Lecture 8: Generative Models (Part 1: GANs)

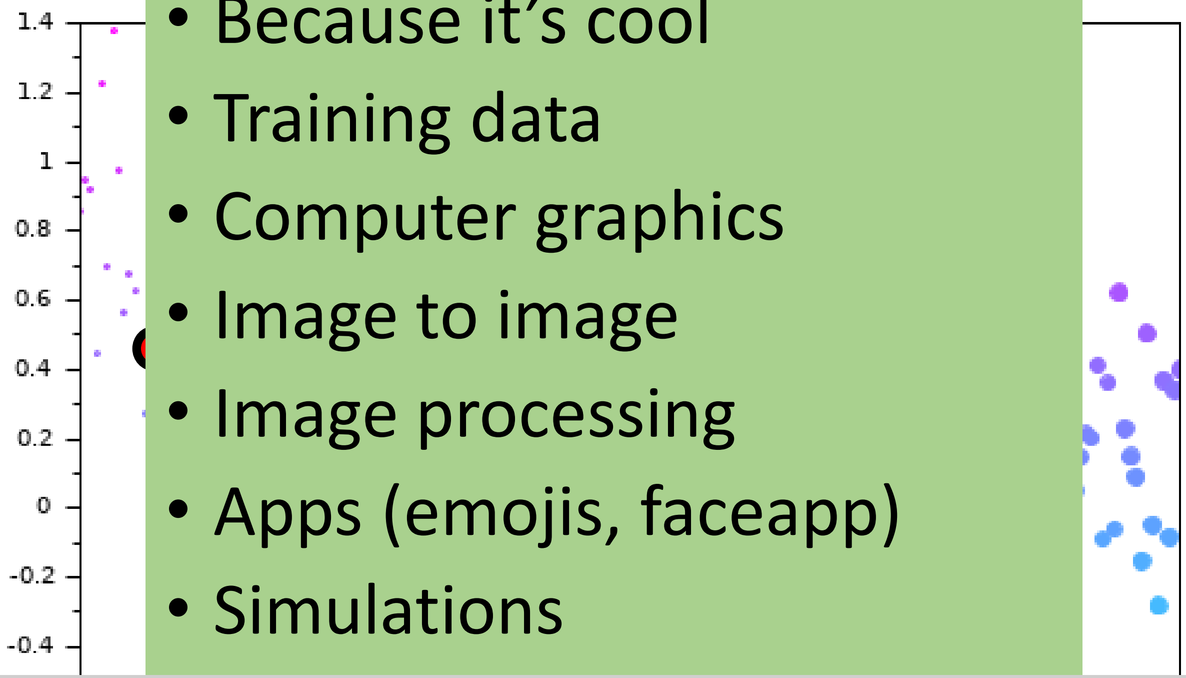
Agenda (Today: GANs)

1. Goal, motivation, Basic methods
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)

Tutorial (next week): Non-Adversarial methods (VAE, VQVAE, IMLE)

Goal

Images
Text
Audio
Video
Whatever…



Why?

- Because it's cool
- Training data
- Computer graphics
- Image to image
- Image processing
- Apps (emojis, faceapp)
- Simulations

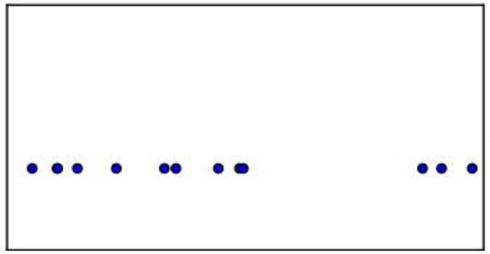
The correct approach when there is more than one valid solution!

Generative methods:

- Parametric distribution estimation (e.g. GMM)
- Autoregressive models (e.g. RNN, Causal CNN, Tranformer)
- Latent space mapping (VAE, GAN, more)

Basic approach:

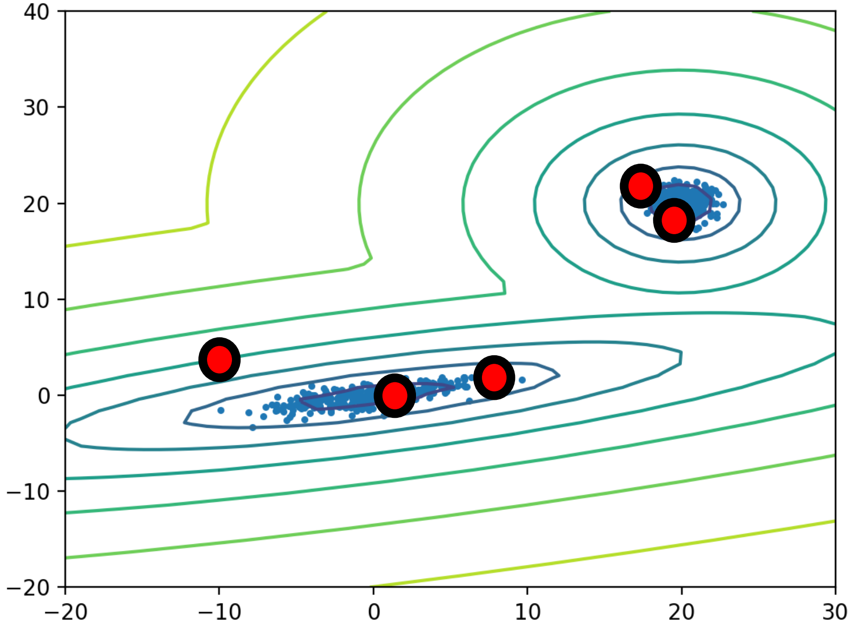
- Density estimation



(Goodfellow 2016)

Example: GMM

Step 1: observe a set of samples



Step 2: assume a GMM model

$$p(x|\theta) = \sum_i \pi_i \mathcal{N}(x|\mu_i, \Sigma_i)$$

Step 3: perform maximum likelihood learning

$$\max_\theta \sum_{x^{(j)} \in \text{Dataset}} \log p(\theta|x^{(j)})$$

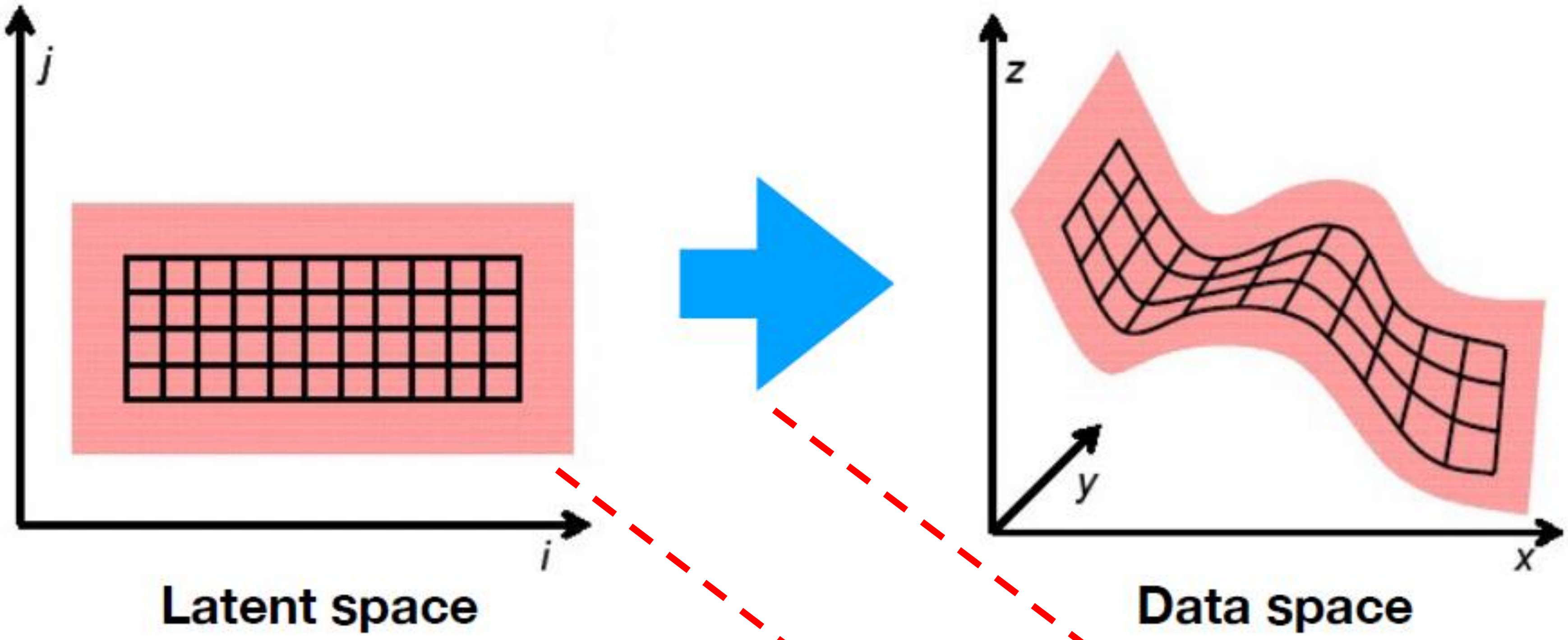
Step 4: Sample

Elements in the slide taken from NVIDIA CVPR'18 GANs tutorial

Agenda

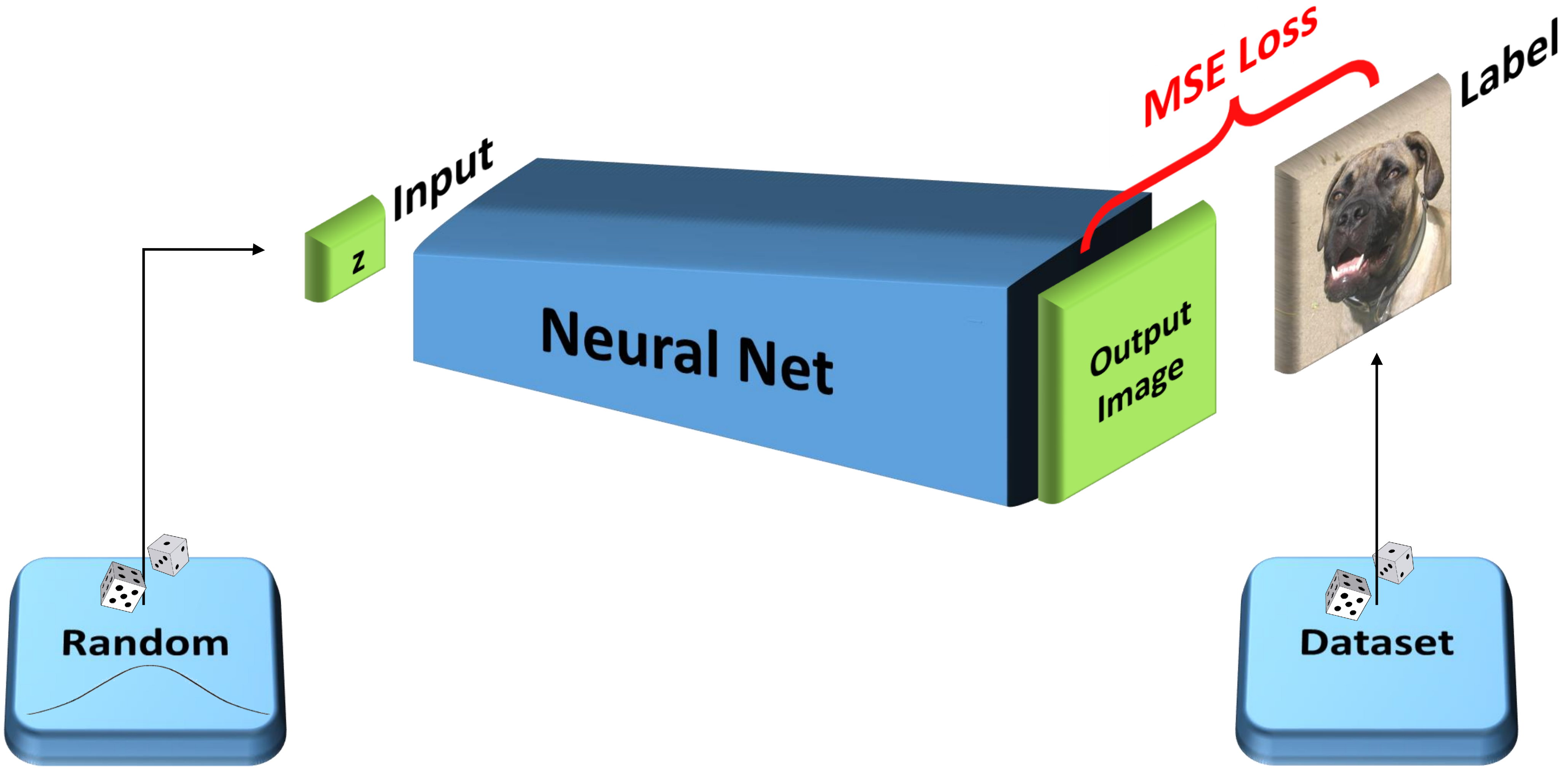
1. ~~Goal, motivation, Basic methods~~
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)

Latent space mapping approach



Data likelihood: $p_\theta(x) = \int p_\theta(z) p_\theta(x|z) dz$

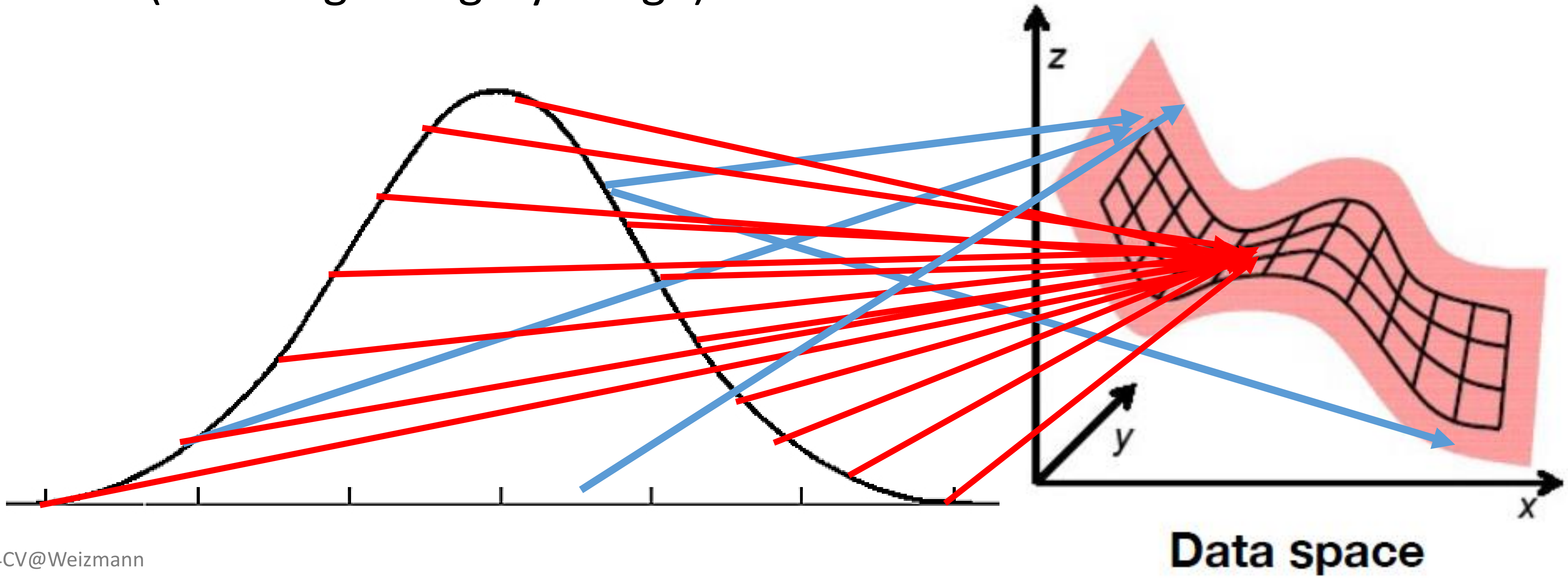
How about this idea for a generative model?



