Monitoring Social Insect Activity with Minimal Human Supervision

Tarun Sharma*+[1], Julian M. Wagner*[1], Sara Beery[2], William B. Dickson[1], Michael H. Dickinson[1], and Joseph Parker[1]

[1]California Institute of Technology, USA
[2]MIT, USA

Abstract

Tracking the behavior of animals and their group dynamics in nature offers a crucial look into the delicate ecological networks that compose wildlife diversity. The velvety tree ant (Liometopum occidentale) is an ecologically dominant ant species found in South Western North America; their extensive foraging activity shapes forest communities, and their nests are a biodiversity hot-spot for a multitude of symbiotic invertebrates (myrmecophiles). Despite their vital role in the ecosystem, their activity is largely unstudied. In this work, we develop a multi-sensor camera trap, named 'Ethocam,' to capture ant behavioral patterns in the field, and combine this technology with a computer vision approach to track colony activity in an undisturbed fashion. We demonstrate an accurate system for counting ants built with minimal human labeling. We show that L. occidentale activity drops rapidly through the morning and study the effect of environmental conditions on ant count. We also report the occurrences of the ants' interactions with other invertebrates in our camera trap data. Together, these findings demonstrate the potential of our system to capture the behavior of Liometopum occidentale as well as its complex associations with various local species including symbionts, potentially at landscape scale. Our study provides proof of concept for the promise of low-cost remote monitoring of social insect populations.

1. Introduction

Studying animal activity patterns in natural habitats is a crucial complement to studying behaviors in the lab, since it is usually impossible to recapitulate the complex web of interspecies associations under laboratory conditions [26][30][21]. Moreover, quantifying animal behavior in natural contexts is critical for understanding how a given species contributes to the dynamics of ecological communities and higher-level ecosystem processes. This is especially true for ecologically important social insects such as ants, bees and termites, where colonies are often large and exert major effects on the habitat at large by way of their associations with other species and impacts on nutrient distribution and habitat structure. In many terrestrial ecosystems, ants in particular are keystone organisms that control the populations of diverse other invertebrates, both via predation and by forging beneficial symbioses with mutualistic herbivores [24]. The ramifications of these interactions can permeate whole communities, even re-configuring the predator-prey dynamics in large mammals [16]. Ecosystems can thus be especially sensitive to changes in the composition of the ant fauna. Human-mediated habitat loss and fragmentation, climate change, and the introduction of exotic species, including invasives, have all been shown to disrupt native ant communities, with often dramatic consequences for the native ecosystem [24].

The ability to monitor the collective behavior and ecological interactions of ants in the field, long term, is hugely desirable [31]. Such data would not only illuminate natural, colony-level behaviors that are impossible to reconstitute in the lab, but could also permit quantification of the ecosystem services that ants provide, as well as lead to predictive models for how ant species may respond to different types of disturbance. Potentially, such a monitoring system could be implemented at a large scale, across multiple colonies within a habitat. To our knowledge, no such automated approach has to date been developed.

The velvety tree ant (*Liometopum occidentale*) is an ecologically dominant ant species in Southern California, forming extensive colonies containing hundreds of thousands to millions of workers. Despite the high prevalence of this ant species in semi-intact ecosystems across its range, little is known about its biology and behavior in the wild [12]. These ants are known to have large foraging areas, are

*These authors contributed equally to this work
+Correspondence to: tsharma@caltech.edu

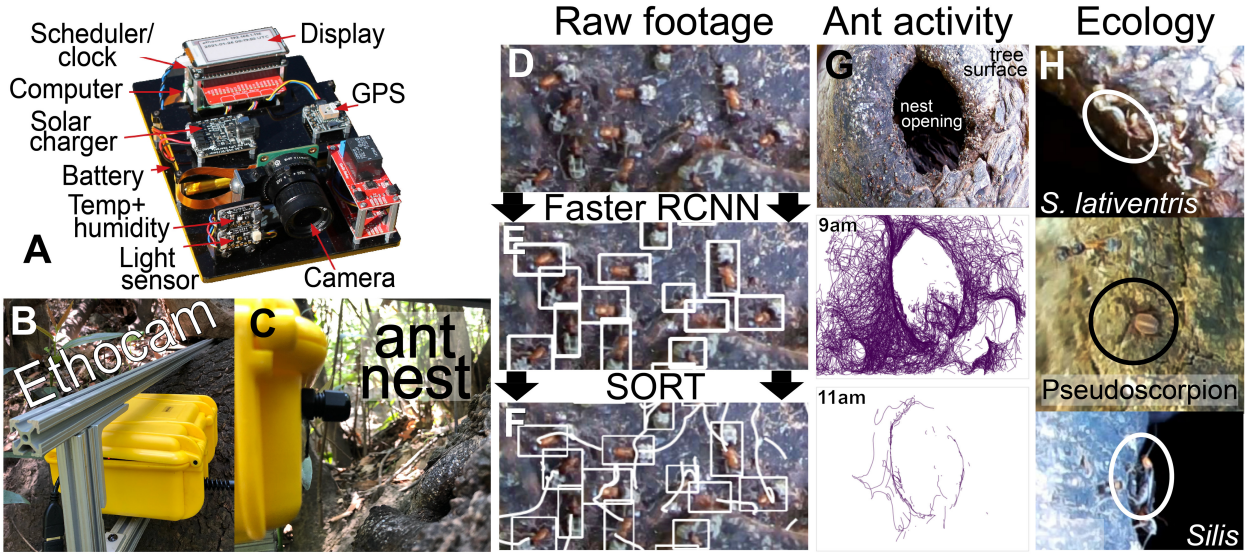

Figure 1. (A). Ethocam setup with various components annotated. (B). Ethocam inside waterproof enclosure attached to custom made 80-20 rail setup for imaging ant nest. (C). Ethocam positioned to be approximately 10 cm away from the surface of the nest. (D-F). Schematic of our machine vision pipeline. (D) Raw data collected from an ant nest was run through Faster R-CNN for detection. (E) The resulting detections were passed through a tracker (F) to generate ant trajectories. (G) Overlaid ant trajectories over the course of 2 minute videos at different time points obtained by our ant tracker (modified SORT) shows ant activity. (H) Examples of three other arthropods captured in raw camera trap footage illustrating ant nests as a biodiversity hub.

