Metric-Based Few-Shot Learning for Pollen Grain Image Classification

Philipp Viertel[a], Matthias König and Jan Rexilius
Campus Minden, Bielefeld University of Applied Sciences, Artilleriestraße 9, Minden, Germany

Keywords: Deep Learning, Few Shot Learning, Computer Vision, Palynology, Pollen Analysis, Image Classification.

Abstract: Pollen is an important substance produced by seed plants. They contain the male gametes which are necessary for fertilization and the reproduction of flowering plants. The scientific study of pollen, palynology, plays a crucial role in a number of disciplines, such as allergology, ecology, forensics, as well as food-production. Current trends in climate research indicate an increasing importance of palynology, partly due to a projected rise in allergies. Pollen detection and classification in microscopic images via deep neural networks has been studied and researched, however, pollen data is often sparse or imbalanced, especially when compared to the number of plant species, which is estimated to be between 330,000 and 450,000, of which only a small percentage is investigated. In this work, we present a solution that does not require a large number of data samples by employing Few-Shot Learning. Our work shows, that by utilizing Prototypical Networks, an average classification accuracy of 90% can be achieved on state-of-the-art pollen data sets. The results can be further improved by fine-tuning the net, achieving up to 98% accuracy on novel classes. To our best knowledge, this is the first attempt at applying Few-Shot Learning in the field of pollen analysis.

1 INTRODUCTION

Palynology is the scientific analysis of pollen grains, which consists of classifying, analyzing, and counting pollen grains to establish their taxonomy. Such tasks are necessary for various disciplines and applications such as medicine, food safety, forensics, botany, and paleoecology. The information that is derived from a pollen analysis can indicate the geographical origin of the sample, the plant family, the grains health status (e.g. abnormal or normal), age, as well as the effects of climate change. Especially the latter is possible due to the resistant hull of pollen, the sporoderm. The need and advantages for an automated solution, such as time, costs, and workload, have been already established in 1996 (Stillman and Flenley, 1996). Typically, an in-depth analysis requires a laboratory environment (e.g. to create a sediment or prepare the pollen sample in other ways) and highly trained palynologists to identify the grains typically via a Light-Microscope (LM). In melissopalynology, which is the scientific study of pollen in honey, the pollen grains have to be classified and counted in order to label the product correctly and to provide allergy-related information. The morphological features are the main distinction between different plants, however, certain pollen grains can have highly similar visual characteristics where a clear identification requires years-long human experience as well as additional extrinsic knowledge, such as season, origin of the sample, and visual reference material.

The application of Deep Learning (DL) methods has made large advantages in recent years and can offer as of 2022 typical solutions to many Computer Vision (CV) problems, such as object detection, classification, and segmentation. Well-established Deep Neural Network (DNN) architectures, such as ResNet (He et al., 2016) and VGG-16 (Simonyan and Zisserman, 2015), however, require a large number of training images, typically provided by data sets such as MSCOCO (Lin et al., 2014) and especially ImageNet (Russakovsky et al., 2015), which contains more than 14,000,000 images and approximately 20,000 classes. With regards to palynology, the availability of large high quality pollen data sets is sparse. Most research work is performed on proprietary data that is not accessible and/or have issues regarding class bias, balance, or overall quality, due to different methods of image acquisition. Only in recent years, a number of quality pollen data sets have been published ((Battiato et al., 2020) (Gonçalves et al., 2016) (Sevillano et al., 2020) (Tsiknakis et al., 2021)) and are publicly accessible. However, the usability in real-world scenarios is often limited, due to the flora of the pollen which is geographically limited and non-uniform im-

[a] https://orcid.org/0000-0002-7274-4290

Viertel, P., König, M. and Rexilius, J.
Metric-Based Few-Shot Learning for Pollen Grain Image Classification.
DOI: 10.5220/0011727900003411
In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, pages 418-425
ISBN: 978-989-758-626-2; ISSN: 2184-4313

