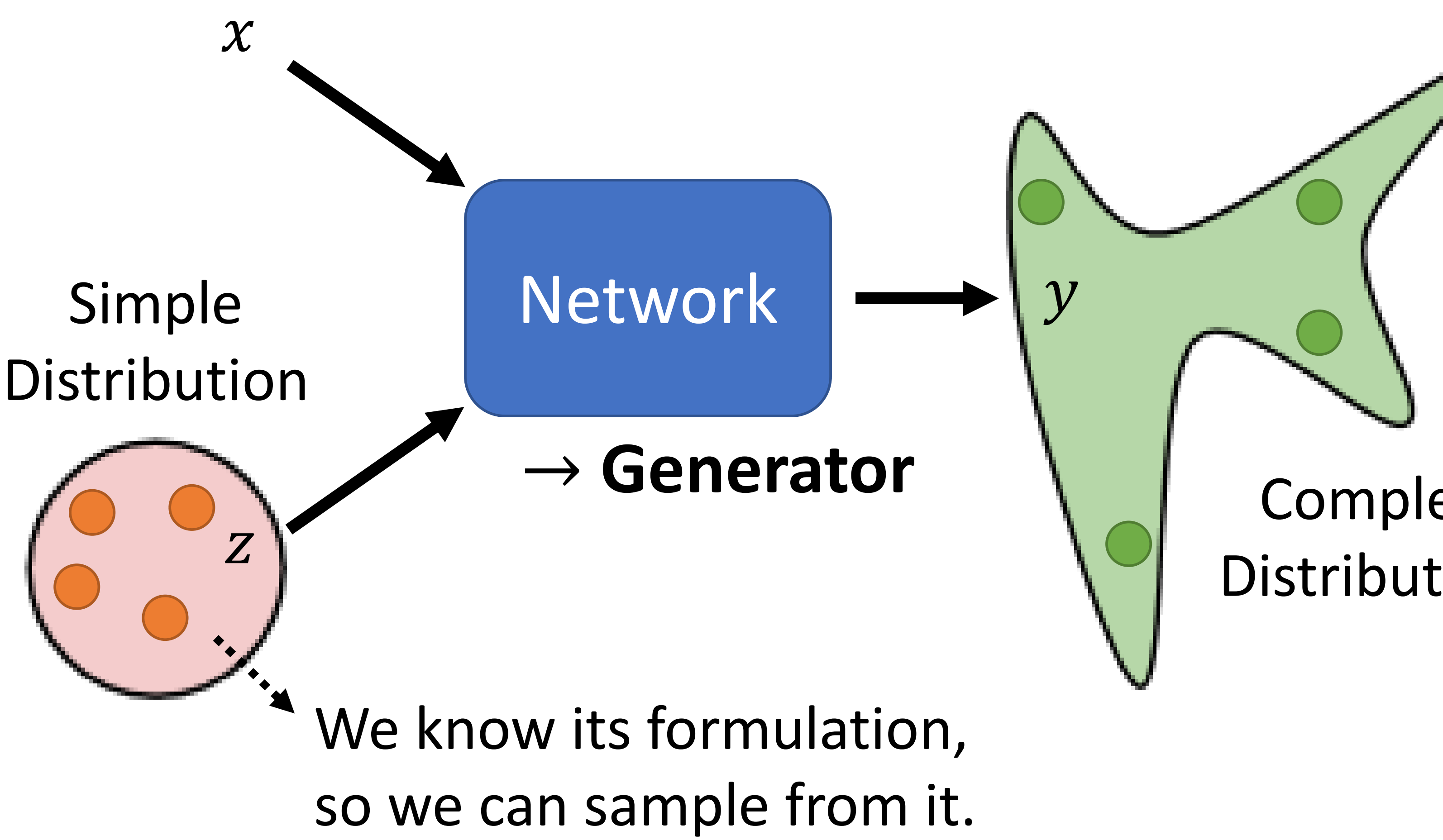
Generation

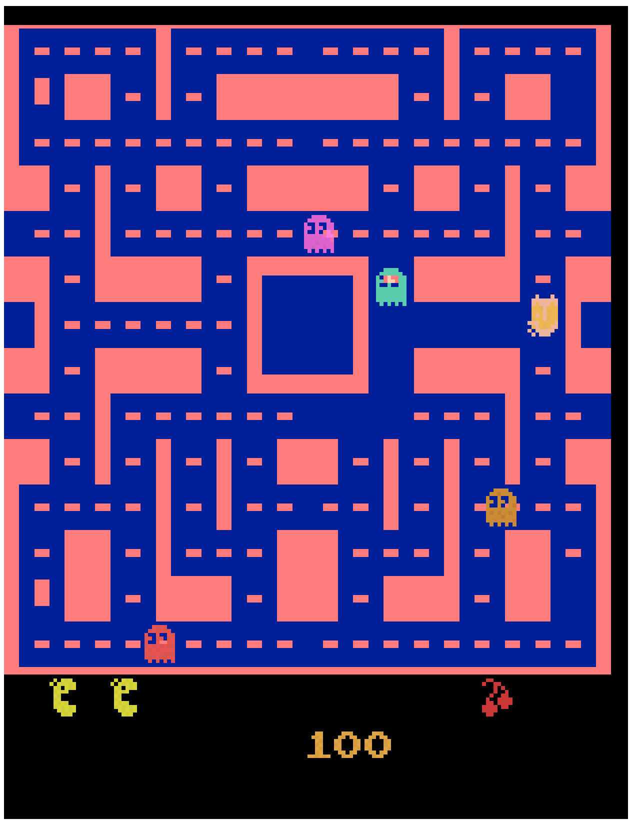
Hung-yi Lee 李宏毅

Network as Generator



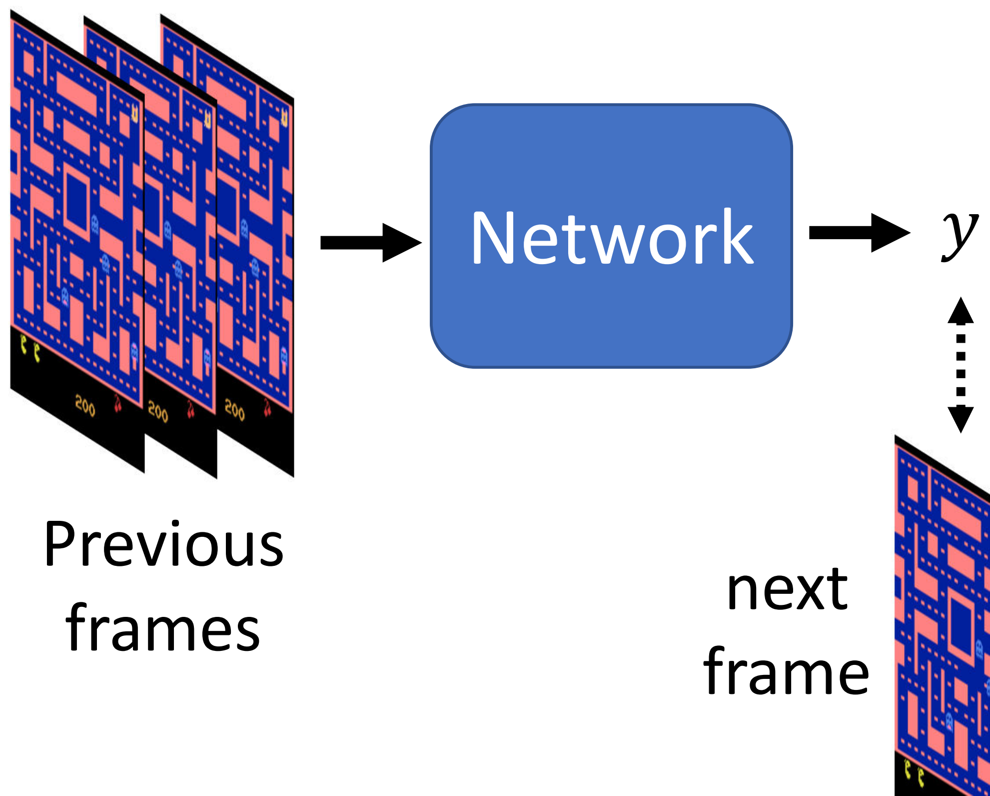
Source: https://github.com/dyelax/Adversarial_Video_G

Why distribution?



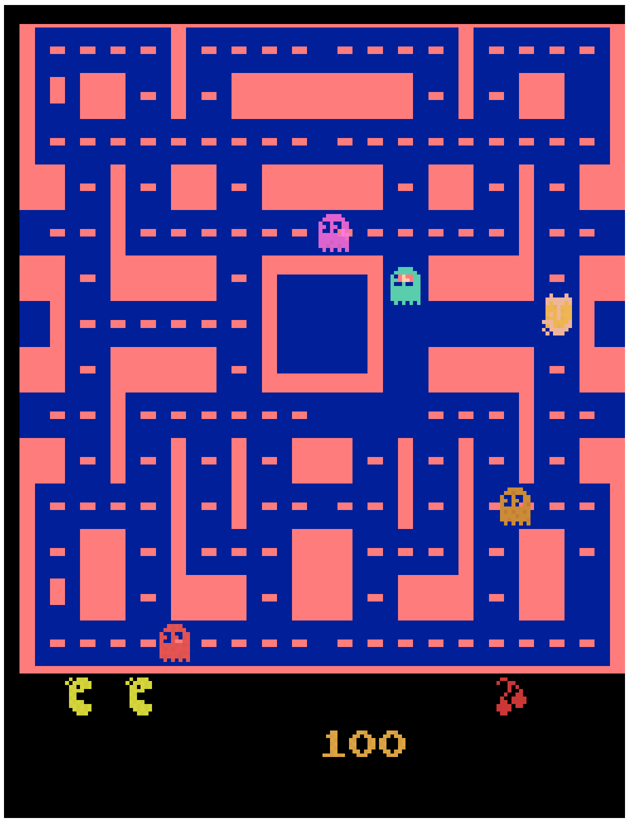
Real Video

<u>Video Prediction</u>



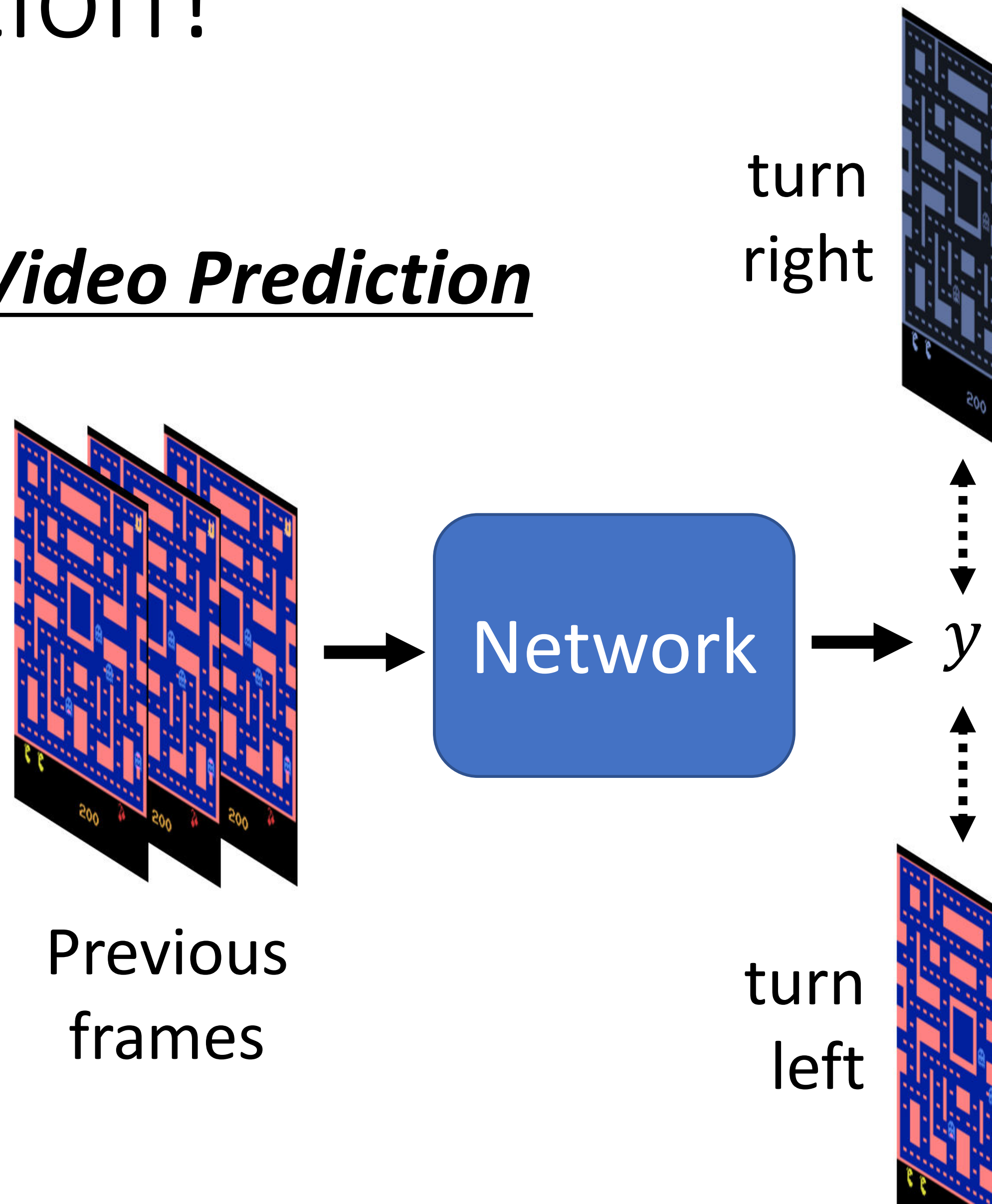
Source: https://github.com/dyelax/Adversarial_Video_G

Why distribution?



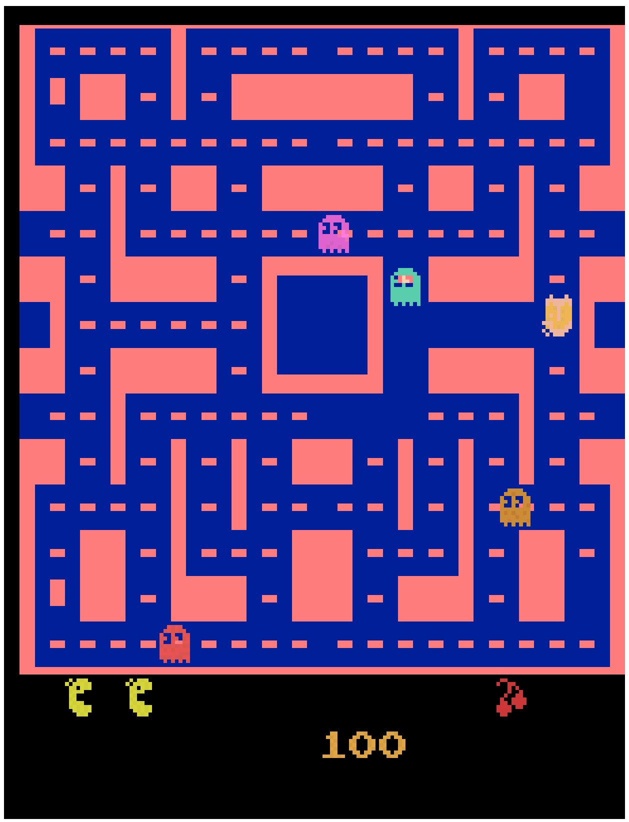
Prediction

Video Prediction



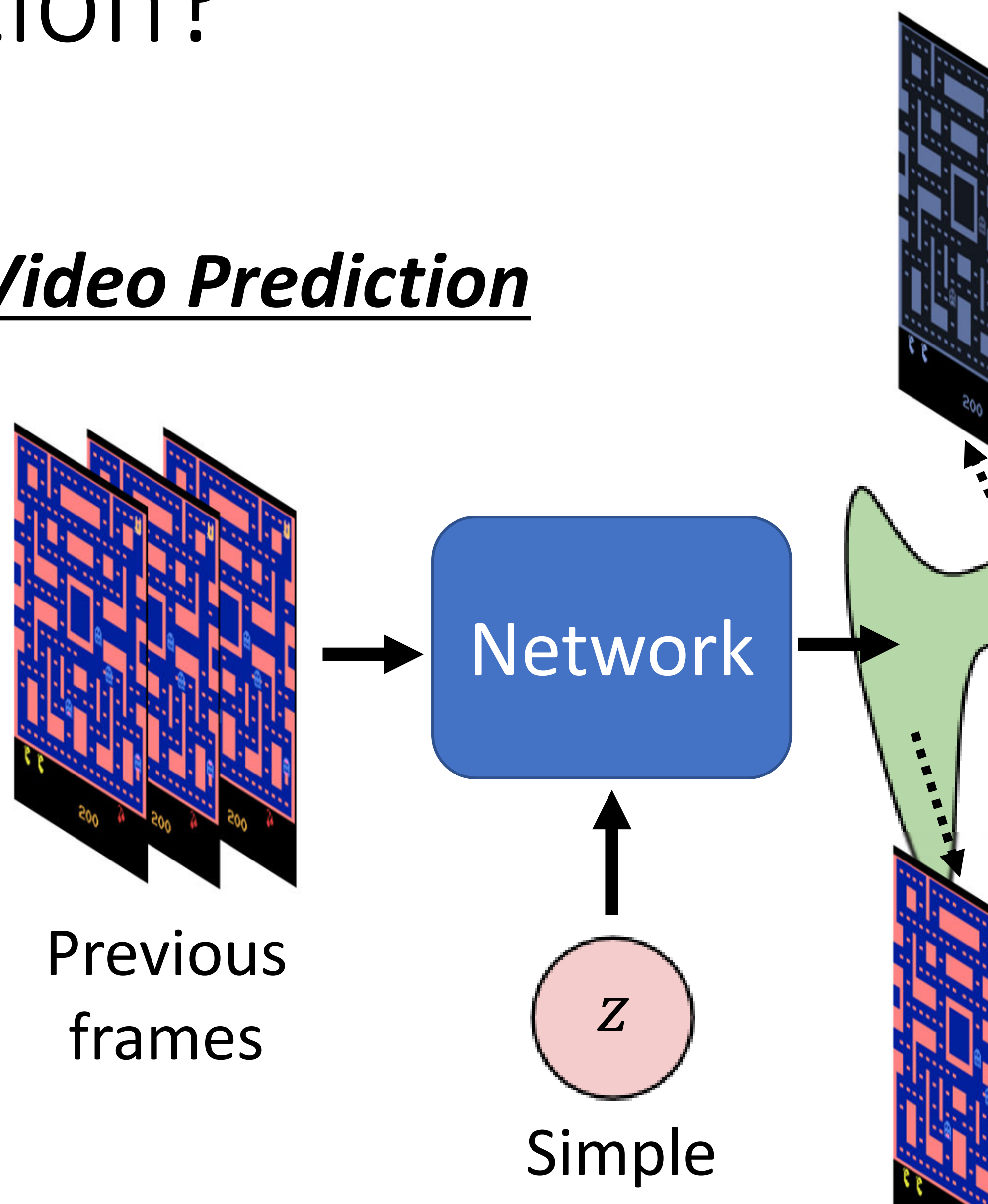
Source: https://github.com/dyelax/Adversarial_Video_G

Why distribution?



Prediction

Video Prediction



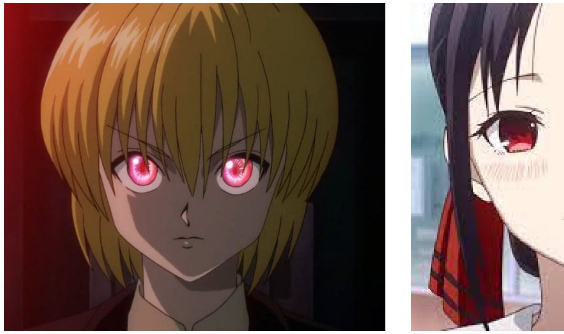
Why distribution?

(The same input has different ou

- Especially for the tasks needs "*creativity*"

<u>Drawing</u>

Character
with red eyes → Network →



<u>Chatbot</u>

你知道輝夜是
誰嗎? → Network

她是秀知院學生會

她開創了忍者時代

Generative Adversarial Network (GAN)

GAN

- How to pronounce “GAN”?

Google 小姐

All Kinds of GAN …

https://github.com/hindupuravinash/the-g

GAN

ACGAN

BGAN

CGAN

DCGAN

EBGAN

fGAN

GoGAN

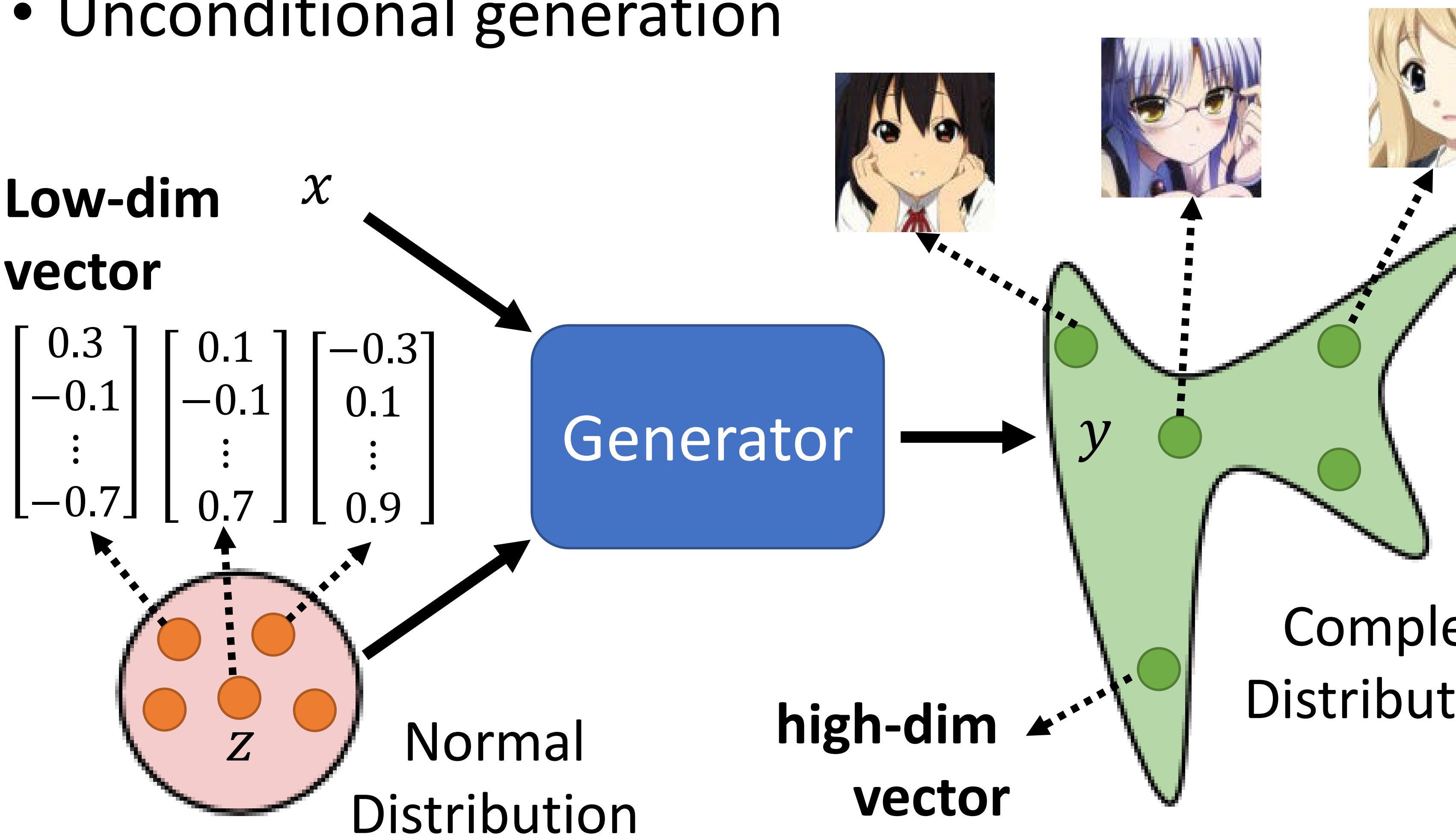
⋮

- SeUDA - Semantic-Aware Generative Adversarial Nets for Unsupervise Segmentation
- SG-GAN - Semantic-aware Grad-GAN for Virtual-to-Real Urban Scene
- SG-GAN - Sparsely Grouped Multi-task Generative Adversarial Networ
- SGAN - Texture Synthesis with Spatial Generative Adversarial Networks
- SGAN - Stacked Generative Adversarial Networks (github)
- SGAN - Steganographic Generative Adversarial Networks
- SGAN - SGAN: An Alternative Training of Generative Adversarial Netwo
- SGAN - CT Image Enhancement Using Stacked Generative Adversarial Segmentation Improvement
- sGAN - Generative Adversarial Training for MRA Image Synthesis Using
- SiftingGAN - SiftingGAN: Generating and Sifting Labeled Samples to In Classification Baseline in vitro
- SiGAN - SiGAN: Siamese Generative Adversarial Network for Identity-P
- SimGAN - Learning from Simulated and Unsupervised Images through
- SisGAN - Semantic Image Synthesis via Adversarial Learning

Mihaela Rosca, Balaji Lakshminarayanan, David Warde-Farley, Shakir Mohamed, "Variational Approaches Encoding Generative Adversarial Networks", arXiv, 2017

Anime Face Generation

- Unconditional generation

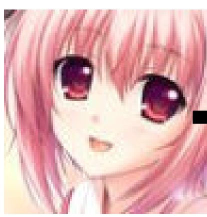


Discriminator

It is a neural netwo
(that is, a function)

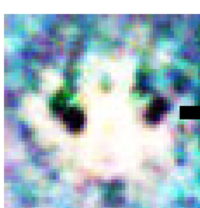
image → Discri-minator →

Scalar: Larger means
smaller value fake.



