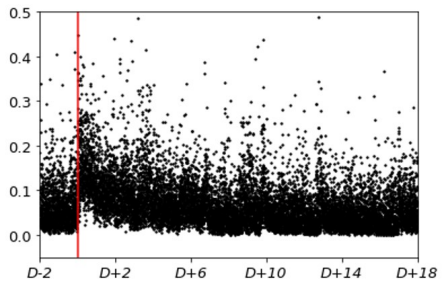

Figure 4: Optical flow weights each sampled at the interval of over 2 minutes for 20 days. The horizontal axis shows the days of observation with the red separator when the *gamer-gates* were intentionally moved. The vertical axis indicates the weight levels, each standardized in $[0,1]$ by the global max and min, omitting extreme outliers for clarity.

m	1	2	4
AUC	0.760 ± 0.016	0.786 ± 0.009	0.787 ± 0.008

Table 1: Average performance when the number of optical flow image frames per input is set to $1, 2$, or 4.

Consequently, a simple model might be built that uses the overall rise of flow weight as the only feature to distinguish the unstable colony from the stable especially at early development of the unstable state. However, following our results, we will provide concrete examples that show the limitations of such a design and the need of more complex models for reliable predictions.

Model Evaluation

Here, we demonstrate the OC performance of our proposed method. We first describe an ablation study to find the best number of optical-flow frames per input. Next, baselines used for comparison are introduced that will help us explore our method's overall reliability and prediction robustness in various time windows during colonial stabilization.

As in previous works (**?**), the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) are measured for each model to reflect the separability between classes. Moreover, the average over three splits is reported with the standard deviation when needed.

Ablation Study: Results from tests with $m \in \{1,2,4\}$ are shown in Table 1. There was an improvement as m increased from 1 to 2, while doubling it to 4 did not offer any benefit. The result may indicate that the observation of one more second does not add significantly more information. Learning IO-GEN could also be more challenging as it is asked to generate longer motional sequences. Thus, m is set to 2 hereafter considering both efficiency and effectiveness of our model.

Baselines: **OFW** uses the temporal optical flow weights to set the best threshold to report the best classification result. **DCAE** is a similar threshold-based method relying on the reconstruction error as the feature of novelty (**?**). **OC-SVM** (**?**) takes the encoder of DCAE to build the One-class SVM on it providing the performance with the best ν parameter. While **DSVDD** here is designed similarly to the description by **?**, the adjustments in our implementation are described in Model Structures & Relevant Parameters above. **GEN** and **N-GEN** are generative models to train a separate classifier as our method. GEN is, however, a standard generative model adopting the feature matching technique in the discriminator network instead without the intervention of DSVDD. N-GEN replaces $\phi(G(z))$ with arbitrary noisy data $\vec{v}' \in \mathbb{R}^{1\times 2048}$ where each element of $\vec{v}'$ is drawn from $N(0,\alpha)$ where α is the global variation of $\vec{v} \sim \phi(G(z))$.

METHOD	AUC
OFW	0.506
DCAE	0.506 ± 0.002
OC-SVM	0.523 ± 0.004
DSVDD	0.762 ± 0.013
GEN	0.587 ± 0.032
N-GEN	0.699 ± 0.006
IO-GEN	$\mathbf{0.786 \pm 0.009}$

Table 2: Average AUC of tested models with the standard deviation as all 18-day unstable observations are considered.

Overall Performance: Table 2 helps estimate overall reliability of each model for the image inputs that can be captured at an arbitrary timing since all samples from unstable colony were included for test. OFW and DCAE suggest the limitation of only relying on thresholding a simplistic one-dimensional signal. In particular, the low accuracy of DCAE implies that precise reconstruction is achieved also for the motions from unseen, unstable colony. Similarly, the OC-SVM can utilize only little benefit from the encoding capability. On the other hand, DSVDD leads at least 45% increase of AUC score simply fine-tuning the encoder part of DCAE because unstable examples are more easily distinguished in the newly learned hyperspheric data description. In addition, our model brings about a further improvement proving that utilizing a subsequent classifier with synthetic examples can be more effective than the distancing heuristic in DSVDD to make full use of multi-dimensional relationships among features. Nevertheless, GEN and N-GEN provide 25% and 11% poorer performance than ours although both also use synthetic data to train a classifier. N-GEN actually performs better than GEN implying that the prior knowledge on data description is useful for effective data synthesis. Still, its insufficient reliability emphasizes the importance of realism in generated datasets as well.

Detection in Different Developmental Phases: Figure 5 displays the performance variation of each model as the tested data from the unstable colony are confined in various temporal windows. Consistent with Fig. 4, the prediction performance generally degrades for later temporal bins because the ant colony is more stabilized. Our framework still indicates the top performance in almost any phase especially presenting the highest margins from DSVDD in a highly ambiguous time period between D+2 and D+10, in

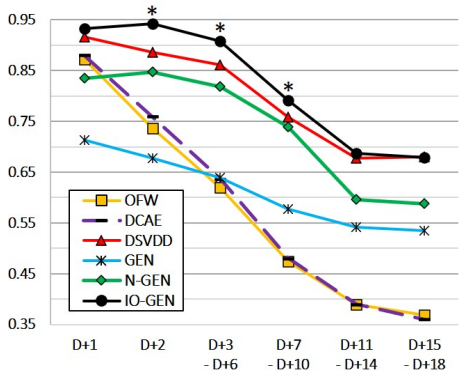

Figure 5: Average AUC changes for predictions within different temporal windows. Asterisks (*) mark statistically significant improvement over DSVDD ($p < .05$).

