Fine-Grained Image Classification for Pollen Grain Microscope Images

Francesca Trenta[1][0000−0003−2524−3837], Alessandro Ortis[1][0000−0003−3461−4679], and Sebastiano Battiato[1][0000−0001−6127−2470]

IPLAB, University of Catania, Catania 95125, Italy
francesca.trenta@unict.it, ortis@dmi.unict.it, battiato@dmi.unict.it

Abstract. Pollen classification is an important task in many fields, including allergology, archaeobotany and biodiversity conservation. However, the visual classification of pollen grains is a major challenge due to the difficulty in identifying the subtle variations between the subcategories of objects. The pollen image analysis process is often time-consuming and require expert evaluations. Even simple tasks, such as image classification or segmentation requires significant efforts from experts in aerobiology. Hence, there is a strong need to develop automatic solutions for microscopy image analysis. These considerations underline the effort to study and develop new efficient algorithms. With the growing interest in Deep Learning (DL), much research efforts have been spent to the development of several approaches to accomplish this task. Hence, this study covers the application of effective Deep Learning methods in combination with Fine-Grained Visual Classification (FGVC) approaches, comparing them with other Deep Learning-based methods from the state-of-art. All experiments were conducted using the dataset Pollen13K, composed of more than 13,000 pollen objects subdivided in 4 classes. The results of experiments confirmed the effectiveness of our proposed pipeline that reached over 97% in terms of accuracy and F1-score.

Keywords: Pollen Classification · Fine-Grained Visualization · Machine Learning.

1 Introduction

With the rapid development of technologies in the field of Artificial Intelligence (AI), image data analysis has attracted much research attention over the last few years. In particular, typical problems in Computer Vision and Machine Learning field are related to image classification tasks. Indeed, image classification embraces several issues including discriminative feature extraction. The rapid emergence in developing such innovative pipeline to solve image classification has led to the spread of AI methods for extracting features from images. In this regard, Deep Learning (DL) approaches, provided a remarkable contribution. In fact, the main advantage of DL methods is the capacity to automatically

learn meaning features from high volume of data, rather than traditional ML solutions which involve the design of hand crafted features. The most valuable techniques include the use of neural networks such as AlexNet [13], ResNet [10], EfficientNet [17], which achieved effective results in a large variety of classification problems. In this work, we proposed an effective pipeline to perform image classification, making use of promising solutions which have reached state-of-art results in wide range of applications. Specifically, this study investigated the problem of classifying pollen grains having similar appearance. The dataset used for the experiments is Pollen13K[1] [2], composed of more than 13,000 pollen objects. In particular, the Pollen13K dataset includes 5 categories of objects. However, we considered the 4 classes and the train/test data splitting used during the International Pollen Grain Classification Challenge 2020. The dataset is publicly available, however, due to the nature of the competition, details about the employed methods are missing [3]. The classification of pollen objects has become a hot research topic in the field of aerobiology. Hence, the automation of pollen classification that could operate largely independently of a human operator would be of great benefit. Motivated by these considerations, we defined an innovative pipeline to improve pollen grains classification by using a Fine-Grained Visual Classification (FGVC) based approach [6]. The methods consists of a progressive training step and the application of a jigsaw patches generator in order to extract information from images at different granularity. We also implemented a Test-Time Augmentation (TTA) method to improve object classification predictions. The paper is organised as follows. In Section 2, we report the most interesting work regarding image classification, outlining the most promising approaches to solve this task. In Section 3, we report the pipeline used to classify pollen images, detailing the approaches for improving classification predictions. Section 4 details the experiments, giving an overview of the methods used, comparing them to other state-of-the-art methods. Section 5 reports the experiments details regarding other DL approaches used to perform a benchmarking evaluation. In Section 6, we discuss the results of the experiments. Finally, Section 7 outlines the conclusions.

2 Related Works

In recent years, there has been a growing interest in implementing effective methods for object classification. In this regard, we provide a brief survey of the recent advances in Deep Learning, outlining the most significant contributions to solving image classification, these approaches can be summarized into three categories, reported as follows.

