Generative Models; Vision & Language

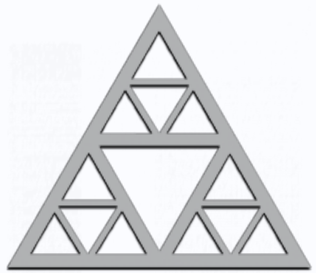
Gül Varol

IMAGINE team, École des Ponts ParisTech

gul.varol@enpc.fr

http://imagine.enpc.fr/~varolg/

@RecVis, 19.11.2024



With many slides from: L. Lazebnik, F.-F. Li, J. Johnson, S. Yeung, F. Fleuret and others

Advanced topics in vision

Previously

1) Instance-level recognition
2) Camera geometry, image processing **[J. Ponce]**
3) Efficient visual search
4) Introduction to deep learning
5) Neural networks for visual recognition
6) Object detection, Segmentation, Human pose

Today

7) Generative models; Vision & language

Next weeks

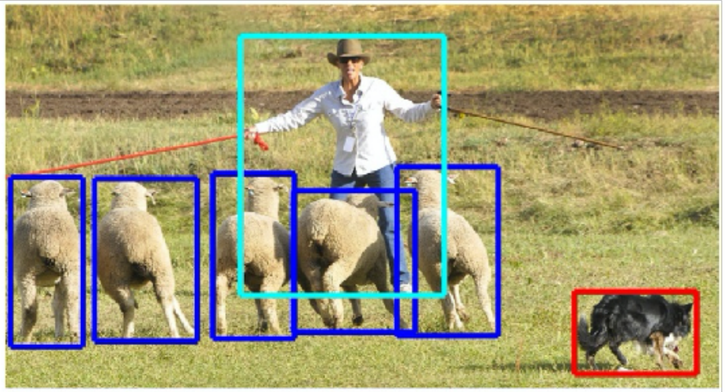
8) Videos (Nov 26) **[Cordelia Schmid]**
9) Vision for robotics (Dec 3) **[Ivan Laptev]**
10) 3D (Dec 10) **[Mathieu Aubry]**

Recap: Visual recognition so far

Image Classification

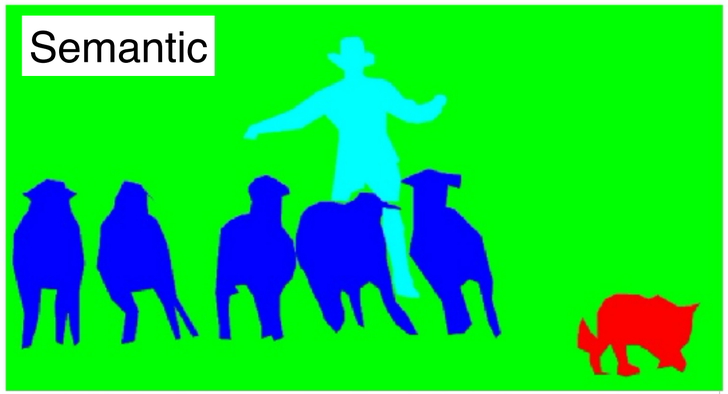
- Class labels

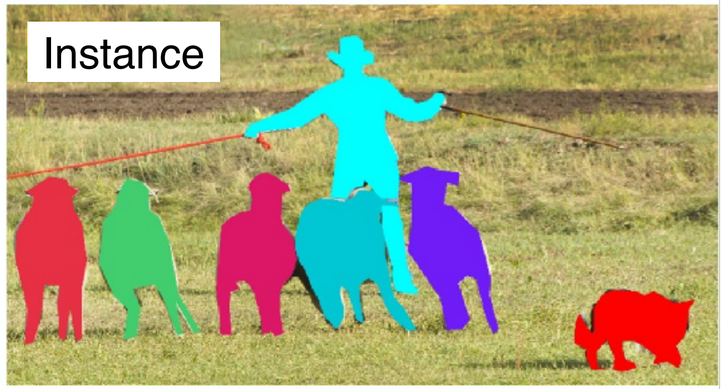
Object Detection



- Bounding box

Segmentation

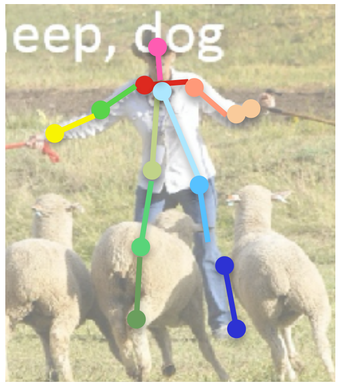




Panoptic,
Promptable …

- Pixel-wise labels

Human Pose



- Structured output

- **Objects in images**
- **Symbolic object categories**

Lin *et al.* “Microsoft COCO: Common Objects in Context”

Today

Analysis

- Class labels
- Bounding box
- Pixel-wise labels
- Structured output

Part 1: Generative models

- Image

Synthesis

Today

Analysis

- Class labels
- Bounding box
- Pixel-wise labels
- Structured output

- **Symbolic (object) categories**

Part 1: Generative models

- Image

Synthesis

Part 2: Vision & Language

- **Free-form text (language)**

Agenda

1. Generative neural networks

- VAE: Variational autoencoders
- GAN: Generative adversarial networks
- Diffusion models

2. Vision & language

- Text-to-image retrieval
- Text-to-image generation
- Image captioning

Agenda

1. Generative neural networks

- VAE: Variational autoencoders
- GAN: Generative adversarial networks
- Diffusion models

2. Vision & language

- Text-to-image retrieval
- Text-to-image generation
- Image captioning

Generative models

- Disclaimer: "Generative" is an overloaded term.
  - In this lecture's context, we are concerned with generating (synthesizing) images with neural networks.

- A lot of buzz around a new term "Generative AI" (same thing)
  = models capable of generating media,
  typically text or images
  based on input text/prompt.

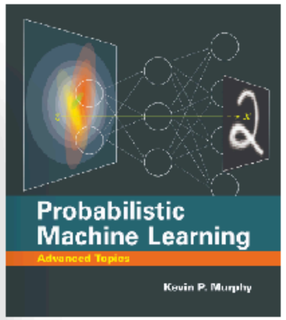
Further reading:

https://probml.github.io/pml-book/book2.html

probml.github.io/pml-book/book2.html

Probabilistic Machine Learning: Advanced Topics

by Kevin Patrick Murphy.
MIT Press, 2023.



Key links

- Short table of contents
- Long table of contents
- Preface
- Draft pdf of the main book, 2023-08-15. CC-BY-NC-ND license. (Please cite the official reference below.)
- Supplementary material
- Issue tracker.
- Code to reproduce most of the figures
- Acknowledgements
- Endorsements

If you use this book, please be sure to cite

```
@book{pml2Book,
author = "Kevin P. Murphy",
title = "Probabilistic Machine Learning: Advanced Topics",
publisher = "MIT Press",
year = 2023,
url = "http://probml.github.io/book2"
}
```

Downloads since 2022-02-28. downloads 158k

Why Generative Models?

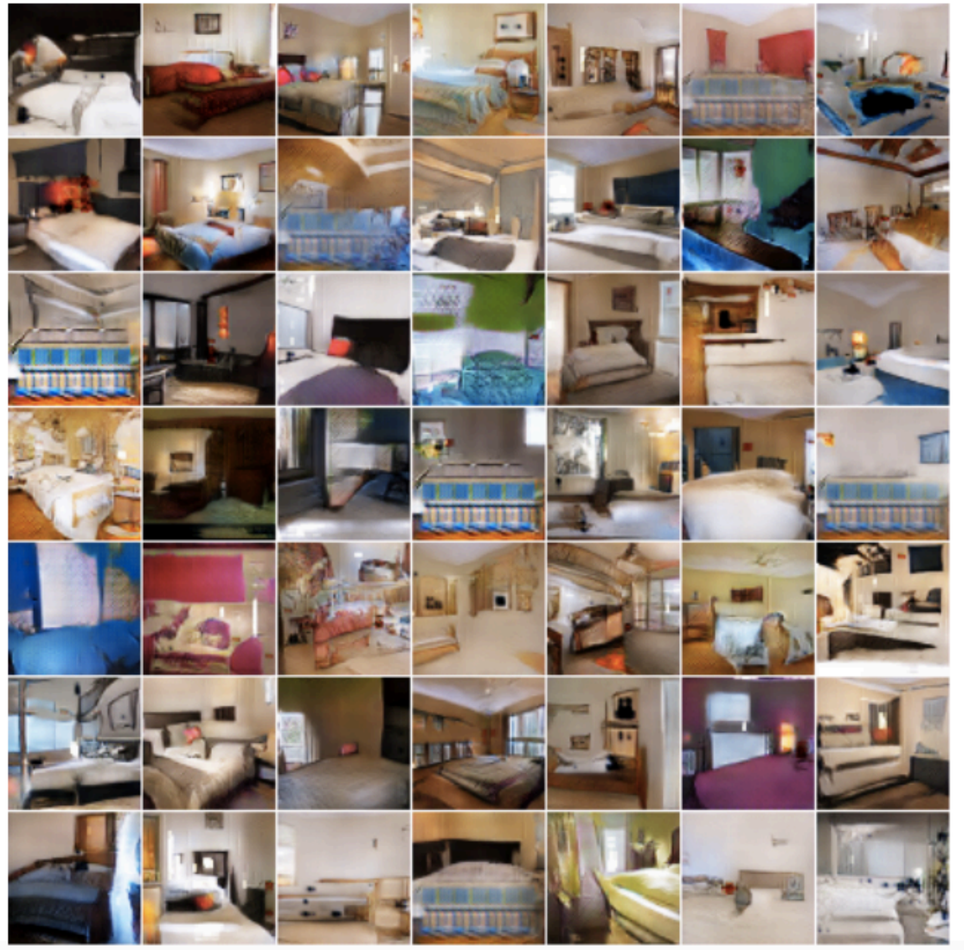
- Creativity/arts, super-resolution,…
- Can create synthetic data for training
- Can provide useful feature representations
- Data compression
- …

Generative (image synthesis) tasks

- Unconditional
- Conditioned on class label
- Conditioned on image
- Conditioned on text
- …

Generative tasks

- **Unconditional** generation: learn to sample from the distribution represented by the training set
  - *Unsupervised learning* task



