

Resultados preliminares de deteção de imagens de pêsegos aplicando o método Faster R-CNN

Por: **Eduardo Assunção¹, Pedro D. Gaspar^{1,2}, Ricardo Mesquita¹, André Veiros¹, Hugo Proença^{1,3}**

¹Universidade da Beira Interior, Covilhã, eduardo.assuncao@ubi.pt, ricardo.mesquita@ubi.pt, andreveiros@hotmail.com, dinis@ubi.pt

²Universidade da Beira Interior, Covilhã, Centre for Mechanical and Aerospace Science and Technologies (C-MAST), Covilhã, dinis@ubi.pt

³IT-Instituto de Telecomunicações, Covilhã, hugomcp@di.ubi.pt

O modelo Faster R-CNN tem grande potencial de aplicação na deteção de pêsegos e poderá vir a ser uma boa ferramenta para estimar a produção em pomares, ajudando no planeamento da colheita e do armazenamento da fruta.

Introdução

A previsão de produção e da necessidade de embalagem no início da campanha agrícola são muitos importantes para os produtores agrícolas e de embalagens. A previsão possibilita o planeamento da colheita, o espaço de trabalho, o armazenamento, apresentando outros benefícios adicionais.

Existem já diversos trabalhos de investigação no campo da previsão de produção de frutos Dorj et al. (2017); Bargoti et al. (2017); Hãni et al. (2019). Hãni et al. (2019) apresentam uma metodologia para deteção, rastreio e contagem de maçãs com o objetivo de estimar a produção. Dorj et al. (2017) desenvolveram um sistema para deteção e contagem de citrinos. Já Bargoti et al. (2017) desenvolveram um sistema para deteção de mangas, maçãs e amêndoas. A etapa de deteção de frutos é uma parte muito importante para um bom desempenho de um sistema de estimação de produção.

Wang et al. (2013) desenvolveram um sistema para estimação de produção de frutos. Para fase de deteção, os autores utilizam um algoritmo tradicional de visão computacional baseado no espaço de cor HSV para segmentação dos frutos de maçãs vermelha e verde. Os métodos do estado-da-arte para deteção de frutos utilizam métodos genéricos de deteção de objetos baseados em redes neuronais artificiais con-

volucionais (CNN) Bargoti et al. (2017), segmentação Puttemans et al. (2016); Bargoti et al. (2017), entre outros. Bargoti et al. (2017) treinaram uma CNN que tem como saída a probabilidade de um pixel da imagem pertencer a um fruto. Esse resultado é utilizado para se obter uma máscara binária e subsequentemente realizar a deteção. Hãni et al. (2019) propõem a utilização da rede neuronal conhecida como U-NET para segmentação de maçãs. A rede foi desenvolvida originalmente para segmentação de imagens médicas. Adicionalmente, as CNNs têm contribuído grandemente para o progresso da deteção e classificação de objetos. Nesse respeito, a rede R-CNN Girshick et al. (2014) e suas derivadas Fast R-CNN Girshick et al. (2015), Faster R-CNN Ren et al. (2015), Mask R-CNN He et al. (2017) e Mask score R-CNN Huang et al. (2019) são as mais relevantes na literatura. Sa et al. (2016) propuseram a utilização da rede Faster R-CNN com o objetivo de detetar pimento, melão, maçã, abacate, manga, morango e laranja. No entanto, apenas as imagens de pimento foram obtidas diretamente no campo. As restantes imagens dos frutos foram obtidas da internet (Google Images). Para obter um bom desempenho utilizando poucas imagens de treino, os autores fazem uso da técnica conhecida como “transfer learning”. Neste caso, o treino da rede não é

