Conditional Generative Adversarial Networks

Jia-Bin Huang
Virginia Tech
ECE 6554 Advanced Computer Vision
Today’s class

• Discussions

• Review basic ideas of GAN

• Examples of conditional GAN

• Experiment presentation by Sanket
Why Generative Models?

• Excellent test of our ability to use high-dimensional, complicated probability distributions
• Simulate possible futures for planning or simulated RL
• Missing data
  • Semi-supervised learning
• Multi-modal outputs
• Realistic generation tasks

(Goodfellow 2016)
Generative Modeling

- Density estimation

- Sample generation

(Goodfellow 2016)
Adversarial Nets Framework

$D(x)$ tries to be near 1

Differentiable function $D$

$x$ sampled from data

$D$ tries to make $D(G(z))$ near 0,
$G$ tries to make $D(G(z))$ near 1

Differentiable function $G$

Input noise $z$

$x$ sampled from model

(Goodfellow 2016)
Training Procedure

• Use SGD-like algorithm of choice (Adam) on two minibatches simultaneously:
  • A minibatch of training examples
  • A minibatch of generated samples

• Optional: run $k$ steps of one player for every step of the other player.

(Goodfellow 2016)
Minimax Game

\[ J^{(D)} = -\frac{1}{2} \mathbb{E}_{x \sim p_{\text{data}}} \log D(x) - \frac{1}{2} \mathbb{E}_z \log (1 - D(G(z))) \]
\[ J^{(G)} = -J^{(D)} \]

- Equilibrium is a saddle point of the discriminator loss
- Resembles Jensen-Shannon divergence
- Generator minimizes the log-probability of the discriminator being correct

(Goodfellow 2016)
Discriminator Strategy

• Optimal discriminator

\[
D(x) = \frac{p_{\text{data}}(x)}{p_{\text{data}}(x) + p_{\text{model}}(x)}
\]

Estimating this ratio using supervised learning is the key approximation mechanism used by GANs.
Non-Saturating Game

\[ J^{(D)} = -\frac{1}{2} \mathbb{E}_{x \sim p_{\text{data}}} \log D(x) - \frac{1}{2} \mathbb{E}_z \log (1 - D(G(z))) \]

\[ J^{(G)} = -\frac{1}{2} \mathbb{E}_z \log D(G(z)) \]

- Equilibrium no longer describable with a single loss
- Generator maximizes the log-probability of the discriminator being mistaken
- Heuristically motivated; generator can still learn even when discriminator successfully rejects all generator samples

(Goodfellow 2016)
Review: GAN

• GANs are generative models that use supervised learning to approximate an intractable cost function

• GANs can simulate many cost functions, including the one used for maximum likelihood

• Finding Nash equilibria in high-dimensional, continuous, nonconvex games is an important open research problem
Conditional GAN

• Learn $P(Y|X)$

[Ledig et al. CVPR 2017]
Image Super-Resolution

- Conditional on low-resolution input image

bicubic (21.59dB/0.6423)
SRResNet (23.53dB/0.7832)
SRGAN (21.15dB/0.6868)
original

[Ledig et al. CVPR 2017]
Image-to-Image Translation

- Conditioned on an image of different modality
- No need to specify the loss function

[Isola et al. CVPR 2017]
Positive examples

Real or fake pair?

\[ \text{D} \]

\[
\begin{array}{c}
\text{G} \\
\text{tries to synthesize fake} \\
\text{images that fool D}
\end{array}
\]

\[
\begin{array}{c}
\text{D} \\
\text{tries to identify the fakes}
\end{array}
\]

Negative examples

Real or fake pair?

\[ \text{D} \]
Label2Image

[Isola et al. CVPR 2017]
Edges2Image

[Isola et al. CVPR 2017]
Generative Visual Manipulation

[Zhu et al. ECCV 2016]
\[ z^* = \arg \min_{z \in \mathcal{Z}} \left\{ \sum_{g} \| f_g(G(z)) - v_g \|^2 + \lambda_s \cdot \| z - z_0 \|^2 + E_D \right\} \]

The image shows a process involving an original photo, projection on a manifold, editing UI, and different degrees of image manipulation. The text provides a mathematical formulation for a problem, likely related to image editing or manipulation, with a solution minimizing a cost function involving a data term and manifold smoothness term. The reference [Zhu et al. ECCV 2016] indicates the source of the formulation.
this small bird has a pink breast and crown, and black primaries and secondaries.

this magnificent fellow is almost all black with a red crest, and white cheek patch.

the flower has petals that are bright pinkish purple with white stigma

this white and yellow flower have thin white petals and a round yellow stamen

[Reed et al. ICML 2016]
Text2Image

Positive samples:
- real image + right texts

Negative samples:
- fake image + right texts
- Real image + wrong texts

[Reed et al. ICML 2016]
StackGAN

This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face

This bird is white with some black on its head and wings, and has a long orange beak

This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments

(a) Stage-I images

(b) Stage-II images

[Zhang et al. 2016]
Plug & Play Generative Networks

<table>
<thead>
<tr>
<th>redshank</th>
<th>ant</th>
<th>monastery</th>
</tr>
</thead>
<tbody>
<tr>
<td>volcano</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Nguyen et al. 2016]
Video GAN


[Vondrick et al. NIPS 2016]
Generative Modeling as Feature Learning

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chance</td>
<td>0.9%</td>
</tr>
<tr>
<td>STIP Features [10]</td>
<td>43.9%</td>
</tr>
<tr>
<td>Temporal Coherence [3]</td>
<td>45.4%</td>
</tr>
<tr>
<td>Shuffle and Learn [42]</td>
<td>50.2%</td>
</tr>
<tr>
<td>VGAN + Random Init</td>
<td>36.7%</td>
</tr>
<tr>
<td>VGAN + Logistic Reg</td>
<td>49.3%</td>
</tr>
<tr>
<td>VGAN + Fine Tune</td>
<td>52.1%</td>
</tr>
<tr>
<td>ImageNet Supervision [37]</td>
<td>91.4%</td>
</tr>
</tbody>
</table>

(a) Accuracy with Unsupervised Methods

(b) Performance vs # Data

(c) Relative Gain vs # Data

[Vondrick et al. NIPS 2016]
Shape modeling using 3D Generative Adversarial Network

[Wu et al. NIPS 2016]
Things to remember

• GANs can generate sharp samples from high-dimensional output space

• Conditional GAN can serve as general mapping model X→Y
  • No need to define domain-specific loss functions
  • Handle one-to-many mappings
  • Handle multiple modalities