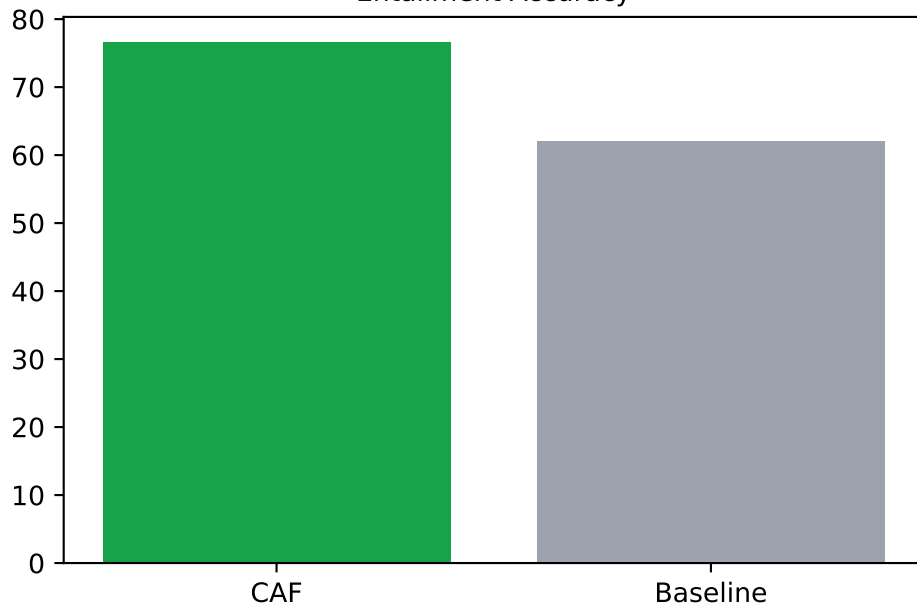
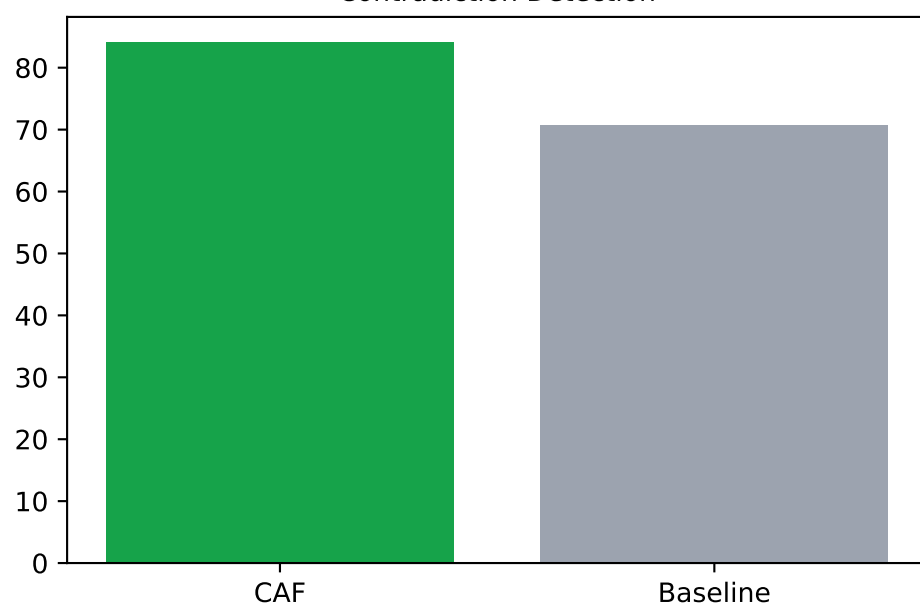


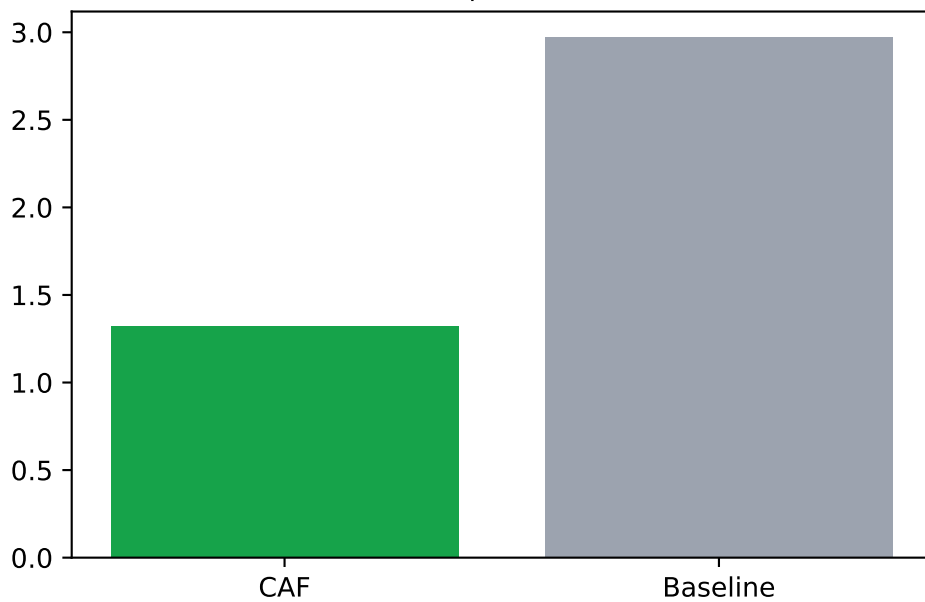
Entailment Accuracy



Contradiction Detection



Inference Depth (lower better)



Semantic Invariance

