SOTA Models by Objectives

VLMs' Tasks from Basic to Advanced

- Core Vision Tasks (Basic Visual Recognition)
- Vision-Language Alignment (Zero-shot Learning)
- Captioning & OCR (Image Description and Reading)
- Visual Question Answering (Vision-Language Reasoning)
- Instruction Following & Visual Dialogue (Advanced Multimodal Interaction)