→ Discri-minator → 1.0

→ Discri-minator →



→ Discri-minator → 0.1

→ Discri-minator →

Basic Idea of GAN

Genera

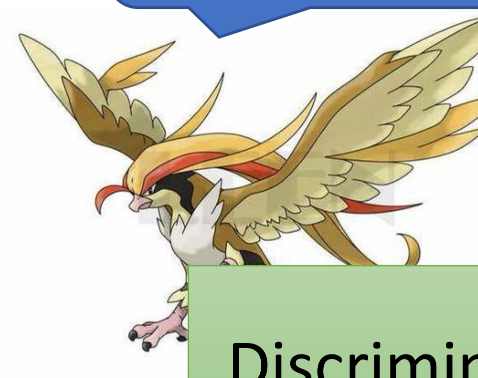
Brown

veins

Butterflies are not brown

Butterflies do not have veins

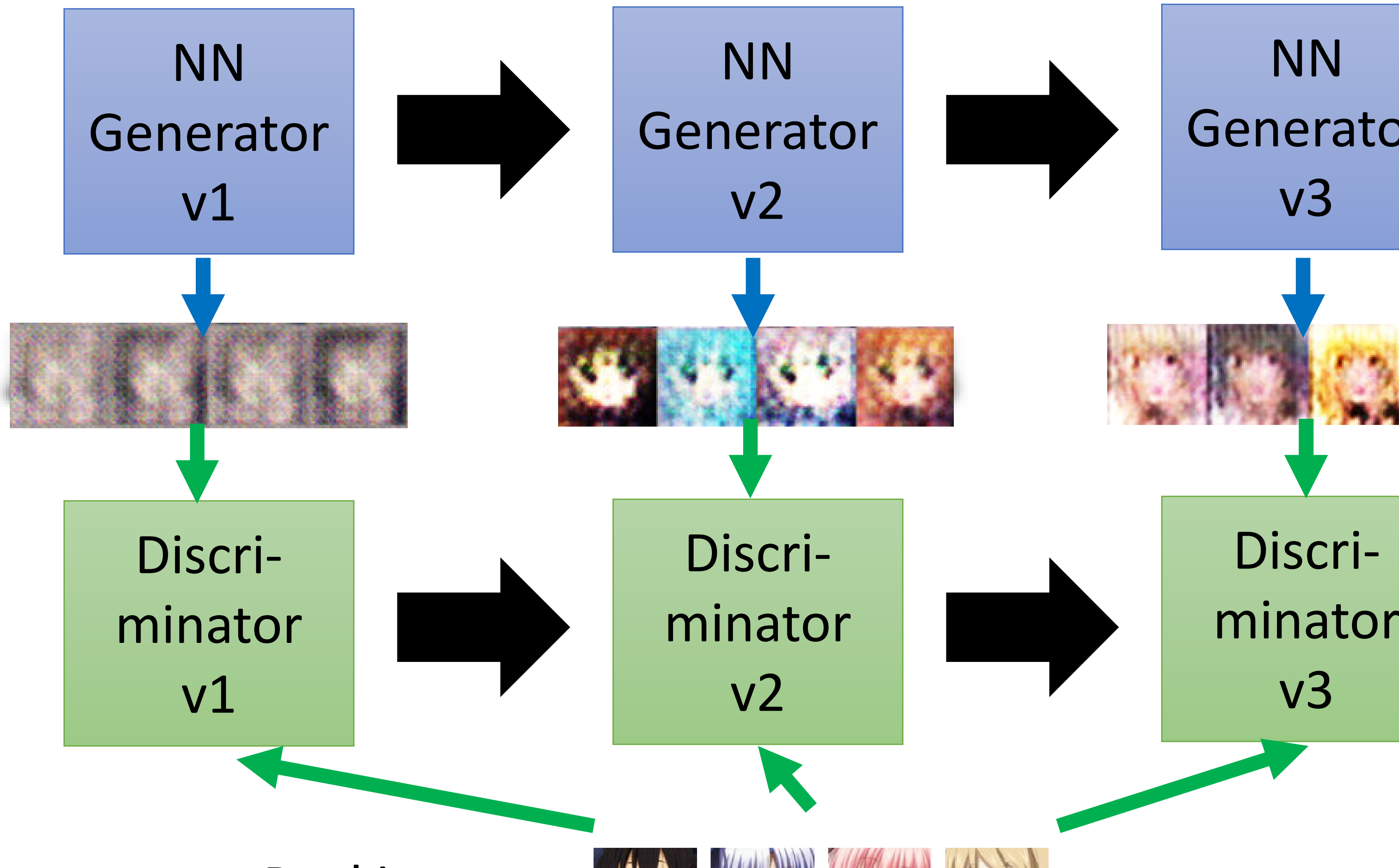
....



Discrimin

Basic Idea of GAN

This is where the term
"***adversarial***" comes



Basic Idea of GAN

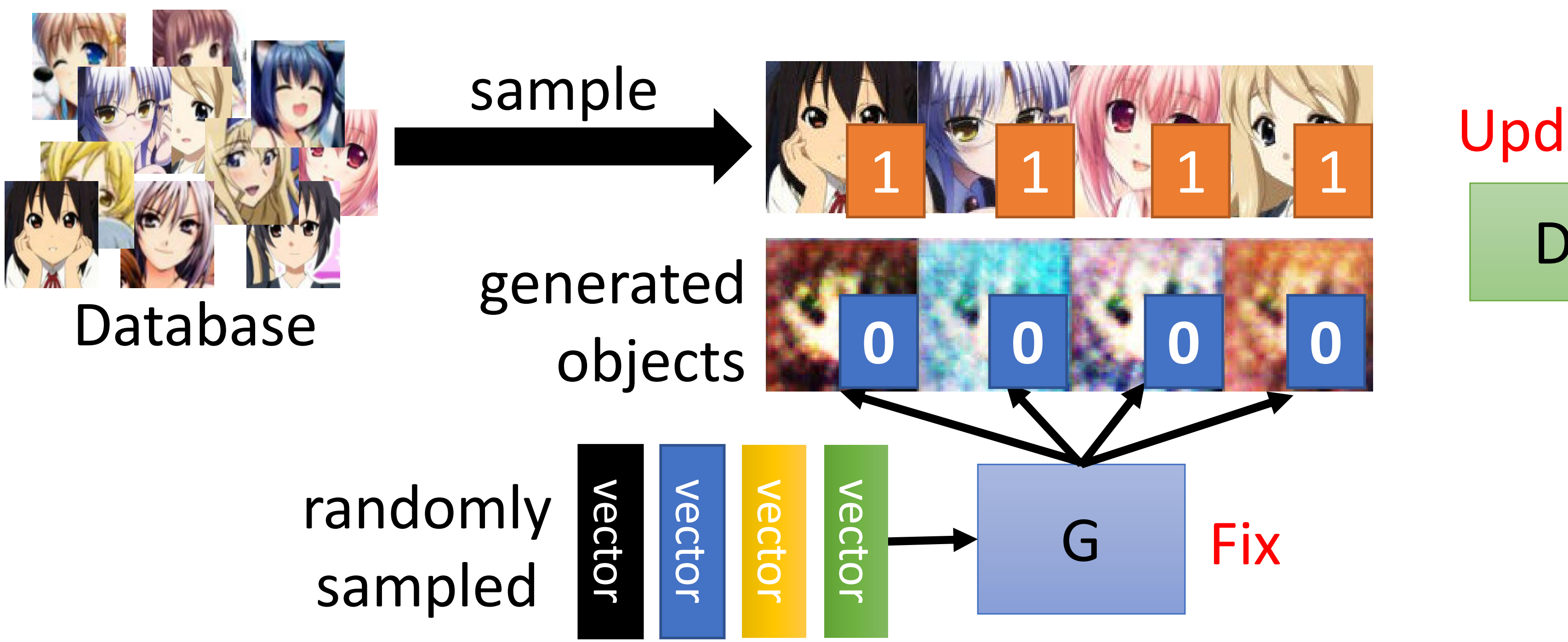
- 寫作敵人，唸做朋友

Algorithm

- Initialize generator and discriminator

- In each training iteration:

Step 1: Fix generator G, and update discriminator D



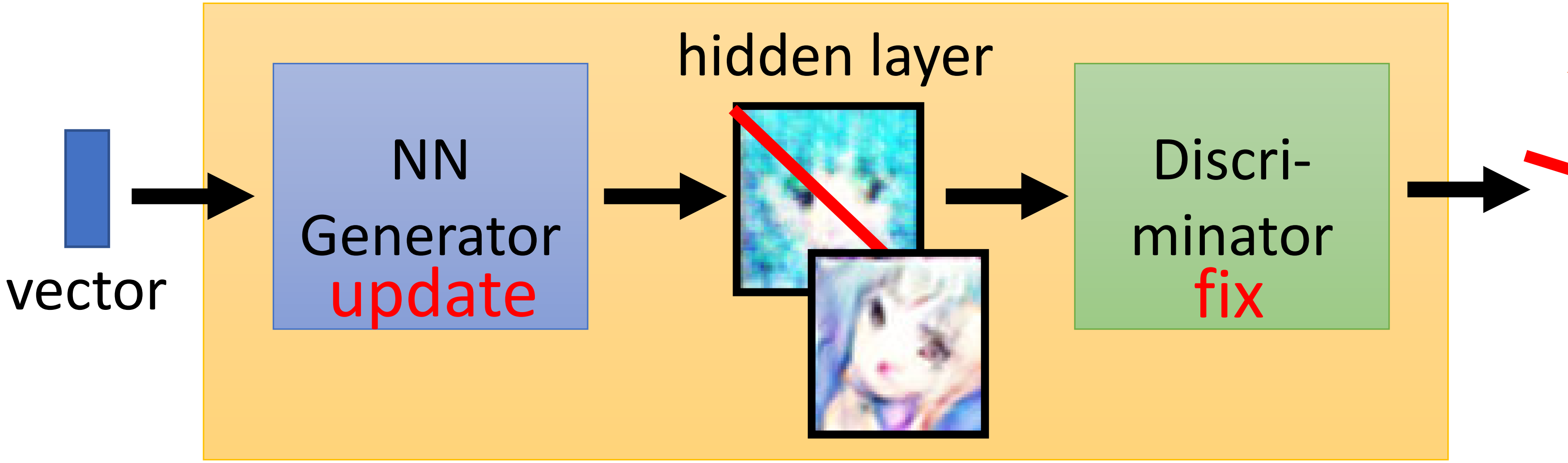
Discriminator learns to assign high scores to real ob

Algorithm

- Initialize generator and discriminator G D
- In each training iteration:

Step 2: Fix discriminator D, and update generator G

Generator learns to "fool" the discriminator

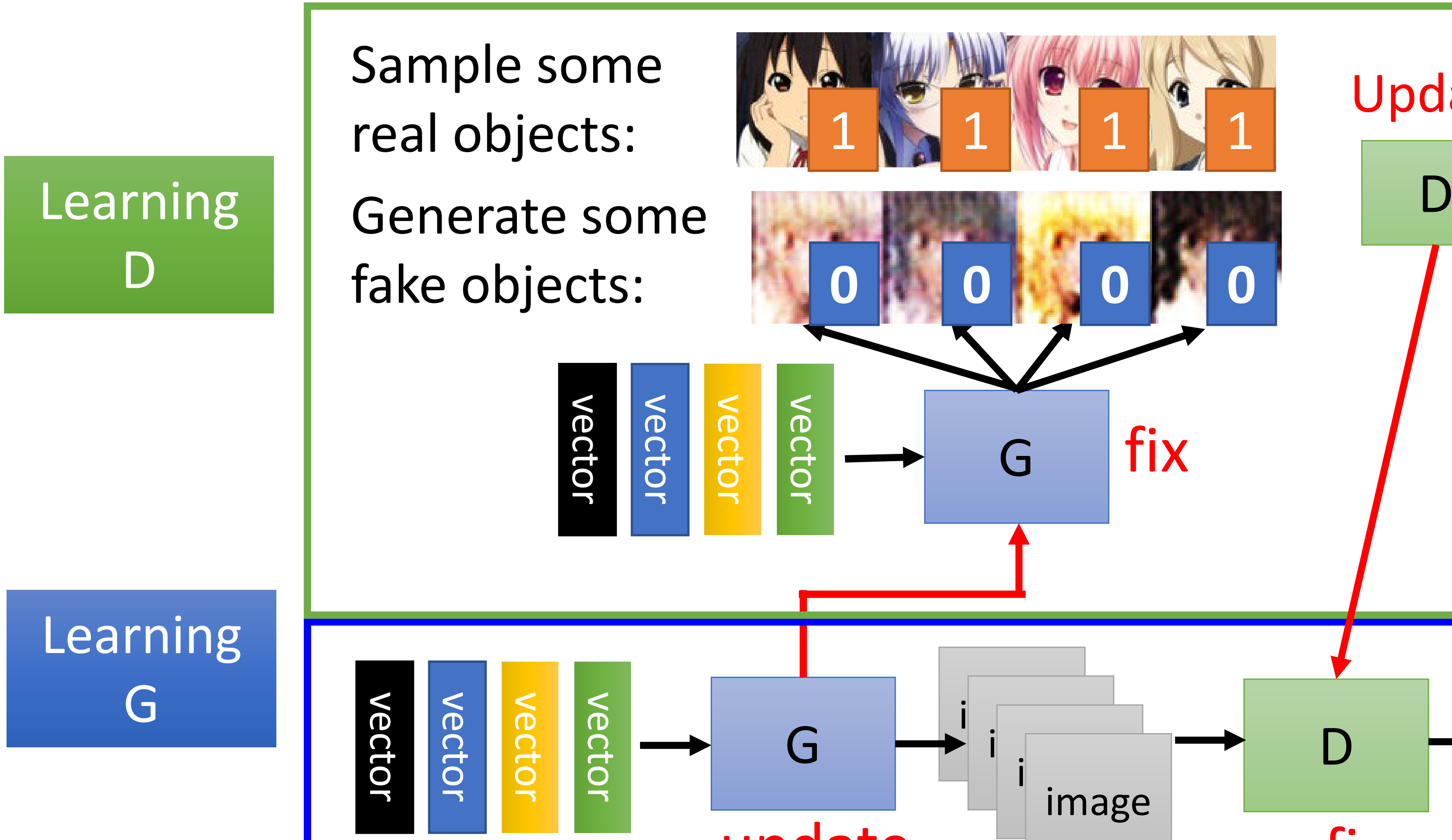


large network

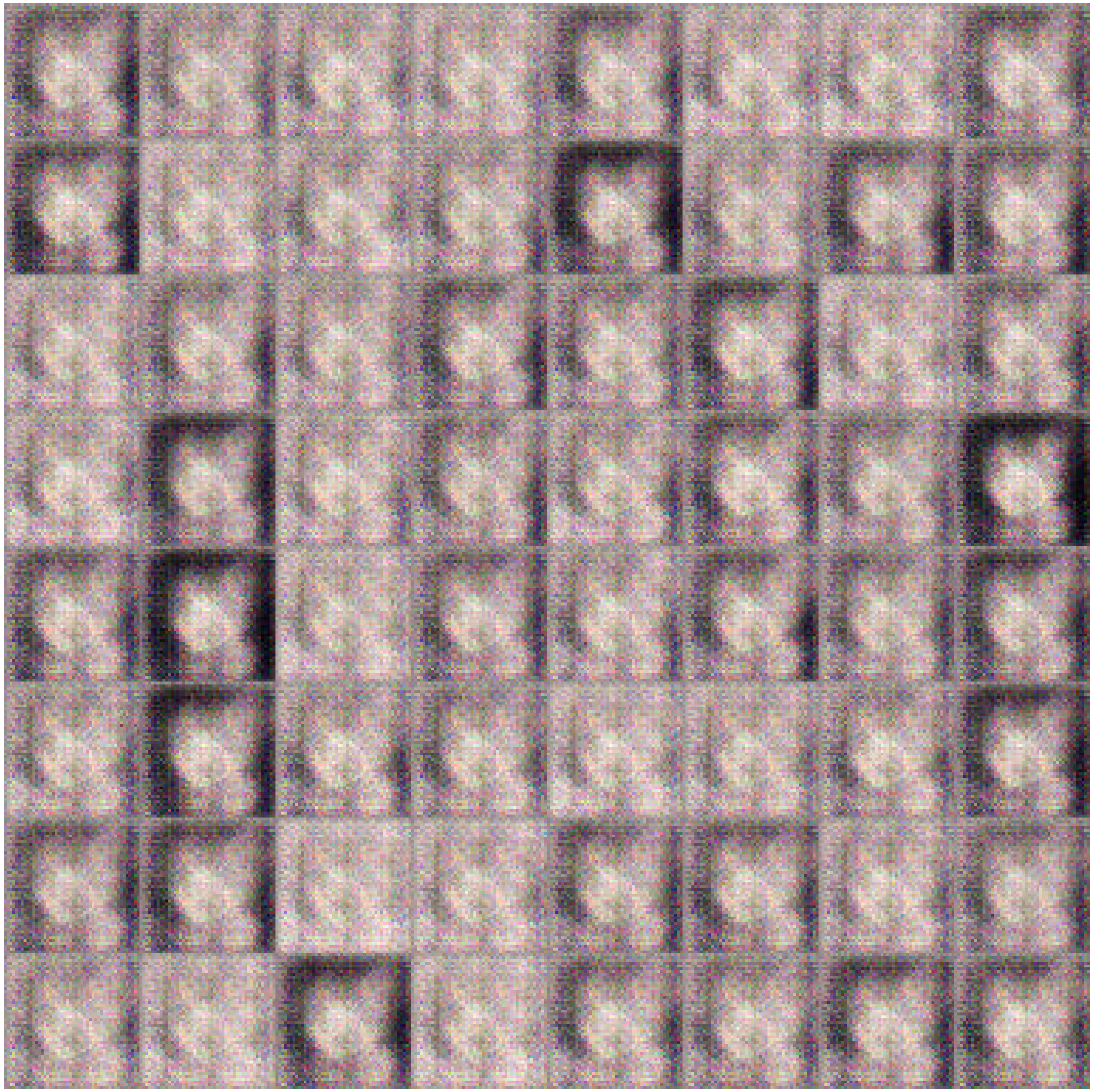
Algorithm

- Initialize generator and discriminator

- In each training iteration:



Anime Face Generation



100 updates

Anime Face Generation

1000 updates

Anime Face Generation

2000 updates

Anime Face Generation

5000 updates

Anime Face Generation

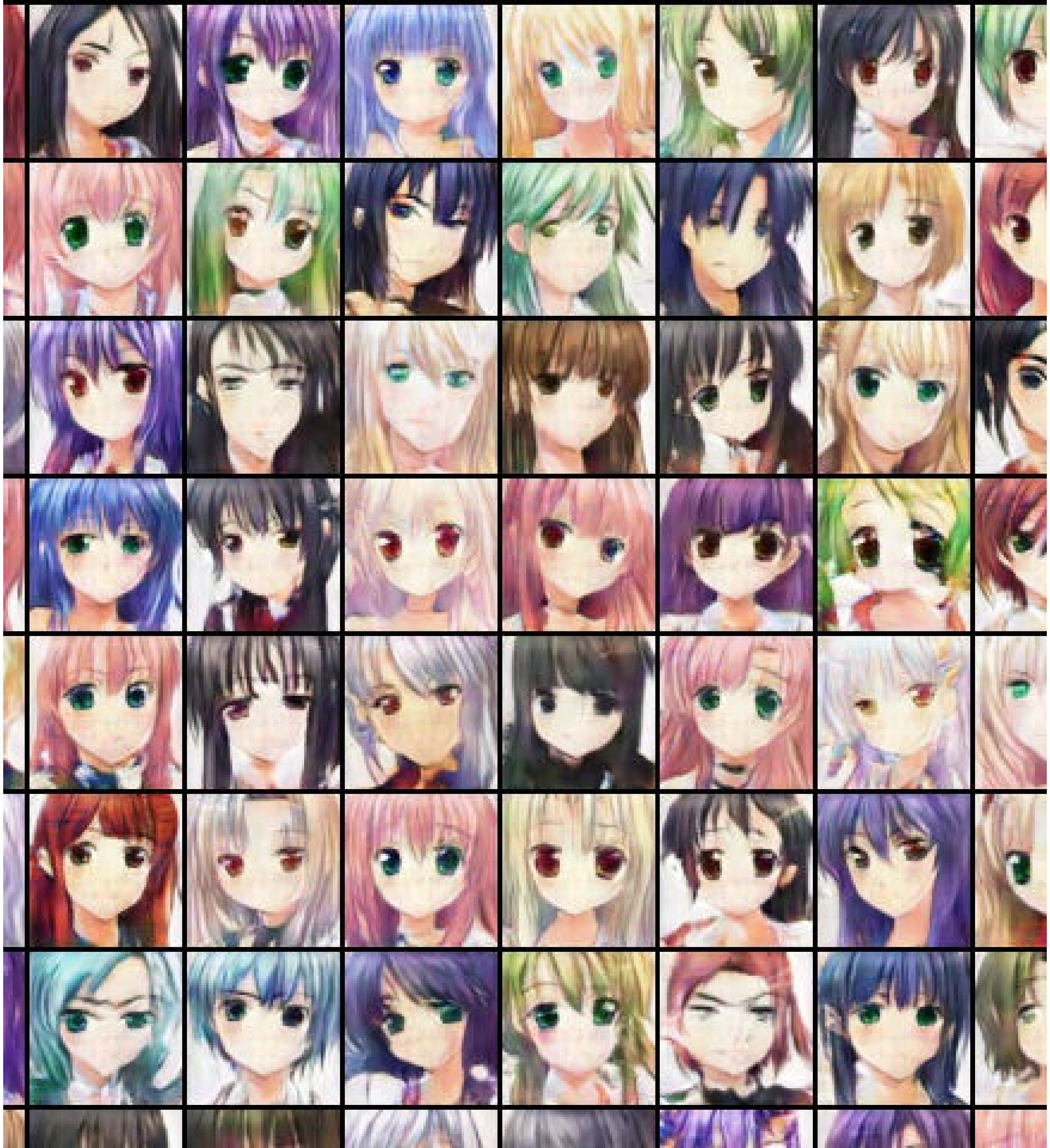
10,000 updates

Anime Face Generation

20,000 updates

Anime Face Generation

50,000 updates



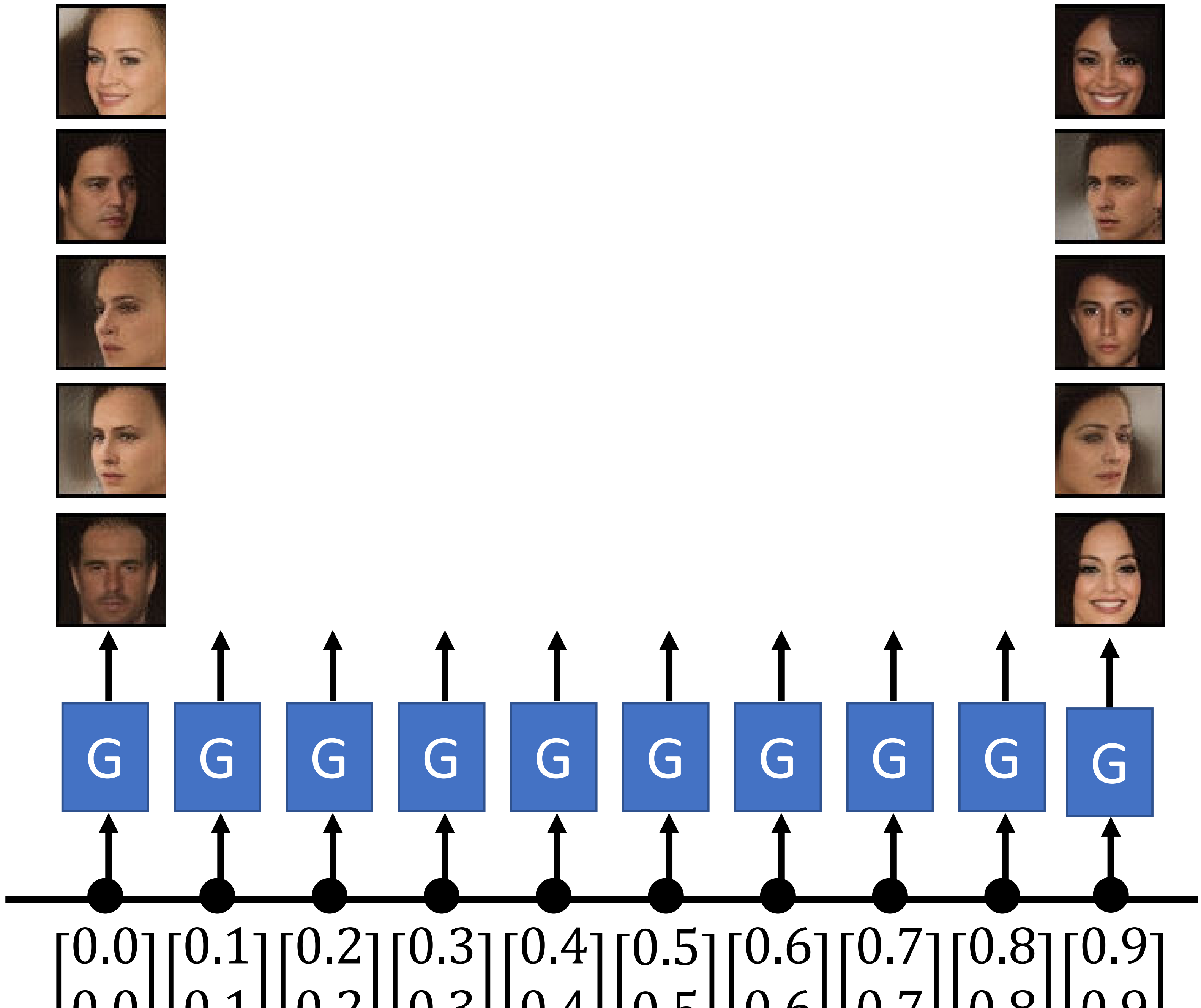
The faces generated machine.

