Document downloaded from:

http://hdl.handle.net/10251/201636

This paper must be cited as:

López García, F.; Valiente González, JM.; Escriche Roberto, MI.; Juan-Borras, MDS.; Visquert Fas, M.; Atienza-Vanacloig, V.; Agustí-Melchor, M. (2023). Classification of Honey Pollens with ImageNet Neural Networks. Lecture Notes in Computer Science. 14185:192-200. https://doi.org/10.1007/978-3-031-44240-7_19

The final publication is available at

https://doi.org/10.1007/978-3-031-44240-7_19

Copyright Springer-Verlag

Additional Information

Classification of Honey Pollens with ImageNet Neural Networks

Fernando López García[1], José Miguel Valiente González[1], Isabel Escriche Roberto[2], Marisol Juan Borrás[2], Mario Visquert Fas[2], Vicente Atienza Vanacloig[1] and Manuel Agustí Melchor[1]

[1] Instituto de Automática e Informática Industrial, Universitat Politècnica de València, Valencia, Spain
[2] Instituto de Ingeniería de Alimentos para el Desarrollo, Universitat Polèctinca de València, Valencia, Spain
flopez@disca.upv.es

Abstract. The classification of honey pollen grains is done to classify honey according to its origin which is of great importance in terms of marketing. This visual work is currently done by human specialists counting and classifying the pollen grains in microscopic images. This is a hard and time-consuming task. Thus, automated methods are required to overcome the limitations of the conventional procedure. This paper deals with the automatic classification of honey pollens using five representative Neural Networks coming from the ImageNet Challenge: VGG16, VGG19, ResNet50, InceptionV3 and Xception. The ground truth is composed of 9983 samples of 16 different types of pollens corresponding to citrus and rosemary pollens and its companions. The best result was obtained with the InceptionV3 network, achieving and accuracy of 98.15%, which is an outstanding result.

Keywords: Pollen Classification, ImageNet Challenge, Deep Learning, Convolutional Neural Networks.

1 Introduction

Palynology is the study of plant pollen, spores and certain microscopic plankton organisms. A specific and important task within this science is recognizing plants from the pollen grains, these are widely used as fingerprint. In apiculture, this classification is used to identify nectar sources. In archaeology, pollen fossils are analysed to reconstruct ecological and climate conditions during past periods. Pollen is also involved in the search of oil and gas for commercial purposes. And in forensic analysis and pollen forecasting. Accurate identification of pollen types is a relevant issue in all these scenarios. Thus, automated methods for pollen identification are required to overcome the limitations of the conventional procedure [1] which is a hard task where human specialists classify and count pollen samples from microscopic images. Many industries rely on the accuracy of this manual classification process, which is reported to be around 67% [2]. In this paper, we focus on the apiculture sector where pollen classification is performed to identify the nectar source of the honey.

Automated pollen classification started in the later decades of the 20th century. But it has been in the 21th century where more progress has been made in this field, helped by the powerful increase in computational capacities. Previous approaches are summarized in [3] and [4]. They can be divided into image-based and non-image based methods [2]. Non-image based methods use alternative characteristics, for example fluorescence, Fourier-Transform infrared, and Raman spectroscopy. Image-based methods typically involve defining and extracting discriminant features from pollen images, followed by sorting via statistical or machine learning classifiers. These image-based methods fall into three different categories based on the type of used features [2]: discriminant features are visual/geometrical (e.g., shape, symmetry, diameter, etc.); discriminant features are texture-based (e.g., grey-level co-occurrence matrices, entropy features etc.); and a combination of the two approaches.

However, a new approach has emerged to deal with image-based pollen classification. This is Deep Learning, a method that has shown great effectiveness in other areas. This new approach uses a model that determines and extracts the features itself, rather than being defined by human specialists.