realizado totalmente do início, ou seja, aproveitam as informações de um treino anteriormente já realizado noutra base de dados (ImageNet). O presente estudo baseia-se neste estudo científico como base de trabalho. Bargoti et al. (2017) também utilizam o método Faster R-CNN para detecção de frutos, porém exploram diferentes tipos de “transfer learning”, avaliam o desempenho com diferentes quantidades de imagens de treino e diferentes tipos de aumento de dados. Recentemente, uma nova ramificação na detecção de objetos, também baseada nas CNNs e conhecida como YOLO, foi utilizada pela primeira vez com o objetivo de detetar frutos de manga Koirala et al. (2019). Por ser um detetor de apenas um estágio, os autores conseguiram um resultado de detecção mais rápido (aproximadamente 3 vezes), para uma mesma taxa de acerto, comparado com o método Faster R-CNN. Estes métodos de detecção fazem uso das cores dos objetos (dos frutos) e/ou de suas formas. Nesse contexto, a utilização de alguns desses sistemas apresentados em Dorj et al. (2017), Bargoti et al. (2017) e Häni et al. (2019) diretamente com o objetivo de detetar frutos diferentes do qual o sistema foi desenvolvido (ex.: abacates, pêssegos, etc.), provavelmente providenciará um desempenho inferior. Esta conclusão advém das cores e formas dos frutos serem diferentes das quais os sistemas originais foram treinados, apesar das cores e formas de alguns frutos serem semelhantes.

Dada a inexistência de estudos científicos com pêssegos, este artigo apresenta os resultados preliminares de detecção de pêssegos aplicando o método de detecção de objetos Faster R-CNN.

Materiais e Métodos

A Faster R-CNN, entre outros modelos baseados em CNN, vieram a substituir os modelos tradicionais na área da visão computacional. Trata-se de um modelo do estado da arte de detecção de objetos e tem como base as redes neurais convolucionais. O modelo tem a versatilidade de detetar múltiplos objetos, bem como informar suas localizações na imagem de entrada.

Os modelos de reconhecimento de objetos do estado-da-arte, no qual inclui a rede Faster R-CNN, utilizam como “espinha dorsal” uma rede CNN. Diferentemente das redes neurais tradicionais (NN), onde a quantidade de camadas e de neurónios (estrutura) são definidos de acordo com o tipo de problema a ser solucionado, as CNNs geralmente têm as suas estruturas fixas e são aplicadas para resolver diversos problemas diferentes, apesar de as estruturas dessas redes puderem ser modificadas (customizadas). Existem diversos tipos de CNNs, como por exemplo a VGG16 Liu et al. (2015), Resnet He et al. (2016), Inception Szegedy et al. (2016), entre outras.

No modelo Faster R-CNN, as saídas das camadas de convolução são denominadas de “mapas de ca-

racterísticas”. Numa segunda fase, utiliza-se a última camada de convolução como entrada da “Rede de proposta de regiões”, no qual produzem as regiões de interesses (ROIs – Region Of Interest) que contém os possíveis objetos. Essas ROIs são utilizadas na última fase do modelo de detecção para classificação dos objetos, bem como para um possível ajuste de posição do enquadramento da detecção do objeto na imagem de entrada. A (fig. 1) mostra um esquema simplificado do método Faster R-CNN. A (fig. 2) mostra um esquema simplificado do método Faster R-CNN no contexto de detecção de pêssegos.

Para realização dos testes é utilizada a interface de programação de aplicação (API) TensorFlow model. (2019). Essa API implementa alguns modelos de detecção de objeto, tais como Faster R-CNN, R-FCN, SSD. É selecionado o modelo Faster R-CNN com espinha dorsal “Inception_v2”, devido ao seu melhor desempenho com respeito a precisão comparado com os modelos R-FCN e SSD Huang et al. (2017). Também é feito uso da técnica “transfer learning”. Isto é, inicia-se a rede com os pesos de uma outra rede já treinada na base de dados COCO. Como recurso computacional foi utilizado um computador desktop com uma placa gráfica (GPU) RTX 2080.

