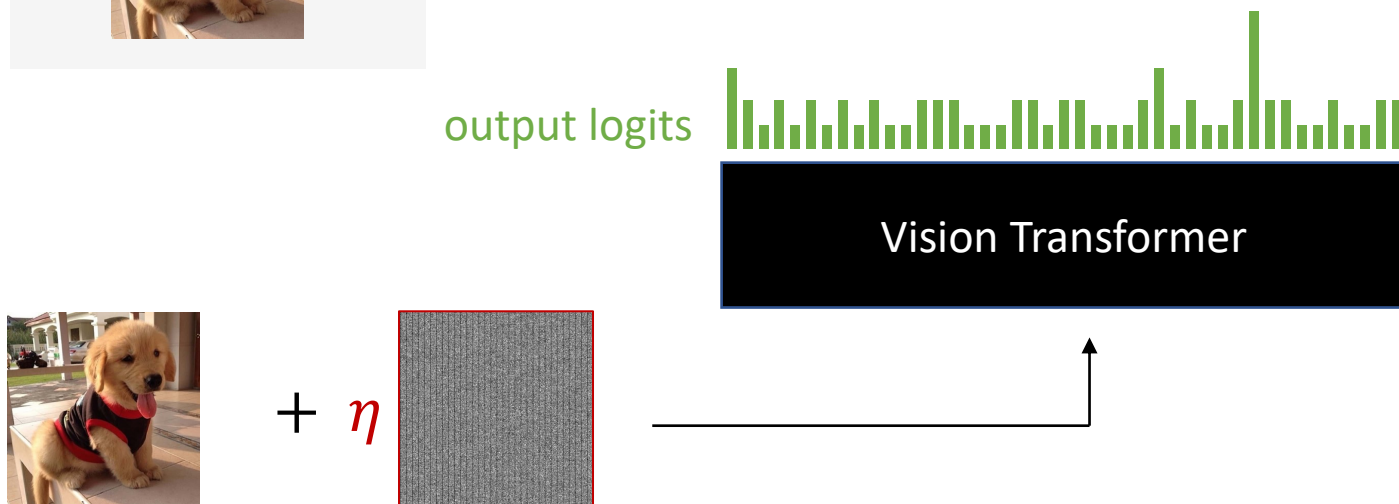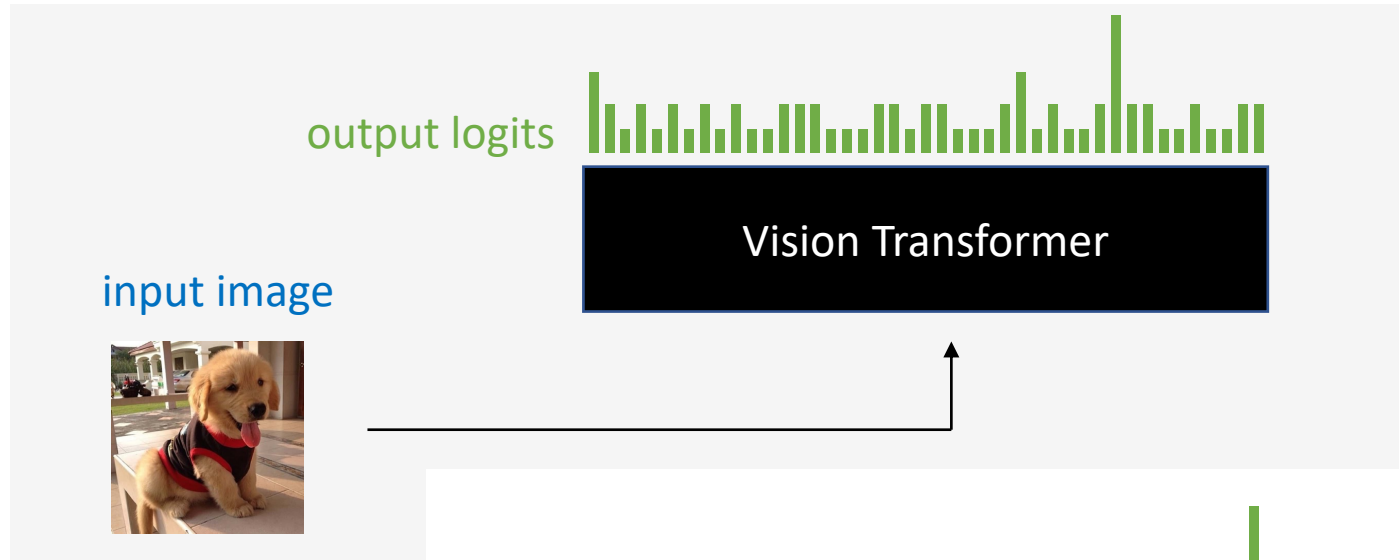$f(\text{🔵}) = f(\text{🔵}) = f(\text{🔵}) = \text{🟢}$

$f(\mathbf{x}) = \beta_1 x_1 + \beta_2 x_2 = y$

$f(\mathbf{x}) = f(\mathbf{x}') = y$

if $\quad \mathbf{x}' = \left(x_1 + \eta, x_2 - \dfrac{\eta \beta_1}{\beta_2} \right),$

$\forall\, \eta \in \mathbb{R} \quad \text{and} \quad \beta_2 \neq 0$

output logits

Vision Transformer

input image

output logits

Vision Transformer

$+\ \eta$

input image perturbed through nullspace for arbitrary η