Eye In-Painting with Exemplar Generative Adversarial Networks

Brian Dolhansky, Cristian Canton Ferrer, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition

INTRODUCTION

- Closed-to-Open eye inpainting
- GANs “IN-PAINT” a person’s eyes without losing the person’s defining features
- Perceptually and semantically plausible results
- Exemplar GANs (ExGANs)
  - Utilizing exemplar information
  - Reference image or perceptual code used as a conditioning example

INTRODUCTION

- Deep Convolutional Neural Networks (DNNs) for inpainting
- Learn to preserve features
  - Missing regions of pictures showing natural scenery
  - Global lighting and skin tone of pictures that require facial transformation
- Can complete images of arbitrary resolutions by filling in missing regions of any shape

INTRODUCTION

Picture with missing regions

In-painted output

S. Iizuka, E. Simo-Serra, and H. Ishikawa. *Globally and locally consistent image completion.* ACM Transactions on Graphics, 36(4), July 2017.

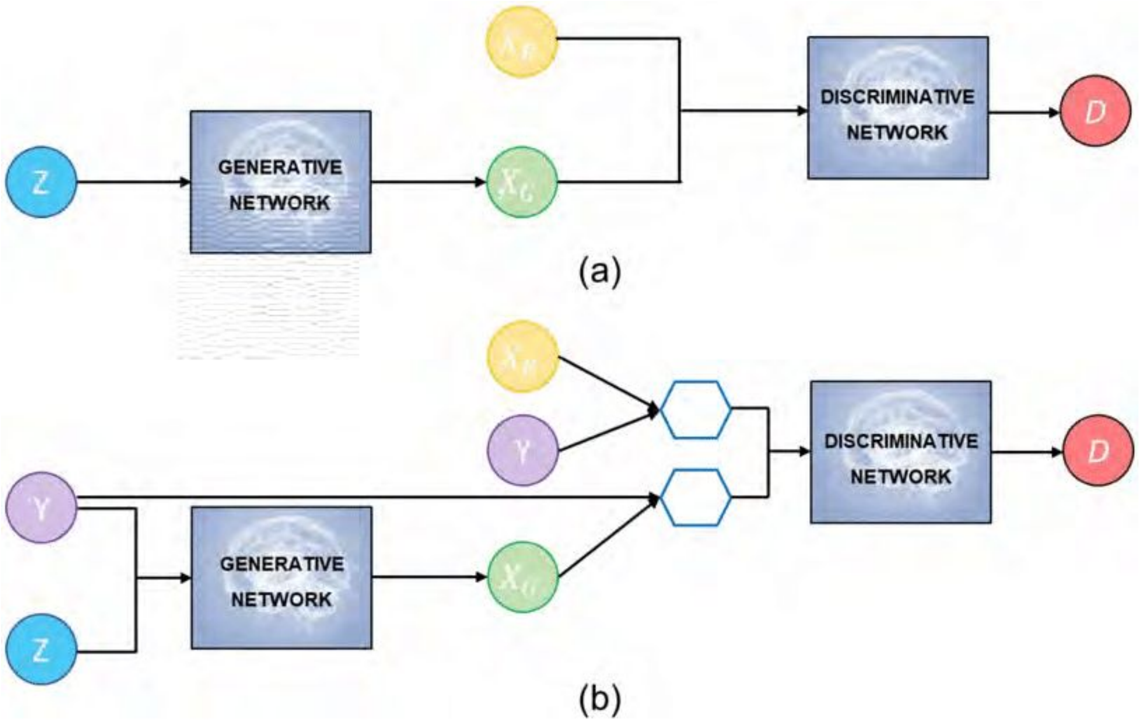
INTRODUCTION

- Downside of using DNNs to solve the problem of facial transformations
    - Don’t preserve identity of the person
    - Only insert a pair of eyes that correspond to similar faces but not to that specific face
    - Defining features on the person’s eyes such as uncommon eye shape will be lost

CONTRIBUTION

- ExGAN based Eye In-Painting
- Similar to Conditional-GANs (CGANs)
    - Add extra information to the generator of the network
    - CGANs vs ExGANs
    - Providing examples to “identifying traits” of the in-painted area in this case the eyes

CONTRIBUTION: CGAN Architecture



CONTRIBUTION

- Provided that multiple images of the object are available at inference time, ExGANs are able to generate semantically and perceptually consistent output
- Examples can be of two types:
    - Raw Images
    - Perceptually-coded sections of images

EXEMPLAR GANS FOR IN-PAINTING

- Using a second source of related information to guide the generator as it creates an image
- A picture of a person in a different pose but with eyes opened
- Use the reference image or the perceptual code generated from that image to “guide” that is inserted in different places when synthesizing results
- Two approaches
    - Reference-based In-Painting
    - Code-based In-Painting