No good!

In expectation: every noise is mapped to every instance

Best L2 solution: All noise is mapped to the mean
(For images: ~ grey image)



Generative Adversarial Networks

# of GAN related papers per year
(Salehi et al.)

Q: What makes a good counterfeiter?

Train G

Q: Who do you train first?

A: Alternate training! G,D,G,D….

Minimax game: Make the best cop do the worst mistake!

Update weights

Don't update weights

$$\min_{G} \max_{D} \{\mathbb{E}_{x \sim p_{data}} \log(D(x)) + \mathbb{E}_{z \sim p_z} log(1 - D(G(z)))\}$$

FAQ1: Why does it work?

- D learns probability! G trains to sample instance with high probability!

- Objective does not determine mapping directly- arrangement of latent space is learned!

- Theory: minimizes JS divergence between generated and real distributions.

FAQ2: Why alternating?

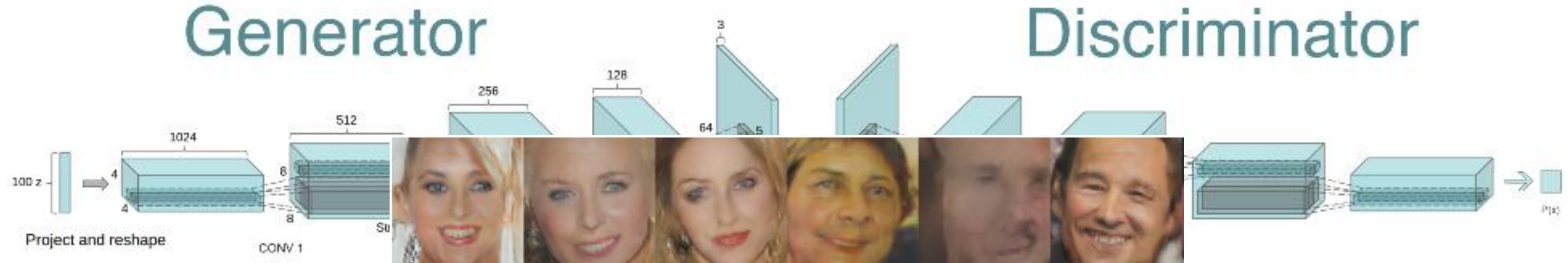
- Gradients are meaningless when game is unbalanced.

- Pre-train D? Negative examples?

- Pre-train G? What loss?
  For G, D is a **learned loss function**

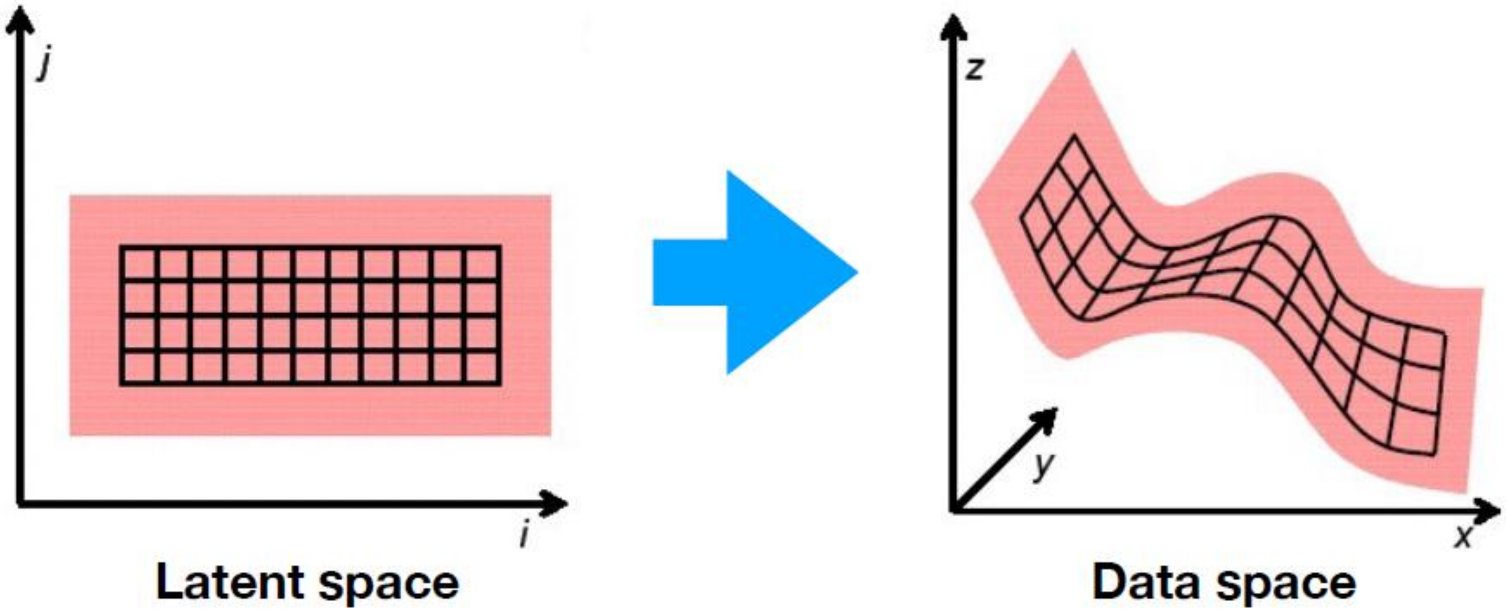
GANs, Goodfellow 2014

DCGAN Radford 2015



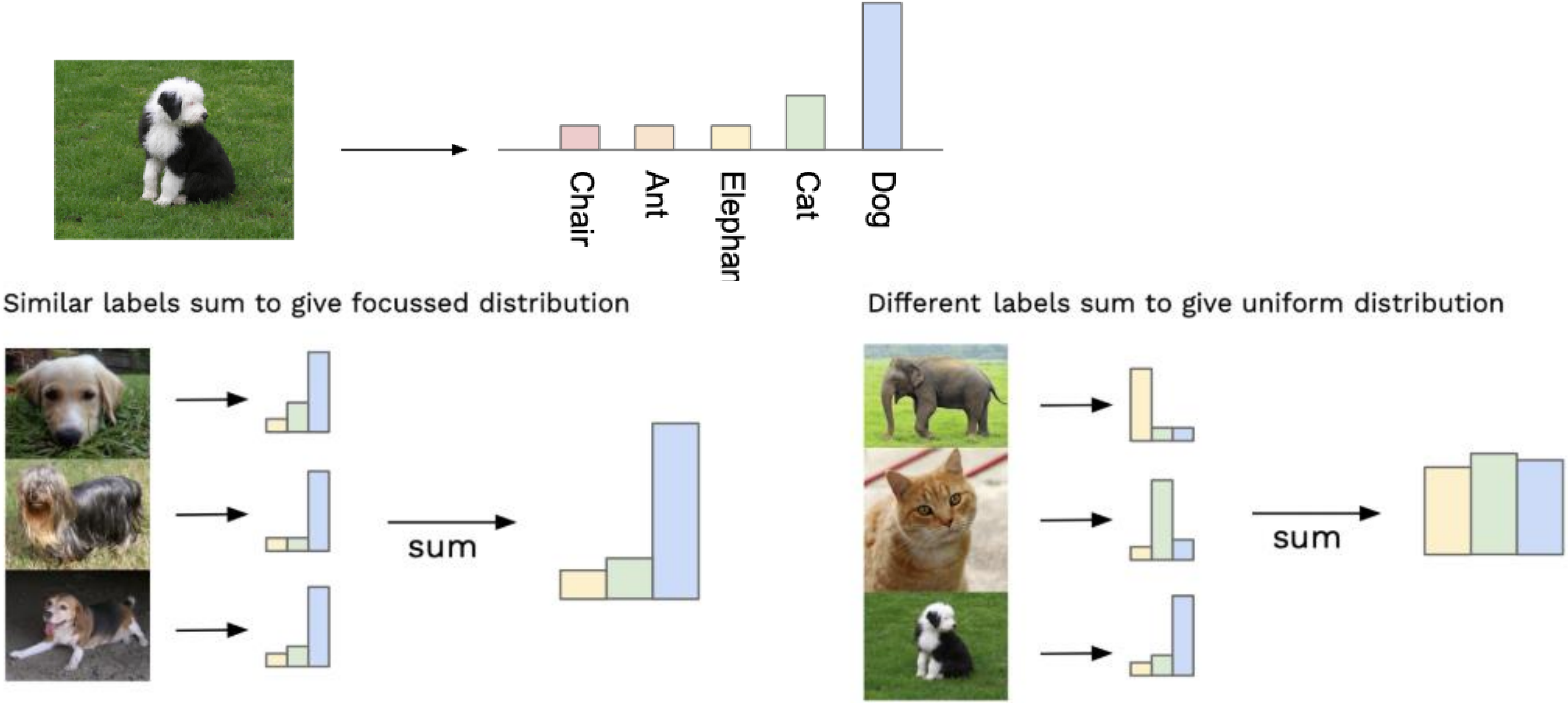
Latent space interpolation

Why does it work?



1. Every point is mapped to a valid example.
2. Network is continuous.

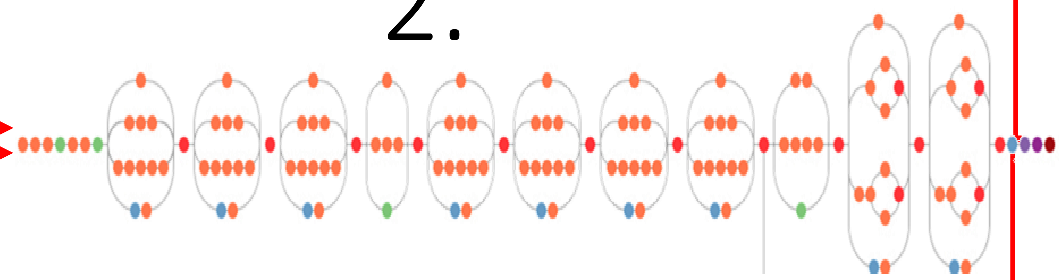
Evaluation metrics: Inception score

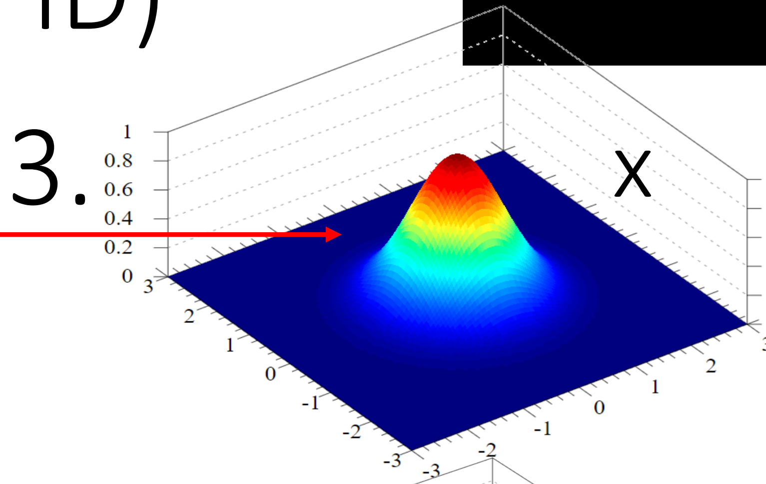


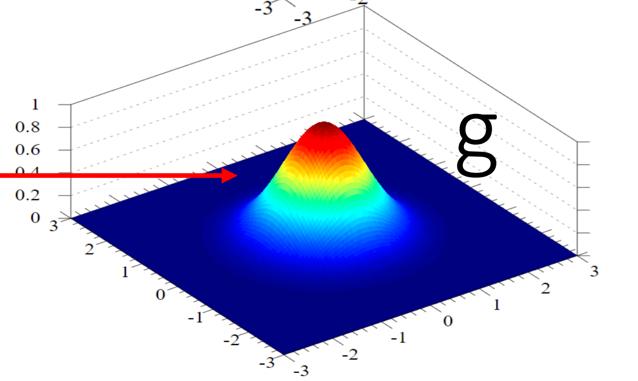
$$\mathbf{IS}(G) = \exp\left(\, \mathbb{E}_{\mathbf{x}\sim p_g}\, D_{KL}(\, p(y|\mathbf{x}) \,\|\, p(y)\,)\,\right)$$

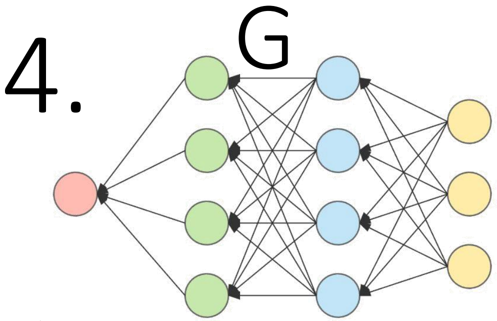
Fréchet Inception Distance (FID)

Depends on the number of samples!