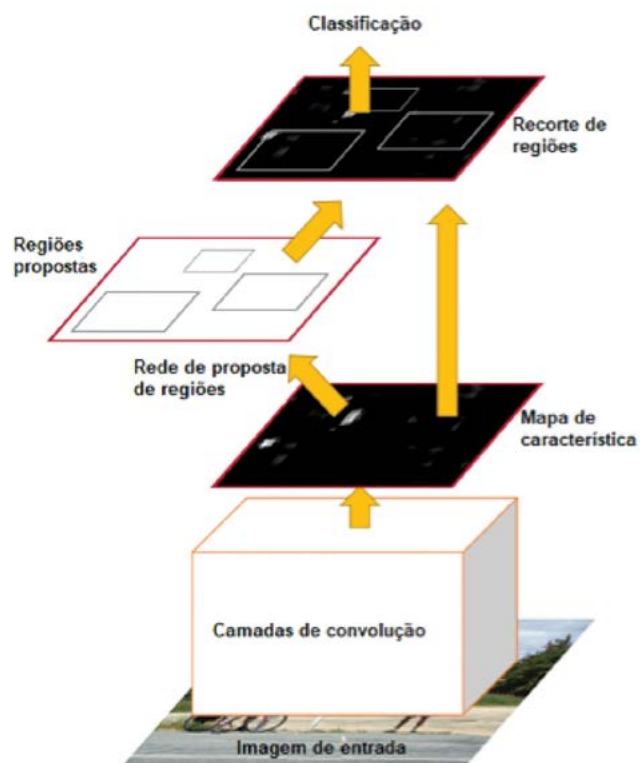


Figura 1 - Esquema simplificado do método Faster R-CNN. Adaptado de Ren et al. (2016).

Para o estudo realizado, uma base de dados de imagens foi obtida num pomar de pessegueiro localizado na Beira Interior (Portugal), mas especificamente a Sul da serra da Gardunha (Soalheira) no ano de 2019. Foi utilizado uma câmara Eken H9R para capturar as imagens. As imagens foram dimensionadas para um tamanho de 2736x1824. Para treino do modelo foram utilizadas 200 imagens de pessegueiros com canais RGB e 40 para teste. É importante salientar que antes do processo de treino existe o processo de anotação que consiste em localizar (com um retângulo), manualmente, cada pêsego nas imagens de treino. Portanto, cada imagem de treino pode ter diversas subimagens. A (fig. 3) mostra um exemplo de uma imagem com anotação.

IoU (Intersection over Union), ilustrado na (fig. 4), é a métrica mais popular como forma de avaliar o desempenho em tarefas de deteção de objetos Rezatofighi

et al. (2019). A mesma é definida como segue:

$$IoU = \frac{A \cap B}{A \cup B}$$

Onde:

I: é a área de interseção;

U: é a área de união;

A: é a área delimitada de deteção do objeto inferido;

B: é a área delimitada real do objeto apresentado como teste.

O valor de IoU, para cada deteção, pode variar de 0 a 1. Deve-se definir um valor de IoU como limite de deteção. Por exemplo, definido o valor de IoU = 0,5 como limite, considera-se como objeto detetado (P) os valores de IoU >= 0,5 quando a previsão corresponde a classe correta, e objeto não detetado (N) os valores de IoU < 0,5.

AP (AP - Average Precision) é uma métrica numérica que pode ser diretamente utilizada para medir o desempenho dos detetores de objetos e tem como base a métrica IoU.

