

Avaliação de Modelos de Detecção de Objetos para Detectar Glomérulos em Imagens Histológicas

Abel Ramalho Galvão
Departamento de Ciências Exatas
Univ. Estadual de Feira de Santana
Feira de Santana, BA, Brasil
Email: abel.ramalho18@gmail.com

Jonathan M. C. Rehem
Univ. Estadual de Feira de Santana
Feira de Santana, BA, Brasil
Email: jonathanmoreirac@gmail.com

Washington L. C. dos Santos
Centro de Pesquisas Gonçalo Muniz
Fundação Oswaldo Cruz
Salvador, BA, Brasil
Email: wlcdossantos@gmail.com

Luciano Rebouças de Oliveira
IVISION Lab
Univ. Federal da Bahia
Salvador, BA, Brasil
Email: luciano.reboucas@gmail.com

Angelo Amâncio Duarte
Departamento de Tecnologia
Univ. Estadual de Feira de Santana
Feira de Santana, BA, Brasil
Email: angeloduarte@uefs.br

Michele Fúlvia Angelo
Departamento de Ciências Exatas
Univ. Estadual de Feira de Santana
Feira de Santana, BA, Brasil
Email: mfangelo@ecomp.uefs.br

Abstract—Glomeruli are renal structures responsible for filtering blood and can be affected by lesions. Currently, computer systems to help identify these lesions have been developed, and thus, the detection of these glomeruli is of great importance. The objective of this work is to evaluate the performance of object detection models for the detection of glomeruli in digital histological images. Three models were evaluated: SM1 (SSD Mobilenet v1), FRR50 (Faster RCNN Resnet 50) and FRR101 (Faster RCNN Resnet 101), of which the FRR50 model obtained the best result, mAP=0.88.

Resumo—Os glomérulos são estruturas renais responsáveis pela filtragem do sangue e podem ser acometidos por lesões. Atualmente, sistemas computacionais para auxiliar na identificação destas lesões têm sido desenvolvidos, e assim, é de grande importância a detecção destes glomérulos. O objetivo deste trabalho é avaliar o desempenho de modelos de detecção de objetos para a detecção de glomérulos em imagens histológicas digitais. Foram avaliados 3 modelos: SM1 (SSD Mobilenet v1), FRR50 (Faster RCNN Resnet 50) e FRR101 (Faster RCNN Resnet 101), dos quais, o modelo FRR50 obteve o melhor resultado, mAP=0.88.

I. INTRODUÇÃO

Uma doença renal pode afetar a capacidade do corpo humano de manter o equilíbrio dos componentes presentes do sangue, permitindo a circulação de impurezas e toxinas, além de afetar os componentes funcionais do sangue, como glóbulos vermelhos, glóbulos brancos e proteínas. Os glomérulos são uma microestrutura de grande importância para o rim. Essa microestrutura é formada por capilares que formam uma espécie de bola de lã encontrada no néfron e é a unidade funcional do rim [1].

As glomerulopatias são doenças que afetam diretamente os glomérulos e dependendo do tipo de lesão e da quantidade de glomérulos afetados, elas podem gerar consequências graves para o funcionamento do organismo, podendo levar o paciente a depender de tratamento permanente ou mesmo levá-lo a óbito [2].

O diagnóstico das glomerulopatias é realizado por meio de critérios clínicos, exames laboratoriais e biópsia. A biópsia

renal, que é o procedimento que gera as imagens histológicas usadas neste trabalho, consiste na extração de um fragmento de 1 a 2cm do rim, por meio de agulha específica ou por meio de cirurgia. Em seguida, o fragmento de tecido é fixado quimicamente, evitando sua decomposição, cortado em cortes de 2-3 μm e colocados em lâminas que serão examinadas por um patologista em microscópio [3]. Atualmente, estas lâminas podem ser digitalizadas ou fotografadas, tornando possível a manipulação através de computadores [4]. Esta evolução permitiu o desenvolvimento de ferramentas que podem auxiliar médicos patologistas na automatização de tarefas que podem abreviar o diagnóstico, contribuindo com a melhora na eficiência do tratamento, além de otimizar o uso do tempo de análise pelos médicos.

