The power of two samples for Generative Adversarial Networks

Sewoong Oh

Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

joint work with
Zinan Lin(CMU), Ashish Khetan(UIUC), Giulia Fanti(CMU)
Generative models

A generative model is a black box that takes a random vector $Z \sim N(0, \mathbf{I}_{k \times k})$ and produces a sample vector $G(Z) \in \mathbb{R}^{1024 \times 1024 \times 3}$.

A generative model is a black box that takes a random vector $Z \in \mathbb{R}^k$ and produces a sample vector $G(Z) \in \mathbb{R}^n$.

---

Generative models

A generative model is a black box that takes a random vector $Z \in \mathbb{R}^k$ and produces a sample vector $G(Z) \in \mathbb{R}^n$

Generative models learn fundamental representations

\[ G(G_1() + Z_{glasses}) = Z_{glasses} \]

[Z space] [Image space]

[DCGAN, Radford et al. 2015]
GAN: a breakthrough in training generative models

2004
Mixed Bernoulli

2007
RBM

2010
DBM

2014
VAE
GAN: a breakthrough in training generative models

2004 2007 2010 2014
Mixed Bernoulli RBM DBM VAE

2014 2015 2017
GAN Progressive GAN VAE
Generative Adversarial Networks (GAN)

\[ Z \xrightarrow{G} X \]

Real data

\[ \text{Discriminator} \ B(X) \]

\[ \text{Generator} \ G(Z) \]

\[ \min_G \ \max_D \ V(G, D) \]
“Mode collapse” is a main challenge

Target samples

Generated samples
“Mode collapse” is a main challenge

Target samples

Generated samples
“Mode collapse” is a main challenge

- “A man in a orange jacket with sunglasses and a hat ski down a hill.”
- “This guy is in black trunks and swimming underwater.”
- “A tennis player in a blue polo shirt is looking down at the green court.”

[“Generating interpretable images with controllable structure”, by Reed et al., 2016]
Lack of diversity is easier to detect if the discriminator sees multiple sample jointly
New framework: PacGAN

- lightweight overhead
- experimental results
- principled
Benchmark tests

<table>
<thead>
<tr>
<th>Mode</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAN</td>
<td>17.3</td>
</tr>
<tr>
<td>PacGAN2</td>
<td>23.8</td>
</tr>
<tr>
<td>PacGAN3</td>
<td>24.6</td>
</tr>
<tr>
<td>PacGAN4</td>
<td>24.8</td>
</tr>
</tbody>
</table>
Benchmark datasets from \textit{VeeGAN} paper

<table>
<thead>
<tr>
<th>Real data</th>
<th>DCGAN</th>
<th>PacDCGAN2</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Real data" /></td>
<td><img src="image2" alt="DCGAN" /></td>
<td><img src="image3" alt="PacDCGAN2" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modes (Max 1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGAN 99.0</td>
</tr>
<tr>
<td>ALI 16.0</td>
</tr>
<tr>
<td>Unrolled GAN 48.7</td>
</tr>
<tr>
<td>\textit{VeeGAN} 150.0</td>
</tr>
<tr>
<td>PacDCGAN2 1000.0</td>
</tr>
<tr>
<td>PacDCGAN3 1000.0</td>
</tr>
<tr>
<td>PacDCGAN4 1000.0</td>
</tr>
</tbody>
</table>
Intuition behind packing via toy example

Target distribution $P$

Generator $Q_1$
with mode collapse

Generator $Q_2$
without mode collapse

$d_{TV}(P, Q_1) = d_{TV}(P, Q_2) = 0.2$
Intuition behind packing via toy example

Target distribution $P$

Generator $Q_1$ with mode collapse

Generator $Q_2$ without mode collapse

$d_{TV}(P \times P, Q_1 \times Q_1) = 0.36$

$d_{TV}(P \times P, Q_2 \times Q_2) = 0.24$
Intuition behind packing via toy example

