# CropFollowRL: Learning Under-Canopy Navigation Policy with Keypoints Abstraction

Arun Narenthiran Sivakumar[1] Ning Wang[1] Felipe Andrade G. Tommaselli[1,2]
Mateus Valverde Gasparino[1] Marcelo Becker[2] Girish Chowdhary[1]
[1]Field Robotics Engineering and Sciences Hub (FRESH), University of Illinois Urbana-Champaign (UIUC)
[2] Dept. of Mechanical Engineering, University of São Paulo (USP), São Carlos

***Abstract*—Under-canopy agricultural robots have shown the potential to enable precise crop monitoring, weeding, spraying, cover crop planting, and plant manipulation tasks without soil compaction. Autonomous navigation under the canopy is challenging due to the tight space available to navigate, frequent occlusion of sensors, and large variability in the appearance of the environment. Prior work focused on developing a learning-based perception system with semantic keypoints for robust under-canopy navigation. Though they demonstrated robust navigation in diverse field conditions, the system is limited by the inability to use failure data and self-improve. We propose CropFollowRL, a real2sim2real system to train a learning-based controller in simulation with reinforcement learning using semantic keypoints as an abstraction. Our evaluation in various simulation environments indicates that our learned controller generalizes well directly to some environments in terms of distance traveled before collisions and shows significant improvement after finetuning in several other environments which indicates potential for sim2real transfer and real-world deployment.**

## I. Introduction

Under-canopy agricultural robots can enable applications such as ultraprecise crop monitoring, weeding, spraying, and cover crop planting which is not possible with existing over-the-canopy equipment like tractors and aerial vehicles [4]. Autonomous navigation under the plant canopy is challenging due to the tight space to navigate between the crop rows (around 17cms of error margin on either side of the robot), inaccuracy in RTK GPS from multipath error, frequent sensor occlusion from leaves, and the large variability in the environment due to changing plant, soil, and weed growth conditions.

Earlier works in under-canopy navigation used 2D LiDAR for perception with classical line fitting approaches [3]. Though this approach works in clean fields with less occlusion and weed growth, it suffers from the lack of semantic information in the sparse point cloud to distinguish the crop rows from occluding leaves and weeds. Visual navigation is a better alternative because of the rich semantic information of the scene captured in the images. However, the classical approach of vision-based row following using segmentation and line fitting requires frequent manual tuning of parameters to deal with the variability in the environment. Learning-based visual navigation can help overcome this challenge but the question becomes: What is the learned output representation from images that enables robust under-canopy visual navigation?

CropFollow [7] and CropFollow++ [6] showed successful demonstrations of learning-based under-canopy visual navigation systems over large distances in diverse field conditions. CropFollow used an end-to-end perception learning approach and directly predicted the states of the robot (heading angle and distance ratio) from the input monocular RGB image. CropFollow++ showed improved robustness by using semantic keypoints as the predicted output representation. The vertices of the triangle formed on the image by the crop row lines and the bottom of the image are the three semantic keypoints. The improved robustness is enabled by adding heuristics that use the uncertainty of the predicted keypoint heatmaps and the known structure of the environment.

Fig. 1: Under-canopy navigation is challenging due to the tight space to traverse between the crop rows, frequent occlusion, and large variability in the appearance of the environment.

Though CropFollow++ showed deployment on multiple under-canopy robots over 25kms, it is limited by the lack of self-learning capability to improve its performance from failures (collision with plants). This is because of the classical model predictive controller (MPC) that is used in this system. Despite the rich information predicted in the keypoint heatmaps, only the maximum intensity pixel coordinates are used to calculate the states (heading angle and distance ratio) used in MPC which is restrictive. We hypothesize that a learning-based differentiable controller with keypoint heatmaps as input can exploit all the information in the heatmaps to enable improved navigation performance and can

