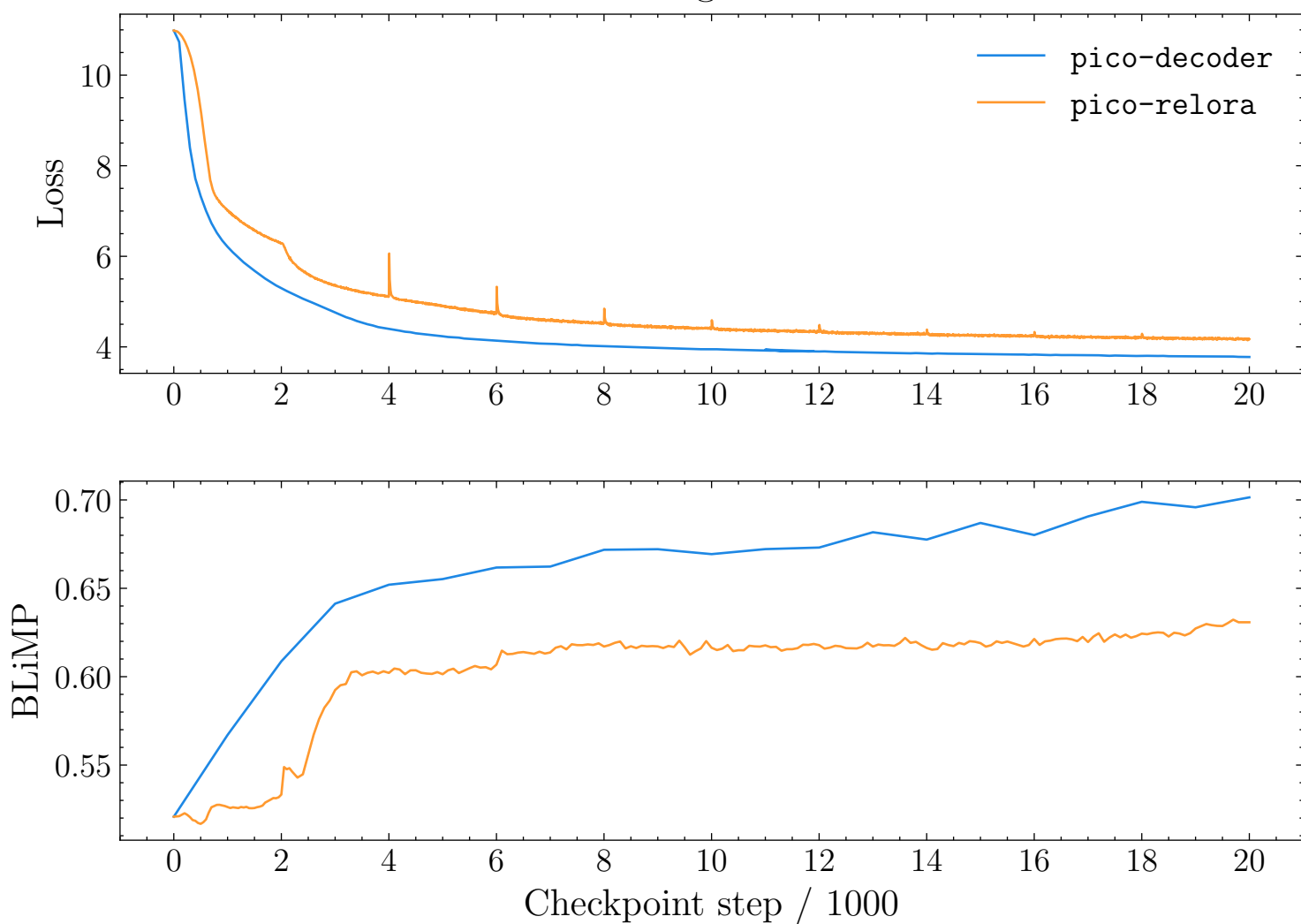
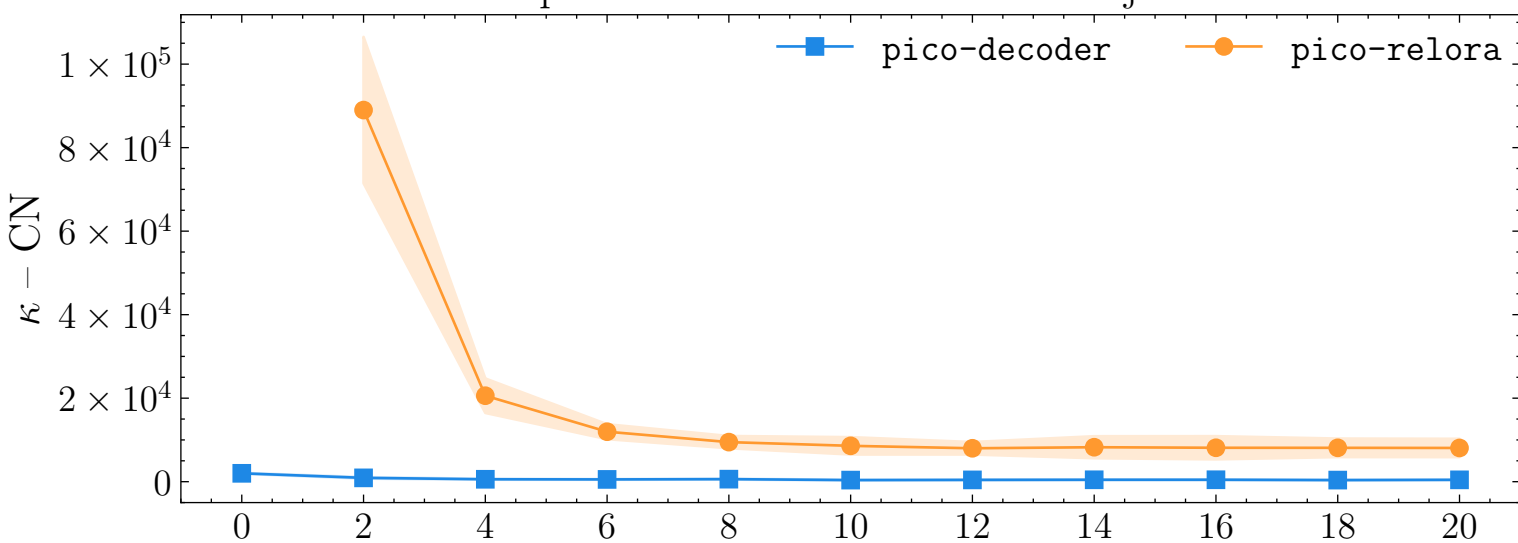


Training Plots



Gradient Condition Number (κ) Plots

Component = Attention Value Projection



Component = Attention Output Projection