Works applying this new approach outperform the traditional methods. Most of these works are summarized in [2] where also a table is provided comparing traditional and Deep Learning approaches. Concerning the Deep Learning methods, we have several works like [5] which achieved a 94% of training accuracy on a dataset of 30 pollen types. Their results are based on the training set and no information is given about how the model behaves with unseen images. The same occurs with [6] and [7]. In the first case they achieved 100% of accuracy on 10 very different pollen grains using transfer learning with the VGG16 network. In the second case they reported 99.8% of accuracy on 5 different types of pollen. In [4] researchers improved classification of pollen grain images of the POLEN23E dataset (30 pollen types) by three different applications of Deep Learning convolutional neural networks achieving a 97% of accuracy. In a recent work [2] they obtain very good results, 98% of accuracy, on the most complete dataset until today, 19,000 samples with 46 different types. They used different techniques of image pre-processing and data augmentation to feed a pre-trained convolutional neural network, retrained by transfer learning to extract features from one of its deepest layers. Moreover, these automatically extracted features are used to perform classification with a linear discriminant classifier. The behaviour of the model is good, giving a 98% of accuracy in unseen sets of images. Finally, also in a recent work [8], they perform an approach similar in part to our approach and use pre-existing convolutional neural networks to classify up to 73 different types of pollens with 2523 samples. They achieve the best accuracy results with the DenseNet-201 (95.7%) and ResNet50 (94.0%) networks.

In this paper we use five pre-existing networks that were developed in the context of the ImageNet Challenge to perform honey pollen classification, specifically on the citrus and rosemary pollens and its companions. The ImageNet Challenge has taken place in recent years and was designed to obtain the best possible results on a database of 1.2 million images corresponding to 1000 different classes. The challenge is oriented to the use of deep learning and there have been several networks that have been presented in these years. We have chosen five of the most representative: VGG16, VGG19, Inception, Xception and ResNet50.

Next section deals with the materials and medthos. Section three shows the experimental work. And the section four presents the summary of the paper.

2 Materials and Methods

2.1 Ground Truth

In total, the ground truth was composed of 9983 samples taken by specialists from microscope images. They also were labelled by specialists. These samples correspond to 16 different types of pollens with a number of samples per type between 70 and 3279, see Table 1. All were samples of Orange Blossom, Rosemary and their companions (Bottom, Bubble, European Olea N.C., Cistus sp. NC, Starch, Brassicaceae, Citrus sp., Echium sp., Legumineuses, Onobrychis sp., Prunus dulcis, Quercus sp., Rosmarinus officinalis, Thymus sp., Taraxacum type, Umbellifers). See Figure 1.

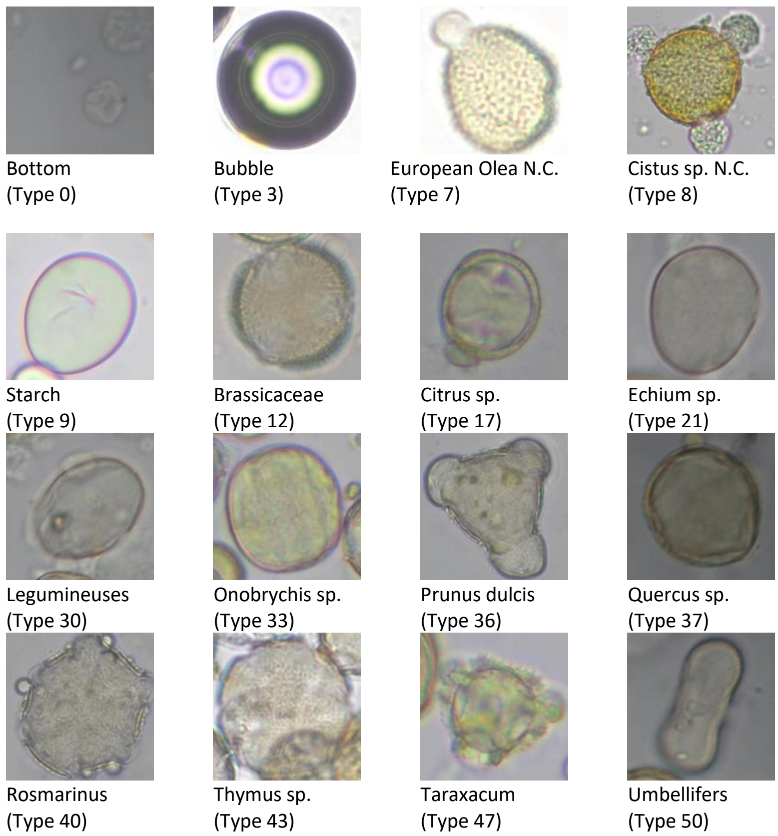

Fig. 1. Examples of the 16 types of studied Pollens.

