Save the meadow birds

Bird nest localization system for autonomous mowing machines

THESIS

submitted to the Delft University of Technology in partial fulfillment
of the requirements for the degree of

Master of Science in Mechanical Engineering

by

Juul Schuurmans

to be defended publicly on Thursday November 30, 2023.

Supervisors:	Dr. ir. Y.B. Eisma,	TU Delft
	Ir. D. Boonstra,	Lely
	Ir. J. Wijkhuizen,	Lely
Thesis committee:	Dr. D. Dodou,	TU Delft
	Dr. J.F.P. Kooij,	TU Delft

TU Delft

Bird nest localization system for autonomous mowing machines

Submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of Master of Science in Mechanical Engineering

Juul Schuurmans - 4560426
November 16, 2023

***Abstract*—Inadvertent bird nest destruction by autonomous mowing machines poses significant threats to the breeding success of meadow birds. Drone-based detection methods represent the current state-of-the-art for bird nest localization to attain mower circumvention. However, they only identify 80% of bird nests with average localization error of 3.344 meters and are restricted to specific application times. This paper introduces alternative, fully automated nest localization systems integrated with autonomous mowers. Two strategies are proposed, 1) Directly detecting bird nests using thermal data, or 2) Indirectly, by tracking birds and extrapolating their trajectories back to their nests using RGBD data. These methods were validated with warmed chicken eggs hidden in grasslands and with drones simulating bird flight. YOLOv8 models were modified for both approaches. The thermal localization method is able to detect all bird nests with an average confidence of 73.4%. It allows for real-time localization and yields one unnecessary nest circumvention for every ten bird nests saved due to false positives. This method is shown to be effective in all breeding season temperatures, both day and night. Conversely, the trajectory extrapolation method detects birds with an average confidence of 82.2% and has localization error of 0.794 meters. Birds taking flight prematurely or from locations other than nests impact the number of bird nests saved and the number of unnecessary circumventions. It is demonstrated that this method fails to detect birds during nighttime. In conclusion, an automated thermal-based localization system integrated with autonomous mowers outperforms both RGBD- and current state-of-the-art drone-based methods. This study highlights therefore the potential of thermal-based solutions for bird nest protection in grasslands.**

***Index Terms*—Meadow birds, Autonomous mower, Nest localization, Object detection, Object tracking.**

I. Introduction

The Netherlands hosts several meadow bird species, including the black-tailed godwit (known as 'grutto' in Dutch), the northern lapwing ('kievit'), the Eurasian oystercatcher ('scholekster'), the northern shoveler ('slobeend') and the common redshank ('tureluur') [1]. Alarmingly, the populations of these species have nearly halved in the last twenty years, putting the entire biodiversity of meadows under significant pressure [2]. Suitable nesting sites are disappearing at a high pace due to urbanization, climate change and intensification of farming activities [3]. Further emphasizing the urgency of this situation, the Dutch Red List of Birds categorizes the shoveler as 'vulnerable' and both the godwit and redshank as 'near threatened' [4]. Protection of these bird species is therefore of utmost importance, starting with safeguarding their breeding success.

Damaging meadow bird nests is for this reason prohibited by Dutch legislation and guidelines have been established for their management [5]. This is vital, since grassland birds all hatch and breed on the ground, making their nests particularly vulnerable to mowing activities [6]. One protection strategy involves locating and documenting the position of nests, after which these sites are protected with no-mow zones of at least 50 square meters. This establishes a safe and food-rich environment for bird chicks before leaving their nests [7]. The methodology and implications of this nest protection approach make up the main focus of this thesis report.

A. Problem statement

Rapid technological advancements in the agriculture domain intensify the threats to meadow birds. As a result, there's growing interest in developing innovative farming solutions that minimize ecological impact. Among the most recent innovations are electrically powered and fully autonomous mowing machines [8]. These machines can independently harvest fresh grass from grasslands, transport it and feed it to dairy cattle, matching the amount of grass distributed to cows' daily nutritional needs [9]. Despite their many advantages, there is an underlying concern: the automated movements of such machinery could unintentionally damage or destruct bird nests.

The protection strategy mentioned in the previous subsection can be integrated with these autonomous mowing machines to overcome this. The bird nests' geographical coordinates can be sent to their software systems, which can then modify their trajectories accordingly. One of the emerging solutions to find these coordinates is the use of drones or unmanned aerial vehicles (UAVs). Various researches have shown that bird surveys using UAVs have a similar number of located nests as manual inspections, while having a much higher speed [10]–[13]. However, the need for separate aerial investigations prior to mowing introduces operational inefficiencies, as mowing cannot be initiated until these investigations are completed. Because their deployment requires operators, they typically occur only once at the start of the breeding season [14]. Consequently, bird nests created after a drone-based localization effort remain at risk of destruction.

B. Research objectives and scope

There is an urgency to protect meadow birds, and given the limitations of existing methods, there arises a need for alternative and innovative solutions. A nest localization system integrated with an autonomous mower could potentially locate nests on its route and seamlessly adapt it accordingly.

This thesis explores the feasibility of such an integrated and automated method. It investigates the challenges associated with nest localization and proposes novel mower-integrated approaches. A further aim is to empirically validate these proposed strategies, quantifying their performance and benchmark them against drone-based methodologies. Combining the research objective and scope, the following research question can be formulated: *Which mower-integrated methods are feasible for automated real-time nest localization of meadow birds in grasslands and how do these methods compare to the conventional drone-based method in terms of number of bird nests saved, accuracy, operational constraints and resources required?* It should be noted that 'resources required' reflects both financial and computational aspects.

C. Outline

Subsequent to this introduction, this thesis continues with a theoretical background, which explores methods for nest localization and outlines the operating principle of object detection models. The following section presents the outcomes of the drone-based method put into practice. Section IV-A and Section IV-B describe the methodology of two proposed methods for automated nest localization. Section V follows with the empirical findings derived from these methods. A discussion of these findings is presented in Section VI. Finally, conclusions are provided in Section VII.

II. Theoretical background

A. Definitions

This research focuses on various methods for localization of bird nests in grasslands. **Bird nest localization** involves determination of a nest's geographical position, often represented in terms of global coordinates. A method's effectiveness in this task is quantified by its localization error, which is defined as the Euclidean distance in meters between a nest's actual and its predicted location [15]. Lower error distances indicate higher **localization accuracy**. Furthermore, this study introduces the **Nest Protection Error Rate (NPER)** as a measure of the effectiveness of nest protection. It quantifies the ratio of incorrectly identified nests which aren't present, resulting in unnecessary no-mow zones, against the number of nests that are correctly identified.

Object detection is required before localization can take place, and is focused on recognizing objects in their environment without determining their geographic coordinates yet [10].

B. Key concepts

Based on studies about manual localization of bird nests, there are roughly two search approaches [16]:

1) directly identifying nests in grasslands by structurally searching every area;
2) tracing birds as they fly away from or to their nests, hence indirectly localizing nests.

These methods remain the most accessible and commonly employed practices among bird-watchers globally [17]. Recent technological advancements have opened up possibilities for automation of these methods, which will be further elaborated upon in the following subsections.

1) Direct localization of bird nests: Although this method yields the most direct results, manually surveying entire grasslands at once does not only consume significant time and labor but also introduces the risk of disturbing nests. As mentioned before, aerial counting methods have been investigated to overcome these difficulties.

Literature has highlighted the effectiveness of drones equipped with visual and thermal cameras for bird nest localization. Because bird eggs have an average temperature of 37.7°C, they are warmer than their surroundings and can be detected by thermal cameras [18]. Integrating both visual and thermal imagery has proved to greatly enhance the detection of bird nests, especially for camouflaged bird nests not discernible with visual cameras alone [12], [13], [19].

For example, agricultural collectives Lopikerwaard and Den Hâneker both employ the drone-based approach for counting and localization of nests. The drone finds heat spots possibly caused by bird nests, prompting subsequent manual inspections of these areas to confirm the identity of the heat spots [14]. Engagement in these inspections led to additional insights, which are elaborated upon in more detail in Section III. The drone-based methodology is adaptable to a nest localization system integrated with an autonomous mower. A thermal camera positioned at its front and directed to the surface can systematically examine its immediate surroundings for heat spots indicating nests. Such a setup benefits from closer proximity to the ground with respect to aerial inspections, possibly increasing detection accuracy [20].

2) Birds' trajectory extrapolation: Manually locating bird nests can also take place by observing birds as they fly away when a disturbance approaches and investigating the area where they departed from. As meadow birds occupy their nests on average $76.9 \pm 2.7\%$ of the time during the breeding season, this could be a feasible method for bird nest localization [21].

Although literature about automation of this process is unavailable, it is possible to attach a RGB depth (RGBD) camera to a mowing machine, providing both color and depth information. This camera can determine its distance to birds flying up from the grass as the machine comes too close [22]. Their trajectories can be traced back to the ground, identifying areas with a higher probability of containing a nest.

C. Object detection

Whether one is analyzing heat spots that suggest the presence of bird nests or examining flying birds using RGB data, the ability to detect objects is fundamental. This involves the use of models to generate bounding boxes around objects and labeling them in the corresponding category [23], [24]. The objective is to develop models that account for variations in position, size and orientation, as well as differences in viewpoints, lightning and occlusion [25].

Convolutional neural networks (CNNs) are the state-of-the-art for high quality object detection in real-time [26]. Based on deep-learning, they identify features from images by themselves [24]. Object detection in RGB images relies on visual features such as colors, gradients and textures, whereas detection in thermal images primarily focuses on contrasts caused by temperatures differences. Training images improve the CNN by iteratively adjusting its weights to minimize the difference between predicted outputs and ground truth [27]. A validation dataset is typically used to assess performance on unseen data, and adapt the model accordingly. Final evaluation is carried out on a completely separate test set [24].

Bird nests typically measure less than 30 centimeters, classifying them as small objects [1]. Small objects detection remains challenging for many detectors due to their limited visual information and resolution, next to their lack of contextual

cues [28]. They are also easily occluded by other objects, further complicating their perception [29]. Yet, multiple overviews of the state-of-the-art for small object detection found that the You Only Look Once (YOLO) algorithms demonstrate an optimal balance between detection accuracy and speed for small objects [26], [30], [31]. While two-stage detectors provide higher accuracy, their slower processing speed limits real-time application. Other one-stage detectors showed lower accuracy compared to YOLO in small object detection. Notably, Li et al. and Hong et al. specifically evaluated multiple object detection models for real-time visual bird and bird nest detection [10], [11]. They confirmed the state-of-the-art performance of the YOLO algorithms in this context. Based on their findings as well as literature on small object detection, the YOLO algorithms are deemed the most suitable choice for the objectives of this research.

1) The YOLO algorithms: The YOLO algorithms achieve their fast detection speed by processing input images in one pass, predicting bounding boxes and their respective class probabilities simultaneously [32]. This approach allows for real-time object detection while maintaining a reasonable level of accuracy. YOLOv8 is the most recent of the YOLO series and outperforms all previous models in both speed and accuracy on numerous datasets, including those with small objects [33]–[35]. Consequently, YOLOv8 will be used for object detection in the continuation of this study.

Five variants of YOLOv8 are available: Nano, Small, Medium, Large, and Extra Large. YOLOv8n (Nano) has the smallest number of parameters and is therefore the fastest to train, while YOLOv8x (Extra Large) is the most accurate yet the slowest among them. The implementations of these models are based on [36].

YOLOv8 is associated with various hyperparameters affecting the model's train time, computational costs, accuracy and its ability to transfer to new, unseen data. The number of epochs specifies how many times the entire training dataset is passed through the model, while the batch size defines the number of training images that are passed simultaneously. The batch size primarily impacts memory usage and training duration [37]. The image size determines the dimensions input images are resized to, before passing through the network. Smaller images reduce computational costs and training time, while larger ones preserve details and improve detection, especially in the context of small objects [38].

How much the weights are updated during training is determined by the learning rate. Higher learning rates can lead to faster convergence but possibly to a less optimal solution, whereas lower learning rates offer more accurate convergence but might require more training epochs [39]. The initial learning rate *lr0* and final learning rate *lr0 * lrf* guide the model during training. Finally, the minimum detection confidence threshold does not affect the model's training but determines the minimum confidence score that a detection must have to be considered valid.

2) Training evaluation: Precision *P* and recall *R* are two performance metrics commonly used in the evaluation of detection models [10]. Precision, defined as in equation 1, is the ratio of true positives (TP) among all predicted bounding boxes, both correct and incorrect. A higher precision value indicates fewer false positives (FP).

$$P = \frac{TP}{TP + FP} \tag{1}$$

Recall on the other hand measures the ratio of correct predictions with respect to the ground truth bounding boxes. Recall is calculated with equation 2. A higher recall value indicates a lower number of false negatives (FN).

$$R = \frac{TP}{TP + FN} \tag{2}$$

Following precision and recall, another metric is the False Discovery Rate *FDR*, which is defined as the proportion of false positives among the the total number of predictions [40]. The *FDR* should ideally be low and is defined as in equation 3.

$$FDR = \frac{FP}{TP + FP} \tag{3}$$

While FDR is related to NPER in measuring false positives, NPER focuses more on practical implications as not all false positives yield equal numbers of unnecessary nest circumventions.

The trade-off between precision and recall can be visualized with the precision-recall (PR) curve. The average precision (AP) of each class is the

area under the PR-curve [41]. The mean Average Precision (mAP) calculates the mean AP-value of all classes.

The precision and recall metrics are closely related to the Intersection over Union (IoU) value [42]. IoU measures the area of overlap and the area of union between a predicted bounding box and the ground truth. Bounding boxes with an IoU above a certain threshold are considered TP, while those below would result in FP [42]. The mAP50-95 metric averages the mAP values for a range of IoU thresholds from 0.50 to 0.95. mAP50-95 provides a rigorous evaluation of the model's accuracy, as it requires highly accurate bounding boxes.

Next to these metrics, the bounding box loss is of importance, which quantifies the difference between predicted and ground truth bounding box coordinates using the mean squared error [36]. By monitoring the behavior of training and validation losses overfitting can be detected. If the model starts to overfit, the training loss will continue to decrease, while the validation loss might stop decreasing or even start to increase [43]. This indicates that the model performs increasingly better on the training data but fails to generalize to new data [44], [45].

D. Object tracking

The second method for nest localization employs birds' trajectory extrapolation. At the foundation of trajectory generation lies object tracking; locating a moving object or multiple objects over successive frames in a video, thereby constructing a trajectory [46]. There are multiple algorithms available for object tracking, but one of the easiest to implement and computationally efficient approaches is Simple Online and Realtime Tracking (SORT) [47]. The algorithms relies on bounding boxes retrieved from an object detection model, combined with a Kalman filter to predict locations of objects in subsequent frames [48]. It associates new detections with existing tracks by calculating the IoU [49]. Its main limitations are its inability to handle occlusions and reappearances, as SORT might assign a new trajectory ID. In the context of birds flying up from the grass, occlusion won't frequently occur as meadows often offer limited obstacles. The SORT algorithm used in this research is retrieved from [49].

E. Reformulation of research question

After exploring possible methods for autonomous mower-integrated bird nest localization, the formulated research question can be redefined: *In terms of number of bird nests saved, localization accuracy, operational constraints and resources required, which method — thermal camera based, RGBD camera based, or drone-based — proves to be the most effective for automated real-time bird nest localization in grasslands?*

III. Benchmark method

A. Methodology

The conventional drone-based method was put into practice on April 18th 2023 at the grasslands behind Blokland Dairy in Goudriaan, The Netherlands. Bird nest localization was carried out with a DJI Mavic 3 Thermal drone, especially designed for surveying nature [50]. This drone is equipped with three cameras; a 48 Megapixel Wide visual camera with 24 mm lens, a 12 Megapixel Zoom visual camera with 162 mm lens and a thermal camera with field of view (FOV) of 60° × 49° (H × V) and resolution of 640 × 512 pixels. It also features Real Time Kinematic (RTK), which enables centimeter accurate positioning. Its flight time with a full battery is 45 minutes.

Two volunteers of collective Den Hâneker helped out with the bird localization effort. One operated the drone and investigated the thermal camera's live stream to find heat spots, after which the Zoom visual camera was used to identify them. If a possible bird nest was detected, the second volunteer examined the corresponding location in the grassland. After confirmation of a nest, its heat spot was marked on the drone's map and its coordinates were calculated.

To ensure the largest temperature difference between bird eggs and their surroundings, the drone took flight over the grasslands precisely at sunrise, at 6 o'clock in the morning. Earlier in the morning would yield larger differences in temperatures hence better visibility of nests, but flying a drone before sunrise is not permitted by Dutch law [51]. The grasslands of Blokland Dairy have an area of approximately 16 hectares and investigating them with the drone took around 3 hours.

Figure 1: Two godwit nests as found during the drone localization effort at Blokland Dairy.

B. Results and insights

Six nests were found during the localization effort; five godwit nests and one from a shoveler, as presented in Figure 1. The drone software's localization error was on average 3.344 $\pm$ 0.773 meters, as determined in Table IX in Appendix IX. Heat spots merely indicated possible locations and the nests still had to be found within this area, while being extremely careful not to destroy the nests.

Once the nests were discovered in the grasslands, their positions were more accurately mapped using a Trimble R2 GNSS receiver, with up to centimeter accuracy [52]. These locations were forwarded to the software of the autonomous mower operating at Blokland Dairy, which then adapted its mowing route to circumvent these areas, as illustrated in Figure 11 in Appendix IX.

Even with the relatively low temperature of the surroundings, approximately two thirds of possible bird nest locations turned out to be false positives, hence NPER equals 2. These heat spots were caused by bare patches of grass, fresh poop or molehills emitting heat and couldn't be identified with the Zoom visual camera alone. The exact number of bird nests that remained undetected during this localization effort was unknown. However, collective Lopikerwaard employs a similar nest localization method and is able to detect approximately 80%, confirmed through manual inspections of the meadows.

While the drone-based approach offers significant advantages in terms of speed compared to manual inspections, it still requires a large time investment due to limited battery life, high number of false positives and low localization accuracy. Moreover, this method is limited to specific application times like sunrise for localization, which becomes progressively earlier as the breeding season advances. It has to be carried out before the mowing process and only happens once at the beginning of the breeding season, meaning that bird nests created after a drone-based localization effort would still be destructed. It can be concluded that there is substantial room for improvement within this benchmark method.

IV. Methodology

The two proposed methods for automated bird nest localization have been implemented for comprehensive comparison, after which they were extensively tested and evaluated.

