**Architecture Comparison: Training Loss** Dynamic (gamma=0.5) MoD (gamma=0.5) 10<sup>1</sup> Fraining Loss (Log Scale)  $2 \times 10^{0}$ 100 200 300 400 500 700 600 **Training Step**