Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

PREPRINT RESEARCH PAPER

Towards edge processing of images from insect camera traps

Kim Bjerge[1] | **Henrik Karstoft**[1] | **Toke T. Høye**[2,3]

[1]Department of Electrical and Computer Engineering, Aarhus University, Aarhus C, Denmark

[2]Department of Ecoscience, Aarhus University, Aarhus C, Denmark

[3]Arctic Research Centre, Aarhus University, Aarhus C, Denmark

Correspondence
Corresponding author Kim Bjerge, Finlandsgade 22, 8200 Aarhus C, Denmark
Email: kbe@ece.au.dk

Abstract

Insects represent nearly half of all known organisms, with nocturnal insects being particularly challenging to monitor. Computer vision tools for automated monitoring have the potential to revolutionize insect study and conservation. The advancement of light traps with camera-based monitoring systems for insects necessitates effective and flexible pipelines for analysing recorded images. In this paper, we present a flexible and fast processing pipeline designed to analyse these recordings by detecting, tracking and classifying nocturnal insects at the taxonomic ranks of order, suborder as well as the resolution of individual moth species. The pipeline consists of four adaptable steps. The first step detect insects in the camera trap images. An order and suborder classifier with anomaly detection is proposed to filter dark, blurry or partly visible insects that will be uncertain to classify correctly. A simple track-by-detection algorithm is proposed to track the classified insects by incorporating feature embeddings, distance and area cost. We evaluated the computational speed and power performance of different edge computing devices (Raspberry Pi's and NVIDIA Jetson Nano) and compared various time-lapse strategies with tracking. The minimum difference was found for 2-minute time-lapse intervals compared to tracking with 0.5 frames per second, however, for insects with fewer than one detection per night, the Pearson correlation decreases. Shifting from tracking to time-lapse monitoring would reduce the amount of recorded images and be able to perform edge processing of images in real-time on a camera trap with Raspberry Pi. The Jetson Nano is the most energy-efficient solution, capable of real-time tracking at nearly 0.5 fps. Our processing pipeline was applied to more than 3.4 million images recorded at 0.5 frames per second from 12 light camera traps during one full season located in diverse habitats, including bogs, heaths and forests.

KEYWORDS
Biodiversity monitoring, Camera trap, Computer Vision, Edge processing, Insect classification, Tracking

1 | INTRODUCTION

Insects, comprising over a million described species, represent nearly half of all known organisms. Also, insects constitute the most diverse group of animals make up 50% of the animal total biomass (Bar-On et al. 2018), play vital roles in terrestrial ecosystems and hold significant economic importance, serving as both agricultural pests, natural enemies and pollinators. Changes in the abundance of insects could have cascading effects through the food web, suggesting that nocturnal insects are a very relevant group of insects to monitor more effectively in the context of global change of climate and habitats (Wagner et al. 2021). Conventional insect trapping techniques, as outlined by 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 is ethically problematic and poses a threat to rare insect species due to the fatal nature of the trapping process.

Data on insect populations is notably sparse due to limited resources, the vast number of species, and the high level of expertise required to study them (Didham et al. 2020). 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) in both real-time scenarios (Bjerge et al. 2021a, Sittinger et al. 2024, Ratnayake et al. 2021) and offline analysis of images from time-lapse (TL) cameras (Geissmann et al. 2022, Bjerge et al. 2023a). Automated insect camera traps, coupled with

 © 2024 Copyright Holder Name

FIGURE 1 Camera trap with UV light to attract and monitor nocturnal insect.

data-analysing algorithms rooted in computer vision and deep learning, could therefore serve as invaluable tools for monitoring insect trends and elucidating the underlying drivers (Barlow and O'Neill 2020, Høye et al. 2021). Animal species recognition from camera traps is a well-established problem within the computer vision community (Oliver et al. 2023), with common challenges including poor lighting, occlusion, camouflage and blur (Beery et al. 2018). However, working with insects presents unique challenges that are not encountered with traditional camera trap systems designed for large animals. For instance, while traditional camera trap images might occasionally capture a target species, nearly every image from an insect camera trap contains insects. This is especially true when using motion detection during nights of high activity, where hundreds of nocturnal insects can be visible in a single image.

Nocturnal insects are difficult to monitor, however, camera-based light traps (Bjerge et al. 2021b, Korsch, D., Bodesheim, P., Denzler, J. 2021) and advancement of standardized hardware and framework for image-base monitoring of nocturnal insects (Roy et al. 2024) paves the way for increased temporal coverage and resolution in insect monitoring. Automated monitoring of moths has been evaluated by comparing traditional lethal methods with light-based camera traps (Möglich et al. 2023). This first proof of concept demonstrated that automated moth traps depict large magnitude events such as phenological patterns just as well as conventional, lethal traps.

Camera trapping methods based on time-lapse recordings can generate millions of images, especially when using sampling intervals of seconds or a few minutes. However, as these tools become more widely applied, they are likely to generate large amounts (terabytes to petabytes) of image data per year and storing all the data may not be feasible or even sensible. It is possible that a reduced framerate will yield comparable results, but rare taxa are less likely to be detected as frame rate is reduced. The alternative is to process images during recording and only store processed data and possibly a subset of the raw images.

In this work, we propose a flexible and fast processing pipeline to analyse image recordings from insect camera traps by detecting, tracking and classifying nocturnal insects at the taxonomic ranks of order, suborder and moth species. We demonstrate the efficacy of the proposed pipeline by evaluating its speed performance on three different edge computing devices, including Raspberry Pi 4, Raspberry Pi 5 and NVIDIA Jetson Nano. Our pipeline supports various time-lapse strategies and real-time tracking. To our knowledge, no other studies on insect monitoring have been published that cover an experiment with 12 traps recording images over more than two thousand nights across two years.

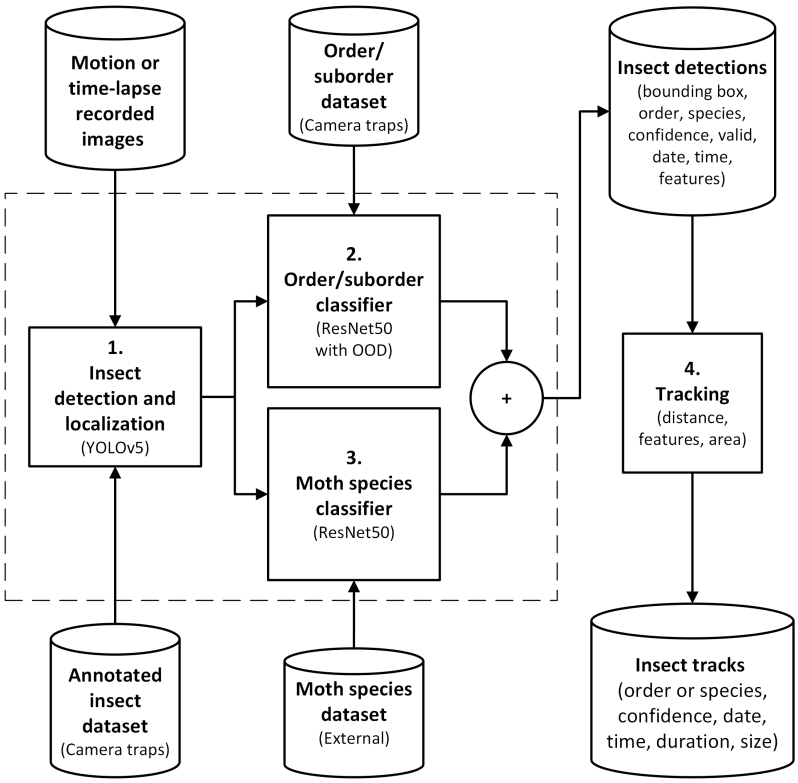


FIGURE 2 Processing pipeline to localize, classify and track insects from motion or time-lapse recorded camera trap images. In steps 1. and 2. the detected insects are classified in order or suborder parallel with moth species classification. Classification results are concatenated "+" providing insect information for the final tracking step performed for motion with a high sampling rate of 0.5 fps.

2 | MATERIALS AND METHOD

2.1 | Data collection

Automated light traps with cameras were constructed with standard components consisting of a Raspberry Pi 4, Brio Camera (Logitech 2021), power controller, UV light and light ring as proposed by Bjerge et al. (2021b). A solid-state drive (500 GB) was connected to the Raspberry Pi to store the captured images. The mechanical design was improved and the background light table was replaced with a plastic plate covered with a white fabric shown in Figure 1.

12 camera traps were placed across three different locations in Denmark during the 2022 and 2023 summer seasons. These locations encompassed diverse habitats, including bogs, heaths and forests. Four of the traps were powered by solar panels, charge regulators and batteries (12V) all others were supplied through mains power (220V). The camera traps were activated in the period from 11 p.m. to 3 a.m. each night. This limitation is primarily due to the power of the battery and solar panel being able to monitor the whole season of insect activity from April to the end of October.