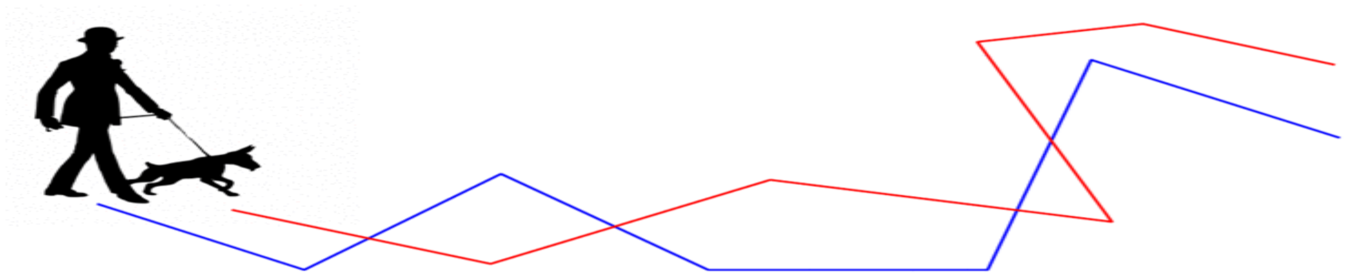








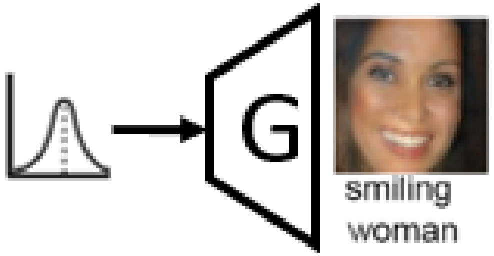
5. $$\mathtt{FID}(x,g) = ||\mu_x - \mu_g||_2^2 + \mathrm{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x\Sigma_g)^{\frac{1}{2}})$$

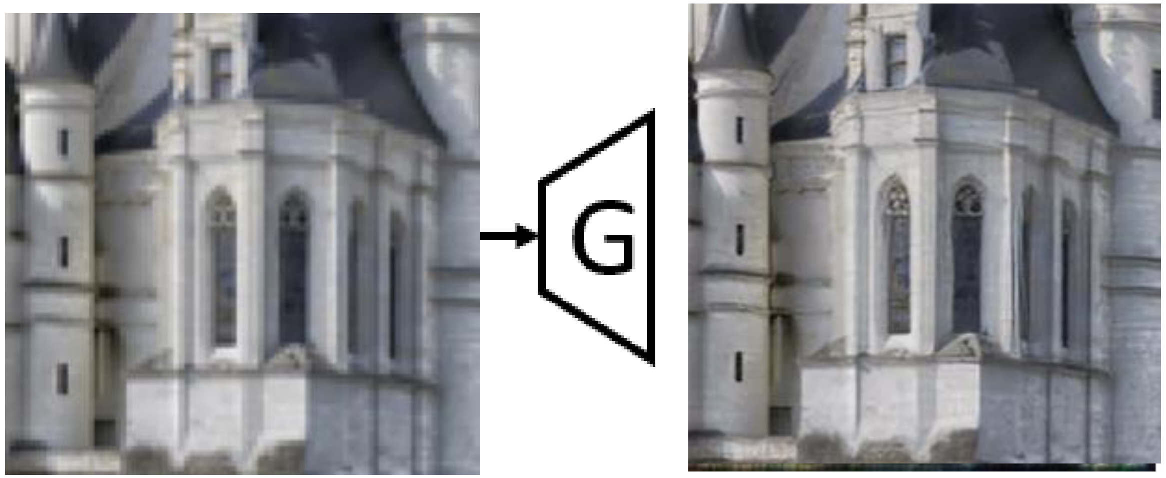


Agenda

1. ~~Goal, motivation, Basic methods~~
2. ~~Introduction to GANs (basic setup, intuition, eval measures)~~
3. $\boxed{\text{Image to image (pix2pix, CycleGAN)}}$
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)

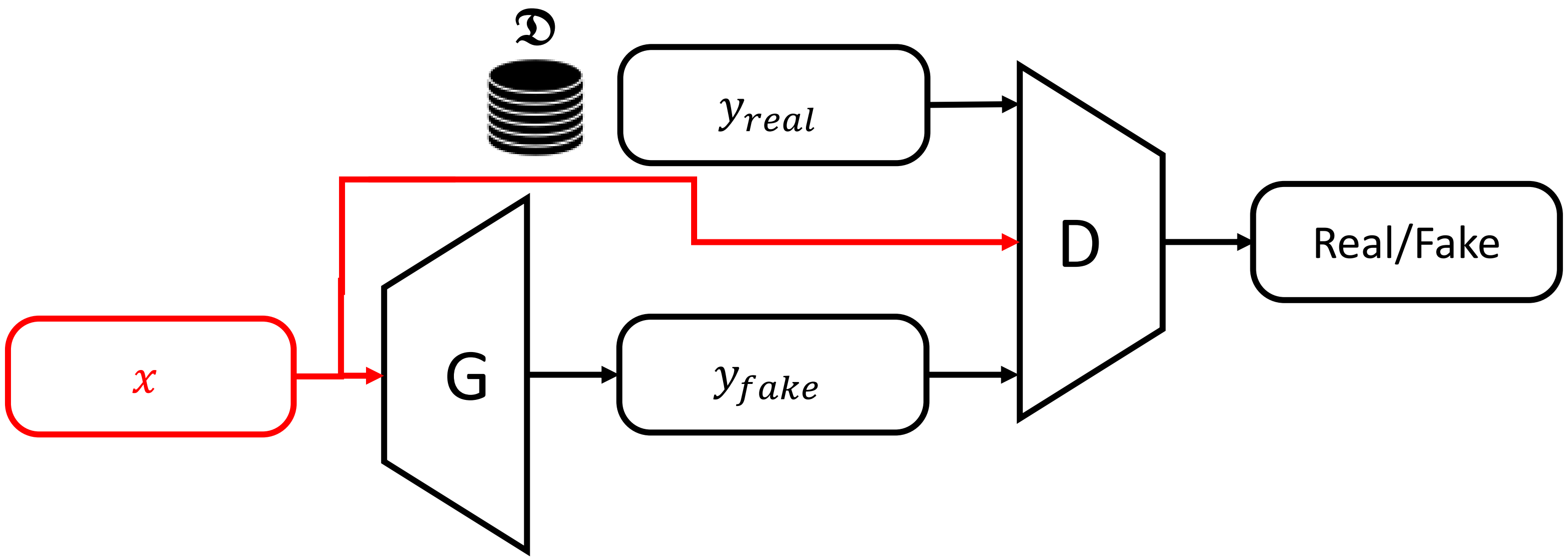
Image to Image translation





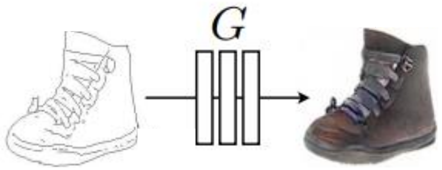
Isola et al. Nov2016

Conditional GAN



$$\mathcal{L}_{C-GAN} = \min_{G} \max_{D} \mathbb{E}[\log D(y, x)] + \mathbb{E}[\log(1 - D(G(x), x))]$$

Pix2Pix



$$\mathcal{L}_{C-GAN} = \min_{G} \max_{D} \mathbb{E}[\log D(y, x)] + \mathbb{E}[\log(1 - D(G(x), x))]$$

$$\mathcal{L}_{L1} = \|y - G(x, z)\|_1$$

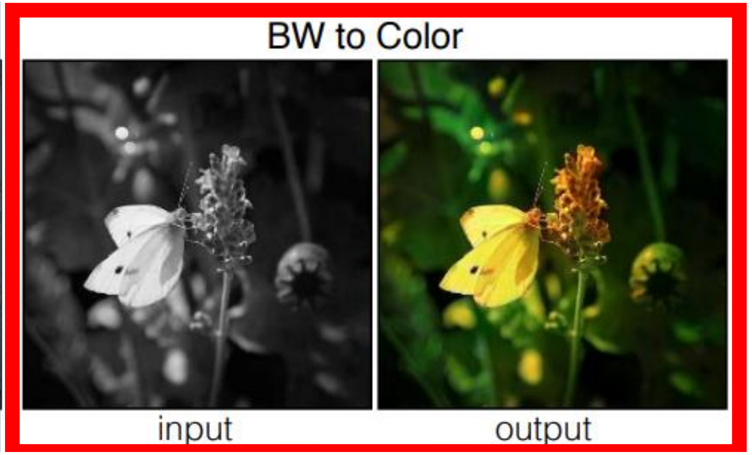
$$Objective = \mathcal{L}_{C-GAN} + \lambda \cdot \mathcal{L}_{L1}$$

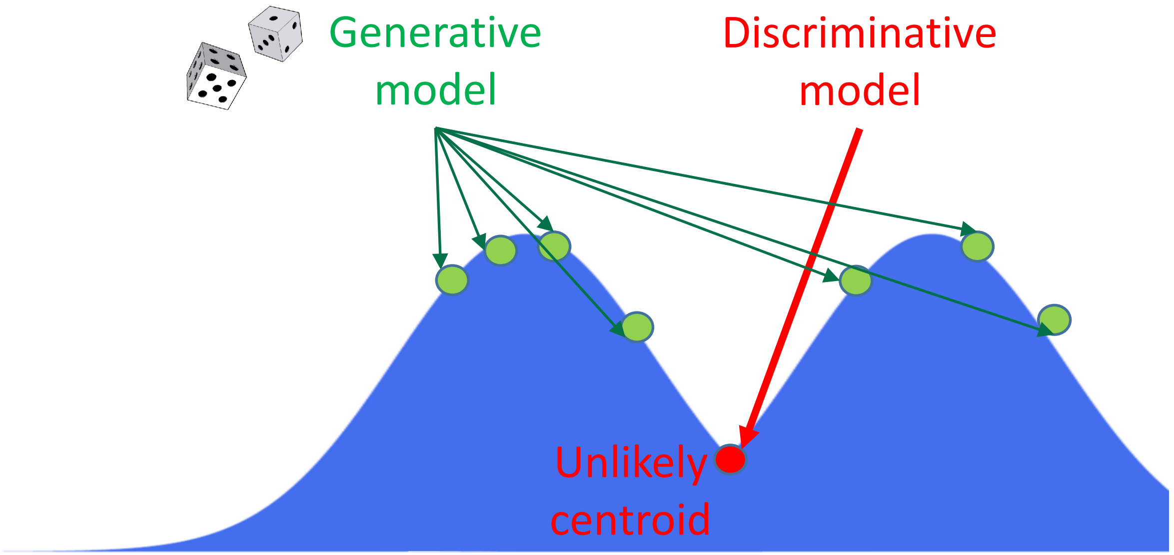
Isola et al. Nov2016

Slide by Sefi Bell-Kligler & Akhiad Bercovich

Generative VS Discriminative

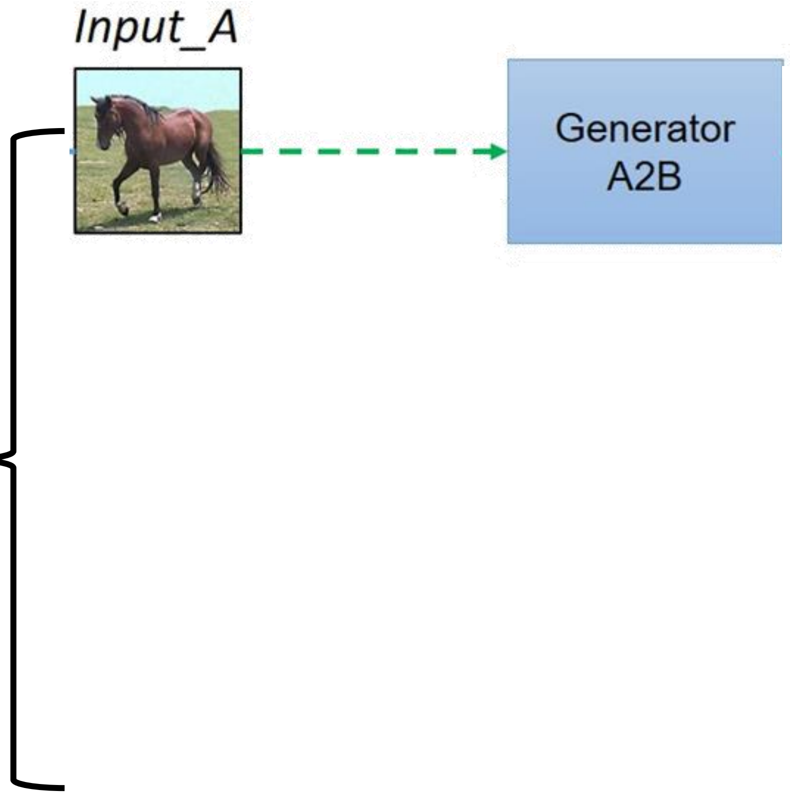
What would happen if we train regular supervised mapping?





True AI needs no explicit supervision

CycleGAN (Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros)



Cycle-consistency Loss $\boldsymbol{L_1}$

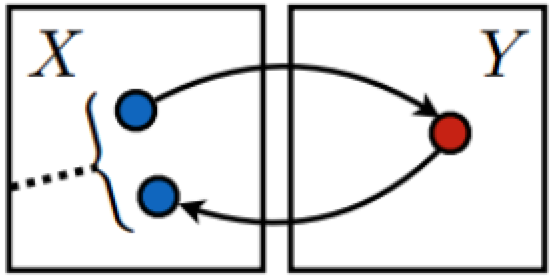


Figure from https://hardikbansal.github.io/CycleGANBlog/

CycleGAN (Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros)

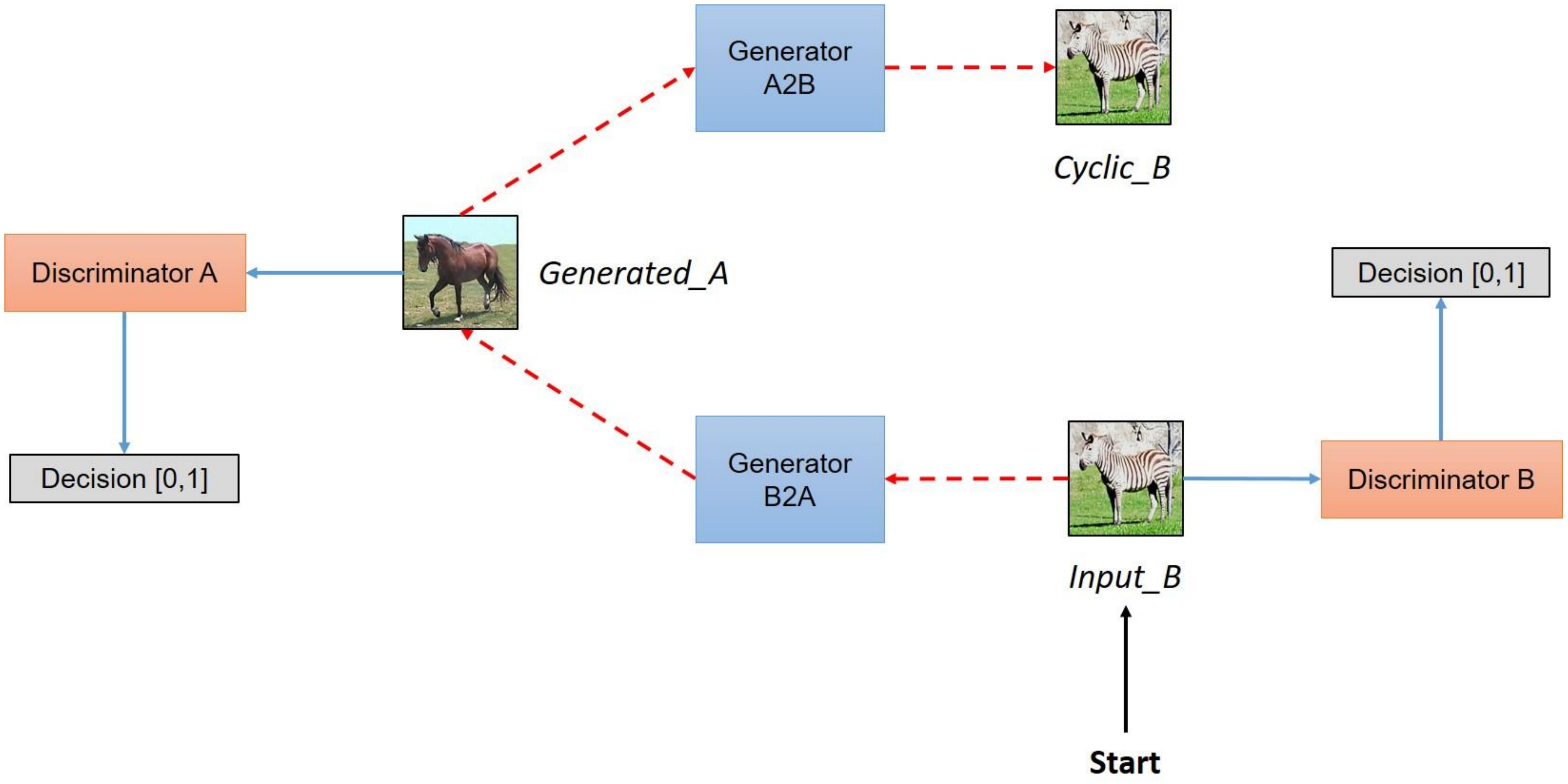
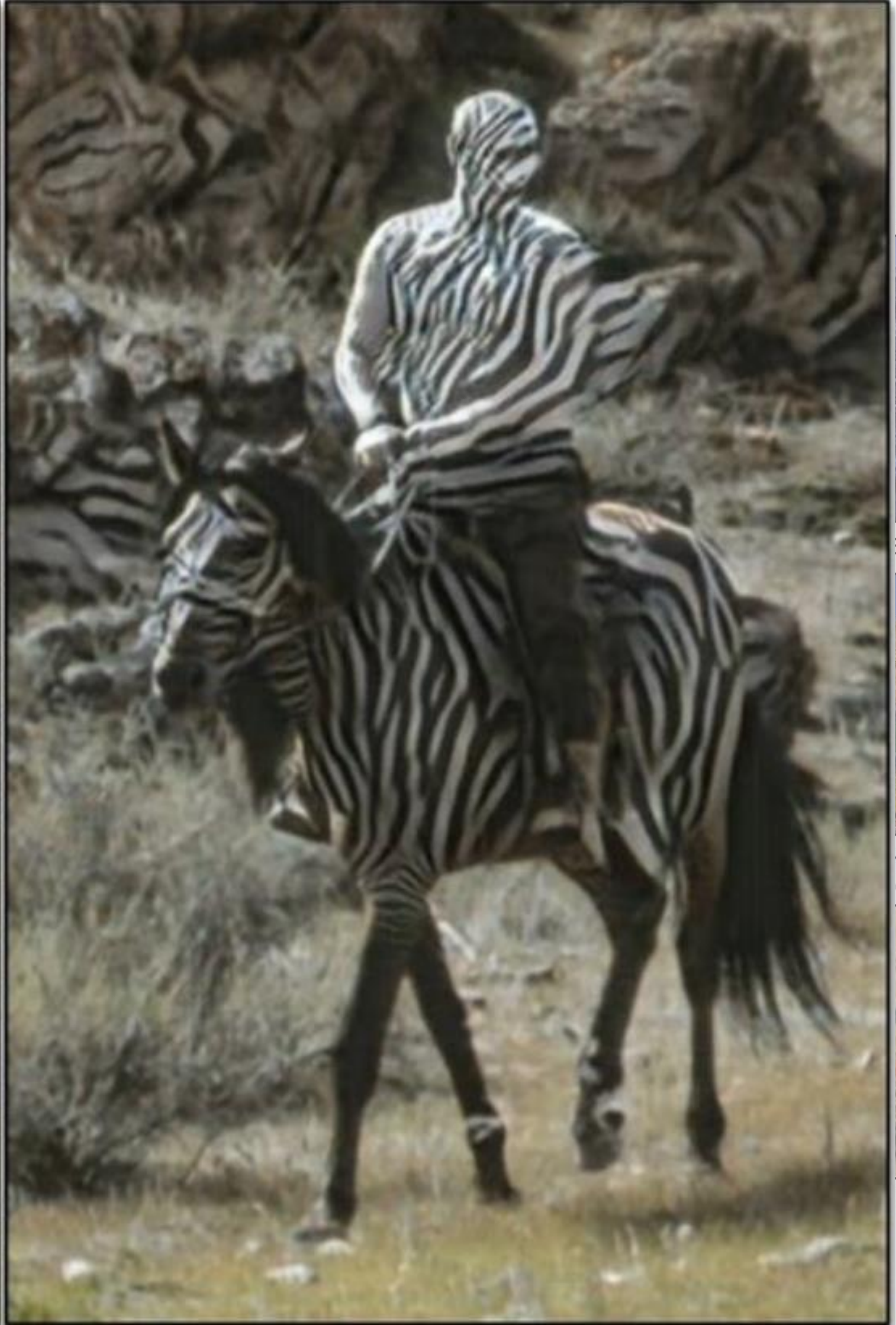


