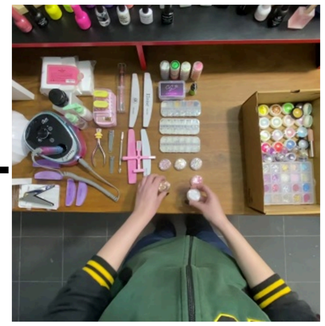
Action-Centric Features

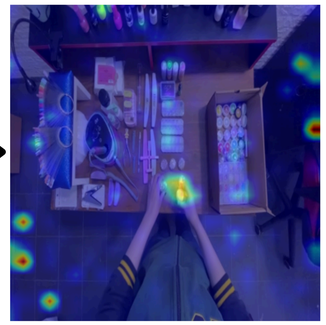
"Organize supplies..."



SigLIP

Semantic Meaning ✓

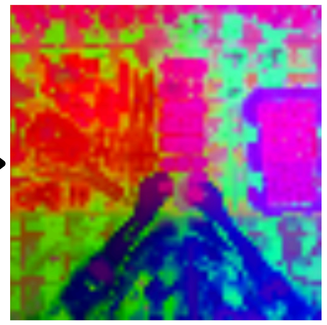
Action-Centric ✗



DINOv2

Spatial Detail ✓

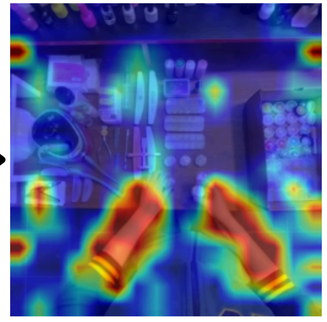
Action-Centric ✗



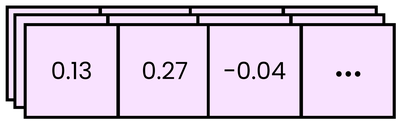
CAIP

Manipulation-relevant features ✓

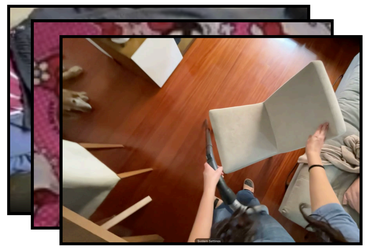
Action-Centric ✓



Contrastive Action-Image Pre-training



Hand Pose Actions



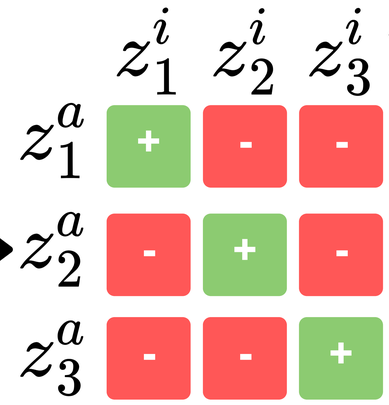
"Vacuum floor..."

Action Encoder

Image-Text Encoder

Image-Text Embedding

Action Embedding



Contrastive Training

Downstream Policy Performance