cada split da validação cruzada no conjunto original separados por configuração

As Figuras 4.8 e 4.9 mostram os resultados de *loss* para a melhor (4B) e a pior (1A) redes, respectivamente. Em ambos os casos, observa-se um comportamento oscilatório nas curvas. Com respeito à configuração 1A, nota-se que a curva de teste permanece abaixo da de treino, atingindo seu mínimo na época 75. A configuração 4B, entretanto, é marcada por uma grande região de sobreajuste (*overfitting*), em que se observa o valor de *loss* no conjunto de teste aumentar. Isto significa que o treinamento poderia ter sido encerrado antes que se entrasse nessa região, em cerca de 1/8 do treinamento. Este comportamento indesejado é diferente do que se observou no primeiro cenário: enquanto há queda, ainda que não monotônica, na curva de treino, a curva de teste passa a ascender. Objetivamente, em que pese esta configuração tenha sido a melhor, grande parte do treinamento foi feita de maneira desnecessária e, além disso, o modelo treinado passa a ficar sobreajustado ao conjunto de treino, perdendo a capacidade preditiva sobre o conjunto de teste. Esta diferença no sobreajuste observada entre os cenários pode ser explicada pelo efeito de *data augmentation*, no sentido de que o modelo passa a se ajustar às alterações feitas no conjunto de treino, conforme se constata pela curva de *loss*. Em resumo, na configuração 1A, observa-se que a predição no conjunto de teste se deu de forma mais fácil do que a no conjunto de treino,

tendo em vista que esta ficou acima daquela. Na configuração 4B, além da pronunciada região de sobreajuste, a defasagem entre as curvas de treino e teste se torna gradativamente maior, fato que reforça a necessidade de se armazenar os pesos das melhores épocas no decorrer do treinamento.

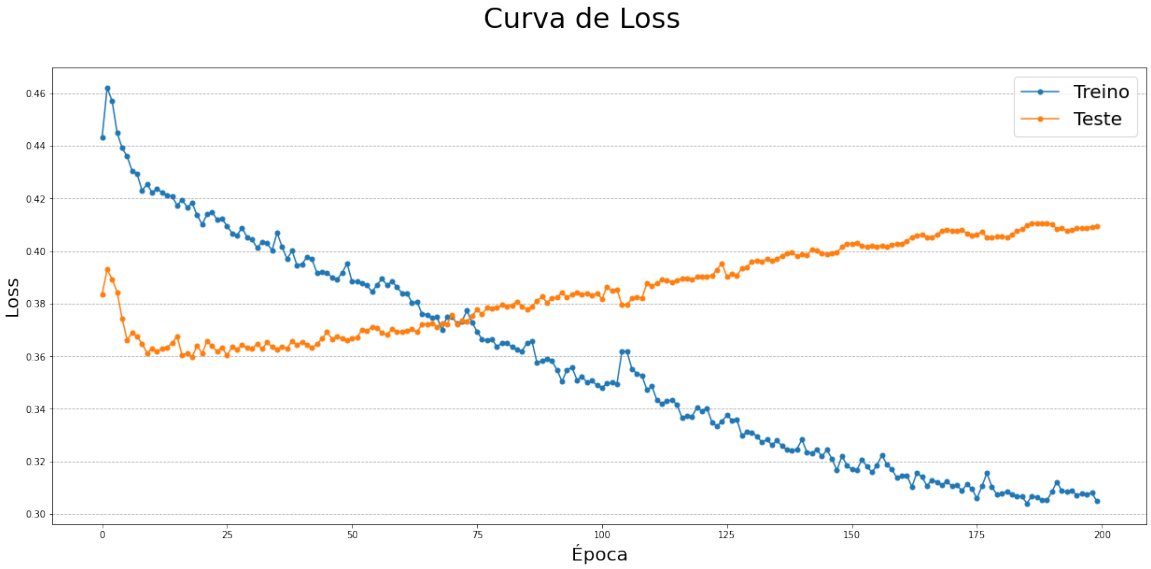


Figura 4.8: Curva de Loss para a rede de melhor desempenho

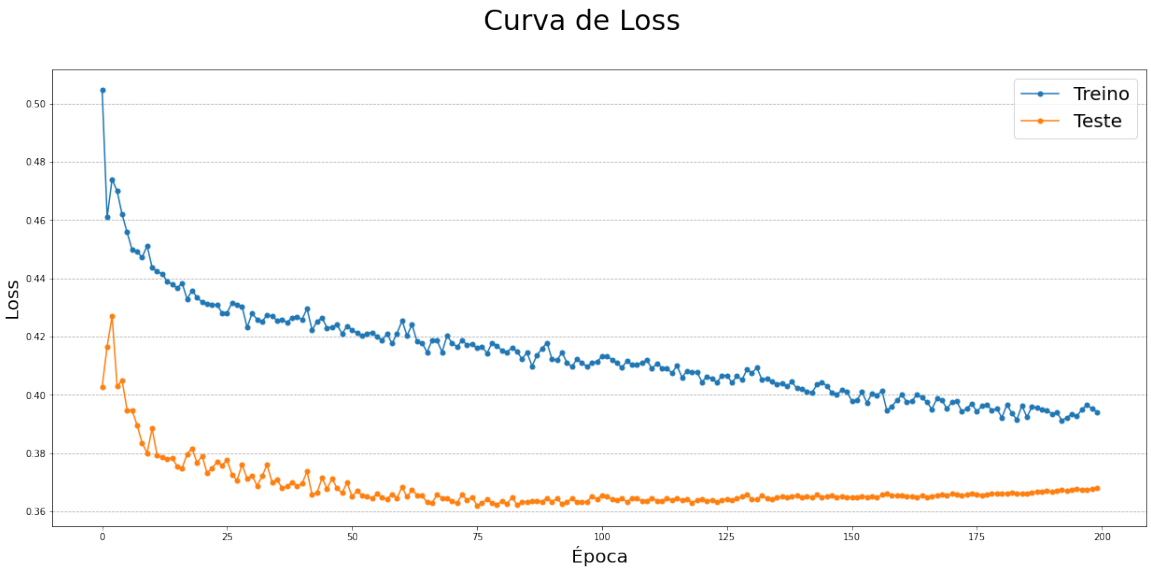


Figura 4.9: Curva de Loss para a rede de pior desempenho

A Figura 4.10 mostra as curvas de AP para a melhor e pior redes. O sobre-ajuste referido na análise de *loss* se reflete de forma nítida no comportamento do AP: apesar de ter obtido resultado superior, a configuração 4B passa a decair e termina o treinamento abaixo da curva 1A. Não obstante, também se observa que a curva 1A sofre uma acomodação após cerca de um quarto do treinamento. Em termos objetivos, o valor de AP é máximo nas épocas 18 (4B) e 125 (1A), resultado que expressa perda da capacidade de aprendizado aludida na análise das curvas de *loss*.

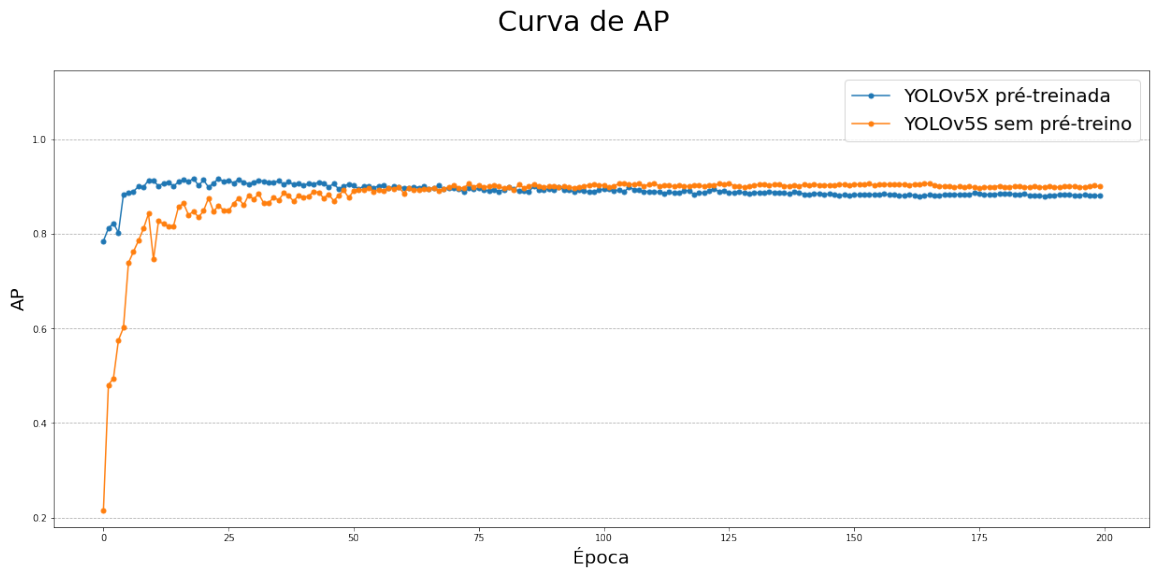


Figura 4.10: Curvas de AP da melhor e da pior redes

As detecções em uma imagem padrão geradas a partir da extração dos pesos das redes 4B e 1A são mostradas na Figura 4.11. Esta imagem é a mesma mostrada no cenário 1. As detecções refletem a pequena diferença de desempenho neste cenário com conjunto de dados aumentados, que pode ser constatada pela Tabela 4.2.

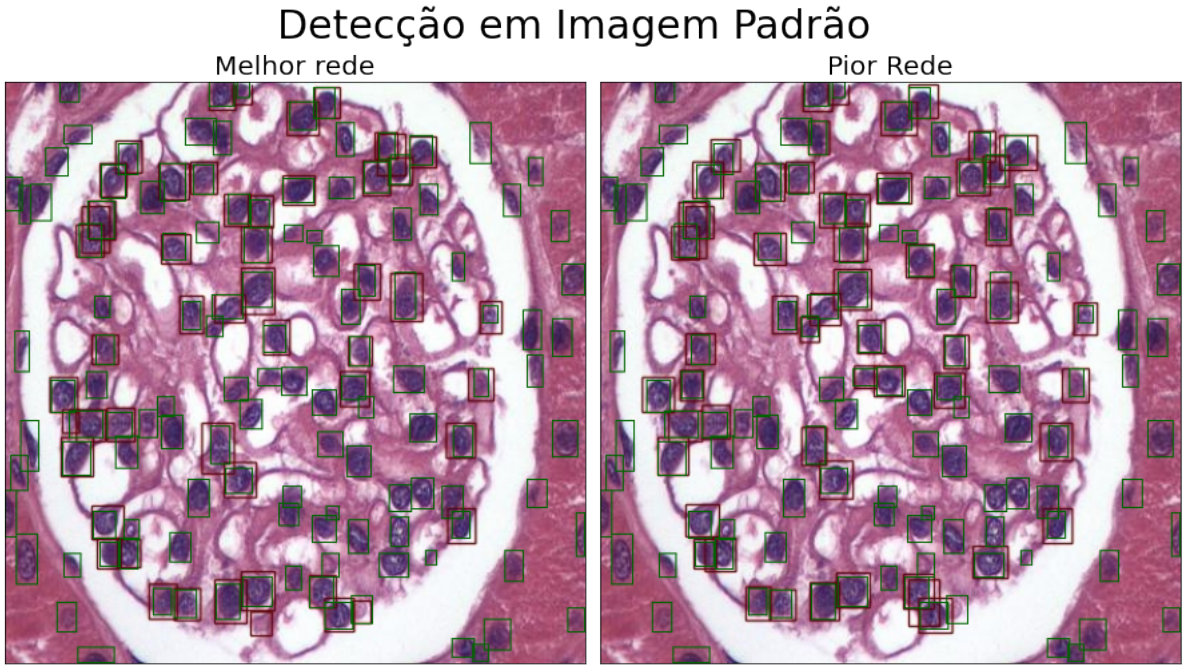


Figura 4.11: Comparação das detecções da melhor e da pior redes em uma imagem padrão

A Figura 4.12 mostra detecções da rede 4B. As imagens mostradas se referem àquelas de melhor e pior AP. As duas imagens diferem daquelas observadas no primeiro cenário. De forma semelhante àquele cenário, observa-se grande número de alvos na imagem de pior resultado.

Figura 4.12: Comparação das detecções da melhor e da pior redes em uma imagem padrão

A Tabela 4.2 mostra os resultados de diversas métricas para o conjunto de núcleos aumentados. A coluna época se refere àquela em que se observou o melhor resultado segundo os critérios definidos na Seção 3.6. Novamente, assim como se observou no primeiro cenário, o emprego de redes pré-treinadas precipitou a ocorrência da época de melhor resultado: na maior parte dos casos, ela ocorreu antes do primeiro quarto do treinamento quando havia pré treino, em contraste com a ocorrência após o segundo quarto do treinamento nos outros casos. Um efeito deste fato também é constatado pela análise das curvas de *loss* da melhor e da pior redes, como feito anteriormente.

Configuração	P	R	AP	F1	$\overline{AP}$	Época
# 1A	0.915	0.887	0.906	0.901	0.463	165
# 2A	0.913	0.891	0.908	0.902	0.469	115
# 3A	0.914	0.895	0.912	0.904	0.476	128
# 4A	0.924	0.889	0.915	0.907	0.478	83
# 1B	0.916	0.886	0.904	0.901	0.466	42
# 2B	0.918	0.889	0.906	0.903	0.473	55
# 3B	0.929	0.893	0.918	0.911	0.475	23
# 4B	0.923	0.894	0.916	0.908	0.481	23

Tabela 4.2: Métricas de teste de cada configuração, obtidas na melhor época durante o treinamento

4.2 Base anotada de podócitos

4.2.1 Cenário 1: conjunto de imagens original

Neste cenário, são avaliados os resultados do conjunto original de podócitos. A Figura 4.13 mostra os resultados de AP obtidos aplicando-se o método de validação cruzada *5-fold* ao conjunto de treino. Tomando-se um recorte por arquitetura, observa-se que todos os *splits* do grupo com pré-treino geraram resultados superiores aos do grupo sem pré-treino. Além disso, é possível notar que há discrepância maior em um dos *splits* de todas as oito configurações. Por meio de inspeção, observou-se que todos os casos se referem a um mesmo *split* (k=3), o que sugere um desbalanceamento no conjunto de treino. Em termos do valor médio de AP, constata-se que a rede 4B obteve o melhor resultado.

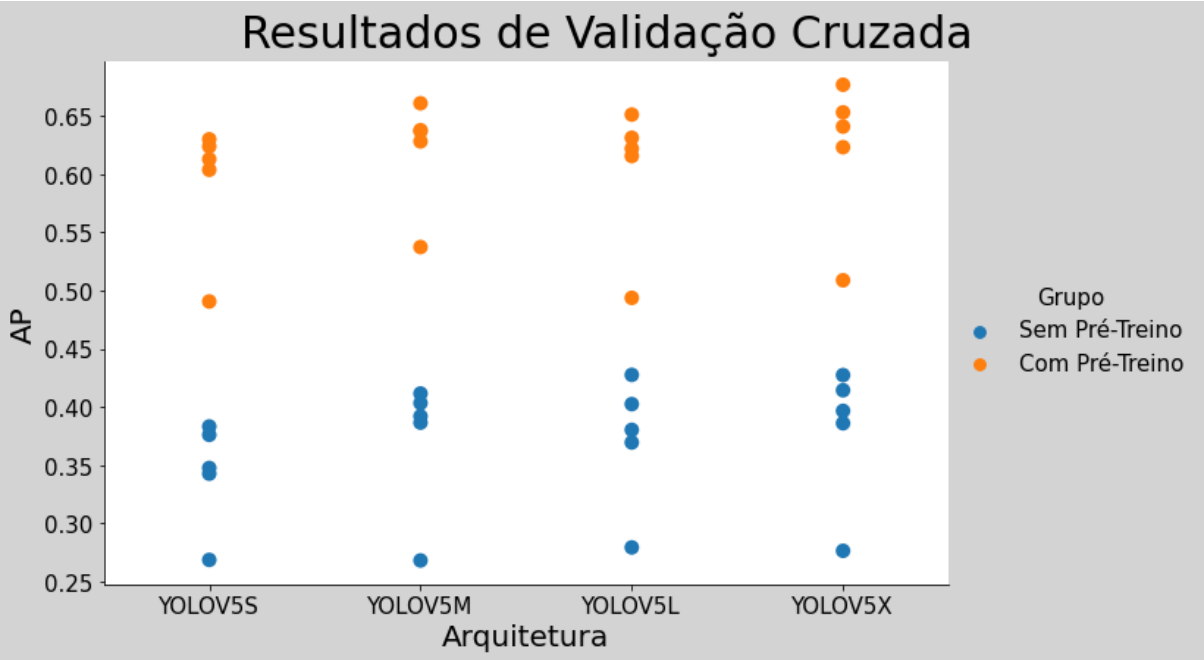


Figura 4.13: Gráfico de dispersão dos valores de AP de cada split da validação cruzada no conjunto original separados por configuração

As Figuras 4.14 e 4.15 mostram as curvas de *loss* de treino e teste para a melhor (2B) e a pior (1A) redes, respectivamente. Novamente, em todos os casos se observa um comportamento oscilatório, em maior ou menor grau. Em relação à configuração 2B, destaca-se o fato de ocorrer sobreajuste. Nesse sentido, deve-se ressaltar a importância de se armazenar os pesos ao longo do treinamento, em vez de fazê-lo somente ao fim da última época. Além disso, a partir da época 101, a curva de teste se mantém acima da de treino e atinge seu mínimo na época 98. Em contraste com a configuração 2B, a 1A tem sua curva de teste acima da de treino em somente dez épocas. Portanto, não se observa sobreajuste, apenas pontos isolados em que o valor de teste fica acima. Desta forma, o ponto de mínimo de ambas as curvas, apesar do comportamento oscilatório, ocorre nas últimas épocas do treinamento, 168 e 194, para treino e teste, respectivamente.

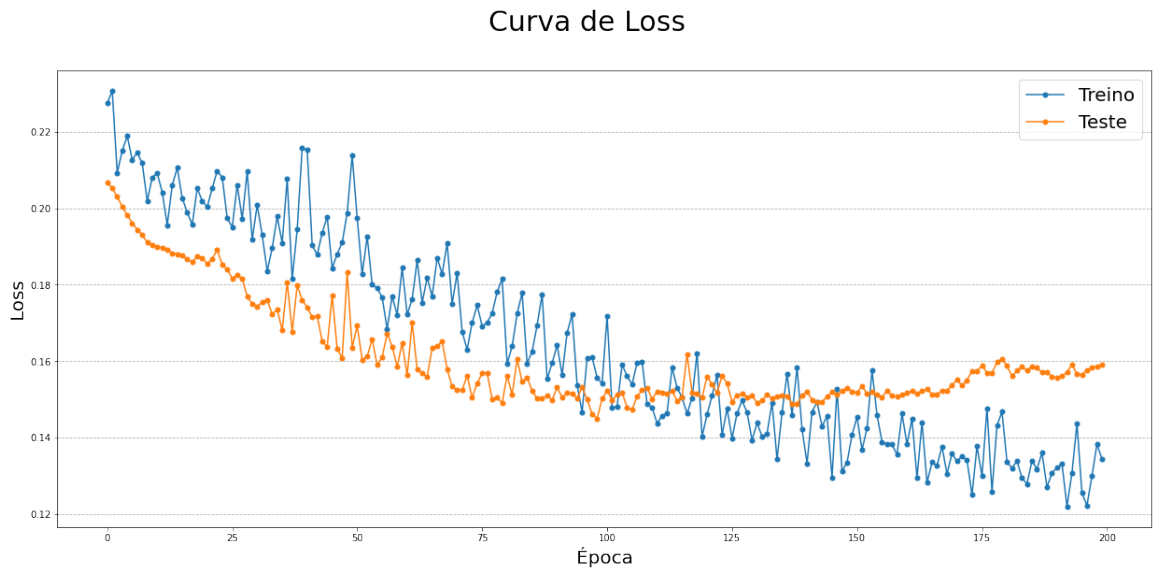


Figura 4.14: Curva de Loss para a rede de melhor desempenho

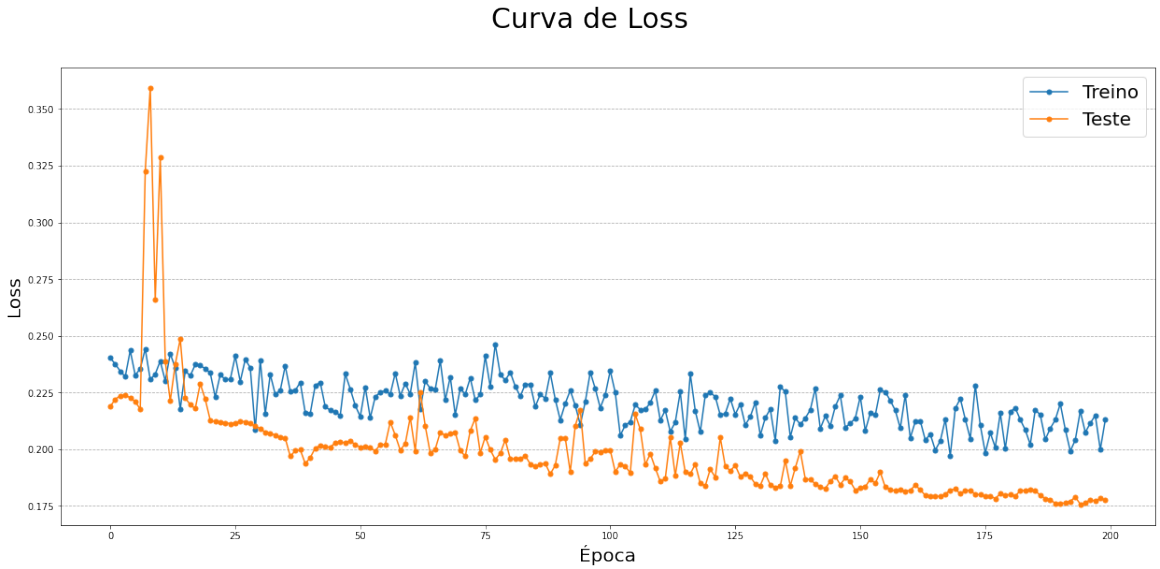


Figura 4.15: Curva de Loss para a rede de pior desempenho

A Figura 4.16 mostra a curva de AP obtida no conjunto de teste ao longo das 200 épocas de treinamento para a melhor e a pior redes. Em ambos os casos, observa-se um comportamento não-monotônico, em que pese a curva da melhor rede tenda a se assentar durante o treinamento (visualmente, observa-se que isso ocorre após n=100). Além disso, a melhor rede, mais profunda, tem resultados significativamente superiores logo nas primeiras épocas. Uma forma de constatar isto é tomar a razão entre o valor de AP da melhor e da pior época. Excluindo-se as divisões por zeros, que indicam as épocas em que a pior rede teve AP=0, esta razão atinge valor máximo logo começo do treinamento (n=10). Outro ponto que vale destacar é que os valores máximos em cada curva não são atingidos na última época, sinal do comportamento oscilante de aprendizado das redes. Com efeito, o valor máximo de AP em cada uma delas é atingido em épocas bastantes

distintas: na época n=98 para a melhor rede e na época n=180 para a pior rede.

