Quasi-Unsupervised Fruit Counting Using Domain Adaptation and Spatial Consistency

1st Enrico Bellocchio
Department of Engineering
University of Perugia
Perugia, Italy
enrico.bellocchio@gmail.com

2nd Gabriele Costante
Department of Engineering
University of Perugia
Perugia, Italy
gabriele.costante@unipg.it

3rd Silvia Cascianelli
Department of Engineering "Enzo Ferrari"
University of Modena and Reggio Emilia
Modena, Italy
silvia.cascianelli@unimore.it

4th Mario Luca Fravolini
Department of Engineering
University of Perugia
Perugia, Italy
mario.fravolini@unipg.it

5th Paolo Valigi
Department of Engineering
University of Perugia
Perugia, Italy
paolo.valigi@unipg.it

***Abstract*—Autonomous robotic platforms can be effectively used to perform automatic fruits yield estimation. To this aim, robots need data-driven models that process image streams and count, even approximately, the number of fruits in an orchard. However, training such models following a supervised paradigm is expensive and unpractical. Extending pre-trained models to perform yield estimation for a completely new type of fruit is even more challenging, but interesting since this situation is typical in practice. In this work, we combine a State-of-the-Art weakly-supervised fruit counting model with an unsupervised style transfer method for addressing the task above. Experiments on datasets collected in four different orchards show that the proposed approach is more accurate than the supervised baseline methods.**

***Index Terms*—Agricultural Automation, Robotics in Agriculture and Forestry, Visual Learning**

I. Introduction

Among the orchard management-related processes, yield estimation plays a crucial role in harvesting operations planning and income prevision. State-of-the-Art works [1], [2] already demonstrated how to make yield sampling cost-effective by using autonomous robots to collect images of the orchard. However, fruit counting is still an open issue. Recently proposed methods [3], [4] rely on heuristics and strong supervision for fruit detection and counting. This limits the application of such methods in real agricultural contexts because of the required costs and time. For this reason, we follow the trend of applying the weak supervision paradigm for yield estimation [5]. This reduces the labelling work and time required to train the counting model. Although being a more flexible approach, neither the weakly supervised paradigm guarantees that the counting model would work for fruits different from those used in training. However, the ability to generalize on different types of orchards is interesting from a commercial point of view.

In the sight of this, in this work we propose to combine the recently proposed weakly supervised fruit counting model presented in [5] with an unsupervised style transfer method for

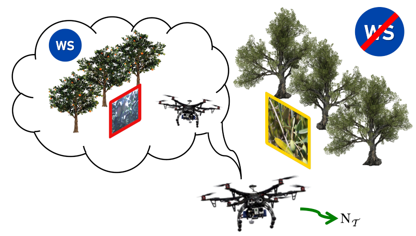

Fig. 1: After being trained on source orchard images with weak ground truth (WS in the picture) and transformed target orchard images with no ground truth, QU-COUNT can directly count fruits from target orchard images.

tackling the yield estimation problem in scenarios in which no ground truth is available. In this sense, our proposed approach is quasi-unsupervised (see Fig. 1). In particular, we jointly train a Cycle-Generative Adversarial Network (C-GAN) [6] to transform images from a source orchard $\mathcal{S}$ into a different target orchard $\mathcal{T}$, and a Presence-Absence Classifier (PAC) to discriminate between images that contain fruits and images that do not contain fruits. The trained PAC generates the weak supervision signal for the counting block, that is now able to work directly on images from the target orchard. The experimental results obtained on four different orchards demonstrate that the proposed approach gives performance comparable to a fully-supervised approach. We name our approach QU-COUNT.

II. Proposed Approach

Ideally, we want a network able to learn what and how to count by only relying on fruit presence-absence labels. As shown in [5], this could be achieved with a multi-branch architecture whose optimization objective is constrained by a Presence-Absence Classifier (PAC). In this work, we follow the strategy proposed in [5]. In particular, we rely on a Multi-branch Counting CNN (MBC-CNN), whose branches operate on different image tiles extracted from the original image at three different scales, *i.e.*, the full image, and the 4 and

16 non-overlapping crops. Each branch regresses the number of fruits for a given tile. During the optimization phase, the count estimates at each level are summed, and a constraint is imposed to ensure that the total count at each scale is consistent with the others.

We design the QU-COUNT approach with the capability to *adapt* the PAC trained on $\mathcal{S}$ to $\mathcal{T}$ without the need for any supervision information associated with the target domain. To do this, we take advantage of the recent Cycle Generative Adversarial Network (GAN) architecture [6] to achieve domain translation. The Cycle GAN (C-GAN) is combined during the optimization phase to adapt the PAC to the target domain. The aim of Cycle GAN is to learn two mapping functions $M : \mathcal{S} \rightarrow \mathcal{T}$ and $N : \mathcal{T} \rightarrow \mathcal{S}$ that translate images from one domain into the other. However, we argue that translating images with the C-GAN framework is not sufficient to guarantee good performance on the target domain. Hence, we devise an objective function whose aim is to adapt the PAC to $\mathcal{T}$, imposing a cross-entropy loss functions on the three images produced during the $\mathcal{S} \rightarrow \mathcal{T} \rightarrow \mathcal{S}$ cycle.

