



Preliminary results of peach detection in images applying convolutional neuronal network

Eduardo Assunção - eassunca@gmail.com

Universidade da Beira Interior, Covilhã

Hugo Proença - hugomcp@di.ubi.pt

IT-Instituto de Telecomunicações, Universidade da Beira Interior, Covilhã

André Veiros - andreveiros@hotmail.com

Universidade da Beira Interior, Covilhã; Centre for Mechanical and Aerospace Science and Technologies (C-MAST), Covilhã

Ricardo Mesquita - ricardo.mesquita@ubi.pt

Universidade da Beira Interior, Covilhã; Centre for Mechanical and Aerospace Science and Technologies (C-MAST), Covilhã

Pedro D. Gaspar - dinis@ubi.pt

Universidade da Beira Interior, Covilhã; Centre for Mechanical and Aerospace Science and Technologies (C-MAST), Covilhã

Abstract

The fruit detection part is very important for a good performance in a yield estimation system. This paper presents the preliminary results using the object detection Faster R-CNN method in the peaches images. The aim is evaluate the method performance in the detection of peach RGB images. Images acquired in an orchard were used. Although this method of object detection has been applied in other studies to detect fruits, according to the literature, it has not been used to detect peaches. The results, although preliminary, show a great potential of using the method to detect peach.

Keywords

Peach, Image detection, Convolutional neuronal network

Resultados Preliminares de Detecção de Imagens de Pêssegos Aplicando Rede Neuronal Convolutacional

1. 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 [1], [2], [3]. Häni *et al.* [1] apresentam uma metodologia para detecção, rastreio e contagem de maçãs com o objetivo de estimar a produção. Dorj *et al.* [2] desenvolveram um sistema para detecção e contagem de citrinos. Já Bargoti *et al.* [3] desenvolveram um sistema para detecção de mangas, maçãs e amêndoas. A etapa de detecção de frutos é uma parte muito importante para um bom desempenho de um sistema de estimação de produção. Os métodos do estado-da-arte para detecção de frutos utilizam métodos genéricos de detecção de objetos baseados em redes neuronais artificiais convolucionais (CNN) [3], segmentação [4], [5], entre outros. 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 [1], [2] e [3] diretamente com o objetivo de estimar a produção de 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 desenvolvidos, 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 e tem como base o trabalho apresentado por Sa *et al.* [6].

2. Estado da Arte

Wang *et al.* [7] desenvolveram um sistema para estimação de produção de frutos. Para fase de detecção, os autores utilizam um algoritmo tradicional de visão computacional baseado no espaço de cor HSV para segmentação dos frutos de maçãs vermelha e verde. Com o advento das CNN, alguns autores têm desenvolvido métodos com o objetivo de segmentar imagens de frutos para posterior detecção. Bargoti *et al.* [5] treinam 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 detecção. Häni *et al.* [1] propõe a utilização da rede neuronal conhecida como U-NET para segmentação de maçãs. Essa rede foi desenvolvida originalmente para segmentação de imagens médicas. Adicionalmente, as CNNs têm contribuído grandemente para o progresso da detecção e classificação de objetos. Nesse respeito, a rede Region with Convolutional Neural Network (R-CNN) [9] e suas derivadas Fast R-CNN [10], Faster R-CNN [11], Mask RC-NN [12] e Mask score R-CNN [13] são as mais relevantes na literatura. Sa *et al.* [6] propôs a utilização da rede Faster R-CNN com o objetivo de detectar 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.* [3]

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 detectar frutos de manga [14]. Por ser um detector 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.

3. Metodologia

Nesta secção é apresentado o método de detecção Faster R-CNN, base deste trabalho, assim como todos os detalhes relacionados com a aquisição e processamento de imagens.

3.1. Detecção de Fruto Utilizando Faster R-CNN

O método Faster R-CNN, entre outros, veio 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 detectar múltiplos objetos de classes diferentes, bem como informar suas localizações na imagem de entrada.

Diferentemente das redes neurais tradicionais (NN), onde a quantidade de camadas e de neurônios (estrutura) são definidos de acordo com o tipo de problema a ser solucionado, as CNN, como por exemplo a VGG16 [16], Resnet [17], Inception [18], geralmente têm as suas estruturas fixas e são aplicadas para resolver diversos problemas diferentes. No modelo Faster R-CNN, utilizam-se as camadas de convolução dessas redes pré-treinadas (CNN) com a finalidade de extrair as características da imagem de entrada. As saídas dessas camadas são denominadas de “mapas de características”, no qual a última camada é utilizada como entrada da segunda fase (Rede de proposta de regiões) do método Faster R-CNN. A saída da rede de proposta de regiões são as possíveis regiões de interesses (ROIs - Region Of Interest) que contém objetos. Essas ROIs são utilizadas na última fase do modelo para classificação dos objetos, bem como para um possível ajuste de posição do enquadramento da detecção do objeto na imagem de entrada. A Figura 1 mostra um esquema simplificado do método Faster R-CNN.