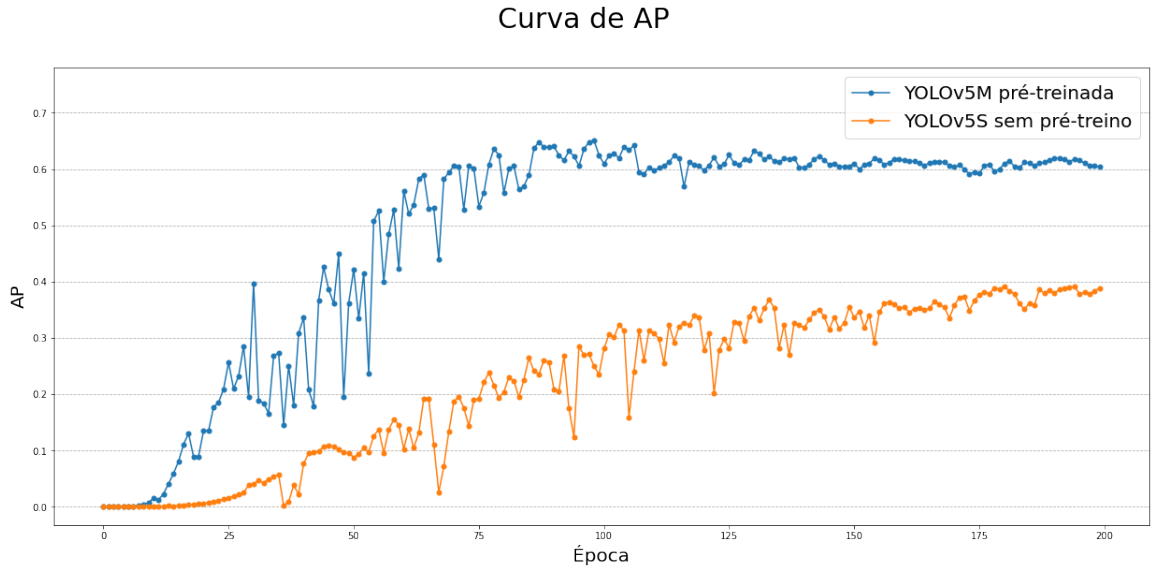


Figura 4.16: Curvas de AP da melhor e da pior redes

A Figura 4.17 mostra as detecções geradas a partir dos melhores pesos da melhor e pior redes em uma imagem padrão, que também será utilizada no cenário 2. De especial, nota-se a diferença entre falsos negativos: cinco para a melhor rede, em contraste com os 10 da pior. Além disso, destacam-se os três falsos positivos na região externa do glomérulo, no canto superior direito da imagem, que ocorreram na pior rede, mas foram corretamente ignoradas pela melhor rede.

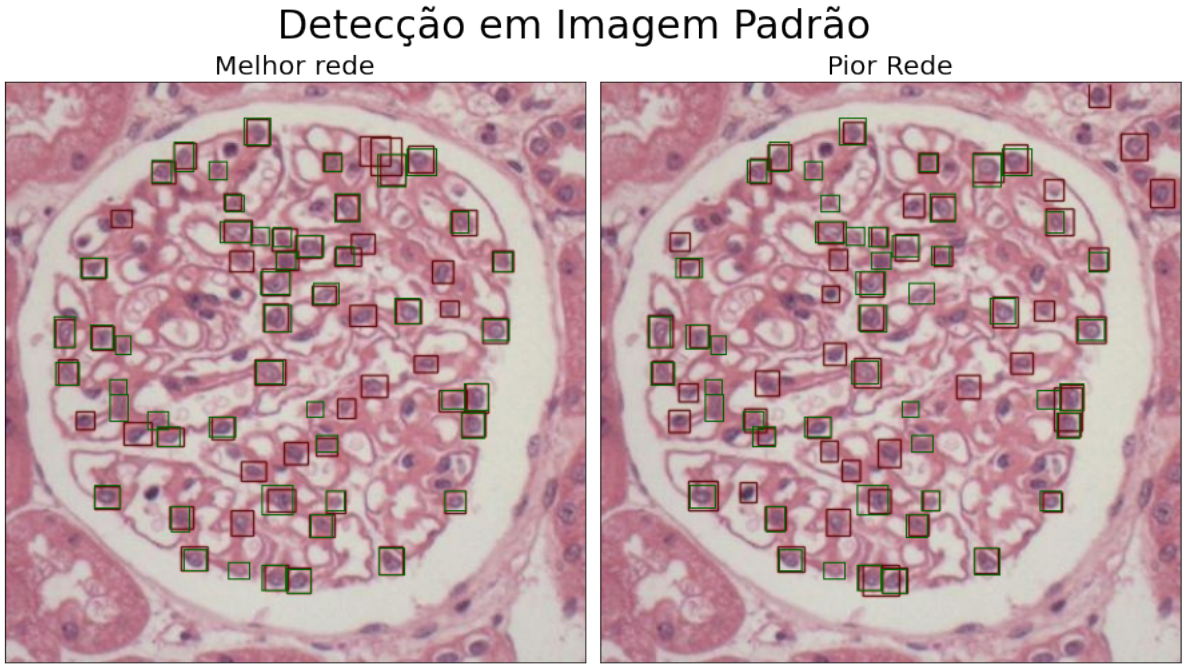


Figura 4.17: Comparação das detecções da melhor e da pior redes em uma imagem padrão

A Figura 4.18 mostra a melhor e a pior detecções da melhor rede lado a lado. Observa-se que a imagem de pior resultado contém grande quantidade de falsos negativos. Além disso, destacam-se a diferença de intensidade dos núcleos em cada imagem, além da direção da inclinação do eixo do glomérulo.

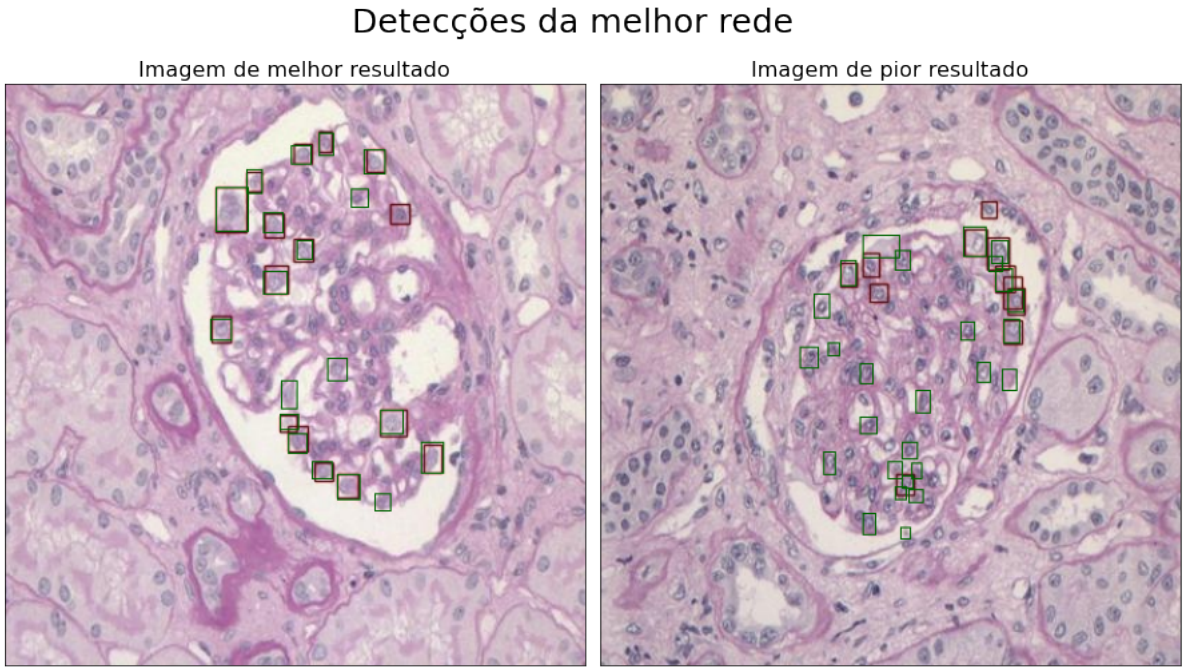


Figura 4.18: Comparação entre as imagens de melhor e de pior resultado para a rede de melhor desempenho

A Tabela 4.3 mostra as métricas obtidas no conjunto de teste para cada uma das configurações definidas na Seção 3.5. A coluna época se refere à época de melhor resultado durante o treinamento, cujo critério de definição é exposto na Seção 3.6. Deve-se notar, como ocorre neste caso, que nem sempre a melhor configuração terá todas as melhores métricas. Por exemplo, a melhor rede, 2B, apresenta o segundo melhor AP. Além disso, constata-se que somente em uma configuração se atingiu o melhor resultado na última época, o que justifica armazenar os pesos de melhor resultado ao longo do treinamento, em vez de armazenar somente o último peso. Um outro ponto importante é notar o efeito do *transfer learning*. As redes com sufixo B, nas quais se empregam pesos iniciais de redes pré-treinadas, a convergência ao melhor resultado ocorre antes, portanto, em menos tempo. De fato, as redes treinadas "do zero"atingem os melhores resultados a partir da época 180, enquanto que as pré-treinadas atingem-no até a época 150.

Configuração	P	R	AP	F1	$\overline{AP}$	Época
# 1A	0.394	0.498	0.391	0.440	0.183	180
# 2A	0.462	0.481	0.412	0.471	0.202	191
# 3A	0.473	0.493	0.458	0.483	0.224	180
# 4A	0.429	0.517	0.420	0.469	0.210	199
# 1B	0.625	0.641	0.631	0.632	0.341	150
# 2B	0.693	0.557	0.648	0.618	0.359	97
# 3B	0.703	0.603	0.652	0.649	0.359	93
# 4B	0.700	0.574	0.632	0.631	0.353	112

Tabela 4.3: Métricas de teste de cada configuração, obtidas na melhor época durante o treinamento

4.2.2 Cenário 2: conjunto de imagens aumentado

A Figura 4.19 mostra os valores de AP obtidos em cada *split* da validação cruzada para o conjunto aumentado de podócitos. Deve-se ter em conta que em alguns casos a diferença de valores é tão pequena que aparenta haver menos do que cinco amostras de cada configuração, pois há interseção entre os círculos. Em relação ao valor médio, somente na versão L a rede pré-treinada teve AP inferior (0.651 contra 0.653 da rede sem pré-treino). Levando-se em conta todas as configurações, a de maior AP foi a configuração 4B, o que corresponde à versão X, como havia ocorrido também no cenário 1. Em todas as configurações, observa-se um ponto discrepante, de máximo, que se refere a um *split* específico (k=1).

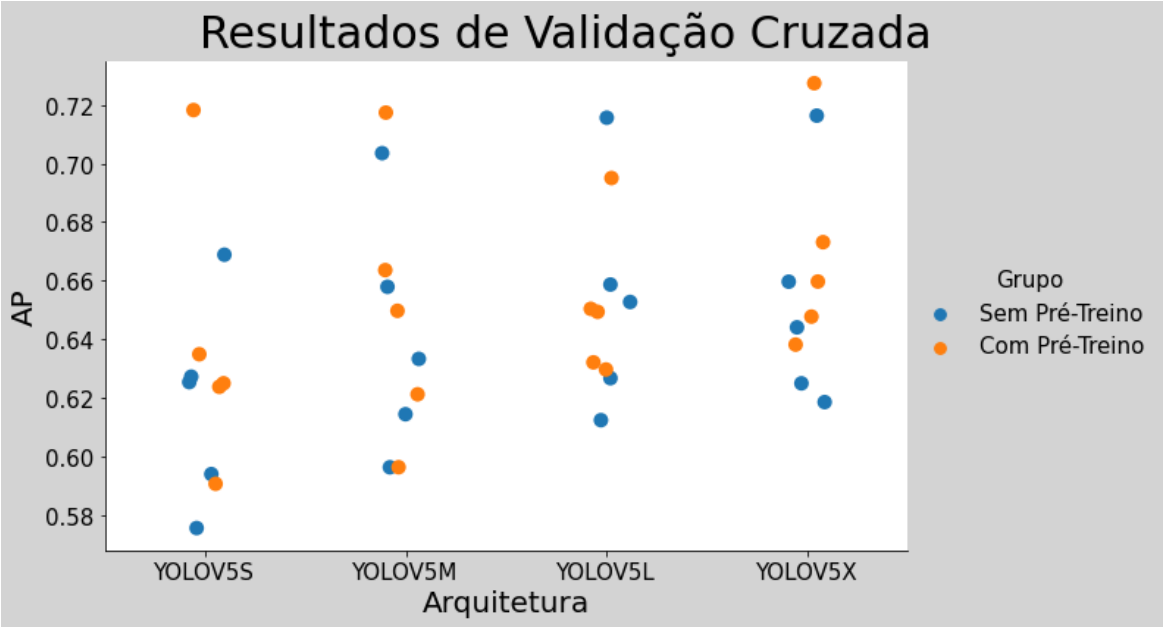


Figura 4.19: Gráfico de dispersão dos valores de AP de cada split da validação cruzada no conjunto aumentado separados por configuração

As Figuras 4.20 e 4.21 mostram as curvas de *loss* para a melhor (3B) e a pior (1B) redes,

respectivamente. Em comum às duas configurações, observa-se aumento do valor de teste logo nas primeiras épocas, algo que não ocorreu no cenário 1. Tendo em vista a região de sobreajuste, deve-se esperar que o melhor resultado seja obtido logo nas primeiras épocas. Em relação a 3B, a curva de teste fica acima da de treino em 187 épocas e tem ponto de mínimo, como já observado, logo na nona época. Deve-se notar que a curva de treino tem menor oscilação, o que pode ser correlacionado ao tamanho do conjunto de dados empregado, e atinge valor mínimo na época 198. Algo similar ocorre com a configuração 1B: a curva de teste atinge seu mínimo no começo do treinamento (n=8), enquanto que a de treino, no final (n=184).

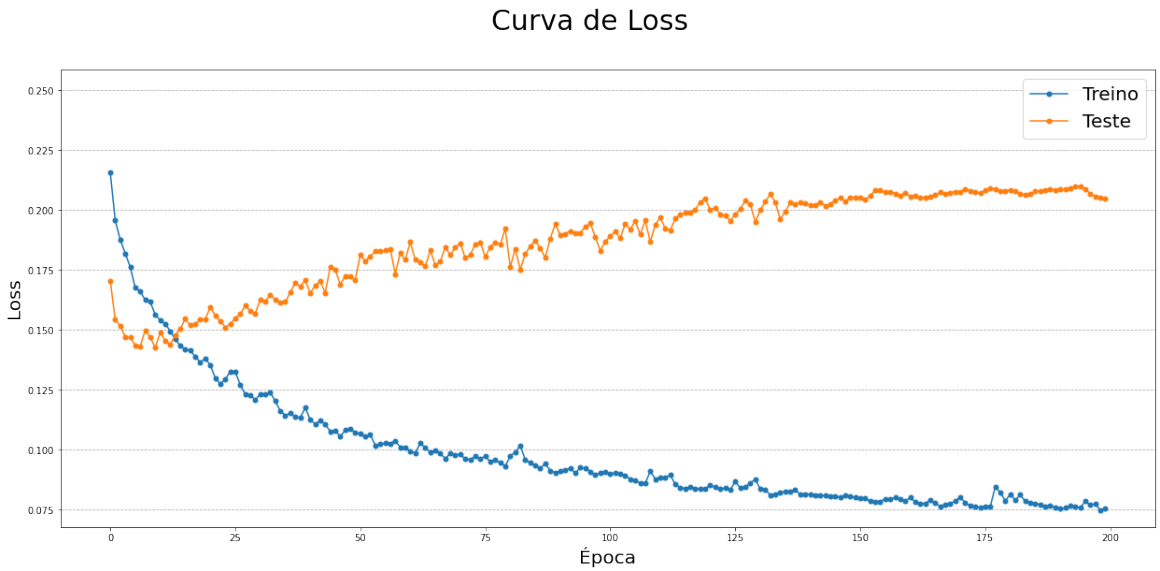


Figura 4.20: Curva de Loss para a rede de melhor desempenho

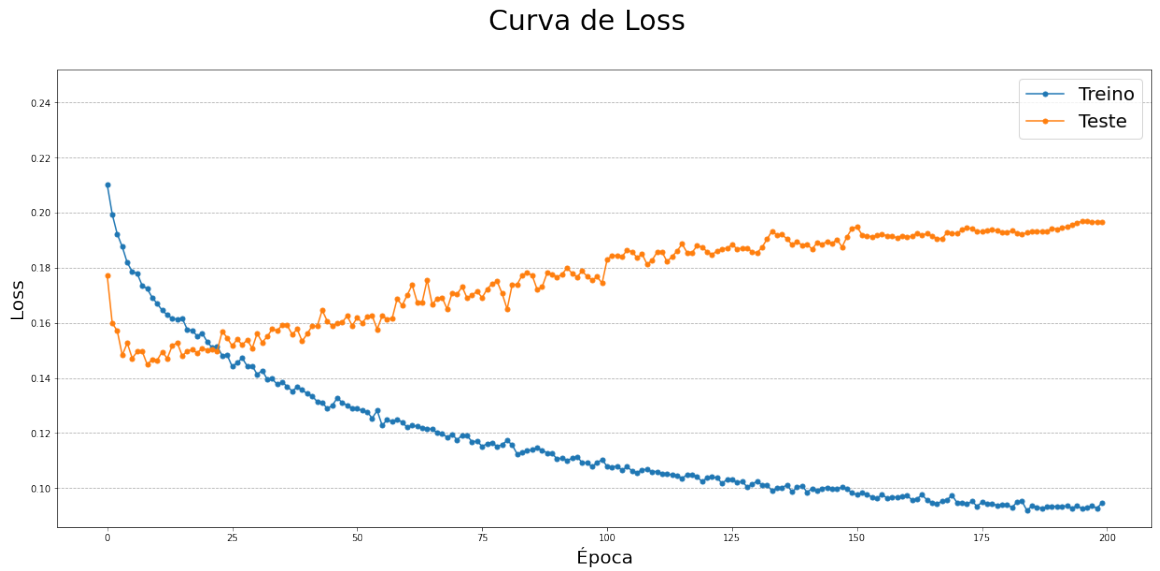


Figura 4.21: Curva de Loss para a rede de pior desempenho

A Figura 4.22 mostra as curvas de AP para a melhor (3B) e a pior (1B) redes. Algumas

características podem ser constatadas a partir das curvas de *loss*, como o ponto de picos, para os dois casos, atingido logo nas primeiras épocas. Também não se nota uma diferença significativa no resultado das duas configurações, como ocorreu no cenário 1. De fato, a maior razão entre o valor de AP da rede 3B em relação a 1B ocorreu logo na segunda época, em que o valor da primeira é pouco mais de 50 % maior que o da segunda. Na melhor rede, atinge-se o valor de máximo, 0,702, na época 9, enquanto que na pior, 0,666, na época 10. Deve-se destacar que, em ambos os casos, as curvas não só se estabilizam, mas passam a decair gradualmente, reflexo do sobreajuste analisado anteriormente.

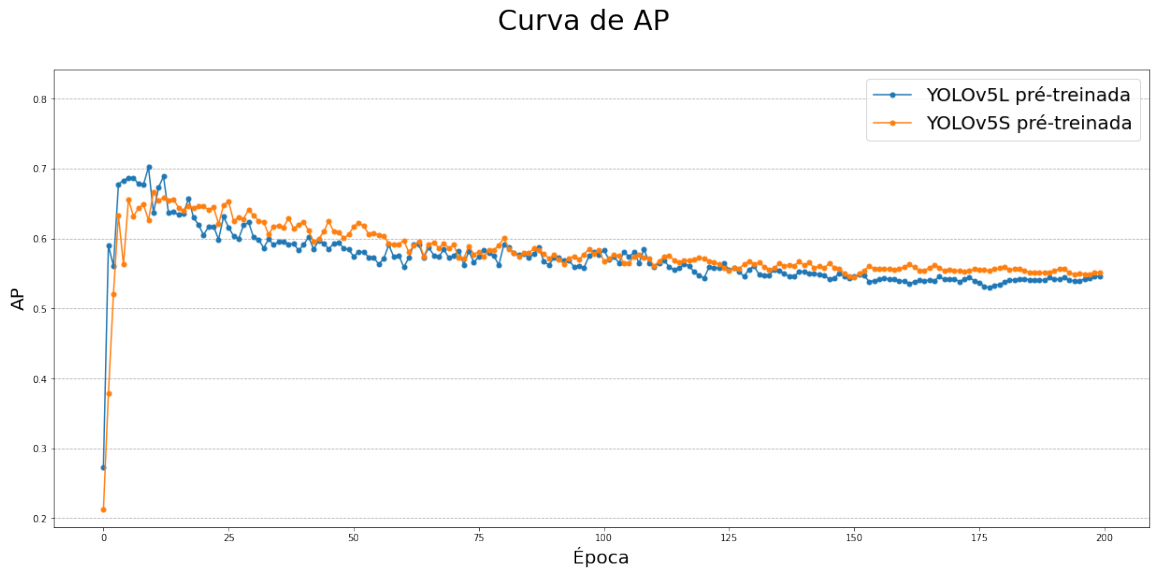


Figura 4.22: Curvas de AP da melhor e da pior redes

A Figura 4.23 traz um comparativo das detecções da melhor e da pior redes em uma imagem padrão.Em ambos os casos, pode-se observar que as detecções contaram com alguns falsos negativos. Apesar disso, as os podócitos corretamente identificados se ajustam bem, de modo geral, às anotações do patologista (*ground-truth boxes*). Pode-se notar que a rede 1B apresenta um falso positivo fora da região glomerular, o que não ocorre com a 3B.