A motion program (Motion 2021) running on the Raspberry Pi 4 was installed to capture a sequence of images whenever a movement was detected in the camera view. The maximum frame rate was limited to 0.5 fps. On warm summer nights with a high level of insect activity, more than 6,000 images were captured per night. In 2022 and 2023 more than 5 million images with the pixel size of 3840 $\times$ 2160 (11 pixels/*mm*) were recorded. As supplement to the motion recorded images, a time-lapse approach was used to saved an image every 10 minutes independent of insect activity.

2.2 | Processing pipeline

For insect monitoring, Multiple Object Tracking (MOT) would be relevant, especially for fast video recording. MOT uses Computer Vision to estimate trajectories for objects of interest presented in a sequence of images, especially videos with high frame rates. Many MOT methods have been developed such as two-step methods named tracking-by-detection (TBD) (Wojke et al. 2018, Shao et al. 2018) or one-step methods also named joint detection and tracking (JDT) (Kieritz et al. 2018, Zhang et al. 2021). The one-step method reduces inference time by a single model that reuses the backbone network for both object detection and re-ID feature extraction. Most MOT methods require annotated tracking datasets, which can be demanding to create.

We aimed to create a flexible pipeline that can be used for both processing with and without tracking depending on the chosen time-lapse sampling interval. Therefore, the TBD approach is chosen since it is flexible and can be realized without any annotated tracking dataset. The proposed processing pipeline is shown in Figure 2. The pipeline is designed to prioritize flexibility above efficiency.

The first step in our pipeline concerns insect detection and localization. Here, the model is trained to detect insects of interest by ignoring dirt and small or blurry insects. An annotated insect dataset is created based on collected images from camera traps. This step could be replaced with a more generic detector and localizer trained on images from various backgrounds with insects.

The second and third step concerns the classification of insects in taxonomical ranks of two separate models that can be executed in parallel. The order/suborder classifier is trained on the camera trap data. Here, we have sorted the insects in order and suborders based on the content of the recorded images. We have incorporated an anomaly detector into the classifier to filter insects which have class scores that are outside the distribution of the created order/suborder dataset. These outliers could be partly visible or blurry insects, or they could be representatives of unseen classes of insects.

The third step implements the moth species classifier trained on external data from GBIF with all moth species known to be present in the region where the trap is located. Here, we have used the moth species classifier published by Rolnick, David; Bundsen, Michael; Jain, Aditya; Cunha (2023).

The output from the insect detector, localizer, order/suborder classifier and moth species classifier is a list of insect detections with additional information about the trap, image, bounding box coordinates, confidence, validity, date, time and embedding features.

The fourth tracking step is based on track-by-detection by using the bounding boxed and embedding features to create a final list of insect tracks with information about predicted insect order/suborder, species, confidence, size, date, arrival time and duration seen by the camera.

The source code for the pipeline is available at Github[†]. Each step in the pipeline is described in the following with focus on the new contributions for anomaly detection and the simple flexible tracking of insects.

2.3 | Insect detection and localization

Image object detection methods based on deep learning rely solely on spatial image information to extract features and detect regions of objects in the image. You-only-look-once (Redmon et al. 2016) (YOLO) is a one-stage object detector and one of the fastest object detectors important for processing millions of images or deployed on edge computers. In our work, YOLOv5 (Glenn Jocher 2020) with a backbone of CSPDarknet53 was evaluated.

In the paper Bjerge et al. (2023a) different YOLOv5 architectures are evaluated finding that YOLOv5m6 with 35.7 million parameters is the optimal model to detect and classify small insect species. To improve performance and speed up training, the YOLOv5m6 is pre-trained on the COCO dataset (Lin et al. 2015). In this work, we have trained YOLOv5m6 and YOLOv5s6 on the dataset described in section 2.7.

2.4 | Order and suborder classifier with anomaly detection

Images were cropped and resized to 128×128 pixels, matching the dimensions used for the moth species classifier. 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. We selected a batch size of 256 for training our models since it is faster to update

[†] https://github.com/kimbjerge/MCC24-trap

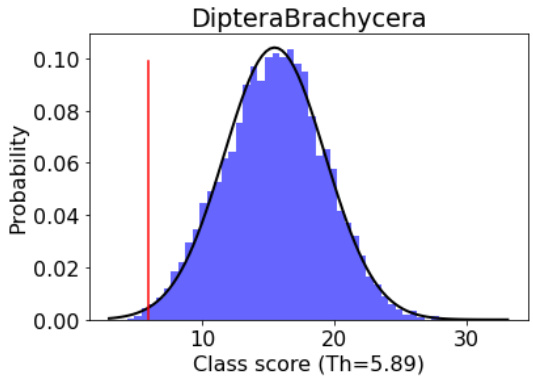


FIGURE 3 Probability density function for the output scores for Diptera Brachycera and the chosen threshold for uncertain anomalies.

and results in less noise, than smaller batch sizes. The Adam optimizer with a fixed learning rate of $1.0 \cdot 10^{-4}$ was chosen based on earlier published experiments (Bjerge et al. 2023b). We have trained ResNet50v2 (He et al. 2016) to classify insects in order or suborder defined by the 16 classes as described in section 2.7. The ResNet50v2 as fine-tuned using pre-trained weights from ImageNet (Russakovsky et al. 2015).

2.4.1 | Out-of-distribution detection

The methodology of out-of-distribution detection (Bulusu et al. 2020) and threshold-based anomaly tagging is employed to identify instances of "anomalies" such as uncertain classifications. In our application, these instances may manifest as debris, obscured or partially visible insects, or those exhibiting blurriness, characteristics that are not represented in the insect order training dataset.

Often, softmax is the last layer in a classification neural network, where the maximum value determines the predicted class. Here, we instead will analyse the output distribution without the softmax layer to determine the anomalies and predict the classes. The distribution of the output x for each predicted j^{th} class follows a normal distribution $x_j \sim \mathcal{N}(\mu, \sigma^2)$.

An example of the output distribution is shown in Figure 3 which is generated on the training dataset of samples (Diptera Brachycera) for correct classified inputs. If the output value x_j is below a threshold of $th = \mu - 2.5\sigma$, we will tag the input as an anomaly. That means when new unknown inputs are presented for the trained network, and the output lies below the threshold, it will be classified as an "Uncertain" prediction.

The output scores x_j are finally mapped to a probability $F(x_j)$ by estimating the cumulative distribution function as the integral of the probability density function given by.

$$F(x_j \| \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_j} e^{-\frac{(x_j-\mu)^2}{2\sigma^2}} dx_j \tag{1}$$

2.5 | Moth species classifier

We are proposing to use the moth species classifier from the AMI data companion code base (Jain et al. 2024, Rolnick, David; Bundsen, Michael; Jain, Aditya; Cunha 2023), trained on GBIF data encompassing 2,530 moth species native to the United Kingdom (UK) and Denmark. The model is tested on a dataset with West-European moths recorded with AMI traps in Denmark and UK with a F1-score of 0.784 on a dataset with 4,066 images of 244 different moth species (Jain et al. 2024). All models within the AMI data companion code base are trained using the ResNet50 architecture. Looking ahead, we anticipate the availability of additional models covering diverse regions worldwide in the future, further enhancing the applicability and scope of moth species classification.

2.6 | Tracking

Our tracking algorithm was extended by comparing feature embeddings from the order/suborder classifier for the tracking algorithm proposed by Bjerge et al. (2021a). The Hungarian Algorithm is the chosen method for finding the optimal assignment for a given cost matrix. In this application, the cost matrix should represent how likely it was that an insect in the previous image had moved to a given position in the current image. The cost function was defined as a weighted cost of embeddings similarity, distance and area of matching bounding boxes in the previous and current image. The Euclidean distance D between center position (x, y) in the two images was calculated as follows.

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{2}$$

This distance was normalized according to the diagonal of the image I:

$$D_{max} = \sqrt{(I_{height})^2 + (I_{width})^2} \tag{3}$$

The area cost was defined as the cost between the area A of bounding boxes:

$$A_{cost} = \frac{A_{min}}{A_{max}} \tag{4}$$

The similarity of feature embeddings was defined as the cosine similarity between embeddings $\boldsymbol{E}$ of classified insects:

$$E_{cost} = \frac{\boldsymbol{E_1} \cdot \boldsymbol{E_2}}{\|\boldsymbol{E_1}\| \|\boldsymbol{E_2}\|} \tag{5}$$

A final cost function in Equation (6) was defined with a weighted cost of distance W_d, embeddings W_e and weighted cost of area W_a.

$$Cost = \frac{D}{D_{max}} W_d + (1 - E_{cost}) W_e + (1 - A_{cost}) W_a \tag{6}$$

A cost threshold was established to determine whether suggestive insect detections should be associated. Subsequently, a track was formulated, stipulating a minimum of two detections per track. For each track, information such as the start date and time, duration, number of detections and average size was recorded. Of particular importance was the recording of the predominant insect order or moth species, along with the accuracy percentage of its classification. Ultimately, a valid track was delineated as one wherein over 50% of the detections corresponded to the predominant classification and comprised at least three detections or persisted for a duration exceeding four seconds.