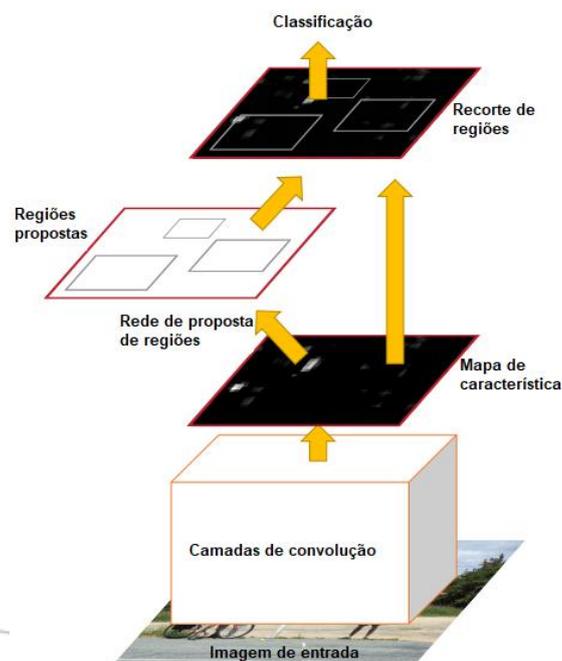


Figura 1 - Esquema simplificado do método Faster R-CNN. Adaptado de <https://arxiv.org/pdf/1506.01497v3.pdf>

O trabalho apresentado neste artigo, é um problema de detecção de objetos de apenas duas classes, pêsego e não-pêsego. A Figura 2 mostra um esquema simplificado do método Faster R-CNN no contexto de detecção de pêsegos.

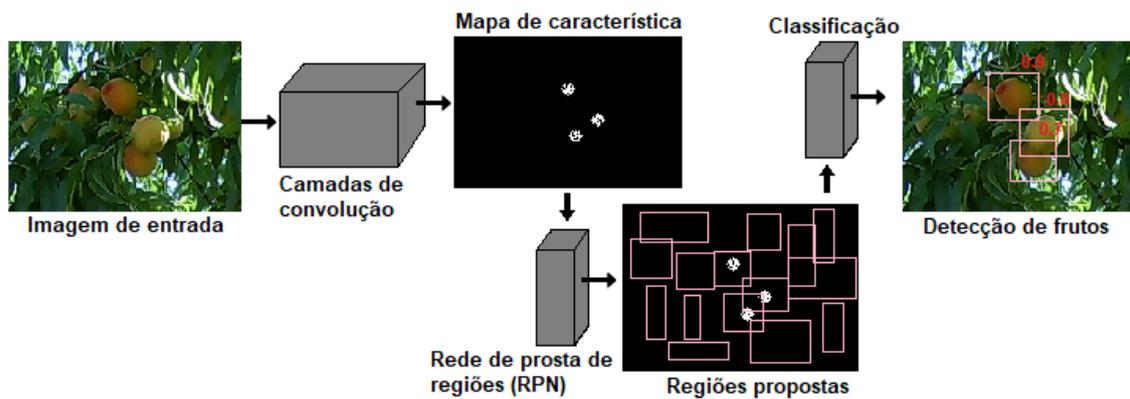


Figura 2 - Esquema simplificado do método Faster R-CNN no contexto de detecção de pêsegos.

No presente estudo, a rede neuronal base que compõe o método de detecção Faster R-CNN é a ZF-Net [15]. Os parâmetros para treino e teste são os mesmos definidos em [11], incluindo a técnica de transferência de aprendizagem com a base de dados ImageNet.

3.2 Configuração do trabalho experimental

Para o estudo realizado, uma base de dados de imagens foi obtida em pomares de pessegueiro localizados na Beira Interior (Portugal), mas especificamente a Sul da serra da Gardunha (Soalheira). Foi utilizado uma câmara Eken H9R para capturar as imagens. O treino foi realizado em 50 imagens (RGB) de pessegueiros com um total de 1100 sub-imagens anotadas de pêsegos e 5 imagens para teste. A Figura 3 mostra um exemplo de uma imagem de treino.



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

4. Resultados Experimentais

Nas Figura 4 e Figura 5 são apresentados exemplos de resultados de deteção.