speculated to form massive supercolonies spreading multiple kilometers, and may maintain nest locations in trees for several years [33]. Species of the genus *Liometopum*, including *L. occidentale*, are also known to engage in a diversity of interspecies relationships, including trophic mutualisms with hemipteran bugs. Colonies of *L. occidentale* ant are also targeted by several species of socially parasitic "myrmecophile" beetles of the family Staphylinidae (rove beetles) [12][23]. Previous reports suggest that activity patterns of *Liometopum* are equally diurnal and nocturnal [28] with temperature determining activity level. Here, we leverage *L. occidentale*'s stable nest locations, large colony size, rich inter-species associations and vital ecological role to pioneer a field-based ant tracking method.

Our system detects and tracks the activity of *L. occidentale* ants at colony locations using modern computer vision techniques and is scalable to habitat-wide monitoring. In this proof-of-concept study, we use an in-house designed and built camera trap to record high-quality videos of an ant nest entrance for 2 minutes every hour throughout the day. Using a combination of traditional image processing techniques for weak supervision combined with a small set of human annotations, we train and evaluate a computer vision pipeline consisting of Faster R-CNN [27] for object detection and a modified version of SORT [2] for multi-object tracking. We successfully track ant trajectories over the day and obtain a measure of ant count. We evaluate the performance of our method both in and out of distribution and show that our method closely follows ground truth ant count trends. We also show that our method predicts ant counts as well as single-frame human annotations (that is annotations without using temporal information). Our methods are sufficient to extract novel scientific insight on the behavior of *L. occidentale* in their natural habitat. In particular, we demonstrate that our method generates robust estimates of ant counts in the field with small amounts of initial training data.

1.1. Related work

Detection. Object detection seeks to localize objects of a certain category or set of categories in images, and is a highly-studied challenge in the computer vision community [13]. In this work, we rely on the well-established Faster R-CNN [27] two-stage object detection architecture with a ResNet50 backbone [11]. Our system is modular, allowing for drop-in replacement of newer architectures [8, 18, 29] as-needed.

Multi-Object Tracking. Multi-object tracking is a notoriously difficult problem, requiring robust detections in often crowded scenes, management of occlusions, and re-identification of individuals returning to view [15, 20]. Canonical multi-object tracking challenges use pedestrian data [7], and algorithmic approaches range from simple IOU overlap in subsequent frames to assign ID [2] to methods using object appearance through time [32] to learning

whole graph structures to generate tracks [4]. Methods using feature information require objects to have visually distinguishable appearance, which make their value tentative for groups of genetically similar sister ants which look nearly identical. Hence, we employed the simple and well-performing SORT algorithm [2] for our ant tracking task.

Tracking Social Insects. Previous approaches to tracking social insects in a lab setting are predominantly marker-based [22][6], which is not feasible for natural colonies with tens of thousands of individuals. [14] developed an ant detector plus tracker using framewise detections with Mask-RCNN, and subsequent tracking by minimizing an optimal transport cost function between consecutive frames. The tracker maintains identities by minimizing the cost, based on spatial distance and appearance, of each ant in frame K with all other ants in frame K+1. This work is designed for tracking foraging paths of carpenter ants (*Camponotus rufipes*) at night, under consistent IR light. These methods are insufficient to handle the fluctuation in ambient lighting and massive changes in ant density and trajectory direction seen at *L. occidentale* nests. [5] propose a method for tracking ants at nests. They use a ResNet to obtain appearance features and combine appearance and motion features to obtain the final trajectories. Instead of collecting their own data they use short stock videos and images obtained online from random ant nests, and their approach relies on a large number of hand-labeled training examples. In contrast, we show accurate ant counts even in out-of-distribution videos with fewer than a thousand hand labeled in-distribution training images using transfer learning for detection and using tracking to eliminate false positives. [25] train a U-Net architecture to produce a density map given an image containing a cluster of monarch butterflies, and subsequently obtain a count by integrating over the density map. This approach is not suitable for our data as it is sometimes difficult to see individual ants in a single frame due to motion blur. Our separate detection + tracking approach allows us to use temporal information to resolve such cases.

