

Evaluation of Training Strategies for *Ambrosia* Pollen Detection on Outdoor Images

Ivana Miličić¹, Nina Pajević¹, Dimitrije Stefanović¹, Marko Panić¹,
Branko Šikoparija², Roland Sarda-Estève³, Benjamin Guinot⁴, Nemanja Djurić⁵

¹Center for Information Technologies, BioSense Institute, Novi Sad, Serbia

²Center for Biosystems, BioSense Institute, Novi Sad, Serbia

³Laboratoire des Sciences du Climat et de l'Environnement (CEA/LSCE), Saint-Aubin, France

⁴OBERON Science, Villard-Bonno, France

⁵Aurora Innovation, Pittsburgh, PA, USA

Abstract—Automatic pollen detection has recently gained significant research attention due to its importance for public health. Numerous computer vision methods have been proposed for this task, with the majority relying on two-stage object detection models applied to optical microscopy images. However, two-stage approaches generally suffer from higher inference times, making them less suitable for real-time pollen monitoring tasks. Moreover, model performance on measurements from outdoor data samples remains largely unexplored because they introduce additional challenges, including visually complex backgrounds and the presence of other non-pollen particles such as dust, spores, and debris. In this work, we investigate the limits of the one-stage detection model YOLOv11 on *Ambrosia* pollen detection in images of outdoor and laboratory-prepared samples, acquired with the AeroTape device. We systematically evaluate training strategies, analyzing the effects of using either laboratory or outdoor images, their combinations, and also the use of data augmentation. The results show that incorporating augmented lab-prepared pollen images into training achieves the best mAP@50-95, highlighting the benefits of controlled laboratory data and one-stage detectors for improved generalization.

Index Terms—*Ambrosia* pollen, outdoor image, AeroTape

I. INTRODUCTION

Detection and quantification of pollen have traditionally relied on the standard volumetric method [1], which requires manual identification of each grain under a microscope. This procedure is highly labor-intensive and time-consuming, often causing delays of at least 36 hours before the data becomes available. Such latency is particularly problematic for highly allergenic taxa such as *Ambrosia*, which is known to trigger allergic rhinitis and asthma [2]. *Ambrosia* is typically active from August to October, with peak concentrations occurring mid-September. Timely response and appropriate therapy are therefore crucial for mitigating its health effects. Various automated approaches have been proposed to overcome these limitations, relying on optical microscopy, holographic imaging, or electrical signals [1], [3]. Most research primarily focused on classification tasks, employing both traditional machine learning algorithms, such as support vector machines

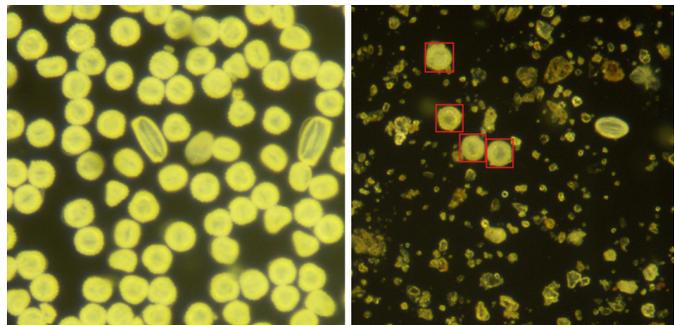


Figure 1: Images of *Ambrosia* pollen acquired in lab (left) and outdoor setting with pollen denoted in red (right)

[4], and deep learning models, including convolutional neural networks (CNNs) [5]. For detection tasks, two-stage models such as Faster R-CNN and RetinaNet have been investigated [4], [5]. Although these methods achieved promising accuracy (mAP@0.5 > 87%), their relatively long inference times limit their applicability for real-time monitoring. On the other hand, one-stage models such as YOLO [6] provide faster inference. However, to date, only a single study investigated the application of YOLOv7 for pollen detection [4], highlighting the need for further research.

