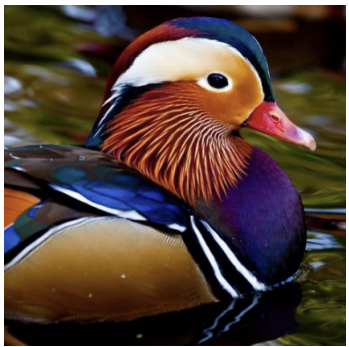
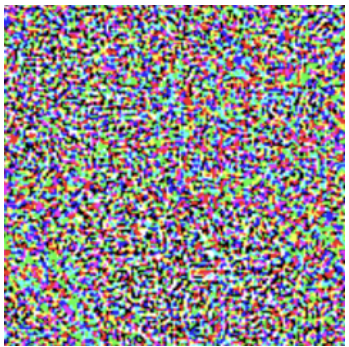
Benign example
"duck": 42.0% confidence



$\boldsymbol{x}$

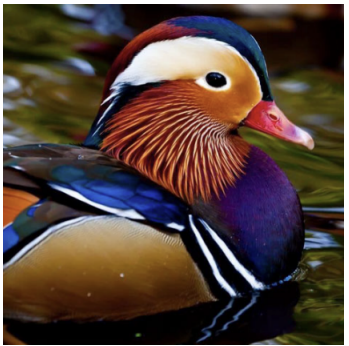
$+ \ \epsilon \times$



$\text{sign}\left(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x})\right)$

$=$

Adversarial example
"duck": 0.01% confidence
"platypus": 15.2% confidence



$\boldsymbol{x} + \ \epsilon \ \text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}))$