Synthetic Image Generation

with a focus on GANs

Koray Ulusan

University of Tübingen

koray.ulusan@student.uni-tuebingen.de

ProSeminar: Deep Neural Networks
4.12.2023

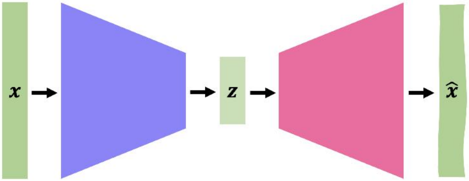
Overview

1. Earlier Approaches
2. Generative Adversarial Networks (GANs)
3. Modern Methods
4. Ethical Aspects
5. Q&A

Image Generation Before GANs

Deep generative models

- Deep Belief Networks [Hinton et al., 2006]
- VAEs [Kingma and Welling, 2013]



AutoEncoder

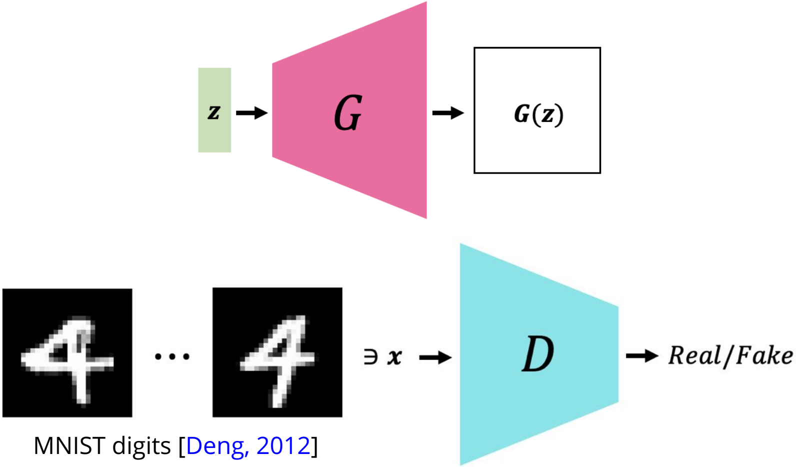
Slice of generated samples [Kingma and Welling, 2013]

Image Classification Models

- GoogLeNet [Szegedy et al., 2014]
- VGG [Simonyan and Zisserman, 2014]
- ResNet [He et al., 2015]

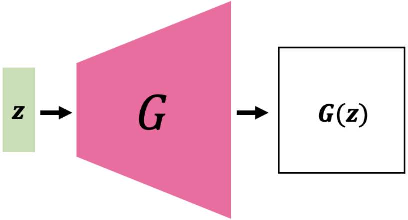
Generative Adversarial Networks (GANs)

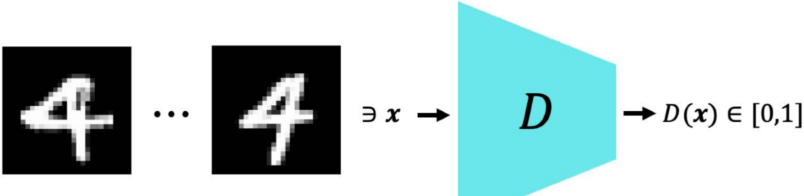
- Generative Adversarial Nets [Goodfellow et al., 2014]



MNIST digits [Deng, 2012]

Value Function





MNIST digits [Deng, 2012]

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

Training Procedure I

GAN training algorithm [Goodfellow et al., 2014]

for number of training iterations **do**

for k steps **do**

- Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
- Sample minibatch of m examples $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$ from data generating distribution $p_{\text{data}}(\boldsymbol{x})$.
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right].$$

end for

- Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right).$$

end for

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

Training Procedure II

for number of training iterations **do**

  for k steps **do**

  - Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
  - Sample minibatch of m examples $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}\}$ from data generating distribution $p_{\text{data}}(\boldsymbol{x})$.
  - Update the discriminator by ascending its stochastic gradient:

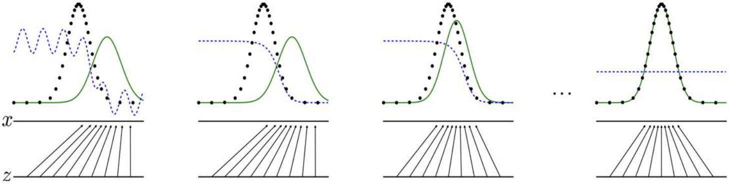
$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right) \right].$$

  end for

- Sample minibatch of m noise samples $\{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(m)}\}$ from noise prior $p_g(\boldsymbol{z})$.
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right).$$

end for



Several steps of training [Goodfellow et al., 2014]

