Artificial Life and Robotics (2022) 27:393–400
https://doi.org/10.1007/s10015-022-00753-y

ORIGINAL ARTICLE

Beyond tracking: using deep learning to discover novel interactions in biological swarms

Taeyeong Choi[1] · Benjamin Pyenson[2] · Juergen Liebig[2] · Theodore P. Pavlic[3]

Received: 31 August 2021 / Accepted: 9 February 2022 / Published online: 23 March 2022

Abstract

Most deep-learning frameworks for understanding biological swarms are designed to fit perceptive models of group behavior to individual-level data (e.g., spatial coordinates of identified features of individuals) that have been separately gathered from video observations. Despite considerable advances in automated tracking, these methods are still very expensive or unreliable when tracking large numbers of animals simultaneously. Moreover, this approach assumes that the human-chosen features include sufficient features to explain important patterns in collective behavior. To address these issues, we propose training deep network models to predict system-level states directly from generic graphical features from the entire view, which can be relatively inexpensive to gather in a completely automated fashion. Because the resulting predictive models are not based on human-understood predictors, we use explanatory modules (e.g., Grad-CAM) that combine information hidden in the latent variables of the deep-network model with the video data itself to communicate to a human observer which aspects of observed individual behaviors are most informative in predicting group behavior. This represents an example of augmented intelligence in behavioral ecology—knowledge co-creation in a human–AI team. As proof of concept, we utilize a 20-day video recording of a colony of over 50 *Harpegnathos saltator* ants to showcase that, without any individual annotations provided, a trained model can generate an "importance map" across the video frames to highlight regions of important behaviors, such as dueling (which the AI has no a priori knowledge of), that play a role in the resolution of reproductive-hierarchy re-formation. Based on the empirical results, we also discuss the potential use and current challenges to further develop the proposed framework as a tool to discover behaviors that have not yet been considered crucial to understand complex social dynamics within biological collectives.

Keywords Deep learning in behavioral ecology · Swarm behavior · Explainable AI · Augmented intelligence · Knowledge co-creation

1 Introduction

Deep Convolutional Neural Networks (DCNNs) have been widely adopted as the primary backbone of data-driven frameworks to solve complex problems in computer vision including object classification or detection and recognition of human actions [19, 24, 25]. The nature of their

This work was presented in part at the joint symposium with the 15th International Symposium on Distributed Autonomous Robotic Systems 2021 and the 4th International Symposium on Swarm Behavior and Bio-Inspired Robotics 2021 (Online, June 1–4, 2021).

✉ Theodore P. Pavlic
tpavlic@asu.edu

Taeyeong Choi
tchoi@lincoln.ac.uk

Benjamin Pyenson
bpyenson@asu.edu

Juergen Liebig
jliebig@asu.edu

[1] Lincoln Institute for Agri-food Technology, University of Lincoln, Riseholme Park, Lincoln LN2 2LG, UK

[2] School of Life Sciences, Social Insect Research Group, Arizona State University, Tempe, AZ 85287, USA

[3] School of Computing and Augmented Intelligence, School of Complex Adaptive Systems, School of Sustainability, and School of Life Sciences, Social Insect Research Group, Arizona State University, Tempe, AZ 85287, USA