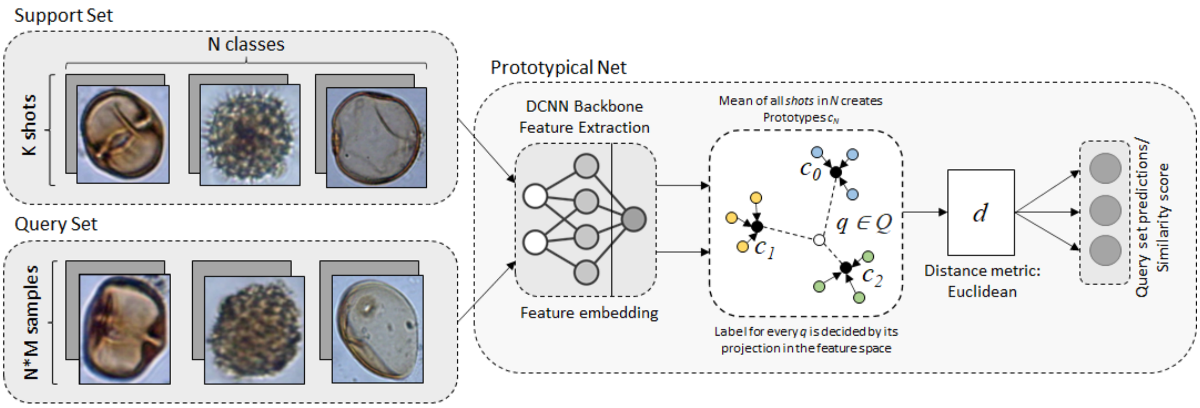

Figure 1: Overview of our employed method with a typical FSL task split, comprised of a support and query set. The Feature extractor has the last layer removed, so that it produces a one-dimensional embedding which is then fed into the PN algorithm. Predictions for each image in the query set are generated by comparing each feature vector to the class prototypes c, which are based on the mean of all k shots from N-classes. The distances are measured via the Euclidean distance.

age acquisition methods remain an obstacle. For future prospects, it is important to notice, that from around 300,000 to 450,000 plant species, only about 10% have been investigated (Scotland and Wortley, 2003).

The problem of data scarcity can be approached in a multitude of ways: data augmentation, e.g. the generation of synthetic pollen images with Generative Adversarial Networks (GAN) (Viertel et al., 2021), handmade feature engineering strategies to define the morphological, textural, and color-based traits, as well as Domain Adaptation (DA) techniques. Another method to handle sparse data is Few-Shot Learning (FSL). It differs from the well-established deep Convolutional Neural Network (CNN) approach by requiring only one to five images per class (or even none in Zero-Shot Learning). While traditional object classification methods produce a classification score in the last layer, FSL methods usually utilize the feature extraction method but rely on other algorithms to actually predict a class label, e.g. metric-learning-based methods, which create an embedding, i.e. feature vector, for each image and calculates the distance in the feature space via a specific distance measure. During inference, these embeddings are compared and the closest match yields the corresponding class label. The feature extractor relies on established CNN architectures, such as the ones mentioned above.

In this work, we will utilize a metric-based approach, i.e. Prototypical Networks (Snell et al., 2017), to evaluate the possibility of classifying pollen grain images from three state-of-the-art data sets. We will show, that even without out-of-domain knowledge, a baseline model performs around 90% on 5-way 5-shot tasks and the results can be further increased via Fine-Tuning. Although a metric-based algorithm is not the only method, in this work, we chose this approach, due to its well-established success in application and inductive approach. Varying methods for tackling FSL problems are discussed in Section 2. The method selected for this work is elaborated on in-depth in Section 3, together with the utilized data and backbone model as well as the results of our experiments. Finally, in Section 4 we will summarize our findings and discuss the applicability of our method, its drawbacks, and how they could be addressed in future works. We believe, that this work is an important step in realizing a feasible solution to an automated pollen classification system, without requiring large quantities of labelled pollen grain images. To the best of the authors knowledge, this is the first attempt at utilizing FSL in the domain of pollen analysis.

