A deep learning pipeline for time-lapse camera monitoring of floral environments and insect populations

Kim Bjerge[a,*], Hjalte M. R. Mann[b], Toke T. Høye[b,c], Henrik Karstoft[a]

[a] *Department of Electrical and Computer Engineering, Aarhus University, Finlandsgade 22, 8200 Aarhus N, Denmark*
[b] *Department of Ecoscience, Aarhus University, C.F. Møllers Allé 4-8, 8000 Aarhus C, Denmark*
[c] *Arctic Research Centre, Aarhus University, Ole Worms Allé 1, 8000 Aarhus C, Denmark*

Abstract

Insects play pivotal roles in various fields, prompting the integration of cameras and computer vision to develop novel monitoring approaches in agriculture, ecology, evolution and biodiversity monitoring. This paper presents a comprehensive methodology to monitor insects in the wild and to quantify floral cover as a key resource affecting the abundance and diversity of insects. We apply the methods across more than 10 million images recorded over two years using 48 insect camera traps placed in three main habitat types. The proposed deep-learning pipeline estimates flower cover and detects and classifies arthropod taxa from time-lapse recordings. Semantic and color segmentation are combined to estimate the percent cover of flowers of different colors. Arthropod including insect detection incorporates motion-informed enhanced images and object detection, followed by filtering stationary objects to minimize double counting of non-moving animals and erroneous plant detections. This filtering approach has been demonstrated to significantly decrease the incidence of false positives, particularly in the case of arthropods, occurring in only 3% of the captured images. The final step involves grouping arthropods into 19 taxonomic classes. The classification model achieved an 80% classification average precision. As expected during the beginning and end of the season, reduced flower cover correlates with a noticeable lack of insects.

The artificial intelligence pipeline utilizes DeepLabv3 for semantic segmentation, You-Only-Look-Once (YOLOv5) for object detection, and EfficientNet for classification. This proposed engineering method offers a cost-effective approach to monitoring flower cover and penology and diverse arthropod taxa in natural environments using time-lapse camera recordings.

Keywords: Classification, Computer Vision, Biodiversity, Deep Learning, Semantic Segmentation

*Corresponding author
Email address: kbe@ece.au.dk (Kim Bjerge)

1. Introduction

Arthropods including insects, constituting the most diverse group contributing to 50% of the animal biomass (Bar-On et al., 2018), play vital roles in terrestrial ecosystems and hold significant economic importance, serving as both agricultural pests and pollinators. Conventional insect trapping techniques, as outlined in Montgomery et al. (2021), involve the manual collection and sampling of insects. Invariably, insects are also sacrificed and the work of manual enumeration and taxonomic identification by human experts necessitates specialised knowledge and is very labour intensive. Moreover, this method poses a threat to rare insect species due to the fatal nature of the trapping process.

The advent of automated monitoring technologies, employing computer vision and deep learning, has brought about a revolution to insect surveys (van Klink et al., 2022; Lima et al., 2020; Besson et al., 2022). Computer vision methods, demonstrating substantial promise in both real-time scenarios (Bjerge et al., 2021a,b) and offline analysis of images from time-lapse (TL) cameras (Geissmann et al., 2022; Bjerge et al., 2023a), have significantly advanced insect monitoring capabilities. Automated insect camera traps, coupled with data-analyzing algorithms rooted in computer vision and deep learning, serve as invaluable tools for comprehensively monitoring insect trends and elucidating the underlying drivers (Barlow and O'Neill, 2020; Høye et al., 2021).

Automation is crucial for achieving spatiotemporal upscaling of global insect monitoring and facilitating efficient pest management (Preti et al., 2021). The increasing utilization of deep-learning methods for classification in the realm of entomology (Høye et al., 2021), underscores the need for large training datasets to ensure robust predictions. However, the diverse nature of insects, coupled with highly variable abundance, often results in unbalanced datasets with numerous classes, posing a limiting factor in the application of deep learning to entomological contexts. As these methods mature and become integral components of monitoring programs, the demand for robust techniques to detect, filter and classify species in natural environments becomes increasingly paramount.

The majority of insects, particularly pollinators, are drawn to flowering plants, making it crucial to monitor them in conjunction with arthropod abundance. Carnivores, like spiders, are also presumed to be indirectly attracted to flowers due to the expected increase in insect presence during the flowering season. The flower coverage observed from the camera perspective is expected to influence pollinators and should be adjusted to correlate with measurements of insect abundance and activity. Hence, it is crucial to include an estimation of flower cover in the recorded images as part of the image processing pipeline.

The automation of arthropod detection poses a substantial challenge due to the swift movement of insects, the ephemeral nature of their environmental interactions (such as pollination events) and the diminutive sizes of the animals (Høye et al., 2021; Xia et al., 2018). Furthermore, insects and other invertebrates may be occluded by flowers or leaves, complicating the separation of objects of interest from the surrounding

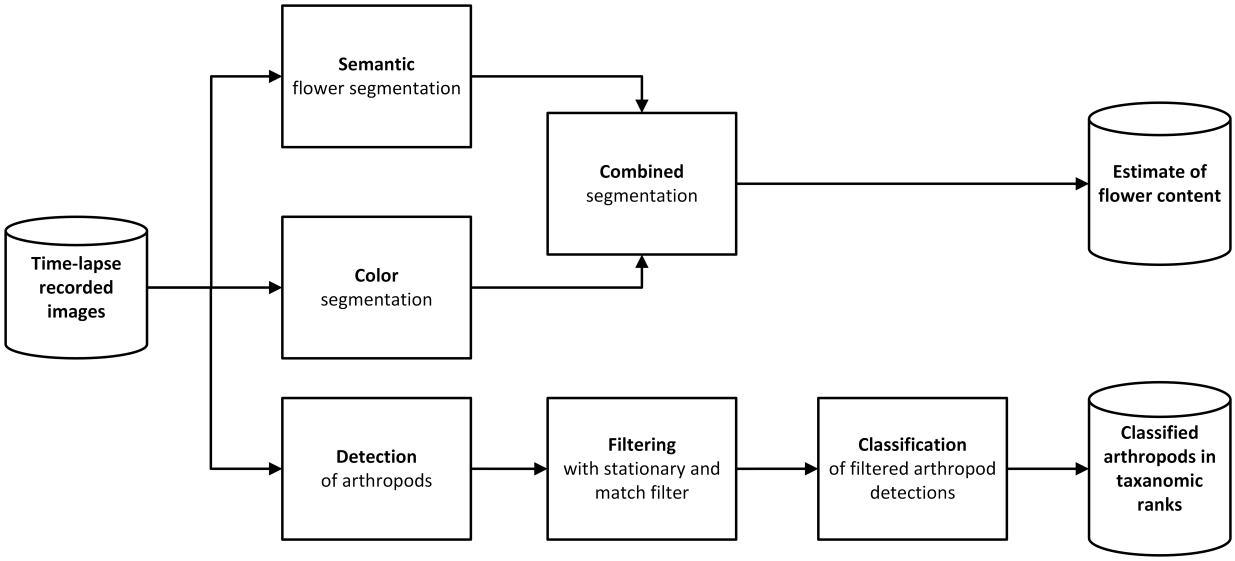


Figure 1: Shows the processing pipeline to estimate flower cover and classify taxa of arthropods in time-lapse image recordings.

natural vegetation. An object in an image that resembles an insect for the model, e.g. a small part of the vegetation, may lead to multiple false detections in consecutive images Bjerge et al. (2021b).

Images captured by fixed-position cameras may not capture distinguishing features or provide sufficient resolution for confident species identification, particularly for those with analogous morphologies. Typically, the working distance of the camera to the floral scene is in the area of 50 cm which means that arthropods constitute only a small portion of each image, often lacking the details necessary for precise species identification. Therefore, it is essential to filter false positive detections and categorise arthropods into taxonomic groups based on the image information content captured in the current study.

