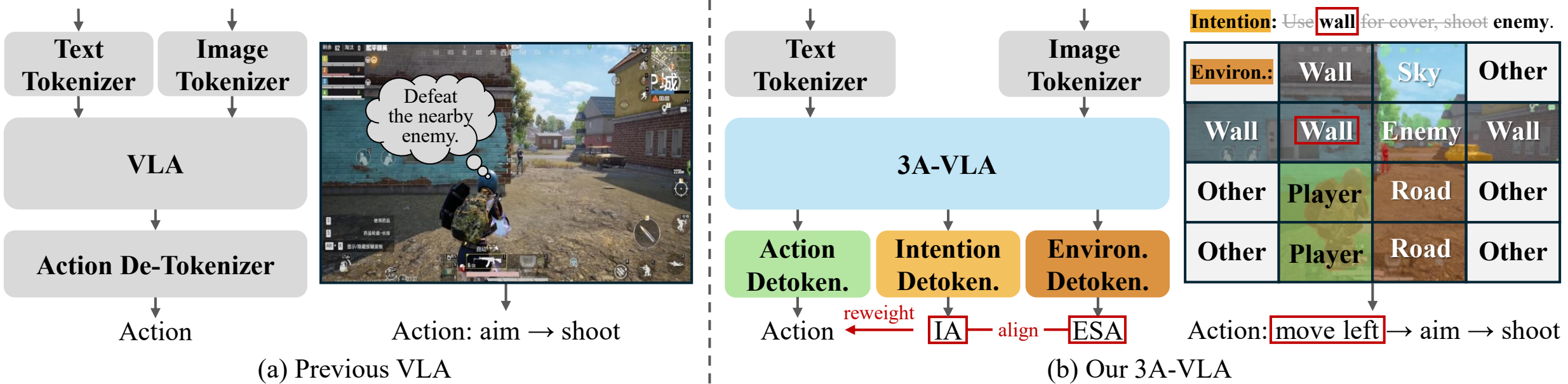

Text Tokenizer
Image Tokenizer
VLA
Action De-Tokenizer
Action
Defeat the nearby enemy.
Action: aim → shoot
(a) Previous VLA
Text Tokenizer
Image Tokenizer
3A-VLA
Action Detoken.
Intention Detoken.
Environ. Detoken.
reweight
Action
IA
align
ESA
Intention: Use wall for cover, shoot enemy.
Environ.:
Wall
Sky
Other
Wall
Wall
Enemy
Wall
Other
Player
Road
Other
Other
Player
Road
Other
Action: move left → aim → shoot
(b) Our 3A-VLA