LaND: Learning to Navigate from Disengagements
06/02/2022

1 Authors, Conference/Journal, Year, Num Citations

- Authors: Gregory Kahn, Pieter Abbeel, Sergey Levine (Berkeley AI Research)
- Conference: ICRA (Best Field Robotics Paper Finalist)
- Year: 2021
- Num Citations: 8
- Readability: 8/10

2 Problem Statement

- How to make a robot that can learn to navigate using minimal sensory input and less(or no) labeled data ?

3 Contribution

- Proposes an algorithm for learning to navigate from disengagements (LaND). Disengagement signal refers to the human operator's intervention when robot takes undesirable action. The key idea is that if robot can successfully learn to execute the actions that avoid disengagement, then the robot will successfully complete the task.
- Shows that disengagement provides a strong signal for autonomous mobile robots to learn to navigate.
- Shows that their algorithm is able to outperform then state-of-the art methods in a real world navigation environment.

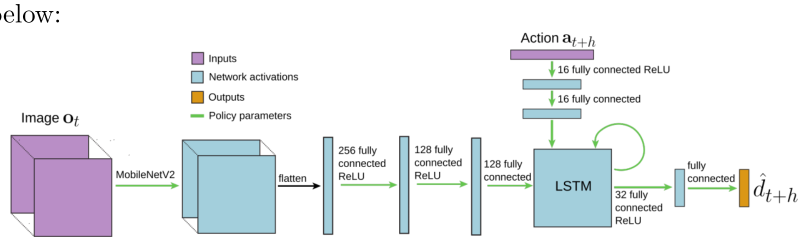
4 How is the work different from related works?

- Other navigation algorithms, build a map of the world, localizing the robot within the map and following paths within the map towards desired destination.
- Another class of learning algorithms uses environmental clues such as depth, objects to learn to navigate. These algorithms are expensive to train due to the labelling cost.

- Another class of algorithms is based on imitation, in which a policy is trained to mimic an expert.
- Sim-to-real is another class of algorithm, in which policy is trained in a simulated environment and executed in the real world.
- LaND algorithm uses no labeled data other than disengagement data. And it learns from off-policy data.

5 Main Idea

- Testing autonomous mobile robots in real world is a necessary aspect of developing autonomous navigation system. During testing human observer oversees the navigation system, whenever robot takes an undesirable action the human operator disengages the robot's autonomy system. These disengagements provide learning signal by which the robot can learn to navigate.
- Data Collection: Collects data, observations(camera image) o_t, action a_t and binary disengagement signal d_t at each time step t are saved after every δx meters into the dataset D.
- Predictive Model:
  - Learned predictive model will take input observation (o_t), a sequence of H future actions $a_{t:t+H} = (a_t, a_{t+1}, .., a_{t+H-1})$ and outputs a sequence of future disengagement probabilities. Mathematically defined as $f_\theta(o_t, a_{t,t+H}) \rightarrow \hat{d}_{t,t+H}$.
  - The learning function $f_\theta(.)$ is parameterized by θ and is defined as below:



  - Loss function: $L(\theta; D) = \sum_{o_t, a_{t,t+H}, d_{t,t+H}} \sum_{h=0}^{H-1} L^{CE}(\hat{d}_{t+h}, d_{t+h})$ s.t $\hat{d}_{t,t+H} = f_\theta(o_t, a_{t,t+H})$, where L^{CE} is cross-entropy loss.
  - Trained using SGD algorithm and oversampling positive disengagement signals.
- Planning and Control:

- Planning objective is encoded with the following cost function:
  $C(\hat{d}_{t,t+H}, a_{t,t+H}) = \sum_{h=0}^{H-1} \hat{d}_{t+h} + \alpha||a_{t+h} - g||_2^2$
  where the first term encourages the robot to avoid disengagement and second term makes robot to navigate towards desired goal, α is a control hyper-parameter.
- At planning step robot solves $a^*_{t,t+H} = \arg\min_{a_{t,t+H}} C(\hat{d}_{t,t+H}, a_{t,t+H})$
  s.t. $\hat{d}_{t,t+H} = f_\theta(o_t, a_{t,t+H})$
- To solve the above objective function authors uses zeroth order stochastic optimizer [4] (non-gradient) algorithm. (TODO: I am not able to understand the algorithm)