2.7 | Datasets

The time-lapse recorded images were annotated to generate two distinct datasets aimed at facilitating insect localization followed by classification within taxonomic orders and suborders. This two-step strategy was adopted to manage the challenge of curating well-balanced dataset encompassing annotated insect taxa across the entirety of the images. This approach not only addresses the complexity of dataset creation but also enables flexibility in the processing pipeline.

For localization, 777 images were selected and annotated skipping very small and blurry insects from being annotated with bounding boxes. The full dataset is shown in Table 1.

Through the analysis of detected and cropped insect images, we successfully identified 15 prevalent taxa groups across various orders and suborders, including spiders, beetles, craneflies, lacewings, harvestmen, caddisflies, as well as both macro and micro moths (Lepidoptera). Example images of these taxa are provided in Figure 4. The creation of our dataset followed an iterative process aimed at achieving optimal balance based on insights gleaned from the recorded images. The resulting dataset, presented

TABLE 1 Dataset for detection and localization

Dataset	Images	Labels
Training	700	5,335
Validation	77	482

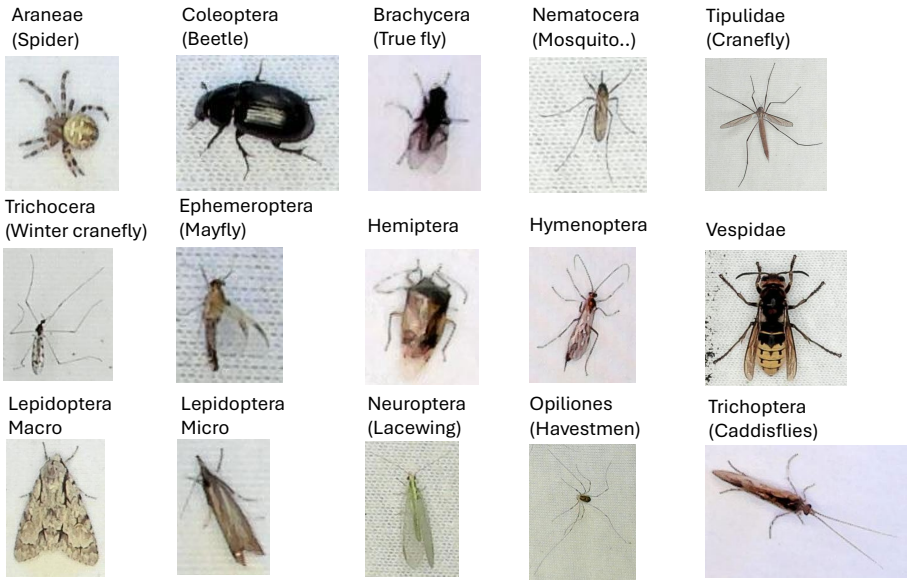


FIGURE 4 Example of samples from the 15 taxa of insect used to classify nocturnal insects in order or suborder.

in Table 2, is organized according to the hierarchical rank of orders and suborders. Furthermore, recognizing the inadvertent presence of vegetation such as leaves and flowers within the images, an additional class dedicated to vegetation was incorporated into the dataset to ensure comprehensive coverage of the observed ecological elements. The dataset was split in 80% for training and 20% for validation of the classifier.

3 | RESULTS

3.1 | Insect detection and localization

The precision, recall and F1-score of the fist stage of the pipeline is listed in Table 3. It is observed that the large YOLOv5m6 model has the highest F1-score, however, the number of parameters for YOLOv5m6 is 35.7M compared to 12.6M for YOLOv5s6. Experiments with newer models, such as YOLOv8m, did not improve the F1-score, indicating that the annotated dataset needs

TABLE 2 Dataset of image samples for order/suborder classification collected during 2022 and 2023 from time-lapse images with 10-minute intervals.

Taxa in Order [Suborder]	Total samples	Validation (20%)
Araneae	2,037	408
Coleoptera	2,384	477
Diptera Brachycera	12,303	2,461
Diptera Nematocera	26,890	5,378
Diptera Tipulidae	1,216	244
Diptera Trichocera	1,664	333
Ephemeroptera	10,147	2,030
Hemiptera	4,897	980
Hymenoptera Other	1,661	333
Hymenoptera Vespidae	529	106
Lepidoptera Macros	18,675	3,735
Lepidoptera Micros	28,141	5,629
Neuroptera	1,106	222
Opiliones	615	123
Trichoptera	11,978	2,396
Vegetation	892	179
Total	125,135	25,035

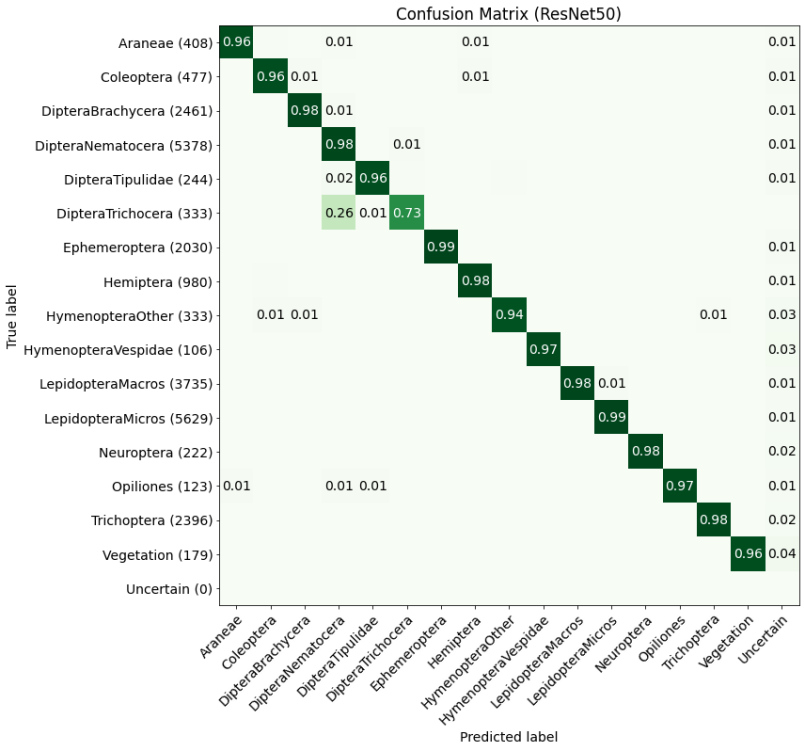


FIGURE 5 Confusion matrix for the order/suborder classifier with out-of-distribution detection of samples marked as the uncertain class.

to be refined and expanded. The discrepancies between model predictions and annotated insects are particularly influenced by the subjective criteria for excluding small insects in the annotated images.

3.2 | Order and suborder classifier with anomaly detection

The precision, recall and F1-score of the second stage of the pipeline are summarized in Table 4. The table shows the results for the ResNet50v2 model validated without the anomaly threshold detector and where the uncertain samples are removed. There is a small increase in all metrics indicating that by removing samples predicted as "Uncertain" the classifier improves by accepting that 1.0% of the true positive samples are ignored.

TABLE 3 Validation results for insect detection and localization on dataset with 482 labels.

Metric	YOLOv5m6	YOLOv5s6
Precision	0.923	0.938
Recall	0.919	0.886
F1-score	0.921	0.911

TABLE 4 Validation metrics for order/suborder classifier. ResNet50v2 without uncertain are the metrics where 235 samples are removed by the anomaly threshold detector.

	ResNet50v2	ResNet50v2 without uncertain samples
Samples	25,034	24,799
Precision	0.974	0.977
Recall	0.965	0.971
F1-score	0.970	0.974

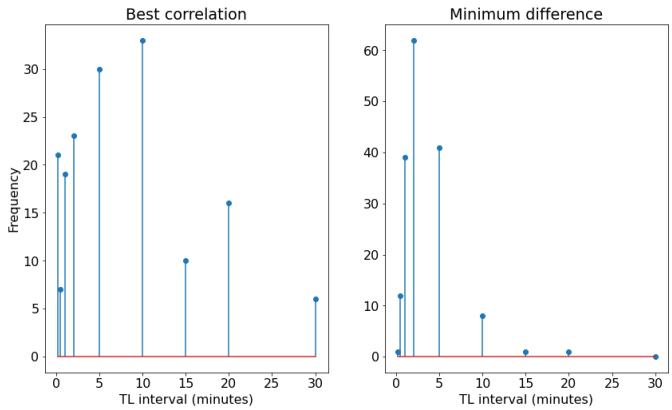


FIGURE 6 Show the frequency of best correlation and minimum difference between tracking (0.5fps) and different time-lapse intervals. Each observation concerns the best correlation or minimum difference for one trap and one of the 15 insect taxa.

The confusion matrix for the classifier is shown in Figure 5, here we have added the uncertain class for predictions below the anomaly threshold. High values are observed in the diagonal of the matrix indicating accurate classification. However, difficulties are observed in classifying winter craneflies (Diptera Trichocera) from mosquitoes (Diptera Nematocera), this is due to a visually similar appearance and possible errors in the dataset for these two taxa of insects.