Na Figura 4 pode ser notado que o modelo tem um bom desempenho, inclusive para frutos agrupados. Todavia, apresenta um falso positivo (Fp) no canto inferior esquerdo e alguns falsos negativos (Fn). Os Fp e Fn estão representados em caixas de cor amarela, sendo que os Fn foram desenhados manualmente como ilustração.

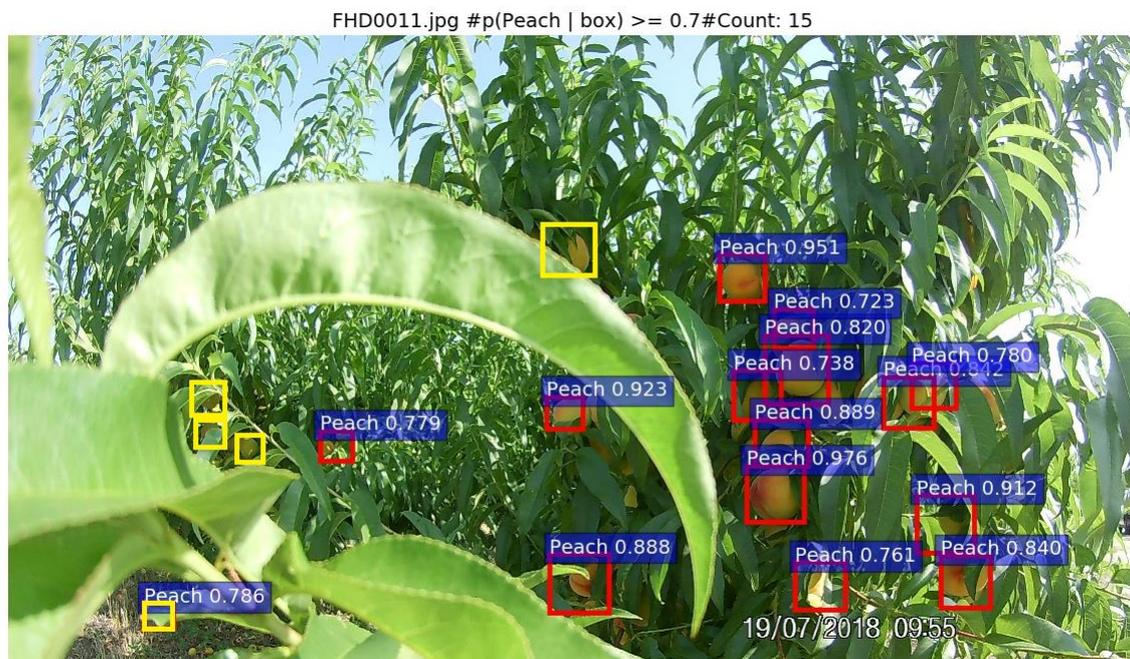


Figura 4 - Exemplo de resultado de deteção.

Na Figura 5 encontra-se um resultado de teste numa imagem adquirida relativamente mais perto dos frutos, com grande concentração de frutos agrupados e grande quantidade pêssegos oclusos. Pode ser verificado que para o problema de agrupamento e oclusão, o modelo teve um bom desempenho. Porém, com relação à detecção dos pêssegos relativamente grandes, o modelo teve um baixo desempenho.

Estes resultados evidenciam a necessidade de um maior número de imagens de treino e simultaneamente a consideração de condições mais similares nas imagens. É importante salientar que as imagens foram obtidas em ambiente real e sujeitas às condições de luminosidade natural existente, o que se configura como uma fonte de variabilidade nos resultados.

