Look and Answer the Question: On the Role of Vision in Embodied Question Answering

Nikolai Ilinykh and Yasmeen Emampoor and Simon Dobnik

Centre for Linguistic Theory and Studies in Probability
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg, Sweden
nikolai.ilinykh@gu.se, gusemampya@student.gu.se,
simon.dobnik@gu.se

INLG 2022

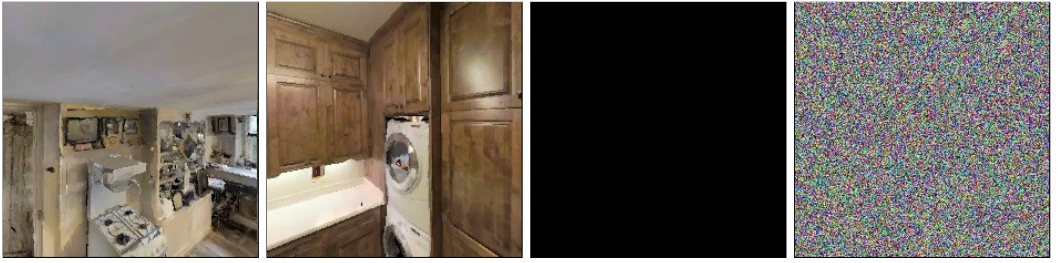
Embodied Question Answering and importance of vision

GÖTEBORGS UNIVERSITET

We take the task of EQA, in which an agent has to navigate
in the house environment, find the target object and answer a question about it.

Our goal is to see how much vision is needed and used by the model to answer the questions.
We challenge the model and permute its vision.
Two experiments: (i) is vision needed for the task? (ii) how much is learned from vision?

What are the "visual permutations"?



Some of the results:

- The model learns low-level visual information and struggles to capture more fine-grained deeper visual understanding
  - Why so? Many problems with the images, questions, models...

Some of the results:

GÖTEBORGS UNIVERSITET

- The model learns low-level visual information and struggles to capture more fine-grained deeper visual understanding
  - Why so? Many problems with the images, questions, models...

- Overall, EQA is an interesting task that is highly different from the established captioning and VQA tasks and therefore introduces novel challenges related to language use in interactive environments

GÖTEBORGS UNIVERSITET

For more, visit our poster =)