Generative tasks

- Generation conditioned on **class label**

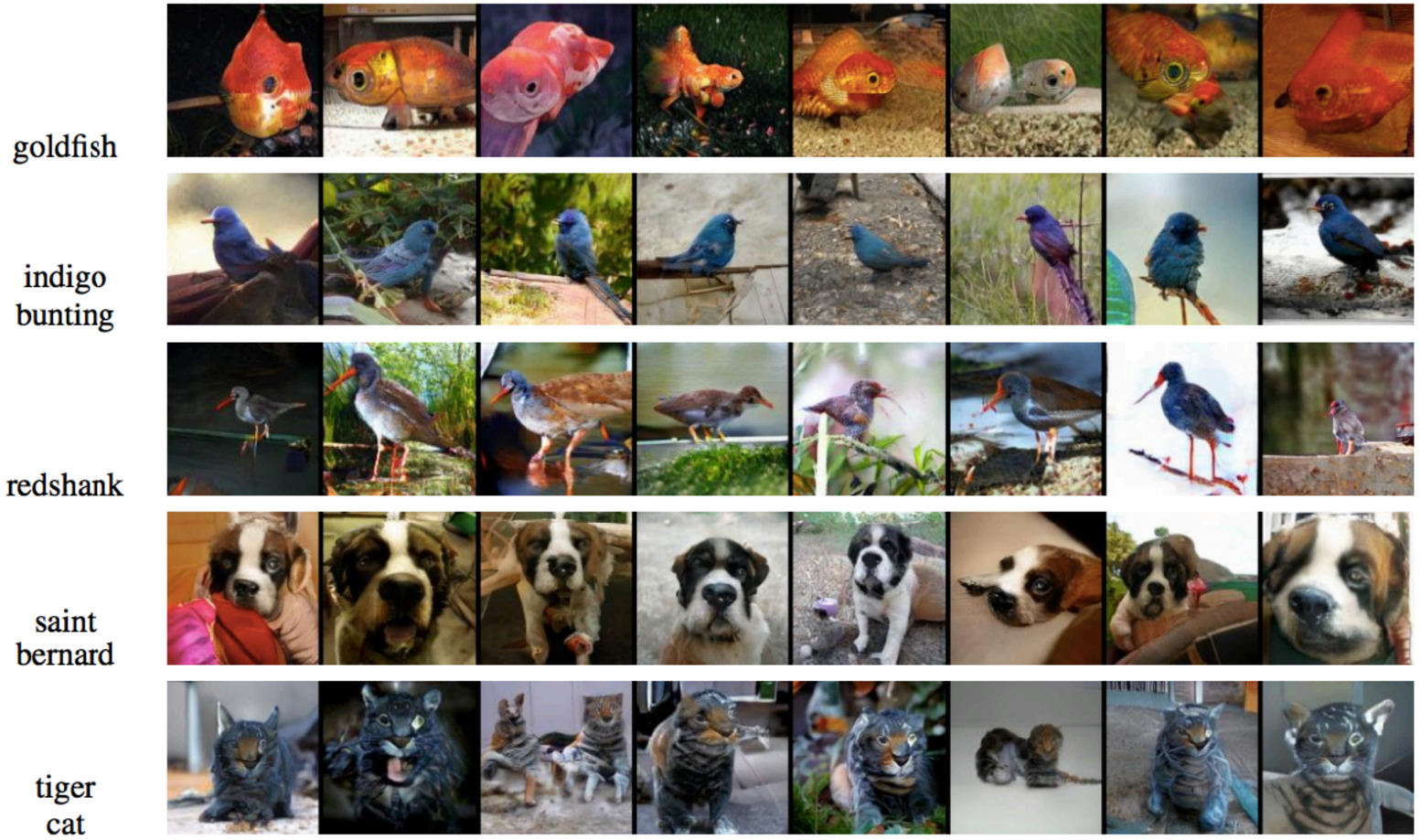
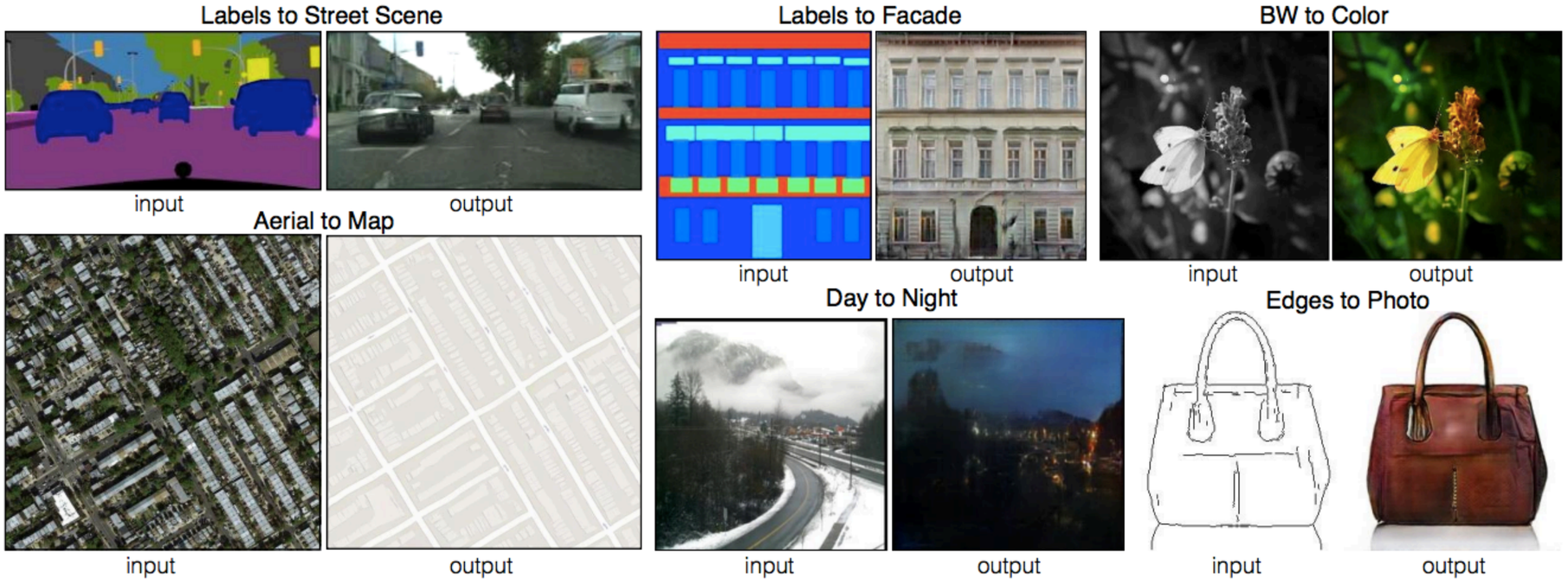


Figure source

Generative tasks

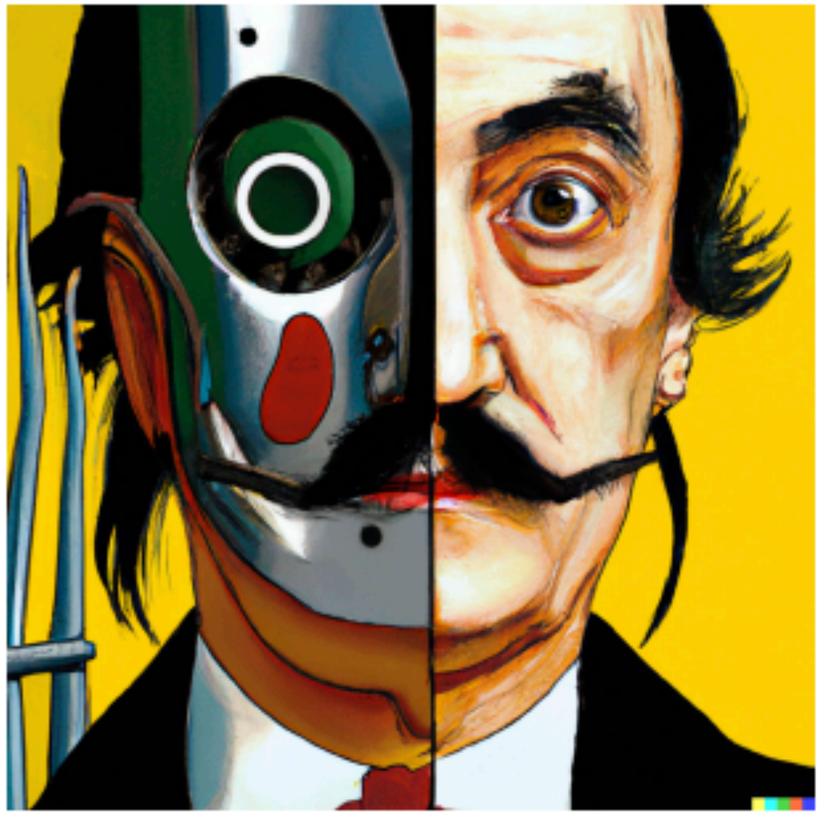
- Generation conditioned on **image**
  or *image-to-image translation*



P. Isola, J.-Y. Zhu, T. Zhou, A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, CVPR 2017

Generative tasks

- Generation conditioned on **text**



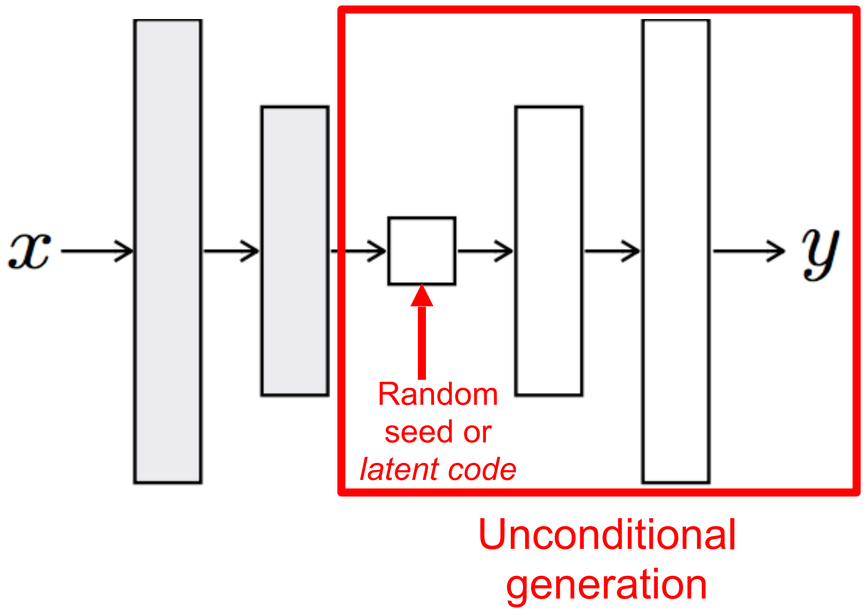
Vibrant portrait painting of Salvador Dali with a robotic half face

