ForeSI: Success-Aware Visual Navigation Agent: Supplementary Material

Mahdi Kazemi Moghaddam, Ehsan Abbasnejad, Qi Wu, Javen Qinfeng shi and Anton Van Den Hengel
The Australian Institute for Machine Learning
The University of Adelaide
mahdi.kazemimoghaddam, ehsan.abbasnejad, qi.wu01, javen.shi and anton.vandenhengel@adelaide.edu.au

1. Sub-Goal Identification

In Figure 1 [1] we separately show how our agent identifies a sub-goal in hindsight after an episode is completed. The agent takes the actions as described in the main paper until reaching the final state, e.g. either the "STOP" action is selected or the maximum number of permitted actions are exhausted. Then, it selects the state with the maximum attention weight $\boldsymbol{\alpha}_j$ as the sub-goal and adds the tuple of $(\mathbf{s}_0, \mathbf{g}_\tau, \hat{\mathbf{s}}_\tau)$ to the replay buffer. The replay buffer is, then, used to train the sub-goal generation module, as described in the main paper.

Furthermore, note that in our sub-goal selection mechanism we add position embedding to each state representation $\mathbf{s}_i$ to maintain the order of the observed states in the sub-goal selection, inspired by [3].

2. Sub-Goal Identification: Qualitative Results

In Figure 2 we present qualitative results of our sub-goal identification method as described in the main paper. In that figure, random successful trajectories from the unseen test set are visualised. The goal state is shown with a green frame and the selected sub-goal state is shown with a red frame. We can see that the sub-goal state highly correlates with the real goal state. In all those episodes our agent learns to identify the sub-goals that have a few common features: (1) the target object is clearly visible; (2) the goal state is reachable with very few steps; (3) from the sub-goal to the goal state the agent needs to take simple actions, usually just *MoveAhead*.

3. What to Imagine

In Table 1 we present more detailed ablation study results. The details of the methods in Table 1 are presented in the main paper; here of special interest is the performance on longer trajectories. We can see that in all the methods compared in Table 1 the performance trend for longer trajectories follows the trend for the short ones, except when we compare **Ours-ATT** with **Ours-ForeSI**. Specifically, **Ours-ATT** performs better than **Ours-ForeSI** on longer trajectories. We believe that this is mainly because it can be more helpful for longer trajectories to imagine a combination of the future states to arrive at rather than a single state that may require a long trajectory to take. In contrast, for shorter trajectories, a single fully identifiable state imagination is more helpful.

[1] Note the trajectories are sub-sampled for visualisation purposes.

4. Explicitly Structured Imagination

As briefly described in the main paper, we compare the imagination capability of our simple method with a more complex reconstruction-based method. We use a Conditional Variational Auto-Encoder (C-VAE) [2] for this experiment. Our C-VAE consists of 6 fully-connected layers with the following number of neurons, respectively: [512+64, 512, 256, 128, 32+22, 256, 512, 512] with a latent dimension of 32. We use $ReLU$ non-linearity on all layers except the last layer for which we use $Tanh$. We use $1e^{-4}$ as the KL divergence regularisation rate.

5. Qualitative Results

In this section, we provide further key insights into how our agent performs in different navigation scenarios and what its main strengths and some weaknesses are hoping to encourage future research for further improvements.

In Figure 4 we compare two sample trajectories between our method and the baseline method [4]. In Figure 4 (a) we observe that forward modelling helps our agent identify the target object while the baseline method disregards the observed target and stops at a random location after a few

Method	SPL	SR	SPL>5	SR>5
A3C+ORG [1]	37.5	65.3	36.1	54.8
Ours-RND	37.57	64.8	35.13	53.07
Ours-INT	37.78	63.8	35.0	53.0
Ours-ATT	37.76	65.4	**38.25**	**56.54**
Ours-ForeSI	**38.66**	**67.6**	36.85	56,11

Table 1. Various ablation studies for our ForeSI.