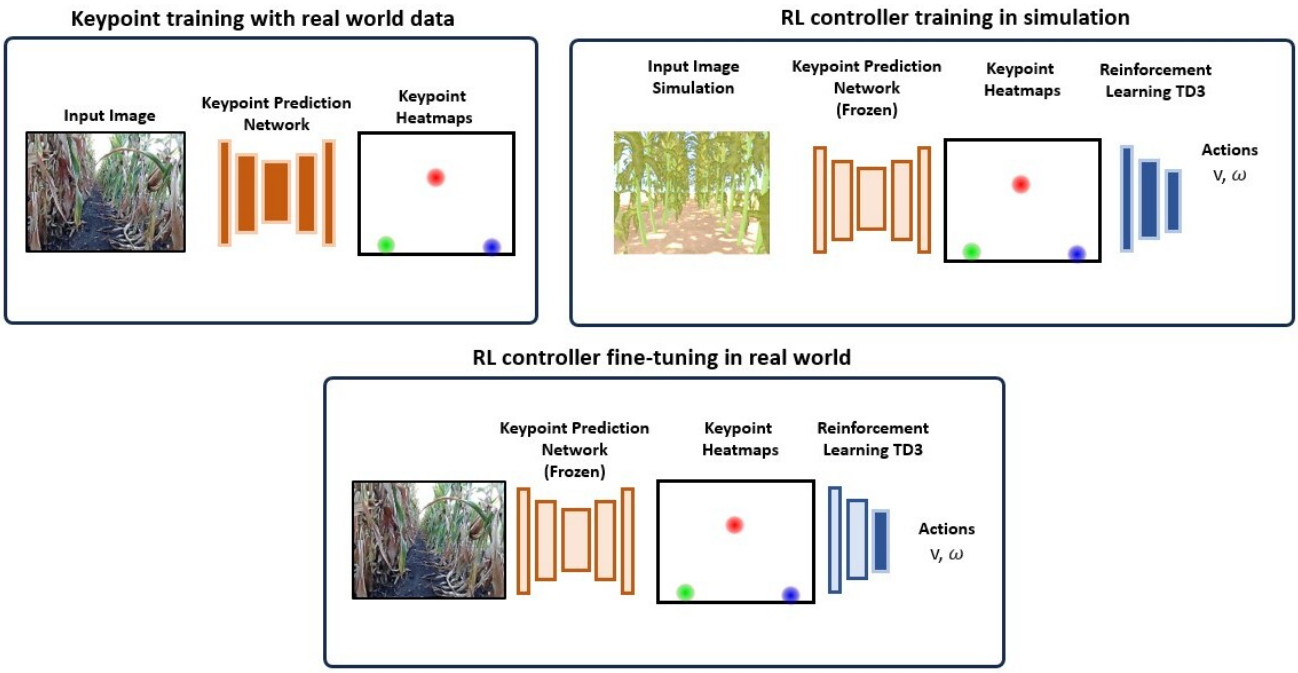


Fig. 2: CropFollowRL overview. We first train the perception neural network that predicts semantic keypoints with real-world data. We then transfer this perception system to simulation and train a neural network controller using TD3 off-policy RL algorithm. We propose to transfer this controller for real-world deployment using fine-tuning with small amount of real world data.

also have the self-learning capability to improve from failures using reinforcement learning (RL).

Because of the limitation of large sample complexity in model-free RL algorithms, we propose to use a real2sim2real approach similar to [5]. We use the keypoint prediction model trained in CropFollow++ as an abstraction to train an RL controller in simulation and transfer this learned policy to the real world. In this initial work, we show successful training of the proposed keypoint heatmap-based policy using Twin Delayed Deep Deterministic Policy Gradient (TD3) off-policy RL algorithm [2] in simulation. In real-world under-canopy environments, the movement of the robot along the crop rows and the collision with plants are the only reliable sources of reward/feedback. We train our proposed controller using only this feedback in one simulation environment. Our evaluation results in simulation show that our learned controller generalizes well to some environments directly and in some other cases it shows significant improvement after minimal finetuning indicating potential for real-world deployment.

