🔵 Vision–Language Adapter Models

Category	Model	Key Idea
VL Pretraining	VL-BERT	Joint vision–language transformer; early multimodal fusion
VL Pretraining	UNITER	Universal image–text representation learning
VL Pretraining	ViLBERT	Two-stream co-attention transformer
VL Pretraining	VisualBERT	Shared embedding for image+text
VL Pretraining	LXMERT	Cross-modal encoder with alignment tasks
VL Pretraining	Pixel-BERT	Pixel-level visual features + BERT
VL Transformer	METER	Modular multimodal transformer
VL Contrastive	ALBEF	Align-before-fuse paradigm
VL Generative	BLIP	Vision-language bootstrapped training
VL-LLM Connector	BLIP-2	Q-Former as lightweight adapter

🟣 Explicit Multimodal Adapter Methods

Category	Model	Key Idea
Adapter-based PEFT	AdapterFusion	Combine multiple adapters for fusion
Multimodal Adapters	MAD-X	Multilingual & domain adapters
Modality Adaptive	MAT	Adapters for modality/domain variance
VL Adapter	M-Adapter	Adapter modules inside ViT/VL models
CLIP Adapter	CLIP-Adapter	Fine-tune CLIP via lightweight adapters
Few-shot Vision	TIP-Adapter / TIP-Adapter-F	Cache-based adapter for CLIP few-shot
Multimodal Unified	VLMo	MoE + adapter architecture
Fine-grained VL	FILIP	Patch–word fine-grained matching
Knowledge Adapter	K-Adapter	Inject factual/structural knowledge via adapters
Cross-modal Adapter	X-Adapter	Align modalities using cross-modal adapters
Prompt Adapter	Prompt-Adapter	Use prompts as soft adapters
Self-supervised	SS-Adapter	Self-supervised multimodal adapter

🟢 LLM + Vision Adapter

Category	Model	Key Idea
LLM Adapter	LLaMA-Adapter V1/V2	Adapter blocks enabling vision input
Tiny LLM Adapter	LLaMA-Adapter Turbo	Efficient adapter for small LLMs
Vision–LLM	LLaVA	Simple image projection adapter + LLaMA
Vision–LLM	MiniGPT-4	Q-Former + adapter bridge
Multi-sensory	ImageBind	Bind TTA–IMU using projection adapters
Instruction Tuning	InstructBLIP	Adapter-based alignment tuning
Multimodal LLM	Chimp	Multi-modal LLM with lightweight adapters
Multimodal LLM	mPLUG-Owl	Vision-LLM with cross-modality adapters
Flamingo-style	OpenFlamingo	Gated cross-attention + adapter-like connectors

🟠 Robotics / Multimodal RL Adapters

Category	Model	Key Idea
Robotics Multimodal	RT-X	Large-scale robot training via adapters
Cross-modal Policy	RoboCLIP Adapter	Use CLIP + adapter for robot manipulation
Multi-input RL	Prompt-based Robot Adapter	Prompt-based multimodal fusion
VL RL	CLIPort	Adapter-like modules for manipulation tasks

🔴 Diffusion + Multimodal Adapter Models

Category	Paper / Model	Key Idea
T2I Adapter	T2I-Adapter	Add modality adapters to stabilize diffusion control
Multi-input T2I	OpenT2I Adapter	Extended multi-modality control adapter
ControlNet	ControlNet-Adapter	Adapter-like conditional injection
Unified Control	Uni-ControlNet	Multi-modality control in one model
Image Adapter	IP-Adapter	Image→text latent adapter; SOTA for face/style alignment
Personalization	DreamBooth-LoRA	Adapter-style fine-tuning on diffusion