We have evaluated the order/suborder classifier with anomaly detection on the 10-minute time-lapse recordings from 2022 and 2023. This is done by selecting up to 200 randomly classified insects of the 16 taxa of categories above and below the threshold learned from the output distribution of the dataset. We have manually verified by visual inspection the classified insects above and below the anomaly threshold. The results listed in Table 5 show that in total 92.3% of the insects are classified above the threshold with a precision of 96.8%. The remaining 7.7% insect detections classified as uncertain anomaly have a similar high precision of 95.9%. Spiders (Araneae) are the group of insects with the lowest precision of 83% above the threshold. This is due to many of the false positive detections being spider webs or dirt since the training data does contain spiders with prey and more blurry and unclear objects.

TABLE 5 Number of classified insects taxa above and below anomaly threshold (TH) with precision based on maximum 200 random selected predictions. Above precision is the precision for classified insect taxa. Below precision is the precision for uncertain classified insects.

Order/suborder category	Above TH	Below TH	Above (%)	Above precision (%)	Below precision (%)
Araneae	4,957	864	85	83	100
Coleoptera	2,896	351	89	96	98
Diptera Brachycera	23,941	1,850	93	97	98
Diptera Nematocera	171,855	8,760	95	92	98
Diptera Tipulidae	1,858	56	97	98	100
Diptera Trichocera	7,070	26	100	100	96
Ephemeroptera	25,423	7,374	78	98	100
Hemiptera	8,008	345	96	92	100
Hymenoptera Other	6,906	182	97	100	92
Hymenoptera Vespidae	542	91	86	100	91
Lepidoptera Macros	46,425	2,871	94	100	89
Lepidoptera Micros	93,870	6,006	93	97	98
Neuroptera	1,094	60	95	100	95
Opiliones	673	50	93	97	98
Trichoptera	64,715	8,935	88	98	100
Vegetation	2,050	951	68	100	96
Total/average	462,283	38,772	92.3	96.8	95.9

3.3 | Computational speed and power usage

The speed performance of our pipeline was evaluated across various edge computing platforms. These included a standard computer equipped with an Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz and an NVIDIA TITAN X Pascal GPU, as well as the NVIDIA Jetson Nano (JN) and Raspberry Pi 4 (RP4) (both with 4 GB of memory) and Raspberry Pi 5 (RP5) (with 8 GB of memory). An additional 4 GB swap file was required to execute the processing pipeline on the Jetson Nano for nights with more than 60 insects. This was necessary because the classification is performed in batches, processing all insects detected in one image simultaneously to enhance performance. The three edge computing devices were selected since they are roughly at the same price, with the JN being the most expensive at about twice the price of RP5 since JN also has a NVIDIA Tegra X1 GPU computer. We have tested the pipeline processing 6,271 images from one night with high insect activity. Detailed time performance and power metrics are provided in Table 6. During the test, an external SSD drive with images was connected to the Raspberry Pi and Jetson Nano computers consuming 0.9W of the idle and active power.

3.4 | Recording statistics

12 traps were operational a total of 1,399 nights during the first year 2022, where 3,428,430 images were captured with 45,285,526 insect detections and 94% were above the anomaly threshold and used for time-lapse analysis. 97% of the detections contribute to 1,177,728 valid insect tracks with a duration of more than 4 seconds and a classification certainty of >50% for the classified insect. A summary of detailed statistics for all traps in 2022 are found in Appendix Table A1.

3.5 | Tracking and time-lapse sampling

Tracking was evaluated by creating videos with insect classification and colored tracks as illustrated in Appendix Figure A1. Video sequences with many fast-moving insects of similar classes did have some difficulties, see example video[‡] with moderate insect activity. Here, it was observed that an insect track will occasionally be associated with a wrong insect. This problem could be minimized by lowering the cost threshold, but it requires a higher sampling rate to ensure the distance an insect has moved between two frames is short.

An evaluation was made by comparing insect abundance by using tracking of insects in images recorded with 0.5 fps compared to selecting time-lapse images recorded by different time intervals. For each trap and insect order/suborder, the abundance was calculated using tracking and the result was compared with a time-lapse approach. The time-lapse approach did not include tracking since large sample intervals were used. Here, we used localization and classification with anomaly detection by removing the detections with a classification score below the learned threshold described in section 2.4.1. We have simulated time-lapse sampling of 10s, 30s, 1m, 2m, 5m, 10m, 15m, 20m and 30m by only including image detections with these intervals.

The mean absolute difference between the time series of tracks and time-lapse insect detections is shown in Figure 7. The red lines show the average difference for all traps for each insect category. We assume the minimum difference would be the optimal

TABLE 6 Processing performance on standard computer, Jetson Nano (JN) and Raspberry Pi's (RP) of four hours recording with 6.271 images with an average of 35-37 insect detections per image.

Platform	YOLOv5 model	Detected insects	Detector sec/img	Classifiers ms/det.	Tracker ms/img	Average sec/img	Total hours	Power (W) idle/active
Standard	s6	221,732	0.018	7.5	15	0.296	0.52	40/90
Standard	m6	231,835	0.032	6.7	15	0.295	0.51	40/90
JN	s6	221,732	0.351	57.6	197	2.58	4.50	3.7/10.2
JN	m6	231,835	0.848	58.8	197	3.22	5.61	3.7/10.2
RP5	s6	221,732	1.97	193	88	8.87	15.45	3.5/10.4
RP5	m6	231,835	4.30	189	88	11.4	19.82	3.5/10.4
RP4	s6	221,732	5.30	464	169	21.9	38.14	3.0/6.8
RP4	m6	231,835	9.99	484	169	28.1	48.87	3.0/6.8

‡ https://www.youtube.com/watch?v=HzOCYlgnhlE

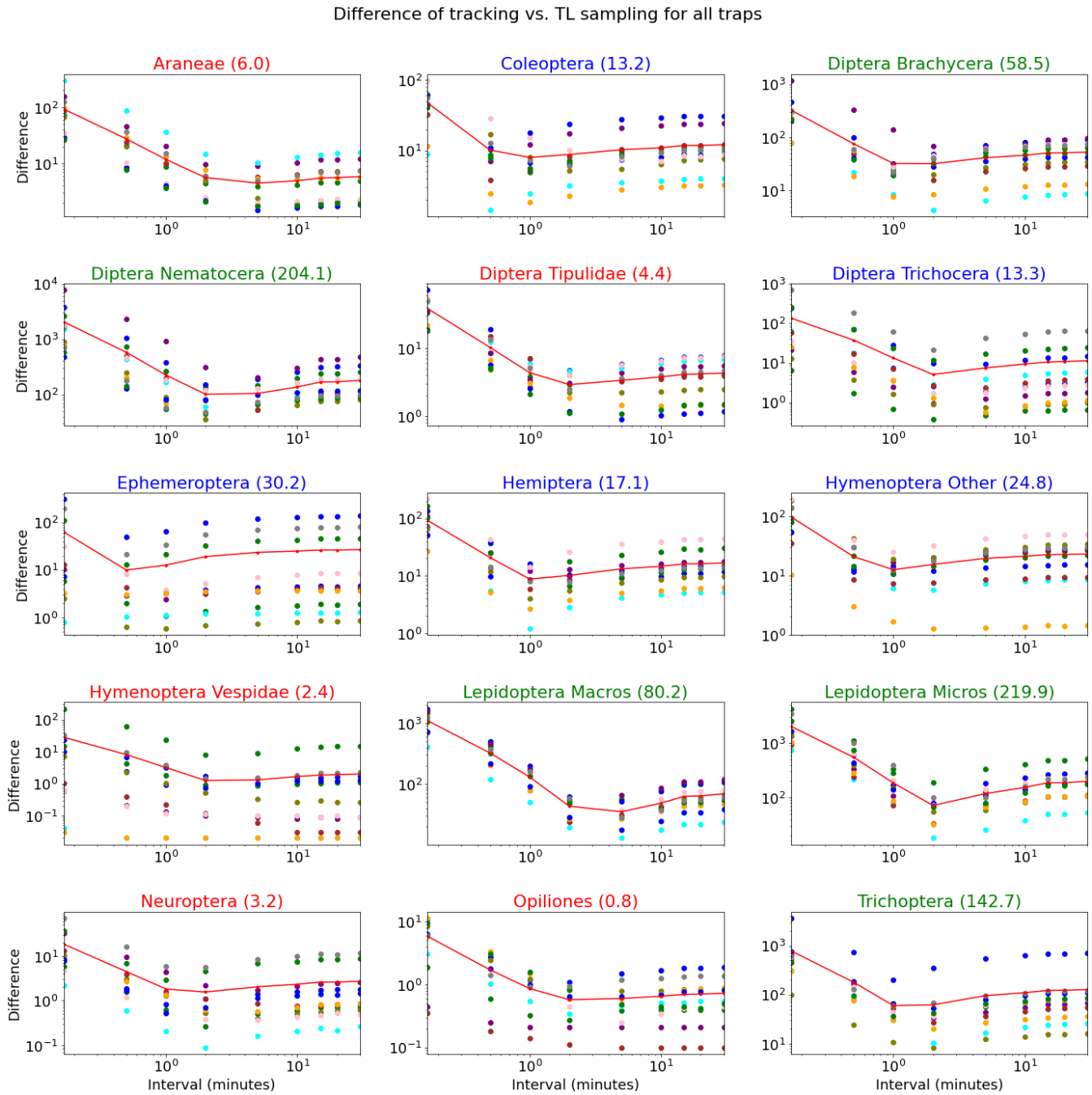


FIGURE 7 Shows the mean absolute difference between tracks and time-lapse (TL) intervals (10s-30m) for all nocturnal insects and traps (Difference marked with "•" of different colors for each trap). The average difference for all traps is shown with a red line. The values in parenthesis are the average number of tracks per night. Green indicates taxa with a high abundance above 50 tracks per day. Blue indicates taxa between 10-50 tracks per night and red below 10 tracks per night.

time-lapse sampling interval for measuring a close approximation to tracking. For short TL intervals (below minimum), the absolute difference is high since there are more detections than tracks. For longer TL intervals (above minimum) the absolute difference starts to increase again indicating more tracks than detections are encountered.

