Image Processing for Improving Detection of Pollen Grains in Light Microscopy Images

Elżbieta Kubera
University of Life Sciences in Lublin
Lublin, Poland *elzbieta.kubera@up.lublin.pl*

Alicja Wieczorkowska
Polish-Japanese Academy of Information Technology
Warsaw, Poland *alicja@poljap.edu.pl*

Agnieszka Kubik-Komar
University of Life Sciences in Lublin
Lublin, Poland *agnieszka.kubik@up.lublin.pl*

Krystyna Piotrowska-Weryszko
University of Life Sciences in Lublin
Lublin, Poland *krystyna.piotrowska@up.lublin.pl*

Agata Konarska
University of Life Sciences in Lublin
Lublin, Poland *agata.konarska@up.lublin.pl*

Abstract

In this paper, we address the problem of preparing data for training detectors to identify transparent objects in light microscopy images. To this end, we propose using blends of reference images and monitoring background, instead of time-consuming labelling of monitoring data. This approach allowed us to avoid the need to involve a palynologist in the preparation of the training data while also ensuring 100% correct ground-truth labels. The statistical analysis of the deep learning results confirms that the results obtained for blends only are more stable, and in some cases surpass the results obtained for the training set with some labelled monitoring data added to reference images and monitoring background.

Keywords: Pollen monitoring, image recognition, machine learning, biological pollution.

1. Introduction

Pollen-related allergies are a serious problem, as estimates suggest that one in four people suffer from them. The costs associated with treatment and missed workdays caused by severe allergic reactions can reach 150 billion euros per year [5]. Palynoecological studies indicate that pollen grains in urban areas and polluted industrial regions can change their structure and even exacerbate allergic reactions [23].

Most of springtime allergies are caused by tree pollen [1]. To forecast pollen counts, pollen monitoring is performed worldwide, and pollen concentration is calculated. This monitoring is based on counting pollen grains under a microscope in material acquired from pollen traps, which aspire air with various particles. Hirst-type pollen traps, commonly used in pollen monitoring, aspire 10 liters of air per minute, thus mimicking human breathing [10]. The material collected in the trap contains pollen grains, but also dust, fungal spores, and other particles that fall on the adhesive tape passing in the slit of the trap. This material is then analyzed under an optical microscope by a palynologist, i.e. a specialist in pollen research.

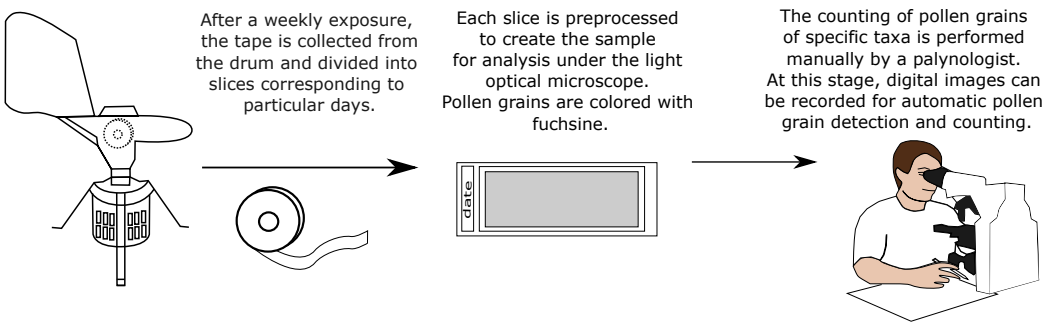


Fig. 1. The procedure of performing pollen monitoring by a palynologist. It consists of capturing airborne biological particles, sample preparation, and counting.

In Central and Eastern Europe, the pollen grains of $Alnus$ (alder), $Betula$ (birch), and $Corylus$ (hazel) are common allergy triggers. These pollen grains are very similar in shape and size, making them extremely difficult to distinguish. Furthermore, their pollen seasons partially overlap, so the pollen grains of these taxa can be found together in the analyzed material, which may contain numerous pollen grains, even more than 1000 grains per taxon daily, as is the case with $Alnus$ and $Betula$. This is why we decided to address the problem of automated counting of pollen grains from monitoring images for these taxa.

Palynologists classify each grain in the monitoring samples, and count grains for all taxa. If it is not possible to recognize the taxon, because it is unknown to the palynologist, damaged, or overlaps with other objects, it is classified as $Varia$. The expected outcome of pollen monitoring is the number of pollen grains counted per taxon; the exact location of each pollen grain is less important. The data acquisition procedure in pollen monitoring is shown in Fig. 1.