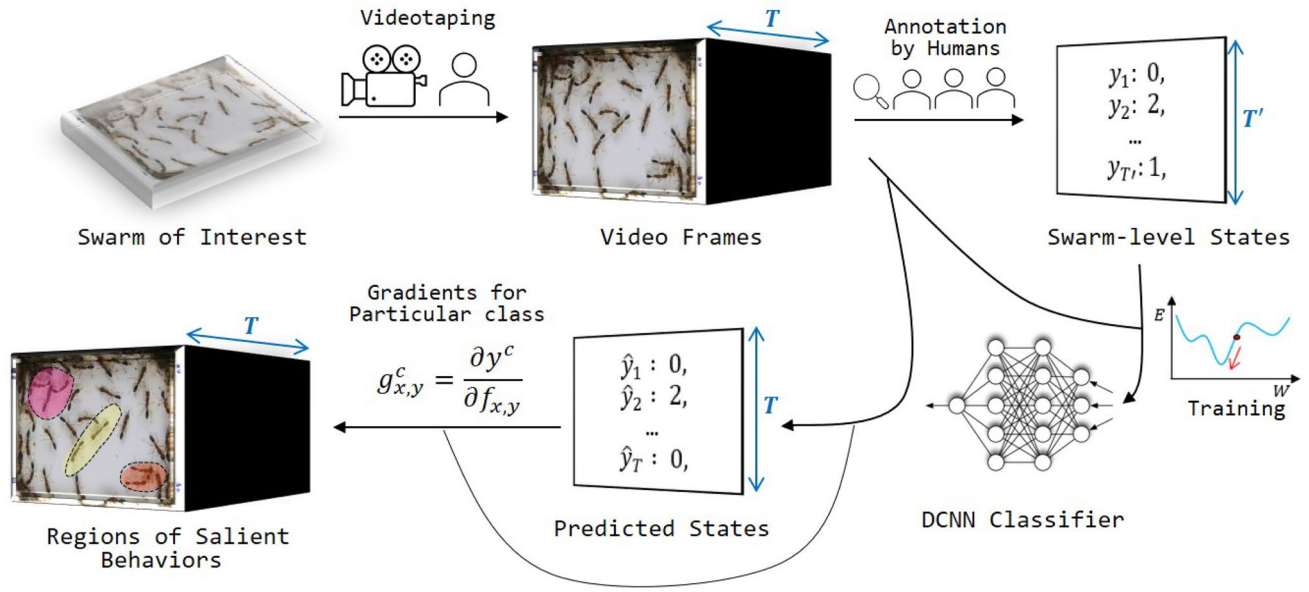

Fig. 1 Proposed usage of DCNNs, trained to predict global state of the swarm system from the entire view and later examined to reveal key local observations by utilizing the computed gradient between the learned local feature and the prediction output in the DCNN classifier

multi-layer structure has a powerful ability to automatically learn to identify key local features (e.g., edges) from raw pixels of images and combine into more meaningful concepts (e.g., pointy ears) to produce a final prediction output (e.g., dog), as the data is processed from the lowest layer through the higher ones [7]. Consequently, if the target video data contains global information of biological swarms, lower-level visual properties such as locations, motions, and interactions of the entities could automatically be identified throughout the hierarchical layers of a DCNN during the training process. However, deep learning in behavioral biology has mostly been limited to building perceptive models to localize particular body parts of each entity to generate another input to a subsequent analysis model to capture motional concepts of individuals and perform a prediction for the entire swarm based on them [1, 8, 14, 21].

There can be two significant challenges in deep-learning approaches based on first localizing human-chosen features in video data: (1) obtaining the individual feature labels can require a significant amount of human effort especially when a large group of individuals are monitored simultaneously, and (2) the choice of features relies heavily on prior knowledge of human experts in the biological system. To address these issues, as visualized in Fig. 1, we here suggest training the deep-network models to predict system-level states directly from generic graphical features from the entire view, which can be relatively inexpensive to gather, and examine the salient behavioral regularities discovered in the trained intermediate layers by using gradient-based explanation modules (e.g., Grad-CAM [23]). In other words, our proposal is to make more use of the aforementioned potential of DCNN to automatically discover fine-grained, individual-level motional patterns highly associated with macroscopic swarm properties so that the predictive model can later be queried about what these patterns are without being constrained by prior knowledge from human experts.

In fact, such explanation-enabled designs of deep neural network have been actively investigated in a number of machine learning fields, such as computer vision [11, 23], natural language processing [10, 18], and medical diagnosis [9, 15], so as to ensure the credibility of the trained model by visualizing the fine-grained features in the space of pixels or words. To the best of our knowledge, however, we are the first to utilize a similar functionality in biological system to map the abstract output of trained deep classifiers to the behavioral level of individual entities. Furthermore, our aim is not only at increasing transparency of the predictive model but also at realizing the knowledge co-creation process for scientific discovery in which human observers can learn from the visual explanations the behavioral factors to the animal collective.

Specifically, in this paper, we propose the use of the explainable module Grad-CAM (Fig. 2) for biological research. Extending our previous work [4], we utilize a 20-day video recording of a colony of 59 *Harpegnathos saltator* ants to demonstrate that without any individual annotations provided as input, the trained classifier of social stability in ant colonies can also generate an "importance map" across video frames to selectively highlight regions of

Fig. 2 Example of Grad-CAM in which the key regions are highlighted for class "Elephant"[5]

interactions (e.g., dueling) as potentially important drivers of colony state.

