

The image features three large word clouds in the background, each representing a different dataset. The top-left cloud is for COCO-Aff, listing items like laptop, toothbrush, bowl, skateboard, etc. The bottom-left cloud is for LVIS-Aff, listing items like spoon, fork, knife, chair, etc. The bottom-right cloud is for LS3D-Aff, listing items like dining table, couch, book, bed, etc. Overlaid on these word clouds is a photograph of a person walking down a city street, carrying a backpack and a bag, illustrating the concept of carrying items.

(b)

Afford-X

"X" = anything
anywhere

187M params

2.38 FPS inference

bell pepper

drink

(c) Generalizable affordance reasoning under diverse world

carry items in

dry your body with

take small items in

pack for travel with

access social media with

step on

sit comfortably on

ter with

display time with

spread butter with

cut hair with

Flexible tool selection

Preference for reasoning