A manual counting of pollen grains for each taxon is a tedious and time-consuming task, thus researchers and companies undertake an effort to automate this task [4], [11]. Nevertheless, manual counting is still a common and reliable procedure [21], and even when deep learning is applied to facilitate pollen counting, the participation of palynologists may be necessary [3]. For example, the authors in [3] present a web-based application for pollen count for Mediterranean taxa, based on deep learning using YOLOv7 (they followed the success of our works based on YOLO [13, 14]), but human review is applied in this approach to check whether semi-automated labelling is correct. The authors of [12] used automatically generated, commercially-labelled data, and additionally used manual corrections to the pollen taxa, as well as a manually created test set of bounding boxes and pollen taxa. In [7], the authors based their work on a set of slides prepared in the laboratory, where each slide contains grains of only one pollen type, to avoid problems with class labelling (this approach was also used in our previous works [13, 14]). In [18], the problem of artificial intelligence-based classification of pollen grains is addressed, but for pollen that can be found in honey-based products.

1.1. Problem statement

Automatic counting of pollen grains for each target taxon is not an easy task. Although it is not necessary to determine the exact location of each pollen grain, the algorithm must be trained on labelled data. To avoid time-consuming manual labelling by palynologists, reference images with pollen grains of one taxon only can be used in training, see Fig. 2a. As we can see, this image has a uniform background and pollen grains of round shape, for which the marking of the bounding boxes does not require a palynologist's skills. However, monitoring images have a much more complex background, see Fig. 2b, and they can also contain other round objects, e.g. air bubbles, see Fig. 2d. Additionally, monitoring images may contain pollen grains of various

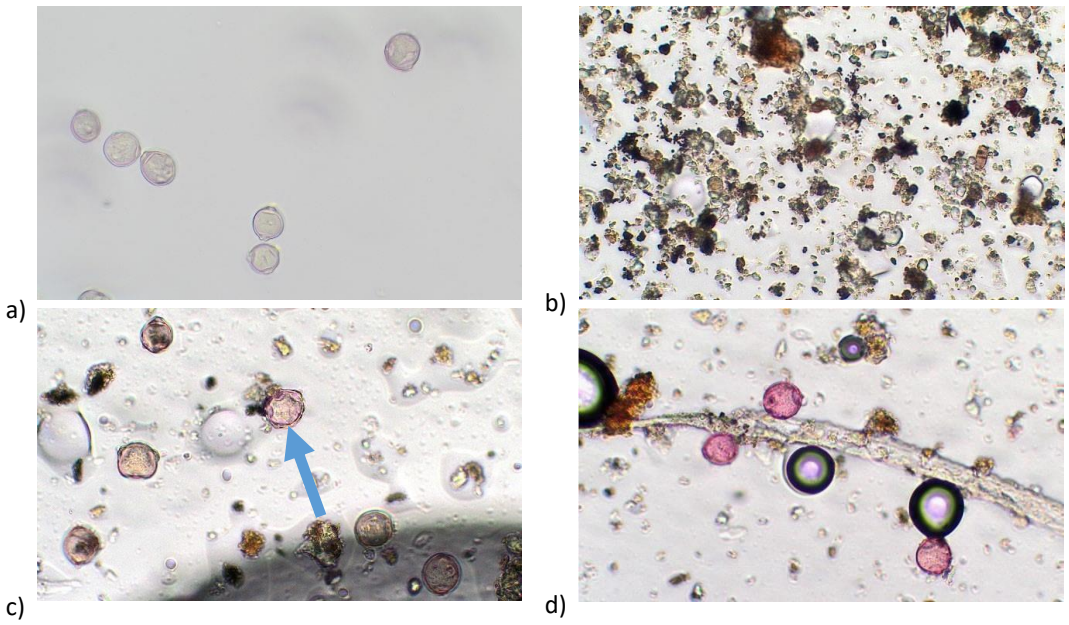


Fig. 2. Images used in our experiments: a) reference image, $Betula$ pollen grains; b) monitoring image without pollen (i.e. monitoring background, dust only); c) monitoring image, $Betula$ (5 pollen grains) and $Alnus$ (the grain marked with an arrow); d) monitoring image, air bubbles and $Betula$ pollen grains

taxa, see Fig. 2c. As we can see, pollen grains of various taxa are similar. We should also keep in mind that pollen grains are 3-dimensional and semitransparent objects, so they can look different depending on their position and microscope focus. In our work, we assume that counted pollen grains are in focus, as we obtain images provided this way from palynologists, but the problem of out-of-focus images is also addressed in automated pollen monitoring research [6].