Convolutional neural networks (CNNs) have emerged as a powerful tool in deep learning for the localization and classification of insects and various organisms. Existing research has explored insect localization and classification through CNNs in diverse contexts (Hansen et al., 2020; Kasinathan et al., 2021; Kittichai et al., 2021; Xia et al., 2018), addressing tasks ranging from categorizing museum specimens to detecting insects in field crops like rice, soybeans and other agricultural settings. Notably, Bjerge et al. (2023a) achieved high accuracy in insect localization and classification, albeit in a study limited to a specific location with low biodiversity and a restricted number of species. Research has also delved into classifying insects into taxonomic ranks, such as order, family and genus (Ong and Hamid, 2022; Xia et al., 2018; Bjerge et al., 2023c). This approach proves particularly relevant for monitoring sites characterised by high species diversity. However, the challenge persists in collecting sufficient training samples to encompass the vast array of species within diverse ecosystems.

While previous studies have proposed insect detection and tracking methods in real-time and offline

settings (Bjerge et al., 2021a; Ratnayake et al., 2021), these works primarily focus on continuous video recordings, posing challenges for long-term, in-field camera recordings. Such continuous recording requires substantial hardware capabilities for data processing and storage than time-lapse camera trapping.

Surprisingly, none of the aforementioned studies address the combined monitoring of flowers with arthropod detection and classification. Conversely, various publications have focused on flower and plant segmentation (Nilsback and Zisserman, 2010; Aydin and Uğur, 2011; Al-Vshakarji et al., 2017; Inthiyaz et al., 2018; Lu et al., 2022). Many of these proposals utilise traditional computer vision methods, such as colored texture feature representation, color clustering, histogram-based thresholding, watershed segmentation and foreground/background segmentation using shape measurements. However, most methods concentrate on segmenting individual leaves and flowers in datasets with clearly defined, visible, single, or few flowers. This emphasis on well-defined backgrounds may not effectively address the diversity and complexity of backgrounds encountered in monitoring arthropods interacting with flowers in natural environments.

Our paper tackles the intricate task of estimating flower cover and identifying and counting arthropods in insect camera trap time-lapse sequences, which may become more abundant in the future. We introduce an extensive image processing pipeline that harnesses the power of deep learning techniques to monitor insects and flowers through time-lapse camera recordings in natural settings. This pipeline integrates innovative approaches for arthropod detection, classification and flower cover estimation, providing a resilient and effective solution for biodiversity monitoring. The efficacy of our approach is demonstrated through rigorous evaluation on a substantial dataset.

2. Method

This work aims to introduce a comprehensive pipeline shown in Figure 1 for the estimation of flower cover, detecting and classifying arthropods in image recordings obtained from a natural environment. The images were captured using time-lapse cameras specifically chosen for monitoring the vegetation including a mix of 12 species of *Sedum* plants and their associated arthropod activity. The deliberate selection of *Sedum* plants is driven by their low vegetation, which facilitates effective camera focus on both flowering plants and visiting arthropods, encompassing insects and pollinators.

2.1. Estimating flower cover in images

Numerous arthropods exhibit an attraction to flowering plants, particularly insect pollinators such as Hymenoptera and Lepidoptera, encompassing honeybees, bumblebees, hoverflies and butterflies. Consequently, we predict that the cover of flowers is correlated with the occurrence of visiting insects and other arthropods and we employed a hybrid deep learning approach by quantifying both semantic and color segmentation. Semantic segmentation, as reviewed by Guo et al. (2018) and Yu et al. (2018), poses a challenging problem

Figure 2: Example of image with flowers and a binary black and white flower mask

in computer vision. The objective is to assign a label to each pixel in an image, ensuring that pixels with the same label share specific characteristics of information within each region. Semantic segmentation, a subset of image segmentation, involves labelling every individual pixel in an image according to its semantic class or region. In our context, we classify each pixel as belonging to either background vegetation or flowers. This approach is simpler and more flexible than establishing a semantic class for each individual flower species.

In semantic segmentation, an inherent challenge lies in accurately labelling regions. To address this challenge, we leveraged the color features of plants and flowers to automate the labelling process through the creation of a color mask for flowers. Initially, we generated a flower mask encompassing yellow, white and red/pink flowers. This involved converting the image into pixel values represented in the Hue, Saturation, Value (HSV) format. Samples of the three different colored flowers were utilised to determine a threshold for the HSV values, effectively separating flowers from plants. An example of threshold color mask is shown in Figure 2. However, color segmentation is susceptible to variations in illumination and changes in the background plant's appearance, particularly during the later stages of the season when leaves may transition to brownish or reddish hues. Consequently, the dataset used for training and testing underwent manual sorting and verification, as detailed in Section 3.1.

Subsequently, employing the dataset of flower masks, we trained a DeepLabV3 (Chen et al., 2018) semantic segmentation model. The trained model produces a segmented image, distinguishing between background plants and flowers. To generate a final flower color segmented image, we combined these segmented images by applying a logical AND operation to the mask images, as illustrated in Figure 3. The estimated flower cover was quantified as the percentage of flower pixels relative to the entire image.

2.2. Arthropod detection, filtering and classification

The image pipeline shown in Figure 1 consists of three stages, commencing with the detection (A) of arthropods, followed by filtering (B) of duplicate and false detections and concluding with the classification (C) of arthropods into taxonomic ranks.

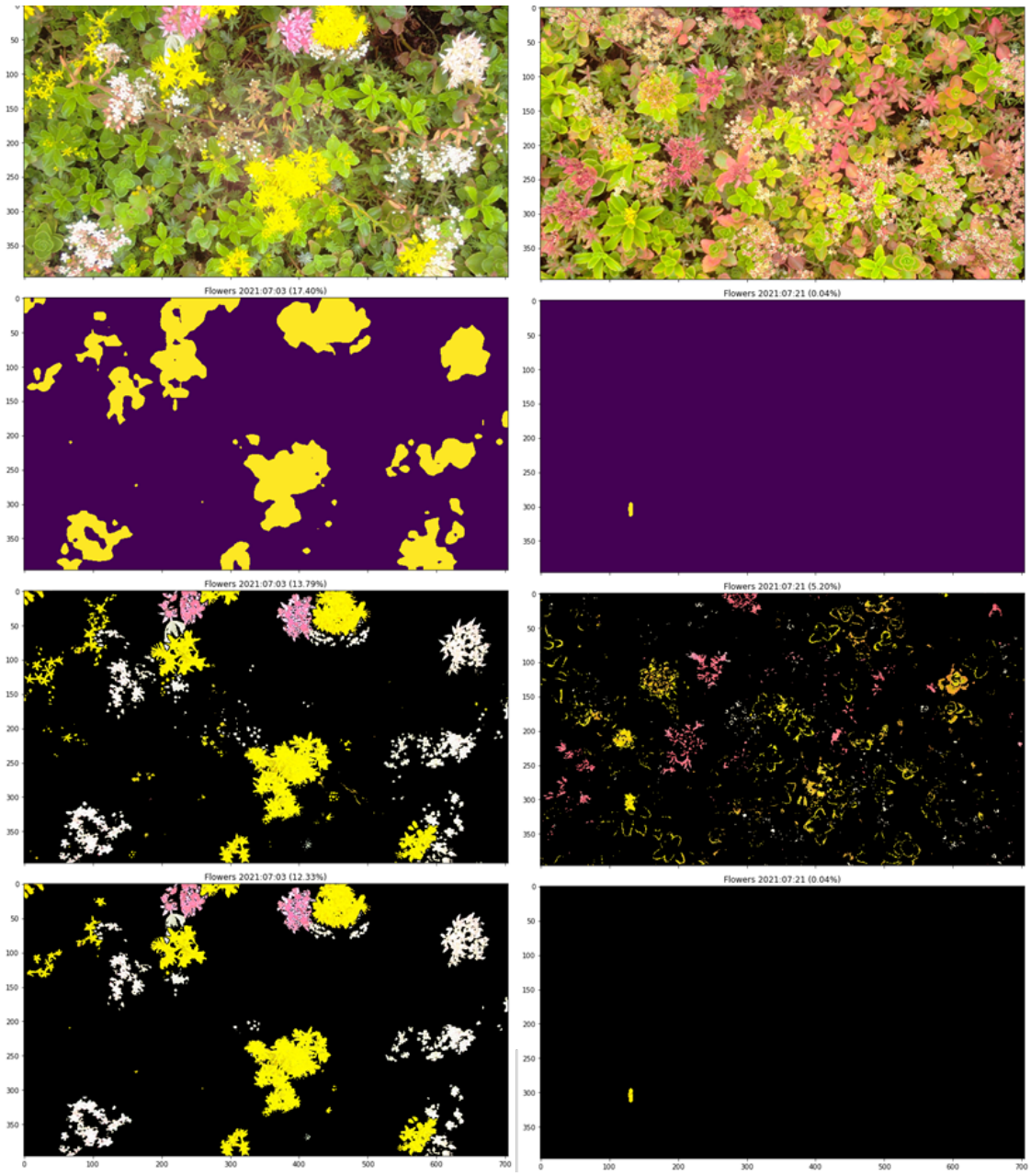


Figure 3: Shows examples of flower segmented images. Top image shows orgininal images of *Sedum* plants. Second images from top shows the semantic segmented images with DeepLab. Second images from the bottom show the color segmented images. The bottom images shows the final combined semantic and color segmented images.