Results

interpolation in z-space [Goodfellow et al., 2014]

Generated samples (using FC or Conv). Green ones are closest images in CIFAR-10 [Goodfellow et al., 2014]

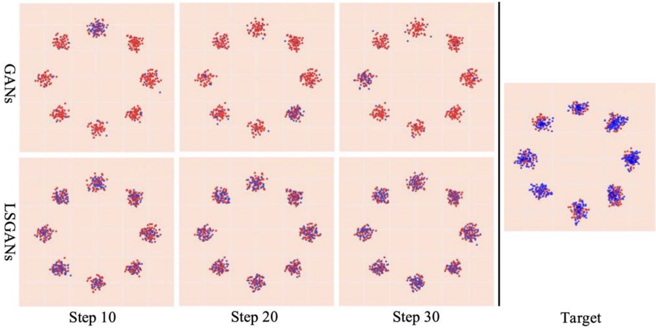
- Fréchet inception distance (FID) [Heusel et al., 2017]

Common Problems

- Mode collapse
- Vanishing gradients
- (non-) convergence

Apple

Orange



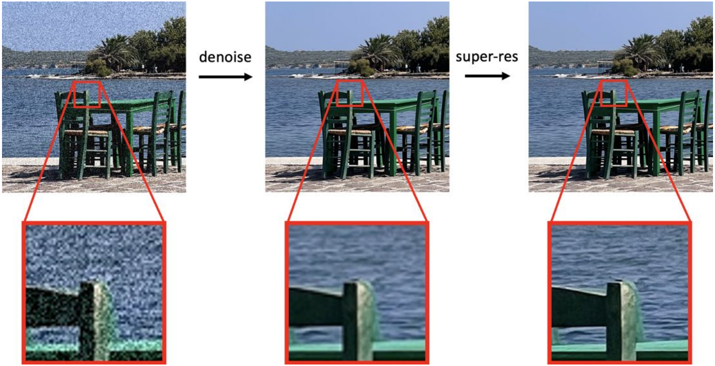
Regular GANs and LSGANs on CIFAR-10 [Shamsolmoali et al., 2020]

GAN Variants I

What can GANs do?

- image super-resolution
- image denoising
- image inpainting
- image fusion

[Shamsolmoali et al., 2020]



GAN Variants II

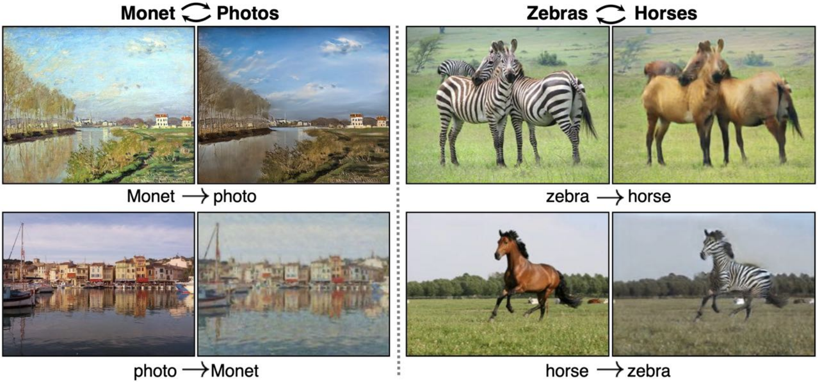
- image captioning
- text-to-image translation
- Multi-stage methods: StackGAN [Zhang et al., 2016]

Generated samples of flowers by StackGAN [Shamsolmoali et al., 2020]

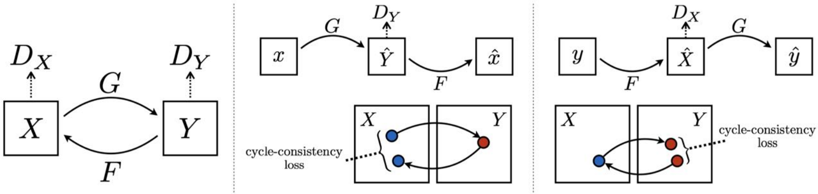
CycleGan I

- Adversarial Domain Adaptation:

  Unpaired Image-to-Image Translation using
  Cycle-Consistent Adversarial Networks [Zhu et al., 2017]