In our previous research [14, 15], we first worked with reference images only, using deep neural networks, and then also with monitoring images, manually labelled. In this work, we would like to take the next step towards automation, namely avoiding the costly manual labelling of monitoring data. To this end, we decided to experiment with blends of reference images and monitoring background images, to check if such images can replace labelled monitoring images. This can save palynologists the work of preparing data to obtain ground truth for automatic counting via deep neural networks and prevent errors in labelling.

2. Materials and methods

Our data include images of pollen grains of $Alnus$, $Betula$, and $Corylus$, both from reference samples and monitoring images. In addition, some of the monitoring images do not contain pollen grains, and these images were used as a monitoring background. We used similar data in our previous works [14, 15], but the data in these works included images in two different scales, namely 400x and 600x magnification. In this work, all images are on the same scale, with 400x magnification. Therefore, the data are more uniform, but also with fewer details available. Additionally, we removed less saturated images of $Alnus$ pollen grains, thus obtaining even more uniform data.

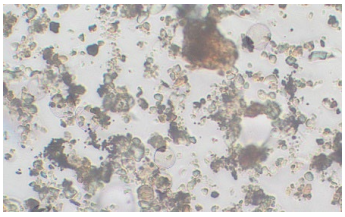
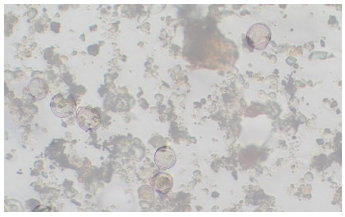
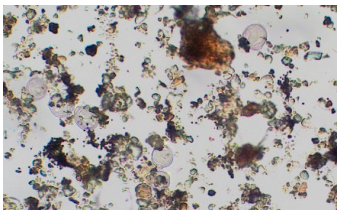
a)  b)  c) 

Fig. 3. Sample images obtained via blending a reference image and an image of the monitoring background without pollen: a) blend consisting of 50% of the reference image value and 50% of the background value in each pixel, b) blend consisting of 25% of the reference image value and 75% of the background value in each pixel, c) blend with minimum of the reference and background image value in each pixel

2.1. Blending methods

Images can be merged in various ways. For instance, objects (e.g. pollen grains) from one image can be cut out and put in front of another image (e.g. background). Another option is to use blends, where the pixel values of a combined image are calculated as combinations of the corresponding pixel values of the input images. Since pollen grains are transparent, we decided to use this approach and prepare blends of images to be used in the training and validation of detectors. In this way, we simulate the overlap of transparent pollen grains (taken from reference images with a clean background) with dirt and other objects from the monitoring background. Blending is performed in RGB (red, green, blue). We prepared the following blends:

- 50% of the reference image value and 50% of the background image value in each pixel,
- 25% of the reference image value and 75% of the background image value in each pixel,
- minimum of the reference and background image value in each pixel.

Each blended image was obtained as a result of blending one reference image and one randomly selected monitoring background image. Therefore, we can rely on the labelling available for the reference images. In this way, we can train detectors even without using labelled monitoring data in training.

The sample images obtained by blending are shown in Fig. 3. Summing increases the lightness, and since the background in both reference and monitoring images is light gray, i.e. of high RGB values, we decided to use blends with a lower proportion of reference images with pollen grains (besides standard blends with 50%/50% proportion), as this way pollen grains are more noticeable in blends.

2.2. Train, Validation, and Test Sets

Our training data contain reference images with clean and uniform backgrounds, and pollen representing only one taxon in one picture. As we could see in Fig. 2, such data are very different from monitoring images, which must be manually labelled to obtain ground-truth data. Furthermore, monitoring images may contain many small rounded objects similar to target classes (i.e. pollen grains of $Alnus$, $Betula$, and $Corylus$), including air bubbles. This is why we decided to investigate whether it is possible to train detectors without labelled monitoring data and whether the obtained results will be comparable. Therefore, we prepared several sets of data for the training and validation of the detectors (see Tab. 1), while our test data contain only monitoring images. Our training and validation datasets are described below.

Table 1. Number of images and pollen grains in train, validation, and test sets. Background represents images without pollen, monitoring represents images with pollen (ref, mon, bg, bub, bln_{50}, bln_{25}, min - reference, monitoring, background, bubbled images, and blends 50/50, 25/75 and minimum, respectively)

