Universiteit
Leiden
The Netherlands

Bachelor Computer Science & Datascience and Artificial Intelligence

Investigation of existing pipeline
for lower magnification datasets

Linde Sluimer

Supervisors:
Lu Cao & Nemi Dorst

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

04/15/2026

Abstract

Accurate monitoring of airborne pollen concentrations is important for public health, particularly for the large share of the population that suffers from pollen-related allergies. Existing brightfield microscopy methods rely on labor intensive manual counting. Traditionally, plastic slides are used instead of now conventional glass slides in this study. However, these images appear blurry due to the lower refractive index of plastic. Reducing the magnification resolves the blur and significantly reduces the file size per slide.

A pre-trained YOLO model, originally trained on high magnification glass-slide images of *Betula pendula* (silver birch) pollen, is fine-tuned in two rounds for the new imaging conditions. The finetuned model is applied to a pollen dispersal experiment and accurately reproduces the expected decrease in pollen concentration with increasing distance from a birch tree. These findings demonstrate that a YOLO-based detection pipeline can be successfully adapted to a low-cost, lower magnification imaging setup.

Contents

1 Introduction

When spring arrives, and even as early as winter, large quantities of airborne pollen are released by trees across Europe. For a majority of the population, this triggers allergic rhinitis, commonly known as hay fever. [BWT+19]. In early spring, this is mainly caused by *Betula pendula* (silver birch), the most allergenic tree genus in Northern and Central Europe. The exposure period has been extending due to climate change driven warming [BWT+19]. Especially in cities, which are on the forefront of rising temperatures, pollen allergy symptoms among residents are on the rise, partly because interactions between air pollution and pollen grains may intensify allergenic effects [SVSM18]. Therefore, accurate monitoring of airborne pollen concentrations is increasingly important for public health.

Current monitoring relies on trained experts who manually examine microscopy slides under high magnification and count pollen grains by hand. This process is time consuming, and increasingly challenging due to the large numbers of slides produced by samplers such as the Pollensniffer introduced by de Weger et al. [dMd+20]. Automating this counting step using deep learning would remove this bottleneck.

Plastic slides are used instead of glass slides, however, plastic has a higher refractive index than glass, which causes a loss of resolution and fine detail at high magnification brightfield microscopy [Jon12], making the slides blurry at 400× magnified. The proposed solution is to reduce the microscope magnification to 100×, at which the resolution produces focused images of pollen through the plastic medium. This change also reduces the file size per slide from approximately 5 GB to 350 MB, saving storage space and accelerating model inference. This is per TIFF file, meaning that for the entire Z stack, this difference is 20 times greater. This significantly affects both the storage requirements for the data and the model's processing speed. However, a lower magnification could result in some details being lost, meaning that the model might be less able to detect small differences between pollen grains. Because the existing YOLO model was trained exclusively on 400× glass slide images, this magnification change places the new input data outside the model's training distribution and makes direct application unreliable. Fine-tuning is therefore required to improve model performance and to make the pipeline functional on this lower magnification dataset.

This leads to the research question of this thesis:

> *How can we improve a deep learning pipeline for pollen recognition for use on a lower magnification dataset?*

To answer this question, multiple objectives are pursued. The model will be evaluated using the images at 100× magnified. After assessing baseline performance of the pre-trained YOLO model under these conditions. The model will be Finetuned for the new magnification scale. Additionally we address a second challenge including blue-stained pollen and blue backgrounds. These blue situations originate from a border created on the slides. In some cases, the ink bled onto the pollen grains, causing them to turn blue. In other cases, the ink ran, resulting in parts of the slide having a blue background where the pollen grains still needed to be identified.

2 Background information

2.1 Pollen in Public Health

When spring arrives, many people start to suffer from hay fever. This is studied in the field of palynology, which is the study of pollen and spores. Pollen comes mainly from plants that are not pollinated by insects. These pollen are spread by the wind and thus end up in the air. Pollen in the air is considered the primary aeroallergen responsible for allergic reactions.
The birch homologous group consists of Birch and other related trees of the families Betulaceae and Fagaceae including, among others, hazel, oak and chestnut. Biedermann and Winther [BWT+19] have found that in Northern Central Europe, Birch is the most dominant tree pollen and is also a major cause of allergenic rhinitis and possibly asthma symptoms. Due to climate change, there has been a significant increase in the amount of birch pollen in recent years, and the period of exposure has also become longer.
The amount of green space in cities is decreasing however, pollen allergies and allergic respiratory symptoms are on the rise among urban residents, even more so than among people living in rural areas. The reason for this is that there is an interaction between chemical air pollutants and pollen grains. This can lead to various effects, such as an increased release of pollen allergens as found by Sedghy, Varasteh and Moghadam[SVSM18].

2.2 Pollen recognition

The field of palynology has a wide range of applications, however most pollen identification is still being done by well-trained experts. For this reason, several systems have been developed in recent years to automate the classification and counting of pollen. Most pollen are spherical and have a diameter of 18 to 60 µm as found by Hyde and Adams [HA58], making it uniform for sampling purposes. For example, Allen and Hodgson [AHMF08] have introduced a system that produces results similar to that of a human expert. Of the three pollen species examined, the classifier has identified 90% correctly.

2.2.1 Sampling pollen

Plants can use various methods to disperse their pollen. For higher plants this is done by the wind or insect vectors that transport pollen for pollination. About ten percent of higher plants use wind for pollination, and these are called anemophilous. When a sample is collected from the air, it must be analyzed microscopically.
The method used to sample the pollen is the portable Pollensniffer introduced by de Weger [dMd+20]. Prior to this technique, static samplers were used, such as the Cyclone sampler introduced by Emberlin [Emb95]. The portable pollen sampler is 5-6 times more efficient in collecting pollen than a static sampler. The pollensniffer uses a system similar to that of the classic Hirst pollen traps. A vacuum mechanism is used to draw the airborne particles onto a strip coated with petroleu1m jelly to trap pollen. The pollen are stained pink to enhance the visibility of the pollen. Following this, the slides are manually analyzed using brightfield microscopy.

2.2.2 Pollen identification

Several factors are important for identifying pollen. Most airborne pollen consists of single grains, known as monads. Sometimes there are multiple grains, with the most common being the 4-grain tetrad. Most wind pollinated grains are relatively the same size, between 20 to 60 µm. Pollen grains are mostly elliptical or spherical, however different shapes such as triangular or wedge-shapes exist. Surface structures are very important for identifying pollen, with pores being the most important. The mainly circular pores occur in varying numbers on the grains. These are most commonly monoporate and triporate. Pores themselves can also look very different in species, as found by Weber [Web98]. In addition to pores, pollen grains can also be recognized by their surface texture and internal characteristics. The outer wall, the exine, and the inner wall, the intine, can also distinguish different types of pollen.

2.2.3 Speckhofje

The dataset used for this project was provided by Nemi Dorst, the PhD student responsible for this project. The dataset contains seven slides from which TIFF files were created, and 20 TIFF files were created per slide in the z direction using a microscope. The slides are air samples from the Speckhofje street. There are many birch trees on this street, and one sample was taken directly under the tree, while the rest of the slides were taken 1 meter, 5 meters, 10 meters and 25 meters away from the tree. The last was taken on another road called Van t'Hoffstraat.