A close up of a handpalm with leaves growing from it

A. Ramesh et al. Hierarchical text-conditional image generation with CLIP latents. 2022

Designing a network for generative tasks

1. We need an architecture that can generate an image
   - Recall upsampling architectures for dense prediction



Designing a network for generative tasks

1. We need an architecture that can generate an image
   - Recall upsampling architectures for dense prediction

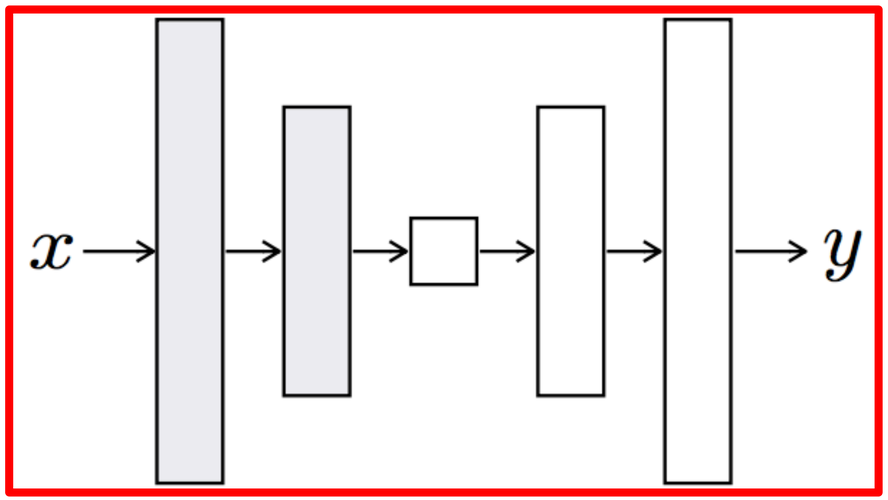


Image-to-image translation

Designing a network for generative tasks

1. We need an architecture that can generate an image
   - Recall upsampling architectures for dense prediction

2. We need to design the right loss function and training framework

Learning to sample

- Given training data, generate new samples from same distribution

Training data $x \sim p_{\text{data}}$

Generated samples $x \sim p_{\text{model}}$

We want to learn p_{model} that matches p_{data}

Adapted from Stanford CS231n

Agenda

1. Generative neural networks

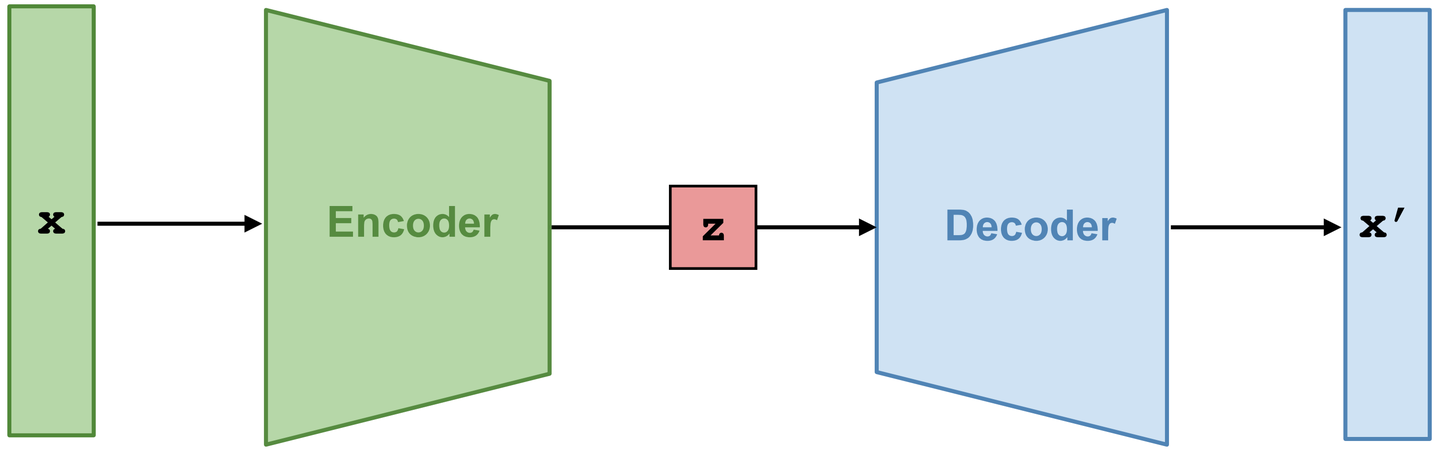
- VAE: Variational autoencoders
- GAN: Generative adversarial networks
- Diffusion models

2. Vision & language

- Text-to-image retrieval
- Text-to-image generation
- Image captioning

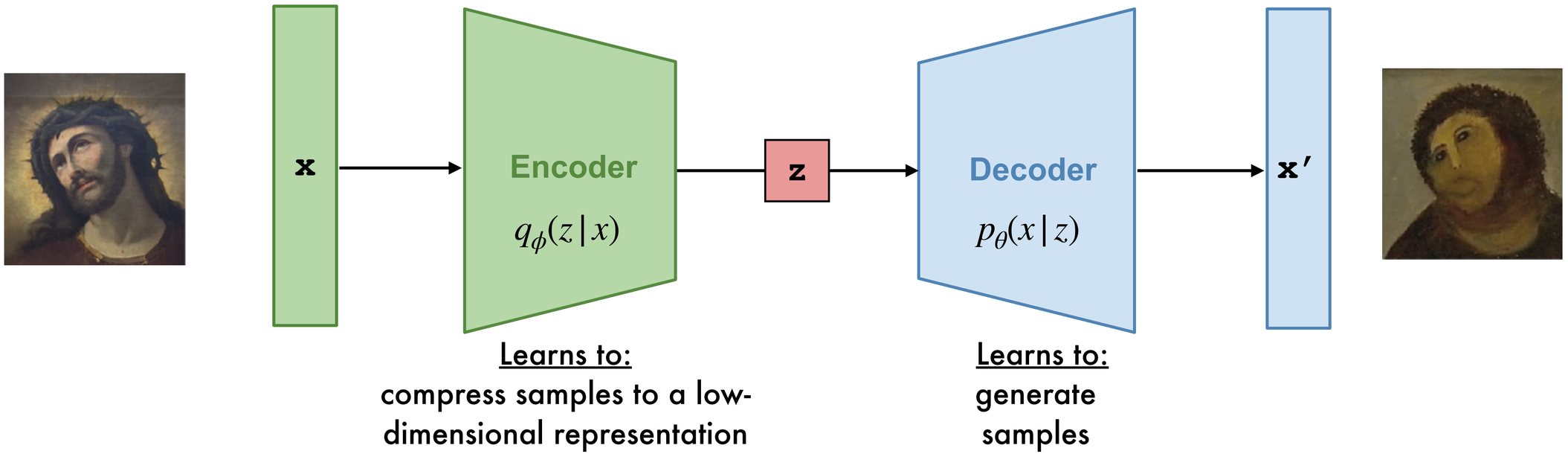
What is an autoencoder?

- Encoder + Decoder with a bottleneck z
- Reconstruction loss (x', x)



Variational Autoencoders (VAEs)

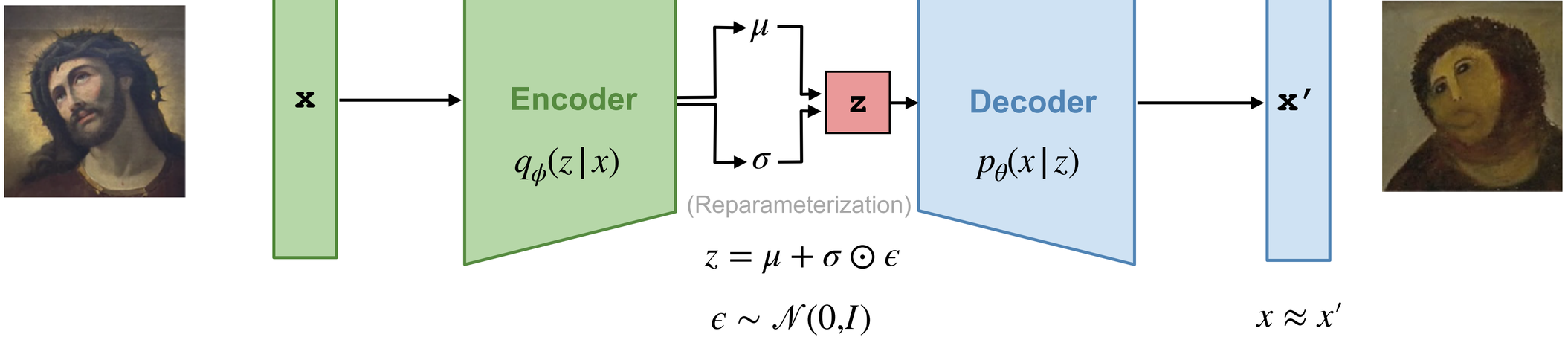
- Autoencoder with structured bottleneck
- At training time, jointly learn *encoder* and *decoder* by maximizing a variational bound on the data likelihood, with 2 loss terms: (1) KL divergence and (2) Reconstruction



D. Kingma and M. Welling, Auto-Encoding Variational Bayes, ICLR 2014

Variational Autoencoders (VAEs)

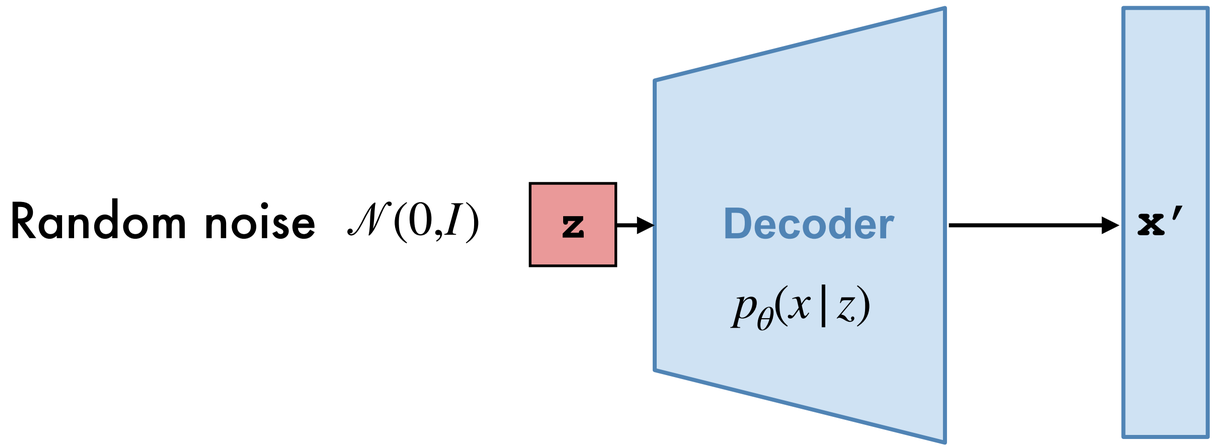
- Autoencoder with structured bottleneck
- At training time, jointly learn *encoder* and *decoder* by maximizing a variational bound on the data likelihood, with 2 loss terms: (1) KL divergence and (2) Reconstruction



D. Kingma and M. Welling, Auto-Encoding Variational Bayes, ICLR 2014

Variational Autoencoders (VAEs)

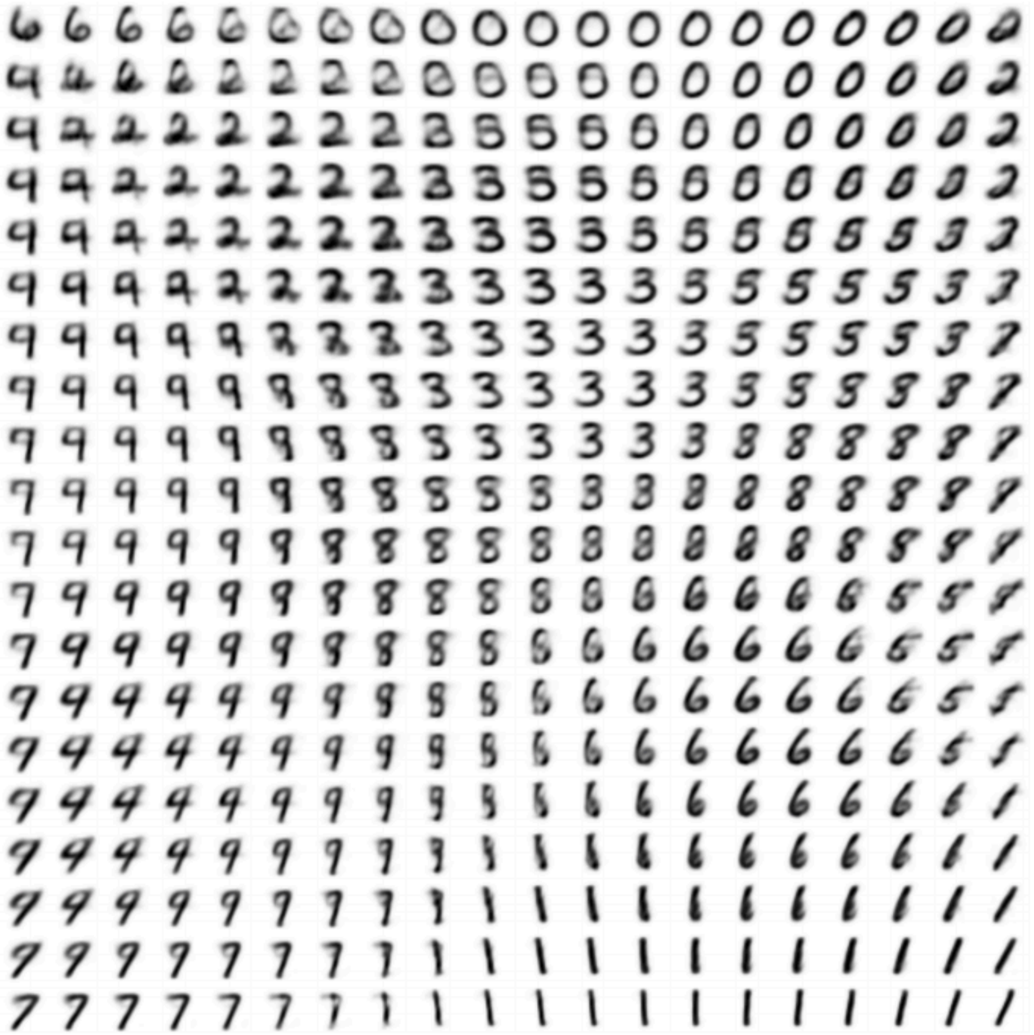
- Autoencoder with structured bottleneck
- At training time, jointly learn *encoder* and *decoder* by maximizing a variational bound on the data likelihood, with 2 loss terms: (1) KL divergence and (2) Reconstruction
- **At test time**, discard encoder and use *decoder* to sample from the learned distribution