III. Experiments

The approach is evaluated on four different datasets containing images of three fruit species: two sets contain almond images while the other ones provide olive and apple frames. Apples and one of the almond dataset are presented in [3]. The olive images and the second almond set are presented in [5]. To highlight the advantages introduced by QU-COUNT, we compare it against different baselines. The first one is the work introduced by [3], we name this method Bargoti et al. We also compare QU-COUNT against WS-COUNT, a state-of-the-art weakly-supervised approach presented in [5]. Another important baseline is the MBC-CNN optimized with the PAC trained on the source domain, this approach is referred to as MBC-CNN+PAC$_{\mathcal{S}}$. We also consider a baseline that does not adapt the PAC, and we refer to it as MBC-CNN+GAN. In each test, we compute the Root Mean Square Error (RMSE) between the predicted count values and the ground truth ones.

A. Results

The quantitative results are provided in TABLE I. Each one of the table blocks refers to the set of tests associated with one of the four possible target domains. The approach proposed by Bargoti et al. [3] achieves the lowest errors in all the experiments where the source and the target domains are the same, with the only exception of the $OL \rightarrow OL$ test, where WS-COUNT obtains slightly better performance. These results are expected since the approach from [3] uses strong supervision signals (*i.e.*, fruit bounding boxes) during training, hence, fruit detection is more robust. On the other hand, WS-COUNT and MBC-CNN+PAC$_{\mathcal{S}}$ score similarly. However, when a different target domain comes into play, the error of the aforementioned approaches rapidly increases in almost all cases. As an instance, consider, the $AP \rightarrow AL$ and $OL \rightarrow AL$ experiments, where the RMSE of Bargoti et al. [3],

Target domain $\mathcal{T}$: Almonds (AL)				
Approach	**AL → AL**	**AL$_{IS}$ → AL**	**AP → AL**	**OL → AL**
Bargoti et al.	**2.19**	**3.24**	6.69	6.40
WS-COUNT	4.39	4.9	6.71	6.93
MBC-CNN+PAC$_{\mathcal{S}}$	3.15	4.07	6.62	6.64
MBC-CNN+GAN	n.a.	6.63	4.93	5.55
QU-COUNT	n.a.	4.14	**3.69**	**4.20**
Target domain $\mathcal{T}$: AlmondsISAR (AL$_{IS}$)				
Approach	**AL$_{IS}$ → AL$_{IS}$**	**AL → AL$_{IS}$**	**AP → AL$_{IS}$**	**OL → AL$_{IS}$**
Bargoti et al.	**1.33**	**2.18**	3.11	3.45
WS-COUNT	2.2	2.58	3.27	3.2
MBC-CNN+PAC$_{\mathcal{S}}$	1.96	2.43	3.45	3.62
MBC-CNN+GAN	n.a.	3.35	2.69	2.79
QU-COUNT	n.a.	2.49	**2.49**	**2.62**
Target domain $\mathcal{T}$: Apples (AP)				
Approach	**AP → AP**	**AL → AP**	**AL$_{IS}$ → AP**	**OL → AP**
Bargoti et al.	**1.55**	3.83	**2.32**	3.72
WS-COUNT	2.58	4.66	4.67	4.36
MBC-CNN+PAC$_{\mathcal{S}}$	2.36	4.98	4.75	4.82
MBC-CNN+GAN	n.a.	3.00	3.38	3.51
QU-COUNT	n.a.	**2.81**	2.56	**2.60**
Target domain $\mathcal{T}$: Olives (OL)				
Approach	**OL → OL**	**AL → OL**	**AL$_{IS}$ → OL**	**AP → OL**
Bargoti et al.	2.32	4.16	3.35	4.34
WS-COUNT	**2.21**	5.75	5.81	5.13
MBC-CNN+PAC$_{\mathcal{S}}$	2.24	5.97	5.57	4.72
MBC-CNN+GAN	n.a.	4.55	4.77	4.87
QU-COUNT	n.a.	**3.49**	**2.80**	**3.32**

TABLE I: Quantitative results achieved by the different approaches in various source-target tests. Each refers to tests on a given target domain. Reults are expressed with RMSE metric.

WS-COUNT [5] and MBC-CNN+PAC are fairly high, *i.e.*, the count is more than 6.00 units different from the ground truth one. Conversely, QU-COUNT provides lower errors, which proves that adapting the PAC gives considerable benefits.

IV. Conclusion

In this work, we introduced QU-COUNT, a novel framework whose objective is to perform fruit counting in scenarios for which no knowledge is available by exploiting a source domain where only weak presence-absence labels are given.

Future works will focus on improvements of the image translation process to ease the adaptation of the PAC and make it more robust to scenarios that considerably differ from the source one.

References

[1] V. Duggal, M. Sukhwani, K. Bipin, G. S. Reddy, and K. M. Krishna, "Plantation monitoring and yield estimation using autonomous quadcopter for precision agriculture," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 5121–5127.

[2] G. Riggio, C. Fantuzzi, and C. Secchi, "A low-cost navigation strategy for yield estimation in vineyards," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2200–2205.

[3] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 3626–3633.

[4] N. Häni, P. Roy, and V. Isler, "A comparative study of fruit detection and counting methods for yield mapping in apple orchards," *arXiv preprint arXiv:1810.09499*, 2018.

[5] E. Bellocchio, T. A. Ciarfuglia, G. Costante, and P. Valigi, "Weakly supervised fruit counting for yield estimation using spatial consistency," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2348–2355, 2019.

[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2242–2251.