

Query Text: "this is a brown leather chair located under the first dark brown table immediately inside the door"

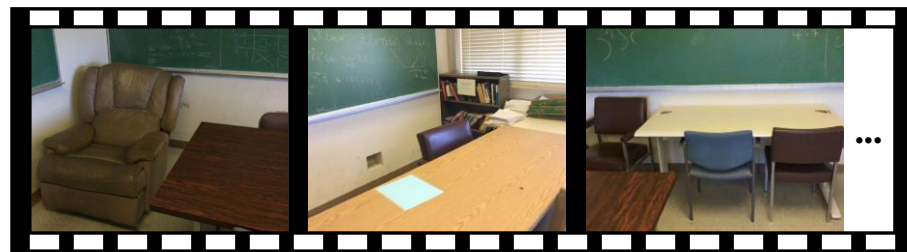
Object 1 is a chair at (x_1, y_1, z_1) ...
Object 2 is a table at (x_2, y_2, z_2) ...

3D-based Model

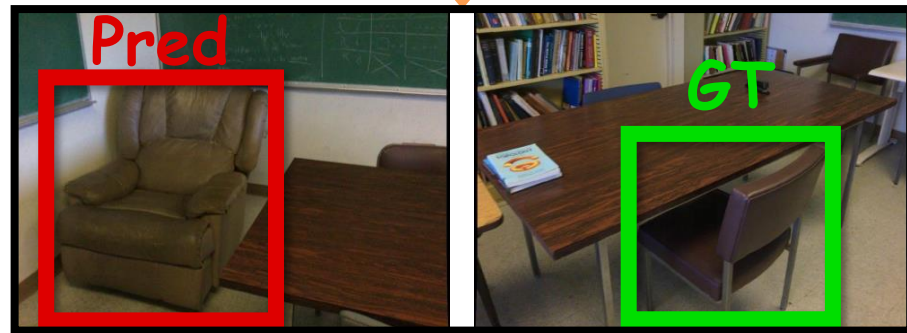


Location: ✓ Attribute: ✗

(a) 3D-based Method



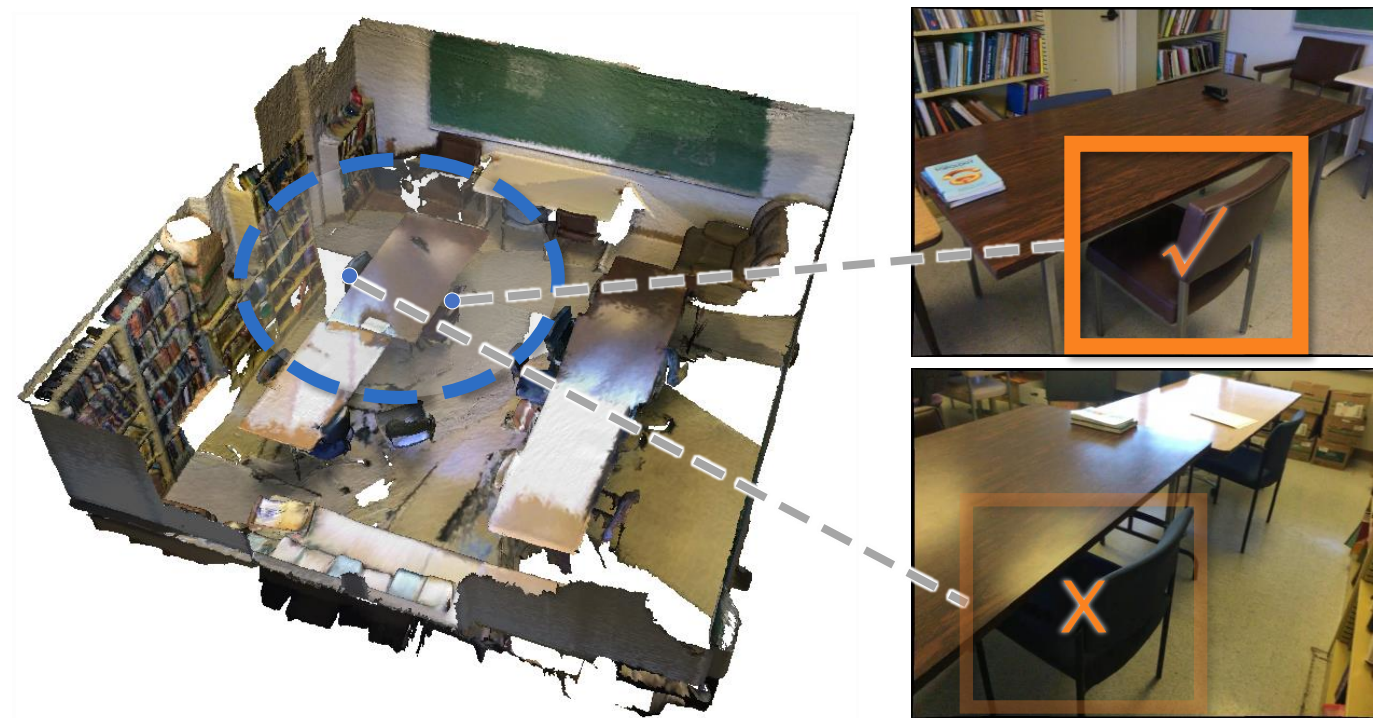
2D-based Model



Location: ✗ Attribute: ✓

(b) 2D-based Method

Spatial Localization → Semantic Verification



Location: ✓ Attribute: ✓

(c) Our proposed SPAZER