An Introduction to Generative Adversarial Networks

Sagie Benaim
Tel Aviv University
Generative Modeling: Density Estimation

Training Data

Density Function
Generative Modeling: Sample Generation

Training Data (CelebA)  

Sample Generator (Karras et al, 2017)
Adversarial Nets Framework

\[ D(x) \text{ tries to be near 1} \]

Differentiable function \( D \)

\( x \) sampled from data

\[ D \text{ tries to make } D(G(z)) \text{ near 0, } \]

\( G \) tries to make \( D(G(z)) \) near 1

\[ D \]

\( x \) sampled from model

Differentiable function \( G \)

\[ \text{Input noise } z \]

(Goodfellow et al., 2014)
Self-Play

1959: Arthur Samuel’s checkers agent

(Silver et al, 2017)
Progress on Face Generation

(Brundage et al, 2018)
BigGAN – Late 2018
From GAN to BigGAN

• Depth and Convolution
• Class-conditional Generation
• Wasserstein GAN
• Self Attention
• BigGAN
No Convolution Needed to Solve Simple Tasks

Original GAN, 2014
Depth and Convolution for Harder Tasks

Original GAN (CIFAR-10)

No convolution

One convolutional layer

DCGAN (ImageNet)

Many convolutional layers (Radford et al, 2015)
From GAN to BigGAN

• Depth and Convolution
• Class-conditional Generation
• Wasserstein GAN
• Self Attention
• BigGAN
Class-Conditional GANs

(Mirza and Osindero, 2014)
AC-GAN: Specialist Generators

(monarch butterfly, goldfinch, daisy, redshank, grey whale)

(Odena et al, 2016)
SN-GAN: Shared Generator

Gray whale

Welsh springer spaniel

Persian cat

(Miyato et al, 2017)
From GAN to BigGAN

- Depth and Convolution
- Class-conditional Generation
- Wasserstein GAN
- Self Attention
- BigGAN
Wasserstein GAN

- Wasserstein Distance: Minimum cost of transporting mass in converting the data distribution $q$ to the data distribution $p$. 
GAN:

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

\[ \nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log \left(1 - D(G(z^{(i)}))\right) \text{ or } \nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} -\log D(G(z^{(i)})) \]

Real image \( x \)

\( z \sim \mathcal{N}(0, 1) \)

\( z \sim \mathcal{U}(-1, 1) \)

Generator

Discriminator

\( D \)

Cost
WGAN

\[ \nabla_w \left[ \frac{1}{m} \sum_{i=1}^{m} f_w(x^{(i)}) - \frac{1}{m} \sum_{i=1}^{m} f_w(g_\theta(z^{(i)})) \right] \]

\[ -\nabla_\theta \frac{1}{m} \sum_{i=1}^{m} f_w(g_\theta(z^{(i)})) \]
<table>
<thead>
<tr>
<th>Discriminator/Critic</th>
<th>Generator</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GAN</strong></td>
<td><strong>GAN</strong></td>
</tr>
<tr>
<td>( \nabla_{\theta_a} \frac{1}{m} \sum_{i=1}^{m} [\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))] )</td>
<td>( \nabla_{\theta_a} \frac{1}{m} \sum_{i=1}^{m} - \log D(G(z^{(i)})) )</td>
</tr>
<tr>
<td><strong>WGAN</strong></td>
<td><strong>WGAN</strong></td>
</tr>
<tr>
<td>( \nabla_{w} \frac{1}{m} \sum_{i=1}^{m} [f(x^{(i)}) - f(G(z^{(i)}))] )</td>
<td>( \nabla_{\theta} \frac{1}{m} \sum_{i=1}^{m} - f(G(z^{(i)})) )</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
w & \leftarrow w + \alpha \cdot \text{RMSProp}(w, g_w) \\
w & \leftarrow \text{clip}(w, -c, c)
\end{align*}
\]
From GAN to BigGAN

- Depth and Convolution
- Class-conditional Generation
- Wasserstein GAN
- Self Attention
- BigGAN
Self-Attention

Use layers from Wang et al 2018
From GAN to BigGAN

• Depth and Convolution
• Class-conditional Generation
• Wasserstein GAN
• Self Attention
• BigGAN
BigGAN