Figure from https://hardikbansal.github.io/CycleGANBlog/

Photo



Agenda

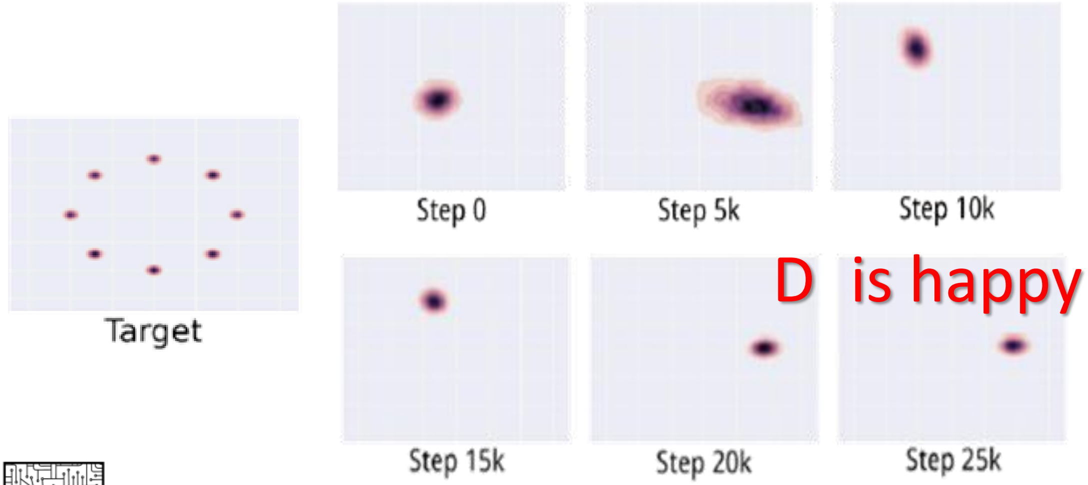
1. ~~Goal, motivation, Basic methods~~
2. ~~Introduction to GANs (basic setup, intuition, eval measures)~~
3. ~~Image to image (pix2pix, CycleGAN)~~
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)

Training GANs is hard

- Stability

Tra
GA
Be

- Mode collapse



D is happy!

- GANs can over-train

- Lo

Efficient training tricks (the basic ones)

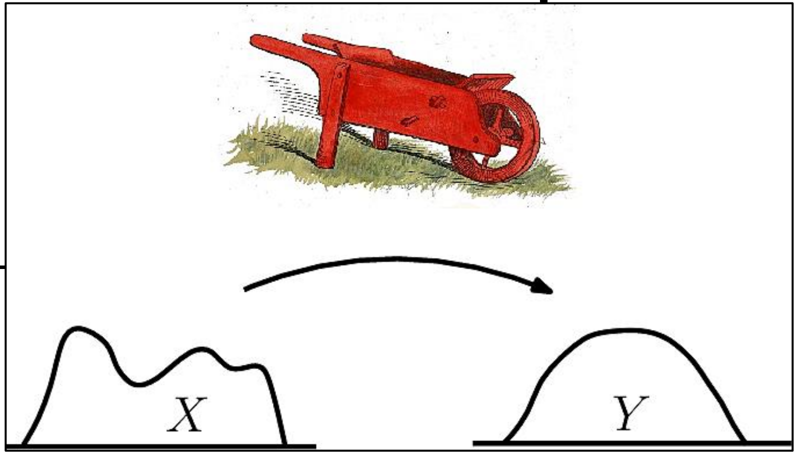
- Batch discrimination Salimans, Goodfellow et al. Improved techniques for training GANs 2015
- Use LeakyReLU in D
- Use BatchNorm (possible- InstanceNorm)
- Apply Spectral Norm
- Modified objective/loss?

Types of GAN losses - Wasserstein GAN

Discriminator

GAN $$\max_D E_{x\sim p_X}[\log D(x)] + E_{z\sim p_Z}[\log(1 - D(G(z)))]$$

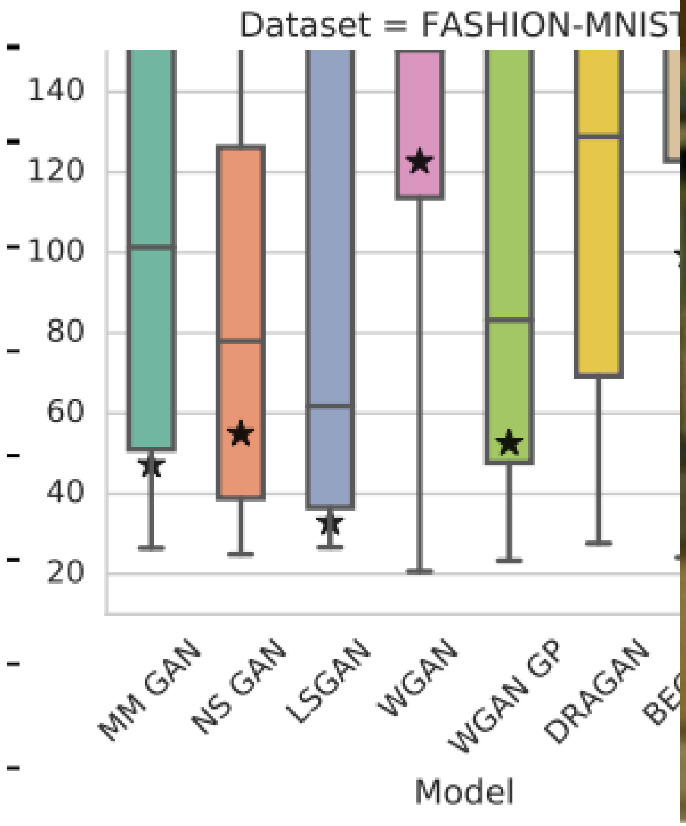
WGAN $$\max_D E_{x\sim p_X}[D(x)] - E_{z\sim p_Z}[D(G(z))]$$

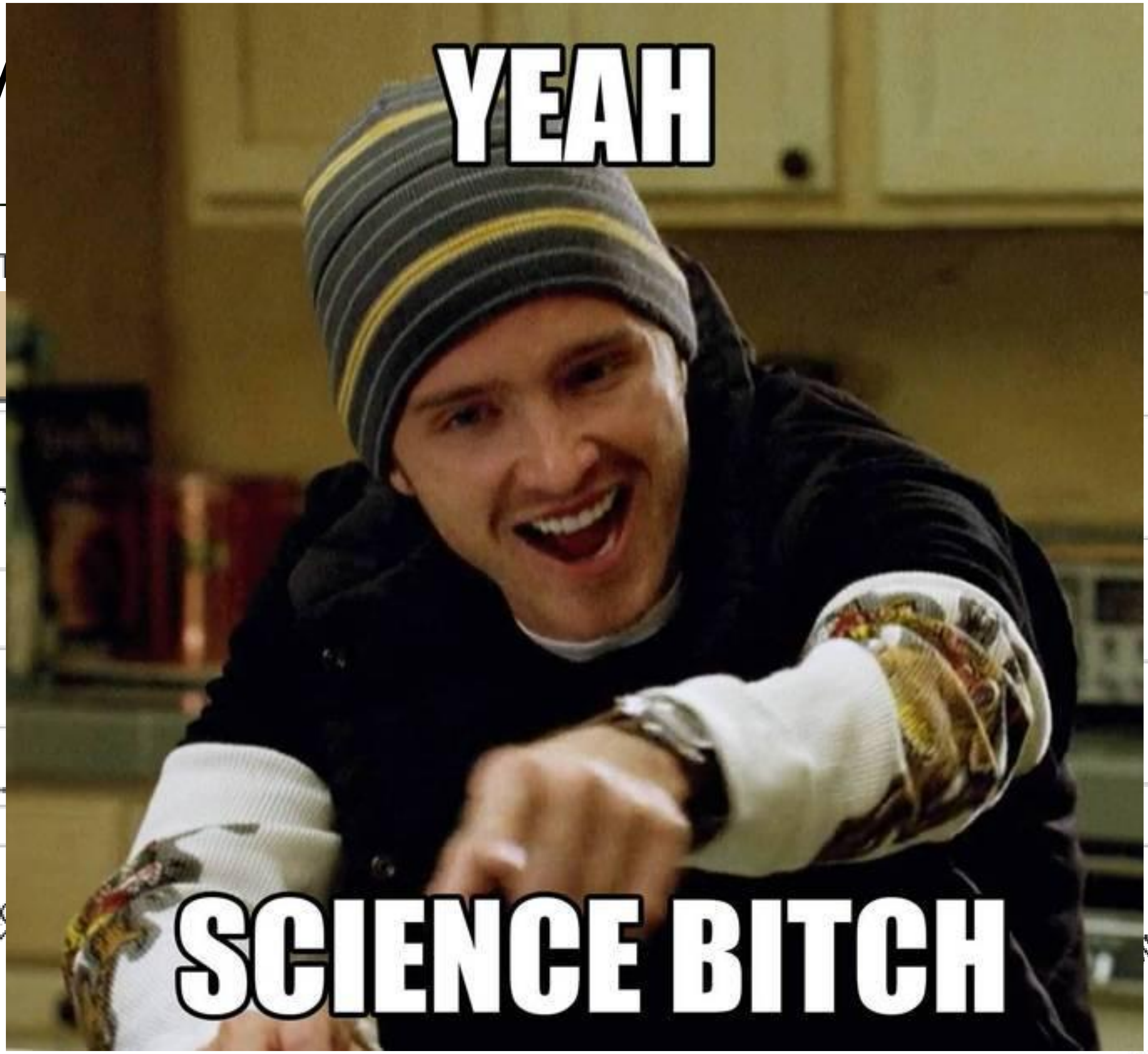


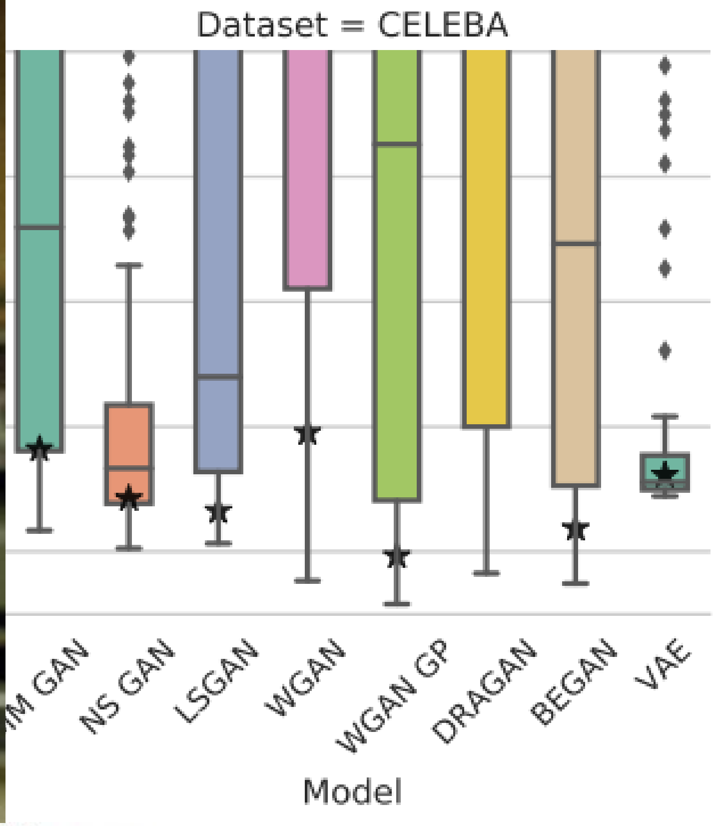
WGAN: minimize earth mover distance between p_X and $p_{G(Z)}$

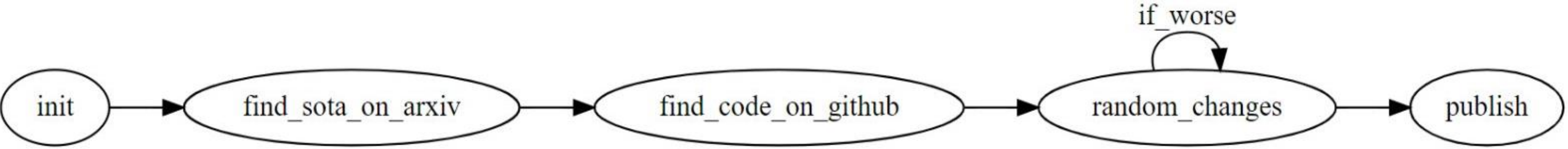
$$EM(p_X, p_{G(Z)}) = \inf_{\gamma \in \prod(p_X, p_{G(Z)})} E_{(x,y)\sim\gamma}[||x - y||]$$