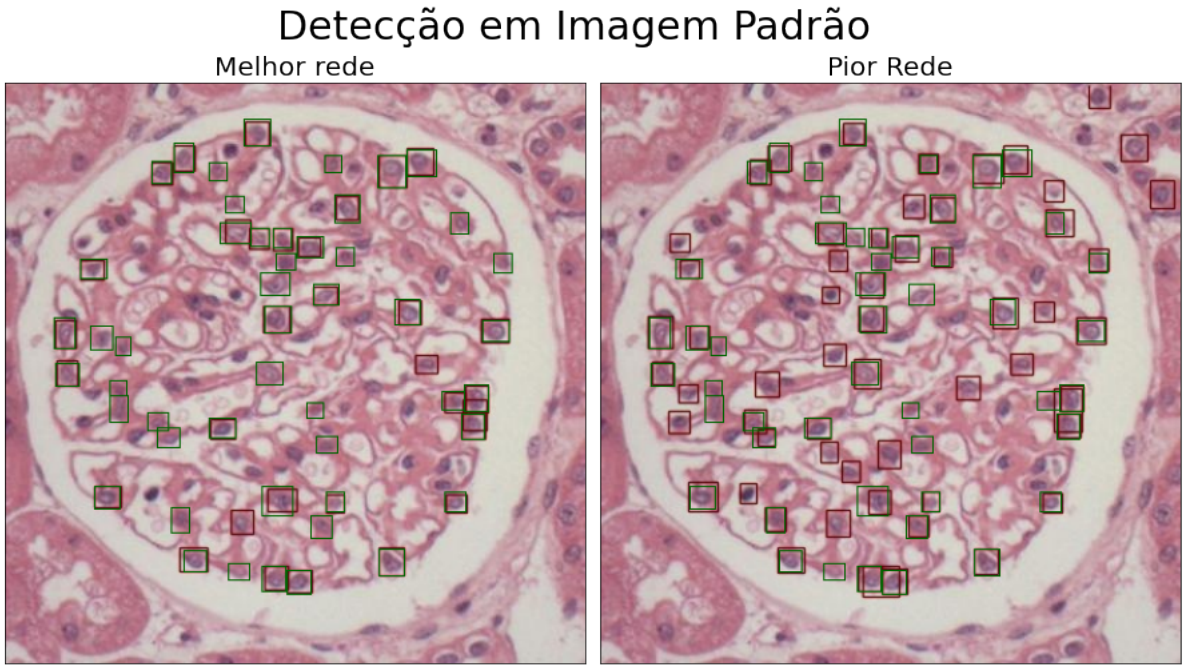


Figura 4.23: Comparação das detecções da melhor (3B) e da pior (1B) redes em uma imagem padrão

A Figura 4.24 mostra a melhor e a pior detecção feitas pela rede 3B. Na melhor imagem, é possível notar somente um falso negativo. Além disso, as detecções da rede parecem bem ajustadas às anotações do patologistas, em quase todos os casos. Tendo em vista que, coincidentemente, as duas imagens são as mesmas do cenário 1, as observações acerca da diferença entre elas - que podem ajudar a compreender a disparidade observada nas detecções - podem ser vistas na discussão feita para aquele cenário. Com relação à pior imagem, as poucas detecções corretas parecem bem ajustadas, mas há diversos falsos positivos e falsos negativos.

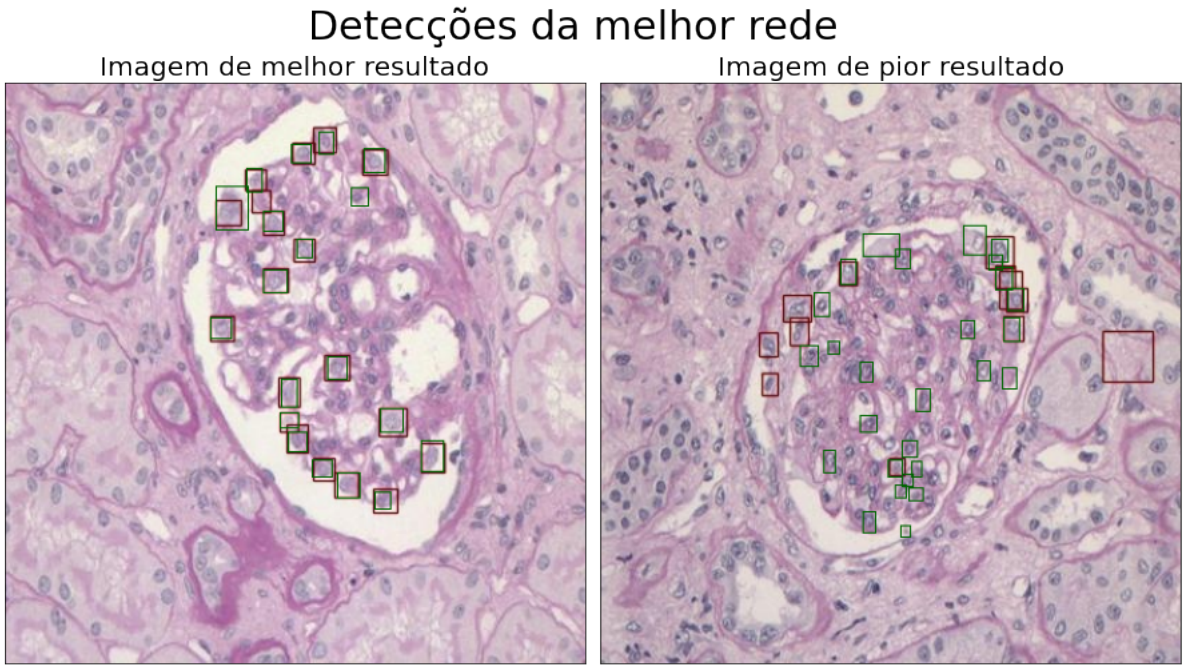


Figura 4.24: Comparação entre as imagens de melhor e de pior resultado para a rede de melhor desempenho

A Tabela 4.4 mostra as métricas obtidas no conjunto de teste para cada uma das configurações. Deve-se notar, como ocorre neste caso, que nem sempre a melhor configuração terá todas as melhores métricas. Novamente, a exemplo do que ocorreu no cenário 1, a melhor rede, 3B, não apresentou todos as maiores métricas. Isto se deve ao fato de que o critério de escolha da melhor rede leva em conta a detecção para limiares de IoU superiores ao padrão, visando marcações mais ajustadas. Além disso, diferentemente do que se observou no cenário 1, a convergência para a melhor época ocorreu em uma etapa inicial do treinamento. Este fato confirma o que se observou nas curvas de *loss* apresentadas, em que se observou uma grande região de sobreajuste. O efeito do *transfer learning* neste cenário foi menor do que o observado no cenário 1. Por exemplo, os valores de AP variam entre 0,68 e 0,72. Ademais, a única diferença que se observa é que as redes pré-treinadas atingem o melhor resultado em épocas inferiores.

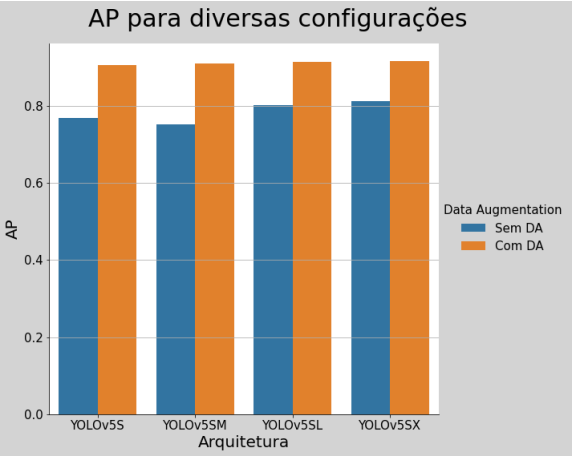
Configuração	P	R	AP	F1	$\overline{AP}$	Época
# 1A	0.681	0.630	0.679	0.654	0.360	30
# 2A	0.680	0.672	0.704	0.676	0.379	30
# 3A	0.681	0.648	0.694	0.664	0.372	33
# 4A	0.731	0.625	0.701	0.674	0.387	35
# 1B	0.677	0.640	0.641	0.658	0.350	29
# 2B	0.702	0.669	0.704	0.685	0.379	9
# 3B	0.712	0.660	0.702	0.685	0.390	9
# 4B	0.709	0.667	0.725	0.687	0.387	6

Tabela 4.4: Métricas de teste de cada configuração, obtidas na melhor época durante o treinamento

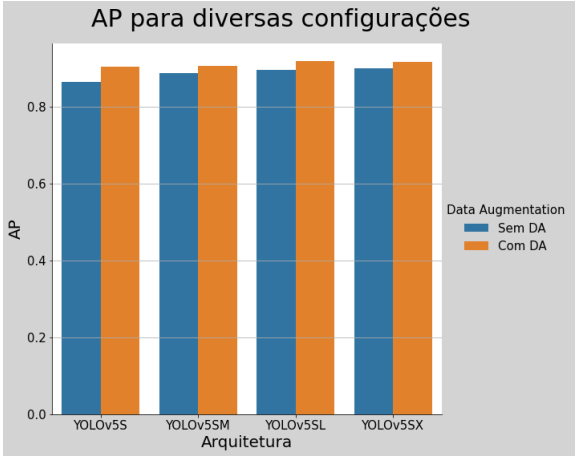
4.3 Discussões

Nas Seções 4.1 e 4.2, fizemos uma análise dos resultados segmentada pela base de dados e pelo cenário observado. Aqui, fazemos uma discussão mais ampla, estendendo a análise a todas as bases e cenários. Em primeiro lugar, houve melhora no desempenho, quando aplicadas, de forma separada ou conjunta, as técnicas de *transfer learning* e *data augmentation*. No caso desta última, deve-se observar que o ganho de desempenho vem a despeito do aumento dos custos computacional e de tempo. Sobre a primeira, observou-se que a aplicação de redes pré-treinadas possibilitou uma redução no número de épocas para que se atingisse o melhor resultado: em média, 39% na base de núcleos e 53% na de podócitos. Além disso, fragmentando a análise aos quatro grupos referentes às condições de presença ou não de *transfer learning* e *data augmentation* de cada base, observamos que as redes mais profundas desempenharam melhor. Em sete das oito possibilidades - considerando as duas bases - o melhor resultado adveio das versões L ou X. Novamente, o resultado superior destas redes é contrabalanceado pelo tempo de processamento que requerem. Com respeito às curvas de *loss*, em ambas as bases pôde-se encontrar um comportamento em comum. No cenário aumentado, em que há 44 cópias para cada imagem original, observa-se que o melhor resultado - e, paralelamente, o ponto de mínimo da curva de *loss* de teste - ocorre logo nas primeiras épocas. Isto se deve ao fato de que o número de iterações feitas, e consequentemente o tempo de computação, por época é muito maior, uma vez que o *batch-size* é definido aprioristicamente. Desse modo, é possível entender a razão pela qual a região de sobreajuste é maior neste cenário, em certos casos ocupando a maior parte do treinamento. Por este motivo, ressalta-se a necessidade de armazenar os melhores pesos ao longo do treinamento. Além disso, de modo geral, as curvas de teste ficaram abaixo das de treino em ambos os cenários, exceto nas regiões de sobreajuste. Isto é um indicativo de que o conjunto de teste é pouco representativo, visto que as predições são feitas com maior facilidade no conjunto de teste.

Anteriormente, os resultados foram dispostos de modo separado por cenários: original ou aumentado. Neste momento, fazemos uma análise combinando os cenários, de modo a interpretar o impacto, em termos de desempenho, da aplicação de *data augmentation*. Neste sentido, a Figura 4.25 mostra o resultado de AP obtido na melhor época de cada rede para a base de núcleos. Os mesmos resultados para a base de podócitos são mostrados na Figura 4.26. Os dados mostram que a expansão do conjunto de treino por meio da criação de imagens artificiais foi capaz de gerar resultados superiores em todos os casos (núcleos e podócitos, redes sem e com pré-treino). Nota-se, entretanto, que a diferença marginal de desempenho é menor no caso das redes pré-treinadas.

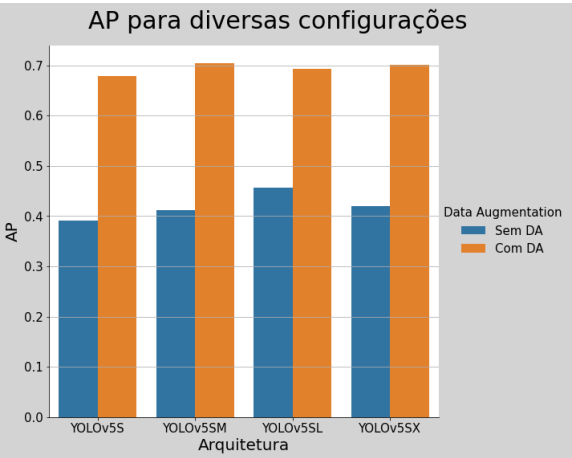


(a) Redes sem pré-treino

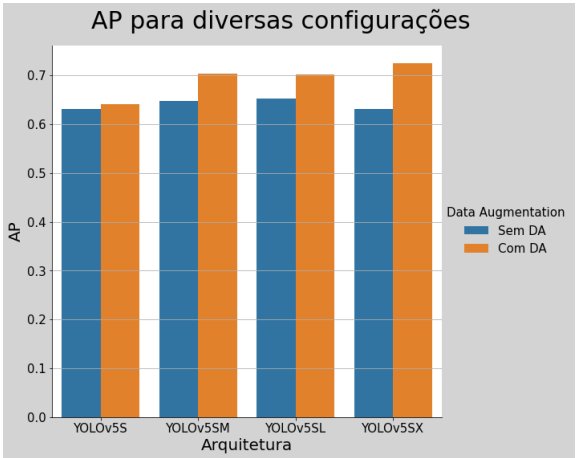


(b) Redes pré-treinadas

Figura 4.25: Resultados de AP na base de núcleos combinando os cenários



(a) Redes sem pré-treino



(b) Redes pré-treinadas

Figura 4.26: Resultados de AP na base de podócitos combinando os cenários

Capítulo 5

Conclusões

Este trabalho se propôs a verificar a possibilidade de se empregar redes neurais convolucionais no processo de detecção de estruturas biológicas, em especial de podócitos, em imagens histológicas de glomérulos. Na base de núcleos, obteve-se precisão, *recall* e AP de 0.922, 0.894 e 0.916 na melhor rede dos modelos finais. Na base de podócitos, obteve-se precisão, *recall* e AP de 0.712, 0.660 e 0.702.

Em todos os cenários, o emprego de redes pré-treinadas e de *data augmentation* - tanto de forma separada, quanto de forma conjunta - proporcionou resultados superiores. Em contrapartida, o treinamento nos conjuntos aumentados representou um custo de tempo maior. Além disso, em razão de o número de épocas ter sido fixado *a priori*, os treinamentos nos conjuntos aumentados estiveram sujeitos a uma grande região de sobreajuste. Em outras palavras, os melhores resultados foram atingidos na fase inicial do treinamento, o que permitiria uma redução do número de épocas. Nos cenários dos conjuntos de imagens originais, por outro lado, os melhores resultados foram atingidos, via de regra, na segunda metade do treinamento. Em todos os casos, o uso de redes pré-treinadas precipitou a época de melhor resultado. A análise das curvas de *loss* permitiu observar duas tendências: em primeiro lugar, o sobreajuste foi significativamente maior no conjunto aumentado; outrossim, as curvas de teste se mantiveram abaixo das de treino excetuando-se, quando houve ocorrência, a região de sobreajuste, o que pode ser um indicativo de um conjunto de dados desbalanceado. Este indicativo também foi observado por meio da validação cruzada, em que alguns *splits* tiveram comportamento discrepante. Por este motivo, ressalta-se a importância de buscar a ampliação da base de dados, de modo a enriquecê-la.

5.1 Perspectivas Futuras

Este trabalho contém diversas limitações. Em primeiro lugar, tendo em vista a quantidade de configurações a serem testadas e o tempo de que se dispunha, não foi possível realizar a otimização de hiper-parâmetros. É comum que se aplique combinações de valores de hiper-parâmetros durante a validação cruzada no conjunto de *treino*, escolhendo-se a média dos resultados em cada *fold* para selecionar os hiper-parâmetros ótimos, que então serão utilizado para o treinamento com o conjunto

completo. Ademais, diferentes arquiteturas poderiam ser testadas neste conjunto de dados, de modo a compará-las com a utilizada neste trabalho. Além disso, poderiam ser testados métodos de deconvolução de cor, de modo que se pudesse mitigar os efeitos da heterogeneidade de corantes existente no conjunto de dados. Técnicas de visualização em redes neurais, como mapas de calor e de saliência poderiam ser empregadas para enriquecer a qualidade da análise sobre as detecções. Futuramente, as anotações poderiam ser atualizadas para que a rede pudesse identificar a condição de cada podócito, se saudável ou lesionado, ou, caso houvesse amostras suficientes, classificá-los pelo tipo de lesão. Neste último caso, seria importante verificar se o conjunto de dados é diverso e balanceado. Por fim, a principal atualização que este trabalho poderia receber é a expansão da base de dados anotada, tendo em vista que trabalhos similares, em segmentação de glomérulos, por exemplo, contam com uma quantidade substantivamente maior de amostras.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] GONZALEZ, R.; WOODS, R. *Digital Image Processing*. Pearson/Prentice Hall, 2008. ISBN 9780131687288. Disponível em: <https://books.google.com.br/books?id=8uGOnjRGEzoC>.

[2] FORSYTH, D.; PONCE, J. *Computer Vision: A Modern Approach. (Second edition)*. Prentice Hall, 2011. 792 p. Disponível em: <https://hal.inria.fr/hal-01063327>.

[3] BALLARD, D.; BROWN, C. *Computer Vision*. Prentice-Hall, 1982. ISBN 9780131653160. Disponível em: <https://books.google.com.br/books?id=EfRRAAAAMAAJ>.

[4] WU, Y. *An Introduction to Computer Vision*. Último acesso em 16/09/21. Disponível em: <http://users.eecs.northwestern.edu/ yingwu/teaching/EECS432/Notes/intro.pdf>.

[5] OUAKNINE, A. *Review of Deep Learning Algorithms for Object Detection*. Último acesso em 16/09/21. Disponível em: <https://medium.com/zylapp/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852>.

[6] NIAZI, M. K. K.; PARWANI, A. V.; GURCAN, M. N. Digital pathology and artificial intelligence. *The Lancet Oncology*, Elsevier BV, v. 20, n. 5, p. e253–e261, maio 2019. Disponível em: <https://doi.org/10.1016/s1470-2045(19)30154-8>.

[7] LITJENS, G. et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, v. 42, p. 60–88, 2017. ISSN 1361-8415. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1361841517301135>.

[8] DENG, S. et al. Deep learning in digital pathology image analysis: a survey. *Frontiers of Medicine*, Springer Science and Business Media LLC, v. 14, n. 4, p. 470–487, jul. 2020. Disponível em: <https://doi.org/10.1007/s11684-020-0782-9>.

[9] DIABETES, N. I. of; DIGESTIVE; (NIDDK), K. D. *Glomerular Diseases*. `https://www.niddk.nih.gov/health-information/kidney-disease/glomerular-diseases#what`. Último acesso em 06/11/21.

[10] MITCHELL, T. *Machine Learning*. McGraw-Hill, 1997. (McGraw-Hill International Editions). ISBN 9780071154673. Disponível em: <https://books.google.com.br/books?id=EoYBngEACAAJ>.

[11] GIRSHICK, R. *Fast R-CNN*. 2015.

[12] REDMON, J. et al. *You Only Look Once: Unified, Real-Time Object Detection*. 2016.

[13] SZEGEDY, C. et al. *Rethinking the Inception Architecture for Computer Vision*. 2015.

[14] BENALI, A. et al. A computerized image analysis system for quantitative analysis of cells in histological brain sections. *Journal of Neuroscience Methods*, v. 125, n. 1, p. 33–43, 2003. ISSN 0165-0270. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0165027003000232>.

[15] CRUZ, L. et al. A statistically based density map method for identification and quantification of regional differences in microcolumnarity in the monkey brain. *Journal of Neuroscience Methods*, v. 141, n. 2, p. 321–332, 2005. ISSN 0165-0270. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0165027004003309>.

[16] BULDYREV, S. V. et al. Description of microcolumnar ensembles in association cortex and their disruption in alzheimer and lewy body dementias. *Proceedings of the National Academy of Sciences*, National Academy of Sciences, v. 97, n. 10, p. 5039–5043, 2000. ISSN 0027-8424. Disponível em: <https://www.pnas.org/content/97/10/5039>.

[17] INGLIS, A. et al. Automated identification of neurons and their locations. *Journal of Microscopy*, v. 230, n. 3, p. 339–352, 2008. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2818.2008.01992.x>.

[18] JAFARI-KHOUZANI, K.; SOLTANIAN-ZADEH, H. Multiwavelet grading of pathological images of prostate. *IEEE Transactions on Biomedical Engineering*, v. 50, n. 6, p. 697–704, 2003.

