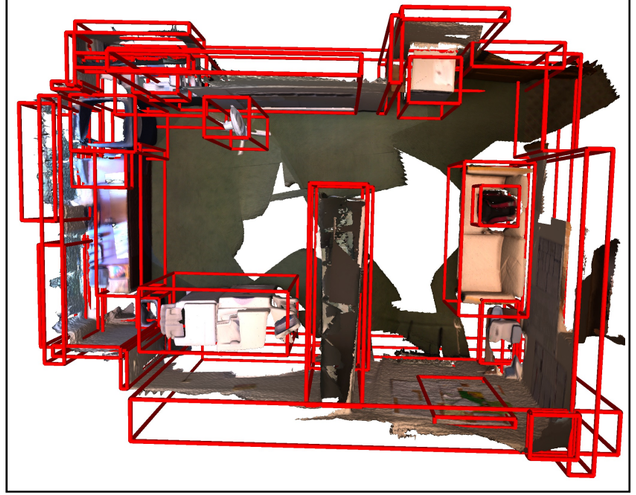
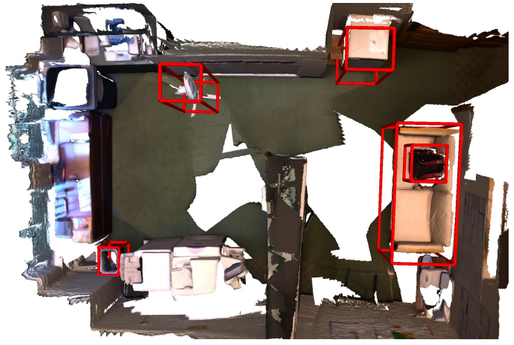
Step1: Scene Graph Construction



Step2: Object Selection



1) Common Object
2) Non-trivial Object
3) Unambiguous object

Step3: Text Generation

Scene type: *Living room*
Selected objects: [*fan*, *backpack, couch*]

↓

Question: "*…The intention is from a first-person perspective… avoid mentioning…. In a living room with fan, backpack, couch, what can you do with each object?"*

↓

Answer:
"[Fan]: I intend to circulate air to ……
[Backpack]: I plan to carry my ……
[Couch]: I want to relax …… "

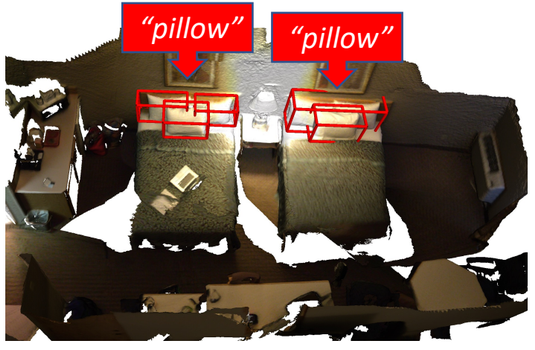
Step4: Data Cleaning

1. Remove garbled text.
2. Remove corrupted characters.
3. Regenerate ambuiguous cases
4. Regenerate repeated cases
5. ……

Target : "*towel***", Number of Targets:** 1
Intention: "*I plan to dry myself after taking a shower or washing my hands.***"**

Target : "*trash can***", Number of Targets: 5**
Intention: "*I intend to dispose of waste materials to keep the hallway tidy.***"**

Target : "*stair rail***", Number of Targets: 2**
Intention: "*I need to steady myself as I go up to the second floor.***"**



Target : "*pillow***", Number of Targets: 5**
Intention: "*I prefer to use it for my head during sleep.***"**