Mapping Navigation Instructions to Continuous Control Actions with Position-Visitation Prediction

Valts Blukis, Dipendra Misra, Ross A. Knepper and Yoav Artzi

Code and simulator: **https://github.com/clic-lab/drif**

Problem Statement

Goal: map instructions, first-person images and pose estimates to actions for control of a dynamic robotic agent

Existing Approaches

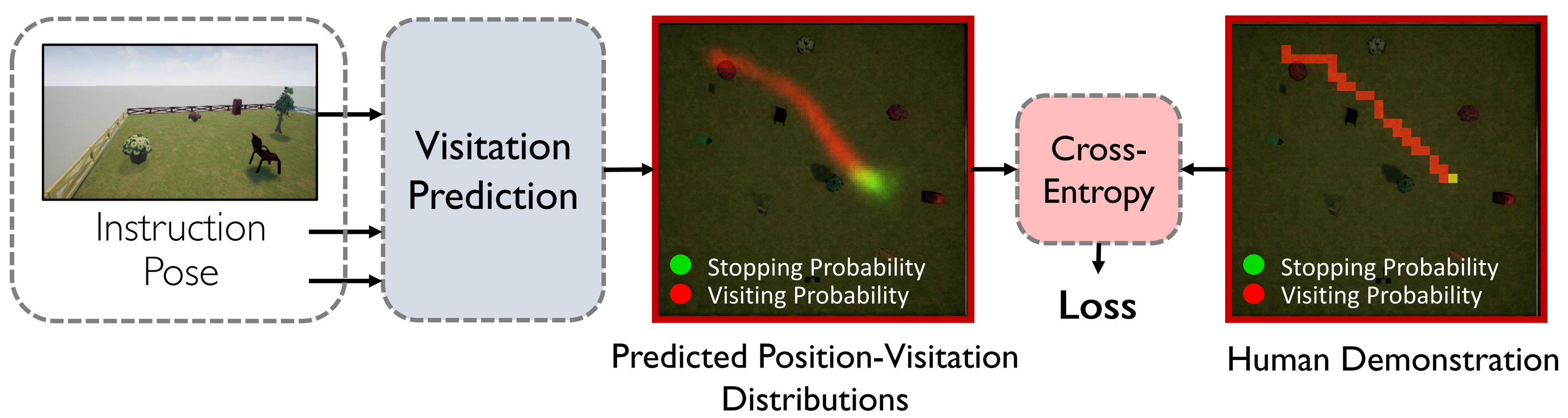
- Decompose problem tn perception, instruction understanding, mapping, planning and control modules. Requires design of intermediate representations. Hard to scale to complex scenarios.
- Single model neural network. Lack interpretability. Require large amount of training data. Complex credit assignment in learning language, control and vision simultaneously.

Our Approach

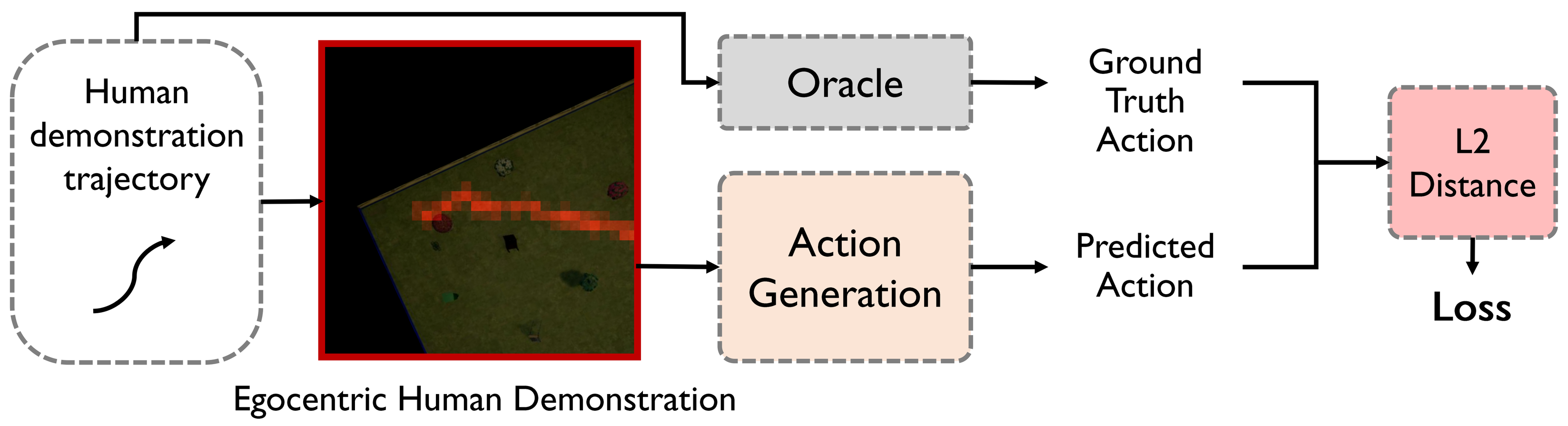
Decompose reasoning and learning in predicting which states to visit (planning) and continuous robot control (action generation).