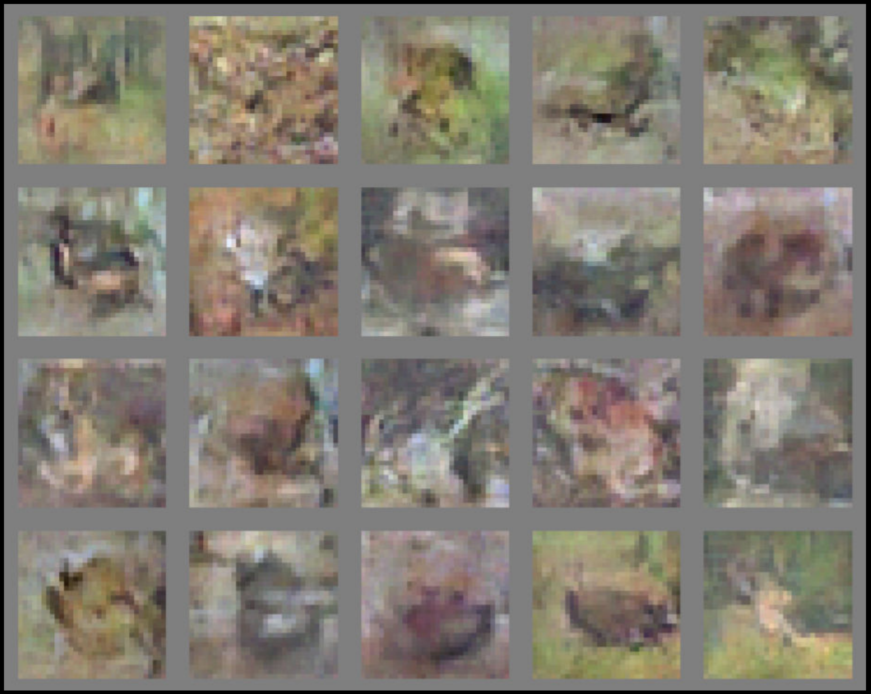
圖片生成
吳宗翰

In 2019, with StyleGAN ……

Source of video:

Progressive GAN


G
G
G
G
G
G
G
G
G
G
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9



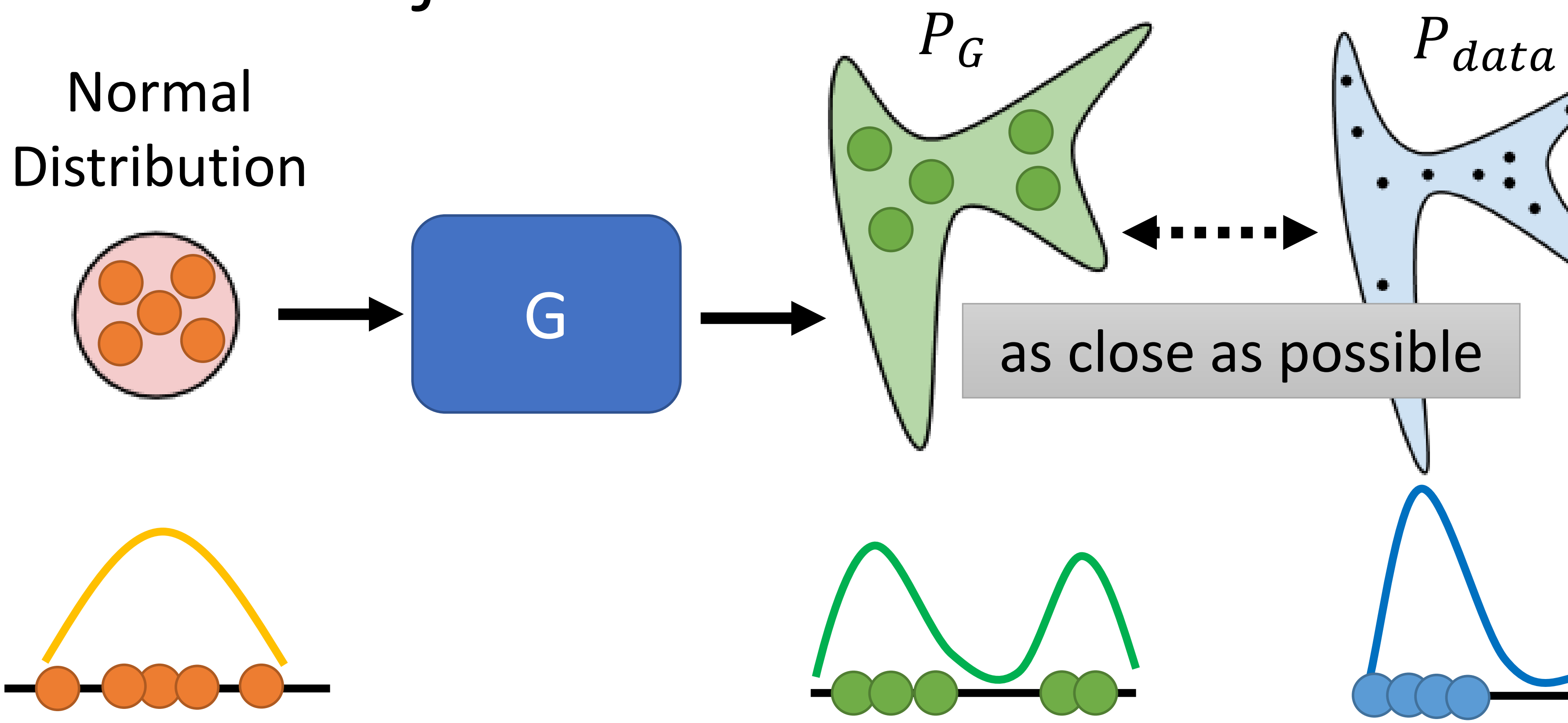
The first GAN

Today …… BigGAN

Theory behind GAN

c.f. $w^*, b^* = argm$

Our Objective



$$G^* = arg\min_G \underline{Div(P_G, P_{data})}$$

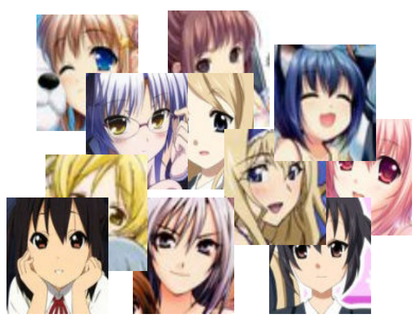
Divergence between distributions P_G and

Sampling is good enough ……

$$G^* = arg\min_G Div(P_G, P_{data})$$

Although we do not know the distributions of P_G and P_{dat}
we can sample from them.

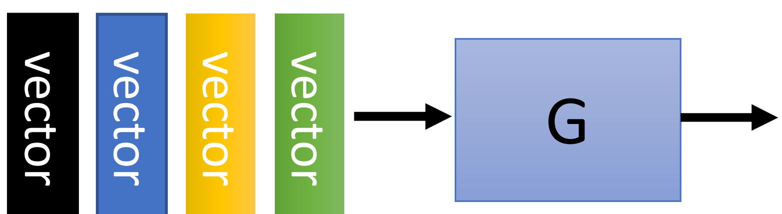
Database

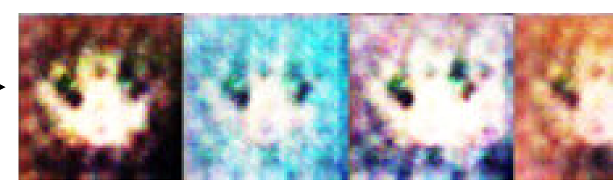


sample

Sampling from P

sample from normal





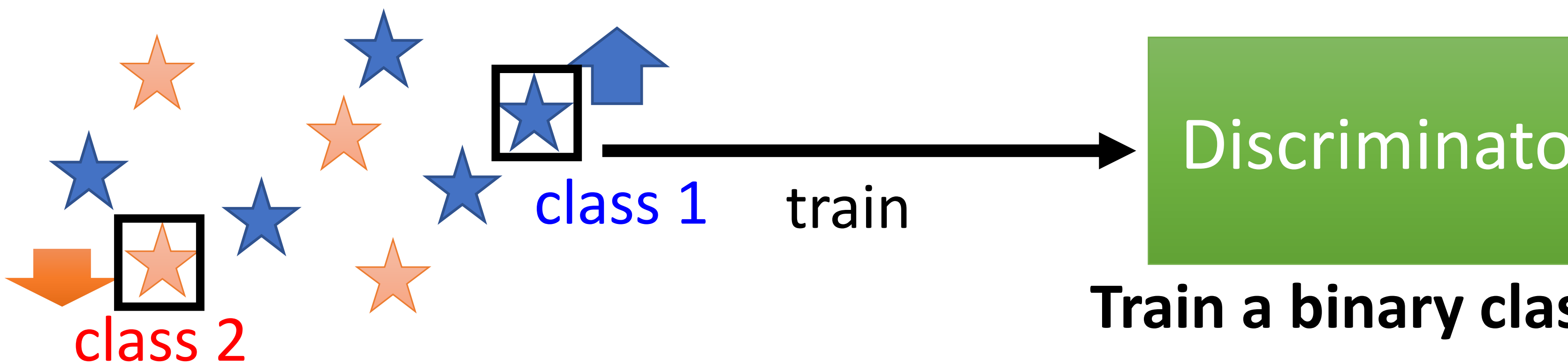
Sampling from

https://arxiv.org/abs/140

Discriminator

$$G^* = arg\min_G Div(P_G, P_{da$$

: data sampled from P_{data} : data sampled from



Train a binary clas

Training: $D^* = arg\max_D \boxed{V(D,G)}$ The value is related to JS divergence.

Objective Function for D

$$V(G,D) = E_{y \sim P_{data}}[logD(y)] + E_{y \sim P_G}[log(1 - D(y$$

$$D^* = arg\max_D \boxed{V(D,G)}$$

=

Training classifier:

Discriminator

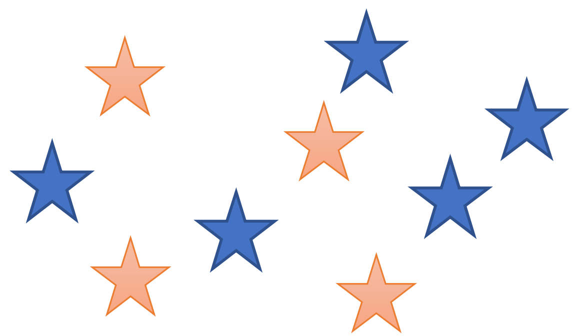
$$G^* = arg\min_G Div(P_G, P_{da}$$

★ : data sampled from P_{data}

★ : data sampled from P_G

Training:

$$D^* = arg\boxed{\max_D V(D,}$$



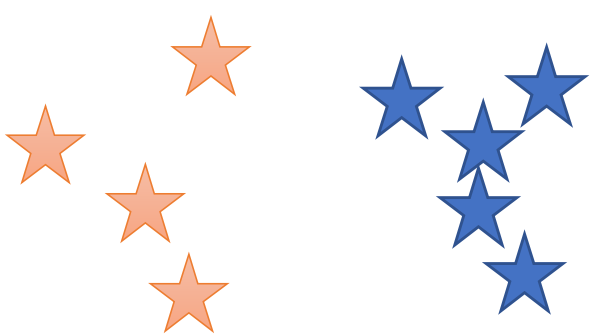
small divergence

train

Discriminato

hard to discrimin

Small $\boxed{\max_D V(D,}$



train

Discriminato

$$G^* = arg\min_G \max_D V(G,D)$$

$$D^* = arg\boxed{\max_D V(D,G)}$$

The maximum objective valu
is related to JS divergence.

- Initialize generator and discriminator
- In each training iteration:

  <u>Step 1</u>: Fix generator G, and update discriminator

  <u>Step 2</u>: Fix discriminator D, and update generator

Can we use other divergence?

Name	$D_f(P\|Q)$	Generator $f(u)$
Total variation	$\frac{1}{2}\int \lvert p(x)-q(x)\rvert \,\mathrm{d}x$	$\frac{1}{2}\lvert u-1\rvert$
Kullback-Leibler	$\int p(x)\log\frac{p(x)}{q(x)}\,\mathrm{d}x$	$u\log u$
Reverse Kullback-Leibler	$\int q(x)\log\frac{q(x)}{p(x)}\,\mathrm{d}x$	$-\log u$
Pearson χ^2	$\int \frac{(q(x)-p(x))^2}{p(x)}\,\mathrm{d}x$	$(u-1)^2$
Neyman χ^2	$\int \frac{(p(x)-q(x))^2}{q(x)}\,\mathrm{d}x$	$\frac{(1-u)^2}{u}$
Squared Hellinger	$\int \left(\sqrt{p(x)}-\sqrt{q(x)}\right)^2\,\mathrm{d}x$	$(\sqrt{u}-1)^2$
Jeffrey	$\int (p(x)-q(x))\log\left(\frac{p(x)}{q(x)}\right)\,\mathrm{d}x$	$(u-1)\log u$
Jensen-Shannon	$\frac{1}{2}\int p(x)\log\frac{2p(x)}{p(x)+q(x)}+q(x)\log\frac{2q(x)}{p(x)+q(x)}\,\mathrm{d}x$	$-(u+1)\log\frac{1+u}{2}+u$ l
Jensen-Shannon-weighted	$\int p(x)\pi\log\frac{p(x)}{\pi p(x)+(1-\pi)q(x)}+(1-\pi)q(x)\log\frac{q(x)}{\pi p(x)+(1-\pi)q(x)}\,\mathrm{d}x$	$\pi u\log u-(1-\pi+\pi u)$
GAN	$\int p(x)\log\frac{2p(x)}{p(x)+q(x)}+q(x)\log\frac{2q(x)}{p(x)+q(x)}\,\mathrm{d}x-\log(4)$	$u\log u-(u+1)\log(u$

Using the divergence you like ☺

https://arxiv.org/abs/1606.00709

Name	Conjugate $f^*(t)$
Total variation	t
Kullback-Leibler (KL)	$\exp(t-1)$
Reverse KL	$-1-\log(-t)$
Pearson χ^2	$\frac{1}{4}t^2+t$
Neyman χ^2	$2-2\sqrt{1-t}$
Squared Hellinger	$\frac{t}{1-t}$
Jeffrey	$W(e^{1-t})+$ [illegible]
Jensen-Shannon	$-\log(2-\exp(t$

GAN is difficult to train ……

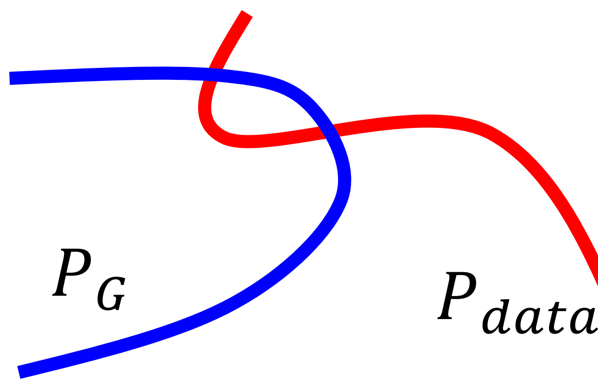
Tips for GAN

JS divergence is not suitable

- In most cases, P_G and P_{data} are not overlapped.
- 1. The nature of data

Both P_{data} and P_G are low-dim manifold in high-dim space.

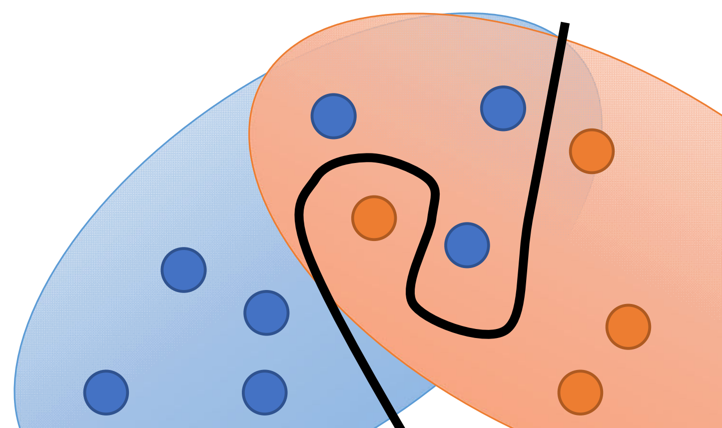
The overlap can be ignored.



- 2. Sampling

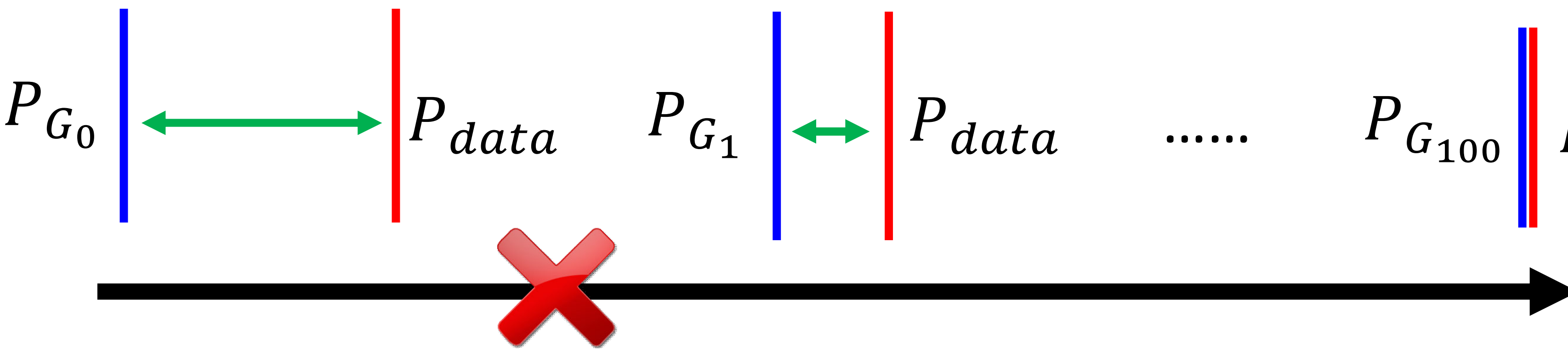
Even though P_{data} and P_G have overlap.

If you do not have enough



What is the problem of JS divergence?

JS divergence is always log2 if two distributions do not o



Equally bad

$$JS(P_{G_0}, P_{data}) = log2$$

$$JS(P_{G_1}, P_{data}) = log2$$

……

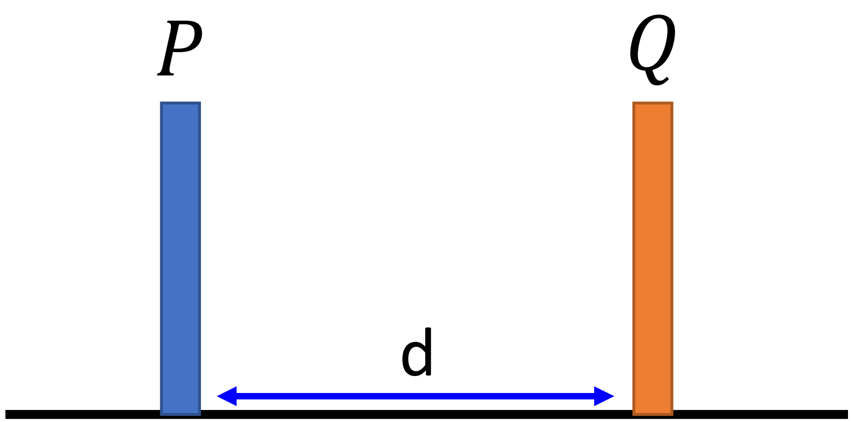
$$JS(P_{G_{100}}, P_{d}$$
$$= 0$$

Intuition: If two distributions do not overlap, binary classif achieves 100% accuracy.

The accuracy (or loss) means nothing during GAN training.

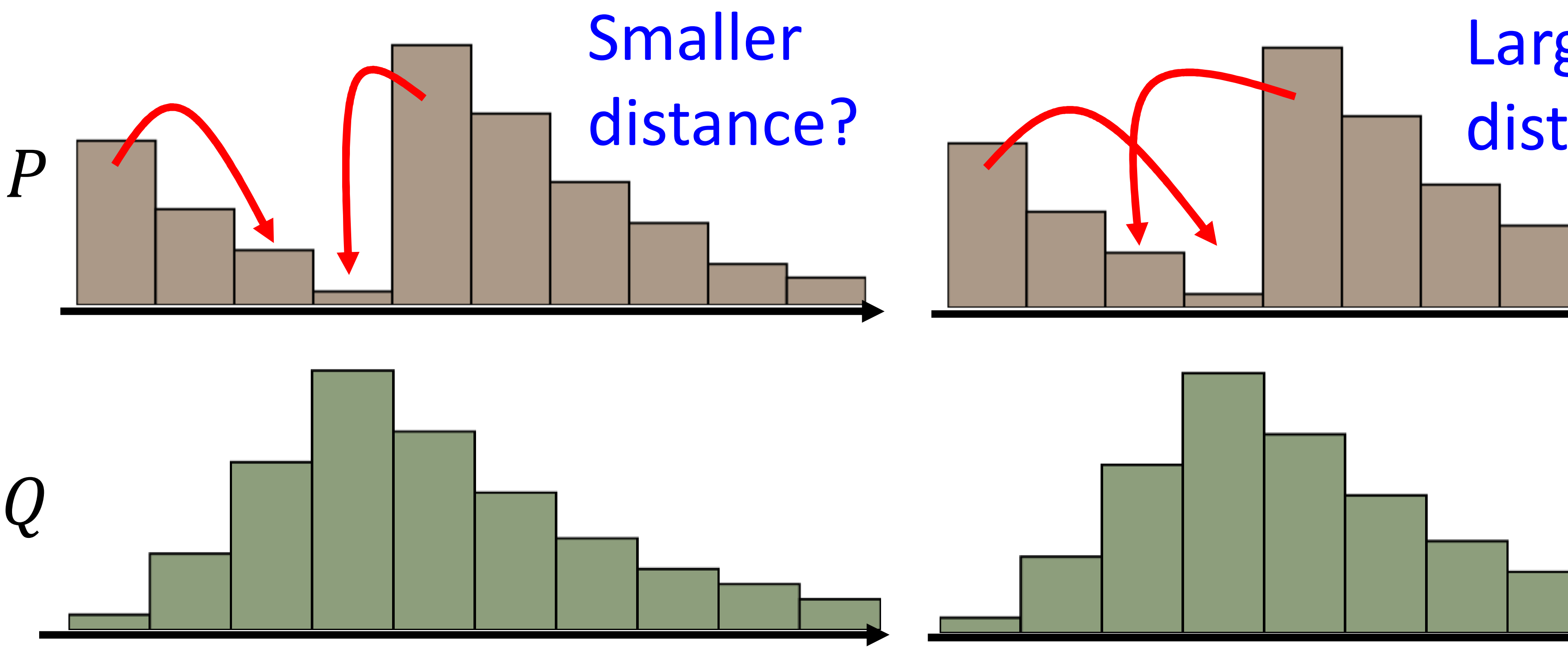
Wasserstein distance

- Considering one distribution P as a pile of earth, and another distribution Q as the target
- The average distance the earth mover has to mov the earth.



$$W(P, Q) = d$$

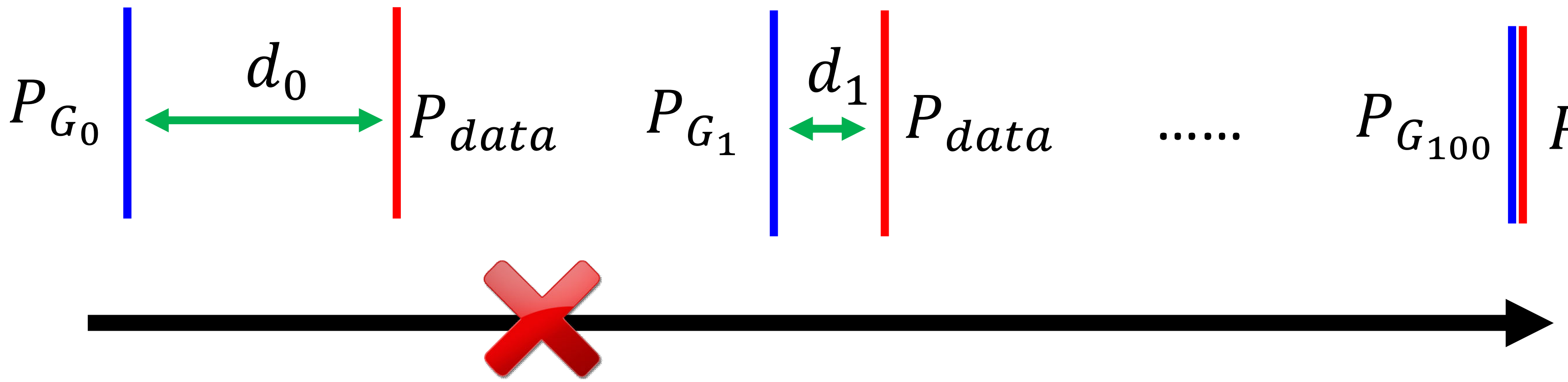
Wasserstein distance



There are many possible “moving plans”.