[19] TABESH, A. et al. Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Transactions on Medical Imaging*, v. 26, n. 10, p. 1366–1378, 2007.

[20] HUANG, P.-W.; LEE, C.-H. Automatic classification for pathological prostate images based on fractal analysis. *IEEE Transactions on Medical Imaging*, v. 28, n. 7, p. 1037–1050, 2009.

[21] DOYLE, S. et al. A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. *IEEE Transactions on Biomedical Engineering*, v. 59, n. 5, p. 1205–1218, 2012.

[22] NAGPAL, K. et al. Publisher correction: Development and validation of a deep learning algorithm for improving gleason scoring of prostate cancer. *npj Digital Medicine*, v. 2, n. 1, 2019.

[23] GALLEGO, J. et al. Glomerulus classification and detection based on convolutional neural networks. *Journal of Imaging*, v. 4, n. 1, 2018. ISSN 2313-433X. Disponível em: <https://www.mdpi.com/2313-433X/4/1/20>.

[24] BUKOWY, J. D. et al. Region-based convolutional neural nets for localization of glomeruli in trichrome-stained whole kidney sections. *Journal of the American Society of Nephrology*,

American Society of Nephrology, v. 29, n. 8, p. 2081–2088, 2018. ISSN 1046-6673. Disponível em: <https://jasn.asnjournals.org/content/29/8/2081>.

[25] BEL, T. de et al. Automatic segmentation of histopathological slides of renal tissue using deep learning. In: TOMASZEWSKI, J. E.; GURCAN, M. N. (Ed.). *Medical Imaging 2018: Digital Pathology*. SPIE, 2018. v. 10581, p. 285 – 290. Disponível em: <https://doi.org/10.1117/12.2293717>.

[26] MARSH, J. N. et al. Deep learning global glomerulosclerosis in transplant kidney frozen sections. *IEEE Transactions on Medical Imaging*, v. 37, n. 12, p. 2718–2728, 2018.

[27] KANNAN, S. et al. Segmentation of glomeruli within trichrome images using deep learning. *bioRxiv*, Cold Spring Harbor Laboratory, 2019. Disponível em: <https://www.biorxiv.org/content/early/2019/02/28/345579>.

[28] GINLEY, B. et al. Computational segmentation and classification of diabetic glomerulosclerosis. *Journal of the American Society of Nephrology*, American Society of Nephrology, v. 30, n. 10, p. 1953–1967, 2019. ISSN 1046-6673. Disponível em: <https://jasn.asnjournals.org/content/30/10/1953>.

[29] HERMSEN, M. et al. Deep learning–based histopathologic assessment of kidney tissue. *Journal of the American Society of Nephrology*, American Society of Nephrology, v. 30, n. 10, p. 1968–1979, 2019. ISSN 1046-6673. Disponível em: <https://jasn.asnjournals.org/content/30/10/1968>.

[30] ALTINI, N. et al. Semantic segmentation framework for glomeruli detection and classification in kidney histological sections. *Electronics*, v. 9, n. 3, 2020. ISSN 2079-9292. Disponível em: <https://www.mdpi.com/2079-9292/9/3/503>.

[31] LIN, G. et al. A hybrid 3d watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, v. 56, p. 23–36, 11 2003.

[32] WÄHLBY, C. et al. Combining intensity, edge and shape information for 2d and 3d segmentation of cell nuclei in tissue sections. *Journal of Microscopy*, v. 215, n. 1, p. 67–76, 2004. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0022-2720.2004.01338.x>.

[33] CHENG, J.; *, J. C. R. Segmentation of clustered nuclei with shape markers and marking function. *IEEE Transactions on Biomedical Engineering*, v. 56, n. 3, p. 741–748, March 2009. ISSN 1558-2531.

[34] AL-KOFAHI, Y. et al. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering*, v. 57, n. 4, p. 841–852, April 2010. ISSN 1558-2531.

[35] VETA, M. et al. Automatic nuclei segmentation in h&e stained breast cancer histopathology images. *PLoS ONE*, v. 8, 2013.

[36] CIRESAN, D. et al. Mitosis detection in breast cancer histology images with deep neural networks. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, v. 16 Pt 2, p. 411–8, 2013.

[37] SIRINUKUNWATTANA, K. et al. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Transactions on Medical Imaging*, v. 35, n. 5, p. 1196–1206, May 2016. ISSN 1558-254X.

[38] XU, J. et al. Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. *IEEE transactions on medical imaging*, v. 35, 01 2016.

[39] WANG, P. et al. Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Signal Processing*, v. 122, p. 1–13, 2016. ISSN 0165-1684. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0165168415003916>.

[40] XING, F.; XIE, Y.; YANG, L. An automatic learning-based framework for robust nucleus segmentation. *IEEE Transactions on Medical Imaging*, v. 35, n. 2, p. 550–566, Feb 2016. ISSN 1558-254X.

[41] KUMAR, N. et al. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, v. 36, n. 7, p. 1550–1560, July 2017. ISSN 1558-254X.

[42] NAYLOR, P. et al. Nuclei segmentation in histopathology images using deep neural networks. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. [S.l.: s.n.], 2017. p. 933–936. ISSN 1945-8452.

[43] NAYLOR, P. et al. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, v. 38, p. 1–1, 08 2018.

[44] LUTNICK B., G. B. G. D. e. a. An integrated iterative annotation technique for easing neural network training in medical image analysis. *Nat Mach Intell 1*, p. 112–119, 2019.

[45] MARASZEK, K. E. et al. The presence and location of podocytes in glomeruli as affected by diabetes mellitus. In: TOMASZEWSKI, J. E.; WARD, A. D. (Ed.). *Medical Imaging 2020: Digital Pathology*. SPIE, 2020. v. 11320, p. 298 – 307. Disponível em: <https://doi.org/10.1117/12.2548904>.

[46] ZIMMERMANN, M. et al. Deep learning–based molecular morphometrics for kidney biopsies. *JCI Insight*, The American Society for Clinical Investigation, v. 6, n. 7, 4 2021. Disponível em: <https://doi.org/10.1172/jci.insight.144779>.

[47] ZENG, C. et al. Identification of glomerular lesions and intrinsic glomerular cell types in kidney diseases via deep learning. *The Journal of Pathology*, v. 252, n. 1, p. e5491, 2020. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/path.5491>.

[48] KHAN, S. et al. *A Guide to Convolutional Neural Networks for Computer Vision*. [S.l.]: Morgan & Claypool, 2018. (Synthesis Lectures on Computer Vision).

[49] GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. MIT Press, 2016. (Adaptive Computation and Machine Learning series). ISBN 9780262337373. Disponível em: <https://books.google.com.br/books?id=omivDQAAQBAJ>.

[50] REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. *ArXiv*, abs/1804.02767, 2018.

[51] BOCHKOVSKIY, A.; WANG, C.-Y.; LIAO, H. Yolov4: Optimal speed and accuracy of object detection. *ArXiv*, abs/2004.10934, 2020.

[52] BERRAR, D. Cross-validation. In: RANGANATHAN, S. et al. (Ed.). *Encyclopedia of Bioinformatics and Computational Biology*. Oxford: Academic Press, 2019. p. 542–545. ISBN 978-0-12-811432-2. Disponível em: <https://www.sciencedirect.com/science/article/pii/B978012809633820349X>.

[53] COCO Dataset. `https://cocodataset.org/#home`. Último acesso em 18/08/21.

[54] PADILLA, R.; NETTO, S. L.; SILVA, E. A. B. da. A survey on performance metrics for object-detection algorithms. In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. [S.l.: s.n.], 2020. p. 237–242.

6. APÊNDICES

6.1 A Operação de Convolução

Conforme mencionado no Capítulo 1, a convolução é uma operação fundamental às redes neurais convolucionais. A Equação 3.1 se refere à operação de convolução com o *kernel* invertido horizontal e verticalmente. Esta inversão é feita para que se possa manter a comutatividade da convolução. Desse modo, a Equação 3.1 pode ser reescrita pela propriedade comutativa como na Equação 6.1. Apesar de esta propriedade ser útil em demonstrações em diversas áreas (Processamento de Sinais, Teoria de Controle), ela é pouco útil em termos práticos em *Machine Learning*.

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n) \tag{6.1}$$

Uma outra forma equivalente e comumente utilizada na implementação de redes neurais é a função de correlação cruzada (*cross-correlation*), mostrada na Equação 6.2. A grande vantagem é que esta forma resulta em uma função idêntica à da convolução, mas sem necessidade de inverter o *kernel*.

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n) \tag{6.2}$$

A Figura 6.1 ilustra a distinção entre correlação e convolução. Esta é uma distinção especialmente importante na literatura de Processamento de Sinais, uma vez que em *Machine Learning* há quem se refira às duas operações indistintamente[49].

Figura 6.1: Distinção entre correlação e convolução[48]

6.2 Funções de Ativação

Nesta seção, apresentamos de forma sucinta algumas funções de ativação. As Equações 6.3-6.5 mostram as funções de ativação sigmoide, tangente hiperbólico e ReLU (Rectifier Linear Unit). Todas elas podem ser utilizadas em camadas intermediárias de CNNs. A escolha de qual deve ser utilizada leva em conta critérios como a resposta ao problema de dissipação do gradiente (*vanishing gradient*).

$$f_s(x) = \frac{1}{1 + e^{-x}} \tag{6.3}$$

$$f_t(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{6.4}$$

$$f_r(x) = \max(0, x) \tag{6.5}$$

A Figura 6.2 mostra o mapeamento das três funções de entrada para pontos no intervalo [-1, 1]. Um ponto importante é que o conjunto Imagem da função ReLU são os reais positivos, enquanto que o das outras duas é o segmento real [0, 1].

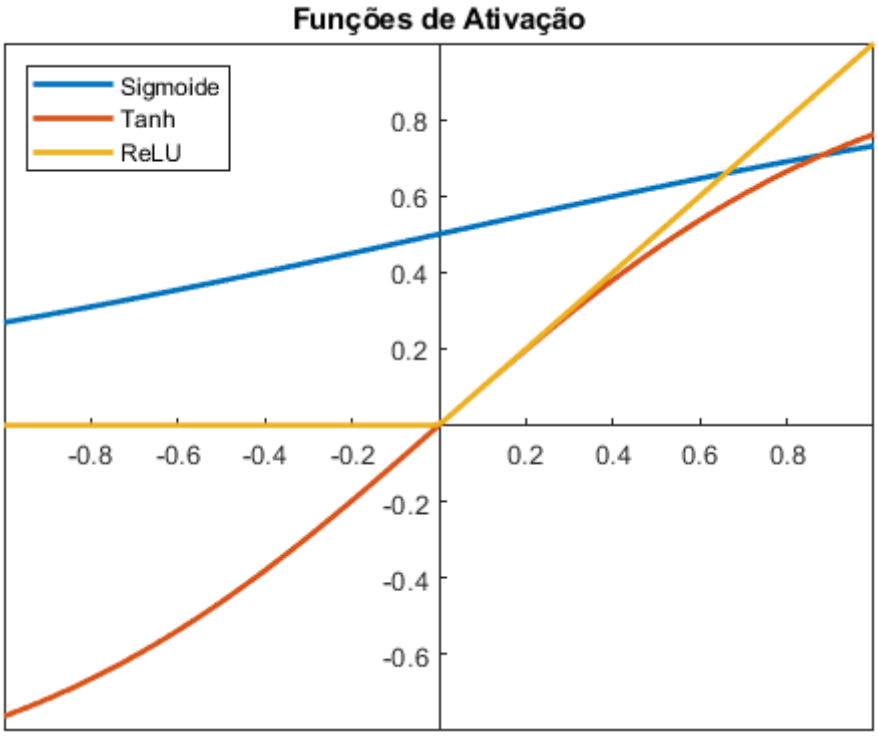


Figura 6.2: Mapeamento de funções de ativação

6.3 Validação Cruzada

6.3.1 Base anotada de núcleos

As Tabelas 6.1 e 6.2 mostram as médias das principais métricas obtidas durante a validação cruzada. Deve-se ressaltar que a métrica de cada *split* se refere à época de melhor resultado. Além disso, ressalta-se que os resultados se referem à aplicação do método *5-fold* ao conjunto de treino.

Configuração	P	R	AP	F1
# 1A	0.809	0.719	0.735	0.761
# 2A	0.820	0.738	0.757	0.777
# 3A	0.830	0.748	0.768	0.786
# 4A	0.842	0.768	0.791	0.803
# 1B	0.889	0.822	0.854	0.854
# 2B	0.911	0.854	0.886	0.881
# 3B	0.924	0.863	0.901	0.892
# 4B	0.927	0.869	0.903	0.897

Tabela 6.1: Média das métricas de validação cruzada para cada configuração do conjunto original

Configuração	P	R	AP	F1
# 1A	0.928	0.864	0.924	0.895
# 2A	0.930	0.870	0.926	0.899
# 3A	0.932	0.875	0.931	0.902
# 4A	0.931	0.874	0.931	0.901
# 1B	0.929	0.860	0.922	0.893
# 2B	0.933	0.866	0.926	0.898
# 3B	0.934	0.870	0.926	0.901
# 4B	0.932	0.873	0.927	0.901

Tabela 6.2: Média das métricas de validação cruzada para cada configuração do conjunto aumentado

6.3.2 Base anotada de podócitos

As Tabelas 6.3 e 6.4 mostram a média das principais métricas obtidas durante a validação cruzada na base de podócitos.

Configuração	P	R	AP	F1
# 1A	0.398	0.434	0.344	0.414
# 2A	0.428	0.430	0.372	0.429
# 3A	0.413	0.462	0.372	0.431
# 4A	0.442	0.436	0.380	0.435
# 1B	0.599	0.583	0.592	0.590
# 2B	0.656	0.564	0.620	0.604
# 3B	0.685	0.541	0.603	0.603
# 4B	0.653	0.587	0.620	0.616

Tabela 6.3: Média das métricas de validação cruzada para cada configuração do conjunto original

Configuração	P	R	AP	F1
# 1A	0.633	0.571	0.618	0.599
# 2A	0.664	0.581	0.641	0.619
# 3A	0.674	0.577	0.653	0.621
# 4A	0.672	0.584	0.653	0.625
# 1B	0.646	0.594	0.639	0.619
# 2B	0.671	0.592	0.660	0.628
# 3B	0.680	0.593	0.651	0.632
# 4B	0.685	0.600	0.669	0.639

Tabela 6.4: Média das métricas de validação cruzada para cada configuração do conjunto aumentado

6.4 Detecções de todas as configurações

6.4.1 Base anotada de núcleos

As Figuras 6.3-6.6 mostram detecções de todas as configurações de redes em uma imagem padrão.

6.4.1.1 Cenário 1: conjunto de imagens original

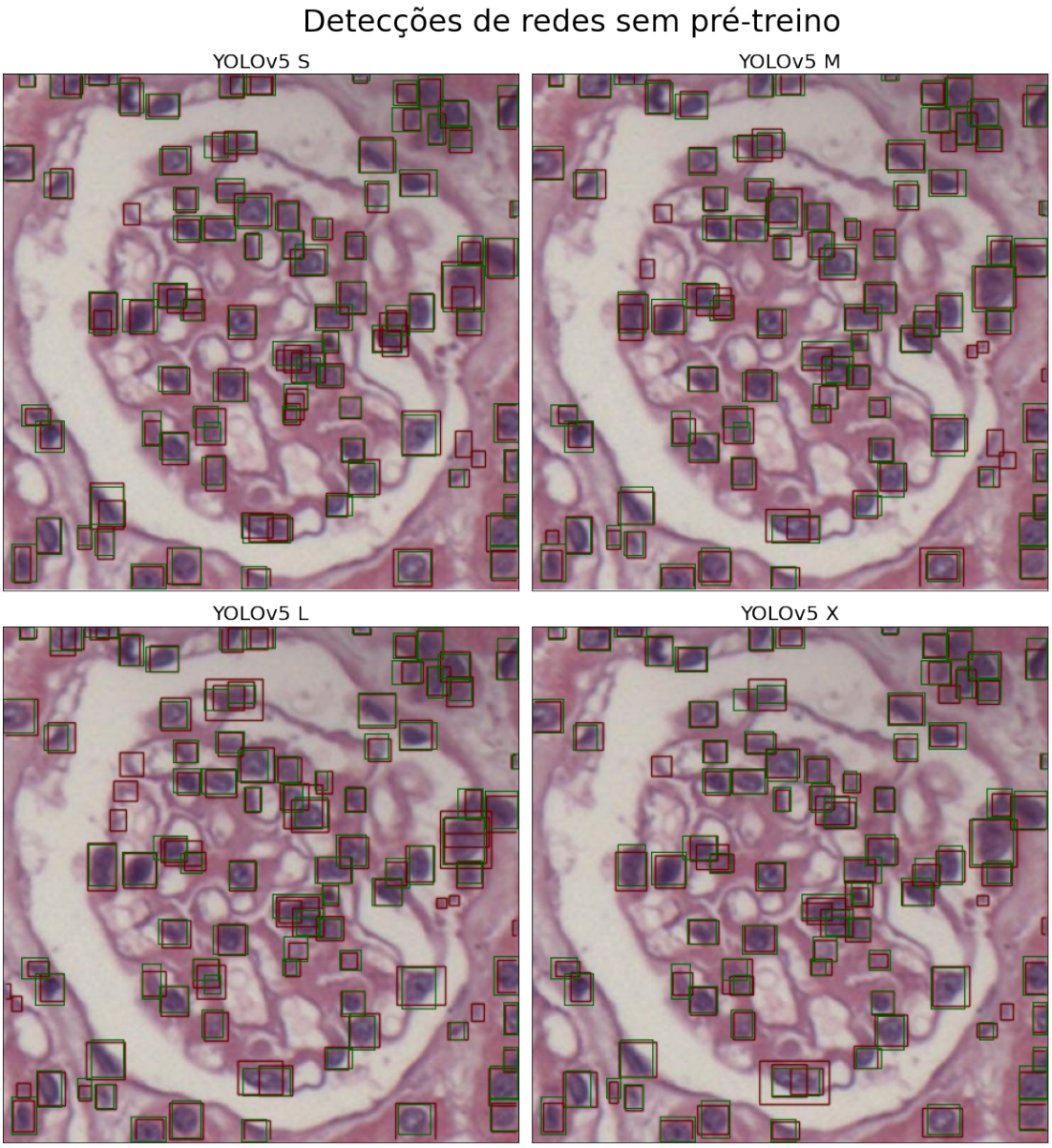


Figura 6.3: Exemplo de detecções das redes sem pré-treino na base de núcleos original

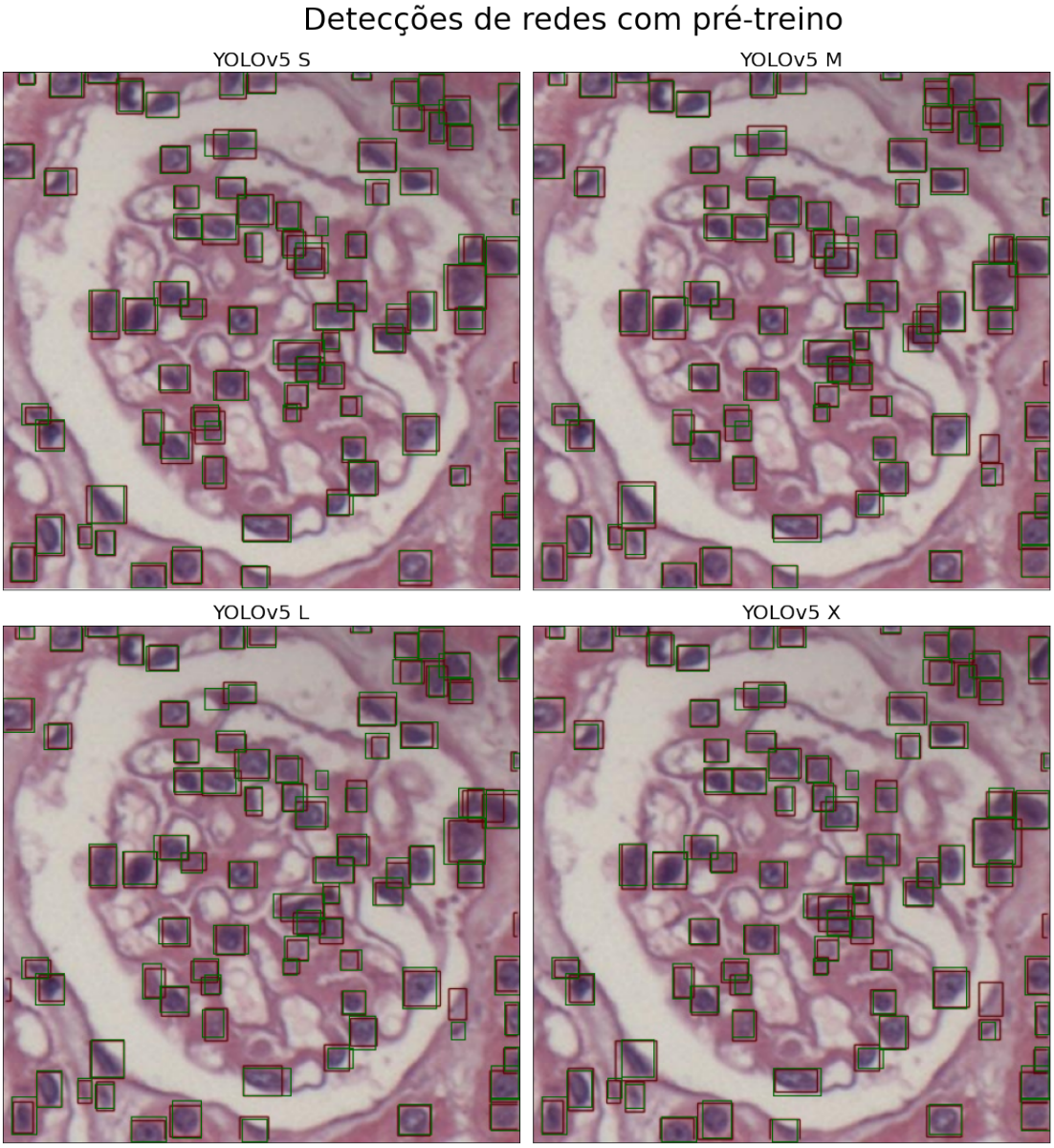


Figura 6.4: Exemplo de detecções das redes com pré-treino na base de núcleos original

