Intelligent Photo Editing
Modifying Input Code

Each dimension of input vector represents some characteristics.

The input code determines the generator output.

Understand the meaning of each dimension to control the output.
### Connecting Code and Attribute

#### CelebA

<table>
<thead>
<tr>
<th>Image</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td>Arched eyebrows, attractive, brown hair, heavy makeup, high cheekbones, mouth slightly open, no beard, pointy nose, smiling, straight hair, wearing earrings, wearing lipstick, young.</td>
</tr>
<tr>
<td><img src="image2.png" alt="Image" /></td>
<td>5 o’clock shadows, attractive, bags under eyes, big lips, big nose, black hair, bushy eyebrows, male, no beard, pointy nose, straight hair, young.</td>
</tr>
</tbody>
</table>
GAN+Autoencoder

- We have a generator (input z, output x)
- However, given x, how can we find z?
  - Learn an encoder (input x, output z)

![Diagram of GAN+Autoencoder](image-url)
Attribute Representation

$$Z_{\text{long}} = \frac{1}{N_1} \sum_{x \in \text{long}} En(x) - \frac{1}{N_2} \sum_{x' \notin \text{long}} En(x')$$

Short Hair  \( x \rightarrow En(x) + Z_{\text{long}} = z' \rightarrow \text{Gen}(z') \)  Long Hair
Photo Editing

https://www.youtube.com/watch?v=kPEIJJJsQr7U
Basic Idea

Why move on the code space?

Fulfill the constraint
Back to $z$

- **Method 1**
  \[
  z^* = \arg \min_z L(G(z), x^T)
  \]
  Difference between $G(z)$ and $x^T$
  - Pixel-wise
  - By another network

- **Method 2**
  as close as possible

- **Method 3**

Using the results from **method 2** as the initialization of **method 1**
Editing Photos

• $z_0$ is the code of the input image

\[ z^* = \arg \min_z U(G(z)) + \lambda_1 \|z - z_0\|^2 - \lambda_2 D(G(z)) \]

Using discriminator to check the image is realistic or not

Does it fulfill the constraint of editing?

Not too far away from the original image

https://www.youtube.com/watch?v=9c4z6YsBGQ0
Image super resolution


![Image Comparison](image.png)

Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4x upscaling]
Image Completion

Demo

Image completion is a very complicated task...

https://www.youtube.com/watch?v=5Ua4NUKowPU