Attempts Towards Robust and Controllable Generation

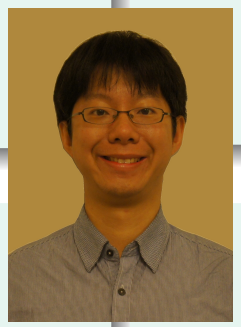
Hsuan-Tien Lin
林軒田

Professor, National Taiwan University

August 27, 2024, Cloud Computing and IoT Association in Taiwan

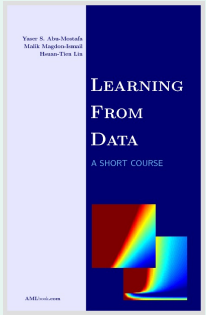
About Me

Professor
National Taiwan University

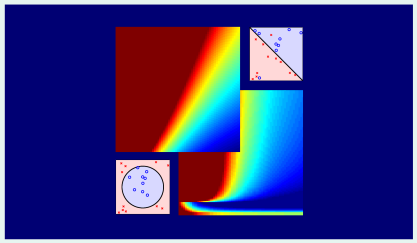


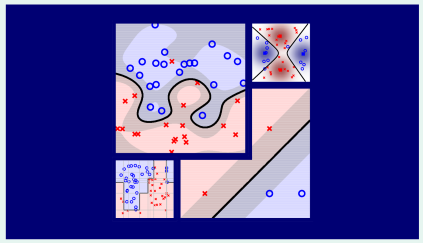
Chief Data Science Consultant
(former Chief Data Scientist)
Appier Inc.

Co-author
Learning from Data



Instructor
NTU-Coursera MOOCs
ML Foundations/Techniques





research goal: making machine more realistic

From Intelligence to Artificial Intelligence

intelligence: thinking and acting smartly

- humanly
- rationally

artificial intelligence: computers ~~thinking and~~ acting smartly

- humanly
- rationally

humanly $\approx$ smartly $\approx$ rationally
—are humans rational? ☺

Humanly versus Rationally

What if your self-driving car decides one death is better than two—and that one is you? (The Washington Post `http://wpo.st/ZK-51`)

> You're humming along in your self-driving car, chatting on your iPhone 37 while the machine navigates on its own. Then a swarm of people appears in the street, right in the path of the oncoming vehicle.