The Pearson correlation was also applied to the time series of insect tracks and time-lapse insect detections. This way, we compare the relative temporal trends in abundance between the two methods, independent of the mean absolute difference. An example of Pearson correlation from four different traps and insect groups is shown in Appendix Figures A2 and A3. It is

observed that the best correlation is achieved by 5-minute sampling intervals except for Vespidae. Here, the best correlation is achieved with 10 sec. The same tendency is observed for all time-lapse sampling intervals compared to tracking although the number of observed detections and tracks differs.

In Appendix Figure A4 we have summarized the Pearson correlation of time-lapse sampling intervals and tracks for the traps and 15 insect orders/suborders of taxa. The correlation is generally high (0.9) for insect groups with high abundance above 50 insect tracks per night. For insect groups with low abundance below 10 insect tracks per night, the correlation drops for sampling intervals above 10 minutes.

The duration an individual insect is visible should also be considered when deciding on a time-lapse interval. Appendix Figure A5 shows the distribution of track duration for each of the 15 insect orders/suborders from the camera trap with high activity. The average duration ranges from 29 to 115 seconds and varies among the different insect groups. Notably, Lepidoptera has the longest track durations compared to most other groups.

Figure 6 summarizes the frequency of minimum difference and best correlation of sample intervals with tracking. It indicates that a time-lapse interval of 2 minutes does in more than 60 situations achieve the best minimum difference between time-lapse detections and tracking. The best Pearson correlation is achieved for 10 minutes, however, 5 and 2-minute TL intervals are also a good choice as an alternative to tracking with 0.5fps. This approach would reduce the amount of recorded images and be able to perform edge processing of images in real-time on the camera trap with Raspberry Pi.

4 | DISCUSSION

This work proposes a deep learning pipeline for monitoring nocturnal insects, encompassing detection, classification, and tracking within selected taxonomic groups, as well as filtering anomalies such as blurry, dark, partly visible, or uncertain insect images. Image recordings were collected from 12 camera traps situated in three different habitats: bogs, heaths and forests.

Interestingly, our findings reveal that the YOLOv5m6 object detector model exhibits a higher insect detection rate, averaging 4.4% higher than YOLOv5s6 similar to observed in (Bjerge et al. 2023a). However, it is noteworthy that while YOLOv5s6 demonstrates a slightly lower recall during evaluation, it boasts faster processing speeds, particularly evident on Raspberry Pi devices. Surprisingly, the performance advantage observed with YOLOv5s6 on Raspberry Pi platforms contrasts with the less significant improvement noted when utilizing GPU acceleration.

For the order/suborder classifier evaluated on time-lapse recordings taken at 10-minute intervals, we identified 7.7% anomalies out of 462,283 insect detections, achieving a high precision of 95.9%, indicating that our anomaly filter effectively removes uncertain classifications. The classification of insects into 16 categories of insect orders/suborders achieved a precision of 96.8%, based on an evaluation of 200 random samples from the time-lapse recordings in 2022 and 2023. These results indicates the the anomaly detector finds and removes uncertain anomalies that may manifest as debris, obscured or partially visible insects, or fall outside the scope of the created dataset's insect orders.

We propose a simple track-by-detection algorithm that matches feature embeddings, distance and area for insect tracking in sequential images at a frame rate of 0.5 fps. The tracking algorithm has been preliminarily evaluated by visually inspecting videos created by the proposed pipeline, however the same algorithm without the embedding feature cost was evaluated in Bjerge et al. (2021b). Even if the amount of data collected is substantial, the frame rate is rather low for tracking, and can sometimes be inaccurate for moving insects of the same species due to the long distances they can travel between frames. Further work is needed to improve this tracking algorithm before conducting a thorough evaluation.

As alternative to the AMI moth species classifier (Jain et al. 2024), researchers may consider employing the European moth species classifier developed by Korsch, D., Bodesheim, P., Denzler, J. (2021), renowned for its high accuracy and trained on a more expansive dataset comprising both species and images. It is noteworthy that the AMI data companion code contains additional trained models capable of classifying moth species from specific regions such as Canada (Quebec/Vermont) and Panama. A single model to perform classification with hierarchical taxonomic ranks has been proposed in other studies (Bjerge et al. 2023b), however creating a dataset that enables the use of such a model will be very demanding.

We compared tracking with a time-lapse sampling approach without tracking. Lower time-lapse sampling intervals (10 seconds to 30 minutes) significantly reduce the amount of collected data. We found a high correlation between tracking and time-lapse sampling, particularly for insect orders/suborders with more than 50 tracks per night. For insects with fewer than 10 tracks per night, the correlation decreases for time-lapse intervals longer than 10 minutes. The best mean absolute difference was found for 2-minute intervals, but for insects with fewer than one detection per night, the correlation decreases.

Edge processing on the JN platform is by far the fastest and most power-efficient since RP5 and JN nearly consume the same amount of power, but JN is more than four times faster processing image with an average of 2.6 seconds. It is observed that RP4 is the slowest, however, the average processing (28.1s) time is still less than 30 seconds and therefore suitable for time-lapse sampling down to intervals of 30 sec. when processing in real-time is performed. The most power-efficient processing platform is the JN computer which on average consumes 26.3 Joule per image (YOLOv5s6) where RP4 consumes 148.9 Joule and RP5 92.3 Joule.

All edge processing platforms are suitable for a time-lapse sampling approach with intervals as short as 10 to 30 seconds. Among these, the JN stands out as the most energy-efficient solution, capable of real-time tracking at nearly 0.5 fps. The JN platform has also been employed for real-time monitoring of diurnal insects, as demonstrated by Bjerge et al. (2021a). Alternatively, the video monitoring platform suggested by Sittinger et al. (2024)) could be considered for deploying our proposed pipeline directly on the Luxonis OAK-1 camera with processing power, though some modifications may be necessary.

AUTHOR CONTRIBUTIONS

All authors have seen and approved the submitted version of the manuscript. All authors have contributed to the work, and all persons entitled to co-authorship have been included. Kim Bjerge: Conceptualization, Investigation, Methodology, Coding, Formal Analysis, Writing — Original Draft, Writing — Review & Editing, Visualization. Henrik Karstoft: Supervision, Writing — Review & Editing, Supervision. Toke T. Høye: Conceptualization, Data curation, Resources, Writing — Review & Editing, Supervision, Funding.

ACKNOWLEDGMENTS

This research was funded by Aage V. Jensen Naturfond grant number N2024-0013 and the European Union's Horizon Europe Research and Innovation programme, under Grant Agreement No. 101060639 (MAMBO). We would also like to acknowledge Nathan Pinoy who has been part of collecting images and contributed to the creation of the order/suborder dataset.

FINANCIAL DISCLOSURE

None reported.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

REFERENCES

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

Barlow, S.E. & O'Neill, M.A. (2020) Technological advances in field studies of pollinator ecology and the future of e-ecology. *Current Opinion in Insect Science*, 38. doi:10.1016/j.cois.2020.01.008.

Beery, S., Van Horn, G. & Perona, P. Recognition in Terra Incognita. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11220 LNCS*, 2018*.

Besson, M., Alison, J., Bjerge, K., Gorochowski, T.E., Høye, T.T., Jucker, T. et al. (2022) Towards the fully automated monitoring of ecological communities. *Ecology Letters*, 25(12). doi:10.1111/ele.14123.

Bjerge, K., Alison, J., Dyrmann, M., Frigaard, C.E., Mann, H.M.R. & Høye, T.T. (2023) Accurate detection and identification of insects from camera trap images with deep learning. *PLOS Sustainability and Transformation*, 2(3). doi:https://doi.org/10.1371/journal.pstr.0000051. URL https://doi.org/10.1371/journal.pstr.0000051

Bjerge, K., Geissmann, Q., Alison, J., Mann, H.M.R., Høye, T.T. & Dyrmann, M. (2023) Hierarchical Classification of Insects with Multitask Learning and Anomaly Detection. *Ecological Informatics*, 77(102278). doi:https://doi.org/10.1016/j.ecoinf.2023.102278.