D. Kingma and M. Welling, Auto-Encoding Variational Bayes, ICLR 2014

Original VAE results

- Learned 2D “manifolds”:



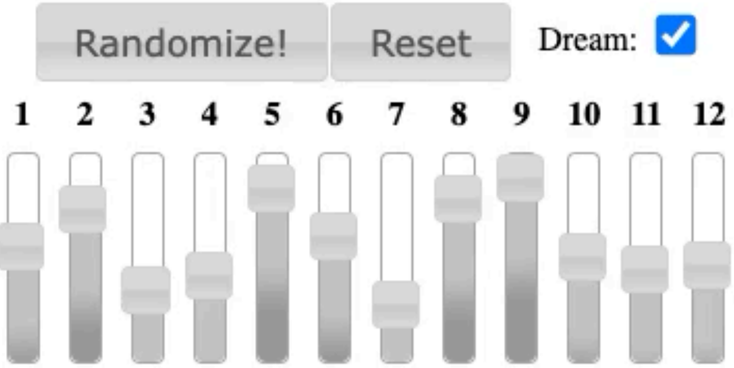
D. Kingma and M. Welling, Auto-Encoding Variational Bayes, ICLR 2014

Variational autoencoders (VAEs)

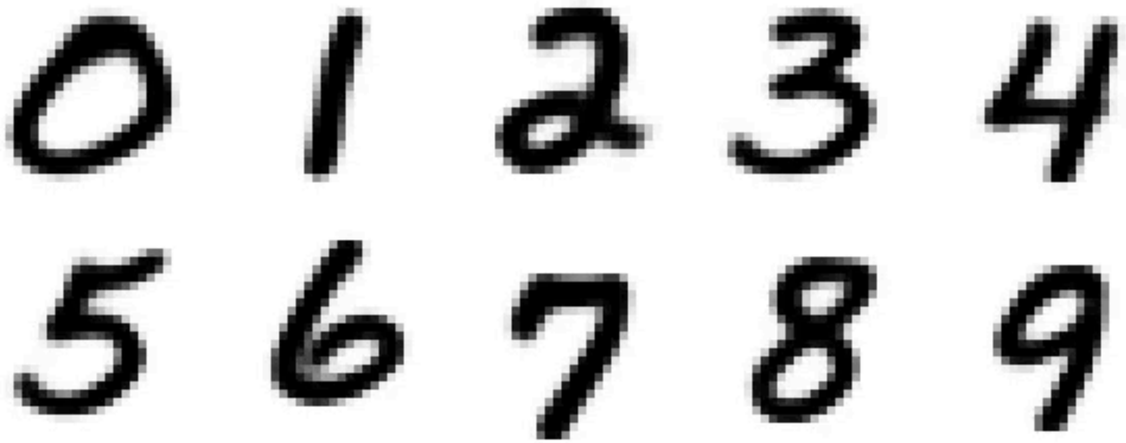
Digit Fantasies by a Deep Generative Model

Instructions:

1. Dream mode: check 'dream' to let the model fantasize digits.
2. Alternatively, you can wiggle the sliders yourselves to wander through **z**-space and observe the effects in **x**-space.



(Latent space) **z**



(Observed space) **x**

http://dpkingma.com/sgvb_mnist_demo/demo.html

Original VAE results

Image source

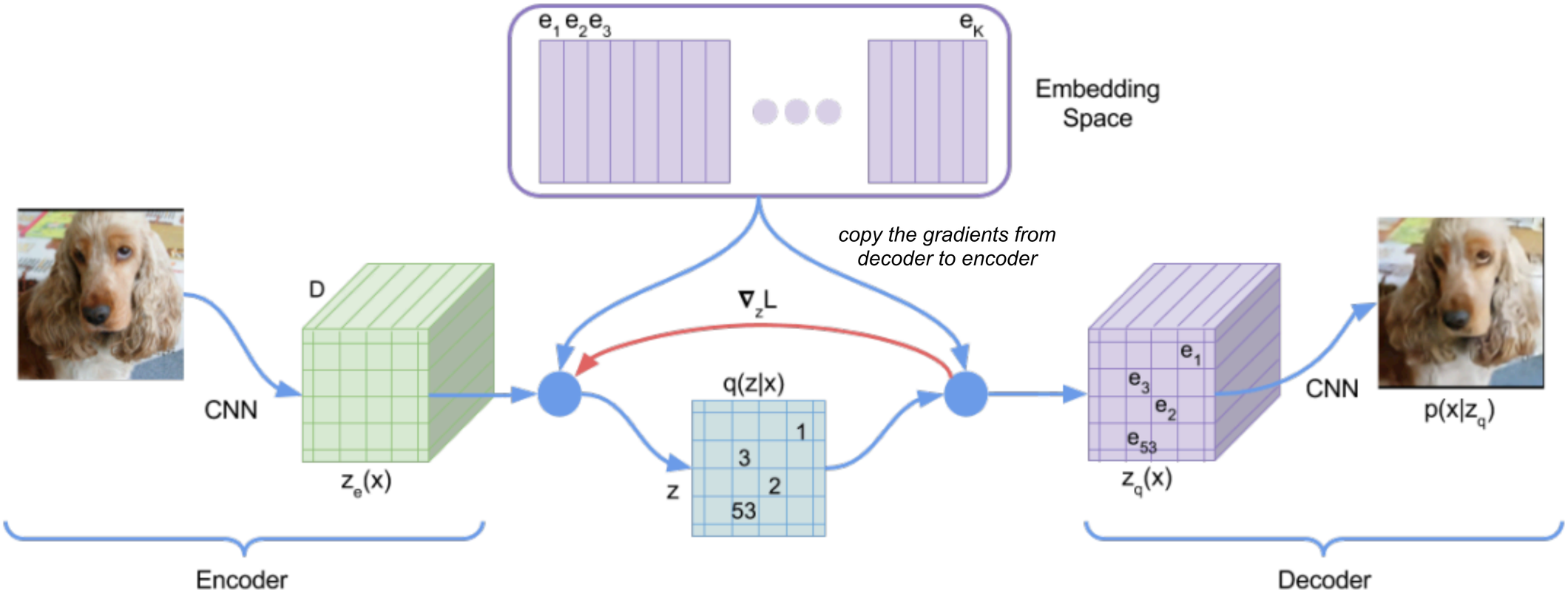
VAE pros and cons

- Pros:
  - Principled mathematical formalism for generative models
  - Allows inference of code given image, can be useful for controlling the latent space
- Cons:
  - Samples blurrier and lower quality compared to GANs
- Active areas of research:
  - More powerful and flexible approximations for relevant probability distributions
  - Combining VAEs and GANs
  - Incorporating structure in latent variables, e.g., hierarchical or categorical distributions

Adapted from J. Johnson

Vector Quantised Variational AutoEncoder (VQ-VAE)

- "We show that a discrete latent model (VQ-VAE) performs as well as its continuous model counterparts in log-likelihood."



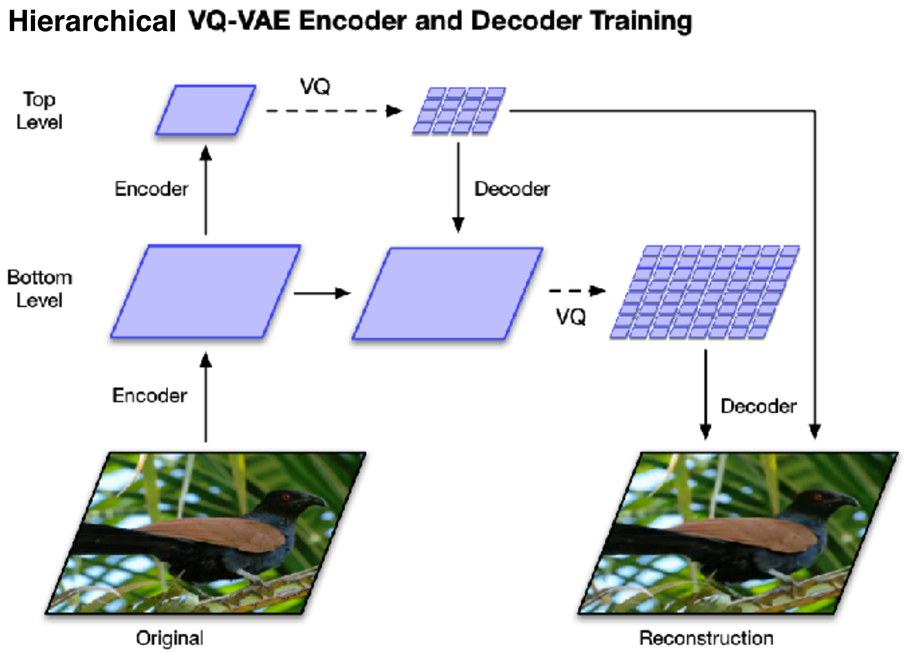
A. van den Oord, O. Vinyals, K. Kavukcuoglu, Neural Discrete Representation Learning, NeurIPS 2017

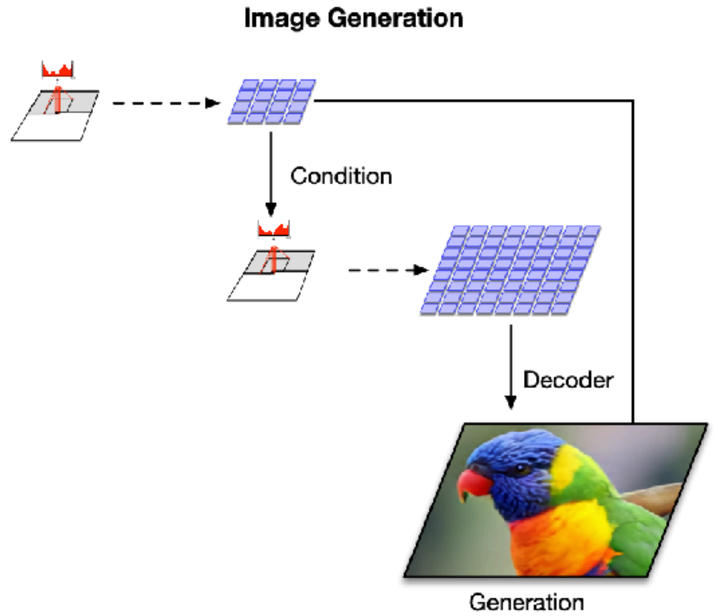
Generating better samples: VQ-VAE-2

- Combining VAE and autoregressive models:

Train a VAE-like model to generate multiscale grids of latent codes

Use a multiscale autoregressive model (PixelCNN) to sample in latent code space





A. Razavi, A. van den Oord, O. Vinyals, Generating Diverse High-Fidelity Images with VQ-VAE-2, NeurIPS 2019

Generating better samples: VQ-VAE-2

- 256 x 256 class-conditional samples, trained on ImageNet:

A. Razavi, A. van den Oord, O. Vinyals, Generating Diverse High-Fidelity Images with VQ-VAE-2, NeurIPS 2019

Generating better samples: VQ-VAE-2

- 256 x 256 class-conditional samples, trained on ImageNet:

A. Razavi, A. van den Oord, O. Vinyals, Generating Diverse High-Fidelity Images with VQ-VAE-2, NeurIPS 2019

Generating better samples: VQ-VAE-2

- 1024 x 1024 generated faces, trained on FFHQ:

A. Razavi, A. van den Oord, O. Vinyals, Generating Diverse High-Fidelity Images with VQ-VAE-2, NeurIPS 2019

Combining VAEs and Transformers: DALL-E

- Train an encoder similar to **VQ-VAE** to compress images to 32x32 grids of discrete tokens (each assuming 8192 values)
- Concatenate with text strings, learn a joint sequential transformer model that can be used to **generate image based on text** prompt

We will come back to text-conditioning.