Learning and Inference

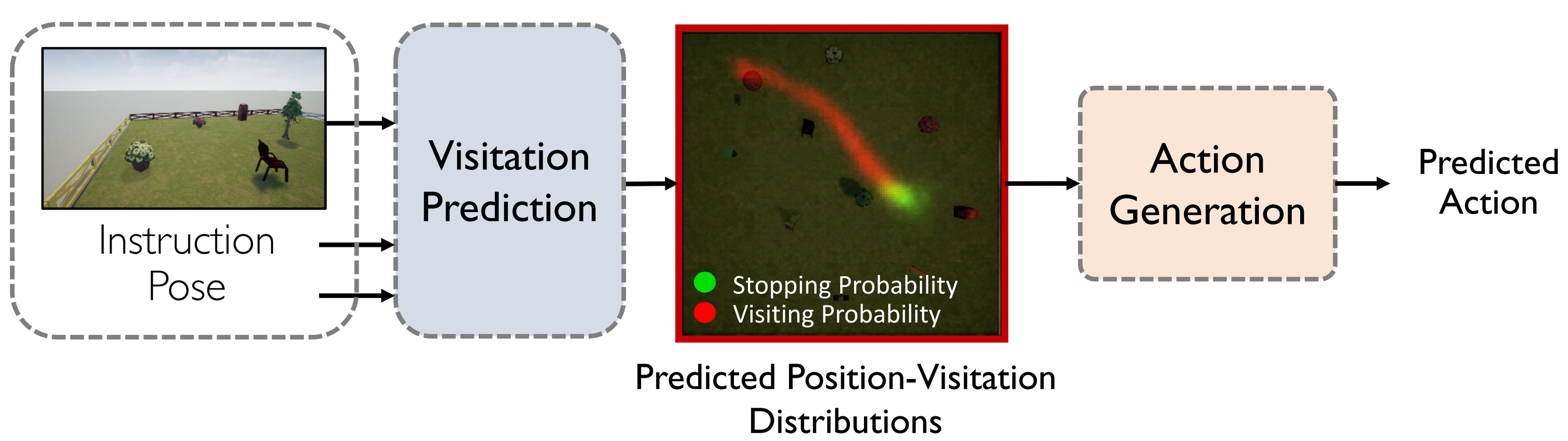
Training Stage A: Supervised learning for visitation prediction



Training Stage B: Imitation learning for action generation



Inference: generate actions from predicted distributions



Reasoning with State-Visitation Distributions

State-Visitation Prediction

1. Predict likely visited states as a state-visitation distribution
2. Generate actions to visit high-probability states

MDP	$\mathcal{S}$ States	$\mathcal{A}$ Actions	R Reward	H Horizon

The **state-visitation distribution** of policy $\pi : \mathcal{S} \rightarrow Pr(\mathcal{A})$ is the probability of visiting state s following policy π from start state s_0 :

$$d_\pi(s|s_0)$$

Predicting a distribution $\hat{d}_\pi(s|s_0)$ over states $\mathcal{S}$ is generally impossible since $\mathcal{S}$ is large.

Solution: predict distribution $\hat{d}_\pi(\tilde{s}|\tilde{s}_0)$ over an approximate state space $\tilde{\mathcal{S}}$, and $\phi : \mathcal{S} \rightarrow \tilde{\mathcal{S}}$ maps states to $\tilde{\mathcal{S}}$.

For a well chosen ϕ, a decomposed policy that follows a state-visitation distribution $\hat{d}_\pi(\tilde{s}|\tilde{s}_0)$ close to $d_\pi(s|s_0)$ of the optimal policy has bounded suboptimality.

