Attempts Towards Controllable Generation

Hsuan-Tien Lin
林軒田

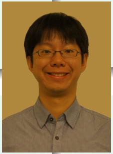
Professor, National Taiwan University

August 10, 2023, NTUILO x MITILP Consortia

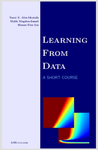
About Me

Professor
National Taiwan University

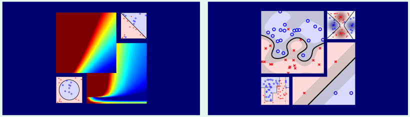
Chief Data Science Consultant
(former Chief Data Scientist)
Appier Inc.



Co-author
Learning from Data



Instructor
NTU-Coursera MOOCs
ML Foundations/Techniques



research goal: making machine more realistic

Outline

Si-An Chen, Chun-Liang Li, and Hsuan-Tien Lin. A unified view of cGANs with and without classifiers. NeurIPS 2021.

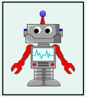
Paul Kuo-Ming Huang, Si-An Chen, and Hsuan-Tien Lin. Score-based conditional generation with fewer labeled data by self-calibrating classifier guidance. Work-In-Progress 2023.

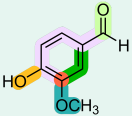
Si-An Chen, Chun-Liang Li, and Hsuan-Tien Lin. A unified view of cGANs with and without classifiers. NeurIPS 2021.

Generative Models

molecular generation

random vector **z**



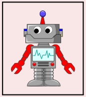


O=Cc1ccc(O)c(OC)c1

prompted drawing

person w/ red eyes

random vector **z**



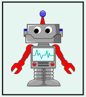
generative models: **creativity** for machines

Generative Adversarial Network (GAN)

random vector $\mathbf{z} \sim$ Gaussian

$\downarrow$

generator



$\downarrow$

generated output $\mathbf{x} \sim P(\mathbf{x})$

generated vector $\mathbf{x}$

$\downarrow$

discriminator

$\downarrow$

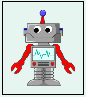
realness of $\mathbf{x}$

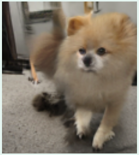
high-quality generation when generator can **fool the discriminator**

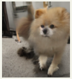
Conditional Generative Adversarial Network (cGAN)

dog random vector **z**

generator





$y =$ **dog** generated $\mathbf{x} =$ 

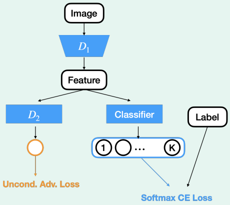
discriminator

realness of $\mathbf{x}|y$

high-quality conditional (i.e. **controllable**) generation when generator can **fool the conditional discriminator**

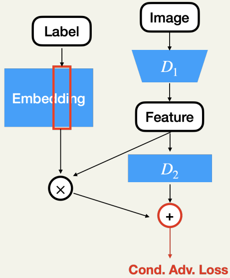
Existing Attempts to Design cGANs

ACGAN (Odena et al., 2017)



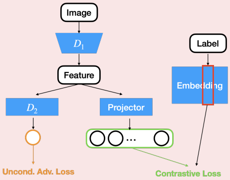
- unconditional discriminator
- classifier

ProjGAN (Miyato et al., 2018)



- conditional discriminator

ContraGAN (Kang et al., 2020)



- unconditional discriminator
- contrastive loss

can we **comprehensively** understand and improve them?

Our Contributions

a comprehensive study of cGANs, which ...

- **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 $p(y|\mathbf{x})$
- unconditional loss: for estimating $p(\mathbf{x})$
- conditional loss: for estimating $p(\mathbf{x}|y)$

—how do they connect?

Bayes Rule

$$\begin{aligned}\log p(\mathbf{x}, y) &= \log p(\mathbf{x}|y) + \underbrace{\log p(y)}_{\text{assumed as known constants}} \\ &= \log p(y|\mathbf{x}) + \log p(\mathbf{x})\end{aligned}$$

unifying view $\Rightarrow$ estimate joint distribution $p(\mathbf{x}, y)$ better **from different angles**

Idea 2: Energy-Based Parameterization

unifying view

estimate $p(\mathbf{x}, y)$ from different angles

energy-based parameterization

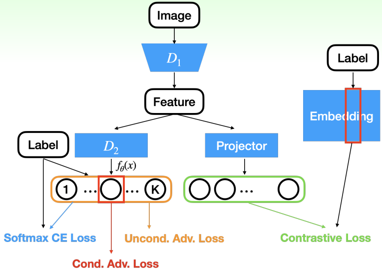
- if $\exp(h_\theta(\mathbf{x}, y)) \propto p(\mathbf{x}, y)$,

$$\log p(\mathbf{x}, y) = \underbrace{h_\theta(\mathbf{x}, y)}_{\text{neural network}} - \log(\text{normalization term})$$

- $\log$(normalization term) $\approx$ contrastive loss (proof omitted ☺)

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

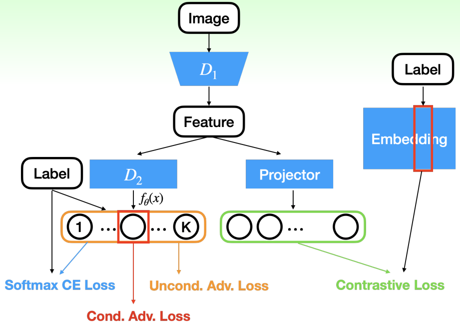
ECGAN



- 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

	# training	# test	# classes	resolution
CIFAR-10	50,000	10,000	10	32x32
Tiny ImageNet	100,000	10,000	200	64x64
ImageNet	1,281,167	50,000	1,000	128x128

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

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

is **complete ECGAN** (w/o contrastive) better?