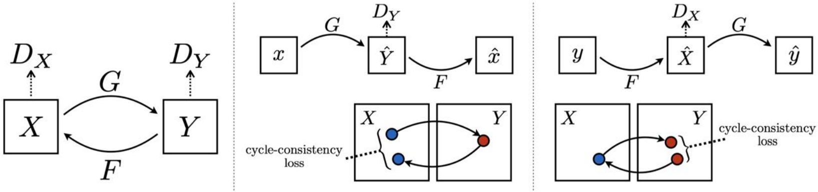
CycleGan II



CycleGAN architecture [Zhu et al., 2017]

“Horse $\rightarrow$ Zebra $\rightarrow$ Horse” cycle [Shamsolmoali et al., 2020]

CycleGan III



$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x)))]$$

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)}[||F(G(x)) - x||_1] + \mathbb{E}_{y \sim p_{data}(y)}[||G(F(y)) - y||_1]$$

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, X, Y) + \lambda \cdot \mathcal{L}_{cyc}(G, F)$$

Note: $\lambda = 10$

State-of-the-Art

- Muse [Chang et al., 2023]
- by Google Research
- Text-To-Image Generation via Masked Generative *Transformers*

- CM3Leon (pronounced "Chameleon") [Yu et al., 2023] (exactly 90 days ago)
- by Meta AI (Facebook)
- *Autoregressive*
- Based on CM3 [Aghajanyan et al., 2022]
  "training ... image tokens (from a VQVAE-GAN) ..."

Ethical Aspects

Chest Radiograph [Desai, 2020]

Blended Signatures [Mattei, 2022]

Obama's Deepfake [MIT video]

"Théâtre D'opéra Spatial" [Roose, 2022]

References I

Aghajanyan, A., Huang, B., Ross, C., Karpukhin, V., Xu, H., Goyal, N., Okhonko, D., Joshi, M., Ghosh, G., Lewis, M., and Zettlemoyer, L. (2022).
Cm3: A causal masked multimodal model of the internet.

Chang, H., Zhang, H., Barber, J., Maschinot, A., Lezama, J., Jiang, L., Yang, M.-H., Murphy, K., Freeman, W. T., Rubinstein, M., Li, Y., and Krishnan, D. (2023).
Muse: Text-to-image generation via masked generative transformers.

Deng, L. (2012).
The mnist database of handwritten digit images for machine learning research.
IEEE Signal Processing Magazine, 29(6):141–142.

Desai, S., B.-A. W. T. A.-S. S. G. K. F. P. R. M. B. G. N. T. P. T. S. K. B. W. P. F. (2020).
hest imaging with clinical and genomic correlates representing a rural covid-19 positive population [data set].

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014).
Generative adversarial nets.
Advances in neural information processing systems, 27.

He, K., Zhang, X., Ren, S., and Sun, J. (2015).
Deep residual learning for image recognition.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017).
Gans trained by a two time-scale update rule converge to a local nash equilibrium.

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006).
A fast learning algorithm for deep belief nets.
Neural computation, 18(7):1527–1554.

References II

Kingma, D. P. and Welling, M. (2013).
Auto-encoding variational bayes.

Mattei, S. E.-D. (2022).
What Are Those Signatures Doing in LensaAI Portraits?
ARTnews.

Roose, K. (2022).
AI-Generated Art Won a Prize. Artists Aren't Happy.
N.Y. Times.

Shamsolmoali, P., Zareapoor, M., Granger, E., Zhou, H., Wang, R., Celebi, M. E., and Yang, J. (2020).
Image synthesis with adversarial networks: a comprehensive survey and case studies.

Simonyan, K. and Zisserman, A. (2014).
Very deep convolutional networks for large-scale image recognition.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2014).
Going deeper with convolutions.

Yu, L., Shi, B., Pasunuru, R., Muller, B., Golovneva, O., Wang, T., Babu, A., Tang, B., Karrer, B., Sheynin, S., Ross, C., Polyak, A., Howes, R., Sharma, V., Xu, P., Tamoyan, H., Ashual, O., Singer, U., Li, S.-W., Zhang, S., James, R., Ghosh, G., Taigman, Y., Fazel-Zarandi, M., Celikyilmaz, A., Zettlemoyer, L., and Aghajanyan, A. (2023).
Scaling autoregressive multi-modal models: Pretraining and instruction tuning.

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., and Metaxas, D. (2016).
Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks.

References III

Zhu, J., Park, T., Isola, P., and Efros, A. A. (2017).
Unpaired image-to-image translation using cycle-consistent adversarial networks.
CoRR, abs/1703.10593.

Questions and Answers

Q&A about:

- something you would like to explore further
- more on GAN training problems
- more on CycleGAN? extensions/limitations
- more on diffusion models
- ...