Car Acting Humanly

to save my (and passengers') life, stay on track

Car Acting Rationally

avoid the crowd and crash the owner for minimum total loss

which is smarter?
—depending on where I am, maybe? ☺

Traditional vs. Modern [My] Definition of AI

Traditional Definition

humanly $\approx$ intelligently $\approx$ rationally

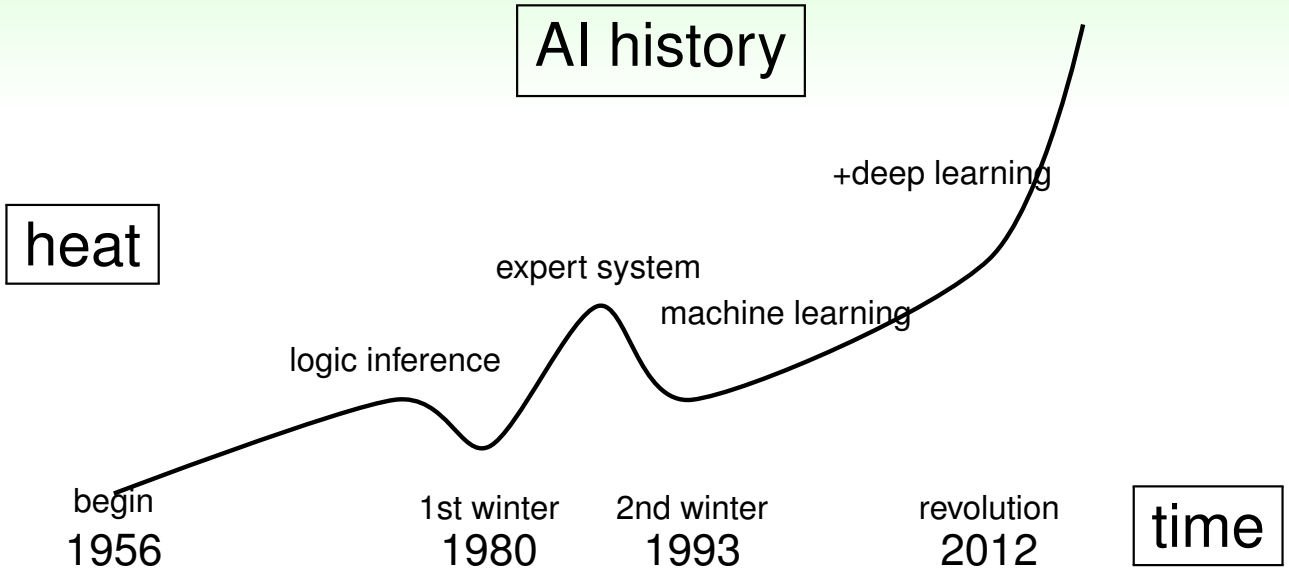
My Definition

intelligently $\approx$ easily

is your smart phone 'smart'? ☺

modern artificial intelligence
$=$ application intelligence

AI Milestones



- first AI winter: AI cannot solve 'combinatorial explosion' problems
- second AI winter: expert system failed to scale

reason of winters: expectation mismatch

What's Different Now?