EXEMPLAR GANS FOR IN-PAINTING: REFERENCE IN-PAINING

- Using a raw image as a reference to guide the process of in-painting
- For each image xi, there is a reference image ri

$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{x}_i,\mathbf{r}_i \sim p_{\text{data}}(\mathbf{x},\mathbf{r})}[\log D(\mathbf{x}_i,\mathbf{r}_i)] + \mathbb{E}_{\mathbf{r_i}\sim p_{\mathbf{r}},G(\cdot)\sim p_{\mathbf{z}}}[\log 1 - D(G(\mathbf{z}_i,\mathbf{r}_i))] + ||G(\mathbf{z}_i,\mathbf{r}_i) - \mathbf{x}_i||_1 \tag{1}$$

EXEMPLAR GANS FOR IN-PAINTING: REFERENCE IN-PAINING

- As compared to vanilla GAN formulation (shown below), both generator and discriminator can take an example as an input

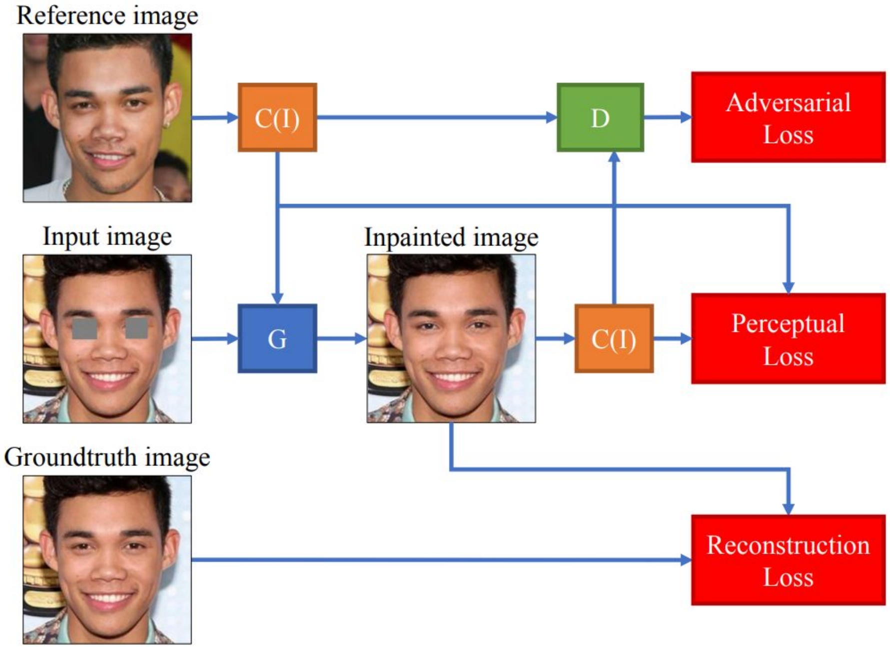
$$\min_G \max_D V(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

EXEMPLAR GANS FOR IN-PAINTING: CODE IN-PAINTING

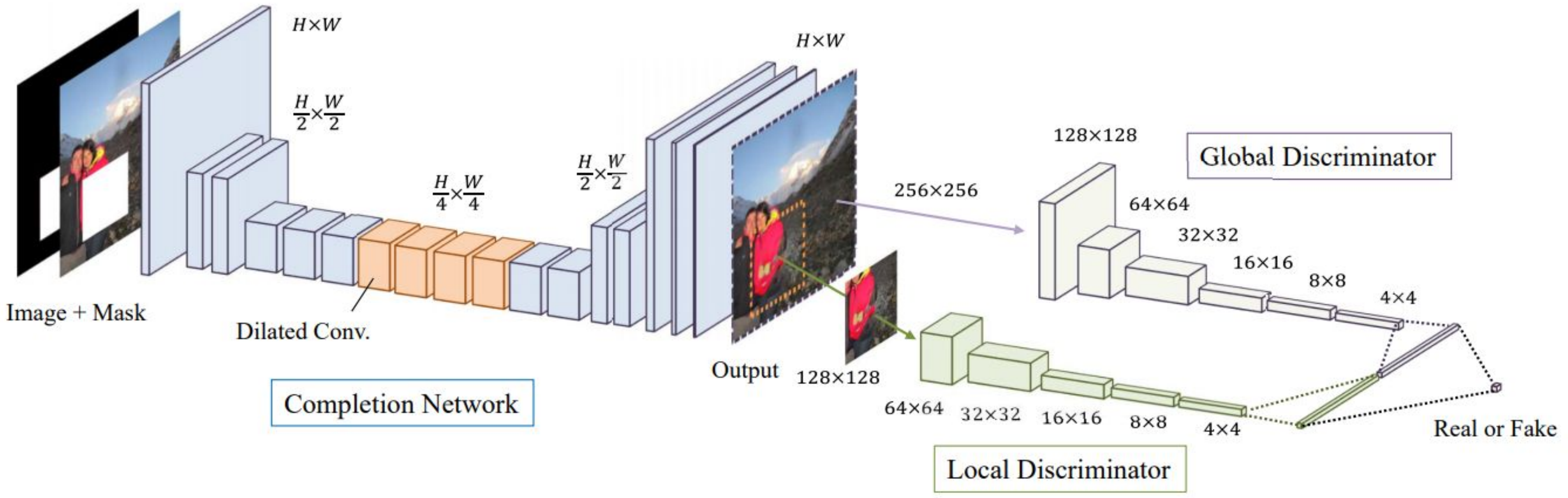
- Compress the area of interest on the image
- Compression function C(r)