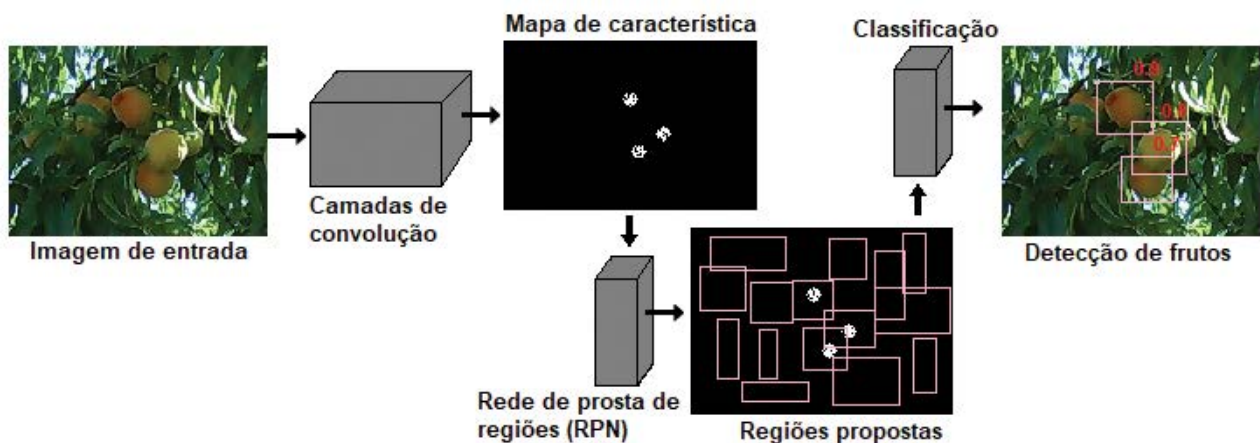


Figura 2 - Esquema simplificado do método Faster R-CNN no contexto de deteção de pêsegos.



Figura 3 - Exemplo de uma imagem de treino com pêsegos anotados. Os retângulos representam as respetivas anotações.

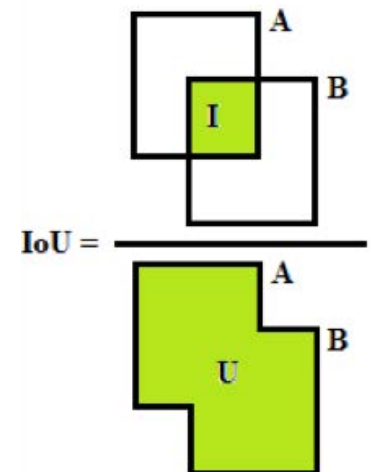


Figura 4 - Ilustração da métrica Intersection over Union.

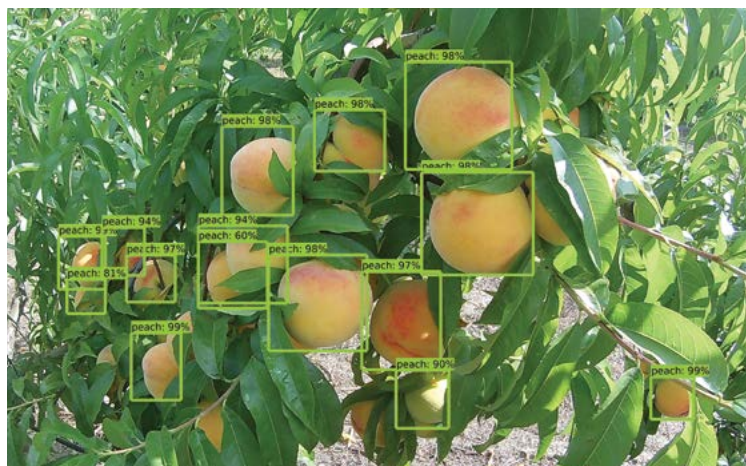


Figura 5 - Exemplo 1 de resultado de detecção.



Figura 6 - Exemplo 2 de resultado de detecção.

Resultados e Discussão

Nesse trabalho foi utilizada a métrica AP para avaliar a detecção dos pêssegos. Foi considerado como detecção o valor de IoU de 0,5, conforme o utilizado na maioria dos trabalhos de detecção de objetos. O valor de AP foi calculado (com IoU = 0,5) para as inferências das detecções das 40 imagens de teste e obtivemos o resultado de 0,87.

O valor obtido de AP de 0,87 com 40 imagens de teste e 200 imagens de treino indica uma boa precisão na detecção, conforme se pode confirmar verificando os resultados visuais nas (fig. 5) e (fig. 6). Ainda assim, alguns trabalhos de detecção de frutos, como por exemplo em Koirala et al. (2019), apresentam resultados de AP de 0,95 para detecção de mangas.

Uma possível alternativa para melhorar o resultado de detecção dos pêssegos apresentado nesse trabalho seria o aumento da quantidade de imagens de treino. Também, pode-se fazer um estudo para verificar a quantidade de oclusão permitida.