More Data

- cheaper storage
- Internet companies

Better Algorithms

- decades of research
- e.g. deep learning

Faster Computation

- cloud computing
- GPU computing

Healthier Mindset

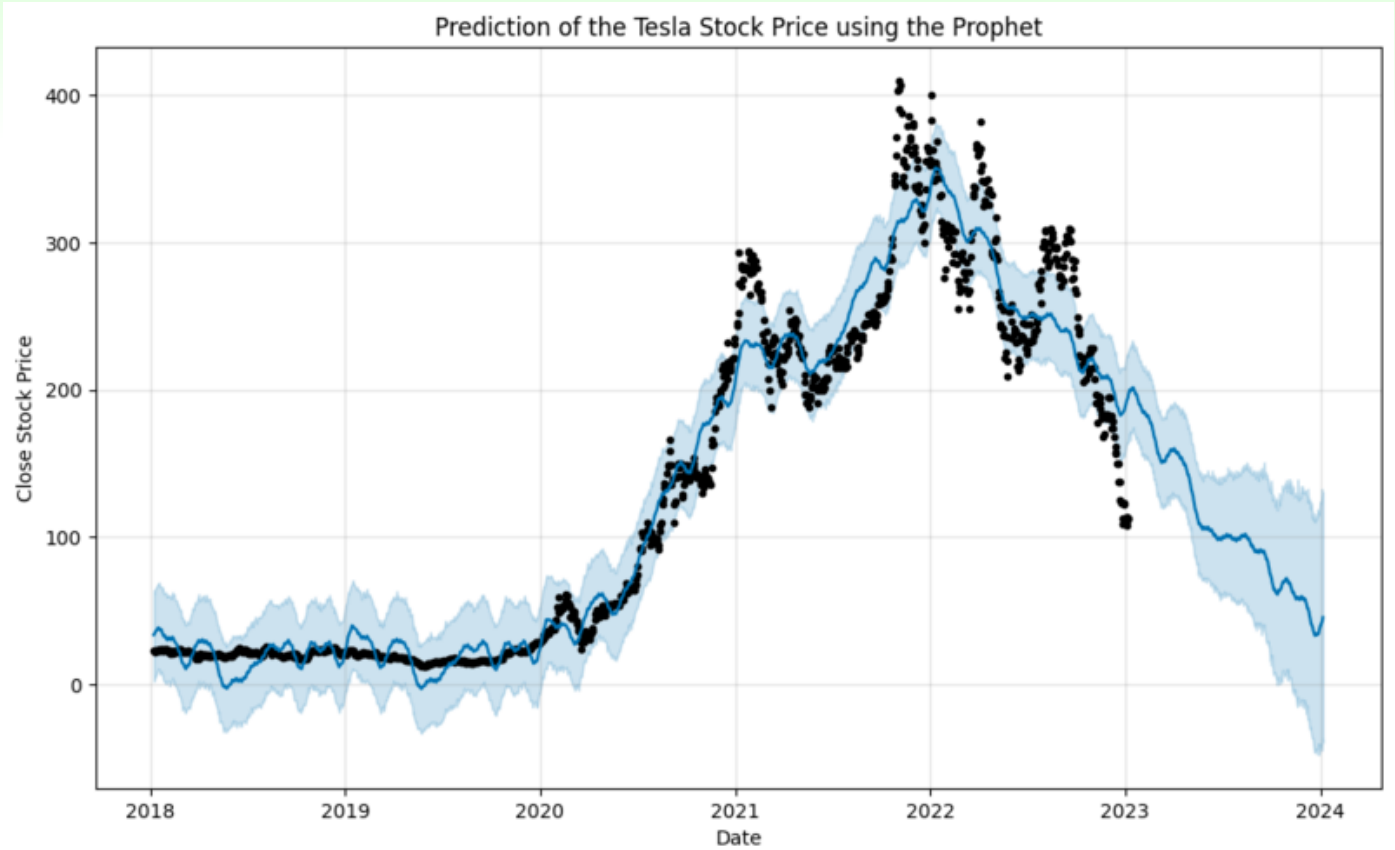
- reasonable wishes
- key breakthroughs

data-enabled AI (with Machine Learning): mainstream nowadays

From AI to GAI: Is This GAI? (1/4)

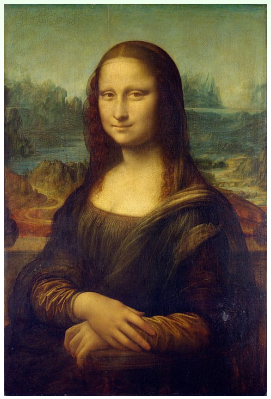
generative,
but arguably no intelligence

From AI to GAI: Is This GAI? (2/4)



predictive intelligence,
but arguably not generative

From AI to GAI: Is This GAI? (3/4)



Leonardo da Vinci,

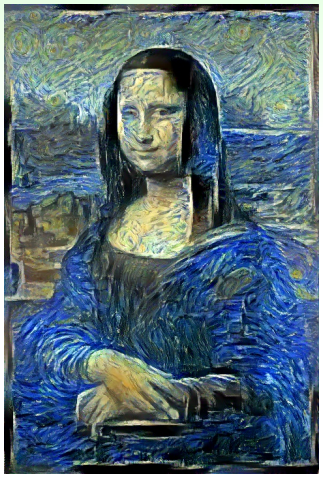
in Public Domain

$+$

Van Gogh,

in Public Domain

$\Rightarrow$



Pjfinlay,

with CC0

all images are downloaded from Wikipedia

generative intelligence,
or just (predictive) image processing?

Properties of Generative AI

Recognitive AI

Listen/Read/Watch

Generative AI

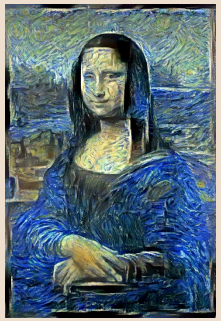
Speak/Write/Draw

Two Properties of Generative AI

variation (creativity)

(Pictures Extracted from Ho et al. for educational purposes)