2.2.4 Slide material

These slides were made with plastic instead of glass. However, it was found that this resulted in blurry microscopic images. This is because glass has a lower refractive index (RI) than plastic. The RI of Pyrex glass is 1.47 and the RI of plastic is between 1.56 and 1.58. A lower refractive index results in better identification and photography, as a result of a decreased loss of fine detail and resolution, as found by Jones [Jon12].

2.3 Pollen types

In this study, we are mostly talking about three different types of allergenic pollen: *B. pendula*, *C. nootkatensis*, and *C. lawsoniana*.

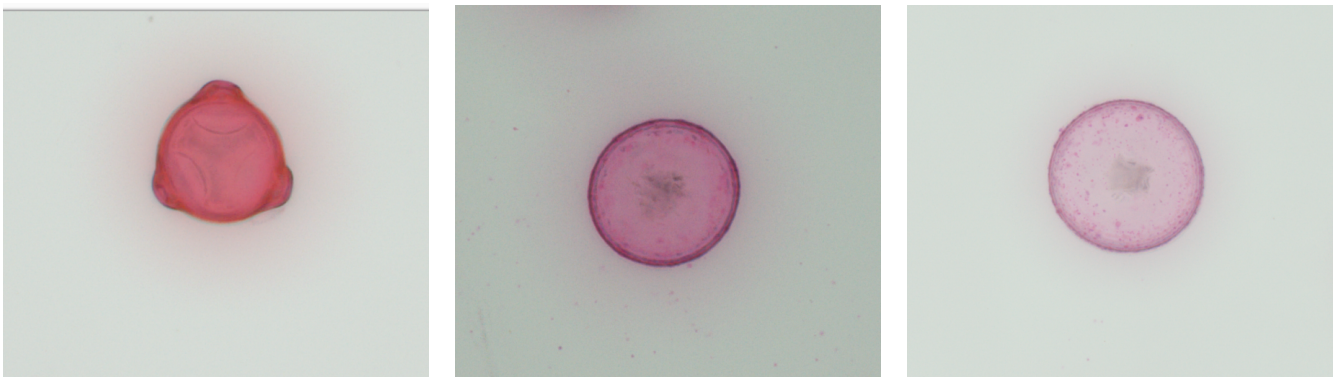
2.3.1 B. pendula

Betula is a genus comprising 50 species. One of these species is *Betula pendula*, also known as *B. pendula*. This pollen comes from the *Betula pendula* Roth tree, which can grow up to 30 metres tall. Birch is a major cause of allergic rhinitis, and recently levels of birch pollen have risen. In addition, the occurrence of pollen sensitivity has also increased shown by Biedermann and Winther [BWT+19] These Birch trees have a wide tolerance to climate and are tolerant to a wide range of soil conditions according to Atkinson [Atk92]. They generally start to flower around 5 to 10 years old. Gimingham found that [Gim84] birch trees have an irregular seed supply because they produce different quantities of seeds each year and the seeds have a short lifespan. Birch produces fruits even in the absence of pollination.

Birch has a very distinctive pore with a thickened surface around the pores as you can see in Figure 1. *B. pendula* is triporate and is between 20 and 35 µm in size as found by Weber [Web98].

2.3.2 C. nootkatensis and C. lawsoniana

The pollination process of Chamaecyparis nootkatensis occurs over a long period of time, where small amounts of pollen are shed at a time. With C. nootkatensis, precipitation is common during pollination. The pollen grains are on average 27 µm in diameter and have a distinct exine and thin intine as found in [OSM80].
The distributions of the Chamaecyparis lawsoniana species are most limited by water as found in [ZH80]. The Port Orford ceder has a paradoxical distribution and is restricted to limited geographical range, but occurring in many habitats. These two pollen grains are difficult to distinguish by sight. However, C. nootkatensis is slightly smaller than the pollen grains of C. lawsoniana.



(a) *B. pendula* (b) C. nootkatensis (c) C. lawsoniana

Figure 1: Three pollen types

2.4 You Only Look Once

The pipeline used in this project was created in advance by Casper Stiekema [Sti25] during his Master's thesis. This model is a YOLO v8 used for pollen recognition. YOLO is used to localize and classify the pollen.
In the field of object detection using deep learning, the You Only Look Once (YOLO) architecture resulted in rapid growth in it's field in 2016 [RDGF16]. YOLO has real-time high-classification performance and has limited but efficient computational parameters. In the field of object detection architectures, there are two distinct approaches: single-stage and two-stage detectors. In the case of two-stage detection, the process is divided into two stages: feature extraction, followed by regression and classification to produce the output. Whilst this can achieve high accuracy, it also places high computational demands on a system, making it less efficient for real-time detection [SI18]. YOLO is a single-stage detector, which means that feature extraction and regression happen in a single pass[DAT22]. Due to the bounding boxes and class probability being predicted concurrently.

2.4.1 Training data

The training data for this YOLO model was collected by gathering slides with one type of pollen. These are called catkin slides. These slides were collected by dragging catkins, the flower clusters of trees, over the slides, causing the pollen to land on them. This was done for all three types of pollen, which were then projected and annotated by hand, and used as training data for the YOLO v8 model.

2.5 Research objective

The research objective of this thesis is to determine how we can use a pre-trained YOLO model for pollen recognition so that blurry microscopic slides can still be utilised. These images were taken using plastic slides instead of glass slides, resulting in blurry slides. The way in which we aim to make these images sharp and usable is by reducing the magnification of the images. By changing the microscope from 400x magnification to 100x magnification, the images are no longer out of focus and their file size is reduced, which is very useful for storage. This reduces the file size per slide from 5 GB to 350 MB, also accelerating model inference. To address this issue, the model is optimized by fine-tuning it for this new situation. This leads to the research question: *"How can we improve a deep learning pipeline for pollen recognition for use on a low-quality dataset?"*

2.6 Thesis overview

This thesis is structured as follows. Chapter 2 provides the theoretical background on deep learning for pollen recognition. Chapter 4 outlines the research methodology, including research design, data collection methods, and data analysis techniques. Chapter 5 presents the results of the study. The findings are explained with figures and other useful illustrations. In Chapter 6, the implications of the findings are explored, and potential explanations are considered. Finally, in Sections 3 and 7, related work and further research related to this project are discussed.

3 Related Work

3.1 Plastic slides

The replacement glass slides with plastic alternatives in brightfield microscopy is of significant practical interest. Various slide and coverslip materials are being researched, but it is a matter of careful consideration between cost, convenience, and optical performance. Glass slides are traditionally used for microscopic slides, but plastic is also an alternative. Plastic is less prone to break than glass, and there are some optical plastics that can be just as good as glass. However, plastic has lower optical clarity than glass and easily scratches as shown by Lawlor [Law19]. Other materials are also being researched to optimize costs and outcomes. Ferreira and Vale [FD22] compare glass, film, and liquid coverslipping methods in order to evaluate their performance.

3.2 Deep learning for pollen recognition

In recent years, there has been significant progress in deep learning, including in combination with pollen identification. Several studies have already been conducted in this area. However, there are limitations to consider. De Geus [GBBS19], for example, faced limitations due to the size of the dataset and taxonomic diversity, but due to recent efforts, an increasing number of datasets covering various species are becoming available. Zhang and Mao [ZM26] has introduced a high-quality dataset comprising more than 140 species with 20 standardized morphological features.
The YOLO architecture is also used as an effective tool for pollen recognition. For example, PollenYOLO was developed by Shi and Hou [SHWL26] to identify pollen in fossil assemblages on the Tibetan Plateau using YOLOv8. It achieved high performance despite challenges such as complex backgrounds.
In addition to YOLO, various other deep learning models are also used to automate pollen recognition. Stiekema [Sti25], for example, compared YOLO with RetinaNet and DINO, on which this study is based. YOLO proved to be the best option. Other models, such as a deep CNN (convolutional neural network), are also used, as demonstrated by Kubera [KKKPWS21]. Another example is Explainable AI, which is used by Shamrat[SIZ24], in addition to this, various other models are used.