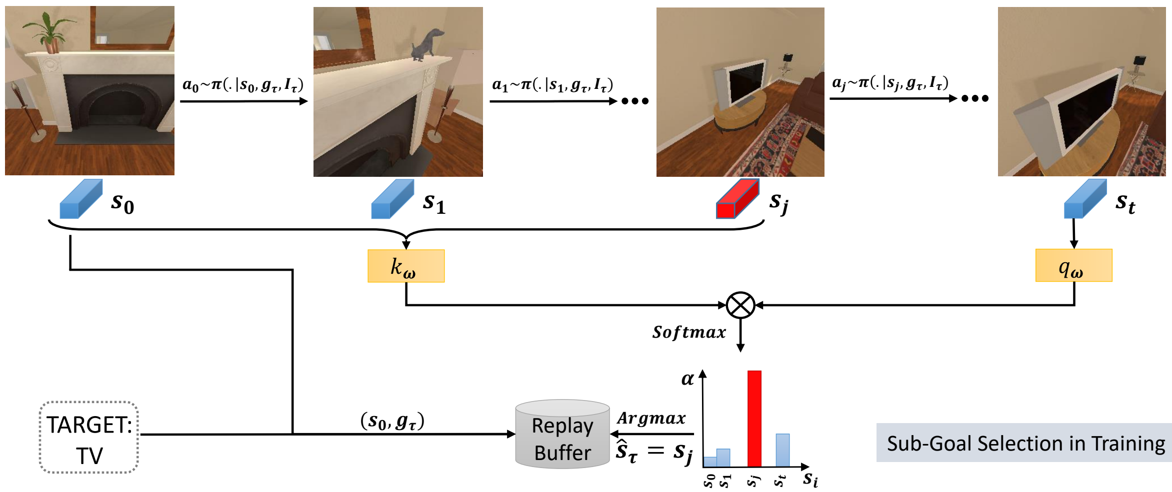

Figure 1. Overview of our sub-goal selection/ identification method showing how our agent selects a sub-goal state to fill the replay buffer during the training; the replay buffer is then used to train our sub-goal generation module as described in the main paper.

