Attempts Towards Controllable Generation

Hsuan-Tien Lin
林軒田

Professor, National Taiwan University

August 10, 2023, NTUILO x MITILP Consortia
About Me

Professor
National Taiwan University

Chief Data Science Consultant
(former Chief Data Scientist)
Appier Inc.

Co-author
Learning from Data

Instructor
NTU-Coursera MOOCs
ML Foundations/Techniques

research goal: making machine more realistic
Outline


Generative Models

Molecular generation

Random vector \( z \)

Prompted drawing

Person with red eyes

Random vector \( z \)

Generative models: creativity for machines
Generative Adversarial Network (GAN)

- random vector $z \sim \text{Gaussian}$
  - generator
    - generated output $x \sim P(x)$
- generated vector $x$
  - discriminator
    - realness of $x$

high-quality generation when generator can fool the discriminator
Conditional Generative Adversarial Network (cGAN)

- **Generator**:
  - Input: Random vector $\mathbf{z}$
  - Output: Generated image $\mathbf{x}$

- **Discriminator**:
  - Input: Image $\mathbf{x}$
  - Output: Realness of $\mathbf{x} | y$

**Equation:**

$$y = \text{dog}\quad \text{generated } \mathbf{x} = \text{dog}\quad \text{realness of } \mathbf{x} | y$$

- **High-quality conditional (i.e. controllable) generation when generator can fool the conditional discriminator**
Existing Attempts to Design cGANs

**ACGAN** (Odena et al., 2017)
- Unconditional discriminator
- Classifier

**ProjGAN** (Miyato et al., 2018)
- Conditional discriminator
- Contrastive loss

**ContraGAN** (Kang et al., 2020)
- Unconditional discriminator
- Contrastive loss

Can we comprehensively understand and improve them?
Our Contributions

- **explains** the designs of classifier loss, unconditional loss, conditional loss, contrastive loss in principle
- **unifies** ACGAN, ProjGAN, ContraGAN to a new architecture ECGAN with the principled explanation
- **achieves** state-of-the-art conditional generation performance with the unification

will describe **key ideas** behind the unification without detailed math
Idea 1: Connecting Loss Terms

- **classifier loss**: for estimating $p(y|x)$
- **unconditional loss**: for estimating $p(x)$
- **conditional loss**: for estimating $p(x|y)$

—how do they connect?

Bayes Rule

$$\log p(x, y) = \log p(x|y) + \log p(y)$$

assumed as known constants

$$= \log p(y|x) + \log p(x)$$

unifying view ⇒ estimate joint distribution $p(x, y)$ better from different angles
Idea 2: Energy-Based Based Parameterization

**unifying view**

estimate $p(x, y)$ from different angles

**energy-based parameterization**

- if $\exp(h_\theta(x, y)) \propto p(x, y)$,

  $$\log p(x, y) = h_\theta(x, y) - \log(\text{normalization term})$$

  neural network

- $\log(\text{normalization term}) \approx \text{contrastive loss}$ (proof omitted 😊)

**ECGAN** (Energy-based Conditional GAN): unifying view + energy-based parameterization
ECGAN

- unifying view: stabilizes estimation by **two angles** instead of one
- energy-based parameterization: explains **contrastive loss** as regularization from **energy normalization**

not just ad-hoc combination, but **principled design from math**
The Unified ECGAN Family

ACGAN (Odena et al., 2017)
1-angle ECGAN without contrastive

ProjGAN (Miyato et al., 2018)
1-angle ECGAN without classifier and contrastive

ContraGAN (Kang et al., 2020)
1-angle ECGAN without classifier

complete ECGAN: 2-angle with classifier (w/ or w/o contrastive)

## Evaluation Setting

### Data

<table>
<thead>
<tr>
<th></th>
<th># training</th>
<th># test</th>
<th># classes</th>
<th>resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIFAR-10</td>
<td>50,000</td>
<td>10,000</td>
<td>10</td>
<td>32x32</td>
</tr>
<tr>
<td>Tiny ImageNet</td>
<td>100,000</td>
<td>10,000</td>
<td>200</td>
<td>64x64</td>
</tr>
<tr>
<td>ImageNet</td>
<td>1,281,167</td>
<td>50,000</td>
<td>1,000</td>
<td>128x128</td>
</tr>
</tbody>
</table>

### Evaluation Metrics

- **Frechet Inception Distance (FID) (↓):** unconditional generation goodness
- **Inception Score (IS) (↑):** unconditional image quality
- **intra-FID (per-class FID) (↓):** conditional generation goodness