The development and evaluation of deep learning models depend heavily on the availability, quantity, and quality of data. Publicly available pollen datasets are predominantly composed of images acquired from laboratory-prepared samples collected under controlled conditions [7]–[10]. However, it is unclear how the performance achieved under these idealized conditions translates to outdoor settings, where environmental variability introduces substantial complexity. To address this gap, a growing trend is emerging toward developing devices capable of quantifying pollen in outdoor environments [11]–[13]. In Figure 1, we illustrate the contrast between images acquired in laboratory and outdoor conditions with the AeroTape device developed by Oberon. In previous work, AeroTape was used with a CNN-based model to detect and quantify

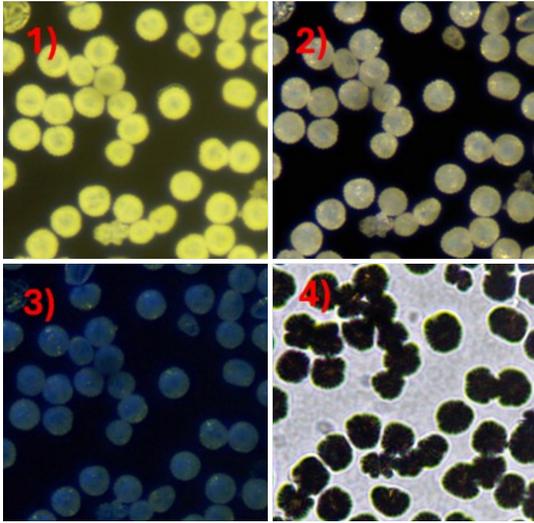


Figure 2: Images of *Ambrosia* pollen across four imaging modalities:(1) epi, (2) lat, (3) lat UV, (4) trans

Ambrosia pollen across multiple monitoring sites in the Lyon region [11]. The authors achieved high detection accuracy with limited false positives, enabling an analysis of the diurnal cycle and tracing the geographical origin of the *Ambrosia* pollen [11]. Our study explores the potential of YOLOv11 [14] for *Ambrosia* pollen detection in outdoor images acquired by the AeroTape device, focusing on this model due to its superior accuracy and speed compared to earlier YOLO versions [14].

II. METHODOLOGY

A. Dataset

Images used in this study were collected with AeroTape, a device equipped with an optical system providing $100\times$ magnification and capable of capturing images at a resolution of 2560×1920 pixels across four modalities [11]. AeroTape is similar to Pollen Sense, an existing commercial system for pollen monitoring, but offers higher spatial resolution [13]. The four modalities, illustrated in Figure 2, are as follows:

- 1) **Epi**: illumination at a right angle;
- 2) **Lateral (lat)**: illumination at an acute angle;
- 3) **Lateral UV (lat UV)**: illumination at an acute angle using UV light;
- 4) **Transversal (trans)**: illumination from beneath the pollen sample surface.

The laboratory dataset comprised all four aforementioned modalities of images acquired on laboratory-prepared samples of *Ambrosia* pollen, with the AeroTape device deployed in controlled laboratory conditions. Since outdoor images were available only in the epi modality, for consistency, we considered only epi images from the laboratory dataset in the following experiments. In total, 21 laboratory images of *Ambrosia* pollen grains were included. The outdoor dataset, in contrast, was acquired using the same device deployed outdoors in Beaurepaire, France, between July 31st and September 6th,

Table I: Dataset stats showing image and label counts

| Dataset | Train | Valid. | Test |
|------------|-------------|----------|------------|
| Laboratory | 303 / 4,303 | 24 / 623 | 59 / 1,351 |
| Outdoor | 705 / 772 | 44 / 46 | 66 / 72 |

2024, corresponding to the peak *Ambrosia* season. The acquisition procedure involved capturing pollen deposited on adhesive tape, illuminated with light beams of varying wavelengths and angles. A total of 1,393 outdoor images were obtained, of which 303 contained *Ambrosia* pollen grains. Images without *Ambrosia* were excluded from further analysis.

Both datasets were split into training, validation, and test sets in a 75 : 10 : 15 ratio. To address the limited dataset size and meet the input resolution required by the used detector, full-size images were divided into patches of 640×640 pixels. Validation and test patches were strictly non-overlapping, while training patches were allowed to overlap by 40% in both height and width. Outdoor images were manually labeled, while laboratory images were labeled using the Segment Anything Model (SAM) [15], followed by manual post-processing by a domain expert to fix errors and eliminate duplicates. The total counts of images and labels are summarized in Table I.

B. Experimental Setup

We used the "nano" version of YOLOv11 (YOLOv11n), containing 2.6 million parameters [16]. The training process was initiated from a model pretrained on the COCO dataset and run for 100 epochs. Data augmentation was applied dynamically during training, with the specific augmentation techniques selected through hyperparameter tuning during 50 epochs using the Ray Tune library [17]. Based on this configuration, the study evaluated the model under six training strategies, designed to systematically examine the effect of laboratory and outdoor images when used as training data:

- 1) **Strategy 1 (S1)**: Training on lab-prepared images;
- 2) **Strategy 2 (S2)**: S1 with data augmentation;
- 3) **Strategy 3 (S3)**: Training on outdoor images;
- 4) **Strategy 4 (S4)**: S3 with data augmentation;
- 5) **Strategy 5 (S5)**: Training on lab and outdoor images;
- 6) **Strategy 6 (S6)**: S5 with data augmentation.