A. Thermal bird nest detection model

1) Field data collection: There were no autonomous mowing machines available for experimentation during the course of this research, as their operations were not to be disrupted. For this reason, an alternative setup was devised. Using a quad with aluminum bars mounted to it, a thermal camera

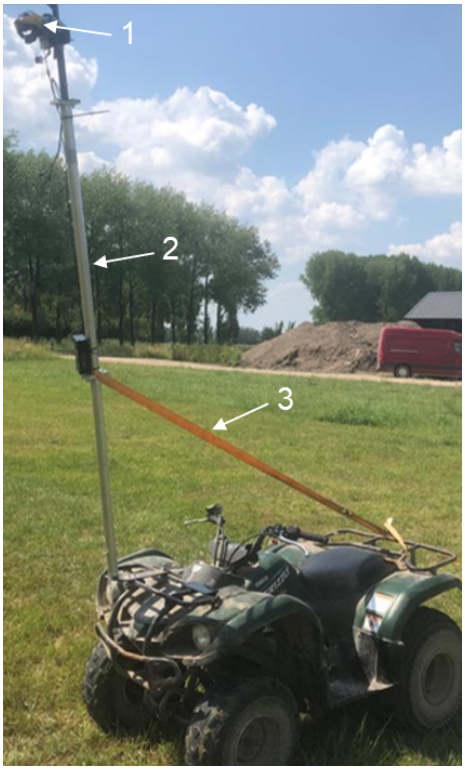


Figure 2: Experimental setup consisting of FLIR thermal camera (1) attached to a quad via aluminum bars (2) and a tension strap (3).

could be attached, creating a mobile system with a designated camera height and angle, see Figure 2. This approach facilitated systematic investigation of various configurations for the thermal camera, without interfering with mowing activities. During measurements, the quad's velocity was set to resemble the speed of an autonomous mower of 1 m/s.

Two thermal cameras were available for experimentation; a Hikvision Thermal & Optical Bi-Spectrum Network Dome Camera DS-2TD1217-2/V1 and a FLIR Vue Pro R 640. The Hikvision camera offers the advantage of integrated thermal and visual imagery, making it possible to verify identities of heat spots and potentially reducing the number of false positives, similar to the benchmark method and research conducted by McKellar et al. [12]. It has a FOV of 90° × 66° (H × V), resolution of 160 × 120 pixels and can measure temperatures up to distances of 15 meters [53]. The FLIR (Forward-Looking Infrared) camera was taken off a drone used for bird nest localization in previous years by collective Den Hâneker. With a lens of 13 mm, the camera provides a FOV of 45° × 37° (H × V) and resolution of 640 × 512 pixels [54]. FLIR cameras share the same operating principle as other thermal cameras, but distinguish themselves through their emphasis on real-time thermal imaging and their superior image quality [55]. During trial measurements, the Hikvision camera demonstrated good potential in terms of field of view. However, its limited resolution proved to be a major drawback, as the small eggs' heat spots were difficult to distinguish from other heat sources. As a result, the Hikvision camera was discarded, and subsequent measurements were carried out with the FLIR camera.

A series of experimental measurements were conducted to find an optimal camera angle and height. Although larger angles with the horizontal provided wider fields of view, nests were only discernible when the camera was directly overhead. This was due to grass covering the eggs, obstructing a clear perspective. Smaller angles on the other hand diminished the distance at which nests could be detected. An angle of 20 degrees with respect to the horizontal yielded the most practical outcomes. Three distinct camera heights were tested: 1.70, 2.65 and 3.71 meters. While the highest option naturally yielded the widest field of view, 2.65 meters was more closely associated to the height of autonomous mowers, offering more insight in the feasibility of this method without having to create extensions on the mower. At a camera height of 2.65 meters and an angle of 20 degrees relative to the horizontal, the camera's field of view measured 2.33 by 2.09 meters (H × V). The FOV's midpoint was situated 1.235 meters ahead of the camera, providing a theoretical distance from which nests can be detected of 2.28 meters.

The FLIR camera used in this study has several features specifically implemented for bird nest detection. Active Contrast Enhancement (ACE) emphasizes temperature differences in captured images, highlighting warm objects while making the environment appear darker. In this research, the ACE value was set to its maximum. The Digital Data Enhancement (DDE) parameter controls the sharpness of the thermal image. It was set to its maximum value by default to accentuate contours, crucial for identification of nests. Additionally, the Smart Scene Optimization (SSE) feature increases visibility of objects with small temperature differences compared to their surroundings, making them appear as if they have a higher temperature than they actually do. In the context of bird nest detection, their exact temperature is of less importance than their visibility, which is why this value was also set to its maximum. One drawback of setting all these features to their highest values is that outcomes of this method might not be generalized to other thermal cameras. Furthermore, it could lead to overemphasis on temperature differences, possibly leading to more false positives.

At the start of measurements in May, half of the bird eggs at Blokland Dairy already hatched and their nests were abandoned. To protect the nests that remained, measurements were not conducted on these nests but they were replicated instead. Grocery store-bought chicken eggs with sizes M and L were used for this purpose, as they are representative of various sizes of meadow bird eggs [56]–[58]. The independent variables in this experimental study were the lateral distance between camera and nest, the number of eggs per nest and the temperature difference of the eggs with respect to their surroundings, see Table I. The nest size variations align

with typical numbers of eggs laid per nesting attempt [59], [60]. The values of the egg temperature difference are explained in Appendix X. The eggs were heated with a water bath at a set temperature, after which they were hidden in the grass in a way that resembled the natural bird nests in Figure 1 as much as possible.

Table I: Independent variables of measurements related to direct bird nest localization with a thermal camera.

Lateral distance to nest [m]	**Number of eggs** [-]	**Egg temperature difference** ΔT [°C]
0	1	17
0.40	2	21
0.80	4	25
1.15	5	29
		33

2) Dataset and annotation: In total, 73 videos with a frame rate of 10 frames per second (fps) were collected. Each video was split up into its frames for annotation purposes, resulting in 7153 images. 2876 of these images were null-images, displaying no bird nests and only background, while the remaining 4277 images contained either one or two bird nests. Five distinct experimental conditions were designated as the test set to ensure it included entirely new data, each condition with four videos corresponding to the number of eggs. Consequently, 3999 images were labeled as train set, 1013 as validation set, and 2141 as test set, achieving a split of train : val : test = 56 : 14 : 30%. Although this distribution wasn't ideal, it was essential to ensure the test set comprised of all egg temperature variations.

The annotation procedure took place manually using the annotation tool Roboflow[1]. An example of the thermal data gathered and the annotation process is shown in Figure 3.

3) YOLOv8 implementation: All models trained in this study utilized an NVIDIA GeForce GTX 1070 Ti GPU with 8 GB of graphics memory. YOLOv8m (Medium) was the largest model that could be trained without running into memory limitations, and a batch size of 16 proved to be the largest feasible size for this model. The three smallest YOLOv8 models - YOLOv8n, YOLOv8s (Small) and YOLOv8m - were trained on the thermal dataset. Using a batch size of 16 for 100 epochs, training times were 1.824 hours for YOLOv8n, 3.416 hours for YOLOv8s and 7.437 hours for YOLOv8m, while their maximum achieved accuracy values were 0.789, 0.802 and 0.810 respectively. Figure 14a in Appendix XII demonstrates a consistent trend of increasing accuracy as model size increases, thereby prompting to consider larger architectures.

The effects of hyperparameter adjustments and data augmentation in the sections below will be examined using YOLOv8n. Unless stated otherwise, the following hyperparameters are used for training: *epochs* = 40, *lr0* = *lrf* = 0.001, *imgsize* = 640 × 480 and *batch* = 32. The learning rates are default values for YOLOv8 [36]. The image size is maintained at its original dimensions to persevere resolution, improving visibility of details. The batch size is set based on GPU limitations. With each new training strategy, a new YOLOv8 model is initialized, hence there is no transfer learning from any other pre-trained models.

Hyperparameter tuning. The impact of different batch sizes on the mAP50-95 metric was assessed while keeping other parameters at their standard values, see Figure 14b in Appendix XII. The effect is, as expected, found to be negligible: the highest mAP50-95 achieved over 40 epochs was 0.763 using a batch size of 64, while the lowest

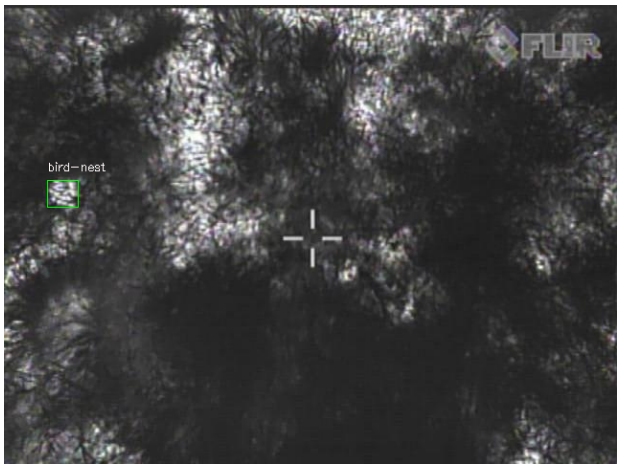


Figure 3: Example of data gathered with FLIR camera, including annotation of a bird nest.

[1]Roboflow, online annotation tool, https://roboflow.com/

value was 0.755 with a batch size of 8, resulting in a mere 1.045% difference. Notably, a model with a batch size of 8 doesn't converge within the specified number of epochs and remains unstable, which explains its lowest mAP-value.

Additionally, two learning rate schedules were studied: linearly decreasing, where *lr0* = *lrf*, and constant, where *lrf* = 1. The results are reported in Figure 14c and 14d in Appendix XII respectively. In both cases, an initial learning rate of *lr0* = 0.001 yielded the best outcomes, with the linear decreasing schedule outperforming the constant one. Therefore, the default learning rates *lr0* = *lrf* = 0.001 were determined to be the optimal ones. Again, the effect of this hyperparameter proved to be minimal, as the difference between best and worst-performing schedules was only 4.858%.

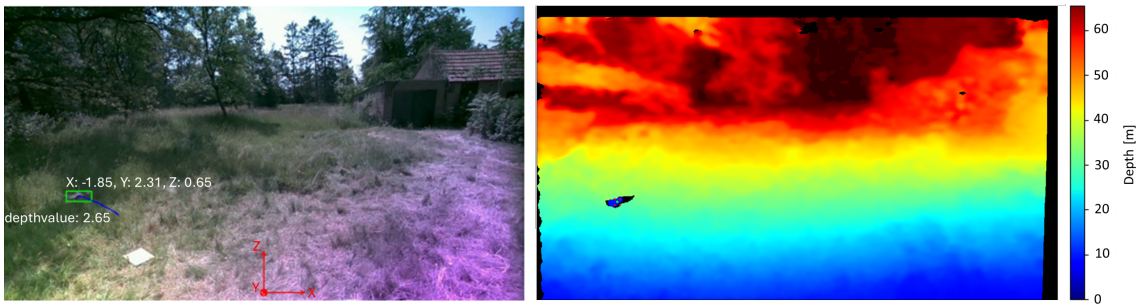
Data augmentation. Augmenting a dataset creates variations of each original training image, introducing new, unseen instances. This added diversity exposes the model to a wider range of scenarios, making it more robust and less sensitive to overfitting [61]. Various augmentation techniques were explored and compared, focusing on highlighting the contrast between the distinctive white heat patterns and the dark background.

To this end, seven different augmentation approaches were investigated and Figure 17 in Appendix XIII presents examples of the augmented training images. A horizontal flip was introduced to increase the number of images available for training. Gaussian noise can improve the robustness of the model to different levels of thermal noise. This augmentation was achieved by adding values sampled from a Gaussian distribution to 5% of the pixels. Adding cutouts to each training image can simulate object occlusion, forcing the model to focus on other features of the objects [62]. Cutouts were introduced with six boxes per images, each sized at 10% of the image dimensions. Applying a blur reduces details and suppresses noise, possibly making features of the heat spots more distinguishable against their background. This effect was achieved using a Gaussian filter with a standard deviation of 5, which controls the amount of blurring applied. With mosaic augmentation, multiple images were combined into one, simulating complex scenes and new interactions between objects, potentially improving small object detection [63]. Variability to image brightness - both brightening and darkening - are meant to make the model more resilient to lightning conditions [61]. Brightness adjustments were made by multiplying all pixel values by either 0.75 or 1.25. Finally, adapting bounding box brightness involves locally scaling each pixel in the bounding box region with either 0.75 or 1.25. This approach proves particularly valuable for objects with different lightning conditions compared to their background. Annotations of original images are transformed for every data augmentation approach, as demonstrated in Figure 17, eliminating the need for manual re-annotation. All values mentioned here are default values for each augmentation technique and provide a solid starting point. Data augmentation produced two new instances for every training image, except for horizontal flip which produced only one, increasing the train dataset from 3999 to 11997 images.

The development of the mAP50-95 value for the various augmentation methods is depicted in Figure 16 in Appendix XIII. The maximum mAP50-95 values over 40 epochs are presented in Table XIII, along with a comparison to no augmentation. Image brightness adjustments proved most effective for the thermal dataset, but Table XIII illustrates the potential of blur, cutouts and noise augmentation techniques. Variations of these approaches were further explored in Figure 18 in Appendix XIII. It is evident from these graphs that default brightness adjustments are the most effective. For blur and noise augmentations, lower parameter values improved results, while a higher number of boxes enhanced the performance of the cutout approach. Still, none surpassed the default brightness adjustments.

The brightness method was explored further by combining it with best-performing variants of blur, cutouts and noise augmentation techniques. Figure 19 in Appendix XIII shows that combinations of brightness augmentation tended to decrease performance on the validation set.

In conclusion, the most effective model for the thermal dataset is YOLOv8m with a batch size of 16 and brightness augmentation by scaling pixel values by either 0.75 or 1.25. Other hyperparameters are set to their standard values.



(a) RGB frame with predicted bounding box, trajectory, depth value and global coordinates, in meters.

(b) Depth frame with center of bounding box and neighbouring pixels.

Figure 4: Data gathered with Intel RealSense D455 Depth Camera.

B. Bird tracking algorithm using depth data

1) Field data collection: Autonomous mowers are typically equipped with two RGB cameras for obstacle avoidance [8]. However, reproducing this setup for experimental purposes proved challenging, requiring two calibrated cameras mounted on a moving vehicle. A ready-to-use RGBD camera was easier to implement and did not call for extensive re-calibration whenever the camera's height or angle was adjusted.

The autonomous mowers' camera setup is reproduced with an Intel RealSense D455 Depth Camera, which combines a RGB camera with two infrared sensors [64]. This camera calculates the distance of each pixel relative to itself, after which it can determine the location of objects in three dimensions. The infrared sensors offer a 87°× 58° (H × V) FOV and operate within a range of 0.6 up to 6 meters, with a depth error less than 2% at 4 m.

Many configurations were examined to optimize the depth camera's FOV. Placing the camera at a greater height intuitively expanded its FOV, yet it simultaneously decreased the range at which depth information could be captured. Additionally, tilting the camera was necessary for optimal visibility in different weather conditions, such as rain and direct sunlight, but also reduced the camera's FOV. Finally, a height of 2 meters and an angle of 22.5 degrees with respect to the horizontal were established as the most suitable configuration. This setup yielded a FOV of 4.11 by 2.40 meters (H × V). The FOV's center was situated 3.20 meters ahead of the camera's own position, theoretically enabling it to detect birds from up to 4.40 meters away.

Table II details the independent variables used in this experimental study. The values for lateral and longitudinal distances aim to assess performance when birds are either flying up inside or outside the depth camera's FOV. It has been confirmed by the volunteers of collective Den Hâneker, who helped out with the benchmark localization effort, that 2 up to 5 meters are realistic values for birds taking flight from their nests when a machine approaches.

As emphasized earlier, the hatching bird nests at the study location were to remain undisturbed throughout the measurement period. Birds flying away from their nests were simulated by a controllable drone, measuring 25 by 7.5 by 6 centimeters, taking off from a designated landing pad. It must also be noted that all measurements were conducted with a stationary camera rather than a moving one. Mounting the camera to a moving vehicle, such as a quad, introduced challenges. The uneven meadow grounds caused many disturbances to the recorded videos, yielding inaccurate depth measurements.

Table II: Independent variables of measurements related to bird trajectory extrapolation with a RGBD camera.

Lateral distance to nest [m]	**Longitudinal distance to nest** [m]	**Direction** [-]
0	2	left
1	3	right
2	4	
4	5	
5		

2) Dataset and annotation: The depth camera captures both visual and depth information, refer to Figure 4 for an example of the data collected. A YOLOv8-based detection model processes the visual data, and, once detections are made, depth information is extracted correspondingly. It's worth

noting that the object detection model is exclusively involved with the RGB dataset.

A total of 61 videos were captured over the course of three different days, featuring three different backgrounds and lightning conditions. The test set consisted exclusively of videos acquired during one of these days, serving to evaluate the model's performance on unseen data. Two additional videos were created, in which the drone took off from shaded areas, which were also included in the train and validation sets. The videos were recorded at 30 fps and split up into their frames for annotation purposes. To maintain variability of the dataset, duplicate or highly redundant frames were excluded, resulting in 3786 images. This dataset consisted of 1253 null-images and 2533 images displaying the drone. Among these, 2132 images were allocated for training, 713 for validation, and 941 for testing, thereby dividing the dataset at 69 : 17 : 14% respectively. The annotation procedure took place manually using the annotation tool Roboflow.

3) YOLOv8 implementation: Similar to the previous method, the effects of hyperparameter tuning and data augmentation techniques were investigated to find the YOLOv8 model that showed optimal performance when trained on the visual dataset. All training strategies were again executed on an NVIDIA GeForce GTX 1070 Ti GPU with 8 GB of graphics memory.

Figure 15a in Appendix XII depicts the outcomes of YOLOv8n, YOLOv8s and YOLOv8m models trained on the visual dataset, with *batch* = 16 to avoid exceeding GPU limitations. The maximum mAP50-95 values achieved in 100 epochs are 0.668 for YOLOv8n, 0.713 for YOLOv8s and 0.719 for YOLOv8m, with respective training times of 0.934 hours, 1.864 hours and 3.976 hours. Optimization of hyperparameters and data augmentation strategies are carried out with YOLOv8n. The following hyperparameters are used, unless stated otherwise: *epochs* = 40, *lr0* = *lrf* = 0.001, *imgsize* = 1280×720 and *batch* = 32.

Hyperparameter tuning. The effect of batch size on the mAP50-95 metric is demonstrated in Figure 15b in Appendix XII. Comparable to the thermal dataset, the influence of the batchsize on training outcomes is small, with a batch size of 64 achieving the highest mAP50-95 value equal to 0.612, and a batch size of 8 achieving 0.580, resulting in a difference of 5.322%. The effects of learning rate schedules are presented in Figures 15c and 15d in Appendix XII. It is evident that a linear decreasing schedule outperforms a constant one, with the smallest learning rates *lr0* = *lrf* = 0.0001 yielding highest mAP50-95 values. All constant learning rate schedules perform exactly the same.

