Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 196 (2022) 364–370

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

Analyzing the Fine Tuning's impact in Grapevine Classification

Gabriel S. Carneiro[a], Ana Ferreira[a], Raul Morais[a,b], Joaquim J. Sousa[a,b,*] and António Cunha[a,b]

[a]*University of Trás-Os-Montes e Alto Douro, Vila Real, 5000-801, Portugal*
[b]*INESC-TEC ' INESC Technology and Science, Porto, 4200-465, Portugal*

Abstract

Wine is one the most important products from Portugal, being the grapevine variety very important to ensure uniqueness, authenticity and classification. In the Douro Demarcated Region, only certain grapevine varieties are allowed, implying the need for an identification mechanism. The ampelographs, professionals that use visual analysis to classify grapevines, are disappearing. In this situation, one possible replacement for ampelographs can be deep learning models. In previous experiments, we successfully classified 12 grapevines varieties, fine-tuning the Xception model, achieving ~0.9 in F1 score, raising the question, "*what is the impact of the fine-tuning layers' configuration in our results?*".

This paper presents an analysis of the impact of different layers' configuration in fine-tuning Xception model to classify 12 grapevine varieties with images acquired in a natural environment. Despite the model achieved F1-score of 0.92 in all configurations, using the Grad-CAM approach, we show that layers' configuration in fine-tuning implies the quality of the models' prediction. As analysis' result, we can see that the model acting as feature extractor and fully fine-tuned obtains similar results in terms of metrics and pixel contribution, and fine-tuning only the last two blocks lead the model to look at more features in the image.

Keywords: convolutional neural network; explainable artificial intelligence; grad-cam; xception

* Corresponding author.
E-mail address: jjsousa@utad.pt

10.1016/j.procs.2021.12.025

1. Introduction

Portugal is known internationally for its wine, either for the variety and for its quality. The grape varieties are among the most relevant factors in the wine production chain [1]. As it directly influences the authenticity and classification of wine, identifying the different vineyard types is fundamental for regulating production.

One of the most used approaches to identify different vineyards is the Ampelography; however, how its basis is the visual analysis, the process becomes subjective. In addition, the analysis can be exposed to interference from environmental, cultural and genetic conditions, which can introduce uncertainty into the identification process [2,3]. Another problem associated is the scarcity of professionals - nowadays, the ampelographs are disappearing.

One great alternative to replace ampelographs is Deep Learning (DL) methods. They emerged in 2012 and are now state-of-the-art in most image classification challenges and can reach performances equal or better than humans. Many works that have been using them for plants identification are referred by Sobha and Thomas [4]. In the grapevine context, there are a few works such as Milella et al. [5], that used deep learning methods for grape-bunch detection and counting based on RGB-D images, Fernandes et al. [6], that create models for identifying 64 different grape-varieties (GV) based on hyperspectral images, and our team initial-work [7], that used a deep learning model for predicting six GV based on images RGB of a single leaf with a white background.

Our initial work, Adão et al. [7] was very successful. We achieved 100% accuracy in predicting one in six GV based on images of single leaves with a white background using deep convolutional models. This experiment proved that deep learning models could successfully identify different grapes-varieties from RGB leave images. After, in recent tests, we classified 12 grapevine varieties utilizing images RGB acquired in a natural environment, achieving ~0.9 in F1-score. Our approach in both cases was fine-tuning the Xception [8] model.

With the rise of transfer learning and fine-tuning, one question emerged [9]: *what are the layers, in pre-trained models, that one can train to get the best result?* Nowadays, there is no general rule or recipe to follow to determine which layers to train [10].

