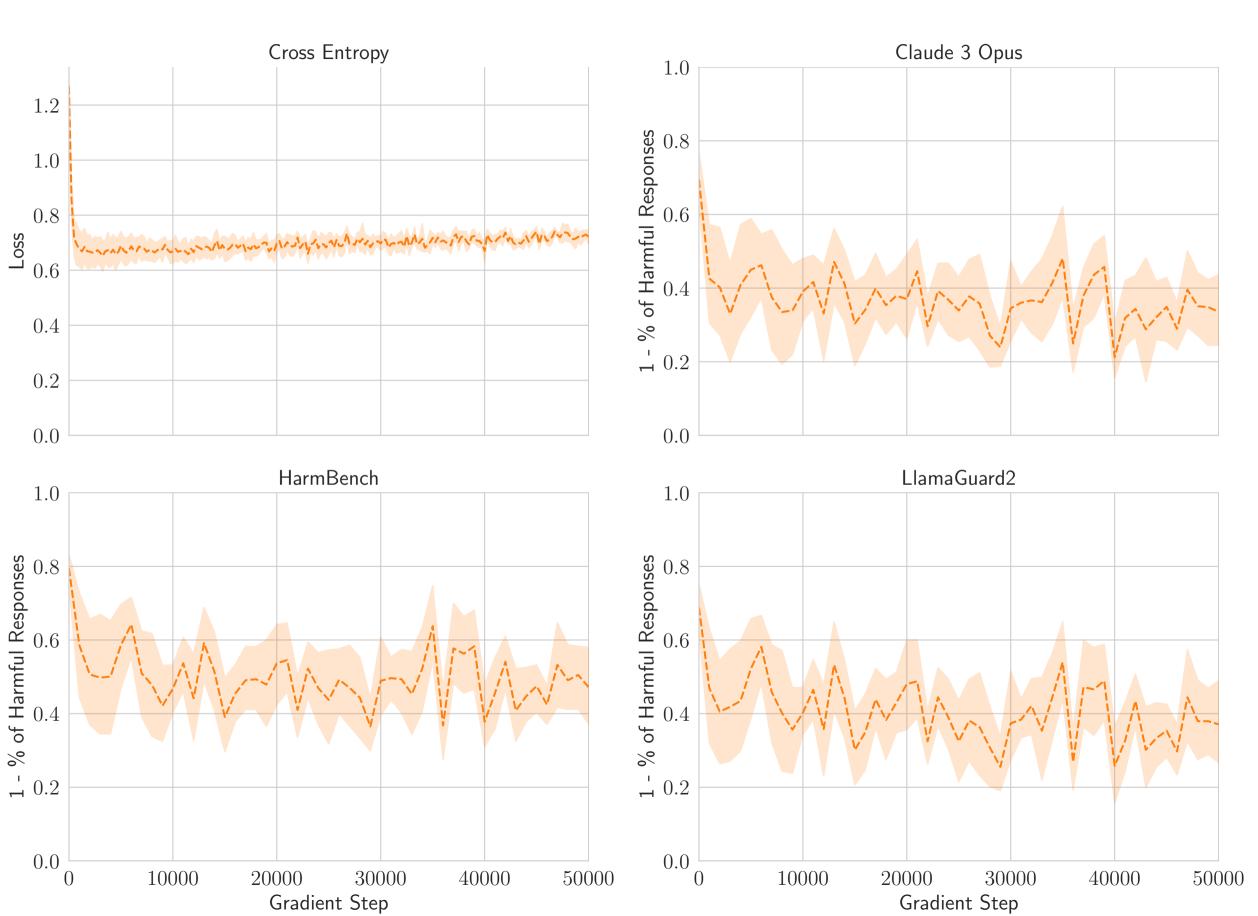
Universality of Image Jailbreaks



Attack Dataset (Train Split)
AdvBench
Anthropic HHH
Same Data Distribution

TrueFalse