DOI: 10.20396/revpibic2720192774

Fruit Detection in Viticulture with Deep Neural Networks

Leonardo L. Souza*, Sandra Avila, Thiago T. Santos

Abstract

We investigate deep neural networks applied to fruit detection in viticulture. We also developed the Embrapa WGISD dataset, composed of images collected in April 2017/2018 at the Guaspari Winery. Annotated manually, the dataset has 5 different varieties of grape: Syrah, Chardonnay, Cabernet Franc, Cabernet Sauvignon and Sauvignon Blanc, totaling 4419 samples of grape bunches. We trained YOLOv2 and YOLOv3 to detect and locate the bunches in the images. YOLOv2 achieved up to 88% accuracy and YOLOv3 up to 92% accuracy. Qualitative tests demonstrated that the YOLOv2 network generalizes better for the dataset used, and the YOLOv3 network provides a better-adjusted location.

Key words:

Grape Detection, Deep Learning, Embrapa WGISD Dataset

Introduction

The mapping and estimation of fruit production is important in agriculture. Through this data the producers can locate and classify orchards problems, determine and apply effective measures and solutions to the problem through better use of available resources.

Recent approaches for the fruit detection step of mapping and estimation are using computer vision techniques based on Deep Neural Networks, the state of the art technique in the computer vision field. To achieve our goal of grape detection through images from a camera potentially embedded in autonomous field vehicle, we've chosen the YOLO (You Only Look Once) architecture [1] in previous work [2], because of it real-time prediction performance. In the current work we use the YOLOv2 [3] and YOLOv3 [4], comparing both networks in a new public more complex dataset, the Embrapa WGISD (Wine Grape Instance Segmentation Dataset).

Results and Discussion

We collected and annotated new images acquired from Guaspari Winery. The dataset (300 images with 5 grape varieties) contains bounding boxes and binary masks labeling the grape bunches. Thus we built the new public Embrapa WGISD Dataset.

The overall YOLO architecture consists in only one network that do the detection and classification in an end-to-end fashion predicting both localization and label of the object that belongs to the bounding box, generating a probability/confidence index for the object.

We trained YOLOv2 (19 convolutional layers) and YOLOv3 (53 convolutional layers) in the same data and train, validation and test split, through the Darknet framework running both networks original implementation with pre-trained weights in the ImageNet dataset.

In Table 1 we show the results for the best parameters used for each network, in Figure 1 we show the detection results for these best network parameters, and in Figure 2 we show bounding boxes examples used.

Table 1. YOLOv2 and YOLOv3 results for test data.

Network	True Positive	False Positive	False Negative	F1
YOLOv2	233	44	121	0.73
YOLOv3	158	31	127	0.66

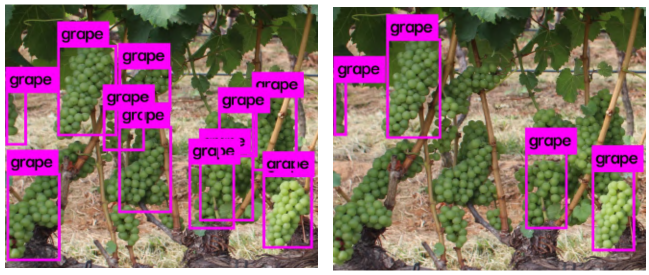

Figure 1. YOLOv2 (left) and YOLOv3 (right).

Figure 2. Bounding boxes examples of augmented data.

Conclusions

YOLOv2 is a better model than YOLOv3 for the amount of training data used, but the bounding boxes generated by the YOLOv3 are much better adjusted than YOLOv2 ones. We observe by the false negative results that the models could have been penalized by some bias in our annotations. Qualitative results show that most of the false negative bounding boxes produced are better than our annotations or were not annotated correctly. In future work we will investigate these questions about the necessary volume of data for YOLOv3 start learning more and producing better results than YOLOv2, and also the possible bias in our dataset.

Acknowledgement

Embrapa/PIBIC/CNPq (125044/2018-6), FAEPEX (3125/17), FAPESP (2017/16246-0), Google LARA 2018, NVIDIA.

[1] Redmon et al., "You Only Look Once: Unified, real-time object detection". In: Computer Vision and Pattern Recognition. 2016. p. 779-788.

[2] Dos Santos et al., "Detecção automática de uvas e folhas em viticultura com uma rede neural YOLOv2". In: Embrapa Informática Agropecuária – Artigo em anais de congresso, 2018.

[3] Redmon, J.; Farhadi, A., "YOLO9000: better, faster, stronger". In: Computer Vision and Pattern Recognition. 2017. p. 7263-7271.

[4] Redmon, J.; Farhadi, A., "YOLOv3: An incremental improvement". arXiv:1804.02767, 2018.