Para auxiliar especialistas da área de patologia renal, um sistema denominado PathoSpotter [5]–[8], vem sendo desenvolvido por meio de uma parceria entre profissionais da área médica e de computação. O PathoSpotter, atualmente, é capaz de classificar glomerulopatias proliferativas com acurácia de 88.3 [9] e glomerulosclerose segmentar com acurácia de 84.8 para imagens com corante HE (Hematoxilina e Eosina) e 81.3 para imagens com corante PAS (*Periodic Acid Schiff*) [10], porém estes classificadores atuam sobre imagens que contém somente um glomérulo por imagem. Assim, em 2021, um trabalho de detecção de glomérulos foi desenvolvido [11].

Rehem *et al.* [11] propuseram uma metodologia para detecção automática de glomérulos em imagens histológicas digitais utilizando técnicas de *deep learning*. Eles usaram os modelos SSD (*Single Shot Detection*) Inception v2 (SI2) e Faster RCNN (*Region-based Convolutional Neural Network*) Inception Resnet v2 (FRI2) para a detecção de glomérulos e obtiveram 0,8831 e 0,8723 de mAP, (*Mean Average Precision*), respectivamente.

Assim, o objetivo deste trabalho é avaliar o desempenho de outros modelos de detecção de objetos, além daqueles já utilizados por Rehem *et al.* [11]), para a detecção de glomé-

culos em imagens histológicas digitais de baixa resolução e de aquisição não padronizada.

II. TRABALHOS RELACIONADOS

Gallego *et al.* [12] propuseram um método de classificação e detecção de glomérulos utilizando imagens WSI (*Whole-Slide Image*). Com o auxílio de redes CNN (*Convolutional Neural Network*) foram capazes de identificar duas classes de objetos, glomérulos e não-glomérulos. O método obteve 0.881 de *Precision*, 1.0 de *Recall* e 0.937 de *F1 Score*.

Simon *et al.* [13] propuseram um método automatizado para a localização de glomérulos em WSI. O método proposto extraiu o vetor de atributos das imagens utilizando mrcLBP (*Multi Radial Color LBP*), que é uma adaptação do descritor visual LBP (*Local Binary Pattern*). Foram utilizadas amostras de humanos, ratos e camundongos, conseguindo mais de 0.9 de precisão, e mais de 0.7 de *Recall*.

Kawazoe *et al.* [14] propuseram a utilização do método *Faster R-CNN-based* em conjunto com o modelo pré-treinado *Inception-ResNet*, para detectar glomérulos a partir de um *dataset* com 33 mil glomérulos anotados em WSI. O trabalho teve como finalidade comparar a metodologia proposta com outros modelos de detecção de glomérulos. O método obteve 0.931 de *Precision*, 0.919 de *Recall* e 0.925 de *F1 Score*.

III. MATERIAIS E MÉTODOS

Para o desenvolvimento deste trabalho está sendo utilizada a linguagem de programação Python, versão 3.6, em conjunto com o *framework* TOD (*Tensorflow Object Detection API*), versão 1.15.0.

Devido à necessidade de alto desempenho computacional, para o treinamento dos modelos foi utilizado um cluster com processador Intel (R) Xeon (R) CPU E5-2630 v4 @ 2,20 GHz, 128 GB de RAM e 4 GPUs Tesla K80. O sistema operacional utilizado foi o Ubuntu Linux, versão 16.04.4 LTS (Xenial Xerus).

O desenvolvimento deste trabalho foi dividido nas etapas de: (a) Definição do *Dataset*; (b) Escolha dos Modelos; (c) Treinamento dos Modelos; (d) Avaliação dos Modelos.

A. Definição do Dataset

Neste trabalho, o *dataset* utilizado foi o mesmo usado por Rehem *et al.* [11]. Ele é formado por um conjunto de 909 imagens fornecidas pelo Centro de Pesquisas Gonçalo Muniz da Fundação Oswaldo Cruz (CpqGM/FIOCRUZ). Tratam-se de imagens de tecido humano com glomérulos saudáveis e acometidos por glomerulopatia membranosa e glomerulosclerose segmentar. As imagens utilizadas são de tecidos corados com H&E (*Hematoxilina e Eosina*) e PAS (*Periodic Acid Schiff*). O método de aquisição das imagens foi através de fotografia digital, utilizando uma câmera fotográfica digital acoplada a um microscópio óptico. A resolução espacial média das imagens foi de 0,798 megapixel, sendo a menor de 0,30 megapixel e a maior resolução de 3,14 megapixels. As imagens foram armazenadas nos formatos: JPEG, GIF e TIFF, todas no modelo de cor RGB (*Red, Green and Blue*).