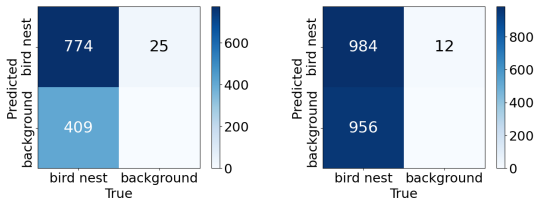
Data augmentation. Given the dataset's uniformity and limited size, characterized by a single drone flying up from the ground against different backgrounds, it becomes crucial to apply data augmentation techniques. Initial training of the model revealed some areas of struggle for object detection, especially when the black drone moved in front of a dark background or when it wasn't directly facing the camera and flying at an angle. Consequently, the data augmentation techniques examined here specifically aim to highlight the contrast between the drone and its background and enable the model to recognize the object in various orientations and sizes.

Gaussian noise, Gaussian blur and variations in image and bounding box brightness were applied to serve the first objective, with their implementation elaborated in Subsection IV-A3. Next to these, the effects of saturation adjustments and grayscale on the model's performance were assessed. Saturation controls the intensity of colors in an image. Adjustments of saturation involve scaling the saturation component of each pixel with a specified factor, with default values equal to 0.50 and 1.50 [65]. Converting images to grayscale completely removes color information, forcing the model to solely focus on intensity contrasts. Furthermore, introducing a horizontal and vertical shear on each bounding box adds variability to the drone's viewpoint, potentially aiding its detection from multiple angles [62]. Each bounding box was subjected to a random shear of $\pm$ 15°, both horizontally and vertically. The effect of mosaic augmentation was also explored, but cutouts weren't as the drone was never occluded by other objects in this dataset. Figure 23 in Appendix XIII displays examples of the augmented training images mentioned here. With data augmentation, two new versions of each training image were created, increasing the train dataset from 2132 to 6396 images.

The effects of the eight data augmentation strategies were explored using the YOLOv8n model and their training outcomes are presented in Figure 20 and Table XIII in Appendix XIII, along with a comparison to no augmentation. Bounding box shear augmentation outperforms the alternative methods, closely followed by brightness, saturation, blur and noise. Variations of the associated parameters of these augmentation techniques were examined to identify configurations that surpass their initial implementations. The effects of these variations can be observed in Figure 23 in Appendix XIII. Both bounding box shear and blur perform best with their default values. Brightness adjustments with lower values yield better outcomes, while scaling the saturation component with 0.25 and 1.75 exceeds the performance of bounding box shear, making this the most effective augmentation technique.

Efforts were made to increase performance by combining the saturation method with the other five best-performing approaches, using their optimal configurations. The results, depicted in Figure 22, clearly demonstrate that combining saturation augmentation with these methods reduces mAP50-95 values.

In conclusion, the best-performing model for this dataset employs a YOLOv8m architecture, batch size of 16, learning rates *lr0* = *lrf* = 0.0001 and saturation augmentation that scales each pixel's saturation component by either 0.25 or 1.75. This particular model will be implemented in the bird tracking algorithm described in Appendix XI.



(a) (b)

Figure 5: Confusion matrices of thermal test set on the left and evaluation nests on the right, with confidence threshold equal to 0.30.

V. Results

This section presents the performance of the two localization methods on their corresponding test data. An independent dataset comprising thermal and RGBD imagery from ten evaluation nests is introduced, detailed in Table III. While the test sets are intended to include unseen data as much as possible, as outlined earlier, this new dataset provides a more comprehensive assessment by incorporating data captured in a different environment. Moreover, it serves as a robustness test, identifying vulnerabilities and evaluating suitability for future real-world applications.

Table III: Characteristics of ten evaluation nests.

Nest number [-]	**Lat. distance to nest** [m]	**Lon. distance to nest** [m]	**Number of eggs** [-]	**Temperature difference** ΔT [°C]
1	0.95	4.90	5	16.5
2	1.20	2.10	1	16.5
3	1.15	2.35	1	20.5
4	2.5	2.10	2	20.5
5	0.90	3.50	4	24.5
6	1.10	1.80	1	24.5
7	1.00	5.30	5	28.5
8	2.46	2.80	4	28.5
9	1.15	2.75	4	32.5
10	0.00	4.90	2	32.5

A. Thermal bird nest detection model

The YOLOv8m model trained on the augmented thermal data, as discussed at the end of Section IV-A, starts to overfit after 63 epochs. The validation loss graph crosses the training loss graph at this point, see Figure 24 in Appendix XIV.

1) Test set: Confusion matrices corresponding to various confidence thresholds are outlined in Figure 26 in Appendix XIV. The number of true negatives remains uncertain, since the background was not annotated. A confidence threshold of 0.30 is selected and its associated confusion matrix is outlined in Figure 5a. This threshold strikes a balance between false positives, false negatives and the distance from the mower a nest is first detected, as presented in Table IV. There are no statistically significant differences in these first detection distances across various confidence thresholds, as determined by a two-sample t-test [66]. The mAP50-95 value on the test set is 0.785 ± 0.165.

The distance a nest is first detected as a function of the confidence threshold is visualized in Figure 27 in Appendix XIV. The Pearson correlation coefficient r is -0.954, and based on a two-tailed t-test, this correlation proved significant

Table IV: Detection results and average first nest detection distance as a function of confidence threshold, for the test set of the thermal bird nest detection model.

Conf. threshold [-]	**True positives** [-]	**False positives** [-]	**False negatives** [-]	**First nest detection** [m]
0.10	787	30	396	1.890 ± 0.655
0.20	787	30	396	1.878 ± 0.652
0.30	774	25	409	1.836 ± 0.636
0.40	753	19	430	1.727 ± 0.734
0.50	727	15	456	1.626 ± 0.747

($p < 0.05$) [67]. With the selected confidence threshold of 0.30, the first nest localization distance is on average 1.836 ± 0.636 meters.

2) Evaluation nests: Figure 29 through Figure 38 in Appendix XV show thermal images of the evaluation nests, along with their corresponding visual images. The performance of the YOLOv8m model applied to thermal data from these evaluation nests is summarized in Figure 5b and for every nest separately in Table V. The average detection confidence of all TP is 0.822 ± 0.094. The *FDR* value equals 0.012. As all FP originate from a single heat spot, NPER equals 0.1. The confidence associated with each detection relative to its distance is presented in Figure 6. The Pearson correlation coefficient r before the dashed line is 0.862, which is statistically significant as determined by a two-tailed t-test [67], [68]. The thermal bird nest detection model processes frames at a speed of 45 fps.

The correlation coefficients r between the independent and dependent variables associated with the thermal measurements are depicted in Table VI. Only the correlation between number of eggs and first detection distance is found to be significant ($p < 0.05$), determined by a two-tailed t-test [67], [69]. The correlation between temperature difference and recall is visualized in Figure 7. Correlations between dependent variables are listed in Table XIV in Appendix XIV.

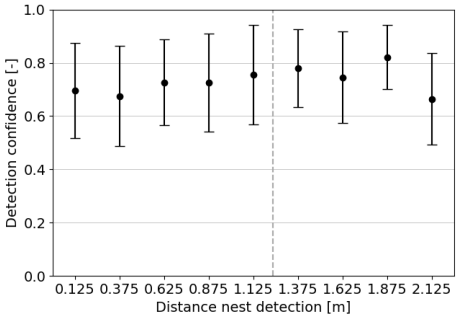


Figure 6: Detection confidence for thermal evaluation nests, thresholded at 0.30, relative to longitudinal distance from the camera. The dashed line represents the location directly beneath the camera.

Table V: Detection results of YOLOv8 model applied to thermal data from evaluation nests with a confidence threshold of 0.30.

Nest number	**True positives**	**False positives**	**False negatives**
1	153	0	34
2	4	0	192
3	360	0	48
4	67	0	146
5	179	0	1
6	36	0	62
7	41	0	33
8	43	0	141
9	66	12	152
10	34	0	144

Table VI: Pearson correlation coefficient r between independent and dependent variables of both thermal test set and thermal evaluation nests, hence n = 30. Statistically significant correlations are underlined.

r	**Lateral distance to nest**	**Number of eggs**	**Temperature difference** ΔT
P	-0.022	-0.049	-0.002
R	-0.074	0.030	0.269
mAP50-95	-0.065	-0.055	0.394
FDR	-0.046	-0.035	0.321
First detection distance	0.298	<u>0.457</u>	-0.102
Detection confidence	-0.375	0.026	-0.048

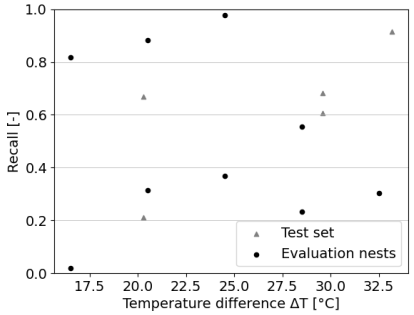
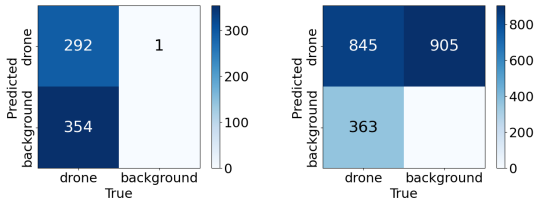


Figure 7: Recall as a function of the temperature difference ΔT between nest and surroundings, with detection confidence threshold equal to 0.30.

B. Bird tracking algorithm using depth data

The YOLOv8m model trained on the visual dataset starts to overfit after 49 epochs. Here, the validation loss starts increasing while the training loss continues to decrease, as presented in Figure 25 in Appendix XIV.

1) Test set: Confidence thresholds ranging from 0.65 to 0.85 were investigated and their respective confusion matrices are depicted in Figure 28 in Appendix XIV. Lower thresholds yield many false positives, while higher thresholds rapidly reduce the number of true positives. A threshold of 0.75 is selected and its associated confusion matrix is outlined in Figure 8a. With this threshold, the mAP-value on the test set is 0.519 $\pm$ 0.152.



(a) (b)

Figure 8: Confusion matrices of visual test set on the left and evaluation nests on the right, with 0.75 confidence threshold

2) Evaluation nests: The confusion matrix corresponding to the evaluation nests is presented in Figure 8b, where *FDR* = 0.517. The detection results of each nest separately are outlined in Table VII. The average detection confidence of all TP is 0.734 $\pm$ 0.172.

The bird tracking algorithm runs at 17 fps. Nests 1 and 10 are excluded from further evaluation as the drone never occurs within the camera's depth range here. Figure 9 illustrates the relationship between the absolute distance to the drone's take-off location and the localization error, with r = 0.124. Only depth values equal to or smaller than 3.2 meters are considered, assessing the bird tracking algorithm's performance rather than the depth camera's capacities. The number of detections available for location predictions is also depicted. The Pearson correlation coefficient r is -0.608 between the number of detections and the localization error. As p = 0.002 based on a two-tailed t-test, this correlation is significant [70]. Correlations between other independent en dependent variables are presented in Table VIII. The correlations between lateral distance and recall R, mAP50-95 and detection confidence are significant based on a two-tailed t-test [67].

The localization error of the evaluation nests is on average 0.794 $\pm$ 0.578 meters for a commonly observed grass length of 20 centimeters. This error is 0.801 $\pm$ 0.520 meters for drone take-off locations inside the camera's FOV, and for those outside the FOV 0.781 $\pm$ 0.641 meters. This difference is not statistically significant [66]. The effect of variations in grass length on the localization error is visualized in Figure 10. The correlation coefficient equals 0.979, which is statistically significant as p = 0.0004. Correlations between dependent variables are listed in Table XV in Appendix XIV.

Table VII: Detection results of YOLOv8 model trained on visual dataset, applied to evaluation nests. The detection confidence threshold is 0.75.

Nest number	True positives	False positives	False negatives
1	129	429	39
2	107	0	1
3	29	0	10
4	206	0	5
5	94	0	76
6	32	0	66
7	40	0	111
8	28	0	6
9	98	0	18
10	82	476	31

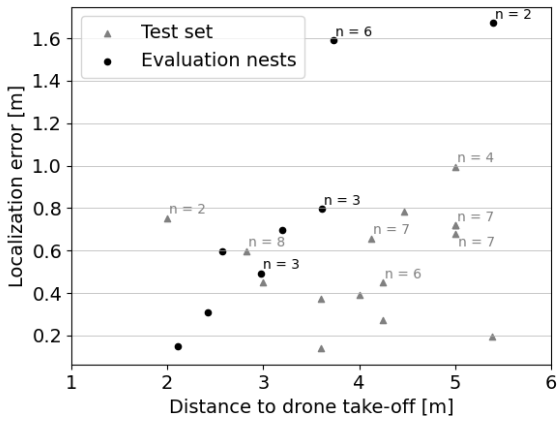


Figure 9: Effect of distance from which the drone takes off on localization error, with a confidence threshold equal to 0.75 and grass length of 0.20 m. Unless stated otherwise, n = 10.

Table VIII: Pearson correlation coefficient r between independent variables and detection results of RGBD test set and evaluation nests, with n = 23. Statistically significant correlations are underlined.

r	**Lon. distance to nest**	**Lat. distance to nest**
P	-0.339	0.033
R	-0.165	-0.519
mAP50-95	-0.080	-0.527
FDR	0.320	-0.254
Localization error	0.402	-0.193
Detection confidence	-0.429	-0.576

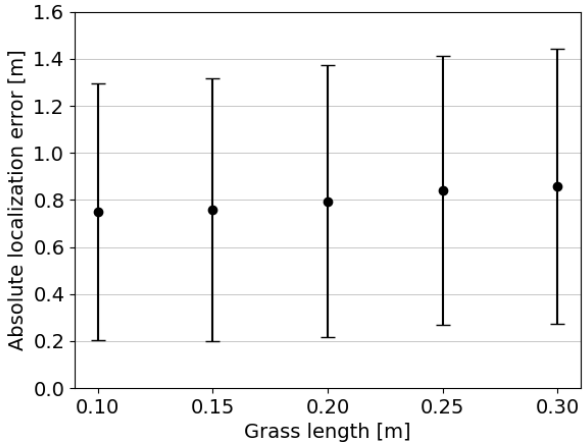


Figure 10: Absolute localization error of evaluation nests as a function of the grass length, for a detection confidence threshold of 0.75.

VI. Discussion

A. Thermal bird nest detection model

1) Bird nests saved: Figure 5b and Table V demonstrate a considerable number of FN as the model sometimes loses sight of nests or fails to recognize them immediately. Nonetheless, the substantial number of TP indicates that the model successfully detects each evaluation nest in numerous frames, enabling effective circumvention. Its average detection confidence 0.734 $\pm$ 0.172 highlights this capability.

From Table V it is clear that all FP are caused by a single heat spot, contributing to an NPER of 0.1. This translates into substantially fewer unnecessary no-mow zones compared to the drone-based method, for which NPER = 2. A solution for further enhancement might involve setting a minimum number of detections for a location to be considered a nest, but this increases the risk of missing nests. For instance, nest 2 is detected in only four frames. To address both the number of FP and FN, gathering more data and obtaining realistic thermal data from actual bird nests is recommended.

2) Localization accuracy: Although localization accuracy for this method cannot be directly determined, the maximum value of the mAP50-95 metric on the test set is 0.785 $\pm$ 0.165, indicating that bounding boxes are predicted close to the ground truth. With the location of bounding boxes within the camera's FOV and the autonomous mower's world coordinates known, localization is straightforward and can achieve a high level of accuracy [71]. To quantify this method's localization error, the bird nest detection model should be trained with and applied to thermal data gathered with an autonomous mower, localizing actual bird nests. The current model allows for an average nest localization distance of 1.836 $\pm$ 0.636 meters. A significant correlation exists between the number of eggs in a nest and the distance it is first detected. Generally, larger nests are associated with greater heat spots and might be detected more easily. Exceptions could be caused by positioning of the eggs and the angle at which the nests are captured, see Figure 35 and Figure 36 in Appendix XV.

Another aspect to consider is the model's confidence score based on nest proximity, as presented in Figure 6. There is a statistically significant correlation for data points occurring before the location directly beneath the camera, marked by the dashed line. This indicates that as nests are closer to the camera and are presumably more visible, the model's detection confidence increases.

3) Operational constraints: There is no significant correlation between temperature difference and recall, observable in Figure 7. This suggests that this nest localization method remains effective across all temperature conditions encountered during the breeding season. Although this absence of correlation might initially seem counter-intuitive, several reasons might explain this observation. The model has been extensively trained on a large dataset comprising nests with a wide range of temperatures across a variety of scenarios, improving the model's robustness [72]. Moreover, the ACE and SSE features of the FLIR camera potentially made

all heat spots appear similarly white, regardless of their actual temperature. This may have made the model rely more on other features for detection, such as nest coverage and the specific shape of heat spots.

While thermal cameras can be used both day and night, as demonstrated by Gauthreaux et al., all training and evaluation data of this approach is gathered during warm sunny days [73]. This method's effectiveness in various weather conditions is to be determined.

It should also be noted that meadow birds do not breed their eggs the first days after laying, which means they aren't warmed up and hard to detect [16].

4) Resources required: The thermal camera used in this study comes with a considerable cost, making this method financially less attractive [54]. On the other hand, it requires less resources than the drone-based method, as this approach doesn't need a drone nor two operators. Research should be conducted into the compatibility of more affordable thermal cameras, assessing how differences in thermal resolution and absence of ACE, DDE and SSE features might impact performance. As measurements conducted with the Hikvision thermal camera were discarded, this camera's resolution might serve as a minimum baseline.

The detection model's operating speed of 45 fps highlights its real-time capabilities [74]. With the thermal camera capturing videos at 10 fps, this could allow for implementation of this method on more affordable and less powerful hardware, potentially reducing financial costs.

B. Bird tracking algorithm using depth data

1) Bird nests saved: The average confidence of detections is 0.822 $\pm$ 0.094. While the *FDR* is notably high at 0.517, these FP don't result in unnecessary nest circumventions, hence NPER = 0. The tracking algorithm removes all trajectories from which the first and last bounding boxes have an IoU that is too high, eliminating standing still objects and false positive detections.

It should be noted that unnecessary circumventions can still occur if birds take flight from locations other than nests. The bird density in the Netherlands is on average 0.77 for every hectare of meadow during the breeding season [1], [75]. Assuming all fly up when a mower approaches, this could potentially lead to numerous unnecessary no-mow zones, greatly affecting this method's efficiency and complicating the mowing process.