2 RELATED WORK

Although automated pollen classification is an established research field, with solutions ranging from Feature Engineering-based Machine Learning (ML) to deep neural network applications (Viertel and König, 2022), the crucial problem of data sparsity and the prospect of constantly adding new specimen requires particular attention. FSL is a paradigm suitable for problems where large numbers of samples are difficult or impossible to acquire. Early proposals were made in the 2000s, based on Bayesian networks (Fei-Fei et al., 2003) (Fei-Fei et al., 2006), utilizing only one to five training samples. In general, the methodological research can be categorized in three groups: metric-based, model-based, and optimization-based methods. Unlike traditional CNN methods, FSL methods usually split the data for a task into a support (consisting of K images for N classes) and query set (unlabelled images, that are compared to the support set). The **metric-based** approach aims at comparing two samples in a latent (or metric) space. The idea is, that samples with the same label are closer to each other than samples with differing labels. An early representative of this idea was published in 2015, with (pseudo)-siamese networks (Koch et al., 2015) (Zagoruyko and Komodakis, 2015) for one-shot tasks, trained, and evaluated on the Omniglot data set (Lake et al., 2015). The output of two identical networks are jointly trained with a relationship function. It produces the probability of two images belonging to the same class. During inference, the input is compared to all examples from the support set, i.e. every single image for each class candidate. It does that by encoding the images into feature vectors (embeddings). The generated embeddings are compared pair-wise, i.e. the L1 distance is calculated and consequently transformed into a probability. Matching Networks (MN) (Vinyals et al., 2016) work similarly by creating an embedding for each image. The embeddings are typically generated by a CNN, the feature extractor. MNs use the Cosine distance between the embeddings to calculate the similarity relationship in the embedding space. (Vinyals et al., 2016) also introduced the miniImageNet data set, which is a commonly used data set for benchmarking the performance of new FSL methods. It is smaller than ImageNet but more complex than CIFAR10, and therefore, fit for rapid experimentation. Prototypical Networks (PN) (Snell et al., 2017) extend the idea of MNs with two major changes: creating mean label-based embeddings, i.e. a prototype feature vector for every class c by using the mean vector of the embedded samples in c and utilizing Euclidean distance instead of Cosine, which increased the accuracy on miniImageNet benchmarks. (Sung et al., 2018) proposes the addition of a relation module, therefore called Relation Network (RN). The feature extractor does not generate one-dimensional vectors but instead feature maps, which are concatenated and fed into the relation module to produce classification results. However, it is regression-based contrary to PNs and MNs. **Optimization-based methods** are represented by the Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017) algorithm and its variations. MAML is model and task-agnostic and aims at training a model (via meta-learning) by a number of gradient descent updates to adapt quickly in learning new tasks. Such algorithms are referred to as gradient-based and model-agnostic, since it puts no constraints on the choice of the model architecture except to be optimizable by a gradient-based optimizer. Further improvements and alternative approaches have been introduced since, such as Reptile (shortest descent) (Nichol et al., 2018) and iMAML (MAML with implicit gradients) (Rajeswaran et al., 2019), where each method aims at improving the core idea by modifying the gradient descent steps. **Model-based approaches** encompass CNN architectures that are primarily designed for the paradigm of FSL, such as (Santoro et al., 2016). FSL methods are usually inductive, i.e. the goal is to learn a generalized model that can be applied onto unseen data and predict the correct labels. In FSL, that is typically done by observing one instance after the other from the query set, transductive methods, however, observe the query set in its entirety in one episode. The data is therefore not split into training, validation, or testing, instead the model uses labelled and unlabelled data together (Zhu and Ghahramani, 2002). It uses a small subset of labelled data points to propagate them to the unlabelled points. However, this implies that such models cannot be used to predict new data and new input requires a re-training of the entire model.

Since this is the first work exploring FSL in pollen analysis, practical research work is limited to familiar fields in micro and cellular biology, specifically histopathology. (Medela et al., 2019) uses a Deep Siamese Neural Network with a VGG16 backbone. The authors fine-tuned it via a data set of colon tissue images and evaluate it on a smaller set of colon, lung, and breast tissue images. A balanced accuracy of 90% is achieved and outperforms the transfer learning approach that obtains only 73%. While the FSL approach required only 20 samples per class, the transfer learning approach would require 600 images (130 samples per class) to reach 81%. (Li et al., 2021) propose a two-stage deep adaptive few example learning network for cell counting. At first, a pre-trained regression network is fine-tuned with a small set of novel medical images. In the second stage, an attention module is used to correlate features and their bounding boxes by exposing the model during testing to cell examples. Evaluation on three cell data sets shows that the proposed method outperforms baseline approaches. Anomaly or artifact detection in histopathological images via PNs is successfully evaluated in (Shaikh et al., 2022), again showing that it outperforms standard transfer learning approaches. (Walsh et al., 2022) investigates the possibility of utilizing FSL for automated human cell classification and utilize best-practice methods. The domain and its inherent issues described are similar to the ones encountered in palynology; the dependence on professionals to classify biological material and the sparsity of data. The authors trained nine different FSL methods on miniImageNet and evaluate the performance on two human cell data sets to identify best method candidates. Additionally, the authors varied the backbone architectures and training schemes to evaluate potential performance benefits. For all experiments, the authors used 5-way 5-shot sampling. The two top-performing methods were Reptile and EPNet (Rodríguez et al., 2020), achieving approx. 40% and 45% accuracy, respectively. EPNets backbone (WideResNet) was changed to EfficientNetV2 (Tan and Le, 2021), ResNet-18, and DenseNet (Iandola et al., 2014). However, none of these changes improved the accuracy compared to EPNets original backbone. The results indicate that a high performance on miniImageNet does not guarantee a similar performance on out-of-domain data.