Strategies S1 and S2 only included laboratory images in the validation set; S3 and S4 used outdoor images, while S5 and S6 utilized a mix of laboratory and outdoor images. All six strategies were evaluated on the same outdoor test set using precision, recall, and mean average precision with an IoU threshold set to 0.5 (mAP@50), and average mAP with a threshold in a range from 0.5 to 0.95 (mAP@50-95). In Table II, we show tuned hyperparameters of YOLOv11n for the appropriate validation sets. During the evaluation phase, the detection confidence threshold was set to 0.7.

III. RESULTS AND DISCUSSION

The model was trained under six strategies designed to examine which experimental setup yields the best performance. As seen in Table III, combining outdoor and laboratory images

Table II: Optimal training hyperparameters

| Parameters | Lab | Outdoor | Lab + Outdoor |
|--------------------|---------|---------|---------------|
| Learning rate (lr) | 0.00723 | 0.01085 | 0.01299 |
| Weight decay | 0.00045 | 0.00051 | 0.00032 |
| Warm-up epochs | 3.69422 | 4.50391 | 3.6279 |
| Optimizer | AdamW | AdamW | AdamW |
| Saturation | 0.9 | 0.83126 | 0.52659 |
| Value | 0.56883 | 0.75634 | 0.24081 |
| Horizontal flip | 0.29379 | 0.49852 | 0.29163 |
| Scale | 0.52098 | 0.40173 | 0.54584 |
| Mosaic | 0.90164 | 0.64924 | 1.0 |

Table III: Results on the outdoor image test set

| Strategy | mAP@50 | mAP@50-95 | Prec. | Recall | F1 |
|----------|---------------|---------------|---------------|---------------|---------------|
| S1 | 87.35% | 63.53% | 97.17% | 76.38% | 85.60% |
| S2 | 91.14% | 66.35% | 92.95% | 86.11% | 89.44% |
| S3 | 98.19% | 83.87% | 94.59% | 97.22% | 95.92% |
| S4 | 97.51% | 84.65% | 92.11% | 97.22% | 94.63% |
| S5 | 97.74% | 84.50% | 96.22% | 97.22% | 96.72% |
| S6 | 97.47% | 84.71% | 97.84% | 95.83% | 96.83% |

with additional augmentation (as per strategy S6) overall achieved the highest mAP@50-95 and F1 score. Examples of output detections for S6 are shown in Figure 3.

Data augmentation was investigated as a strategy to improve model performance. The same set of augmentation techniques was applied across all strategies. Color-based augmentations (saturation and value) were incorporated to account for simulated changes in brightness and contrast induced by sensor instabilities or changing illumination. Geometric augmentations (horizontal flip and scaling) were selected to introduce variability in object orientation and size. Mosaic augmentation was used to simulate complex backgrounds where pollen grains co-occur with other particles and noise, resembling outdoor conditions. The evaluation on the outdoor test set showed that augmentation contributed to an increase in mAP@50-95 across all considered strategies. The effect of data augmentation can also be observed in the F1 scores between Strategies 1 and 2, as well as between Strategies 5 and 6.

In Section I, we discussed concerns about the generalization of models trained exclusively on laboratory images when applied to outdoor environments. To investigate this, Strategies 1 and 2 were trained solely on laboratory data, with the only difference being that augmentation was applied in Strategy 2. As observed from the evaluation metrics and the confusion matrix (given in Table IV), Strategy 1 produced a considerable number of false positives, indicating that many background elements were incorrectly classified as *Ambrosia*. These results confirm that training on laboratory-only data limits the model performance in outdoor settings, as illustrated in Figure 3b showing the incorrect S1 predictions on an example outdoor test image. On the other hand, when only outdoor training images are used (strategies S3 and S4), we see a clear performance improvement (see mAP@50-95 and F1 scores in Table III), which is expected as in this setup there is no domain shift between train and test domains. However, we can see that combining laboratory and outdoor images (strategies S5 and S6) led to a further boost in performance. Moreover, using

