

SOTA Models for Vision-Language Alignment

Task	Description	SOTA Models
Zero-Shot Classification	Classify images without fine-tuning	CLIP (OpenAI, 2021)
Multimodal Retrieval	Search for images based on text or vice versa	CLIP, BLIP-2

