

MULTI-PURPOSE CHESTNUT CLUSTERS DETECTION USING DEEP LEARNING: A PRELIMINARY APPROACH

Telmo Adão ^{1,2,*}, Luís Pádua ^{1,2}, Tatiana M. Pinho ¹, Jonáš Hruška ²,
António Sousa ^{1,2}, Joaquim João Sousa ^{1,2}, Raul Morais ^{1,2} and Emanuel Peres ^{1,2}

¹ Centre for Robotics in Industry and Intelligent Systems (CRIIS), INESC Technology and Science (INESC-TEC), Porto, Portugal -
(telmoadao,tatianap,luispadua,arms,jjsousa,rmorais,eperes)@utad.pt

² Engineering Department, School of Science and Technology, University of Trás-os-Montes e Alto Douro, Vila Real, Portugal -
jonash@utad.pt

Commission VI, WG VI/4

KEY WORDS: Chestnut, Chestnut Tree, Chestnut Detection, Convolutional Neural Networks, CNN, Deep Learning, DL, Xception, Rough Segmentation, Tiling Segmentation.

ABSTRACT:

In the early 1980's, the European chestnut tree (*Castanea sativa*, Mill.) assumed an important role in the Portuguese economy. Currently, the Trás-os-Montes region (Northeast of Portugal) concentrates the highest chestnuts production in Portugal, representing the major source of income in the region (€50M-€60M).

The recognition of the quality of the Portuguese chestnut varieties has increasing the international demand for both industry and consumer-grade segments. As result, chestnut cultivation intensification has been witnessed, in such a way that widely disseminated monoculture practices are currently increasing environmental disaster risks. Depending on the dynamics of the location of interest, monocultures may lead to desertification and soil degradation even if it encompasses multiple causes and a whole range of consequences or impacts. In Trás-os-Montes, despite the strong increase in the cultivation area, phytosanitary problems, such as the chestnut ink disease (*Phytophthora cinnamomi*) and the chestnut blight (*Cryphonectria parasitica*), along with other threats, e.g. chestnut gall wasp (*Dryocosmus kuriphilus*) and nutritional deficiencies, are responsible for a significant decline of chestnut trees, with a real impact on production. The intensification of inappropriate agricultural practices also favours the onset of phytosanitary problems. Moreover, chestnut trees management and monitoring generally rely on in-field time-consuming and laborious observation campaigns. To mitigate the associated risks, it is crucial to establish an effective management and monitoring process to ensure crop cultivation sustainability, preventing at the same time risks of desertification and land degradation.

Therefore, this study presents an automatic method that allows to perform chestnut clusters identification, a key-enabling task towards the achievement of important goals such as production estimation and multi-temporal crop evaluation. The proposed methodology consists in the use of Convolutional Neural Networks (CNNs) to classify and segment the chestnut fruits, considering a small dataset acquired based on digital terrestrial camera.

1. INTRODUCTION

The chestnut (*Castanea sativa*, Mill.) agro-ecosystem is of great social, economic and landscape importance in north-eastern Portugal, namely to Trás-os-Montes, due to the various resources associated with this crop (e.g. fruit and wood production and mushroom harvesting (Baptista et al., 2010)).

The increasing industry and consumer-grade demand at an international level is the reason behind chestnut cultivation area has been over explored, leading to monoculture practices that represent a great environmental risk, eventually, with an implicated economic impact as well. Desertification and soil degradation are the main concerns in this topic.

Particularly in Trás-os-Montes, threats related to phytosanitary problems of biotic and abiotic origins, such as the chestnut ink disease (*Phytophthora cinnamomi*), and the chestnut blight (*Cryphonectria parasitica*), among others - e.g. chestnut gall wasp (*Dryocosmus kuriphilus*) and nutritional deficiencies - are responsible for a significant decline of chestnut trees with a direct impact on production, despite cultivation area growth (Martins et al., 2015).

Regarding to chestnut trees management and monitoring, both generally rely on costly and time-consuming in-field observation

campaigns. To mitigate the risks associated to this culture and current exploration practices, it is of high importance establishing an effective management and monitoring process to ensure crop cultivation sustainability, preventing at the same time risks of desertification and land degradation.

