Which Tasks Need Vision vs LLM Finetuning?

State
- Full
- Partial

Vision-favored synergy

LLM-favored synergy

Vision helps LLM hurts

Both hurt

LLM helps Vision hurts

Zoom-In Puzzle

Colorization

Patch Reassembly

Video Unshuffle

Mental Rotation 3D (Cube)

Matchstick Rotation

Fetch Pick-Place

Fetch Reach

Mental Rotation 3D (Objaverse)

Jigsaw

Maze 2D

Counting

Referring Dot-Pointing

Sliding Block

Mental Rotation 2D

Matchstick Equation

Maze 3D

X-axis: LLM Gain

Y-axis: Vision Gain