Training Data Augmentation. A number of data augmentation strategies have been proposed over the years [15]. The most used techniques encompass the application of simple geometric transformations, such as horizontal flipping, color space augmentations, and random cropping, devoting to increasing the

[1] more details are available on the dataset website: https://iplab.dmi.unict.it/pollengraindataset/dataset

amount of training data for neural networks [13]. Recent works have shown that forming new artificial samples by combining two or more images from training data can lead to significant improvements in the performance of neural networks. Modern works include approaches such as MixUp [21], CutMix [20], and CutOcclusion [9]. In [21], the authors propose an innovative approach for creating a new example by performing a weighted linear interpolation of two existing images. In [20], the authors implemented a method to encourage the model to focus on less prominent parts of an image. The strategy is based on replacing a region from an image with patches from another one. The added patches further improve localization capability of the model by identifying the object considering a partial view. Penghui et al. [9] define a novel method for data training augmentation forcing the neural network to pay more attention to the surrounding area of a given object by including an occlusion region into an image. Specifically, the proposed approach was designed for the classification of pollen grains. In particular, the authors demonstrated that the most discriminating parts rely on the surrounding area of pollen object. Therefore, they introduced black patches around the center point of image in order to force the network to learn information from pollen wall and aperture area.

Fine-Grained Visual Classification. In recent years, the most promising solutions were devoted to the analysis of the granularity of images. Although neural networks such as AlexNet [13], ResNet [18], etc. have achieved remarkable results in image classification task, these models often fail to discriminate objects presenting a limited intra-class variation. For this reason, the key for improvement is represented by FGVC-based approaches. For example, Chen et al. [5], defined a method to "destruct" and "reconstruct" images. With regard to the "destruction" stage, the authors subdivided input images into k patches. Then, they shuffled them in order to create a new sample. For "construction", the authors implemented a region alignment mechanism to force the model to restore the spatial layout of image regions. The main advantage of this method is the capability of the model to pay more attention on local parts of the images than global features.

Tricks for improving classification predictions. One of the most used techniques for improving class predictions is Test Time Augmentation (TTA). This approach has been applied to several works including [16], where the authors proposed a test augmentation by applying horizontal flipping to input images, and [12], where the authors propose a test time augmentation method based on dynamically selecting transformations according to the loss function. Basically, the idea behind TTA method is based on performing a data augmentation on the test set in order to create different variants of the same image and perform the prediction on them. In general, a system of soft voting is implemented to determine which prediction is the most voted in order to assign a certain label. Several works propose the average of the resulting predictions or the sum of the probabilities to determine the confidence of the model.

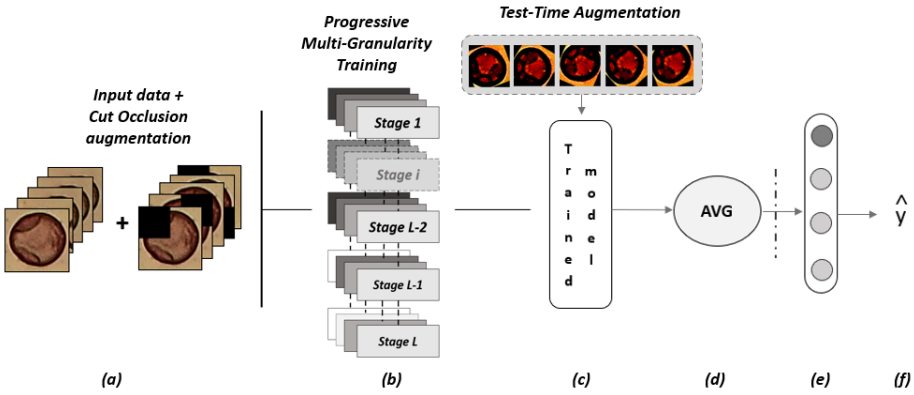

Fig. 1. The overall pipeline. (a) Input data consisting of pollen grains from Pollen13K and augmented dataset with Cut Occlusion. (b) Training performed using Progressive Multi-Granularity strategy. (c) Test-Time Augmentation. (d) Average calculation of predictions. (e) Max value of each predictions. (f) Predicted label.