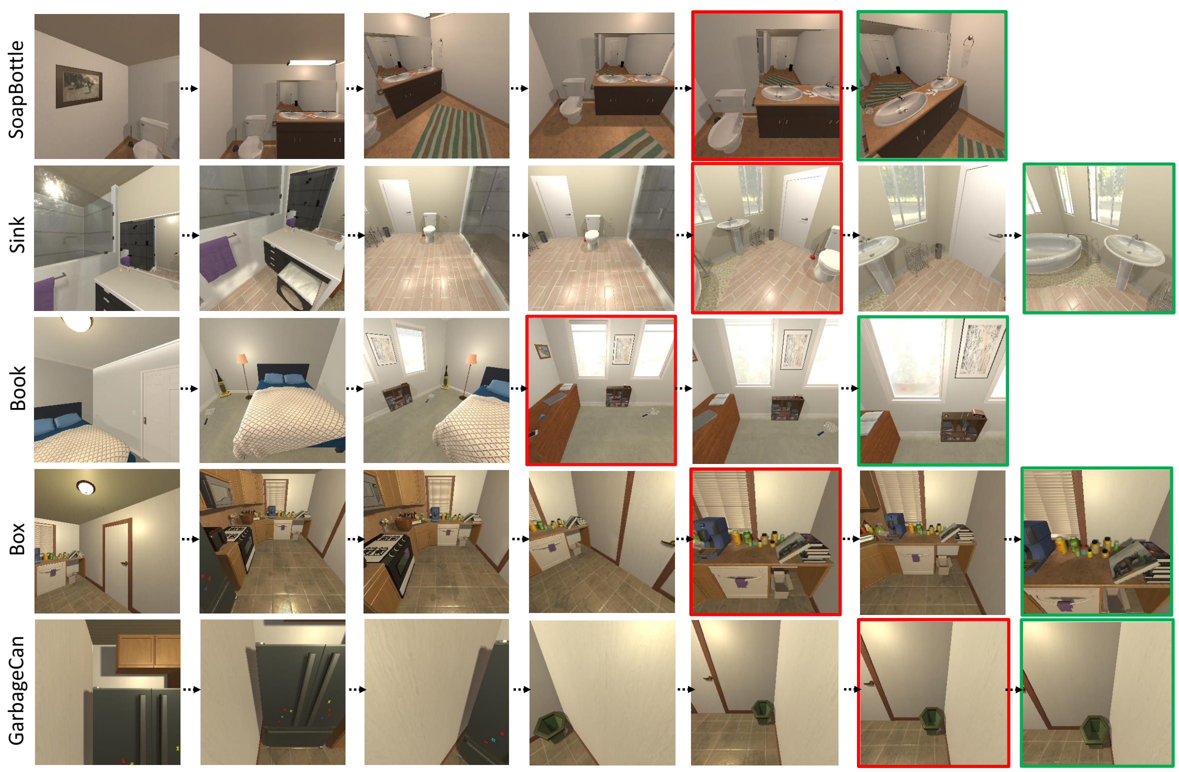

Figure 2. Sample trajectories from the unseen test set showing sub-goal selection results using our method. The **red** frame shows the selected sub-goal state and the **green** frame the final successful goal state.

searching steps. While being eventually successful, here our agent misses the target once and returns to a similar state again. This might be due to the fact that our sub-goal selection method might not have observed the target in the very first few steps and thus can still be further improved. A similar scenario can be observed in Figure 4 (b), where the laptop is visible in the agent's observation but the agent has to take a few extra steps and return to the laptop again. This has been observed multiple other times in the test trajectories. Despite the improvement that our method presents in SPL 4 provides more insights into how we might further improve our sub-goal selection.

In Figure 3 we present two sample trajectories that show how our agent is able to take the shortest path to the target object. In contrary, the baseline method [4] either dismisses the target object and takes the wrong trajectory in Figure 3 (a), or cannot stop at the right location in Figure 3 (b) and fails. Those two samples show that our method has the potential to address the two major problems with the current previous state-of-the-art methods.

In Figure 5 we present sample failure cases of our method where the baseline agent is able to complete the trajectories successfully. In Figure 5 (a), our agent fails to detect the "alarm clock" despite being visible in the first few state observations and instead moves towards the place where a bedside is usually located. This can be due to the fact that during the training the agent has mostly observed the "alarm clock" closer to the bed hence our agent imagines a state closer to the bed. In Figure 5 (b), although our agent takes the correct trajectory that leads to the target, "plant", it fails to stop close enough. On the contrary, the baseline method [4] stops within the 1-meter proximity and is successful.