Bjerge, K., Mann, H.M. & Høye, T.T. (2021) Real-time insect tracking and monitoring with computer vision and deep learning. *Remote Sensing in Ecology and Conservation*, 8(6). doi:10.1002/rse2.245.

Bjerge, K., Nielsen, J.B., Sepstrup, M.V., Helsing-Nielsen, F. & Høye, T.T. (2021) An automated light trap to monitor moths (Lepidoptera) using computer vision-based tracking and deep learning. *Sensors (Switzerland)*, 21(2). doi:10.3390/s21020343.

Bulusu, S., Kailkhura, B., Li, B., Varshney, P.K. & Song, D. (2020) Anomalous Example Detection in Deep Learning: A Survey. *IEEE Access*, 8. doi:10.1109/ACCESS.2020.3010274.

Didham, R.K., Basset, Y., Collins, C.M., Leather, S.R., Littlewood, N.A., Menz, M.H. et al. (2020) Interpreting insect declines: seven challenges and a way forward. *Insect Conservation and Diversity*, 13(2). doi:10.1111/icad.12408.

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

Glenn Jocher (2020) *You Only Look Once Ver. 5 (YOLOv5) on Github*. https://github.com/ultralytics/yolov5 [Accessed: 21.05.2024].

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*. Vol. 2016-Decem*, 2016*.

Høye, T.T., Ärje, J., Bjerge, K., Hansen, O.L.P., Iosifidis, A., Leese, F. et al. (2021) Deep learning and computer vision will transform entomology. *Proceedings of the National Academy of Sciences*, 118(2). doi:10.1073/pnas.2002545117.

Jain, A., Cunha, F., Bunsen, M.J., Cañas, J.S., Pasi, L., Pinoy, N. et al. (2024) Insect identification in the wild: The ami dataset,.doi:arXiv:2406.12452.

Kieritz, H., Hubner, W. & Arens, M. (2018) Joint detection and online multi-object tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2018-June. doi:10.1109/CVPRW.2018.00195.

Korsch, D., Bodesheim, P., Denzler, J. (2021) Deep Learning Pipeline for Automated Visual Moth Monitoring: Insect Localization and Species Classification. *INFORMATIK 2021. Gesellschaft für Informatik, Bonn.*,, 443–460.doi:https://dx.doi.org/10.18420/informatik2021-036. URL https://dl.gi.de/handle/20.500.12116/37700

Lima, M.C.F., Leandro, M.E.D.d.A., Valero, C., Coronel, L.C.P. & Bazzo, C.O.G. (2020) Automatic detection and monitoring of insect pests - A review. *Agriculture (Switzerland)*, 10(5). doi:10.3390/agriculture10050161.

Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J. et al. (2015) Microsoft COCO: Common Objects in Context. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*,.

Logitech (2021) *Brio ultra hd pro webcam.* https://www.logitech.com/en-us/products/webcams/brio-4k-hdr-webcam.960-001105.html [Accessed: 21.05.2024].

Möglich, J.M., Lampe, P., Fickus, M., Younis, S., Gottwald, J., Nauss, T. et al. (2023) Towards reliable estimates of abundance trends using automated non-lethal moth traps. *Insect Conservation and Diversity*,(April 2022), 1–11. doi:10.1111/icad.12662.

Montgomery, G.A., Belitz, M.W., Guralnick, R.P. & Tingley, M.W. (2021) Standards and Best Practices for Monitoring and Benchmarking Insects. *Frontiers in Ecology and Evolution*, 8. doi:10.3389/fevo.2020.579193.

Motion (2021) *Motion an open source program that monitors video from cameras.* https://motion-project.github.io/ [Accessed: 21.05.2024].

Oliver, R.Y., Iannarilli, F., Ahumada, J., Fegraus, E., Flores, N., Kays, R. et al. (2023) Camera trapping expands the view into global biodiversity and its change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378(1881). doi:10.1098/rstb.2022.0232.

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

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016*.

Rolnick, David; Bundsen, Michael; Jain, Aditya; Cunha, F. (2023) *AMI data companion*. https://github.com/RolnickLab/ami-data-companion [Accessed: 21.05.2024].

Roy, D.B., Alison, J., August, T.A., Bélisle, M., Bjerge, K., Bowden, J.J. et al. (2024) Towards a standardized framework for of AI-assisted, image-based monitoring of nocturnal insects. *Philosophical Transactions B*, 379(20230108). doi:https://doi.org/10.1098/rstb.2023.0108.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S. et al. (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3). doi:10.1007/s11263-015-0816-y.

Shao, S., Zhao, Z., Li, B., Xiao, T., Yu, G., Zhang, X. et al. (2018) CrowdHuman: A Benchmark for Detecting Human in a Crowd. *arXiv*,, 1–9. URL http://arxiv.org/abs/1805.00123

Sittinger, M., Uhler, J., Pink, M. & Herz, A. (2024) Insect detect: An open-source DIY camera trap for automated insect monitoring. *PLoS ONE*, 19(4 April), 1–28. doi:10.1371/journal.pone.0295474. URL http://dx.doi.org/10.1371/journal.pone.0295474

van Klink, R., August, T., Bas, Y., Bodesheim, P., Bonn, A., Fossøy, F. et al. (2022) Emerging technologies revolutionise insect ecology and monitoring. *Trends in Ecology and Evolution*, 37(10), 872–885. doi:10.1016/j.tree.2022.06.001.

Wagner, D.L., Grames, E.M., Forister, M.L., Berenbaum, M.R. & Stopak, D. (2021) Insect decline in the Anthropocene: Death by a thousand cuts. *Proceedings of the National Academy of Sciences of the United States of America*, 118(2). doi:10.1073/PNAS.2023989118.

Wojke, N., Bewley, A. & Paulus, D. (2018) Simple online and realtime tracking with a deep association metric. *Proceedings - International Conference on Image Processing, ICIP*, 2017-September. doi:10.1109/ICIP.2017.8296962.

Zhang, Y., Wang, C., Wang, X., Zeng, W. & Liu, W. (2021) FairMOT: On the Fairness of Detection and Re-identification in Multiple Object Tracking. *International Journal of Computer Vision*, 129(11). doi:10.1007/s11263-021-01513-4.

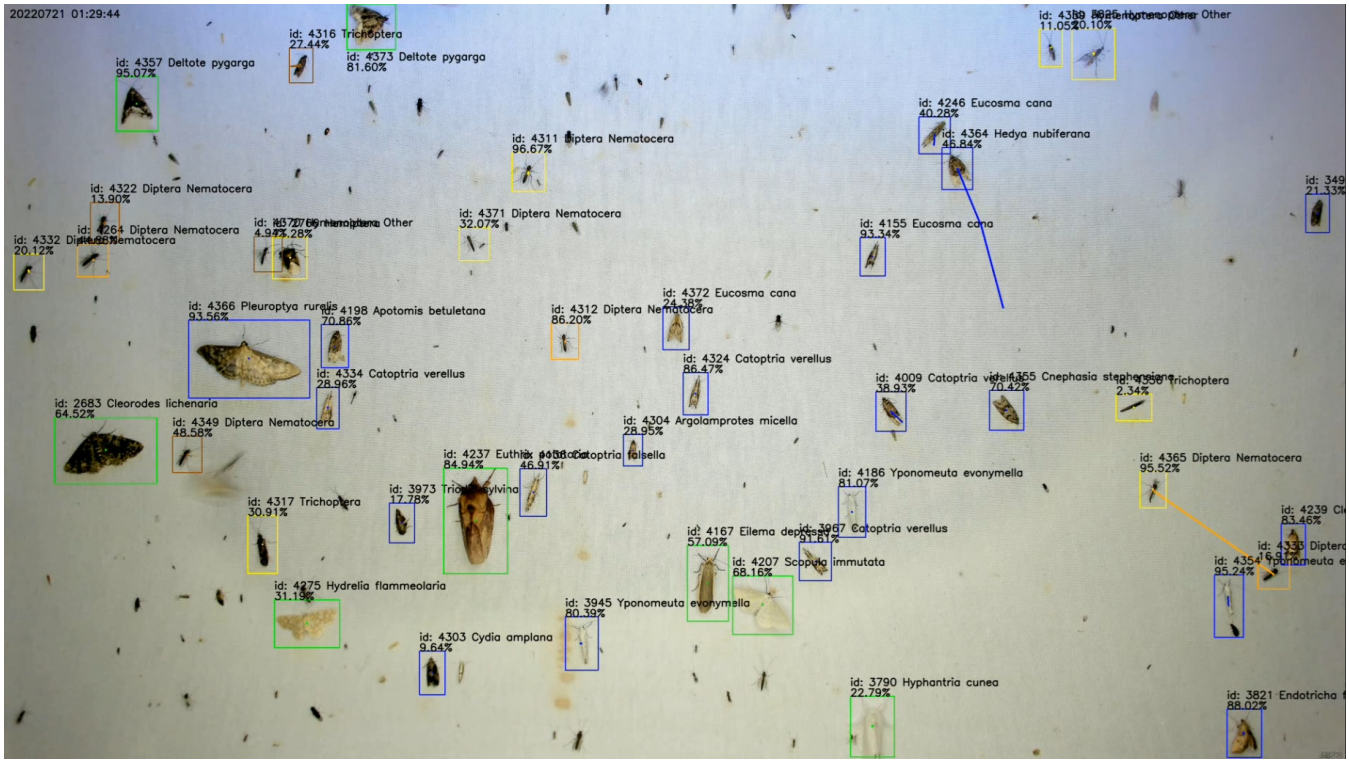


FIGURE A1 Example of image from camera trap processed with the pipeline to track and classify insects in taxa of order/suborder and moth species.

