

Vision Tasks Hierarchy

Classification "What is in the image?"

Output: Single label "This is a cat"

Add location

Detection "What and where?"

Output: Boxes + labels [cat @ (x,y,w,h)]

Add pixel precision

Segmentation "Exact pixel boundaries?"

Output: Pixel masks Every pixel labeled

Increasing Complexity → More output detail, harder task