Going for 2D or 3D? Investigating various Machine Learning Approaches for Peach Variety Identification

Anna Wróbel[1], Gregory Gygax[1], Andi Schmid[2] and Thomas Ott[1]

[1] Zurich University of Applied Sciences ZHAW, School of Life Sciences and Facility Management, Wädenswil, Switzerland
`wrob@zhaw.ch, gygg@zhaw.ch, ottt@zhaw.ch`
[2] Realisation Schmid, Scharans, Switzerland
`andi@realisation-schmid.ch`

Abstract. Machine learning-based pattern recognition methods are about to revolutionize the farming sector. For breeding and cultivation purposes, the identification of plant varieties is a particularly important problem that involves specific challenges for the different crop species. In this contribution, we consider the problem of peach variety identification for which alternatives to DNA-based analysis are being sought. While a traditional procedure would suggest using manually designed shape descriptors as the basis for classification, the technical developments of the last decade have opened up possibilities for fully automated approaches, either based on 3D scanning technology or by employing deep learning methods for 2D image classification. In our feasibility study, we investigate the potential of various machine learning approaches with a focus on the comparison of methods based on 2D images and 3D scans. We provide and discuss first results, paving the way for future use of the methods in the field.

1   Introduction

The identification of plant species and varieties has always been an important skill in human culture and a driving factor for agricultural success. The diversity of crops is important for resilient cultivation and ecosystems as well as healthy nutrition, but has come under pressure, e.g. due to monoculture farming methods. Thus, various initiatives aim to preserve the diversity of crops. For specific crops such as peaches or apples the task of identifying varieties requires very specific expertise, and even experts can get ambiguous results. Therefore, genetic identification has become a major tool for variety identification [1]. However, DNA analysis takes time, is relatively expensive and is not entirely unambiguous [2].

Obviously, machine learning methods offer an alternative for the identification of crop varieties. It does not come as a surprise that the identification and classification of different plant species has a long tradition in the literature of neural network applications and machine vision (e.g., [3, 4]). However, the identification of varieties within

a species is a difficult problem as subtle differences can point to relevant discriminating features [5]. Given its spectacular success and development over the past few years, it is tempting to use deep neural network (DNN) technology based on 2D images for this task [6]. Traditionally, however, for many crops manually calculated 3D shape descriptors have been used by experts since relevant differentiating features of varieties seem to be mainly encoded in the 3D structure of the plants or their seeds or stones respectively [7]. As the advances in technology render an automated approach using 3D scanning feasible, some authors have suggested using 3D scans in combination with targeted feature engineering as a basis for crop identification, e.g. [8].

In this contribution, we present the first, to the best of our knowledge, study to investigate and compare the potential of 2D image-based and 3D scanning-based machine learning approaches for identifying varieties of peaches. As with all crops, the identification problem is of great importance for peach breeding. Peach varieties are expressed in differences in the structure of the peach stones. The goal of the study is to pre-evaluate the potential of different methods that can offer breeders a cheap and fast tool which can complement, if not replace, DNA analysis. Based on the established literature in the field of plant identification, we decided to focus on two different methodological lines. We evaluated the performance of convolutional neural networks (CNN) based on 2D images and we assessed several classification methods (support vector classifier SVC, random forest classifier, linear discriminant analysis LDA, k-nearest neighbor classifier KNN) in combination with 3D descriptors gained from a Fourier analysis of the 3D scans.

Finally, in regard to future practical usage in the field, we focus on "cheap" equipment. Thus, spectral imaging technology was not considered.

2   Materials and Methods

2.1   Peach Sample Preparation, Data Acquisition and Preprocessing

A selection of eight representative varieties of peaches with a total of 190 different fruits were used as the data basis for this study. After carefully cleaning and drying the stones of the peaches, 3D scans and 2D images were taken, according to the protocols described below.