Set	Data for:	Images		Pollen grains of:		
	Training	with pollen	without pollen	$Alnus$	$Betula$	$Corylus$
$RefBg_{Train}$	reference + background	244 ref	32 bg	370	312	316
$RefBgMon_{Train}$	reference + background + monitoring	256 = 244 ref + 12 mon	32 bg	388	328	336
$RefBgBln_{Train}$	reference + background + blends	280 = 244 ref + 12 bln_{50} + 12 bln_{25} + 12 min	32 bg	415	363	382
$BgBlnBub_{Train}$	all blends + background + background with air bubbles	732 = 244 bln_{50} + 244 bln_{25} + 244 min	47 = 32 bg + 15 bub	1110	936	948
	Validation					
$RefBg_{Val}$	reference + background	66 ref	7 bg	100	96	59
$RefBgMon_{Val}$	reference + background + monitoring	69 = 66 ref + 3 mon	7 bg	103	96	61
$BlnBg_{Val}$	all blends + background	198 = 66 bln_{50} + 66 bln_{25} + 66 min	7 bg	300	288	177
	Testing					
Mon_{Test}	Monitoring images	110	0	53	215	55

The $RefBgBln_{Train}$ set contains 36 blends, added to the set of 244 reference and 32 background images. These blends were obtained using our 3 blending methods for 12 reference images and 12 monitoring background images not used earlier in training or validation. The $BgBlnBub_{Train}$ set contains the same 32 background images as before, and blends obtained using our 3 blending methods for each reference image and a randomly selected background image from the $RefBg_{Train}$ set, plus 15 background images with air bubbles (not used before). The $BlnBg_{Val}$ set contains background images and blends created using reference and background images from the $RefBg_{Val}$ set.

The results obtained for training the detectors using the $RefBg_{Train}$ set will constitute the baseline. We did not prepare any training set with reference images only, as they are too different from the monitoring data, but we added the monitoring background without pollen, as such images do not need labelling of the target pollen grains.

To compare the results obtained with and without using monitoring images, we also used some monitoring images in training and validation, namely 12 in training and 3 in validation, together with 244 reference images and 32 background images (less than 5% of training data). This is because we had very few labelled monitoring images, and saved most of them for tests, thus expecting lower, but reliable results.

The monitoring images used in training included pollen grains of $Alnus$, $Corylus$, and $Betula$. A random selection of monitoring images for validation ended with the selection of only $Alnus$ and $Corylus$ pollen grains; $Betula$ pollen was represented by reference images and blends only. In the test set, the $Betula$ class is the most represented one (up to 12 grains in one image), because it is most commonly found in pollen traps in its pollen season. At the same time, there were no more than 7 grains of $Corylus$ in one image and 4 of $Alnus$.

Monitoring data were used in only one training set, without blends, as our goal was to investigate whether costly monitoring data could be replaced by blends.

2.3. Object Detection in Images

Object detection in images generally uses one-stage and two-stage methods; more stages can also be used. All methods aim to find the location of objects of interest and classify them into predefined classes. Two-stage methods determine regions of interest (RoI) in the first stage and then classify objects in each RoI in the second stage. One-stage methods are usually faster. A simplified approach has also been recently used, without hand-designed processing [2].

Deep neural networks are commonly used for object detection in images. One-stage methods include such methods as YOLO (You Only Look Once, [22]), or RetinaNet [17]. YOLO uses a fully convolutional neural network (CNN) to predict the bounding boxes and class probabilities of objects in images. RetinaNet uses a unified (single) network composed of a backbone network and two task-specific subnetworks. The two-stage approach includes R-CNN methods, i.e. Regions with CNN features [9]. A simplified approach called DEtection TRansformers (DETR) [2] uses a transformer encoder-decoder architecture; DETR-based models achieve good performance in object detection tasks.

Models used in object detection within images are often evaluated on publicly available datasets. The Common Object in Context COCO [16] is widely used as benchmark data in object detection tasks in images [20], and models pre-trained on these data are also available.

In our work, we decided to use YOLOv8 as an object detection model, since we used YOLO (earlier versions) in our previous work, and recently popular transformer architectures are much slower to train; they are computationally more intensive. We used YOLOv8 models pre-trained on COCO as a starting point for further training, following the transfer-learning approach.

2.4. Evaluation measures

Various measures can be used to evaluate results, including general measures such as precision and recall, but measures dedicated to object detection in images are more frequently used. These measures are often calculated on the basis of the bounding boxes of the objects. For example, the box mAP@.5 (mean Average Precision) measure is calculated based on the predicted and ground-truth bounding boxes of the detected objects. First, Intersection over Union (IoU) is calculated as the overlap between the predicted and true bounding boxes, and then IoU$\geq$0.5 means that it is a hit, otherwise it is a fail. Therefore, box mAP@.5 represents mean average precision with bounding boxes with IoU$\geq$0.5 classified as positives. Average precision is calculated as an area under the precision-recall curve for each class separately.

Measures based on higher IoU levels can also be used if a precise location of the object of interest is important. In our case, the location of the detected object is not important, thus a standard box mAP@.5 measure is used. Additionally, we used average precision calculated for each class separately.

