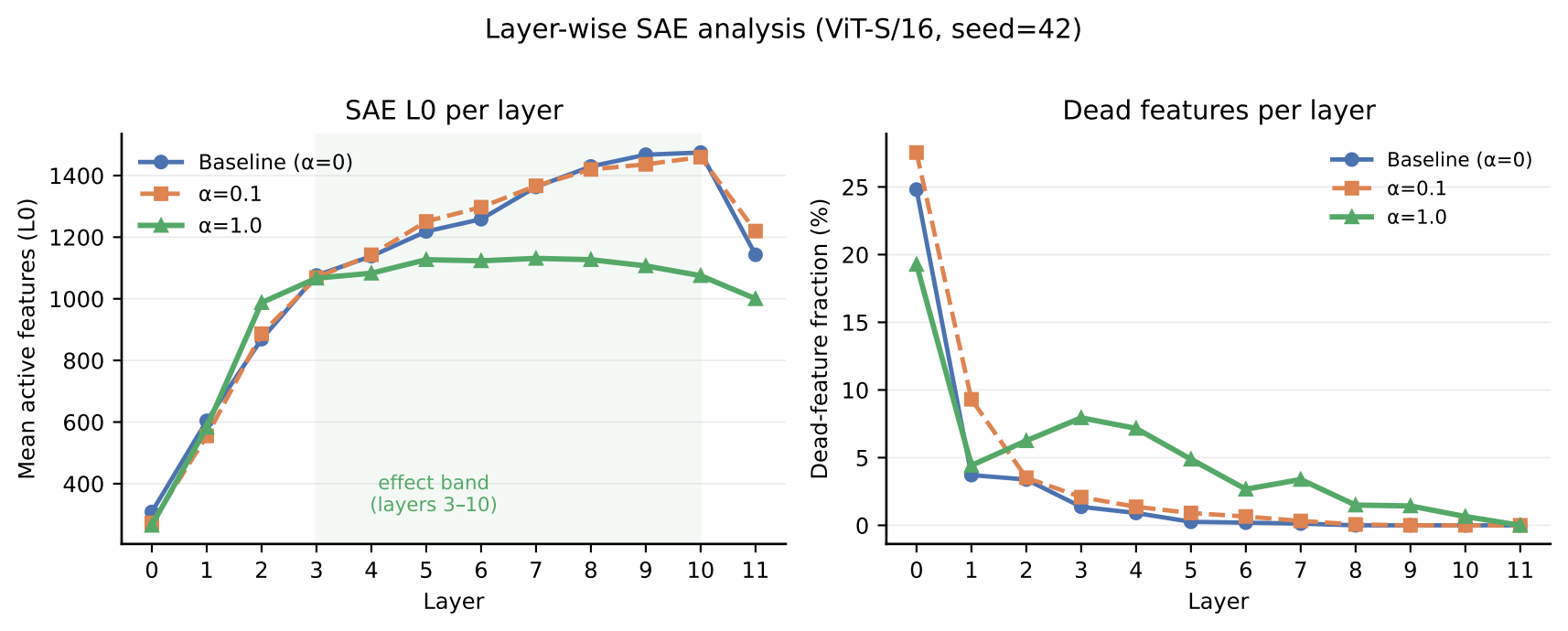

Layer-wise SAE analysis (ViT-S/16, seed=42)
SAE L0 per layer
Mean active features (L0)
Layer
Baseline (α=0)
α=0.1
α=1.0
effect band
(layers 3–10)
Dead features per layer
Dead-feature fraction (%)
Layer
Baseline (α=0)
α=0.1
α=1.0