## II. CropFollowRL overview

Our proposed architecture consists of two modules - a perception system trained to predict three semantic keypoints and a learning-based controller that uses the semantic keypoint heatmaps as input and outputs linear and angular velocity action commands. We use the same semantic keypoint prediction network from CropFollow++ [6] which is trained on a large and diverse real world under-canopy dataset. We transfer this perception system to simulation and train a neural network-based controller with TD3 off-policy RL algorithm. We evaluate this trained controller in various other crop environments in simulation with and without fine-tuning. These modules are explained in detail below.

### A. Perception with semantic keypoints

Our perception system is a fully convolutional network that takes as input an RGB image and outputs three semantic keypoints corresponding to the vertices of the triangle formed on the image by the crop row lines and the bottom of the image. The output heatmaps are 1/4th the spatial resolution of the input image. A spatial softmax layer is applied to the final layer.

### B. RL controller training in simulation

We use a gazebo environment with simulated plants to train the RL controller. We use one type of corn plant as a training environment and seven diverse validation environments (comprised of two types of corn plants, three types of sorghum plants, and two types of tobacco plants). We transfer the perception system to simulation to predict the keypoint heatmaps. Note that this perception network remains frozen and its weights are not updated while training the controller.

*1) TD3 algorithm:* We use TD3 off-policy RL algorithm to train the controller. TD3 is used in environments with continuous action spaces like our problem setting. It builds on the Deep Deterministic Policy Gradient (DDPG) algorithm with a few modifications to address some of the issues in DDPG. It uses two Q-functions instead of one in DDPG to

minimize the issue of overestimation bias. The Bellman error to update the Q-functions is formulated by treating the smaller of the two Q-values as the target. Since TD3 is a deterministic algorithm, a noise distribution is added to the action to enable exploration during training. The action with added noise is clipped to ensure the actions are within the bounds. Also, TD3 updates the policy network less frequently than Q-function networks.

*2) Network architecture and training:* Our actor network and two Q-function networks use fully connected layers. The three keypoint heatmap outputs are flattened to a vector of 13440 dimensions and the action from the previous time step is concatenated to this to create a 13442 dimensional state vector. A $\tanh$ activation layer is used in the output of the actor network. We adapt the TD3 implementation from [1] for our problem.

To train this policy network, we use the collision with plants as a negative reward and the distance traveled along the length of the crop rows as a positive reward. The full reward expression $r_t$ can be visualized in figure 1:

$$r_t = \begin{cases} -100 & \text{if collision} \\ -3 \cdot |v_x - 1| - |w_z| + \frac{\delta_x}{10} & \text{otherwise} \end{cases} \quad (1)$$

The reward function $r_t$ is associated with action $a_t$. The expression depends on linear velocity $v_x$, angular velocity $w_z$ around the z-axis, and $\delta_x$, which signifies movement along the x-axis. We made this choice for the reward since these are the only reliable sources of feedback available during real-world deployment in under-canopy conditions.

We use a corn plot simulated with one type of corn plant model as our training environment.

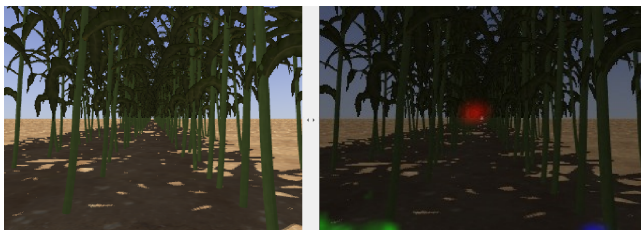

Fig. 3: We visualize the RGB image and the corresponding keypoint heatmap for a sample image from the environment used for training the RL controller.

## III. Preliminary Results and Discussion

We trained the above-mentioned policy network with TD3 algorithm in one corn environment until convergence of average Q-value (shown in Fig. 4). To evaluate this policy, we chose seven diverse environments available to us in simulation corresponding to two types of corn plants, three types of sorghum plants and two types of tobacco plants. Fig 5. shows the RGB image and the corresponding keypoint heatmaps (superimposed on the input RGB image) for a sample image from each validation environment. The varying levels of noise in the keypoint heatmaps as seen in this visualization is an indicator of the varying difficulty in generalization for the trained controller.