□

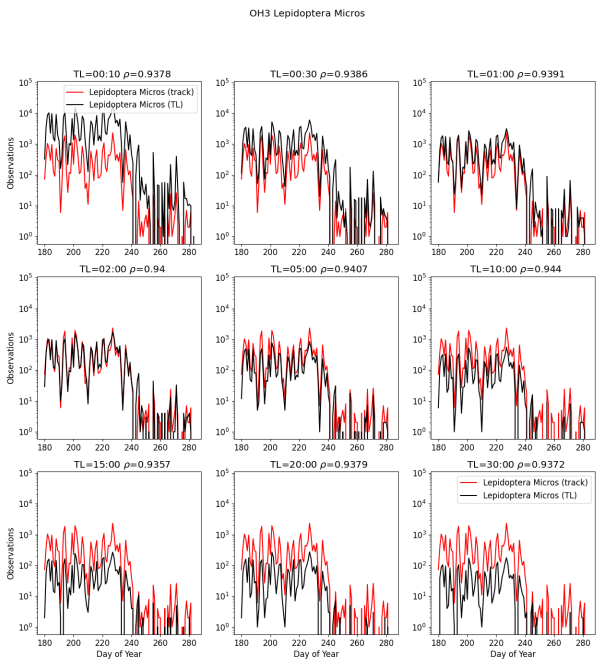
APPENDIX

A SUPPLEMENTARY MATERIAL

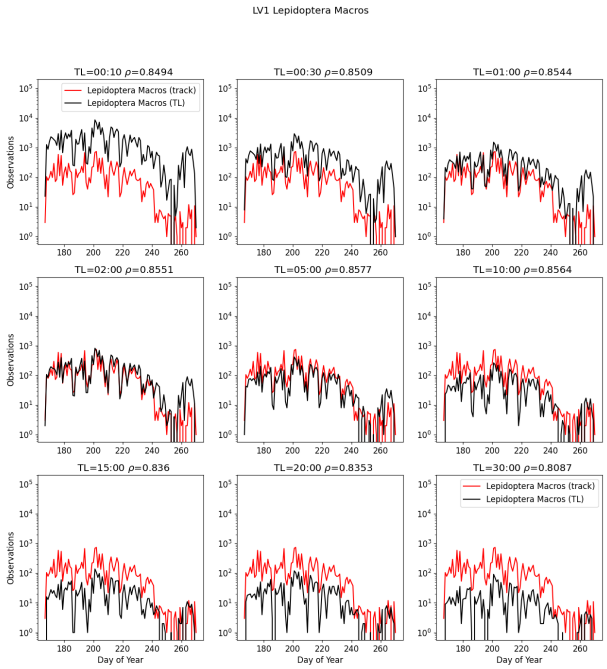
Figure A1 shows an image recorded from a trap processed by the processing pipeline to detect, track and classify insects in taxa of order/suborder and moth species. Figure A2 and A3 shows the abundance measured using tracking of images recorded with 0.5fps or filtered detections and classifications with time-lapse intervals ranging from 10 seconds to 30 minutes. The Pearson correlation between tracking and filtered classifications with varying TL intervals is shown in the title of each plot. Figure A4 shows the Pearson correlations between tracks and time-lapse intervals for all nocturnal insect and traps.

Table A1 contains list of the statistics of operational nights, number of recorded images, number of insect detections and insect tracks. Trap SS3 only collected images in the first 11 nights before tilted over by a wild boar. Trap OH1 was not operational at the start of the season due to technical issues. The power cable to trap LV1 was broken by humans driving a forest machine at the end of the monitoring season. Traps OH1-OH4 where powered by a 175W solar panel and 12V/100Ah battery.

Figure A5 shows the histograms for the distribution of track duration for the 15 different orders or suborders of insect taxa from one of the traps with the highest and most diverse insect activity.

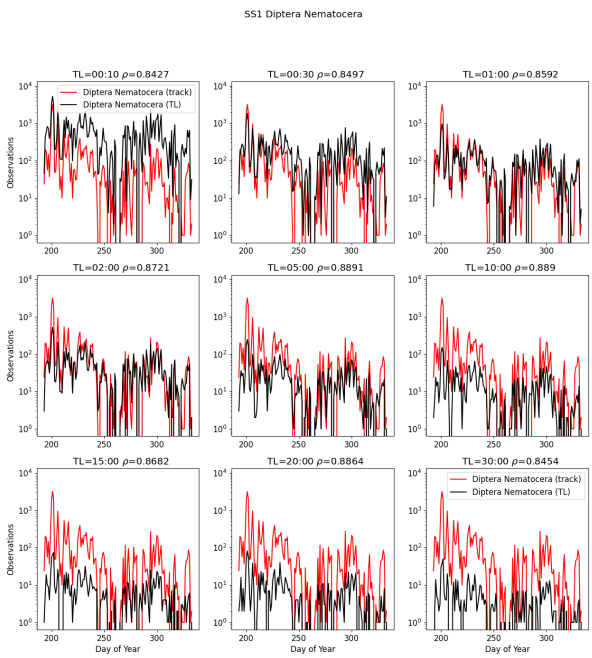


(a) Pearson correlation for micro moths (Camera trap OH3).
Best correlation of 0.9407 with 5 minutes interval.

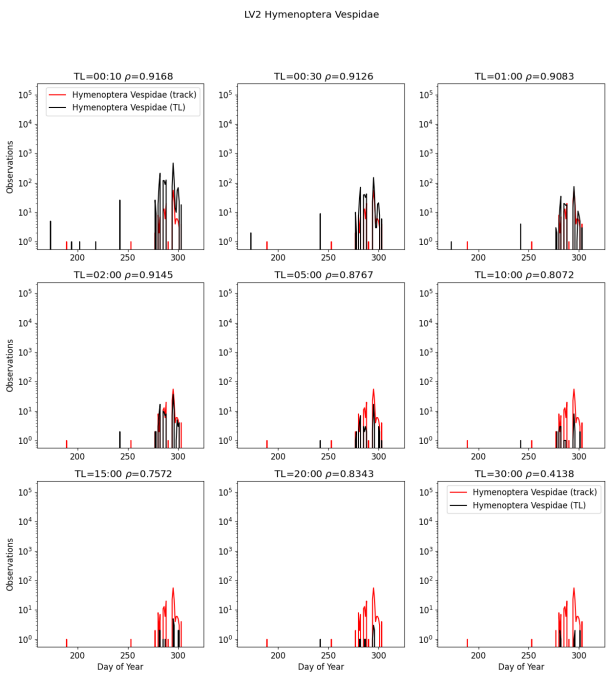


(b) Pearson correlation for macro moths (Camera trap LV1).
Best correlation of 0.8577 with 5 minutes interval.

FIGURE A2 Shows the number of observations at several different time-lapse (TL) intervals (10s-30m) and tracking (track) for Lepidoptera micros and macros at example traps.



(a) Pearson correlation for mosquitoes (Camera trap SS1).
Best correlation of 0.8891 with 5 minutes interval.



(b) Pearson correlation for wasps (Camera trap LV2).
Best correlation of 0.9168 with 10 seconds interval.

FIGURE A3 Shows the number of observations at several different time-lapse (TL) intervals (10s-30m) and tracking (track) for Diptera Nematocera and Hymenoptera Vespidae at example traps.

