Photoshop 2.0: Generative Adversarial Networks for Photo Editing

Homero Roman, Michelle Zhang, Brandon Yang
CS231N Spring 2017

Background

- Generative Adversarial Networks (GANs) generate images from a min-max game between generator and discriminator
- Recent research has focused on improving GAN architectures for training and image quality, such as WGAN
- We focus on building an architecture to directly train autoencoders for GANs to better understand the learnt latent space representation of these networks
- Previous research into conditional GANs conditions GAN training on labeled data for more specific image generation,
- Preliminary results in the DCGAN paper illustrate the potentials of latent space traversal and selective dropout for tuning image generation and semantic understanding

Dataset

- The **celebA** dataset contains over 200,000 face images of various celebrities
- Labeled with 40 different binary attributes for facial features, including hair color, gender, relative age, and whether or not the person is smiling.

Models

- To train a GAN auto-encoder we first trained the DCGAN model for 25 epochs and used the trained generator weights as our encoder
- We experimented with three approaches to encoder networks with the L2 similarity metric
  - Fully connected encoder model baseline
  - Deep convolutional encoder
  - Transfer learning encoder with weights from the trained discriminator from the DCGAN model
- We then experimented with different loss functions to produce more realistic encodings: L1, L2, and SSIM

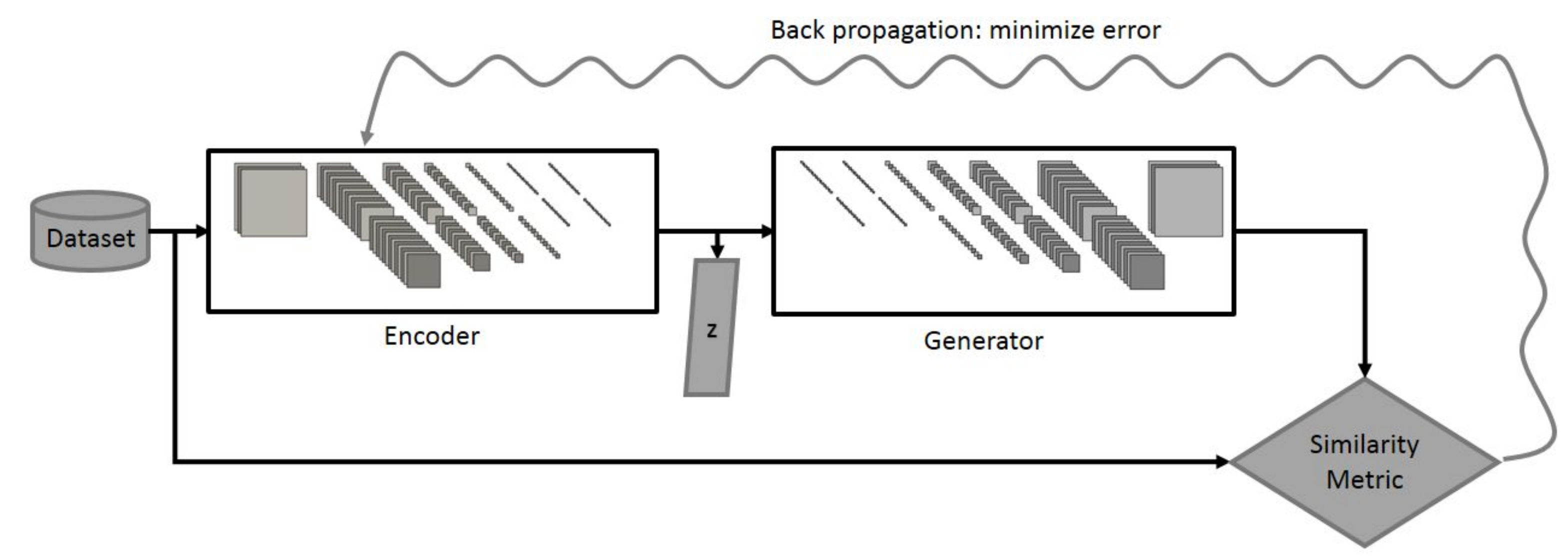


Figure 1: GAN Autoencoder and Generator network overview.