2. Methods

2.1. Data collection

Data were recorded using an in-house camera trap dubbed 'Ethocam.' The Ethocam is a cheap (∼$250) and open-sourced setup consisting of a 12.3 MP raspberry pi HQ camera, temperature, humidity and light level sensors, GPS, power management unit WittiPi with a 10000mAh rechargeable battery, solar charger, relays for external light control and an e-ink persistent display. The source code and design files are available at https://github.com/willdickson/ethocam. The Ethocam schematic is shown in Figure 1A. To optimize video for ant detection, we tested a range of camera parameters and external lighting, first in the lab, capturing images of ants on bark in a plastic tub (we will refer to this data subset as DATA-LAB), then tuned for lighting variations by placing the tub and camera outside in natural light (we will refer to this data subset as DATA-NL for "natural light"), and made final adjustments at the ant nest field site (we will refer to this data subset as DATA-NEST). The final set of parameters used by the camera were a frame rate of 30 FPS, a bitrate of 20000000, auto white balance, a dynamic range compression set to high, auto exposure and ifx set to denoise. The camera was placed roughly 10cm from the ant nest entrance and was supported with a custom 80-20 aluminum rail. Ultimately, there was some motion blur when the ants were moving quickly, but the proximity of the camera to the ants allowed us to successfully detect and track the ants. The ant nest we used for our proof-of-concept study was in a hollowed-out bay tree, slightly off the path at Chaney trail in the Angeles National Forest. We collected 2 minute long videos every hour for 41 hours from the ant nest (13 hours from May 13th, 2021, and 14 hours each day on July 17th and 19th, 2021). As we did not use any external lighting in this pilot study, only footage between 7am - 7pm from May, and 6:20am - 7:20pm from July, were usable. Collecting nighttime data remains an area of future work.

2.2. Detection

Generating Weak Supervision. To reduce the need for hand-labeling, we use traditional computer vision techniques to provide initial weak supervision. We use color channel information to locate the ants' distinctive orange-brownish thorax. For each frame, we first perform a histogram equalization of the value channel in the HSV color space, then convert back to the RGB. We then create a mask by subtracting the red and blue channels and manually threshold to handle lighting variation using data collected at 3 different timepoints within the DATA-NL subset - 10:30am, 1pm and 7pm. We repeat the same for the red and green channels, and take the intersection of the two. We then use contour detection on this mask and OpenCV's [3] boundingrect function to draw bounding boxes around the ants. This method fails when the thorax is not in view (*e.g.* when the ants are in a crevice in the bark). To address this, we add an adaptive background subtraction using OpenCV's MOG detector, and clean the results using morphological opening and closing. Bounding boxes are again obtained via boundingrect. We ensemble the two approaches to increase our recall and remove overlapping boxes (IoU ≥ 0.3), giving preference to the boxes extracted using the RGB method as it qualitatively provided more precise localization.

The detections from this simple method work well for DATA-NL, where there is little background texture or light-

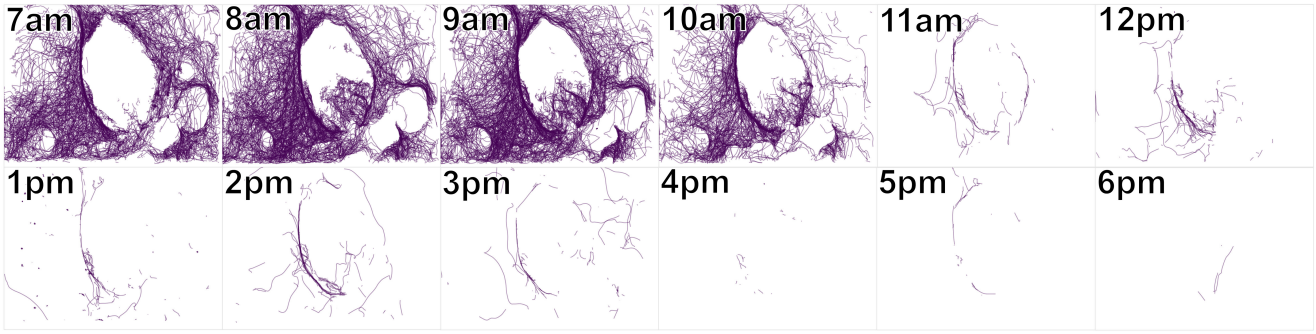

Figure 2. Activity maps produced at different times of day based on the output of the tracker.

ing variation. We found qualitatively that the performance declines significantly when moving to DATA-NEST for a few reasons: 1. The ant nest entrance was more complex than the bark from the DATA-NL set (the ant nest tree was highly textured and also redder in color causing the RGB method to perform poorly), 2. there were moving rays of sunlight at different locations (because of leaves in the canopy above) and 3. wind led to some slight camera shake which interfered with background subtraction.

To build a more robust, generalizable model to handle diverse data at the nest site, we use our traditional computer vision approach to extract approximate ant locations from DATA-NL for weak supervision of a Faster R-CNN ant detection model. We extract weakly-supervised bounding boxes from 10,500 frames across DATA-NL, and use a 75-25 random train-test split (note that here the test data is also weakly-labeled). We train a Faster R-CNN with a Resnet-50 [27] backbone, starting with Imagenet weights, on this data for 5 epochs using the Detectron2 library [34].

Fine-tuning on DATA-NEST. The DATA-NEST set contains 2 minutes of video collected every hour between 7am-7pm on May 13th and between 6:20am-7:20pm on July 17th and July 19th, 2021. We manually annotated 507 frames - 200 "dense frames" and 307 "sparse frames" where dense frames have a large number of ants (videos between 7am-10am) and sparse frames have few ants (1-5 ants, videos between 11am-7pm). All hand annotations were generated with data from May 13th, hence we term the May subset of the data as DATA-NEST-ID (ID for in-distribution), and the set of videos entirely unseen by our model as DATA-NEST-OOD (OOD for out-of-distribution). We make this distinction in order to ensure we are evaluating our system as it is intended to be used, where the evaluation data will shift from the training distribution both visually and in ant density over time (real-world applications almost always encounter distribution shift in deployment, a known challenge for automated monitoring approaches [1, 17]). DATA-NEST-ID annotations are shuffled and randomly divided into a train, val and test set using a 60-20-20 split. This resulted in 119, 43, 38 dense frames and 185, 58, 64 sparse frames in the training, validation and test sets respectively. Starting from our weakly-supervised DATA-NL-trained Faster R-CNN model, we fine-tune on this small manually labeled dataset to transfer our model to DATA-NEST. We do not see training and validation loss diverge, indicating either little-to-no overfitting or, more likely, a large amount of similarity between the randomly-subsampled training and validation data. Our operating point for use in downstream tracking is selected using precision-recall curves on the DATA-NEST-ID test set.