3. Experiments and Results

In our experiments, we used Ultralytics YOLOv8.0.122 [22] and NVIDIA GeForce RTX 3060 to obtain pollen grain detectors for microscope images acquired from pollen monitoring. YOLOv8 training was performed for 3 different random seed values, namely 3, 4, and 5, for each training data set. Each model was trained for 500 epochs, to check how the detection results change for consecutive epochs. We selected the best model (indicated by the best results for the validation set used) from these 500 epochs, but the results for the last model are also shown for comparison. We present results using the box mAP@.5 measure, most commonly used for the evaluation of object detection in images. The results of our experiments on the test set are presented in Tab. 2. It should be noted that these results are calculated as the average of per-class results, thus the results for each class equally contribute to the box mAP@.5 measure, no matter how many objects are in each class.

The construction of the training and validation set that allows obtaining the best results for the test set is a challenging task. However, the results presented in Tab. 2 show that we can obtain better results than the baseline using the proposed sets, but usually not with the model from the last epoch of the detector training. Therefore, the use of validation sets to select the best model among all epochs of training is a better solution than taking the last model. Various validation sets indicate different models as the best ones.

We selected 3 models (detectors) for more detailed analysis, namely:

Model 1 - the model obtained for $RefBg_{Train}$ used as a train set and $RefBg_{Val}$ as a validation set, i.e. with simple and similar train and validation sets, containing reference images and background without manually labelled pollen grains from monitoring data (baseline);

Model 2 - the model obtained for $RefBgMon_{Train}$ used as a train set and $RefBgMon_{Val}$ as a validation set, i.e. with manually labelled monitoring data (costly to obtain, but necessary when ground-truth pollen monitoring data are needed);

Model 3 - the model obtained for $BgBlnBub_{Train}$ used as a train set and $BlnBg_{Val}$ used as a validation set; this is our best model, which contains blended images and monitoring background.

We selected these models to compare the proposed blend-based approach, represented by Model 3, with both the baseline model (Model 1) and the model obtained using manually labelled monitoring data (Model 2). In our previous work, we found that the addition of monitor-

Table 2. Box mAP@.5 obtained for the test set Mon_{Test} in our experiments. The best results are shown in bold. The results corresponding to the models selected for further analysis are underlined.

Box mAP@.5 for:	$RefBg_{Val}$	$RefBgMon_{Val}$	$BlnBg_{Val}$	Last model
$RefBg_{Train}$ (baseline)	61.90%	61.90%	**63.23%**	60.10%
$RefBgMon_{Train}$	60.27%	**62.70%**	61.07%	49.10%
$RefBgBln_{Train}$	61.13%	60.10%	**63.57%**	63.47%
$BgBlnBub_{Train}$	67.07%	63.63%	**68.03%**	65.97%

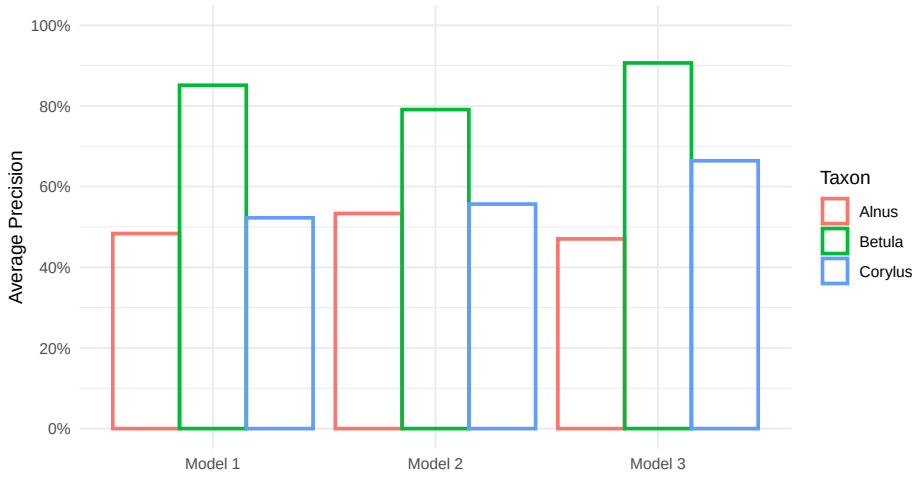


Fig. 4. Average precision results for the selected models for each class.

ing data to reference images in training improves the detection results significantly [15]. However, a specialist is needed to label pollen grains in the monitoring data. This choice of models and the comparative analysis of their results allow investigation of whether manually labelled monitoring data can be replaced with blends of the reference data and monitoring background without decreasing the detection quality.