O *dataset* foi dividido em:

1) *Dataset de treino*: 509 imagens anotadas, das quais 209 imagens contêm glomérulos saudáveis, 150 glomérulos afetados por glomerulopatias membranosas e 150 imagens contendo glomérulos afetados por glomerulosclerose segmentar.

2) *Dataset de validação*: Para verificar o desempenho dos modelos durante o treinamento, o *dataset* de validação utiliza 200 imagens, das quais 100 imagens contêm glomérulos saudáveis, 50 imagens contêm glomerulopatias membranosas e 50 imagens contendo glomerulosclerose segmentar.

3) *Dataset de testes*: este *dataset* não é utilizado durante o treinamento do modelo, e sim após o treino. Seu objetivo é verificar a detecção de objetos com amostras nunca antes vistas por ele. Para isso, estão sendo utilizadas 200 imagens, sendo elas, 100 imagens contendo glomérulos saudáveis, 50 imagens contendo glomerulopatias membranosas e 50 imagens contendo glomerulosclerose segmentar.

B. Escolha dos Modelos

Os modelos escolhidos fazem parte da coleção de modelos pré-treinados Model Zoo, os quais foram treinados com os conjuntos de dados das principais competições de detecção de objetos, como COCO Dataset (*Common Objects in Context*) [15].

A escolha dos modelos baseou-se nas informações sobre precisão e tempo de processamento fornecidas pelo COCO e são eles: SSD Mobilenet v1 (SM1), Faster RCNN Resnet 50 (FRR50), Faster RCNN Resnet 101 (FRR101), SSD Resnet 50 (SR50), e RFCN Resnet 101 (RR101)

C. Treinamento dos Modelos

Inicialmente, manteve-se a configuração padrão dos modelos com os 200 mil ciclos de treinamento. O objetivo foi comparar o tempo de treinamento e as diferenças de acurácia dos modelos (Tabela I). Até o momento, foram treinados apenas três modelos, o SSD Mobilenet V1, Faster RCNN Resnet 50 e o RCNN Resnet 101.

Após o treinamento dos modelos com a configuração de ciclos padrão, acrescentou-se 50 mil ciclos para verificar o comportamento do desempenho dos modelos. O tempo de treino dos modelos com a adição de ciclos pode ser observado na Tabela II.

Tabela I
TEMPO DE TREINAMENTO DOS MODELOS COM 200 MIL CICLOS

Modelo	Tempo de processamento
SSD Mobilenet v1	1 dia, 9 horas, 38 minutos e 35 segundos
Faster RCNN Resnet 50	1 dia, 0 horas, 22 minutos e 21 segundos
Faster RCNN Resnet 101	1 dia, 4 horas, 14 minutos e 19 segundos

Tabela II
TEMPO DE TREINAMENTO DOS MODELOS COM 250 MIL CICLOS

Modelo	Tempo de processamento
SSD Mobilenet v1	1 dia, 20 horas, 17 minutos e 01 segundo
Faster RCNN Resnet 50	2 dias, 5 horas, 0 minutos e 25 segundos
Faster RCNN Resnet 101	2 dias, 8 horas, 11 minutos e 23 segundos

D. Avaliação dos modelos

O protocolo escolhido para a avaliação dos modelos foi o protocolo *COCO* (*Common Objects in Context*), o qual traz como métricas de avaliação o mAP, limiar de corte 0.5 à 0.95 com passo 0.05; $mAP@0.5$ e $mAP@0.75$, limiar de corte igual a 0.5 e 0.75; *AP Across Scales*, mAP por tamanho de objetos (*Small*, *Medium* e *Large*); *AR* (*Average Recall*), para detecção de objetos por imagem e; *AR Across Scales*, com *Recall* máximo por tamanho de objetos.