$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{x}_i, \mathbf{c}_i \sim p_{\text{data}}(\mathbf{x},\mathbf{c})}[\log D(\mathbf{x}_i, \mathbf{c}_i)] + \mathbb{E}_{\mathbf{c_i} \sim p_{\mathbf{c}}, G(\cdot) \sim p_{\mathbf{z}}}[\log 1 - D(G(\mathbf{z}_i, \mathbf{c}_i))] + ||G(\mathbf{z}_i, \mathbf{c}_i) - \mathbf{x}_i||_1 + ||C(G(\mathbf{z}_i, \mathbf{c}_i) - \mathbf{c}_i||_2 \quad (2)$$

EXEMPLAR GANS FOR IN-PAINTING



EXEMPLAR GANS FOR IN-PAINTING: MODEL ARCHITECTURE



EXEMPLAR GANS FOR IN-PAINTING: MODEL ARCHITECTURE

- Adversarial Loss: in-painted image vs reference image
  - Global Adversarial Loss: enforcing overall semantic consistency
  - Local Adversarial Loss: ensures detail consistency and sharpness
- Perceptual Loss: optional loss measuring distance of generated image to the original reference
- Reconstruction Loss: in-painted image vs ground truth image

RESULTS

(a) (b) (c) (d)

Figure 4: Comparison between (a) ground truth, (b) non-exemplar and (c, d) exemplar-based results. An ExGAN that uses a reference image in the generator and discriminator is shown in column (c), and an ExGAN that uses a code is shown in column (d).

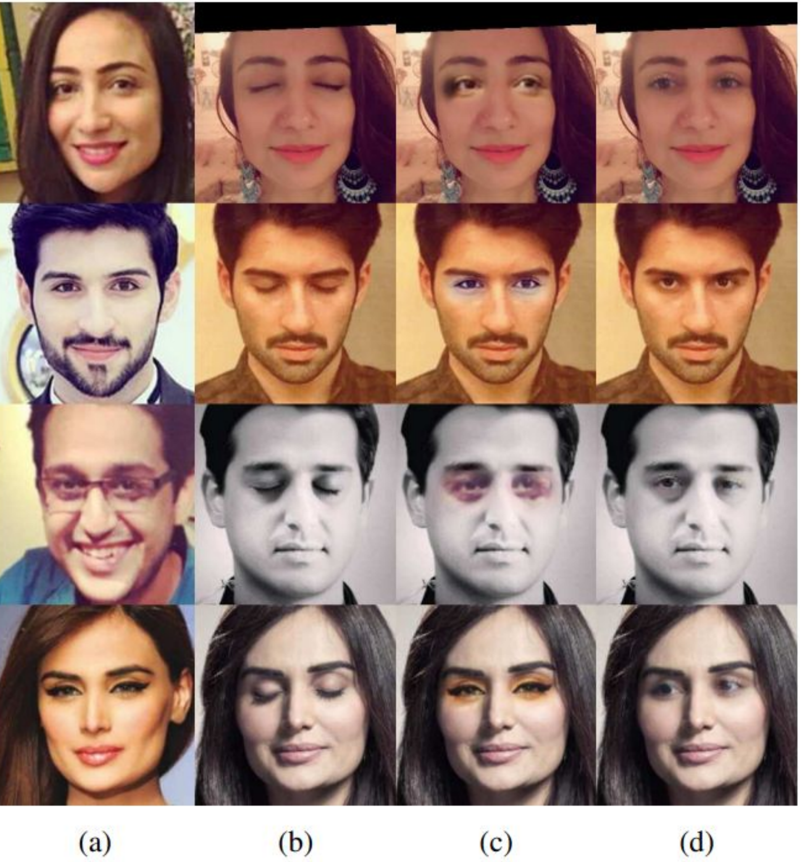


Figure 1: Comparison between the commercial state of the art eye opening algorithm in Adobe Photoshop Elements [1] (c) and the proposed ExGAN technique (d). The exemplar and original images are shown in (a) and (b), respectively.

RESULTS

THANK YOU :)