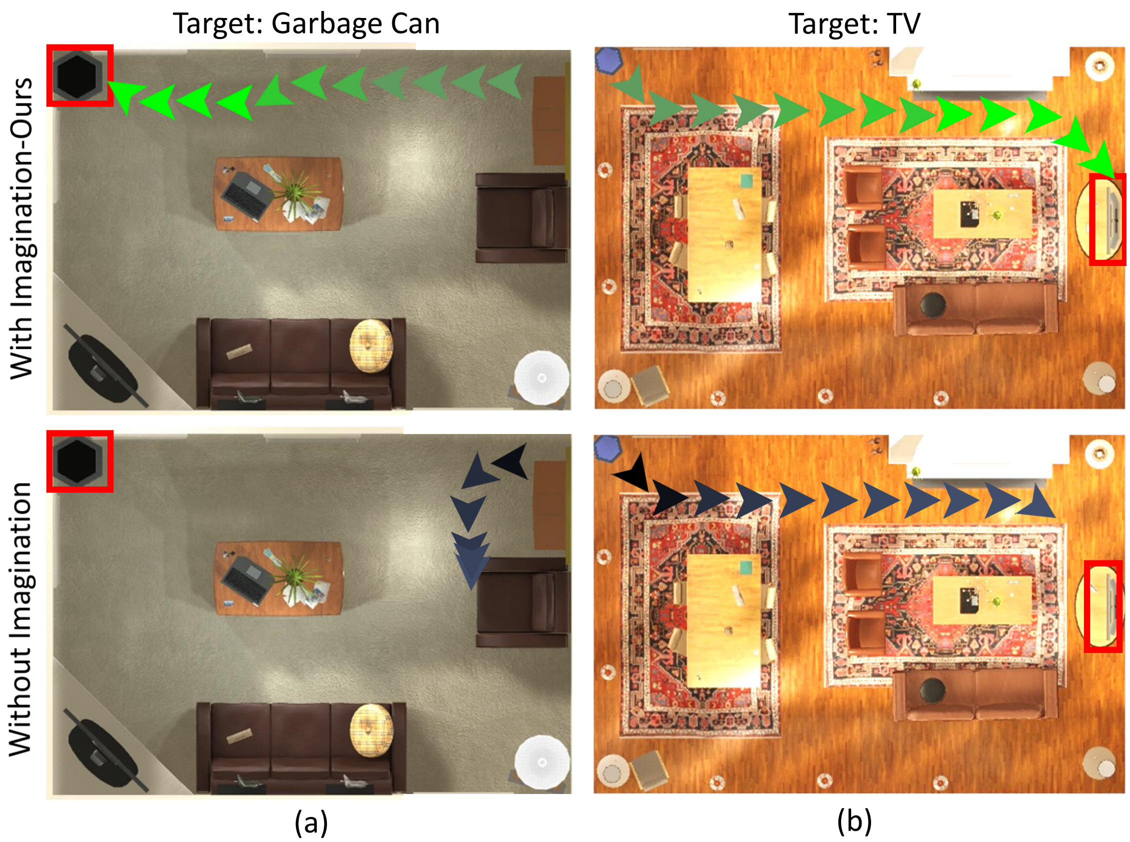

Figure 3. Our ForeSI agent has learnt to imagine the optimal sub-goal hence it takes a near-optimal trajectory while the baseline method fails.

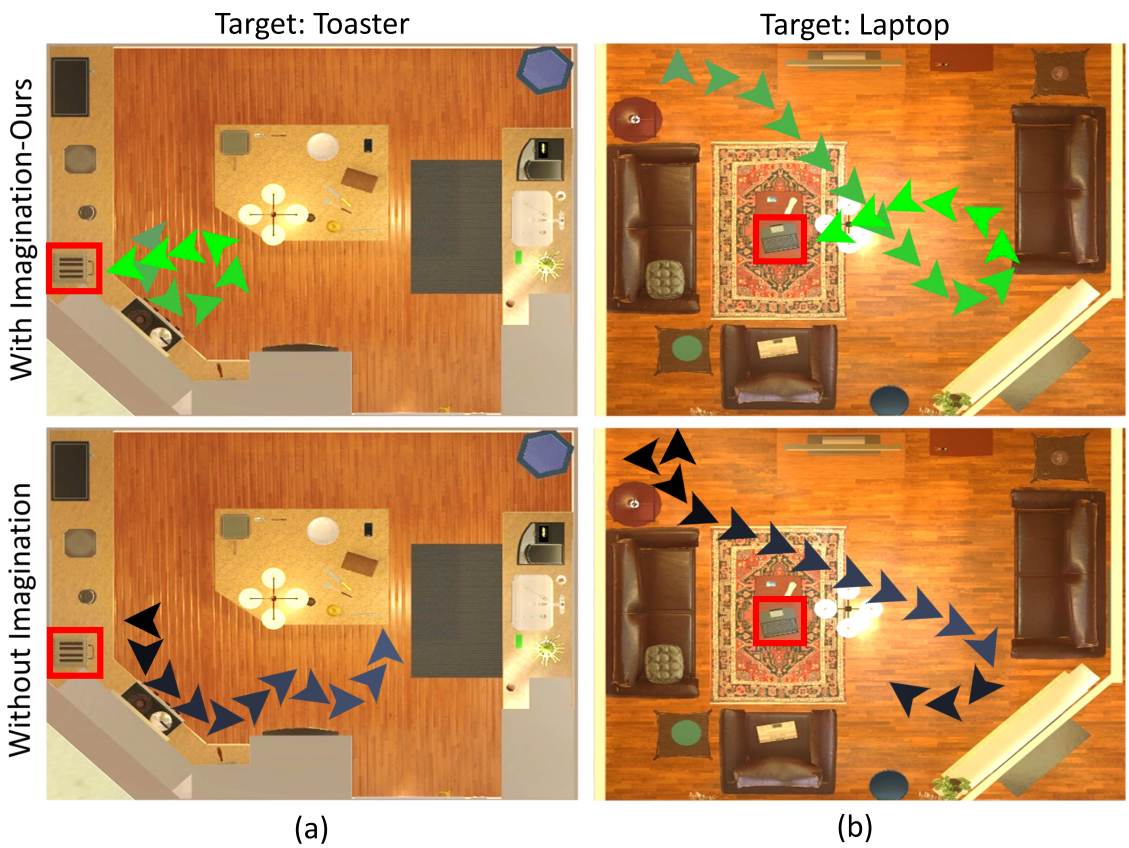

Figure 4. Although our agent achieves an overall higher success rate, the length of the trajectory could still be more optimal.