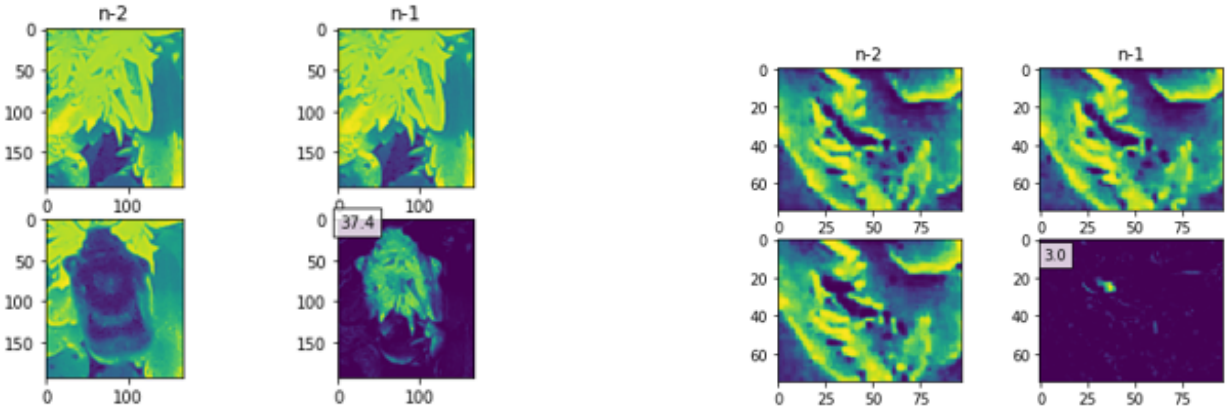
While deep learning networks have the capability to integrate both detection and classification within the same model, we opted to divide the task into two distinct models. This approach offers increased flexibility, particularly because the specific taxa of arthropods present in the recordings are unknown until the detection phase is concluded. Additionally, training a generic detection model simplifies the annotation process, as classifier training can be conducted on grouped and sorted image crops of arthropods.

A - Detection. Initially, arthropods were detected using a YOLOv5 (Redmon and Farhadi, 2018; Terven and Cordova-Esparza, 2023) model trained based on the methodology proposed in (Bjerge et al., 2023a,b). YOLOv5 provides the capability of models with different neural network architectures and image sizes. The smaller models allow faster execution. To improve performance and speed up training, the YOLOv5 is pre-trained on the COCO (Lin et al., 2015) dataset. We fine-tuned the pre-trained models and compared the capability of YOLOv5m6 to detect arthropods with a resized image size of 1280×768 pixels.

A dataset of annotated arthropods with motion-information-enhanced images was created as described in Section 3. Bjerge et al. (2023b) propose motion-informed enhancement of pre-processed images, where the YOLOv5 detector improves the average $F1$-score from 0.49 to 0.71 on a time-lapse recorded dataset in a setup similar to described in our work. Our annotated dataset with various arthropods was used to fine-tune a pre-trained YOLOv5m6 model on the COCO (Lin et al., 2015) dataset with 35.7M parameters. The model was chosen because their size allows it to be run with the required memory and achieving a high accuracy on most reasonably priced GPUs and achieved the best performance by (Bjerge et al., 2023a).

B - Filtering. Detecting and localizing small arthropods within intricate and dynamic scenes of natural vegetation poses a challenge, especially when they may be occluded by flowers or leaves. The complexity lies in distinguishing the objects of interest from the natural vegetation. Despite the feasibility of achieving the detection and classification of arthropods *in situ*, the aforementioned challenges can cause models to erroneously identify background elements as objects of interest, particularly in the case of time-lapse images. A significant portion of these images may not contain any arthropods, resulting in false positive detections of plant parts in the background. To address this issue, especially given the temporal context of time-lapse images, we introduce a filtering mechanism. This filter named stationary aims to mitigate false detections by identifying and removing instances that occur in successive recorded images. It operates by disregarding detections found in the same position (distance less than 15 pixels) within the last 15 minutes, thereby reducing the likelihood of double-counting arthropods and eliminating some of the false positives.

In addition, we incorporated a matching filter designed to eliminate detections when no discernible difference is detected within the last three successive images of the recordings. This filter effectively eliminates detections of non-moving arthropods or false positives in the background. The matching filter utilises the normalised sum of squared differences eq. (2) between the detected bounding box and the average of the



(a) True positive bumblebee detection

(b) False positive background detection

Figure 4: The left images shows three successive images of the bounding box for the detected bumblebee. With $NSSD = 37.4$ indicating a true positive object detection. The right images shows three successive images of the bounding box for a false positive detection. With $NSSD = 3.0$, the detection associated with an $NSSD$ close to zero will be eliminated.

corresponding detected bounding boxes in the preceding two recorded images eq. (1).

$$B_{avg}(u,v) = \frac{G_{n-2}(u,v) + G_{n-1}(u,v)}{2} \tag{1}$$

here, (u, v) represents the pixel coordinates within the bounding box of the detected object. G_{n-1} and G_{n-2} denote the detected bounding box in two preceding grayscale images. B_{avg} signifies the average grayscale bounding box of the two preceding images.

$$NSSD_n = \frac{\sum_{u,v}(B_{avg}(u,v) - G_n(u,v))^2}{\sqrt{\sum_{u,v} B_{avg}(u,v)^2 \cdot G_n(u,v)^2}} \tag{2}$$

G_n refers to the detected bounding box in the current grayscale image. The normalised sum of squared differences ($NSSD_n$) provides a value indicating the similarity between the background bounding box B_{avg} and the grayscale detected bounding box G_n. A value of $NSSD_n$ close to zero indicates a match of the detection with preceding images and suggests that the detection may correspond to a non-moving arthropod or a false positive background detection. In such cases, the detection is flagged for elimination. The match filtering with $NSSD$ is illustrated in Figure 4.

C - Classification. In the classification problem within deep learning, the data is categorised into different classes, such as sorting arthropod images according to taxonomic ranks. The classes are defined by a number of different labelled categories in a dataset for training and test of the deep learning model. Here, we have sorted the detected and filtered arthropods in classes based on a manual study of the actually detected objects creating a dataset with 19 classes as described in Section 3. We have evaluated ResNet50v2 (He

et al., 2016) and EfficientNetB4 (Tan and Le, 2019) and found that EfficientNetB4 gave the best performance. We achieved an $F1$-score of 0.811 with EfficientNetB4 and 0.775 with ResNet50v2 on an initial dataset with 11 classes of arthropods.

The training on the datasets was performed using data augmentation, including image scaling, horizontal and vertical flip and adding color jitter for brightness, contrast and saturation. Data augmentation mitigates overfitting by increasing the diversity of the training data. We selected a batch size of 32 for training our models since it is faster to update and results in less noise, than smaller batch sizes. The accuracy of the models on the training and validation datasets was computed after each epoch. The Adam optimizer with a fixed learning rate of $1.0 \cdot 10^{-4}$ was finally chosen based on earlier experiments (Bjerge et al., 2023c; Smith, 2018)

The EfficientNetB4 and ResNet50v2 were modified and trained with transfer learning using pre-trained weights from ImageNet (Russakovsky et al., 2015). Transfer learning involves training a model on data from a source domain $T_S = P(y|X)$ and then transferring it to a target domain $T_T = P(y|X)$ – typically with less data available. In this case, ImageNet contains 1,000 classes with 1,281,167 images for training and 50,000 for testing, while the arthropod dataset only contains 19 classes with 13,812 images. In another study, models trained with transfer learning outperformed models trained with random weights with an increase of 4% in the $F1$-Score at the species level (Bjerge et al., 2023c).