Using the “moving plan” with the smallest average distance t
define the Wasserstein distance.

What is the problem of JS divergence?



$JS(P_{G_0}, P_{data})$
$= log2$

$JS(P_{G_1}, P_{data})$
$= log2$

......

$JS(P_{G_{100}}, P_{d}$
$= 0$

$W(P_{G_0}, P_{data})$
$= d_0$

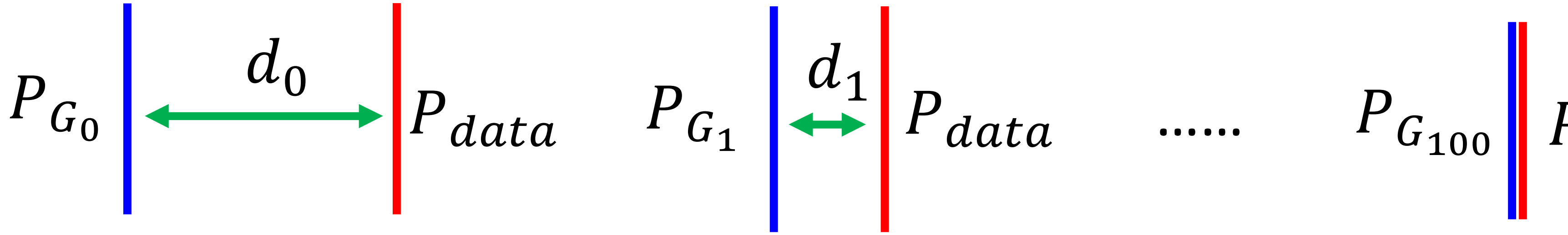
$W(P_{G_1}, P_{data})$
$= d_1$

......

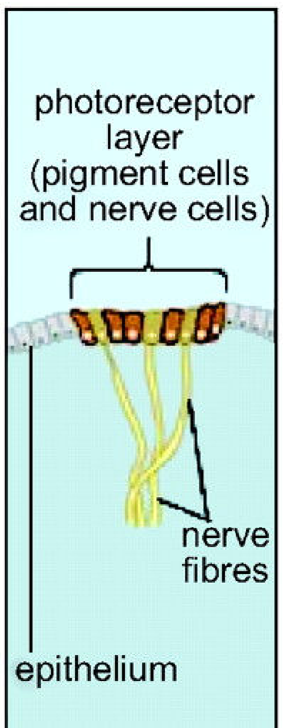
$W(P_{G_{100}}, P_{d}$
$= 0$

Better!

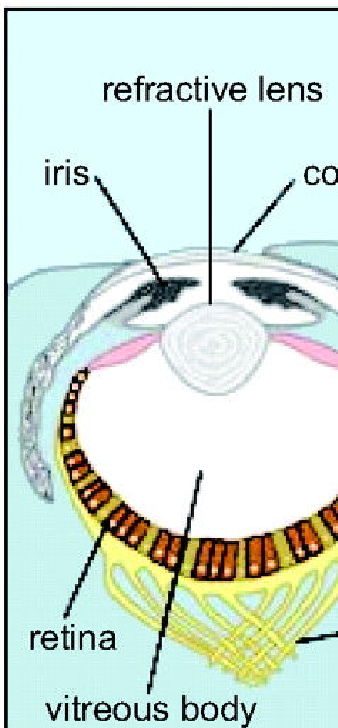
What is the problem of JS divergence?



pigment spot
(limpet, *Patella*)



Complex eye
(octopus)



https://arxiv.org/abs/170

WGAN

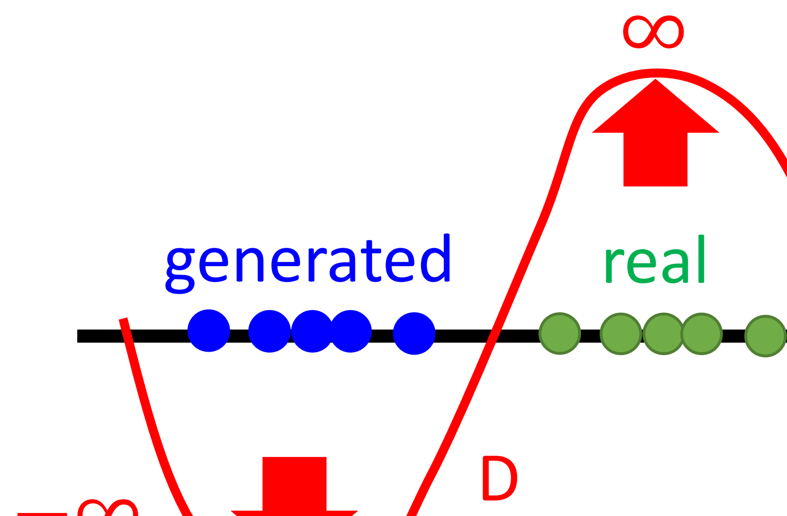
Evaluate Wasserstein distance between P_{data} and

$$\max_{\underline{D\in 1-Lipschitz}} \{E_{y\sim P_{data}}[D(y)] - E_{y\sim P_G}[D(y)]\}$$

D has to be smooth enough. How to fulfill this constr

Without the constraint, the training of D will not converge.

Keeping the D smooth forces D(y) become ∞ and $-\infty$



$$\max_{\boxed{D\in 1-Lipschitz}}\{E_{y\sim P_{data}}[D(y)] - E_{y\sim P_G}[D(y)]\}$$

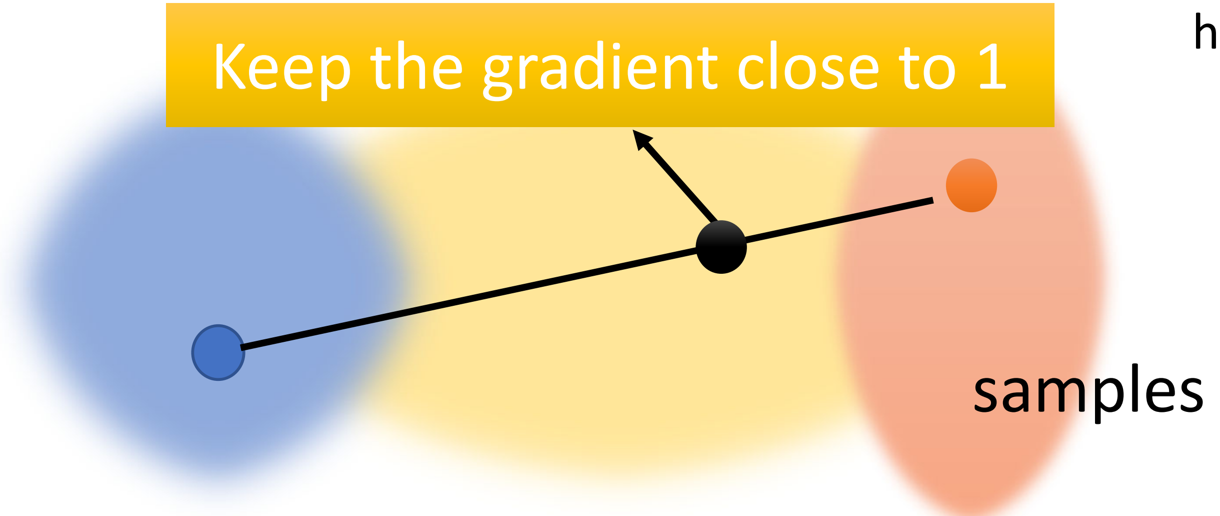
- Original WGAN → Weight

  Force the parameters w between c and -c

  After parameter update, if w > c, w = c; if w < -c, w

- Improved WGAN → Gradient Penalty

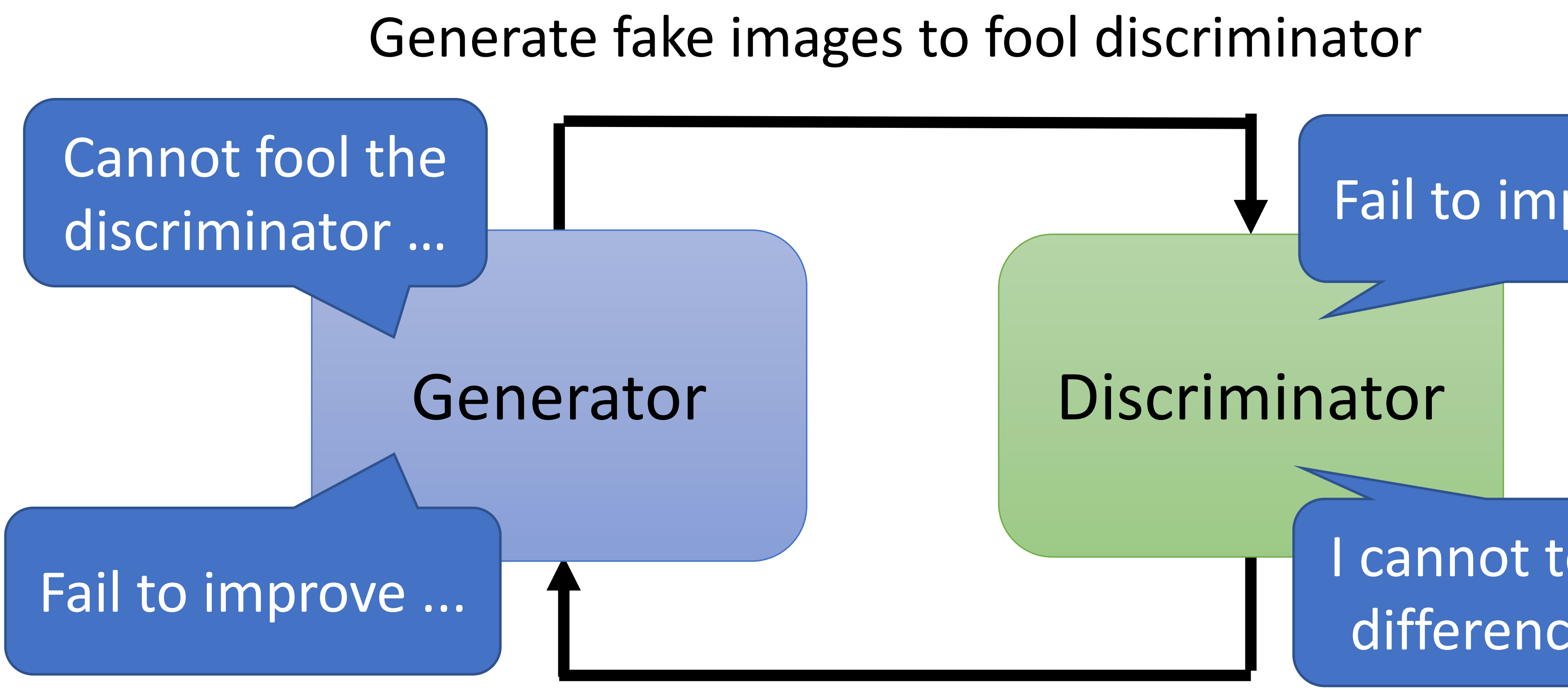
https://arxiv.org/abs/1704



- Spectral Normalization → Keep gradient norm smaller than 1 everywhere

GAN is still challenging …

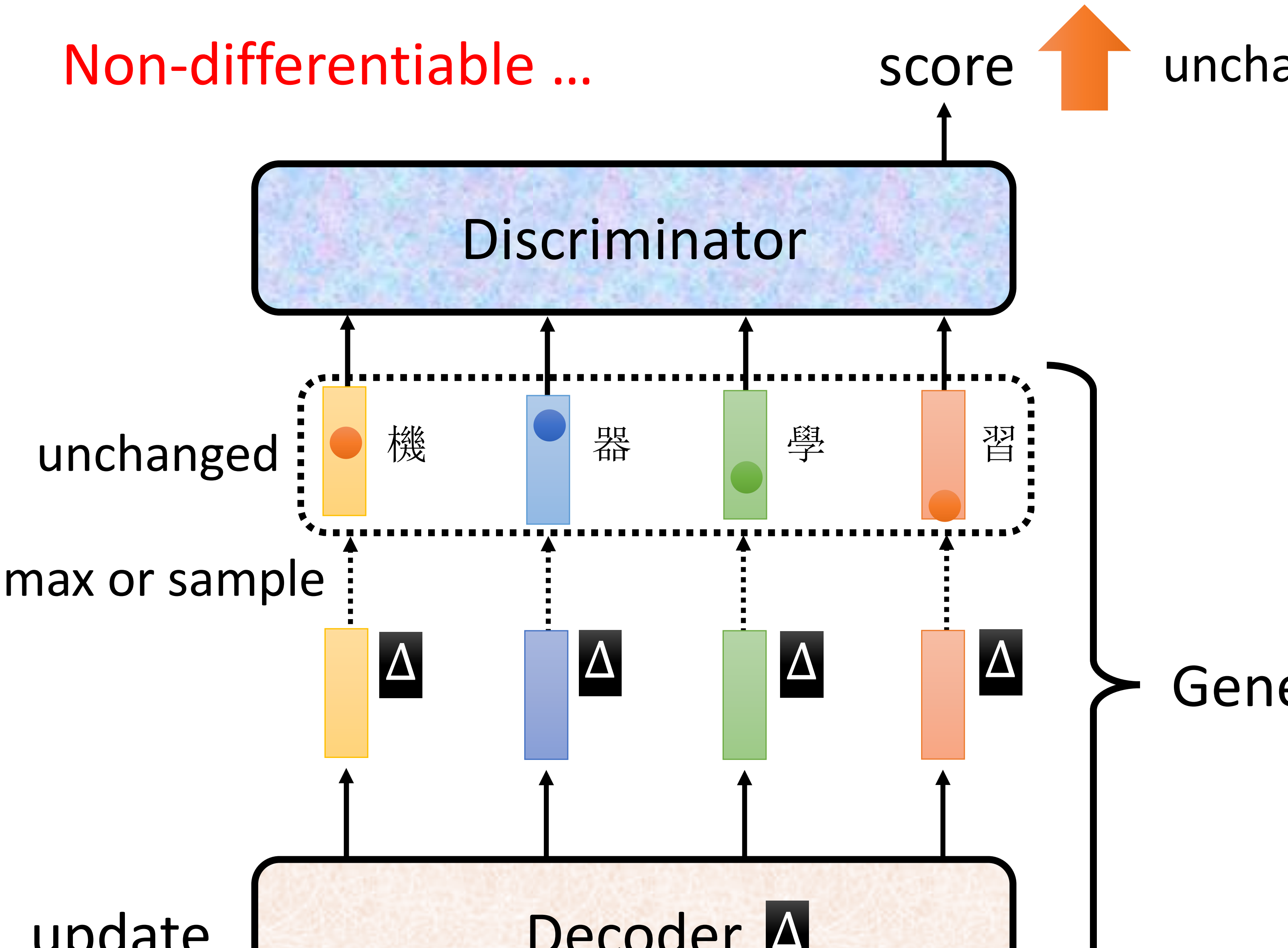
- Generator and Discriminator needs to match each other (棋逢敵手)



More Tips

- Tips from Soumith
  - https://github.com/soumith/ganhacks
- Tips in DCGAN: Guideline for network architecture design for image generation
  - https://arxiv.org/abs/1511.06434
- Improved techniques for training GANs
  - https://arxiv.org/abs/1606.03498
- Tips from BigGAN
  - https://arxiv.org/abs/1809.11096

GAN for Sequence Generation

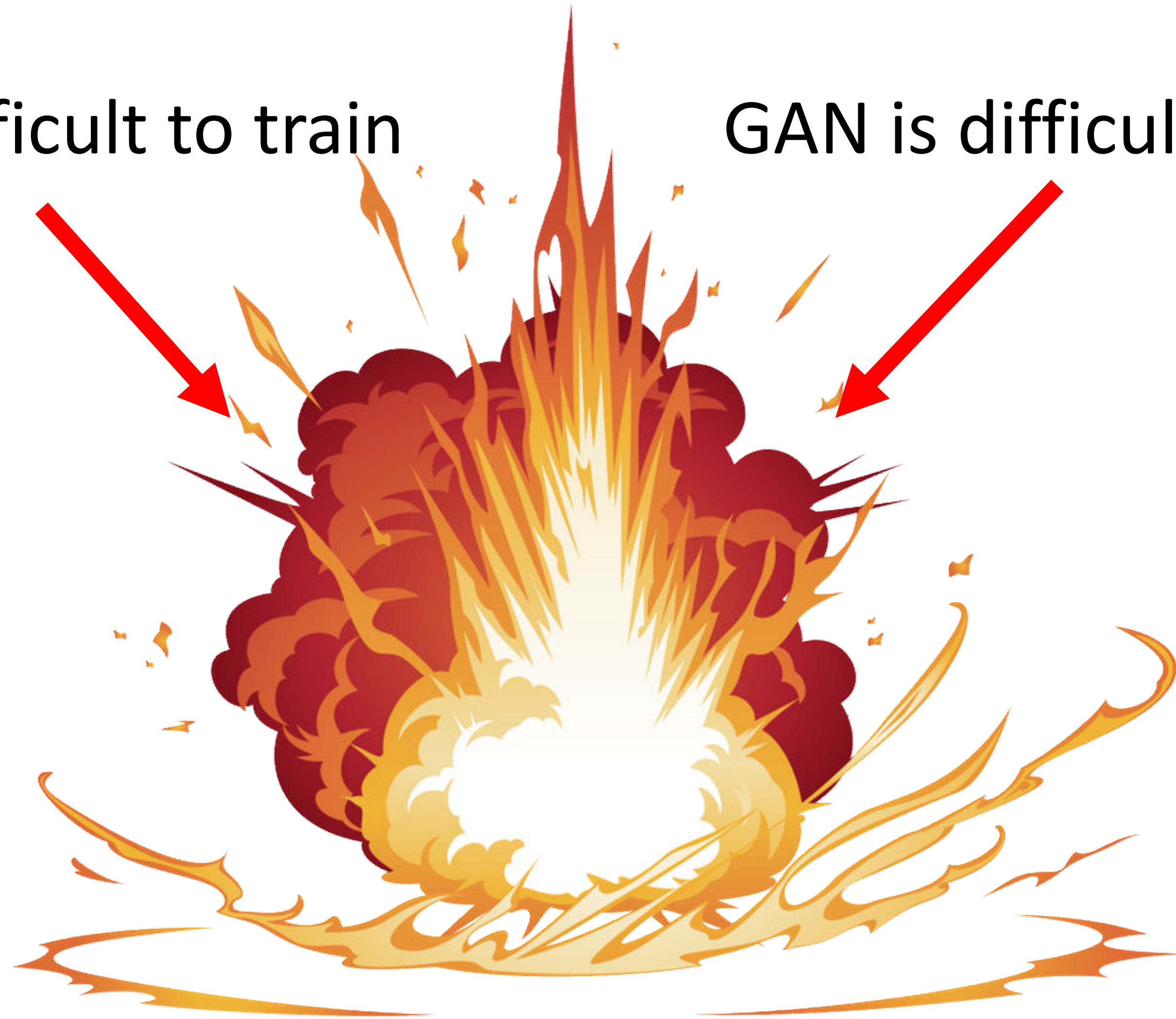


GAN for Sequence Generation

Reinforcement learning (RL) is involved ..

RL is difficult to train

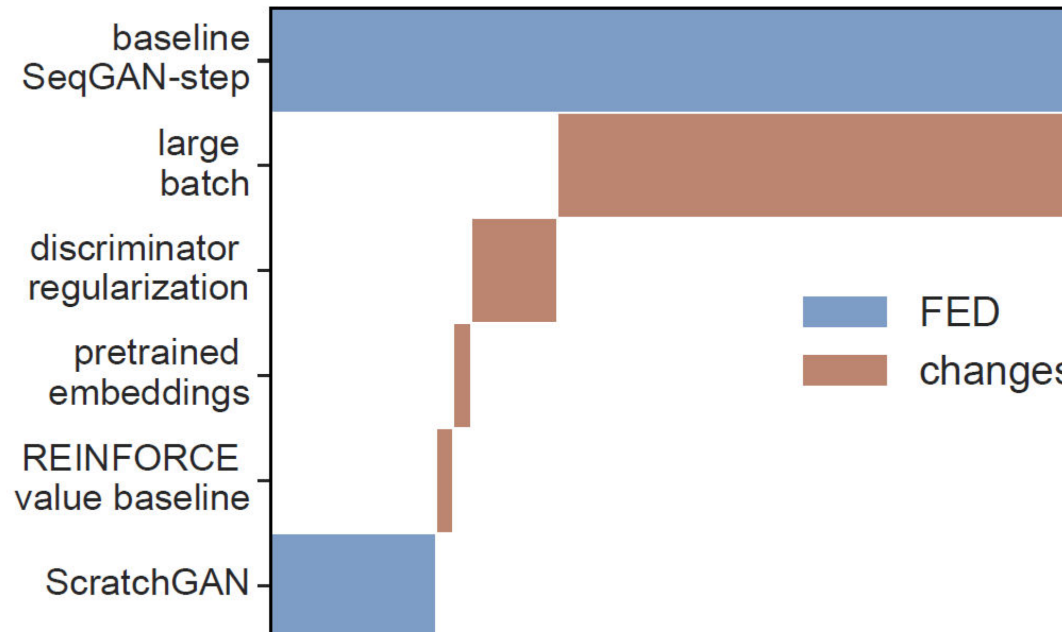
GAN is difficult to tr



Sequence Generation GAN (RL+GAN)

GAN for Sequence Generation

- Usually, the generator are fine-tuned from a model learne by other approaches.
- However, with enough hyperparameter-tuning and tips, ScarchGAN can train from scratch.



