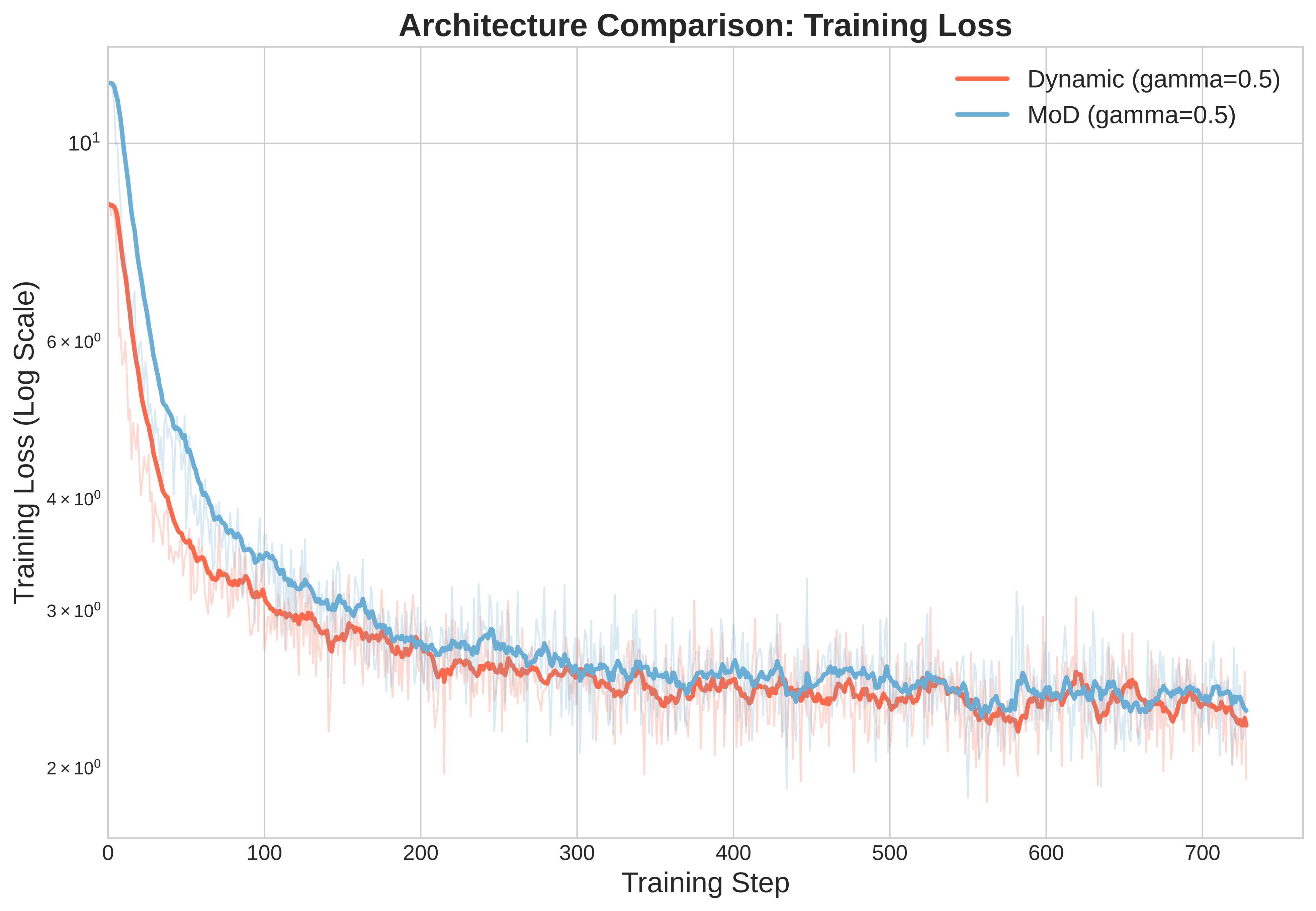
Architecture Comparison: Training Loss
Dynamic (gamma=0.5)
MoD (gamma=0.5)
Training Loss (Log Scale)
10^1
6×10^0
4×10^0
3×10^0
2×10^0
0
100
200
300
400
500
600
700
Training Step