---

is **complete ECGAN** better?
### Comparison to Existing cGANs

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Backbone</th>
<th>method</th>
<th>FID (↓)</th>
<th>IS (↑)</th>
<th>Intra-FID (↓)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIFAR-10</td>
<td>DCGAN</td>
<td>ACGAN</td>
<td>32.507 ± 2.174</td>
<td>7.621 ± 0.088</td>
<td>129.603 ± 1.212</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ProjGAN</td>
<td>21.918 ± 1.580</td>
<td>8.095 ± 0.185</td>
<td>68.164 ± 2.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ContraGAN</td>
<td>28.310 ± 1.761</td>
<td>7.637 ± 0.125</td>
<td>153.730 ± 9.965</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>ECGAN-UC</strong></td>
<td><strong>18.035 ± 0.788</strong></td>
<td><strong>8.487 ± 0.131</strong></td>
<td><strong>59.343 ± 1.557</strong></td>
</tr>
<tr>
<td></td>
<td>ResGAN</td>
<td>ACGAN</td>
<td>10.073 ± 0.274</td>
<td>9.512 ± 0.050</td>
<td>48.464 ± 0.716</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ProjGAN</td>
<td>10.195 ± 0.203</td>
<td>9.268 ± 0.139</td>
<td>46.598 ± 0.070</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ContraGAN</td>
<td>10.551 ± 0.976</td>
<td>9.087 ± 0.228</td>
<td>138.944 ± 12.582</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>ECGAN-UC</strong></td>
<td><strong>9.244 ± 0.062</strong></td>
<td><strong>9.651 ± 0.098</strong></td>
<td><strong>43.876 ± 0.384</strong></td>
</tr>
<tr>
<td></td>
<td>BigGAN</td>
<td>ACGAN</td>
<td>8.615 ± 0.146</td>
<td>9.742 ± 0.041</td>
<td>45.243 ± 0.129</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ProjGAN</td>
<td>8.145 ± 0.156</td>
<td>9.840 ± 0.080</td>
<td>42.110 ± 0.405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ContraGAN</td>
<td>8.617 ± 0.671</td>
<td>9.679 ± 0.210</td>
<td>114.602 ± 13.261</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>ECGAN-UC</strong></td>
<td><strong>7.942 ± 0.041</strong></td>
<td><strong>10.002 ± 0.120</strong></td>
<td><strong>41.425 ± 0.221</strong></td>
</tr>
<tr>
<td>Tiny ImageNet</td>
<td>BigGAN</td>
<td>ACGAN</td>
<td>29.528 ± 4.612</td>
<td>12.964 ± 0.770</td>
<td>315.408 ± 1.171</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ProjGAN</td>
<td>28.451 ± 2.242</td>
<td>12.213 ± 0.624</td>
<td>242.332 ± 11.447</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ContraGAN</td>
<td>24.915 ± 1.222</td>
<td>13.445 ± 0.371</td>
<td>257.657 ± 3.246</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>ECGAN-UC</strong></td>
<td><strong>18.780 ± 1.291</strong></td>
<td><strong>17.475 ± 1.052</strong></td>
<td><strong>204.830 ± 5.648</strong></td>
</tr>
</tbody>
</table>

**is complete ECGAN (w/o contrastive) better?** **YES**
State-of-the-art Performance (2021)

<table>
<thead>
<tr>
<th>Method</th>
<th>FID(↓)</th>
<th>IS(↑)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BigGAN*</td>
<td>24.68</td>
<td>28.63</td>
</tr>
<tr>
<td>ContraGAN*</td>
<td>25.16</td>
<td>25.25</td>
</tr>
<tr>
<td>ECGAN-UC</td>
<td>30.05</td>
<td>26.47</td>
</tr>
<tr>
<td>ECGAN-UCE</td>
<td>12.16</td>
<td>56.33</td>
</tr>
<tr>
<td>ECGAN-UCE (40k step)</td>
<td>8.49</td>
<td>80.69</td>
</tr>
</tbody>
</table>

Table 6: Evaluation on ImageNet128×128. (*: Reported by StudioGAN.)

