

Supplementary Materials for Learning to Terminate in Object Navigation

Yuhang Song *

*Department of Computer Science
University of Liverpool, UK*

SGYSON10@LIVERPOOL.AC.UK

Anh Nguyen

*Department of Computer Science
University of Liverpool, UK*

ANH.NGUYEN@LIVERPOOL.AC.UK

Chun-Yi Lee

Department of Computer Science National Tsing Hua University,

CYLEE@CS.NTHU.EDU.TW

1. Judge Model Prediction Analyse

```
task data ['DeskLamp|-01.31|-01.23|-01.00'] -----
{'action': 'LookDown'}|0.0106
*****nb_out_train tensor([[0.6099, 0.3901],
 [0.5985, 0.4015],
 [0.6225, 0.3775],
 [0.5623, 0.4377],
 [0.6225, 0.3775],
 [0.5623, 0.4377],
 [0.6552, 0.3448],
 [0.6306, 0.3694],
 [0.5725, 0.4275],
 [0.6302, 0.3698],
 [0.1645, 0.8355],
 [0.1645, 0.8355],
 [0.3014, 0.6986],
 [0.3244, 0.6756],
 [0.3244, 0.6756],
 [0.3014, 0.6986],
 [0.3014, 0.6986],
 [0.3014, 0.6986],
 [0.3014, 0.6986],
 [0.1645, 0.8355],
 [0.5636, 0.4370],
 [0.5565, 0.4435],
 [0.5565, 0.4435],
 [0.5565, 0.4435],
 [0.4416, 0.5584],
 [0.6861, 0.3939],
 [0.6495, 0.3505],
 [0.6495, 0.3505],
 [0.6021, 0.3979],
 [0.1014, 0.8986],
 [0.4060, 0.5940]], device='cuda:0', grad_fn=<SoftmaxBackward0>)
*****nb_true tensor([0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0,
 0, 0, 0, 0, 0, 0, 1, 1], device='cuda:0')
```

Figure 1: Judge Model Prediction. An example of judge model inference after 2000 updates during training, with ground truth and predictions. Where the ground truth 0 stands for time steps that should terminate.

2. Judge Model Prediction Analyse

Given that our RL branch aligns with the baseline model and our contribution lies in the introduction of the judge model, we drew direct comparisons with the baseline. An ablation study on the essential components is detailed in Table 1.

* The author is affiliated with the Department of Computer Science, National Tsing Hua University as well in a dual Ph.D. program.

Method	w/ Transformer		w/o Focal Loss		w/o Judge Model		DITA (Ours)	
	SR(%)	SPL(%)	SR(%)	SPL(%)	SR(%)	SPL(%)	SR(%)	SPL(%)
All	55.6	15.8	58.5	20.1	65.3	21.1	71.4	21.6
L \geq 5	36.8	15.5	38.4	17.6	50.5	20.9	57.9	22.2

Table 1: An ablation study on the judge model. In this table, ‘w/ Transformer’ denotes a variant that employs a Transformer as the feature encoder. On the other hand, in the variant ‘w/o Focal Loss,’ we replaced the focal loss in the judge model with cross-entropy loss for comparison.