2.3. Performance metrics for testing

To evaluate model performance, the precision, recall and $F1$-score metrics were chosen for evaluation the final detection, filtering and classification pipeline. These metrics are based on true positive (TP), false positive (FP) and false negative (FN) detections and classification. Recall and precision were used in conjunction to obtain a complete picture of the model's ability to find all arthropods and classify them correctly. To balance precision and recall, we used the $F1$-score. The $F1$-score is calculated as the harmonic mean of precision and recall, given in eq. (5) and it provides a balance between the two metrics. It prioritises the importance of detecting and classifying the classes correctly.

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

$$recall = \frac{TP}{FN + TP} \tag{4}$$

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{5}$$

Dataset	Images	Arthropods	Background
Training	6,901	4,770	2654
Testing	766	525	299

Table 1: Dataset used to train the detector of arthropods. It shows the number of images with number of annotated arthropods and background images without arthropods.

Figure 5: Example images of plant parts and arthropods for classes listed in Table 2

3. Experiment and datasets

Images were captured using Wingscapes (Wingscapes, 2024) cameras, which recorded time-lapse images at 60-second intervals between June and September 2020 and 2021, spanning from 6 AM to 10 PM. The images encompassed *Sedum* plants, positioned at three distinct locations — urban, agriculture and grassland — each featuring four camera sites. All sites were located in Eastern Jutland, Denmark. At each camera site, two plots measuring $120 \times 80cm$ were cultivated with *Sedum* plants and monitored by two cameras. Hence, four closely located cameras were deployed at each location and site, resulting in a total of $3 \times 4 \times 4 = 48$ cameras.

The primary objective was to showcase the feasibility of automated camera monitoring for live insects and flowers *in situ* on a comprehensive temporal and spatial scale. In total 5,932,597 images were recorded from the 48 cameras in the summer 2020 and total 4,524,328 images were recorded from the 48 cameras in the summer 2021. For model training and validation, only images from 2021 have been used.

Label ID	Class name	Order	Family	Genus/Species	Samples
0	Ladybirds	Coleoptera	Coccinellidae		370
1	Beetles	Coleoptera			366
2	Plants	Crassulaceae			619
3	Bumblebees	Hymenoptera	Apidae	Bombus	616
4	Hoverflies	Diptera	Syrphidae		1,517
5	Butterflies	Lepidoptera			903
6	Spiders	Aranaea			1,842
7	Ants	Hymenoptera	Formidicidae		608
8	Flies	Diptera			663
9	True bugs	Hemiptera			638
10	Isopods	Isopoda			851
11	Unspecified				1,544
12	Hymenoptera	Hymenoptera			372
13	Grasshoppers	Orthoptera			165
14	*R. fulva*	Coleoptera	Cerambycidae	*Rhagonycha fulva*	820
15	Satyrines	Lepidoptera	Nymphalidae	Satyrinae	686
16	Small Tortoiseshell	Lepidoptera	Nymphalidae	*Aglais urticae*	620
17	Dragonflies	Odonata			249
18	Honeybees	Hymenoptera	Apidae	*Apis melifera*	363

Table 2: The 19 classes of arthropods and plant parts used to train the classifier with in total 13.812 sample images. The classes of arthropods are grouped by different taxonomical ranks. The class "Unspecified" is a mixture of different species with insufficient samples to create separate classes. The taxonomic rank of order, family, genus and species are listed for each class given by class name and a label ID. The splitting of the dataset is 20% for testing and 80% for training.

3.1. Datasets with flower masks

Daily, at 12:00, an image was chosen from each of the 48 cameras in 2021, wherein a color mask, as detailed in Section 2.1, was applied to generate a binary mask for flowers in yellow, white and pink hues, as depicted in Figure 2 and 3. Images were specifically selected if they exhibited more than 1% pixel area coverage of flowers, yielding 1,635 images featuring flower masks. Subsequently, images with false positive flower masks were eliminated, resulting in a refined dataset comprising 507 images with accurately segmented flowers. To augment the dataset, 25% of the images with false positive flower masks were reintroduced, with corrections made by a corrected black flower mask. The resultant dataset, featuring correct flower masks, encompasses a total of 768 images.

3.2. Datasets with arthropods

Despite the large full image resolution (4224×2375 or 6080×3420 pixels), each insect observation covered only a small portion of the full image, as the recording was conducted at a distance to monitor a designated patch of vegetation. Bjerge et al. (2023a) introduced a YOLOv5 model trained for insect detection on *Sedum* plants; however, this model only identified 26,987 detections in the recorded images equal to only 0.7% of the 3.9 million images. We utilised this model to generate a dataset comprising annotated bounding boxes for arthropods. The model yielded positive detections and false negative, encompassing instances of plant

parts and animals, such as snails and birds. Subsequently, these detections of plant parts and animals were excluded and removed from the final dataset. A total of 6,901 images were meticulously selected, manually corrected and annotated with bounding boxes outlining arthropod locations. To enhance the detection of small objects, we employed motion-informed image enhancements, following the methodology proposed by Bjerge et al. (2023b). The resulting dataset, comprising annotated arthropods, is detailed in Table 1. Examples of background images and annotated arthropods are shown in Appendix A.

A YOLOv5 model without motion-informed image enhancement was trained to detect arthropods and found 351,505 detections in image recordings from 2021. These detections were cropped and manually selected and sorted into 19 classes of taxonomic ranks as shown in Table 2. The images are semi-randomly chosen from the model detections trying to have as many images of individuals in each class. An example of sample images for each class is shown in Figure 5. The class of unspecified arthropods (Label ID 11) contains images of species with only few images insufficient to create separate classes. A class (Label ID 2) with false positive plant parts was also included in the group of classes. The resulting dataset is unbalanced ranging from 165 to 1842 images in each class. To facilitate model training, squared crops of arthropod images were resized to 224×224 pixels.

4. Results and discussion

4.1. Estimating flowers in images

Utilizing the dataset outlined in Section 3.1, we trained a DeepLabV3 (Chen et al., 2018) semantic segmentation model. The optimal model was identified after 24 epochs, demonstrating the lowest test loss (0.00675), an $F1$-score of 0.505 and an Area Under Curve (AUC) of 0.888. Subsequently, we combined the semantic-segmented images with flower color-segmented images through a logical AND operation on the mask images. The efficacy of this final model was assessed by randomly selecting 500 images from the 2020 and 2021 datasets. Among these, 15 images from 2020 and 151 images from 2021 exhibited false positive instances with flower segmentation.

Table 3 presents the percentage of flower pixel area for false positives (FP) and true positives (TP). Training the model on images from 2021 resulted in a notably high precision of 98.1%. However, when evaluating the model on images recorded in 2020, the precision drops to 80.7%. This discrepancy can be attributed to the fact that more images in 2020 contained plants with a different appearance than those in 2021. A few examples can be found in Appendix A, where plants are overexposed or exhibit a more reddish hue, especially late in the season. This change is most certainly due to drought for much of the 2020 season that has led to changes in the color profile of the plants. To enhance performance, the inclusion of images from both 2020 and 2021 in the training set for the semantic segmentation model could have been considered.

Year	FP pixel area (%)	TP pixel area (%)	Precision
2020	223.0	932.6	0.807
2021	8.04	413.9	0.981

Table 3: Shows the false positive (FP) and true positive (TP) area of flowers in percentage of pixels out of 500 random selected images in 2020 and 2021. The precision of flower detection is calculated based on the FP and TP pixel areas.

Model	Precision	Recall	$F1$-score	mAP
YOLOv5	0.933	0.901	0.917	0.943
MIE	0.890	0.908	0.908	0.935

Table 4: Shows the precision, recall, $F1$-score and mean Average Precision (mAP) of the YOLOv5 models trained without and with motion-informed enhancement (MIE).