In previous work (Marques et al., 2019), an approach based on digital image processing for macro detection and monitoring of chestnut trees was proposed, through multiple parameters estimation such as tree identification and counting, individual extraction of tree height, tree crown diameter and area features, using remote sensed imagery acquired by unmanned aerial vehicles. Good accuracy and root mean square error (RMSE) correlations were achieved. Currently, the challenge of performing a more fine-grained analysis through chestnut fruit detection is being pursued. Considering that digital image analysis typically requires an extended knowledge on the subject under study (environment, involved entities, eventual interactions, etc.), and on image manipulation operations, which usually integrate a burdensome development process that, in the end of the day, is bounded to a very particular class of problems, Deep Learning (DL) techniques must not be overlooked. DL techniques are becoming increasingly popular for forestry and agricultural applications (Kamilaris and Prenafeta-Boldú, 2018),

* Corresponding author

due to their flexible adaptability for characterizing different problems in the most varied contexts, requiring minimal changes regarding implementation, wherein the quality of the datasets constitute the key for success. Thus, one of the most prominent motivation for such study is admittedly related to advances in neural networks, namely in what concerns to backpropagation. It is, also, a goal assessing the application of such class of approaches in the identification of elements of interest in rough conditions, as it is the case of the chestnut fruit still hanging in the tree, whose color and texture might be easily confounded with background, depending on its phenological state.

This study presents an automatic method that allows to perform chestnut trees clusters identification, a key-enabling task towards the achievement of important goals such as production estimation and multi-temporal crop evaluation. The proposed methodology consists in the use of Convolutional Neural Networks (CNNs) to classify and segment the chestnut fruits, considering a small dataset acquired based on digital terrestrial camera.

2. BACKGROUND

DL is a modern and promising technique for image processing and data analysis, whose application domains include (but are not limited to) plant recognition, leaf and crop type classification and plant diseases detection, among other useful operations with benefits for natural environments mainly populated by vegetation (Kamilaris and Prenafeta-Boldú, 2018).

Grinblat et al. (2016) used CNNs to identify plants through leaf veins morphology. Similar research interests were pursued by Lee et al. (Lee et al., 2015), with 44 species of the Royal Botanic Gardens (Kew, England). Hall et al. (Hall et al., 2015) evaluated the robustness of traditional hand-crafted leaves features and showed classification performance improvement by combining CNNs with such features, while Reyes et al. (2015) proposed a fine-tuning strategy as a solution to transfer learned recognition capabilities from general domains to the specific challenge of plant identification task. The possibility of increasing the number of hidden layers within CNNs was explored to boost discriminative power, with remarkable results on plant phenotyping classification, more specifically, focusing root and shoot features (Pound et al., 2017). Semantic segmentation using deep convolutions was applied by Mortensen et al. (Mortensen et al., 2016), who resorted to close-up images of crops, data augmentation and manual annotation of 7 classes, among them barley, weed and soil. The potential of CNNs for segmenting plant species was emphasized. DL-based solutions towards plant recognition include (Lee and Yoon, 2019; Zhu et al., 2018). Also, efforts to achieve computer-based recognition of plant diseases have been developed (e.g. Sladojevic et al., 2016; Mohanty et al., 2016, Amara et al., 2017 and Toda and Okura, 2019).

In the specific case of fruit detection/recognition/counting, whose addressing problems recurrently include occlusion, depth variation, and uncontrolled illumination and high color similarity between fruit/foilage, a few solutions can also be found. An approach combining two CNNs to detect and count apples and oranges was proposed by Chen et al. (2017). Labelling was done through a crowdsourcing tool that enables to quickly produce hand-made datasets. Fruit detection in orchards, including mangoes, almonds and apple was a challenge embraced by S. Bargoti and Underwood (2017), who used a tiling approach to deal with the hundreds of fruits that can be present in a single image and a Faster Region-CNN (R-CNN) for estimation purposes. Apples and mangoes could be accurately detected. These cultures earned attention in two other (apparently) previous works (Suchet Bargoti and Underwood, 2017; Stein et al., 2016). (Rahneemoonfar and Sheppard, 2017) used synthetic data to predict real fruits, resorting to an Inception-ResNet

architecture. Accuracies for both synthetic and real data were above 90%. Fruit detection based on Faster R-CNN and using RGB and NIR proximity imagery was a concern for Sa et al. (2016).

