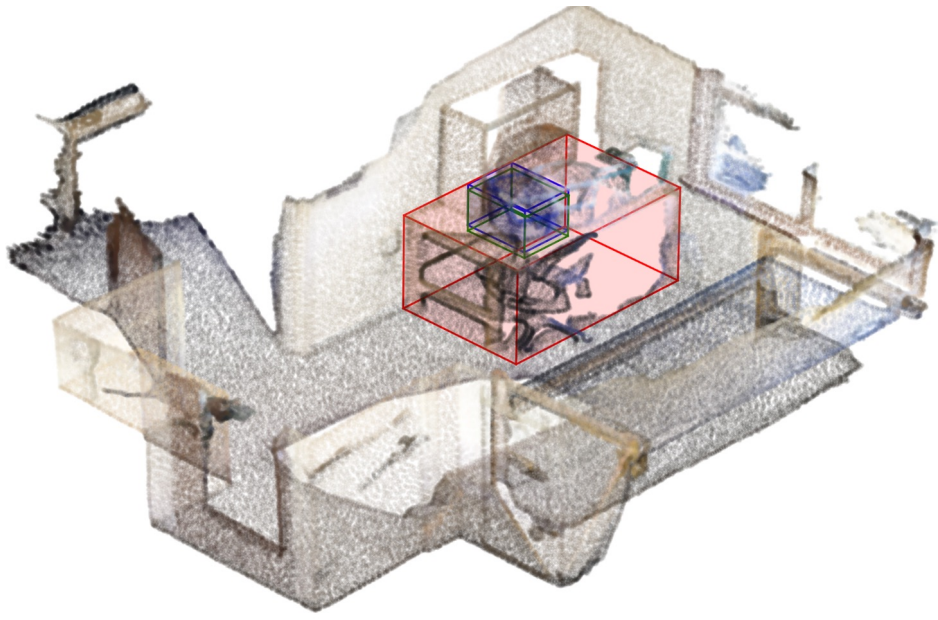


Visual Grounding

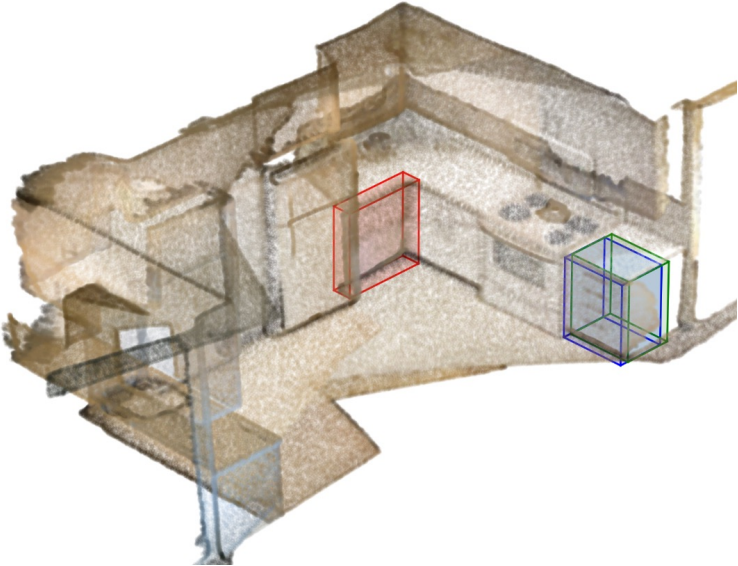


Question: Identify the object according to the following description. A small blue cabinet. there is a chair and bed near it.

GT: Blue box.

Baseline: Red box.

Ours: Green box.



Question: Identify the object according to the following description. There is a rectangular white kitchen cabinet. it is to the right of a stove.

GT: Blue box.

Baseline: Red box.

Ours: Green box.

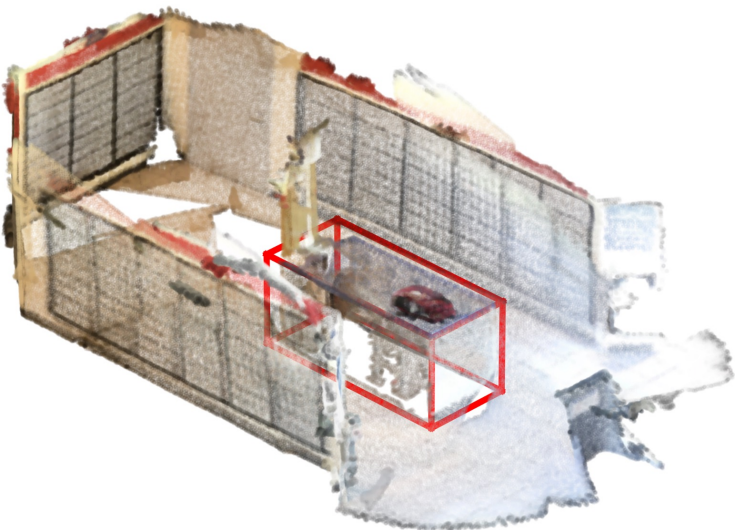
Object Captioning



GT: This is a rectangular ottoman. The ottoman is in front of the sofa.

Baseline: This is a round end table. It is to the left of the couch. (**CIDER: 0.01**)

Ours: This is a brown ottoman. It is in front of a sofa. (**CIDER: 3.52**)



GT: The cabinet is in the middle of the room. The cabinet is a white box with a blue rectangular top.

Baseline: This is a long counter. it is behind a backpack. (**CIDER: 0.00**)

Ours: The cabinet is in the center of the room. the cabinet is a white rectangular prism. (**CIDER: 3.07**)

Question Answering

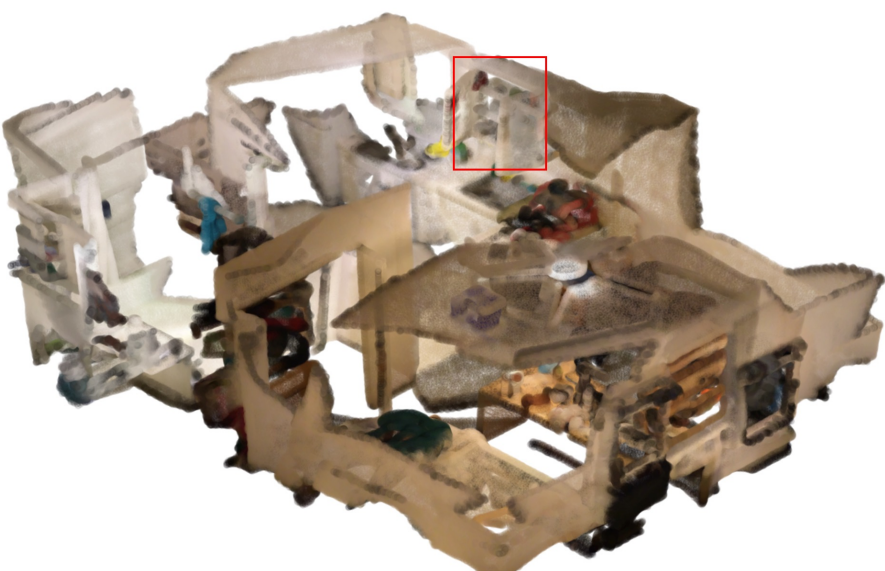


Question: How many windows? Answer the question simply.

GT: 4.

Baseline: 2.

Ours: 4.



Question: How many levels does the shelf have? Answer the question simply.

GT: 3.

Baseline: 6.

Ours: 3.