4 Methods

The following section briefly explains the methods used in this project. It describes how the data is collected into slides and processed so that it can ultimately be used as input data for the model. It also explains how the blurry data can still be used and to what extent the model needs to be adapted for this purpose.
To accomplish all of this, NVIDIA CUDA [NVI26] was used to harness the computing power of the GPUs. This was done on a Linux computer to facilitate the implementation of various packages to improve YOLO models using CUDA.

4.1 Sample retrieval

Pollensniffer samples were taken at different distances from Himalayan birch (*Betula utilis*) trees at various locations and days in Leiden. 69 slides were collected in total, of which 7 were scanned. Six of these were collected in Speckhofje (Figure 2b), a private courtyard with one birch tree. The last slide was collected in Van 't Hoffstraat, where the street is lined with birch trees, of which three were in bloom at the time. Sampling locations are visualized in Figure 2a, where Speckhofje is in the center of Leiden and Van 't Hoffstraat is south of the city center. The samples were taken with the Pollensniffer in April 2025.

(a) Leiden sample locations

(b) Speckhofje sample distances

Figure 2: Sampling locations

4.2 Data preparation

To counteract the blurriness of the images at 400x magnified, all slides for this experiment were scanned by a brightfield microscopy scanner, Zeiss Axioscan 7, at 100x magnified. This resulted in a magnification that was four times lower, meaning fewer fine details are visible, but enough

information is retained for object detection. The reduction also ensures that the size of the images drops from around 5 GB per slide to approximately 350 MB. This is a major advantage for slide storage and improves the model's processing speed.

4.3 Preprocessing

The data consists of microscopic slides with a z-stack of 20. This means that multiple 2D photos were taken with the microscope at different focal planes, depths, along the z-axis. These scans were saved as TIFF files for high resolution and small data loss. Before a slide can be used as input in the existing YOLOv8 model, it was first projected. This means that this z-stack of 20 was projected so that it becomes one sharp image. This is a minimum projection, where the lowest voxel is selection from the 20 depth layers for each pixel, and one output image is created. The minimum projection was used, since Casper Stiekema [Sti25] showed that the minimum projection performs best in this case comparing to grayscale, standard deviation and extended focus.
In some cases, the slide scanner had scanned the slide and saved it in tiles. If this was the case, the tiled TIFF files were merged into complete slides before they could be projected.

4.4 Evaluate model

The performance of the model on smaller images had to be evaluated before determining what adjustments were needed to make it work better with the new magnification. The model's output was compared with manual birch pollen counts, based on the number of *Betula* pollen grains identified by an expert palynologist. The model was also visually assessed to check for any common misidentifications.

4.5 Finetuning

To prepare the ground truth data for the model, pollen grains were manually annotated with LabelImg [Tzu15]. Bounding boxes were be annotated on the tiles for each pixel. One of three classes will be assigned to each pixel so that this data can be used to fine-tune the model. The image tiles, in combination with the corresponding bounding box data, were be split 70/30. This means that 70% of the tiles become training data and 30% would become validation data. The fine-tuning was done with 1-fold training using around 1300 pollen from a slide as data. Optuna was used to optimize the hyperparameters as seen in Table 1. To test the results of the fine-tuning, the other slides were used in conjunction with metrics to evaluate this.

4.6 Model comparison

The performance of the model and different finetuned versions on smaller images were evaluated. This was done by comparing the results of manually counting the pollen with the number of classified pollen grains per slide to evaluate the model's accuracy. In addition, visual comparison was used to evaluate the results and observing how big the difference in pollen recognition is for the model. The models were also compared using performance metrics for YOLO models.

Parameter	Value
Optimizer	AdamW
Initial learning rate	0.001
LR schedule	Cosine annealing
Weight decay	0.0005
Warmup epochs	5
Batch size	4
Image size	1024×1024
Max epochs	100
Early stopping patience	20

Table 1: Hyperparameters used for YOLOv8 fine-tuning.

These metrics include Precision, which indicates how accurately objects were detected.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall measures the model's ability to identify all instances of objects in the images. F1 is the harmonic mean of precision and recall.

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F\_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

Mean Average Precision reflects the model's accuracy in identifying and classifying objects within an image. This was calculated using IoU, which calculates the area of overlap between two bounding boxes of the ground truth and the prediction. And the AP, which is the area under the precision-recall curve, was calculated at a specific threshold.

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}$$

$$\text{AP} = \frac{1}{11} \sum_{t \in \{0,\, 0.1,\, \ldots,\, 1.0\}} \max_{\tilde{r} \geq t} P(\tilde{r})$$

$$\text{mAP}_{50} = \frac{1}{C} \sum_{c=1}^{C} \text{AP}_c \Big|_{\text{IoU}=0.5}$$

$$\text{mAP}_{50:95} = \frac{1}{10} \sum_{t \in \{0.50,\, 0.55,\, \ldots,\, 0.95\}} \text{mAP@}t$$

5 Results

The following section presents the results of how the model is adapted to work with smaller datasets and optimized for this purpose.

5.1 Projections

Per slide the minimum projections were created from the 20 TIFF files in the z-direction. The result of one of the slides can be seen in Figure 3. The corners of the slides have not been saved because the manually counted pollen grains are only counted within the blue circle, thereby keeping the size of the TIFF files as small as possible.