Regarding CNN architectures, a considerable amount of them have been proposed: VGG - named after Visual Geometry Group laboratory - (Simonyan and Zisserman, 2014), ResNet (He et al., 2015), Inception family (Szegedy et al., 2014; Szegedy et al., 2015; Szegedy et al., 2016), Google's MobileNet (Howard et al., 2017), Xception (Chollet, 2016) and DenseNet (Huang et al., 2016). Segmentation networks oriented to object detection include - but are not confined to - Fully Convolutional Network (FCN) (Long et al., 2014), ParseNet (Liu et al., 2015), U-Net (Ronneberger et al., 2015) and Feature Pyramid Network (FPN) (Zhao et al., 2016).

Following DL advances, many tools that have been developed to facilitate the interface of programmers and users with the modelling of problems through convolutional neural networks, such as TensorFlow ("TensorFlow," 2015), Theano ("Welcome - Theano 0.9.0 documentation," 2007), Keras ("Keras Documentation," 2015), deep learning Matlab toolbox ("Deep Learning Toolbox," 2016), Caffe ("Caffe | Deep Learning Framework," 2013), PyTorch ("PyTorch," 2016) or even user-friendly studios such as Deep Cognition ("Deep Cognition," 2018).

The following sections focus the proposal of a preliminary methodology consisting in the use of Convolutional Neural Networks (CNNs) to classify and roughly segment the chestnut clusters, considering a small dataset acquired based on digital terrestrial camera.

3. DEEP LEARNING-BASED APPROACH TOWARDS CHESTNUT FRUIT DETECTION

After imagery gathering, and dataset preparation - involving manual segmentation, rough tiling and training/validation sets splitting - an Xception architecture was prepared to learn and predict chestnut clusters still hanging in the tree. The whole process is depicted in Figure 1.

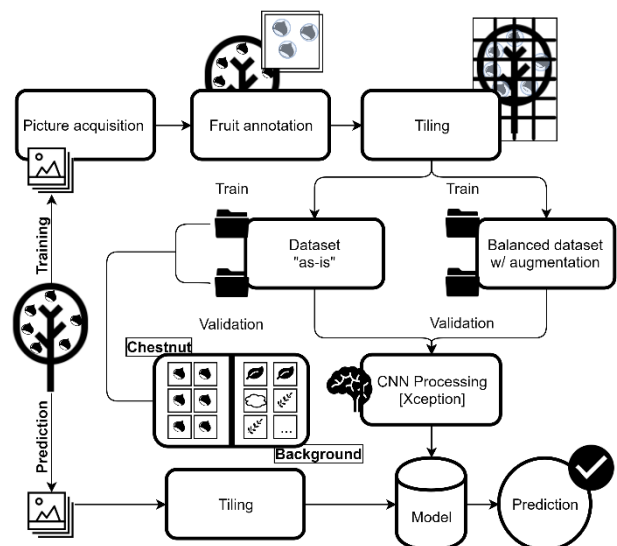


Figure 1. Pipeline of the process adopted to apply DL in chestnut clusters segmentation. After proper pictures acquisition, fruit annotation and tiling, datasets are set up to comply with CNN requirements and submitted to Xception processing for the learning stage. After model output, pictures for prediction can be submitted to estimate the presence/absence of

chestnuts, which implies a preliminary tiling operation.

3.1 Tree models used for gathering imagery

The dataset was created using a Canon EOS 600D, with a CMOS of 18 MP resolution. A focal distance of 23 mm was used. The acquisition was performed from September to October 2018 in two different chestnut plantations, one within the campus of the University of Trás-os-Montes e Alto Douro and another in plantation located in a rural area (see previous work of Pádua et al., 2018, for more information on this plantation).

From the gathered images, two were chosen for CNN training and validation purposes, while a third one was left out for testing CNN performance. Used images original dimensions are 3260x2480 pixels (96dpi).

It is noteworthy to highlight that, typically, chestnut clusters are mainly concentrated in the extremities of trees' branches, which makes it viable the use of imagery for crown analysis towards fruit detection. On the other hand, fruit clusters and background similarities in terms of colour and tonality can constitute a real challenge due to their proneness to be misclassified, besides the reduced size of the element of interest.