Training language GANs from Scratch

https://arxiv.org/abs/

Generative Models

- This lecture: Generative Adversarial Network (GA

Full version

https://www.youtube.com/playlist?list=PLJV_el3uVTsMq

More Generative Models

Variational Autoencoder (VAE)

https://youtu.be/8zomhgKrsmQ

FLOW-based Model

https://youtu.be/uXY18nzdSs

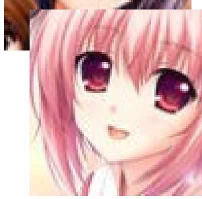
Possible Solution?

$$\begin{bmatrix} 0.3 \\ -0.1 \\ \vdots \\ -0.7 \end{bmatrix}$$

$$\begin{bmatrix} 0.1 \\ -0.1 \\ \vdots \\ 0.7 \end{bmatrix}$$

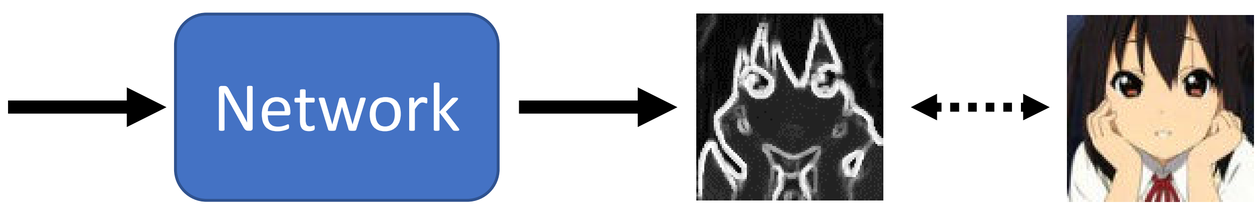
$$\begin{bmatrix} -0.3 \\ 0.1 \\ \vdots \\ 0.9 \end{bmatrix}$$

$$\begin{bmatrix} 0.7 \\ 0.1 \\ \vdots \\ -0.9 \end{bmatrix}$$



$$\begin{bmatrix} -0.1 \\ 0.8 \\ \vdots \\ 0.8 \end{bmatrix}$$

$$\begin{bmatrix} 0.3 \\ -0.1 \\ \vdots \\ -0.7 \end{bmatrix}$$

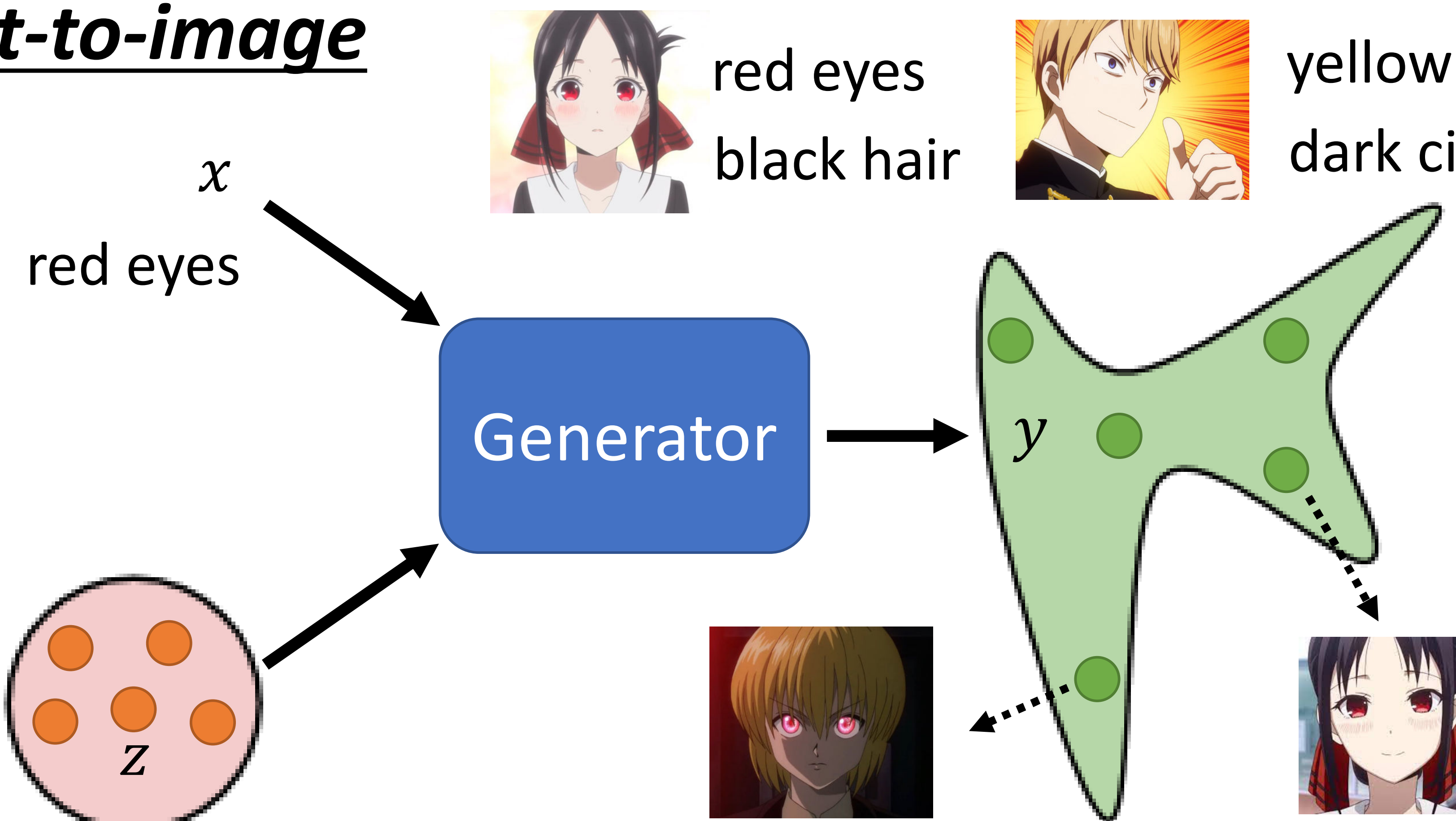


Using typical learning approach

Generative Latent Optimization (GLO), https://arxiv.org/abs/1707.0

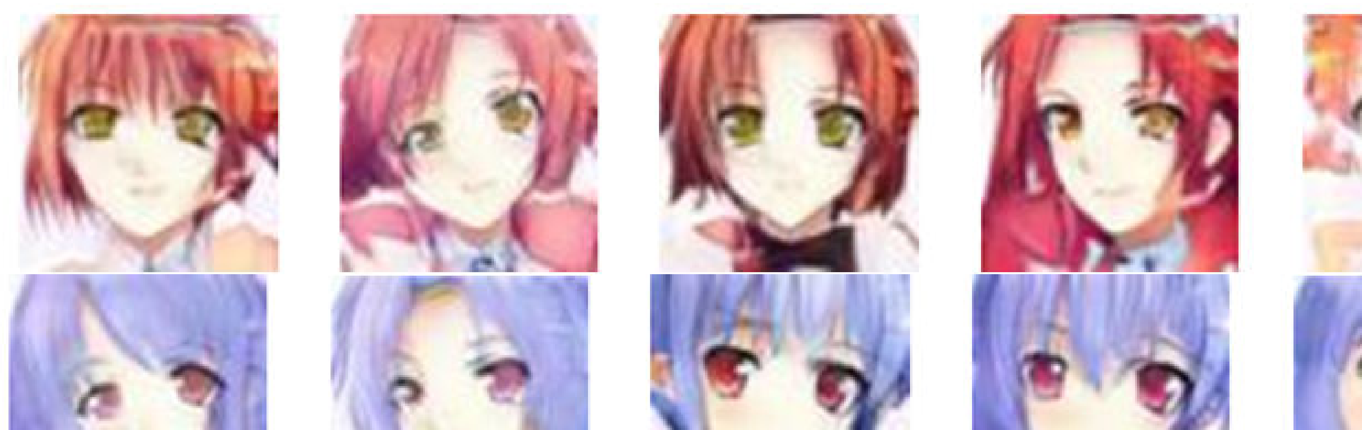
Conditional Generation

Text-to-image



red hair,
green eyes

blue hair,



Conditional GAN

x: Red eyes

Normal distribution z

G

Image

$y = G($

y is real image or not

y

D
(original)

scalar

Generator will learn to generate realistic image

But completely ignore th input conditions.

Real images:  1

Generated images: Image 0

https://arxiv.org/abs/16

Conditional GAN

x: Red eyes → G

Normal distribution z → G

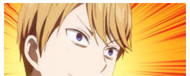
G → Image $y = G($

y → D (better)

x → D (better)

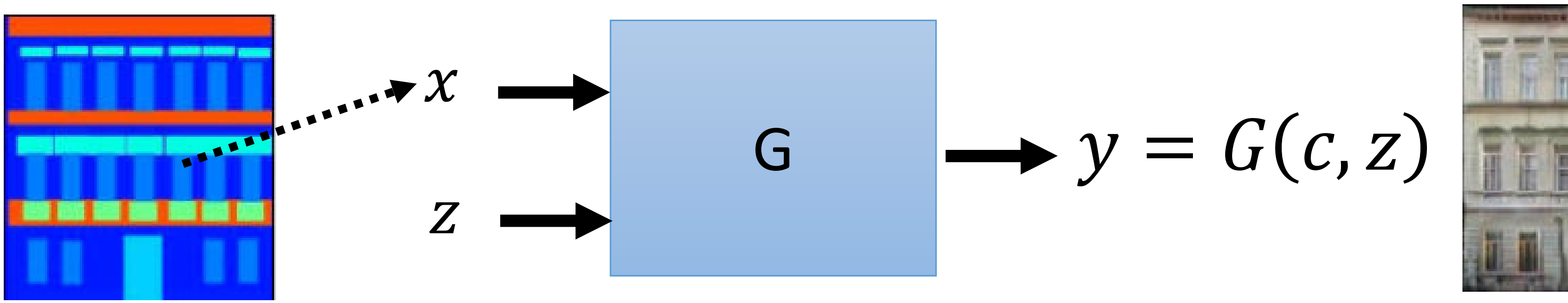
D (better) → scalar ⇢ y is realistic or not + x and y are matched

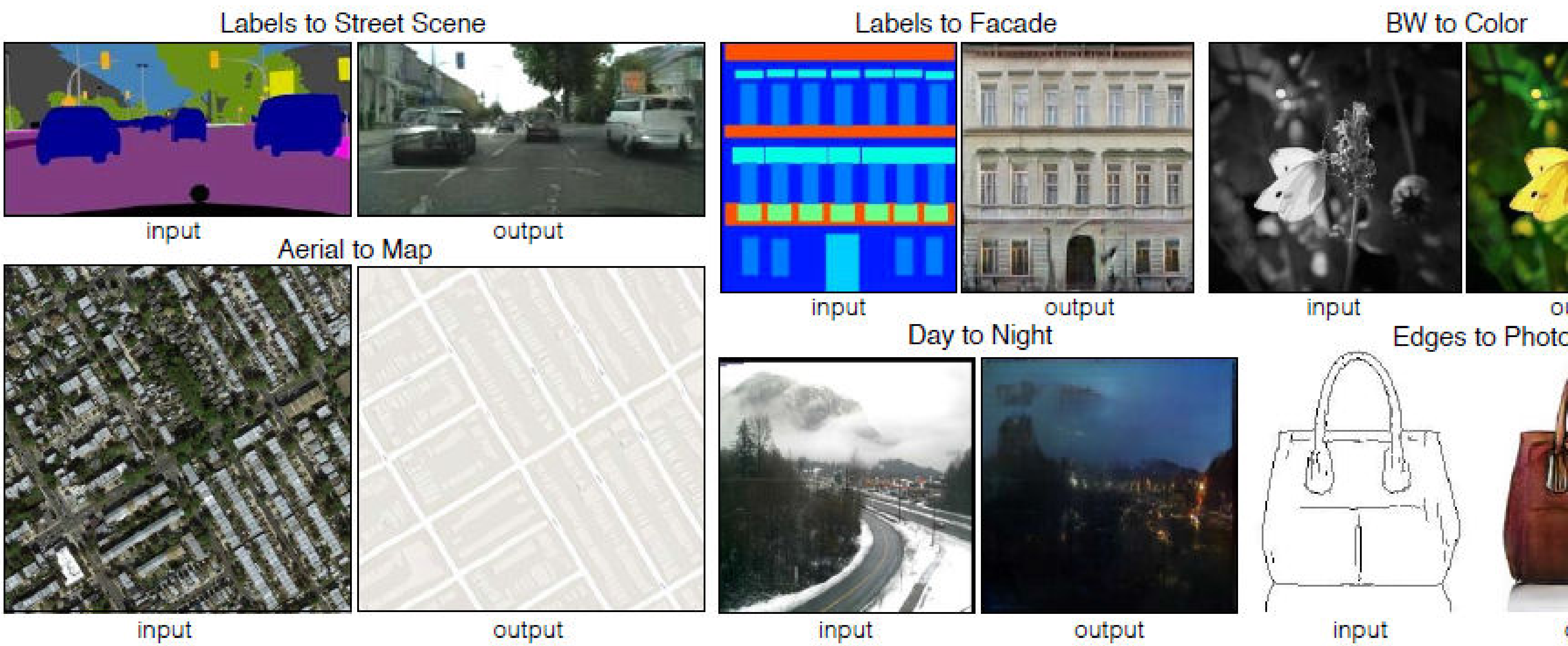
True text-image pairs: (red eyes, ) 1

(red eyes, ) 0 (red eyes, Image) 0

https://arxiv.org/abs/1

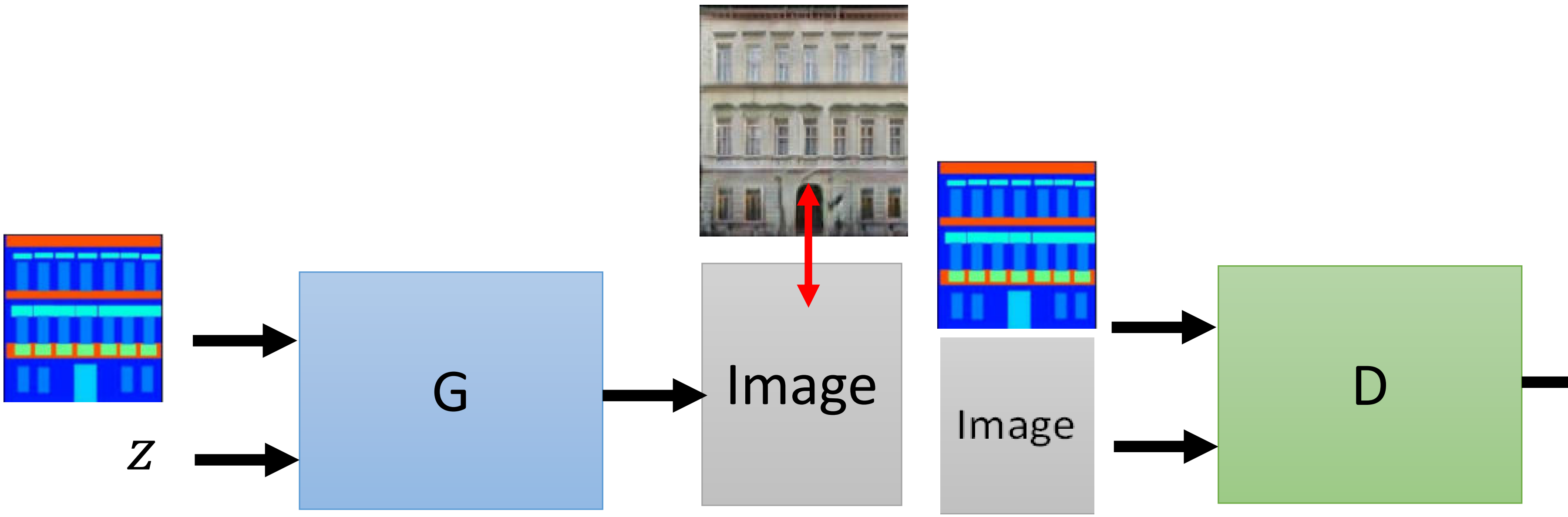
Conditional GAN



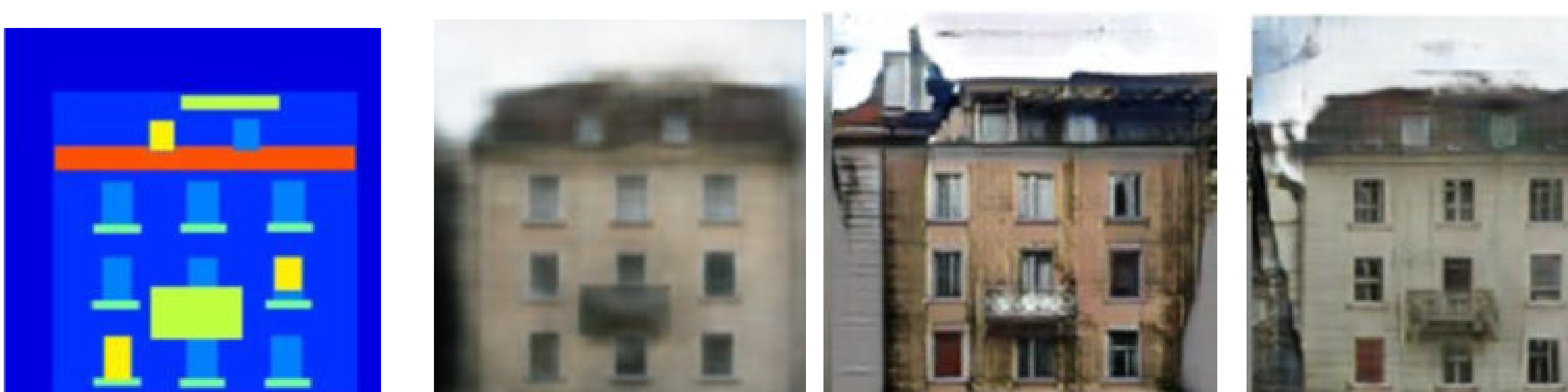


https://arxiv.org/abs/1

Conditional GAN

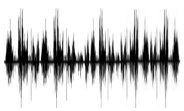


Testing:

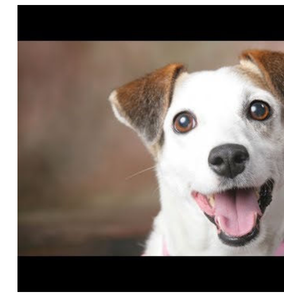


https://arxiv.org/abs/180

Conditional GAN



x: sound → G → Image



"a dog barking sound"

Training Data Collection

Conditional GAN

The images are generate
Chia-Hung Wan and Sh
Re Chuang

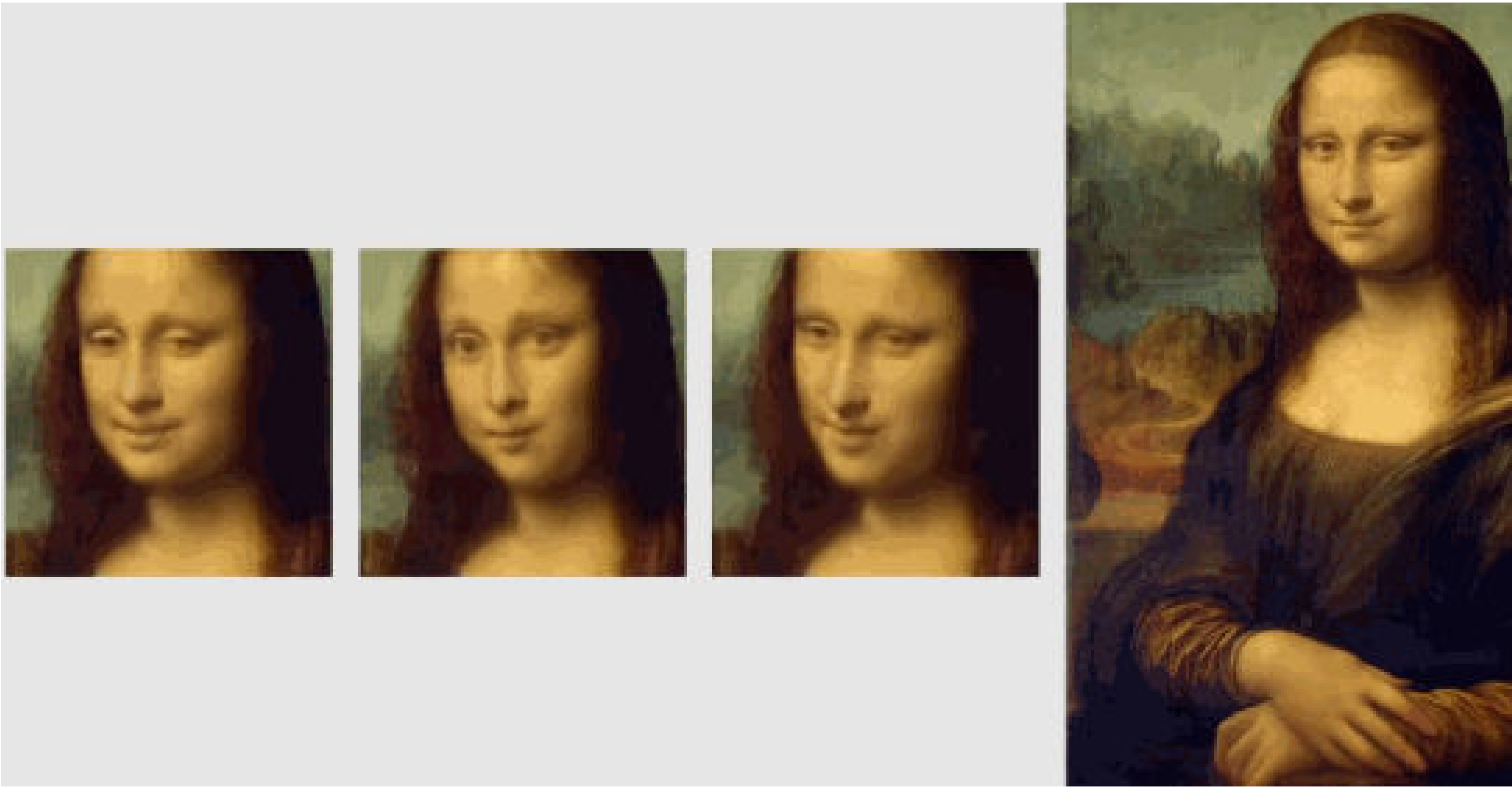
https://wjohn1483.github.io/audio_to_scene/index.html

