Artificial Intelligence Research and Development
K. Trejo et al. (Eds.)

doi:10.3233/FAIA250602

Continuous Object Orientation Estimation of Cotton Crops Using Deep Convolutional Neural Networks

Antonio MARTINEZ [a,1], Sergi FOIX [a]
[a] *Institut de Robotica i Informatica Industrial, CSIC- UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain*
ORCiD ID: Sergi Foix https://orcid.org/0000-0001-9249-6696

Abstract. Deep Convolutional Neural Networks (DCNNs) have shown an interesting performance in accurately estimating the continuous orientation of objects, successfully predicting angles ranging from 0 to 360 degrees based on RGB 2D image inputs to enhance manipulation. However, these types of solutions have not been explored for agriculture scenarios. In this study, we propose a real-world application using previously presented methods to estimate cotton crop orientation, which provides key information for a Next Best View (NBV) algorithm, enhancing decision-making with respect to the selection for the next optimal end-effector (EE) position for efficient and successful harvesting.

Keywords. Orientation Estimation, DCNN, Cotton Crops, Automated Agriculture

1. Introduction

The use of deep convolutional neural networks (D-CNN) in computer vision tasks has yielded significant results in a wide range of problems. Including image classification [1], object detection [2], semantic segmentation [3] and object orientation estimation. In terms of the last one, some interesting works have been presented in recent years, such as the case of Orient-Anything [4], which predicts the orientation of 3D objects from a single image using a large synthetic dataset and models orientation as probability distributions. Another important study focuses on vehicle/pedestrian orientation estimation achieved through two different methods, minimizing the loss of L2 or the angular difference in the unit circle, and discretizing the orientations and then applying a mean shift algorithm to recover a continuous estimate [5].

An area that has benefited from these advancements is agriculture, helping to move toward the automation of crucial tasks like fruit counting [6], crop and weed detection [7], and plant disease detection [8]. Apart from those tasks, another crucial field of study with large real-world implications is the fruit orientation estimation, which can be really promising for automatic harvesting systems. Some exploration has been carried out in the field with fruits such as apples [9]. However, our study focuses on a very par-

[1]Corresponding Author: Antonio Martínez Fernández, antonio.martinez.f@upc.edu.

ticular type of crop, never explored before, the cotton crop, which counts with its own challenges. Like the plant topology and the existence of asymmetries in the fruit, besides not having a lot of texture or characteristic features that can help to determine its orientation.

The principal objective of this work is to determine whether, through previously presented methods based on D-CNN, it is possible or not to estimate the projected orientation of the cotton crop over the 2D image plane of the camera sensor. The development of this work is part of a major project[2], in which one of its goals is to create an NBV algorithm, based on maximization of the information gain, to optimize the harvesting of cotton crops. By adding the estimate of the cotton bolls' orientation into our NBV decision cost function, we expect to enhance the final harvesting manipulation approach.

2. Method

2.1. Labeling Tool

Since no existing tools met the project requirements, a GUI application was developed in Python using Tkinter. The idea was to create some simple intuitive GUI tool, replicating existing software such as LabelStudio. Annotations are created by dragging a vector in the principal area. Once saved, a .txt file is created with the same name as the original image. This .txt file presents the format visible in the **Figure 1**, where the different parts of the GUI can be appreciated.

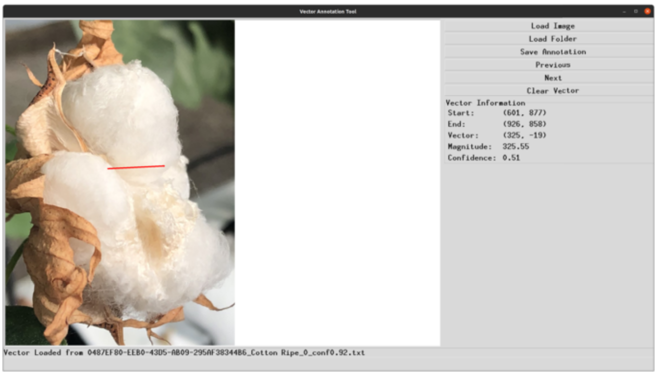

Figure 1. Image of the GUI app designed for 2D vector annotation, the red vector displayed represent the corresponding orientation of the image with the vector information on the right side.