2.3. Tracking

To estimate ant counts across each video from frame-wise ant detections, we implement a tracking module - a modified version of the multi-object tracker SORT [2]. SORT uses a Kalman filter based on linear velocity changes and Hungarian algorithm for ID assignment. We generate new detections by linearly interpolating bounding box coordinates over detection gaps in tracks generated by SORT. This SORT+interpolation adaptation allows us to use video information to reduce false negatives and improve detection, while the ID maintained across a track allows us to capture trajectory information for each ant. In some lighting conditions, particularly in DATA-NEST-OOD, parts of the bark of tree roughly resemble the shape of an ant, resulting in repeated, stationary false positives. In order to address this, a track is considered a stationary false positive if it moves less than a threshold distance of 10 pixels (frames are 640 x 480) for more than 100 frames. To account for ants that stop moving temporarily, we check if the furthest distance moved by the entire track is less than 20 pixels.

2.4. Statistical analysis

We perform a regression analysis in order to study the effects of the recorded environmental variables on ant count. We fit a linear regression model using the mean ant counts, calculated per video, as the dependent variable, and temperature, humidity, light level, time and day of collection (1, 2 or 3) as the independent variables. We could not use

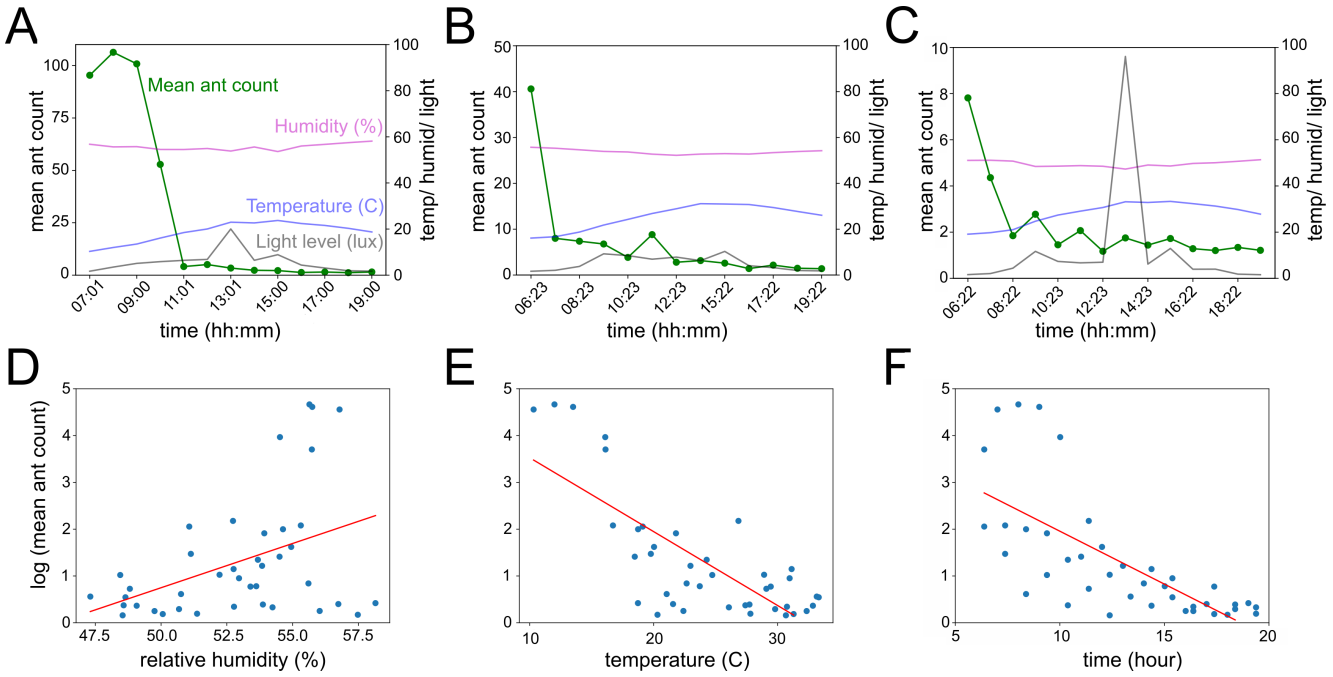

Figure 3. (A). Mean ant count along with recorded environmental factors for data collected on 13th May 2021. (B). Mean ant count along with recorded environmental factors for data collected on 17th July 2021. (C). Mean ant count along with recorded environmental factors for data collected on 19th July 2021. (D). Log of mean ant count plotted against humidity: log(mean ant count)=0.19×(humidity)-8.71 (R^2=0.17, p<0.01). (E). Log of mean ant count plotted against temperature: log(mean ant count)=-0.16×(temperature)+5.10 (R^2=0.58, p<0.001). (F). Log of mean ant count plotted against time: log(mean ant count)=-0.23×(time)+4.22 (R^2=0.49, p<0.001).

a mixed effects model to account for day of collection as a random effect as we only have 3 random effect levels (3 days) as opposed to the minimum of 5-6 required for statistically significant variance [10], and hence use day of collection as another independent variable.