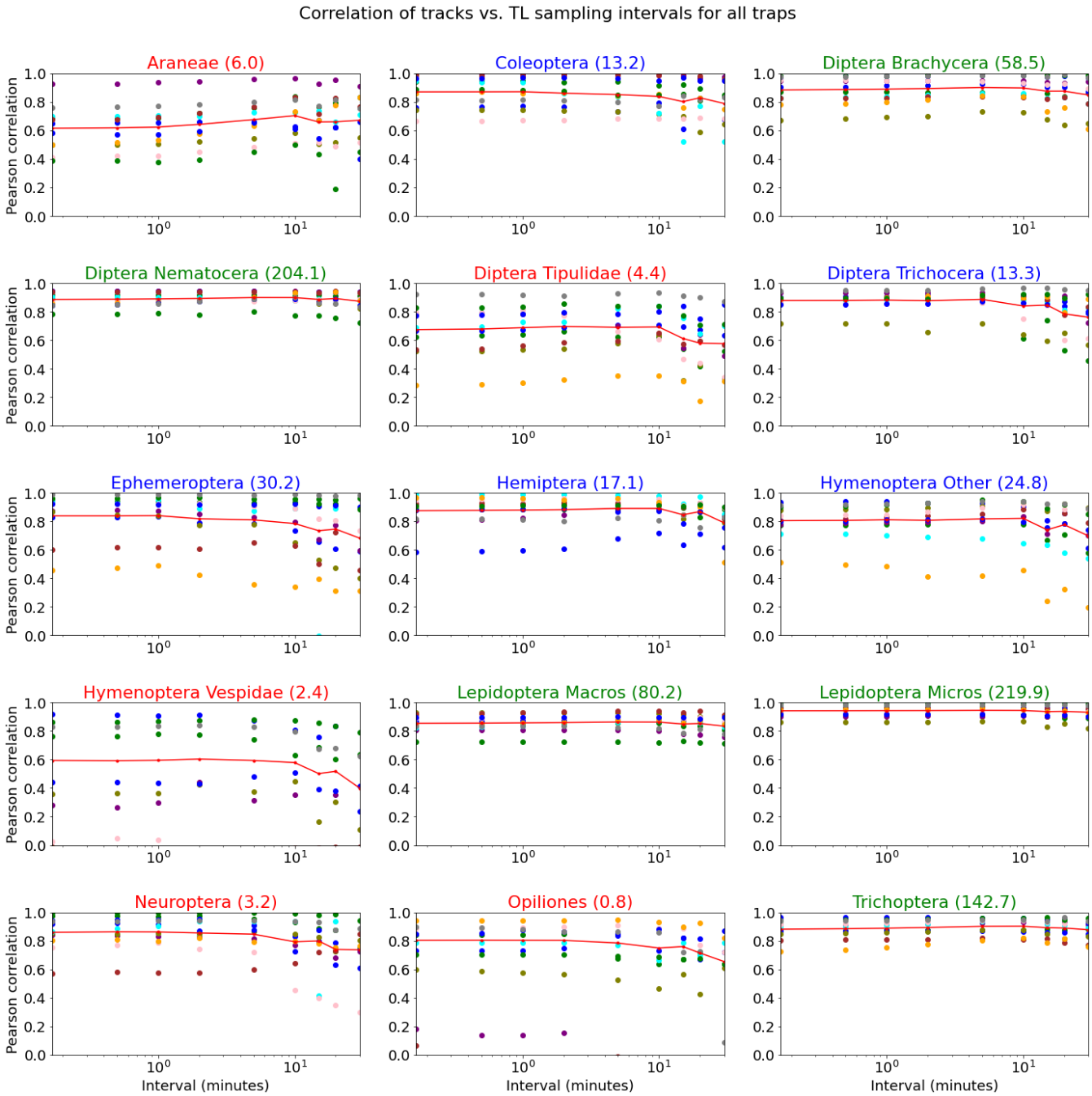


FIGURE A4 Pearson correlation between tracks and time-lapse (TL) intervals (10s-30m) for all nocturnal insects and traps (Correlation marked with "•" of different colors for each trap). The average Pearson correlation is shown with a red line. The values in parenthesis are the average number of tracks per night. Green indicates taxa with a high abundance above 50 tracks per day. Blue indicates taxa between 10-50 tracks per night and red below 10 tracks per night.

TABLE A1 Statistics for processing all images recorded in 2022 from 12 camera traps. The detections above the anomaly threshold (>TH) indicate the percentage of valid detections. Valid tracks are defined for insects visible more than 4 sec. and classified for the same taxa in more than 50% of the detections in a track.

Trap	Days	Files	Detections	Detections >TH (%)	Tracks	Valid tracks	Detections in tracks (%)	Valid tracks (%)
LV1	103	361,429	5,046,324	96	151,174	125,420	97	83
LV2	151	464,950	6,742,175	94	201,241	167,879	97	83
LV3	135	371,851	8,894,099	94	196,214	165,082	97	84
LV4	151	309,317	2,327,043	94	76,128	62,495	97	82
OH1	82	131,441	959,849	94	32,031	26,588	96	83
OH2	118	208,048	2,027,460	94	66,450	53,052	96	80
OH3	116	300,387	3,488,553	93	113,569	91,076	96	80
OH4	111	246,971	1,387,302	91	52,089	41,767	95	80
SS1	140	292,991	5,196,560	93	233,952	193,596	96	83
SS2	140	324,401	33,92,246	95	115,447	93,216	96	81
SS3	11	50,879	1,011,345	93	37,089	30,059	96	81
SS4	141	365,765	4,809,570	95	153,823	127,498	97	83
T/A	1399	3,428,430	45,282,526	94	1,429,207	1,177,728	97	82

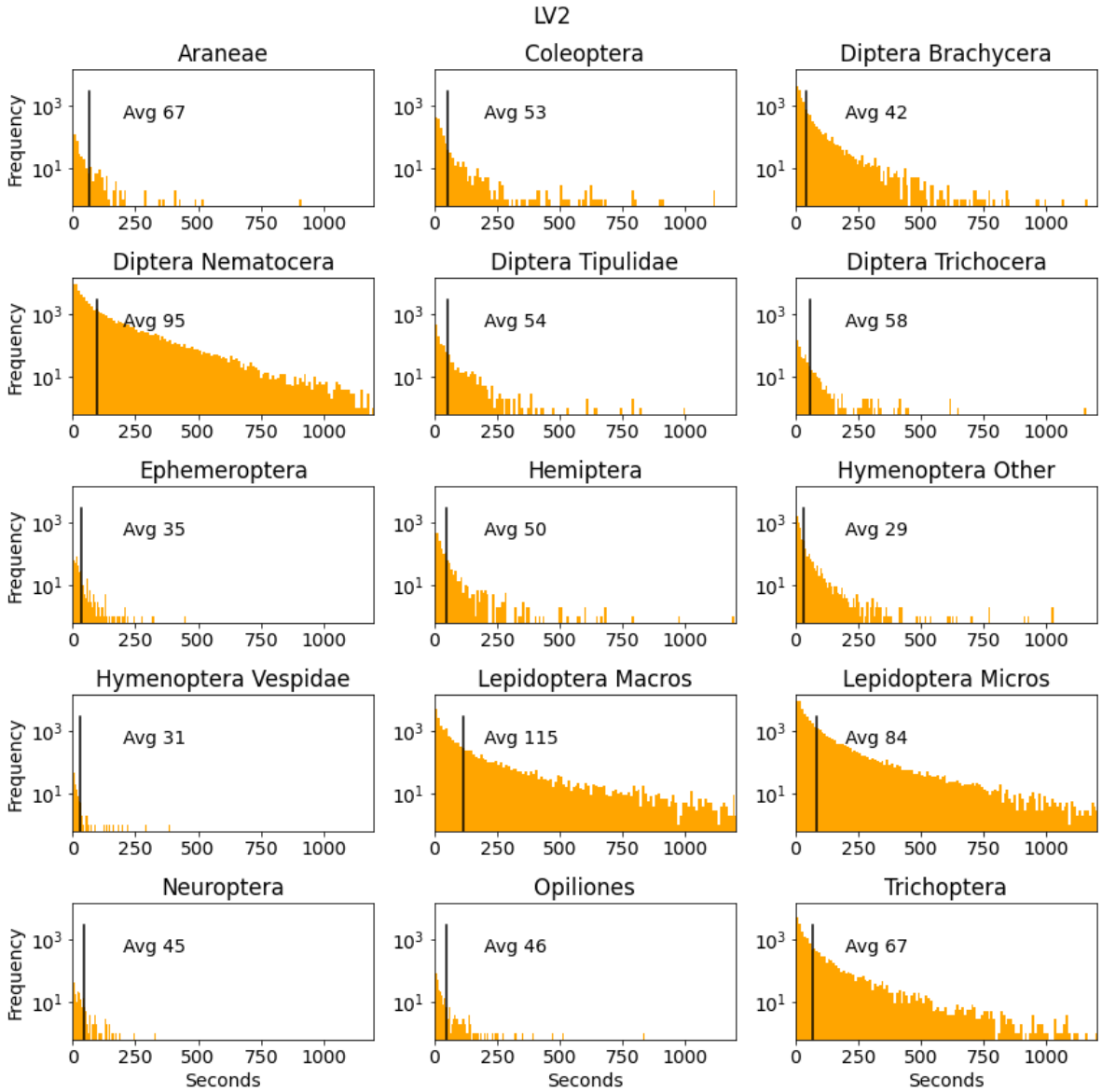


FIGURE A5 Shows the distribution of insect track durations for the 15 orders/suborders of insects and the average duration of tracks in seconds from camera trap LV2. The horizontal lines show the average time of tracks in seconds. The x-axis is limited to 20 minutes, however, some insects stayed in the camera view for the whole night.