Fig. 4 shows per-class average precision for these three models (with bounding boxes with IoU$\geq$0.5 classified as positives, see Section 2.3). As we can see, the best results are obtained for $Betula$, which is the largest class in our test data that comes from pollen monitoring. Therefore, $Betula$ pollen is best recognized, and since $Betula$ produces the most allergenic tree pollen in North, Central, and Eastern Europe, we can conclude that our models allow good detection of the most important allergen.

3.1. Statistical analysis

To compare statistically the quality of the selected models, we performed the tests also for the subsets of the test set. We divided the test set into 23 subsets, with each subset containing pollen of all taxa (although no image contained all taxa). Therefore, the precision of the obtained bounding boxes can be calculated for each class in each subset. Since we have 3 models, and each model was trained 3 times (using 3 seeds), we obtained:

$$23(\text{subsets}) \times 3(\text{models}) \times 3(\text{seeds}) = 207 \text{ values for each taxon.}$$

The distribution of average precision for each taxon within each model was significantly different from the normal distribution. Therefore, the statistical comparison of selected models

Table 3. Statistics for average precision for each taxon for the selected models.

Model	taxon	Mean	Minimum	Maximum	Std.Dev.
Model 1	*Alnus*	0 .566	0 .000	0 .995	0 .343
	Betula	0 .657	0 .000	0 .995	0 .368
	Corylus	0 .570	0 .000	0 .995	0 .378
Model 2	*Alnus*	0 .685	0 .000	0 .995	0 .303
	Betula	0 .617	0 .000	0 .995	0 .319
	Corylus	0 .742	0 .000	0 .995	0 .264
Model 3	*Alnus*	0 .623	0 .000	0 .995	0 .292
	Betula	0 .646	0 .000	0 .995	0 .383
	Corylus	0 .793	0 .332	0 .995	0 .222

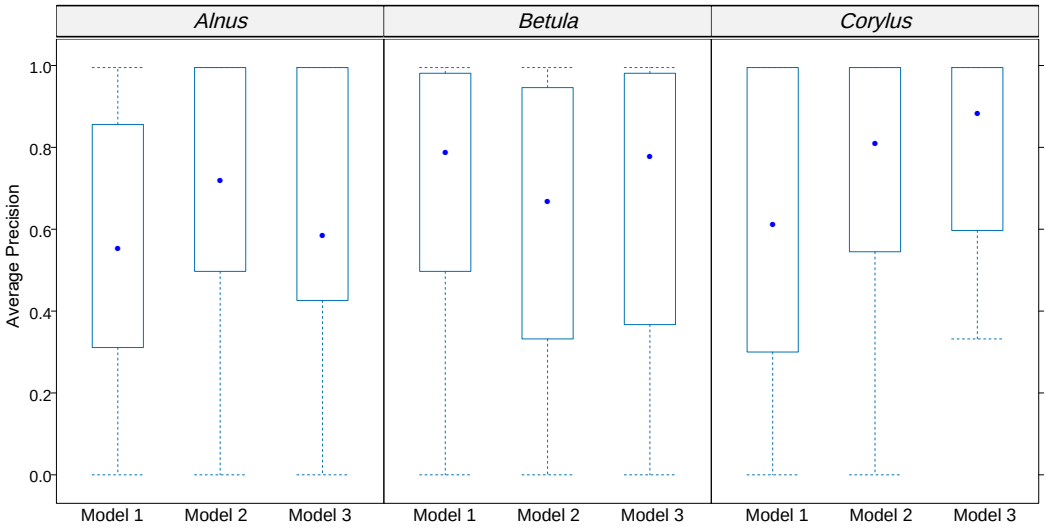


Fig. 5. The box plots of the average precision values for each taxon.

was made using the non-parametric test for repeated measurements, based on the relative treatment effects (RTE). RTE is an estimated probability that a randomly chosen observation from the whole dataset results in a smaller value than a randomly chosen observation from the studied group. Thus, the higher the RTE values for a given group of results, the better the results of this group compared to the other groups.

Each test subset was used as input for each model version (built with seeds 3, 4, and 5) of our detectors, i.e., of Model 1, Model 2, and Model 3. Therefore, we used LD-F2 design for experiments with two sub-plot factors [19], treating our 3 models (3 levels) and their versions obtained via the training with 3 seeds (3 levels) as the first and second sub-plot factor variables, respectively. We tested the null hypothesis H_0 that the observations from the studied groups have the same distribution, against the alternative one (H_1) that the distributions of observations from different groups are not equal. In addition, we applied multiple comparisons with the Holm-Bonferroni adjustment to the p-value [8], which allows controlling Type I error inflation.