3 METHOD

3.1 Data

Three data sets were chosen for evaluation and training: NzPollen (Sevillano et al., 2020), POLEN23E (Gonçalves et al., 2016), and CPD-1 (Tsiknakis et al., 2021). All of these data sets contain a large number of classes, i.e. at least 20 different pollen classes. Figure 2 shows examples from each data set. Although POLEN23E contains only 805 images, it has a balanced distribution of images (35 w/ 23 classes). Its pollen are from the Brazilian Savannah and the images were captured via a digital Bresser LCD microscope at 40 × magnification and subsequently segmented. The pollen in CPD-1, collected in Crete, Greece, were dyed with fuchsin to increase the visibility of textural features and captured at 400 × magnification[1]. The pollen were recorded in a dry state, which is important to notice, due to the harmomegathic effect of pollen grains; depending on their state, dry or hydrated, the morphology, and therefore the visual characteristics, can change completely. This can make pollen from the same class, depending on their state, incomparable. The set contains 20 classes, with 4,025 images and 22 to approx. 700 per class. The images in NzPollen, pollen from plants in New Zealand and the Pacific region, were captured with a dark-field microscope (DFM). This allows for the creation of high-contrast images of translucent samples and thus achieving the same effect as with the application of dye. DFMs utilize a condenser lens, which redirects the light away from the objective lens. As CPD-1, NzPollen is also imbalanced. It contains 19,667 images with 46 classes. Each class is made up of 45 to approx. 1,500 images. The information is summed up in Table 1.

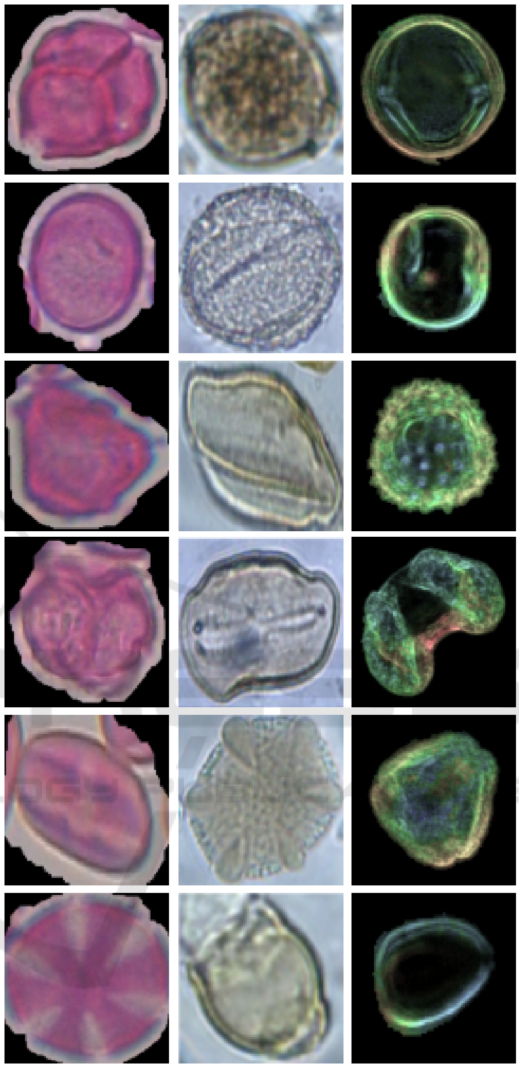

Figure 2: Various pollen grain samples from all three data sets. From left to right: CPD-1, POLEN23E, and NzPollen.

3.2 Prototypical Networks

Metric-based FSL tasks can be described as N-way K-shot classification tasks. N indicates the number of classes in the support set S and K the number of samples (shots) per class k, with K being typically $<= 10$. In addition to the support set, FSL tasks require a query set Q. Q contains a number of unlabelled images, for which the model has to predict the correct label by comparing the class prototypes, calculated from S, with each feature vector in Q. A PN is typically trained via the Meta-Learning paradigm

[1] 400 × is a typical magnification strength. It is in the recommended range set by the German DIN NORM 10760 to analyse and count pollen grains in a honey sediment analysis for producing a correct label.