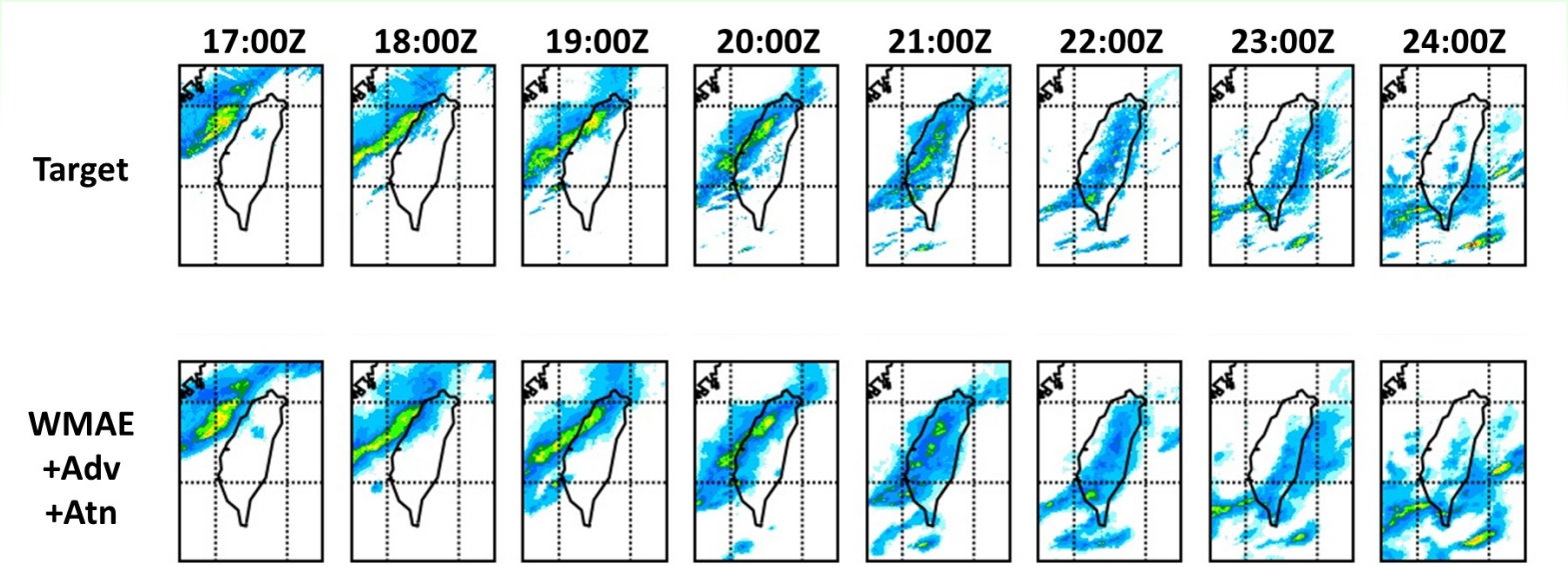
complexity (structure)



(Pictures Licensed under CC0 on Wikipedia)

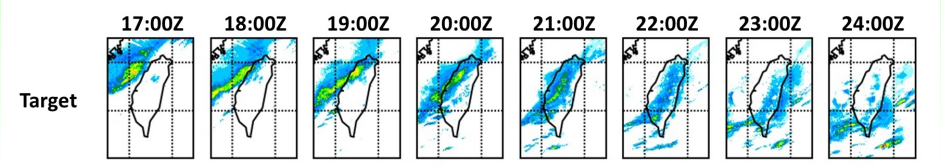
Generative AI :
complex outputs with variations

From AI to GAI: Is This GAI? (4/4)

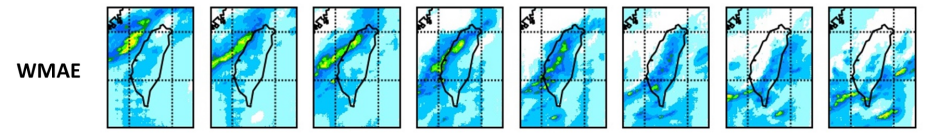


predictive: time-series prediction;
generative: complex output;
or does it matter? ☺

A Story on Modern Generative AI



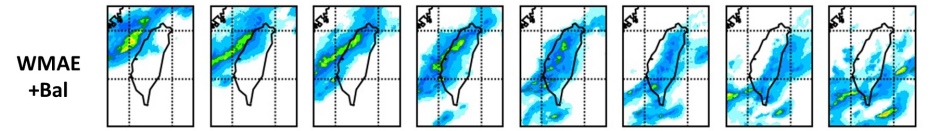
let's start with multi-pixel regression



appears always raining, why?

