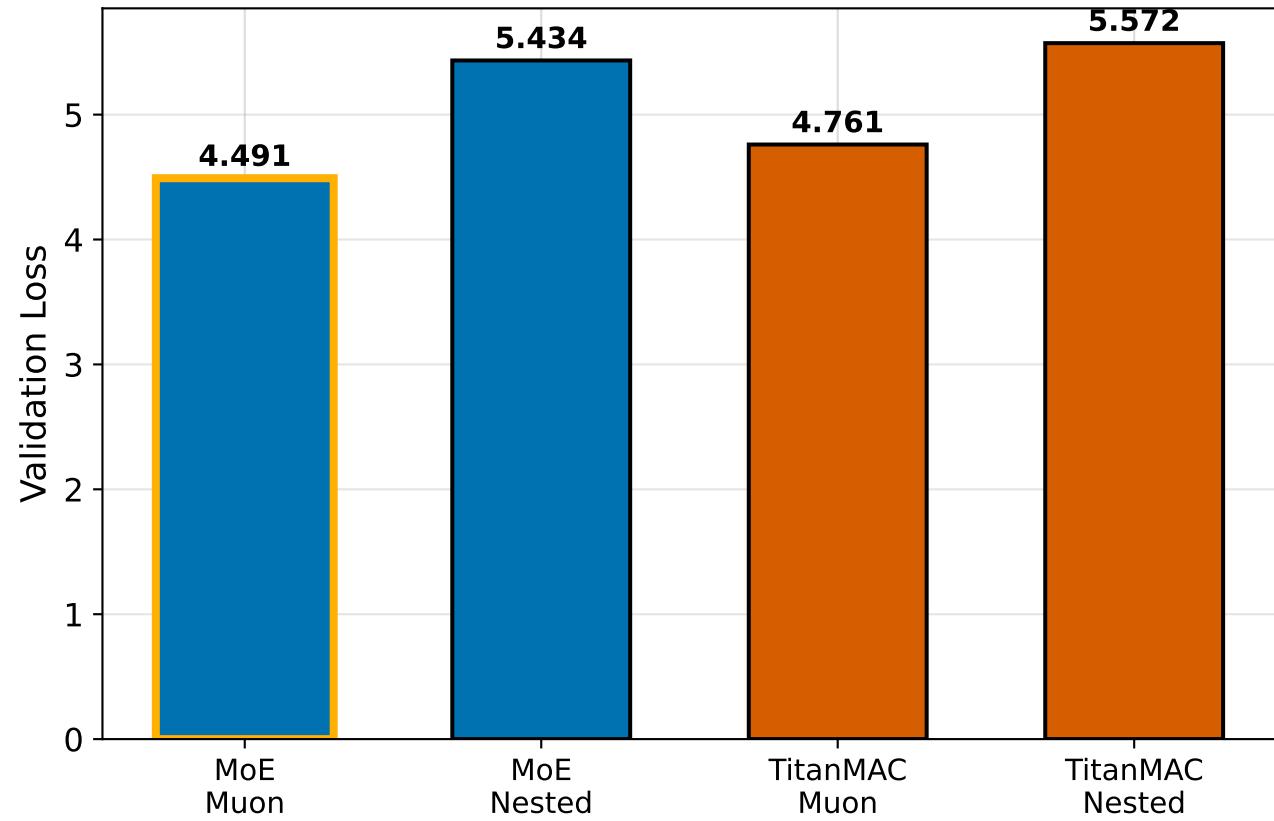
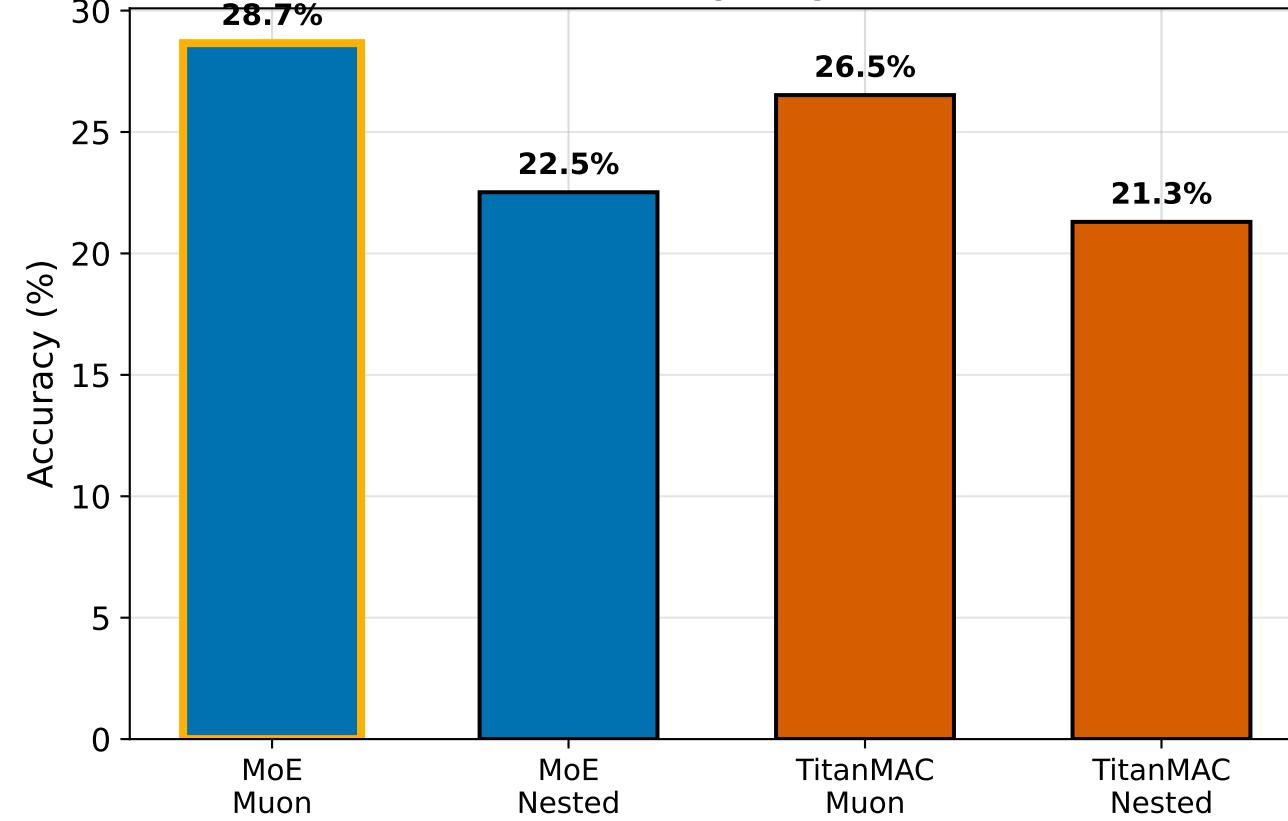