FHD0032.jpg #p(Peach | box) >= 0.7#Count: 17

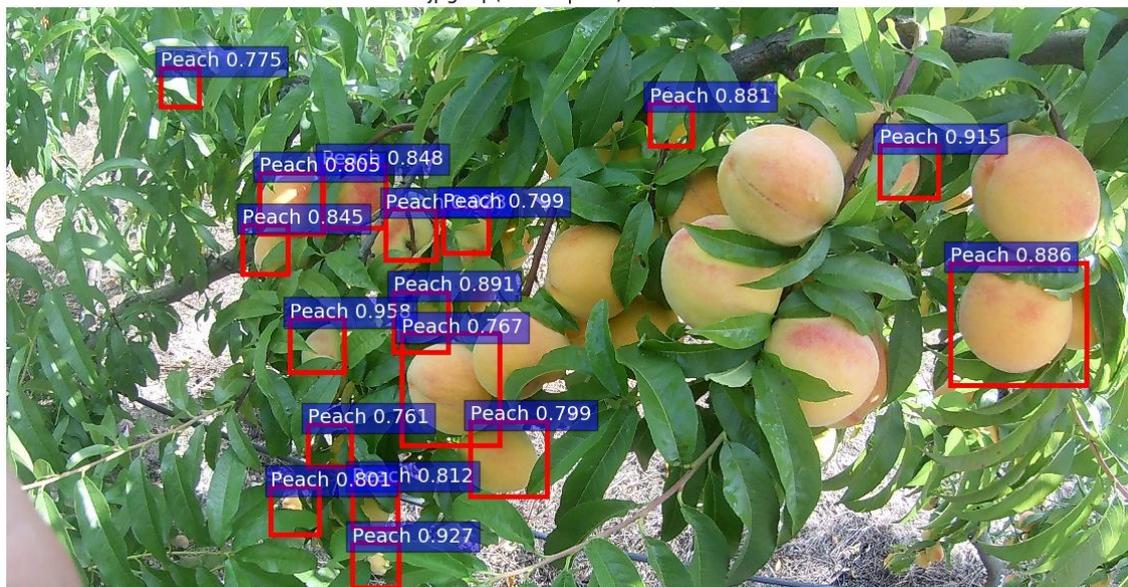


Figura 5 - Exemplo de resultado de detecção.

5. Conclusão

Neste artigo é apresentado um resultado preliminar para detecção de imagens de pêsegos 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. O modelo foi capaz de detectar bem os frutos que tenham uma gama de tamanho médio. Ou seja, os tamanhos que não estejam nos extremos (muito pequenos ou muito grandes). O modelo precisa ser melhorado com relação à escala para detectar imagens de pêsegos grandes. Esta melhoria pode ser realizada ajustando o tamanho das imagens de entrada da rede neuronal convolucional, bem como dos parâmetros do modelo de detecção. Os resultados mostraram um grande potencial da aplicação do modelo Faster RCNN para detecção de pêsegos, com o objetivo de implementar um sistema de estimação de produção de pêsegos em pomares.

Agradecimentos

Este trabalho de investigação é financiado pelo projeto PrunusBot - Sistema robótico aéreo autónomo de pulverização controlada e previsão de produção frutícola, Operação n.º PDR2020-101-031358 (líder), Consórcio n.º 340, Iniciativa n.º 140, promovido pelo PDR2020 e co-financiado pelo FEADER e União Europeia no âmbito do Programa Portugal 2020.

Referências

[1] N. Häni, P. Roy, and V. Isler, "A comparative study of fruit detection and counting methods for yield mapping in apple orchards," *Journal of Field Robotics*, vol. 0, no. 0. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21902>

- [2] U.-O. Dorj, M. Lee, and S.-s. Yun, "An yield estimation in citrus orchards via fruit detection and counting using image processing," *Computers and Electronics in Agriculture*, vol. 140, pp. 103-112, 08 2017
- [3] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 3626-3633
- [4] S. Puttemans, Y. Vanbrabant, L. Tits, and T. Goedemé, "Automated visual fruit detection for harvest estimation and robotic harvesting," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Dec 2016, pp. 1-6
- [5] S. Bargoti and J. P. Underwood, "Image segmentation for fruit detection and yield estimation in apple orchards," *Journal of Field Robotics*, vol. 34, no. 6, pp. 1039-1060, 2017. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21699>
- [6] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. Mccool, "Deepfruits: A fruit detection system using deep neural networks," *Sensors*, vol. 16, p. 1222, 08 2016
- [7] Q. Wang, S. Nuske, M. Bergerman, and S. Singh, "Automated Crop Yield Estimation for Apple Orchards". Heidelberg: Springer International Publishing, 2013, pp. 745-758.3
- [8] R. Girshick, "Fast r-cnn," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ser. ICCV '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 1440-1448. [Online]. Available: <http://dx.doi.org/10.1109/ICCV.2015.169>
- [9] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580-587, 2013
- [10] R. B. Girshick, "Fast r-cnn," *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440-1448, 2015.
- [11] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1137-1149, 2015
- [12] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask r-cnn," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980-2988, 2017
- [13] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask scoring r-cnn," *ArXiv*, vol. abs/1903.00241, 2019
- [14] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, "Deep learning for real-time fruit Detection and orchard fruit load estimation: benchmarking of 'mangoyolo'," *Precision Agriculture*, Feb 2019. [Online]. Available: <https://doi.org/10.1007/s11119-019-09642-0>
- [15] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision - ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 818-833
- [16] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, Kuala Lumpur, 2015, pp. 730-734.

[17] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 770-778

[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 2818-2826