(StudioGAN: https://github.com/POSTECH-CVLab/PyTorch-StudioGAN)

deep mathematical understanding  
⇒ state-of-the-art performance
Generated Images from ECGAN

high-quality **controllable** generation somewhat achieved
Half Summary

*a comprehensive study of cGANs, which ...*

- **explains** the designs of classifier loss, unconditional loss, conditional loss, contrastive loss in principle
  - connecting three loss terms with Bayes rule
  - applying energy-based models to reveal contrastive loss

- **unifies** ACGAN, ProjGAN, ContraGAN to a new architecture **ECGAN** with the principled explanation

- **achieves** state-of-the-art conditional generation performance with the unification

fundamental research on **math principles** $\Rightarrow$ **controllability**

Score-based Generative Model (SGM)

**Generated image** $\mathbf{x} \leftarrow \mathbf{z}$

**Forward SDE** (data $\rightarrow$ noise)

$$\mathbf{x}(0) \quad \mathbf{dx} = f(\mathbf{x}, t) dt + g(t) d\mathbf{w} \quad \mathbf{x}(T)$$

**Reverse SDE** (noise $\rightarrow$ data)

$$\mathbf{x}(0) \quad \mathbf{dx} = [f(\mathbf{x}, t) - g^2(t) \nabla_{\mathbf{x}} \log p_t(\mathbf{x})] dt + g(t) d\mathbf{w} \quad \mathbf{x}(T)$$

(Song et al., ICLR 2021)

High-quality generation when **score function** can be estimated.
Conditional SGM

SGM

- **high-quality unconditional generation** when $\nabla_x \log p(x)$ can be estimated

Conditional SGM

- **high-quality conditional generation** when $\nabla_x \log p(x|y)$ can be estimated

Hello Again, Bayes Rule

$$\nabla_x \log p(x|y) = \underbrace{\nabla_x \log p(x)}_{\text{unconditional score}} + \underbrace{\nabla_x \log p(y|x)}_{\text{classifier gradient}} - \underbrace{\nabla_x \log p(y)}_{0}$$

simple CGSGM

by classifier guidance + unconditional SGM

Simple CGSGM

\[ \nabla_x \log p(x|y) = \nabla_x \log p(x) + \nabla_x \log p(y|x) \]

unconditional score  
classifier gradient

Pros

- easy reuse of well-trained unconditional SGM
- naturally applicable to semi-supervised data (few labeled data)

Cons

overfitting classifier

\[ \implies \]  
bad conditional score

\[ \implies \]  
bad conditional generation

but few labeled data \( \implies \) overfitting classifier?!

Key Idea: Align Classifier with Unconditional SGM

energy-based parameterization $\exp(h_\theta(x, y)) \propto p(x, y)$

**classifier**: approximate $p(y|x)$ by

$$\frac{\exp(h_\theta(x, y))}{\sum_k \exp(h_\theta(x, k))}$$

**unconditional SGM**: approximate $\nabla_x \log p(x)$ by

$$\nabla_x \log \frac{\sum_k \exp(h_\theta(x, k))}{p(x)}$$

$$= \nabla_x \log \sum_k \exp(h_\theta(x, k))$$

$$- \nabla_x \log(\text{normalization})$$

classifier can **cosplay as** (be cast to) unconditional SGM

can **regularize** classifier by its unconditional SGM loss to avoid overfitting (proof omitted 😊)
Comparison to Original CGSGM

with merely 5% of labeled data

<table>
<thead>
<tr>
<th>Original CGSVM</th>
<th>Our Improved CGSFM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-FID (↓) 31.17</td>
<td>Intra-FID (↓) 18.95</td>
</tr>
</tbody>
</table>

ours: better quality & more accurate
Short Take-Home Messages

- creativity can go **wild**
  —regularization by another view helps
  - ECGAN: another view by decomposing joint probability
  - improved CGSGM: another view by casting classifier as unconditional SGM
- most importantly, **math helps! 😊**
  —more efforts on fundamental research needed
  - Bayes rule helps
  - energy-based parameterization helps

Thank you! Questions?

H.-T. Lin (NTU)

Attempts Towards Controllable Generation
Appendix: Image Citations

Except for images from our papers, other images are cited as follows:


- 🧵: CC-0 from https://bioicons.com/icons/cc-0/Molecular_modelling/Simon_Dürrr/smiles.svg

- 📸: CC-0 from https://commons.wikimedia.org/wiki/File:Demonic_possession.jpg

- 👉: CC-0 from https://commons.wikimedia.org/wiki/File:Detective.svg

- 🔍: Figure 1 from Song et al., ICLR 2021 https://openreview.net/forum?id=PxTIG12RRHS