Additionally, meadow bird species like the godwit and oystercatcher fly away prematurely when disturbed, sometimes even with the disturbance still a few hundred meters away [17]. This behavior introduces the risk of still inadvertently mowing nests, which can also occur when birds are not on their nests. As the nest occupancy of meadow birds is on average 76.9 $\pm$ 2.7% during the day, almost a quarter of nests is still at risk of destruction [21]. If a bird feels threatened and does not take flight at all, this method would also fail to ascertain the location of its nest [16].

2) Localization accuracy: The average localization error of this method is 0.794 $\pm$ 0.578 meters, with no statistical significant difference between take-off locations inside or outside the camera's FOV. This error is notably lower than the one of the drone-based method and shows less variation. There is a significant correlation between grass length and localization error, suggesting that grass length has a crucial impact on localization accuracy.

In Figure 9 localization errors range from 0.139 to 1.672 meters. Excluding data points where $n < 10$, the algorithm can localize nests from distances of 2.110 up to 5.385 meters with an average error of 0.419 $\pm$ 0.223 meters. This indicates an improved detection range compared to the previous method and suggests that the actual detection range could be even larger, as the current data represents only the boundaries of the tested range. In Figure 9 there is a significant negative correlation between the number of data points available for extrapolation and the localization error. The number of data points is equal to the amount of detections within the camera's depth range. This is determined by the lateral distance, as evidenced by the significant correlations between lateral distance and both recall and detection confidence. The lateral distance affects the angle at which the drone is perceived by the camera, influencing its visibility and attributing to this correlation.

Volunteers from collectives Lopikerwaard and Den Hâneker highlighted that localization accuracy may inherently be limited for this method. Certain

meadow bird species, such as the lapwing and the redshank, tend to walk several meters from their nests before taking flight to lead predators away [16]. No literature is available about typical distances meadow birds might travel before flying or how this behavior varies among different species. Further observational studies are required to assess how this behavior impacts localization accuracy.

3) Operational constraints: Reliable bird detection in visual data is not possible during the night, hence bird nest localization can only be carried out during daytime [73]. It is recommended to collect RGBD data in various weather conditions and conduct additional research into this method's effectiveness in these conditions. Moreover, as mentioned in the previous section, meadow birds do not breed their eggs the first days after laying [16]. During this period, their nests cannot be localized using this method.

4) Resources required: It was initially assumed that this method could easily be implemented in the available RGB cameras of autonomous mowers, but insights of this research underscored their limitations. They are tilted too much downwards to provide proper field of view, capturing only birds directly in front of the mower. In addition, the cameras' frame rate of 10 fps limits the number of frames a bird would be visible, potentially affecting the ability to extrapolate its trajectory. For these reasons, it is recommended to employ a separate RGBD camera, specifically for the purpose of bird trajectory extrapolation. When mounting to an autonomous mower, this study's findings can be taken into account for optimal camera configurations. Notably, the financial costs of the RGBD camera used in this research are substantially lower than those of the thermal camera and the drone-based methods [64].

This method, operating at 17 fps, demands more computational capacity compared to the thermal approach, which runs at 45 fps on the same hardware. While both involve object detection, this method also identifies tracks, extracts depth data and extrapolates trajectories. With its processing speed lower than the video capturing rate of 30 fps, its real-time applications might be limited [74]. This raises the need for more powerful hardware, increasing financial costs.

C. Limitations and future considerations

In this research, alterations and simplifications are made in an effort to replicate eggs, nests and breeding birds. Store-bought chicken eggs probably have other physical and thermal characteristics than bird eggs, but no literature is available about their differences. Bird eggs are warmed up by parent birds, which differs from how the chicken eggs were heated with water at a specific temperature. Moreover, as the birds fly away when a mower arrives, it is uncertain how quickly they cool down. While temperature differences were kept as realistic as possible during the experiments, it should be noted that cooling rates and core temperatures between bird eggs and chicken eggs might differ. Additionally, the bird nests in this study were made by humans in an effort to replicate birds' nest building techniques as illustrated in Figure 1. Naturally, there are some differences between the two, observable in Appendix XV. Gathering thermal data of actual bird nests is essential to investigate the real-world capabilities of the bird nest localization model.

Flying birds were replicated by a drone, which has a fixed shape and is therefore easier to detect, in contrast to birds which appear different each time they flap their wings [76], [77]. The drone's flying pattern is also substantially distinctive with respect to birds', as can be observed in Figure 13in Appendix XI [22]. This raises the need for re-examination of this method using data from actual birds, investigating their flying patterns and corresponding regression models. This could potentially have a large impact on localization accuracy.

Other limitations were introduced by the thermal camera mounted to bars connected to a quad. This setup intuitively gives more camera movements than when attached to a mower, as the quad is more sensitive to small hills in the meadows and the bars amplified these movements. This resulted in videos with more motion and possibly less clear heat spots. Data should therefore be gathered with a thermal camera mounted on an autonomous mower to assess the model's real-world performance.

Feasibility of the trajectory extrapolation method was tested using a static camera, which eliminates a lot of challenges that moving cameras face [71]. With a camera mounted on a moving mower, a more extensive bird detection model has to be trained to

account for variations in size, orientation and viewpoint of birds [78]. More importantly, trajectories might appear differently when the camera itself is moving. Other limitations were introduced by the restricted depth range of the RGBD camera used in this study. Using depth cameras with larger ranges could extend the localization distance and improve accuracy.

Finally, the SORT algorithm cannot handle occlusions; if two birds would overlap it loses track of their trajectories [47]. Given that the visual data examined in this study never features more than one flying object, this limitation isn't a problem here. Nonetheless, more advanced and robust trackers should be explored for utilization in real-world scenarios [79], [80].

While enhancing the dataset can significantly improve the models' performance, further optimization should also be considered. Experimenting with different architectures, hyperparameters and augmentation approaches than discussed in this research could yield better outcomes [10]. Other parameters values and augmentation combinations should also be investigated [11], [81].

Despite this study's focus on individual methods, future research might examine combinations of thermal and RGBD cameras. Such a combined system could benefit from an increased distance at which nests can be detected, while simultaneously reducing the number of false positives.

VII. Conclusion

This research can be concluded by answering the research question: *In terms of number of bird nests saved, localization accuracy, operational constraints and resources required, which method — thermal camera based, RGBD camera based, or drone-based — proves to be the most effective for automated real-time bird nest localization in grasslands?*

Integrating autonomous mowers with a thermal camera has proven effective for real-time bird nest detection, achieving an average detection confidence of 0.734. However, the current setup yields an NPER of 0.1 due to false positives, leading to 1 unnecessary circumvention for every 10 nests. This approach offers the potential for highly accurate localization and is effective across all temperature conditions present during the breeding season, both day and night. Its performance in conditions other than sun remains uncertain. It can localize nests from an average distance of 1.836 ± 0.636 meters to the mower. Alternative thermal cameras should be explored to make this method more financially appealing.

The bird trajectory extrapolation method inherently performs well in terms of number of nests detected. Yet, the overall number can be compromised by the absence of birds from their nests. The tracking algorithm effectively filters out false positives generated by the detection model, hence NPER = 0, but unnecessary no-mow zones can still be caused by birds flying up from locations other than nests. Bird nest locations can be predicted with an average localization error of 0.794 ± 0.578 meters and from 2.110 up to 5.385 meters distance. This method is limited to daytime operations and its real-time applications are constricted. Implementing a separate RGBD camera for trajectory extrapolation is recommended, yet financial resources required are considerably lower than those of the other two methods. At the same time, its computational resources needed are higher.

With the drone-based method, approximately 80% of bird nests are saved with NPER of 2. The average localization error is 3.344 ± 0.773 meters, which is considerably higher than the other two methods. This method is constrained to specific application times such as sunrise. Furthermore, localization must be executed every time before mowing, to ensure all newly created bird nests are protected. This method requires more resources than the other two, including a drone and two operators for the localization effort.

In summary, an automated thermal-based mower-integrated localization system outperforms both the RGBD camera and drone-based approaches in number of bird nests saved, localization accuracy and operational constraints, but not in financial resources required. Its effectiveness and accuracy in detecting bird nests make it a promising solution for real-time bird nest localization in grasslands, but challenges remain due to unnecessary no-mow zones caused by false positives. Real-world evaluations using thermal data from autonomous mowers detecting actual bird nests in various weather conditions are recommended.

VIII. Acknowledgments

Firstly, I would like to thank Yke Bauke Eisma from Delft University of Technology for his daily supervision and insightful comments throughout my graduation. Thank you to Dirk-Jan Boonstra and Jasper Wijkhuizen from Lely for their enthusiasm about this project, interesting conversations on a day-to-day basis and continuous feedback during the coding and writing of this thesis. I owe many thanks to Tom Kortleve and Arjan de Jong from agricultural collective Den Hâneker, who helped me with the drone-based localization effort and generously lent me their thermal camera. Without their belief in this project, it would not have been possible. Their support with the fieldwork, expert knowledge on the subject and inexhaustible eagerness to answer my questions were invaluable contributions. I would also like to thank Korstiaan Blokland and Karel van den Berg for defining this research topic and their help with the practical execution of it. Finally, I would like to thank everyone at Blokland Dairy for their assistance with the measurements and for graciously welcoming me into their home.

References

[1] E. Kleyheeg, T. Vogelzang, I. van der Zee, and M. van Beek, *Boerenlandvogelbalans 2020*. Sovon Vogelonderzoek Nederland and LandschappenNL, 2020.

[2] A. Boele, J. Vergeer, J. Van Bruggen, *et al.*, *Broedvogels in Nederland in 2022*. Nijmegen: Sovon Vogelonderzoek Nederland, 2023.

[3] N. van Eekeren, J. de Wit, C. Versteeg, *et al.*, *Winst Weidevogels; Weidemaatregelen voor (functionele agro-)biodiversiteit*. Louis Bolk Instituut, 2022.

[4] BirdLife International, *European Red List of Birds 2021*. Luxembourg: Publications Office of the European Union, 2021.

[5] Ministerie van Economische Zaken en Klimaat, *Wet natuurbescherming, Artikel 3.1. Beschermingsregime soorten Vogelrichtlijn*, ID BWBR0037552, valid from July 1st 2021 to the present. [Online]. Available: https://wetten.overheid.nl/BWBR0037552/2021-07-01/0#Hoofdstuk3_Paragraaf3.1.

[6] Collectief Alblasserwaard/Vijfheerenlanden, *Beheerpakketten 2023-2028*, IntelRealSense/librealsense, 2023.

[7] W. Teunissen, C. Klok, D. Kleijn, and H. Schekkerman, *Factoren die de overleving van weidevogelkuikens beïnvloeden*. Beek-Ubbergen: Vogelonderzoek Nederland, 2008.

[8] Lely. "Lely is introducing a unique concept for autonomous feeding with fresh grass." (2020), [Online]. Available: https://www.lely.com/press/2020/10/06/lely-introducing-unique-concept-autonomous-feeding/ (visited on 04/17/2023).

[9] Lely. "The Lely Exos fresh grass concept." (2020), [Online]. Available: https://www.lely.com/exos/ (visited on 04/17/2023).

[10] S. J. Hong, Y. Han, S. Y. Kim, A. Y. Lee, and G. Kim, "Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery," *Sensors*, vol. 19, no. 7, 2019. DOI: 10.3390/s19071651.

[11] H. Li, Y. Dong, Y. Liu, and J. Ai, "Design and implementation of UAVs for bird's nest inspection on transmission lines based on deep learning," *Drones*, vol. 6, no. 9, pp. 252–270, 2022. DOI: 10.3390/drones6090252.

[12] A. McKellar, N. Shephard, and D. Chabot, "Dual visible-thermal camera approach facilitates drone surveys of colonial marshbirds," *Remote Sensing in Ecology and Conservation*, vol. 7, no. 2, pp. 214–226, 2021. DOI: 10.1002/rse2.183.

[13] M. Israel and A. Reinhard, "Detecting nests of lapwing birds with the aid of a small unmanned aerial vehicle with thermal camera," in *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2017, pp. 1199–1207. DOI: 10.1109/ICUAS.2017.7991393.

[14] R. Buiter, *Liever dronen dan lopen*, Vogelbescherming, 2023. [Online]. Available: https://www.vogelbescherming.nl/actueel/bericht/liever-dronen-dan-lopen#:~:text=Broedende%20sterns%20en%20andere%20kolonievogels,dan%20er%20in%20werkelijkheid%20zijn (visited on 08/08/2023).

[15] L. Zhang, F. Peng, P. Cao, and W. Ji, "An improved three-dimensional DV-Hop localization algorithm optimized by adaptive cuckoo search algorithm," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 13, no. 2, pp. 102–118, 2017. DOI: 10.3991/ijoe.v13i02.6358.

[16] A. van Paassen and O. Vloegraven, *Handvat voor Weidevogelbescherming*. Stichting Landschapsbeheer Nederland, 1995.

[17] B. R.G., M. F., and N. J., *Handleiding Sovon nestonderzoek*. Nijmegen: Sovon Vogelonderzoek Nederland, 2020.

[18] I. Lamprecht and E. Schmolz, "Thermal investigations of some bird nests," *Thermochimica Acta*, vol. 415, no. 1-2, pp. 141–148, 2004. DOI: 10.1016/j.tca.2003.08.029.

[19] A. Santangeli, Y. Chen, E. Kluen, R. Chirumamilla, J. Tiainen, and J. Loehr, "Integrating drone-borne thermal imaging with artificial intelligence to locate bird nests on agricultural land," *Scientific Reports*, vol. 10, 2020. DOI: 10.1038/s41598-020-67898-3.

[20] E. Faye, O. Dangles, and S. Pincebourde, "Distance makes the difference in thermography for ecological studies," *Journal of Thermal Biology*, vol. 56, 2016. DOI: 10.1016/j.jtherbio.2015.11.011.

[21] F. Jongbloed, H. Schekkerman, and W. Teunissen, "Verdeling van de broedinspanning bij kieviten," *Limosa*, vol. 79, no. 2, 2006.

[22] H. Ling, G. Mclvor, G. Nagy, *et al.*, "Simultaneous measurements of three-dimensional trajectories and wingbeat frequencies of birds in the field," *Journal of the Royal Society Interface*, vol. 15, 2018. DOI: 10.1098/rsif.2018.0653.

[23] P. Rajeshwari, P. Abhishek, P. Srikanth, and T. Vinod, "Object detection: An overview," *International Journal of Trend in Scientific Research and Development (IJTSRD)*, vol. 3, no. 3, pp. 1663–1665, 2019. DOI: 10.31142/ijtsrd23422.

[24] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, 2023. DOI: 10.48550/arXiv.1905.05055.

[25] P. Adarsh, P. Rathi, and M. Kumar, "YOLO v3-Tiny: Object detection and recognition using one stage improved model," in *6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 687–694. DOI: 10.1109/ICACCS48705.2020.9074315.

[26] N. Nguyen, T. Do, T. Ngo, and D. Le, "An evaluation of deep learning methods for small object detection," *Journal of Electrical and Computer Engineering*, 2020. DOI: 10.1155/2020/3189691.

[27] A. Dhillon and G. Verma, "Convolutional neural network: A review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, pp. 85–112, 2020. DOI: 10.1007/s13748-019-00203-0.

[28] C. Chen, M. Liu, O. Tuzel, and J. Xiao, "R-CNN for small object detection," in *Computer Vision – ACCV 2016*, S. Lai, V. Lepetit, K. Nishino, and Y. Sato, Eds., 2016, pp. 214–230. DOI: 10.1007/978-3-319-54193-8_14.

[29] A. Benjumea, I. Teeti, F. Cuzzolin, and A. Bradley, "YOLO-Z: improving small object detection in yolov5 for autonomous vehicles," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021. DOI: 10.48550/arXiv.2112.11798.

[30] Y. Liu, P. Sun, N. Wergeles, and Y. Shang, "A survey and performance evaluation of deep learning methods for small object detection," *Expert Systems with Applications*, vol. 172, 2021. DOI: 10.1016/j.eswa.2021.114602.

[31] C. Y. Wang, A. Bochkovskiy, and H. Y. Mark Liao, *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*, 2022. DOI: 10.48550/arXiv.2207.02696.

[32] 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)*, 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.

[33] H. Lou, X. Duan, J. Guo, *et al.*, "DC-YOLOv8: Small-size object detection algorithm based on camera sensor," *Electronics*, vol. 12, no. 10, 2023. DOI: 10.3390/electronics12102323.

[34] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, "UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios," *Sensors*, vol. 23, no. 16, 2023. DOI: 10.3390/s23167190.

[35] J. Terven and D. Cordova-Esparza, *A comprehensive review of YOLO: From YOLOv1 and beyond*, 2023.

[36] G. Jocher, A. Chaurasia, and J. Qiu, *YOLO by Ultralytics*, version 8.0.0, 2023. DOI: 10.5281/zenodo.3908559. [Online]. Available: https://github.com/ultralytics/ultralytics.

[37] I. Kandel and M. Castelli, "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset," *ICT Express*, vol. 6, no. 4, pp. 312–315, 2020. DOI: https://doi.org/10.1016/j.icte.2020.04.010.

[38] E. Chai, L. Ta, Z. Ma, and M. Zhi, "ERF-YOLO: A YOLO algorithm compatible with fewer parameters and higher accuracy," *Image and Vision Computing*, vol. 116, 2021. DOI: 10.1016/j.imavis.2021.104317.

[39] D. Fu, L. Gao, T. Hu, S. Wang, and W. Liu, "Research on safety helmet detection algorithm of power workers based on improved YOLOv5," *Journal of Physics: Conference Series*, vol. 2171, 2022. DOI: 10.1088/1742-6596/2171/1/012006.

[40] J. Svozil and K. Baerenfaller, "Chapter twenty-one - a cautionary tale on the inclusion of variable posttranslational modifications in database-dependent searches of mass spectrometry data," in *Proteomics in Biology, Part B*, ser. Methods in Enzymology, A. Shukla, Ed., vol. 586, Academic Press, 2017, pp. 433–452.

[41] K. Velumani, R. Lopez-Lozano, S. Madec, *et al.*, "Estimates of maize plant density from UAV RGB images using Faster-RCNN detection model: Impact of the spatial resolution," *Plant Phenomics*, May 2021. DOI: 10.34133/2021/9824843.

[42] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 658–666.

[43] S. Salman and X. Liu, *Overfitting mechanism and avoidance in deep neural networks*, Computing Research Repository, 2019. DOI: abs/1901.06566.

[44] V. Silva, R. d. Pinho, M. K. Allahdad, J. Silva, M. J. Ferreira, and L. Magalhães, "A robust real-time leather defect segmentation using YOLO," in *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*, 2023. DOI: 10.23919/CISTI58278.2023.10211894.

[45] B. Sabiri, B. El Asri, and M. Rhanoui, "Mechanism of overfitting avoidance techniques for training deep neural networks.," in *Proceedings of the 24th International Conference on Enterprise Information Systems (ICEIS 2022)*, vol. 1, 2022, pp. 418–427.