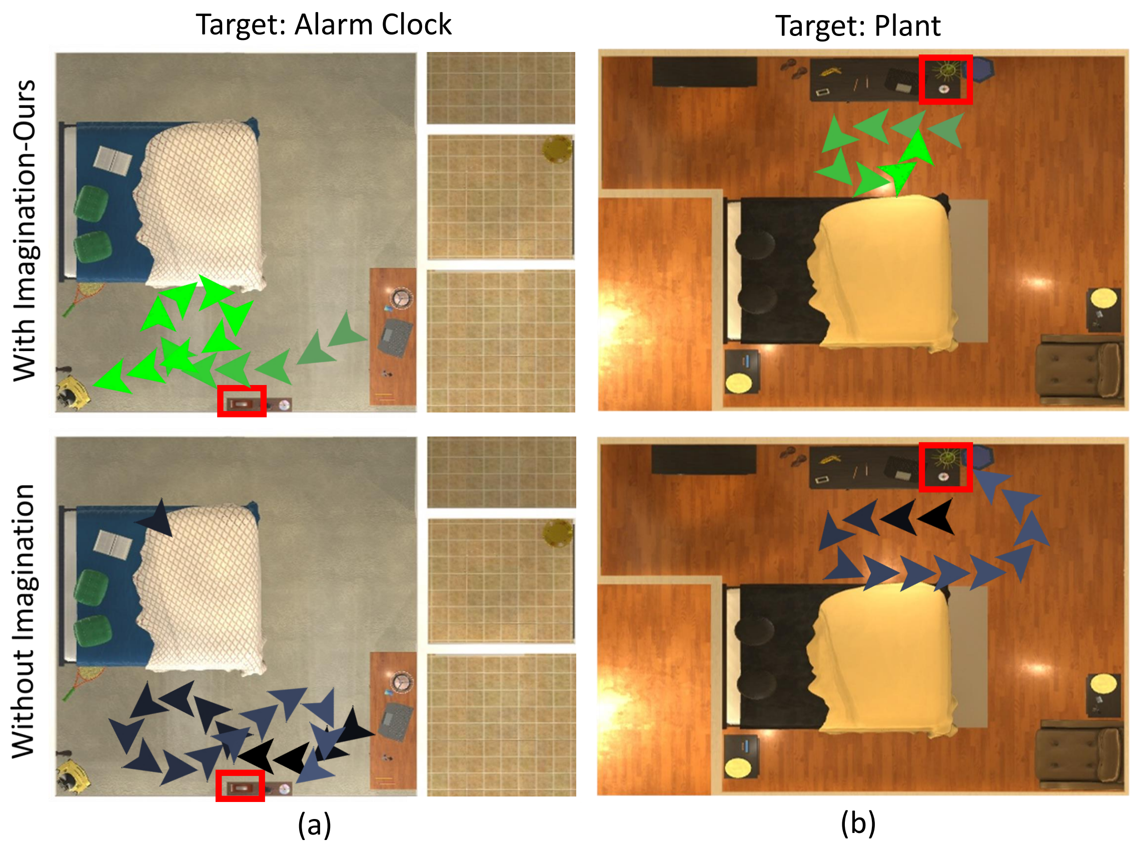

Figure 5. Failure cases of our method; **(a)** shows a sample failure that might happen due to imagining the wrong state (here a bedside near the bed where the alarm clock normally is located) and **(b)** happens due to early stopping.

References

[1] Heming Du, Xin Yu, and Liang Zheng. Learning object relation graph and tentative policy for visual navigation, 2020.

[2] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *Advances in neural information processing systems*, pages 3483–3491, 2015.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[4] Mitchell Wortsman, Kiana Ehsani, Mohammad Rastegari, Ali Farhadi, and Roozbeh Mottaghi. Learning to learn how to learn: Self-adaptive visual navigation using meta-learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6750–6759, 2019.