2D Images and Imaging Protocol. For 2D images a commercially available camera (Sony DSC-HX300) was used. Objects were placed in a light tent, always in the same place. Pictures of a resolution of 5184x2920 pixels were taken with the same camera and light settings. Every object was flipped in a repeatable manner resulting in 6 samples per object. The images were further preprocessed in order to obtain centered 256x256 images. Each picture was cropped based on the weighted center of mass and then resized to size 256x256 padded with edge pixels values in height. It was necessary as some of the objects were much longer than wide. The aspect ratio of the images remained unchanged.

3D Scanner and Scanning Protocol. For 3D scanning a commercially available scanner (PT-M scanner from Isra Vision) was used. For scanning, objects were placed in the middle of a turning table. The scanner was set to turn the turntable automatically 16 times. After a full turn, the object was flipped around its longest axis and scanned again using 16 turntable steps. The scans were automatically aligned by the software that comes with the scanner. In case of gross errors (misalignments), the data were discarded and the object was scanned again. The resulting scans were in a form of a collection of meshed points in 3D space. They were exported in the STL format.

As the orientation in space of this representation is not identical for every scanned object, each object was first shifted such that its center of gravity lied in the origin of the 3D coordinate system. Then, it was rotated such that its principal rotation axes were aligned with the axes of the coordinate system. For the first part, the center of gravity was defined as the component-wise mean of every vector. In the next step, the components of the inertia tensor were calculated (assuming the same weight for every point). The inertia tensor is described by the symmetric matrix

$$I = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix} \tag{1}$$

with $I_{l \neq k} = -\sum x_l x_k$ and $I_{11} = \sum(x_2^2 + x_3^2)$, $I_{22} = \sum(x_1^2 + x_3^2)$, $I_{33} = \sum(x_2^2 + x_1^2)$.

The data were rotated such that the principal axes of the inertia tensor, i.e. the eigenvectors, align with the axes of the coordinate system (see Fig.1). As the points are not equally spaced, the alignments are not perfect and there can be some variation between scans. Furthermore, since only the axes are aligned and the stones are not symmetric, not all of the stones have the same orientation. Therefore, every object was flipped once around every axis, such that 4 samples per scan were created (original and rotated around the three axes). The data were then transformed from vectors to a spatial grid with binary encoding cells, representing the surface of the stone. The resolution of the grid was 0.1 mm.
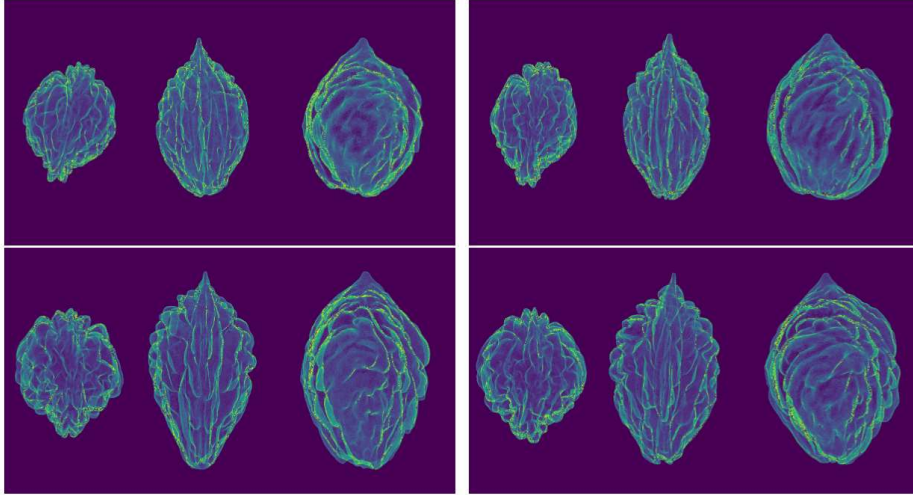
Fig. 1. Projections of 3D scans onto each principal axis for two different varieties. In each row, two different stones of the same variety are shown.

The data basis that was used for the classification tasks is summarized in Table 1. For the classes, i.e. peach varieties, an encoding scheme used by the breeder A. Schmid was applied. Additionally, 3 varieties have specific names.

Table 1. Overview data basis (* for class 929 one stone had to be excluded from the scans due to damaged data).