The rest of this paper is organized as follows. In Sect. 2, we explore related literature and the distinction of our work. Then, we generally formalize the proposed method with DCNNs in Sect. 3. Section 4 offers an introduction of *Harpegnathos saltator* ants and the video recording procedure performed for data collection. We then present more details about actual implementation in Sects. 5, and 6 shows the experimental results with qualitative examples to support our proposed framework. Finally, in Sect. 7, we summarize our research and discuss challenges and future work for further development as a useful instrument in research of collective behavior.

2 Related work

2.1 State modeling of swarms

Precise state assessment in collective systems can allow for autonomous monitoring of complex social interactions to better inform intervention strategies. Therefore, a number of data-driven approaches have been proposed to build models of focal systems of a large group to accurately discern irregularity from available observational data. For example, Mehran et al. [13] designed a computer vision algorithm to detect otherwise cryptic cues of group-level panic in observed behaviors in a human crowd. Similar approaches have also been applied to allow individuals within a systems to achieve situational awareness and adapt their local behaviors to meet global needs. In particular, Choi et al. [2, 3] designed a mobile robot that could accurately infer remote events encountered at distal ends of its team outside of sensing range; as a result, the robot could react and move in a fashion complementary to its remote teammates, hastening the achievement group-level mission objectives.

Behavioral ecologists, however, have focused more on simulating state evolution of social systems by modelling individual interactions as mathematical, stochastic processes in response to proximal information. To be specific, self-organizing positional dynamics [6, 20], collective nest choice [17], and social hierarchy reformation [22] are all explained primarily by relatively simple equations representing local interactions among individuals. These models fit the representative data from real observations generally well, but the atomic behaviors taken into account for prediction are limited to prior knowledge that human experts can offer and incorporate into the abstract model structures.

To consider more dense information, DCNNs are often employed as part of a tracking system [1, 14] to extract spatial coordinates of each individual entity from video recordings. We claim that the DCNNs are not being used to their full potential in this approach. Instead of using the DCNNs to identify novel features for analysis, the DCNNs extract data already known by humans to likely be of importance, and those data are then further analyzed by other means. Although results from this approach are likely to be readily explainable as they make use of features already identified by human researchers, the tracking process discards significant amounts of data and potentially informative features. To tackle this issue, our proposal is to utilize DCNNs beyond tracking to independently discover salient behaviors for swarm-level states directly from generic graphical inputs, and human scientists intervene afterwards to examine the discoveries to potentially generate new knowledge of the biological system.

2.2 Use cases of explainable AI systems

In general, explainability is implemented in deep learning models to help users build trust in the predictive outcomes. As demonstrated by Selvaraju et al. [23], for instance, particularly challenging patterns from images can be identified to draw more attention of researchers to those cases for further investigation. In addition, predictive outcomes could be reconfirmed by human doctors before final decision in medical scenarios [15], in which false diagnosis can cause an irrecoverable risk to patients. Similarly, if classification results are produced along with human interpretable explanations in text, end users tend to more readily trust the model [18].

A closely related application to our work is anomaly detection [11], in which anomalous parts of industrial products are automatically localized in image data while explaining the peculiarity from normal structures seen during training. Though our model uses optical flows as input to handle behavioral features of ants, novel shapes of flows will

similarly be identified to localize the ants whose behaviors have been deemed to be anomalous. In a similar sense, the prioritization of ants could be analogous to the feature-selection process where an explanatory module reveals dominant features for prediction of a particular class [9]. Most distinctively, however, we introduce a novel application to complex biological systems, in which local interactions continuously occur over time leading the entire system to present unique social states throughout large-scale time periods.

3 Proposed framework

Rather than training on small-scale features of individuals in videos, our approach trains a DCNN to predict coarse-grained, large-scale state class c from representations of generic features from video data. Any discrete, large-scale property can be used, such as whether a crowd [13] is about to riot. We use hierarchy state $c \in \{\text{Stable}, \text{Unstable}\}$ for our example system (described in Sect. 4), with the goal of predicting the current hierarchy state based on short intervals of video of the system [4]. Our n-layer classifier consists of m two-dimensional convolutional layers $\phi_{1 \le \ell \le m}$ followed by other types $\psi_{m+1 \le \ell' \le n}$, such as recurrent or fully connected layers to ultimately produce the likelihood of each state class y^c. Convolutional layers are used as feature extractors in this architecture since each output f_{ij} at ϕ_ℓ can compactly encode the local observation in a larger region ("receptive field") at previous layers $\phi_{\ell'' < \ell}$; i.e., a change in f_{ij} can imply the amplification or decrease of the motion pattern observed in the corresponding region.