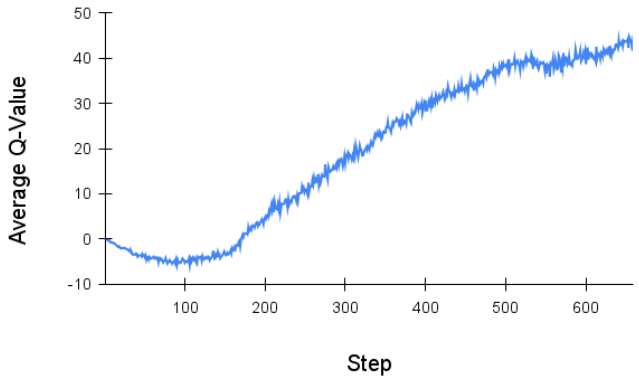


Fig. 4: Average Q-value during training in the corn simulation training environment is shown here. We can see the convergence of the average Q-value at the end of training.

### A. Evaluation Metric

We used the distance traveled before the collision as the quantitative metric for evaluation. We first calculated this metric by evaluating the generalization performance of the policy trained in one corn environment. We then also fine-tuned this policy to each validation environment with a small amount of data and recalculated the same metric.

### B. Validation Results

Fig 6. shows the distance traveled before collision before and after fine-tuning for all seven validation environments mentioned above. Based on the results in Fig 6. there are three types of environments:

- The controller trained in one corn environment directly generalizes well (more than 30m traveled before the collision). Sorghum #1 environment belongs to this category. This indicates that the distribution of input keypoint heatmaps is similar in the training environment and this validation environment. We also see only minimal improvement in the metric after finetuning.
- The trained controller does not show good performance during evaluation indicating lack of generalization. However, after finetuning, there is a significant improvement in the distance traveled before the collision. Corn #1, Corn #2, and Tobacco #2 belong to this category. Among these, Corn #1 especially shows a significant jump in performance after finetuning (from 9.07m to 25.44m) The lack of good performance could be due to a shift in the distribution of input keypoint heatmaps to the controller. But the heatmaps might still contain useful information for control and hence the jump in performance after finetuning.
- The validation performance is poor before as well as after finetuning (Sorghum #2, Sorghum #3, Tobacco #1). This might indicate that the keypoint prediction module trained with only real-world data might not be generalizing well

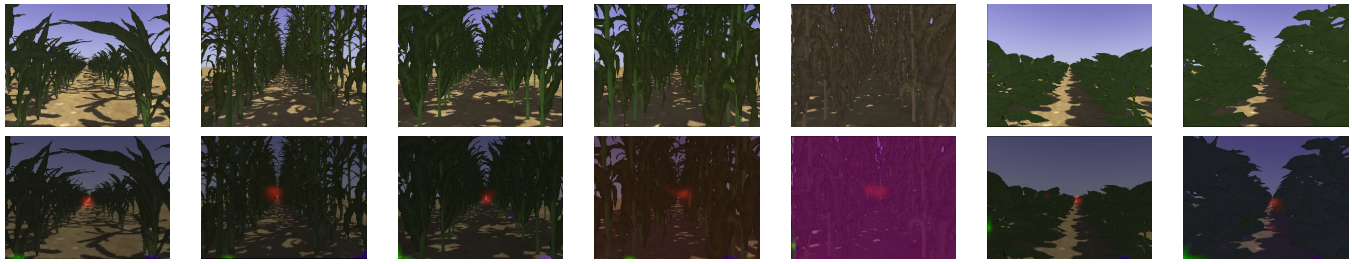

Fig. 5: We visualize the RGB (top row) and corresponding keypoint heatmaps (bottom row) for a sample image from each of the evaluation environments in simulation. The first two columns represent corn, columns 3-4 represent sorghum and the last two columns represent tobacco. Note the varying levels of noise in the heatmap across these different environments.

to these simulation environments and hence the predicted keypoint heatmaps do not contain enough information for the controller to learn a good mapping from heatmap space to action space.

