Airborne pollen grain detection from partially labelled data utilising semi-supervised learning

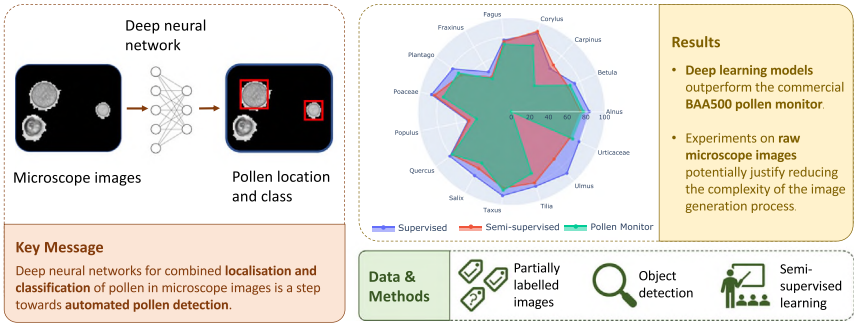
Benjamin Jin [a], Manuel Milling [a], Maria Pilar Plaza [b,c], Jens O. Brunner [d,e,f,g], Claudia Traidl-Hoffmann [b,c], Björn W. Schuller [a,h], Athanasios Damialis [b,i,*]

[a] Chair of Embedded Intelligence for Health Care & Wellbeing, Faculty of Applied Computer Science, University of Augsburg, Augsburg, Germany
[b] Department of Environmental Medicine, Faculty of Medicine, University Clinic of Augsburg & University of Augsburg, Augsburg, Germany
[c] Institute of Environmental Medicine, Helmholtz Center Munich, German Research Center for Environmental Health, Augsburg, Germany
[d] Chair of Health Care Operations/Health Information Management, Faculty of Business and Economics, University of Augsburg, Augsburg, Germany
[e] Decision Science, Healthcare Department of Technology, Management, and Economics, Technical University of Denmark, Denmark
[f] Region Zealand, Data & Development Support, Denmark
[g] Health Care Operations/Health Information Management, Faculty of Business and Economics, Faculty of Medicine, University of Augsburg, Germany
[h] GLAM – The Group on Language, Audio & Music, Imperial College, London, UK
[i] Terrestrial Ecology and Climate Change, Department of Ecology, School of Biology, Faculty of Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece

HIGHLIGHTS

- We used deep-learning neural networks to detect pollen grains in microscope images.
- Our algorithm was compared against the built-in of an automated device BAA500.
- We achieved an unweighted average F1 score of 76.9 % vs. 61.3 % of the BAA500.
- Deep learning pollen detection contributes further to operational pollen monitoring.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Pavlos Kassomenos

Keywords:
Aerobiology
Automatic detection
Object detection
Semi-supervised learning
Deep learning
Pollen taxonomy

ABSTRACT

Airborne pollen monitoring has been conducted for more than a century now, as knowledge of the quantity and periodicity of airborne pollen has diverse use cases, like reconstructing historic climates and tracking current climate change, forensic applications, and up to warning those affected by pollen-induced respiratory allergies. Hence, related work on automation of pollen classification already exists. In contrast, detection of pollen is still conducted manually, and it is the gold standard for accuracy. So, here we used a new-generation, automated, near-real-time pollen monitoring sampler, the BAA500, and we used data consisting of both raw and synthesised microscope images. Apart from the automatically generated, commercially-labelled data of all pollen taxa, we additionally used manual corrections to the pollen taxa, as well as a manually created test set of bounding boxes and pollen taxa, so as to more accurately evaluate the real-life performance. For the pollen detection, we employed two-stage deep neural network object detectors. We explored a semi-supervised training scheme to remedy the partial labelling. Using a teacher-student approach, the model can add pseudo-labels to complete the labelling during training. To evaluate the performance of our deep learning algorithms and to compare them to the commercial algorithm of the BAA500, we created a manual test set, in which an expert aerobiologist corrected automatically annotated labels. For the novel manual test set, both the supervised and semi-supervised approaches clearly outperform the commercial algorithm with an F1 score of up to 76.9 %

Abbreviations: CNN, convolutional neural network; IoU, intersection over union; mAP, mean average precision.

* Corresponding author at: Department of Ecology, School of Biology, Faculty of Sciences, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece.
E-mail addresses: maria.plaza@uni-a.de (M.P. Plaza), dthanos@bio.auth.gr (A. Damialis).

compared to 61.3 %. On an automatically created and partially labelled test dataset, we obtain a maximum mAP of 92.7 %. Additional experiments on raw microscope images show comparable performance for the best models, which potentially justifies reducing the complexity of the image generation process. Our results bring automatic pollen monitoring a step forward, as they close the gap in pollen detection performance between manual and automated procedure.

1. Introduction

A large proportion of the general population displays allergic reactions to airborne pollen. Affected people can experience severe impairment due to pollen-induced allergic rhinitis which is often accompanied by allergic asthma (Bergmann et al., 2016; Meltzer et al., 2009; Brożek et al., 2017). The symptoms include nasal obstruction, rhinorrhoea (runny nose), sneezing, itching of the nose, and post-nasal drainage (Bousquet et al., 2001). Increasing the burden, pollen seasons in northern mid-latitudes start earlier due to climate change (Cissé et al., 2022). As a result, the prevalence of allergic respiratory diseases increases (Cissé et al., 2022; Damialis et al., 2019). Effective countermeasures can be immunotherapy, pharmacotherapy, or prophylaxis by limiting exposure to pollen (Bousquet et al., 2001). As for the latter, digital health systems can support individuals in assessing their risk and adapting their daily plans accordingly.

To illustrate, digital health systems have already been introduced in other medical use cases such as depression recognition (see https://deprexis.com/ or https://www.novego.de/) and treatment of diabetes (see https://www.emperra.com/en/esysta-product-system/). In both cases, these systems monitor diurnal conditions, increase the patient's awareness, and support healthy decisions. In the case of digital health systems for patients experiencing allergic rhinitis due to exposure to airborne pollen, two components are crucial: the provision of real-time information and accurate pollen monitoring.

Provisioning of real-time information has been implemented in the form of web applications (e.g. https://pollenmonitor.com and https://www.pollensense.com). Pollen monitoring has traditionally been performed by manual expert labour which is still considered the gold standard with regard to the accuracy (Oteros et al., 2020). However, the effort of manual labour limits the realisation of genuine real-time monitoring. Various approaches to automating the monitoring process have been proposed: currently, they are distinguished into two groups, those based on digital microscopy and image recognition, and those using light-induced fluorescence (Buters et al., 2022). The former category of monitoring process, which is also the oldest and might be considered as the most developed so far, can be further distinguished between approaches for mere classification (assigning the correct pollen class for an image with a single pollen grain) on which there already exists extensive work (Viertel and König, 2022) and approaches for detection (finding the locations of potentially multiple pollen grains in an image and assigning the correct class to each of them). Detection describes a more holistic approach to pollen monitoring as entire microscope images can be fed into detection algorithms. Therefore, pollen detection includes the additional step of cropping raw microscope images, which is the precondition of pollen classification and thus seems more relevant for real-world application. Leaps in progress in the field of object detection (with applications such as autonomous driving, robot vision, and video surveillance) as described in Zou et al. (2019) and, likewise, domain knowledge transfer, promise to close the gap in quality between manual and automatic pollen monitoring.