Peach variety	Number of stones	Number of images	Number of scans
146	25	150	100
303	25	150	100
929	22	132	84*
930	25	150	100
349	18	108	72
101 zephir	25	150	100
102 nectaross	25	150	100
103 sweet dream	25	150	100
Total	**190**	**1140**	**756**

2.2 CNN Approach for 2D Images

Considering the small data set size, transfer learning based on a pretrained convolutional net was used for the classification of the images [9]. We evaluated several image classification models, including InceptionV3, Resnet50, MobileNetV2, VGG16 and

VGG19 as the convolutional base, all pretrained on the ImageNet dataset [10, 11]. Classification with VGG16 model was the only successful approach. We experimented with adding 2 to 3 dense layers consisting of 128 up to 512 neurons with dropout ranging from 0.2 to 0.5 on top of the nontrainable convolutional base. Adam and SGD optimizers with categorical cross-entropy loss function and accuracy metrics were investigated. In the final architecture, the VGG16 convolutional base was followed by global max pooling, 30% dropout and two dense layers. The first dense layer consisted of 256 neurons with ReLU activation function while the classifier on top of it was an 8-neuron dense layer with a softmax activation function. The final model was compiled with an SGD optimizer with a learning rate of 0.001 and a momentum of 0.9. The target vector was one-hot encoded. Training data were augmented with ImageDataGenerator.

For each training, 10% of the images were used as a test set, another 10% of the remaining samples were used as a validation set and the rest was used as a training set. The model was trained for 100 epochs with an early stopping condition. The performance of the model was assessed using 10-fold cross-validation with a stratified fold split.

The model was built using Python with TensorFlow2 and trained on a Tesla P100 GPU.

2.3   Classification Methods for 3D Scans

We evaluated several different machine learning algorithms in combination with a feature selection procedure based on a 3D Fourier analysis. The evaluated approaches are Linear Discriminant Analysis (LDA), Support Vector Classifier (SVC), Random Forest Classifier (RF) and k-Nearest Neighbors Classifier (KNN). All these approaches were applied after a preceding feature engineering step based on 3D Fourier coefficients. To this end, the spatial grid of the scans was transformed using a fast Fourier transform. From the obtained Fourier domain, the (50, 50, 50) 'corners' of the Fourier spectrum were considered. Since the Fourier spectrum exhibits point symmetry and thus redundancy, only the 4 lower corners were taken into account, the imaginary parts of the coefficients were discarded and only the real parts were used. Important features were then selected using ANOVA by keeping only the frequencies with a p-value below the 0.9999 quantile of all p-values. In this way, 100 frequencies were selected that were then used as a feature vector. Before training the classifiers, each component was scaled by the z-transform (centered around 0, with std = 1). Only the training data were used for fitting the scaler.

Hyperparameters of the models were tuned with GridSearchCV. For final classification we used LDA with singular value decomposition solver, RF with Gini impurity criterion and KNN with 5 nearest neighbors and Euclidean distance. The SVC model achieved the highest classification accuracy with radial basis function kernel type, regularization parameter increased from 1 to 40 and kernel coefficient gamma set to 0.01.

The models were built using Python with scikit-learn and trained on an Intel Xeon-based cluster computing node. They were evaluated using 10-fold cross-validation with a stratified fold split.

3 Results

The results for all the methods based on 10-fold cross-validation are summarized in Table 2. Generally, the accuracy of the best methods is around 90% with the best 3D-based methods slightly above (92.2% for LDA and 91.9% for SVM) and the 2D CNN method slightly below this value (89.2%). In comparison, the accuracy of the RF and KNN models is significantly lower (84.1% and 83.1% respectively).