4.2. Detecting, filtering and classifying arthropods

A YOLOv5 model was trained on the detection dataset presented in Table 1, with and without motion-informed enhancement (MIE). The results in Table 4 demonstrate a high $F1$-score for both models. The precision is 93.3% for the YOLOv5 model without MIE, compared to 89.0% for the model with MIE. It is worth noting that a similar study by Bjerge et al. (2023b) reported comparable results, indicating that the performance increase with MIE is particularly evident on time-lapse test images not included in the training and validation datasets. The model trained with MIE showed in general a higher number of detections than the model trained with color images, indicating more insects were found and detected on a time-lapse test dataset created from different sites and weeks not included in the training and validation dataset.

While the precision of arthropod detection is quite high at 90%, this still results in a considerable number of false positive predictions when dealing with millions of time-lapse images. The magnitude of this issue is illustrated in Table 5, where a substantial number of false positive plant detections are observed. These are subsequently filtered using a sequence of stationary and matching filters applied after the YOLOv5 MIE arthropod detector.

The stationary detections encompass both false plant detections and non-moving arthropods, aiming to minimize the double counting of the same individual. However, if an animal has moved within the camera's field of view, it may be counted multiple times. The matching filter further mitigates false positive plant detections, particularly in 2020, where 90% of the 181,928 detections pertain to plant parts. Notably, detections filtered by the matching filter may include arthropods that blend with the background, resulting in their removal.

The purpose of the next processing pipeline step were to sort the filtered detections in classes of arthropod taxa. A EfficientNetB4 model was trained on the arthropod dataset, which consists of 19 different classes, as detailed in Section 3.2. The dataset was partitioned into 80% for training and 20% for validation. The model achieved an overall average $F1$-score of 0.81, as depicted in Table A.1. Analysing the confusion

Filter	Year	Detection	Plants (%)
Stationary	2020	179,027	79
Matching	2020	181,928	90
Stationary	2021	245,404	55
Matching	2021	81,321	69

Table 5: Shows the filtered detections by the stationary and matching filters of the image recordings in 2020 and 2021. The percentages of plants are evaluated on 200 randomly selected detections from both 2020 and 2021. Here the remaining percentages of detections were identified as arthropods.

matrix in Figure 6, it becomes evident that certain class names, namely plants (2), butterflies (5), true bugs (9), unspecified (11) and Hymenoptera (12), exhibit a subpar accuracy, falling below 70%. The discrepancy in accuracy for plants (2) is attributed to YOLOv5 arthropod detections that could not be effectively filtered by the stationary and matching filter. The diminished accuracy in identifying butterflies (5) stems from numerous images featuring butterflies with folded wings, posing a challenge for classification. The unspecified (11) class comprises a mix of arthropods that could not definitively be grouped into any other class. Consequently, many arthropods are erroneously classified into this unspecified category, such as 49% of Hymenoptera (12). True bugs (9) exhibit low accuracy due to misclassifications as various other classes, including beetles, plants, spiders, grasshoppers and unspecified categories.

After the detection and filtering of the 2020 and 2021 recordings, predictions were generated using the trained EfficientNetB4 model. The count of predictions for each of the 19 classes is presented in Table 6. A validation test was conducted on 200 randomly selected predictions from each class. Any class predictions that could not be confidently verified were manually labeled as false positives. For the unspecified class, only plant parts were treated as false positive predictions, resulting in a higher precision than that achieved on the validation dataset in Table A.1. Given that the EfficientNetB4 model was trained on data recorded in 2021, a higher precision of 80% was attained for data from 2021 compared to the 60% precision in 2020. Classes with the best performance, exhibiting a precision above 60%, include ladybirds (0), plants (2), hoverflies (4), spiders (6), ants (7), flies (8), Hymenoptera (12), grasshoppers (13) and R. fulva (14). Notably, butterflies (5), bumblebees (3), true bugs (9), satyrines (15) and dragonflies (17) show precisions below 30% in 2020. In the analysis of 200 random test samples, it is particularly plant parts that are misclassified in the mentioned classes.

In the years 2020 and 2021, we identified approximately 300.000 arthropods among the more than 10 million recorded images. The inherent challenge with time-lapse recordings in our scenario lies in the relatively low occurrence of arthropods compared to the vast number of images, leading to a significant number of false positive plant detections compared to the number of positive arthropod detections. With a precision of 89% measured on the validation dataset in table 4 there should in theory be 11% false positive detections which will result in more than 1 million false plant detections in the recordings from 2020 and

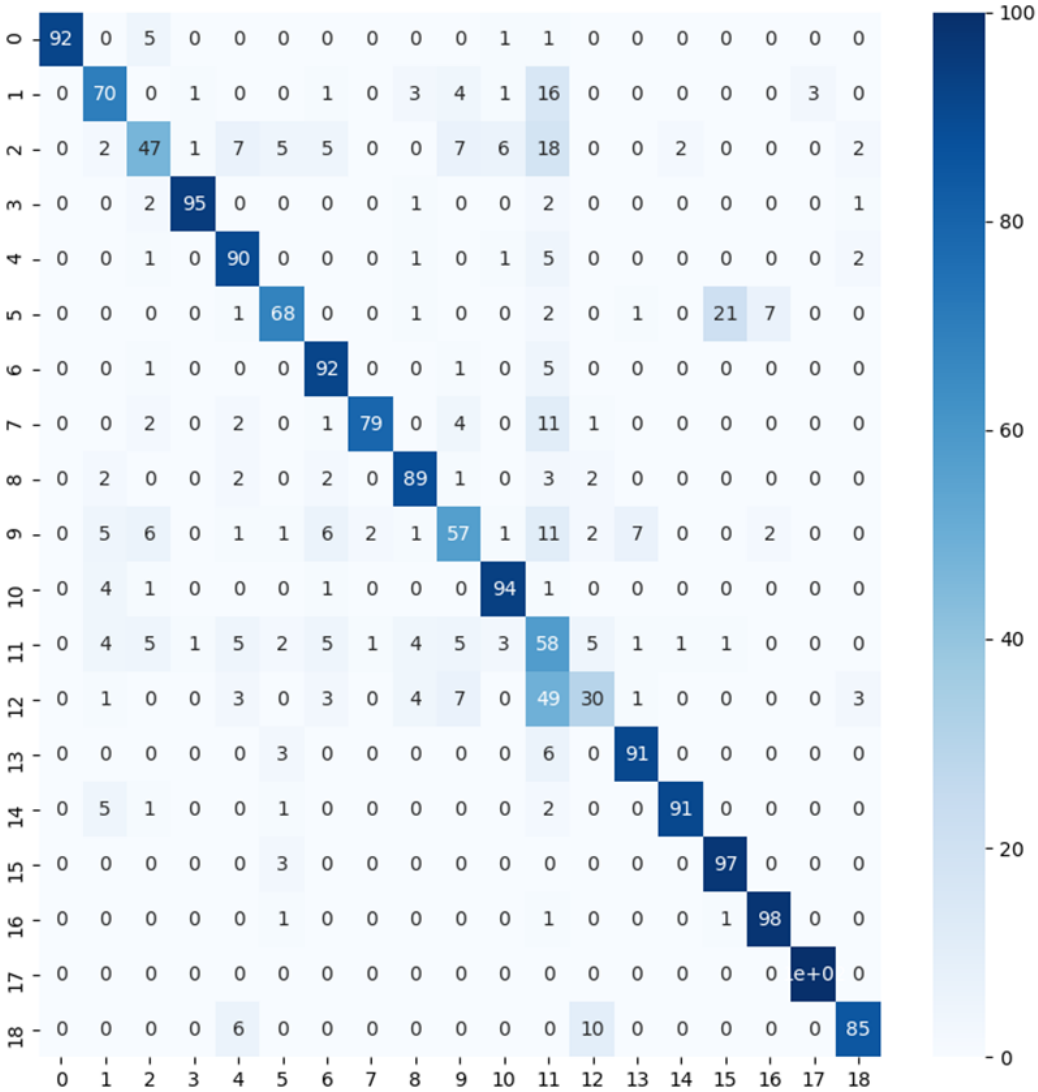


Figure 6: Confusion matrix on the validation dataset of the trained EfficientNetB4 model. The X-axis is the ID of the predicted labels. The Y-axis is the ID of the true labels. The numbers in the table are the accuracy percentage related to true labels. A high diagonal value indicates a high accuracy.