Ideally, pollen monitoring applications include real-time information from a large number of high-quality automated monitoring stations. The information of such a network can also enhance accurate pollen forecast, similar to current weather prognosis. An example of the creation of a pollen monitoring network is the electronic pollen information network (see https://epin.lgl.bayern.de/pollenflug-aktuell) that has been deployed in Bavaria, Germany by the Bavarian Health and Food Safety Authority. Beyond the medical application, pollen monitoring can also be applied in further use cases including tracking of climate-induced change of pollen prevalence (Lind et al., 2016), reconstruction of historic climates and environmental change (Zhang et al., 2017; Ivanov et al., 2002), and even the solving of forensic issues (Mildenhall et al., 2006).

In this work, we focus on the improvement of automated pollen detection, expanding on previous research, such as in Gallardo-Caballero et al. (2019) and based on data, which has partly been used for classification tasks before (Schiele et al., 2019; Schäfer et al., 2021). Our main contributions here are the training and evaluation of two-stage object detectors on a dataset with pollen grain images sampled over the period of two years. To our knowledge, our training dataset (manually classified image library) is the largest worldwide, with more than 100,000 images identified. Furthermore, we train and evaluate our models on raw microscope images from a single focal range that are less sharp than previously used images. And last but not least, we propose a novel semi-supervised training scheme to deal with the specific challenges of an automatically created dataset that was only partially corrected by aerobiologist experts.

2. Material and methods

We used microscope images created by BAA500 commercial pollen monitors located at two different sites in Augsburg, Germany. For 2016, the images were created at the Bavarian Environment Agency located at 48°19′33.6″N, 10°54′10.8″E and 1.5 m above ground level. For 2018, the images were created at the Chair of Environmental Medicine at 48° 23′ 04.15″N, 10° 50′35.95″E at 4 m above ground level. (Smaller) parts of the 2016 data have been analysed in previous research on automated pollen classification (Schiele et al., 2019; Schäfer et al., 2021). The BAA500 can produce microscope images at a varying frequency. Typically, the commercial standard model uses 3-hourly intervals. The pollen monitor exercises two main tasks: creating images of particles that are currently in the air and localising and identifying pollen grains in the created images.

The process of the pollen monitor to create the images is described in Heimann et al. (2009). The pollen monitor creates stacks of images for different focal ranges. Afterwards, two different processes aggregate the stacks into a single image. The first process selects the raw image from the stack with maximum sharpness. The second process synthesises the entire stack of raw images into a single synthesised image with optimum sharpness derived from the single images. For the second process, an additional non-optional step deletes parts of the image that the proprietary algorithm of the pollen monitor deems background.

2.1. Datasets

We trained our models on two datasets. For the characteristics of the datasets refer to the upper part of Table 1. Both datasets derived from the images created by the pollen monitor. 2016 + 2018Augsburg15 consists of synthesised and raw images which could both be retrieved from the pollen monitor. For 2016Augsburg15 only synthesised images were available. Both datasets contain 15 distinct pollen taxa that were also previously selected by Schiele et al., 2019 and Schäfer et al. (2021). The 15 pollen taxa include most common pollen types in the atmosphere and some of the most important allergenic pollen taxa (Kolek et al., 2021).

The images in Fig. 1 show the advantages and disadvantages of the two types of output images. On the one hand, the raw images keep information from all parts of the image. As a result, a raw image is not dependent on the proprietary algorithm of the pollen monitor to correctly identify which parts of the image contain pollen grains — which is potentially error-prone. As highlighted in the top centre of the images in Fig. 1, pollen grains

Table 1
Overview of the characteristics of our datasets. Training datasets are shown in the upper part, whereas datasets used for selection and evaluation are shown in the bottom part. The abbreviation Syn. denotes synthesised images. The imbalance ratio is shown under ρ. Note that due to technical problems, images are not available in the time frame from 09/05/2018 to 05/07/2018, depicted in 2016 + 2018Augsburg15.

	Period of time	Type	#Bounding boxes	#Images	ρ
2016Augsburg15	04/11/2015–05/10/2016	Syn.	30,785	11,822	64.4
2016 + 2018Augsburg15	01/01/2016–31/12/2016, 02/01/2018–26/12/2018	Raw & Syn.	253,794	88,131	53.6
PoMoVal	Sampled throughout 2016	Syn.	10,258	3974	70.0
PoMoTest	Sampled throughout 2016	Syn.	10,257	4051	61.5
ManTest	Sampled throughout 2018	Raw & Syn.	3791	976	39.3

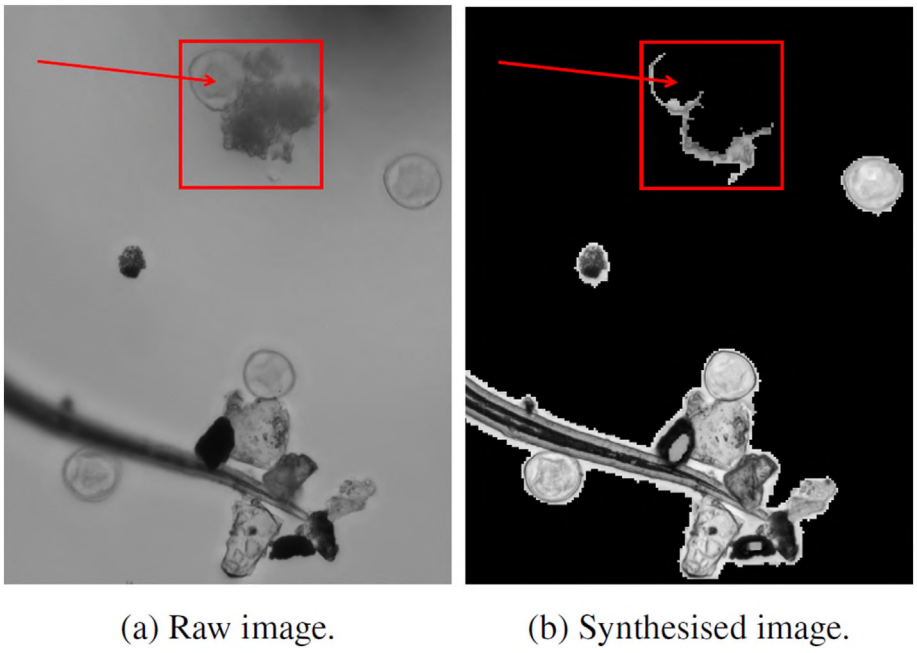


Fig. 1. Examples of raw and synthesised images created by the BAA500. Highlight at a location of an additional pollen grain which is deleted in the synthesised image. Since the particles can be at different ranges from the camera, not all objects can be sharp for the raw image.

can be deleted from the synthesised image by the pollen monitor algorithm because the algorithm does not detect the pollen grain. Since the pollen monitor algorithm does not edit the raw images, those pollen grains will still appear in the raw images. The synthesised images, on the other hand, have the advantage of combining information from all images of the stack. As a result, the synthesised images are generally sharper (compare the images in Fig. 1). Since raw images contain information from a single focal range, some pollen grains merely appear as blurry, unidentifiable, round objects.