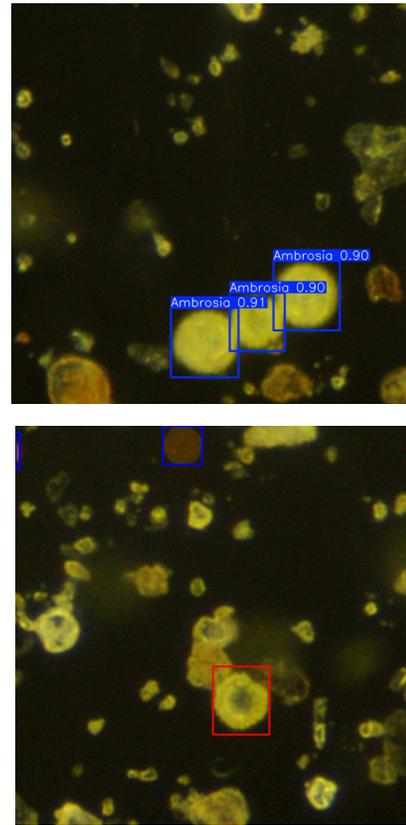


Figure 3: Examples of YOLOv11n detections: (a) correct detections using S6, (b) incorrect detection using S1; red boxes indicate ground truth, while blue represent predictions

this setup together with augmentation (as per S6) achieved the highest F1 and mAP@50-95, demonstrating that including augmented laboratory samples alongside outdoor images improves generalization to real-world conditions. Lastly, looking at the confusion matrices for strategies S1 and S6 (as shown in Tables IV and V, respectively), we observe a reduction in false positives and a higher number of true positives when using S6, albeit some false negatives remain.

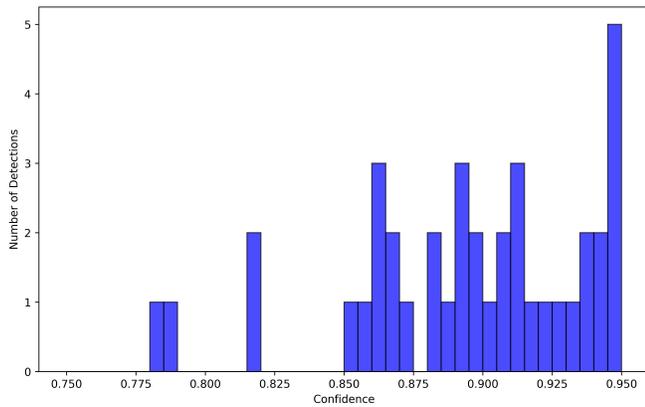
Table IV: Confusion matrix for S1

| | Ambrosia (Pred) | Background (Pred) |
|-------------------|-----------------|-------------------|
| Ambrosia (True) | 55 | 3 |
| Background (True) | 17 | 0 |

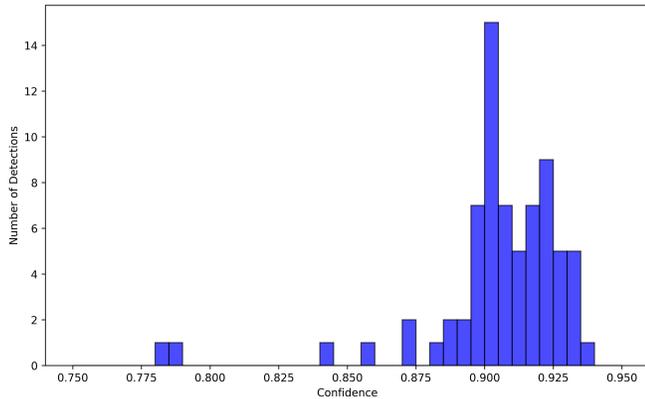
Table V: Confusion matrix for S6

| | Ambrosia (Pred) | Background (Pred) |
|-----------------|-----------------|-------------------|
| Ambrosia (GT) | 70 | 4 |
| Background (GT) | 2 | 0 |

To further examine model behavior, we generated confidence histograms for strategies S1 and S6, shown in Figure 4. The histogram for S1 shows a larger number of highly confident detections, clustered above 0.9. However, the confusion matrix (Table IV) indicates that many of these predictions are incorrect, with 17 false positives and only 55 true positives.



(a) Strategy S1



(b) Strategy S6

Figure 4: Distribution of confidence scores for S1 and S6

This highlights that a model trained solely on laboratory images tends to assign high confidence even to wrong detections in outdoor settings, leading to poor generalization. In contrast, S6 produces fewer detections above the 0.7 confidence threshold, with the histogram showing a more balanced distribution of confidence values. Yet, the confusion matrix (Table V) reveals a much better trade-off: 70 true positives and only 2 false positives. This indicates that including outdoor data during training leads to a better-calibrated model, providing more reliable detections. Hence, S6 demonstrates improved selectivity and robustness compared to the S1 strategy.

