Computer Vision Task Hierarchy

Classification

Input: Image

What is this?

Output: **Single Label** "Dog"

Add localization

Detection

Input: Image

What + Where?

Output: **Labels + Boxes** [("Dog", x1,y1,x2,y2)]

Pixel-level precision

Segmentation

Input: Image

Exact shape?

Output: **Pixel Masks** Every pixel labeled