Our results in the second category of environments show great potential for sim2real transfer by fine-tuning with a small amount of data. As long as the predicted keypoint heatmaps contain useful signal, our proposed RL system can be used to self-improve during operation in real fields.

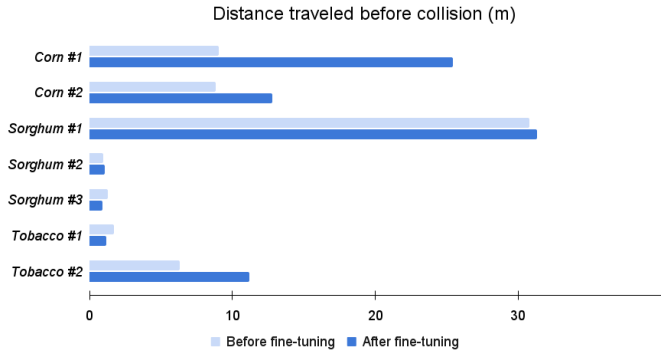


Fig. 6: Distance traveled before collision in meters for various evaluation environments simulation before and after minimal fine-tuning. The significant increase in distance after fine-tuning can be seen in three of the seven environments.

## IV. Conclusion

We have demonstrated the feasibility of training a learning based policy for under-canopy navigation with RL using only the reliable sources of reward available in real field conditions. Our preliminary results in simulation show promise for sim2real transfer and self-improvement in the real world. Further work exploring other choices of network architectures, off-policy algorithms and finetuning the policy network can improve the performance of the proposed system.

## Acknowledgments

This work was supported in part by NSF STTR #1951250, NSF NRI 2.0 NIFA #2021-67021-33449, AI-FARMS #1024178, NSF-USDA COALESCE #2021-67021-34418, USDA grants iCOVER(#NR233A750004G066) and iFARM(#2022-77038-37306). Felipe Tommaselli's research internship visit to UIUC was funded by BEPE fellowship from FAPESP (No . 2022/08330-9 and 2023/15926-8). We also acknowledge CNPQ support in LabRoMs research group (No 308092/2020-1).

Link to videos showing evaluation in validation environments before and after fine-tuning https://uofi.box.com/s/09xeqz4wekoiujke7fea0msdipiaoj9e

## References

[1] Reinis Cimurs, Il Hong Suh, and Jin Han Lee. Goal-driven autonomous exploration through deep reinforcement learning. *IEEE Robotics and Automation Letters*, 7(2):730–737, 2022. doi: 10.1109/LRA.2021.3133591.

[2] Scott Fujimoto, Herke Hoof, and David Meger. Addressing function approximation error in actor-critic methods. In *International conference on machine learning*, pages 1587–1596. PMLR, 2018.

[3] Vitor AH Higuti, Andres EB Velasquez, Daniel Varela Magalhaes, Marcelo Becker, and Girish Chowdhary. Under canopy light detection and ranging-based autonomous navigation. *Journal of Field Robotics*, 36(3):547–567, 2019.

[4] Wyatt McAllister, Denis Osipychev, Adam Davis, and Girish Chowdhary. Agbots: Weeding a field with a team of autonomous robots. *Computers and Electronics in Agriculture*, 163:104827, 2019.

[5] Matthias Müller, Alexey Dosovitskiy, Bernard Ghanem, and Vladlen Koltun. Driving policy transfer via modularity and abstraction. *arXiv preprint arXiv:1804.09364*, 2018.

[6] Arun N Sivakumar, Mateus V Gasparino, Michael McGuire, Vitor AH Higuti, M Ugur Akcal, and Girish Chowdhary. Lessons from deploying cropfollow++: Under-canopy agricultural navigation with keypoints. *arXiv preprint arXiv:2404.17718*, 2024.

[7] Arun Narenthiran Sivakumar, Sahil Modi, Mateus Valverde Gasparino, Che Ellis, Andres Eduardo Baquero Velasquez, Girish Chowdhary, and Saurabh Gupta. Learned visual navigation for under-canopy agricultural robots. *arXiv preprint arXiv:2107.02792*, 2021.