(a) a tapir made of accordion. a tapir with the texture of an accordion.

(b) an illustration of a baby hedgehog in a christmas sweater walking a dog

(c) a neon sign that reads "backprop". a neon sign that reads "backprop". backprop neon sign

A. Ramesh et al., Zero-Shot Text-to-Image Generation, ICML 2021, https://openai.com/blog/dall-e/

Agenda

1. Generative neural networks

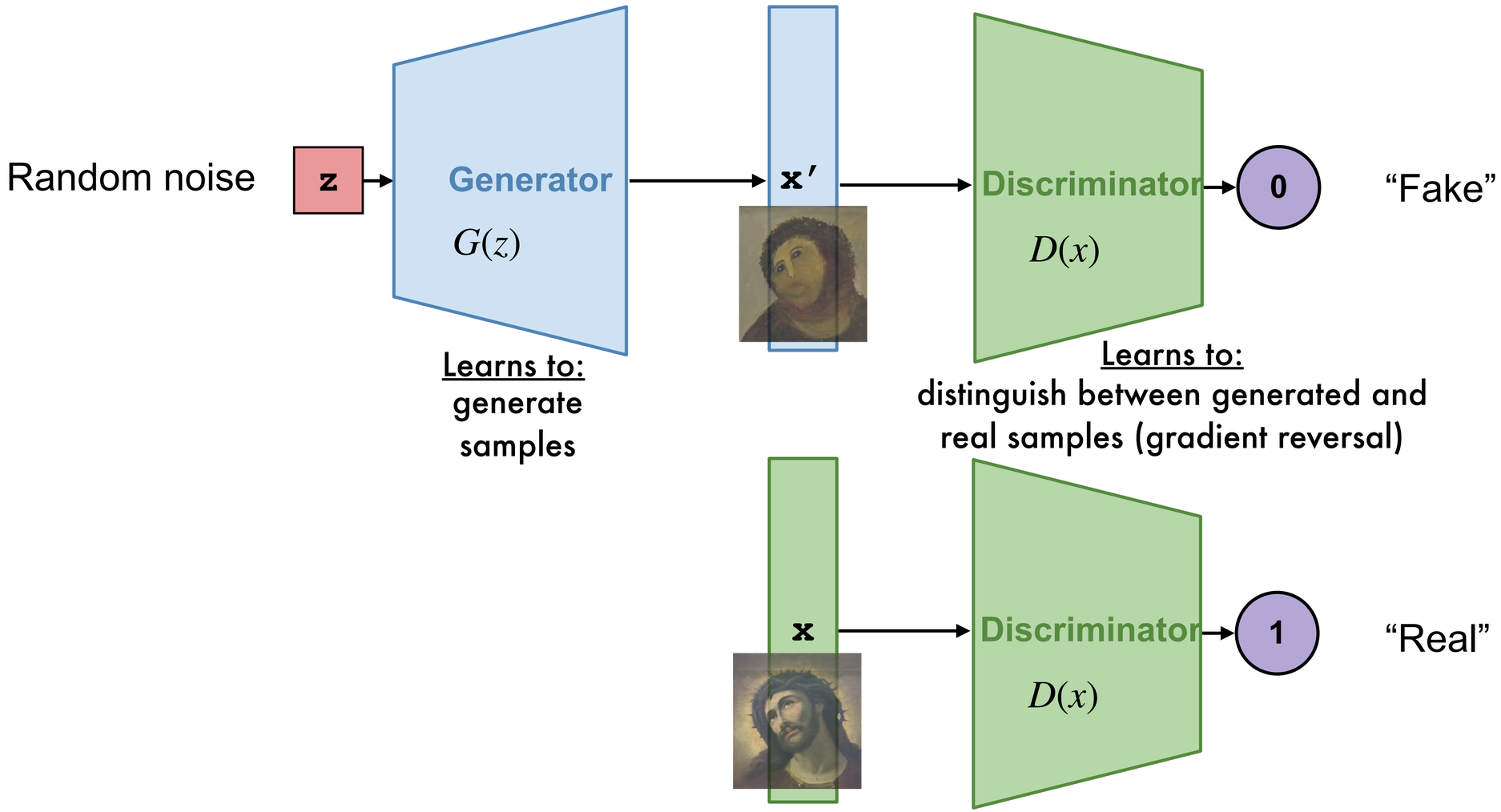
- VAE: Variational autoencoders
- GAN: Generative adversarial networks
- Diffusion models

2. Vision & language

- Text-to-image retrieval
- Text-to-image generation
- Image captioning

Generative Adversarial Networks (GANs)

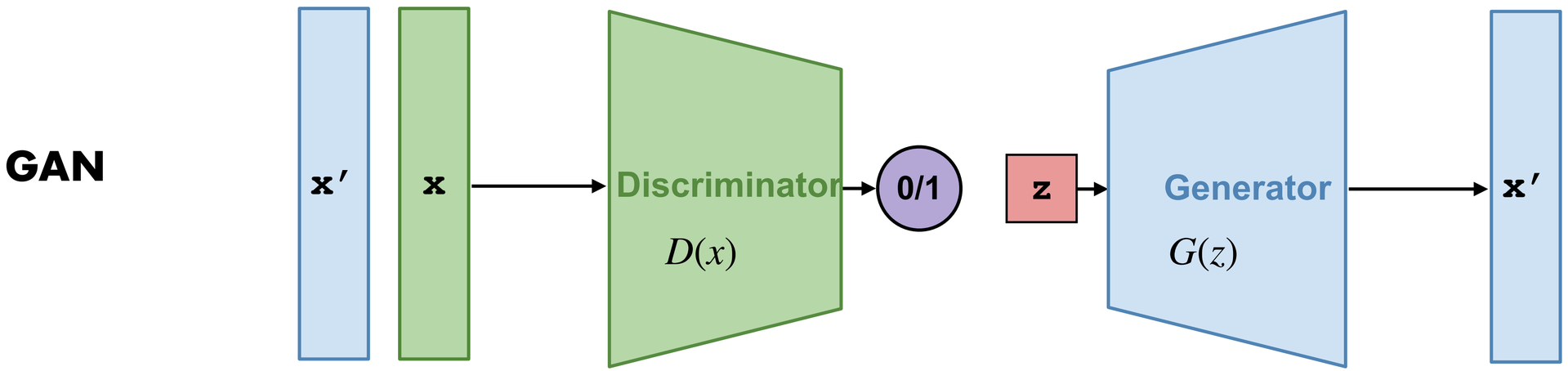
- Train two networks with opposing objectives:



I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, NeurIPS 2014

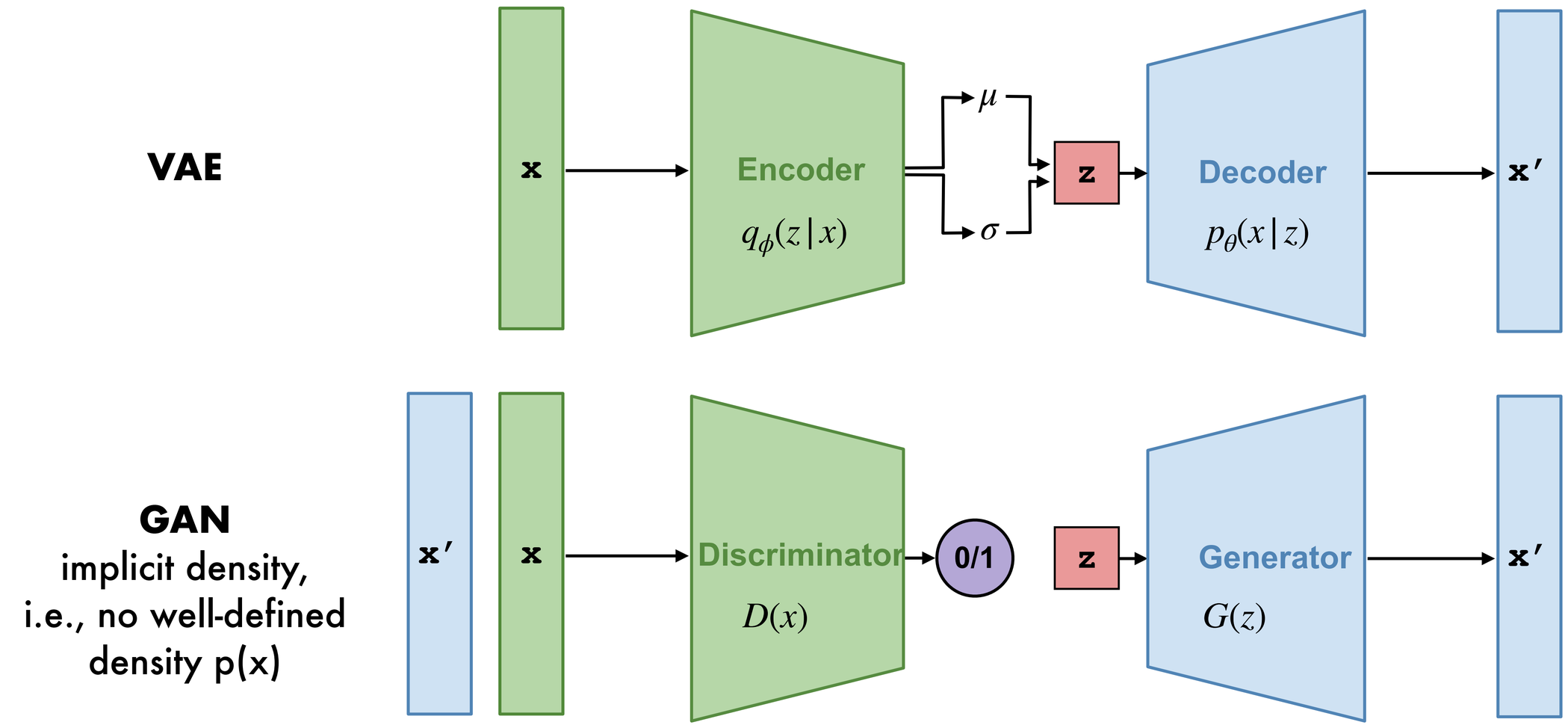
Generative Adversarial Networks (GANs)

- Train two networks with opposing objectives:



I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, NeurIPS 2014

Generative Adversarial Networks (GANs)