YES

State-of-the-art Performance (2021)

Method	FID(↓)	IS(↑)
BigGAN*	24.68	28.63
ContraGAN*	25.16	25.25
ECGAN-UC	30.05	26.47
ECGAN-UCE	12.16	56.33
ECGAN-UCE (40k step)	**8.49**	**80.69**

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

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

deeper mathematical understanding
$\Longrightarrow$ state-of-the-art performance

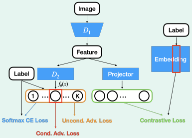
Generated Images from ECGAN

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**
$\Longrightarrow$ **controllability**

Outline

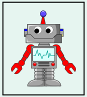
Si-An Chen, Chun-Liang Li, and Hsuan-Tien Lin. A unified view of cGANs with and without classifiers. NeurIPS 2021.

Paul Kuo-Ming Huang, Si-An Chen, and Hsuan-Tien Lin. Score-based conditional generation with fewer labeled data by self-calibrating classifier guidance. Work-In-Progress 2023.

classifier guidance. Work-In-Progress 2023.

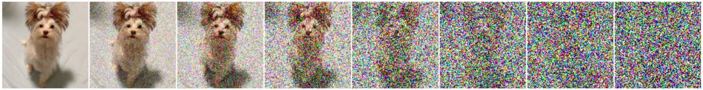
Score-based Generative Model (SGM)

SGM

generated image **x** $\leftarrow$  $\leftarrow$ random image **z**

Forward SDE (data → noise)

$$\mathbf{x}(0) \longrightarrow \mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w} \longrightarrow \mathbf{x}(T)$$



score function

$$\mathbf{x}(0) \longleftarrow \mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g^2(t)\boxed{\nabla_{\mathbf{x}} \log p_t(\mathbf{x})}\right] \mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}} \longleftarrow \mathbf{x}(T)$$

Reverse SDE (noise → data)

(Song et al., ICLR 2021)

high-quality generation
when **score function** can be estimated

Conditional SGM

SGM

high-quality unconditional generation when $\nabla_{\mathbf{x}} \log p(\mathbf{x})$ can be estimated

Conditional SGM

high-quality conditional generation when $\nabla_{\mathbf{x}} \log p(\mathbf{x}|y)$ can be estimated

Hello Again, Bayes Rule

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

simple CGSGM
by classifier guidance + unconditional SGM

Simple CGSGM

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

Pros

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

Cons

overfitting classifier
$\Longrightarrow$ bad conditional score
$\Longrightarrow$ bad conditional generation

but **few labeled data** $\Rightarrow$ overfitting classifier?!

Key Idea: Align Classifier with Unconditional SGM

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

classifier:
approximate $p(y|\mathbf{x})$ by

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

unconditional SGM:
approximate $\nabla_x \log p(\mathbf{x})$ by

$$\nabla_{\mathbf{x}} \log \underbrace{\frac{\sum_k \exp(h_\theta(\mathbf{x}, k))}{\text{normalization}}}_{p(\mathbf{x})}$$

$$= \nabla_{\mathbf{x}} \log \sum_k \exp(h_\theta(\mathbf{x}, k)) \underbrace{-\nabla_{\mathbf{x}} \log(\text{normalization})}_{0}$$

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

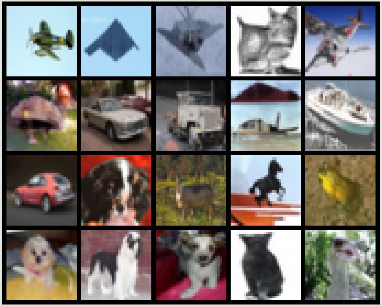
can **regularize** classifier by its unconditional SGM loss to avoid overfitting (proof omitted ☺)

Comparison to Original CGSGM

with merely 5% of labeled data

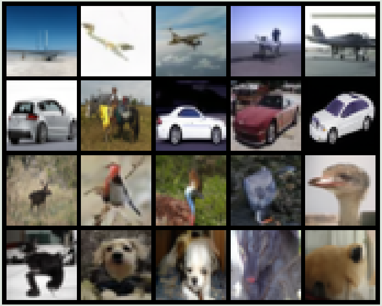
Original CGSVM

Intra-FID (↓) 31.17



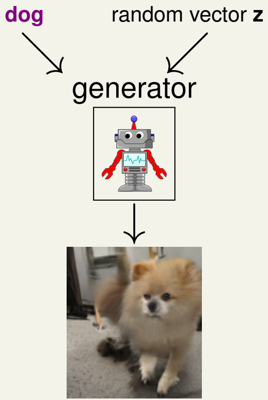
Our **Improved** CGSFM

Intra-FID (↓) 18.95



ours: better quality & more accurate classifier guidance. Work-In-Progress 2023.

Short Take-Home Messages

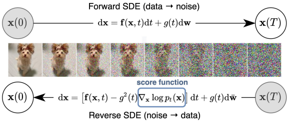


- creativity can go **wild**
  —regularization **by another view** helps
  - ECGAN: another view by **decomposing joint probability**
  - improved CGSGM: another view by **casting classifier as unconditional SGM**
- most importantly, **math helps**! ☺
  —more efforts on **fundamental research** needed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

Thank you! Questions?

Appendix: Image Citations

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

- : Free Content Use from `https://pixabay.com/vectors/robot-machine-technology-science-312566/`
- : CC-0 from `https://bioicons.com/icons/cc-0/Molecular_modelling/Simon_Dürr/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`