For explanation of what visual regions are most important to the predictive model, Grad-CAM [23] is employed on K two-dimensional output feature maps, each denoted as $f^k \in \mathbb{R}^{h \times w}$, at a convolutional layer ϕ_ℓ to finally calculate the "importance map" M^c over the original input for a particular state class c. In the technical aspect, ϕ_ℓ can be an arbitrary layer satisfying $\ell \in \{1, 2, \ldots, m\}$, but the layer ϕ_ℓ close to ϕ_m is typically chosen to access more abstract features with wider receptive fields than the ones available at lower layers $\phi_{\ell' < \ell}$. For brevity, we denote ϕ to be the chosen convolutional layer in the following descriptions.

To generate the importance map M^c, we first obtain the gradient g^c of the output y^c (e.g., $c = \textit{Unstable}$) with respect to each feature map f^k from ϕ, i.e., $g^c_{ij} = \partial y^c / \partial f^k_{ij}$. Therefore, $g^c_{ij} > 0$ implies that enhancing the observational pattern encoded by f^k_{ij} increases the predicted likelihood of class c—the discovered pattern is "salient" for class c—and $g^c_{ij} \le 0$ implies that the observation is considered irrelevant to the prediction of class c. Then, for each feature map f^k, Grad-CAM then uses this quantity to gain the averaged importance $a^c_k = (1/Z) \sum_i \sum_j g^c_{ij}$ (where Z is a normalization constant). Finally, the importance map M^c is computed by the rectified weighted summation of feature maps, as in:

$$M^c = \Gamma\left(\sum_k a^c_k \odot f^k \right), \tag{1}$$

where $\odot$ is the element-wise multiplication, and linear rectifier $\Gamma(a) = a$ for $a > 0$ and $\Gamma(a) = 0$ otherwise, which ensures that only the features that bring positive impacts on y^c are considered, as Selvaraju et al. [23] designed in their original Grad-CAM module [23]. In Sect. 6, we also introduce a more restrictive Γ' that gates only the top 5% values so as to strictly verify whether key behaviors are effectively highlighted with the highest level of confidence. Also, M^c can be spatially upsampled to fit the original image of a desired size for visualization purpose.

4 *Harpegnathos saltator* ant-colony testbed

Following our previous work [4], a colony of Jerdon's jumping ant, *Harpegnathos saltator*, is utilized as a testbed to validate whether our proposed framework can reveal salient behavioral patterns. A conspicuous transient "unstable" state can be induced in this system through the removal of identified egg layers ("gamergates") [16] that triggers a hierarchy reformation process among female workers, whose body lengths are typically 18–20 mm. During this process, aggressive interactions—e.g., dueling, for which two ants alternatively lunge back and forth whilst drumming their antennae (Fig. 4a) [22]—can be readily observed for several weeks until several workers activate their ovaries and start to lay eggs as part of the gamergate replacement process, causing the colony to recover its nominal stable state [12]. We apply our framework to this system by building a binary-state classifier on the stability of the colony. We use the resulting deep-network model to identify important behaviors of interest and validate whether dueling is discovered without a priori knowledge of it. Other behaviors identified by the system may then warrant further investigation by human researchers.

4.1 Video data from colonies undergoing stabilization

As shown in Fig. 3, each 20-day video was taken with an overhead camera to observe 59 *H. saltator* ants in plaster nests covered with glass. Due to a foraging chamber outside the view of the camera, not all ants are always visible, and some paralyzed crickets, their preferred food, can be carried into the view. We artificially disturbed the reproductive hierarchy by removing all four preidentified gamergates after the second day of recording and further observed the

Fig. 3 Colony of 59 *H. saltator* ants as a testbed. Ants can access a hidden foraging chamber through the south tunnel to bring paralyzed crickets for food