3.2 Dataset preparation

First task towards dataset preparation implied the manual production of ground-truth masks, to identify/classify the fruits of interest among the 3 images composing referenced imagery. An example depicting a mask overlaid with transparency upon a chestnut tree picture is presented in Figure 2, wherein brown spots represent fruit clusters annotation. Then, algorithmically, images and masks were properly sliced into tiles of 71x71, the minimum input dimensions supported by Xception. Regarding mask images, granular classification was converted into rough, inasmuch as tiles matching with more than 2.5% of white pixels in the original ground-truth - marking the elements of interest - were completely turned into white, fostering dataset enrichment through a wide variety of conditions that can be characterizing chestnut clusters (completely framed, partially visible, with more or less background, several development states, etc.). Two classes are divided according to masks: one denoting background and another specifying chestnut clusters. This process is essentially characterized by the creation of two virtual folders within the operative system. 3060 tiles resulted from this operation, 2470 for background and 590 for chestnut.



Figure 2. Example of a chestnut picture with the fruits annotated with brown blobs, manually drawn.

To comply with the common Deep Learning practices, the aforementioned group of tiles was properly split in a couple of datasets, each one composed of training (70% of the available images) and validation (30% of the available images). One of the datasets relies in the post-tiling operation, which consists in an unbalanced knowledge base, with much more images for background than for chestnuts. To tackle with an eventual off-balance training, another dataset based on the former was produced, but considering a data augmentation step to even both chestnut and background elements. More specifically, the unbalanced dataset was augmented with 1940 images, totalizing 5000 of them, 2500 per class. Data augmentation consisted in synthetically replicating existing imagery with transformations that included different rotations, contrasts/brightness, vertical/horizontal mirroring and scale variations.

3.3 CNN architecture

Xception (Chollet, 2016) combines point-wise convolutions followed by depth-wise separable convolutions and residual connections (inspired by ResNet (He et al., 2015)). By exchanging the order of operations relatively to original depth-wise separable convolution - that starts by convoluting the channels and then the pixels - and, also, by removing intermediate rectified linear units (ReLU's) non-linearity, Xception reached state-of-the-art performances in tests with ImageNet dataset (Russakovsky et al., 2014). Thereby, Xception architecture was adopted to assess chestnut fruit training and detection estimation. The optimizer used on this network was the stochastic gradient descent (SGD), which already shown good results in previous studies, comparatively to learning rate adaptive managers (Adão et al., 2019).

The next section will focus the tools used to put into practice the proposed chestnut detection approach.

4. MATERIALS AND METHODS

To prepare and produce the datasets used in this work, image editing/manipulation operations were carried out with a specialized software. Also, programming tools were employed to handle DL-related activities. Free and open-source software was selected.

4.1 Image editing and manipulation

For dataset preparation and production, Paint.NET (Rick Brewster's project as a student, Pullman, Washington, United States of America) was used. It consists in a handy set of image processing tools supporting multi-layer capabilities. It offers the possibility of upgrading functionalities resorting to plugins that can be freely developed by the community.

Masks' manual delineations to segment chestnut fruits were carefully carried out, by taking advantage of multi-layer capabilities provided by Paint.NET, as well as the possibility for parameterizing overlay transparency for individual layers. To sum up, for each chestnut image, mask production can be performed by adding another layer to the top of the image that, basically, consists in a black canvas where in white stains can be drawn to annotate fruit clusters.

4.2 Programming tools

Programming activities were done in Python ("Welcome to Python.org," 2018), namely, the tiling-based dataset production. DL implementation resources relied in Keras ("Keras Documentation," 2015) supported by Tensorflow ("TensorFlow," 2015) backend. Keras' Xception baseline

implementation was adopted as CNN, to generate the models for prediction purposes. During testing estimations with unseen data, several parameters were monitored to assess resulting Xception-based models, namely the ones presented in the next subsection.

4.3 Evaluation metrics

Widely accepted assessment indicators were used as evaluation metrics, more specifically, true positive/negative (TP , TN) and false positive/negative (FP , FN) tiles counting, as well as Dice coefficient and Jaccard Index. The formulas for latter pair of metrics are presented in equations (1 and 2), under a set theory perspective, wherein DC and JI represent Dice Coefficient and Jaccard Index, both operating with two sets specified by X and Y .

$$DC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

$$JI = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

4.4 Relevant hardware details

Regarding the computational resources, the hardware that is worth to highlight is related with the DL-based activities, in which two graphical cards NVidia Quadro M4000 (Nvidia Corporation, Santa Clara, California, USA) - 1664 CUDA cores, 8GB of GDDR5 memory and a maximum bandwidth of 192GB/s - working in cooperation were used, allowing to speed-up both training/validation and testing procedures.

