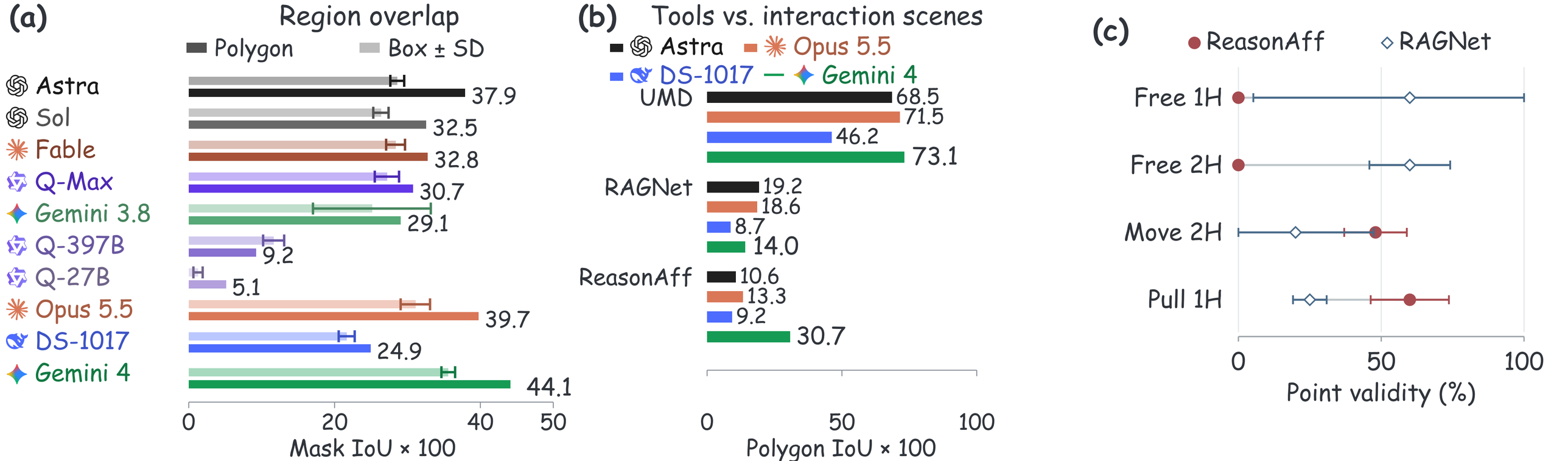

(a)
Region overlap
Polygon
Box ± SD
Astra
Sol
Fable
Q-Max
Gemini 3.8
Q-397B
Q-27B
Opus 5.5
DS-1017
Gemini 4
37.9
32.5
32.8
30.7
29.1
9.2
5.1
39.7
24.9
44.1
0
20
40
50
Mask IoU × 100
(b)
Tools vs. interaction scenes
Astra
Opus 5.5
DS-1017
Gemini 4
UMD
68.5
71.5
46.2
73.1
RAGNet
19.2
18.6
8.7
14.0
ReasonAff
10.6
13.3
9.2
30.7
0
50
100
Polygon IoU × 100
(c)
ReasonAff
RAGNet
Free 1H
Free 2H
Move 2H
Pull 1H
0
50
100
Point validity (%)