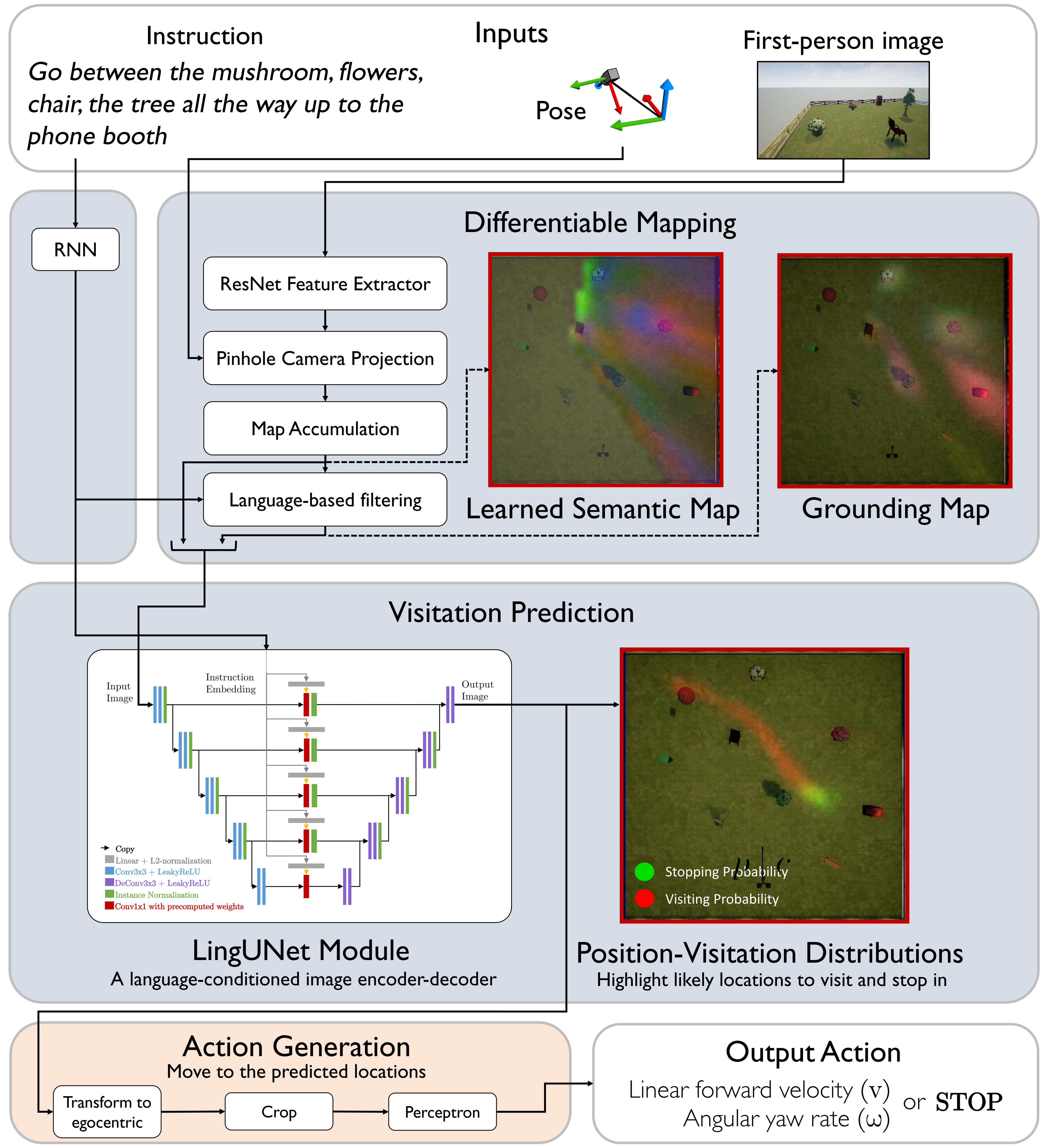
Navigation with Position-Visitation Prediction

$\tilde{\mathcal{S}}$ is the set of discrete **positions** in the world.

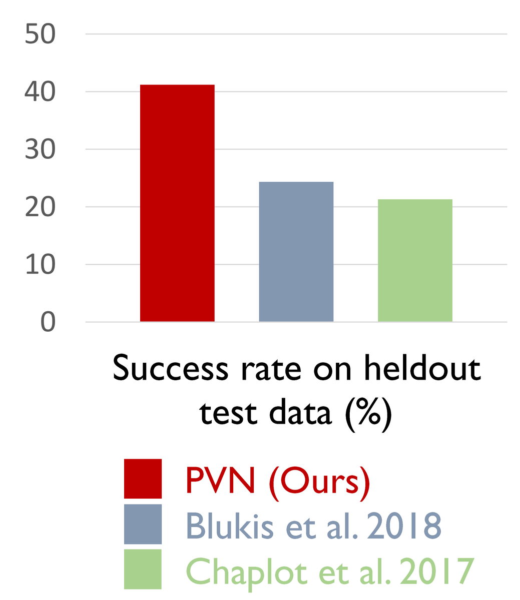
Given context C containing images, poses and an instruction, predict **position-visitation distributions:**

$d^p_\pi(\tilde{p} \mid c, s_0)$ - probability of visiting position $\tilde{p}$.

$d^g_\pi(\tilde{p} \mid c, s_0)$ - probability of stopping at position $\tilde{p}$.



Evaluation and Results



Experimental Setup:

- Realistic, simulated quadcopter powered by Microsoft AirSim.
- Crowdsourced instruction and trajectory data from LANI dataset.

Results:

- **41.21%** instruction following success rate.
- Robust to test-time viewpoint and dynamics changes.
- Full-observability maximum success rate: **60.59%.**

Execution Examples

Engineered instructions from a fixed starting position | Development Set Instructions