3 Method and Materials

3.1 *CutOcclusion*: Training data augmentation

In [9], data augmentation is performed by operating Cut Occlusion strategy. As mentioned previously, the strategy was shown to be effective for pollen object classification. Inspired by the results of Penghui et al. [9], we reproduce the Cut Occlusion strategy in order to create new instances of the training data. The main advantage of this approach consists in avoiding some parts of the images by means of black patches in order to help the model to extract discriminative features from pollen wall. This strategy can bring substantial improvements for the pollen classification task, where extracting discriminative features from aperture area of pollen grain instance seems to be more important than concentrating on the center area of the pollen object.

3.2 Pipeline

In Fig. 1, the full pipeline is depicted. The architecture design was firstly introduced in [6], which tackles the image classification task by introducing a novel approach based on Progressive Multi-Granularity training strategy (PMG). As discussed, pollen objects, from the Pollen13K dataset, belonging to different classes, present a similar appearance. In addition, objects in the same category could report a varying appearance. Therefore, applying a fine-grained visual classification approach, taking advantage of local features information, could lead to remarkable improvements. The framework consists of two main components: (a) a progressive training method to add new layers during training process in

order to extract discriminative features from images with different granularities. Hence, the process starts at low stage and progressively include new layers. (b) a jigsaw patches generator [19] to capture local information from images. In [6], the authors used ResNet50 as backbone. In our study, we use ResNet101 [10] as feature extractor. For each layer $\mathcal{L}$ of the feature extractor, a new convolution layer is added taking as input the feature maps from the output of intermediate layers that is transformed into a vector representation. Then, classification modules are added to calculate the probability distribution between classes. Finally, the outputs of the last levels are concatenated.

Progressive Training. This technique allows to train the model starting from the low stage and then adding new layers. The advantage of this technique consists in forcing the model to learn discriminative information from local details rather than focusing on global information. The loss cross entropy function $\mathcal{L}_{CE}$ is applied to the output of each stage and the output of the concatenated features.

jigsaw puzzle generator. To train the PMG model, we define a set of jigsaw puzzle permutations. This approach has been widely employed to find the multi-granularities of the images during training stage. Given an image x, it can be subdivided into k patches. Then, the patches are shuffled randomly and merged together into a new image x'.

3.3 Test Time Augmentation.

In order to boost the prediction accuracy, we implemented a strategy called Test Time Augmentation (TTA), which consists of applying data augmentation techniques to the test set in order to improve the prediction of a given class of objects. For this reason, we create several variants of the same image, applying a horizontal or vertical flip, standard color augmentation, or other geometric transformations. We computed a prediction for each of these images. Then, we average these predictions and calculate the max value in order to obtain which prediction has the highest confidence score. Finally, we computed the predicted class for the analysed object. By applying this strategy, we avoid the uncertain of the model by averaging the predictions and averaging the error. In this context, we created 5 different variants for each single image of the test set by applying a horizontal flip, a vertical flip and a random rotation. The rotation angle ranges from -90° to 90°.In addition, as in the training set, we performed an image resize of 550×550 pixels and a centre crop of 448×448 pixels. Moreover we normalized data setting with a mean and a standard deviation of 0.5.

3.4 Dataset

The dataset Pollen13K [2] is composed of $13,416$ objects divided into 5 categories, respectively: *Coryllus Avellana* (well-developed), *Coryllus Avellana* (anomalous), Alnus, *Cuprissaceae*, and Debris. However, considering the small number of observations related to *Cupressaceae* class (43), we did not include them in

the dataset used for the experiments. Hence, the dataset is composed of images depicting one pollen type among the 4 mentioned categories, these patches have been manually labelled by experts in the field of aerobiology. The dataset includes: (1) 84×84 RGB images for each segmented object, for each of the four categories; (2) binary masks for single object segmentation (84×84 resolution); (3) segmented versions of the patches obtained by applying the segmentation mask and padding the background with all green pixels (84×84 resolution).