Table 1: Composition of the three data sets used in this work in comparison. The NzPollen set was recorded via a DFM, the others with a LM. The CPD-1 set is colored with Fuchsin, which is a common dye used in palynology that colors the pollen grains pink and highlights the textural features. DFM images tend to provide the same effect.

Data set	**Classes**	**Total number of imgs.**	**Imgs. p. class**
NzPollen	46	19,667	45 to ~1,500
POLEN23E	23	805	35
CPD-1	20	4,025	22 to ~700

which consists of a fixed number of episodes, where in each training iteration N classes with K labels are randomly selected and serve as the support set S, which can be defined as: $S = \{(x_i, y_i), ..., (x_N, y_N)\}$, where x_i, y_i are the input-output pairs. $x_i \in \mathbb{R}^D$ is the D-dimensional feature vector, with $y_i \in \{1, ..., K\}$ being the corresponding label. Q is defined analogous: $Q = \{(x_j, y_j)\}$. This method simulates the known training/test split. The goal of Meta-Learning is to optimize the net, i.e. the creation of fitting embeddings. This can be done by replacing the Fully-Connected output layer with a Flatten layer to reshape the output into a one-dimensional vector encoding. This output is fed into the PN which creates the class prototypes c_k, i.e. the feature vectors by averaging the embeddings from all images (x_i) of the specific class. We can define this as S_k, the set of support images with label k, then c_k is defined as:

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\phi(x_i) \tag{1}$$

with $f_\phi : \mathbb{R}^D \to \mathbb{R}^M$ being the embedding function, with the learnable parameter ϕ, that computes the M-dimensional representation of each class. The embedding for each Q image is compared to the class prototypes via a Euclidean distance measure d. The largest probability indicates the predicted label for $\hat{y}$. This can be defined as:

$$p(\hat{y} = k \mid \hat{x}, S) = \frac{\exp\left[-d(f_\phi(\hat{x}), c_k)\right]}{\sum_{k'} \exp\left[-d(f_\phi(\hat{x}), c_{k'})\right]} \tag{2}$$

To get the class predictions, a softmax is performed on the computed distances. The shortest distance to a prototype c indicates the sample belonging to that specific class. The original learning phase consists of a log-softmax loss $J(\phi) = -log(p_\phi(y = k|x)))$ of the true class k via stochastic gradient descent. The architecture of this method is displayed in Figure 1.

In order to optimize the method, we evaluated a number of CNNs to improve the accuracy. State-of-the-art pretrained models on ImageNet, without additional out-of-domain training, were evaluated on all three data sets. We selected a number of nets that performed well in related use cases (see Section 2), as well as different sizes of ResNet to validate the perception that deep nets perform worse in FSL. The complete results are shown in Table 2. First, we performed the experiments as 5-way 5-shot tasks for 1,000 sampled tasks. We assume that the candidates perform similarly for larger Ns. Although most FSL methods utilize miniImageNet as well as small CNN architectures, our approach differs in both points: the image size in miniImageNet is 84×84, while e.g. CPD-v1 and POLEN23E have a mean image size of

Table 2: Comparison between different feature extractors. Each CNN backbone+PN is pretrained on ImageNet and evaluated without out-of-domain training on all three data sets. There is no network that performed best across all data sets.

Feature Extractor	**5-way 5-shot (Acc.%)**		
	Nz-Pollen	**POLEN 23E**	**CPD-1**
ResNet-18	93.5	90.4	78.2
ResNet-34	93	92	80
ResNet-50	93.2	92.1	81
ResNet-101	91	90	84
EfficientNetV2s	86.1	77	80.2
DenseNet-121	94.4	89.7	81.3
WideResNet-50	93.9	92	82.2
SqueezeNet 1.1	90.3	79.1	63.2
ConvNeXt tiny	93.3	90.4	83.3
GoogLeNet	95.4	90.1	80.5

Table 3: Performance based on miniImageNet compared to ImageNet. Each trained model was evaluated without out-of-domain knowledge on all three data sets. Preprocessing was applied to all data sets accordingly. The model trained on miniImageNet performs significantly worse than the one on ImageNet.

Base training w/ ResNet-18	**5-way 5-shot (Acc.%)**		
	Nz-Pollen	**POLEN 23E**	**CPD-1**
ImageNet	93.5	90.4	78.2
miniImageNet	76.2	79	60

Table 4: Baseline performance (w/o Fine-Tuning) of WideResNet-50+PN with different task sizes. A larger N decreases the performance, due to more prototypes in the feature space.