Label ID	Class name	Predictions 2020	Predictions 2021	Precision 2020	Precision 2021
0	Ladybirds*	556	580	0.64	0.88
1	Beetles	5,478	4,813	0.58	0.70
2	Plants*	34,537	19,027	0.90	0.92
3	Bumblebees	1,216	1,634	0.35	0.80
4	Hoverflies*	15,437	6,620	0.90	0.82
5	Butterflies	5,749	3,218	0.07	0.56
6	Spiders*	16,048	23,862	0.72	0.80
7	Ants*	2,867	6,412	0.92	0.94
8	Flies*	8,708	13,162	0.98	0.98
9	True bugs	6,207	8,946	0.29	0.46
10	Isopods*	8,708	4,380	0.80	0.80
11	Unspecified*	68,351	38,664	0.78	0.90
12	Hymenoptera*	411	2,556	0.78	0.88
13	Grasshoppers*	2,320	1,462	0.76	0.70
14	*R. fulva**	7,949	3,708	0.66	0.74
15	Satyrines	2,303	3,173	0.25	0.80
16	Small Tortoiseshell	158	691	0.50	0.98
17	Dragonflies	385	285	0.07	0.81
18	Honeybees	964	509	0.48	0.70
	Total/Average	187,866	143,702	0.60	0.80

Table 6: Shows the number of classified predictions of 19 categories of arthropods in 2020 and 2021. Precision 2020/2021 is a test performed on 200 random predictions selected from each class after detection and filtering on the entire image collection in 2020 and 2021. The total number of predictions and the average precision in 2020 and 2021 are listed at the bottom of the table. The best performing classes with a precision above 60% in 2020/2021 are marked with an '*'.

2021. If we combine the estimates of plant detections in table 4 and table 6 an approximately 550.000 false plant detections. Therefore the filtering approach we propose is crucial for mitigating this issue by reducing the number of false positive plant detections before classification into different taxa of arthropods.

4.3. Evaluation of flower estimation and arthropod counting

Figure 7 illustrates the temporal progression of the processing pipeline results for a designated camera positioned in a grassland environment throughout the year 2021. At the onset and conclusion of the season, characterised by diminished flower coverage, there is a noticeable absence of insects. The prevalence of spiders mirrors the abundance of other arthropods, consistent with their role as predators. Notably, the utilization of YOLOv5 with MIE leads to the detection of a greater number of small arthropods, including spiders and ants. The same tendency was observed by Bjerge et al. (2023b) where the models trained MIE achieved a high F1-score on a time-lapse test dataset.

Figure 8 depicts the temporal dynamics observed in two cameras situated within agricultural and grass-

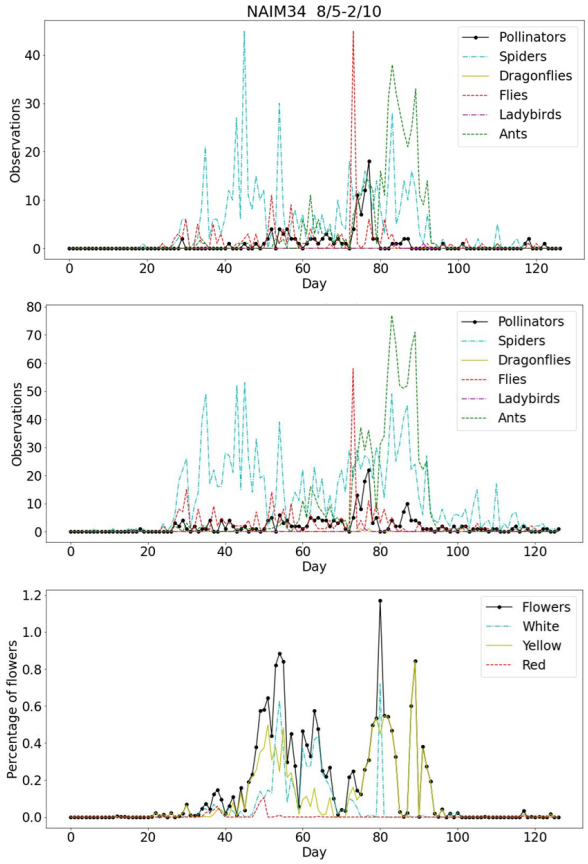


Figure 7: Shows the analytical outcomes derived from a particular camera (NAIM34) situated in a grassland setting during the year 2021. The upper image illustrates the outcome of the arthropod detection, filtering and classification pipeline using YOLOv5 without MIE for a specific set of classes. Pollinators encompass bumblebees, hoverflies, honeybees and butterflies (Satyrines and S. Tortoiseshell). The middle image displays the pipeline results with YOLOv5 and MIE. The lower image exhibits the flower cover, highlighting white, yellow and red flowers.

land settings throughout the year 2021. The figures distinctly show an increased abundance of pollinators during the flowering phase. This consistent trend is corroborated by two additional camera locations in the urban and a habitats, as depicted in fig. 8b. Notably, non-pollinators like spiders, flies, isopods and ants exhibit activity even outside the flowering period.

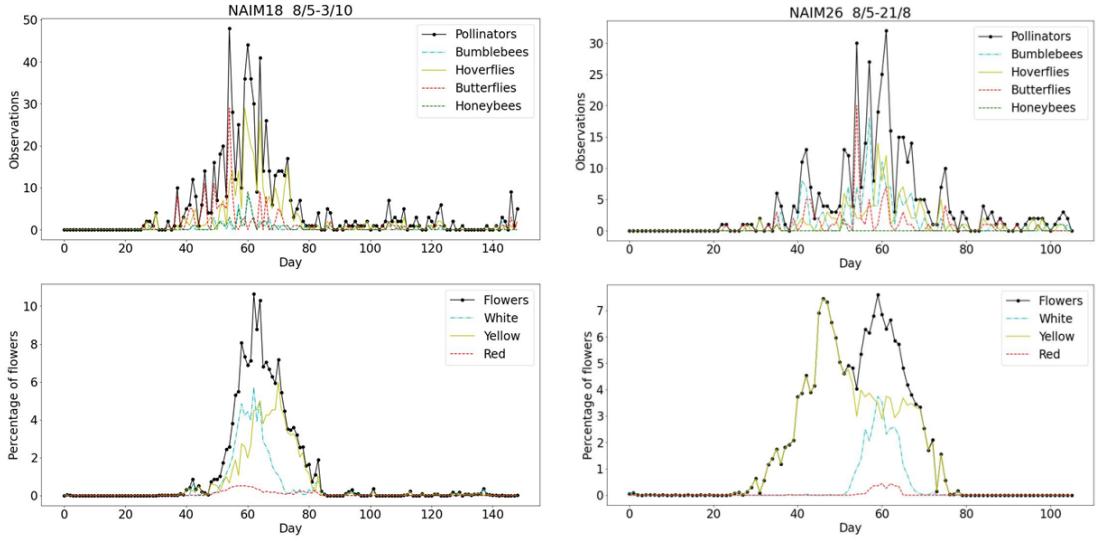
Our approach holds immediate applicability to image-based insect monitoring in natural settings, serving purposes in agriculture and biodiversity conservation (Ratnayake et al., 2021; Bjerge et al., 2021a; Preti et al., 2021). We have showcased its capability in estimating flower cover and establishing a discernible correlation between pollinator abundance and recorded flower cover. Our method can generate relevant information regarding flower phenology by analyzing color variations over time and species abundance as studied by Mann et al. (2022). This allows for the exploration of insect preferences for different flower species.