4 Experiments

To evaluate the performance of the proposed pipeline, rigorous experiments are performed on two image datasets: Pollen13K [2] and Augmented Pollen13K. We implemented the proposed approaches as well as several state-of-the-art approaches on these datasets. The experimental settings are given in the following subsections.

Proposed pipeline: CutOcclusion + PMG + TTA. In order to boost the predictions, we performed dataset augmentation by using Cut Occlusion strategy [9]. We inserted occlusions around the center of pollen area. In particular, we created 4 different variants for each image of the training set. With regard to PMG method [6], all experiments were conducted using PyTorch [14] over a cluster of GPU NVIDIA® T4. We employed ResNet101 [10] as backbone. All setting are indicated in [6], where $S = 3$, $\alpha = 1$, and $\beta = 2$. In addition, the input images were resized to 550×550 pixels and randomly cropped by 448×448 pixels. A random horizontal flipping is applied for data augmentation for training data. We use Stochastic Gradient Descent (SGD) [11] optimizer and batch normalization as the regularizer. We train the model for 100 epochs. The batch size was set to 16. Moreover, we used a weight decay of 0.0005 and a momentum of 0.9. Finally, we performed TTA algorithm to improve the performance of our proposed pipeline.

Other approaches. We use the Pytorch [14] Deep Learning library for performing the experiments related to other advanced DL networks: ResNet101 and Residual Attention Network (ResAttNet). We resized input images by 256×256 pixels. A 224×224 center crop is sampled from an augment image, applying geometric transformations. The network is trained using Stochastic Gradient Descent (SGD) with a momentum of 0.9. We set initial learning rate to 0.0001, decaying learning rate by a factor of 0.1 every 7 epochs. We set the number of epochs to 100. All the experiments use a batch size of 16. With regard to the methods based on CutMix strategy [20], we set hyperparameters values to $\beta = 1.0$ and *cutmix probability* to 0.5.

WRS method. In this study, we evaluate the performance of the aforementioned algorithms using the Pollen13K dataset [2]. To the best of our knowledge, it represents the public dataset with the largest number of pollen objects, with more than $13,000$ objects. However, the Pollen13K dataset consists of imbalanced classes since the largest class consists of the objects from class Alnus

Method	Accuracy	F1 (weighted)	F1 (macro)
WRS + ResNet101	92.667 %	92.899 %	88.233 %
WRS + ResAttNet	83.776 %	84.497 %	77.627 %
WRS + CutMix + ResNet101	93.922 %	94.046 %	90.097 %
PMG [6]	**96.384 %**	**96.349 %**	**93.585 %**
Method + TTA	Accuracy	F1 (weighted)	F1 (macro)
WRS + ResNet101 + TTA	94.626 %	94.702 %	91.266 %
WRS + ResAttNet + TTA	83.626 %	84.748 %	80.416 %
WRS + CutMix + ResNet101 + TTA	95.228 %	95.280 %	92.581 %
CutOcclusion + PMG + TTA (Proposed)	**97.087 %**	**97.050 %**	**94.726 %**

Table 1. Comparison between DL approaches. On the top part of the table, we reported evaluation results without applying TTA. On the bottom part of the table, we reported results by applying TTA method.

(8, 216 objects in the train set). Other classes include a total number of objects less than 1, 600. Motivated by these issues, we provided an effective solution by implementing the Weighted Random Sampler (WRS) function to deal with imbalanced dataset and preventing overfitting problems. Hence, one of the proposed solution is to oversample minority classes [4]. By applying this technique, we balanced batches of data. As a result, during training time, the model will not concentrate significantly on one class over another and risks of overfitting are reduced. Basically, the WRS method uses the array of weights which corresponds to weights given to each class. The goal is to assign a higher weight to the minor class, providing a more robust classification. Finally, we evaluate the performance of each classifier by also using the weighted and macro F1 score, which represent two more reliable performance metrics than accuracy. The weighted F1 score function calculates the F1 metrics for each class, and their average weighted by support (i.e., the number of true instances for each class). The F1 macro score computes the F1 for each label and returns the average.