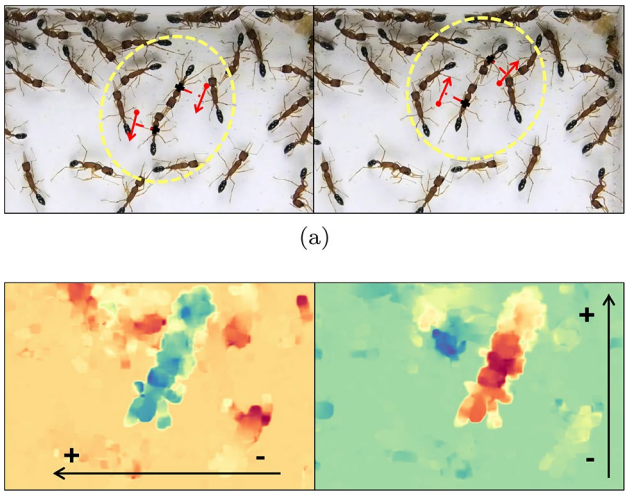

Fig. 4 **a** Example of two consecutive RGB frames cropped around a dueling interaction in yellow circle, in which red arrows visualize the moving directions of the two engaged ants at each time instant; **b** horizontal and vertical optical flow vectors generated from **a**, in each of which red (blue) are the regions of movement in the positive (negative) direction along the corresponding axis. Each flow vector has been normalized for better visualization (colour figure online)

process of hierarchy reformation until aggressive interactions almost disappeared in the last several days of the 20-day videotaping period. Therefore, the video frames of the first 2 days are annotated with c = Stable, while the later ones of 18 days are all with c = Unstable.

We follow the preprocessing method in [4] to extract from consecutive frames their optical flow, for which a pair of vectors encodes the horizontal and vertical transient movements from the input sequence (e.g., Fig. 4) [13]. Two optical flows in spatial resolution of 64×64 were computed every 2 min to use as an input x to the model, as each was obtained from two consecutive RGB frames 0.5 s apart in times. More details of the dataset are available online.[1]

5 Implementation of DCNNs with Grad-CAM

We use a classifier from our previous work [4] for the one-class classification task, in which the model trained only with observations from the first 2-day stable colony is to detect the unstable state of the colony later. We use the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) to measure classifier performance; this classifier metric ranges from 1.0 (perfect classification) to 0.5 (performance at chance) to 0.0 (reversed classification). That colony-state classifier has an overall AUC-ROC performance of 0.786 using only two consecutive optical flows as input for each prediction. Moreover, as shown in Table 1, the colony-state predictions during the early period of first 6 days after the reproductive hierarchy is disturbed have high AUC–ROC scores (0.909–0.933) on average [4], indicating that the micro-scale graphical features identified by the deep network may be strong predictors of macro-scale state dynamics.

More specifically, the classifier we use has four 2D convolutional layers $\phi_{1:4}$ with 2D max pooling between consecutive layers, and six other types of layers $\psi_{5:10}$ follow to produce the estimated likelihood of unstable colony state. As described in Sect. 3, we then employ Grad-CAM on the feature maps from ϕ_4. For each generated importance map M^c, bicubic interpolation is applied to match the size of the frame image to overlay. Our codes are available online.[2]

Table 1 Temporal performance of state detector developed in [4] while the tested colony was stabilized for 18 days—i.e., $D+1$–$D+18$

	$D+1$	$D+2$	$D+3$–$D+6$	$D+7$–$D+10$	$D+11$–$D+14$	$D+15$–$D+18$
AUC–ROC	0.933 ± 0.027	0.943 ± 0.014	0.909 ± 0.014	0.792 ± 0.013	0.688 ± 0.017	0.678 ± 0.022

For each period, the average AUC–ROC score from three individual executions is reported with the standard deviation

[1] https://github.com/ctyeong/OpticalFlows_HsAnts

[2] https://github.com/ctyeong/BeyondTracking

Fig. 5 Heatmaps from Grad-CAM at two arbitrary times as the rectifier function Γ is applied. Central areas appear more positively influential (red) than the edges (yellow) (colour figure online)

6 Results and model validation

As discussed in Sect. 4, we validate our approach by confirming that dueling behavior between ants is identified by the AI as strongly related to the unstable colony state. A model that can detect dueling with no prior knowledge of the behavior may identify other behavioral patterns that warrant further investigation.

Figure 5 displays the heatmaps produced by the initial application of Grad-CAM with rectifier Γ. Grad-CAM identifies that the central area is more critical than the boundaries, and this general pattern is consistent over time despite changes in ant behaviors. This visualization indicates that, for the purpose of identifying changes in colony hierarchical state, the neural network has learned to ignore interactions near boundaries and instead focuses on interactions in the center of the area. Although this pattern matches intuition from human observations of these ants, it is too coarse to identify important behaviors.