Adapted from: https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

GAN objective

- Discriminator $D(x)$ outputs the probability that the sample x is *real.*
- We want $D(x)$ to be close to 1 for real data and close to 0 for fake.
- Expected conditional log likelihood for

real and generated data:

$$= \mathbb{E}_{x \sim p_{\text{data}}} \log D(x) + \mathbb{E}_{z \sim p} \log(1 - D(G(z)))$$

We seed the generator with noise z
drawn from a simple distribution p
(Gaussian or uniform)

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, NeurIPS 2014

GAN objective

$$V(G, D) = \mathbb{E}_{x \sim p_{\text{data}}} \log D(x) + \mathbb{E}_{z \sim p} \log(1 - D(G(z)))$$

- The discriminator wants to correctly distinguish real and fake samples:

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

- The generator wants to fool the discriminator:

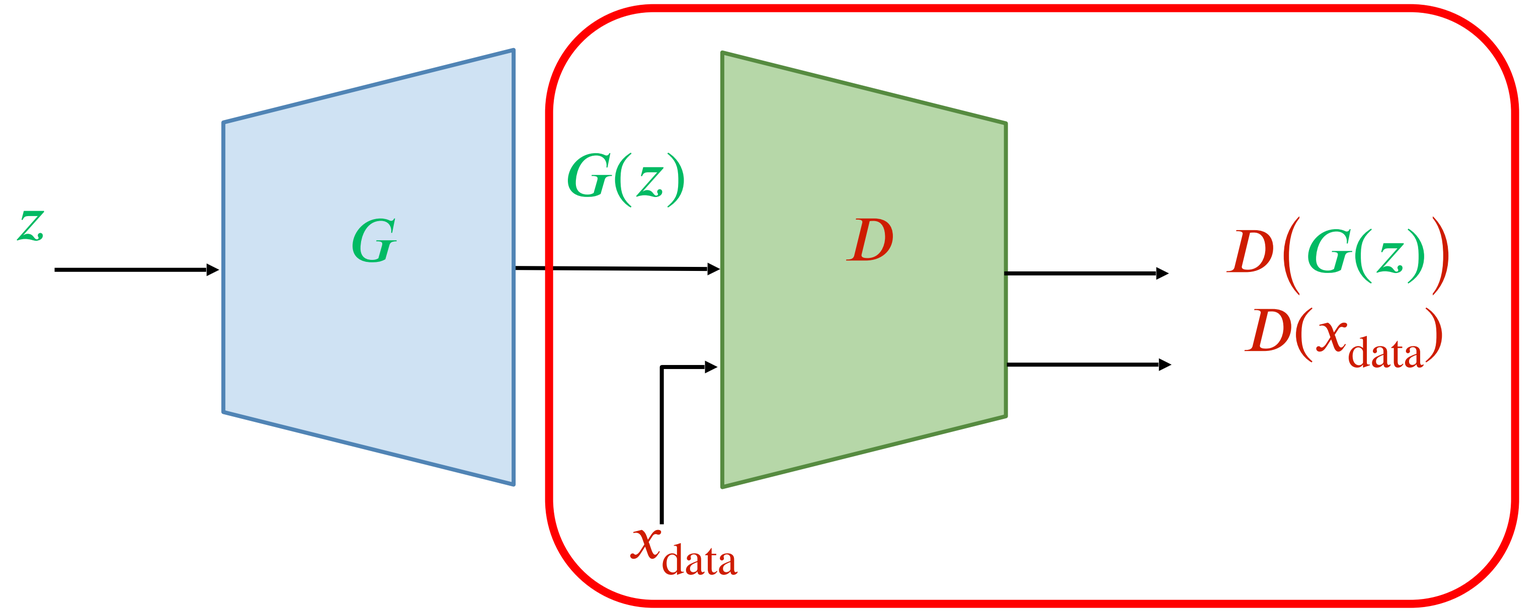
$$G^* = \arg\min_G V(G, D)$$

- Train the generator and discriminator jointly in a *minimax game*

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, NeurIPS 2014

GAN: Schematic picture

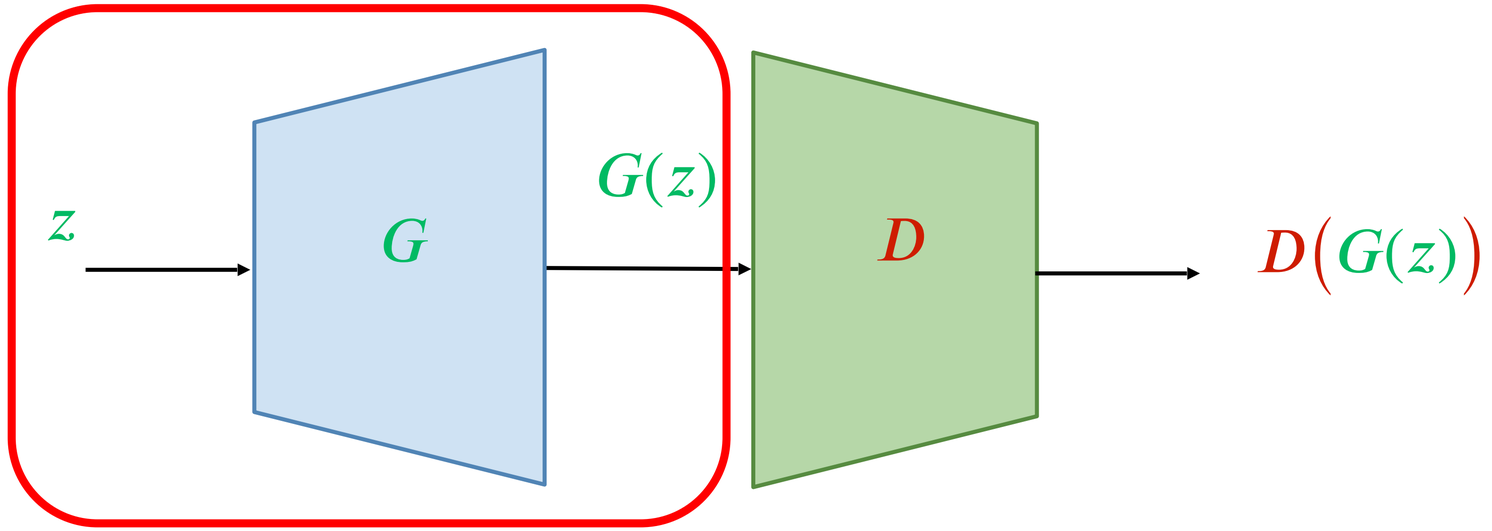
- Update discriminator: push $D(x_{\text{data}})$ close to 1 and $D\big(G(z)\big)$ close to 0
  - The generator is a "black box" to the discriminator



I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, NeurIPS 2014

GAN: Schematic picture

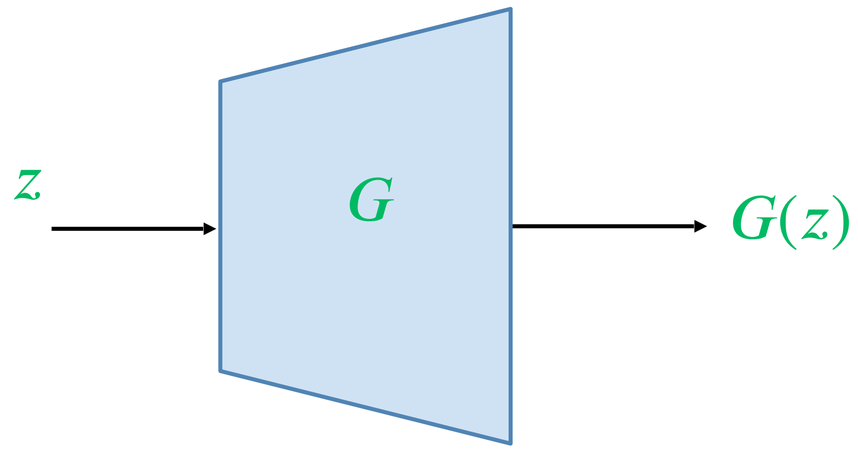
- Update generator: increase $D\big(G(z)\big)$
  - Requires back-propagating through the composed generator-discriminator network (i.e., the discriminator cannot be a black box)
  - The generator is exposed to real data only via the output of the discriminator (and its gradients)



I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, NeurIPS 2014

GAN:

- Test time – the discriminator is discarded

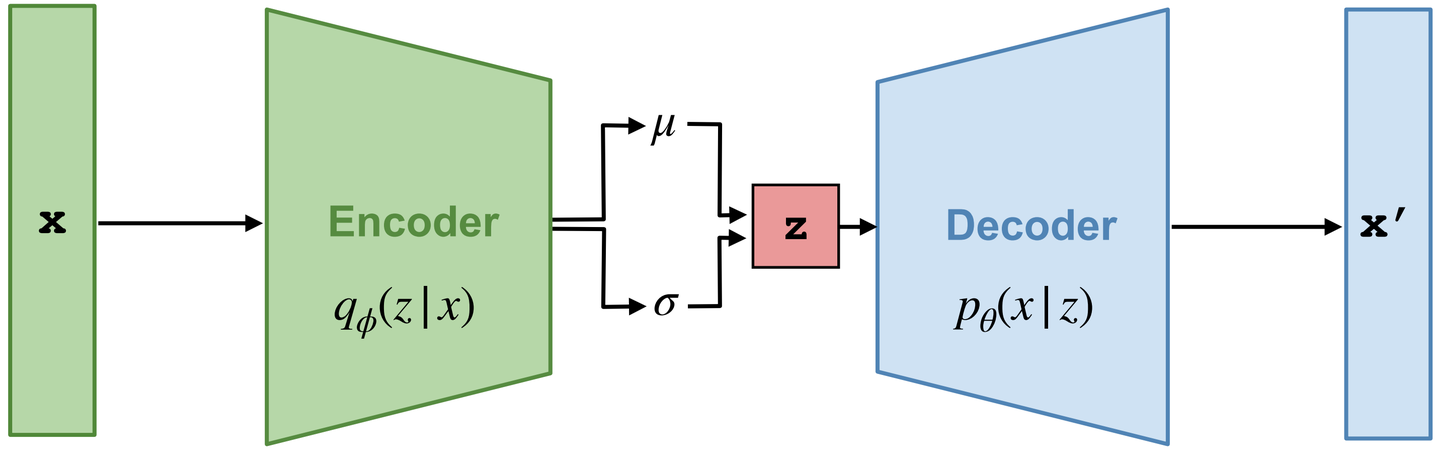


I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, NeurIPS 2014

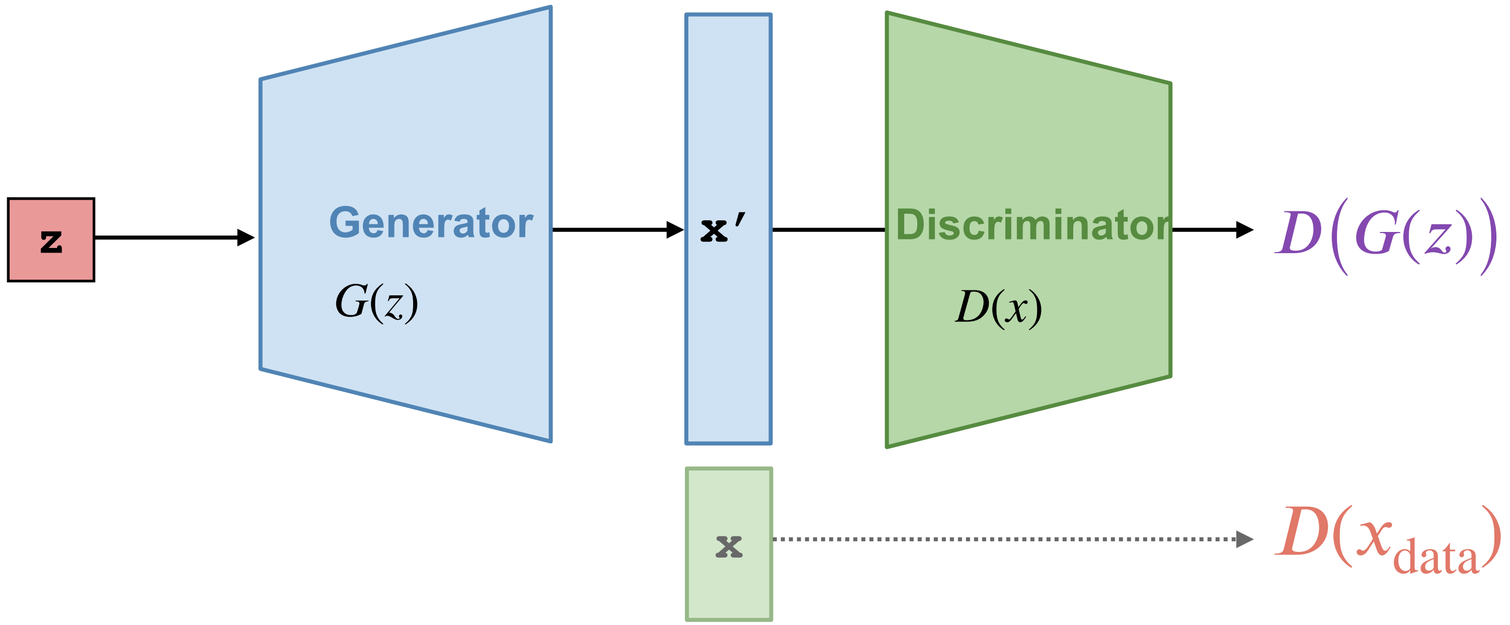
VAEs vs. GANs

VAEs offer more control over the latent space, explicit encoding

VAE training



GAN training
implicit density,
i.e., no well-defined
density p(x)

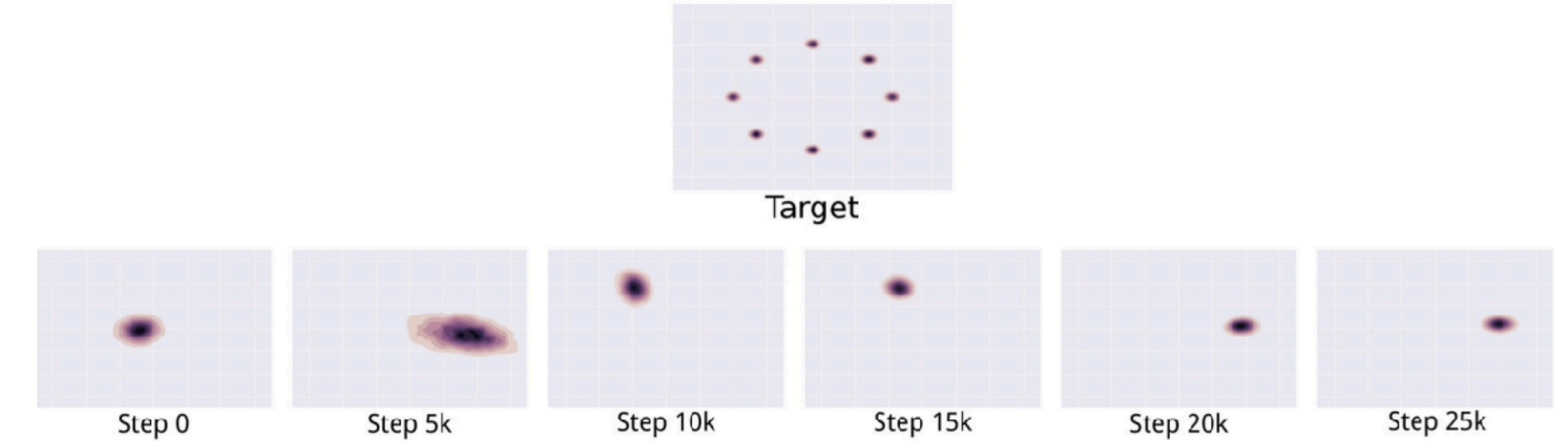


Problems with GAN training

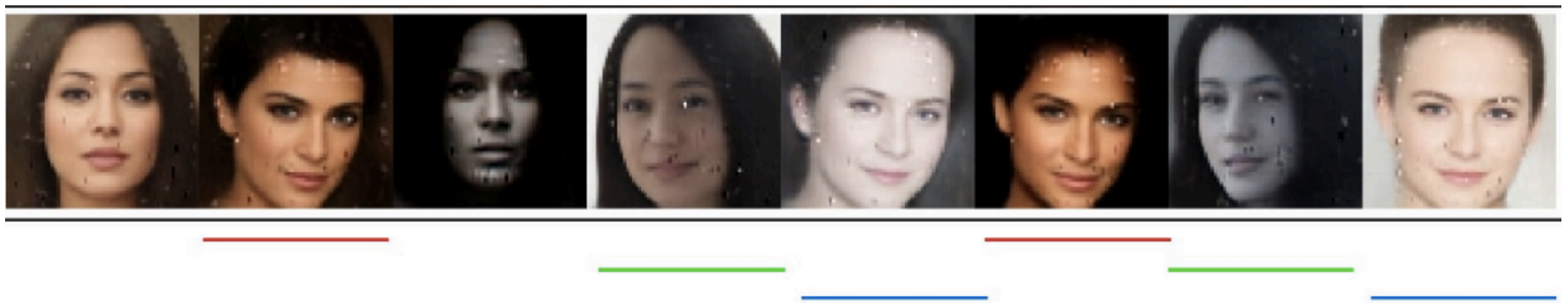
- Stability
  - Parameters can oscillate or diverge, generator loss does not correlate with sample quality
  - Behavior very sensitive to hyperparameter selection

Problems with GAN training

- Mode collapse
  - Generator ends up modeling only a small subset of the training data



Source

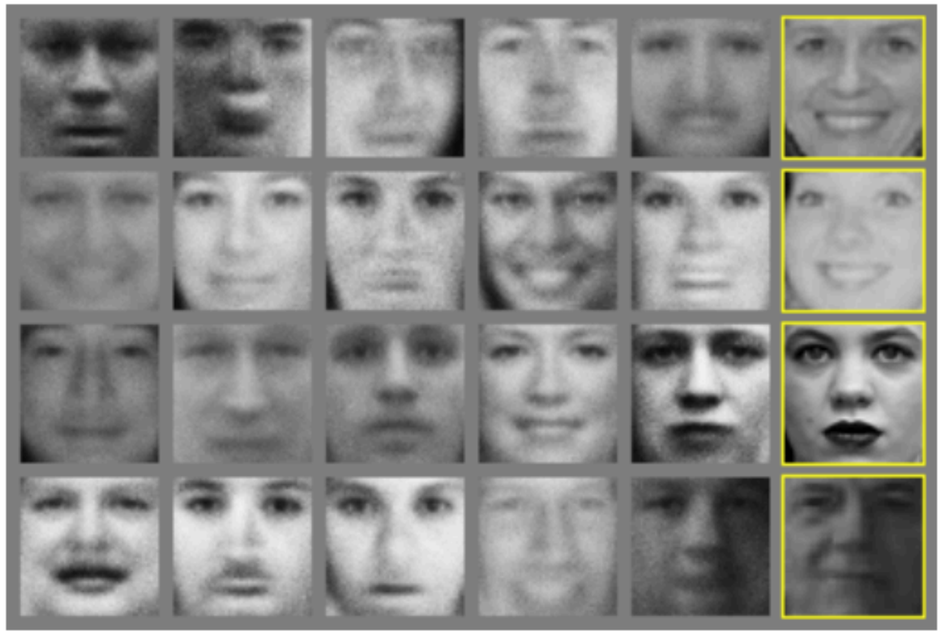


Source

Original GAN results

MNIST digits

Toronto Face Dataset

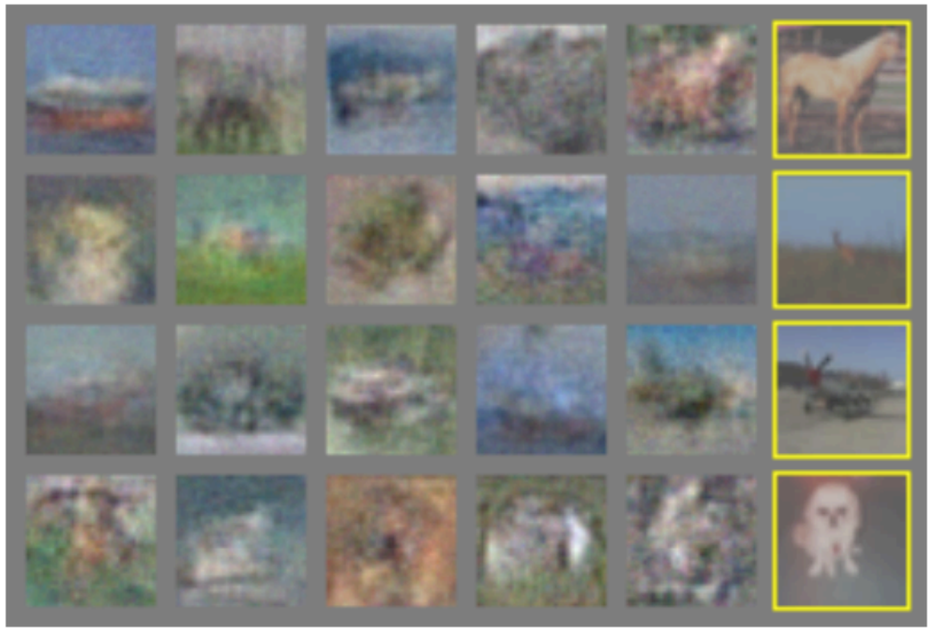


Nearest real image for
sample to the left

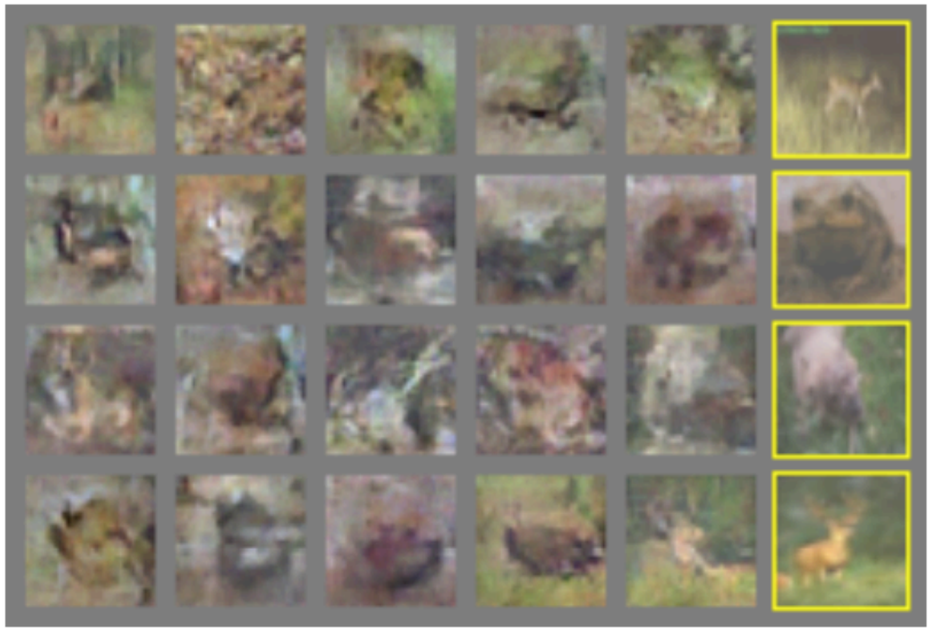
I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, NeurIPS 2014