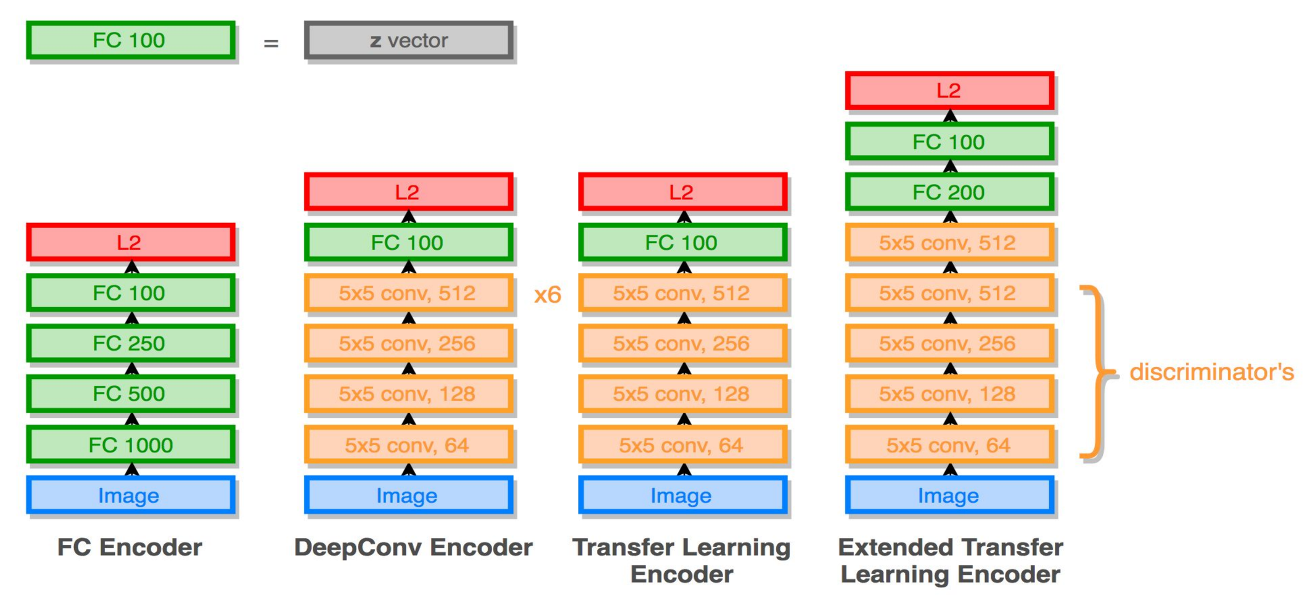
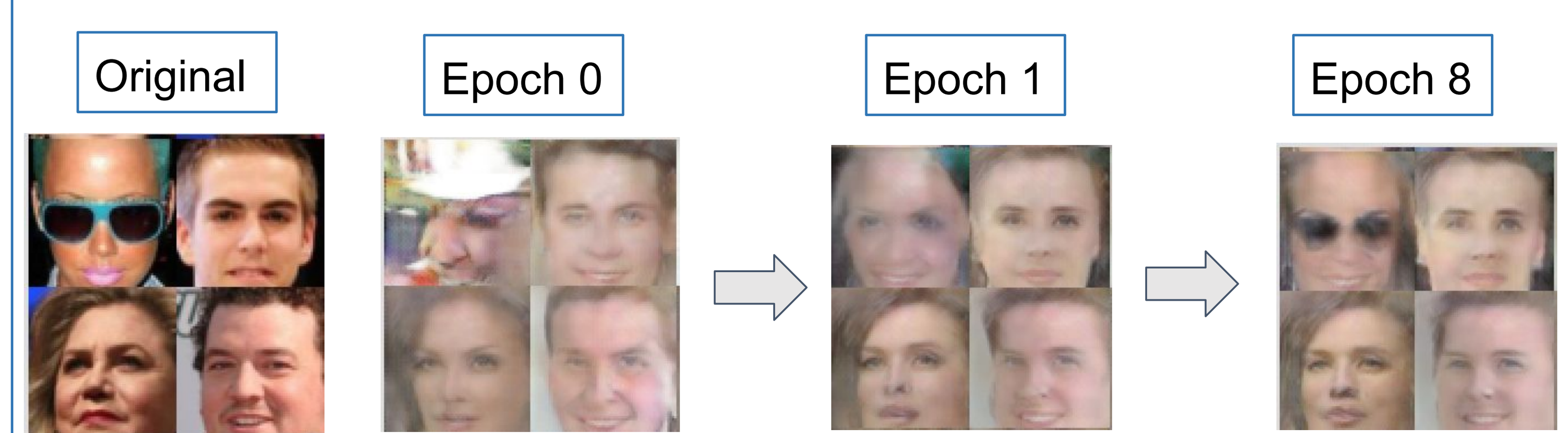