3. Results

3.1. Detection

To evaluate the performance of our preliminary, motion- and color-based method we manually counted the number of successfully detected ants and misses across 100 randomly selected frames from the DATA-NL set. Out of 970 ant occurrences, this traditional approach missed 67 ants and only had 3 false positives resulting in a recall of 0.93.

Our weakly-supervised model trained using manually-thresholded motion- and color- supervision on DATA-NL is evaluated vs. the weak labels on a held-out test set. The precision-recall curve has a high AUC of 0.91 with a maximum precision and recall corresponding to a threshold of 0.65.

Table 1 shows the AP (11 point interpolation method with 0.5 IoU) calculated using the weakly supervised model (trained on DATA-NL) and the model fine tuned on DATA-NEST-ID, on both the DATA-NL and DATA-NEST-ID datasets. While we see good results of the weakly supervised model on DATA-NL (AP=0.89) and the fine tuned model on DATA-NEST-ID (AP=0.79), we see poor generalizations of the models across the two datasets.

We visually analyze results and failure modes in Figure 4C. We find that most false negatives are quite small and difficult to distinguish without a motion signal. The annotations were done on higher resolution 1920x1080 frames within a video (model input resolution is 640x480), providing both motion signal and higher resolution for human annotation. We also see failures related to challenging lighting conditions, particularly as it gets dark later in the evening.

3.2. Tracking

In order to robustly assess the performance of our tracker, we manually annotated the trajectories of all ants in the first 500 frames of one dense high-activity video (9am video from May 13th). The first 196 of these 500 frames overlap with the densely-annotated frames from the DATA-NEST-ID detection train/val/test splits. This represents 64325 hand-labeled detections corresponding to tracks for 315 individual ants. Linearly interpolating to fill in missing detections within a SORT track improves the tracking metrics across the board. With our SORT+interpolate we achieve 26% (82/315) mostly tracked (ID label maintained for 80% of a given individual's track), and only have 14% (43/315) mostly lost tracks (individual tracked for less than 20% of

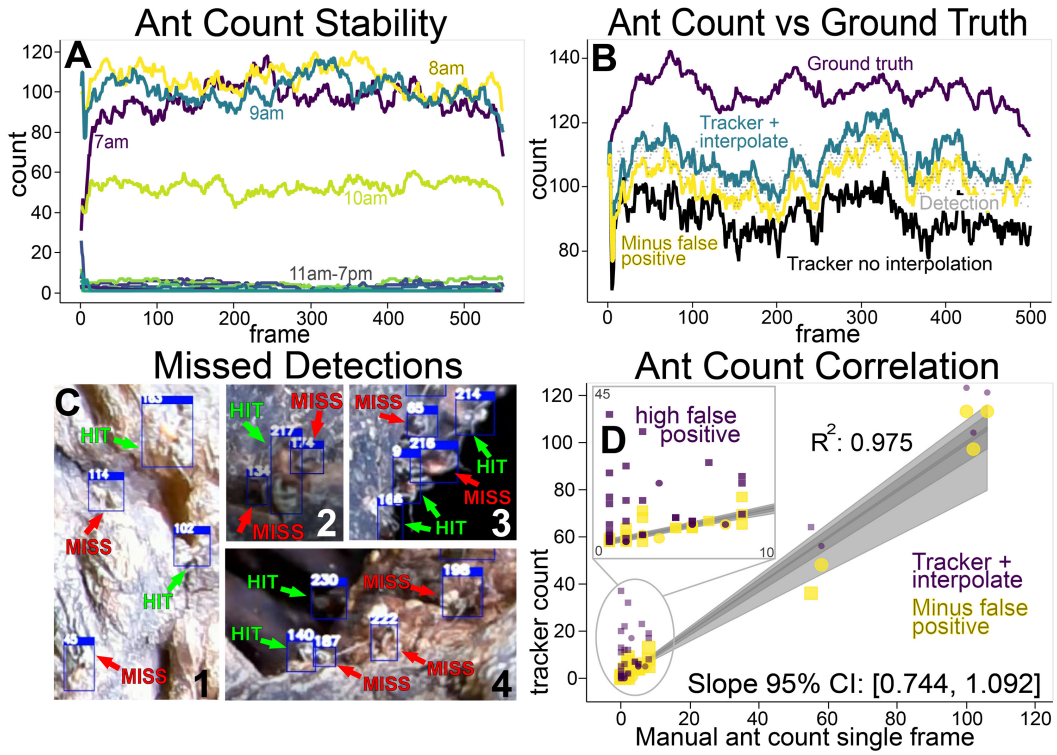

Figure 4. (A) Stability of ant count throughout two minute videos generated via our detector+tracker approach. (B) Comparison of ant counts in ground truth data (purple) versus raw detection counts from our fine tuned model (grey dots), plain SORT output (black), SORT+interpolation (green), or SORT+interpolation minus false positives (filtering)(yellow). Our approach consistently under-counts ants due to several difficult annotation types. (C) Examples of missed detections, likely due to 1) overexposure, 2) ant mostly occluded, 3) previously occluded ant freshly entering frame and 4) bark providing effective camouflage. (D) Human generated ant counts vs SORT+interpolation counts (one frame per video). Squares are from the DATA-NEST-OOD set, circles from DATA-NEST-ID. Yellow dots are after filtering out false positives, purple before filtering. Human ant count was lower than from the ground truth data in (B) since only a single frame without time series of ant movement was used for count. Our algorithm (which has temporal information) performs as well as a human without the timeseries, as indicated by a slope which is not significantly different than 1 (non-parametric bootstrapping on slope was used to generate a confidence interval). DATA-NEST-OOD had a large number of false positives from lighting conditions not similar to the training set (D upper left inset, purple squares), which we filtered out based on low movement to help generalization.

their ground truth track). We have 421 ID swaps across the evaluation set, which contains 315 ground truth IDs. We also use the HOTA metric to evaluate our methods, which correlates well with human perception of tracking success [19]. This metric provides a single number that summarizes how well the generated trajectory tracks align while also docking for failed detections. The HOTA score for our tracking approach + interpolation is 40.4, whereas SORT without interpolation scores 38.1. These metrics indicate that our tracker is able to maintain a large number of nearly full-length trajectories and performs well on the DATA-NEST-ID set.