The descriptive characteristics are presented in Tab. 3. The best results were obtained for Model 3 for the *Corylus* class, with an average precision of almost 0.8; this result also had the smallest dispersion and the smallest standard deviation. The lowest average precision (0.566) was obtained for Model 1 and the *Alnus* class. The box plots presented in Fig. 5 show the distribution of these results across the studied models.

Table 4. The results of the ANOVA-type test for average precision of the investigated taxa and models. Versions represent model versions obtained for three seeds (3,4, and 5).

		Statistic	df	p-value
Average precision for $Alnus$	Model	1.305	1.935	0.271
	Version	1.414	1.894	0.243
	Model:Version	1.945	2.894	0.122
Average precision for $Betula$	Model	0.630	1.693	0.507
	Version	0.467	1.419	0.561
	Model:Version	3.681	3.265	0.0093**
Average precision for $Corylus$	Model	4.983	1.745	0.0096**
	Version	4.248	1.815	0.017*
	Model:Version	2.228	3.077	0.081

* significance level $\alpha = 0.05$, ** significance level $\alpha = 0.01$

Table 5. Multiple comparisons of box AP for $Corylus$ distribution between the studied models with Holm-Bonferroni adjustment (p_{adj}).

Comparison	p-value	p_{adj}
Model1 – Model2	0.002**	0.006**
Model1 – Model3	0.005**	0.0098**
Model2 – Model3	0.432	0.432

** significance level $\alpha = 0.01$

ANOVA-type statistic was used to verify the hypothesis of the equality of the distribution of the average precision values obtained for each taxon for the studied models and their three training versions (for three seeds). The results are presented in Tab. 4. As we can see, there are no significant differences between the distributions of the average precision values for $Alnus$ and $Betula$ for the studied models. For $Alnus$ and $Corylus$ we cannot reject the null hypothesis of the equality of the distributions of the average precision values for the Model:Version interactions.

We conclude that the distribution of the average precision values for $Corylus$ significantly differs for the studied models. It is therefore worth investigating which models show significant differences in the distributions of the average precision values. These comparisons are shown in Tab. 5. As we can see, there are no significant differences between Model 2 and Model 3. This is a very satisfactory conclusion, as the preparation of the data for Model 3 does not require the manual labelling by a specialist.

We also tested the significance of differences in the distribution of the average precision values for the $Betula$ taxon between the training versions for each model. We found that there are significant differences between these versions for Model 2 (p-value=0.006), whereas Model 1 and Model 3 are more stable (p-value=0.142, and 0.845 respectively), i.e. there are no significant differences between different versions of these two models.

4. Summary and Conclusions

In this paper, we addressed the problem of preparing data sets for training detectors for pollen monitoring. Manual pollen grain counting is a tedious task, performed by specialists, as is the labelling of ground-truth data. Moreover, manual labelling of monitoring data may produce

errors. This is why we suggest using image augmentation in the form of the blends of reference images and monitoring background, which does not require labelling by a palynologist. The goal of this work was to investigate whether the use of blends instead of a small set of manually labelled monitoring data does not deteriorate or even improves the obtained detection results.

We cannot compare our results with other works (and it was not our main goal) because other published results were calculated for different data sets, but we are aware that the results obtained for taxa with less similar pollen grains are usually higher.

We prepared several versions of train and validation sets, and analyzed the results. No statistically significant differences were found between the model trained on blended images (Model 3) and the model trained with a small number of labelled monitoring images added to the train set (Model 2). These detectors give statistically better results than the baseline model. Additionally, the results obtained for Model 3 are significantly more stable than the results for the other studied models, thus we can recommend Model 3 as the most preferable one in the pollen grain detection task.

Furthermore, we note that $Betula$ pollen grains were the easiest to detect for the detectors investigated. This is a good result, as $Betula$ pollen causes most spring allergic reactions to tree pollen.

In our previous work, we used train data enriched with less saturated images for $Alnus$, thus the train set was less uniform, and we obtained even better results [15]. Additionally, two different magnification scales were used, i.e. 600x and 400x, whereas our data represented 400x magnification only, with the 600x magnified images re-scaled to 400x magnification. This supports the conclusion that the more diverse and detailed the data, the better the results.

In future work, we intend to explore other blending methods as well. We also plan to use DETR-based detectors, as it is another promising approach to detecting objects in images, and continue our work towards automated counting of the detected pollen grains.