In addition to creating images of pollen grains, the pollen monitor also detects pollen grains in the form of bounding boxes with a class label identifying the pollen grain in an automated fashion. For the classification, morphological and textural characteristics are used to determine the label. The manufacturer claims precision and recall of more than 90 % for allergologically more relevant pollen types (*Alnus*, *Ambrosia*, *Artemisia*, *Betula*, *Corylus*, Poaceae, and *Taxus*) and allergologically less relevant (*Acer*, *Carpinus*, *Quercus*, and *Salix*) (Heimann et al., 2009).

In practice, the labelling of the bounding boxes by the pollen monitor can be improved. To do so, the class labels of the bounding boxes provided by the pollen monitor were manually corrected by expert aerobiologists.

In contrast, the locations of the bounding boxes are not manually corrected. Therefore, the locations of the bounding boxes exhibit several problems. For one, multiple overlapping pollen grains can be combined into a single bounding box. Additionally, pollen grains can be combined with dust or other unidentified particles. Finally, some pollen grains do not have a bounding box (see Fig. 2). All the above lead to a consistent

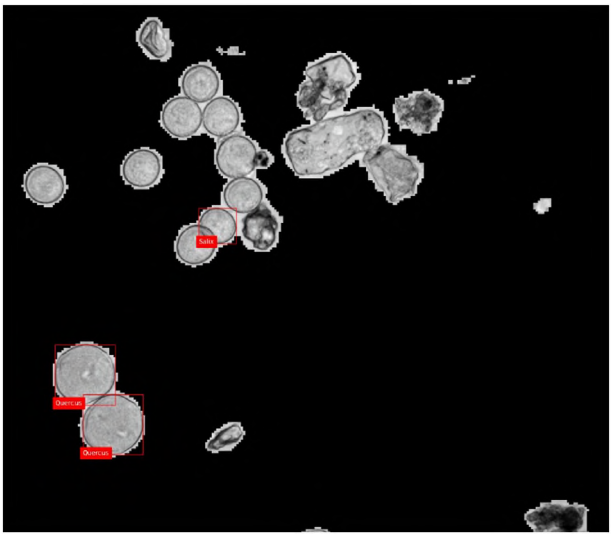


Fig. 2. Example of an image where the pollen monitor pollen detection algorithm only finds a fraction of the existing pollen grains. The bounding boxes derived from the pollen monitor with the manually corrected label are depicted with red boxes.

underestimation of the total amount of pollen counted. Especially for images with many pollen grains, the pollen monitor struggles to locate all samples. Therefore, we call the data partially labelled. To deal with the above problem, we cleaned the dataset in such a way that only the bounding boxes which contain a single pollen grain remain. With this process, we make sure that the pollen detection algorithm is only trained to detect high-quality bounding boxes, even though we concede a penalty if the algorithm predicts a pollen, whose bounding box has been removed. Due to the challenges with missing bounding boxes, we expect our algorithms thus to be pushed towards a bias of false negatives over false positives, i.e., detections of pollen are likely to be correct, but not all pollen grains will be detected.

Another general challenge with the datasets is the distribution of samples for the classes. As shown in Table 1, all datasets exhibit a notable class imbalance. The class imbalance for the different datasets is measured with the imbalance ratio ρ defined in Johnson and Khoshgoftaar (2019). The imbalance ratio shows the ratio between the number of samples in the class with the most samples and the class with the least samples. Thus, a high ρ shows a high imbalance.

For overviews on the distribution of pollen grain classes for the datasets, please refer to Supplementary Material, Figs. S1, S2, and S3.

2.2. PoMoVal, PoMoTest, and ManTest

For the selection of models and for a first evaluation of the models, we used the types of images that are in 2016Augsburg15. As a result, the training on 2016Augsburg15 and 2016 + 2018Augsburg15 is likely to excel on these datasets. However, as another result, the datasets also suffer from the same bias to the pollen monitor and the derived challenges described in Section 2.1. Therefore, we call these datasets **Po**llen **Mo**nitor **Val**idation (PoMoVal) and **Po**llen **Mo**nitor **Test** (PoMoTest) in the following. PoMoVal and PoMoTest differ slightly when used for models that are trained on 2016Augsburg15 and 2016 + 2018Augsburg15, respectively. When used for models trained on 2016 + 2018Augsburg15, 81 (or about 2 %) of the 4051 images are omitted to ensure comparability between experiments on raw and on synthesised images.

To accurately measure the real-world performance of a model, we created an additional **Man**ual **Test** set (ManTest) with additional bounding boxes for pollen grain samples that the pollen monitor did not detect. The annotations were performed by expert aerobiologists in the current team by use of a customised annotation tool. 74 % of the bounding boxes were correctly created by the pollen monitor, the remaining 26 % were added manually with the used annotation tool. For an overview of the pollen frequency and the distribution of the added manual labels please refer to Fig. S3.

Despite the manual effort, the resulting test set still exhibits properties which make pollen detection challenging for evaluated models on both the raw and the synthesised images:

1. The bounding box location distribution of the manual test set is different from 2016Augsburg15 and 2016 + 2018Augsburg15. While the former contains bounding boxes for pollen grain samples at the edges of the image, the latter do not.
2. Previously existing bounding boxes can contain unidentifiable content in the raw images, because they are not sharp in all parts of the image.
3. Additional pollen grain samples can exist without bounding boxes in synthesised images, as the manual labelling was done on the raw images where parts of the image can be blurry.
4. Bounding boxes can be in blackened parts of the synthesised image.

Some of the challenges arise from the fact that raw images were chosen for the labelling process, since the synthesised images do not contain any information in the blackened parts of the images.

2.3. Data augmentation and pre-processing

To prevent overfitting, we employed data augmentation to artificially inflate the existing dataset. For the experiments, we applied different configurations of the following data augmentations. For geometric augmentations we use: horizontal flip, vertical flip, centre crop, and rotation. When used, each of the augmentations was applied with a probability of 0.5. For some of the geometric augmentations, there are implications that are not immediately obvious. The centre crop augmentation adds bounding boxes at the edges of the image. These bounding boxes often only contain parts of a whole pollen grain. Such bounding boxes do not appear in 2016Augsburg15 or in 2016 + 2018Augsburg15. The same holds true for rotation augmentation since bounding boxes can be partly rotated outside of the image. A second implication of the rotation augmentation is slightly less accurate bounding boxes because simply rotating the bounding boxes leads to boxes that are not aligned with the image frame. Therefore, the rotated bounding boxes are enclosed in bigger – hence, less accurate – but aligned bounding boxes.