Taking into account that the top layers in a Convolutional Neural Network (ConvNet) are sensitive to semantics, while more intermediate layers are specific to low-level features [11], many works explored different ways to do fine-tuning. To verify the impact of fine-tune the two last convolutional blocks of VGG-16, Yin et al. [12] showed a gain of 6% more accuracy when compared with a VGG-16 model working as a feature extractor (without fine-tuning). Grega et al. [10] developed a method called Evolution Based Fine-Tuning to select which layers should be trained for a target dataset, outperforming compared methods by a margin between 4.45% and 32.75%. Guo et al. [13] proposed too an adaptative technique called AdaFilter, using a Recurrent Neural Network to select which part of the convolutional to train, based on the activation of the previous layer.

In this paper, we present an analysis of the fine-tuning impact in the classification of grapevines, using 12 grape varieties present in the Douro Valley in Portugal based on images of grapevine leaves with the natural background. We fine-tuned the Xception model with weights pre-trained on ImageNet in three different layer configurations, comparing the results with the model acting as a feature extractor. To compare the results, we used accuracy and F1-score metrics and provided predictions explanations based on Grad-CAM method [14].

The remaining of this paper is organized as follows: Section 2 presents the used dataset, the experiments conducted and the metrics employed to evaluate our results; Section 3 presents and discusses the results achieved; and Section 4 closes the paper, presenting main conclusions and future work.

2. Methods

To evaluate the impact of fine-tuning in a deep learning-based system for the classification of GV based on grapevine images, we follow the pipeline shown in Figure 1.

First, we collected data for 12 grape varieties – Codega (CD), Rabigato (RG), Malvasia Fina (MF), Tinta Amarela (TA), Malvasia Preta (MP), Tinta Barroca (TB), Malvasia Rei (MR), Tinta Roriz (TR), Moscatel Galego (MG), Tinto Cao (TC) Mourisco Tinto (MT), and Touriga Nacional (TN) – and organized it into a trainable dataset. Then, we trained and evaluated 4 deep learning models, varying the trained layers in each model, and finally, we analyzed the results.

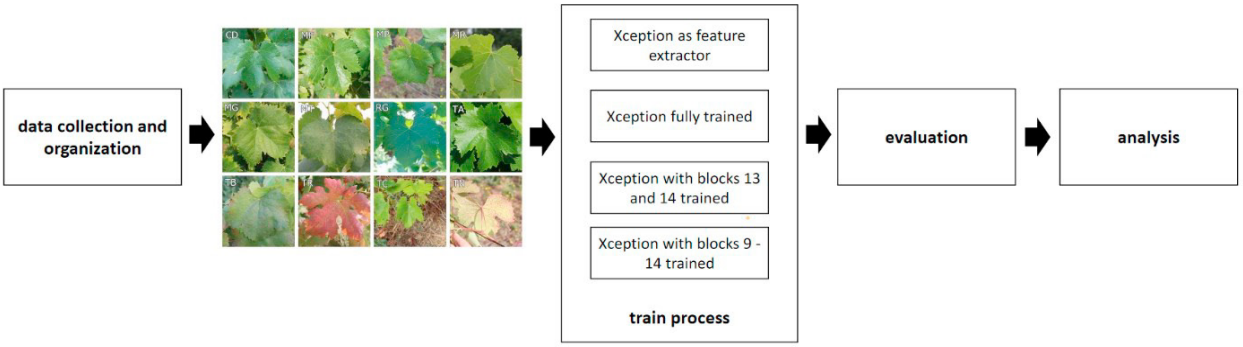

Figure 1. Methodology Pipeline

2.1. Data collection and preparation

To create the dataset, we collect grapevines images for each type in different stages of leaves growing, establishing two main gridlines: a) only one grapevine-type per image and b) the plat should fill most of the image. We look for images that allow an expert to identify the grapevine by observation. The images were taken with a Canon EOS 600D camera, equipped with a 50 mm *f*/1.4 lens, and 18 megapixels.

The images were redistributed randomly into training, validation and test sets, with the proportion 70%, 20% and 10%, respectively. We applied data augmentation to the training set, generating 10 images for each in the training set, applying rotations, shifts, variations in brightness and horizontal/vertical flips. The final dataset has 6718 training images, 132 validation images and 72 test images. For the train set, the quantity of samples per class varies, meaning that the set is unbalanced.

