Supplementary Material: Room-and-Object Aware Knowledge Reasoning for Remote Embodied Referring Expression

Chen Gao[1]*, Jinyu Chen[1]*, Si Liu[1]†, Luting Wang[1], Qiong Zhang[3], Qi Wu[2]
[1]Institute of Artificial Intelligence, Beihang University
[2]The University of Adelaide [3]Xiaomi AI Lab, Xiaomi Inc

1. Quantitative Results

Direction-Aware Loss (DAL). We conduct experiments about the weighting factor λ_{dir} to analyse the effect of DAL, which is shown in Tab. 1. On val-seen, varying λ_{dir} from 2 to 10 does not have great impact. Setting $\lambda_{dir} = 5$ achieves obvious improvement (SPL from 51.79% to 56.02%). More importantly, on val-unseen, comparing to without-DAL (*i.e.*, $\lambda_{dir} = 0$), setting λ_{dir} to 10 significantly increases SPL from 8.11% to 11.05% and RSR from 7.91% to 10.04%. Note that TL also decreases, which further illustrates our model benefits from the DAL, resulting in a shorter trajectory.

Distance-aware Policy. An evaluation of our proposed distance-aware policy is shown in Tab. 2. For this experiment, we vary hyperparameter w from 0 to 10, where $w = 0$ means not applying this policy. We observe that the policy continuously reduces TL as w increases (more obvious on val-unseen). On val-unseen, when $w = 2$, TL significantly drops from 37.09m to 22.37m. Though SR also decreases, the decline (from 19.91% to 17.30%) is smaller than that of TL, leading to a 1.42% SPL improvement. However, as w gets larger ($w = 10$), the decrease rate of SR becomes greater than TL, thus SPL starts to descend. On val-seen, since TL is already short, adjusting w does not affect SR, TL, or SPL to a great extent.

2. Qualitative Results

Navigation Visualisation. More navigation visualisations are shown in Fig. 1, Fig. 2, Fig. 3, Fig. 4 and Fig. 5.

Firstly, the visualisation of our CKR model on four viewpoints are shown in Fig. 1, which demonstrate our model can make proper decisions in various scenarios.

Secondly, two navigation samples of our CKR model are shown in Fig. 2 and Fig. 3, which verify our model conducts reasonable goal-oriented exploration.

Thirdly, the comparisons between our proposed CKR model and navigator-pointer [1] model are shown in Fig. 4 and Fig. 5, where [1] fails and the CKR successes. These results demonstrate the superiority our model.

Object-Entity Reasoning. We visualise more about the top-10 relevant categories in ConceptNet and learned model (in Tab. 3) to examine the necessity about applying internal

λ_{dir}	Val-Seen			Val-Unseen		
	SPL↑	TL↓	RSR↑	SPL↑	TL↓	RSR↑
0	51.79	12.58	34.14	8.11	17.60	7.91
2	53.02	**11.57**	40.34	**11.90**	**14.59**	8.80
5	**56.02**	11.86	39.26	9.76	14.90	8.87
10	51.39	12.39	36.24	11.05	16.36	**10.04**

Table 1. **Direction-Aware Loss (DAL)**: Results with various weighting factor λ_{dir}, illustrating the DAL consistently improve the performance.

w	Val-Seen			Val-Unseen		
	SR↑	SPL↑	TL↓	SR↑	SPL↑	TL↓
0	57.27	53.14	12.66	**19.91**	10.56	37.09
0.5	57.13	53.40	12.24	19.37	11.13	30.26
1	57.27	53.57	12.16	19.14	11.84	26.26
2	57.41	**53.67**	11.99	17.30	**11.98**	22.37
10	**57.48**	53.71	**11.85**	16.56	11.89	**20.35**

Table 2. **Distance-aware Policy**: On val-unseen, the policy significantly reduces TL and slightly hurts SR, leading to higher SPL.

Category	Most relevant categories	
	ConceptNet	Learned
map	street sign, instructions, ...	bed, couch, blinds ...
outlet	drain	chandelier
log	woods	chairs
word	roman numeral	picture
design	lettering, sculpture, structure	couch, chairs, phone
scale	numbers, ramp, foot	bed, bathroom, table
stick	rod	vase
rose	bud, ...	bed, ...

Table 3. Illustration of the difference between general-level commonsense and domain-specific learned knowledge.

*Equal contribution
†Corresponding author (liusi@buaa.edu.cn)

Figure 1. Visualisation of the agent behaviours on four viewpoints.

Figure 2. Visualisation of the agent behaviours on a trajectory.

