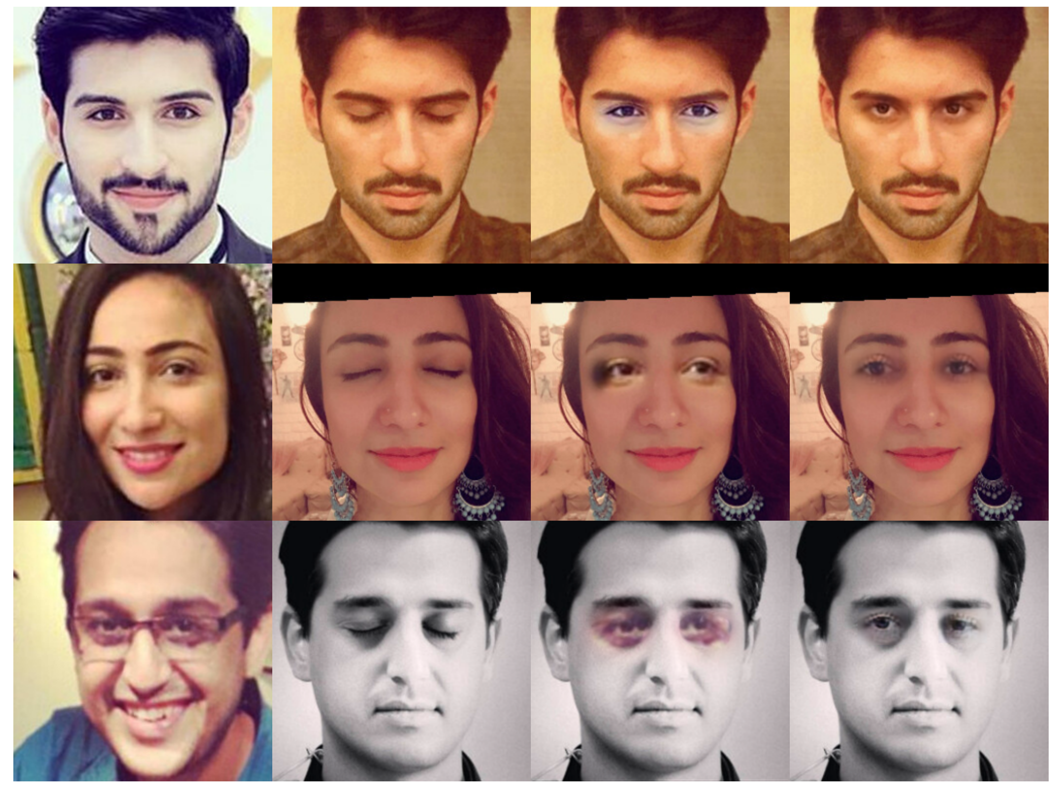
Eye In-Painting with Exemplar Generative Adversarial Networks

Brian Dolhansky and Cristian Canton Ferrer, Facebook, {bdol, ccanton}@fb.com

Overview

We introduce a novel approach to in-painting where the identity of the object to remove or change is preserved and accounted for at inference time: Exemplar GANs (ExGANs). ExGANs are a type of conditional GAN that utilize exemplar information to produce high-quality, personalized in-painting results. We propose using exemplar information in the form of a reference image of the region to in-paint, or a perceptual code describing that object. We show that using exemplars produce better results than existing in-painting GANs or other deterministic approaches.