[46] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-Learning-Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2012. DOI: 10.1109/TPAMI.2011.239.

[47] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 3464–3468. DOI: 10.1109/ICIP.2016.7533003.

[48] A. Bathija and G. Sharma, "Visual object detection and tracking using YOLO and SORT," *International Journal of Engineering Research Technology (IJERT)*, vol. 8, no. 11, 2019.

[49] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 3464–3468. DOI: 10.1109/ICIP.2016.7533003.

[50] DRONExpert Netherlands. "Natuurdrone 2023." (n.d.), [Online]. Available: https://dronexpert.nl/natuurdrone-2023/ (visited on 09/22/2023).

[51] Ministry of Infrastructure and Water Management, *Drones for EU-residents*. [Online]. Available: https://english.ilent.nl/topics/rpas---remotely-piloted-aircraft-systems-drones.

[52] Trimble Geospatial. "Trimble r2." (n.d.), [Online]. Available: https://geospatial.trimble.com/en/products/hardware/trimble-r2#:~:text=The%20Trimble%C2%AE%20R2%20GNSS,submeter%20to%20centimeter%20positioning%20accuracy. (visited on 09/12/2023).

[53] Hikvision, *Thermal Optical Bi-Spectrum Network Turret Camera*, DS-2TD1217-2/V1, 2019. [Online]. Available: https://www.hikvision.com/my/products/Thermal-Products/Security-thermal-cameras/Turret-series/DS-2TD1217-2-V1/ (visited on 05/22/2023).

[54] Teledyne FLIR, *FLIR Vue Pro R Radiometric Drone Thermal Camera*, Model: Vue Pro R 640, $45\mathrm{deg}\ FOV$, $13mm$, $9Hz$. [Online]. Available: https://www.flir.eu/products/vue-pro-r/?model=436-0023-00S&vertical=suas&segment=oem (visited on 05/22/2023).

[55] A. Kirimtat, O. Krejcar, A. Selamat, and E. Herrera-Viedma, "FLIR vs SEEK thermal cameras in biomedicine: Comparative diagnosis through infrared thermography," *BMC Bioinformatics*, vol. 21, no. 2, 2020. DOI: 10.1186/s12859-020-3355-7.

[56] M. A. Verhoeven, A. J. Loonstra, A. D. McBride, *et al.*, "Variation in egg size of black-tailed godwits," *Ardea*, vol. 107, no. 3, pp. 291–302, 2020. DOI: 10.5253/arde.v107i3.a7.

[57] T. Lislevand, I. Byrkjedal, T. Borge, and G. Sætre, "Egg size in relation to sex of embryo, brood sex ratios and laying sequence in northern lapwings (Vanellus vanellus)," *Journal of Zoology*, vol. 267, no. 1, pp. 81–87, 2006. DOI: 10.1017/S0952836905007260.

[58] British Trust for Ornithology. "Shoveler." (n.d.), [Online]. Available: https://www.bto.org/understanding-birds/birdfacts/shoveler (visited on 06/13/2023).

[59] D. Téo, J. B. Hulscher, and M. Kersten, "Egg size, egg composition and reproductive success in the oystercatcher Haematopus ostralegus," *International Journal of Avian Science*, vol. 142, no. 4, pp. 603–613, 2000. DOI: 10.1111/j.1474-919X.2000.tb04460.x.

[60] British Trust for Ornithology. "Black-tailed Godwit." (n.d.), [Online]. Available: https://www.bto.org/understanding-birds/birdfacts/black-tailed-godwit (visited on 06/13/2023).

[61] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, 2019. DOI: 10.1186/s40537-019-0197-0.

[62] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, 2022, International Conference on Intelligent Engineering Approach(ICIEA-2022). DOI: 10.1016/j.gltp.2022.04.020.

[63] F. Dadboud, V. Patel, V. Mehta, M. Bolic, and I. Mantegh, "Single-Stage UAV Detection and Classification with YOLOV5: Mosaic Data Augmentation and PANet," in *2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2021. DOI: 10.1109/AVSS52988.2021.9663841.

[64] Intel Corporation, *Intel RealSense D400 Series*, tech. rep. No. 337029-013, 2023.

[65] V. Uchôa, K. Aires, R. Veras, A. Paiva, and L. Britto, "Data augmentation for face recognition with CNN transfer learning," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 143–148. DOI: 10.1109/IWSSIP48289.2020.9145453.

[66] N. Cressie and H. Whitford, "How to use the two sample t-test," *Biometrical Journal*, vol. 28, no. 2, pp. 131–148, 1986. DOI: 10.1002/bimj.4710280202.

[67] E. I. Obilor and E. Amadi, "Test for significance of Pearson's correlation coefficient ( )," 2018.

[68] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: Appropriate use and interpretation," *Anesthesia Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018. DOI: 10.1213/ANE.0000000000002864.

[69] P. M. Sedgwick, "Pearson's correlation coefficient," *British Medical Journal*, vol. 345, 2012. DOI: 10.1136/bmj.e4483.

[70] E. S. Pearson, "The test of significance for the correlation coefficient," *Journal of the American Statistical Association*, vol. 26, no. 174, pp. 128–134, 1991. DOI: 10.1080/01621459.1931.10503208.

[71] L. Lecrosnier, R. Khemmar, N. Ragot, *et al.*, "Deep learning-based object detection, localisation and tracking for smart wheelchair healthcare mobility," *International Journal of Environmental Research and Public Health*, vol. 18, no. 91, 2021. DOI: 10.3390/ijerph18010091.

[72] S. Larson, A. Zheng, A. Mahendran, *et al.*, "Iterative feature mining for constraint-based data collection to increase data diversity and model robustness," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds., Association for Computational Linguistics, 2020, pp. 8097–8106. DOI: 10.18653/v1/2020.emnlp-main.650.

[73] S. A. Gauthreaux and J. W. Livingston, "Monitoring bird migration with a fixed-beam radar and a thermal-imaging camera," *Journal of Field Ornithology*, vol. 77, no. 3, pp. 319–328, 2006. DOI: 10.1111/j.1557-9263.2006.00060.x.

[74] D. Andreev, "Real-time frame rate up-conversion for video games: Or how to get from 30 to 60 fps for "free"," in *ACM SIGGRAPH 2010 Talks*, Los Angeles, California: Association for Computing Machinery, 2010. DOI: 10.1145/1837026.1837047.

[75] R. Slaterus, *Weidevogels in het weidevogelkerngebied in 2016*. Nijmegen: Sovon Vogelonderzoek Nederland, 2016.

[76] B. Tobalske, "Biomechanics of bird flight," *Journal of Experimental Biology*, vol. 210, no. 18, pp. 3135–3146, 2007. DOI: 10.1242/jeb.000273.

[77] S. Dhawan, "Bird flight," *Sadhana*, vol. 16, no. 4, pp. 275–352, 1991. DOI: 10.1007/BF02745345.

[78] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: A survey," *Computer Science Review*, vol. 28, pp. 157–177, 2018. DOI: j.cosrev.2018.03.001.

[79] X. Chen, H. Pu, Y. He, *et al.*, "An efficient method for monitoring birds based on object detection and multi-object tracking networks," *Animals*, vol. 13, no. 10, 2023. DOI: 10.3390/ani13101713.

[80] S. Pal, A. Pramanik, J. Maiti, and P. Mitra, "Deep learning in multi-object detection and tracking: State of the art," *Applied Intelligence*, vol. 51, 2021. DOI: 10.1007/s10489-021-02293-7.

[81] S. Prasanna and M. El-Sharkawy, "Hyperparameter optimization for object detection network," in *Proceedings of Seventh International Congress on Information and Communication Technology*, X.-S. Yang, S. Sherratt, N. Dey, and A. Joshi, Eds., Singapore: Springer Nature Singapore, 2023, pp. 761–776.

[82] KNMI. "6-hourly data of soil temperatures station 120, De Bilt." (n.d.), [Online]. Available: https://www.knmi.nl/nederland-nu/klimatologie/bodemtemperaturen (visited on 07/17/2023).

[83] GitHub, *Projection in RealSense SDK 2.0*, IntelRealSense/librealsense, 2023. [Online]. Available: https://github.com/IntelRealSense/librealsense/wiki/Projection-in-RealSense-SDK-2.0#point-coordinates (visited on 09/15/2023).

IX. Drone-based localization outcomes

Nest	Location determined with drone	Location determined with Trimble receiver	Distance [m]
shoveler 1	N 51.90374309° E 4.92008089°	N 51.90376133° E 4.92012383°	3.562
godwit 2	N 51.903888513° E 4.919905304°	N 51.90387011° E 4.91994242°	3.267
godwit 3	N 51.904664473° E 4.920275351°	N 51.90465069° E 4.92021147°	4.643
godwit 4	N 51.905253597° E 4.920622616°	N 51.90522808° E 4.92061128°	2.942
godwit 5	N 51.904720632° E 4.921487012°	N 51.90469067° E 4.92148222°	3.348
godwit 6	N 51.904224333° E 4.922332735°	N 51.90423203° E 4.92236389°	2.302
		Average distance	3.344

Table IX: Nest locations at Blokland Dairy.

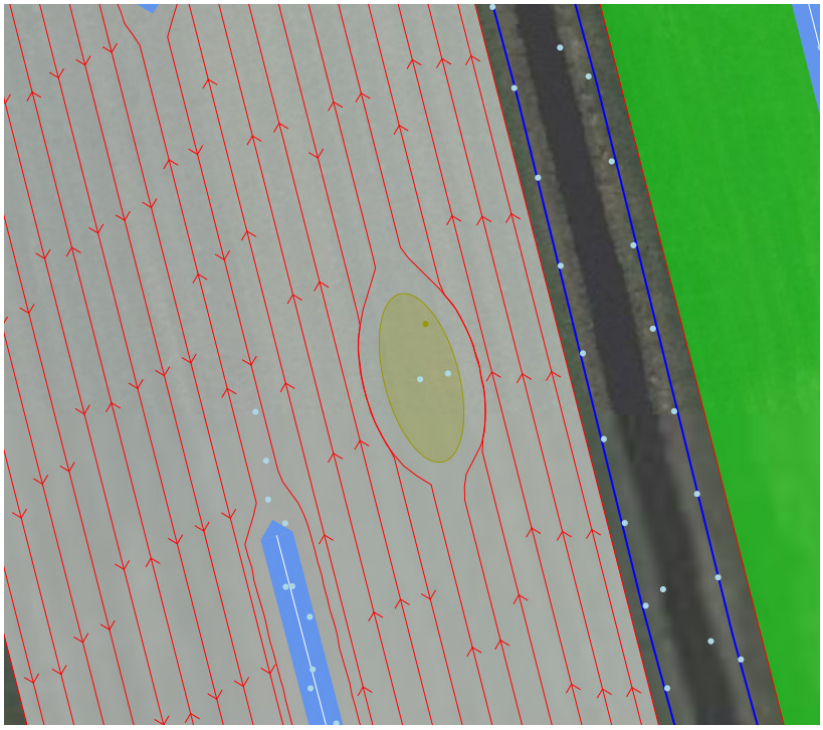


Figure 11: Autonomous mower's route modification around bird nest, with yellow no-mow zone.

X. Egg temperature difference as independent variable

Images captured by a thermal camera are primarily affected by the temperature of the soil at the surface [13]. When the drone-based localization effort was conducted in April, the soil temperature was considerably lower compared to when experiments were concluded in June. The Royal Netherlands Meteorological Institute (KNMI) records the soil temperature every six hours at their weather station in De Bilt, see Table X, which is their closest station to Goudriaan [82]. Given that the soil temperature varies by only 0.5°C across all their stations in the Netherlands, it can assumed that the soil temperatures were relatively similar in De Bilt and Goudriaan.

Table X: Average soil temperatures in De Bilt during all months of the breading season, recorded throughout the day. Data retrieved from [82].

Month	**Average soil temperature**				**Average temperature**
	at 06:00h [°C]	**at 12:00h** [°C]	**at 18:00h** [°C]	**at 24:00h** [°C]	[°C]
April	8.9	9.4	10.6	9.9	9.7
May	15.3	16.5	18.1	16.8	16.7
June	19.9	21.1	22.0	21.2	21.0

Bird eggs have an average temperature of 37.7°C when brooded by a parental bird [18]. The temperature difference ΔT between each of the ground temperatures in Table X and a bird egg is depicted in Table XI. Considering these temperature differences, values as displayed in Table I were selected for the measurements. Larger temperature differences with the surroundings indicate colder days in April and nighttime scenarios, while smaller differences can simulate warmer days in June or midday conditions.

During the measurements conducted on June 22nd, 27th, and 28th, the average ground temperature was 23.1°C. To simulate the various environmental conditions, the grocery store-purchased eggs were subjected to controlled temperatures. These temperatures were equal to the egg temperature differences from Table I plus the ground temperature.

Table XI: Temperature difference ΔT between soil temperatures depicted in Table X and bird eggs.

Month	**Average temperature difference ΔT**				**Average temperature difference**
	at 06:00h [°C]	**at 12:00h** [°C]	**at 18:00h** [°C]	**at 24:00h** [°C]	[°C]
April	28.8	28.3	27.1	27.8	28.0
May	22.4	21.2	19.6	21.0	21.0
June	17.8	16.6	15.7	16.5	16.7

XI. Bird tracking algorithm

The Intel RealSense D455 Depth Camera captures both RGB and depth information. The local coordinate system of this depth camera is depicted in Figure 12.

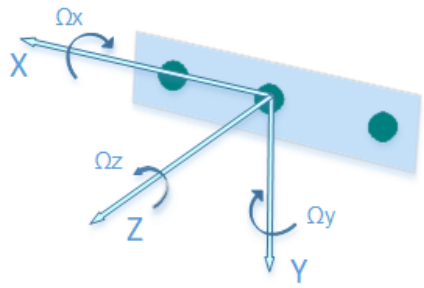


Figure 12: Local coordinate system of Intel D455 Depth Camera, retrieved from [64].

The tracking algorithm is made up of five parts, which include (1) Object detection, (2) Object tracking, (3) Depth retrieval, (4) Global coordinate calculation and (5) Extrapolation of trajectories.

Object detection. The modified YOLOv8 model is employed in the tracking algorithm. In each frame of the visual video, detections are made and stored in a data frame; each row corresponds to a detection, with frame number and xmin, ymin, xmax and ymax coordinates of a bounding box, along with a confidence score.

Object tracking. The SORT algorithm matches new detections to already existing tracks based on an IoU threshold of 0.3 [49]. Its adds a track ID to each row of the data frame. If the IoU of a track's first and last bounding boxes is higher than 0.5, the track is assumed to be caused by a stationary object or false positive and discarded.

Depth retrieval. Remaining detections are grouped by frame number, and while looping through the frames, depth information is retrieved for all detections associated to a frame. As the visual information is captured with a RGB camera and the depth information with two separate infrared sensors, their coordinate systems are distinct. The relationship between the separate 3D coordinate systems is described by the camera extrinsics, containing a rotation and translation matrix. The `rs.rs2_project_color_pixel_to_depth_pixel` function of the librealsense library aligns the pixels from the visual detection with those in the depth stream [83]. The captured depth information provides the distance to the bounding box' center of each detection. By integrating this depth data with the intrinsic parameters of the camera it is possible to map 2D pixel locations into corresponding 3D point locations within the camera's local coordinate system. This transformation is achieved through the use of the `rs2_deproject_pixel_to_point` function of the librealsense library, which also provides undistortion of the depth image. This results in 3D coordinates of the center of each predicted bounding box, within the camera's local coordinate system.

The center of each detection might occasionally lack a corresponding depth value or may be located just beside the drone, resulting in a very high depth value. To address these inconsistencies, depth values of four surrounding pixels are also retrieved, as illustrated in Figure 4. If the measured depth value of the bounding box center exceeds a certain threshold or does not exist, depth values of these four neighboring pixels are assessed. All values above the threshold are discarded, and the depth of the bounding box's center is determined by averaging the remaining values. If all depth values are above the threshold, no

depth information can be associated with the predicted bounding box. The threshold is set to 3.2 meters, based on some trial-and-error tests.

```
if depth_value > threshold:
    valid_depth_values = [depth_value for depth_value in [depth_value1,
    depth_value2, depth_value3, depth_value4] if 0 < depth_value <= threshold]

    if valid_depth_values:
        avg_depth_value = sum(valid_depth_values) / len(valid_depth_values)
        depth_value = avg_depth_value
```

Global coordinate calculation. Until now all derived 3D coordinates are presented within the local coordinate system of the camera, as depicted in Figure 12. The camera configuration used during measurements is achieved by first introducing a translation of two meters in the negative local Y-direction and then a rotation around the X-axis of negative 22.5°. These operations can be described by their respective translation and rotation matrices, as in equation 4, where α = -22.5° and h = 2 meters.

$$\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(\alpha) & -sin(\alpha) & 0 \\ 0 & sin(\alpha) & cos(\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -h \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4}$$

The calculated local 3D point coordinates can be transformed to global coordinates using the inverse of the matrices in equation 4. The origin of this global coordinate system is then located directly beneath the camera, with its Y-axis directed into the ground and the Z-axis along the surface. For ease of interpretation, an additional rotation of 90° around the X-axis is introduced, allowing the Z-axis to orient upwards and the Y-axis to align with the ground. This global coordinate system is depicted in Figure 4. Each frame of the visual stream is now associated with a bounding box around the drone, a distance relative to the camera and a set of 3D coordinates within the global coordinate system, see Figure 4. The trajectory is updated with every frame a detection is made.

Extrapolation of trajectories. Finally, the global X-, Y- and Z-coordinates of the drone in each frame can be used to predict the location it came from. All detections with a depth value higher than 3.2 meters are disregarded, and all detections with a Z-value below a set grass length as well. This is to simulate a drone appearing from the grass, as the need for a landing pad made its departure less bird-like.

