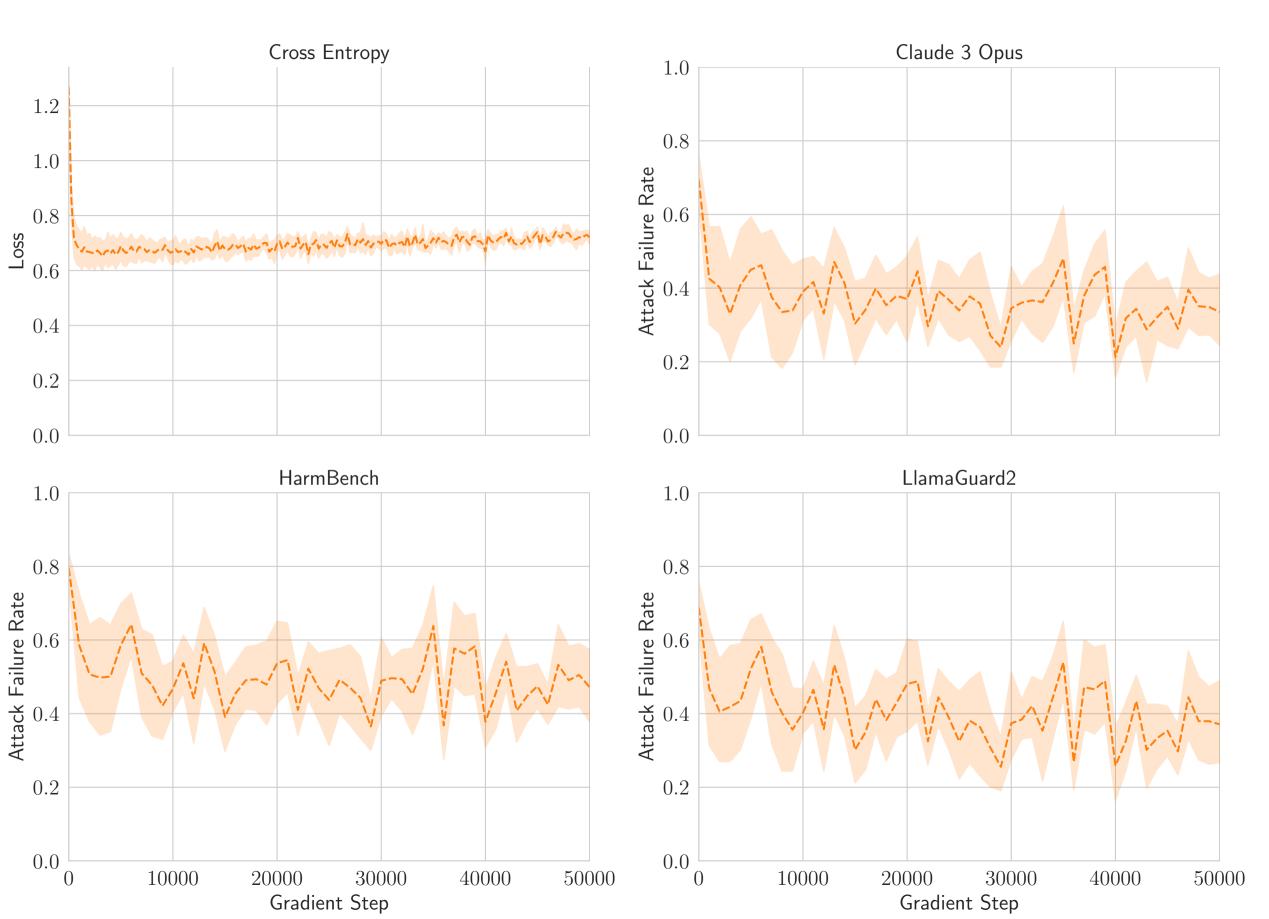
Universality of Image Jailbreaks



Attack Dataset (Train Split)

— AdvBench

Anthropic HHHSame Data Distribution

— True --- False