	Nz-Pollen	**POLEN 23E**	**CPD-1**
10-way 5-shot (Acc.%)	89.5	86	74
20-way 5-shot (Acc.%)	83.4	78.55	66.2

132×132 and 285×278, respectively. Therefore, our preprocessing and model evaluation is based on ImageNet, i.e. 224×224. In the worst case, up-scaling can increase the amount of redundant information, however, down-scaling can possibly lead to information loss. In the case of pollen, visual distinctions can be subtle and their classification can benefit from any visual information available, as shown in conventional pollen object classification (Sevillano et al., 2020). To support this decision, we trained a PN with ResNet-18 on miniImageNet. We used 64 classes for training, 16 for validation, and 20 for testing (each class contains 600 images; 60,000 images in total). We performed episodic training for 200 epochs, with 500 tasks per epoch consisting of 5-way 5-shot tasks and $Q = 10$. For testing we used 1,000 tasks with the same shape. We achieved an accuracy of 68% on the test set after 130 epochs. We used a learning rate of 0.01 and a momentum of 0.9. After 100 epochs we reduced the learning rate by a factor of 10. However, our results show that it performs significantly worse than a pre-trained model on ImageNet: On NzPollen, POLEN23E, and CPD-1 it yielded an accuracy of 76.2% (-17.3) , 79% (-11.4), and 60% (-18.2), as shown in Table 3.

As stated earlier, all tests are performed as 5-way 5-shot tasks. Such a task is shown in Figure 3. When we changed the tasks to 10-way and 20-way, the accuracy dropped, as shown in Table 4. This problem is rather obvious, it is a well-encountered problem with many clustering algorithms, such as k-means; the larger the number of prototypes is, the more difficult it becomes to create a distinct, clear assignment based on a distance metric.

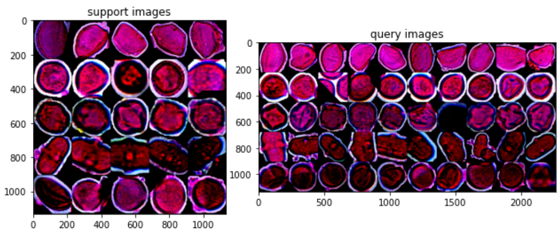

Figure 3: 5-Way 5-shot task from the CPD-1 data set. The query sets Q consists of 10 images, for each class that is represented in the support set S. Preprocessing was applied according to the CNN architecture and training: resizing, using bilinear interpolation, and normalized pixel values.

3.3 Fine-Tuning for Unseen Classes

Although the results without any out-of-domain knowledge are already of high quality, we evaluated the possibility of increasing the results via fine-tuning. Since the domain shift is significant, we decided to update the entire network in combination with a low learning rate of 0.0001 and the Adam adaptive gradient optimizer. We used the NzPollen data and split the 46 classes into 30 for training, 8 for validation, and 8 for testing. The typical Meta-Learning approach works by emulating the evaluation method, i.e. splitting the training and test examples into S and Q to recreate the behavior for a N-way k-shot task during training time. This is done in episodes, coining the term episodic-training. However, recent research (Laenen and Bertinetto, 2021) suggests that this is not optimal. Classical, non-episodic training, by simply utilizing a cross-entropy loss on the meta-training classes, i.e. the dedicated training set, performs better. The authors state, that the separation between S and Q during the episodes negatively affects the distances, which are contributing to the loss. Therefore, we used a non-episodic approach for training our model with the aforementioned loss-function. We choose ResNet-18, since it performed already well on the data set and is smaller in design than e.g. DenseNet-121. We trained 100 epochs and validated after every 10th epoch on the validation set. The 30 training classes contain 11,994 images. During validation 500 5-way 5-shot tasks were sampled from the pool of 8 classes with 3,831 images. For testing, we performed 1,000 tasks of the same shape from the remaining 8 testing classes, in total 3,842 images. On this data, an accuracy of 98% was achieved. Furthermore, we increased N to 8, the max. number of classes in the test set, which also yielded 98% accuracy. The original authors reported the same accuracy of 98% in their work (Sevillano et al., 2020), however, via different means: by averaging a 10-fold cross-validation (90% and 10% of the images for each class for training and testing, respectively). In contrast to our experiments, the imbalanced classes were also filled with augmented images to match the number of images of the class with more samples. Due to the split and without counting augmentation, each fold contained approx. 17,700 training images (from all 46 classes). It is important to notice, that during training, our network was not exposed to any classes that were in the test set. All of the 8 test classes during testing are unknown to the net.