5. Conclusion and future work

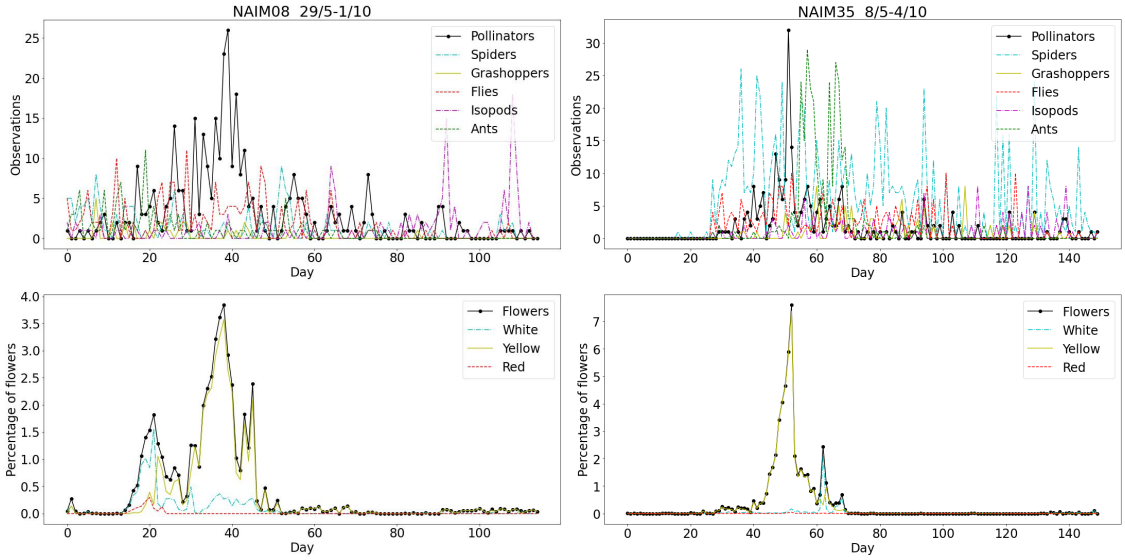
In conclusion, our proposed methodology presents a comprehensive pipeline for estimating flower cover and detecting arthropods in natural flowering environments. The utilization of semantic and color segmentation provides a straightforward approach to estimating flower cover in images. Given the low arthropod content within complex vegetation, our solution addresses the impact of false positives on detection precision. The incorporation of stationary and matching filters effectively eliminates non-moving false plant detections and reduces the occurrence of double-counting individual arthropods. Over half of the detections involve false positives or non-moving animals, underscoring the importance of our filtering mechanisms. The pipeline culminates with a trained CNN classifier that categorizes arthropods into selected taxa based on our experiment's actual recordings.

Our comprehensive evaluation, conducted across 48 cameras in three distinct habitats with over 10 million images, resulted in datasets used to train a YOLOv5 object detector and an EfficientNetB4 classifier. The pipeline exhibited an average precision of 80%, with certain categories, such as hoverflies, ants, and flies, achieving precision levels exceeding 90%. Furthermore, our study confirmed a strong correlation between insect pollinators and flowering plants.

Future efforts should focus on collecting additional time-lapse images with annotated arthropods to train a generic YOLO detector using motion-informed enhanced images for broader applicability across monitoring sites. Expanding the dataset to include more taxa and increasing samples within each class can enhance classifier performance. Moreover, replacing the unspecified arthropods class with an anomaly detector, achieved by thresholding CNN classifier output scores, could further refine the classification process. An ecological analysis of the results is left to future studies outside the scope of this paper.



(a) The upper images illustrates the outcome of the pollinators encompass bumblebees, hoverflies, honeybees and butterflies (Satyrines and S. Tortoiseshell). The lower images exhibits the flower coverage, highlighting white, yellow and red flowers.



(b) The upper images illustrates the outcome of selected arthropods with high precision and pollinators. The lower images exhibits the flower coverage, highlighting white, yellow and red flowers.

Figure 8: Shows the analytical outcomes derived from four cameras situated in an urban (NAIM08), agriculture (NAIM18) and grassland (NAIM26, NAIM35) setting during the year 2021.

CRediT authorship contribution statement

Kim Bjerge: Conceptualization, Data curation, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. Hjalte M.R. Mann: Data curation, Writing – review & editing. Toke T. Hoye: Conceptualization, Data curation, Validation, Funding acquisition, Project administration, Writing – review & editing. Henrik Karstoft: Supervision, Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgments

The work was supported by the European Union's Horizon Europe Research and Innovation programme, under Grant Agreement No. 101060639 (MAMBO). We would like to give thanks to the companies Nature Impact and Byggros that have provide *Sedum* plants for all 48 camera setups.

Data availability

Data will be made available on request.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT in order to improve formulations in writing. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Figure A.1: Full size images of background images with killer slugs (*Arion lusitanicus*) but without annotations of arthropods.

Appendix A. Supplementary material

In Figure A.1, full-sized (6080 × 3420 pixels) images are shown with examples of background images without annotations. Examples of annotations from part of full-sized images are shown in Figure A.2. Examples of false positive flower segmentations are shown in Figure A.3 from images recorded in 2020. The model precision, recall and $F1$-score are shown in Table A.1 evaluated on 20% of the arthropods dataset.

Figure A.2: Part of full-size images with annotated arthropods (Honeybee, hoverflies and butterfly).

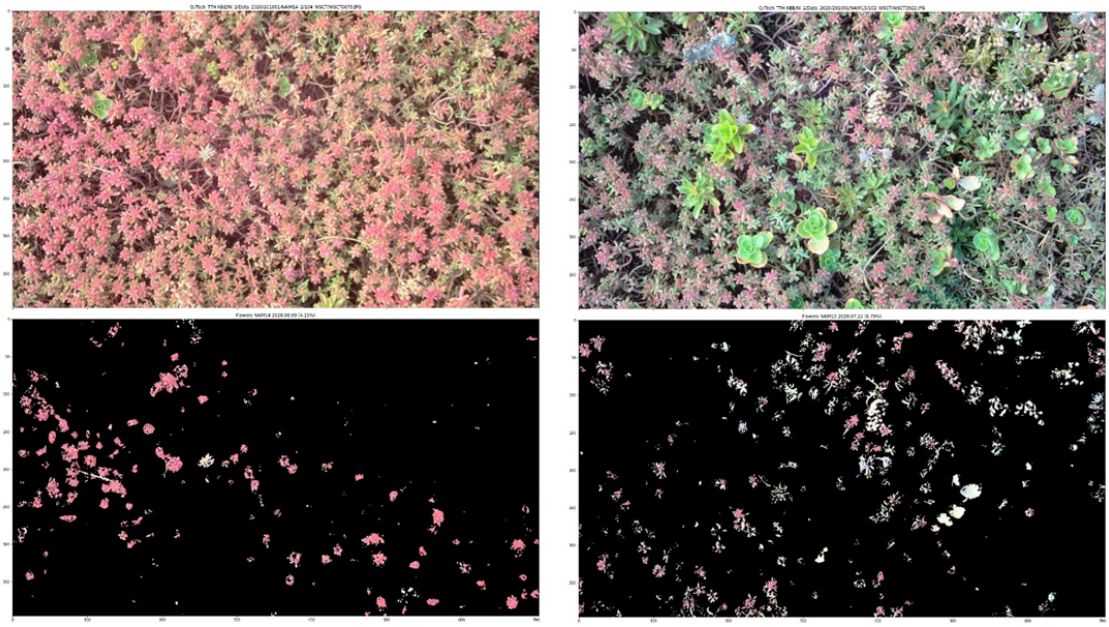


Figure A.3: Example of false positive combined flower segmentations in images recorded in 2020.

Label ID	Class name	Precision validation	Recall validation	$F1$-score validation	Validation dataset
0	Ladybirds	0.97	0.92	0.94	74
1	Beetles	0.57	0.70	0.63	73
2	Plants	0.60	0.47	0.53	123
3	Bumblebees	0.95	0.95	0.95	123
4	Hoverflies	0.88	0.90	0.89	303
5	Butterflies	0.87	0.68	0.76	180
6	Spiders	0.90	0.92	0.91	368
7	Ants	0.95	0.79	0.86	121
8	Flies	0.84	0.89	0.86	132
9	True bugs	0.64	0.57	0.60	127
10	Isopods	0.88	0.94	0.91	170
11	Unspecified	0.55	0.58	0.57	308
12	Hymenoptera	0.43	0.30	0.35	74
13	Grasshoppers	0.67	0.91	0.77	33
14	*R. fulva*	0.96	0.91	0.94	164
15	Satyrines	0.76	0.97	0.85	137
16	Small Tortoiseshell	0.88	0.98	0.93	124
17	Dragonflies	0.96	1.00	0.98	49
18	Honeybees	0.82	0.85	0.84	72
	Average	0.80	0.81	0.80	

Table A.1: The EfficientNetB4 model validated on 20% of the dataset with the 19 classes of arthropods and plant parts in Table 2. The precision, recall and $F1$-score are show evaluated on the validation dataset.

