

 $10^{4}$ 

'llava-lrv+7b' Attacked Model = 'one-stage+7b', 'reproduction-llava-v15+7b', 'train-1.5-epochs+7b', 'train-2-epochs+7b', 'train-3-epochs+7b', 'llava-lvis4v+7b', 'llava-lvis4v+7b',

Evaluated Model

'llava-lvis4v-lrv+7b' 'one-stage+7b'

'train-1.25-epochs+7b'

'reproduction-llava-v15+7b'

'train-2-epochs+7b'

'train-1.5-epochs+7b'

'train-3-epochs+7b'

In ensemble

False True

 $\begin{array}{c} 4 \times 10^{-1} \\ 4 \times 10^{-1} \\ 3 \times 10^{-1} \\ 2 \times 10^{-1} \end{array}$  $10^{3}$ Gradient Step