- Sound-to-image

Louder

Conditional GAN

Talking Head Generation



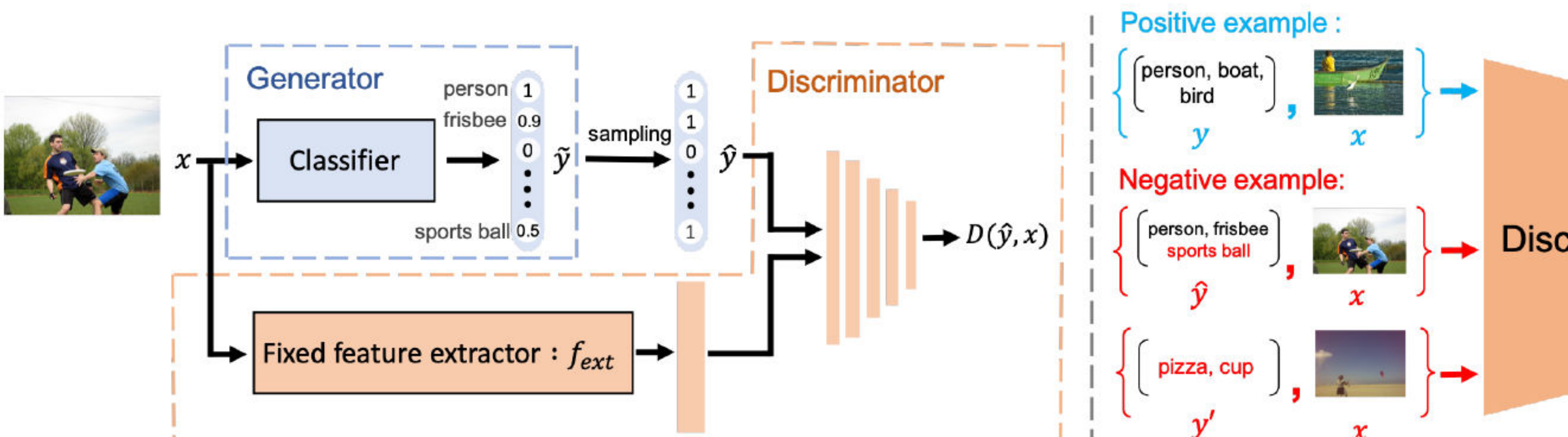
Conditional GAN

Multi-label Image Classifier = Conditional Generato

Input condition

person, sports ball,
baseball bat, baseball glove

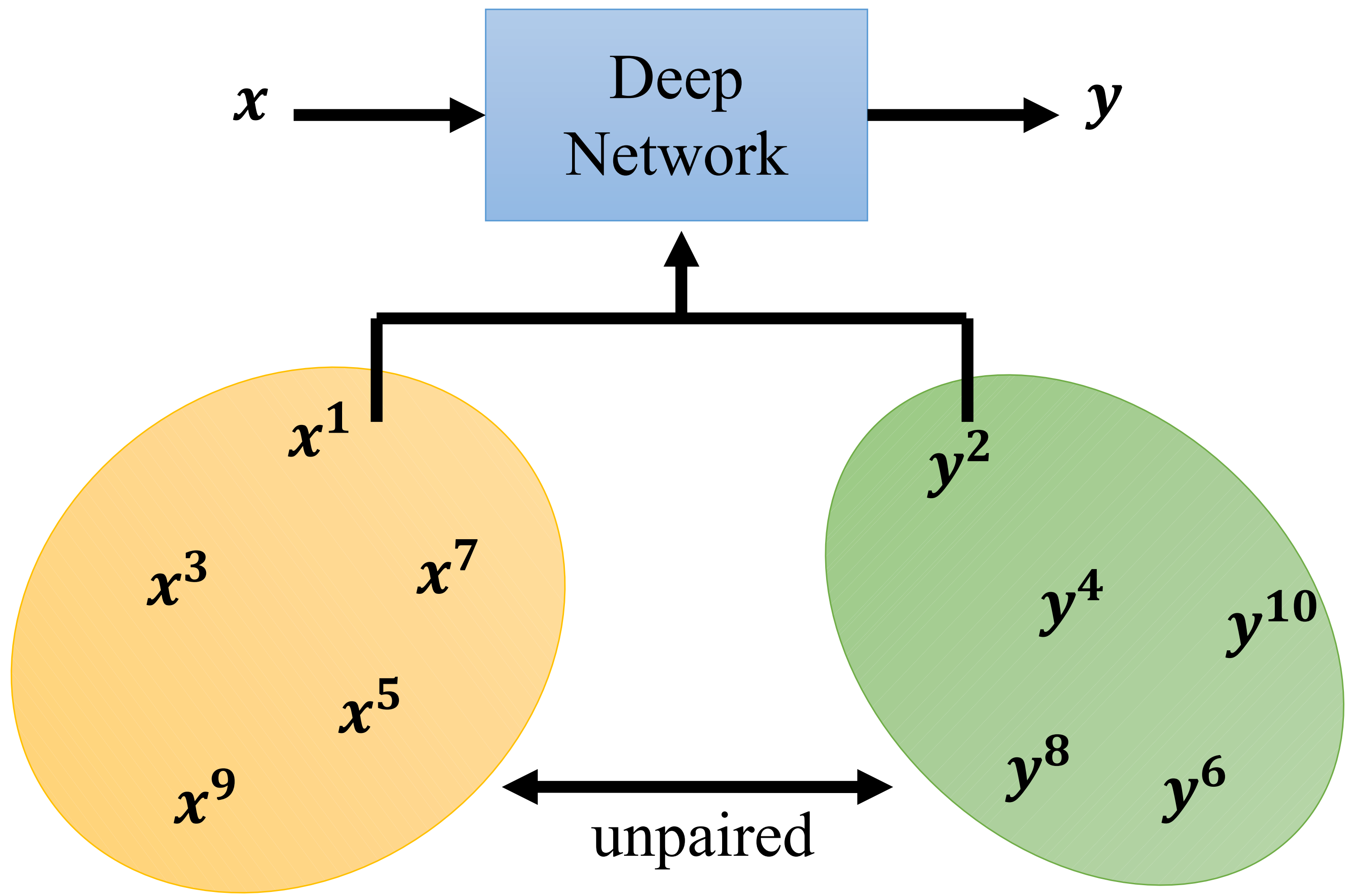
Generated output




Q&A

Learning from Unpaired Data

Learning from Unpaired Data

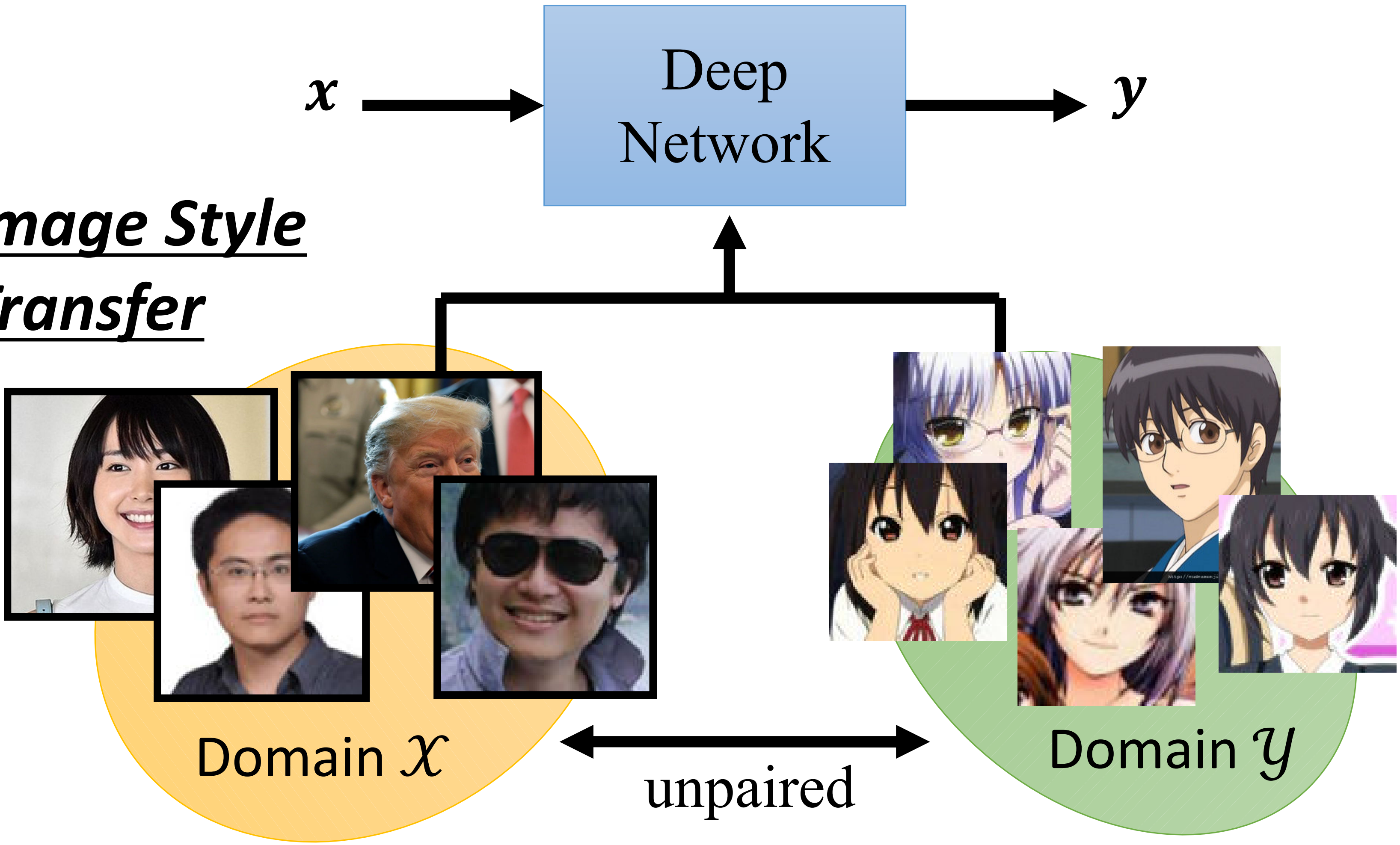


HW3: pseudo labeling
HW5: back translation

Still need **some** paired data

Learning from Unpaired Data

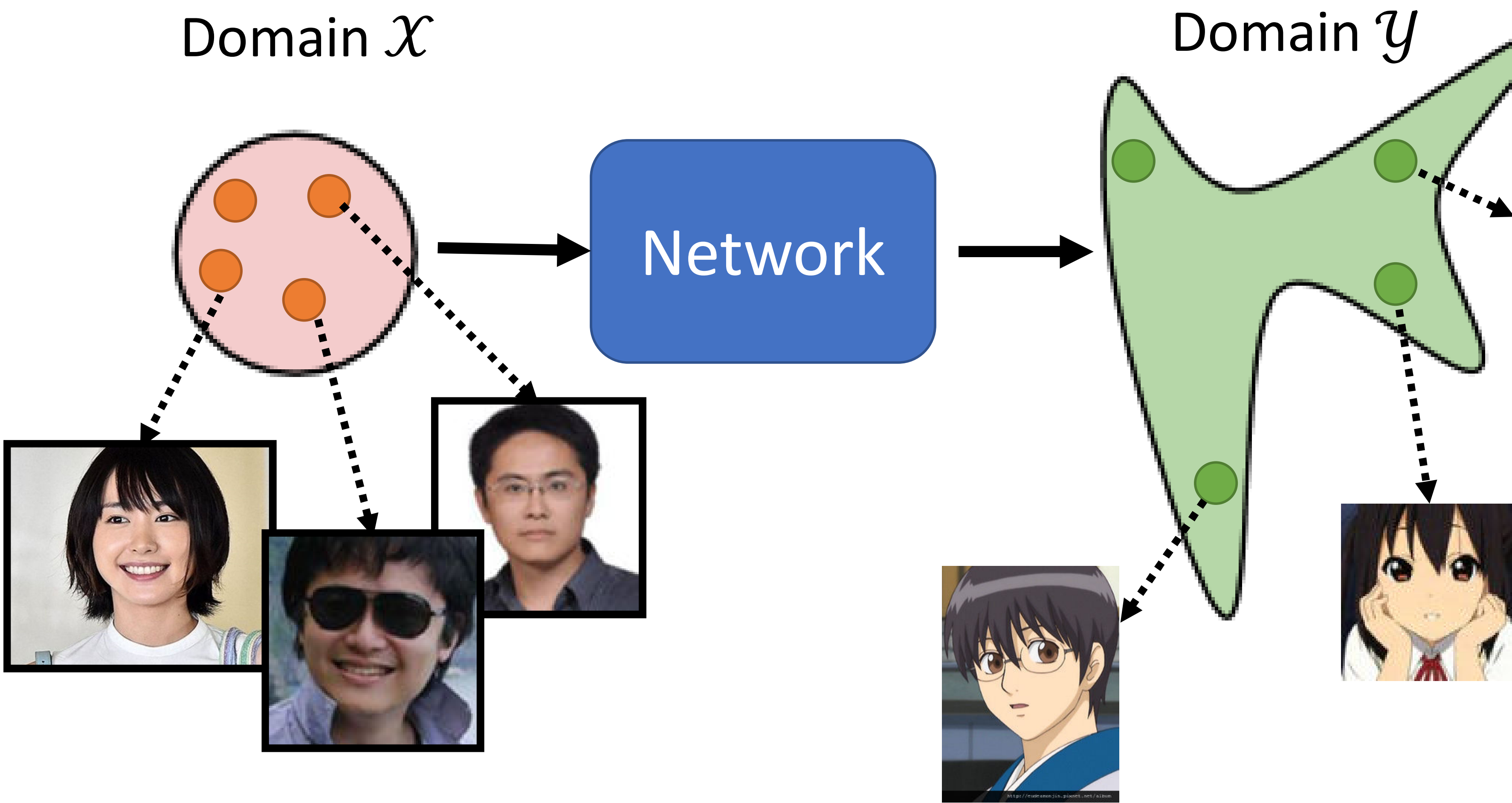
Image Style Transfer



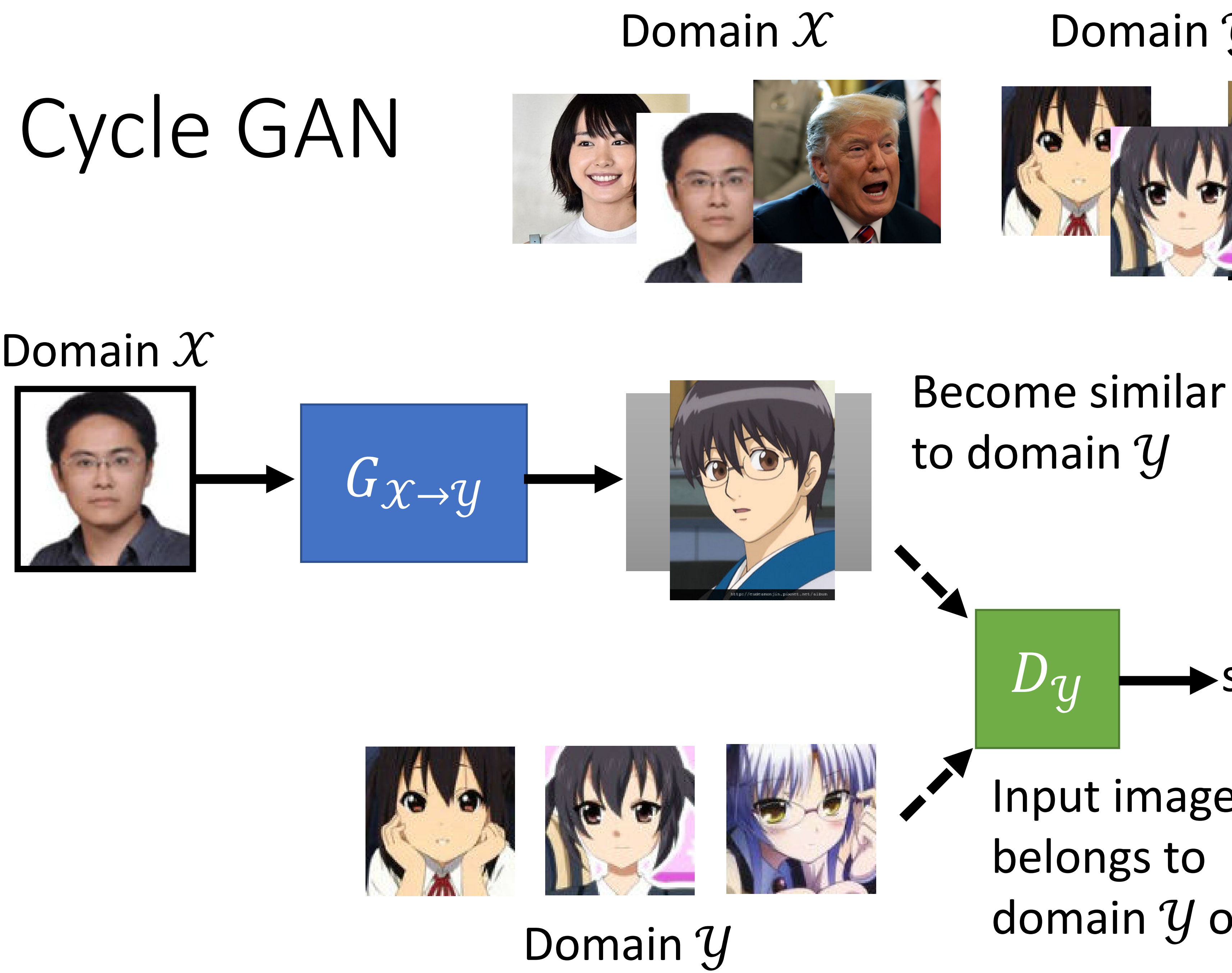
Can we learn the mapping without any paired data?

Unsupervised Conditional Gene

Learning from Unpaired Data

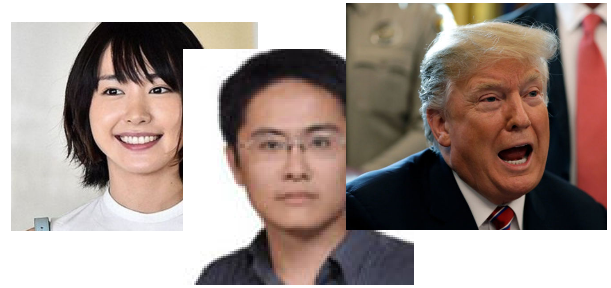


Cycle GAN


Domain $\mathcal{X}$
Domain
Domain $\mathcal{X}$
$G_{\mathcal{X}\rightarrow\mathcal{Y}}$
Become similar
to domain $\mathcal{Y}$
$D_{\mathcal{Y}}$
Input image
belongs to
domain $\mathcal{Y}$
Domain $\mathcal{Y}$

Cycle GAN

Domain $\mathcal{X}$



Domain

Domain $\mathcal{X}$

ignore input

$G_{\mathcal{X}\rightarrow\mathcal{Y}}$

Become similar

to domain $\mathcal{Y}$

$D_{\mathcal{Y}}$

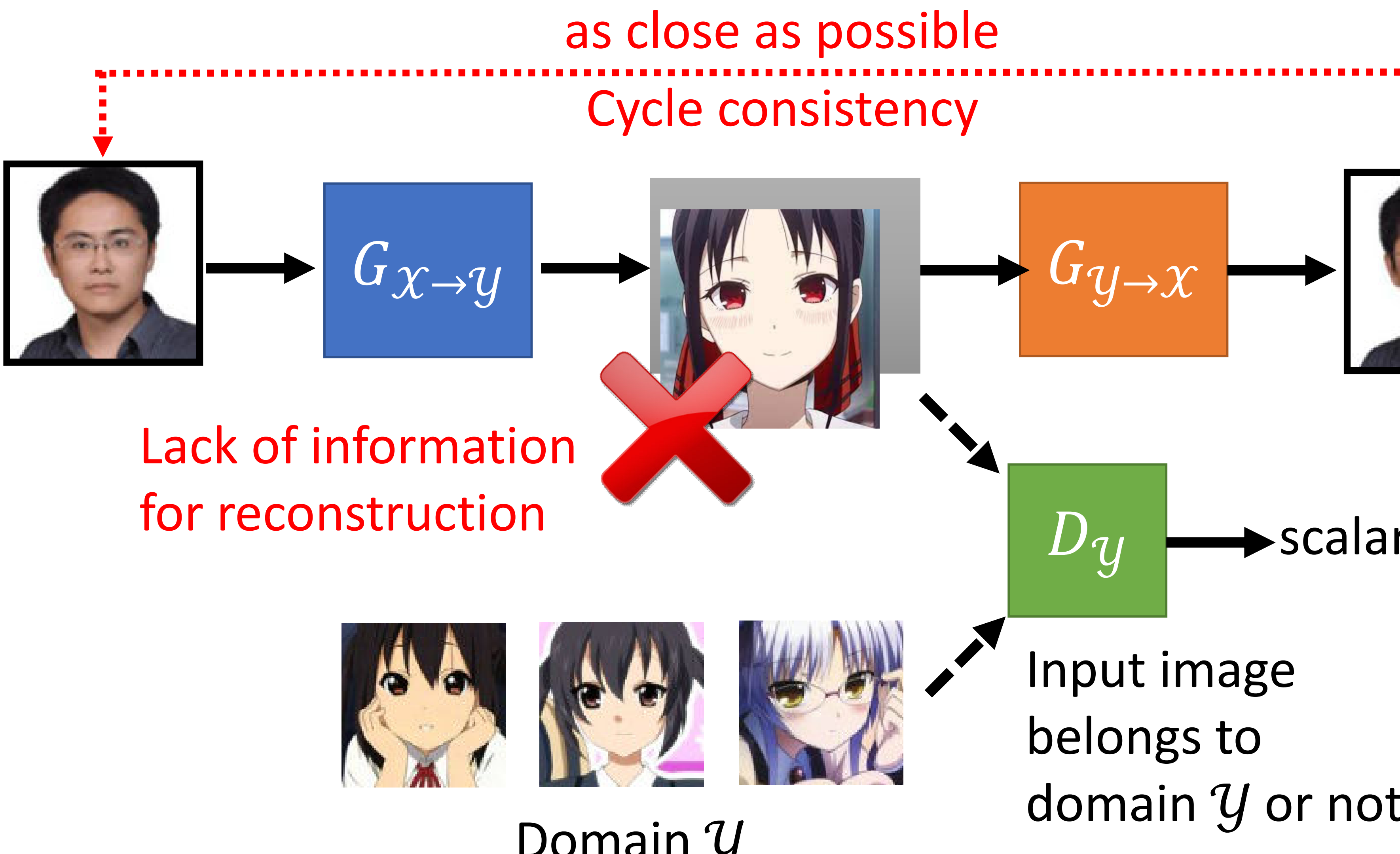
Domain $\mathcal{Y}$

Input image

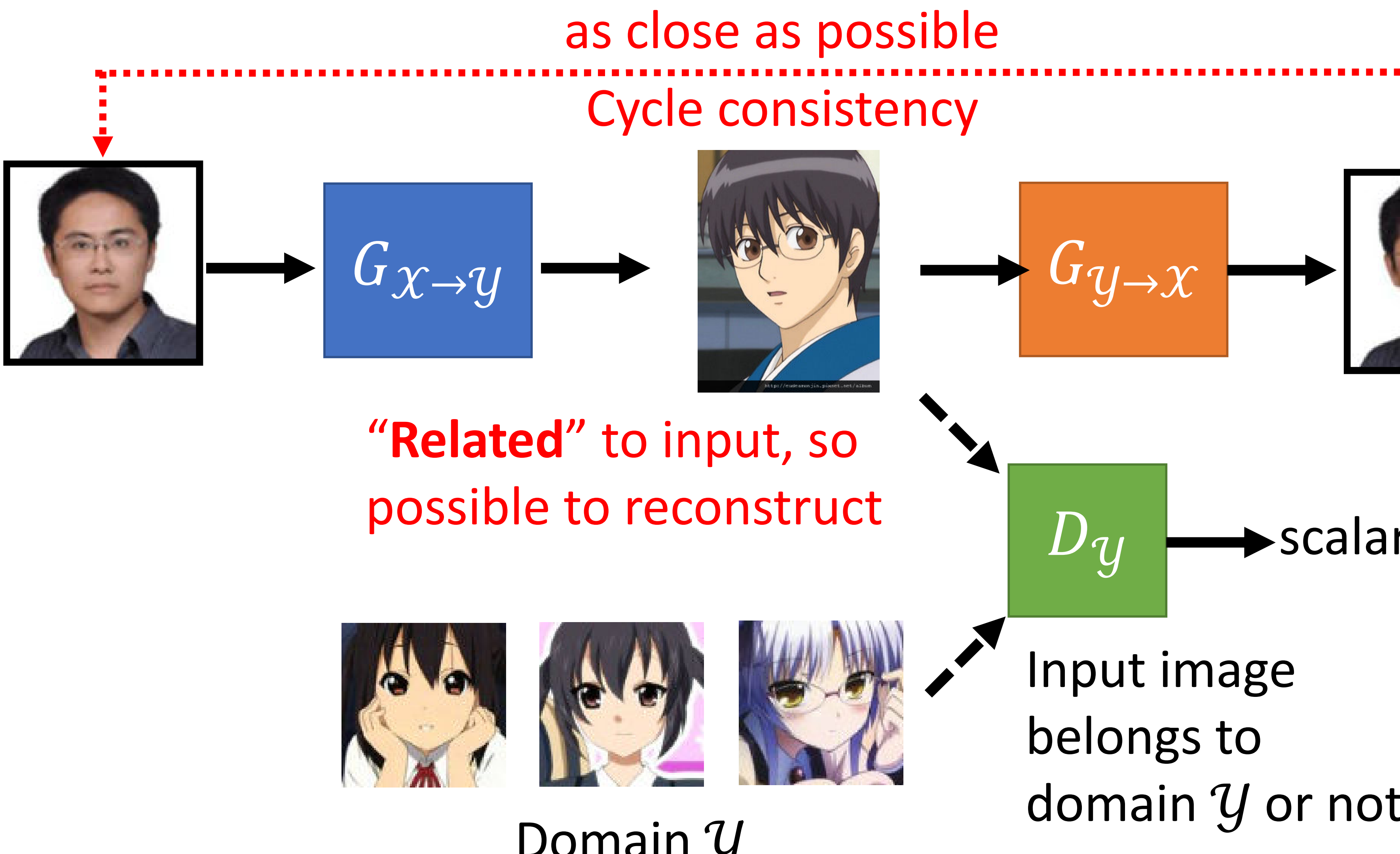
belongs to

domain $\mathcal{Y}$ o

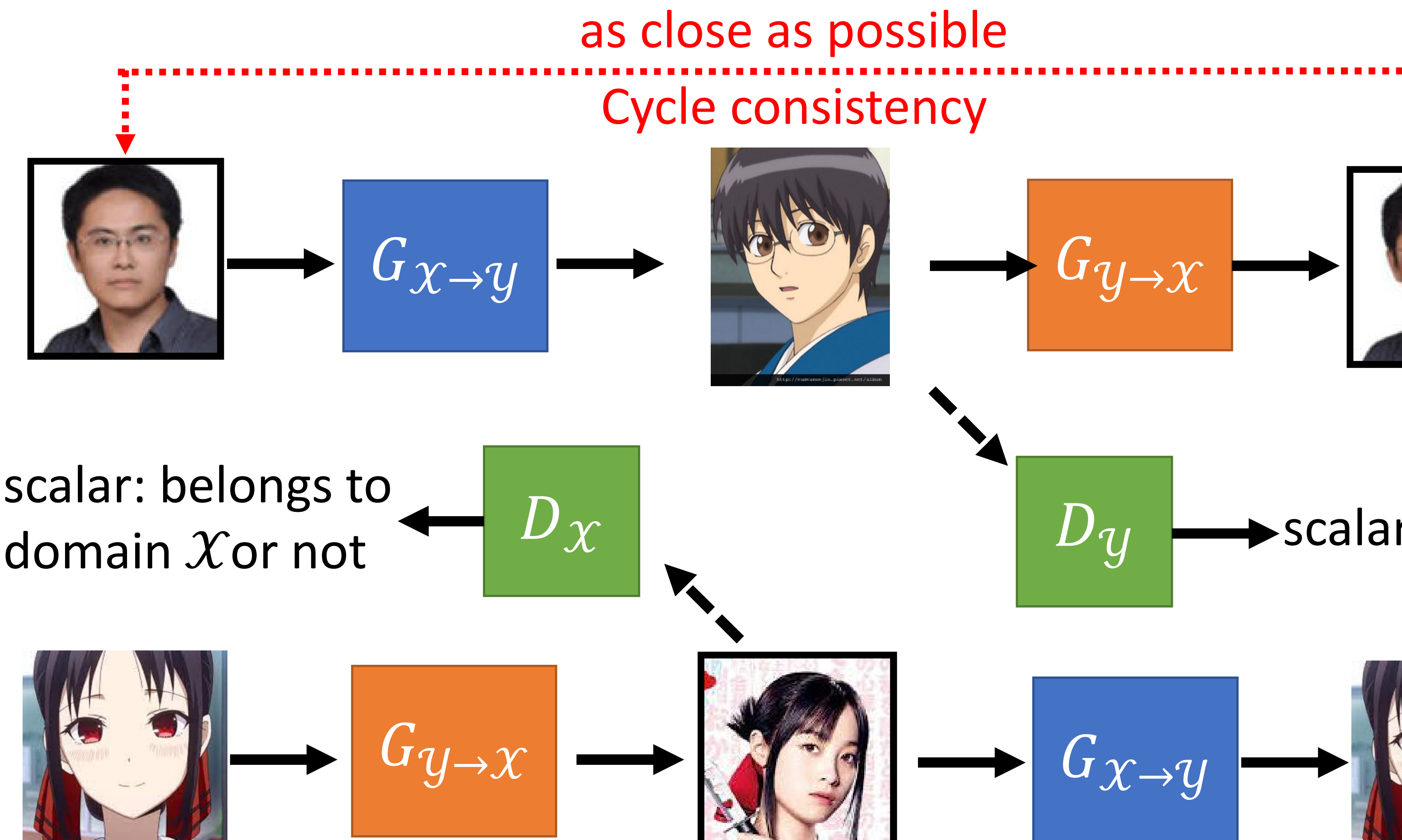
Cycle GAN


as close as possible
Cycle consistency
$G_{\mathcal{X}\to\mathcal{Y}}$
$G_{\mathcal{Y}\to\mathcal{X}}$
Lack of information
for reconstruction
$D_{\mathcal{Y}}$
scalar
Input image
belongs to
domain $\mathcal{Y}$ or not
Domain $\mathcal{Y}$