Results and discussion, based on the aforementioned assessment parameter, are presented in the following section.

5. RESULTS AND DISCUSSION

Training/validation results, as well as assessment with unseen data - i.e., a chestnut tree wittingly reserved for CNN approach evaluation - are presented in this section.

5.1 Training/validation stage

Training/validation stage was set to perform during a maximum of 20 epochs, with a batch size of, also, 20, for the both non-augmented and augmented datasets that went through Xception CNN processing and model outputting. Only two classes were considered for a binary problem: background and chestnut. Steps per epoch were set according to the ratio between the number of training images and batch size. Validation steps were defined by the same rule, i.e. number of validation images divided by the batch size. Early stop monitors were defined with a patience of 5 epochs relatively to loss validation parameter, which stops the training in case of consecutive learning stagnation.

The training/validation plots of the learning stage seems to point out a steady convergence, even though there is still some space for loss and accuracy improvement. The non-augmented dataset was the first used to shape up a Xception-based model. Regarding training behaviour parameters, it started with a loss of 0.41 and ended with 0.29. Accuracy started growing from 0.83 until it reached 0.88. Analysing validation behaviour parameters, accuracy raised from 0.27 until 0.87, in the last epoch; loss gradually improved from 4.88 until 0.31. Due to the early stop monitor call-back, learning process was stopped at the 12th epoch, in which the progress was perceived as stuck. The augmented dataset made the Xception learning process slightly improve its accuracy and loss. Regarding training, accuracy improved from 0.73 until 0.91, while loss decreased from 0.53 to 0.21. Looking at validation, accuracy progressed from 0.63 until

0.91; loss gradually decreased from 0.61 until 0.23. The last epoch of training was the 16th, when the patience of the early stop monitor reached its limit for learning stagnation. Figure 3 shows the learning plots for both datasets.

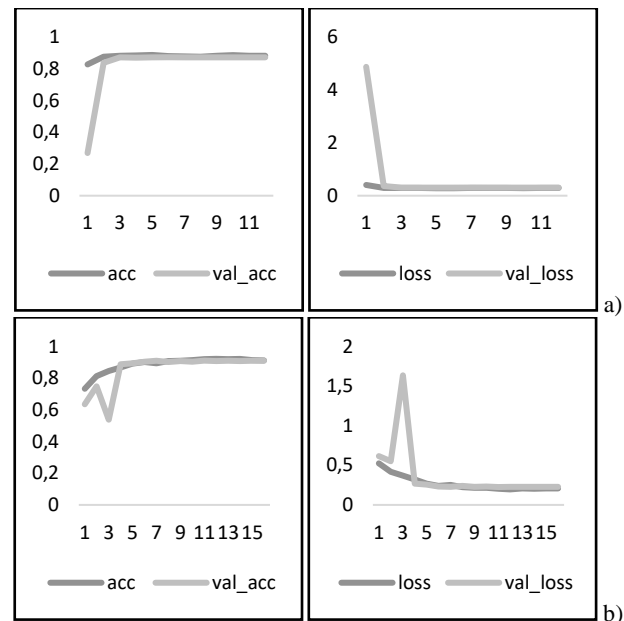


Figure 3. CNN training/validation results for a) dataset without augmentation and b) dataset with augmentation. A convergence tendency can be observed for both.

5.2 Chestnut detection in unseen data

To proceed with the actual tests to assess CNN models – resulting from both previously mentioned datasets (non-augmented/unbalanced and augmented/balanced) used for training – regarding chestnut fruit detection capabilities, a single image without participation in the training stage was used. It was decomposed into 1530 position-aware tiles of 71x71 that were submitted to the referred models individually and in two different moment, whose estimation results were provided in the form of rough tiled-based masks, as it is shown by Figure 4.