2.2. Model Training and Fine Tuning

We decided to use a state-of-the-art deep learning model with transfer learning - we chose the Xception model [4]. The weights of ImageNet were used, replacing the classifier at the top of the network. To train the models, we froze the network weights until the new classifier converged. In the end, we fine-tuned the model, unfreezing all the weights.

The Xception model was fine-tuned in three different layer configurations. The experiments' summarization is shown in Table 1. In the first experiment using fine-tuning, we train the entire model in the second phase (Experiment 2). Considering that layers more in the top of ConvNet represent high-level features, in Experiment 3, we train all the layers in the 13 and 14 blocks, the model's more-top blocks. After, we enter more in the model's middle training the blocks between 9 and 14 in Experiment 4.

Table 1. Experiments' Summarization.

Experiment	Trained Blocks in Fine Tuning
Experiment 1 (Exp. 1)	-
Experiment 2 (Exp. 2)	All
Experiment 3 (Exp 3.)	13, 14
Experiment 4 (Exp 4.)	9, 10, 11, 12, 13, 14

As the classifier, we choose a dense layer with 40 neurons, with activation function Rectified Linear Units (ReLU), followed by a Dropout of 25% and a Dense layer, with 12 neurons with a SoftMax activation function. The classifier

was attached to the network through an Average Global Pooling at the output of the convolutional part and for the classifier, as in the standard Xception model.

All the models were trained for 100 epochs, with a batch size of 12, and with Early Stop with 15-season patience. We used the Stochastic Gradient Descent (SGD) as the optimizer, initiated with a learning rate of 0.1 and step decay in five epochs for the first train phase. For the fine-tuning phase, we changed the learning rate to 0.0001, keeping SGD and step decay configurations. The Focal Loss [15] was used with default parameters to deal with the problem of unbalanced classes during training (code available at [16]).

All the models used the Keras Framework with TensorFlow as backend, running in the Google Collaboratory.

2.3. Model Evaluation

To evaluate the models, we use the usual metrics for classification problems: the accuracy and, as this is an unbalanced dataset, the F1-score.

Additionally, to interpret the results obtained by the classifier, we used the Grad-CAM [14] method that allows visualizing a heatmap of the positive pixels' contribution to each class. Aiming view pixels with high contribution, we define a threshold where only pixels with contribution greater 10% than the maximum pixel's contribution in the image were shown in obtained heatmaps.

3. Results and Discussion

The results achieved by all experiments are 0.92 of F1-Score and 0.92 of accuracy. The confusion matrix (CM) for each experiment can be seen in Figure 2.

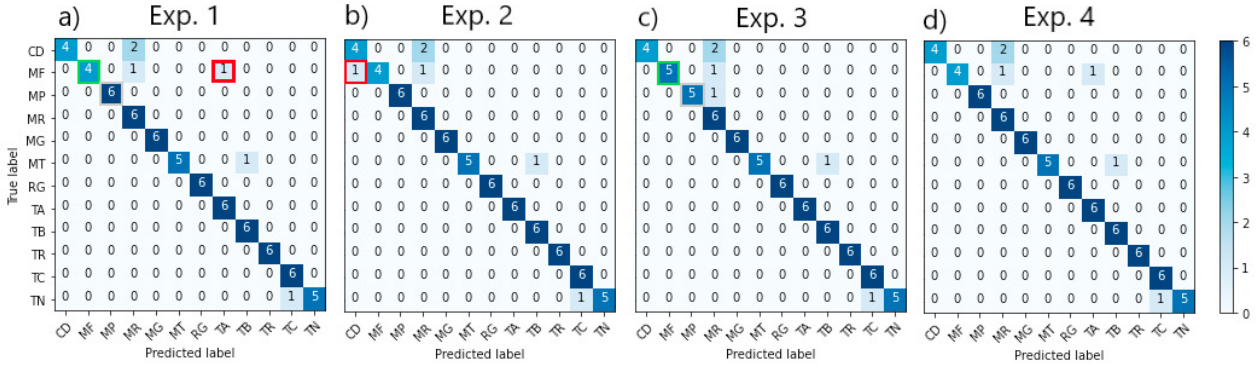

Figure 2. Confusion Matrix for each experiment (the changes are marked with squares)