KG reasoning to conduct domain-specific knowledge reasoning. For example, the 'rose' category is related to 'bud' in general (ConceptNet), which is not useful for the specific REVERIE task. However, after training with internal KG reasoning, 'rose' is related to 'bed', where 'rose' is usually nearby as a decoration. These results demonstrate the indoor domain knowledge is effectively learned.

Failure Cases. We visualise a failure case in Fig. 6. The agent stops at a wrong room, which is the same as described in the instruction.

Figure 3. Visualisation of the agent behaviours on a trajectory.

References

[1] Yuankai Qi, Qi Wu, Peter Anderson, Xin Wang, William Yang Wang, Chunhua Shen, and Anton van den Hengel. Reverie: Remote embodied visual referring expression in real indoor environments. In *CVPR*, 2020. 1, 3, 4

Figure 4. Navigation samples of [1] and our method, where [1] fails and our method successes.

Figure 5. Navigation samples of [1] and our method, where [1] fails and our method successes.

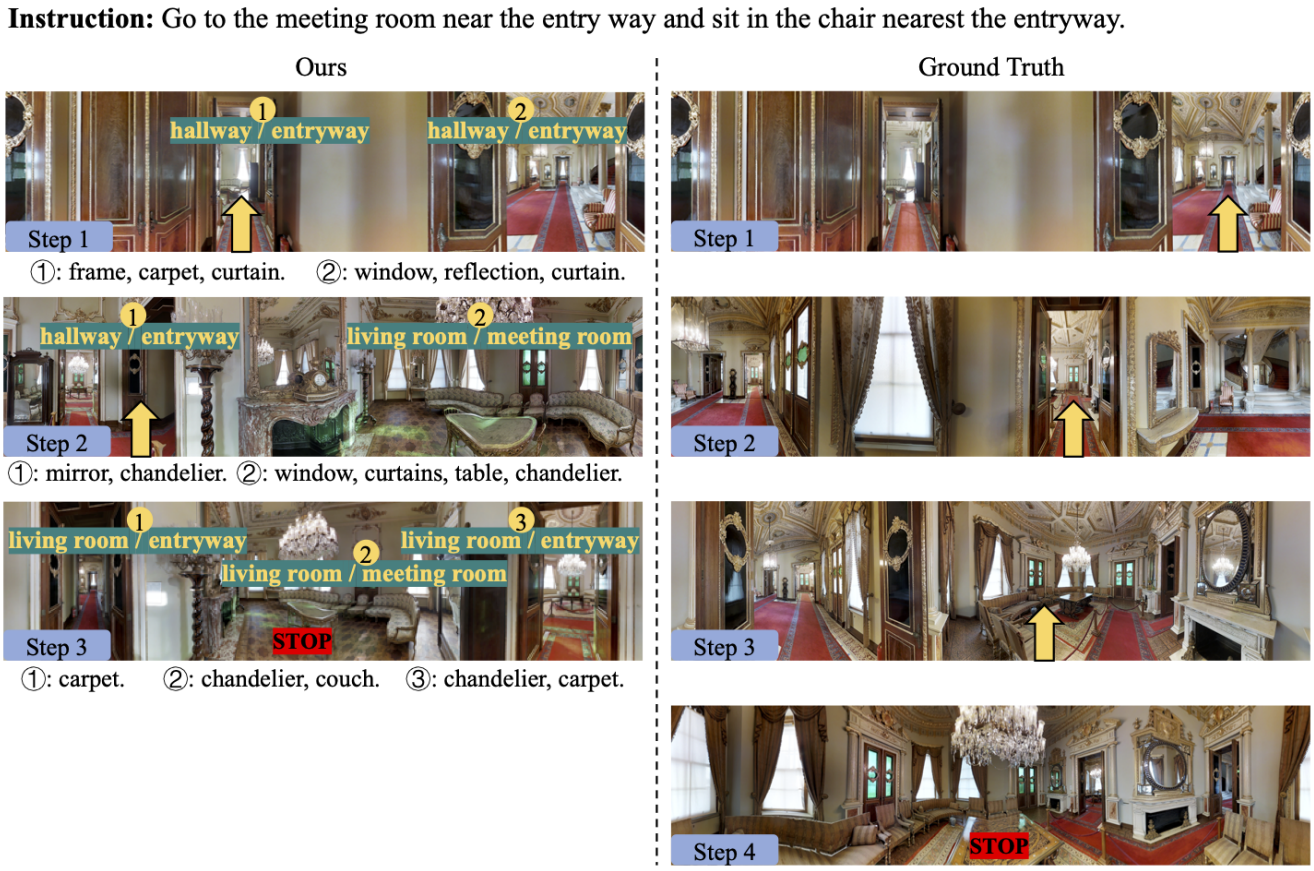

Figure 6. Failure case.