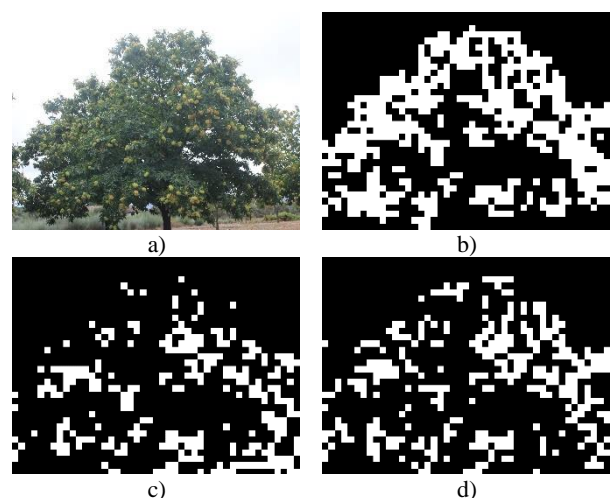


Figure 4. Testing image for assessment purposes: a) is the RGB image of the chestnut used for assessment; b) represents tile-based rough ground-truth; c) is the mask generated by the estimation using a Xception model trained with the non-augmented/unbalanced dataset; d) is the mask generated by the estimation

using a Xception model trained with the augmented/balanced dataset.

Quantitative metrics are provided by Table 2, wherein one can conclude about the importance of DL mechanisms such as data augmentation and balancing, especially in situations of dataset scarcity, as it is the case in this study. As it is clarified, the model based on the non-augmented dataset presents a much lower Dice-Coefficient (0.59) than the model built out of the augmented dataset (0.70).

DS	Estimation Data					
	TP(<i>t</i>)	FP(<i>t</i>)	TN(<i>t</i>)	FN(<i>t</i>)	DC	Jl
1	222	313	995	307	0.59	0.41
2	301	254	975	228	0.70	0.54

Table 1. Results of the estimations made using the models trained with both non-augmented/unbalanced (DS 1) and augmented/balanced datasets (DS 2). *TP*, *FP*, *TN* and *FN* correspond to *True positive*, *False Positive*, *True Negative* and *False Negative*, respectively; all of them are measured in numbers of pixels. *DC* stands for *Dice Coefficient*, while *Jl* represents *Jaccard Index*, both presented as percentual quantities. The *t* designation should be read as *tiles number*. *DS* is a short form of dataset.

Comparatively to other studies focusing fruit detection, this work seems to have much space for improvement, starting by the datasets strengthening. Also, the rough segmentation proposed in this approach can be perfectly replaced by fine-grained emerging strategies such as Feature Pyramid Networks (FPN). Another aspect worth of attention would be to test complementary digital image processing algorithms to address contrast stimulation with the goal (of attempting) to mitigate chestnuts and background visual resemblance, in certain conditions and environments.

6. CONCLUSIONS AND FUTURE WORK

A preliminary study focusing the application of CNNs to detect chestnut clusters hanging in the respective trees species was addressed in this paper. For this purpose, a small dataset composed of digital photographs of three chestnut trees within Trás-os-Montes region was prepared to comply with CNN requirements. Imagery's chestnuts were manually annotated – based on segmentation masks – and, then, considering that guidelines, two of those images were algorithmically split into tiles that, in turn, were distributed by training and validation sets, each one integrating two groups of this binary problem, i.e., chestnut and background. Also, data augmentation was performed to generate pseudo-new cases for both of the involved classes and balance datasets, since background had much more examples than chestnut. A third image was saved for assessment purposes.

Using an Xception optimized by an SGD, training stages were performed to build models based on the set-up datasets, followed by assessment predictions on unseen data.

Regarding the accuracies of the tested models, values were not so high (0.59 and 0.70 Dice-Coefficient rates), but some conclusions can be drawn: dataset augmentation and balancing are useful tools to improve models' learnability, mainly in situations of data scarcity, as it is the case in this work. Also, the apparent convergence of the models at the end of training stage does not ensure reliable results, especially in problems that can be characterized by an infinity of conditions, as it typically occurs in natural environments.

To conclude, chestnut detection is not a problem with a straightforward solution, due to the characteristics of this fruit, i.e., its resembling features with background, namely, color and tonality, not to mention the several phenological states that it goes through and its recurrently reduced size. Therefore, future work must encompass a wider study with a broader dataset acquired in the most varied conditions and fruit development stage, and techniques for data stratification. Other approaches focusing fine-granularity segmentation – e.g. U-Net or Feature Pyramid segmentation – must be considered as well for upcoming studies. The results obtained in this study also encourages to explore the effectiveness of this approach to be applied in high-resolution imagery acquired from unmanned aerial vehicles, since this remote sensing platforms provide a quick way to assess either the overall status of chestnut plantations or some particular close-range aspects. Moreover, the approach proposed in this paper seems to be promising to foster the automatic detection of damages induced by chestnut gall wasp in the tree buds.