Comparing the results of predictions, we can see that, against the equality of the metrics, applying fine-tuning modifies the obtained results. The first case was the changing of the class in predicted images in the class MF if we compare the Exp. 1' CM and the Exp. 2' CM (red squares in Figure 2 (a) and (b)). In Figure 3 (a), we can see the Grad-CAM heatmaps and scores for both involved classes. Note that the score is distributed because, in both experiments, these classes are the top-2 greatest scores. Furthermore, the contribution of the pixels was mostly inverted, meaning that the most pixels that contributed for the class TA in the Exp 1, in Exp 2 contributed to the class CD.

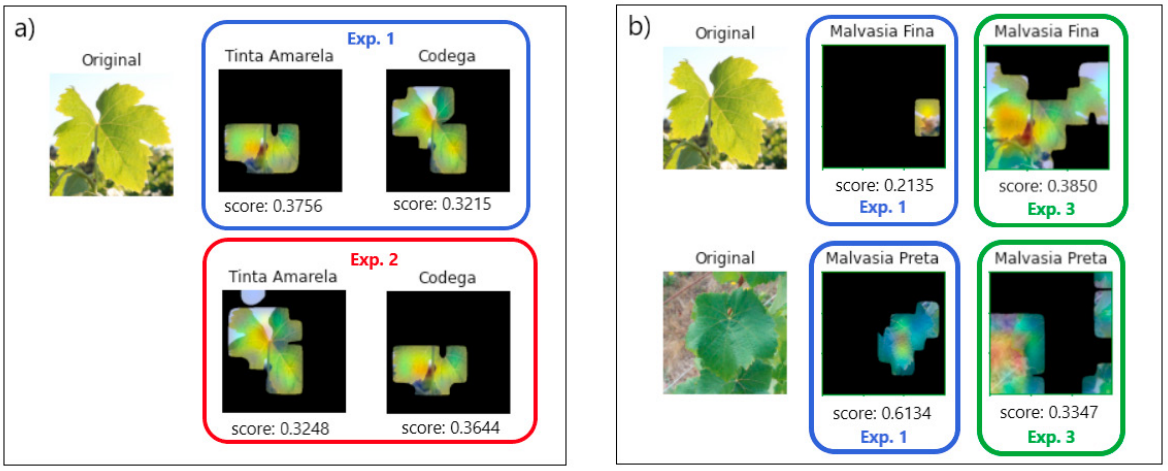

Figure 3. a) Comparison between heatmaps generated by Exp. 1 and Exp. 2 for the same image; b) Comparison between heatmaps generated by Exp. 1 and Exp. 3 for two different images.

In the Exp. 3 CM, we can see that the model decreases the error for the class MF (green squares in Figure 2 (a) and (c)) but increase the error for class MP (grey squares in Figure 2 (a) and (c)), keeping the accuracy and F1 score. As we can see in the first image shown in Figure 3 (b), the fine-tuning of the blocks 13-14 increase de contribution of the pixels, leading the model to consider one greater region, but in the second sample, in Figure 3 (b), we can see that model used background region to make the decision. Note that in both cases, the given scores to involved classes are small, considering that a big score is greater than 0.7, and in the second image, the fine-tuning decreased almost by 50%.

We identify two other interesting cases when analyzing the heatmaps. There are cases, e.g. Figure 4 (a), where the model increase pixel contribution of and softmax scores when fine-tuning only the last blocks, but the pixel contribution decrease when other blocks more on the middle are trained.