Target distribution $P$

$P \times P$

$P \times P$

$P \times P$

$P \times P$

Generator $Q_1$

with mode collapse

$Q_1 \times Q_1$

$Q_1 \times Q_1$

$Q_1 \times Q_1$

$Q_1 \times Q_1$

$d_{TV}(P \times P, Q_1 \times Q_1) = 0.36$

Generator $Q_2$

without mode collapse

$Q_2 \times Q_2$

$Q_2 \times Q_2$

$Q_2 \times Q_2$

$Q_2 \times Q_2$

$d_{TV}(P \times P, Q_2 \times Q_2) = 0.24$
Evolution of TV distances

Total variation
$\text{TV}(P^m, Q^m)$

Through packing, the target-generator pairs are expanded over the strengths of the mode collapse.
Evolution of TV distances through the prism of packing

Through packing, the target-generator pairs are expanded over the strengths of the mode collapse
\[ d_{TV}(P^m, Q^m) \]

\[
\max_{P,Q} \quad \min_{P,Q} \quad d_{TV}(P^2, Q^2) \\
\text{subject to} \quad d_{TV}(P, Q) = \tau
\]

- we focus on \( m = 2 \) for this talk
Definition [mode collapse region]

We say a pair \((P, Q)\) of a target distribution \(P\) and a generator distribution \(Q\) has \((\varepsilon, \delta)\)-mode collapse if there exists a set \(S\) such that

\[
P(S) \geq \delta, \quad \text{and} \quad Q(S) \leq \varepsilon.
\]
We say a pair \((P, Q)\) of a target distribution \(P\) and a generator distribution \(Q\) has \((\varepsilon, \delta)\)-mode collapse if there exists a set \(S\) such that

\[
P(S) \geq \delta, \quad \text{and} \quad Q(S) \leq \varepsilon.
\]
We say a pair \((P, Q)\) of a target distribution \(P\) and a generator distribution \(Q\) has \((\varepsilon, \delta)\)-mode collapse if there exists a set \(S\) such that

\[
P(S) \geq \delta, \quad \text{and} \quad Q(S) \leq \varepsilon.
\]
We say a pair \((P, Q)\) of a target distribution \(P\) and a generator distribution \(Q\) has \((\varepsilon, \delta)\)-mode collapse if there exists a set \(S\) such that

\[
P(S) \geq \delta, \quad \text{and} \quad Q(S) \leq \varepsilon.
\]
Intuition from Blackwell

Definition [mode collapse region]

We say a pair \((P, Q)\) of a target distribution \(P\) and a generator distribution \(Q\) has \((\varepsilon, \delta)\)-mode collapse if there exists a set \(S\) such that

\[ P(S) \geq \delta \quad \text{and} \quad Q(S) \leq \varepsilon. \]
**Definition [mode collapse region]**

We say a pair \((P, Q)\) of a target distribution \(P\) and a generator distribution \(Q\) has \((\epsilon, \delta)\)-**mode collapse** if there exists a set \(S\) such that

\[
P(S) \geq \delta , \quad \text{and} \quad Q(S) \leq \epsilon .
\]

---

**Diagram**

- **Target distribution** \(P\)
- **Generator** \(Q_1\) with mode collapse

- \(\delta\) and \(\epsilon\) are shaded regions.
- \(R(P, Q_1)\) is the region of interest.
We say a pair \((P, Q)\) of a target distribution \(P\) and a generator distribution \(Q\) has \((\varepsilon, \delta)\)-mode collapse if there exists a set \(S\) such that

\[
P(S) \geq \delta, \quad \text{and} \quad Q(S) \leq \varepsilon.
\]
We say a pair \((P, Q)\) of a target distribution \(P\) and a generator distribution \(Q\) has \((\varepsilon, \delta)\)-mode collapse if there exists a set \(S\) such that

\[
P(S) \geq \delta, \quad \text{and} \quad Q(S) \leq \varepsilon.
\]
Intuition from Blackwell