Table 2. Accuracy of different methods based on 10-fold cross-validation

Method	Accuracy (mean +/- stdv)
2D CNN	0.892 +/- 0.036
3D SVC	0.919 +/- 0.052
3D LDA	0.922 +/- 0.054
3D Random Forest	0.841 +/- 0.095
3D KNN	0.831 +/- 0.084

To further invest and understand these results we take a look at the normalized confusion matrices, averaged over the 10-fold cross-validation (Fig.2: 2D-based CNN; Fig.3: 3D-based methods)
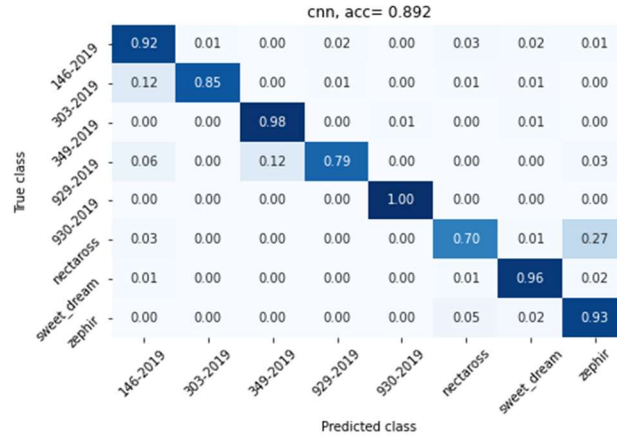

Fig. 2. CNN normalized confusion matrix averaged over all 10 folds of the k-fold cross-validation with a stratified fold split.
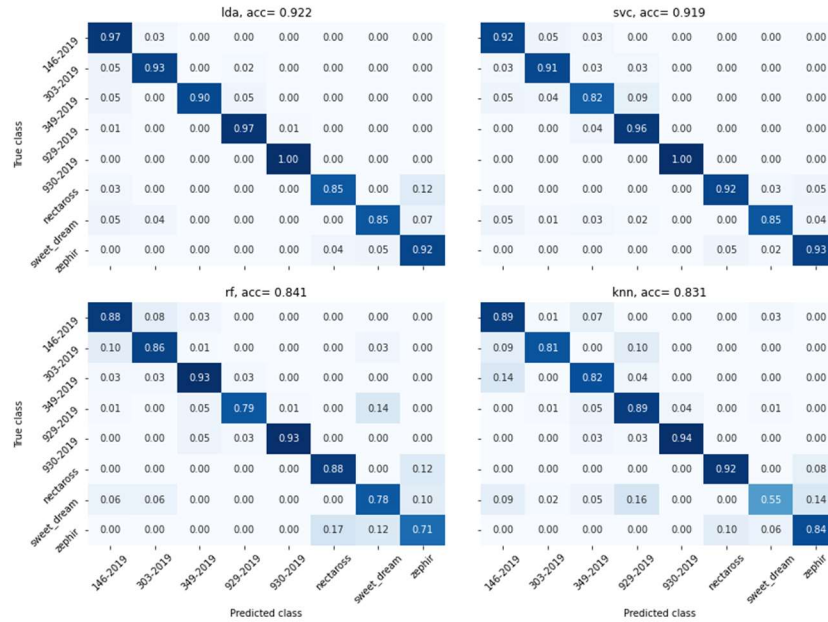
Fig. 3. Normalized confusion matrices for the 3D-based machine learning methods (from top left to bottom right: LDA, SVC, RF, KNN). Each matrix is averaged over all 10 folds of the k-fold cross-validation with a stratified fold split.

For the 2D-based CNN approach, the main difficulty seems to occur for the discrimination of the variety *nectaross* from *zephyr* as in average 27% of images of the *nectaross* class are classified as *zephyr* (Fig. 2). This problem seems to be much less pronounced for the discrimination based on 3D scanning (Fig. 3). In particular, in the case of SVC, small misclassification errors seem to occur for various classes in a rather arbitrary fashion, not hinting at a specific problem of two classes. A possible explanation for the problem of discriminating *nectaross* from *zephyr* in the 2D case is revealed when looking at the actual stones and their respective images (Fig. 4). The stones of the two varieties exhibit a similar structure. They mainly differ in size. The size information is, however, lost for the 2D images as they are automatically rescaled. In turn, the analysis of the Fourier-based feature vectors in the 3D approach shows that the low frequencies describing the coarse-grained structure of the stones play an important role for the classification. Hence, size is a feature that is definitely exploited in the 3D-based approach.

Fig. 4. Example of the 2 sorts that are misclassified most frequently. On the left *nectaross*, on the right *zephyr*.