We thus applied a filtered rectifier Γ′ to only visualize regions of the top-5% positive gradients to identify the most dramatic responses in the generated heatmap to the ant motions, which resulted in more refined identifications of regions of importance. Figure 6 shows examples of dueling interactions detected by these highest gradients. Given that the deep network was not provided coordinates of the ants nor prior behavioral models of dueling, it is not surprising that the highlighted regions do not precisely identify specific ants in the interactions. Nevertheless, the network identifies general regions in close proximity to important behaviors. In particular, in Fig.6c, d, more than two ants were engaged in dueling, but the detection region dynamically moved around them while they actively participated. These results support that the trained model has not overfit trivial attributes such as brightness or contrast of video but learned from ant behaviors themselves.

Figure 7a also shows the case where two duelers are captured as intended while other active ants who are simply showing swift turns nearby each other without direct interaction are ignored by our model. This indicates that the DCNN classifier does not blindly take any type of movement into account for prediction; only relevant patterns are prioritized as features to utilize. Similarly, in Fig. 7b, two dueling ants are detected among a group of other non-dueling neighbors that are presenting rapid changes in motion and orientation.

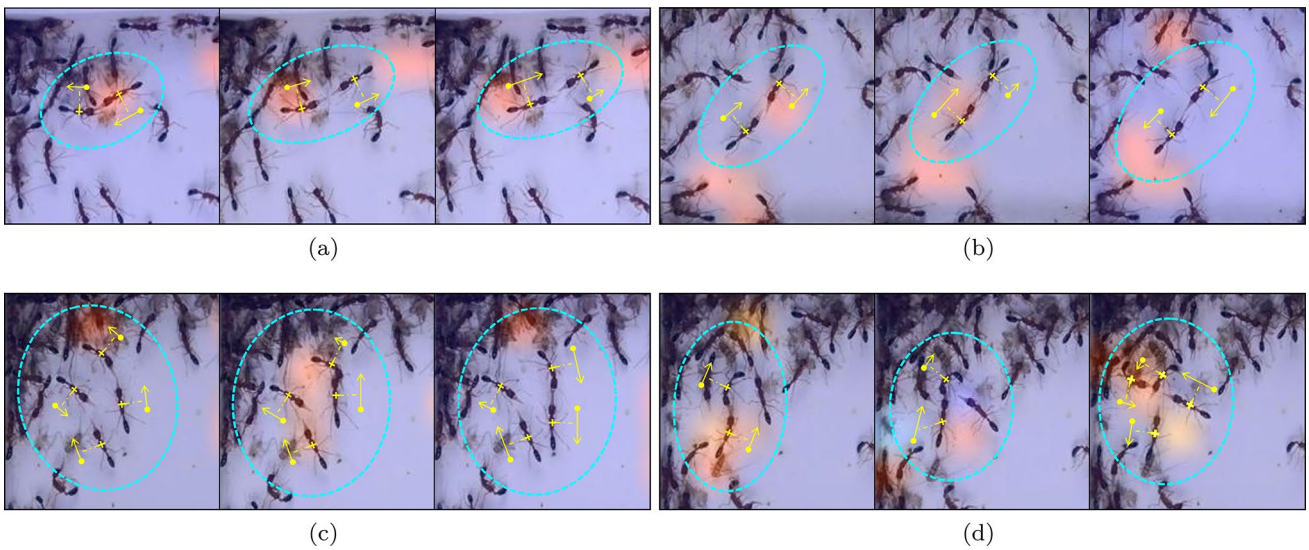

Fig. 6 Dueling examples captured by the modified rectifier Γ′, which only visualizes the top 5% impactful regions. Each sequence displays three consecutive frames cropped around the interaction for clarity, in which engaged ants are within blue circles, and each yellow arrow indicates the motional direction and speed of the corresponding ant with its tip and length, respectively (colour figure online)