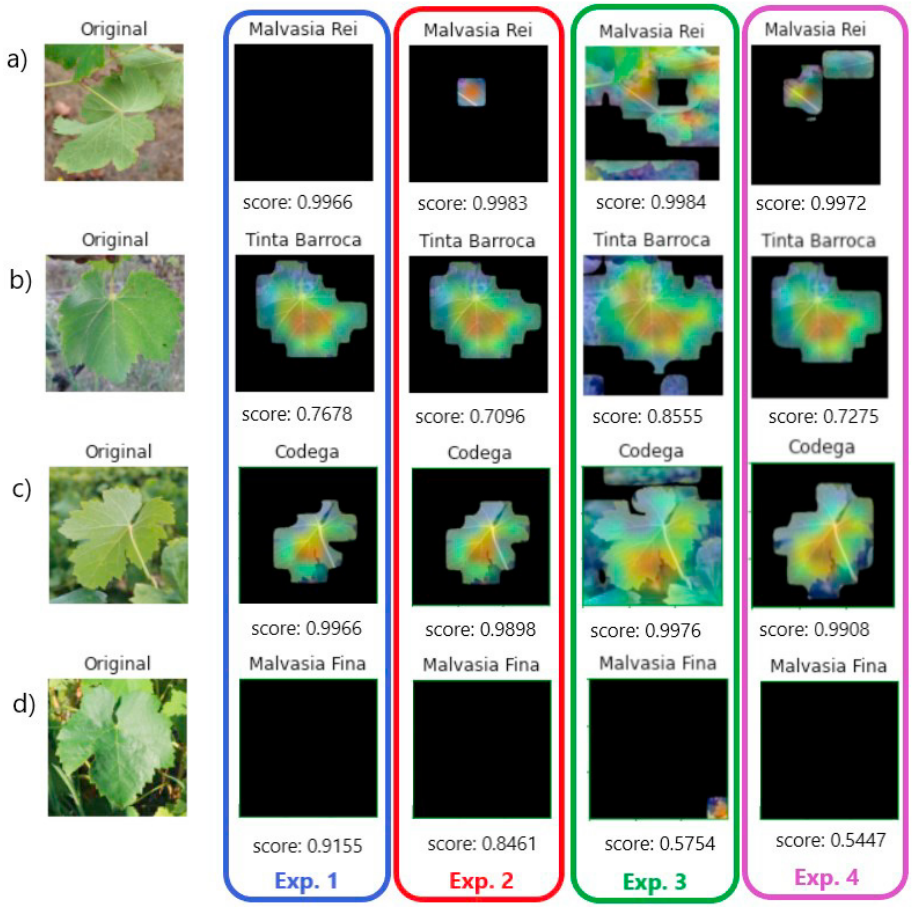

Figure 4. Heatmaps and scores examples generated for all the models for the same samples.

Note that in all Figure 4 examples, although Exp. 3 increase the pixel contribution, it does not imply an improvement in the prediction. In Figure 4 (d), despite some pixels contribute to the prediction, if compared with the results for the same sample in other experiments, the score of the class was decreased. In some cases, this can lead the model to error. If we consider de features that the model was using for prediction, in Figure 4 (c) in the Exp3. background-pixels contributed so much if compared with the results obtained for the same sample in Exp. 1, being that some pixels of background contributed more than leaf's pixels. This situation implies directly in the confidence of the model.

One can observe there are samples where exists few changes in the region of contribution if comparing the model acting as feature extractor (Exp. 1) and the model totally fine-tuned (Exp 2), e.g. Figure 4 (c) and (d).

4. Conclusion

Wine is one the most important products from Portugal, being the grapevine variety very important to ensure uniqueness, authenticity and classification. In Douro Demarcated Region only certain grapevine varieties are allowed, implying the need for an identification mechanism.

This work presented a deep learning model to classify 12 different grapevine varieties using images acquired in a natural environment. We fine-tuned the Xception model and, using the Grad-CAM method, we analyzed the impact of different fine-tuning configurations, changing the layers that were trained. All the models obtained an F1-score of 0.92. As analysis' result, we can see that the model acting as feature extractor and fully fine-tuned obtain similar

results in terms of metrics and pixel contribution, and fine-tuning only the last two blocks led the model to look at more features in the image, but not necessarily increase the model performance and turn it more reliable.