IV. RESULTADOS PARCIAIS E DISCUSSÃO

Ao final da etapa de treinamento dos modelos já avaliados, podemos observar que os resultados obtidos na Tabela III demonstram ser bastante promissores. Diferente de trabalhos encontrados na literatura [12]–[14], neste, não houve uma padronização na aquisição das imagens utilizadas para treinar os modelos de detecção de objetos. Analisando-se a Tabela III, o modelo FRR50 obteve mAP de 0.88, mesmo resultado obtido pelo modelo SI2, utilizado por Rehem *et al.* [11]

Para verificar o desempenho dos modelos, foram adicionados mais 50 mil ciclos. Ao comparar os resultados obtidos da validação dos modelos com os diferentes ciclos configurados, podemos observar que em SM1 e em FRR101 houve uma queda no desempenho, enquanto que no modelo FRR50, houve uma pequena melhora. É interessante notar que o modelo SI2, proposto por Rehem, obteve um resultado melhor que os demais modelos apresentados e, o modelo FRI2 obteve um resultado igual ao modelo FRR50, apresentado neste trabalho.

Ao analisar os resultados dos testes finais, podemos observar que houve uma diferença significativa no desempenho geral (*mAP*) dos modelos. O FRR50 (0.88) demonstrou-se melhor que o SM1 (0.78), que por sua vez, apresentou resultados melhores que o FRR101 (0.54). Ao observar o *mAP* por corte de IoU (*Intersection Over Union*), $mAP@.50$ e $mAP@.75$, onde a *boundary box* sobrepõe 50% ou 75% da área da caixa *ground-truth*, vemos que houve um desempenho próximo em $mAP@.50$ nos modelos SM1 (0.90) e FRR50 (0.97) e um tanto distante em FRR101 (0.54), onde o SM1 apresenta um melhor resultado. Em $mAP@.75$, o modelo FRR50 possui um desempenho bem melhor (0.96) quando comparado com os outros dois modelos (SM1 - 0.88 e FRR101 - 0.56).

Observando o desempenho por escala (*mAP Small*, *mAP Medium*, *mAP Large*) não houveram objetos *Small* (-1000) (<32x32 *pixels*). Os objetos *Medium* (>32x32 & <96x96 *pixels*) em geral, obtiveram um resultado muito inferior ao desempenho geral (*mAP*) enquanto que os objetos *Large* obtiveram um desempenho mais próximo, tendo superado o desempenho geral nos modelos SM1 (0.814) e FRR101 (0.57).

Ao observar o *recall* por número de objetos, em imagens de um único glomérulo (*AR@1*) os modelos SM1 e FRR50 obtiveram desempenhos próximos (0.74 e 0.73 respectivamente, sendo o modelo SM1 o melhor), enquanto que o modelo FRR101 obteve um desempenho bem abaixo (0.55). Em imagens que contém até 10 glomérulos (*AR@10*), o FRR50 obteve o melhor desempenho (0.91) seguido por SM1 (0.81) e FRR101 (0.61). Como não houveram imagens com

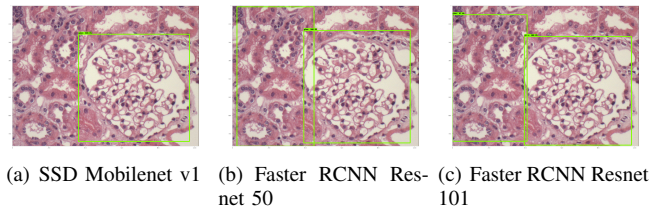


Figura 1. Detecção de glomérulos normais.

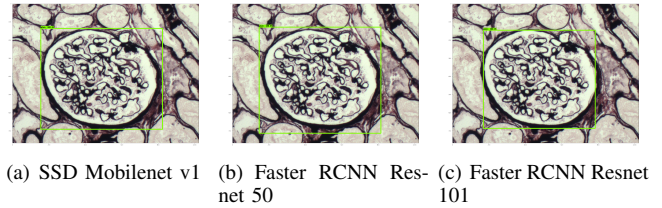


Figura 2. Detecção de glomérulos afetados por glomerulopatias membranasas.

mais de 10 glomérulos, os resultados de *AR@100* refletem os resultados de *AR@10*.

Ao dividir o *AR@100* por escala (*AR@100 Large*, *AR@100 Medium*, *AR@100 Small*) pode-se observar que não houveram imagens *Small* (-1000) com mais de 10 glomérulos, enquanto que imagens *Medium* e *Large* apresentaram uma diferença bastante significativa nos desempenhos, sendo que em imagens *Large* a detecção é melhor.

