Computer Vision (CV)

- CV is the feild of AI foucused on enabling machines to understand and interprit Visual data Like: (imagrs, videos, 3D scenes) with the goal to mimic or surpass human ability to do things Like: (recognize objects, understand scenes, track movement)

- Why: self driving cars, facial recognition, img recognition, etc

- key tasks in CV

1) imagr classification: input img → output class, Models: ResNet, NN

2) Object Detection: Find and locate objs Like faces im imgs, Models: CNN

3) segmentation: Semantic segmentation: color each pixel based on its class
instance segmentation: Seperate indivisual objs, Models: R-CNN

4) img generation: Create new img from scrach with conditions (GanAE)
• techniques: Diffusion, GAN • Models: DALLE, Wan2, SD

Others) img to img translation, 3D vision, Video understanding, img enhancement

- Concepts

• Pixels: smallest unit of img, in RGB or Grayscale (B/w)
• Features: Patterns like edges, textures, shapes extracted from img
• Convolution: Proctss of sliding filters over imgs to detect Features
• Feature Maps: Outputs after convolution layers in NN
• Pooling: Downsampling to focus on important info/efficiency. Ex Max/avg pooling
• Normalization: Scaling pixel values to stabalize learning. Ex Div 255 to get [0-1]