Types of G



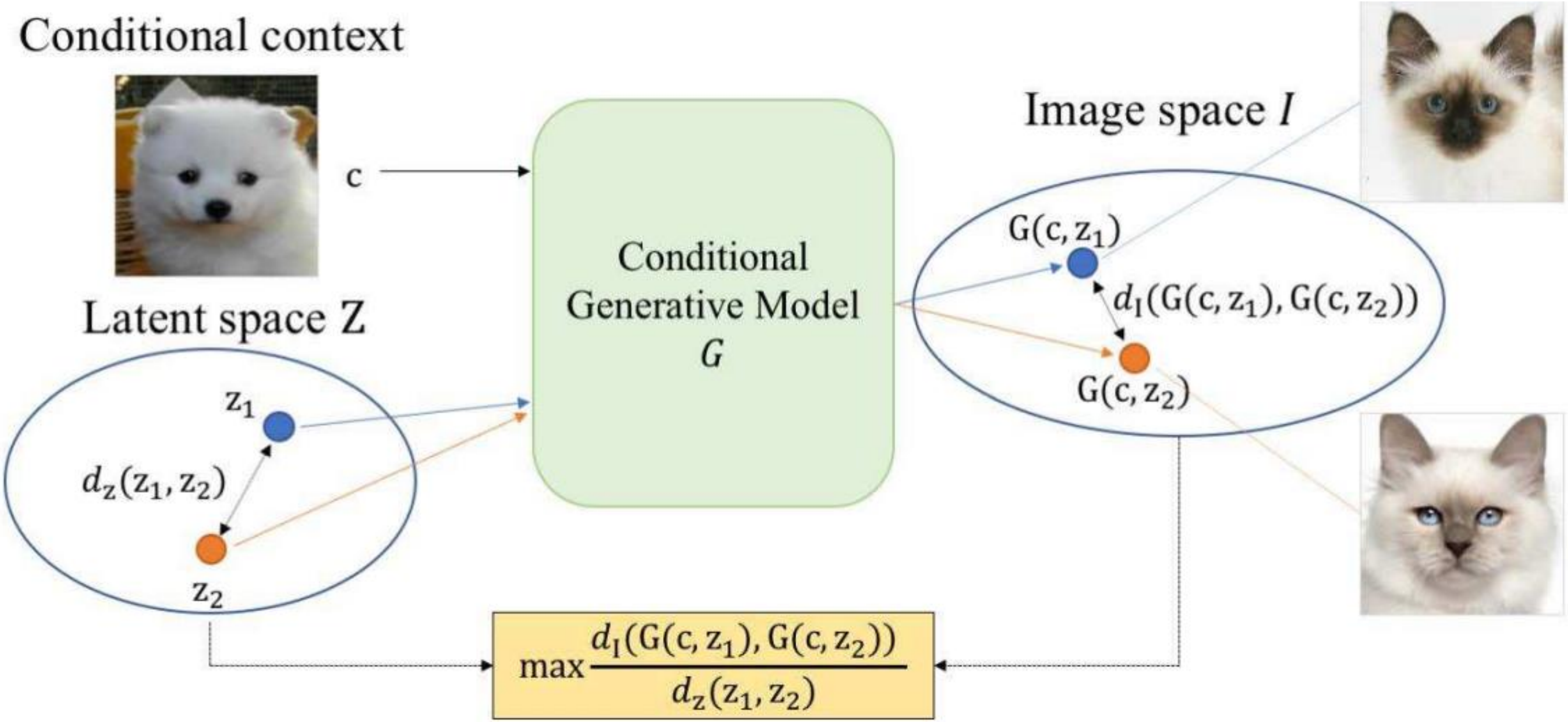






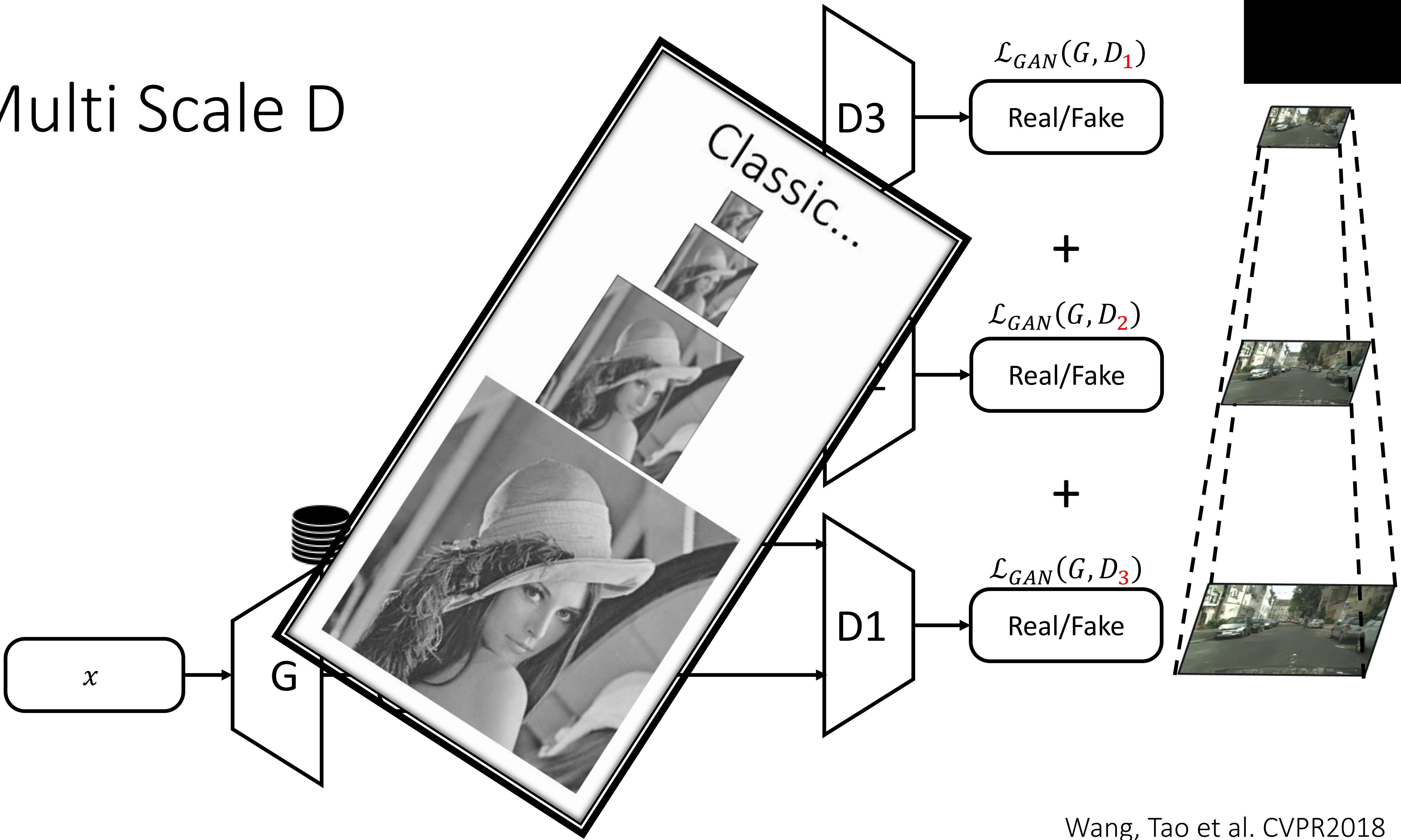
Mark Saroufim

Mode-Seeking GAN (Mao&Lee et al; CVPR’18)



WAIC

Multi Scale D



Wang, Tao et al. CVPR2018

Pix2Pix

Slide by Sefi Bell-Kligler & Akhiad Bercovich

Pix2Pix HD

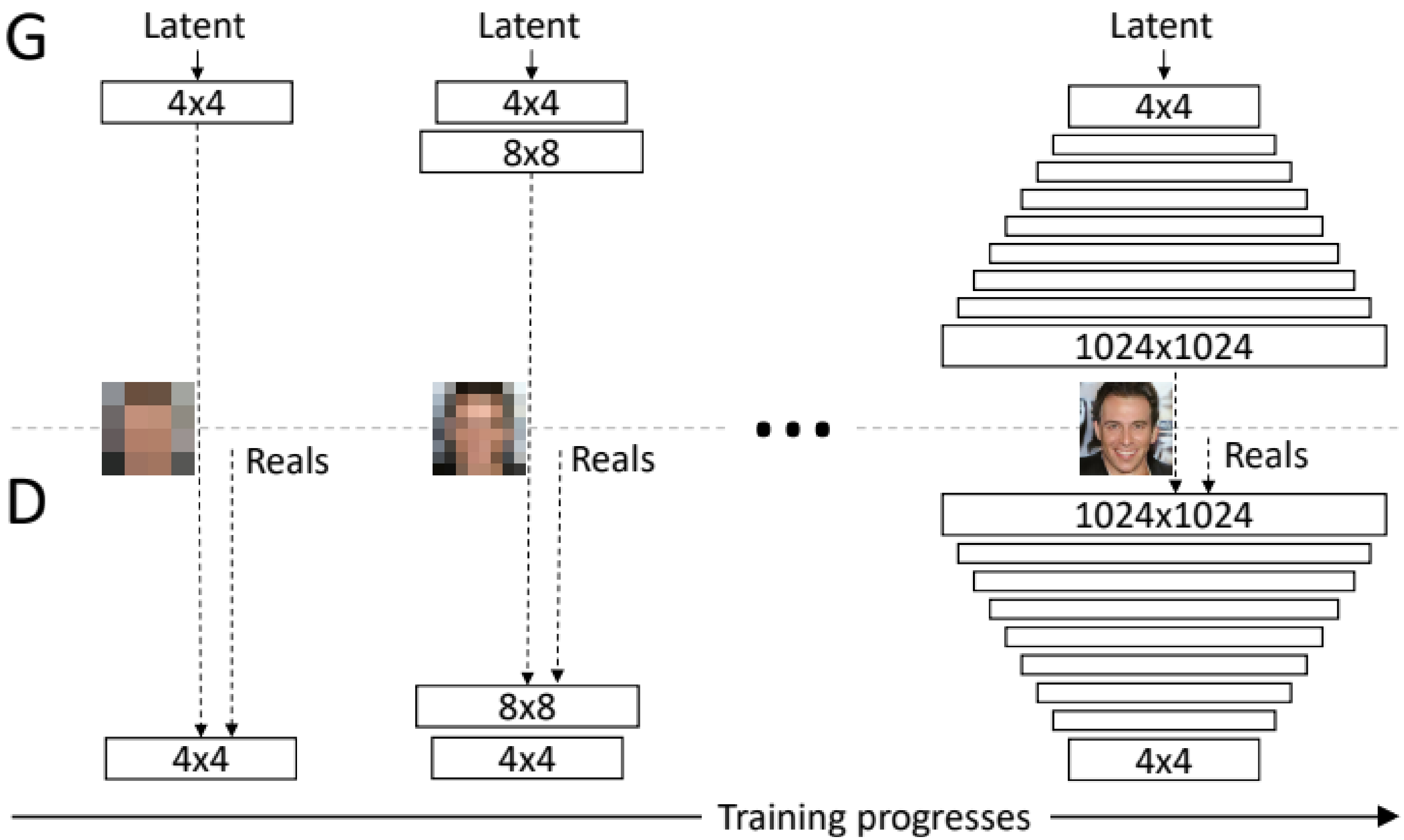
Wang, Tao et al. CVPR2018

Slide by Sefi Bell-Kligler & Akhiad Bercovich

Agenda

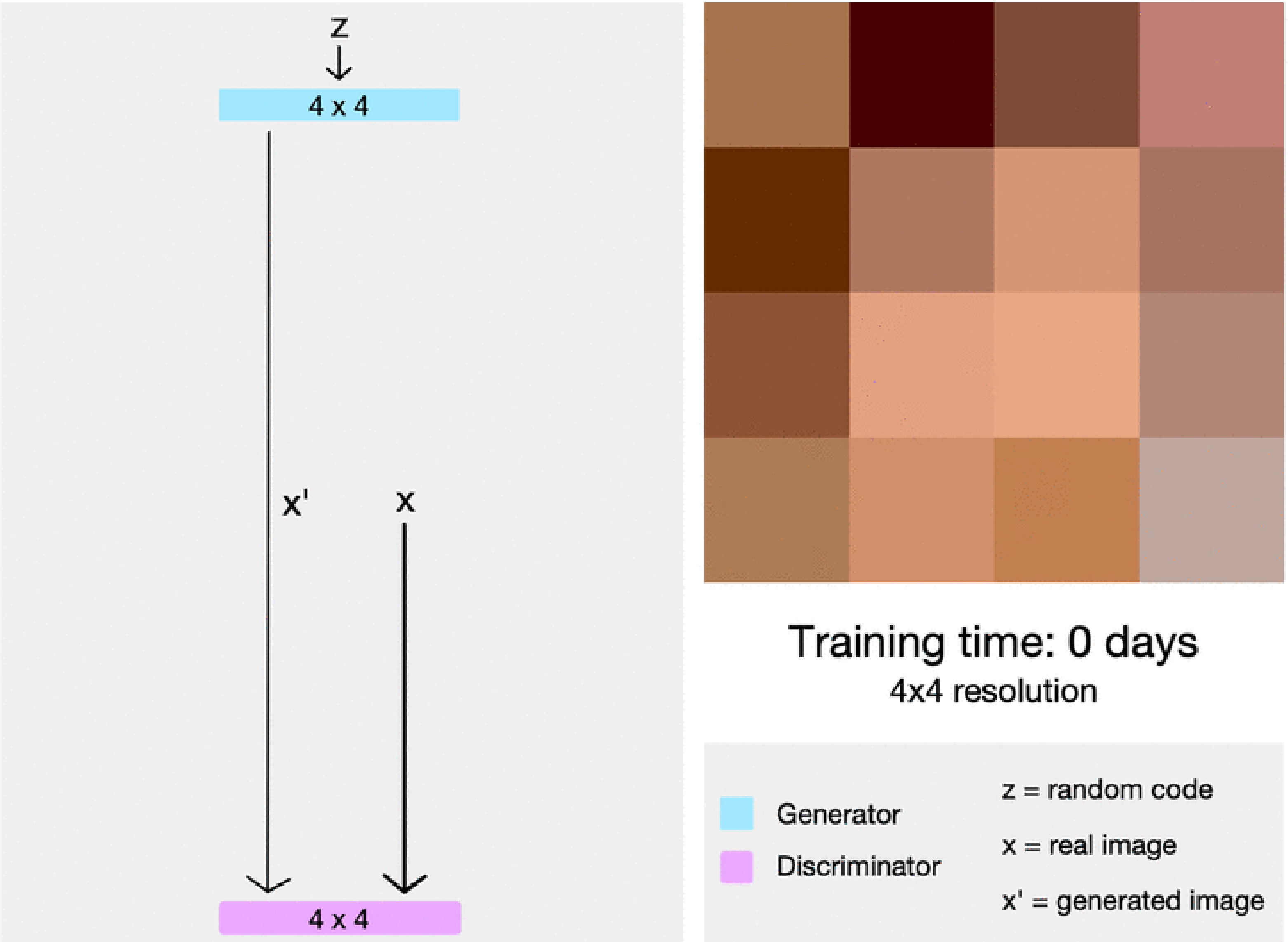
1. ~~Goal, motivation, Basic methods~~
2. ~~Introduction to GANs (basic setup, intuition, eval measures)~~
3. ~~Image to image (pix2pix, CycleGAN)~~
4. ~~Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)~~
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)