Conclusão

Neste artigo são apresentados resultados preliminares para detecção de imagens de pêssegos utilizando o modelo de detecção de objetos Faster R-CNN. Os resultados do modelo mostraram um desempenho relativamente bom, inclusive para frutos agrupados e oclusos. Uma possível alternativa para melhorar o resultado de detecção dos pêssegos apresentado nesse trabalho, seria o aumento da quantidade de imagens de treino e também definir um melhor limite na oclusão dos frutos. Os resultados mostraram um grande potencial da aplicação do modelo Faster R-CNN para detecção de pêssegos, com o objetivo de implementar um sistema de estimativa de produção em pomares. ■

Referências

Häni, N., Roy, P. & Isler, V. 2019. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. *Journal of Field Robotics* 1-20.
 Dorj, U.O., Lee, M. & Yun, S.-s. 2017. An yield estimation in citrus orchards via fruit detection and counting using image processing. *Computers and Electronics in Agriculture* 140:103-112.
 Bargoti, S. & Underwood, J. 2017. Deep fruit detection in orchards. *International Conference on Robotics and Automation (ICRA)* 3626-3633.

Puttemans, S., Vanbrabant, Y., Tits, L. & Goedemé, T. 2016. Automated visual fruit detection for harvest estimation and robotic harvesting. *Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)* 1-6.

Bargoti, S. & Underwood, J.P. 2017. Image segmentation for fruit detection and yield estimation in apple orchards. *Journal of Field Robotics* 34:1039-1060.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T. & Mccool, C. 2016. Deepfruits: A fruit detection system using deep neural networks. *Sensors* 16:1222.

Wang, Q., Nuske, S., Bergerman, M. & Singh, S. 2013. Automated Crop Yield Estimation for Apple Orchards. *Springer Tracts in Advanced Robotics* 88.

Girshick, R. 2015. Fast r-cnn. *International Conference on Computer Vision (ICCV)* 1440-1448.

Girshick, R.B., Donahue, J., Darrell, T. & Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. *Conference on Computer Vision and Pattern Recognition* 580-587.

Girshick, R.B. 2015. Fast r-cnn. *International Conference on Computer Vision (ICCV)* 1440-1448.

Ren, S., He K., Girshick R.B. & Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Transactions on Pattern Analysis and Machine Intelligence* 39: 1137-1149.

He, K., Gkioxari, G., Dollár, P. & Girshick, R. 2017. Mask R-CNN. *IEEE International Conference on Computer Vision (ICCV)* 2980-2988.

Huang, Z., Huang, L., Gong, Y., Huang, C. & Wang, X. 2019. Mask Scoring R-CNN. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 6402-6411.

Koirala, A., Walsh, K.B., Wang, Z., & McCarthy, C. 2019. Deep learning for real-time fruit Detection and orchard fruit load estimation: benchmarking of 'mangoyolo'. *Precision Agriculture* 1107-1135.

Zeiler, M.D. & Fergus, R. 2014. Visualizing and understanding convolutional networks. *Computer Vision - ECCV* 818-833.

Liu, S. & Deng, W. 2015. Very deep convolutional neural network based image classification using small training sample size. *IAPR Asian Conference on Pattern Recognition (ACPR)* 730-734.

He, K., Zhang, X., Ren, S., & Sun, J. 2016. Deep residual learning for image recognition. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770-778.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. 2016. Rethinking the inception architecture for computer vision. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2818-2826.

TensorFlow model API. 2019. <https://github.com/tensorflow/models>.

COCO dataset. 2019. <http://cocodataset.org/#home>.
 Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, Ian., Wojna, Zbigniew., Song, Y., Guadarrama, S. & Murphy K. 2017. Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 3296-3297.

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I. & Savarese S. 2019. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 658-666.

Nota: Trabalho apresentado no II Simpósio Ibérico de Engenharia Hortícola, realizado de 4 a 6 de Março de 2020, em Refóios do Lima, na Escola Superior Agrária do Instituto Politécnico de Viana do Castelo.