4   Conclusions and Future Work

Species and variety identification is an important problem in crop breeding, which can potentially benefit greatly from machine learning-based pattern recognition methods. We investigated several machine learning approaches for peach variety identification based on 2D images and 3D scans of peach stones. The goal of the study was to learn more about the potential, strengths and weaknesses of different approaches for this particular problem. Our findings can be summarized as follows.

Despite a relatively small data basis with 190 peach stones, a 2D-based CNN approach looks promising. In fact, when looking at the images, an accuracy of nearly 90% seems surprising and speaks for the potential of the method. On the one hand, many stone sorts differ in color which is beneficial for this classification method. On the other hand, the method does not make use of size information of the stones, which can be considered a drawback. However, with a larger data basis and perhaps a different imaging protocol, even better results can be expected.

Among the approaches based on 3D scans of the stones, LDA and SVC were the most successful ones with an accuracy even larger than 90%. We used an automated feature engineering preprocessing based on 3D Fourier analysis and an ANOVA-based feature selection, resulting in feature vectors that mainly describe the coarse-grained structure of the stones. This automated preprocessing can be challenged and leaves room for improvement. An alternative could be offered by volumetric CNN directly applied on the 3D scans. The first attempt with this idea, however, has not yet shown conclusive results and has not been included in this study.

In conclusion, both 2D and 3D based methods showed promising accuracies on the basis of a limited data set. We are confident that an accuracy of 95% can be achieved. This will provide a basis for stable applications in the field, offering an alternative to DNA analysis. As 2D and 3D methods exploit to some extent different features for the classification, a combined approach could be beneficial. From an overall methodological perspective, the case of peach stone classification could thus be an ideal playground for combining 2D and 3D images. Further investigations are ongoing. Following approaches are taken into consideration: volumetric CNNs, multi-view CNNs aggregating 2D projections of the 3D scans, combination of the two [12] or a combination with 2D images.

Acknowledgements

References

1. Arús, P., Verde, I., Sosinski, B. Zhebentyayeva, T., Abott A. G.: The peach genome. Tree Genetics and Genomes. 531-547. DOI 10.1007/s11295-012-0493-8 (2012).
2. Singh, B. D., Singh, A. K.: Marker-Assisted Plant breeding: Principles and Practices. Springer Science and Business Media (2015).
3. Wäldchen, J., Mäder, P.: Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. Archives of Computational Methods in Engineering. 25 (2):507–543. doi:10.1007/s11831-016-9206-z (2017).
4. Gan, Y. Y., Hou, C. S., Zhou, T., Xu, S. F.: Plant Identification Based on Artificial Intelligence. Advanced Materials Research, 255-266, pp. 2286-2290 (2011).
5. Wäldchen, J., Rzanny, M., Seeland, M., Mäder, P.: Automated plant species identification – Trends and future directions. PLoS Comput Biol 14(4): e1005993 (2018).
6. Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., Joly, A.: Going deeper in the automated identification of Herbarium specimens. BMC Evolutionary Biology 17:181 (2017).
7. Rivera, A., Roselló, S., Casanas, F., Seed curvature as a useful marker to transfer morphologic, agronomic, chemical and sensory traits from Ganxet common bean (Phaseolus vulgaris L.). Scienta Horticulturae 197, pp. 476-482 (2015).
8. Karasik, A., Rahimi, O., David, M., Weiss, E., Drori, E.: Developement of a 3D seed morphological tool for grapevine variety identification, and its comparison with SSR analysis. Scientific Reports 8:6545 (2018).
9. Soekhoe, D., Putten, P., Plaat, A.: On the Impact of Data Set Size in Transfer Learning Using Deep Neural Networks. 50-60. DOI 10.1007/978-3-319-46349-0_5 (2016).
10. Krizhevsky, A., Sutskever, I., Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 1097–1105 (2012).
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. arXiv:1512.03385v1 (2015).
12. Hedge, V., Zadeh, R.: FusionNet: 3D Object Classification Using Multiple Data Representations. arXiv:1607.05695v3(2016).