Nas Figuras 1, 2 e 3 é possível observar alguns resultados da detecção de glomérulos pelos modelos. A detecção foi realizada usando o conjunto de dados de teste, que foram divididos em 3 categorias: glomérulos normais, como podem ser observados na Figura 1; glomérulos afetados por glomerulopatias membranasas (Figura 2) e; glomérulos afetados por glomerulosclerose segmentar (Figura 3).

As Figuras 1, 2 e 3 mostram que os modelos conseguem detectar glomérulos mesmo em imagens cujos tecidos foram preparados com corantes diferentes.

V. CONCLUSÃO E TRABALHOS FUTUROS

Até o presente momento, constatou-se que o modelo FRR50 obteve um desempenho geral melhor na validação do treinamento em relação aos modelos até então avaliados. Ao final do treinamento de todos os modelos, será realizada uma nova avaliação utilizando também as métricas *Precision*, *Recall* e *F1-Score*.

O projeto continua em andamento e os próximos modelos que serão treinados são: RR101 - RFCN Resnet 101 e SR50 -

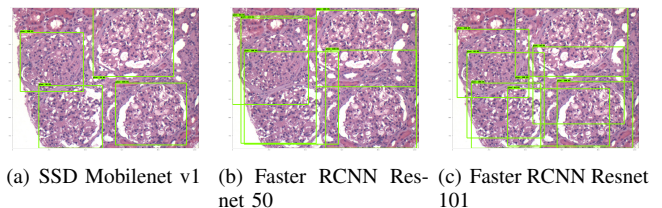


Figura 3. Detecção de glomérulos afetados por glomerulosclerose segmentar.

Tabela III
RESULTADOS OBTIDOS APÓS O TREINAMENTO DOS MODELOS E TESTES FINAIS

Métricas	Validação (200.000 e 250.000 ciclos)			Testes Finais (200.000 e 250.000 ciclos)		
	SMI	FRR50	FRR101	SMI	FRR50	FRR101
<i>mAP</i>	0.80 - 0.78	0.84 - 0.88	0.55 - 0.54	0.80 - 0.78	0.84 - 0.88	0.55 - 0.54
<i>mAP@.50</i>	0.91 - 0.91	0.95 - 0.97	0.83 - 0.82	0.90 - 0.90	0.95 - 0.97	0.82 - 0.82
<i>mAP@.75</i>	0.89 - 0.89	0.94 - 0.97	0.62 - 0.56	0.89 - 0.88	0.94 - 0.96	0.60 - 0.56
<i>mAP Small</i>	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
<i>mAP Medium</i>	0.21 - 0.26	0.60 - 0.72	0.14 - 0.04	0.20 - 0.27	0.54 - 0.72	0.14 - 0.04
<i>mAP Large</i>	0.84 - 0.81	0.87 - 0.88	0.57 - 0.57	0.84 - 0.81	0.85 - 0.88	0.57 - 0.57
<i>AR@1</i>	0.76 - 0.74	0.72 - 0.74	0.55 - 0.55	0.76 - 0.74	0.71 - 0.73	0.56 - 0.55
<i>AR@10</i>	0.83 - 0.81	0.88 - 0.91	0.61 - 0.61	0.82 - 0.81	0.87 - 0.91	0.60 - 0.60
<i>AR@100</i>	0.83 - 0.81	0.88 - 0.91	0.61 - 0.61	0.82 - 0.81	0.85 - 0.91	0.61 - 0.61
<i>AR@100 Small</i>	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
<i>AR@100 Medium</i>	0.22 - 0.29	0.63 - 0.75	0.17 - 0.04	0.22 - 0.29	0.56 - 0.75	0.17 - 0.04
<i>AR@100 Large</i>	0.87 - 0.84	0.63 - 0.95	0.64 - 0.65	0.86 - 0.84	0.91 - 0.93	0.64 - 0.65

SSD Resnet 50. Posteriormente, os resultados serão analisados e comparados.

Ainda como trabalho futuro, serão adicionadas novas imagens, contendo outras lesões. O objetivo será retrainar os modelos com este novo dataset e verificar o desempenho, levando-se em conta uma maior variedade de lesões. Uma avaliação dos modelos por lesão também será realizada.

