**Motivation**

- Vision and Language Navigation – an autonomous agent navigating to a goal location using visual inputs and provided instructions while navigating.
- Navigation without prior global maps
- Generalization to novel environments is a challenge
- Current works enforce unrealistic assumptions i.e. known topology, perfect localization and deterministic navigation

**Robo-VLN Dataset**

- Introduce Robo-VLN - a richer VLN formulation which is defined in continuous environments over long horizon trajectories.
- Robo-VLN provides longer horizon trajectories (4.5x times average number of steps), more visual frames and a balance high-level action distribution compared to discrete VLN settings.
- Robo-VLN computes ground truth oracle feedback controllers in 3D reconstructed environments and obtains navigable instruction-trajectory pairs in continuous environments. The dataset is an extension of VLN-CE/R2R.
- Annotated instructions in the dataset does not describe goals

**Overview**

- a. Continuous VLN
- b. Visual observations
- c. Decoupled reasoning
- d. Hierarchical Imitation

**Architecture**

- The Agent is comprised of a high-level policy and a low-level policy. A layered decision making allows spatially different reasoning at different levels in the hierarchy, hence specializing each policy with a different reasoning abstraction level.

**Experiments**

- We introduced a suit of flat baselines similar to ones used in VLN-CE.
- Sequence to Sequence (Seq2Seq): Encoder-Decoder Architecture
- Cross-Modal Attention (CMA): Aligning instructions with images
- Progress Monitor [1]: Adding auxiliary losses to aid learning
- Flattened hierarchical: Provide sub-goal supervision to flat model

**Hierarchical Cross-Modal Agent**

- Layered two-tiered decision making
- High-level policy performs cross modal reasoning and produces sub-goal
- Low-level policy imitates the controller and translates sub-goal to continuous actions

**Results**

- We lift the agent off the assumptions enforced by discrete action spaces and navigation graph based VLN formulation.
- Provide a suit of baselines in Robo-VLN inspired by recent state of the works in VLN.
- Show that hierarchical approach performs better across all key standard metrics in Robo-VLN.