6.4.1.2 Cenário 2: conjunto de imagens aumentado

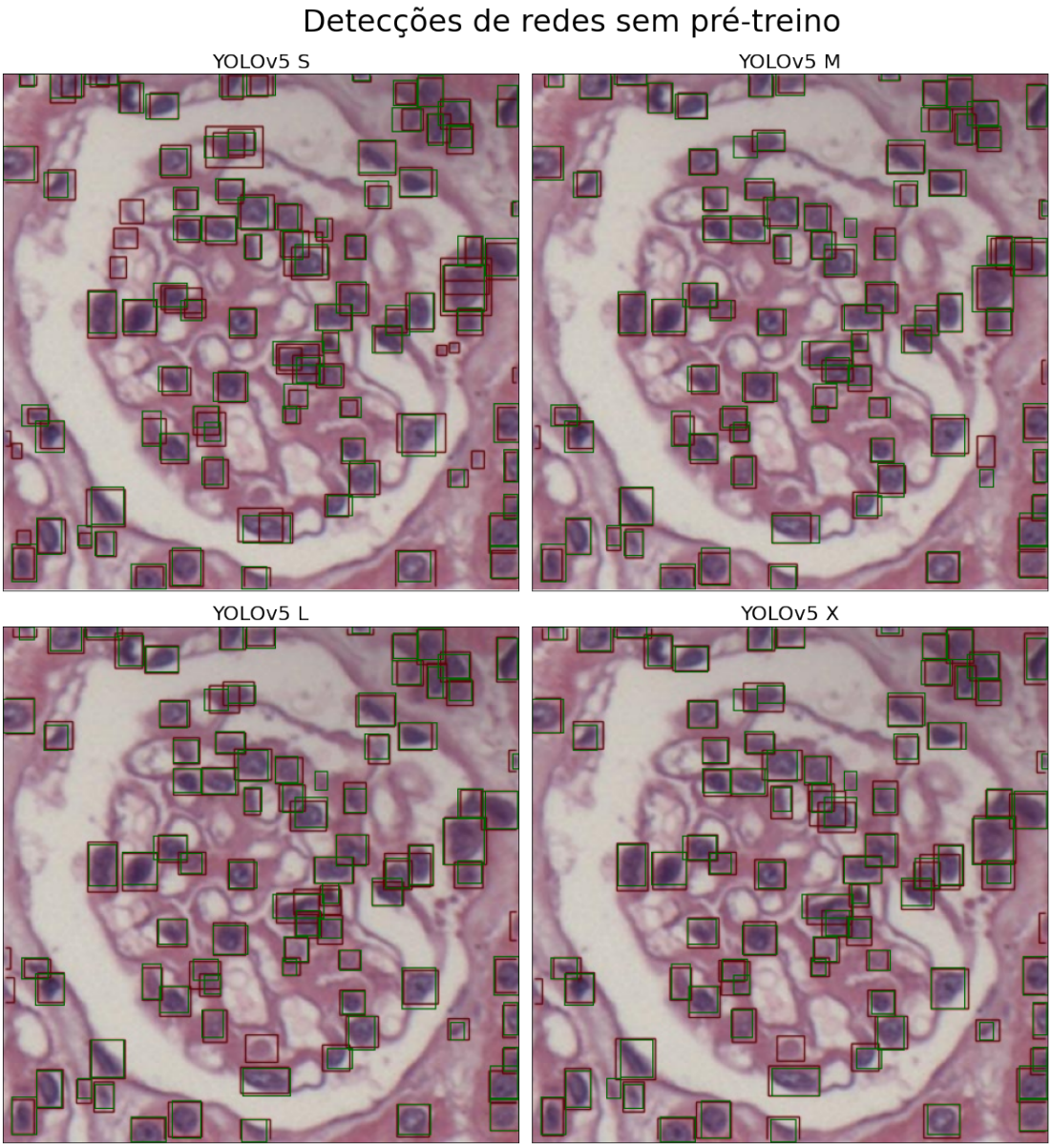


Figura 6.5: Exemplo de detecções das redes sem pré-treino na base de núcleos aumentada

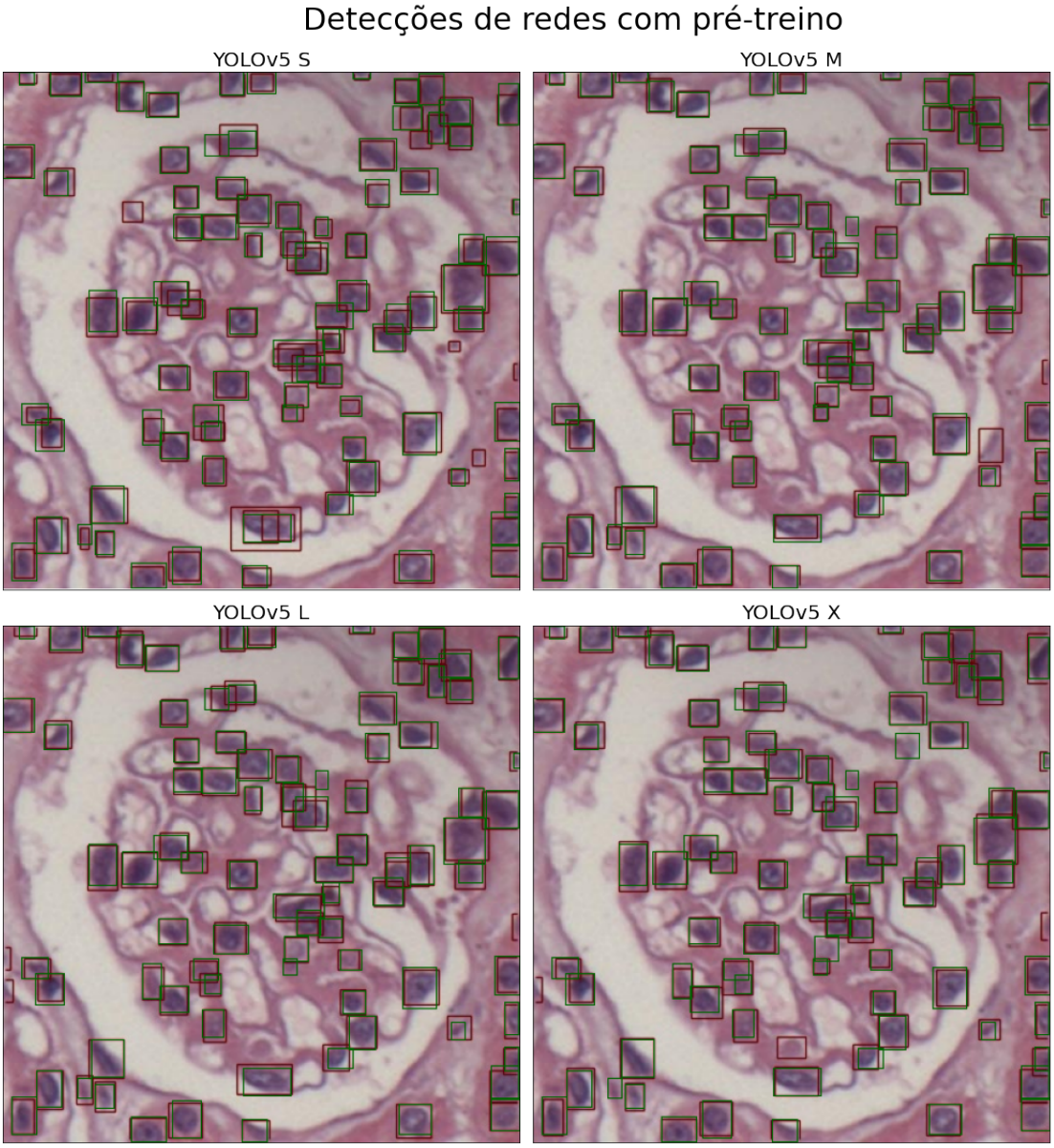


Figura 6.6: Exemplo de detecções das redes com pré-treino na base de núcleos aumentada

6.4.2 Base anotada de podócitos

As Figuras 6.7-6.10 mostram detecções de todas as configurações de redes em uma imagem padrão.

6.4.2.1 Cenário 1: conjunto de imagens original

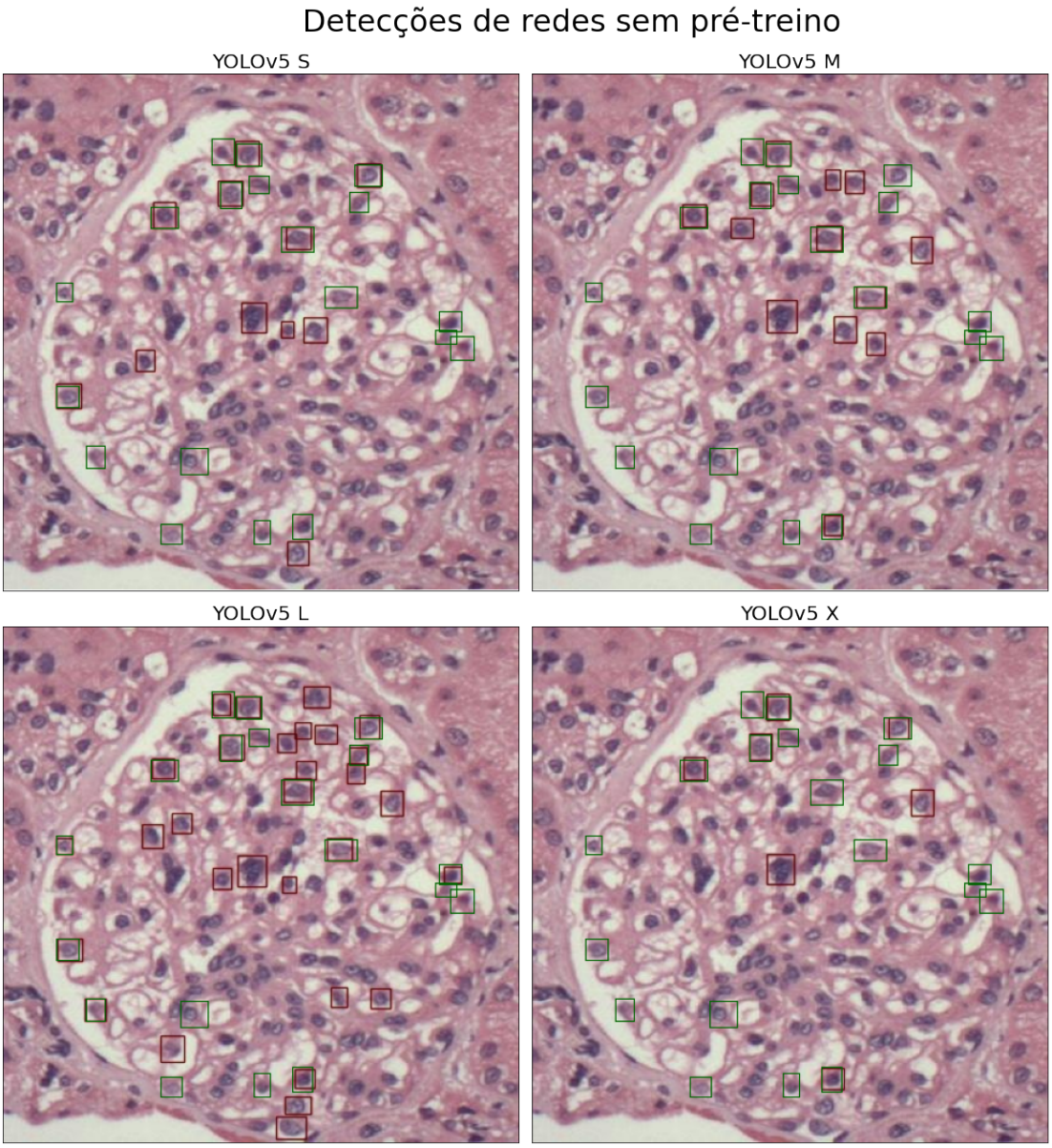


Figura 6.7: Exemplo de detecções das redes sem pré-treino na base de podócitos original

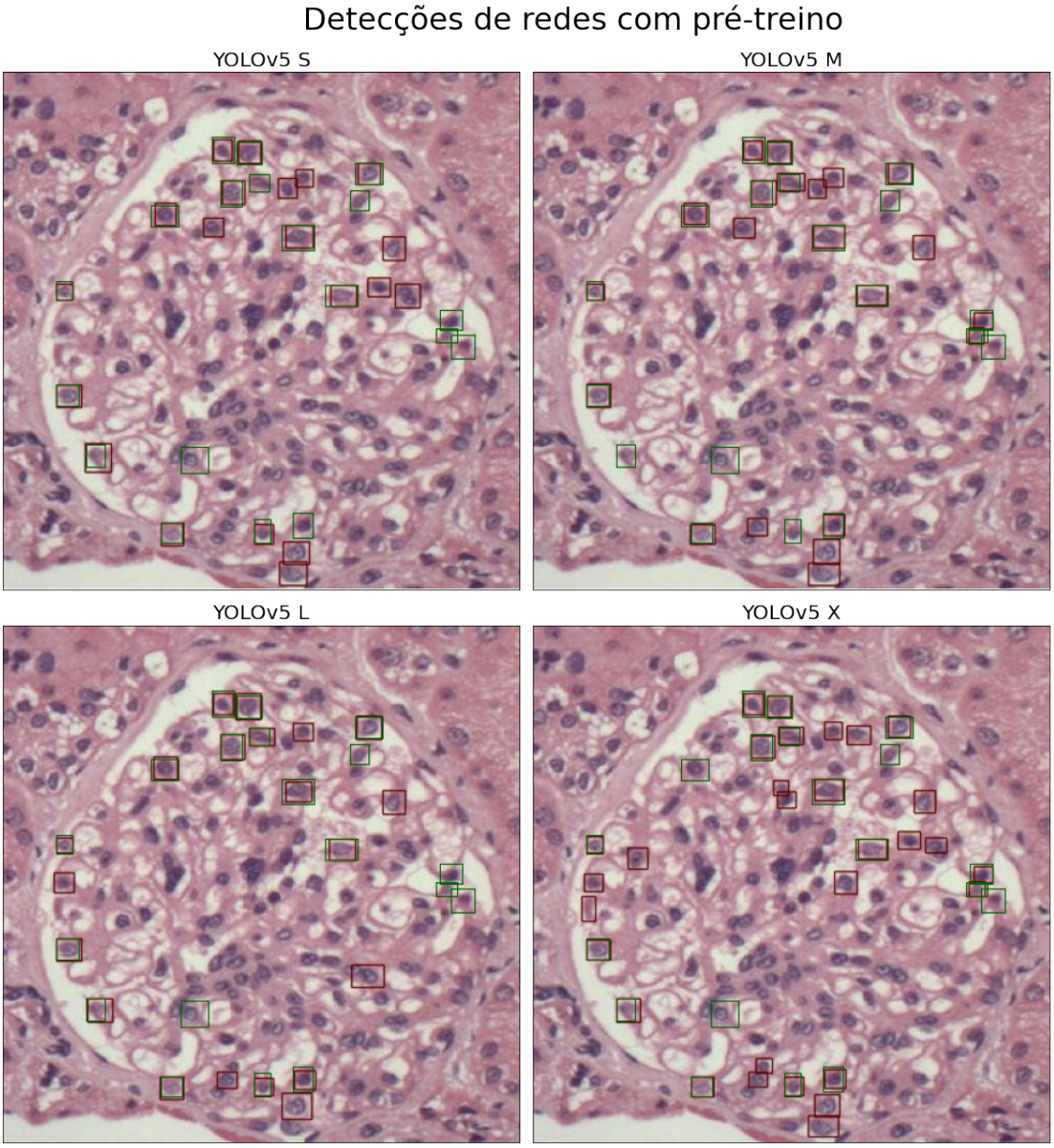


Figura 6.8: Exemplo de detecções das redes com pré-treino na base de podócitos original

6.4.2.2 Cenário 2: conjunto de imagens aumentado

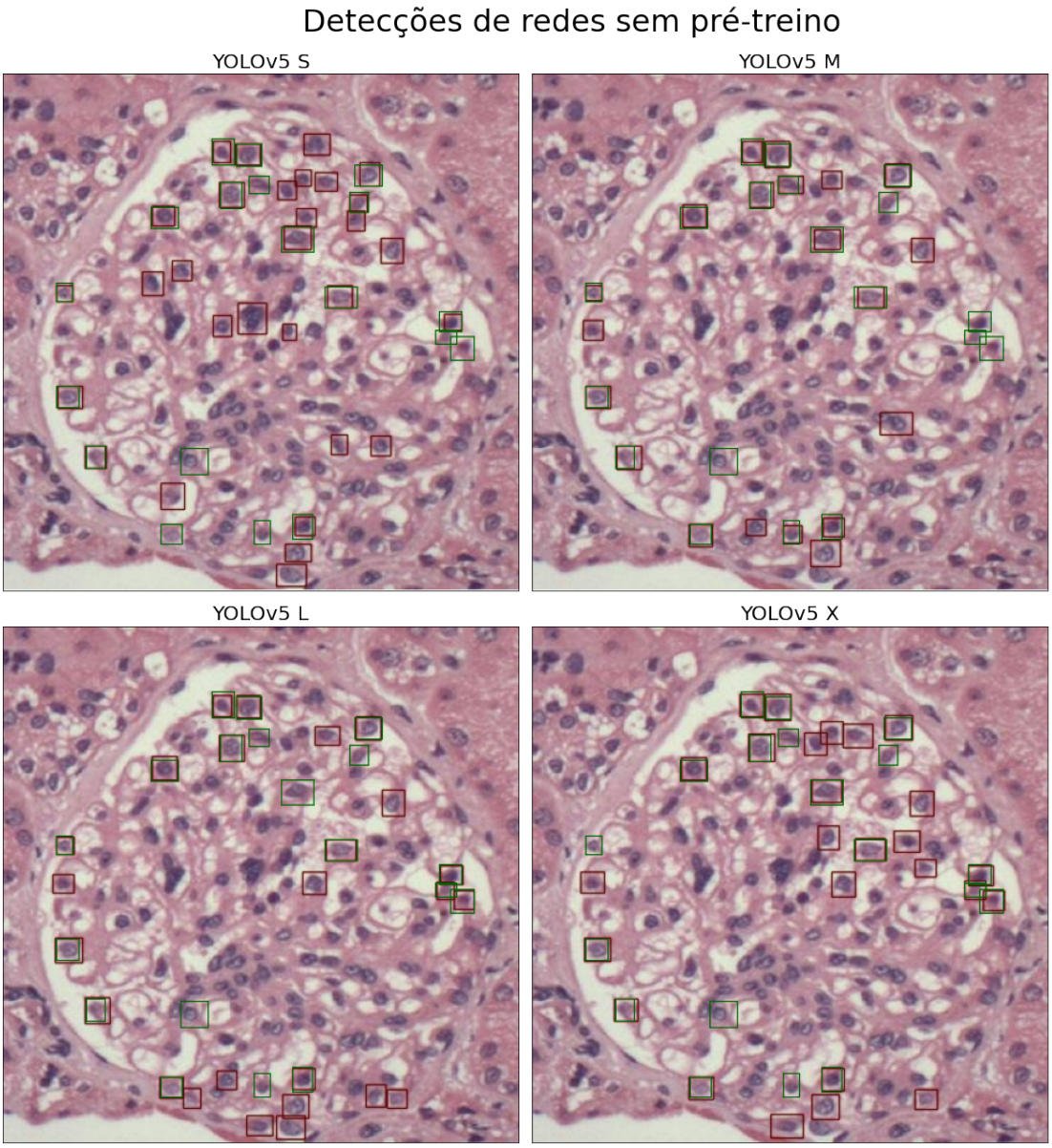


Figura 6.9: Exemplo de detecções das redes sem pré-treino na base de podócitos aumentada

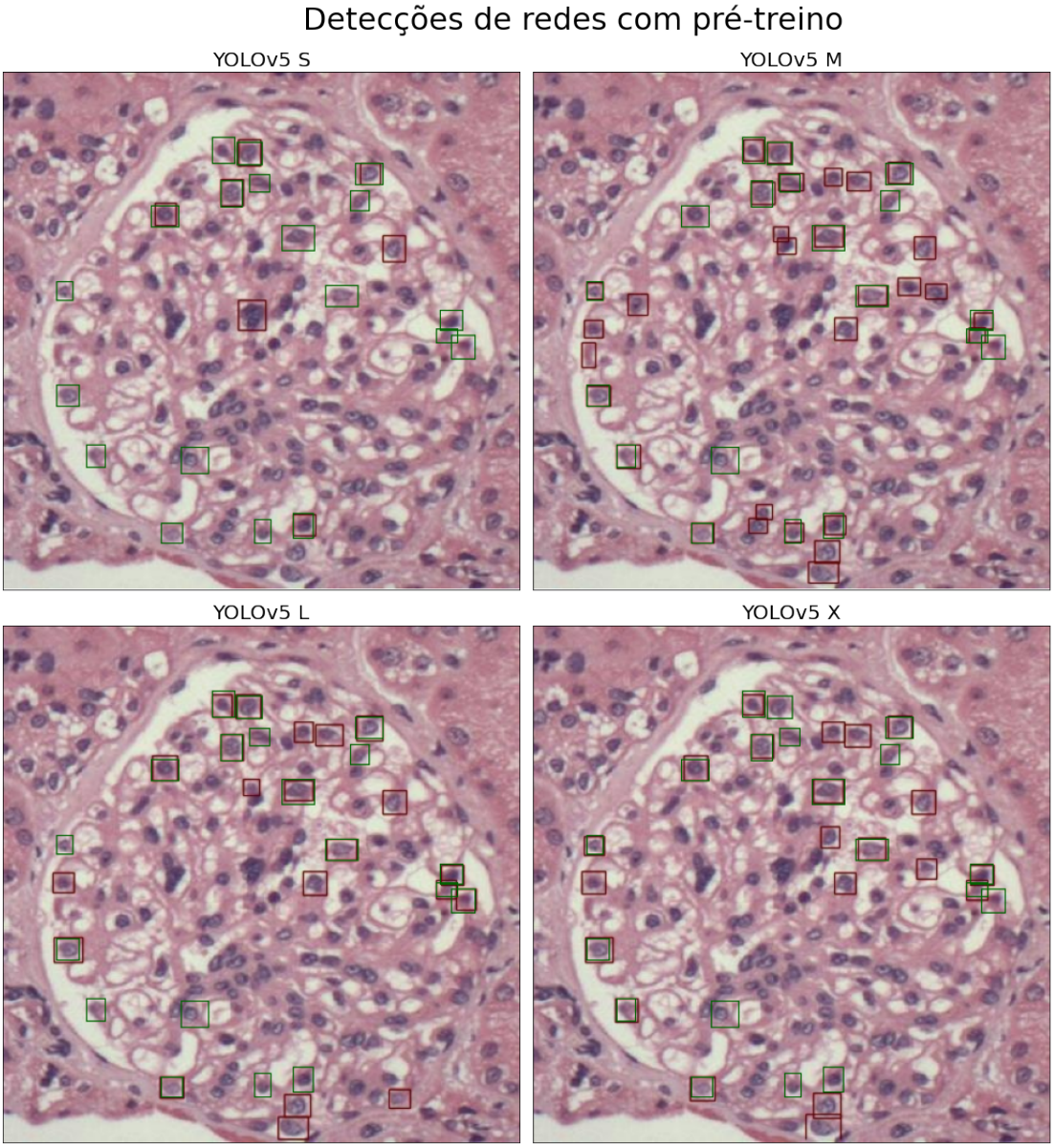


Figura 6.10: Exemplo de detecções das redes com pré-treino na base de podócitos aumentada

6.5 Gráficos envolvendo confiança

6.5.1 Base anotada de núcleos

Um ponto importante ao se fazer detecções com redes neurais é acompanhar o comportamento do nível de confiança. Idealmente, conforme mencionado na Seção 3.6, se deseja aumentar o limiar de confiança, sem que isto incorra em deterioração do *recall*. De forma a se avaliar o comportamento da rede com respeito ao nível de confiança, foram gerados dois gráficos para cada configuração: F1 *versus* confiança e Precisão *versus recall*. Os resultados se referem à última época do treinamento. Além disso, foram gerados gráficos de Precisão *versus* confiança e Recall *versus* confiança, os quais são omitidos em favor da concisão do texto, uma vez que o comportamento

destes se vê refletido na curva F1 *versus* confiança.