References

1. Asthma and Allergy Foundation of America. 2023 Allergy Capitals, https://aafa.org/wp-content/uploads/2023/03/aafa-2023-allergy-capitals-report.pdf. Last accessed 31 January 2024
2. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-End Object Detection with Transformers. https://doi.org/10.48550/arXiv.2005.12872 (2020)
3. Chaves, A.J, Martín, C., Torres, L.L,, Díaz, M., Ruiz-Mata, R., de Gálvez-Montañez, E., Recio, M., Trigo, M.M., Picornell, A.: Pollen recognition through an open-source web-based system: automated particle counting for aerobiological analysis. Earth Sci Inform 17, 699–710 (2024)
4. Clot, B., Gilge, S., Hajkova, L., Magyar, D., Scheifinger, H., Sofiev, M., Bütler, F., Tummon, F. The EUMETNET AutoPollen programme: establishing a prototype automatic pollen monitoring network in Europe. Aerobiologia (2020)
5. European Centre for Medium-Range Weather Forecasts: Forecasting pollen to alleviate allergy suffering, https://stories.ecmwf.int/forecasting-pollen-to-alleviate-allergy-suffering/index.html Last accessed 28 June 2024
6. Filipovych, R., Daood, A., Ribeiro, E., Bush, M.: Pollen Recognition in Optical Microscopy by Matching Multifocal Image Sequences. In: 23rd International Conference on Pattern Recognition (ICPR), pp. 2128-2133, IEEE, Cancún, México (2016)
7. Gallardo, R., García-Orellana, C.J., González-Velasco, H.M., García-Manso, A., Tormo-Molina, R., Macías-Macías, M., Abengózar, E.: Automated multifocus pollen

detection using deep learning. Multimed Tools Appl, https://doi.org/10.1007/s11042-024-18450-2
8. Giacalone, M., Zirilli, A., Cozzucoli, P.C., Alibrandi, A.: Bonferroni-Holm and permutation tests to compare health data: methodological and applicative issues. BMC Medical Research Methodology 18 (2018)
9. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. Tech report (v5). https://arxiv.org/pdf/1311.2524.pdf (2014)
10. Hirst, J.M.: An automatic volumetric spore trap. Ann Appl Biol 39, 257-265 (1952)
11. Jiang, C., Wang, W., Du, L., Huang, G., McConaghy, C., Fineman, S., Liu, Y.: Field Evaluation of an Automated Pollen Sensor. Int. J. Environ. Res. Public Health 19, 6444 (2022)
12. Jin, B., Milling, M., Plaza, M.P, Brunner, J.O, Traidl-Hoffmann, C., Schuller, B.W., Damialis, A.: Airborne pollen grain detection from partially labelled data utilising semi-supervised learning. Sci Total Environ 891, 164295 (2023)
13. Kubera, E., Kubik-Komar, A., Kurasiński, P., Piotrowska-Weryszko, K., Skrzypiec, M.: Detection and recognition of pollen grains in multilabel microscopic images. Sensors 22(7), 2690 (2022)
14. Kubera, E., Kubik-Komar, A., Wieczorkowska, A., Piotrowska-Weryszko, K.,Kurasiński, P., Konarska, A.: Towards Automation of Pollen Monitoring: Image-Based Tree Pollen Recognition. In: Ceci, M., Flesca, S., Masciari, E., Manco, G., Raś, Z.W. (eds.) ISMIS 2022, LNAI, vol. 13515, pp. 219–229. Springer, Cham. https://doi.org/10.1007/978-3-031-16564-1 (2022)
15. Kubera, E., Wieczorkowska, A., Piotrowska-Weryszko, K., Konarska, A., Kubik-Komar, A.: Towards Automation of Pollen Monitoring - Dealing with the Background in Pollen Monitoring Images. In: NFMCP 2023 (to be published)
16. Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P.: COCO - Common Objects in Context. https://arxiv.org/abs/1405.0312 (2014)
17. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal Loss for Dense Object Detection. https://doi.org/10.48550/arXiv.1708.02002 (2018)
18. Mahmood, T., Choi, J., Park, K.R: Artificial intelligence-based classification of pollen grains using attention-guided pollen features aggregation network. J King Saud Univ Comput Inf Sci 35, 740-756 (2023)
19. Noguchi, K., Gel, Y.L., Brunner, E., Konietschke, F.: nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. J. Stat. Soft. 50(12), 1-23 (2012)
20. Object Detection on COCO test-dev, https://paperswithcode.com/sota/object-detection-on-coco. Last accessed 3 Feb 2024
21. Plaza, M.P., Kolek, F., Leier-Wirtz, V., Brunner, J.O., Traidl-Hoffmann, C., Damialis, A.: Detecting Airborne Pollen Using an Automatic, Real-Time Monitoring System: Evidence from Two Sites. Int J Environ Res Public Health 19(4) (2022)
22. Ultralytics YOLOv8 Docs, https://docs.ultralytics.com/. Last accessed 5 Feb 2024
23. Vasilevskaya, N.: Pollution of the Environment and Pollen: A Review. Stresses 2, 515-530 (2022)