(a) Perception of Text-based LLM

IC: IconNet
Y : Yolo
G : Grouding DINO
O : OCR

Tool box

(b) Perception of Multimodal LLM