regression model feels "safer" to predict a bit of rain

let's force no-rain by discretizing regression output

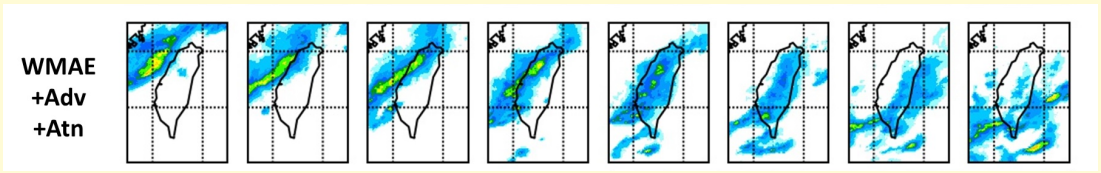


looks unnatural, why?

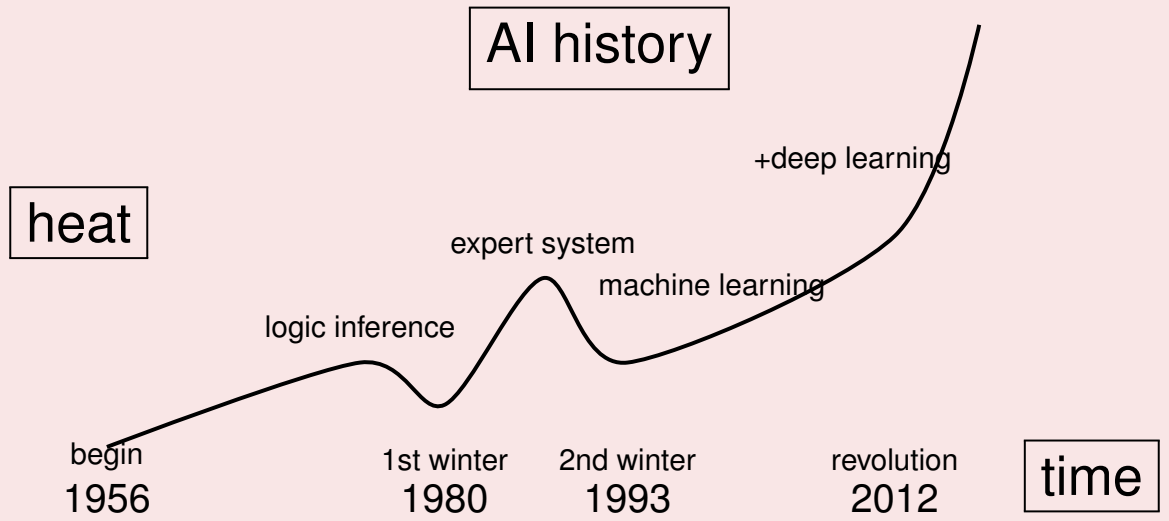
multi-pixel regression + discretization lack details

force human-indistinguishable by generative AI

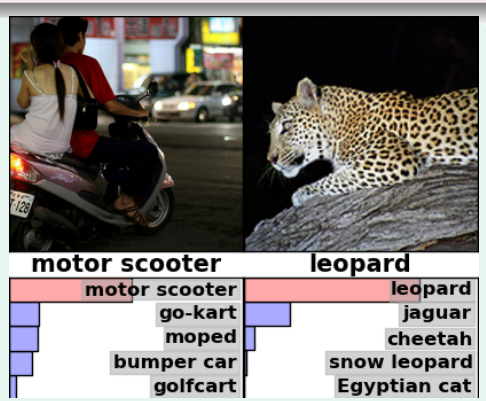
modern generative AI with mixed tools:



History (?): From Recognition to Generation

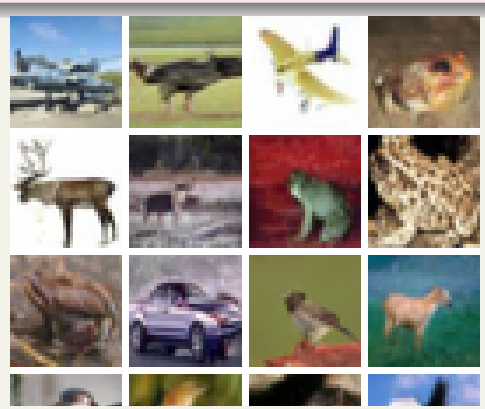


(Krizhevsky et al., 2012) ImageNet classification with deep convolutional neural networks



Picture extracted from the original paper of Krizhevsky et al. for educational purposes

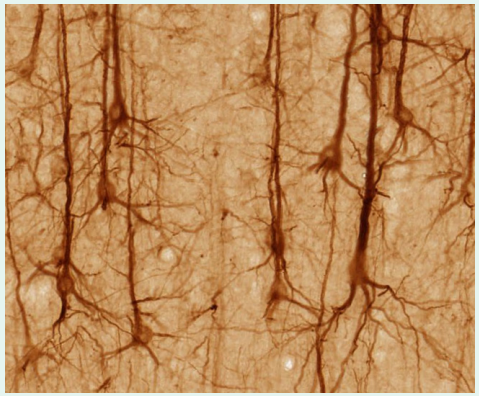
(Ho et al., 2020) Denoising diffusion probabilistic models



Picture extracted from the original paper of Ho et al. for educational purposes

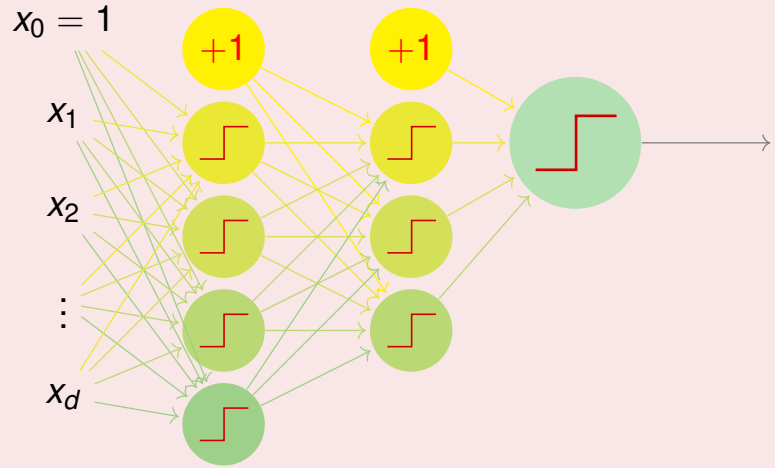
deep learning (neural network) speeds up realizing modern AI

Neural Network: from Bird to Airplanes



by UC Regents Davis campus-brainmaps.org.

Licensed under CC BY 3.0 via Wikimedia Commons



by Lauris Rubenis. Licensed under CC BY 2.0 via https://flic.kr/p/fkVuZX

by Pedro Ribeiro Simões. Licensed under CC BY 2.0 via https://flic.kr/p/adiV7b

neural network: a bio-inspired model

From Wright Flyer (1903) to Commercial Airplanes (1919–)

by Wright Brothers. Licensed under Public Domain via US Library of Congress

by Pedro Ribeiro Simões. Licensed under CC BY 2.0 via `https://flic.kr/p/adiV7b`

we are at wright-flyer-age of (generative) AI

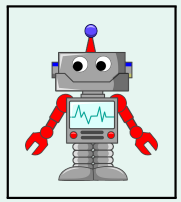
What's Needed before Wider Acceptance

- war? ☺
- technology advancements
  —like lighter materials, more efficient engines, better control
- regulations
  —like laws, licenses, etc.
- trials
  —understanding success and failure cases

will discuss research on better control

Score-based Generative Model (SGM)