Progressive Grow



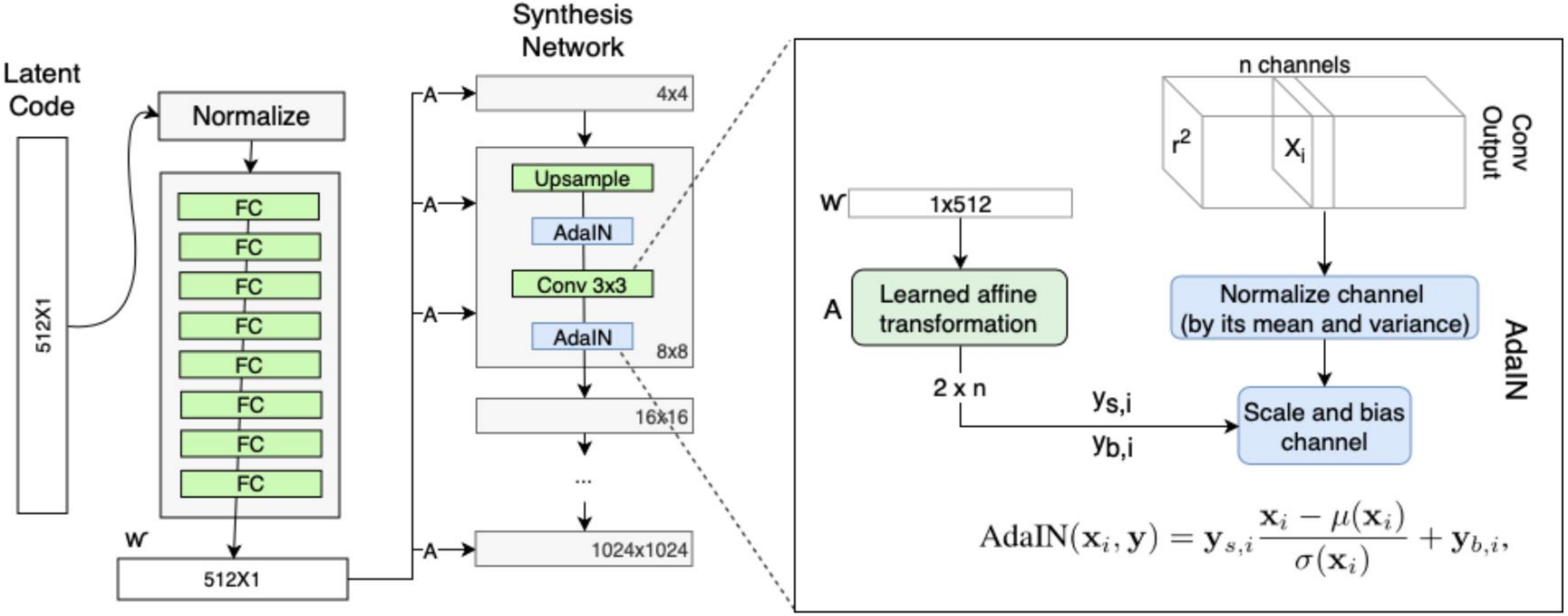
Progressive Growing of GAN, Karras et al., Feb2018

 Slide by Sefi Bell-Kligler & Akhiad Bercovich



Progressive Growing of GAN, Karras et al., Feb2018

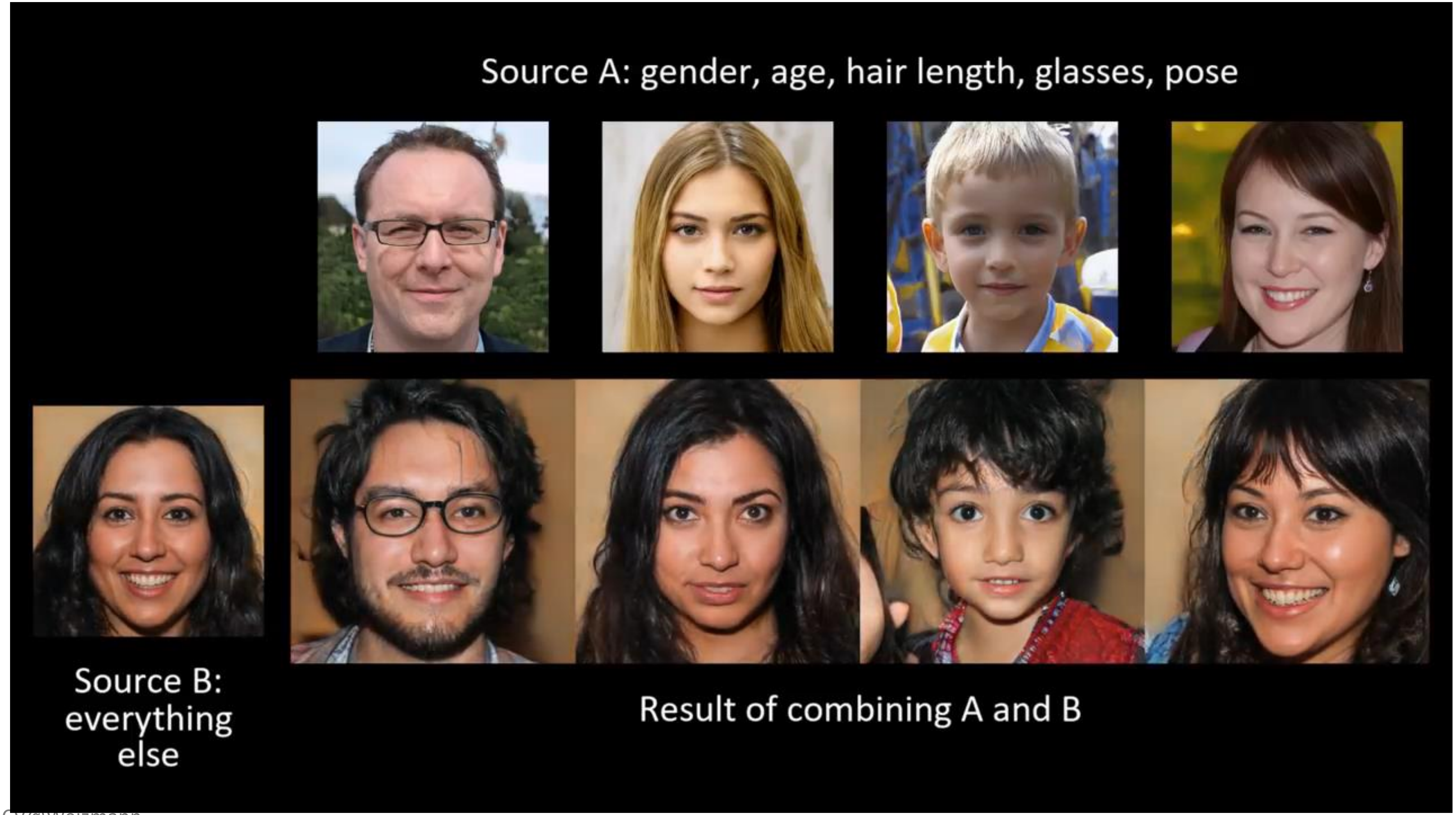
Style Modules (AdaIN)

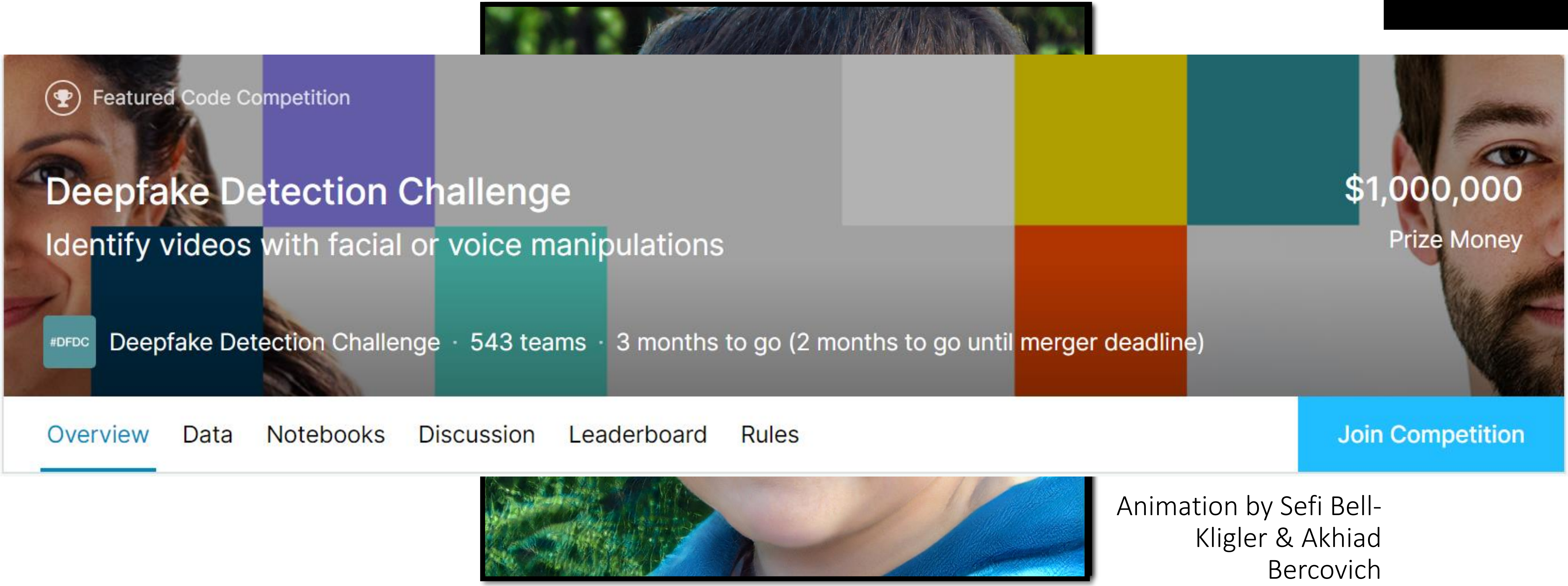


The generator's Adaptive Instance Normalization (AdaIN)

StyleGAN, Karras et al. NVIDIA 2019

Results


Source A: gender, age, hair length, glasses, pose
Source B: everything else
Result of combining A and B



Animation by Sefi Bell-Kligler & Akhiad Bercovich

https://thispersondoesnotexist.com/

https://whichfaceisreal.com/ https://thisxdoesnotexist.com/

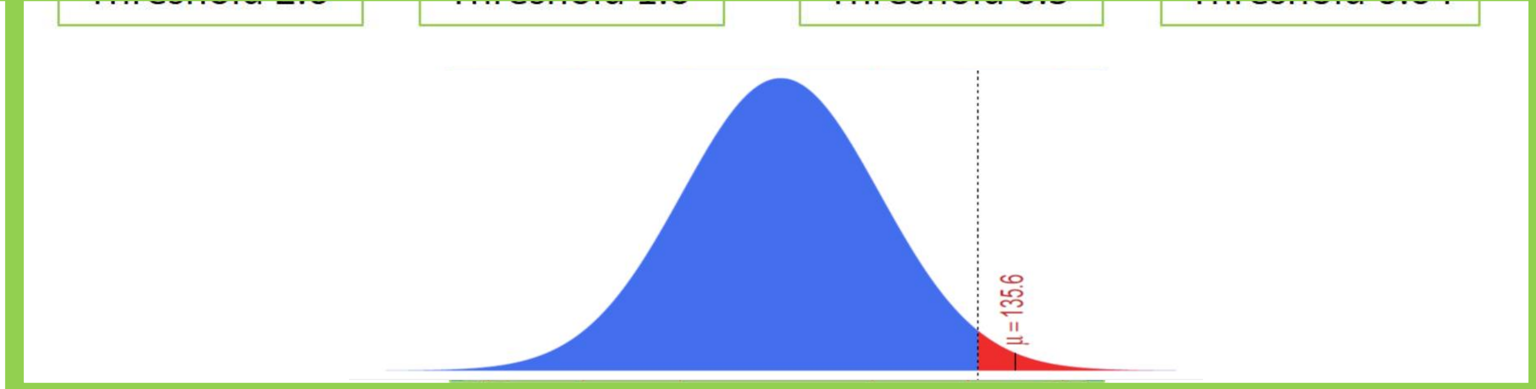
Agenda

1. ~~Goal, motivation, Basic methods~~
2. ~~Introduction to GANs (basic setup, intuition, eval measures)~~
3. ~~Image to image (pix2pix, CycleGAN)~~
4. ~~Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)~~
5. ~~Progressive GANs (PGGAN, StyleGAN)~~
6. $\boxed{\text{Scaling up GANs (BigGAN)}}$
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)

SAGAN, BigGAN- Scaling up GANs

- Class Conditional BatchNorm
- Self-Attention (both D and G)
- Split noise, and use skip connections for its chunks
- Bigger batches and number of channels per layer (and deeper)
- Spectral Norm (both D and G)
- Truncation Trick

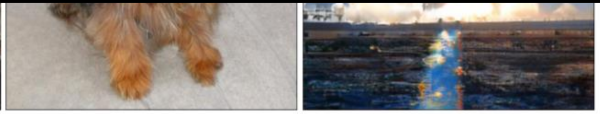
Batch	Ch.	Param (M)	Shared	Skip-z	Ortho.	Itr $\times 10^3$	FID	IS
256	64	81.5	SA-GAN Baseline			1000	18.65	52.52
512	64	81.5	✗	✗	✗	1000	15.30	58.77($\pm$1.18)
1024	64	81.5	✗	✗	✗	1000	14.88	63.03($\pm$1.42)
2048	64	81.5	✗	✗	✗	732	12.39	76.85($\pm$3.83)
2048	96	173.5	✗	✗	✗	295($\pm$18)	9.54($\pm$0.62)	92.98($\pm$4.27)
2048	96	160.6	✓	✗	✗	185($\pm$11)	9.18($\pm$0.13)	94.94($\pm$1.32)
2048	96	158.3	✓	✓	✗	152($\pm$7)	8.73($\pm$0.45)	98.76($\pm$2.84)
2048	96	158.3	✓	✓	✓	165($\pm$13)	8.51($\pm$0.32)	99.31($\pm$2.10)
2048	64	71.3	✓	✓	✓	371($\pm$7)	10.48($\pm$0.10)	86.90($\pm$0.61)



Impressive results?