We find that the filtering approach we propose to reduce stationary false positives in DATA-NEST-OOD reduces the HOTA score on DATA-NEST-ID, mostly due to the loss of ants that remain still through large portions of the video (Table 2). However, the filtering significantly improves our ability to generalize to DATA-NEST-OOD, which had a larger number of false positives since it is not represented in the training data. The increase in counting performance in DATA-NEST-OOD seems to be worth the trade-off in tracking score (see Section 3.3).

In Figure 2 we visualize ant trajectories for DATA-NEST-ID. We see a slight increase in activity from 7am to 8am followed by a slight decrease up to 10am and then a massive decrease from 10am - 11am. The activity remains very low for the rest of the day.

3.3. Counting

We use the number of unique framewise track IDs to obtain ant counts per-frame in DATA-NEST-ID. We compare the counts from our algorithm to the count from the manual track annotations, when available. In addition, to investigate counting performance at different time points with different

Model	Dataset	AP
Weakly supervised	DATA-NL	0.89
Weakly supervised	DATA-NEST-ID	0.05
Fine tuned	DATA-NL	0.28
Fine tuned	DATA-NEST-ID	0.76

Table 1. Detection results on different datasets.

ambient lighting, we manually counted the number of ants in one frame per video from DATA-NEST-ID and DATA-NEST-OOD. Figure 3 shows predicted mean ant count vs time of day for each of the 3 days of field data, which gives us an estimate of stability, along with corresponding temperature, humidity and light level.

Figure 4B shows the evaluation of the counts obtained from our pipeline against the counts obtained by manually annotating the full trajectories from the first 500 frames from a single DATA-NEST-ID video (the 9am from May 13th). The first 196 of these 500 frames are in the dataset used for the train-val-test splits whereas the remaining frames were not seen by the network. We see that the ant count generated by detector + tracker system closely follows the trend in the ground truth, thus serving as a good measure of ant activity. Additionally, we note that, compared to raw detections which fluctuate by 10-20 ants over just a few frames (Figure 4B) in dense videos; our tracker outputs much more consistent counts. Though we lose detections with the raw SORT algorithm, the interpolation step recovers misses and gives consistent counts that outperform the plain detections. Visual analysis (Figure 4C) reveals that the missed detections are difficult cases - partial occlusions, ants freshly entering the frame, etc.

To verify the performance of the system at different time points and more varied conditions, we compared counts from our tracker with human counts on a single frame from every video in DATA-NEST-ID and DATA-NEST-OOD (Figure 4D, Table 2). We see a strong correlation with the human count on DATA-NEST-ID, with a slope not significantly different than one, indicating human-level performance, and a high R^2 value (In-distribution results Table 2). Our raw tracker approach struggles to generalize to the out-of-distribution DATA-NEST-OOD, where we had an uptick in false positive detections (Figure 4D, upper inset, purple points, Out-of-distribution results in Table 2) in frames with lighting not similar to our training data. Since it appears that false positives rather than false negatives hamper generalization, we apply a simple filter to remove detections that scarcely moved through the video. This drastically improves generalization, increasing the R^2 for the DATA-NEST-OOD count regressions from 0.54 to 0.9 and tightening the confidence interval on the slopes (Table 2, Out-of-distribution results). This illustrates that incorporating a simple filter into our pipeline largely improves the generalization issue with our network and allows for robust counts on out-of-distribution data, though quality control will be undertaken on all future data to catch model/data drift.

The discrepancy in performance on the ground truth tracking (where we under-count) compared to human counts (where we match human accuracy) may be due to the different annotation techniques. While annotating tracks we used temporal clues from previous and subsequent frames to find partially occluded or blurred ants whereas the counts were obtained looking at only one frame. Though our tracker includes temporal information, it is difficult for our system to pick out every ant in a dense cluster and we lose the tail of tracks post-occlusion. Although there is room for improvement, our approach succeeds in obtaining near-human measures of ant count with a very small amount of manually-annotated data (507 frames).

3.4. Data Analysis

In order to assess the impacts of different environmental factors on ant count, we fit a linear regression model with temperature, humidity, light level, time and day as the independent variables. We obtain an adjusted R^2 value of 0.42 ($p<0.001$) and find that temperature is the only significant factor ($p<0.05$, df=1, t=-2.60). We however can not make any claims on the effect of temperature on ant count as a variance inflation factor analysis shows serious (>4) multicollinearity [9] (VIF(temperature) = 6.48, VIF(humidity) = 9.22, VIF(day) = 5.5). This makes sense due to the small dataset size. In order to fully tease apart the effects of temperature and circadian rhythm (time), we need to collect more data over varying conditions. We will address this in future work and leave this manuscript as proof-of-concept of our methodology. Figures 3D, 3E and 3F show the log of mean ant counts from all three days of collection, plotted against humidity, temperature and time respectively.