Table 1. Number of samples per type.

Type	Number	Type	Number
Type 0	772	Type 30	112
Type 3	117	Type 33	306
Type 7	722	Type 36	179
Type 8	197	Type 37	1029
Type 9	233	Type 40	837
Type 12	3279	Type 43	599
Type 17	341	Type 47	70
Type 21	372	Type 50	818

2.2 ImageNet Networks

We used up to five pre-existing networks coming from the ImageNet Challenge: ResNet50, Xception, VGG19, VGG16, InceptionV3.

ResNet50 is a variant of ResNet model which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer [9]. The ResNet50 architecture contains the following elements: First of all, an input image of 224 x 224 target size. Behind, a convolution layer (size 64) with a stride of size 2, then a max pooling with also a stride size of 2. Subsequently, 3 convolution layers repeated 3 times (sizes 64, 64, 256 respectively). After this we could see 3 convolution layers repeated 4 times (sizes 128, 128, 512 respectively). Then other 3 convolution layers repeated 6 times (sizes 256, 256, 1024 respectively). After those 3 more convolution layers repeated 3 times again (sizes 512, 512, 2048 respectively). Finally, there are an average pool, then a fully connected layer and at the end a SoftMax function. ResNet50 introduces a new neural network layer, the residual block, whose aim is to address the degradation problem observed while training the networks [8]. It gives us a total of 50-layer Deep Convolutional Network.

We used the InceptionV3 network that comes from Google's Inception Convolutional Neural Network (as a third edition) [10]. It was introduced during the ImageNet Recognition Challenge. It was committed on allowing deeper networks while also keeping the number of parameters from growing too large. The InceptionV3 architecture contains the following elements: First of all, an input image of 299 x 299 target size. Behind, a convolution layer (size 32) with a stride of size 2, a convolution layer (size 32) and a convolution layer (size 64). Then a MaxPool layer with a stride of size 2. After that, a convolution layer (size 64) with a stride of size 2, a convolution layer (size 80) and a convolution layer (size 192). Then, the architecture has three inception modules placed. The first module carries out convolution on an input using filters (sizes 1×1, 3×3, and 3×3) followed by MaxPool (same for the others modules). The outputs are concatenated and go through to the following inception module. In the second module, a grid reduction technique is applied whose purpose is to diminish the number of parameters to become the model computationally less expensive. The process uses 1×n and n×1 convolutions instead of n×n convolutions. Last inception module takes after the second, it allows expanded the filter bank outputs to promote high dimensional representations.

Finally, we would observe other MaxPool layer with a stride of size 2. Then a fully connected layer and at the end a SoftMax function.

We have also used the Xception network which is another deep convolutional neural network architecture that involves depth wise Separable Convolutions[11]. It was developed by Google researchers. It proposes an advanced deep convolutional neural network architecture based on Inception network, where Inception modules have been replaced with deeper separable convolutions. The Xception architecture contains the following elements: First of all, an input image of 299 x 299 target size following by operations of batch normalization and ReLU. After this, the architecture has three blocks in sequence carrying out convolution, batch normalization, ReLU, and MaxPool operations.

VGG16 and VGG19 are convolutional neural networks models proposed by K. Simonyan and A. Zisserman from the University of Oxford in [12], which have 16 and 19 layers respectively. A crucial thing about VGG16 and VGG19 is that instead of having a large number of hyper-parameters they focused on using convolution layers of 3x3 filter with a stride 1 and always used the same padding and a MaxPool layer of 2x2 filter of stride 2. It follows this sequence of convolution and MaxPool layers consistently overall the architecture. In the end it has 2 Fully Connected layers followed by a SoftMax for output.