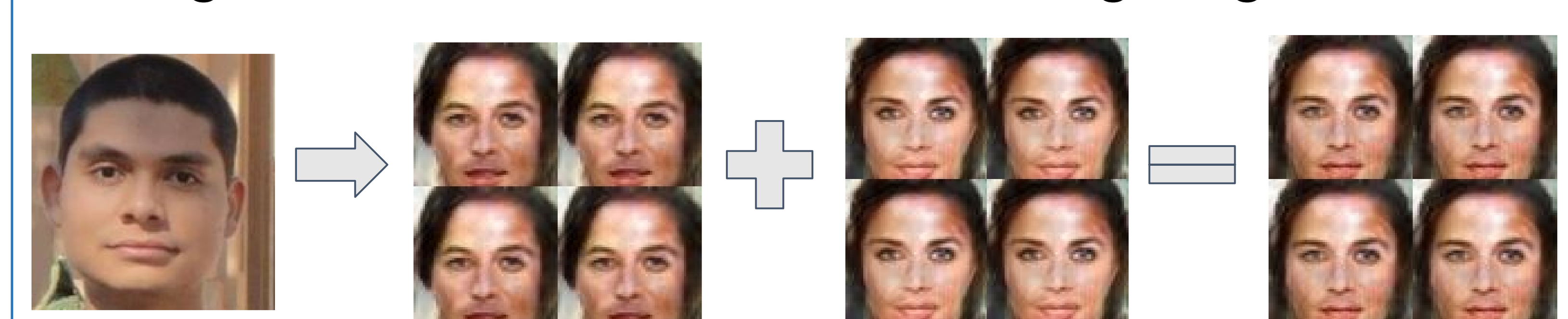
Figure 2: Encoder network architectures.

Qualitative Results

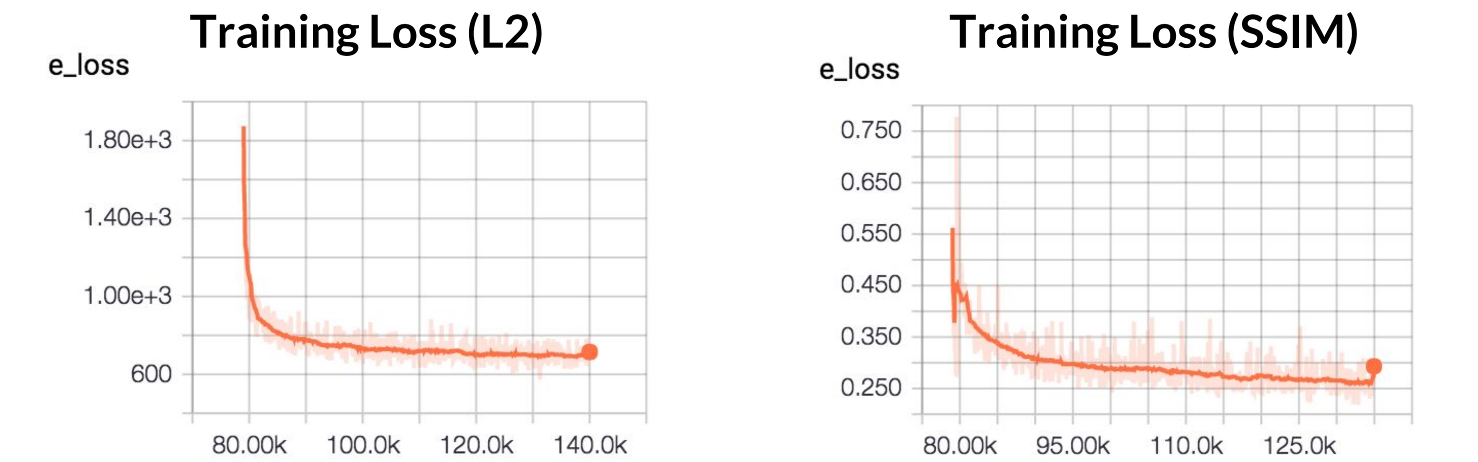
Training the extended transfer learning encoder :



Using GAN autoencoder to create smiling image



Evaluation



$$L_2(x,y)=\sum_{j=1}^{n}(x_j-y_j)^2$$

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)}$$

Figure 3: Training loss curves for transfer learning encoder.

	L2 Loss	L1 Loss	SSIM Loss
FC Encoder	668.223	-	-
DeepConv Encoder	525.860	-	-
Transfer Learning Encoder	641.906	1990.323	0.293
Extended Transfer Learning Encoder	701.273	-	-

Figure 4: Loss results.

Conclusions and Future Work

- GAN autoencoders can be effectively created by transfer learning from the discriminator network
- Manipulating encoded images in the latent vector space with simple vector operations to generate images
- Our model seems limited by inherent biases in the generator and dataset (female faces), which can perhaps be addressed with different generator approaches
- In the future, we want to explore better quantitative metrics for generated image quality and possible end to end architectures for image processing
- Explore different operations in the latent space and selective dropout and effects on generated images

References

1. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. 2014.
2. G. Heinrich. Photo editing with generative adversarial networks, Apr. 2017.
3. T. Kim. Neural face. https://github.com/carpedm20/DCGAN-tensorflow, 2016.
4. A. B. L. Larsen, S. K. Sønderby, and O. Winther. Autoencoding beyond pixels using a learned similarity metric. CoRR, abs/1512.09300, 2015.
5. Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, Dec. 2015.
6. A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR, abs/1511.06434, 2015.
7. T. White. Smile vector.