• Scalability: GANs benefit dramatically from scaling. Two architectural changes that improve scalability.
• Robustness: Fine control of the trade-offs between fidelity and variety is possible via the “truncation trick”
• Stability: Devises solutions that minimize the instabilities in Large Scale GANs

Figure 1: Class-conditional samples generated by our model.
Applying GANs

- Semi-supervised Learning
- Model-based optimization
- Extreme personalization
- Program synthesis
Image to Image Translation
<table>
<thead>
<tr>
<th></th>
<th>Supervised</th>
<th>Unsupervised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unimodal</td>
<td>Pix2pix, CRN, SRGAN</td>
<td>DistanceGAN, CycleGAN, DiscoGAN, DualGAN, UNIT, DTN, StarGAN, OST</td>
</tr>
<tr>
<td>Multimodal</td>
<td>pix2pixHD, BicycleGAN</td>
<td>MUNIT, Augmented CycleGAN</td>
</tr>
</tbody>
</table>
Paired

\[ x_i \]

\[ y_i \]

Unpaired

\[ X \]

\[ Y \]
Fully Supervised: pix2pix

Conditional GAN

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G).$$

[Isola et al., CVPR 2017]
Unsupervised: Circular GANs

**DiscoGAN**: “Learning to Discover Cross-Domain Relations with Generative Adversarial Networks”. Kim et al. ICML’17.


Cycle-Consistent Adversarial Networks

[Mark Twain, 1903]

[Zhu et al., ICCV 2017]
Cycle Consistency Loss

\[ \|F(G(x)) - x\|_1 \quad \|G(F(y)) - y\|_1 \]

See similar formulations [Yi et al. 2017], [Kim et al. 2017]
Collection Style Transfer

Photograph
@ Alexei Efros

Monet

Van Gogh

Cezanne

Ukiyo-e
DistanceGAN

• A pair of images of a given distance are mapped to a pair of outputs with a similar distance.
• $|x_i - x_j|_1$ and $|G(x_i) - G(x_j)|_1$ are highly correlated.

Benaim et al., NIPS 2017
Motivating distance correlations

Analysis of CycleGAN’s horse to zebra results

Benaim et al., NIPS 2017
Mode Collapse

• GAN:

  \[ G_{AB} \]

  \[ G_{AB} \]

  \[ \text{Vehicle} \]

  \[ \text{Head} \]

  \[ \text{Vehicle} \]

  \[ \text{Head} \]

  \[ \text{Vehicle} \]

  \[ \text{Head} \]

Cycle:

  \[ G_{AB} \]

  \[ G_{BA} \]

  \[ L_{\text{CONST}} \]

  \[ \text{Vehicle} \]

  \[ \text{Head} \]

  \[ \text{Vehicle} \]

  \[ \text{Head} \]

  \[ \text{Vehicle} \]

  \[ \text{Head} \]

  \[ \text{Vehicle} \]

  \[ \text{Head} \]

Benaim et al., NIPS 2017
More than 2 domains

Choi et al., CVPR 2018
More than 2 domains

Choi et al., CVPR 2018
Modeling multiple possible outputs
BiCycleGAN [Zhu et al., NIPS 2017] (c.f. InfoGAN [Chen et al. 2016])

MAD-GAN [Ghosh et al., CVPR 2018]
Sketch to Image Translation

(a) edges ↔ shoes
(b) edges ↔ handbags

Huang et al., ECCV 2018
Animal Image Translation

(a) house cats → big cats
(b) big cats → house cats
(c) house cats → dogs
(d) dogs → house cats
(e) big cats → dogs
(f) dogs → big cats

Huang et al., ECCV 2018
Content Transfer?
One Shot?

• Not only are we unsupervised, but we have only a single sample in the input domain!

Applications Beyond Computer Vision

• Many other Vision Applications: Photo Enhancement, Image Dehazing
• Medical Imaging and Biology [Wolterink et al., 2017]
• Voice conversion [Fang et al., 2018, Kaneko et al., 2017]
• Cryptography [CipherGAN: Gomez et al., ICLR 2018]
• Robotics
• NLP: Unsupervised machine translation.
• NLP: Text style transfer.
• …
Thank You! Questions?