Furthermore, we employed augmentation that changes the value of pixels in the image. Here, we use sharpness adjustment augmentation for the semi-supervised model, with a probability of 0.25. We also tried out solarisation and colour jitter augmentations in preliminary experiments, but decided against their usage due to inferior performance.

2.4. Supervised model

As a baseline, we trained a set of supervised learning algorithms as this is the most straightforward approach for most object detection tasks. Here, we used an adapted implementation of a Faster R-CNN (Ren et al., 2017) from torchvision (Paszke et al., 2019). For the backbone convolutional neural network (CNN), we used pre-trained models obtained from PyTorch Image Models (Wightman, 2019). For future work, the written experiment code allows for the flexible integration of the plethora of pre-trained models that are available in Wightman (2019) and will presumably work with future models with minimal effort, too. On top of the backbone, a feature pyramid network (Lin et al., 2017a) was employed to exploit the information encoded at different scales of the backbone CNN. The resulting feature maps were used by the region proposal network to identify objects in different scales.

2.5. Semi-supervised model

To tackle the challenge of missing bounding boxes, we used a semi-supervised teacher-student model. The goal is to jointly learn to detect correct bounding boxes as they appear in the ground truth labelling of the dataset and be able to add bounding boxes in order to detect additional pollen grains that are missing in the labelling of an image. As a result, the training configuration was slightly different from the traditional approach in semi-supervised learning, which often operates on a smaller dataset consisting of labelled images and a larger unlabelled dataset. An overview of the proposed architecture can be found in Fig. 3.

In more detail, both the teacher and the student consist of a Faster R-CNN as described in Section 2.4. However, only the student is trainable, while the teacher's parameters are frozen. For the training of the student, first the teacher predicts bounding boxes for the images that are fed to the student.

The predicted bounding boxes with their class labels, so-called pseudo labels (Lee, 2013), were filtered to obtain bounding boxes with high confidence only.

The filtering was done twice. First, we discarded region proposals with objectness scores smaller than 0.9, i.e., if the probability assigned by the model to indicate if a region proposal contains a pollen grain is smaller than 0.9. And, second, we discarded final output bounding boxes with confidence scores smaller than 0.9. Furthermore, bounding boxes with an intersection over union (IoU) greater than 0.7 compared to any of the existing ground truth bounding boxes were discarded, such that multiple detections of the same pollen grain were not encouraged. IoU is a metric measuring to what extent two bounding boxes overlap with each other. To aggressively remove overlapping boxes, we do not take the predicted class into account in this step.

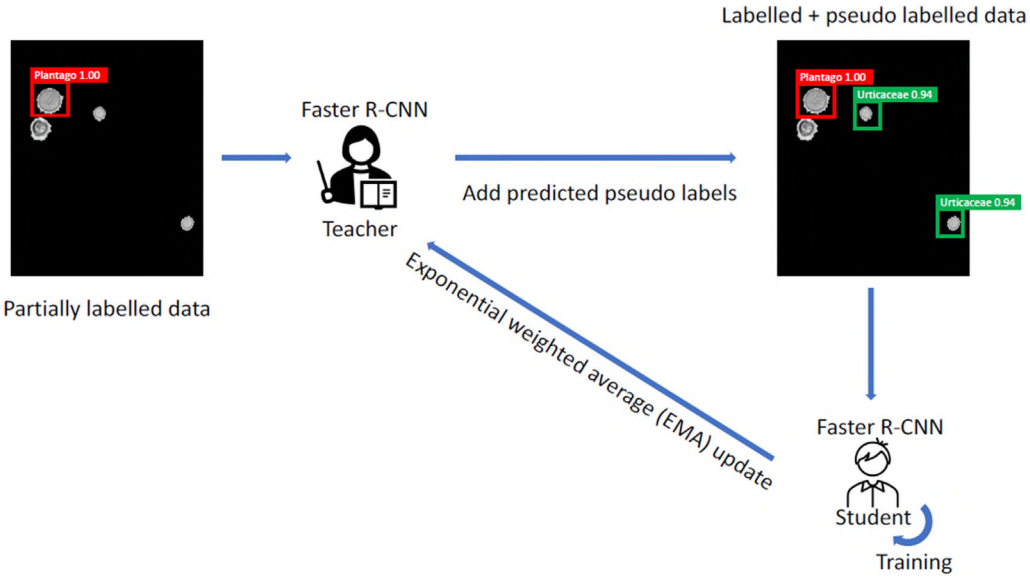


Fig. 3. Overview of our semi-supervised learning approach. A teacher model enriches the available labelled data with pseudo labels. The student trains on the dataset labels in conjunction with the pseudo labels. The teacher parameters are an exponentially weighted average of the student parameters.

The image with the ground truth bounding boxes and the pseudo labels was forwarded to the student to use for training according to the supervised learning paradigm. However, teacher and student used differently augmented input images, with the student having an additional sharpness adjustment augmentation applied. This practice of weak augmentation for the teacher in conjunction with stronger augmentation for the student has been called consistency regularisation (Sohn et al., 2020) and has previously been described in Bachman et al. (2014); Sajjadi et al. (2016); Laine and Aila (2017). We employed geometric augmentations for teacher and student. The student input images are augmented with an additional adjustment of the sharpness for stronger augmentation.

Furthermore, for the classification of the region of interest heads, a new loss function as described in Xu et al. (2021) was used. The loss comprises a supervised loss and an unsupervised loss. Here, we base both losses on the focal loss (Lin et al., 2017b). In addition, the loss function leverages confidence scores from the teacher model for the predicted bounding boxes of the student to weigh the loss derived from the student boxes.

To update the teacher, we used an exponential weighted average of the parameters of the student as described in Tarvainen and Valpola (2017).

Our experiments are implemented with Python and the neural network code is based on PyTorch (Paszke et al., 2019). Furthermore, we used PyTorchLightning (Falcon et al., 2019) to avoid boilerplate code for the training implementation. The code for the experiments is publicly available (supervised at https://github.com/MilesGrey/pollen-detection and semi-supervised at https://github.com/MilesGrey/ssl-pollen-detection).

3. Results

3.1. Metrics

We evaluated the performance of our models with two distinct types of metrics: the F1 score, which shows the capabilities of assigning a bounding box to the correct pollen class in terms of precision and recall, and the mean average precision (mAP), commonly used for object detection, which measures overall detection performance (Liu et al., 2020; Jiao et al., 2019; Zaidi et al., 2022). For the latter, predicted bounding boxes are compared to the ground truth box for each class individually and correct detections are considered if the IoU between predicted and ground truth bounding box is above a certain threshold, otherwise, they are assigned to the background class. For our work, we chose a threshold of 0.5 (denoted as mAP@0.5), which is on the lower end of commonly used thresholds, since a pixel-accurate location of the bounding boxes is not considered important for the task at hand.

