

# Generative Visual Manipulation on the Natural Image Manifold

Qilei Zhang

Jul 2 2018

## Abstract

*Realistic image manipulation is challenging because it requires modifying the image appearance in a user-controlled way, while preserving the realism of the result. Unless the user has considerable artistic skill, it is easy to fall off the manifold of natural images while editing.*

## 1. Introduction

Visual communication is a sad one-sided nature. Generally, information is perceived by visual form, but only a few have enough talents to express themselves effectively. Even in the most mundane task, this imbalance will emerge [1].

For most people, even simple image processing in PS image processing software is difficult to overcome. Any imperfect editor instantly makes the image totally unrealistic. In other words, the classic visual manipulation paradigm does not prevent users from shedding from natural images [2].

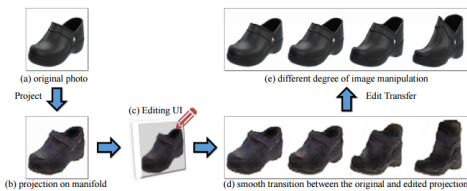


Figure 1. We use generative adversarial networks (GAN) to perform image editing on the natural image manifold. We first project an original photo (a) onto a lowdimensional latent vector representation (b) by regenerating it using GAN. We then modify the color and shape of the generated image (d) using various brush tools (c) (for example, dragging the top of the shoe). Finally, we apply the same amount of geometric and color changes to the original photo to achieve the final result (e). See our interactive image editing demo on Youtube.

## 2. Prior Work

Image editing is a recognized field in computer graphics, where input images are manipulated to achieve user specified specific targets. Examples of basic editing include global or partial change of the color property of an image. More advanced editing methods, such as image warpage or structured image editing, intelligently adjust the pixels in the image after user editing. When impressive results are made in the hands of experts, the results of these types of methods fail to look like real images.



Figure 2. GAN as a manifold approximation. (a) Randomly generated examples from a GAN, trained on the shirts dataset; (b) random jittering: each row shows a random sample from a GAN (the first one at the left), and its variants produced by adding Gaussian noise to  $z$  in the latent space; (c) interpolation: each row shows two randomly generated images (first and last), and their smooth interpolations in the latent space.

Common artifacts include unrealistic colors, exaggerated stretching, obvious repetition and excessive smoothing. This is because they rely on low-level principles and do not capture higher level information about natural images [3].

## References

- [1] I. I. A. Groen, S. Ghebreab, V. A. F. Lamme, and H. S. Scholte. Low-level contrast statistics are diagnostic of invariance of natural textures. *Frontiers in Computational Neuroscience*, 6(6):34, 2012. 1
- [2] R. Nikolaidis, B. Walker, and G. Weinberg. Generative musical tension modeling and its application to dynamic sonification. *Computer Music Journal*, 36(1):55–64, 2012. 1
- [3] M. Palmiero, V. Cardì, and M. O. Belardinelli. The role of vividness of visual mental imagery on different dimensions of creativity. *Creativity Research Journal*, 23(4):372–375, 2011. 1