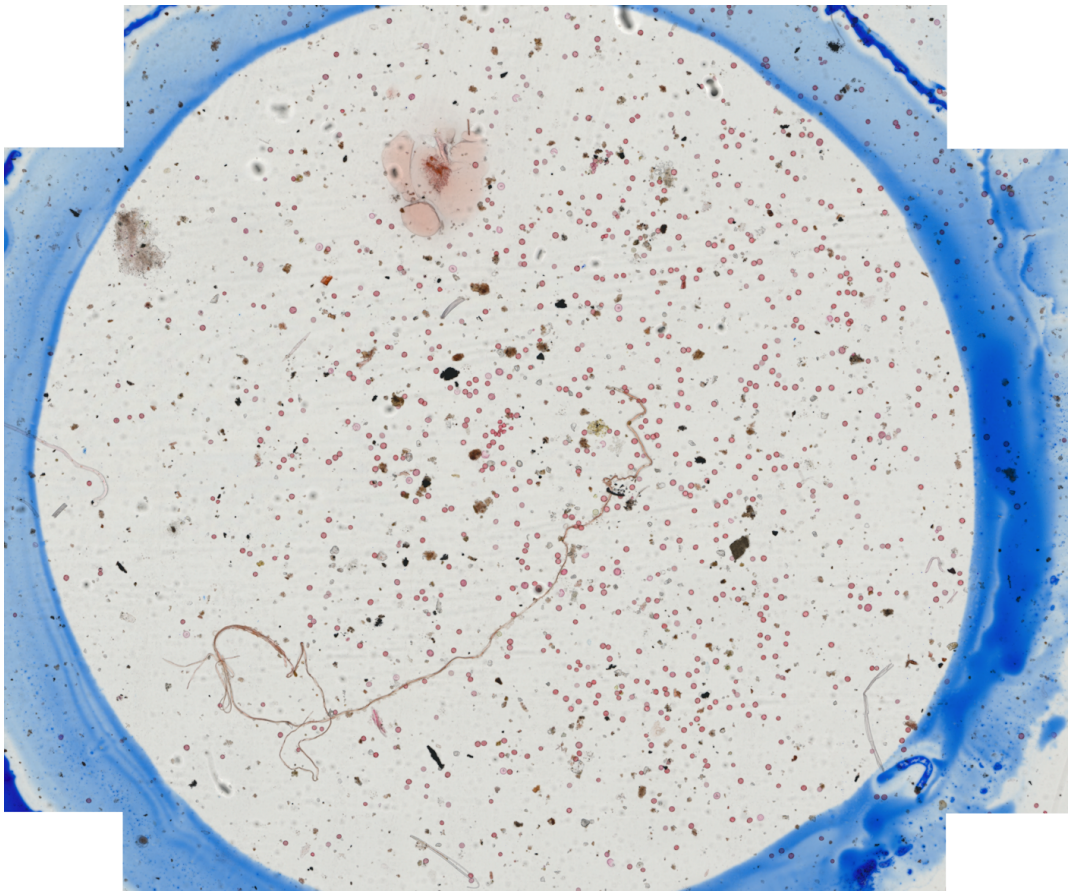


Figure 3: Minimum projection of Slide 209

5.2 Model evaluation

The projections of the TIFF files were used as input for the model. The model was used without any fine-tuning or post-processing to assess how it performs in the new situation with images reduced from 400x magnified to 100x. Figure 4 shows a comparison of the model's output in the new magnification scale with the weekly manually counted number of *B. pendula* present on those slides. The difference between the manual count and the model's prediction varies considerably

from slide to slide. The slides are labeled with an 'm' to indicate the distance from the tree and a 'd' to indicate the day on which the sample was taken. The street sample was taken in the other street, Van 't Hoffstraat. The difference is not consistent and does not indicate whether the model works well or identifies a large number of *B. pendula*. On average, the model identifies 29% more pollen as *B. pendula* than the actual amount. This means that more objects are recognized as pollen and therefore incorrectly identified.

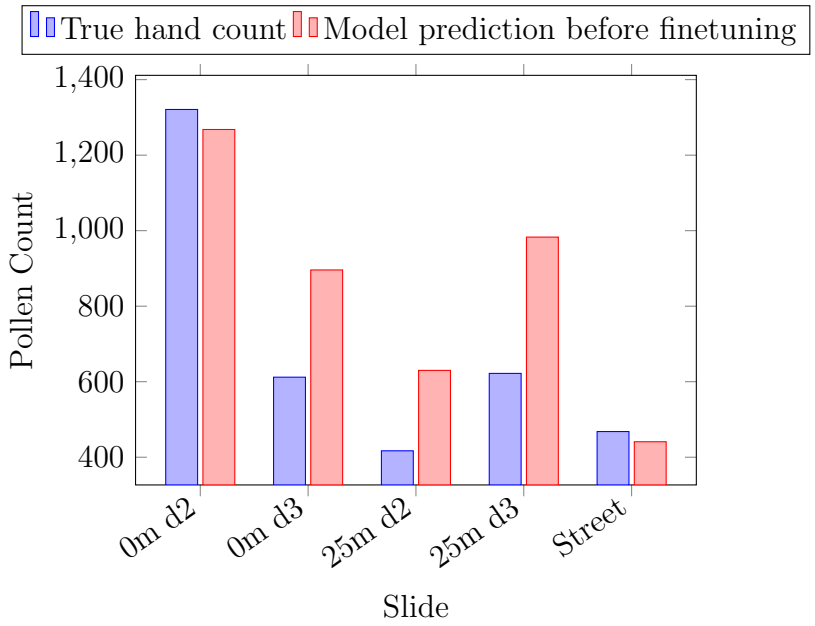


Figure 4: Manual pollen counts compared to the model predictions before finetuning.

Figure 5 shows a zoomed-in section of the slide, which displays the output of the model in the new situation. This output demonstrates that the model identifies debris, brown or white objects, as pollen. Some other types of pollen are misidentified as *B. pendula*. It also shows that some *B. pendula* pollen grains are not identified, particularly the smaller ones. In some cases pollen grains that are correctly identified have a lower confidence score than debris.
From these results, we concluded that the model needed to be improved. Small pollen grains needed to be recognized more frequently, whereas debris and other types of pollen needed to be identified less frequently as *B. pendula*. This meant that the model should be fine-tuned so that it could recognize the pollen grains in the new magnification conditions.

5.3 Model finetuning

5.3.1 New magnification

The model was fine-tuned using the new magnification data by converting a single slide into training and validation data. This slide was divided into 240 images, which were split 70/30, resulting in 168 training images and 72 validation images.
The output of the YOLO model following the fine-tuning is shown in Figure 6b and can be visually compared with Figure 6a before the finetuning to see the improvement in the model. As can be seen, the number of debris objects identified as clusters has decreased, and more small clusters are

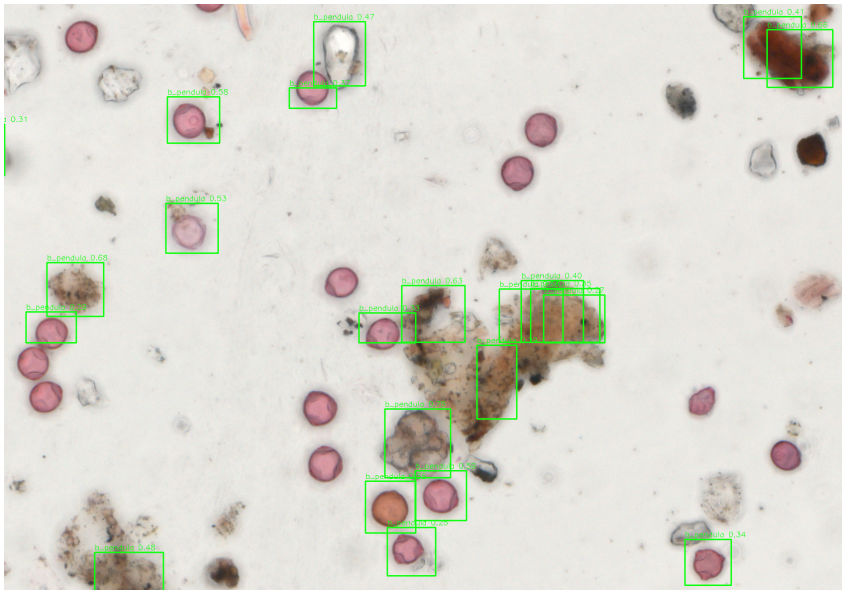
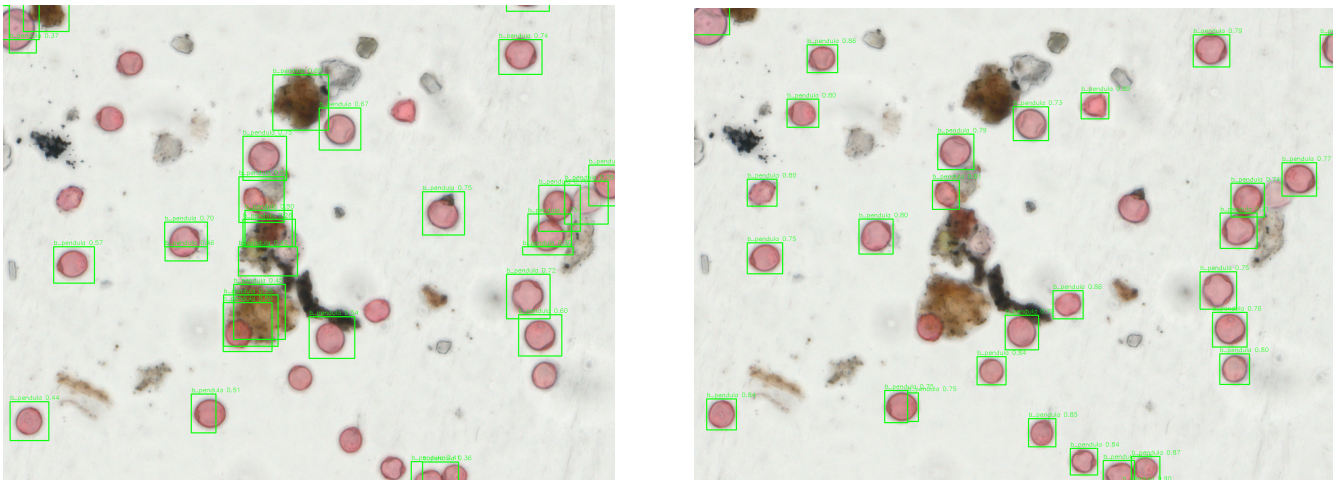