Definition [mode collapse region]

We say a pair \((P, Q)\) of a target distribution \(P\) and a generator distribution \(Q\) has \((\varepsilon, \delta)\)-mode collapse if there exists a set \(S\) such that

\[ P(S) \geq \delta, \quad \text{and} \quad Q(S) \leq \varepsilon. \]
Intuition from Blackwell

**Definition [mode collapse region]**

We say a pair \((P, Q)\) of a target distribution \(P\) and a generator distribution \(Q\) has \((\epsilon, \delta)\)-**mode collapse** if there exists a set \(S\) such that

\[
P(S) \geq \delta, \quad \text{and} \quad Q(S) \leq \epsilon.
\]

![Diagram showing target distribution \(P\), generator distribution \(Q_2\), and the region \(\mathcal{R}(P, Q_2)\) with \(d_{TV}(P, Q_2) = 0.2\).](image-url)
Upper bound

\[ \max_{P,Q} d_{TV}(P^2, Q^2) \]

subject to

\[ d_{TV}(P, Q) = \tau \]
Upper bound

\[ \max_{P,Q} d_{TV}(P^2, Q^2) \]

subject to

\[ d_{TV}(P, Q) = \tau \]

\[ \mathcal{R}(P, Q) \subseteq \mathcal{R}_{outer}(\tau) \]
Upper bound

\[
\max_{P, Q} d_{TV}(P^2, Q^2) \\
\text{subject to } d_{TV}(P, Q) = \tau
\]

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}_{\text{outer}}(\tau)
\]
Upper bound

\[
\max_{P,Q} d_{TV}(P^2, Q^2) \quad \text{subject to} \quad d_{TV}(P, Q) = \tau
\]

\[
R(P, Q) \subseteq R_{\text{outer}}(\tau) \\
R(P^2, Q^2) \subseteq R(\hat{P}_{\text{outer}}, \hat{Q}_{\text{outer}})
\]

Blackwell’s theorem

\[
R(P, Q) \subseteq R(P', Q') \quad \Rightarrow \quad R(P^2, Q^2) \subseteq R(P'^2, Q'^2)
\]
Upper bound

\[
\max_{P,Q} d_{TV}(P^2, Q^2)
\]

subject to

\[
d_{TV}(P, Q) = \tau
\]

such that

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}_\text{outer}(\tau)
\]

\[
\mathcal{R}(P^2, Q^2) \subseteq \mathcal{R}(P^2_{\text{outer}}, Q^2_{\text{outer}})
\]

\[
d_{TV}(P^2, Q^2) \leq d_{TV}(P^2_{\text{outer}}, Q^2_{\text{outer}}) \quad 1-(1-\tau)^2
\]

Blackwell’s theorem

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}(P', Q') \quad \Rightarrow \quad \mathcal{R}(P^2, Q^2) \subseteq \mathcal{R}(P'^2, Q'^2)
\]
Lower bound

\begin{align*}
\min_{P, Q} & \quad d_{TV}(P^2, Q^2) \\
\text{subject to} & \quad d_{TV}(P, Q) = \tau
\end{align*}

\(d_{TV}(P^2, Q^2)\)
Lower bound

\[
\text{min}_{P,Q} \quad d_{TV}(P^2, Q^2)
\]

subject to

\[
d_{TV}(P, Q) = \tau
\]

\[
\mathcal{R}_{\text{inner}}(\tau, \alpha) \subseteq \mathcal{R}(P, Q)
\]
Lower bound

$$
\min_{P,Q} \quad d_{TV}(P^2, Q^2)
\text{subject to} \quad d_{TV}(P, Q) = \tau
$$

$$
R_{\text{inner}}(\tau, \alpha) \subseteq R(P, Q)
$$
Lower bound

\[
\min_{P,Q} d_{TV}(P^2, Q^2) \\
\text{subject to } d_{TV}(P, Q) = \tau
\]

\[
\mathcal{R}_{inner}(\tau, \alpha) \subseteq \mathcal{R}(P, Q) \\
\mathcal{R}(P_{inner}^2, Q_{inner}^2) \subseteq \mathcal{R}(P^2, Q^2)
\]