2.2. Orientation Estimation

Regarding orientation estimation, the input images are expected to be the output of a prior detection model, such as YOLO. This has already been explored in our project, for the detection and localization of cotton fruits [10]. The idea is to minimize the noise that

[2]DEMETER 5.0 project PLEC2022-009289 funded by MCIU/ AEI /10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR"

could be introduced by including larger portions of the original image. Before feeding the images to the network, we resize them to the canonical input size of the corresponding network. All prediction units are trained by back-propagation.

2.2.1. Implemented Approach

The angle description of an image is presented as a vector $v = (x, y)$. It was trained as a regression layer for the 2D orientation vector. Two different Loss functions methods were tested, in the first place using the Huber loss, which combines the best properties of Mean Squared Error (MSE) and Mean Absolute Error (MAE), making it less sensitive to outliers in data than the squared error loss. This loss function can be described for a batch of size N as:

$$\ell(x,y) = L = \{l_\delta(x_1,y_1),\ldots,l_\delta(x_N,y_N)\}^T, \tag{1}$$

$$\text{where} \quad l_\delta(a,b) = \begin{cases} \frac{1}{2}(a-b)^2, & \text{if } |a-b| \le \delta \\ \delta\left(|a-b| - \frac{1}{2}\delta\right), & \text{otherwise} \end{cases} \tag{2}$$

Being δ the threshold at which to change between delta-scaled L1 and L2 loss.

The other loss function tested consisted of directly computing the angular losses. This type of loss function has also been explored in other works [5]. This loss function can be expressed as follows:

$$L(v_g, v) = 1 - \cos(\theta) = 1 - \frac{v_g \cdot v}{|v_g||v|} = 1 - \frac{x_g x + y_g y}{\sqrt{x^2+y^2}} \tag{3}$$

3. Experiments

3.1. General View

A total of 602 images were successfully labeled using the orientation labeling tool,and in addition to preventing overfitting and improving the robustness of the model, some data augmentation techniques were employed. First, mirroring was applied along the x- and y-axes. Additionally, rotations from 45° to 315°, in steps of 45°, were used for augmentation, resulting in a final dataset of 1146 images, this dataset was then divided into train, validate, and test datasets, with a proportion of 70,15 and 15%, respectively.

To evaluate the performance of the proposed method, multiple experiments were conducted using different feature extraction architectures, mainly the Resnet architecture family and EfficientNet, since both have shown a good balance between inference speed and performance [11], making both good candidates for this type of real-time solution. Different versions of both families were tested to find the one with the best results. Furthermore, some hyperparameter tuning was included, including variables like batch size, learning rate, number of epochs.

An early stopping mechanism was also implemented to stop the training if no validation loss improvement was perceived after a particular number of epochs. This number of epochs is defined as patience, and after a different test was decided to use eight epochs, providing a decent tradeoff between accuracy and overfitting avoidance.

3.2. Results

To evaluate the global performance of the model, we decided to use the Mean Angle Error (MAE) on the test dataset, which the model has never seen before. Apart from evaluating the model's accuracy within a margin 45 degrees, making it acceptable for our desired application due to the discretization in the NBV approach. With these margins, the final next-best-view pose would likely remain unchanged. After hyperparameter tuning trying to find the best combination, we came up with the values presented in **Table 1**, with these values different versions of the mentioned architectures were tested. The results show a general good performance using the Huber loss function in all architecture versions, as can be appreciated in **Figure 2**, with around 5 degrees of difference between the simplest versions of the models and the more complex ones. Overall, ResNet101 and EfficientNetB3 achieved the lowest MAE, while ResNet50 and EfficientNetB2 slightly outperformed them in terms of accuracy within a 45-degree error margin. With respect to the use of the angular error as a loss function, it did perform slightly worse, with an average of 2.04 degrees more in the MAE and 6.78 percentuale points in the accuracy under 45 degrees.