ACKNOWLEDGEMENTS

The authors would like to acknowledge project “CHIC – Cooperative Holistic View on Internet and Content” (N° 24498), financed the European Regional Development Fund (ERDF) through COMPETE2020 - the Operational Programme for Competitiveness.

The research activities of Luís Pádua were funded by the Portuguese Science Foundation (SFRH/BD/139702/2018).

REFERENCES

- Adão, T., Pinho, T., Ferreira, A., Sousa, A., Pádua, L., Sousa, J., Sousa, J.J., Peres, E., Morais, R., 2019. Digital Ampelographer: a CNN based preliminary approach, in: Progress in Artificial Intelligence: 19th EPIA Conference on Artificial Intelligence, EPIA 2019. Presented at the EPIA Conference on Artificial Intelligence, Springer, Vila Real, Portugal.
- Amara, J., Bouaziz, B., Algergawy, A., 2017. A Deep Learning-based Approach for Banana Leaf Diseases Classification, in: BTW.
- Baptista, P., Martins, A., Tavares, R.M., Lino-Neto, T., 2010. Diversity and fruiting pattern of macrofungi associated with chestnut (*Castanea sativa*) in the Trás-os-Montes region (Northeast Portugal). *Fungal Ecol.* 3, 9–19. <https://doi.org/10.1016/j.funeco.2009.06.002>
- Bargoti, S., Underwood, J., 2017. Deep fruit detection in orchards, in: 2017 IEEE International Conference on Robotics and Automation (ICRA). Presented at the 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 3626–3633. <https://doi.org/10.1109/ICRA.2017.7989417>
- Bargoti, Suchet, Underwood, J.P., 2017. Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards. *J. Field Robot.* 34, 1039–1060. <https://doi.org/10.1002/rob.21699>
- Caffe | Deep Learning Framework [WWW Document], 2013. URL <https://caffe.berkeleyvision.org/> (accessed 4.25.19).
- Chen, S.W., Shivakumar, S.S., Dcunha, S., Das, J., Okon, E., Qu, C., Taylor, C.J., Kumar, V., 2017. Counting Apples and Oranges With Deep Learning: A Data-Driven Approach. *IEEE Robot. Autom. Lett.* 2, 781–788. <https://doi.org/10.1109/LRA.2017.2651944>
- Chollet, F., 2016. Xception: Deep Learning with Depthwise Separable Convolutions. ArXiv161002357 Cs.