6.5.1.1 Cenário 1: conjunto de imagens original

As Figuras 6.11 e 6.18 mostram as curvas de F1 *versus* confiança e Precisão *versus recall* referentes à base de núcleos original

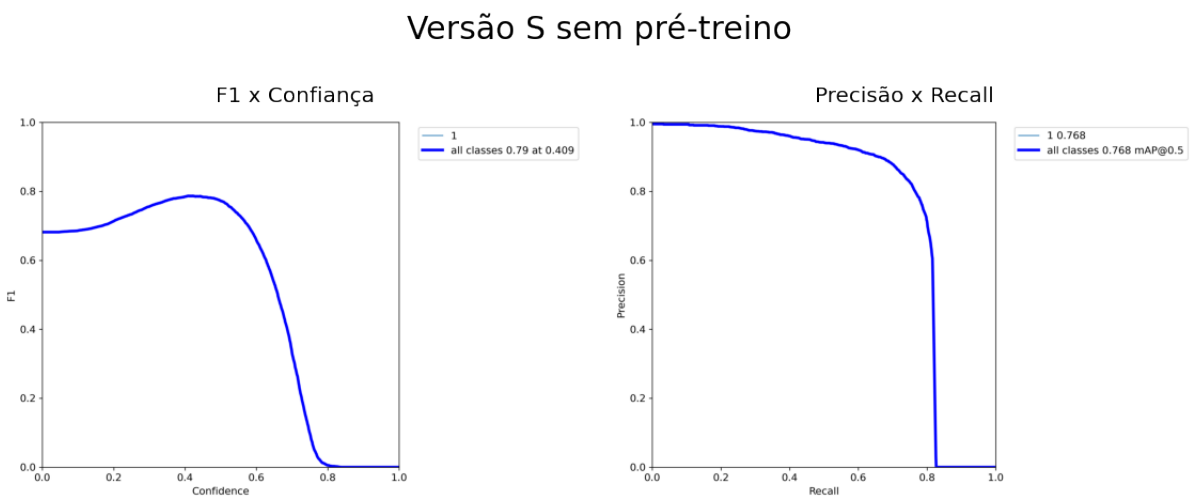


Figura 6.11: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 1A na base de núcleos original

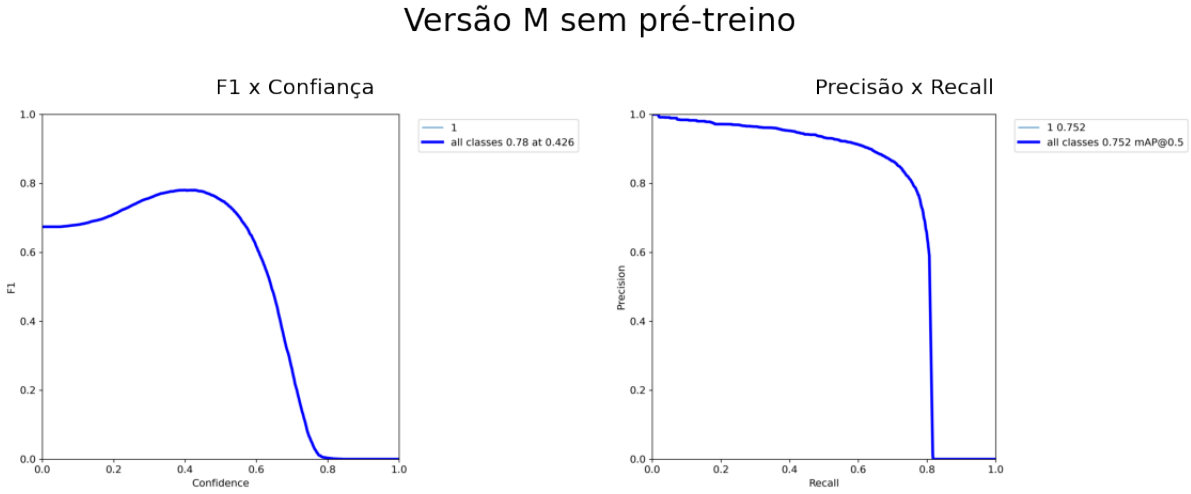


Figura 6.12: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 2A na base de núcleos original

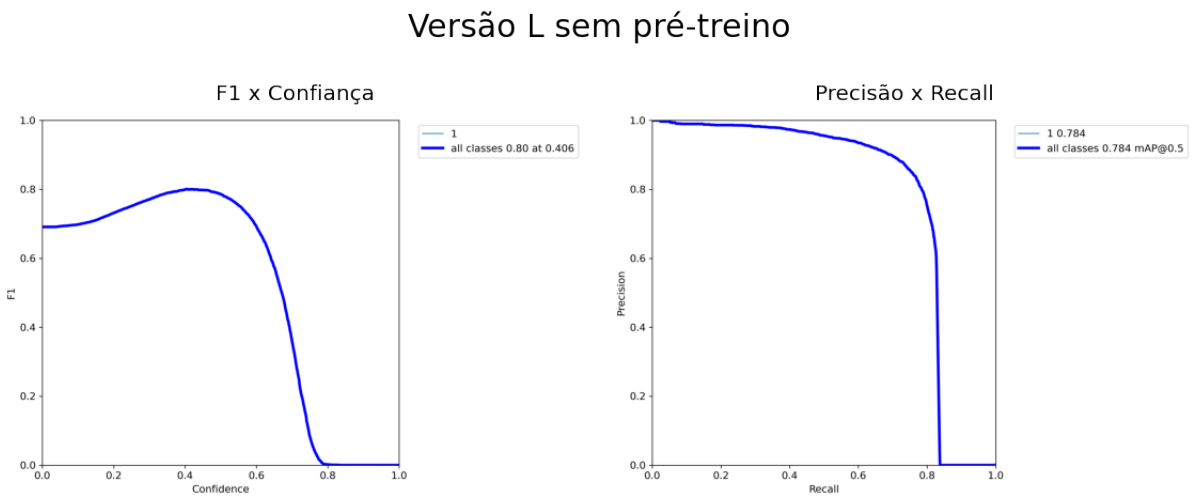


Figura 6.13: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 3A na base de núcleos original

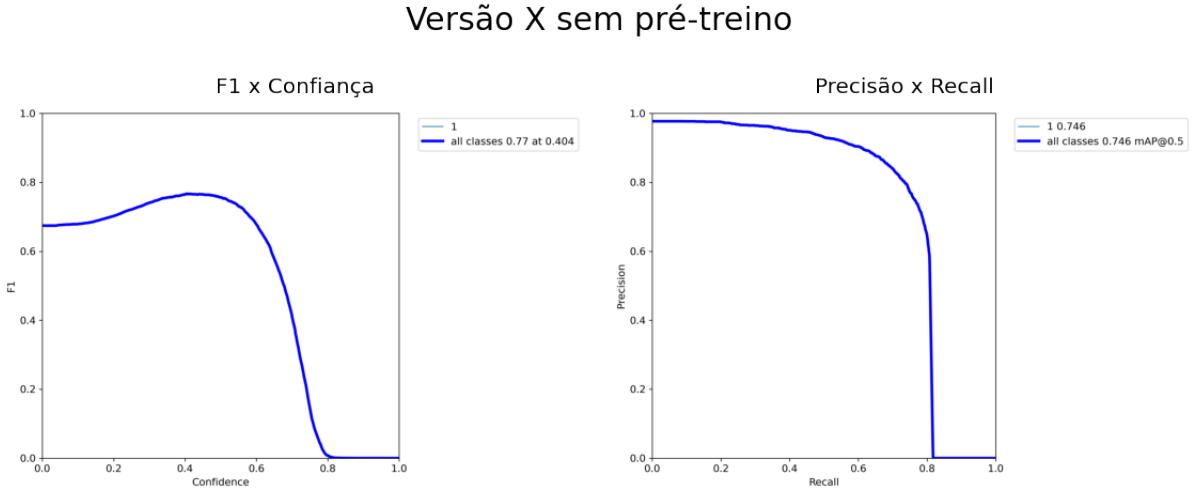


Figura 6.14: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 4A na base de núcleos original

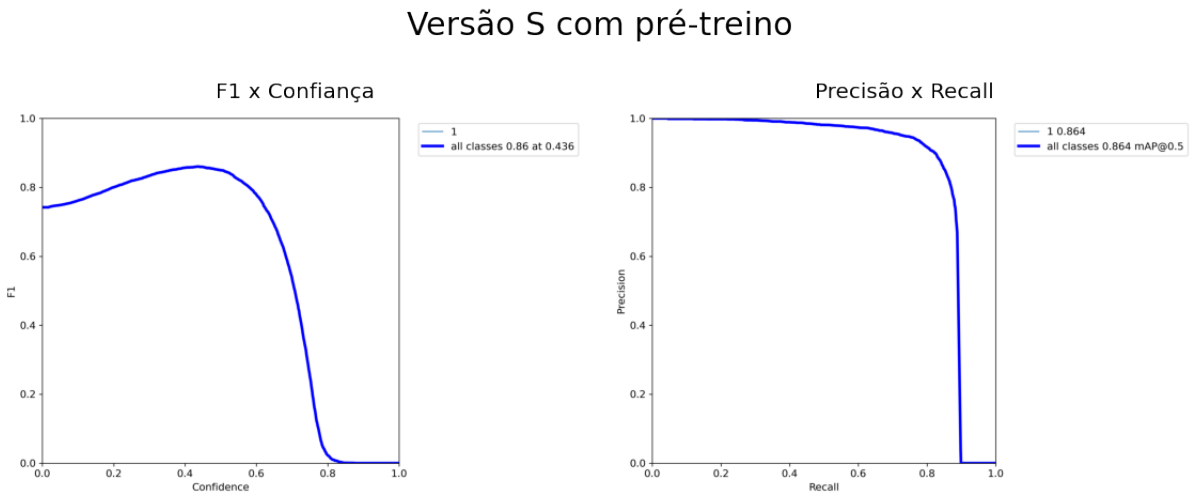


Figura 6.15: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 1B na base de núcleos original

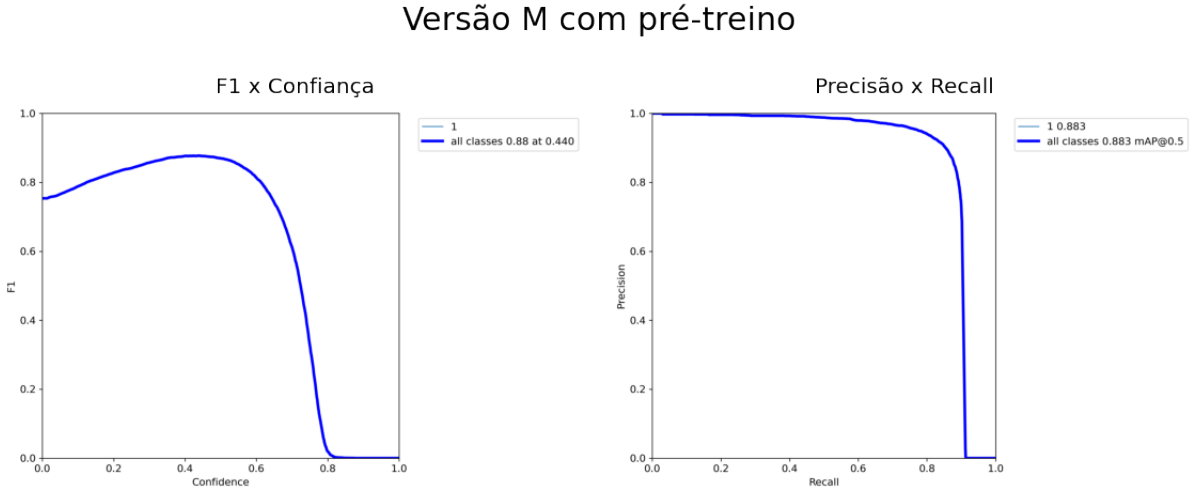


Figura 6.16: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 2B na base de núcleos original

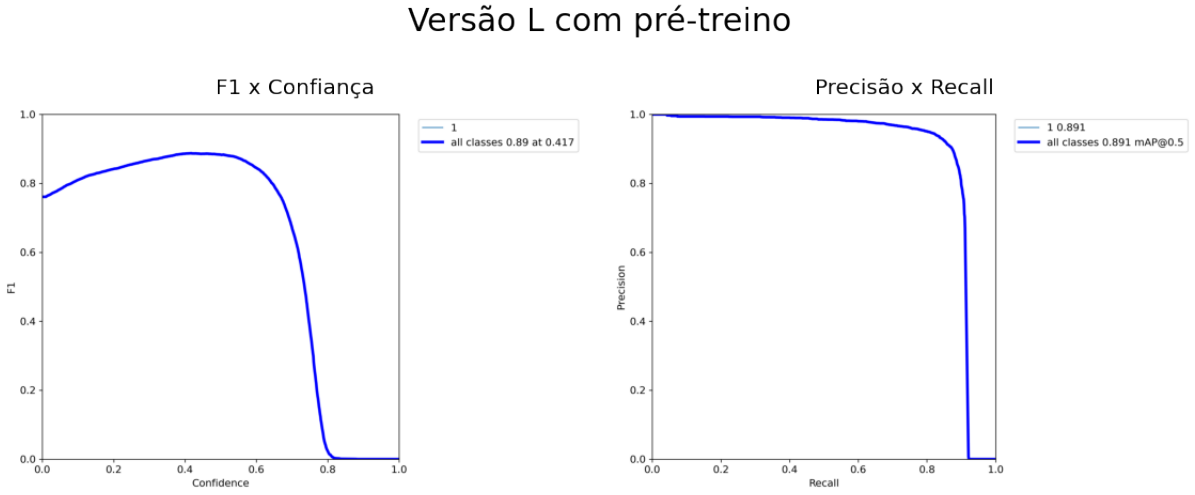


Figura 6.17: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 3B na base de núcleos original

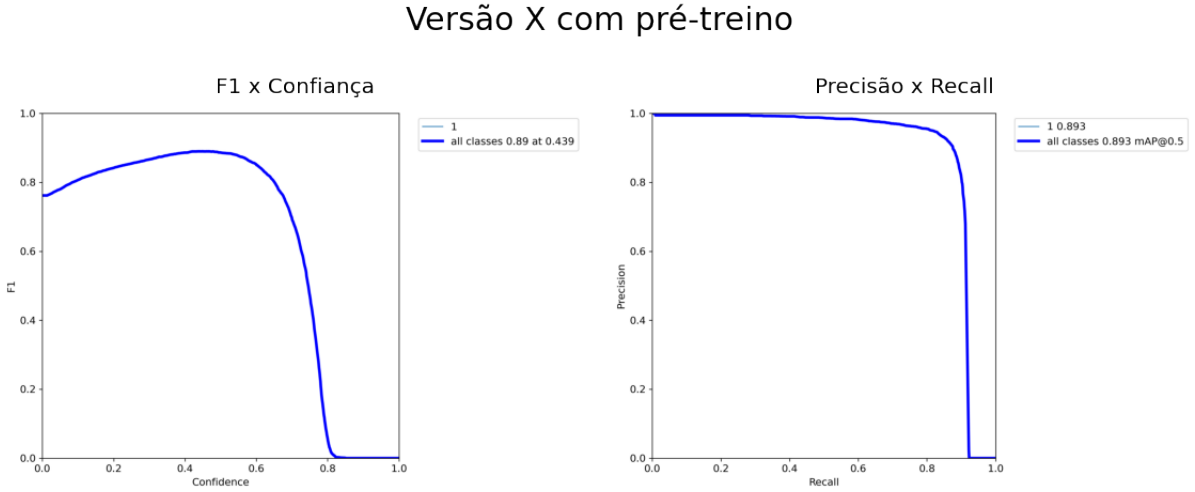


Figura 6.18: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 4B na base de núcleos original

6.5.1.2 Cenário 2: conjunto de imagens aumentado

As Figuras 6.19-6.26 mostram as curvas de F1 *versus* confiança e Precisão *versus recall* referentes à base de núcleos original

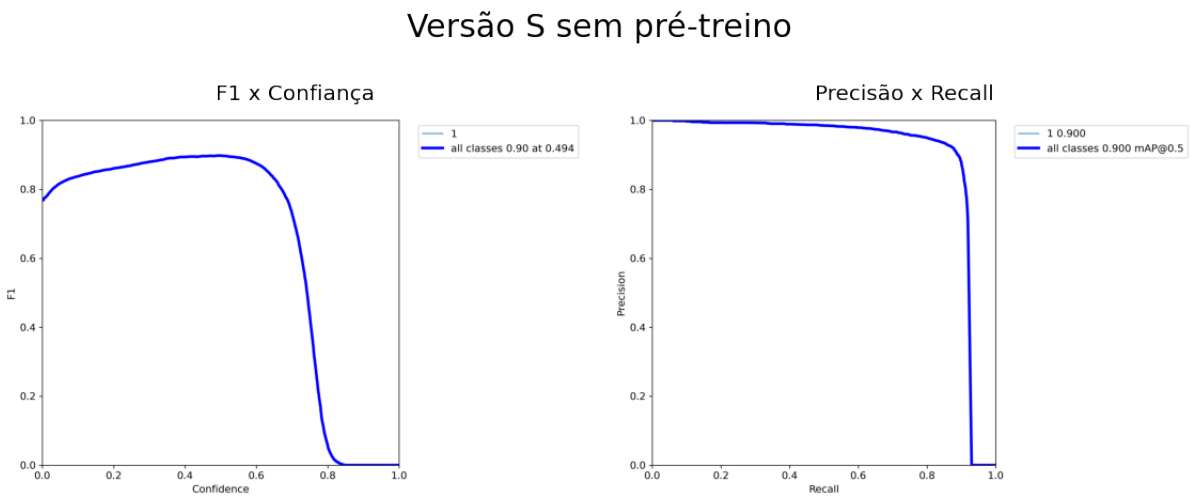


Figura 6.19: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 1A na base de núcleos aumentada

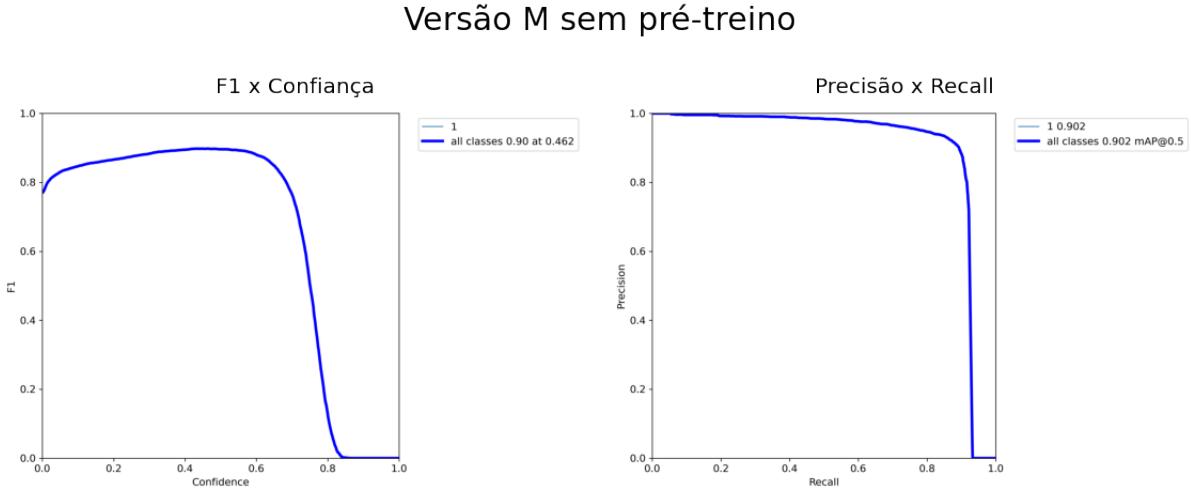


Figura 6.20: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 2A na base de núcleos aumentada