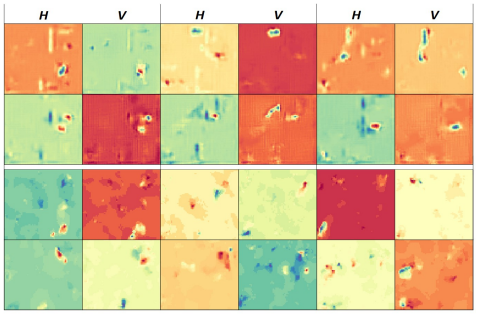

Figure 6: Optical flow examples: (top two rows) Six synthesized pairs from IO-GEN; (bottom two rows) Six real examples. Each (H-V) pair show horizontal and vertical motions, respectively, for which pixels are normalized in each image.

which the proportion of stable observations dramatically increased. As expected from Fig. 4, OFW and DCAE highly depend on the timing of application because their scores are close to that of DSVDD early while lower even than 0.5 after D+6. If the initial social transition is less conspicuous, possibly due to a smaller population, these models may perform poorly because of less intense competition caused. Moreover, the results from GEN and N-GEN reemphasize the insufficiency of solely relying on realism or spatial characteristics of produced features when training generative models. In particular, as illustrated in Fig. 2, GEN produces fake samples that closely resemble the ones of stable state, and so the biased classifier leads to the worst performance in the early stages (∼D+2) when colonial instability was highest.

Model Properties

Figure 6 compares synthetic optical flows from IO-GEN to real optical flows; the generated optical flows are visually similar to real flows. Furthermore, Fig. 7a illustrates that the lowest distance distribution to $\vec{c}$ is measured with IO-GEN, as designed, whereas GEN behaves similarly to the stable dataset. Figure 7b finally shows the predictive outcomes of Classifier, which are likelihoods of unstable state. With the *label switch*, the confidence becomes positively correlated with the distance to $\vec{c}$ viewing inner outliers as samples from the most stable colony. Clear differences between classes imply that learned knowledge to discriminate stable and more-stable states in DSVDD can be transferred for classification of another pair as stable or unstable.

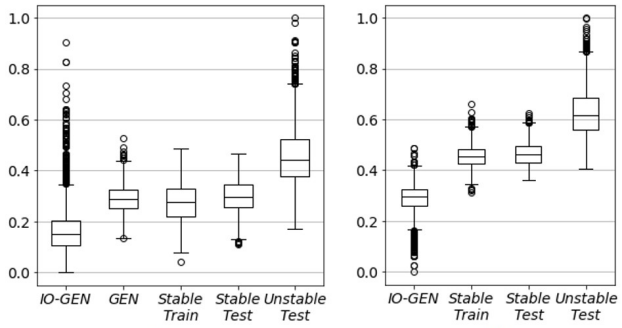

Figure 7: For different types of data: On left, normalized Euclidean distances to $\vec{c}$ in feature description $\mathscr{F}$ of DSVDD. On right, predicted likelihoods from Classifier.

Conclusion

We have introduced a novel generative model IO-GEN that can utilize a pre-trained DSVDD and a separate classifier to successfully solve the OC problem. Our framework has been applied to 20-day video data from an entire society of 59 *H. saltator* ants to identify a colony's stable or unstable state only from a 1-second motional sequence. Experiments have shown that the classifier trained with the synthetic data from IO-GEN outperforms other state-of-the-art baselines at any temporal phase during social stabilization.

Our future directions include a graphical user interface for this method that acts as a tool for biologists that can propose frames or individuals (regions of interest) implicated as being crucial in the evolution of social state. To implement this, an additional module can be built to monitor and visualize the levels of gradient passing from spatio-temporal behavioral features to the final decision output (**?**).

Acknowledgments

Support provided by NSF PHY-1505048 and SES-1735579.

Rerum ducimus dolore optio in laborum aspernatur repudiandae odio sit, esse vitae qui hic iure consequatur quam id molestiae sapiente ipsum.Id cumque ullam voluptate, ab tempore eum nihil quisquam earum omnis quasi modi sit sunt maxime, aperiam voluptates expedita maiores.Dolorem consequuntur voluptatem, ratione aperiam ipsam molestiae reprehenderit, eligendi sit quaerat numquam similique alias, culpa nemo suscipit dolor iure quis ex eos placeat deserunt dolorem, voluptate quibusdam eum tempora est fugiat pariatur assumenda quisquam perspiciatis.Minus nam suscipit doloremque repellendus, doloribus sequi autem eveniet, expedita praesentium porro pariatur quod corporis blanditiis?Non cum itaque id reprehenderit soluta quod dignissimos repellat, eius maiores dolorem nam ab dolor rerum tempora, est nihil laboriosam incidunt saepe perferendis