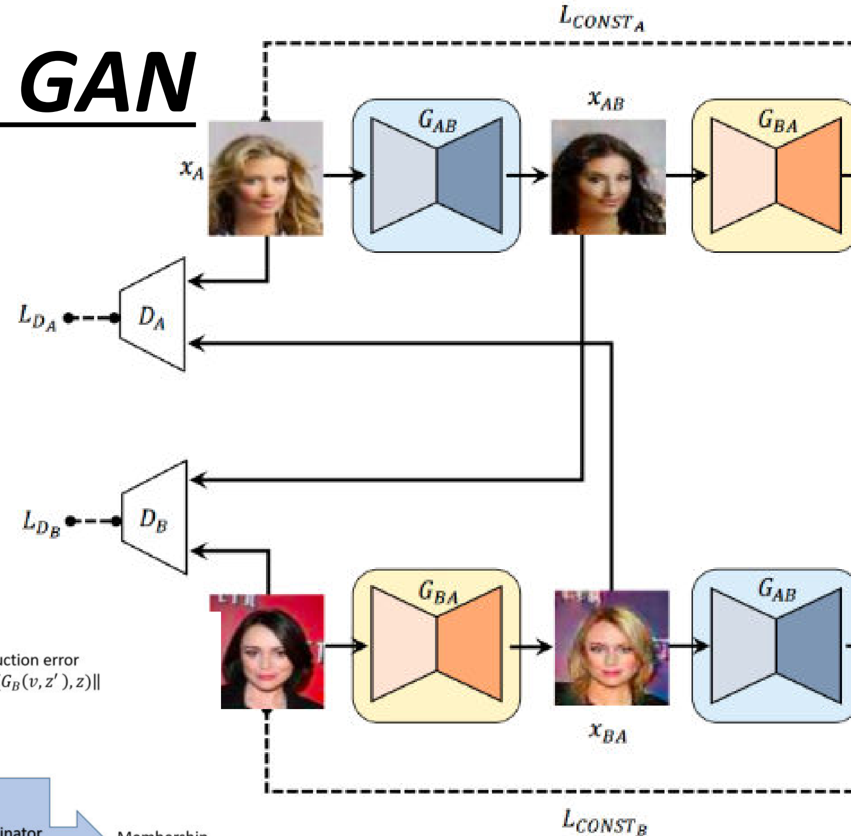
Cycle GAN


as close as possible
Cycle consistency
$G_{\mathcal{X}\to\mathcal{Y}}$
$G_{\mathcal{Y}\to\mathcal{X}}$
“**Related**” to input, so possible to reconstruct
$D_{\mathcal{Y}}$
scalar
Input image belongs to domain $\mathcal{Y}$ or not
Domain $\mathcal{Y}$

Cycle GAN


as close as possible
Cycle consistency
$G_{\mathcal{X}\to\mathcal{Y}}$
$G_{\mathcal{Y}\to\mathcal{X}}$
scalar: belongs to domain $\mathcal{X}$ or not
$D_{\mathcal{X}}$
$D_{\mathcal{Y}}$
scalar
$G_{\mathcal{Y}\to\mathcal{X}}$
$G_{\mathcal{X}\to\mathcal{Y}}$

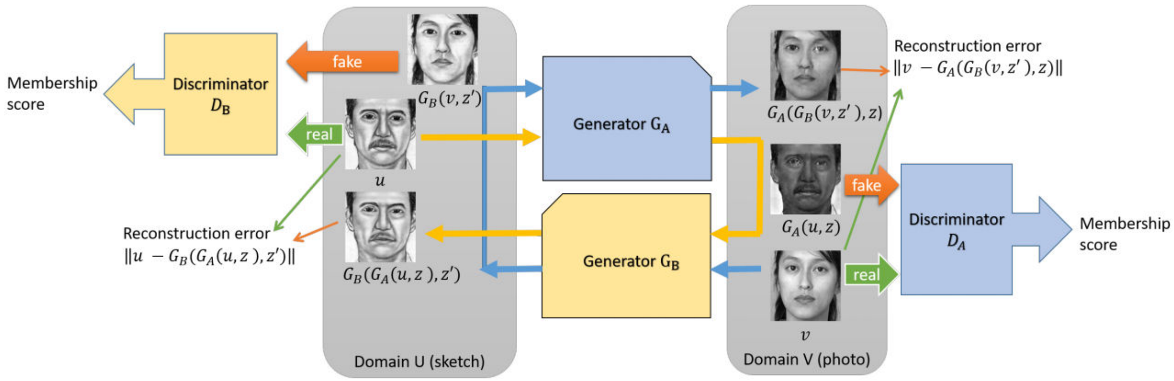
Disco GAN



https://arxiv.org/abs/1703.05192

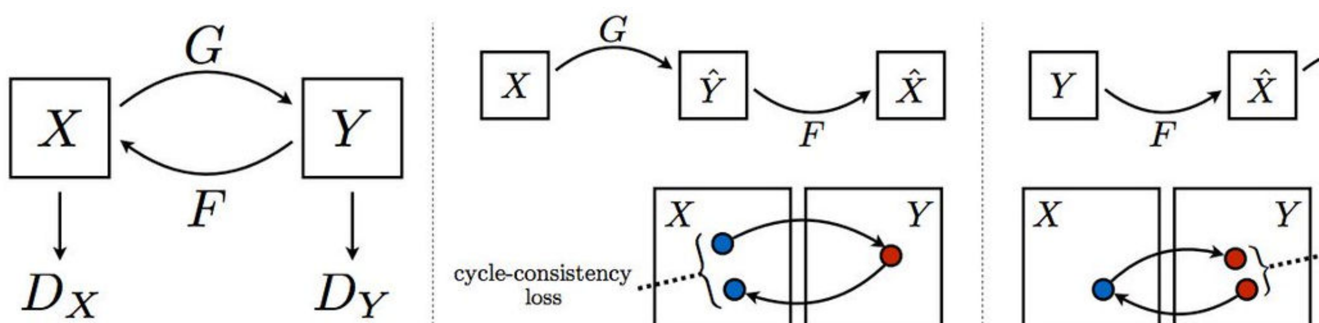
Dual GAN

https://arxiv.org/abs/1704.02510



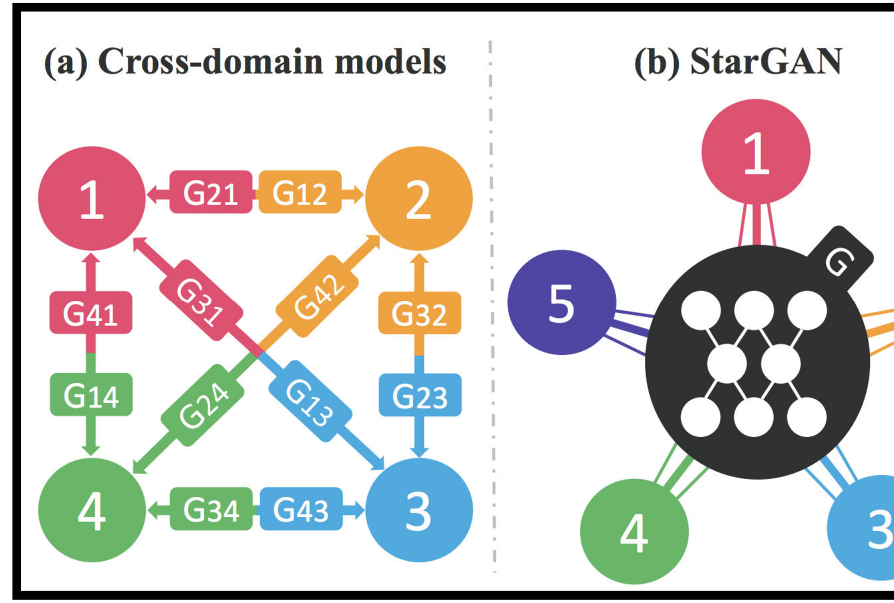
Cycle GAN

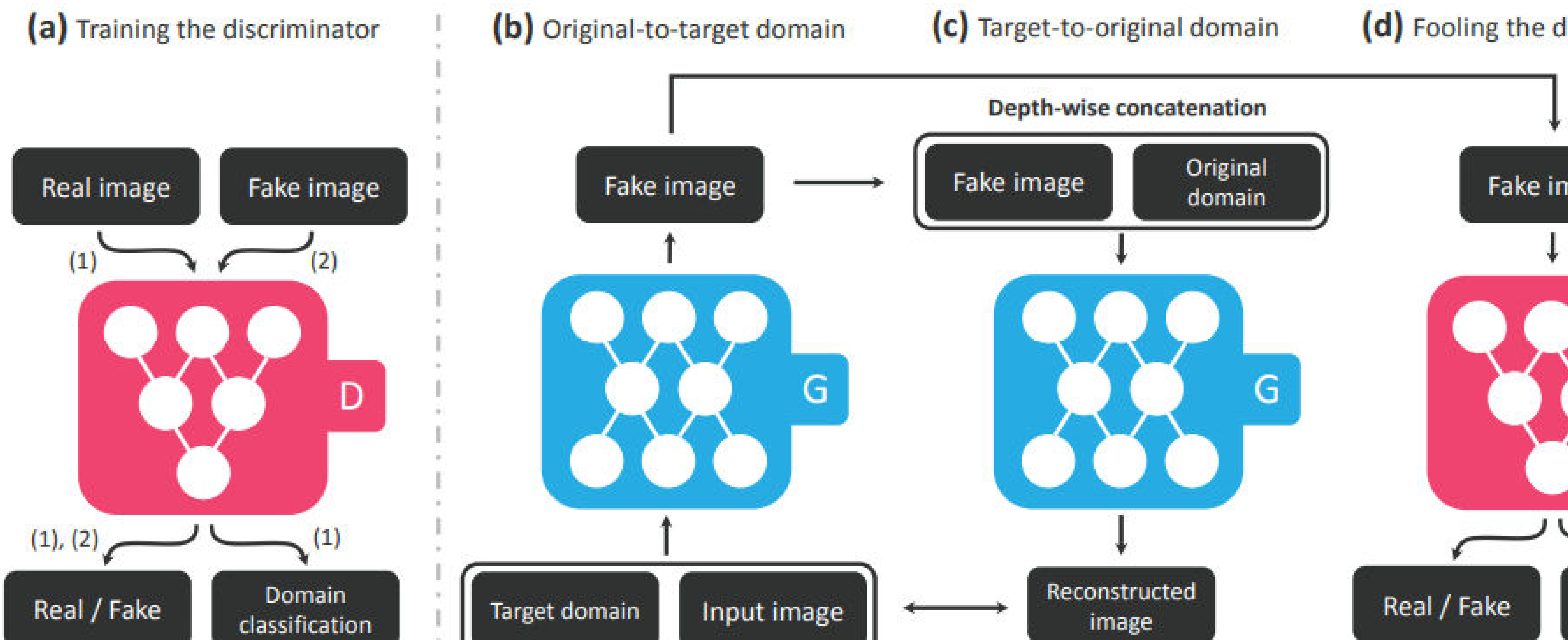
https://arxiv.org/abs/1703.105



StarGAN

https://arxiv.org/abs/1711.09020



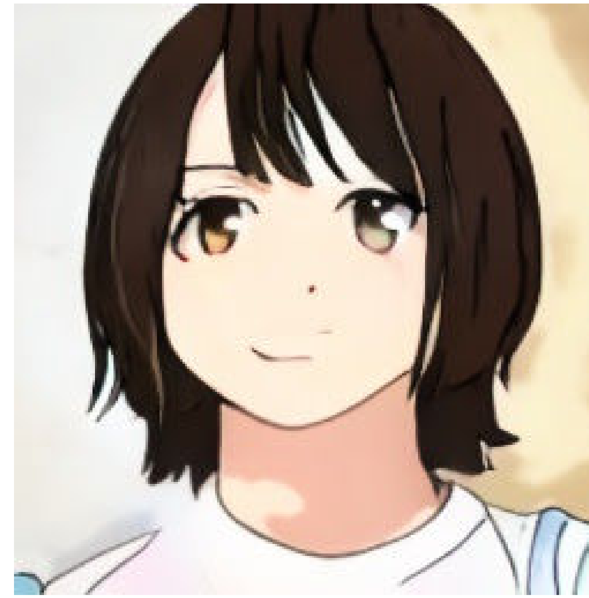



Q&A

SELFIE2ANIME

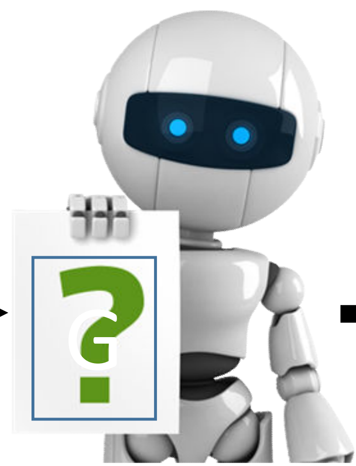
https://selfie2anime.c

https://arxiv.org/abs/1907.10



Text Style Transfer

你真笨
(negative)

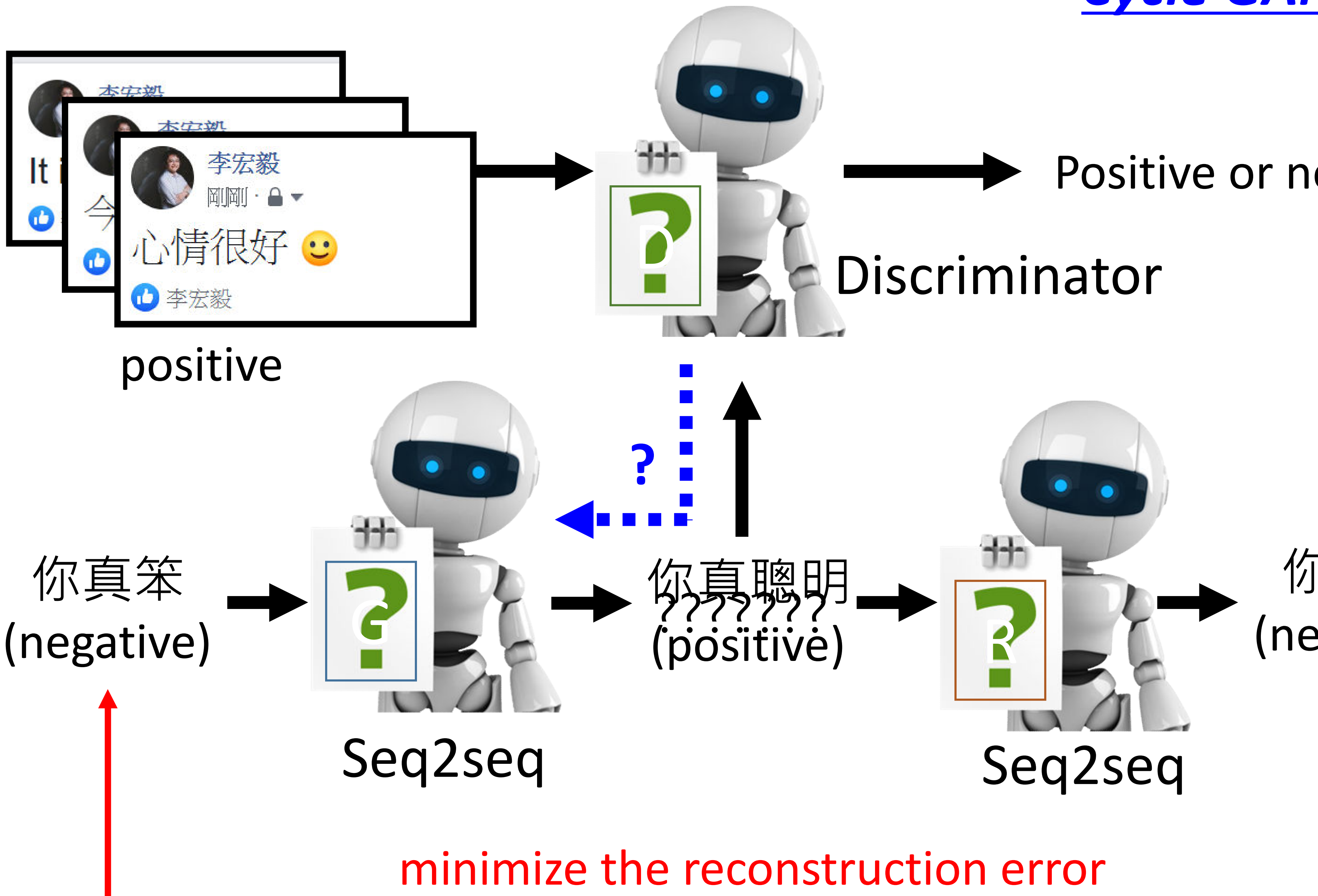


你真聰明
(positive)

Seq2seq

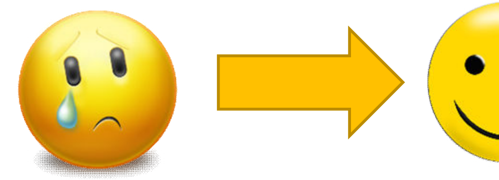
Text Style Transfer

Cycle GAN



感謝 張瓊之 同學提供實

Text Style Transfer



- From negative sentence to positive one

胃疼，沒睡醒，各種不舒服

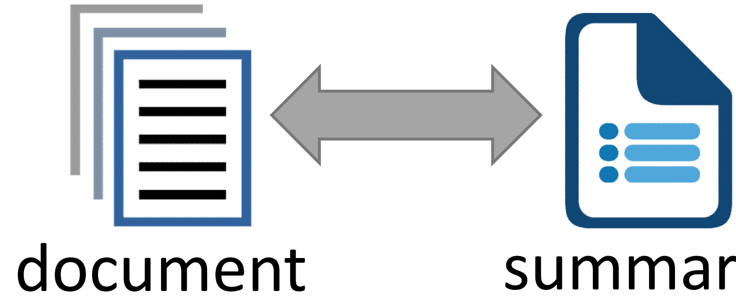
我都想去上班了, 真夠賤的!

暈死了, 吃燒烤、竟然遇到個變態狂

我肚子痛的厲害

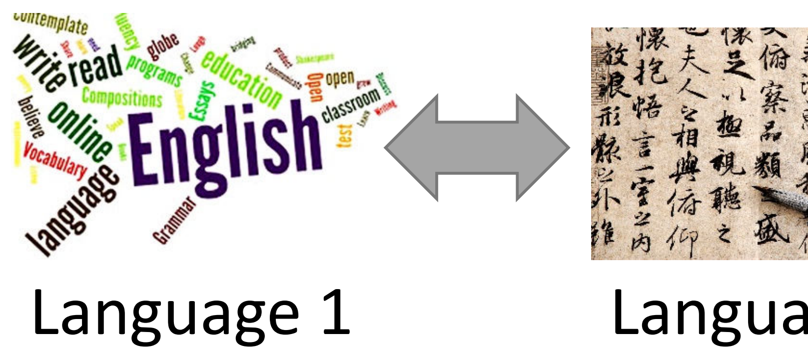
Unsupervised Abstractive Summarization

https://arxiv.org/abs/1810.02851



Unsupervised Translation

https://arxiv.org/abs/1710.04087
https://arxiv.org/abs/1710.11041



Unsupervised ASR

https://arxiv.org/abs/1804.00316
https://arxiv.org/abs/1812.09323
https://arxiv.org/abs/1904.04100

Evaluation of Generatio

Quality of Image

- Human evaluation is expensive (and sometimes unfair/unstable).
- How to evaluate the quality of the generated images automatically?

e.g., Inception net, VGG, etc.