3.2. Training on 2016Augsburg15

The models were trained on 2016Augsburg15 for 40 epochs with a batch size of 4 and the best models during training (measured with the mAP@0.5 performance on PoMoVal) were used for the final evaluation on the test sets. For the training, we chose an Adam optimiser (Kingma and Ba, 2014) with a learning rate of $\gamma = 0.0001$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. While we selected focal as the loss function for the supervised model, we applied an additional unsupervised loss as described in Section 2.5 for the semi-supervised model. Additionally, we set the decay of the exponential weighted average update of the teacher to 0.99 during a ramp-up phase of three epochs and 0.999 afterwards, in our semi-supervised experiments. A ramp-up phase is suggested by Tarvainen and Valpola (2017) because the student improves more quickly at the start of the training. For the backbone, we used an EfficientNetV2 (Tan and Le, 2021) for all experiments. We ran preliminary experiments with the ubiquitous ResNet50 (He et al., 2016) as a backbone which exhibited similar performance to the EfficientNetV2, even though ResNet50 has a much larger number of parameters and thus, requires more heavy computations for training.

The results of the experiments are summarised in Table 2. It is noticeable that the performance evaluation depends largely on the employed test data. Whilst the supervised approach has a slightly better performance on PoMoTest – with a maximum mAP@0.5 of 90.9 % compared to 89.6 % – the semi-supervised approach shows a superior performance on the more relevant manual test set with a maximum mAP@0.5 of 74.6 % compared to 70.9 %.

Similarly, application of crop and rotation data augmentation is detrimental to the evaluation on PoMoTest, but clearly helps the performance of the semi-supervised approach on the manual test set. For the supervised approach however, the crop and rotation data augmentation are detrimental to both test datasets.

3.3. Training on 2016 + 2018Augsburg15

We evaluate the models trained on the 2016 + 2018Augsburg15 dataset similarly to the ones trained on the 2016Augsburg15 dataset. The hyperparameters are also set similarly, but here, we limited the experiments to using both crop and rotation augmentation.

Table 2
Overview of our experiment results with training on 2016Augsburg15 and 2016 + 2018Augsburg15 in terms of mAP@0.5. Supervised training (see Section 2.4) and semi-supervised training (see Section 2.5) are denoted with SL and SSL, respectively.

2016Augsburg15				
Train	Crop + Rot.	mAP@0.5 (%)		
		PoMoVal	PoMoTest	ManTest
SL	No	91.0	90.9	70.9
	Yes	90.2	89.8	70.5
SSL	No	90.0	89.6	71.6
	Yes	86.3	85.2	74.6
Synthesised 2016 + 2018Augsburg15				
Train	Crop + Rot.	mAP@0.5 (%)		
		PoMoVal	PoMoTest	ManTest
SL	Yes	93.1	92.7	77.9
SSL	Yes	88.8	88.0	75.9
Raw 2016 + 2018Augsburg15				
Train	Crop + Rot.	mAP@0.5 (%)		
		PoMoVal	PoMoTest	ManTest
SL	Yes	90.3	90.1	78.3
SSL	Yes	73.5	73.0	74.6

An overview of the mAP@0.5 scores to compare the supervised and semi-supervised models is shown in Table 2. Additionally, we utilise the original (non-corrected) annotations of the pollen monitor as a baseline to compare our experiments against. Unfortunately, mAP scores cannot be calculated for the pollen monitor because the predictions of the pollen monitor do not have confidence levels. Instead, precision, recall, and F1 scores were calculated and compared. For our models, we set the confidence level with maximum F1 score to filter the predictions. The results are shown in Table 3.

3.3.1. Synthesised images

The mAP@0.5 results show that on this dataset, the supervised model has the edge over the semi-supervised model with an absolute difference of about 4.5 % on the PoMoTest set and 2 % on the manual test set. Overall, both approaches, in particular the supervised approach, improve their performance compared to the smaller 2016Augsburg15 dataset.

Expectedly, the pollen monitor's performance is competitive with an F1 score of 83.9 % on the PoMoTest set. With 78.2 %, the semi-supervised model even has a lower F1 score compared to the pollen monitor. The score can be attributed to the lower precision. In contrast, the performance on the manual test set is considerably worse for the pollen monitor with an F1 score of 61.3 %. Especially the recall suffers with only roughly half of the ground truth pollen grains being detected by the pollen monitor (52.5 %). Compared to the pollen monitor, the supervised and the semi-supervised approaches yielded drastically better recall with 70.3 % and 70.5 %, respectively. As a result, both display much better F1 scores with 76.0 % and 71.0 %, as well.

Table 3
Precision, recall, and F1 score to compare our models trained on 2016 + 2018Augsburg15 with the pollen monitor algorithm.

Synthesised images						
Model	PoMoTest			ManTest		
	Prec. (%)	Rec. (%)	F1 (%)	Prec. (%)	Rec. (%)	F1 (%)
Pollen monitor	87.1	81.9	83.9	79.2	52.5	61.3
SL	84.5	92.1	88.1	82.7	70.3	76.0
SSL	74.1	82.8	78.2	71.6	70.5	71.0
Raw images						
Model	PoMoTest			ManTest		
	Prec. (%)	Rec. (%)	F1 (%)	Prec. (%)	Rec. (%)	F1 (%)
Pollen monitor	87.1	81.9	83.9	79.2	52.5	61.3
SL	83.3	87.9	85.5	81.2	73.1	76.9
SSL	68.4	75.1	71.6	75.7	70.0	72.8

3.3.2. Raw images

On the raw images, the mAP@0.5 results on the PoMoTest set for the supervised model were similar but lower with 90.1 %. In contrast, scores for the semi-supervised model were considerably lower with 73.0 % mAP@0.5.

For the manual test set, the mAP@0.5 for the supervised model was slightly higher than for the synthesised images with 78.3 %. While the score on the manual test set is lower for the semi-supervised model with 74.6 % mAP@0.5, the difference was not as large as on the PoMoTest set.

In comparison to the pollen monitor, the result from the supervised model on the PoMoTest set trumps the pollen monitor with 85.5 % compared to 83.9 % F1 score. For the semi-supervised model, the F1 score of 71.6 % suffers on the PoMoTest set derived from the pollen monitor. For the manual test set, the best results were produced from the supervised model (with an F1 score of 76.9 % compared to 61.3 % of the pollen monitor). The semi-supervised model considerably improved the pollen monitor performance with an F1 score 72.8 %, as well.

A comparison of the class-specific performance is shown in Fig. 4. Furthermore, Fig. 5 displays a confusion matrix of the supervised model to show the strengths and weaknesses of the model across classes. Notably, the performance of the models and especially the pollen monitor is volatile for different classes.

4. Discussion

Regarding the adopted methods, we introduced techniques and implementation details to train detection algorithms for pollen grains in microscope images with partially labelled data, such as some of the datasets presented in Section 2.

