Energy-based Generative Adversarial Network

Junbo Zhao, Michael Mathieu, Yann LeCun

NYU, Facebook AI Research

Gio Paik
giopaik@naver.com

GAN: Generative Adversarial Net

- One of the most important technology in Deep Learning History.
- Creates Realistic things from nothing.
- Can create images, musics, poem… almost everything human can create.

Photos of non-existent people created by GAN.

Figure from Tero Karras et al, A Style-Based Generator Architecture for Generative Adversarial Networks, CVPR 2019

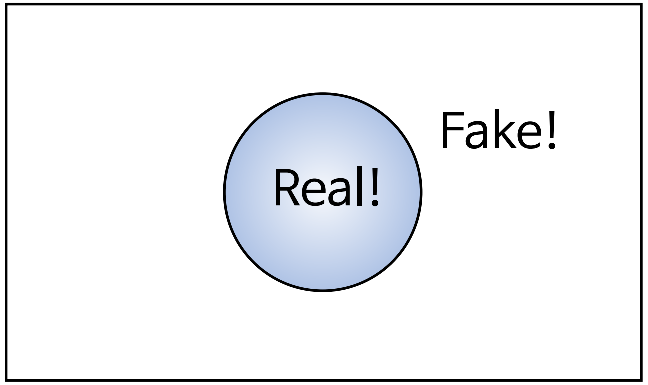
GAN: How it works?

- We have two models:
  - Generator
  - Discriminator
- Generator tries to create fake images looks like real images.
- Discriminator tries to distinguish which image is real and which image isn't.
- This means Generator learns how to create realistic image by the output of Discriminator.
  - If Discriminator returns true to Generator's fake image, It means Generator successfully created realistic image.

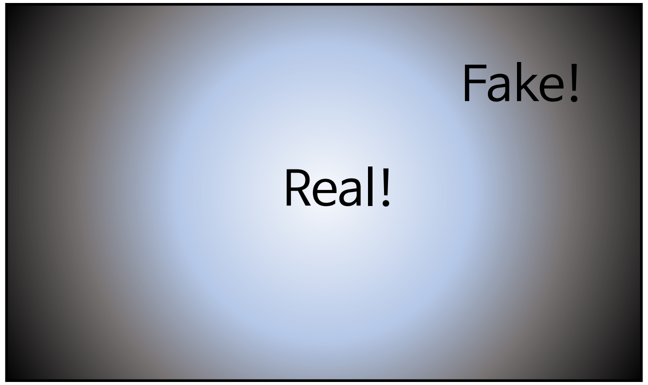
The limitation of GAN

- Discriminator evaluates Generator's image is real or fake.
  - So Discriminator returns a value between 0 to 1.
  - Generator learns real-data's manifold roughly.
  - We call this kind of GANs as probabilistic GANs.
- What if Discriminators output could be more vary?
  - Generator can learn real-data's Manifold well!

Probabilistic GANs

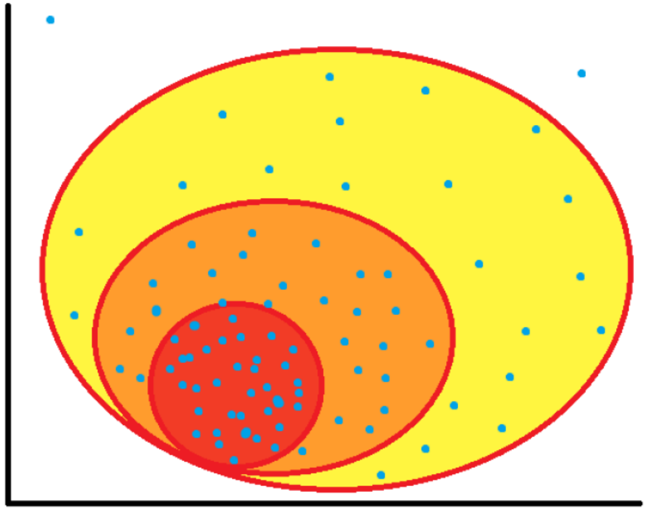


Energy Based GANs



Energy Based Model

- We map every point of input space as single scalar.
- Energy Function returns low energy in high density region, and high energy in low density region.
- Good Generative model should create image like real (looks similar with original data)



Discriminator as an Energy Function

- In probabilistic GANs, Discriminator tries to classify $G(z)$ to y.
- In EBGAN, Discriminator return the energy of create image $\mathrm{G}(z)$.
- If energy of $\mathrm{G}(z)$ is high, it means $\mathrm{G}(z)$ doesn’t look like real.
- If energy of $\mathrm{G}(z)$ is low, it means $\mathrm{G}(z)$ looks like one of the real images.

Auto-Encoder Discriminator

- The Discriminator D is structured as an auto−encoder.

$$D(x) = \left\| Dec\big(Enc(x)\big) - x \right\|$$

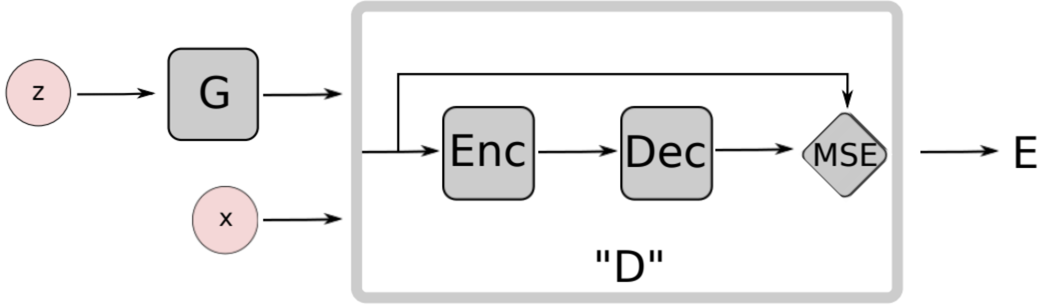


Figure 1: EBGAN architecture with an auto-encoder discriminator.

- x is a data sample.
- D trained on data set X will reconstruct data x well, So MSE will be minimized.
- But D can’t reconstruct the inputs that are far from X,So MSE will be maximized.

Figure from the paper

Regularizing Auto Encoder

- Auto Encoder model in D may attributes 0 energy to the whole space.
- In order to avoid this problem, we should push the model to give higher energy to points outside the data manifold.
- In this paper, they used Repelling Regularizer.

Loss Functions

- Given a positive margin m, a data sample x and a generated sample $G(z)$, The loss functions of D and G is:

$$\mathcal{L}_D(x,z) = D(x) + \max\Big(0, m - D\big(G(z)\big)\Big)$$
$$\mathcal{L}_G(z) = D\big(G(z)\big)$$

- Minimizing $\mathcal{L}_G$ is similar to maximizing the second term of $\mathcal{L}_D$.
- It has the same minimum but non-zero gradients when $D\big(G(z)\big) \geq m$.

Energy-based Generative Adversarial Network

Junbo Zhao, Michael Mathieu, Yann LeCun

NYU, Facebook AI Research

Thank you for watching!

Gio Paik
giopaik@naver.com