Figure 5: Model output Slide 130 before finetuning

now being recognized. After fine-tuning on the new magnification data, the amount of overlapping bounding boxes has also decreased, due to the model learning what pollen look like in the new magnification, placing a more confident, well-localized box per grain rather than several uncertain ones.



(a) Slide 200 before finetuning

(b) Slide 200 after finetuning

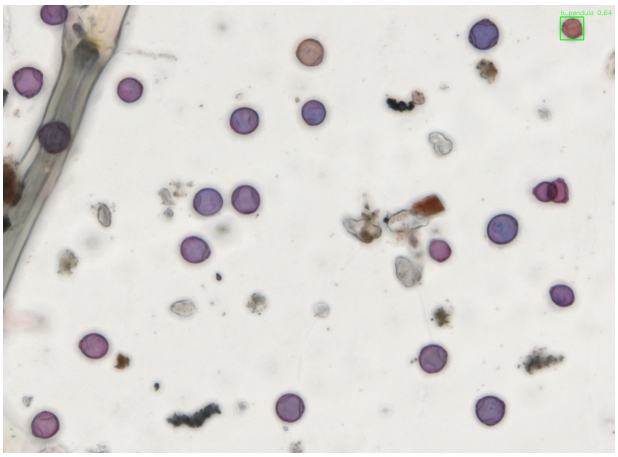
Figure 6: Visual comparison of the output of the model before and after finetuning

After analyzing the output data following fine-tuning with one slide, we concluded that the model had improved, however a number of additional scenarios needed to be included in order to optimize the model. On a number of slides, the blue circle, intended to indicate a boundary within which the pollen grains should be recognized, overlapped some of the pollen grains. These pollen grains were not identified by the model, as were slides where the circle had faded and had therefore partially

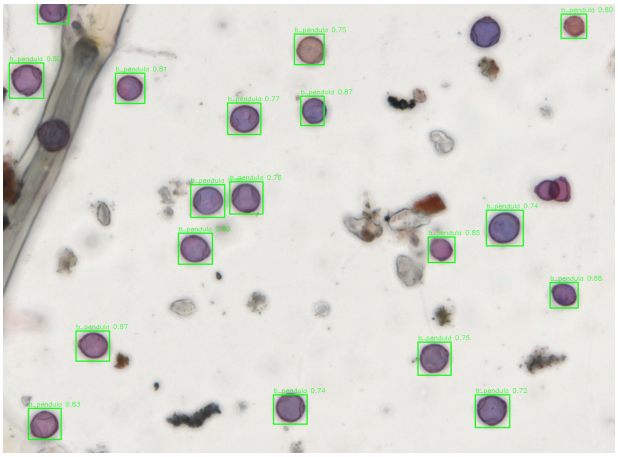
blended into the background. Both of these situations need to be taken into account during the fine-tuning process, as they occur frequently. Therefore, a second round of fine-tuning was required.

5.3.2 Expanded with blue

After fine-tuning the model for the new situation, it was also necessary to add additional scenarios to improve the model. The blue pollen on certain slides were not recognized and had to be included in the fine-tuning. Ten tiles were added where the background has light blue stripes and 30 tiles of images containing blue backgrounds. This was combined with the training and validation data from the fine-tuning, resulting in 196 training images and 84 validation images.
Figure 7a shows a comparison of the model's output in the scenario where fine-tuning was performed solely on the new magnification. Figure 7b illustrates how the model has improved following the inclusion of the blue scenarios during fine-tuning, resulting in the recognition of more blue pollen being identified as *B. pendula*.



(a) Slide 130 after finetuning in the new magnification



(b) Slide 130 after finetuning with blue pollen and background

Figure 7: Visual comparison of the output of the finetuned model with new magnification and expanded with blue

The confidence threshold was determined for the fine-tuned model by identifying the confidence threshold at which the F1 score was highest. This was the case at 0.8. However, after combining this with visual inspection, it was concluded that at a confidence threshold of 0.8, many blue pollen grains were no longer identified. At a confidence threshold of 0.6, pollen from other species and objects that were not *B. pendula* were identified, hence it was concluded that 0.7 is the best confidence threshold for this fine-tuned model. This ensured the greatest possible retention of the blue pollen identified, whilst minimizing the identification of debris as *B. pendula*.
The difference between the manually counted total and the model's count is shown in Figure 8. In general, the model identifies more *B. pendula* than are actually present on the slides, on average 7.1% more. Following visual inspection, it was concluded that this is because, in some cases, a pollen grain has two bounding boxes. Other pollen species are also sometimes misidentified as *B. pendula*.

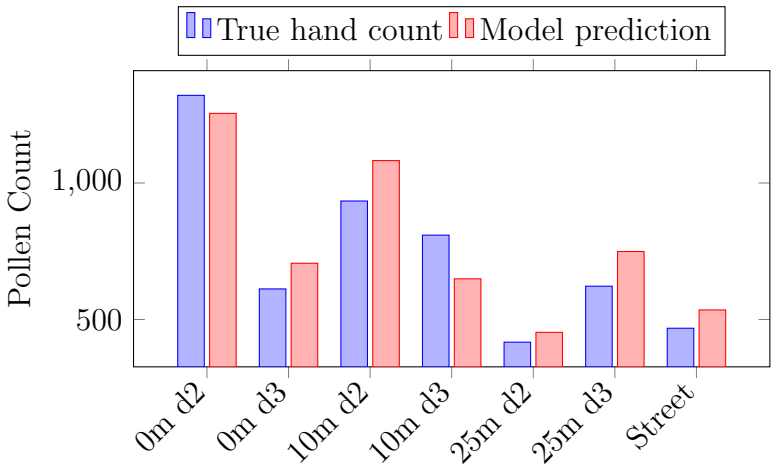


Figure 8: Hand counted pollen compared to the model output

Figure 9 shows the detection density averaged across all slides. These detections occurred mainly within the innermost circle of the slides. This indicates that, apart from a few outliers, the model detects objects within the blue border and hardly at all on or outside it.