4. Discussion

Monitoring the behavior of ecologically dominant ant species like *Liometopum occidentale* in the wild promises rich insights into biological communities. Using computer vision approaches, we develop a pipeline to capture data, detect, track and then analyse behavior at the ant nest. Automating these steps using computer vision and GPU accelerated computing, as opposed to manually observing activity, makes it possible to not only analyze extensive data (weeks to years) but also allows robust, quantitative metrics like accurate ant count and ant trajectory maps. Such data is prohibitively laborious to obtain with human annotation. Our data acquisition strategy allows us to collect a wealth of useful environmental factors such as temperature, humidity and ambient light level at the collection site.

With our proof-of-concept dataset we report a number of interesting observations which underscore the potential of our system: **i.** We observe a massive change in overall ant

Tracker	In-distrib. HOTA	In-distrib. Count R^2	In-distrib. Count slope 95% CI	Out-of-distrib. Count R^2	Out-of-distrib. Count slope 95% CI
SORT	38.1	0.99	0.71-0.95	0.58	0.11-1.0
SORT+Interp	40.4	0.98	0.96-1.2	0.54	-0.14-1.5
SORT+Interp+Postproc.	34.9	0.99	0.84-1.1	0.9	0.63-1.2

Table 2. HOTA and detector ant count correlation to hand labeled counts for our tracking approaches.

activity over the course of the day and quantify ant counts between 7am-7pm in May and between 6:20am-7:20pm in July. For data collected on May 13th, we qualitatively see that ant activity starts to increase again at 7:30pm, suggesting imaging nighttime activity using infrared lighting may offer further insights which we will explore in future work. We see significantly lower ant counts in the data collected in July as compared to May. This may be due to seasonal and temperature changes, but further analysis and monitoring over longer time horizons is needed. **ii.** We also observe ants performing an excavation behavior, bringing pieces of debris from inside to the edge of the ant nest to discard. **iii.** In addition to the ants themselves, we observe other athropods (Figure 1H): two instances of the symbiotic beetle *Sceptobius lativentris*, which steals ant pheromones via grooming to move around freely inside the ant nest; a pseudoscorpion, which may also be a symbiont of these ants; and a member of the non-symbiotic beetle genus *Silis*, which, in contrast to *Sceptobius* was attacked by worker ants and dragged into the nest (*Liometopum occidentale* are known to be omnivores). It is exciting that we see such diverse behaviors and interactions in our proof-of-concept dataset, a total video time of 82 minutes over the course of 3 days, demonstrating the value of imaging animals in the wild in an undisturbed manner. Further data collection with longer video times over larger time scales along with data collection at night will provide further, previously-inaccessible information for this little-studied ant species and its associated arthropod community. We generated numerous ant annotations with a semi-automated approach to train a network that performed well for DATA-NL. We then leveraged this network to obtain initialization weights to train a model to detect ants in DATA-NEST. We achieve good performance on DATA-NEST with a small number of manually annotated frames (507). Since the detector relies only on information in a single frame, we were further able to improve the detection performance and collect video-level counts using our modified SORT tracker. The tracker uses a Kalman filter constant velocity model along with interpolation to recover ant detections missed by the network. The ID assignment using the Hungarian algorithm enables us to produce trajectories for individual ants, giving activity maps. Although we see from Figure 4B that our ant count undershoots human counts given a full video as context, our counts so approach the accuracy of a human given only single frames for annotation (Figure 4D). Qualitatively, raw videos clearly demonstrates that ant activity drops rapidly over the course of the day, and our tracking shows the same trend (Figure 2, Figure 3). Using a tracker stabilizes the ant counts (4A) during the videos, as seen in Figure 4B, where the raw detections change in count by 10-20 ants within a few frames.

Although there is room to improve the accuracy of our models, our system nevertheless provides an indication on the key focal times for future Ethocam acquisition, i.e. times of high activity. We captured multiple instances of various other species interacting with the ants in our small data set. In future, we will investigate species classification approaches to automatically recognize symbionts and other arthropods, although we recognize that the large imbalance between ant and symbiont sightings will prove a challenge. If non-ant arthropods occur in-frame frequently (as our pilot data imply), and can be accurately detected, further work could include automated investigation of inter-species interactions involving *L. occidentale*.

In summary, we have successfully developed a data collection paradigm and computer vision methodology to extract quantitative activity information from natural ant colonies. We successfully tuned camera parameters and positioning to capture high quality ant nest data; we then generated detections on a per-frame basis with semi-automated and manually generated annotations. Subsequently, we tracked ants to improve detection, obtained trajectories and counted ants, and validated these counts and tracker performance both in and out of distribution. We found that *L. occidentale* activity drops rapidly over the course of the day and picks up again at night. We observed interesting trends in ant activity potentially resulting from different environmental factors and we look forward to collecting more data over longer time horizons to explore these further. This pilot study illustrates the potential of our system for further applications. Our low-cost, open source hardware and software together show promise for quantitatively observing the networks of interactions in natural populations. Quantitative measures will help elucidate complex, environmentally influenced animal behaviors and provide insight on how invasive species and our changing climate affect the rich, native ant biodiversity.