References

Al-Vshakarji NM, Kassim YM, Palaniappan K. Unsupervised learning method for plant and leaf segmentation. In: Proceedings - Applied Imagery Pattern Recognition Workshop. volume 2017-October; 2017. doi:10.1109/AIPR.2017.8457935.

Aydin D, Uğur A. Extraction of flower regions in color images using ant colony optimization. In: Procedia Computer Science. volume 3; 2011. doi:10.1016/j.procs.2010.12.088.

Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. Proceedings of the National Academy of Sciences of the United States of America 2018;115(25). doi:10.1073/pnas.1711842115.

Barlow SE, O'Neill MA. Technological advances in field studies of pollinator ecology and the future of e-ecology. Current Opinion in Insect Science 2020;38. doi:10.1016/j.cois.2020.01.008.

Besson M, Alison J, Bjerge K, Gorochowski TE, Høye TT, Jucker T, Mann HM, Clements CF. Towards the fully automated monitoring of ecological communities. Ecology Letters 2022;25(12). doi:10.1111/ele.14123.

Bjerge K, Alison J, Dyrmann M, Frigaard CE, Mann HMR, Høye TT. Accurate detection and identification of insects from camera trap images with deep learning. PLOS Sustainability and Transformation 2023a;2(3):1–18. doi:10.1371/journal.pstr.0000051.

Bjerge K, Frigaard CE, Karstoft H. Object detection of small insects in time-lapse camera recordings. Sensors 2023b;23(16). URL: https://www.mdpi.com/1424-8220/23/16/7242. doi:10.3390/s23167242.

Bjerge K, Geissmann Q, Alison J, Mann HM, Høye TT, Dyrmann M, Karstoft H. Hierarchical classification of insects with multitask learning and anomaly detection. Ecological Informatics 2023c;77:102278. URL: https://www.sciencedirect.com/science/article/pii/S1574954123003072. doi:https://doi.org/10.1016/j.ecoinf.2023.102278.

Bjerge K, Mann HM, Høye TT. Real-time insect tracking and monitoring with computer vision and deep learning. Remote Sensing in Ecology and Conservation 2021a;doi:10.1002/rse2.245.

Bjerge K, Nielsen JB, Sepstrup MV, Helsing-Nielsen F, Høye TT. An automated light trap to monitor moths (Lepidoptera) using computer vision-based tracking and deep learning. Sensors (Switzerland) 2021b;doi:10.3390/s21020343.

Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence 2018;40(4). doi:10.1109/TPAMI.2017.2699184.

Geissmann Q, Abram PK, Wu D, Haney CH, Carrillo J. Sticky Pi is a high-frequency smart trap that enables the study of insect circadian activity under natural conditions. PLoS Biology 2022;20(7):1–26. doi:10.1371/journal.pbio.3001689.

Guo Y, Liu Y, Georgiou T, Lew MS. A review of semantic segmentation using deep neural networks. International Journal of Multimedia Information Retrieval 2018;7(2). doi:10.1007/s13735-017-0141-z.

Hansen OL, Svenning JC, Olsen K, Dupont S, Garner BH, Iosifidis A, Price BW, Høye TT. Species-level image classification with convolutional neural network enables insect identification from habitus images. Ecology and Evolution 2020;10(2). doi:10.1002/ece3.5921.

He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2016. doi:10.1109/CVPR.2016.90.

Høye TT, Ärje J, Bjerge K, Hansen OLP, Iosifidis A, Leese F, Mann HMR, Meissner K, Melvad C, Raitoharju J. Deep learning and computer vision will transform entomology. Proceedings of the National Academy of Sciences 2021;118(2). doi:10.1073/pnas.2002545117.

Inthiyaz S, Madhav BT, Kishore PV. Flower image segmentation with PCA fused colored covariance and gabor texture features based level sets. Ain Shams Engineering Journal 2018;9(4). doi:10.1016/j.asej.2017.12.007.

Kasinathan T, Singaraju D, Uyyala SR. Insect classification and detection in field crops using modern machine learning techniques. Information Processing in Agriculture 2021;8(3). doi:10.1016/j.inpa.2020.09.006.

Kittichai V, Pengsakul T, Chumchuen K, Samung Y, Sriwichai P, Phatthamolrat N, Tongloy T, Jaksukam K, Chuwongin S, Boonsang S. Deep learning approaches for challenging species and gender identification of mosquito vectors. Scientific Reports 2021;11(1). doi:10.1038/s41598-021-84219-4.

van Klink R, August T, Bas Y, Bodesheim P, Bonn A, Fossøy F, Høye TT, Jongejans E, Menz MH, Miraldo A, Roslin T, Roy HE, Ruczyński I, Schigel D, Schäffler L, Sheard JK, Svenningsen C, Tschan GF, Wäldchen J, Zizka VM, Åström J, Bowler DE. Emerging technologies revolutionise insect ecology and monitoring. Trends in Ecology and Evolution 2022;37(10):872–85. doi:10.1016/j.tree.2022.06.001.

Lima MCF, Leandro MEDdA, Valero C, Coronel LCP, Bazzo COG. Automatic detection and monitoring of insect pests - A review. Agriculture (Switzerland) 2020;10(5). doi:10.3390/agriculture10050161.

Lin TY, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Zitnick CL, Dollár P. Microsoft COCO: Common Objects in Context. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015;arXiv:1405.0312v3.

Lu Y, Young S, Wang H, Wijewardane N. Robust plant segmentation of color images based on image contrast optimization. Computers and Electronics in Agriculture 2022;193. doi:10.1016/j.compag.2022.106711.

Mann HM, Iosifidis A, Jepsen JU, Welker JM, Loonen MJ, Høye TT. Automatic flower detection and phenology monitoring using time-lapse cameras and deep learning. Remote Sensing in Ecology and Conservation 2022;8(6). doi:10.1002/rse2.275.

Montgomery GA, Belitz MW, Guralnick RP, Tingley MW. Standards and Best Practices for Monitoring and Benchmarking Insects. Frontiers in Ecology and Evolution 2021;8. doi:10.3389/fevo.2020.579193.

Nilsback ME, Zisserman A. Delving deeper into the whorl of flower segmentation. Image and Vision Computing 2010;28(6). doi:10.1016/j.imavis.2009.10.001.

Ong SQ, Hamid SA. Next generation insect taxonomic classification by comparing different deep learning algorithms. PloS one 2022;17(12):e0279094. URL: http://dx.doi.org/10.1371/journal.pone.0279094. doi:10.1371/journal.pone.0279094.

Preti M, Verheggen F, Angeli S. Insect pest monitoring with camera-equipped traps: strengths and limitations. Journal of Pest Science 2021;94(2):203–17. URL: https://doi.org/10.1007/s10340-020-01309-4. doi:10.1007/s10340-020-01309-4.

Ratnayake MN, Dyer AG, Dorin A. Tracking individual honeybees among wildflower clusters with computer vision-facilitated pollinator monitoring. PLOS ONE 2021;16(2):1–20. URL: https://doi.org/10.1371/journal.pone.0239504. doi:10.1371/journal.pone.0239504.

Redmon J, Farhadi A. YOLOv3: An incremental improvement. 2018. arXiv:1804.02767.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision 2015;115(3). doi:10.1007/s11263-015-0816-y.

Smith LN. A disciplined approach to neural network hyper-parameters: Part 1 – Learning rate, batch size, momentum, and weight decay. arXiv 2018;:1–21URL: http://arxiv.org/abs/1803.09820. arXiv:1803.09820.

Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. In: 36th International Conference on Machine Learning, ICML 2019. volume 97; 2019. p. 6105–14.

Terven J, Cordova-Esparza D. A comprehensive review of YOLO: From YOLOv1 and beyond. 2023. arXiv:2304.00501.

Wingscapes . Wingscapes TimelapseCam Pro. 2024. URL: https://www.wingscapes.com/timelapsecam-pro.

Xia D, Chen P, Wang B, Zhang J, Xie C. Insect detection and classification based on an improved convolutional neural network. Sensors (Switzerland) 2018;doi:10.3390/s18124169.

Yu H, Yang Z, Tan L, Wang Y, Sun W, Sun M, Tang Y. Methods and datasets on semantic segmentation: A review. Neurocomputing 2018;304. doi:10.1016/j.neucom.2018.03.037.