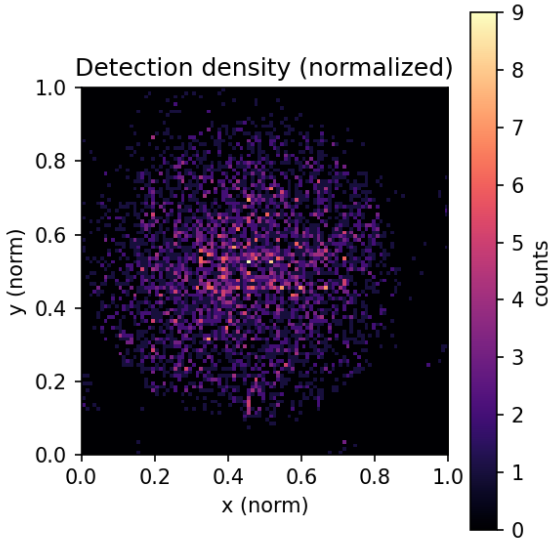


Figure 9: Detection density(normalized)

5.4 Model comparison

The comparison of the finetuned YOLO models is shown in Table 2 and visually in Figure 10. The model is essentially non functional on the data without any finetuning. A precision of 0.039 means that around 1 in 25 detections is actually a correct *B. pendula* grain and the rest are debris, other species, or noise. One round of fine-tuning on a single slide produces a dramatic jump across every

metric. However, precision and recall are nearly equal and are around 0.36, which indicates the model is still mediocre on a dataset that includes blue scenarios. When blue scenarios were added in the finetuning dataset, all evaluation metrics improved. This means that more predictions are correct and there are fewer false positives. The recall is nearly 1, which means that almost all pollen grains are detected. The F1 score is a combination of these two, and it has more than doubled, indicating that the model has improved. The average IoU shows that the bounding boxes are drawn accurately around the pollen grains. The mAPs have both improved slightly, but they were already good after a single round of fine-tuning. By adding the blue scenarios to the training data, the model has improved because it can now recognize *B. pendula* in a wider variety of situations.

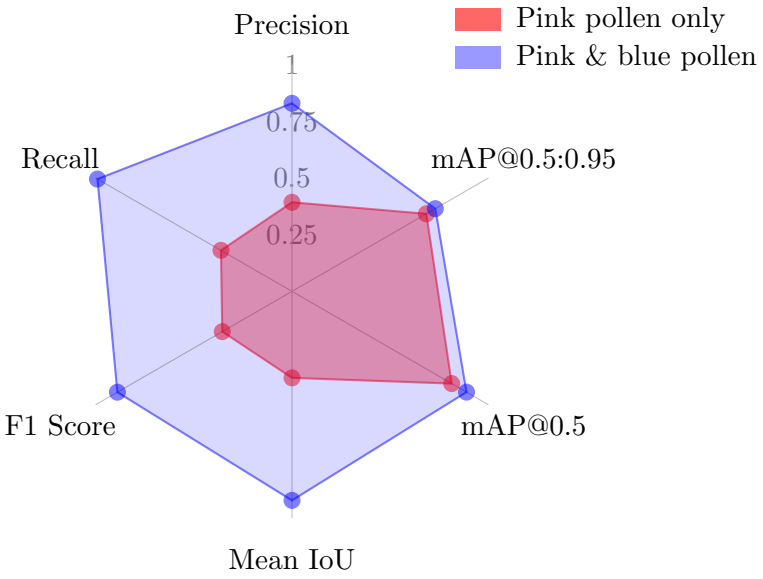


Figure 10: Radar chart with model's evaluation metrics

Model	Precision	Recall	F1 Score	Mean IoU	mAP50	mAP50-95
Pretrained YOLOv8	0.0388	0.1659	0.0579	0.1533	0.2229	0.1733
Finetuned with pink pollen	0.3929	0.3619	0.3551	0.3815	0.8134	0.6841
Finetuned with pink and blue	0.8293	0.9905	0.8892	0.9216	0.8885	0.7295

Table 2: Model performances compared

Visually the model has also improved, by recognizing more blue-colored *B. pendula*. If you look at Figures 6 and 7, you can see that the model has improved in both cases because it now recognizes more *B. pendula*. In both cases, the fine-tuning has led to improvements where they were needed. To further compare the models figure 11 shows how the models perform compared to the actual ground truth. Without fine-tuning, the model exhibits many unpredictable discrepancies. Sometimes the pollen count is very high due to the large amount of debris being misidentified as pollen, as the model is not accustomed to noisy input. At other times, the count is much lower because the model is not yet familiar with the new situations. After the first round of fine-tuning, this has already become much less extreme. However, it can be seen that on slide Street; at Van t'Hoffstraat, almost half of the pollen grains are recognized. There are many blue pollen grains on this slide.

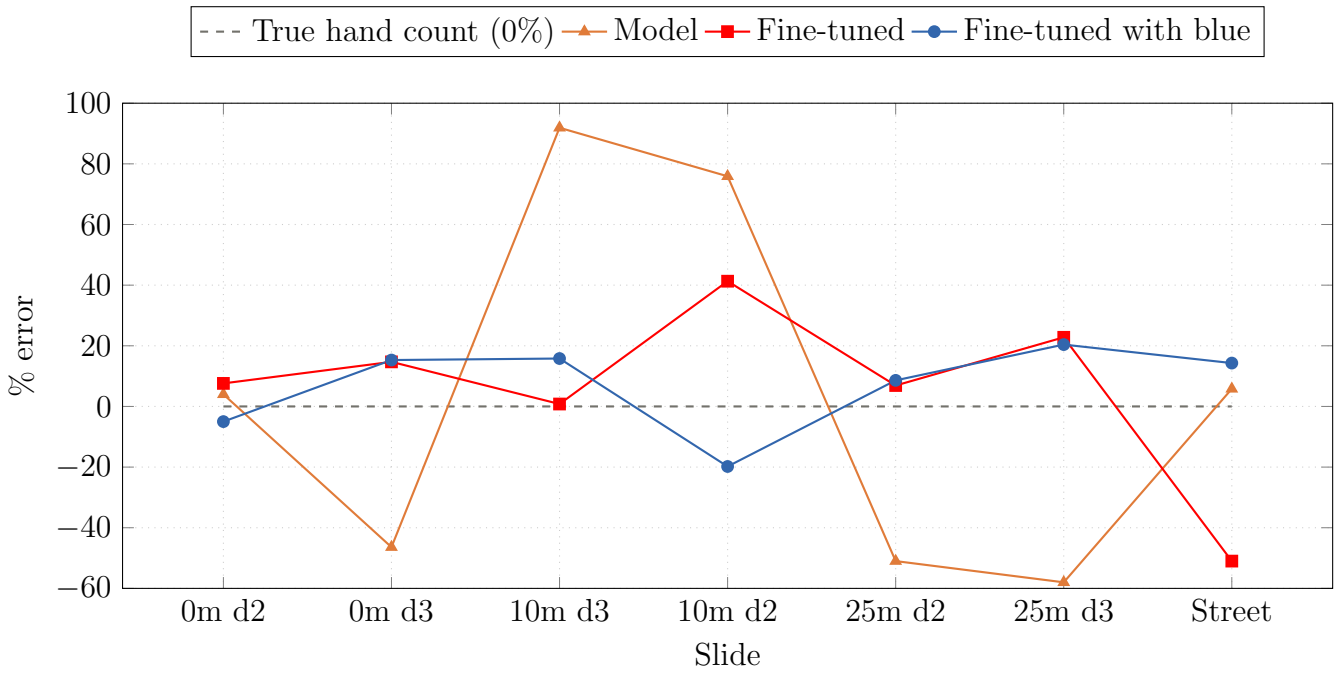


Figure 11: error percentage relative to true count

We can therefore conclude that the recognition of the blue pollen grains has improved by adding the blue scenarios to the fine-tuning. If we compare the mean of absolute error values of the two fine-tuned models, we see that this has decreased from 20.7% to 14.2%. This means that the model has improved by adding the extra blue scenarios to the training data.