IV. CONCLUSION

This study evaluated the detection of *Ambrosia* pollen in outdoor images using the YOLOv11n model under six different training strategies, investigating the effect of augmentation and of using laboratory and outdoor images during training. The results highlight that incorporating lab images into the training set substantially enhanced model performance, underscoring the value of leveraging controlled data to mitigate the challenges posed by outdoor variability. Data augmentation also contributed positively to performance, as reflected by mAP@50-95; however, the combination of laboratory images and augmentation yielded the highest improvements. In gen-

eral, the results strongly suggest the importance of combining data sources with augmentation techniques to improve the detection performance in outdoor conditions.

ACKNOWLEDGMENT

The authors acknowledge the support of the EU Horizon Europe project SYLVA (grant agreement no. 101086109), as well as of the Ministry of Science of the Republic of Serbia, Technological Development and Innovations (grant agreement no. 451-03-136/2025-03/200358).

REFERENCES

- [1] S. Brdar, M. Panić, P. Matavulj *et al.*, “Explainable ai for unveiling deep learning pollen classification model based on fusion of scattered light patterns and fluorescence spectroscopy,” *Scientific Reports*, vol. 13, p. 3205, 2023.
- [2] S. S. S. Y. e. a. Schaffner, U., “Biological weed control to relieve millions from ambrosia allergies in europe,” *Nature Communications*, vol. 11, no. 1, p. 1745, 2020.
- [3] E. Sauvageat, Y. Zeder, K. Auderset, B. Calpini, B. Clot, B. Crouzy, T. Konzelmann, G. Lieberherr, F. Tummon, and K. Vasilatou, “Real-time pollen monitoring using digital holography,” *Atmospheric Measurement Techniques*, vol. 13, pp. 1539–1550, 2020.
- [4] A. Chaves, C. Martín, L. Torres *et al.*, “Pollen recognition through an open-source web-based system: automated particle counting for aerobiological analysis,” *Earth Science Informatics*, vol. 17, pp. 699–710, 2024.
- [5] R. Gallardo, C. García-Orellana, H. González-Velasco *et al.*, “Automated multifocus pollen detection using deep learning,” *Multimedia Tools and Applications*, vol. 83, pp. 72 097–72 112, 2024.
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [7] N. Khanzhina *et al.*, “Combating data incompetence in pollen images detection and classification for pollinosis prevention,” *Computers in Biology and Medicine*, vol. 140, p. 105064, 2022.
- [8] V. Sevillano and J. Aznarte, “Improving classification of pollen grain images of the polen23e dataset through three different applications of deep learning convolutional neural networks,” *PLoS ONE*, vol. 13, no. 9, p. e0201807, 2018.
- [9] G. Astolfi and A. Barbosa Gonçalves, “Pollen73s,” <https://doi.org/10.6084/m9.figshare.12536573.v1>, 2020, dataset.
- [10] S. Battiato, A. Ortis, F. Trenta, L. Ascari, M. Politi, and C. Siniscalco, “Pollen13k: A large scale microscope pollen grain image dataset,” in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, October 2020, pp. 2456–2460.
- [11] J. Truskina, V. Krotov, Y. Prat, D. Filippi, B. Šikoparija, D. O’Connor, B. Guinot, D. Baisnee, and R. Sarda-Esteve, “Real-time detection and quantification of ragweed pollen by optical microscopy and artificial intelligence,” 2024, unpublished.
- [12] “Bioscout website,” <https://www.bioscout.com.au/>.
- [13] Q. Farooq, R. López-Orozco, M. Martínez-Bracero, C. Galán, and J. Oteros, “Intercomparison studies and potential of pre-load algorithms for training and validating automated sampling systems in aerobiology,” Vienna, Austria, 2025, eGU25-4712.
- [14] R. Khanam and M. Hussain, “Yolov11: An overview of the key architectural enhancements,” 2024. [Online]. Available: <https://arxiv.org/abs/2410.17725>
- [15] A. Kirillov, E. Mintun, N. Ravi *et al.*, “Segment anything,” in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2023, pp. 3992–4003.
- [16] G. Jocher and J. Qiu, “Ultralytics yolo11,” 2024. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [17] European Committee for Standardization, “EN 16868:2019 - Ambient air - Sampling and analysis of airborne pollen grains and fungal spores for networks related to allergy - Volumetric Hirst method,” Standard, Brussels, Belgium, 2019, eN 16868:2019.