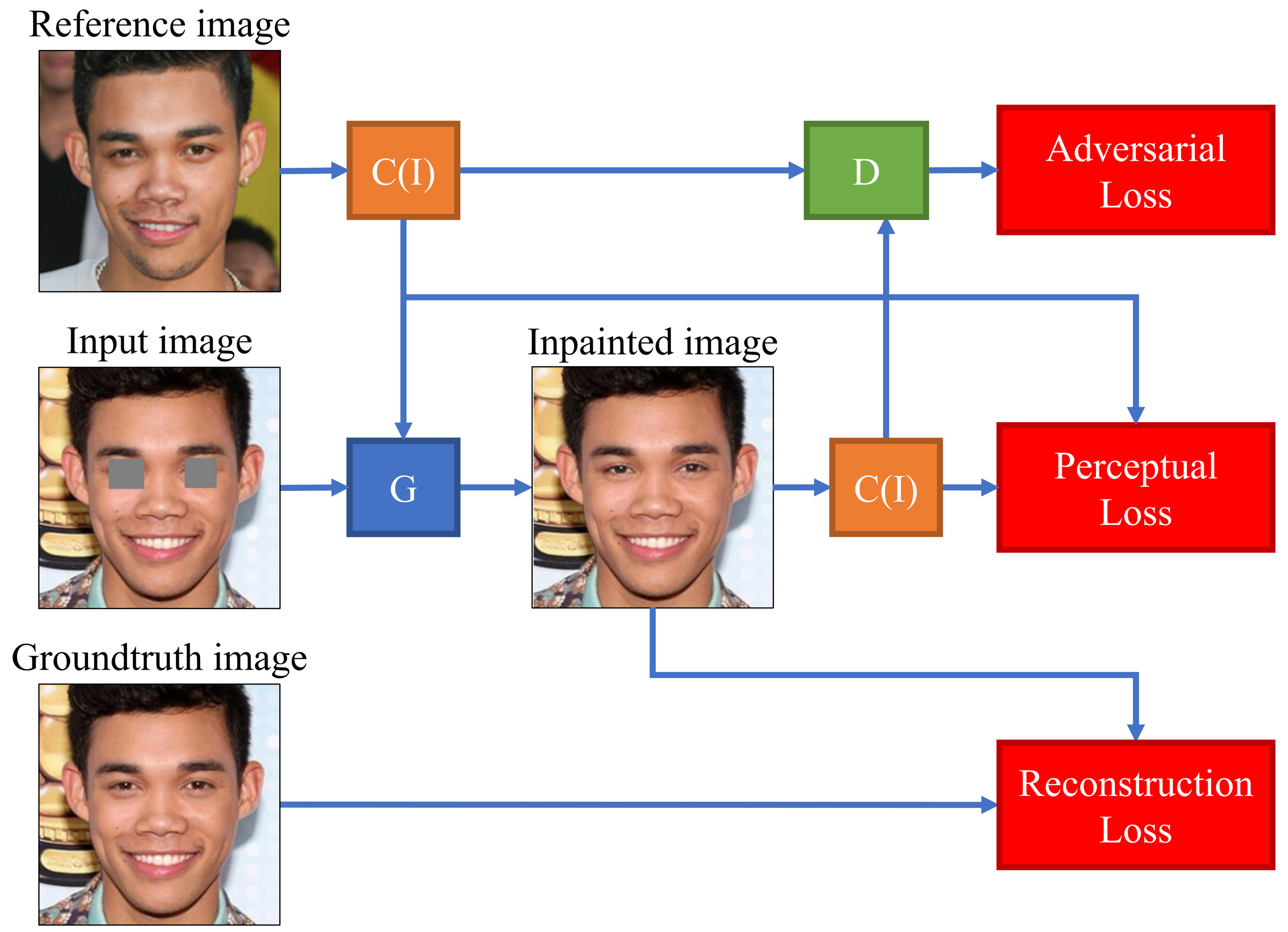
Motivation



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

Existing approaches to opening closed eyes do not account for all of lighting, pose, and overall coloring. Most importantly, general GANs do not account for the imdividual identity in an image to in-paint. Here, (a) are reference images, (b) are images to in-paint, (c) are eyes that have been opened with Adobe Photoshop Elements, and (d) are results produced with our method.