5 Results and Discussion

This section presents the results obtained for a pool of Deep Learning algorithms, providing also a benchmarking evaluation of the performance of the techniques herein proposed for pollen grains classification. In Table 1, results show that PMG [6] method lead to a boost of the prediction accuracy than other methods. The main advantage of this approach include the analysis of images with different granularities. Basically, it forces each stage of the network to focus on local features rather than concentrating on global information. Furthermore, a jigsaw generator perform an image splitting into several patches during the training phase, providing discriminative information at the specific granularity level. Although the method based on Residual Attention Network (ResAttNet) produces remarkable results, it fails to outperform the other proposed methods. In both cases, we observed that the CutMix-based approach leads to better classification results than the pre-trained model (ResNet101) without using this strategy as

data augmentation. According to these results, the method that yields to good results, both in terms of accuracy and F1 scores, is the approach based on progressive training and the use of jigsaw generator, i.e., PMG. With the attempt to further improve performance of the PMG model, we defined a data augmentation technique, based on Cut Occlusion, and a Test Time Augmentation to achieve higher accuracy during inference. The proposed pipeline yield to better results than other methods, confirming the effectiveness of the proposed framework. This strategy tends to improve the classification results and obtain results consistent with state-of-the-art. We reported the achieved results in Table 1. As observed, the TTA strategy provides reliable results in terms of accuracy and F1-score (weighted and macro) than other methods where this strategy was not applied to. With regard to ResAttNet, the accuracy value is decreased compared to the value from previous experiment. Instead, the F1-score metrics improve their value. In general, we observe that Deep Learning methods in combination with TTA strategy yields to better results than methods not using Test Time Augmentation strategy.

Misclassification: In Fig. 2, we report some examples of misclassification obtained by the standard PMG [6] algorithm without using training augmentation and TTA approach. As observed, the standard PMG approach [6] misclassifies objects of class 1 with objects of class 3 and vice versa. In fact, the object from these classes present similar characteristics, leading to a challenging image classification task. Furthermore, objects belonging to class 4 are indicated as class 3 objects. Probably, it depends on the presence of other objects within the image which fools the model, forcing it to extract their features and leading to a misclassification. However, the implementation of Cut Occlusion strategy allows us to avoid these objects, encouraging the network to focus on the object depicted at the center of the image, reporting better classification predictions.

Comparisons with previous studies. We also reported comparison between our proposed approach and previous studies for the classification of pollen grains. Fang et al. [7] propose a blending strategy consisting of a Destruction and Construction Learning architecture [5] and DenseNAS [8] output vectors to be used as the input of a Random Forest Classifier, which performs the final classification. Penghui Gui et al. [9] generated a number of images by applying the cut occlusion approach. The trained model is based on ResNet101. In our previous study [1], we investigated the performance of several Machine Learning approaches, such as AlexNet, SmallerVGGNet, etc. Fang et al. [7] leads to the best results in terms of accuracy (97.539) and F1-score (97.510), whereas Penghui Gui et al. achieved an accuracy of 97.290 and an F1-score of 97.260. Our previous method [1] achieved an accuracy of 89.730 % and an F1-score of 89.140%. In our experiments by using the proposed pipeline, we achieved an accuracy of 97.087 % and an F1-score (weighted) of 97.050 %, providing results similar to [7] and [9]. We also performed cross-validation, achieving good results in terms of accuracy (96.5%) and F1-score (96%). Although the experiments suggest that the algorithm we proposed provides better results, an accurate validation was

