Lecture 8: Generative Models (Part 1: GANs)

Just kidding—They don’t exist
Agenda (Today: GANs)

1. Goal, motivation, Basic methods
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)

Tutorial (next week): Non-Adversarial methods (VAE, VQVAE, IMLE)
Goal

Images
Text
Audio
Video
Whatever...

Why?

- Because it’s cool
- Training data
- Computer graphics
- Image to image
- Image processing
- Apps (emojis, faceapp)
- Simulations

The correct approach when there is more than one valid solution!
Generative methods:

• Parametric distribution estimation (e.g. GMM)

• Autoregressive models (e.g. RNN, Causal CNN, Transformer)

• Latent space mapping (VAE, GAN, more)
Basic approach:

- Density estimation
Example: GMM

Step 1: observe a set of samples

Step 2: assume a GMM model

\[ p(x | \theta) = \sum_i \pi_i \mathcal{N}(x | \mu_i, \Sigma_i) \]

Step 3: perform maximum likelihood learning

\[ \max_{\theta} \sum_{x^{(j)} \in \text{Dataset}} \log p(\theta | x^{(j)}) \]

Step 4: Sample
Agenda

1. Goal, motivation, Basic methods
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)
Latent space mapping approach

Data likelihood: \( p_\theta(x) = \int p_\theta(z)p_\theta(x|z)dz \)
How about this idea for a generative model?
No good!

In expectation: every noise is mapped to every instance

Best L2 solution: All noise is mapped to the mean
(For images: ~ grey image)
Generative Adversarial Networks

# of GAN related papers per year (Salehi et al.)
Q: What makes a good counterfeiter?

A: Can fool a good cop

Minimax game: Make the best cop do the worst mistake

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

Q: Who do you train first?

A: Alternate training! G,D,G,D,...
FAQ1: Why does it work?

• D learns probability! G trains to sample instance with high probability!

• Objective does not determine mapping directly- arrangement of latent space is learned!

• Theory: minimizes JS divergence between generated and real distributions.
FAQ2: Why alternating?

• Gradients are meaningless when game is unbalanced.

• Pre-train D? Negative examples?

• Pre-train G? What loss?
  For G, D is a learned loss function
GANs,
Goodfellow
2014
Latent space interpolation
Why does it work?

1. Every point is mapped to a valid example.
2. Network is continuous.
Evaluation metrics: Inception score

\[ IS(G) = \exp \left( \mathbb{E}_{x \sim p_a} \ D_{KL} \left( p(y|x) \parallel p(y) \right) \right) \]
1. Fréchet Inception Distance (FID)

Depends on the number of samples!

2. \( \text{FID}(x, g) = \|\mu_x - \mu_g\|^2 + \text{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{1/2}) \)
Agenda

1. Goal, motivation, Basic methods

2. Introduction to GANs (basic setup, intuition, eval measures)

3. Image to image (pix2pix, CycleGAN)

4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)

5. Progressive GANs (PGGAN, StyleGAN)

6. Scaling up GANs (BigGAN)

7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)
Image to Image translation

Isola et al. Nov 2016

Slide by Sefi Bell-Kligler & Akhiad Bercovich
Conditional GAN

\[ \mathcal{L}_{C-GAN} = \min_G \max_D \mathbb{E}[\log D(y, x)] + \mathbb{E}[\log(1 - D(G(x), x))] \]

Slide by Sefi Bell-Kligler & Akhiad Bercovich
Pix2Pix

\[ \mathcal{L}_{C-GAN} = \min_{G} \max_{D} \mathbb{E}[\log D(y, x)] + \mathbb{E}[\log(1 - D(G(x), x))] \]

\[ \mathcal{L}_{L1} = \|y - G(x, z)\|_1 \]

**Objective** = \( \mathcal{L}_{C-GAN} + \lambda \cdot \mathcal{L}_{L1} \)
Generative VS Discriminative
What would happen if we train regular supervised mapping?

What would happen if we train regular supervised mapping?
True AI needs no explicit supervision
CycleGAN (Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros)

Cycle-consistency Loss $L_1$

Figure from https://hardikbansal.github.io/CycleGANBlog/
CycleGAN (Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros)

Figure from https://hardikbansal.github.io/CycleGANBlog/
Agenda

1. Goal, motivation, Basic methods
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)
Training GANs is hard

- Stability
- Mode collapse
- GANs can over-train

ADDICTS: BEFORE AND AFTER

ALCOHOL  HEROIN

COCAINE  Training GANs

D is happy!
Efficient training tricks (the basic ones)

- Batch discrimination  Salimans, Goodfellow et al. Improved techniques for training GANs 2015
- Use LeakyReLU in D
- Use BatchNorm (possible- InstanceNorm)
- Apply Spectral Norm
- Modified objective/loss?
Types of GAN losses  - Wasserstein GAN

\[
\text{GAN} \quad \max_D E_{x \sim p_X} [\log D(x)] + E_{z \sim p_Z} [\log (1 - D(G(z)))]
\]

\[
\text{WGAN} \quad \max_D E_{x \sim p_X} [D(x)] - E_{z \sim p_Z} [D(G(z))]
\]

WGAN: minimize earth mover distance between \(p_X\) and \(p_{G(Z)}\)

\[
EM(p_X, p_{G(Z)}) = \inf_{\gamma \in \Pi(p_X, p_{G(Z)})} E_{(x,y) \sim \gamma} [||x - y||]
\]
Types of GAN losses

Are GANs Created Equal?