6 Evaluation

- Evaluation Metric: average distance travelled before disengagement. Author accepts that using this metric for measuring progress is debatable. But there is a general consensus on good autonomous system has less disengagements.
- They showed that with more disengagement data the robot is able improve

Method	Avg. distance until disengagement (meters)
LaND	101.2
LaND with finetuning	218.3

TABLE II: Experimental demonstration of our LaND improving as more data is gathered. We first collected data using our LaND control policy on 1.3 km of never-before-seen sidewalk. We then finetuned our method on this additional data, and evaluated this finetuned model on the same 1.3 km of sidewalk. Our method improves by over 2× when finetuned on the collected data, showing that our approach is able to continue to improve as more data is gathered.

its policy.

- Compares with two state-of-the-art navigation algorithms:
  - Behaviour cloning: Imitation learning approach [3], [1]
  - [2]'s reinforcement learning algorithm
- Performance comparison:

Method	Avg. distance until disengagement (meters)
Behavioral cloning (e.g., [20], [22], [25])	13.4
Kendall et. al. [36]	2.0
LaND (ours)	**87.5**

TABLE I: Experimental evaluation on 2.3 km of never-before-seen sidewalks (Fig. 1). Our LaND approach is better able to navigate these sidewalks, travelling $6.5\times$ further before disengagement compared to the next best method.

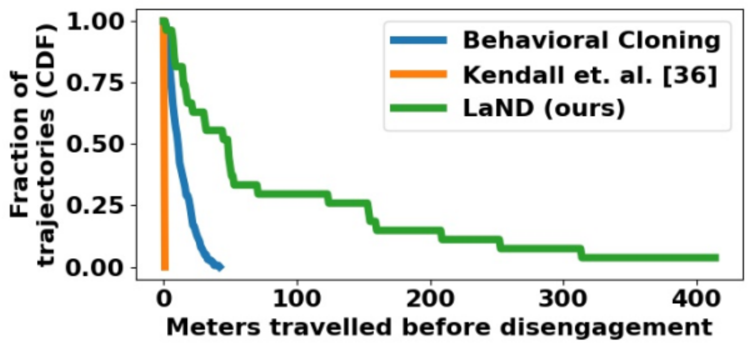


Fig. 7: Experimental evaluation on 2.3 km of never-before-seen sidewalks (Fig. 1). The plot shows the fraction of trajectories—defined as a continuous episode of engaged autonomy—which travelled a certain distance before a disengagement. Methods closer to the top right are better because this indicates a longer distance travelled before disengagement. Our LaND approach is able to travel farther before disengagement: 33% of the trajectories travelled further than 50 meters, including a trajectory of over 400 meters. In contrast, none of the prior methods were able to travel more than 50 meters before disengagement.

7 Conclusion

- Research Direction: How to combine ideas from imitation learning and disengagement signal to learn better control policies?

References

[1] Felipe Codevilla, Matthias Müller, Alexey Dosovitskiy, Antonio M. López, and Vladlen Koltun. End-to-end driving via conditional imitation learning. *ICRA*, abs/1710.02410, 2018.

[2] Alex Kendall, Jeffrey Hawke, David Janz, Przemyslaw Mazur, Daniele Reda, John-Mark Allen, Vinh-Dieu Lam, Alex Bewley, and Amar Shah. Learning to drive in a day. *ICRA*, abs/1807.00412, 2019.

[3] Antonio Loquercio, Ana I. Maqueda, Carlos R. del Blanco, and Davide Scaramuzza. Dronet: Learning to fly by driving. *IEEE Robotics and Automation Letters*, 3(2):1088–1095, 2018.

[4] Anusha Nagabandi, Kurt Konolige, Sergey Levine, and Vikash Kumar. Deep dynamics models for learning dexterous manipulation. *CORL*, abs/1909.11652, 2019.