Blackwell’s theorem

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}(P', Q') \\
\Rightarrow \mathcal{R}(P^2, Q^2) \subseteq \mathcal{R}(P'^2, Q'^2)
\]
Lower bound

\[ \min_{P, Q} d_{TV}(P^2, Q^2) \]

subject to

\[ d_{TV}(P, Q) = \tau \]

\[ \mathcal{R}_{inner}(\tau, \alpha) \subseteq \mathcal{R}(P, Q) \]

\[ \mathcal{R}(P^2_{inner}, Q^2_{inner}) \subseteq \mathcal{R}(P^2, Q^2) \]

\[ \min_{\alpha} d_{TV}(P^2_{inner}, Q^2_{inner}) \leq d_{TV}(P^2, Q^2) \]

Blackwell’s theorem

\[ \mathcal{R}(P, Q) \subseteq \mathcal{R}(P', Q') \]

\[ \Rightarrow \mathcal{R}(P^2, Q^2) \subseteq \mathcal{R}(P'^2, Q'^2) \]
\[ d_{TV}(P^m, Q^m) \]

\[
\text{max} / \text{min}_{P,Q, P,Q} \quad d_{TV}(P^2, Q^2)
\]

subject to \[ d_{TV}(P, Q) = \tau \]
\[
d_{TV}(P^m, Q^m)
\]

subject to

\[
d_{TV}(P, Q) = \tau
\]

no \((\varepsilon_0, \delta_0)\)-mode collapse
Upper bound without \((\varepsilon_0, \delta_0)\)-mode collapse

\[
\max_{P,Q} d_{TV}(P^2, Q^2) \\
\text{subject to} \\
d_{TV}(P, Q) = \tau \\
\text{no } (\varepsilon_0, \delta_0)\text{-mode collapse}
\]
Upper bound without \((\varepsilon_0, \delta_0)\)-mode collapse

\[
\max_{P,Q} d_{TV}(P^2, Q^2)
\]

subject to \(d_{TV}(P, Q) = \tau\)

no \((\varepsilon_0, \delta_0)\)-mode collapse

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}_{\text{outer}}(\tau, \varepsilon_0, \delta_0, \alpha)
\]
Upper bound without \((\varepsilon_0, \delta_0)\)-mode collapse

\[
\max_{P, Q} \ d_{TV}(P^2, Q^2)
\]

subject to
\[
d_{TV}(P, Q) = \tau
\]
no \((\varepsilon_0, \delta_0)\)-mode collapse

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}_{\text{outer}}(\tau, \varepsilon_0, \delta_0, \alpha)
\]
Upper bound without \((\varepsilon_0, \delta_0)\)-mode collapse

\[
\max_{P, Q} \quad d_{TV}(P^2, Q^2)
\]

subject to
\[
d_{TV}(P, Q) = \tau
\]

no \((\varepsilon_0, \delta_0)\)-mode collapse

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}_{\text{outer}}(\tau, \varepsilon_0, \delta_0, \alpha)
\]

\[
\mathcal{R}(P^2, Q^2) \subseteq \mathcal{R}(P_{\text{outer}}^2, Q_{\text{outer}}^2)
\]

Blackwell’s theorem

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}(P', Q')
\]

\[
\Rightarrow \quad \mathcal{R}(P^2, Q^2) \subseteq \mathcal{R}(P'^2, Q'^2)
\]
Upper bound without \((\varepsilon_0, \delta_0)\)-mode collapse

\[
\max_{P,Q} \quad d_{TV}(P^2, Q^2) \\
\text{subject to} \quad d_{TV}(P, Q) = \tau \\
\text{no } (\varepsilon_0, \delta_0)\text{-mode collapse}
\]

