Transfer From VLM to New VLM Eval VLM Llama2 Chat 7B + SigLIP DeepSeek-VL Chat 7B + SigLIP&SAM-B  ${\sf Gemma\ Instr\ 2B + SigLIP}$ Gemma Instr 8B + CLIP  ${\sf Gemma\ Instr\ 8B+SigLIP}$ Gemma Instr 2B + CLIPGemma Instr 2B + DINOv2&SigLIP Llama2 7B + SigLIP Llama2 Chat 7B + CLIP Llama2 Chat 7B + DINOv2/SigLIP — Llama2 7B + DINOv2&SigLIP Llama3 Instr 8B + DINOv2/SigLIP Llama2 7B + CLIP (Control) — Llama2 7B + CLIP Llama2 7B + DINOv2&SigLIP (Control) —— Llama2 13B + CLIP 0.0 Image Initialization — False 1.0 LLAVAv1.5 7B + CLIP (Repro)  $Llama2\ 13B + DINOv2\&SigLIP\ Llama2\ 13B + DINOv2\&SigLIP\ (Control)$ Llama2 13B + CLIP (Control)LLAVAv1.513B + CLIP (Repro) Llama $2\ 13B + CLIP$ 0.0 Llama2 13B + SigLIP (Control)Llama2 7B + CLIP (Control) Llama $2\ 13B + SigLIP$ Llama2.7B + CLIPLlama2 7B + DINOv2&SigLIP Llama2 7B + DINOv2&SigLIP (Control) 1.0 0.0 Llama2.7B + SigLIPLlama2 7B + SigLIP (Control)Llama2 Chat 7B + CLIPLlama2 Chat 7B + DINOv2/SigLIPLlama2 Chat 7B + SigLIPLlama3 Instr 8B + CLIP 0.0 $1.0\,\mathsf{Llama3\ Instr\ 8B} + \mathsf{DINOv2/SigLIP}$ Mistral Instr v0.2 7B + CLIP Mistral Instr v0.2 7B + DINOv2/SigLIP Mistral Instr v0.2 7B + SigLIP Llama3 Instr 8B + SigLIP Qwen VL Chat  $0.0_{0}^{-}$ 40000 20000 40000 20000 40000 40000 40000 40000 20000 20000 20000 20000 Gradient Step Gradient Step Gradient Step Gradient Step Gradient Step Gradient Step