Generative Visual Manipulation on the Natural Image Manifold

Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman* and Alexei A. Efros
UC Berkeley *Adobe

Image Manipulation is Hard!

The Lack of "Safety Wheels":

- any less-than-perfect edit immediately makes the image look unrealistic.
- classic visual manipulation paradigm does not prevent the user from "falling off" the manifold of natural images

Realism CNN — Predict / Improve — Image Editing Model

[Zhu et al. 15']: $M = \{x \mid P(real|x) = 1\}$

A desired output

- stay close to the input.
- satisfy user's constraint.
- Lie on the **natural image manifold** $M = \{x|x = G(z)\}$

Learning Natural Image Manifold

Generative Adversarial Network (GAN)
[Goodfellow et al. 14'] [Radford et al. 15']

$G(z): z \rightarrow x$

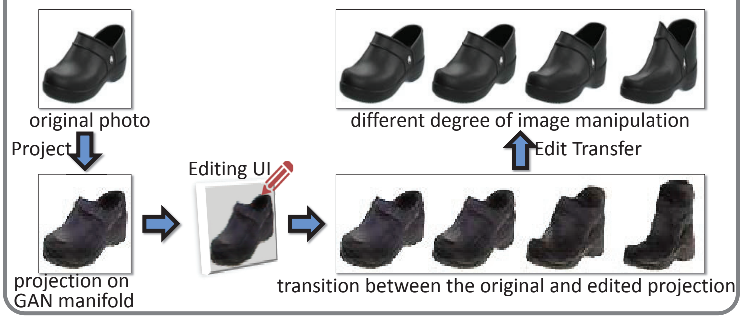
$N(0, \mathrm{I}), \mathrm{Unif}[-1,1]^d$ Natural images

Generator $x = G(z)$
Generate an image sample x from a latent vector z

Discriminator $D(x) \rightarrow 0/1$
Distinguish between real images and fake samples


Linear Interpolation in z space

Overview



Manipulating the Latent Vector

Objective: $z^* = \arg\min_{z \in \mathbb{Z}} \Big\{ \underbrace{\sum_g \mathcal{L}_g(G(z), v_g)}_{\text{data term}} + \underbrace{\lambda_s \cdot \|z - z_0\|_2^2}_{\text{manifold smoothness}} \Big\}.$

constraint loss L_g — user guidance image

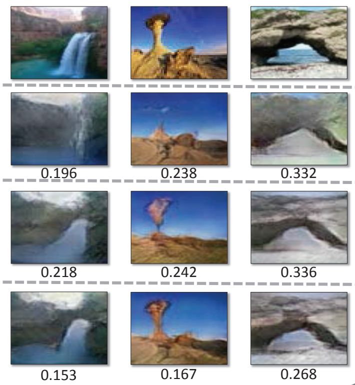
Projecting an Image onto the Manifold

Input: real image x^R
Output: latent vector z

Optimization
$z^* = \arg\min \mathcal{L}(G(z), x^R)$

Inverting Network $z = P(x)$
$\theta_P^* = \arg\min_{\theta_P} \sum_{x_n^R} \mathcal{L}(G(P(x^R; \theta_P)), x^R)$

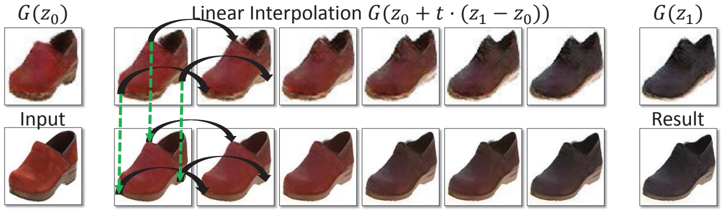
Hybrid Method
Use the network as initialization for the optimization problem



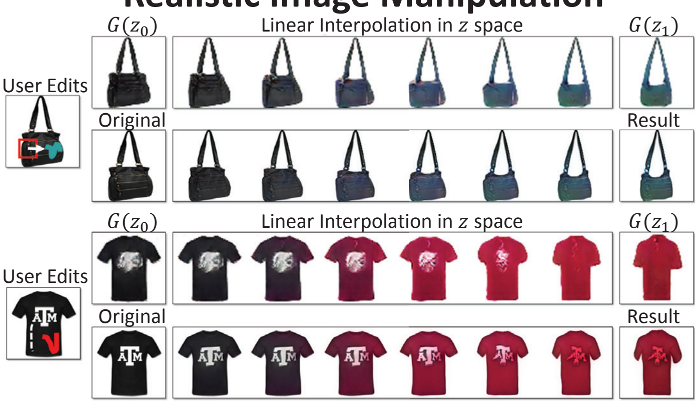
Edit Transfer

Motion (u, v)+ Color ($A_{3\times4}$): estimate per-pixel geometric and color variation

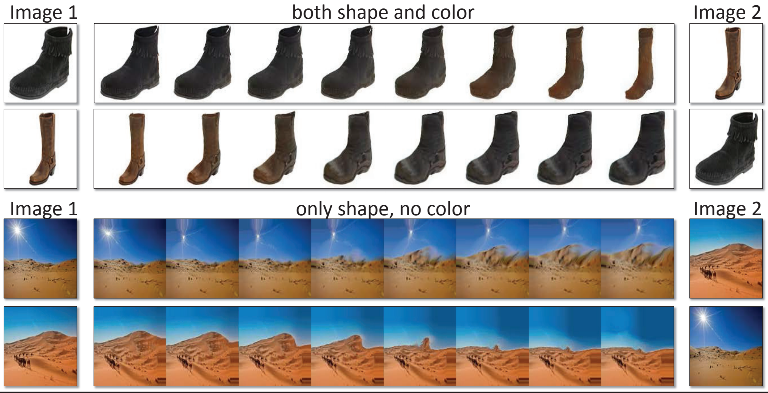
$$\iint \underbrace{\|I(x,y,t) - \boxed{A} \cdot I(x+\boxed{u}, y+\boxed{v}, t+1)\|^2}_{\text{data term}} + \underbrace{\sigma_s(\|\nabla u\|^2 + \|\nabla v\|^2)}_{\text{spatial reg}} + \underbrace{\sigma_c \|\nabla A\|^2}_{\text{color reg}} dx dy$$



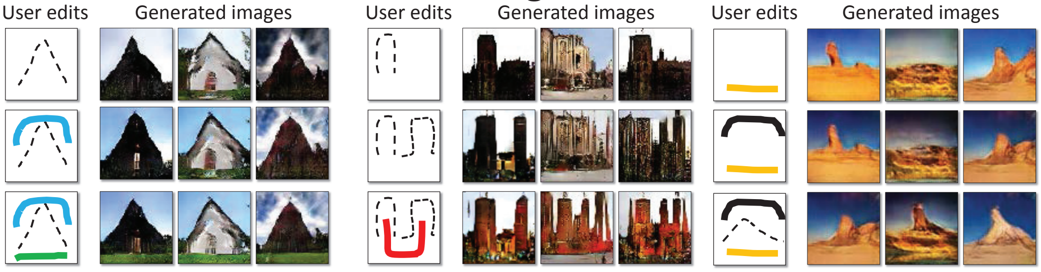
Realistic Image Manipulation



Generative Image Transformation



Interactive Image Generation



Reference

[1] Zhu et al. Learning a Discriminative Model for the Perception of Realism in Composite Images. ICCV 2015.

[2] Goodfellow et al. Generative Adversarial Nets. NIPS 2014

[3] Radford et al. Unsupervised representation learning with deep convolutional generative adversarial networks. ICLR 2016