



Original

“Is there a **baseball glove** in the Image?”



VLM

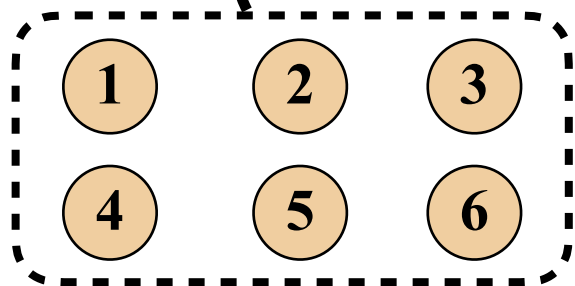


No

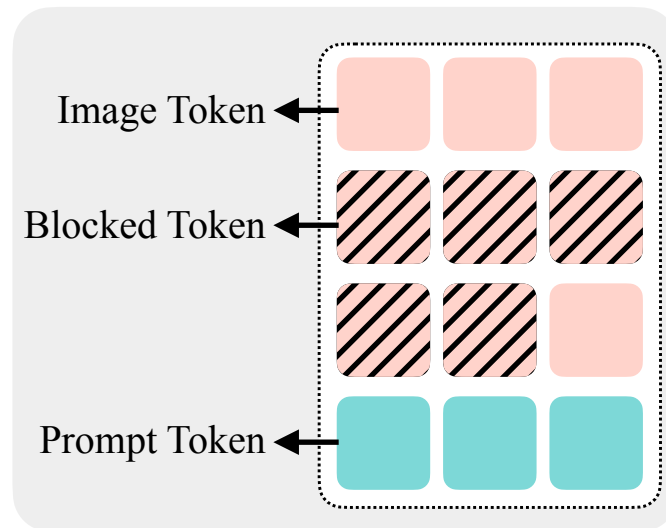
Yes



?



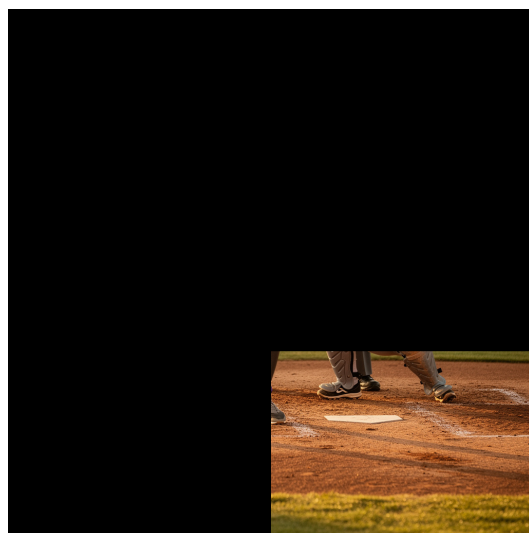
Global Level Intervention



① Token Mask



② Noise Blur



③ Black

Entity Level Intervention



④ Black Mask



⑤ Black Box



⑥ Entity Swap