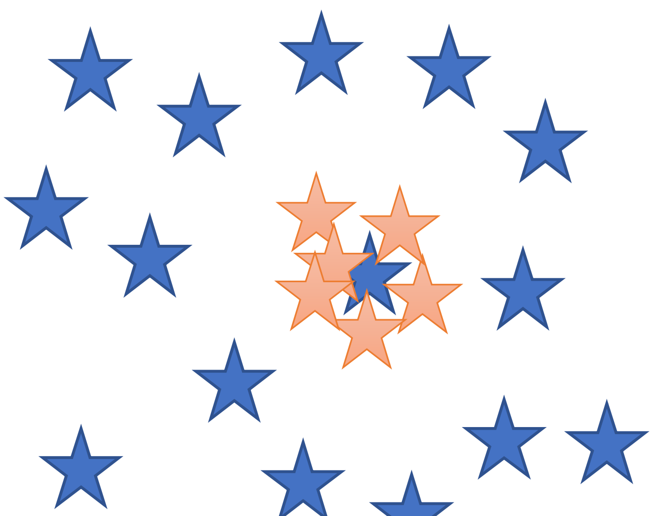
Concentrated distribu

means higher visual c

Diversity - Mode Collapse

: real data

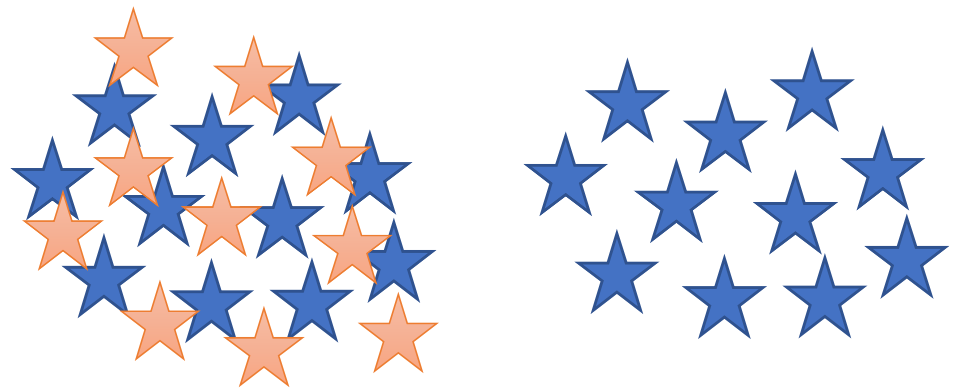
: generated data



Diversity - Mode Dropping

★ : real data

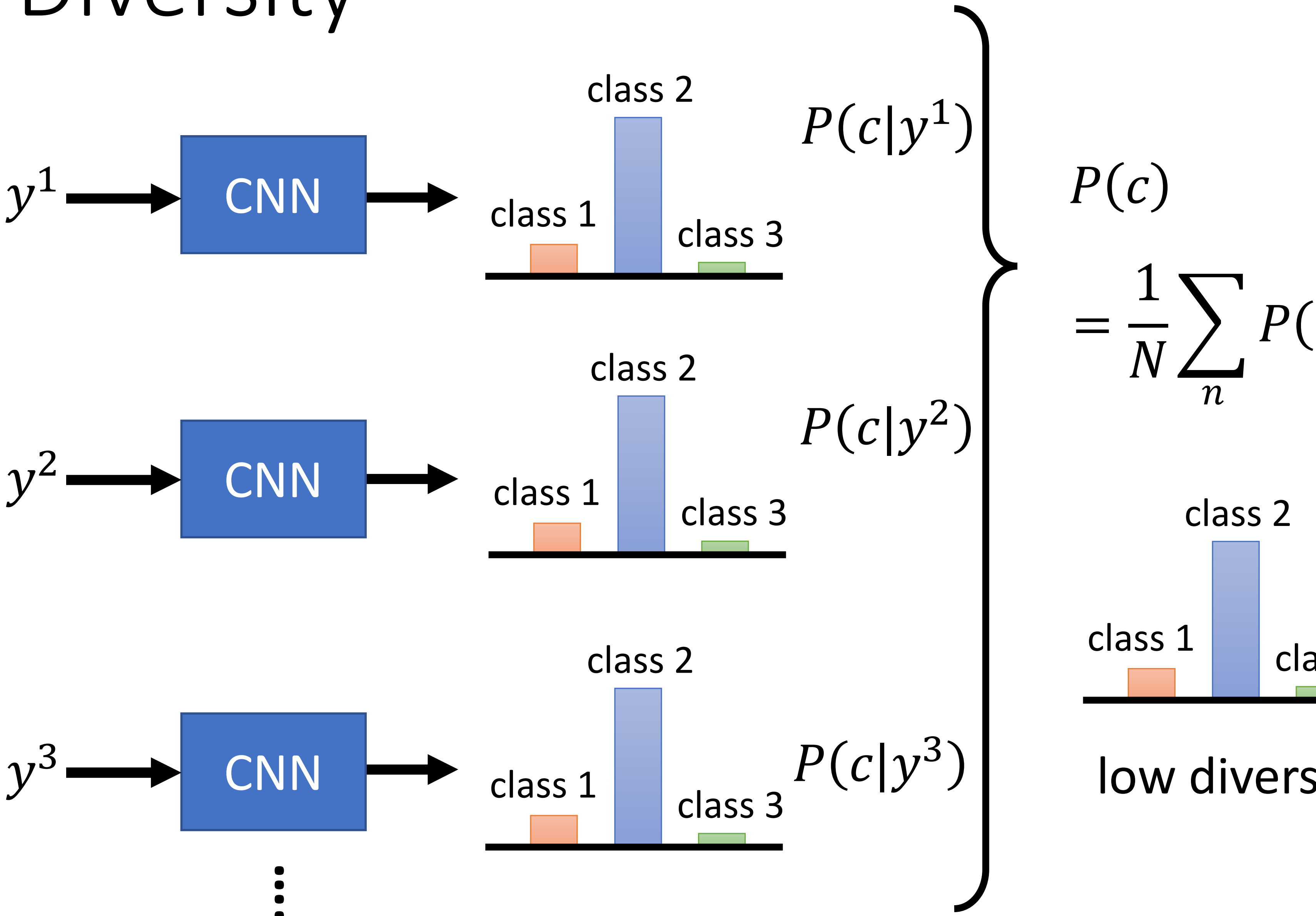
★ : generated data



Generator
at iteration t

Generator
at iteration t+1

Diversity

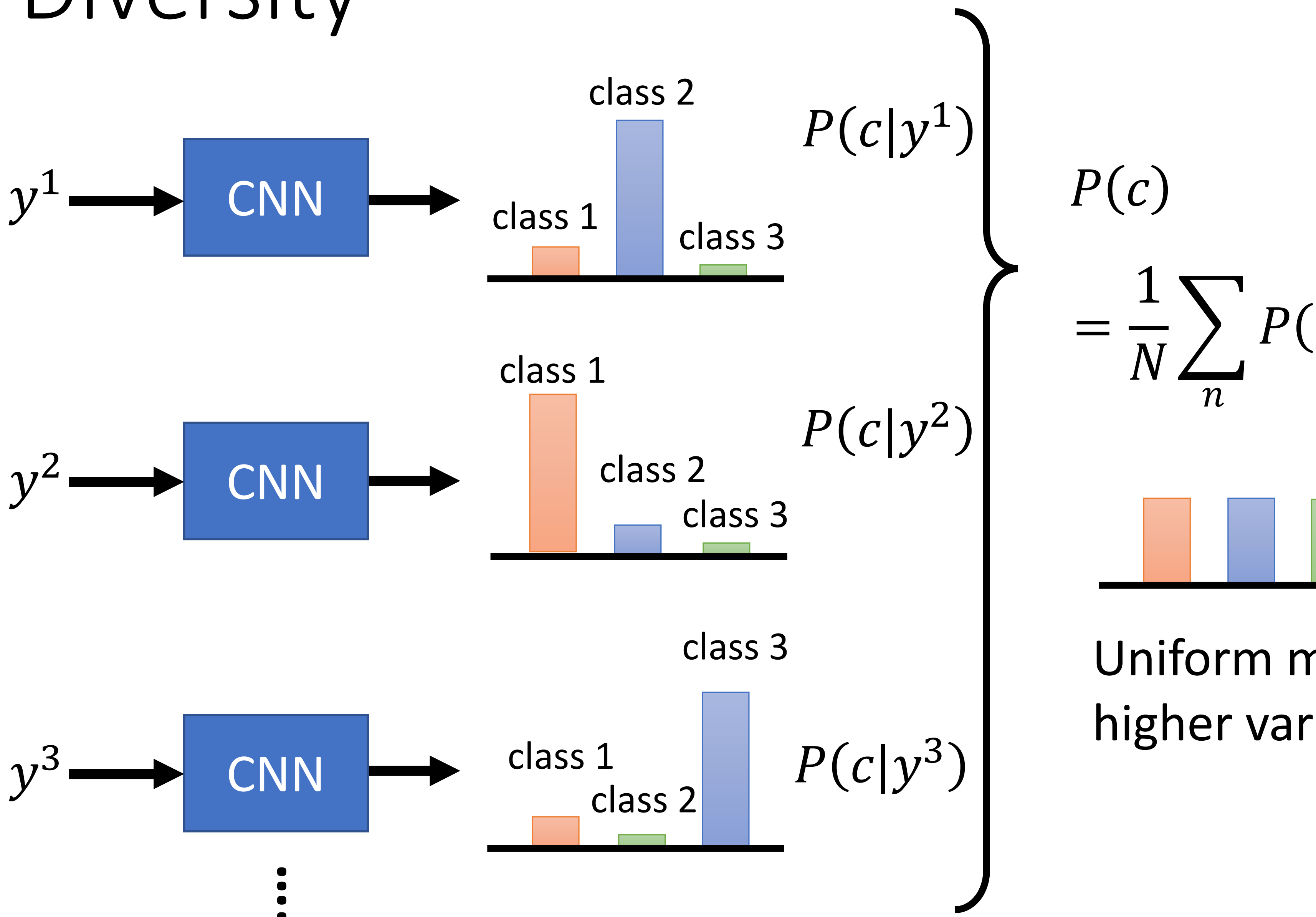


y^1
CNN
class 1
class 2
class 3
$P(c|y^1)$
y^2
CNN
class 1
class 2
class 3
$P(c|y^2)$
y^3
CNN
class 1
class 2
class 3
$P(c|y^3)$
$P(c)$
$= \frac{1}{N}\sum_n P($
class 1
class 2
cla
low divers

Diversity

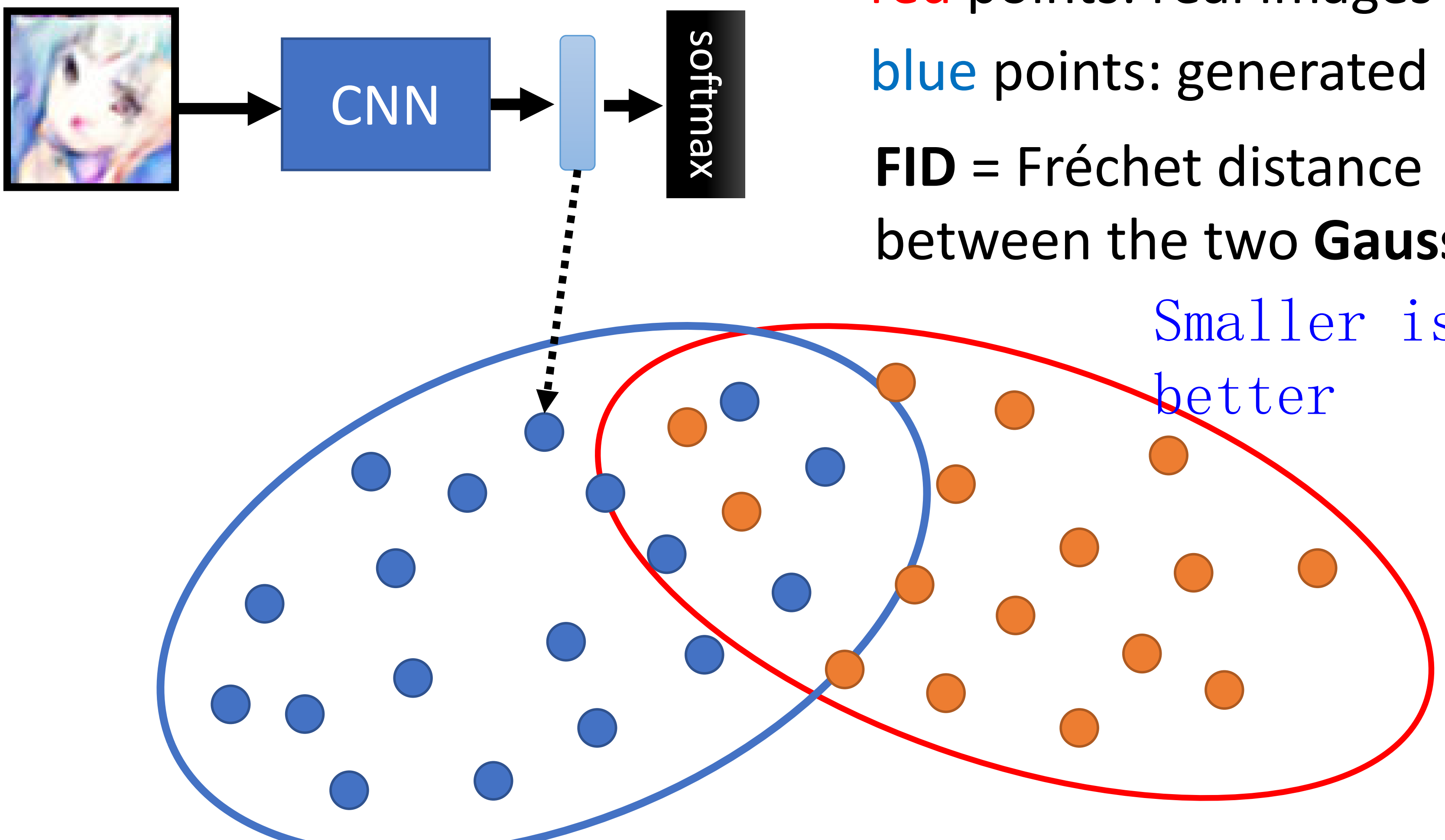
Inception Score (IS):

Good quality, large diversity → Large



https://arxiv.org/pdf/1706.0850

Fréchet Inception Distance (FID)



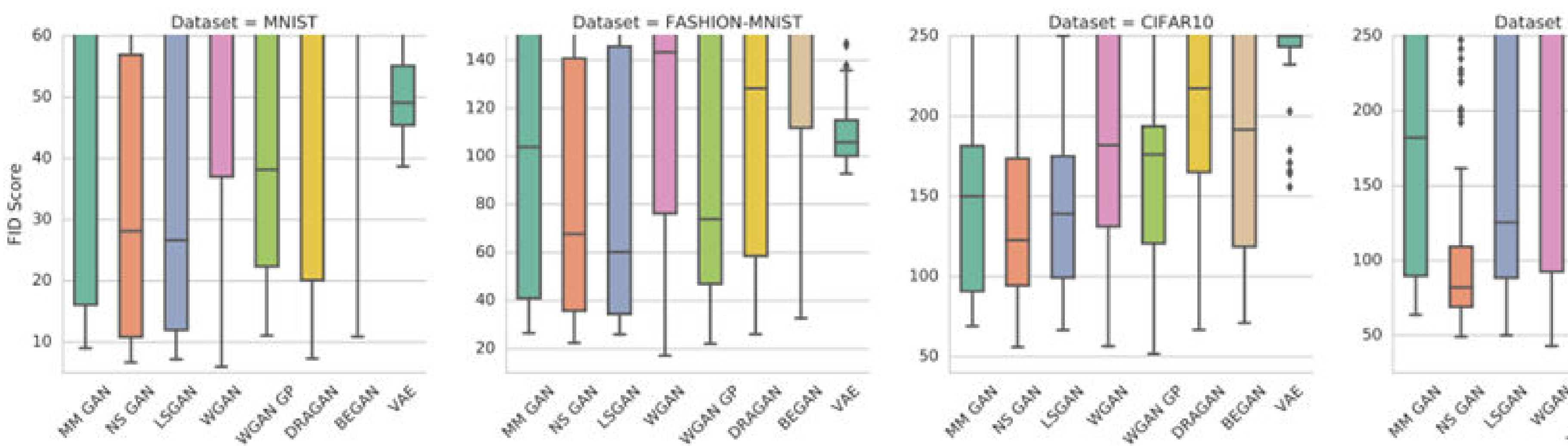
red points: real images

blue points: generated i

FID = Fréchet distance between the two **Gauss**

Smaller is better

GAN	Discriminator Loss	Generator Loss
MM GAN	$\mathcal{L}_{\mathrm{D}}^{\mathrm{GAN}} = -\mathbb{E}_{x\sim p_d}[\log(D(x))] + \mathbb{E}_{\hat{x}\sim p_g}[\log(1 - D(\hat{x}))]$	$\mathcal{L}_{\mathrm{G}}^{\mathrm{GAN}} = -\mathcal{L}_{\mathrm{D}}^{\mathrm{GAN}}$
NS GAN	$\mathcal{L}_{\mathrm{D}}^{\mathrm{NSGAN}} = \mathcal{L}_{\mathrm{D}}^{\mathrm{GAN}}$	$\mathcal{L}_{\mathrm{G}}^{\mathrm{NSGAN}} = \mathbb{E}_{\hat{x}\sim p_g}[\log(D(\hat{x}$
WGAN	$\mathcal{L}_{\mathrm{D}}^{\mathrm{WGAN}} = -\mathbb{E}_{x\sim p_d}[D(x)] + \mathbb{E}_{\hat{x}\sim p_g}[D(\hat{x})]$	$\mathcal{L}_{\mathrm{G}}^{\mathrm{WGAN}} - = \mathcal{L}_{\mathrm{D}}^{\mathrm{WGAN}}$
WGAN GP	$\mathcal{L}_{\mathrm{D}}^{\mathrm{WGAN}} = \mathcal{L}_{\mathrm{D}}^{\mathrm{WGAN}} + \lambda\mathbb{E}_{\hat{x}\sim p_g}[(\|\|\nabla D(\alpha x + (1-\alpha\hat{x})\|\|_2 - 1)^2]$	$\mathcal{L}_{\mathrm{G}}^{\mathrm{WGAN}} = -\mathbb{E}_{\hat{x}\sim p_g}[D(\hat{x})]$
LS GAN	$\mathcal{L}_{\mathrm{D}}^{\mathrm{LSGAN}} = -\mathbb{E}_{x\sim p_d}[(D(x) - 1)^2] + \mathbb{E}_{\hat{x}\sim p_g}[D(\hat{x})^2]$	$\mathcal{L}_{\mathrm{G}}^{\mathrm{LSGAN}} = -\mathbb{E}_{\hat{x}\sim p_g}[(D(\hat{x}$
DRAGAN	$\mathcal{L}_{\mathrm{D}}^{\mathrm{DRAGAN}} = \mathcal{L}_{\mathrm{D}}^{\mathrm{GAN}} + \lambda\mathbb{E}_{\hat{x}\sim p_d+\mathcal{N}(0,c)}[(\|\|\nabla D(\hat{x})\|\|_2 - 1)^2]$	$\mathcal{L}_{\mathrm{G}}^{\mathrm{DRAGAN}} = -\mathcal{L}_{\mathrm{D}}^{\mathrm{NS\ GAN}}$
BEGAN	$\mathcal{L}_{\mathrm{D}}^{\mathrm{BEGAN}} = \mathbb{E}_{x\sim p_d}[\|\|x - \mathrm{AE}(x)\|\|_1] - k_t\mathbb{E}_{\hat{x}\sim p_g}[\|\|\hat{x} - \mathrm{AE}(\hat{x})\|\|_1]$	$\mathcal{L}_{\mathrm{G}}^{\mathrm{BEGAN}} = \mathbb{E}_{\hat{x}\sim p_g}[\|\|\hat{x} - \mathrm{AE}$



FIT: Smaller is better

Are GANs Created Equal? A Large-Scale S

https://arxiv.org/pdf/1511.0

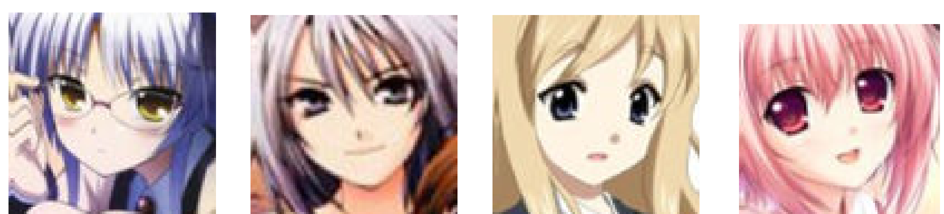
We don’t want memory GAN.

Real Data

Generated Data

Same as real data …

Generated Data



Simply flip real data …

To learn more about evaluation …

	Measure	Description
Quantitative	1. Average Log-likelihood [18, 22]	• Log likelihood of explaining realworld held out/test data using a density estimated from the generated data (*e.g.* using KDE or Parzen window estimation). $L = \frac{1}{N}\sum_i \log P_{model}(\mathbf{x}_i)$
	2. Coverage Metric [33]	• The probability mass of the true data "covered" by the model distribution $C := P_{data}(dP_{model} > t)$ with t such that $P_{model}(dP_{model} > t) = 0.95$
	3. Inception Score (IS) [3]	• KLD between conditional and marginal label distributions over generated data. $\exp\left(\mathbb{E}_{\mathbf{x}}\left[\mathrm{KL}\left(p(y \mid \mathbf{x}) \parallel p(y)\right)\right]\right)$
	4. Modified Inception Score (m-IS) [34]	• Encourages diversity within images sampled from a particular category. $\exp(\mathbb{E}_{\mathbf{x}_i}[\mathbb{E}_{\mathbf{x}_j}[(\mathrm{KL}(P(y\|\mathbf{x}_i)\|\|P(y\|\mathbf{x}_j)))]])$
	5. Mode Score (MS) [35]	• Similar to IS but also takes into account the prior distribution of the labels over real data. $\exp\left(\mathbb{E}_{\mathbf{x}}\left[\mathrm{KL}\left(p(y \mid \mathbf{x}) \parallel p(y^{train})\right)\right] - \mathrm{KL}\left(p(y) \parallel p(y^{train})\right)\right)$
	6. AM Score [36]	• Takes into account the KLD between distributions of training labels vs. predicted labels, as well as the entropy of predictions. $\mathrm{KL}(p(y^{\mathrm{train}}) \parallel p(y)) + \mathbb{E}_{\mathbf{x}}[H(y\|\mathbf{x})]$
	7. Fréchet Inception Distance (FID) [37]	• Wasserstein-2 distance between multi-variate Gaussians fitted to data embedded into a feature space $FID(r,g) = \|\|\mu_r - \mu_g\|\|_2^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{\frac{1}{2}})$
	8. Maximum Mean Discrepancy (MMD) [38]	• Measures the dissimilarity between two probability distributions P_r and P_g using samples drawn independently from each distribution. $M_k(P_r, P_g) = \mathbb{E}_{\mathbf{x},\mathbf{x}'\sim P_r}[k(\mathbf{x},\mathbf{x}')] - 2\mathbb{E}_{\mathbf{x}\sim P_r, \mathbf{y}\sim P_g}[k(\mathbf{x},\mathbf{y})] + \mathbb{E}_{\mathbf{y},\mathbf{y}'\sim P_g}[k(\mathbf{y},\mathbf{y}')]$
	9. The Wasserstein Critic [39]	• The critic (*e.g.* an NN) is trained to produce high values at real samples and low values at generated samples $\hat{W}(\mathbf{x}_{test}, \mathbf{x}_g) = \frac{1}{N}\sum_{i=1}^{N}\hat{f}(\mathbf{x}_{test}[i]) - \frac{1}{N}\sum_{i=1}^{N}\hat{f}(\mathbf{x}_g[i])$
	10. Birthday Paradox Test [27]	• Measures the support size of a discrete (continuous) distribution by counting the duplicates (near duplicates)
	11. Classifier Two Sample Test (C2ST) [40]	• Answers whether two samples are drawn from the same distribution (*e.g.* by training a binary classifier)
	12. Classification Performance [1, 15]	• An indirect technique for evaluating the quality of unsupervised representations (*e.g.* feature extraction; FCN score). See also the GAN Quality Index (GQI) [41].
	13. Boundary Distortion [42]	• Measures diversity of generated samples and covariate shift using classification methods.
	14. Number of Statistically-Different Bins (NDB) [43]	• Given two sets of samples from the same distribution, the number of samples that fall into a given bin should be the same up to sampling noise
	15. Image Retrieval Performance [44]	• Measures the distributions of distances to the nearest neighbors of some query images (*i.e.* diversity)
	16. Generative Adversarial Metric (GAM) [31]	• Compares two GANs by having them engaged in a battle against each other by swapping discriminators or generators. $p(\mathbf{x}\|y=1; M_1')/p(\mathbf{x}\|y=1; M_2') = (p(y=1\|\mathbf{x}; D_1)p(\mathbf{x}; G_2))/(p(y=1\|\mathbf{x}; D_2)p(\mathbf{x}; G_1))$
	17. Tournament Win Rate and Skill Rating [45]	• Implements a tournament in which a player is either a discriminator that attempts to distinguish between real and fake data or a generator that attempts to fool the discriminators into accepting fake data as real.
	18. Normalized Relative Discriminative Score (NRDS) [32]	• Compares n GANs based on the idea that if the generated samples are closer to real ones, more epochs would be needed to distinguish them from real samples.
	19. Adversarial Accuracy and Divergence [46]	• Adversarial Accuracy. Computes the classification accuracies achieved by the two classifiers, one trained on real data and another on generated data, on a labeled validation set to approximate $P_g(y\|\mathbf{x})$ and $P_r(y\|\mathbf{x})$. Adversarial Divergence: Computes $\mathrm{KL}(P_g(y\|\mathbf{x}), P_r(y\|\mathbf{x}))$
	20. Geometry Score [47]	• Compares geometrical properties of the underlying data manifold between real and generated data.
	21. Reconstruction Error [48]	• Measures the reconstruction error (*e.g.* L_2 norm) between a test image and its closest generated image by optimizing for z (*i.e.* $min_{\mathbf{z}}\|\|G(\mathbf{z}) - \mathbf{x}^{(test)}\|\|^2$)
	22. Image Quality Measures [49, 50, 51]	• Evaluates the quality of generated images using measures such as SSIM, PSNR, and sharpness difference
	23. Low-level Image Statistics [52, 53]	• Evaluates how similar low-level statistics of generated images are to those of natural scenes in terms of mean power spectrum, distribution of random filter responses, contrast distribution, etc.
	24. Precision, Recall and F_1 score [23]	• These measures are used to quantify the degree of overfitting in GANs, often over toy datasets.
Qualitative	1. Nearest Neighbors	• To detect overfitting, generated samples are shown next to their nearest neighbors in the training set
	2. Rapid Scene Categorization [18]	• In these experiments, participants are asked to distinguish generated samples from real images in a short presentation time (*e.g.* 100 ms); *i.e.* real v.s fake
	3. Preference Judgment [54, 55, 56, 57]	• Participants are asked to rank models in terms of the fidelity of their generated images (*e.g.* pairs, triples)
	4. Mode Drop and Collapse [58, 59]	• Over datasets with known modes (*e.g.* a GMM or a labeled dataset), modes are computed as by measuring the distances of generated data to mode centers
	5. Network Internals [1, 60, 61, 62, 63, 64]	• Regards exploring and illustrating the internal representation and dynamics of models (*e.g.* space continuity) as well as visualizing learned features

Pros and cons of GAN evaluation measures

Concluding Remarks

- Introduction of Generative Models
- Generative Adversarial Network (GAN)
- Theory behind GAN
- Tips for GAN
- Conditional Generation
- Learning from unpaired data
- Evaluation of Generative Models


Q&A