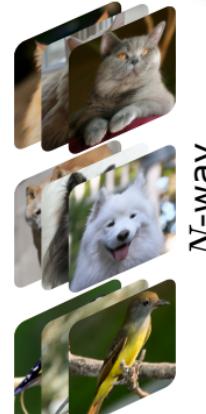


"a photo of a **cat**"
"a photo of a **dog**"
"a photo of a **bird**"



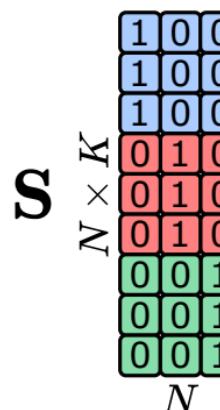
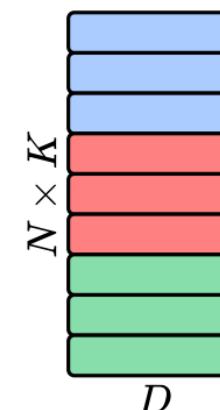
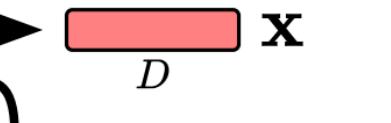
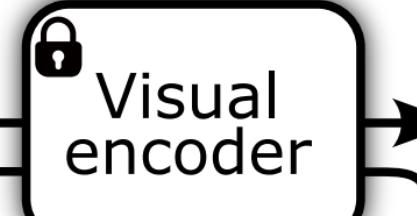
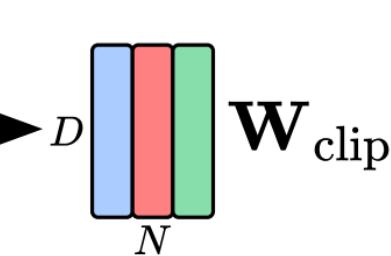
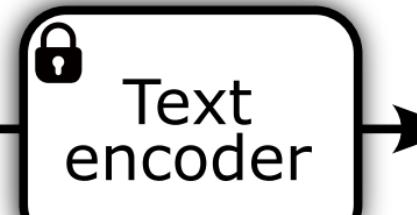
Input

Adaptation data



N -way

K -shot



Zero-shot CLIP
 $\mathbf{x}\mathbf{W}_{clip}$

Tip-Adapter
 $\mathbf{x}\mathbf{W}_{clip}$
 $+ \alpha \exp(-\beta(1 - \mathbf{x}\mathbf{S}^\top))\mathbf{L}$

ProKeR

$\mathbf{x}\mathbf{W}_{clip}$
 $+ \sum_{i=1}^{NK} k_\beta(\mathbf{S}_i, \mathbf{x})(\mathbf{I} + \frac{1}{\lambda} k_\beta(\mathbf{S}, \mathbf{S}))^{-1}(\mathbf{L}_i - \mathbf{S}_i \mathbf{W}_{clip})$

