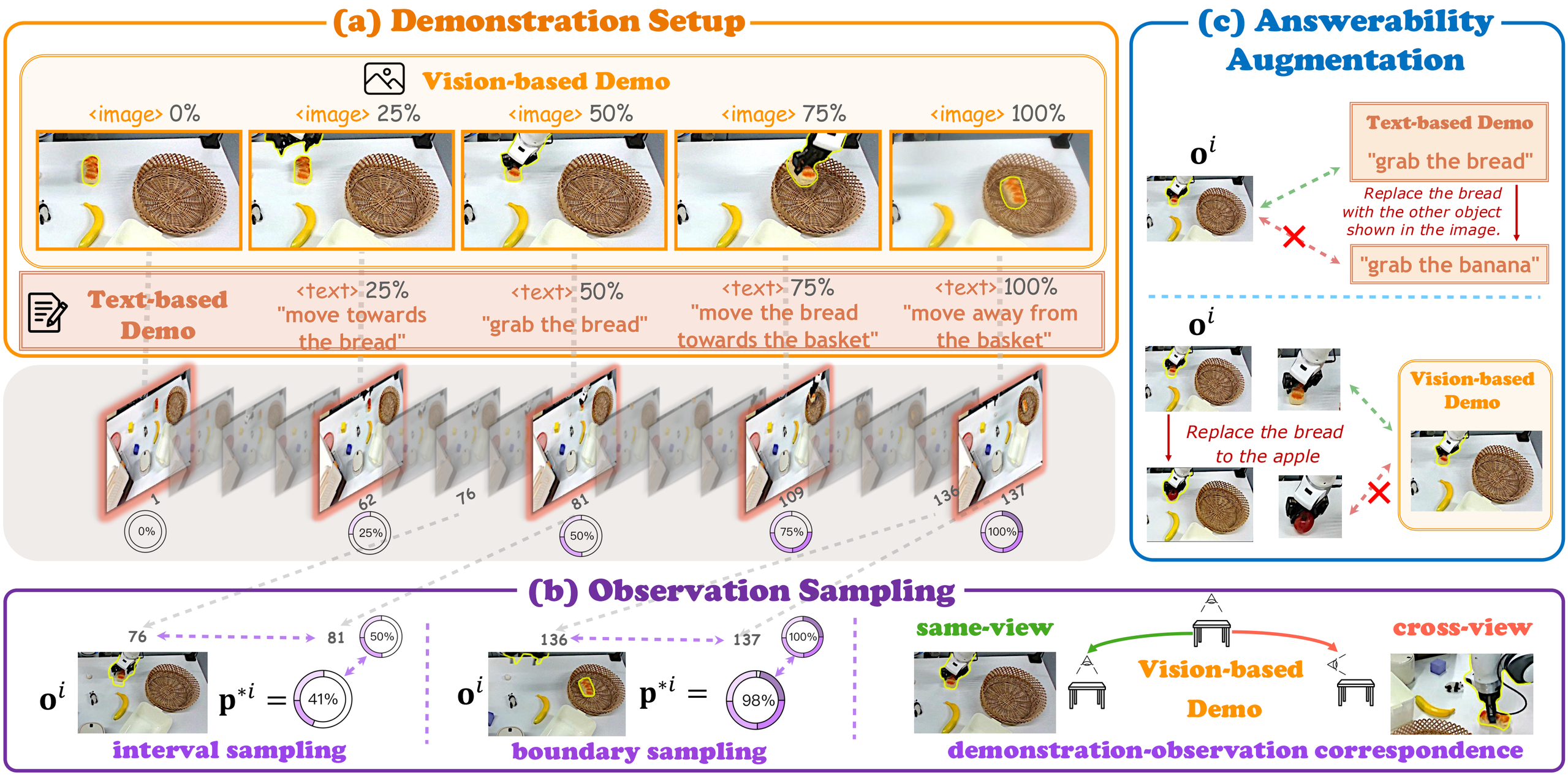
(a) Demonstration Setup
Vision-based Demo
<image> 0%
<image> 25%
<image> 50%
<image> 75%
<image> 100%
Text-based Demo
<text> 25%
"move towards the bread"
<text> 50%
"grab the bread"
<text> 75%
"move the bread towards the basket"
<text> 100%
"move away from the basket"
1
62
76
81
109
136
137
0%
25%
50%
75%
100%
(c) Answerability Augmentation
Text-based Demo
"grab the bread"
Replace the bread with the other object shown in the image.
"grab the banana"
Replace the bread to the apple
Vision-based Demo
(b) Observation Sampling
76
81
50%
41%
interval sampling
136
137
100%
98%
boundary sampling
same-view
cross-view
Vision-based Demo
demonstration-observation correspondence