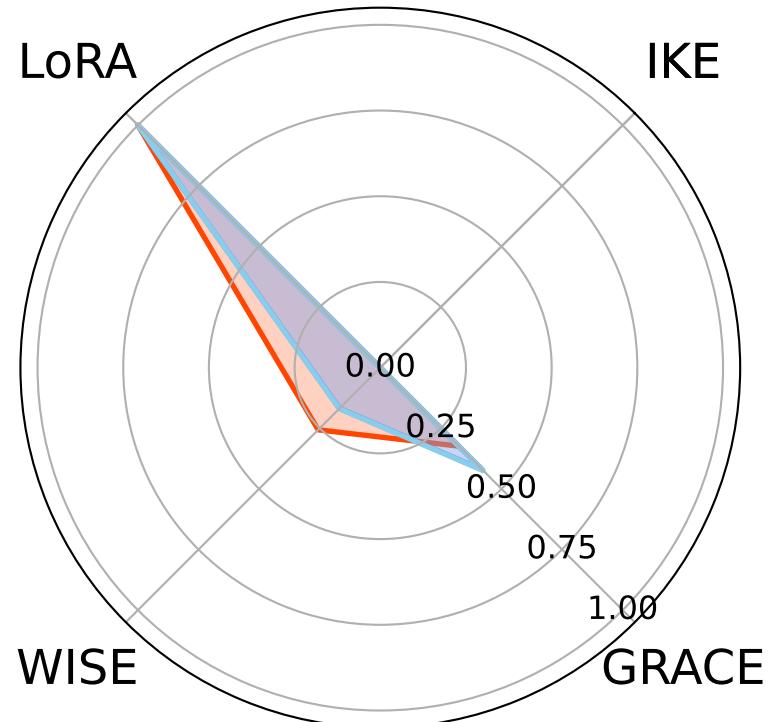
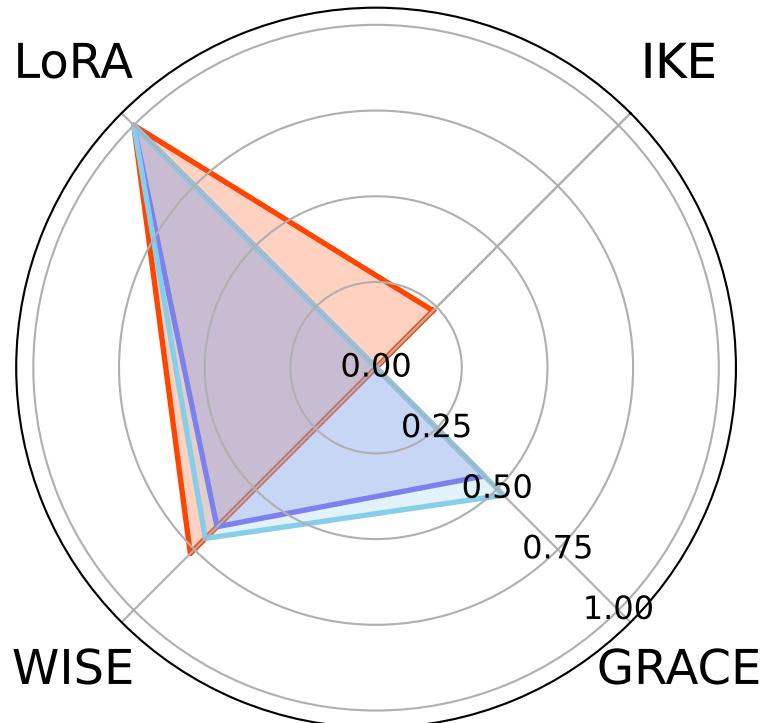


Avg Time per Sample (Normalized) Peak GPU Memory (Normalized)



— LLaVA-Onevision

■ QWen2-VL

— HuaTuoGPT-7B