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}_{\text{outer}}(\tau, \varepsilon_0, \delta_0, \alpha) \\
\mathcal{R}(P^2, Q^2) \subseteq \mathcal{R}(P_{\text{outer}}^2, Q_{\text{outer}}^2) \\
d_{TV}(P^2, Q^2) \leq \max_{\alpha} d_{TV}(P_{\text{outer}}^2, Q_{\text{outer}}^2)
\]

simple to evaluate

Blackwell’s theorem

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}(P', Q') \\
\Rightarrow \quad \mathcal{R}(P^2, Q^2) \subseteq \mathcal{R}(P'^2, Q'^2)
\]
\[ d_{TV}(P^m, Q^m) \]

subject to

\[
\min_{P,Q} \quad d_{TV}(P^2, Q^2)
\]

\[
d_{TV}(P, Q) = \tau
\]

\[(\varepsilon_1, \delta_1)\text{-mode collapse}\]
Lower bound with \((\varepsilon_1, \delta_1)\)-mode collapse

\[
\min_{P,Q} d_{TV}(P^2, Q^2) \quad \text{subject to} \quad d_{TV}(P, Q) = \tau \\
(\varepsilon_1, \delta_1)\text{-mode collapse}
\]
Lower bound with \((\varepsilon_1, \delta_1)\)-mode collapse

\[
\begin{aligned}
\min_{P,Q} & \quad d_{TV}(P^2, Q^2) \\
\text{subject to} & \quad d_{TV}(P, Q) = \tau \\
& \quad (\varepsilon_1, \delta_1)\text{-mode collapse}
\end{aligned}
\]

\[
\mathcal{R}_{\text{inner}}(\tau, \alpha, \varepsilon_1, \delta_1) \subseteq \mathcal{R}(P, Q)
\]
Lower bound with \((\varepsilon_1, \delta_1)\)-mode collapse

\[
\begin{align*}
\min_{P,Q} & \quad d_{TV}(P^2, Q^2) \\
\text{subject to} & \quad d_{TV}(P, Q) = \tau \\
& \quad (\varepsilon_1, \delta_1)\text{-mode collapse}
\end{align*}
\]

\[
\mathcal{R}_{\text{inner}}(\tau, \alpha, \varepsilon_1, \delta_1) \subseteq \mathcal{R}(P, Q)
\]
Lower bound with \((\varepsilon_1, \delta_1)\)-mode collapse

\[
\begin{align*}
\min_{P,Q} & \quad d_{TV}(P^2, Q^2) \\
\text{subject to} & \quad d_{TV}(P, Q) = \tau \\
& \quad (\varepsilon_1, \delta_1)\text{-mode collapse}
\end{align*}
\]

\[
\mathcal{R}_{\text{inner}}(\tau, \alpha, \varepsilon_1, \delta_1) \subseteq \mathcal{R}(P, Q)
\]

\[
\mathcal{R}(P_{\text{inner}}^2, Q_{\text{inner}}^2) \subseteq \mathcal{R}(P^2, Q^2)
\]

Blackwell’s theorem

\[
\mathcal{R}(P, Q) \subseteq \mathcal{R}(P', Q')
\]

\[
\Rightarrow \quad \mathcal{R}(P^{2'}, Q^{2'}) \subseteq \mathcal{R}(P'^{2'}, Q'^{2'})
\]
Lower bound with \((\varepsilon_1, \delta_1)\)-mode collapse

\[
\begin{align*}
\min_{P, Q} & \quad d_{TV}(P^2, Q^2) \\
\text{subject to} & \quad d_{TV}(P, Q) = \tau \\
& \quad (\varepsilon_1, \delta_1)\text{-mode collapse}
\end{align*}
\]