The results of the pollen experiment, showing the amount of *Betula pendula* pollen present at various distances from a birch tree, are shown in Figure 12. On both days, the locations, measured in meters from the tree, are to the east of the tree. It can be seen that on average, the pollen count decreases with distance from the tree, based on the pollen counts recorded on the slides by the fine-tuned model. It is evident that the pollen count decreases as the distance from the tree increases. This is evident from the trend in the true manual pollen count, as well as from the model's results. We could have drawn different conclusions about Day 3, for example. Based on the model, one might observe that the highest pollen concentration was 25 meters from the tree, whereas in reality it was at 10 meters. On day 2, the wind blew from west to east at 5 to 7 km/h. A day later, on day 3, the wind was lighter, at 3 to 5 km/h, and blew from the south-west to the north-east.

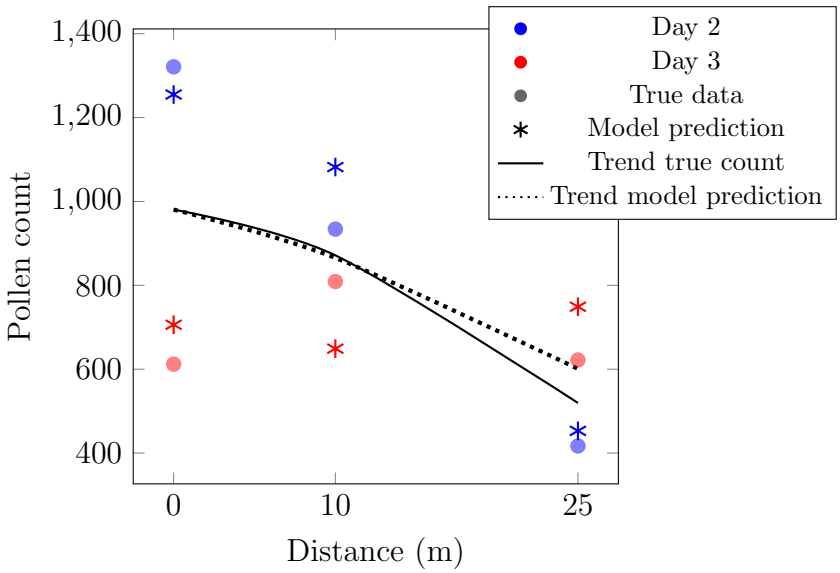


Figure 12: Pollen counts at different distances: true vs. predicted per day with average trends

6 Discussion

6.1 Model performance

The results of this study show that an existing YOLO model trained on pollen images on a glass slide at 400× magnified cannot be directly applied to a 100× magnification of a plastic slide. The change in slide material , but resulted in optical blur at high magnification. Reducing the magnification from 400× to 100× resolved the blurring issue and simultaneously reduced the file size per slide from approximately 5 GB to 350 MB. However, this meant that the input data differed significantly from the original distribution during training. The training images contained pollen grains without noise and debris, resulting in an average over count of approximately 29%, driven primarily by misidentification of debris and other types of pollen as *B. pendula*.
This is the result that we expected from the model when the images deviate from the training distribution. The high level of noise, due to the fact that these data were captured from the air rather than from a single catkin. This together with the new magnification situation where pollen grains are smaller relative to the image frame, results in a less accurate model. Consequently, fine-tuning was not only necessary for optimization, but also a requirement to get the pipeline functioning with the lower magnification dataset.

6.1.1 Finetuning

The first round of fine-tuning was conducted using training and validation data derived from a single 100× magnified slide, which produced a notable improvement in the model's ability to distinguish genuine *B. pendula* pollen grains. The visual inspection of output images confirmed that debris clusters were less frequently assigned bounding boxes, and that smaller pollen grains, previously missed entirely, were more reliably detected.
The second round of fine-tuning, which incorporated slides with blue-stained pollen and light blue background stripes, due to bleeding or staining ink, produced further gains across all evaluation

metrics. In particular, recall reaching excellence (0.991), indicating that nearly all pollen grains present on the slides were detected by the model. The F1 score increased from 0.355 to 0.889, and mean IoU rose from 0.382 to 0.922. The confidence threshold of 0.7 was identified as optimal through a combination of F1-score analysis and visual inspection, balancing the retention of blue-stained pollen against the suppression of false positives from other species and debris. The most notable discrepancy between model and manual counts occurred at the Street location (Van 't Hoffstraat), which contains a high proportion of blue-stained pollen. Prior to the second round of fine-tuning, the model substantially under counted at this location. Following the inclusion of blue scenarios in the training data, the performance improved noticeably.
Despite these improvements, the fine-tuned model still overestimated pollen counts by an average of 7.1% relative to manual counts. Visual inspection suggested two principal sources of error: first, individual pollen grains occasionally received two overlapping bounding boxes, second, non-*B. pendula* pollen species, e.g. *Carpinus*, with similarities to the target species were occasionally misclassified. The first of these problems can be partially resolved by applying post-processing and combining two bounding boxes that are close together, while the second points to the inherent challenge of distinguishing morphologically similar pollen types under low magnification.

6.1.2 Pollen dispersal experiment

The application of the fine-tuned model to the dispersal experiment yielded results broadly consistent with expectations. On both sampling days, pollen counts tended to decrease with increasing distance from the source tree, mirroring the trend observed in the manual counts. Day 2, characterized by stronger westerly winds (5–7 km/h), showed a steeper gradient and closer agreement between model predictions and manual counts than Day 3, where lighter and more variable wind conditions (3–5 km/h, south-westerly) produced a less predictable distribution. Based on the model's predictions, it appears on day 3 that the highest count of B. pendula pollen was recorded 25 meters east of the tree. This suggests that the quantity increases the further away from the tree, but this is not the case when looking at the actual manually counted quantities. Looking at the trends over these two days, it appears that the predictions largely correspond to the actual data, however, this is not exactly the case. This is because data at the same distances cancel each other out, as one is predicted to be higher than the actual figure and the other lower.
It should be noted that the dispersal experiment in this study was conducted with a limited number of sampling points and over only two days, which constrains the conclusions that can be drawn from the spatial trend. Nonetheless, the directional consistency between model predictions and ground truth provides encouraging evidence that automated counting is sufficiently reliable for use in exploratory pollen detection experiments.

6.2 Limitations

Several limitations of the study need to be acknowledged. First, the size of the fine-tuning dataset was comparatively small, the finetuning dataset comprised 280 images consisting of 196 training and 84 validation images. Although the data suggests that the model is sufficient for identifying B. pendula, more data is needed to confirm this with certainty. Second, the evaluation was conducted on slides from a single geographic location and a single B. pendula tree; it is unclear how the model would perform with pollen sourced from trees in different environments.