3 Experimental Work

All those networks were used with four different image datasets built randomly from the ground truth in a 4-fold manner (dataset1, dataset2, dataset3, dataset4), each one with images for training, validation and test distributed in 80% for training, 10% for validation and 10% for testing. Among the parameters to be highlighted we should mention the number of epochs that was 30 and the learning rate was 0.005. We trained all layers because the images of pollens are quite different to those images used in the ImageNet Challenge (dogs, cats, cars, houses, etc.) but used as the initial coefficients of the networks those coefficients obtained for the pre-existing networks on ImageNet. The results can be seen on Table 2.

Table 2. Accuracy results of the ImageNet Networks.

	Dataset1	Dataset2	Dataset3	Dataset4	Average
VGG16	97.33%	97.23%	98.12%	97.83%	97.63%
VGG19	96.84%	97.43%	97.83%	96.94%	97.26%
ResNet50	98.02%	97.23%	96.64%	97.73%	97.41%
Xception	97.63%	97.53%	98.32%	97.33%	97.70%
InceptionV3	97.92%	98.32%	98.42%	97.92%	**98.15%**

The best accuracy result was achieved by the InceptionV3 with an average of 98.15%. The rest of deep learning neural networks achieved also very good results. The ResNet50, Xception, VGG19 and VGG16 resulted in an average of 97.41%, 97.70%, 97.26% and 97.63%, respectively. The difference between networks is only in one point and as we can observe, the VGG16 network achieves a very good result with a difference with regards to InveptionV3 network of only 0.52 points and yet it is much simpler.

In the following figure we can see the Loss vs Accuracy of the dataset1 for the corresponding InceptionV3 ImageNet Network. We can appreciate that there is no overfitting in the training process since the curves of train accuracy and validation accuracy do not separate more than 15%. This also happens in the rest of the networks.

Fig. 2. Loss vs Accuracy for the InceptionV3 Network.

3.1 Results per Types and Multiclass Metrics

In this section we have computed the results of accuracy per types of pollens and networks, and we have computed several multi-class metrics in order to compare the goodness of the classification. The metrics we have used are the Precision, Recall, and F1-Score (English Wikipedia). The latter is the harmonic mean of Precision and Recall. In these cases, a value near to 1 means a good classification while a value near to 0 means a bad classification. We also have computed the multi-class version of Matthews Correlation Coefficient (MCC) (English Wikipedia), which is a metric with possible values between +1 and -1. A coefficient of +1 represents a perfect prediction, 0 no better than random prediction and −1 indicates total disagreement between prediction and observation.

We have computed these metrics and results per types for the 4-fold scheme we have followed in the experiments, that is, we have results for the dataset1, dataset2, dataset3 and dataset4. In Table 3 we show the results of the average of them. We can observe that the results of metrics are correlated with the accuracy obtained in each network, being the best the InceptionV3 network. With regards to the results of accuracy per type, in general, the best results are obtained by the InceptionV3 network. It should be noted that type 30 obtains low results in all networks and that it is the Xception that gives the best result for this type. This happens with other types and other networks, which are better in some cases than InceptionV3, but in average InceptionV3 responds better.

Table 3. Accuracy results per types and networks, total accuracy and multiclass metrics. Average.

