Stealing Generative Style-transfer Models

Can we extract high-fidelity GANs?

Background:

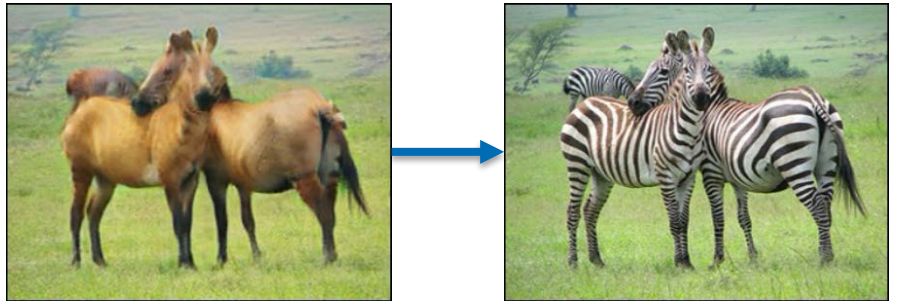
- GANs can generate realistic images
- Can be used to change image style
  - coloring, face filters, style application
- Core feature in generative art and in social media apps

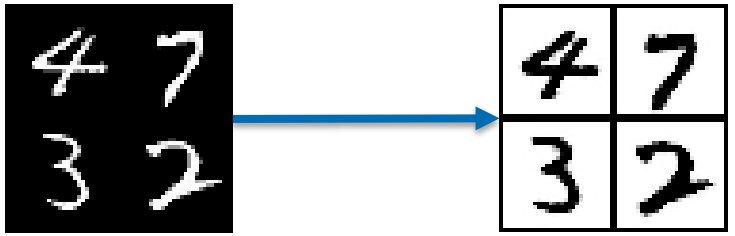
Approach:

- Gather unstyled images from the victim model domain
- Query victim model to obtain styled images
- Train a local GAN that maps raw images to styled images

Properties:

- Copies functionality without secret source style images
- No assumptions about the type of the victim GAN
- Requires data from the same or similar domain





In progress