The model was fine-tuned using data in the new magnification. This training data contained hardly any pollen types other than *B. pendula*, less than 1%. As a result, after fine-tuning, the model only recognizes pollen as *B. pendula* and not the other two types, even though these classes still exist in the model.
Another limitation of the model is that, in some cases, a single pollen grain is assigned two different bounding boxes. This results in a higher pollen count than is actually the case.

7 Further Research

Several directions could extend and improve the work presented in this thesis.

7.1 Expanding the training dataset

The fine-tuning dataset comprised 280 images derived from a small number of slides from two locations. While the results demonstrate clear improvement, a larger and more diverse training set would provide a more reliable model performance. Future work should incorporate slides from different locations and seasonal time points. In order to draw more accurate conclusions from the results, more data points are needed. This will allow for a more accurate comparison between the model's predictions and the manually counted data. Cross-validation across multiple slides, rather than a fixed 70/30 split from a single slide, would give a more robust measure of how well the model performs on genuinely unseen data and reduce the risk of overfitting to the appearance of a particular slide.
The dataset of slides on which the model is run can also be expanded. By doing so, a better well founded conclusion could be drawn as to whether the model can produce comparable results for the pollen experiment and whether it could be used to make predictions.

7.2 Post-processing

Visual inspection revealed that individual pollen grains occasionally receive two bounding boxes, which inflates the total count. A post-processing step that merges bounding boxes whose centers fall within a set distance would suppress these duplicates without requiring additional training data or model changes. The distance threshold could be tuned on a hold-out validation set. This is a low effort solution that could meaningfully reduce the remaining 14.2% mean of absolute error values of the fine-tuned model.

7.3 Multi-species detection

The current model detects only *B. pendula, C. lawsoniana,* and C. nootkatensis.The current model has been trained to recognize three different types of pollen. However, the training data used to fine-tune the model consisted of more than 99.9% *B. pendula* pollen grains. As a result, the model fails to recognize the other two types of pollen in the new magnification, and these are also rarely found in the current test data. Expanding the fine-tuned model to detect and classify multiple allergenic pollen species would greatly increase its usability.

7.4 Alternative projection strategies

The minimum z-projection was selected because it performed best in the original pipeline [Sti25]. However, that finding was established for 400× glass-slide data, and the optimal projection strategy may differ at 100× through plastic. Alternative approaches such as extended depth of focus or maximum intensity projection could be evaluated to determine whether a different strategy better preserves the morphological features of the pollen grains on which the model relies for reliable detection at lower magnification.

7.5 Conclusion

A pipeline can be improved for lower magnification by finetuning. Further optimization can be achieved by expanding the training data to cover visual variety in the new magnification such as debris and different colored pollen. It was shown that finetuning on a single slide is sufficient enough to make the model functional, however full diversity needs to be covered for the model to be working well.

References

[AHMF08] G. P. Allen, R. M. Hodgson, S. R. Marsland, and J. R. Flenley. Machine vision for automated optical recognition and classification of pollen grains or other singulated microscopic objects. In *2008 15th International Conference on Mechatronics and Machine Vision in Practice*, pages 221–226, 2008.

[Atk92] M. D. Atkinson. Betula pendula roth (b. verrucosa ehrh.) and b. pubescens ehrh. *Journal of Ecology*, 80(4):837–870, 1992.

[BWT+19] T. Biedermann, L. Winther, S. J. Till, P. Panzner, A. Knulst, and E. Valovirta. Birch pollen allergy in europe. *Allergy*, 74(7):1237–1248, 2019.

[DAT22] Tausif Diwan, G. Anirudh, and Jitendra V. Tembhurne. Object detection using yolo: challenges, architectural successors, datasets and applications. *Multimedia Tools and Applications*, 82:9243–9275, Aug 2022.

[dMd+20] Letty A. de Weger, Frank Molster, Kevin de Raat, Jeffrey den Haan, Johan Romein, Willem van Leeuwen, Hans de Groot, Marijke Mostert, and Pieter S. Hiemstra. A new portable sampler to monitor pollen at street level in the environment of patients. *Science of The Total Environment*, 741:140404, 2020.

[Emb95] J. Emberlin. Plant allergens on pauci-micronic airborne particles. *Clinical & Experimental Allergy*, 25(3):202–205, 1995.

[FD22] Curado M Polónia A Eloy C. Ferreira D, Vale J. The impact of different coverslipping methods in the quality of the whole slide images used for diagnosis in pathology. *Journal of pathology informatics*, 2022.

[GBBS19] André R. de Geus, Celia A.Z. Barcelos, Marcos A. Batista, and Sérgio F. da Silva. Large-scale pollen recognition with deep learning. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5, 2019.

[Gim84] C. H. Gimingham. Ecological aspects of birch. *Proceedings of the Royal Society of Edinburgh. Section B. Biological Sciences*, 85(1–2):65–72, 1984.

[HA58] H. A. Hyde and K. F. Adams. An atlas of airborne pollen grains, 1958.

[Jon12] Gretchen D. Jones. Pollen analyses for pollination research, unacetolyzed pollen. *Journal of Pollination Ecology*, 9:96–107, Oct. 2012.

[KKKPWS21] Elżbieta Kubera, Agnieszka Kubik-Komar, Krystyna Piotrowska-Weryszko, and Magdalena Skrzypiec. Deep learning methods for improving pollen monitoring. *Sensors*, 21(10), 2021.

[Law19] Dee Lawlor. *Sample Prep*, pages 115–126. Springer International Publishing, Cham, 2019.

[NVI26] NVIDIA. Cuda toolkit documentation, 2026.

[OSM80] John N. Owens, Sheila J. Simpson, and Marje Molder. The pollination mechanism in yellow cypress (chamaecyparisnootkatensis). *Canadian Journal of Forest Research*, 10(4):564–572, 1980.

[RDGF16] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali 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.

[SHWL26] Xuan Shi, Guangliang Hou, Fubo Wang, and Hongyu Li. Pollen-yolo: A deep learning framework for automated pollen identification and its application to palaeoecological reconstruction on the tibetan plateau. *Quaternary*, 9(1), 2026.

[SI18] Petru Soviany and Radu Tudor Ionescu. Optimizing the trade-off between single-stage and two-stage deep object detectors using image difficulty prediction. In *2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 209–214, 2018.

[SIZ24] F. Shamrat, M. Idna Idris, and X. Zhou. Pollennet: A novel architecture for high precision pollen grain classification through deep learning and explainable ai. *Heliyon*, 10, 2024.

[Sti25] Casper Stiekema. One-stage neural network analysis for allergenic tree pollen counting. 2025.

[SVSM18] Farnaz Sedghy, Abdol-Reza Varasteh, Mojtaba Sankian, and Malihe Moghadam. Interaction between air pollutants and pollen grains: The role on the rising trend in allergy, Apr 2018.

[Tzu15] Tzutalin. Labelimg. https://github.com/tzutalin/labelImg, 2015. Accessed: 2026-04-10.

[Web98] Richard W Weber. Pollen identification. *Annals of Allergy, Asthma & Immunology*, 80(2):141–148, 1998.

[ZH80] Donald B. Zobel and Glenn M. Hawk. The environment of chamaecyparis lawsoniana. *The American Midland Naturalist*, 103(2):280–297, 1980.

[ZM26] Teng Zhang and Limi Mao. Deep learning of pollen images under low annotation costs: joint optimization of morphological features and training and prediction strategies. *Review of Palaeobotany and Palynology*, 344:105458, 2026.