When looking at the global coordinates retrieved at each detection, a linear pattern can be distinguished for the first few detections, as depicted in Figure 13. The function `numpy.polynomial.polynomial.polyfit` returns a linear regression model, and the coefficient of a polynomial of degree 1 is found. This polynomial is the least squares fit to the first ten Z-coordinates, if available. A linear regression model is also applied to the X- and Y-coordinates. Subsequently, predictions about the previous and future values of these coordinates can be found using `np.polynomial.polynomial.polyval`. The timestamp at which the Z-regression model crosses Z = 0 is calculated, and this timestamp is used to find the corresponding X- and Y-coordinates at the ground.

```
total_index = len(t)
if total_index <= 10:
    index = total_index
else:
    index = 10

regression_model_z =
np.polynomial.polynomial.polyfit(t_normalized[0:index], z[0:index], deg=1)

z_predicted =
np.polynomial.polynomial.polyval(t_normalized, regression_model_z)

ground_z = 0

equation =
lambda time: np.polynomial.polynomial.polyval(time, regression_model_z)
- ground_z
roots = fsolve(equation)

ground_x = np.polynomial.polynomial.polyval(roots, regression_model_x)
ground_y = np.polynomial.polynomial.polyval(roots, regression_model_y)
```

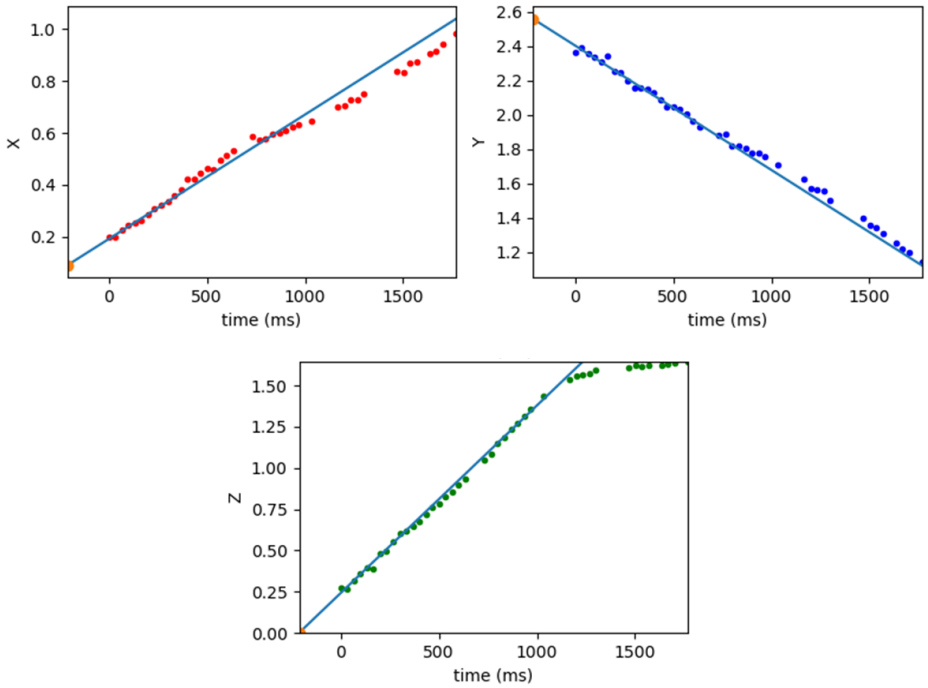
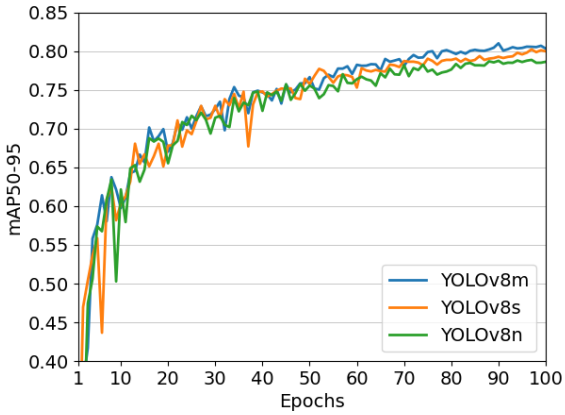
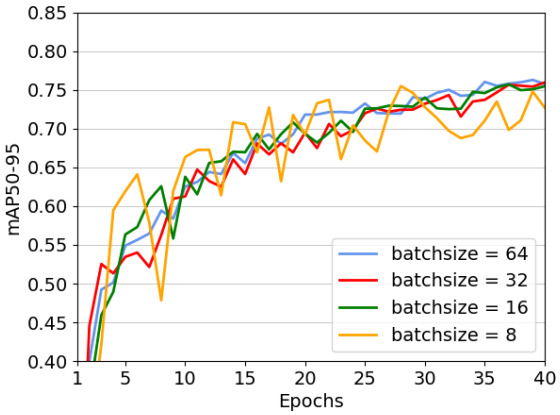


Figure 13: World coordinates of drone departing from X = 0 and Y = 3, with linear regression lines through the first 10 data points. The orange dots indicate the predicted ground timestamp and corresponding coordinates.

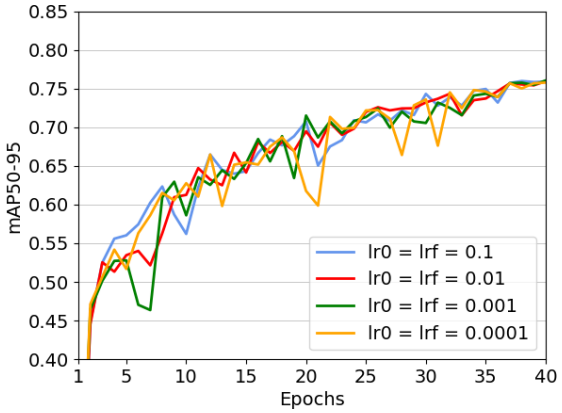
XII. YOLOv8 PARAMETER TUNING



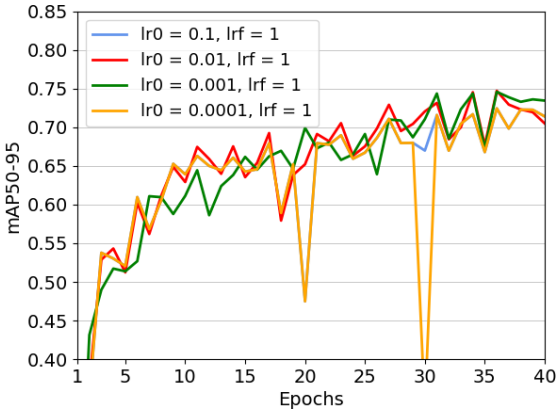
(a) Modelsize, employing a batch size of 16



(b) Batch size

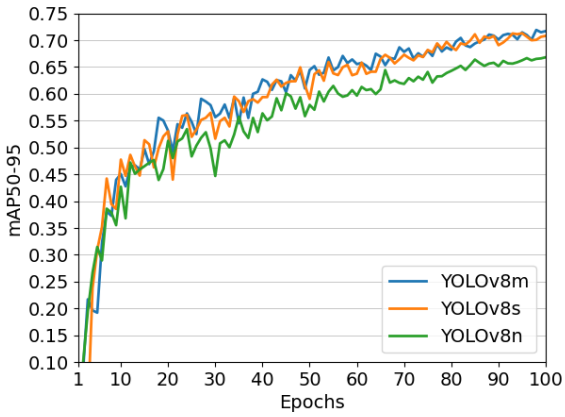


(c) Learning rates with linear schedule

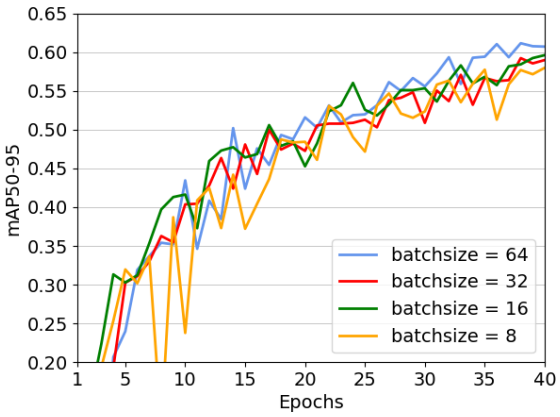


(d) Learning rates with constant schedule

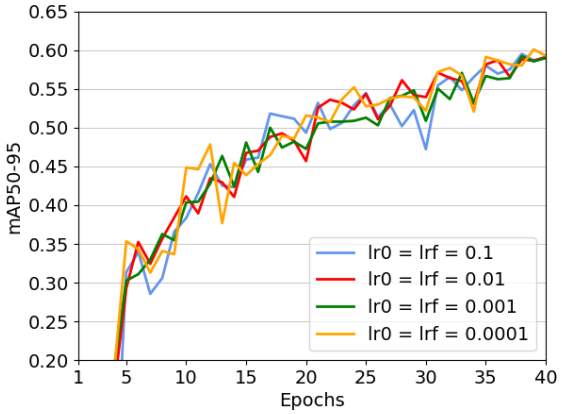
Figure 14: Effect of various parameters on YOLOv8n outcomes trained on thermal dataset. All other parameters are set to their standard values as detailed in Subsection IV-A3.



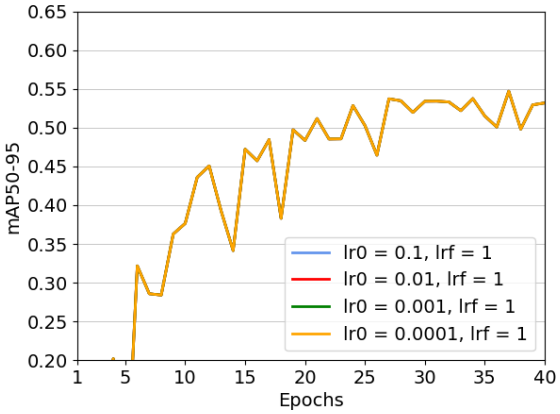
(a) Modelsize, employing a batch size of 16



(b) Batch size



(c) Learning rates with linear schedule



(d) Learning rates with constant schedule

Figure 15: Effect of various parameters on YOLOv8n outcomes trained on visual dataset. All other parameters are set to their standard values as detailed in Subsection IV-B3.

XIII. Data augmentation optimization

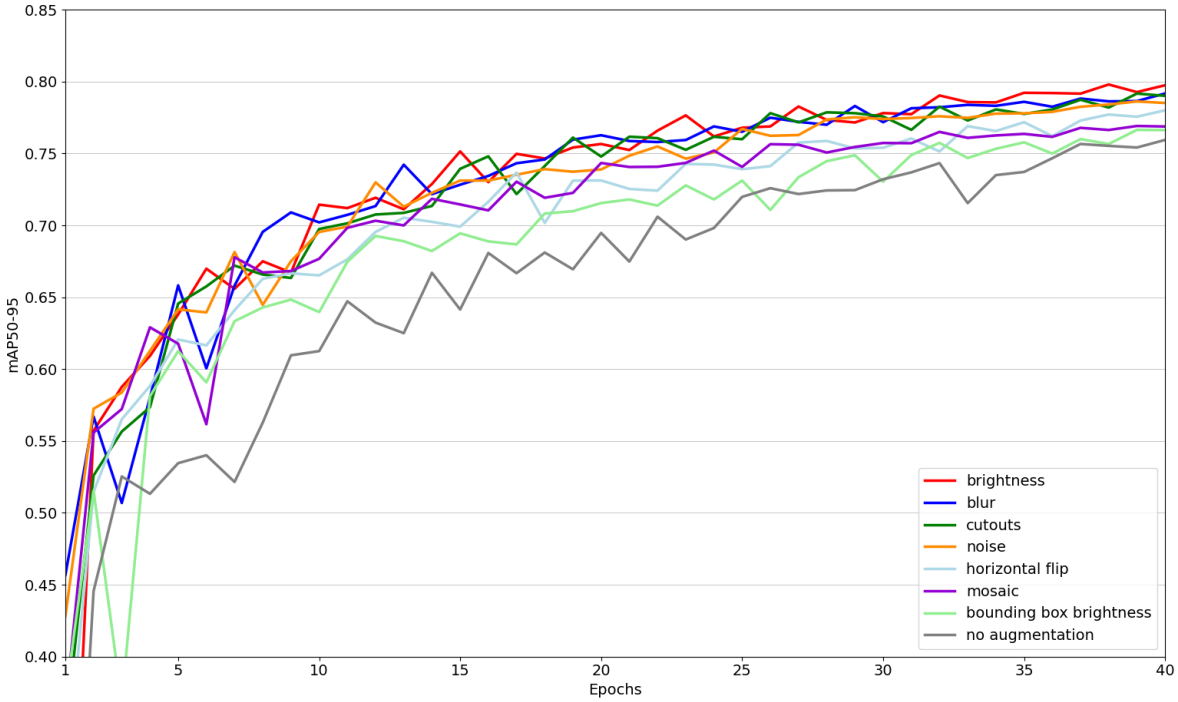
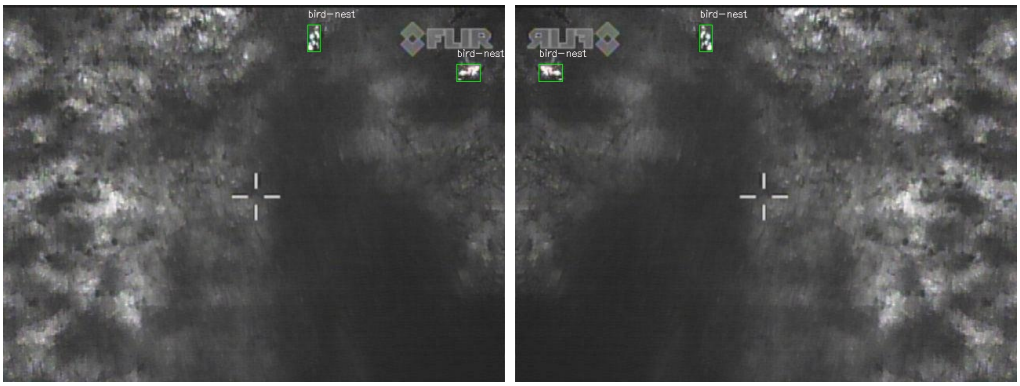


Figure 16: Effect of various data augmentation techniques on detection accuracy of YOLOv8n trained on thermal dataset.

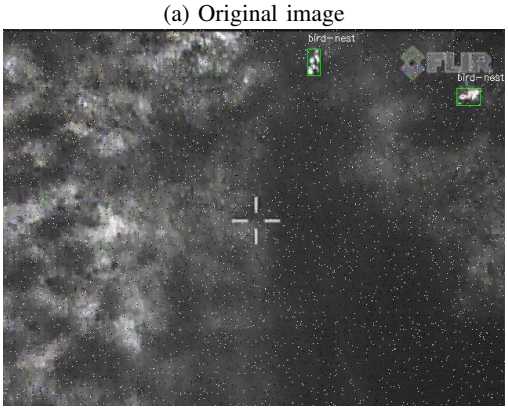
Table XII: Maximum mAP50-95 values of YOLOv8n after 40 epochs of various data augmentation techniques applied to thermal dataset. Their difference compared to no augmentation is also depicted.

Augmentation method	**Maximum mAP50-95 value**	**Difference from no augmentation**	**Percentage difference**
brightness	0.79803	0.03856	4.952
blur	0.79179	0.03232	4.167
cutouts	0.79171	0.03224	4.157
noise	0.78624	0.02677	3.464
horizontal flip	0.78008	0.02061	2.677
mosaic	0.76917	0.00970	1.269
bounding box brightness	0.76647	0.00700	0.917
no augmentation	0.75947	0	0

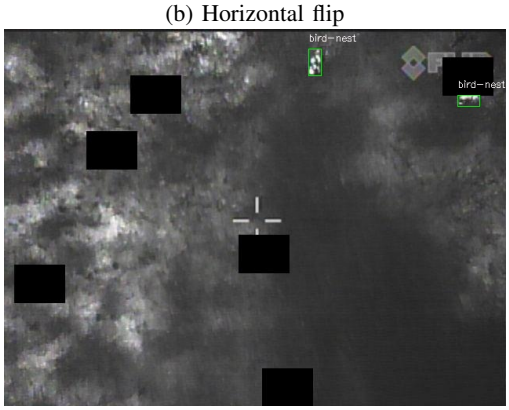


(a) Original image

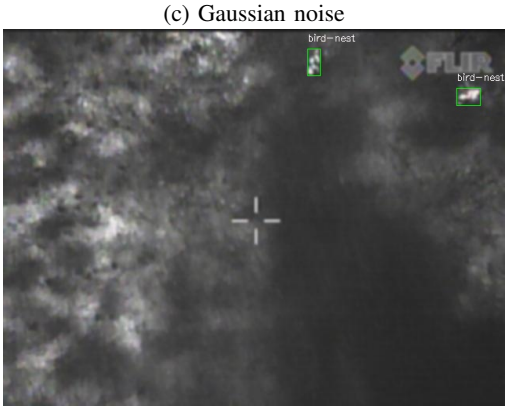
(b) Horizontal flip



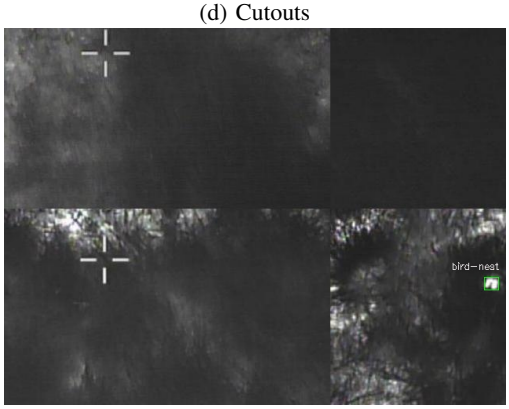
(c) Gaussian noise



(d) Cutouts



(e) Gaussian blurring

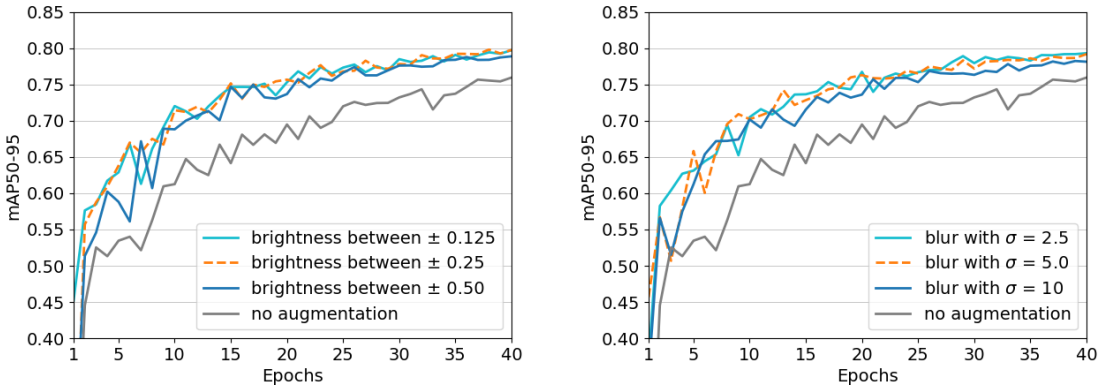


(f) Mosaic

(g) Image brightness

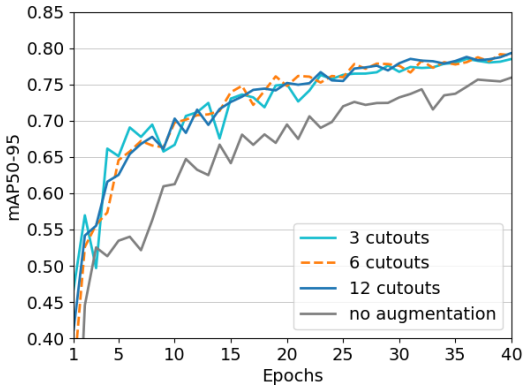
(h) Bounding box brightness

Figure 17: Examples of data augmentation techniques applied to thermal image, including annotations.