Here we employed a multi-approach inter-comparison of pollen detection methods, an outcome quite sought after the last decade. As for the last almost 70 years the ‘gold-standard’ of airborne pollen monitoring is still a rather simple technique of sampling and classifying the pollen captured in the atmosphere, it is obviously necessary to employ the next-generation, automated and as close to real-time as possible. So, here we

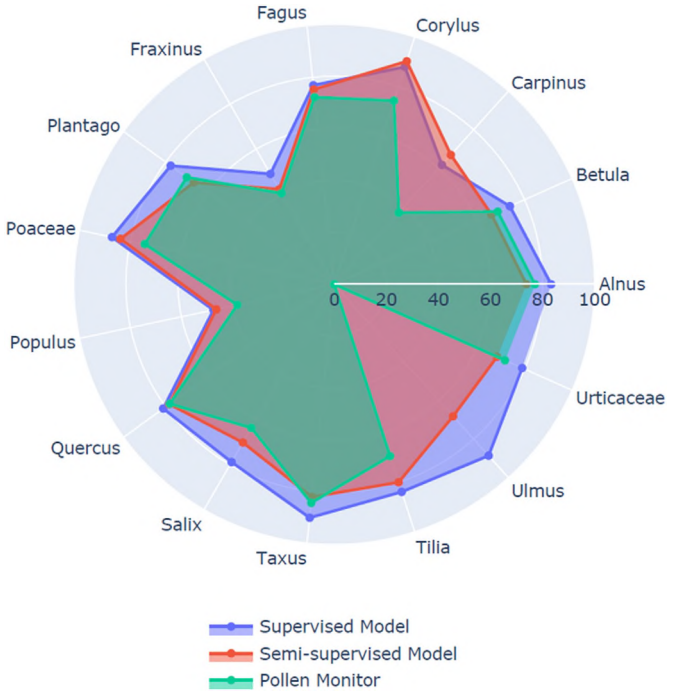


Fig. 4. Depiction of class-specific F1 scores on the manual test set. The supervised and semi-supervised models were trained on raw images of 2016 + 2018Augsburg15.

hypothesised that adoption of cutting-edge, sophisticated methods of pollen detection and classification would contribute significantly to advance the state-of-the-art.

For the strengths of our work, we highlight the performance of our deep neural networks compared to the performance of the algorithm of the pollen monitor currently deployed as part of the electronic pollen information network in Bavaria, Germany. In our experiments, the best models show a better performance on both PoMoTest and ManTest, especially for the latter, as indicated in Table 3. We also found lower detection performance for the pollen monitor than was claimed by the manufacturer, albeit with a different set of pollen taxa for the evaluation.

We also emphasise the performance of the proposed deep learning models on synthesised images in comparison to their performance on raw images, which are available from the current version of the pollen monitor, too.

We find that our best models show comparable performance for both types of images. As a result, simplifying the image generation process of the pollen monitor by reducing the amount of images produced for the synthesised images and by omitting the additional synthesis operation seems like a viable option. Still, for optimal detection performance we believe the optimal solution is to use synthesised images, which do not have the additional step of blackening parts of the images, which has not previously been made accessible by the proprietary software. Nevertheless, more recent versions of the pollen monitor provide these images (Hund GmbH, 2009, 2022), so they might be preferred in future work.

Even though some advantage of the deep learning algorithms seems reasonable as they are presented with corrected pollen classes, the extent of the improvement seems remarkable, since the deep learning algorithms are still based on the original predictions of the pollen monitor and thus have to deal with incomplete data in form of missing bounding boxes. Especially our results on the manual test set imply the potential of deep learning algorithms for automated pollen detection over traditional approaches. This conclusion becomes even more apparent when considering that a large part of the misclassifications on the manual test set results from missed bounding boxes (indicated by the ratio between precision and recall at

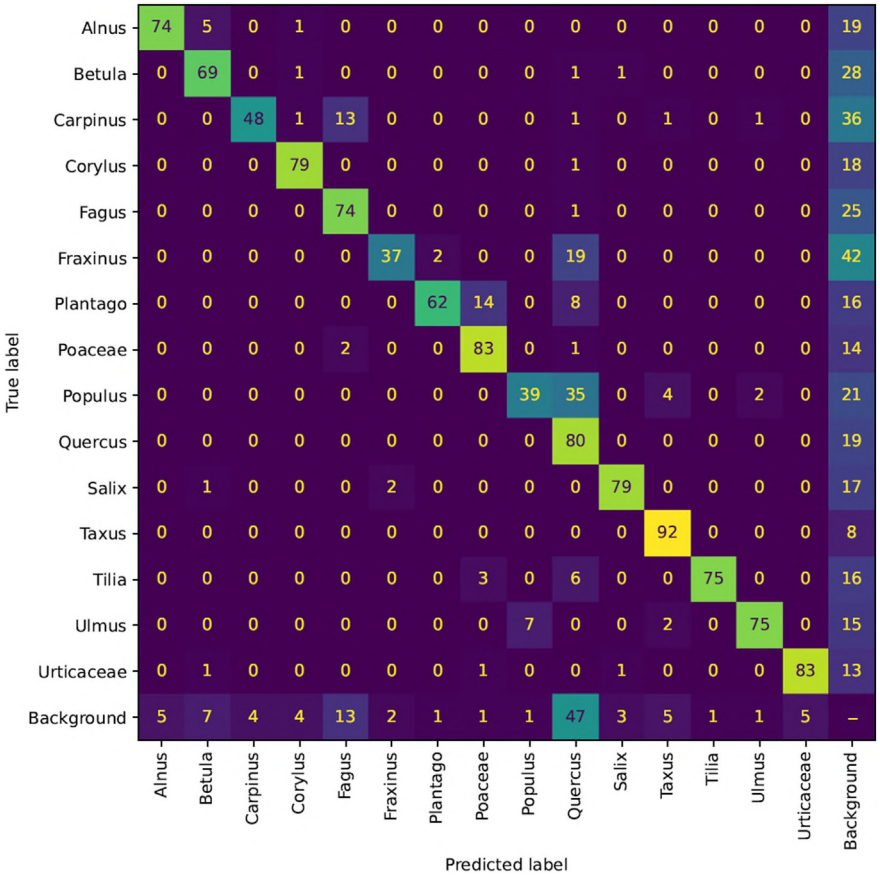


Fig. 5. Confusion matrix of the highest scoring model on the manual test set, the supervised model. The model was trained on raw images of 2016 + 2018Augsburg15.

the point of maximum F1), which we expect to be improved with a higher quality of the labelled data.

Furthermore, our methods can be transferred, i.e., they can be retrained on datasets with different types of pollen. And, our models are flexible as they can be configured to operate on different confidence levels to maximise precision or recall.

As for the differences between the different deep neural networks that we used in our work, we evaluated the performance of semi-supervised training to deal with partially labelled data compared to purely supervised training.

We found that the semi-supervised model (see Section 2.5) shows better performance than the supervised model (see Section 2.4) on the smaller dataset, while the supervised one exhibits better performance on the larger training set. Possible reasons for the different performances include differences in the datasets (difference in size, location of the pollen monitor, annotators). Moreover, deep learning models typically improve their performance when trained on larger training datasets. However, an interesting observation is that the performance of the supervised approach scales better than the semi-supervised approach. Additional techniques such as verifying the quality of the pseudo labels created by the semi-supervised learning method can potentially improve the exploitation of additional data.

The semi-supervised learning methods described are based on a level of distrust of the ground truth labels and thereby enable the model to simultaneously improve the dataset labelling and to exploit the improved labels. For our datasets, the motivation for distrust is straightforward as they have the problem of missing boxes, which we introduced in Section 2.