	Learning Rate	Batch Size	Max. Epochs	Head Drop Rate (%)
Value	1e-4	15	40	50

Table 1. Best values obtained from hyperparameter tuning

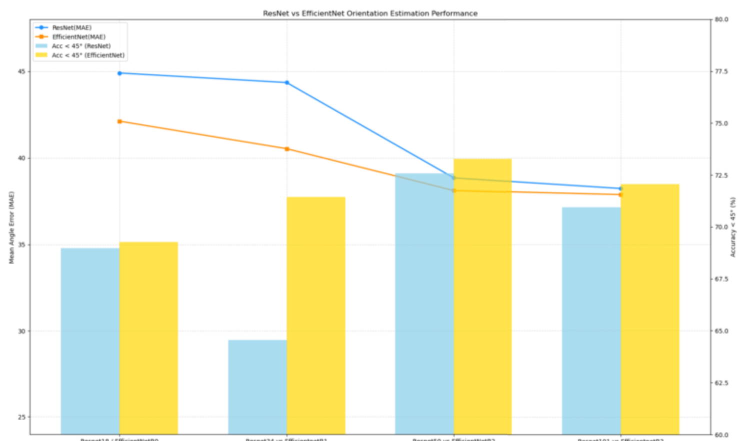

Figure 2. Graphic with overall models performance. Bars representation correspond to the percentage from the total estimations with a lower error margin of 45 degrees. The blue ones corresponds to the family of Resnet while the yellow ones corresponds to the family of EfficientNet. Meanwhile the line graph corresponds to the MAE of each corresponding model (orange one for EfficientNet and blue one for Resnet)

4. Conclusions

This study demonstrates that previously proposed orientation estimation methods can be successfully applied in real-world agricultural contexts. Lightweight D-CNN architectures provide accurate angle estimation, making them suitable for real-time applications in resource-constrained environments. Although manual labeling introduces potential biases, it can still provide sufficient accuracy when no reliable ground-truth data is available.

5. Acknowledgments

This work has been partially funded by the DEMETER 5.0 project PLEC2022-009289 funded by MCIU/AEI/10.13039/501100011033 and by the ”European Union NextGenerationEU/PRTR”.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[2] R. 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, Columbus, OH, USA, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.

[3] E. Shelhamer, J. Long and T. Darrell, ”Fully Convolutional Networks for Semantic Segmentation,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640-651, 1 April 2017, doi: 10.1109/TPAMI.2016.2572683.

[4] Z. Wang, Z. Zhang, T. Pang, C. Du, H. Zhao, and Z. Zhao, Orient Anything: Learning Robust Object Orientation Estimation from Rendering 3D Models. 2024. [Online]. Available: https://arxiv.org/abs/2412.18605 ,doi:10.48550/arXiv.2412.18605

[5] K. Hara, R. Vemulapalli, and R. Chellappa, Designing Deep Convolutional Neural Networks for Continuous Object Orientation Estimation. 2017. [Online]. Available: https://arxiv.org/abs/1702.01499, doi:10.48550/arXiv.1702.01499

[6] N. Hani, P. Roy, and V. Isler, “Apple Counting using Convolutional Neural Networks,” in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Oct. 2018, pp. 2559–2565. doi: 10.1109/iros.2018.8594304.

[7] Hu W, Chen T, Lan C, Liu S, Yin L. SkipResNet: Crop and Weed Recognition Based on the Improved ResNet. Land. 2024; 13(10):1585. https://doi.org/10.3390/land13101585

[8] Pandian JA, Kumar VD, Geman O, Hnatiuc M, Arif M, Kanchanadevi K. Plant Disease Detection Using Deep Convolutional Neural Network. Applied Sciences. 2022; 12(14):6982. https://doi.org/10.3390/app12146982

[9] E. Kok and C. Chen, “Occluded apples orientation estimator based on deep learning model for robotic harvesting,” Computers and Electronics in Agriculture, vol. 219, p. 108781, 2024, doi: https://doi.org/10.1016/j.compag.2024.108781.

[10] G. González, A. Martínez, V. Martínez, S. Foix and G. Alenyà, ”Fast Ready-To-Harvest Cotton Detection and Classification with YOLOv8 in Greenhouse Crops,” 2024 7th Iberian Robotics Conference (ROBOT), Madrid, Spain, 2024, pp. 1-6, doi: 10.1109/ROBOT61475.2024.10796900.

[11] O. Elharrouss, Y. Akbari, N. Almadeed, and S. Al-Maadeed, “Backbones-review: Feature extractor networks for deep learning and deep reinforcement learning approaches in computer vision,” Computer Science Review, vol. 53, p. 100645, Aug. 2024, doi: 10.1016/j.cosrev.2024.100645.