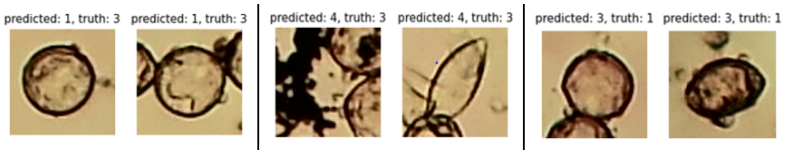

Fig. 2. Example of bad classification performed by PMG. These objects are classified accurately by PMG with training augmentation and TTA.

not performed. On the contrary, our 3-fold cross-validation method has proved to provide more robust results.

6 Conclusions

In this paper, we tackled the problem of the classification of pollen grains by designing an innovative pipeline, consisting of a progressive training strategy and a jigsaw generator to extract information about image granularity. We further applied a data augmentation method to input images of the training set by forming 4 variants of each image. In addition, a Test Time Augmentation (TTA) method was implemented by applying simple geometric transformations to test images, creating 5 variants of each image in order to provide better classification results. The results show that our proposed pipeline has obtained a robust and consistent results with respect to state-of-the-art methods for image classification. One explanation of the performance improvement of the pipeline could depend on the dataset augmentation for both train and test test which generally leads to a better generalization of the model, improving also its performance.

7 Future Works

As future development, we plan to collect more data with the aim of improving the effectiveness of the proposed approach. Specifically, we will address further application in the aerobiology field with special focus to designing more advanced DL pipelines to perform pollen object classification.

References

1. Battiato, S., et al.: Detection and classification of pollen grain microscope images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 980–981 (2020)
2. Battiato, S., et al.: Pollen13k: A large scale microscope pollen grain image dataset. In: 2020 IEEE International Conference on Image Processing (ICIP). pp. 2456–2460. IEEE (2020)

3. Battiato, S., et al.: Pollen grain classification challenge 2020. In: Pattern Recognition. ICPR International Workshops and Challenges. pp. 469–479. Springer International Publishing, Cham (2021)
4. Buda, M., et al.: A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks **106**, 249–259 (2018)
5. Chen, Y., et al.: Destruction and construction learning for fine-grained image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5157–5166 (2019)
6. Du, R., et al.: Fine-grained visual classification via progressive multi-granularity training of jigsaw patches. In: European Conference on Computer Vision. pp. 153–168. Springer (2020)
7. Fang, C., et al.: The fusion of neural architecture search and destruction and construction learning. In: Pattern Recognition. ICPR International Workshops and Challenges. pp. 480–489. Springer International Publishing, Cham (2021)
8. Fang, J., et al.: Densely connected search space for more flexible neural architecture search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10628–10637 (2020)
9. Gui, P., et al.: Improved data augmentation of deep convolutional neural network for pollen grains classification. In: Pattern Recognition. ICPR International Workshops and Challenges. pp. 490–500. Springer International Publishing, Cham (2021)
10. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
11. Kiefer, J., et al.: Stochastic estimation of the maximum of a regression function. The Annals of Mathematical Statistics **23**(3), 462–466 (1952)
12. Kim, I., et al.: Learning loss for test-time augmentation. arXiv preprint arXiv:2010.11422 (2020)
13. Krizhevsky, A., et al.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems **25**, 1097–1105 (2012)
14. Paszke, A., et al.: Automatic differentiation in pytorch (2017)
15. Simard, P.Y., et al.: Transformation invariance in pattern recognition—tangent distance and tangent propagation. In: Neural networks: tricks of the trade, pp. 239–274. Springer (1998)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
17. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. pp. 6105–6114. PMLR (2019)
18. Wang, F., et al.: Residual attention network for image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3156–3164 (2017)
19. Wei, C., et al.: Iterative reorganization with weak spatial constraints: Solving arbitrary jigsaw puzzles for unsupervised representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1910–1919 (2019)
20. Yun, S., et al.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6023–6032 (2019)
21. Zhang, H., et al.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017)