

1. Do you think the framing of the problem (the inputs it needs) is realistic? Why or why not?
2. The proposed approach uses the current image as its state. Is there any information that this state representation does not capture?
3. The proposed approach takes the current image and the target image as inputs and outputs the action to take. What might the model be learning to be able to do this?
4. Parts of the network need to be retrained for each new scene. Do you think this is practical? Why or why not?
5. How would you change the model if a blueprint of the house is also available?