- Deep Cognition [WWW Document], 2018. . DeepCognition.ai. URL <https://deepcognition.ai/> (accessed 4.26.19).
- Deep Learning Toolbox [WWW Document], 2016. URL <https://www.mathworks.com/products/deep-learning.html> (accessed 4.25.19).
- Grinblat, G.L., Uzal, L.C., Larese, M.G., Granitto, P.M., 2016. Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric.* 127, 418–424. <https://doi.org/10.1016/j.compag.2016.07.003>
- Hall, D., McCool, C., Dayoub, F., Sunderhauf, N., Upcroft, B., 2015. Evaluation of Features for Leaf Classification in Challenging Conditions, in: 2015 IEEE Winter Conference on Applications of Computer Vision. Presented at the 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 797–804. <https://doi.org/10.1109/WACV.2015.111>
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. *ArXiv151203385 Cs*.
- Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *ArXiv170404861 Cs*.
- Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2016. Densely Connected Convolutional Networks.
- Kamilaris, A., Prenafeta-Boldú, F.X., 2018. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* 147, 70–90. <https://doi.org/10.1016/j.compag.2018.02.016>
- Keras Documentation [WWW Document], 2015. URL <https://keras.io/> (accessed 7.30.18).
- Lee, J.W., Yoon, Y.C., 2019. Fine-Grained Plant Identification using wide and deep learning model 1, in: 2019 International Conference on Platform Technology and Service (PlatCon). Presented at the 2019 International Conference on Platform Technology and Service (PlatCon), pp. 1–5. <https://doi.org/10.1109/PlatCon.2019.8669407>
- Lee, S.H., Chan, C.S., Wilkin, P., Remagnino, P., 2015. Deep-plant: Plant identification with convolutional neural networks, in: 2015 IEEE International Conference on Image Processing (ICIP). Presented at the 2015 IEEE International Conference on Image Processing (ICIP), pp. 452–456. <https://doi.org/10.1109/ICIP.2015.7350839>
- Liu, W., Rabinovich, A., Berg, A.C., 2015. ParseNet: Looking Wider to See Better. *ArXiv150604579 Cs*.
- Long, J., Shelhamer, E., Darrell, T., 2014. Fully Convolutional Networks for Semantic Segmentation. *ArXiv14114038 Cs*.
- Marques, P., Pádua, L., Adão, T., Hruška, J., Peres, E., Sousa, A., Sousa, J.J., 2019. UAV-Based Automatic Detection and Monitoring of Chestnut Trees. *Remote Sens.* 11, 855. <https://doi.org/10.3390/rs11070855>
- Martins, L., Castro, J.P., Bento, R., Sousa, J.J., 2015. Monitorização da condição fitossanitária do castanheiro por fotografia aérea obtida com aeronave não tripulada. *Rev. Ciênc. Agrár. Soc. Ciênc. Agrár. Port.* 38, 184–190.
- Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant Sci.* 7, 1419–1419. <https://doi.org/10.3389/fpls.2016.01419>
- Mortensen, A.K., Dyrmann, M., Karstoft, H., Jørgensen, R.N., Gislum, R., 2016. Semantic segmentation of mixed crops using deep convolutional neural network. *CIGR-AgEng Conf.* 26-29 June 2016 Aarhus Den. Abstr. Full Pap. 1–6.
- Pádua, L., Marques, P., Adão, T., Hruška, J., Peres, E., Morais, R., Sousa, A., Sousa, J.J., 2018. UAS-based Imagery and Photogrammetric Processing for Tree Height and Crown Diameter Extraction, in: Proceedings of the International Conference on Geoinformatics and Data Analysis, ICGDA '18. ACM, New York, NY, USA, pp. 87–91. <https://doi.org/10.1145/3220228.3220241>
- Pound, M.P., Atkinson, J.A., Townsend, A.J., Wilson, M.H., Griffiths, M., Jackson, A.S., Bulat, A., Tzimiropoulos, G., Wells, D.M., Murchie, E.H., Pridmore, T.P., French, A.P., 2017. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience* 6, 1–10. <https://doi.org/10.1093/gigascience/gix083>
- PyTorch [WWW Document], 2016. URL <https://www.pytorch.org> (accessed 4.26.19).
- Rahmehoonfar, M., Sheppard, C., 2017. Deep Count: Fruit Counting Based on Deep Simulated Learning. *Sensors* 17, 905. <https://doi.org/10.3390/s17040905>
- Reyes, A.K., Caicedo, J.C., Camargo, J.E., 2015. Fine-tuning Deep Convolutional Networks for Plant Recognition. *CLEF Work. Notes* 1391, 9.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv150504597 Cs*.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2014. ImageNet Large Scale Visual Recognition Challenge. *ArXiv14090575 Cs*.
- 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. <https://doi.org/10.3390/s16081222>
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv14091556 Cs*.
- Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D., 2016. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification [WWW Document]. *Comput. Intell. Neurosci.* <https://doi.org/10.1155/2016/3289801>
- Stein, M., Bargoti, S., Underwood, J., 2016. Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. *Sensors* 16, 1915. <https://doi.org/10.3390/s16111915>
- Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2016. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *ArXiv160207261 Cs*.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going Deeper with Convolutions. *ArXiv14094842 Cs*.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the Inception Architecture for Computer Vision. *ArXiv151200567 Cs*.
- TensorFlow [WWW Document], 2015. . TensorFlow. URL <https://www.tensorflow.org/> (accessed 7.30.18).
- Toda, Y., Okura, F., 2019. How Convolutional Neural Networks Diagnose Plant Disease. *Plant Phenomics.* <https://doi.org/10.1155/2019/9237136>
- Welcome — Theano 0.9.0 documentation [WWW Document], 2007. URL <http://deeplearning.net/software/theano/> (accessed 8.16.17).
- Welcome to Python [WWW Document], 2018. . Python.org. URL <https://www.python.org/> (accessed 7.30.18).
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2016. Pyramid Scene Parsing Network. *ArXiv161201105 Cs*.

Zhu, H., Liu, Q., Qi, Y., Huang, X., Jiang, F., Zhang, S., 2018.
Plant identification based on very deep convolutional
neural networks. *Multimed. Tools Appl.* 77, 29779–
29797. <https://doi.org/10.1007/s11042-017-5578-9>

Revised May 2019