ExGAN formulation



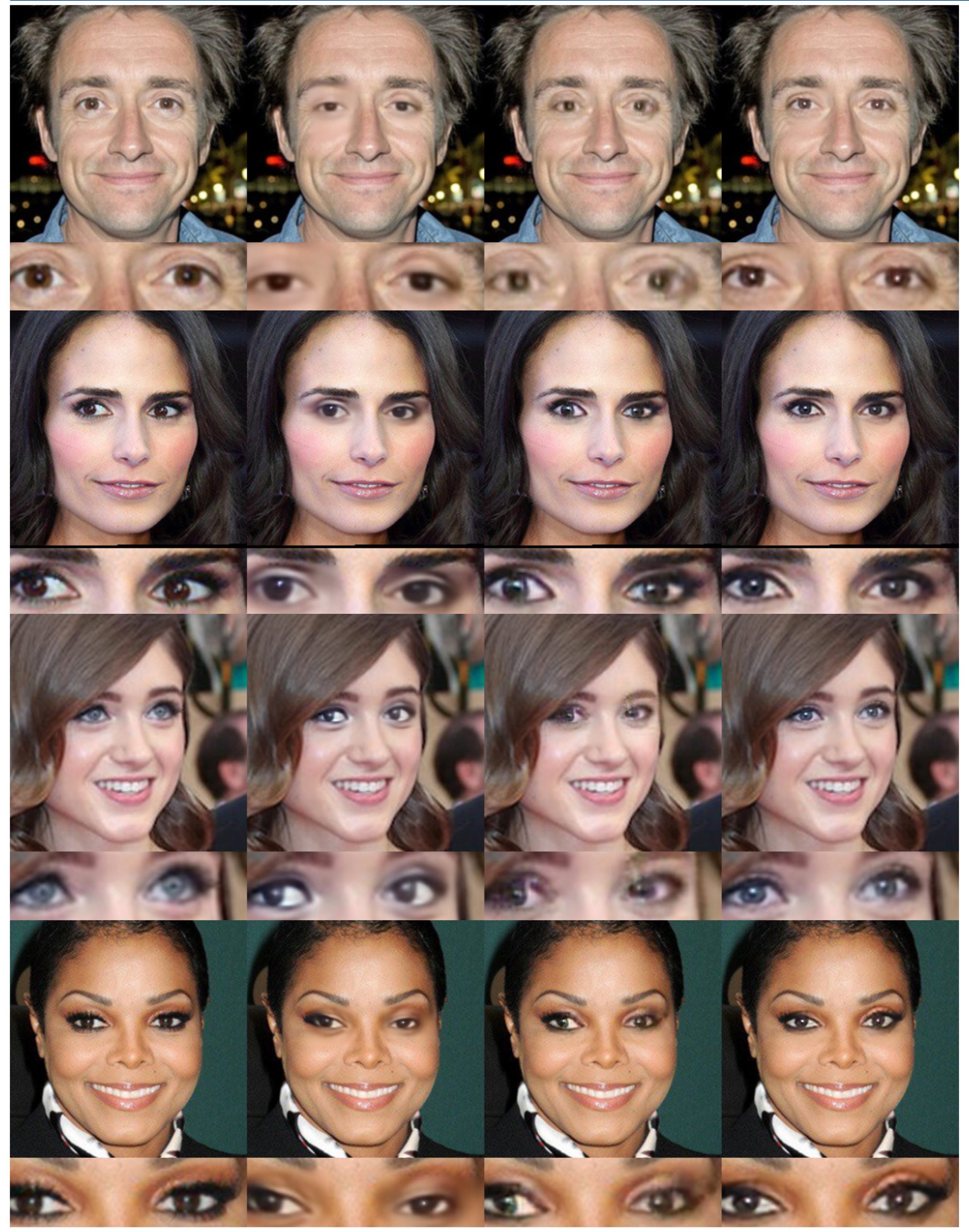
$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{x}_i,\mathbf{r}_i \sim p_{\mathrm{data}}(\mathbf{x},\mathbf{r})}[\log D(\mathbf{x}_i,\mathbf{r}_i)] + \mathbb{E}_{\mathbf{r}_i \sim p_{\mathbf{r}}, G(\cdot)\sim p_{\mathbf{z}}}[\log 1 - D(G(\mathbf{z}_i,\mathbf{r}_i))] + ||G(\mathbf{z}_i,\mathbf{r}_i) - \mathbf{x}_i||_1$$

Eq. 1: Reference-based exemplar inpainting

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

Eq. 2: Code-based exemplar inpainting

Results



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

Direct quality comparison between (a) original image, (b) non-exemplar GAN [1], (c) reference ExGAN and (d) code-based ExGAN. All networks in-painted the original image in (a) with eyes removed.

(a) (b) (c)

More eye-opening results with a reference ExGAN that uses a reference image in (a).

(a) (b) (c)

Justification that FID score is a better proxy for perceptual quality than L1 loss. (a) is the original image. Images in (b) have a higher L1 loss than those in (c), but a lower FID score.

Some failure cases of our model include not in-painting the correct iris color, or failing when an eye is hidden by occlusions.

References

1. *Globally and Locally Consistent Image Completion*, Iizuka et al. SIGGRAPH 2017