However, even for completely manually labelled datasets, single erroneous labels are always possible and can hugely impact model performance (Bulusu et al., 2020). That is why learning models to distrust labelling can be a fertile research field to pursue based on our findings.

Having said the above, we are aware that our study exhibits certain limitations. First, the datasets are dependent on the image generation and annotation of the pollen monitor. All images result from pollen monitors located at two sites, collecting data in the same region in Augsburg, Germany, implying potential biases in the data collection. In order to obtain a more general impression of the generalisability of our models, further data collection and evaluation is necessary. Furthermore, even though the bounding box class labels are corrected by experts – and in the case of the manual test set additional bounding boxes are added – the basis of the data consists of automatically generated bounding boxes by a proprietary algorithm of the pollen monitor (for more details on how the pollen monitor operates, refer to Section 2). As a result, the evaluations on the PoMoTest set are biased by the bounding box creation and even the manual test set might retain some of that bias. To a certain degree, the design of the study therefore encourages approximation to the pollen monitor algorithm. The above actually pinpoint that even though our employed methods look promising, we are confident that if we can surpass the existing shortcomings, we would expect even better results in the automatic pollen monitoring techniques.

5. Conclusion

We applied the Faster R-CNN framework on partially labelled data for the task of airborne pollen grain detection, introducing a novel semi-supervised learning approach for partially labelled data, based on a teacher-student design. We show that our deep learning-based approaches clearly outperform the detection capabilities of the commercial BAA500 pollen monitor on a test data set, which has been manually cleaned by expert aerobiologists-annotators. Furthermore, we showed that the processing of single raw images from the pollen monitor could be a viable option compared to a complex pre-processing step synthesising a stack of images, performed by a proprietary algorithm of the pollen monitor. Future work can explore adapting the proposed semi-supervised approach on the larger 2016 + 2018Augsburg15 dataset to potentially help the model leverage the additional training data more effectively. Beyond, further data collection at different places seems necessary to improve and test the robustness of the algorithm to different environments.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.164295.

CRediT authorship contribution statement

Benjamin Jin: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Manuel Milling:** Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Maria Pilar Plaza:** Investigation, Data curation. **Jens O. Brunner:** Resources, Writing – review & editing, Funding acquisition. **Claudia Traidl-Hoffmann:** Resources, Writing – review & editing, Funding acquisition. **Björn W. Schuller:** Resources, Writing – review & editing, Supervision. **Athanasios Damialis:** Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bachman, P., Alsharif, O., Precup, D., 2014. Ensembles. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2014/file/66be31e4c40d676991f2405aaecc6934-Paper.pdf.

Bergmann, K.C., Heinrich, J., Niemann, H., 2016. Current status of allergy prevalence in Germany. Allergo J. Int. 25, 6–10. https://doi.org/10.1007/s40629-016-0092-6.

Bousquet, J., van Cauwenberge, P., Khaltaev, N., 2001. Allergic rhinitis and its impact on asthma. J. Allergy Clin. Immunol. 108, S147–S334. https://doi.org/10.1067/mai.2001.118891.

Brożek, J.L., Bousquet, J., Agache, I., Agarwal, A., Bachert, C., Bosnic-Anticevich, S., Brignardello-Petersen, R., Canonica, G.W., Casale, T., Chavannes, N.H., Correia de Sousa, J., Cruz, A.A., Cuello-Garcia, C.A., Demoly, P., Dykewicz, M., Etxeandia-Ikobaltzeta, I., Florez, I.D., Fokkens, W., Fonseca, J., Hellings, P.W., Klimek, L., Kowalski, S., Kuna, P., Laisaar, K.T., Larenas-Linnemann, D.E., Lødrup Carlsen, K.C., Manning, P.J., Meltzer, E., Mullol, J., Muraro, A., O'Hehir, R., Ohta, K., Panzner, P., Papadopoulos, N., Park, H.S., Passalacqua, G., Pawankar, R., Price, D., Riva, J.J., Roldán, Y., Ryan, D., Sadeghirad, B., Samolinski, B., Schmid-Grendelmeier, P., Sheikh, A., Togias, A., Valero, A., Valiulis, A., Valovirta, E., Ventresca, M., Wallace, D., Waserman, S., Wickman, M., Wiercioch, W., Yepes-Nuñez, J.J., Zhang, L., Zhang, Y., Zidarn, M., Zuberbier, T., Schünemann, H.J., 2017. Allergic rhinitis and its impact on asthma (ARIA) guidelines—2016 revision. J. Allergy Clin. Immunol. 140, 950–958.. https://www.sciencedirect.com/science/article/pii/S0091674917309193. https://doi.org/10.1016/j.jaci.2017.03.050.

Bulusu, S., Kailkhura, B., Li, B., Varshney, P.K., Song, D., 2020. Anomalous example detection in deep learning: a survey. IEEE Access 8, 132330–132347. https://doi.org/10.1109/ACCESS.2020.3010274.

Buters, J., Clot, B., Galán, C., Gehrig, R., Gilge, S., Hentges, F., O'Connor, D., Sikoparija, B., Skjoth, C., Tummon, F., Adams-Groom, B., Antunes, C.M., Bruffaerts, N., Çelenk, S., Crouzy, B., Guillaud, G., Hajkova, L., Seliger, A.K., Oliver, G., Ribeiro, H., Rodinkova, V., Saarto, A., Sauliene, I., Sozinova, O., Stjepanovic, B., 2022. Automatic detection of airborne pollen: an overview. Aerobiologia https://doi.org/10.1007/s10453-022-09750-x.

Cissé, G., McLeman, R., Adams, H., Aldunce, P., Bowen, K., Campbell-Lendrum, D., Clayton, S., Ebi, K., Hess, J., Huang, C., Liu, Q., McGregor, G., Semenza, J., Tirado, M., 2022. Health, wellbeing, and the changing structure of communities. In: Pörtner, H.O., Roberts, D., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., Rama, B. (Eds.), Climate Change 2022: Impacts, Adaption, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Chapter 7. Cambridge University Press, Cambridge, UK and New York, NY, USA.

Damialis, A., Traidl-Hoffmann, C., Treudler, R., 2019. Climate Change and Pollen Allergies. Springer International Publishing, Cham, pp. 47–66 https://doi.org/10.1007/978-3-030-02318-8_3.

Falcon, W., et al., 2019. PyTorch Lightning. URL https://github.com/PyTorchLightning/pytorch-lightning. https://doi.org/10.5281/zenodo.3828935.