Based on the idea, that the learned features are transferable on the other pollen data sets, due to shared visual characteristics, we evaluated the fine-tuned net on CPD-1 and POLEN23E. However, the accuracy declined, with 69% and 84.2% accuracy, respectively. Decreasing the learning rate from 0.001 to 0.0001 already improved the accuracy significantly, but the results are still subpar when compared to the baseline performance. The results are shown in Ta-

Table 5: Evaluation results of ResNet-18 backbone on all three data sets, with and without Fine-Tuning on nzPollen. All tests are 5-way 5-shot. On 8-way 5-shot tasks (max. number of classes in the test set of nzPollen), the fine-tuned model achieved 98% as well. However, the fine-tuned model on nzPollen does not produce better accuracies when tested on POLEN23E and CPD-1.

	Nz-Pollen	**POLEN 23E**	**CPD-1**
Baseline (Acc.%)	93.5	90.4	78.2
w/ Fine-Tuning (Acc.%)	98 (+4.5)	84.2 (-6.2)	69.4 (-8.8)

ble 5. We deduce the results from the visual differences in image quality, as seen in Figure 2, due to varying capturing methods and use of dye. We assume, the same pattern occurs when training and cross-evaluating with POLEN23E and CPD-1.

4 CONCLUSIONS

In this work, we investigated the applicability of FSL in the field of pollen image classification. It was shown that FSL, specifically PNs, can compete with the results of traditional CNN classification methods, with the advantage of predicting novel classes that were not included in the training process. The baseline models, which have no information about pollen grain images, achieved accuracies up to 95%. Fine-Tuning can increase the accuracy on novel classes up to 98%. The choice of feature extractor cannot be conclusively answered, since each model performed differently, depending on the data set and its specific method of image acquisition and pollen preparation. No model achieved best results on all three sets, we can not recommend a definitive net. Each data set yielded a different result for each backbone. On practical terms, if the acquisition and processing method of pollen grain images are uniform, only a subset of the data would be required for fine-tuning, while novel classes can be classified with a small number of labelled images for the support set. This can drastically reduce the dependency of data being steadily collected and labelled in large quantities.

However, the problem in the feature space for a large number of classes in the support set, i.e. prototypes, requires further attention. The accuracy suffers due to the close proximity of the prototypes. An obvious solution would be to increase the number of images per class in the support set (increasing from 5 to e.g. 10-shots). However, one has to be careful not to leave the paradigm and advantage of FSL, if incorporating and effectively depending on a large amount of data. For future work, the applicability of this method depends on the conditions and requirements that exist for a pollen analysis. Depending on the use-case, the number of class prototypes can be grouped or limited by a set of factors, that predetermine a limited set of candidates. E.g. a typical lab-report for a honey pollen analysis includes grouping by genus (e.g. *Brassica*[2] being the genus, of which rapeseed (*Brassica napus*) is the species). Furthermore, the candidates can be limited due to the geographical origin of the honey sediment and season. This can reduce the number of prototypes in FSL tasks.

REFERENCES

Battiato, S., Ortis, A., Trenta, F., Ascari, L., Politi, M., and Siniscalco, C. (2020). POLLEN13K: A Large Scale Microscope Pollen Grain Image Dataset. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 2456–2460. ISSN: 2381-8549.

Fei-Fei, L., Fergus, and Perona (2003). A Bayesian approach to unsupervised one-shot learning of object categories. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 1134–1141.

Fei-Fei, L., Fergus, R., and Perona, P. (2006). One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Finn, C., Abbeel, P., and Levine, S. (2017). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1126–1135. PMLR. ISSN: 2640-3498.

Gonçalves, A. B., Souza, J. S., Silva, G. G. d., Cereda, M. P., Pott, A., Naka, M. H., and Pistori, H. (2016). Feature Extraction and Machine Learning for the Classification of Brazilian Savannah Pollen Grains. *PLOS ONE*, 11(6):e0157044.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., and Keutzer, K. (2014). DenseNet: Implementing Efficient ConvNet Descriptor Pyramids. arXiv:1404.1869 [cs].

Koch, G., Zemel, R., and Salakhutdinov, R. (2015). Siamese Neural Networks for One-shot Image Recognition. *ICML Deep Learning workshop*.