SGM

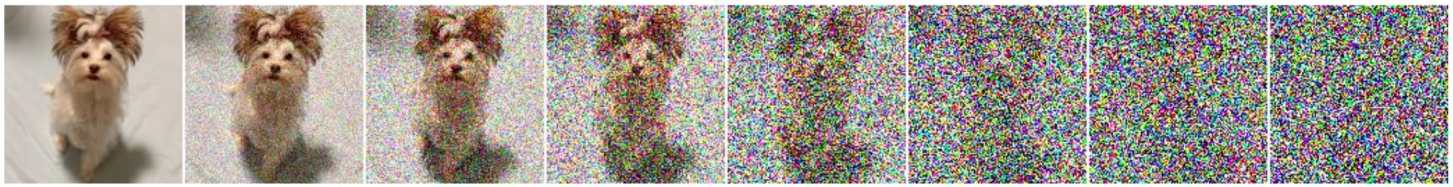


generated image **x** ← ← random image **z**

(Free Content Use from https://pixabay.com/vectors/robot-machine-technology-science-312566/)

Forward SDE (data → noise)

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



score function

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

Reverse SDE (noise → data)

(Figure 1 from Song et al., ICLR 2021)

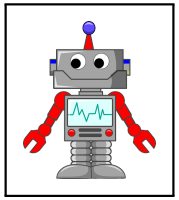
high-quality generation
when score function can be estimated

Conditional SGM

SGM

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

Conditional SGM

generated $\mathbf{x}$ $\leftarrow$  $\leftarrow$ random $\mathbf{z}$ & $y = \text{dog}$

(Free Content Use from https://pixabay.com/vectors/robot-machine-technology-science-312566/)

Conditional SGM

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

Hello, 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

Our Contributions

manuscript: Paul Kuo-Ming Huang, Si-An Chen, and Hsuan-Tien Lin. Semi-Supervised Classifier Guidance with Self-Calibration for Conditional Score-Based Generation.

our impacts: an in-depth study of cSGM, which ...

- makes its classifier design more robust with a novel angle of regularization
- reduces the use of labeled data significantly
- achieves state-of-the-art conditional generation performance in semi-supervised setting

next: our fundamental research attempt

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

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

$$= \nabla_{\mathbf{x}} \log p(\mathbf{x})$$

$$= \nabla_{\mathbf{x}} \log \sum_k \exp(h_\theta(\mathbf{x}, k))$$

$$- \underbrace{\nabla_{\mathbf{x}} \log(\text{normalization})}_{0}$$

HEAVY MATH

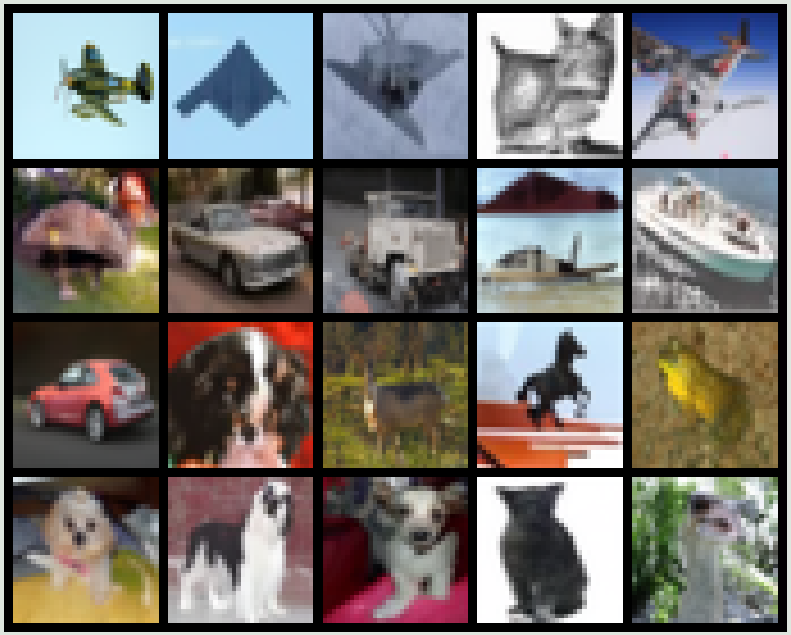
classifier can regularize itself by viewing from angle of unconditional SGM (proof omitted ☺)