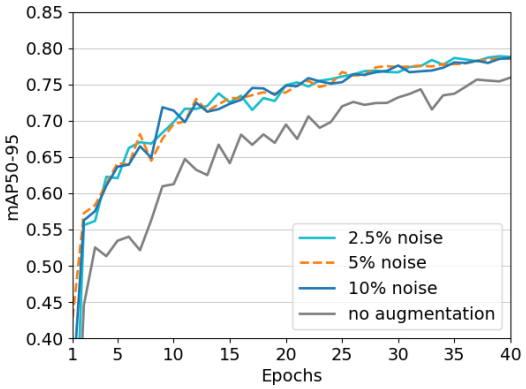


(a) Brightness

(b) Blur



(c) Cutouts



(d) Noise

Figure 18: Effect of parameters associated with data augmentation techniques for YOLOv8n model trained on thermal dataset. Each augmentation method is assessed in three ways; in its default configuration (represented by an orange dotted line), with a lower value for the parameter associated to it (in light blue) and with a higher value (in dark blue).

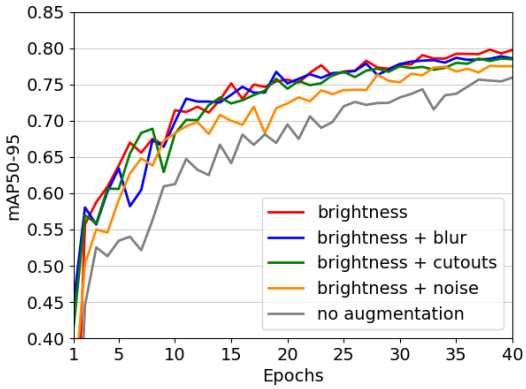


Figure 19: Effect of brightness adjustments between 0.75 and 1.25 combined with best performing variants of other data augmentation methods, for YOLOv8n trained on thermal dataset.

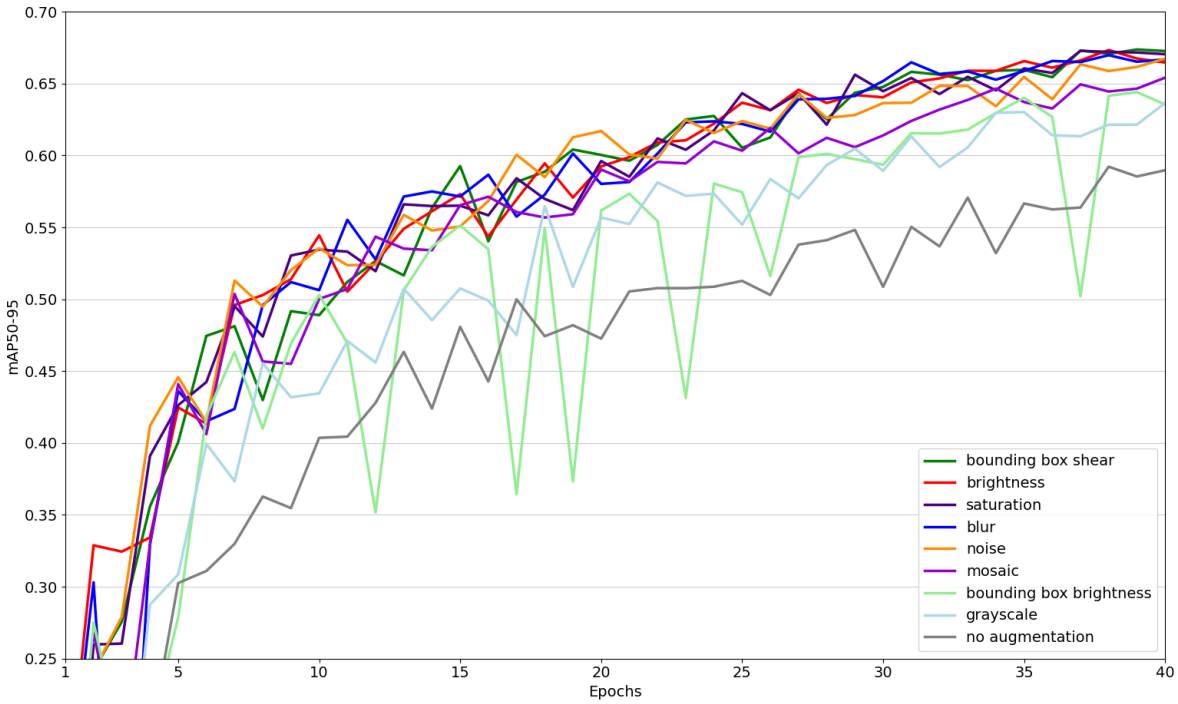
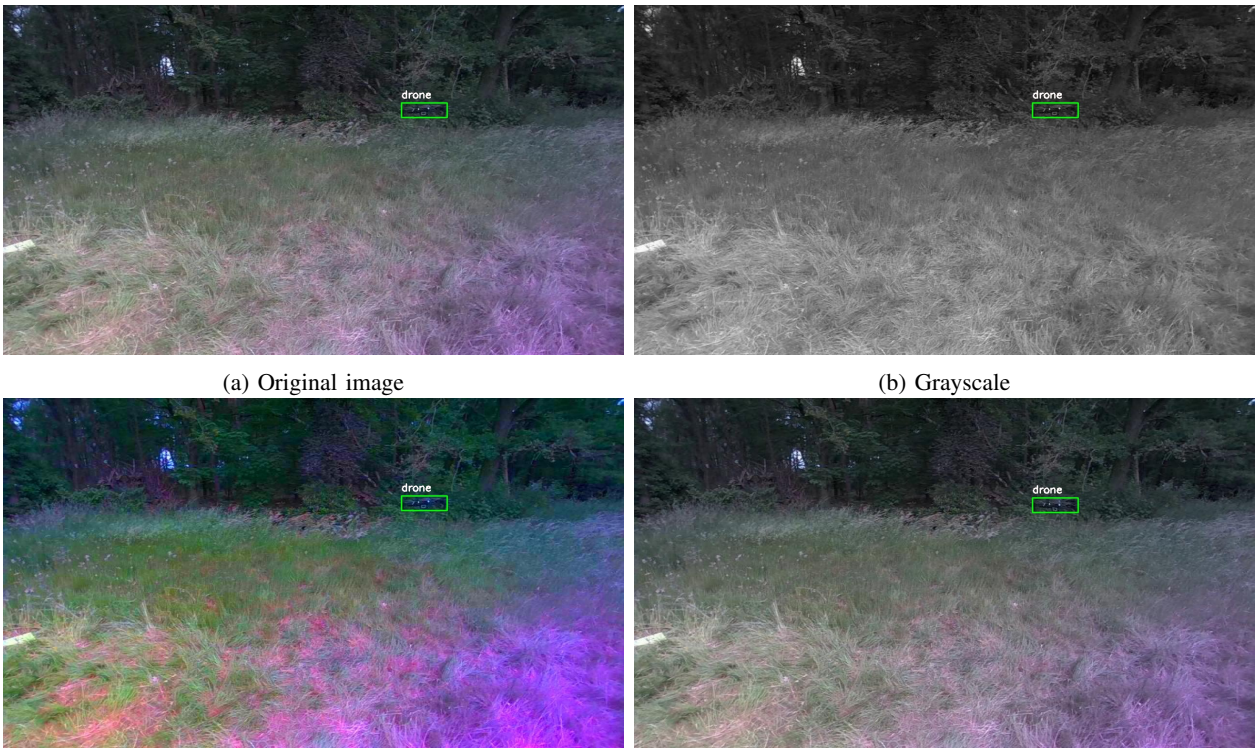


Figure 20: Effect of various data augmentation techniques on YOLOv8n performance trained on visual dataset.

Table XIII: Maximum mAP50-95 values of YOLOv8n after 40 epochs of various data augmentation techniques applied to visual dataset. Their difference compared to no augmentation is also depicted.

Augmentation method	**Maximum mAP50-95 value**	**Difference from no augmentation**	**Percentage difference**
shear	0.67377	0.08154	12.882
brightness	0.67341	0.08118	12.828
saturation	0.67284	0.08061	12.744
blur	0.66981	0.07758	12.294
noise	0.66743	0.07520	11.940
mosaic	0.65419	0.06196	9.942
bounding box brightness	0.64411	0.05188	8.393
grayscale	0.63658	0.04435	7.218
no augmentation	0.59223	0	0



(a) Original image (b) Grayscale

(c) Saturation (d) Bounding box shear

Figure 21: Examples of data augmentation techniques applied to visual image, including annotation.

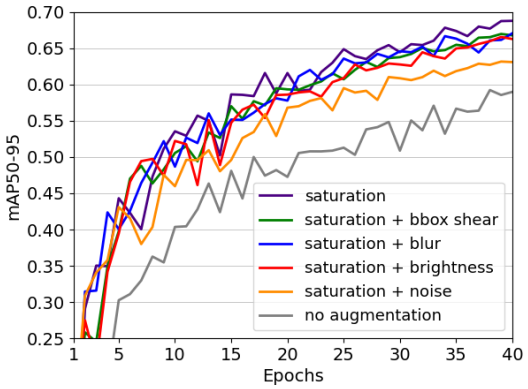
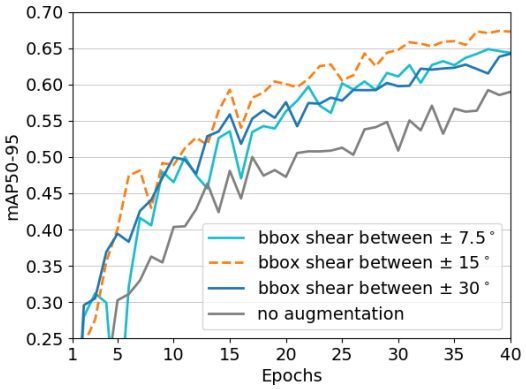
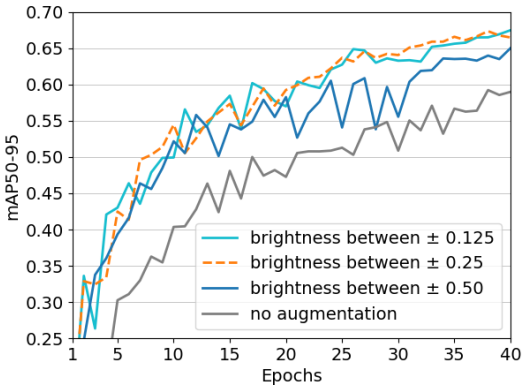


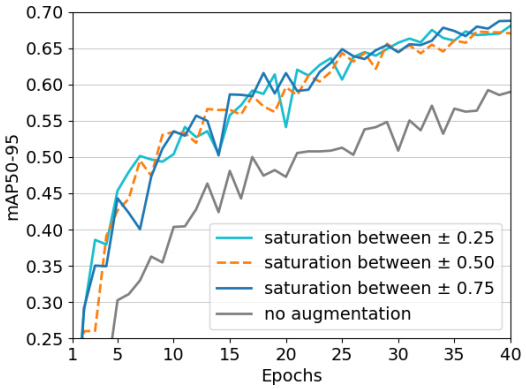
Figure 22: Effect of saturation adjustments between 0.25 and 1.75 combined with best performing variants of other data augmentation methods, for YOLOv8n model trained on visual dataset.



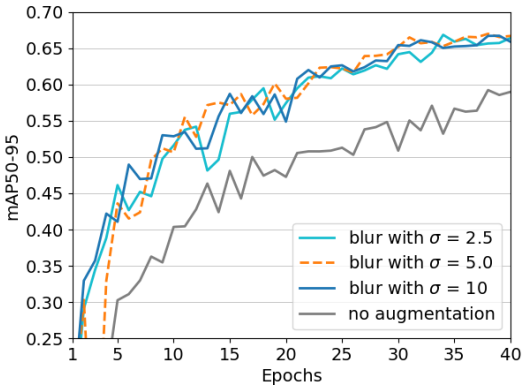
(a) Bounding box shear, both horizontally and vertically



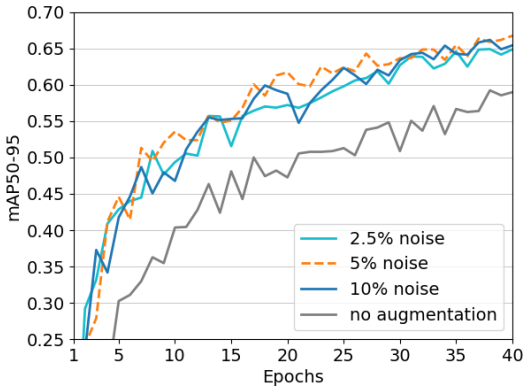
(b) Brightness



(c) Saturation



(d) Gaussian blur



(e) Gaussian noise

Figure 23: Effect of parameters associated with data augmentation techniques on YOLOv8n trained on visual dataset outcomes. Each method is evaluated at its standard configuration (represented by an orange dotted line), a lower value (light blue) and a higher value (dark blue)

XIV. SUPPLEMENTARY RESULTS

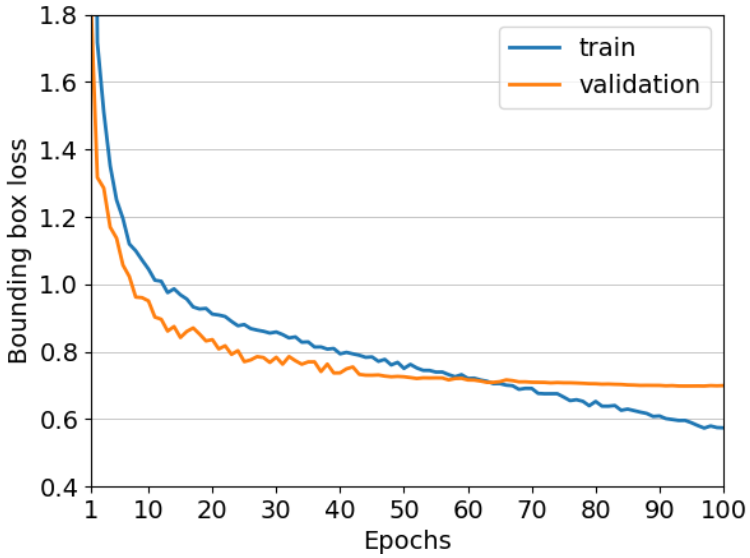


Figure 24: Bounding box loss during training and validation stages of YOLOv8m trained on thermal dataset with brightness adjustments.The graphs intersect at 63 epochs, after which overfitting occurs.

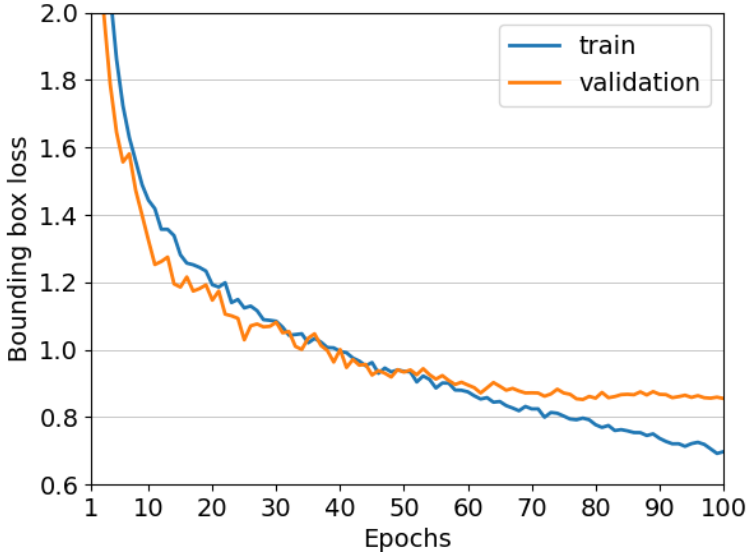
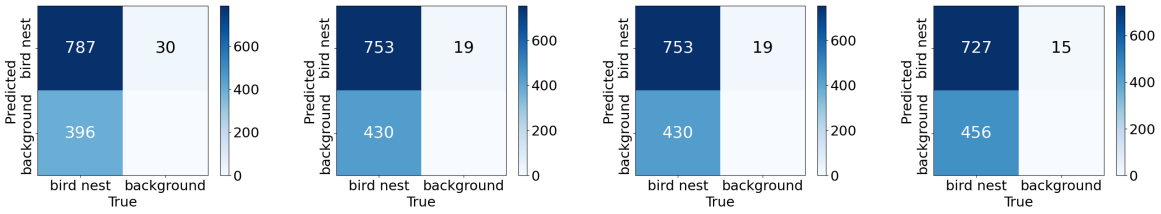


Figure 25: Bounding box loss of YOLOv8m trained on visual dataset with saturation adjustments, during training and validation phases. After 49 epochs, the validation loss starts increasing while the training loss continues to decrease.



(a) Threshold of 0.10 (b) Threshold of 0.20 (c) Threshold of 0.40 (d) Threshold of 0.50

Figure 26: Confusion matrices of thermal test set, with different confidence thresholds for detections.

Table XIV: Pearson correlation coefficient r between dependent variables related to thermal bird nest detection model. Statistical significant correlations are underlined.

r	**P**	**R**	**mAP50-95**	**FDR**	**First detection distance**	**Detection confidence**
P	-	0.223	0.745	-0.343	0.034	0.496
R	0.223	-	0.861	0.112	0.589	0.789
mAP50-95	0.745	0.861	-	0.071	0.083	0.399
FDR	-0.343	0.112	0.071	-	-0.180	0.780
First detection distance	0.034	0.589	0.083	-0.180	-	0.107
Detection confidence	0.496	0.789	0.399	0.780	0.107	-

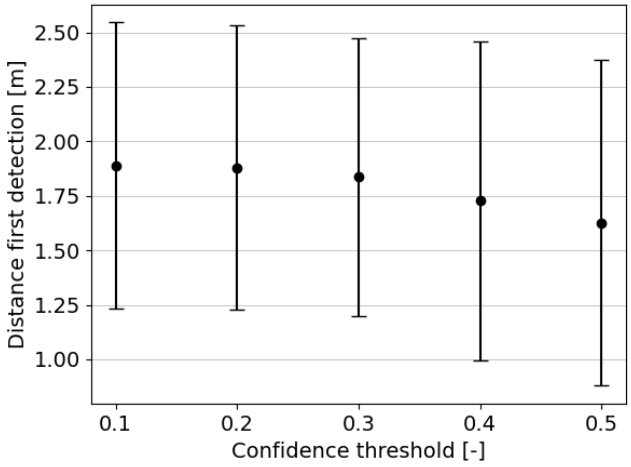
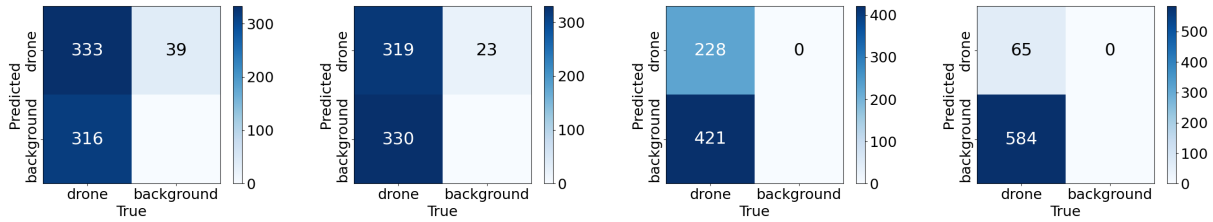


Figure 27: Effect of confidence threshold on the distance a nest is first detected. The Pearson correlation coefficient r equals -0.954, which is proved to be statistically significant ($p < 0.05$).