References

[1] Sara Beery, Grant Van Horn, and Pietro Perona. Recognition in terra incognita. In *Proceedings of the European conference on computer vision (ECCV)*, pages 456–473, 2018. 4

[2] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3464–3468, 2016. 2, 3, 4

[3] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 3

[4] Guillem Brasó and Laura Leal-Taixé. Learning a neural solver for multiple object tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6247–6257, 2020. 3

[5] Xiaoyan Cao, Shihui Guo, Juncong Lin, Wenshu Zhang, and Minghong Liao. Online tracking of ants based on deep association metrics: method, dataset and evaluation. *Pattern Recognition*, 103:107233, 2020. 3

[6] James D Crall, Nick Gravish, Andrew M Mountcastle, and Stacey A Combes. Beetag: a low-cost, image-based tracking system for the study of animal behavior and locomotion. *PloS one*, 10(9):e0136487, 2015. 3

[7] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixé. Mot20: A benchmark for multi object tracking in crowded scenes. *arXiv:2003.09003[cs]*, 2020. arXiv: 2003.09003. 2

[8] Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian. Centernet: Keypoint triplets for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6569–6578, 2019. 2

[9] John Fox. *Applied Regression Analysis and Generalized Linear Models*. Los Angeles: Sage Publications, 2016. 7

[10] Xavier A Harrison, Lynda Donaldson, Maria Eugenia Correa-Cano, Julian Evans, David N Fisher, Cecily ED Goodwin, Beth S Robinson, David J Hodgson, and Richard Inger. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6:e4794, 2018. 5

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 2

[12] Rochelle Hoey-Chamberlain, Michael K Rust, and John H Klotz. A review of the biology, ecology and behavior of velvety tree ants of north america. *Sociobiology*, 60(1):1–10, 2013. 1, 2

[13] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, et al. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7310–7311, 2017. 2

[14] Natalie Imirzian, Yizhe Zhang, Christoph Kurze, Raquel G Loreto, Danny Z Chen, and David P Hughes. Automated tracking and analysis of ant trajectories shows variation in forager exploration. *Scientific reports*, 9(1):1–10, 2019. 3

[15] Sara Bouraya Jr. and Abdessamad Belangour. Multi object tracking: a survey. In *Thirteenth International Conference on Digital Image Processing (ICDIP 2021)*, pages 142 – 152. International Society for Optics and Photonics, SPIE, 2021. 2

[16] Douglas N Kamaru, Todd M Palmer, Corinna Riginos, Adam T Ford, Jayne Belnap, Robert M Chira, John M Githaiga, Benard C Gituku, Brandon R Hays, Cyrus M Kavwele, et al. Disruption of an ant-plant mutualism shapes interactions between lions and their primary prey. *Science*, 383(6681):433–438, 2024. 1

[17] Pang Wei Koh, Shiori Sagawa, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning*, pages 5637–5664. PMLR, 2021. 4

[18] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 2

[19] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixé, and Bastian Leibe. Hota: A higher order metric for evaluating multi-object tracking. *International journal of computer vision*, 129(2): 548–578, 2021. 6

[20] Wenhan Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, Xiaowei Zhao, and Tae-Kyun Kim. Multiple object tracking: A literature review. *arXiv preprint arXiv:1409.7618*, 2014. 2

[21] Therese A Markow. Reproductive behavior of drosophila melanogaster and d. nigrospiracula in the field and in the laboratory. *Journal of Comparative psychology*, 102(2):169, 1988. 1

[22] Danielle P Mersch, Alessandro Crespi, and Laurent Keller. Tracking individuals shows spatial fidelity is a key regulator of ant social organization. *Science*, 340(6136):1090–1093, 2013. 3

[23] Joseph Parker. Myrmecophily in beetles (coleoptera): evolutionary patterns and biological mechanisms. *Myrmecological news*, 22:65–108, 2016. 2

[24] Joseph Parker and Daniel JC Kronauer. How ants shape biodiversity. *Current Biology*, 31(19):R1208–R1214, 2021. 1

[25] Shruti Patel, Amogh Kulkari, Ayan Mukhopadhyay, Karuna Gujar, and Jaap de Roode. Using deep learning to count monarch butterflies in dense clusters. *bioRxiv*, 2021. 3

[26] David J Pritchard, T Andrew Hurly, Maria C Tello-Ramos, and Susan D Healy. Why study cognition in the wild (and how to test it)? *Journal of the Experimental Analysis of Behavior*, 105(1):41–55, 2016. 1

[27] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv:1506.01497*, 2015. 2, 4

[28] Harlow Shapley. Thermokinetics of liometopum apiculatum mayr. *Proceedings of the National Academy of Sciences of the United States of America*, 6(4):204, 1920. 2

[29] Mingxing Tan, Ruoming Pang, and Quoc V. Le. Efficientdet: Scalable and efficient object detection, 2020. 2

[30] Neil D Tsutsui. Wild vs. lab box 1.2 understanding the nature of ant cognition by studying ant cognition. *Field and Laboratory Methods in Animal Cognition: A Comparative Guide*, page 23, 2018. 1

[31] Emma C Underwood and Brian L Fisher. The role of ants in conservation monitoring: if, when, and how. *Biological conservation*, 132(2):166–182, 2006. 1

[32] Gaoang Wang, Yizhou Wang, Haotian Zhang, Renshu Gu, and Jenq-Neng Hwang. Exploit the connectivity: Multi-object tracking with trackletnet. In *Proceedings of the 27th ACM International Conference on Multimedia*, page 482–490, New York, NY, USA, 2019. Association for Computing Machinery. 2

[33] Thea B Wang, Ankur Patel, Francis Vu, Peter Nonacs, et al. Natural history observations on the velvety tree ant (liometopum occidentale): unicoloniality and mating flights. *Sociobiology*, 55(3):787–794, 2010. 2

[34] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019. 4