Comparison to Original CGSGM

with merely 5% of labeled data

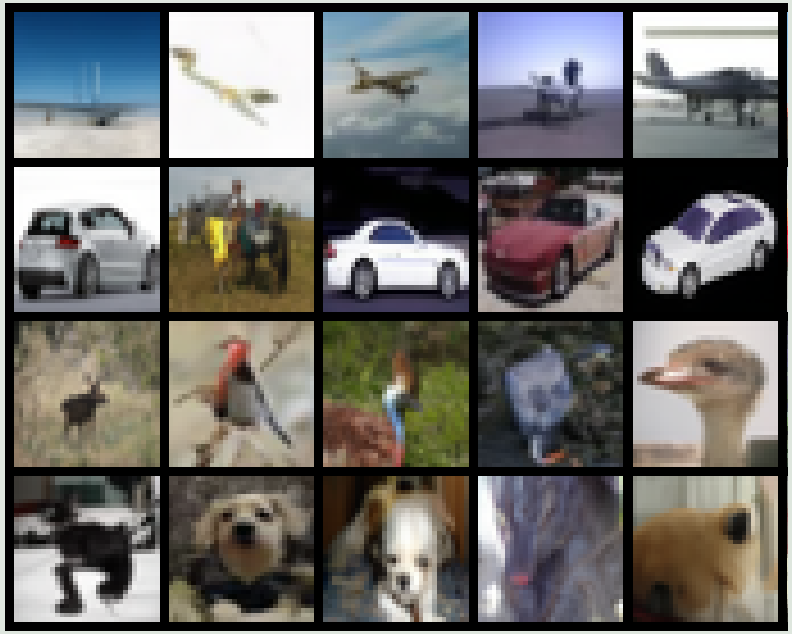
Original CGSVM

Intra-FID (↓) 31.17



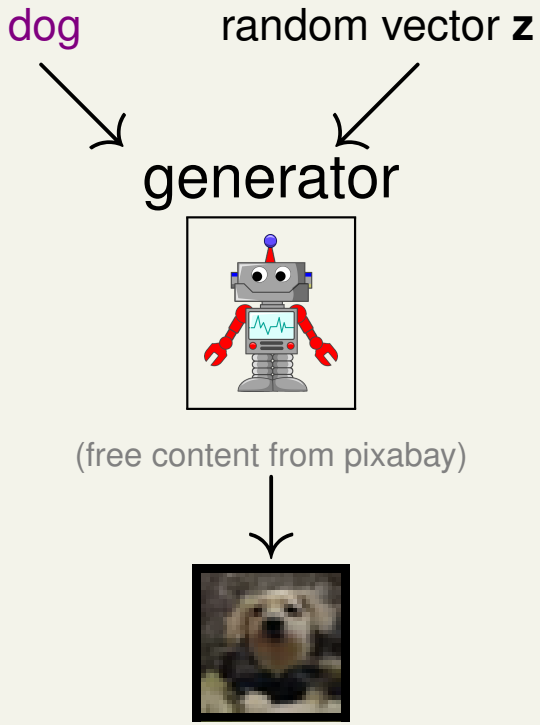
Our Improved CGSFM

Intra-FID (↓) 18.95



ours: better quality & more accurate

Technical Summary



- creativity can go wild
  —regularization by another view helps control
  - improved CGSGM: another view of classifier as unconditional SGM
- most importantly, math helps! ☺
  —more efforts on fundamental research needed
  - energy-based parameterization helps

enough about boring research ☺,
let's share some final wisdom

My Thoughts after Research/(Teaching) Attempts

need research on process

manipulation challenge

generating something is easy;
generating good thing is difficult

—need research on control

certification challenge

trying is easy;
systematic testing is difficult

—need research on evaluation

let's research more to move GAI

to trustworthy commercial tools