Laenen, S. and Bertinetto, L. (2021). On Episodes, Prototypical Networks, and Few-Shot Learning. In *Advances in Neural Information Processing Systems*, volume 34, pages 24581–24592. Curran Associates, Inc.

[2] A genus of plants in the mustard and cabbage family.

Lake, B. M., Salakhutdinov, R., and Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338. Publisher: American Association for the Advancement of Science.

Li, M., Zhao, K., Peng, C., Hobson, P., Jennings, T., and Lovell, B. C. (2021). Deep Adaptive Few Example Learning for Microscopy Image Cell Counting. In *2021 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–7.

Lin, T.-Y., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In *Computer Vision - ECCV 2014*, volume 8693, pages 740–755, Cham. Springer International Publishing.

Medela, A., Picon, A., Saratxaga, C. L., Belar, O., Cabezón, V., Cicchi, R., Bilbao, R., and Glover, B. (2019). Few Shot Learning in Histopathological Images:Reducing the Need of Labeled Data on Biological Datasets. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1860–1864. ISSN: 1945-8452.

Nichol, A., Achiam, J., and Schulman, J. (2018). Reptile: A Scalable Meta-Learning Algorithm.

Rajeswaran, A., Finn, C., Kakade, S. M., and Levine, S. (2019). Meta-learning with implicit gradients. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 113–124, Red Hook, NY, USA. Curran Associates Inc.

Rodríguez, P., Laradji, I., Drouin, A., and Lacoste, A. (2020). Embedding Propagation: Smoother Manifold for Few-Shot Classification. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., editors, *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science, pages 121–138, Cham. Springer International Publishing.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis*, 115(3):211–252.

Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., and Lillicrap, T. (2016). Meta-Learning with Memory-Augmented Neural Networks. In *Proceedings of The 33rd International Conference on Machine Learning*, pages 1842–1850. PMLR. ISSN: 1938-7228.

Scotland, R. W. and Wortley, A. H. (2003). How Many Species of Seed Plants Are There? *Taxon*, 52(1):101–104.

Sevillano, V., Holt, K., and Aznarte, J. L. (2020). Precise automatic classification of 46 different pollen types with convolutional neural networks. *PLOS ONE*, 15(6):e0229751.

Shaikh, N. N., Wasag, K., and Nie, Y. (2022). Artifact Identification in Digital Histopathology Images Using Few-Shot Learning. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4. ISSN: 1945-8452.

Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*.

Snell, J., Swersky, K., and Zemel, R. (2017). Prototypical Networks for Few-shot Learning. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Stillman, E. C. and Flenley, J. R. (1996). The needs and prospects for automation in palynology. *Quaternary Science Reviews*, 15(1):1–5.

Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H., and Hospedales, T. M. (2018). Learning to Compare: Relation Network for Few-Shot Learning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1199–1208. ISSN: 2575-7075.

Tan, M. and Le, Q. (2021). EfficientNetV2: Smaller Models and Faster Training. In *Proceedings of the 38th International Conference on Machine Learning*, pages 10096–10106. PMLR. ISSN: 2640-3498.

Tsiknakis, N., Savvidaki, E., Kafetzopoulos, S., Manikis, G., Vidakis, N., Marias, K., and Alissandrakis, E. (2021). Segmenting 20 Types of Pollen Grains for the Cretan Pollen Dataset v1 (CPD-1). *Applied Sciences*, 11(14):6657. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

Viertel, P. and König, M. (2022). Pattern recognition methodologies for pollen grain image classification: a survey. *Machine Vision and Applications*, 33(1):18.

Viertel, P., König, M., and Rexilius, J. (2021). PollenGAN: Synthetic Pollen Grain Image Generation for Data Augmentation. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 44–49.

Vinyals, O., Blundell, C., Lillicrap, T., kavukcuoglu, k., and Wierstra, D. (2016). Matching Networks for One Shot Learning. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Walsh, R., Abdelpakey, M. H., Shehata, M. S., and Mohamed, M. M. (2022). Automated human cell classification in sparse datasets using few-shot learning. *Sci Rep*, 12(1):2924. Number: 1 Publisher: Nature Publishing Group.

Zagoruyko, S. and Komodakis, N. (2015). Learning to compare image patches via convolutional neural networks. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4353–4361. ISSN: 1063-6919.

Zhu, X. and Ghahramani, Z. (2002). Learning from Labeled and Unlabeled Data with Label Propagation. Technical report, Carnegie Mellon University.