International Journal of Computer Vision (2023) 131:607–625
https://doi.org/10.1007/s11263-022-01721-6

Active Perception for Visual-Language Navigation

Hanqing Wang[1] · Wenguan Wang[2] · Wei Liang[1] · Steven C. H. Hoi[3] · Jianbing Shen[4] · Luc Van Gool[5]

Received: 11 October 2021 / Accepted: 7 November 2022 / Published online: 3 December 2022

Abstract
Visual-language navigation (VLN) is the task of entailing an agent to carry out navigational instructions inside photo-realistic environments. One of the key challenges in VLN is how to conduct robust navigation by mitigating the uncertainty caused by ambiguous instructions and insufficient observation of the environment. Agents trained by current approaches typically suffer from this and would consequently struggle to take navigation actions at every step. In contrast, when humans face such a challenge, we can still maintain robust navigation by actively exploring the surroundings to gather more information and thus make a more confident navigation decision. This work draws inspiration from human navigation behavior and endows an agent with an active perception ability for more intelligent navigation. To achieve this, we propose an end-to-end framework for learning an exploration policy that decides (i) when and where to explore, (ii) what information is worth gathering during exploration, and (iii) how to adjust the navigation decision after the exploration. In this way, the agent is able to turn its past experiences as well as new explored knowledge to contexts for more confident navigation decision making. In addition, an external memory is used to explicitly store the visited visual environments and thus allows the agent to adopt a late action-taking strategy to avoid duplicate exploration and navigation movements. Our experimental results on two standard benchmark datasets show promising exploration strategies emerged from training, which leads to significant boost in navigation performance.

Keywords Visual-language navigation · Active perception · Curriculum reinforcement learning

Communicated by Wenguan Wang and Wei Liang.

A preliminary version of this work has appeared in ECCV 2020 Wang et al. (2020b). Our algorithm implementations are available at https://github.com/HanqingWangAI/Active_VLN.

✉ Jianbing Shen
shenjianbingcg@gmail.com

Hanqing Wang
hanqingwang@bit.edu.cn

Wenguan Wang
wenguanwang.ai@gmail.com

Wei Liang
liangwei@bit.edu.cn

Steven C. H. Hoi
stevenhoi@gmail.com

Luc Van Gool
vangool@vision.ee.ethz.ch

[1] Beijing Institute of Technology, Beijing, China

[2] ReLER Lab, Australian Artificial Intelligence Institute, University of Technology Sydney, Sydney, Australia

[3] Salesforce Research Asia, Singapore, Singapore

[4] The State Key Laboratory of Internet of Things for Smart City, Department of Computer and Information Science, University of Macau, Macau, China

[5] ETH Zurich, Zurich, Switzerland