Gallardo-Caballero, R., García-Orellana, C.J., García-Manso, A., González-Velasco, H.M., Tormo-Molina, R., Macías-Macías, M., 2019. Precise pollen grain detection in bright

field microscopy using deep learning techniques. Sensors (Basel, Switzerland) 19, 3583. https://doi.org/10.3390/s19163583.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 https://doi.org/10.1109/CVPR.2016.90.
Heimann, U., Haus, J., Zuehlke, D., 2009. Op3 — fully automated pollen analysis and counting: the pollen monitor baa500. Proceedings OPTO 2009 & IRS2 2009. AMA Association for Sensors and Measurement, pp. 125–128 https://doi.org/10.5162/opto09/op3.
Hund GmbH, 2009. BAA500 Pollen Monitor. Hund GmbH, URL. https://www.hund.de/images/pdf/Datasheet_BAA_english_12.03.09.pdf.
Hund GmbH, 2022. Pollen Monitoring Systems BAA500e, BAA502. Hund GmbH, URL. https://www.hund.de/images/pdf/Broschre_BAA500eBAA502_V32_english_Einzelseiten.pdf.
Ivanov, D., Ashraf, A., Mosbrugger, V., Palamarev, E., 2002. Palynological evidence for Miocene climate change in the Forecarpathian Basin (Central Paratethys, NW Bulgaria). Palaeogeogr. Palaeoclimatol. Palaeoecol. 178, 19–37.. https://www.sciencedirect.com/science/article/pii/S0031018201003650. https://doi.org/10.1016/S0031-0182(01)00365-0.
Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., Qu, R., 2019. A survey of deep learning-based object detection. IEEE Access 7, 128837–128868. https://doi.org/10.1109/ACCESS.2019.2939201.
Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. J. Big Data 6, 27. https://doi.org/10.1186/s40537-019-0192-5.
Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. https://arxiv.org/abs/1412.6980. https://doi.org/10.48550/ARXIV.1412.6980.
Kolek, F., Plaza, M.P., Charalampopoulos, A., Traidl-Hoffmann, C., Damialis, A., 2021. Biodiversity, abundance, seasonal and diurnal airborne pollen distribution patterns at two different heights in Augsburg, Germany. Atmos. Environ. 267, 118774. https://doi.org/10.1016/j.atmosenv.2021.118774.
Laine, S., Aila, T., 2017. Temporal ensembling for semi-supervised learning. 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings. OpenReview.net, URL . https://openreview.net/forum?id=BJ6oOfqge.
Lee, D.H., 2013. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks. ICML Workshop on Challenges in Representation Learning.
Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017a. Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944 https://doi.org/10.1109/CVPR.2017.106.
Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017b. Focal loss for dense object detection. 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2999–3007 https://doi.org/10.1109/ICCV.2017.324.
Lind, T., Ekebom, A., Alm Kübler, K., Östensson, P., Bellander, T., Lõhmus, M., 2016. Pollen season trends (1973-2013) in Stockholm area, Sweden. PLoS One 11, 1–12. https://doi.org/10.1371/journal.pone.0166887.
Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M., 2020. Deep learning for generic object detection: a survey. Int. J. Comput. Vis. 128, 261–318. https://doi.org/10.1007/s11263-019-01247-4.
Meltzer, E.O., Blaiss, M.S., Derebery, M.J., Mahr, T.A., Gordon, B.R., Sheth, K.K., Simmons, A.L., Wingertzahn, M.A., Boyle, J.M., 2009. Burden of allergic rhinitis: results from the pediatric allergies in America survey. J. Allergy Clin. Immunol. 124, S43–S70.. https://www.sciencedirect.com/science/article/pii/S0091674909008045. https://doi.org/10.1016/j.jaci.2009.05.013.
Mildenhall, D., Wiltshire, P., Bryant, V., 2006. Forensic palynology: why do it and how it works. Forensic Sci. Int. 163, 163–172 URL: https://www.sciencedirect.com/science/article/pii/S0379073806005007. https://doi.org/10.1016/j.forsciint.2006.07.012.
Oteros, J., Weber, A., Kutzora, S., Rojo, J., Heinze, S., Herr, C., Gebauer, R., Schmidt-Weber, C.B., Buters, J.T., 2020. An operational robotic pollen monitoring network based on automatic image recognition. Environ. Res. 191, 110031 URL: https://www.sciencedirect.com/science/article/pii/S0013935120309282. https://doi.org/10.1016/j.envres.2020.110031.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Volume 32. Curran Associates, Inc. , pp. 8024–8035.. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.
Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.
Sajjadi, M., Javanmardi, M., Tasdizen, T., 2016. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. Proceedings of the 30th International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, pp. 1171–1179.
Schäfer, J., Milling, M., Schuller, B.W., Bauer, B., Brunner, J.O., Traidl-Hoffmann, C., Damialis, A., 2021. Towards automatic airborne pollen monitoring: from commercial devices to operational by mitigating class-imbalance in a deep learning approach. Sci. Total Environ. 796, 148932 URL: https://www.sciencedirect.com/science/article/pii/S0048969721040043. https://doi.org/10.1016/j.scitotenv.2021.148932.
Schiele, J., Rabe, F., Schmitt, M., Glaser, M., Haring, F., Brunner, J.O., Bauer, B., Schuller, B., Traidl-Hoffmann, C., Damialis, A., 2019. Automated classification of airborne pollen using neural networks. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE https://doi.org/10.1109/embc.2019.8856910.
Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L., 2020. Fixmatch: simplifying semi-supervised learning with consistency and confidence. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc. , pp. 596–608 URL https://proceedings.neurips.cc/paper/2020/file/06964dce9addb1c5cb5d6e3d9838f733-Paper.pdf.
Tan, M., Le, Q., 2021. Efficientnetv2: smaller models and faster training. In: Meila, M., Zhang, T. (Eds.), Proceedings of the 38th International Conference on Machine Learning, PMLR , pp. 10096–10106 URL https://proceedings.mlr.press/v139/tan21a.html.
Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc. . https://proceedings.neurips.cc/paper/2017/file/68053af2923e00204c3ca7c6a3150cf7-Paper.pdf.
Viertel, P., König, M., 2022. Pattern recognition methodologies for pollen grain image classification: a survey. Mach. Vis. Appl. 33, 18. https://doi.org/10.1007/s00138-021-01271-w.
Wightman, R., 2019. Pytorch Image Models. https://github.com/rwightman/pytorch-image-models. https://doi.org/10.5281/zenodo.4414861.
Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F., Bai, X., Liu, Z., 2021. End-to-end semi-supervised object detection with soft teacher. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3060–3069.
Zaidi, S.S.A., Ansari, M.S., Aslam, A., Kanwal, N., Asghar, M., Lee, B., 2022. A survey of modern deep learning based object detection models. Digital Sign. Proc. 126, 103514 URL: https://www.sciencedirect.com/science/article/pii/S1051200422001312. https://doi.org/10.1016/j.dsp.2022.103514.
Zhang, W., Lu, H., Li, C., Dodson, J., Meng, X., 2017. Pollen preservation and its potential influence on paleoenvironmental reconstruction in Chinese loess deposits. Rev. Palaeobot. Palynol. 240, 1–10.. https://www.sciencedirect.com/science/article/pii/S0034666716302214. https://doi.org/10.1016/j.revpalbo.2017.01.002.
Zou, Z., Shi, Z., Guo, Y., Ye, J., 2019. Object Detection in 20 years: A Survey. https://arxiv.org/abs/1905.05055. https://doi.org/10.48550/ARXIV.1905.05055.