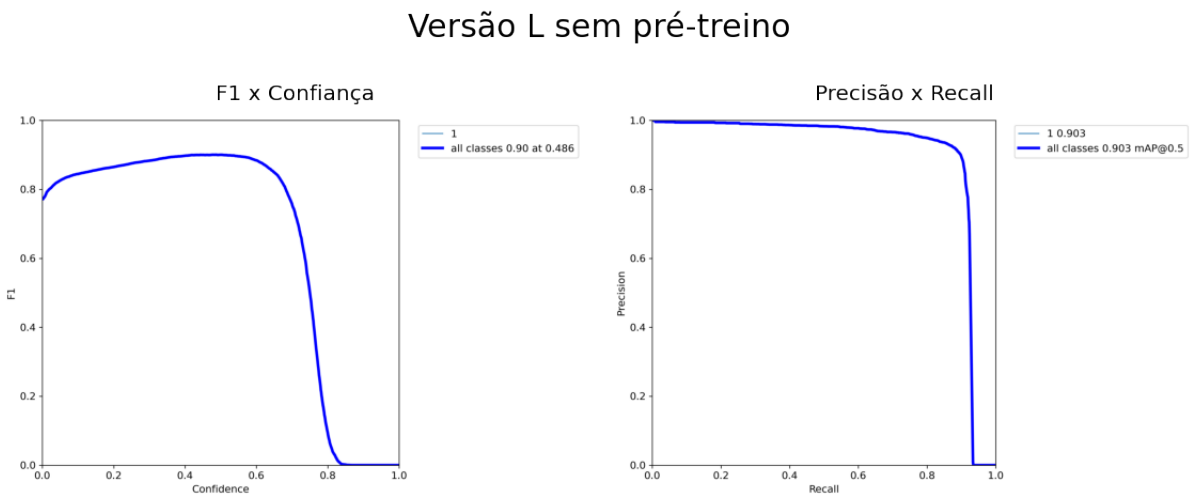


Figura 6.21: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 3A na base de núcleos aumentada

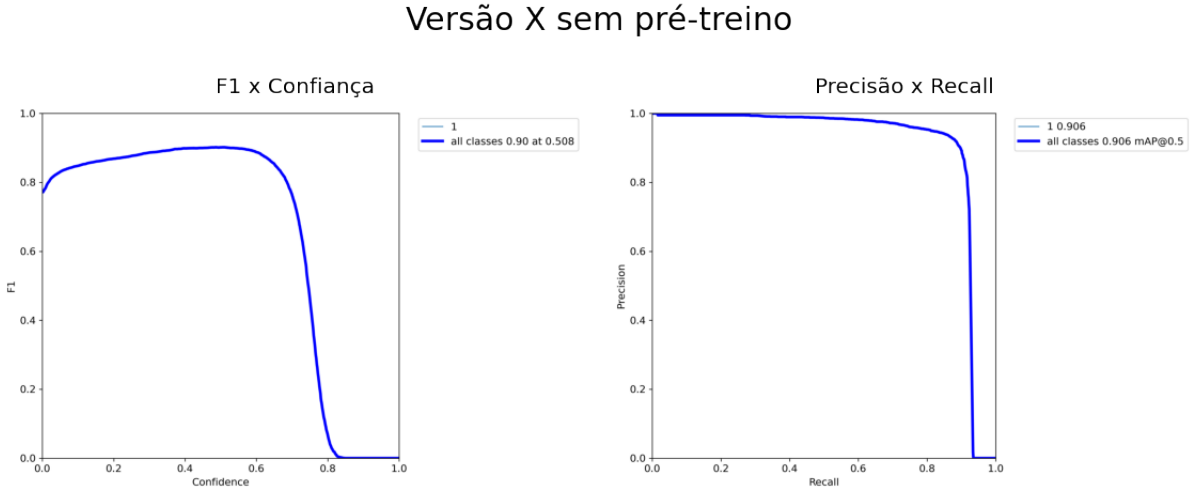


Figura 6.22: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 4A na base de núcleos aumentada

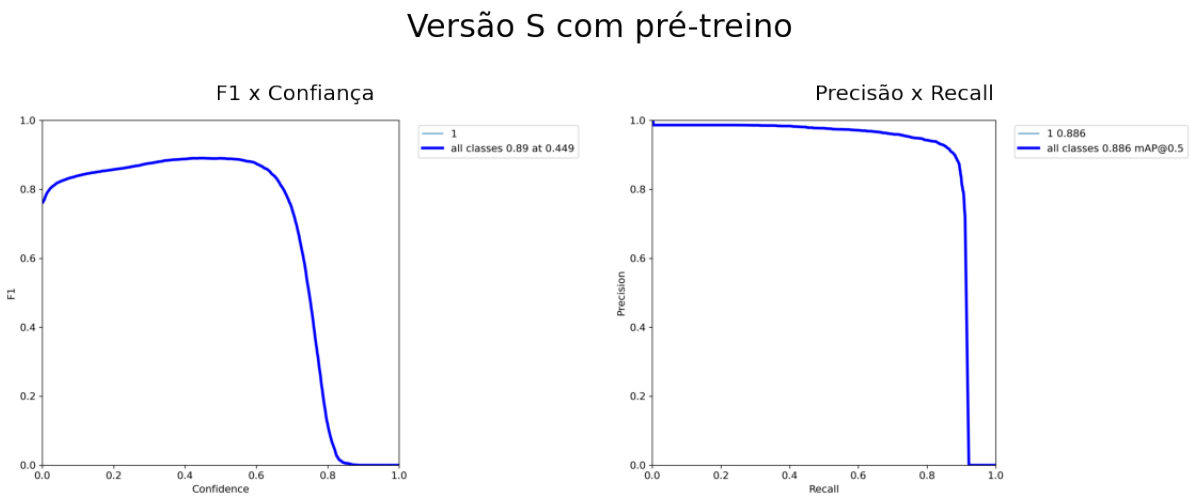


Figura 6.23: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 1B na base de núcleos aumentada

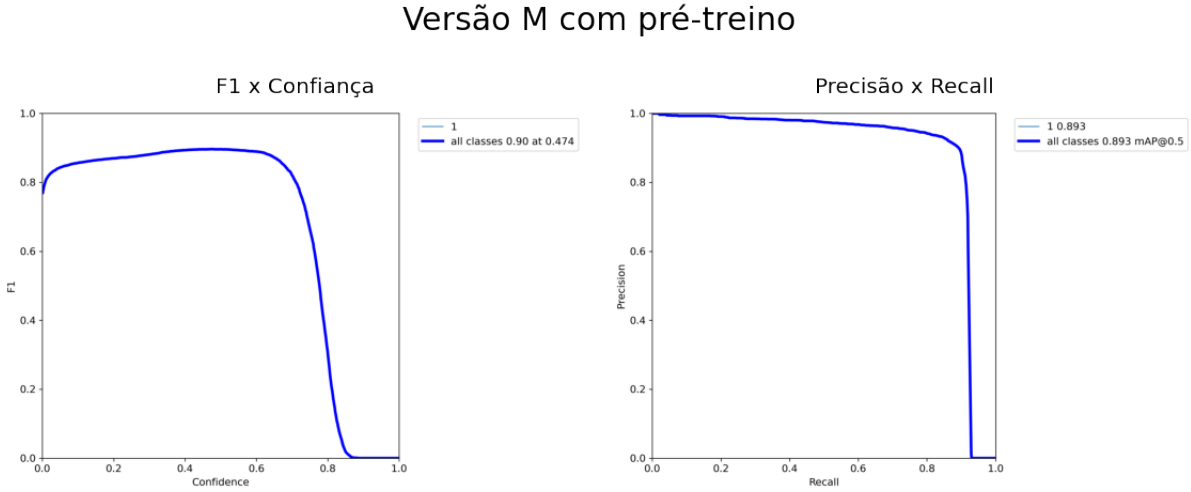


Figura 6.24: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 2B na base de núcleos aumentada

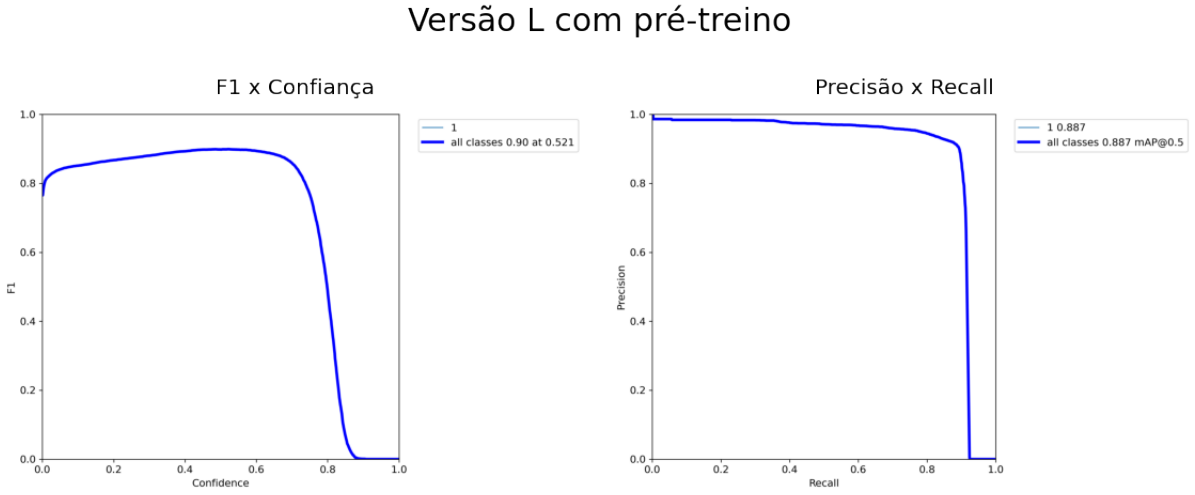


Figura 6.25: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 3B na base de núcleos aumentada

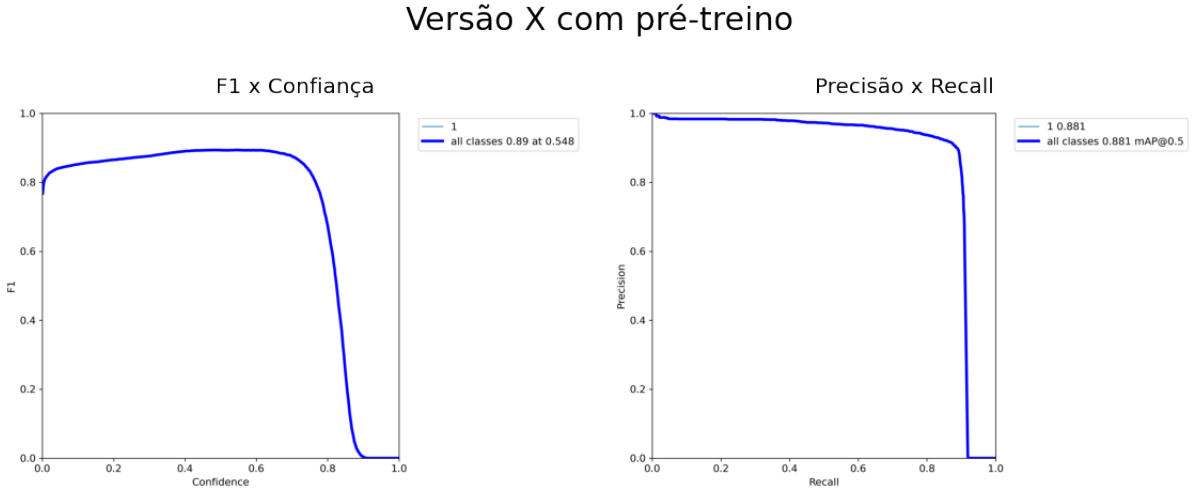


Figura 6.26: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 4B na base de núcleos aumentada

6.5.2 Base anotada de podócitos

6.5.2.1 Cenário 1: conjunto de imagens original

As Figuras 6.27-6.34 mostram as curvas de F1 *versus* confiança e Precisão *versus recall* referentes à base de podócitos original.

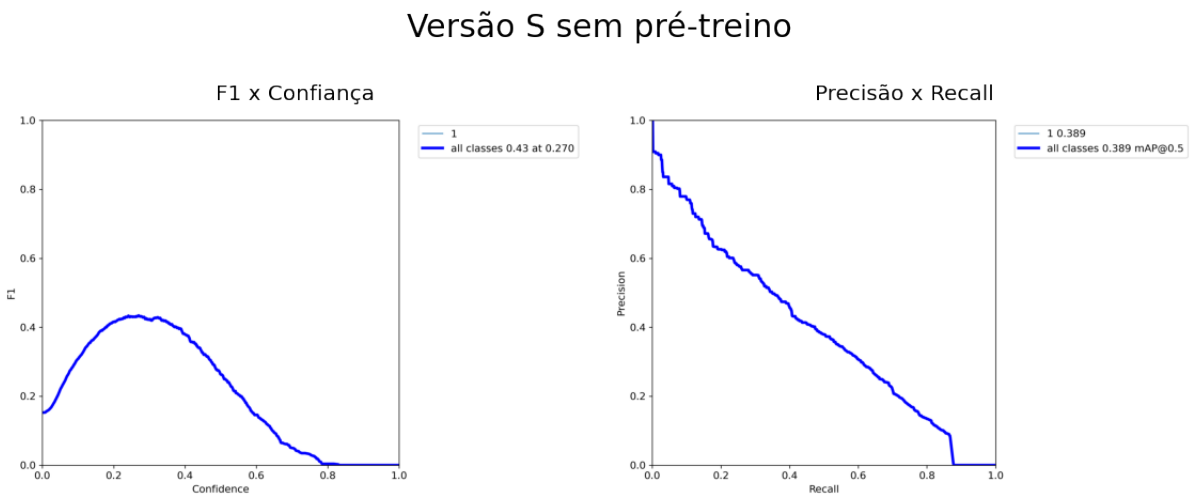


Figura 6.27: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 1A na base de podócitos original

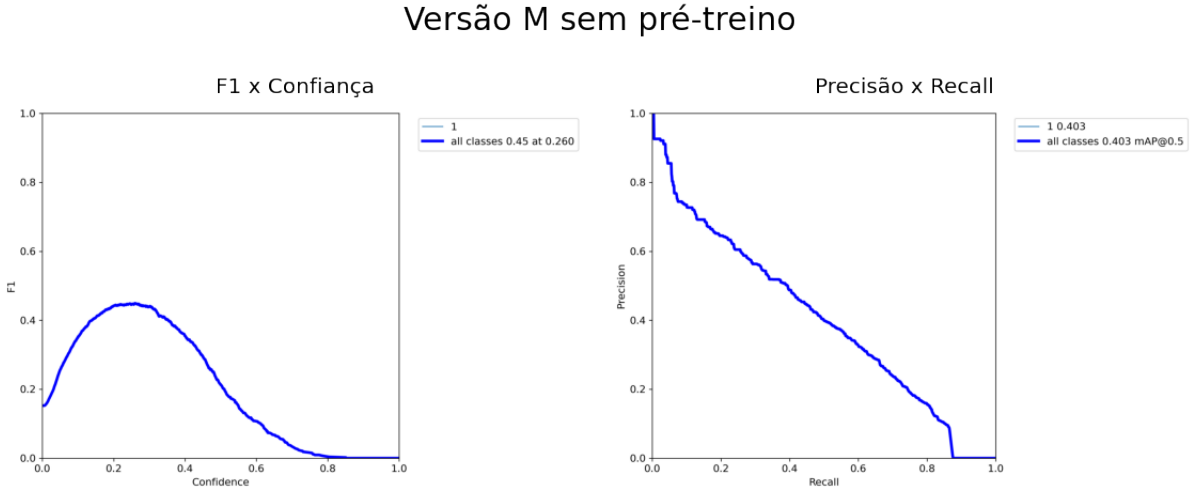


Figura 6.28: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 2A na base de podócitos original

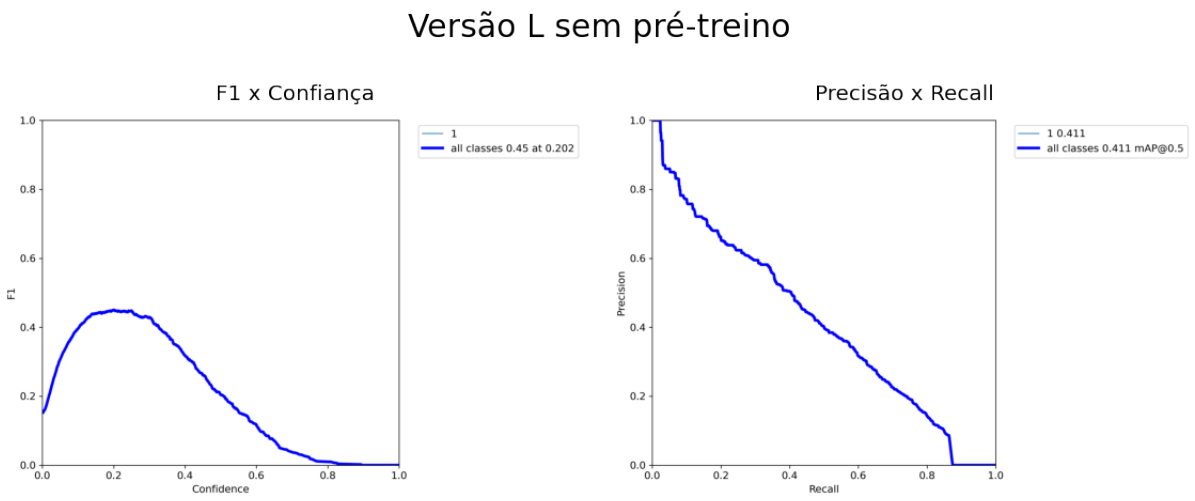


Figura 6.29: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 3A na base de podócitos original

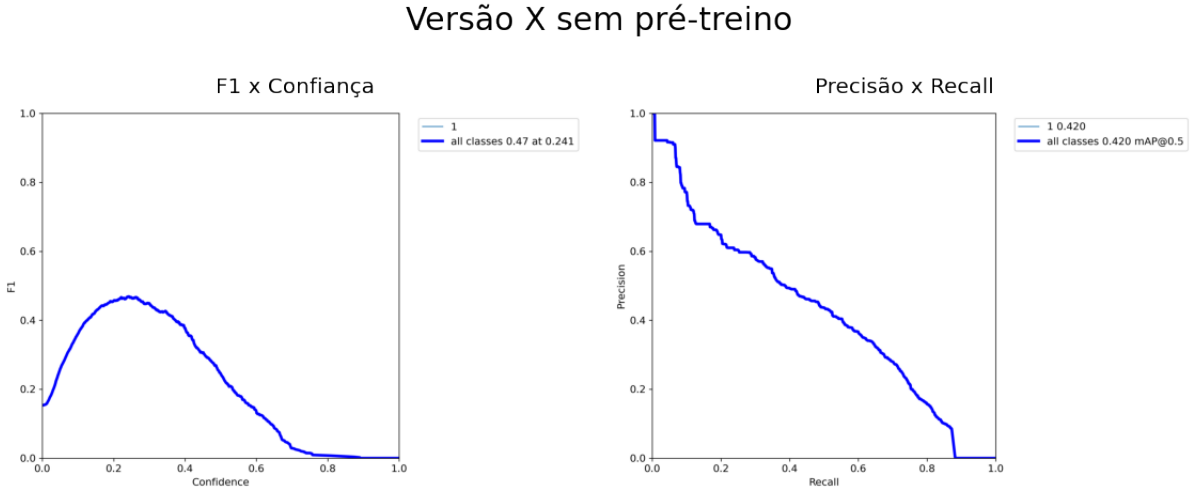


Figura 6.30: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 4A na base de podócitos original

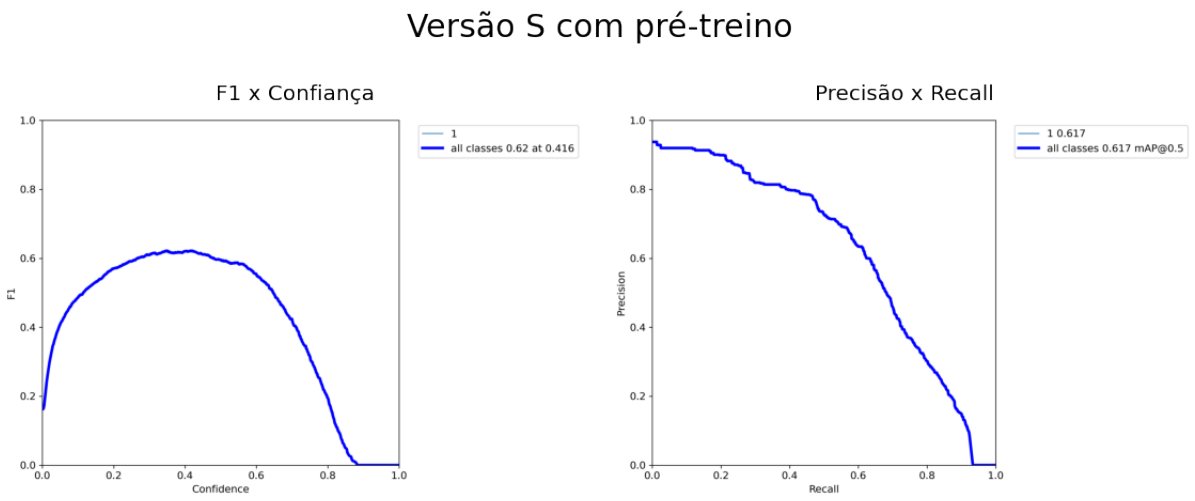


Figura 6.31: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 1B na base de podócitos original