StyleGAN branch	BigGAN Branch
One domain (eg. Faces)	Class conditional (I... et)
Structured data	Ver... ata
Progressive grow	...Attention
Amazing results	...ome of the results are amazing

Much harder

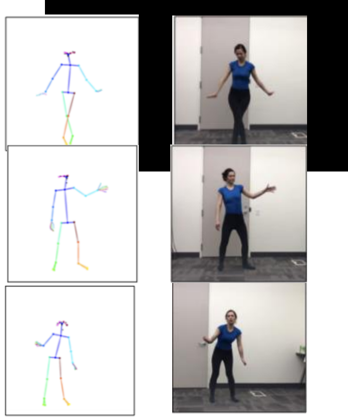


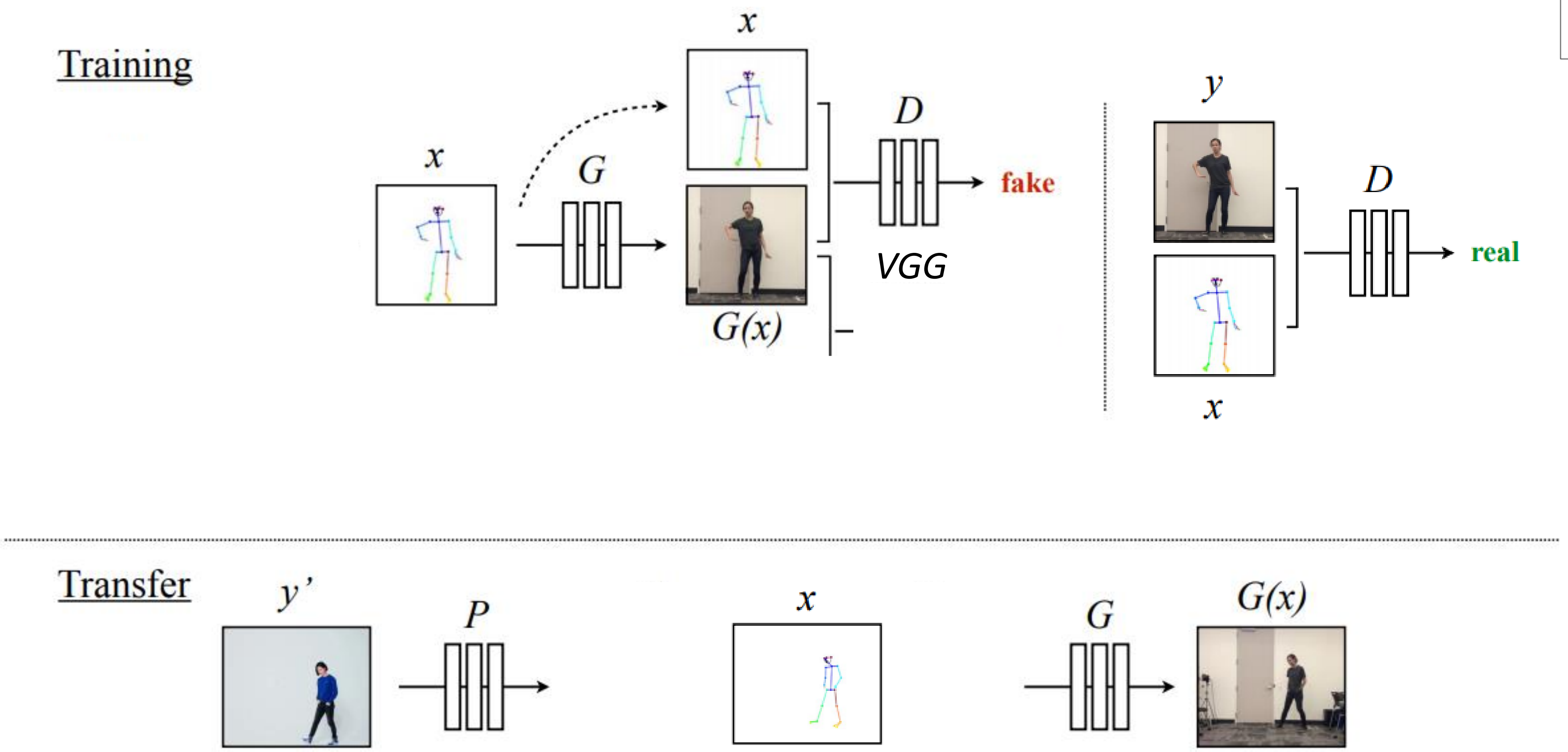
Brock et al. 2019

Agenda

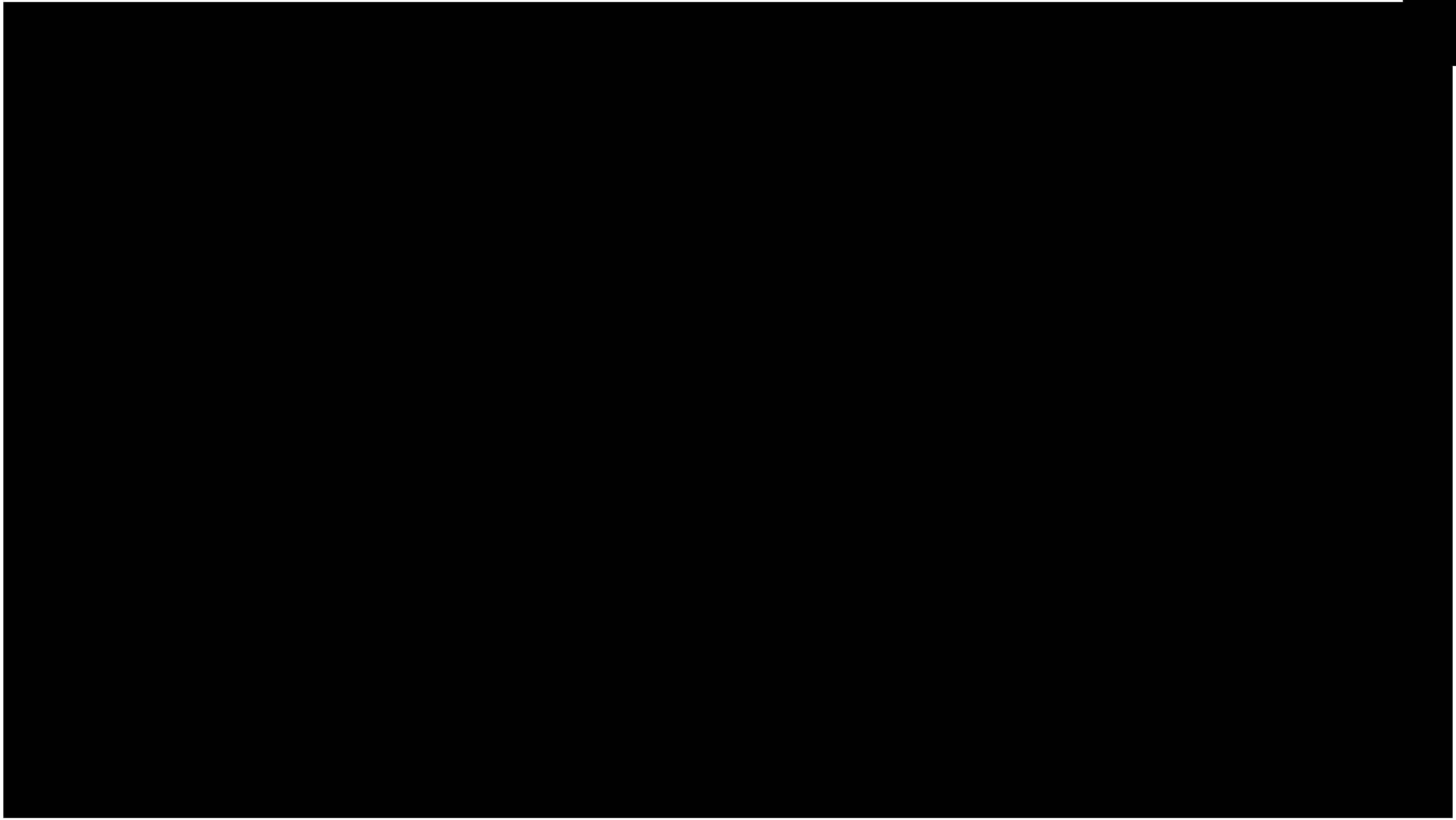
1. ~~Goal, motivation, Basic methods~~
2. ~~Introduction to GANs (basic setup, intuition, eval measures)~~
3. ~~Image to image (pix2pix, CycleGAN)~~
4. ~~Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)~~
5. ~~Progressive GANs (PGGAN, StyleGAN)~~
6. ~~Scaling up GANs (BigGAN)~~
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)

Everybody Dance Now

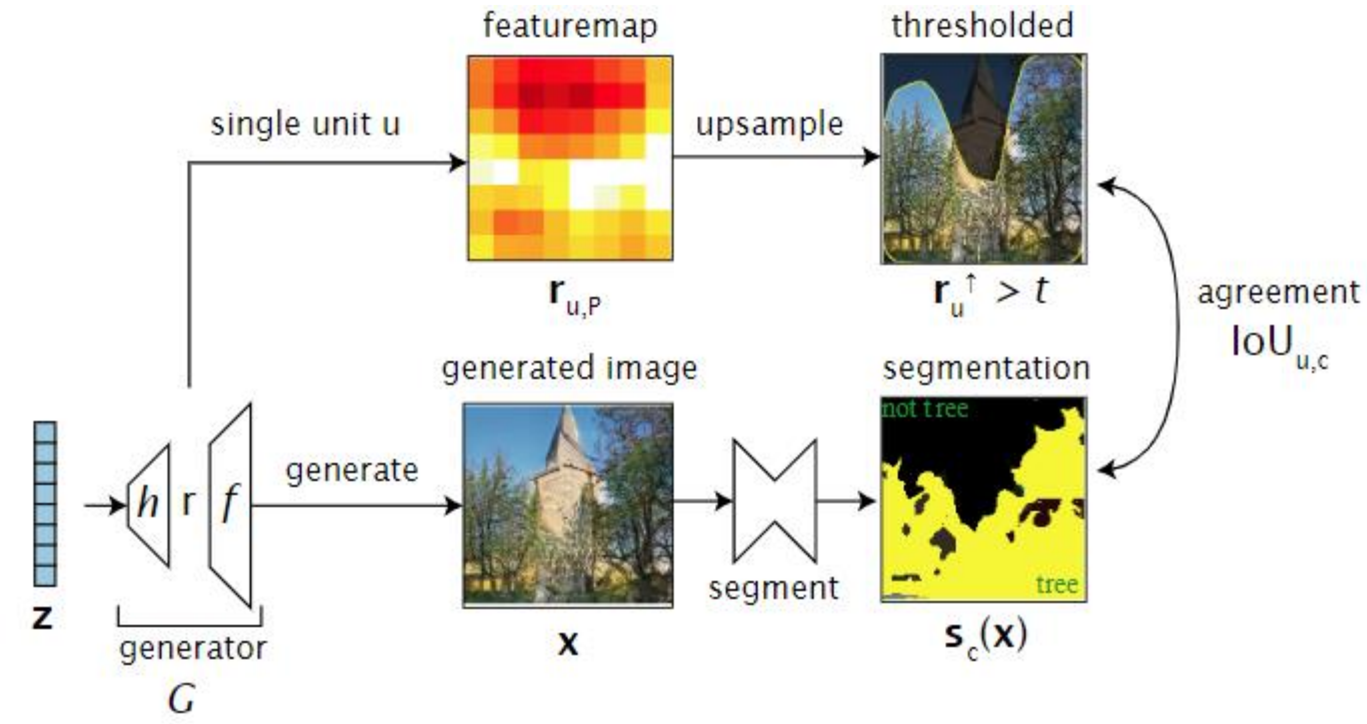


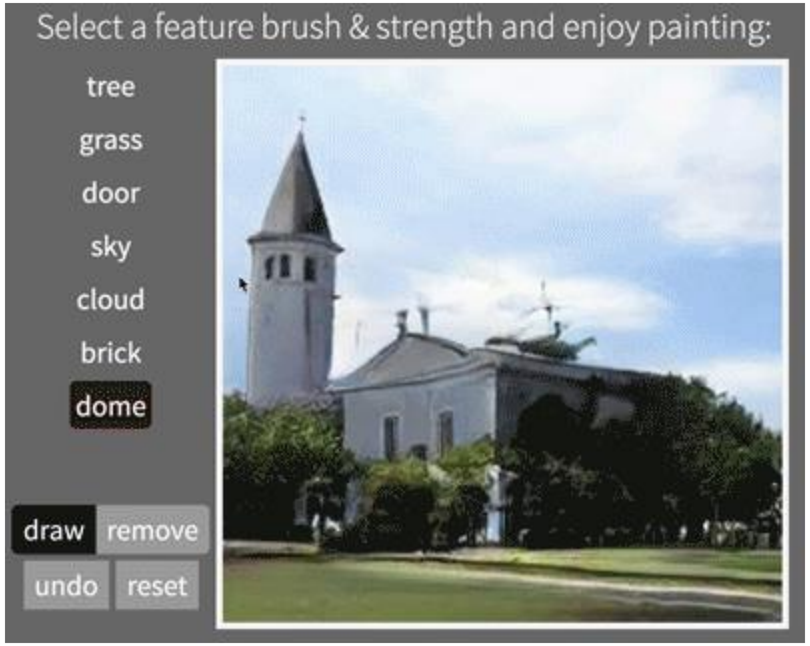


Chan et al., Aug2018



GAN Dissection





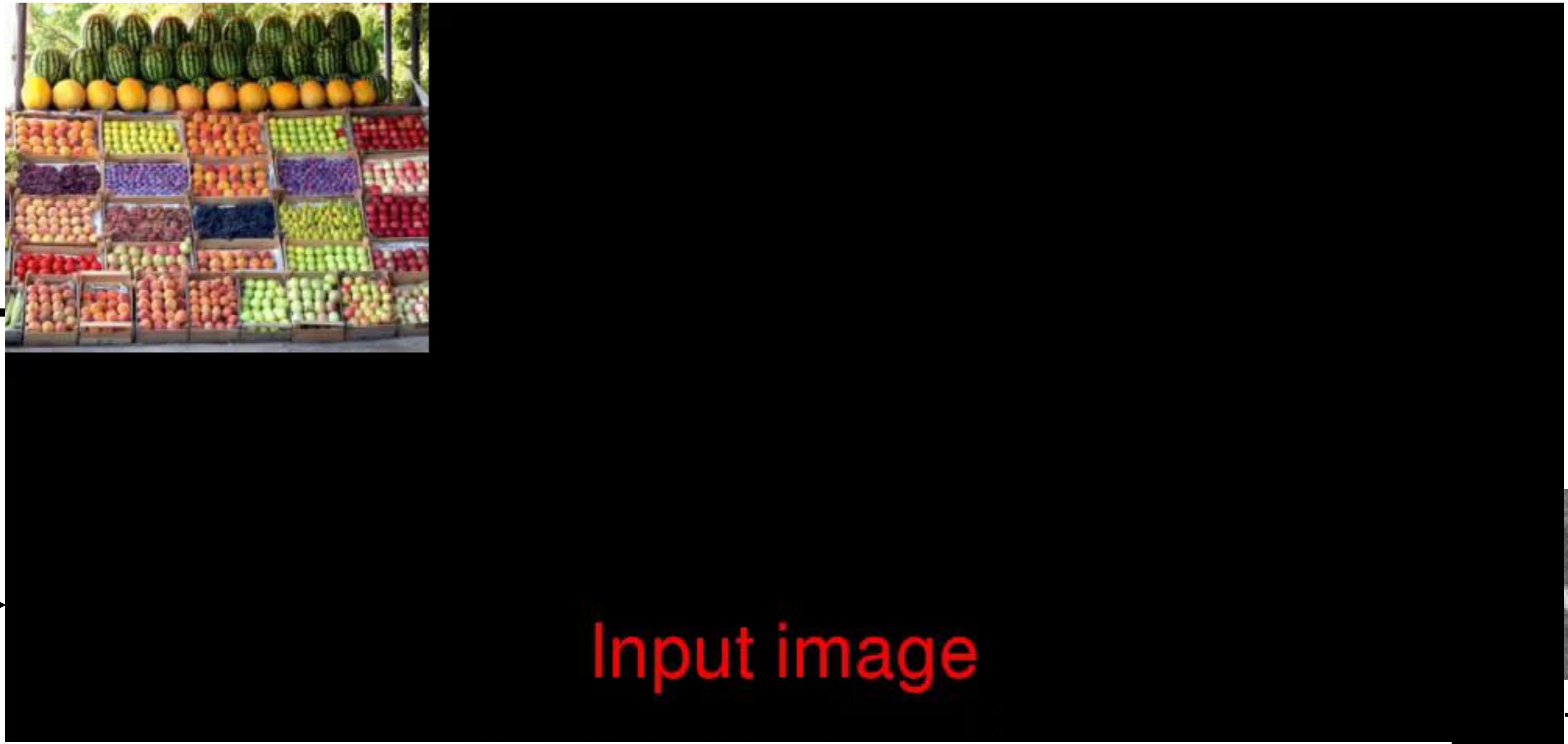
David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, Antonio Torralba