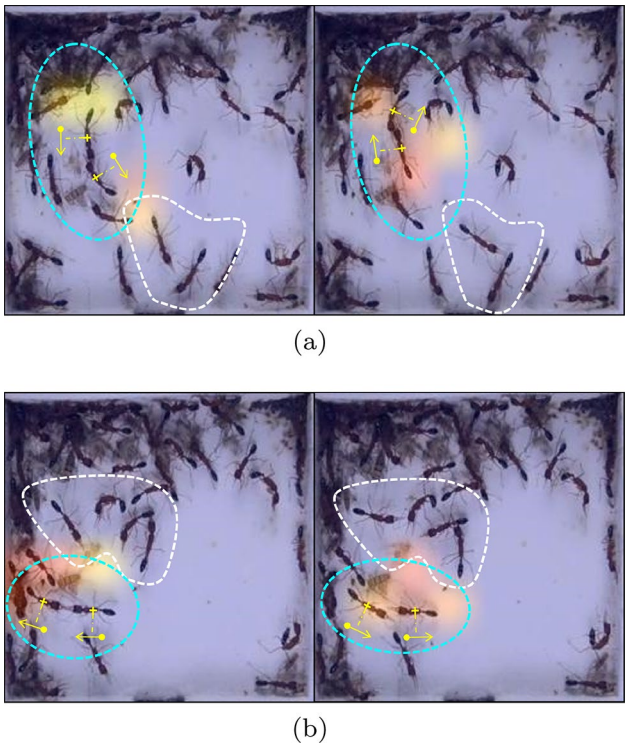

Fig. 7 Two examples in each of which dueling ants are detected (blue dash line) while other active ones are ignored (white dash line) (colour figure online)

This example also demonstrates the ability of our trained model to filter out unimportant motion patterns even when a high degree of motion flow is present.

7 Summary, discussion, and future work

We have proposed a deep-learning pipeline as a tool to uncover salient interactions among individuals in a swarm without requiring prior human knowledge about the behaviors or significant preprocessing effort devoted to individual tracking and behavioral coding. Our experimental results show that a trained classifier integrated with Grad-CAM can localize regions of key individual-scale interactions used by the classifier to make its colony-scale predictions. Validating our approach, identified behaviors, such as dueling, are the same behaviors that have been identified previously by human researchers without the aid of machine learning; however, our classifier discovered them without any prior guidance from humans. Thus, the library of other highlighted patterns from our pipeline can be used to generate new testable hypotheses of individual-to-colony emergence.

Our proposed approach greatly reduces human annotation effort as only macro-scale, swarm-level annotations are used in training. Significant effort is currently being used to develop machine-learning models for the subtask of tracking alone. Our approach suggests that tracking individuals may, in many cases, be an unnecessary step that wastes both computational and human resources. Furthermore, our proposed approach reduces the risk of introducing human bias in the pre-processing of individual-level observations. By examining the resulting set of prioritized micro-scale behaviors, human investigators could both identify individual contributors in specific videos as well as infer novel, generalizable patterns useful for understanding the evolution of global states. Consequently, our example is a model of how human–AI observational teams can engage in knowledge co-creation—each providing complementary strengths and ultimately realizing the vision of augmented, as opposed to purely artificial, intelligence.

An important future direction is to further classify the highlighted patterns automatically discovered by these pipelines. Human behavioral ecologists can discriminate between peculiar interactions (e.g., dueling, dominance biting, and policing [22]) that all may occur during the most unstable phases of reproductive hierarchy formation in *H. saltator* ants. Our method may have the ability to identify these behaviors, but it does not currently cluster similar identified patterns together and generate generalizable stereotypes that would be instructive to human observers hoping to identify these behaviors in their own future observations. Unsupervised learning methods could be adopted as a subsequent module to perform clustering and dimensionality reduction to better communicate common features of clusters, which may include patterns not yet appreciated by human researchers that are apparently useful in predicting swarm behavior.

The example application that has motivated the current work focuses on large-scale phase transitions of groups over time. That is, the approach we have demonstrated is tailored for experimental blocks where each collective generates a single sequence of video data that can be divided into distinct temporal intervals labeled as one state (e.g., “before”) or another state (e.g., “after”). We have used hierarchy formation in ants as a model example, but the same approach could be applied to a range of other changes in social behavior over time, such as understanding the individual contributions to transitions into mobbing or rioting behavior in previously calm human crowds. Furthermore, our approach could also be applied to experiments comparing the behavior of groups under some treatment condition to control groups. For example, if a behavioral ecologist would like to identify candidate behaviors of worker ants that are directed only toward their queen, videos of worker isolates without a queen could be compared to worker isolates with a queen. More subtle differences in the behaviors of the workers may be difficult for an unaided human observer to notice, but a DCNN trained to discriminate among the treatment and control groups may be able to identify these less apparent

behaviors. Thus, our model-free approach can be a precursor to more formal hypothesis testing based on candidate behavioral differences between groups first highlighted by the DCNN.