	VGG16	VGG19	ResNet50	InceptionV3	Xception
type 0	99.68%	99.68%	99.37%	99.68%	100%
type 36	92.54%	92.55%	92.96%	98.75%	96.25%
type 50	97.63%	96.23%	98.80%	97.96%	95.66%
type 21	95.44%	95.55%	98.65%	96.82%	99.34%
type 37	98.12%	98.30%	96.93%	97.38%	97.38%
type 40	99.12%	97.13%	97.17%	98.26%	97.73%
type 9	100%	100%	99.00%	100%	99.00%
type 8	97.83%	98.81%	98.86%	98.86%	100%
type 47	100%	100%	100%	100%	96.88%
type 7	96.98%	96.28%	95.52%	97.60%	98.26%
type 43	95.93%	96.72%	98.75%	96.84%	95.54%
type 33	92.75%	93.93%	88.93%	96.24%	96.85%
type 3	100%	100%	98.21%	100%	100%
type 12	98.71%	98.86%	99.09%	99.16%	98.71%
type 30	84.44%	76.78%	78.44%	83.92%	88.02%
type 17	93.47%	89.27%	92.72%	94.99%	91.54%
Accuracy	0.9763	0.9726	0.9741	0.9815	0.9770
Precision	0.9750	0.9725	0.9750	0.9800	0.9775
Recall	0.9750	0.9725	0.9750	0.9800	0.9775
F1-score	0.9750	0.9725	0.9750	0.9800	0.9775
MCC	0.9669	0.9678	0.9641	0.9736	0.9701

4 Summary

We have studied the use of Convolutional Neural Networks to perform the classification of honey pollens, specifically the rosemary and citrus pollens and its companions, in total, 16 types or classes of pollens. We have used a ground truth of 9983 samples corresponding to these types of pollens. The number of samples vary from type to type between 70 and 3279.

We have used five pre-existing Networks coming from the ImageNet Challenge: VGG16, VGG19, ResNet50, Xception and InceptionV3. We trained all layers starting from the original coefficients of ImageNet. We followed a 4-fold scheme for training and classification and the best result of accuracy was achieved by the network InceptionV3 (98.15%), but the rest of networks obtained also good results. In fact, the VGG16 network with is significantly simpler that InceptionV3 is only 0.52 percentage points from the result of the InceptionV3.

We also studied the accuracy results per type of pollen and network. The best average result was achieved by InceptionV3 network, but in some types other networks performed better. Finally, we computed several multi-class metrics: Precision, Recall, F1-Score and MCC (Matthews Correlation Coefficient). We observed that the results of metrics were correlated with the accuracy achieved in each network, and the best was once again the InceptionV3 network.

Acknowledgment

The authors are grateful for financial support from the "Ministerio de Ciencia e Innovación. Agencia Estatal de Investigación" of Spain under the project PID2019-106800RB-I00 (2019).

References

1. Stillmana, E., Flenley, J. R. The needs and prospects for automation in palynology. Quat. Sci. Rev. 15, 1996, pp. 1-5.
2. Sevillano, V., Holt, K., Aznarte, J. L. Precise automatic classification of 46 different pollen types with convolutional neural networks. PLoS ONE 15(6). 2020.
3. Holt, K. A., Bennett, K. D. Principles and methods for automated palynology. New Phytologist. 203(3), 2014, pp. 735-742.
4. Sevillano, V., Aznarte, J. L. Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks. PLOS ONE. 13(9), 2018.
5. Daood, A., Ribeiro, E., Bush, M. Pollen Grain Recognition Using Deep Learning. In: Advances in Visual Computing. Lecture Notes in Computer Science. Springer. 2016. pp. 321–330.
6. Daood, A., Ribeiro, E., Bush, M. Sequential Recognition of Pollen Grain Z-Stacks by Combining CNN and RNN. In: The Thirty-First International Flairs Conference; 2018.
7. Khanzhina, N., Putin, E., Filchenkov, A., Zamyatina, E. Pollen grain recognition using convolutional neural network. In: 2018 proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018), Bruges (Belgium), April 2018, pp. 25-27.
8. Astolfi, G., Barbosa, A. , Vilharva, G., Brito, F. S., Melo, A. C., Astolfi, N., Takashi, E., Alvarez, M., Pistori, H. POLLEN73S: An image dataset for pollen grains classification. Ecological Informatics 60(2020) 101156.
9. Feng, V. An Overview of ResNet and its Variants. 2017. Available from: https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035
10. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. Rethinking the Inception Architecture for Computer Vision. 2015. Available from: https://arxiv.org/abs/1512.00567
11. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. 2016. Available from: https://arxiv.org/abs/1610.02357
12. Simonyan, K., Zisserman, A. University of Oxford: Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014. Available from: https://arxiv.org/abs/1409.1556