\[
\begin{align*}
\mathcal{R}_{inner}(\tau, \alpha, \varepsilon_1, \delta_1) & \subseteq \mathcal{R}(P, Q) \\
\mathcal{R}(P_{inner}^2, Q_{inner}^2) & \subseteq \mathcal{R}(P^2, Q^2) \\
\min_{\alpha} d_{TV}(P_{inner}^2, Q_{inner}^2) & \leq d_{TV}(P^2, Q^2)
\end{align*}
\]

simple to evaluate

Blackwell’s theorem

\[
\begin{align*}
\mathcal{R}(P, Q) & \subseteq \mathcal{R}(P', Q') \\
\Rightarrow & \quad \mathcal{R}(P^2, Q^2) \subseteq \mathcal{R}(P'^2, Q'^2)
\end{align*}
\]
Achievable TV distances for distributions

with \((\varepsilon_1, \delta_1)\)-mode collapse  
without \((\varepsilon_0, \delta_0)\)-mode collapse

\[d_{TV}(P^m, Q^m)\]

with packing, the discriminator naturally penalizes \((P, Q)\) with severe mode collapses
Could we be cheating (hyper-parameter tuning)?

1. Discriminator size

GAN

minibatch size = 64

D(X)

PacGAN2

minibatch size = 64

D(X₁, X₂)

real 1
real 1
fake 0
real 1
real 1
	×2
Could we be cheating (hyper-parameter tuning)?

1. Discriminator size

![Diagram showing the number of modes captured and the number of parameters in $D(\cdot)$ over different numbers of iterations.](image)

- GAN
- PacGAN2
- PacGAN3
- PacGAN4
Could we be cheating (hyper-parameter tuning)?

2. Minibatch size

minibatch size $= 64$

$D(X)$

GAN

PacGAN2

minibatch size $= 64$

$D(X_1, X_2)$
Could we be cheating (hyper-parameter tuning)?

2. Minibatch size

minibatch size = 64

GAN

minibatch size = 32

PacGAN2
Could we be cheating (hyper-parameter tuning)?

2. Minibatch size

<table>
<thead>
<tr>
<th>Modes</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCGAN</td>
<td>99.0</td>
</tr>
<tr>
<td>PacDCGAN2</td>
<td>1000.0</td>
</tr>
</tbody>
</table>
Theoretical challenges in GAN

Designing Loss

\[ D_{JS}(P, Q) \quad \text{Jansen-Shannon} \]

\[ D_f(P, Q) \quad \text{f-divergence} \]

\[ D_W(P, Q) \quad \text{Wasserstein} \]

[FeiziSuhXiaTse 2017]
### Theoretical challenges in GAN

<table>
<thead>
<tr>
<th>Designing Loss</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{JS}(P,Q)$</td>
<td>Jansen-Shannon</td>
</tr>
<tr>
<td>$D_f(P,Q)$</td>
<td>$f$-divergence</td>
</tr>
<tr>
<td>$D_W(P,Q)$</td>
<td>Wasserstein</td>
</tr>
</tbody>
</table>

\[
D(P^m, Q^m) \succ D(P, Q) \quad (\varepsilon, \delta)-mode\ collapse
\]
Theoretical challenges in GAN

Training Dynamics

[LiMadryPeeblesSchmidt 2017]

Designing Loss

Evaluation

$D_{JS}(P, Q)$  Jansen-Shannon
$D_{f}(P, Q)$  $f$-divergence
$D_{W}(P, Q)$  Wasserstein

$D(P^m, Q^m) \succ D(P, Q)$  $(\varepsilon, \delta)$-mode collapse
Our paper is available at:
https://arxiv.org/abs/1712.04086

All codes for the experiments at:
https://github.com/fjxmlzn/PacGAN

Zinan Lin (CMU)    Ashish Khetan (UIUC)    Giulia Fanti (CMU)