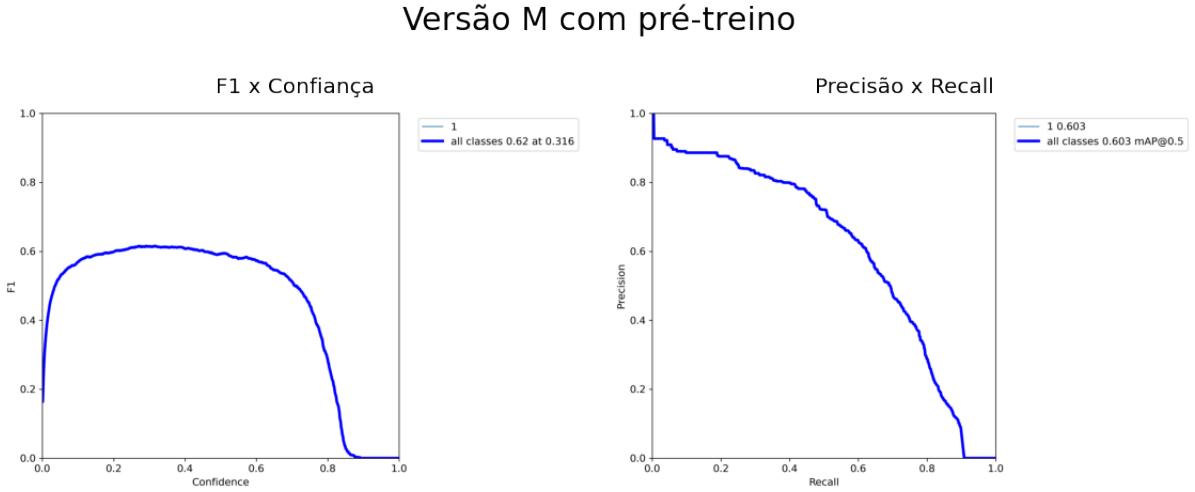


Figura 6.32: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 2B na base de podócitos original

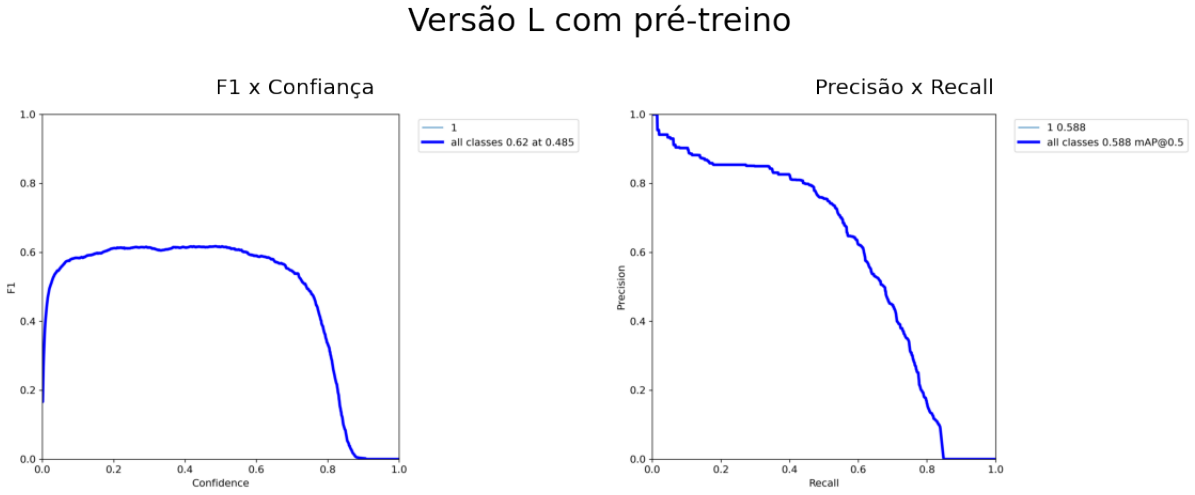


Figura 6.33: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 3B na base de podócitos original

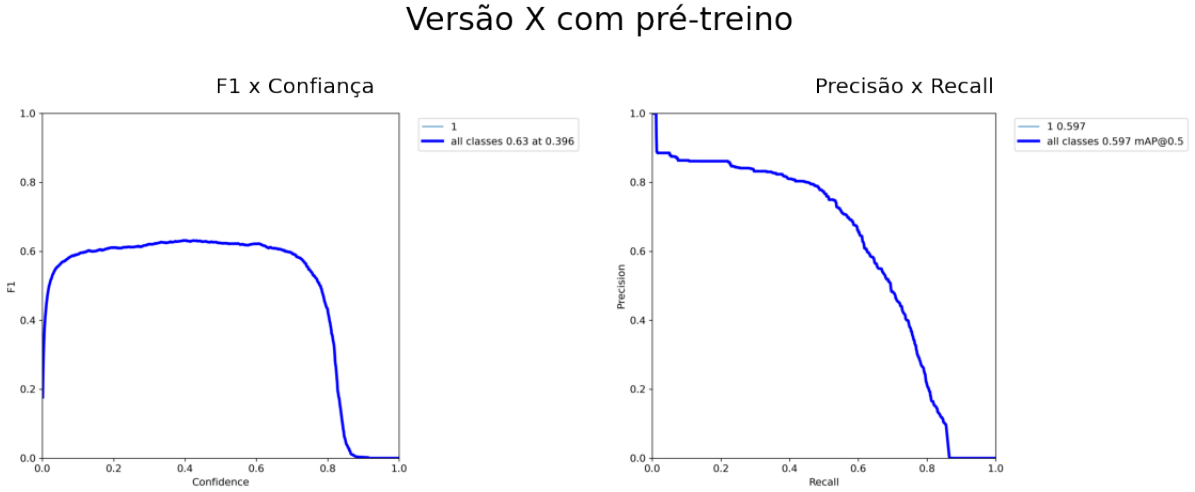


Figura 6.34: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 4B na base de podócitos original

6.5.2.2 Cenário 2: conjunto de imagens aumentado

As Figuras 6.35 e 6.42 mostram as curvas de F1 *versus* confiança e Precisão *versus recall* referentes à base de podócitos aumentada.

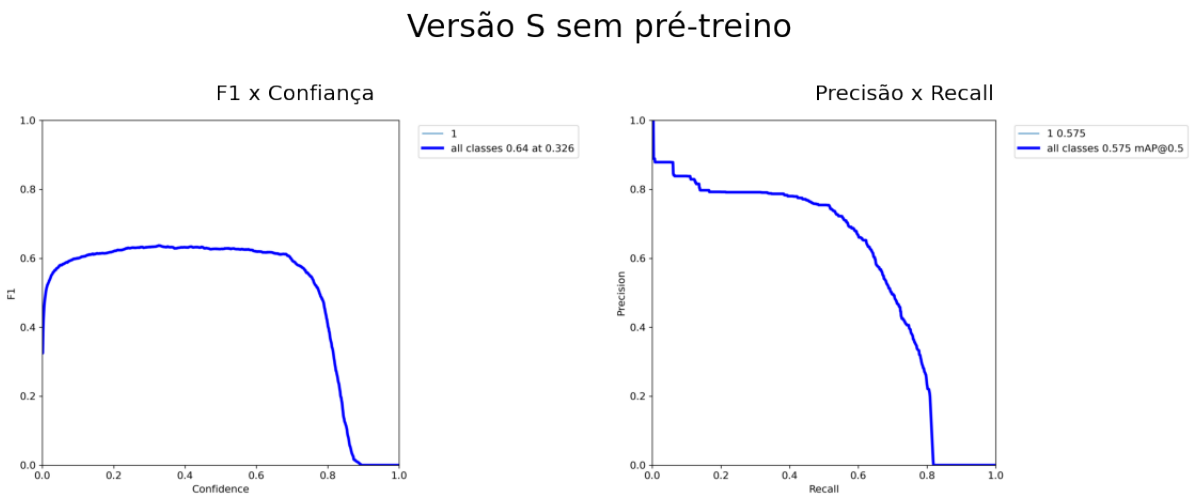


Figura 6.35: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 1A na base de podócitos aumentada

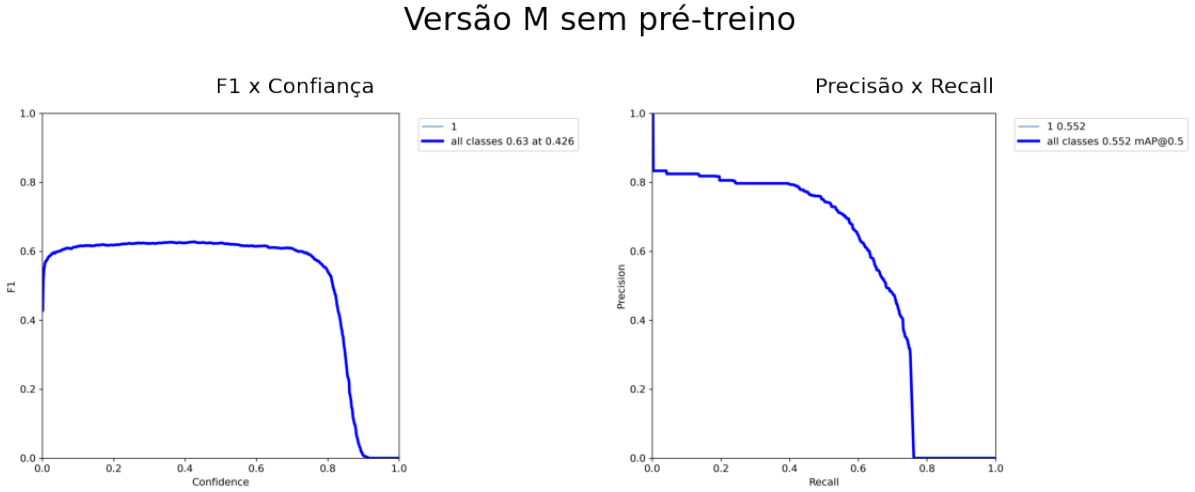


Figura 6.36: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 2A na base de podócitos aumentada

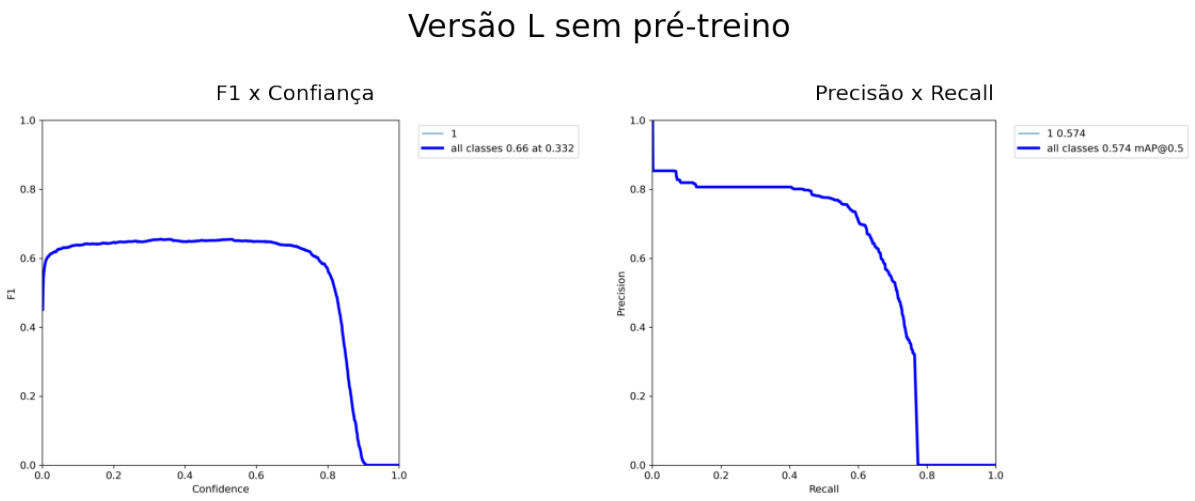


Figura 6.37: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 3A na base de podócitos aumentada

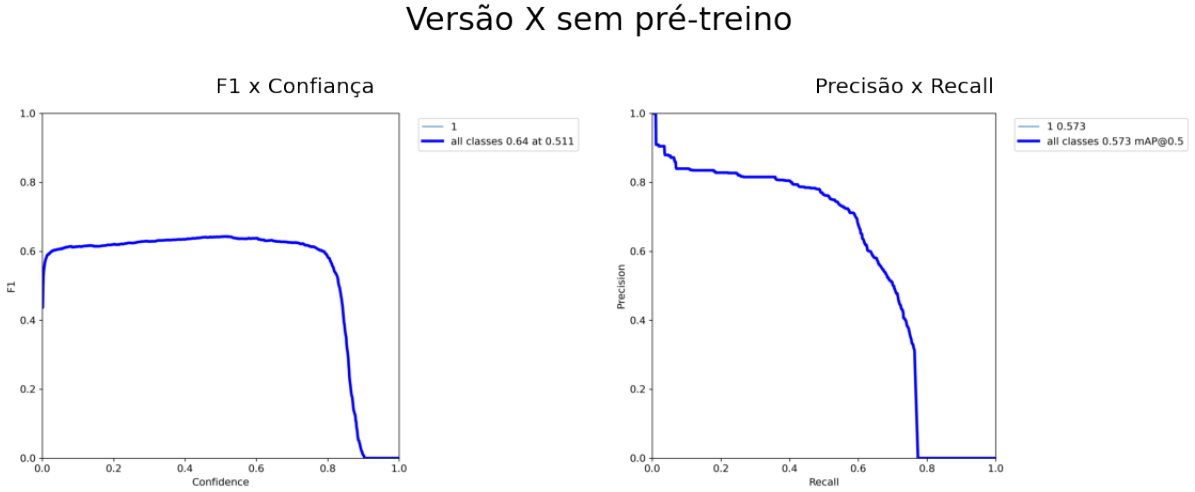


Figura 6.38: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 4A na base de podócitos aumentada

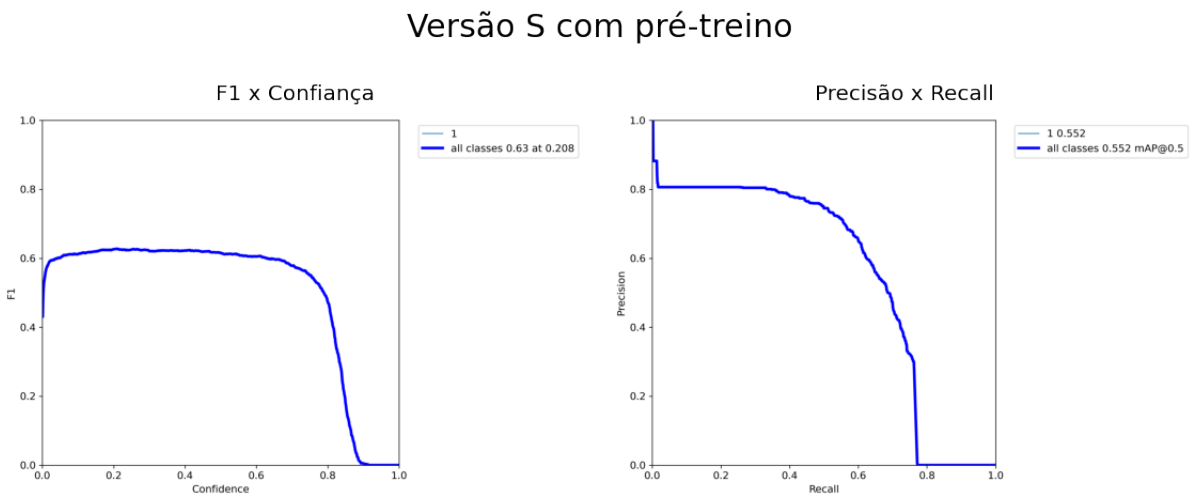


Figura 6.39: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 1B na base de podócitos aumentada

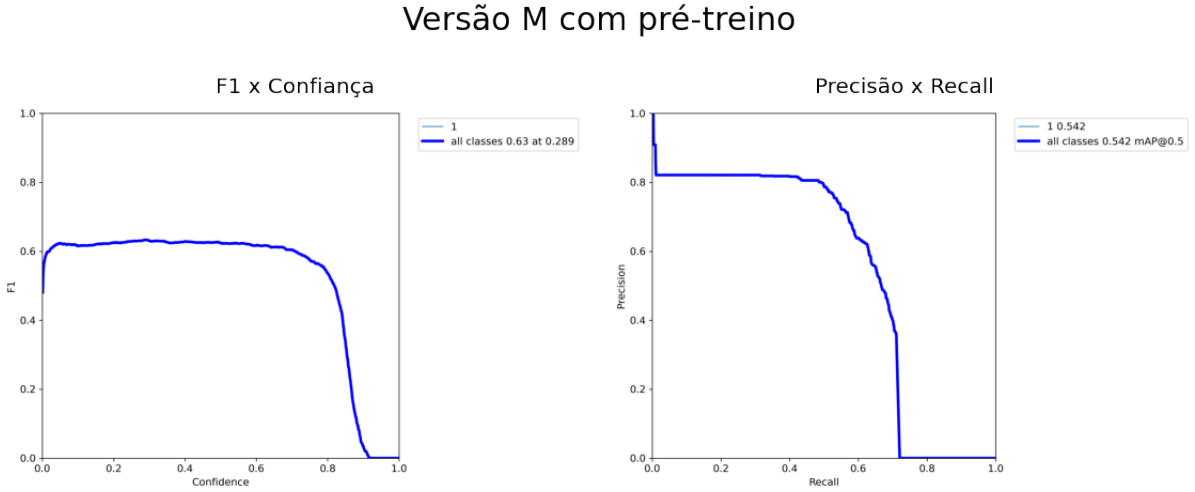


Figura 6.40: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 2B na base de podócitos aumentada

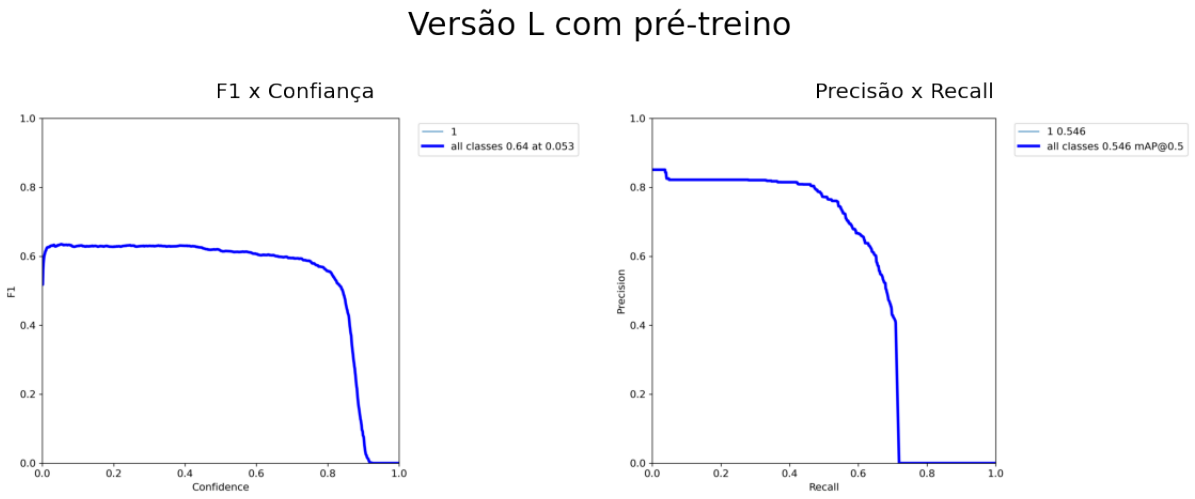


Figura 6.41: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 3B na base de podócitos aumentada

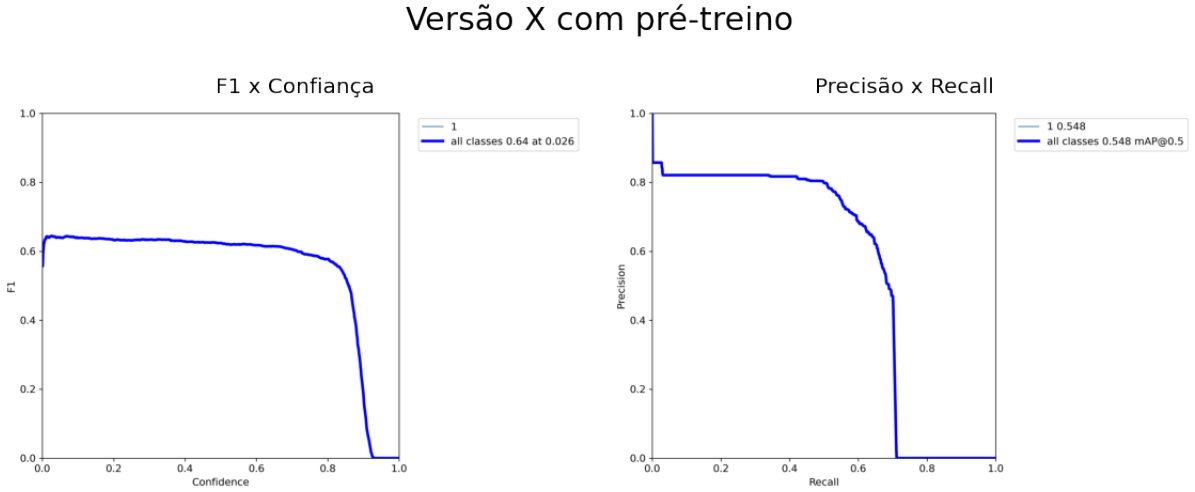


Figura 6.42: Gráficos de F1 x confiança e Precisão x *Recall* para a configuração 4B na base de podócitos aumentada