http://gandissect.res.ibm.com/ganpaint.html?project=churchoutdoor&layer=layer4

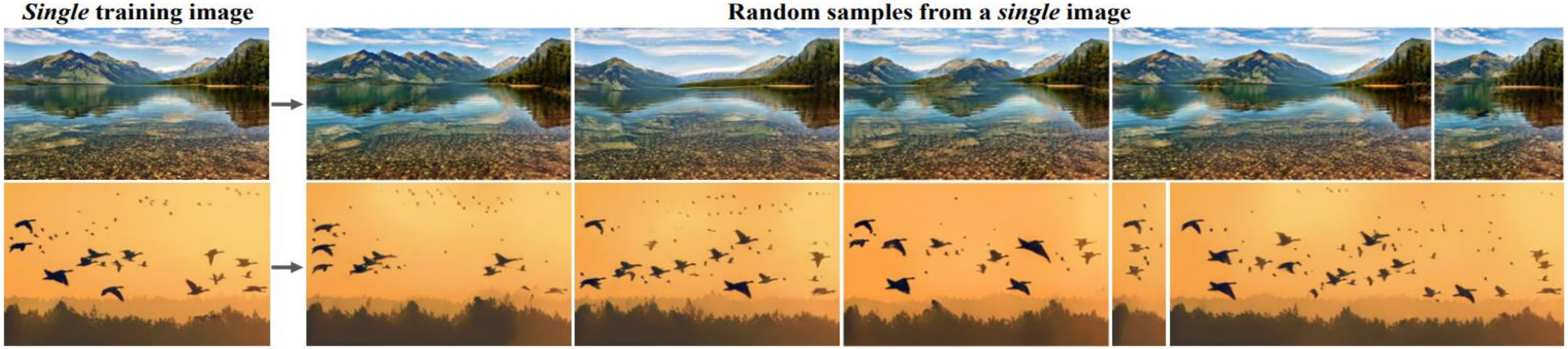
Training a GAN on a single image

InGAN (Shocher, Bagon, Isola, Irani)

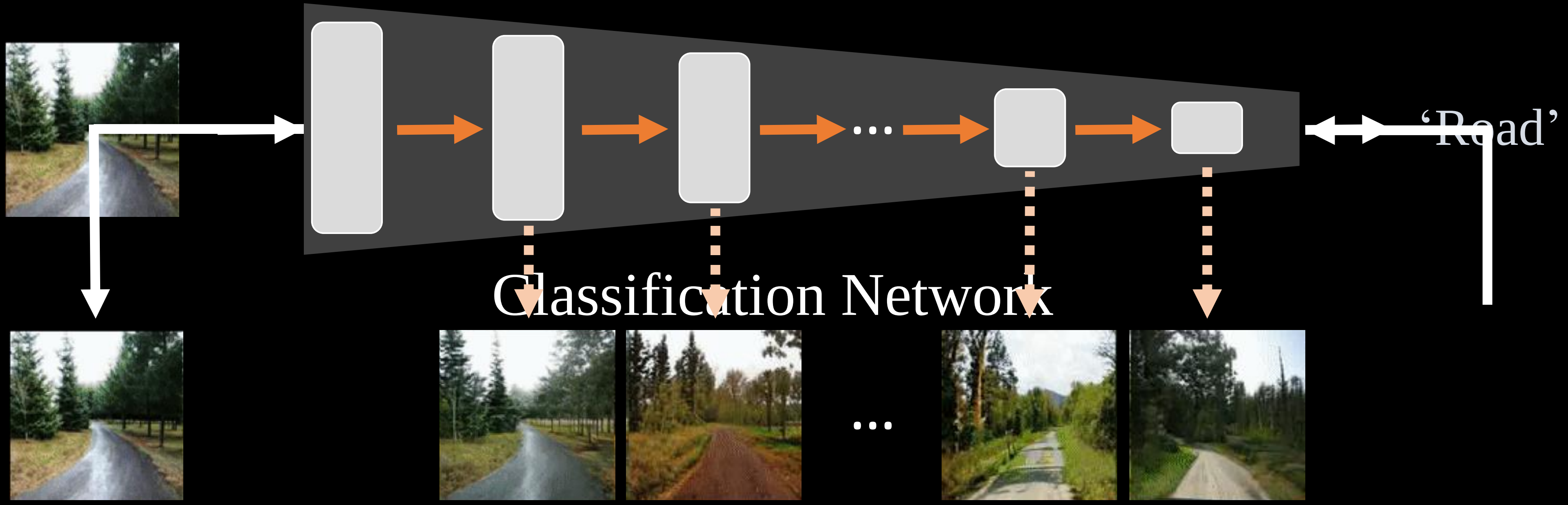
Input/noise



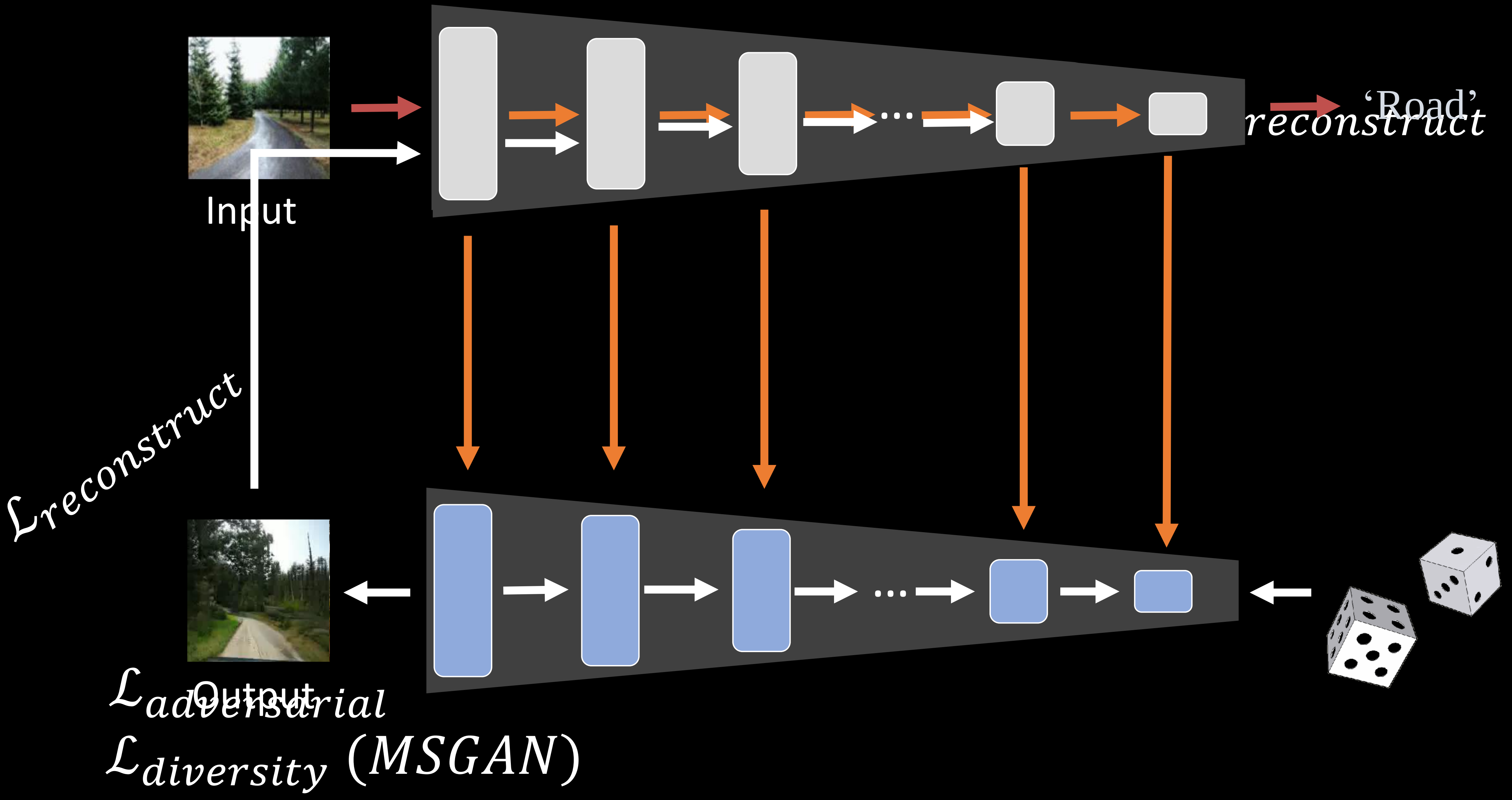
True/False map

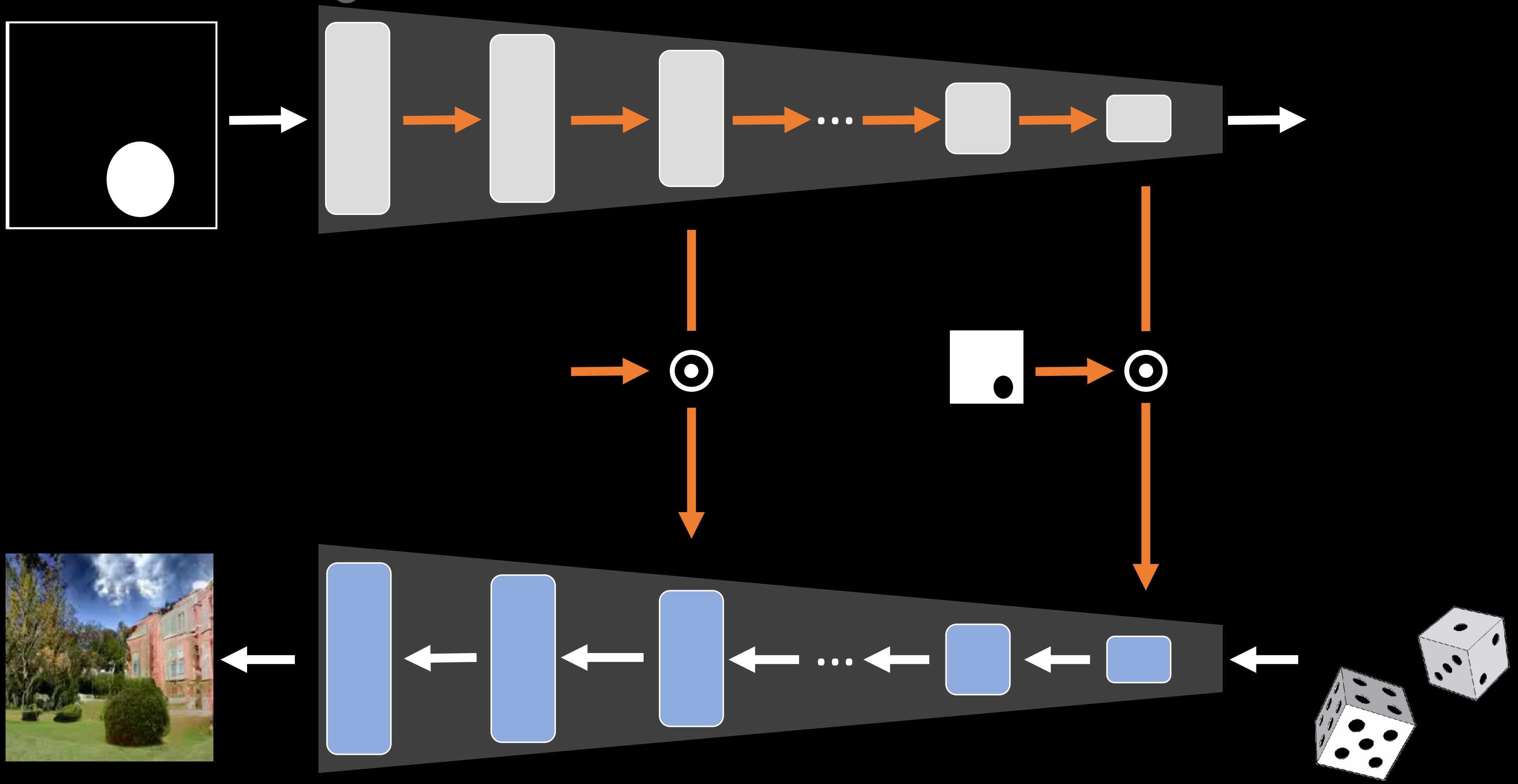


SinGAN (Rott-Shaham, Dekel, Michaeli)



Shocher, Gandelsman, Mosseri, Yarom, Irani, Freeman, Dekel; Semantic Pyramid for Image Generation; CVPR'2020


Input
'Road'
$\mathcal{L}_{reconstruct}$
$\mathcal{L}_{reconstruct}$
Output
$\mathcal{L}_{adversarial}$
$\mathcal{L}_{diversity}$ $(MSGAN)$



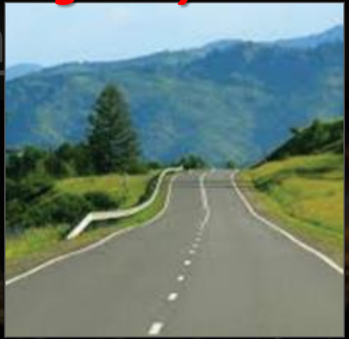
Applications

Semantic Image Compositi

Out of distribution referen

Image re-labeling

Original label: ***Highway***



New label: ***Desert road***

Original label: ***Mountain***

New label: ***Volcano***

Agenda

1. ~~Goal, motivation, Basic methods~~
2. ~~Introduction to GANs (basic setup, intuition, eval measures)~~
3. ~~Image to image (pix2pix, CycleGAN)~~
4. ~~Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)~~
5. ~~Progressive GANs (PGGAN, StyleGAN)~~
6. ~~Scaling up GANs (BigGAN)~~
7. ~~Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)~~

Tricks and Techniques

- Batch discrimination
- Label smoothing
- Use LeakyReLU in D
- Use BatchNorm
- Consider InstanceNorm
- Apply Spectral Norm
- Playing with G:D ratio is bad for you
- Add small noise (1/256) to real examples (because D may recognize quantization)
- Consider WGAN
- Consider LSGAN
- Multiscale D
- VGG loss
- Strict similarity
- Self attention
- Noise skip connections
- Truncation trick
- Conditional BatchNorm
- Mode Seeking GAN

Agenda

1. Goal, motivation, Basic methods
2. Introduction to GANs (basic setup, intuition, eval measures)
3. Image to image (pix2pix, CycleGAN)
4. Improve GAN performance (losses, MSGAN, tricks, pix2pixHD)
5. Progressive GANs (PGGAN, StyleGAN)
6. Scaling up GANs (BigGAN)
7. Special stuff (GAN Dissection, Single Image, Dance transfer, Semantic Pyramid)

Thanks

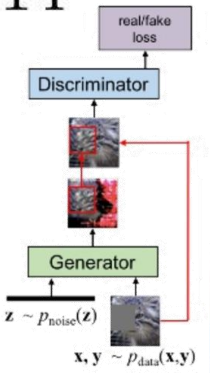
2014

2015

2016

2017

2018



Benny GANs

This week's Tutorial: Akhiad Bercovich  LSTM

Next week's Lecture: Tali Dekel  Self-Supervision

Next week's Tutorial: Niv Granot  Non-adversarial Generative Models