Lucic & Kurach et al. (Google Brain)

2018

Mark Saroufim

Dataset = FASHION-MNIST

Dataset = CELEBA

if_worse

init
→ find_sota_on_arxiv
→ find_code_on_github
→ random_changes
→ publish

Mark Saroufim
Mode-Seeking GAN (Mao&Lee et al; CVPR’18)
Multi Scale D

\[ \mathcal{L}_{GAN}(G, D_1) \]

Real/Fake

\[ \mathcal{L}_{GAN}(G, D_2) \]

Real/Fake

\[ \mathcal{L}_{GAN}(G, D_3) \]

Real/Fake

Wang, Tao et al. CVPR2018
Agenda

1. Goal, motivation, Basic methods
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)
Progressive Growing of GAN, Karras et al., Feb 2018

Progressive Growing of GAN, Karras et al., Feb 2018
Progressive Growing of GAN, Karras et al., Feb 2018
Style Modules (AdaIN)

The generator’s Adaptive Instance Normalization (AdaIN)

StyleGAN, Karras et al. NVIDIA 2019
Results

Source A: gender, age, hair length, glasses, pose

Source B: everything else

Result of combining A and B
Deepfake Detection Challenge
Identify videos with facial or voice manipulations

Deepfake Detection Challenge · 543 teams · 3 months to go (2 months to go until merger deadline)

$1,000,000
Prize Money

Animation by Sefi Bell-Kligler & Akhiad Bercovich

https://thispersondoesnotexist.com/
Agenda

1. Goal, motivation, Basic methods
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)
SAGAN, BigGAN - Scaling up GANs

- Class Conditional BatchNorm
- Self-Attention (both D and G)
- Split noise, and use skip connections for its chunks
- Bigger batches and number of channels per layer (and deeper)
- Spectral Norm (both D and G)
- Truncation Trick
## Impressive results?

<table>
<thead>
<tr>
<th>StyleGAN branch</th>
<th>BigGAN Branch</th>
</tr>
</thead>
<tbody>
<tr>
<td>One domain (eg. Faces)</td>
<td>Class conditional (Imagenet)</td>
</tr>
<tr>
<td></td>
<td>Very versatile data</td>
</tr>
<tr>
<td></td>
<td>Attention</td>
</tr>
<tr>
<td>Structured data</td>
<td></td>
</tr>
<tr>
<td>Progressive grow</td>
<td></td>
</tr>
<tr>
<td>Amazing results</td>
<td>Some of the results are amazing</td>
</tr>
</tbody>
</table>

Some of the results are amazing.
Agenda

1. Goal, motivation, Basic methods
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)
Everybody Dance Now

Training

Transfer
GAN Dissection


http://gandissect.res.ibm.com/ganpaint.html?project=churchoutdoor&layer=layer4
Training a GAN on a single image

InGAN (Shocher, Bagon, Isola, Irani)

*Input/noise*

*Single training image*

Random samples from a single image

SinGAN (Rott-Shaham, Dekel, Michaeli)
\[ \mathcal{L}_{\text{adversarial}} \]
\[ \mathcal{L}_{\text{diversity}} \text{ (MSGAN)} \]

Shocher, Gandelsman, Mosseri, Yarom, Irani, Freeman, Dekel; Semantic Pyramid for Image Generation; CVPR’2020
Applications

Semantic Image Composition

Out of distribution reference

Feature inverting ➔ Generate by reference

Image re-painting

Image re-labeling

Original label: **Highway**

New label: **Desert road**

Original label: **Mountain**

New label: **Volcano**

Applications
Agenda

1. Goal, motivation, Basic methods
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)
Tricks and Techniques

- Batch discrimination
- Label smoothing
- Use LeakyReLU in D
- Use BatchNorm
- Consider InstanceNorm
- **Apply Spectral Norm**
- Playing with G:D ratio is bad for you
- Add small noise (1/256) to real examples (because D may recognize quantization)

- Consider WGAN
- Consider LSGAN
- **Multiscale D**
- VGG loss
- Strict similarity
- Self attention
- Noise skip connections
- Truncation trick
- **Conditional BatchNorm**
- Mode Seeking GAN
Agenda

1. Goal, motivation, Basic methods
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)
Thanks
This week’s Tutorial: LSTM

Next week’s Lecture: Self-Supervision

Next week’s Tutorial: Non-adversarial Generative Models