

RF : \mathbf{w}^{adv} ARF : $\mathbf{w}^{\text{adv}} + \boldsymbol{\alpha}^{\text{adv}^\top} \mathbf{D}_{[3, :]}^{\text{adv}}$ 