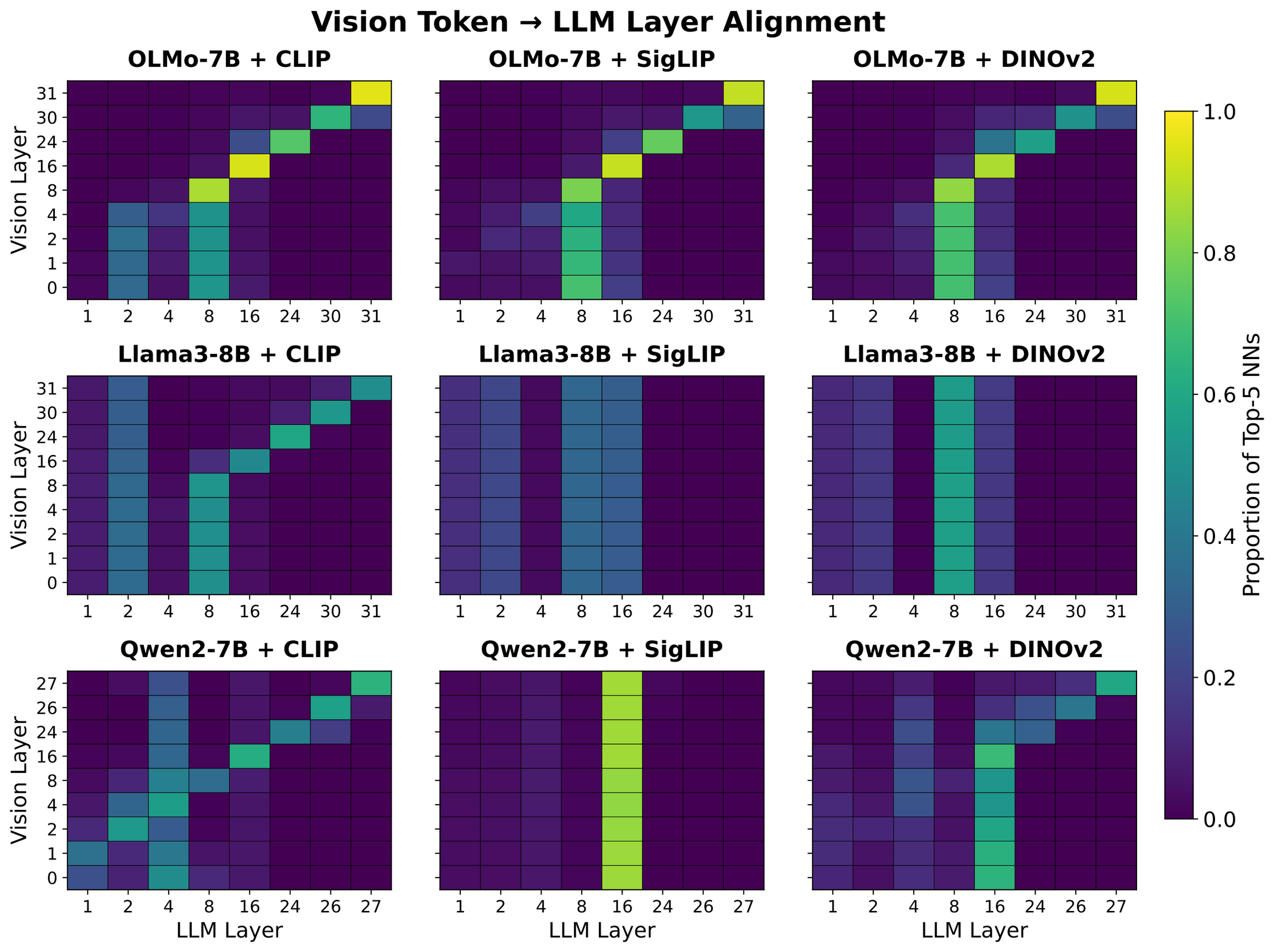
Vision Token → LLM Layer Alignment
OLMo-7B + CLIP
OLMo-7B + SigLIP
OLMo-7B + DINOv2
Llama3-8B + CLIP
Llama3-8B + SigLIP
Llama3-8B + DINOv2
Qwen2-7B + CLIP
Qwen2-7B + SigLIP
Qwen2-7B + DINOv2
Vision Layer
LLM Layer
Proportion of Top-5 NNs
0.0
0.2
0.4
0.6
0.8
1.0