(a) Threshold of 0.65 (b) Threshold of 0.70 (c) Threshold of 0.80 (d) Threshold of 0.85

Figure 28: Confusion matrices of visual test set, with different confidence thresholds for detections.

Table XV: Pearson correlation coefficient r between dependent variables related to bird trajectory extrapolation method. Statistical significant correlations are underlined.

r	**P**	**R**	**mAP50-95**	**FDR**	**Localization error**	**Detection confidence**
P	-	-0.041	0.734	-0.542	0.165	0.136
R	-0.041	-	0.800	0.093	0.269	0.428
mAP50-95	0.734	0.800	-	0.127	0.254	0.440
FDR	-0.542	0.093	0.127	-	0.683	0.580
Localization error	0.165	0.269	0.254	0.683	-	0.025
Detection confidence	0.136	0.428	0.440	0.580	0.025	-

XV. THERMAL AND VISUAL IMAGES OF EVALUATION NESTS

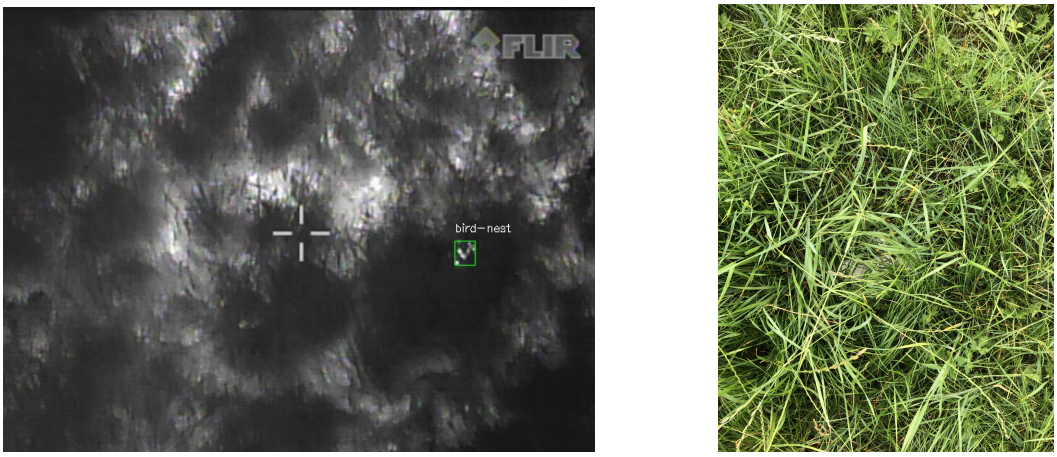


Figure 29: Nest 1, with 5 eggs and ΔT = 14.5°C.

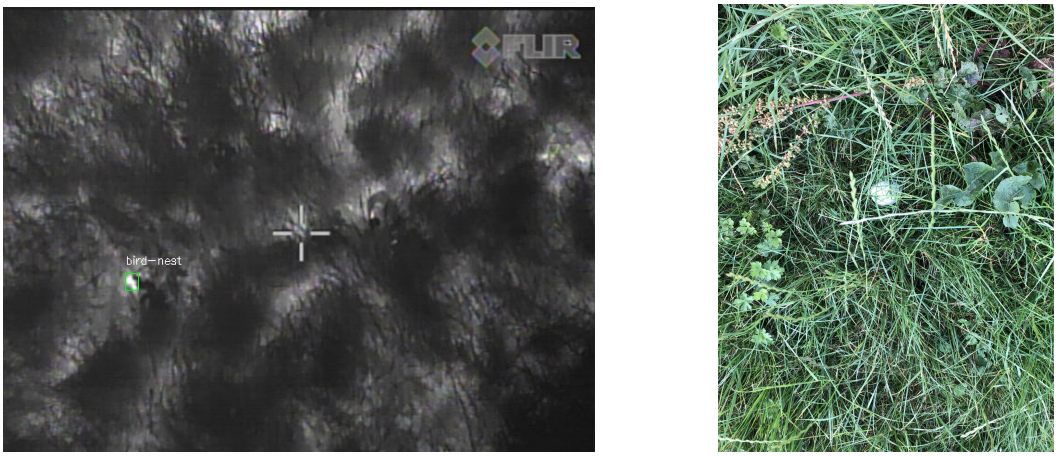


Figure 30: Nest 2, with 1 egg and ΔT = 14.5°C.

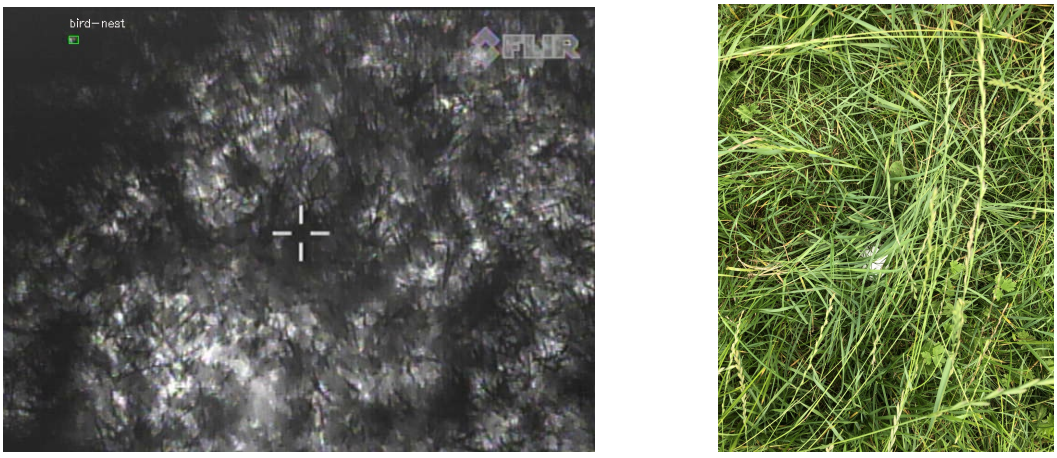


Figure 31: Nest 3, with 1 egg and ΔT = 20.5°C.

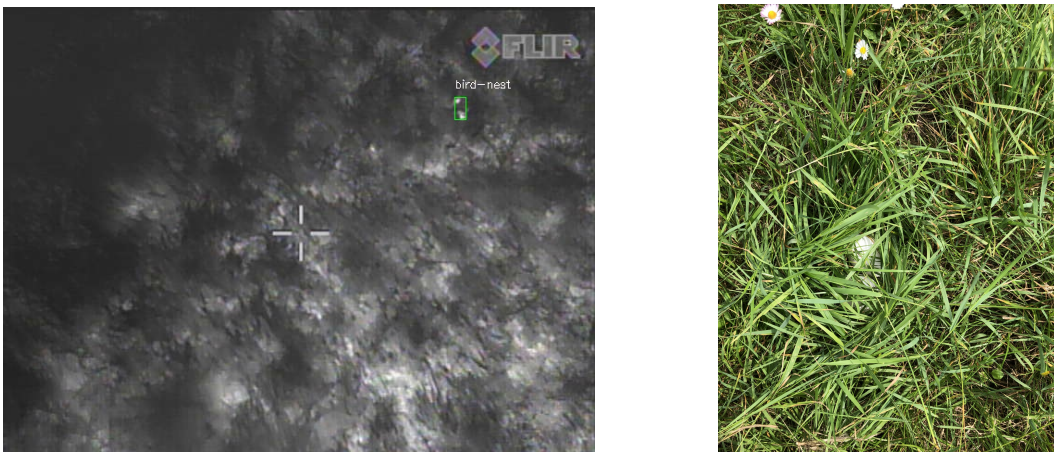


Figure 32: Nest 4, with 2 eggs and ΔT = 20.5°C.

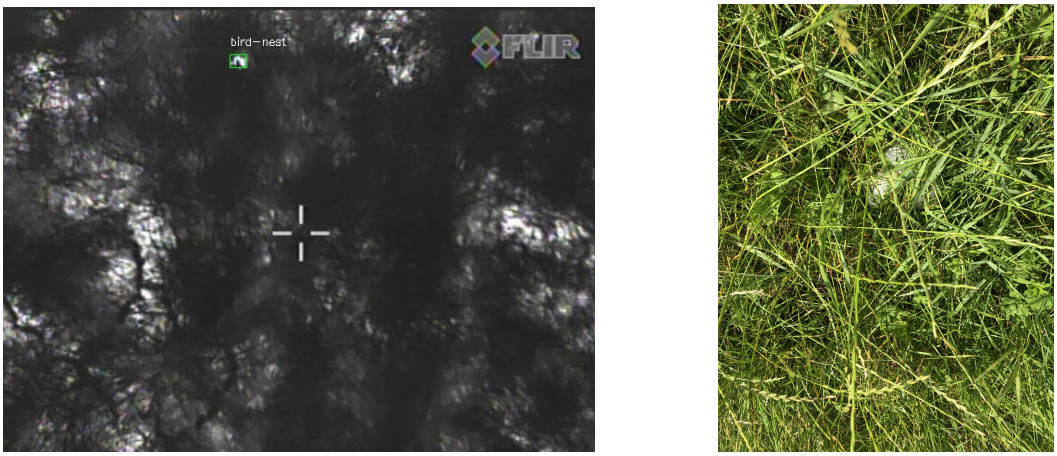


Figure 33: Nest 5, with 4 eggs and ΔT = 24.5°C.

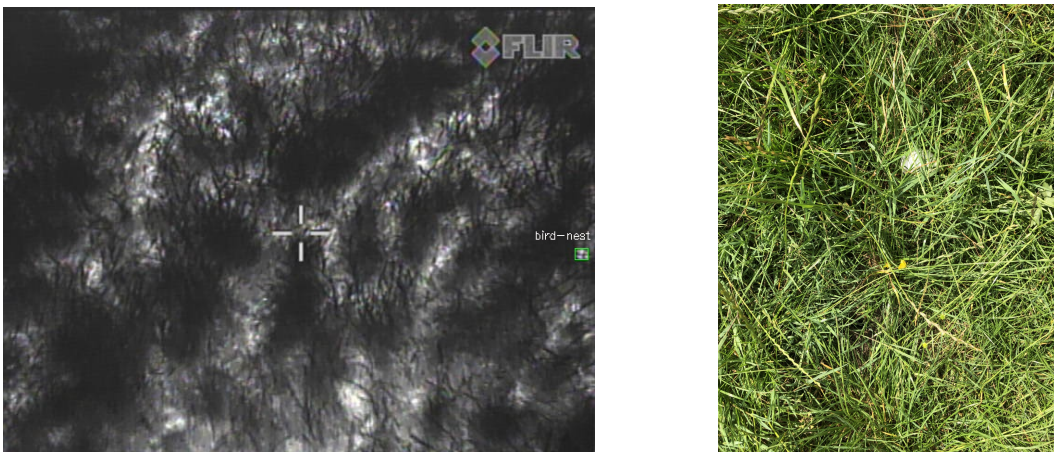


Figure 34: Nest 6, with 1 egg and ΔT = 24.5°C.

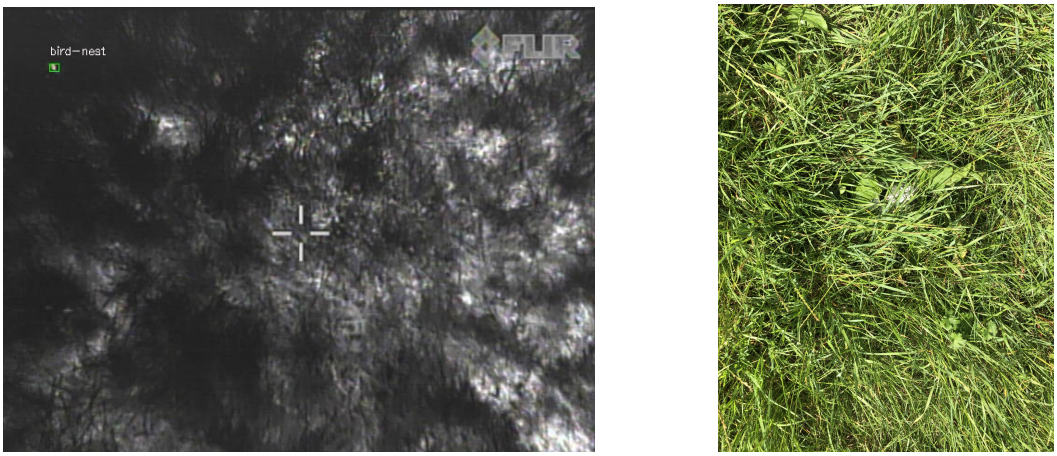


Figure 35: Nest 7, with 5 eggs and ΔT = 28.5°C.

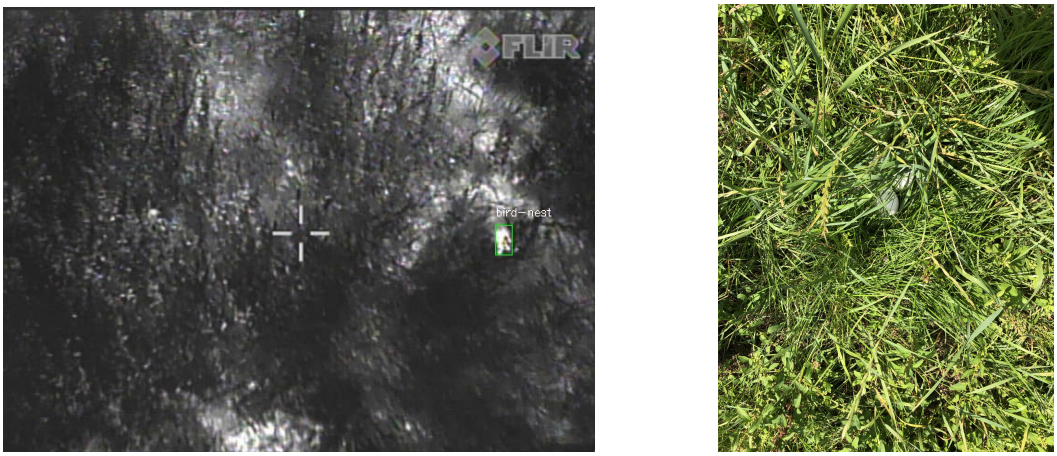


Figure 36: Nest 8, with 4 eggs and ΔT = 28.5°C.

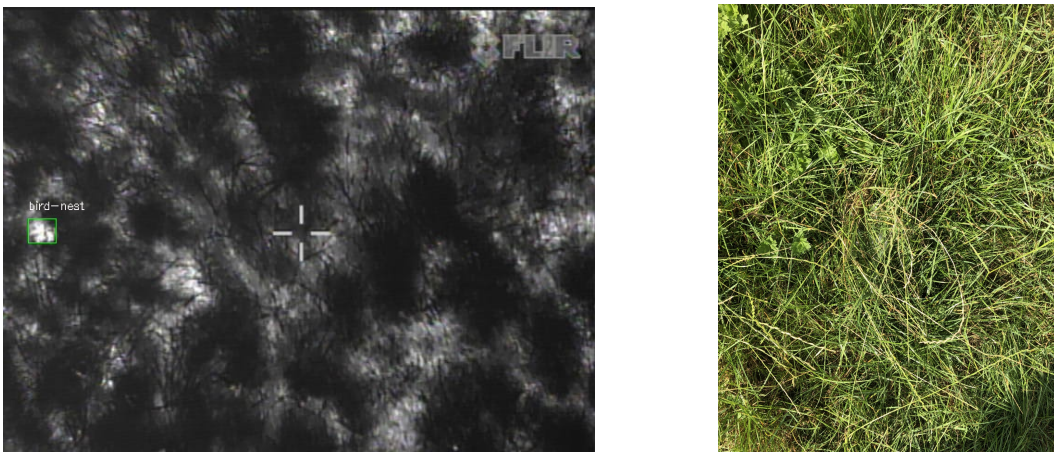


Figure 37: Nest 9, with 4 eggs and ΔT = 32.5°C.

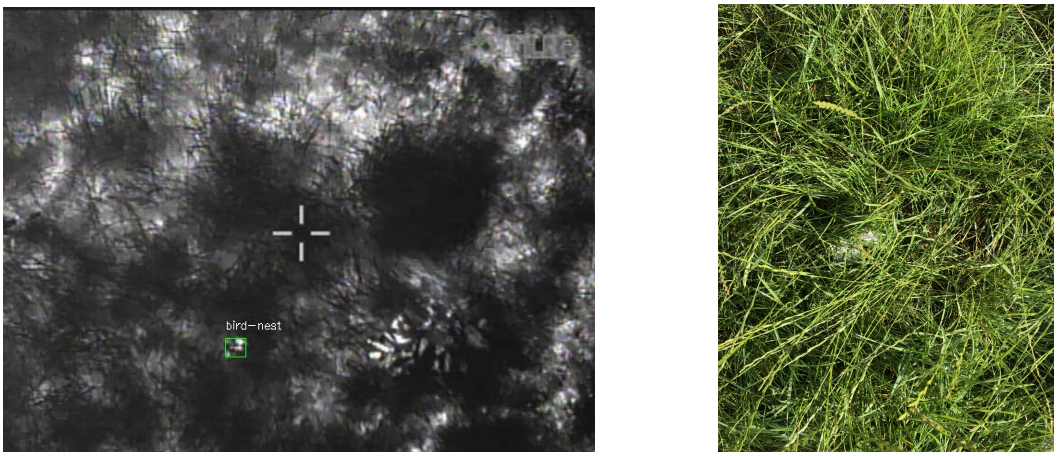


Figure 38: Nest 10, with 2 eggs and ΔT = 32.5°C.