For future work, we are planning to increase the number of samples and classes in the dataset and fine-tuning more models, comparing the results.

Acknowledgements

This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020.

References

[1] Moncayo S, Rosales JD, Izquierdo-Hornillos R, Anzano J, Caceres JO. Classification of red wine based on its protected designation of origin (PDO) using Laser-induced Breakdown Spectroscopy (LIBS). Talanta 2016;158:185–91. https://doi.org/https://doi.org/10.1016/j.talanta.2016.05.059.

[2] Garcia-Muñoz S, Muñoz-Organero G, de Andrés M, Cabello F. Ampelography - An old technique with future uses: the case of minor varieties of Vitis vinifera L. from the Balearic Islands. J Int Des Sci La Vigne Du Vin 2011;45:125–37. https://doi.org/10.20870/oeno-one.2011.45.3.1497.

[3] Tassie L. Vine identification--knowing what you have. Grape Wine Res Dev Corp Â€“Australian Gov GW RDC Innov Network, Greenhill Road Wayv 2010.

[4] Sobha P, Thomas P. Deep Learning for Plant Species Classification Survey, 2019, p. 1–6. https://doi.org/10.1109/ICAC347590.2019.9036796.

[5] Milella A, Marani R, Petitti A, Reina G. In-field high throughput grapevine phenotyping with a consumer-grade depth camera. Comput Electron Agric 2019;156:293–306. https://doi.org/https://doi.org/10.1016/j.compag.2018.11.026.

[6] Fernandes AM, Utkin AB, Eiras-Dias J, Cunha J, Silvestre J, Melo-Pinto P. Grapevine variety identification using “Big Data” collected with miniaturized spectrometer combined with support vector machines and convolutional neural networks. Comput Electron Agric 2019;163:104855. https://doi.org/https://doi.org/10.1016/j.compag.2019.104855.

[7] Adão T, Pinho TM, Ferreira A, Sousa A, Pádua L, Sousa J, et al. Digital Ampelographer: A CNN Based Preliminary Approach. In: Moura Oliveira P, Novais P, Reis LP, editors. Prog. Artif. Intell., Cham: Springer International Publishing; 2019, p. 258–71.

[8] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conf. Comput. Vis. Pattern Recognit., 2017, p. 1800–7. https://doi.org/10.1109/CVPR.2017.195.

[9] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? CoRR 2014;abs/1411.1.

[10] Vrbančič G, Podgorelec V. Transfer Learning With Adaptive Fine-Tuning. IEEE Access 2020;8. https://doi.org/10.1109/ACCESS.2020.3034343.

[11] Zheng L, Zhao Y, Wang S, Wang J, Tian Q. Good Practice in {CNN} Feature Transfer. CoRR 2016;abs/1604.0.

[12] Yin X, Chen W, Wu X, Yue H. Fine-tuning and visualization of convolutional neural networks. 2017 12th IEEE Conf. Ind. Electron. Appl., 2017, p. 1310–5. https://doi.org/10.1109/ICIEA.2017.8283041.

[13] Guo Y, Li Y, Wang L, Rosing T. AdaFilter: Adaptive Filter Fine-Tuning for Deep Transfer Learning. Proc AAAI Conf Artif Intell 2020;34:4060–6. https://doi.org/10.1609/aaai.v34i04.5824.

[14] Selvaraju RR, Das A, Vedantam R, Cogswell M, Parikh D, Batra D. Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization. CoRR 2016;abs/1610.02391.

[15] Lin T-Y, Goyal P, Girshick RB, He K, Dollár P. Focal Loss for Dense Object Detection. CoRR 2017;abs/1708.0.

[16] Zhang C. Multi-class classification with focal loss for imbalanced datasets n.d.