Baseline Comparison: MoE vs TitanMAC (600 steps)

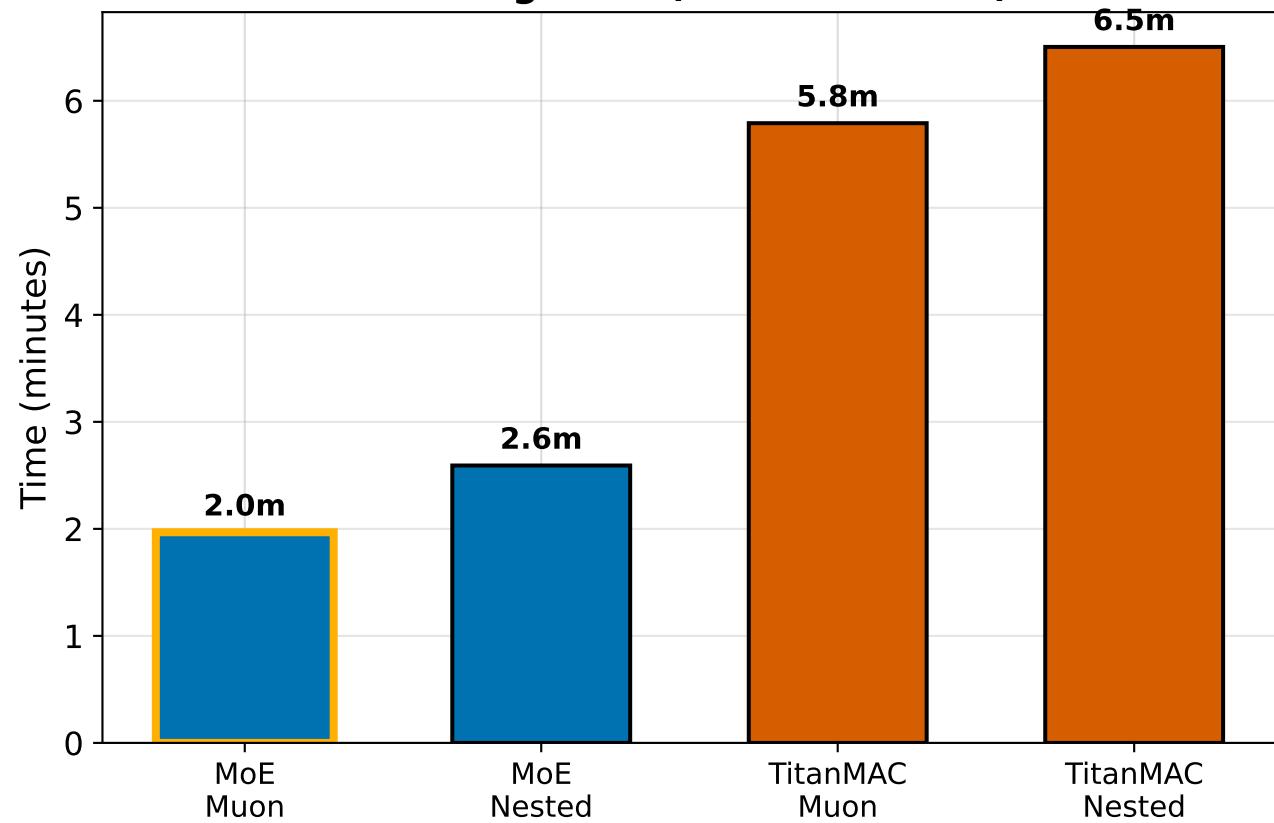
Validation Loss (lower is better)



Validation Accuracy (higher is better)



Training Time (lower is better)



Peak VRAM Usage (lower is better)