References

1. Bozek K, Hebert L, Mikheyev AS, Stephens GJ (2018) Towards dense object tracking in a 2D honeybee hive. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 4185–4193
2. Choi T, Pavlic TP, Richa AW (2017) Automated synthesis of scalable algorithms for inferring non-local properties to assist in multi-robot teaming. In: 13th IEEE Conference on Automation Science and Engineering (CASE 2017), pp 1522–1527
3. Choi T, Kang S, Pavlic TP (2020) Learning local behavioral sequences to better infer non-local properties in real multi-robot systems. In: 2020 IEEE International Conference Robotics and Automation (ICRA 2020), pp 2138–2144
4. Choi T, Pyenson B, Liebig J, Pavlic TP (2021) Identification of abnormal states in videos of ants undergoing social phase change. In: Proceedings of the AAAI Conference on Artificial Intelligencec (AAAI 2021), vol 35. pp 15286–15292
5. Chollet F (2020) Grad-CAM class activation visualization. https://keras.io/examples/vision/grad_cam/
6. Couzin ID, Krause J, James R, Ruxton GD, Franks NR (2002) Collective memory and spatial sorting in animal groups. J Theor Biol 218(1):1–11
7. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge
8. Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, Couzin ID (2019) DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. eLife 8:e47994
9. He T, Guo J, Chen N, Xu X, Wang Z, Fu K, Liu L, Yi Z (2019) MediMLP: using Grad-CAM to extract crucial variables for lung cancer postoperative complication prediction. IEEE J Biomed Health Inform 24(6):1762–1771
10. Lertvittayakumjorn P, Toni F (2019) Human-grounded evaluations of explanation methods for text classification. arXiv preprint arXiv:1908.11355
11. Li CL, Sohn K, Yoon J, Pfister T (2021) CutPaste: self-supervised learning for anomaly detection and localization. In: EEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021), pp 9664–9674
12. Liebig J, Peeters C, Hölldobler B (1999) Worker policing limits the number of reproductives in a ponerine ant. Proc R Soc Lond Ser B Biol Sci 266(1431):1865–1870
13. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. In: Conference on Computer Vision and Pattern Recognition (CVPR 2009), pp 935–942
14. Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW (2019) Using deeplabcut for 3D markerless pose estimation across species and behaviors. Nat Protoc 14(7):2152–2176
15. Panwar H, Gupta P, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Singh V (2020) A deep learning and Grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-scan images. Chaos Solitons Fractals 140
16. Peeters C, Crewe R (1985) Worker reproduction in the ponerine ant *Ophthalmopone berthoudi*: an alternative form of eusocial organization. Behav Ecol Sociobiol 18(1):29–37
17. Pratt SC, Mallon EB, Sumpter DJ, Franks NR (2002) Quorum sensing, recruitment, and collective decision-making during colony emigration by the ant *Leptothorax albipennis*. Behav Ecol Sociobiol 52(2):117–127
18. Rajagopal D, Balachandran V, Hovy E, Tsvetkov Y (2021) SelfExplain: a self-explaining architecture for neural text classifiers. arXiv preprint arXiv:2103.12279
19. Redmon J, Farhadi A (2018) YOLOv3: an incremental improvement. arXiv preprint arXiv:1804.02767
20. Reid CR, Lutz MJ, Powell S, Kao AB, Couzin ID, Garnier S (2015) Army ants dynamically adjust living bridges in response to a cost-benefit trade-off. Proc Natl Acad Sci 112(49):15113–15118
21. Romero-Ferrero F, Bergomi MG, Hinz RC, Heras FJ, de Polavieja GG (2019) idtracker.ai: tracking all individuals in small or large collectives of unmarked animals. Nat Methods 16(2):179–182
22. Sasaki T, Penick CA, Shaffer Z, Haight KL, Pratt SC, Liebig J (2016) A simple behavioral model predicts the emergence of complex animal hierarchies. Am Nat 187(6):765–775
23. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: visual explanations from deep networks via gradient-based localization. In: IEEE International Conference on Computer Vision (ICCV 2017), pp 618–626
24. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
25. Wang L, Xiong Y, Wang Z, Qiao Y, Lin D, Tang X, Van Gool L (2016) Temporal segment networks: towards good practices for deep action recognition. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9912. Springer, Cham, pp 20–36