**Architecture Comparison: Validation Loss** Dynamic (gamma=0.5) MoD (gamma=0.5)  $3.2 \times 10^{0}$ Validation Loss (Log Scale) 3 × 10<sup>0</sup>  $2.4 \times 10^{0}$ 300 500 700 100 200 400 600 **Training Step**