AGRADECIMENTOS

Os autores agradecem a Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) e ao Conselho Nacional de Desenvolvimento Científico (CNPq) pelo auxílio financeiro para o desenvolvimento deste trabalho.

REFERÊNCIAS

- [1] D. C. Meldau, "Rim - Anatomia dos Rins sistema excretor," <https://www.infoescola.com/sistema-urinario/rim/>, 2017, accessed: 2021-07-12.
- [2] I. Rychlík, E. Jančová, V. Tesař, A. Kolský, J. Lácha, J. Stejskal, A. Stejskalová, J. Dušek, and o. b. o. t. C. R. o. R. B. Herout, Vladimír, "The Czech registry of renal biopsies. Occurrence of renal diseases in the years 1994–2000," *Nephrology Dialysis Transplantation*, vol. 19, no. 12, pp. 3040–3049, 10 2004. [Online]. Available: <https://doi.org/10.1093/ndt/gfh521>
- [3] C. Kelly and J. Landman, "Anatomia do trato urinário," *Coleção Netter de ilustrações médicas. 2a ed. Rio de Janeiro: Saunders-Elsevier*, pp. 24–7, 2014.
- [4] A. Belsare and M. Mushrif, "Histopathological image analysis using image processing techniques: An overview," *Signal & Image Processing*, vol. 3, no. 4, p. 23, 2012.
- [5] I. C. de Araújo, L. Schnitman, A. A. Duarte, and W. dos Santos, "Automated detection of segmental glomerulosclerosis in kidney histopathology," in *XIII Brazilian Congress on Computational Intelligence*, 2017, p. 12.
- [6] G. O. Barros *et al.*, "Pathospotter: um sistema para classificação de glomerulopatias a partir de imagens histológicas renais," 2016.
- [7] G. O. Barros, B. Navarro, A. Duarte, and W. L. Dos-Santos, "Pathospotter-k: A computational tool for the automatic identification of glomerular lesions in histological images of kidneys," *Scientific reports*, vol. 7, no. 1, pp. 1–8, 2017.
- [8] P. Chagas, L. Souza, I. Araújo, N. Aldeman, A. Duarte, M. Angelo, W. L. Dos-Santos, and L. Oliveira, "Classification of glomerular hypercellularity using convolutional features and support vector machine," *Artificial intelligence in medicine*, vol. 103, p. 101808, 2020.
- [9] G. O. Barros, B. Navarro, A. Duarte, and W. L. Dos-Santos, "Pathospotter-k: A computational tool for the automatic identification of glomerular lesions in histological images of kidneys," *Scientific reports*, vol. 7, no. 1, pp. 1–8, 2017.
- [10] I. C. de Araújo, L. Schnitman, A. A. Duarte, and W. dos Santos, "Automated detection of segmental glomerulosclerosis in kidney histopathology," in *XIII Brazilian Congress on Computational Intelligence*, 2017, p. 12.
- [11] M. C. R. Jonathan, F. A. Michele, C. Santos, and W. Luís, "Detecção automática de glomérulos em imagens histológicas renais digitais," *Anais do XII Simpósio de Engenharia Biomédica - IX Simpósio de Instrumentação e Imagens Médicas*, 2019. [Online]. Available: <https://doi.org/10.5281/zenodo.3473700>
- [12] J. Gallego, A. Pedraza, S. Lopez, G. Steiner, L. Gonzalez, A. Laurinavicius, and G. Bueno, "Glomerulus classification and detection based on convolutional neural networks," *Journal of Imaging*, vol. 4, no. 1, p. 20, 2018.
- [13] O. Simon, R. Yacoub, S. Jain, J. E. Tomaszewski, and P. Sarder, "Multi-radial lbp features as a tool for rapid glomerular detection and assessment in whole slide histopathology images," *Scientific reports*, vol. 8, no. 1, pp. 1–11, 2018.
- [14] Y. Kawazoe, K. Shimamoto, R. Yamaguchi, Y. Shintani-Domoto, H. Uozaki, M. Fukayama, and K. Ohe, "Faster r-cnn-based glomerular detection in multistained human whole slide images," *Journal of Imaging*, vol. 4, no. 7, p. 91, 2018.
- [15] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.