



(a) Conventional Top-2 Routing

→ (b) + Fine-grained Expert Segmentation

→ (c) + Shared Expert Isolation

(DeepSeekMoE)