Original GAN results

CIFAR-10 (FC networks)



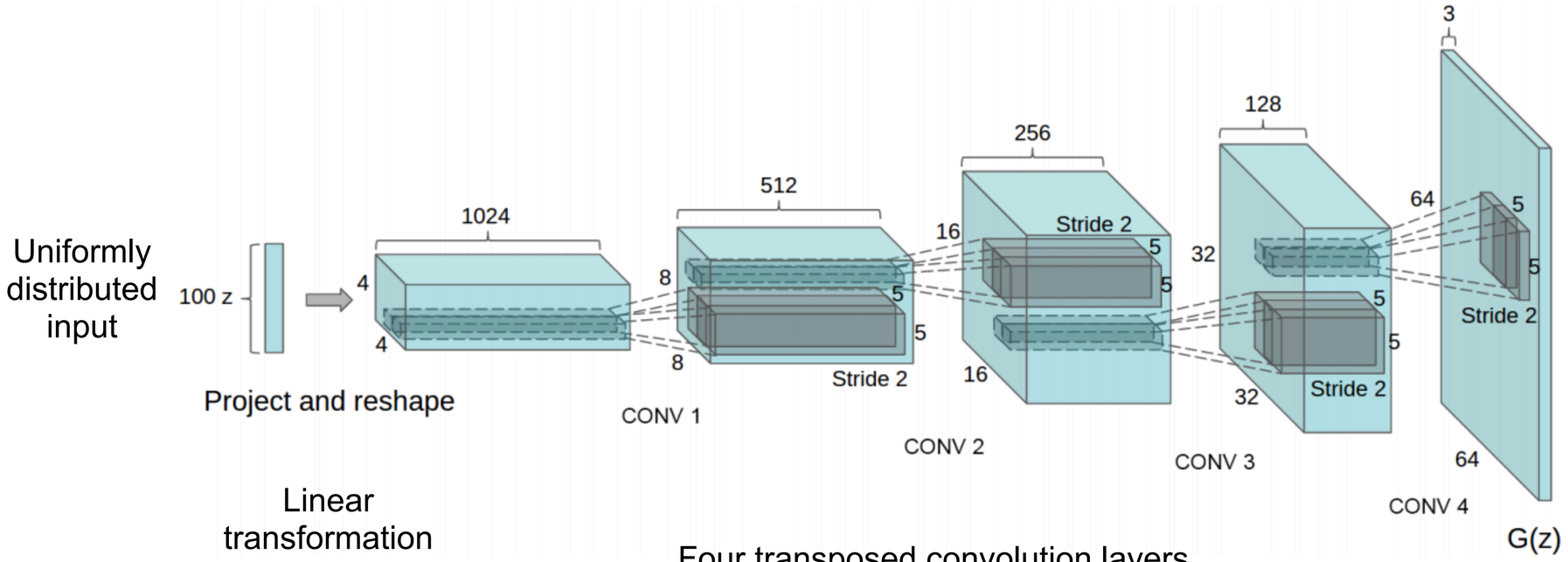
CIFAR-10 (conv networks)



I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, NeurIPS 2014

DCGAN (Deep Convolutional GAN)

- Early, influential convolutional architecture for generator



A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, ICLR 2016

DCGAN

- Early, influential convolutional architecture for generator
- Discriminator architecture (empirically determined to give best training stability):
  - Don’t use pooling, only strided convolutions
  - Use Leaky ReLU activations (sparse gradients cause problems for training)
  - Use only one FC layer before the softmax output
  - Use batch normalization after most layers (in the generator also)

A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, ICLR 2016

DCGAN results

Generated bedrooms after one epoch

A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, ICLR 2016

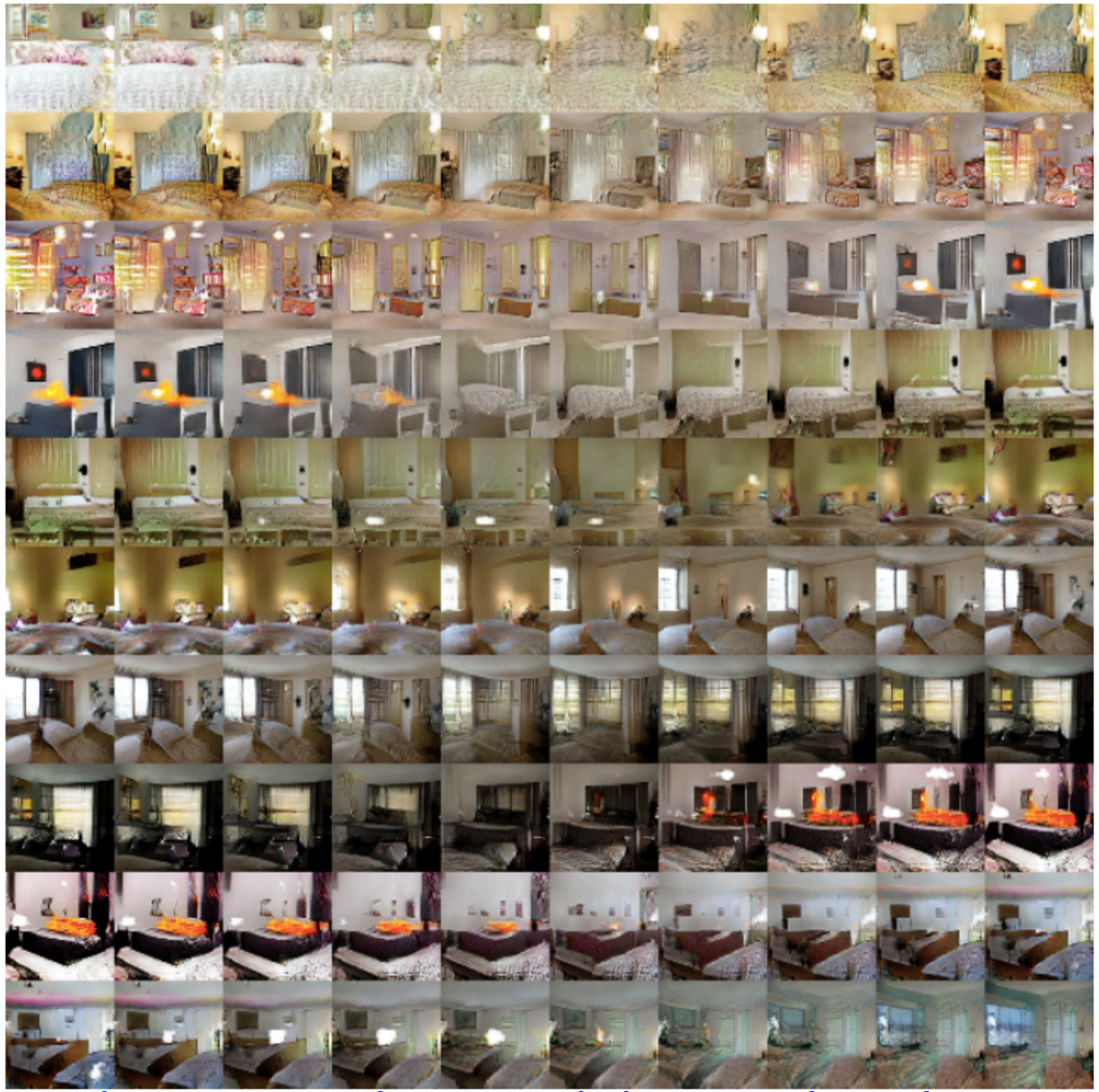
DCGAN results

Generated bedrooms after five epochs

A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, ICLR 2016

DCGAN results

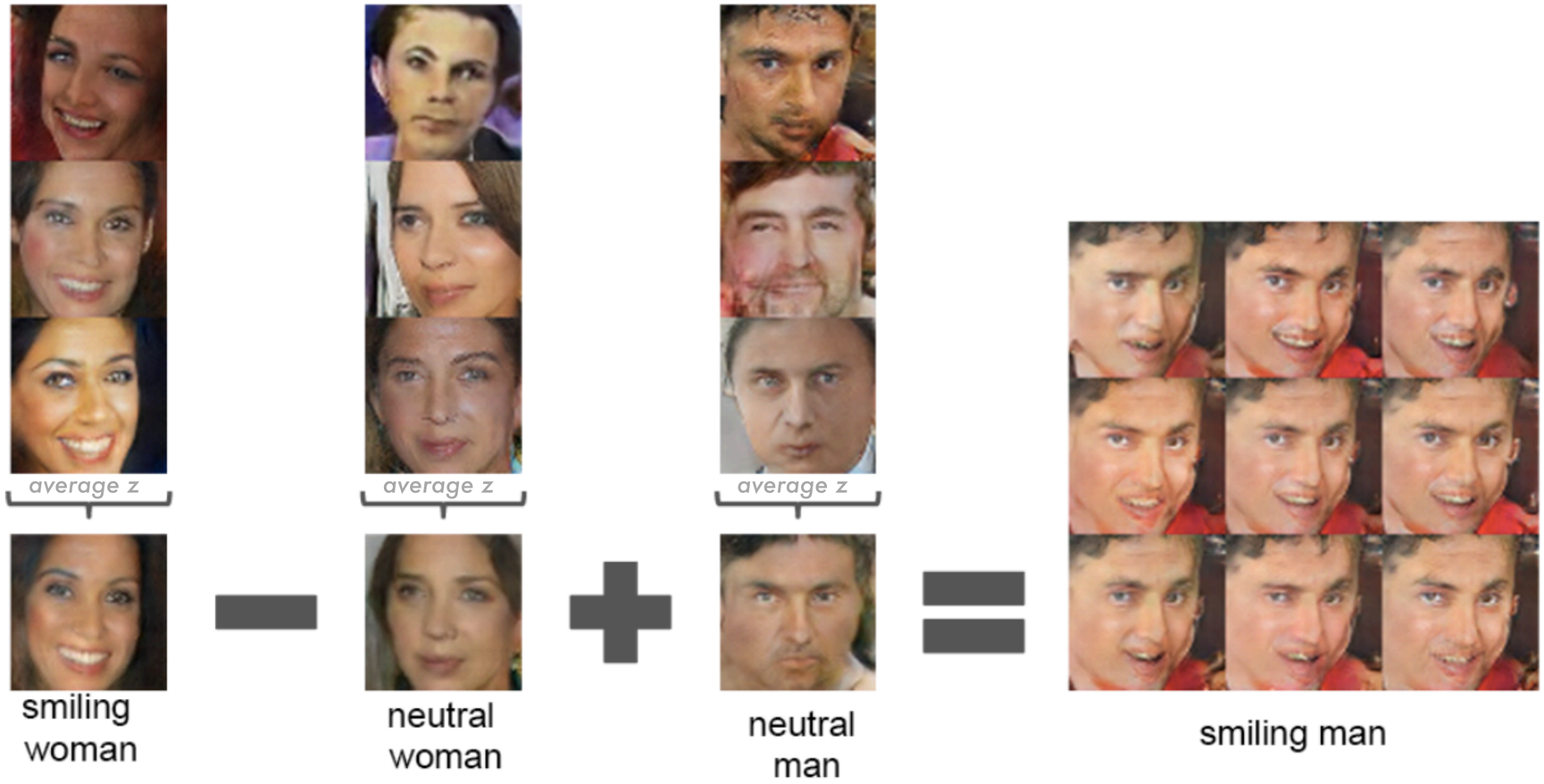
Interpolation between different points in the z space



A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, ICLR 2016

DCGAN results

- Vector arithmetic in the z space

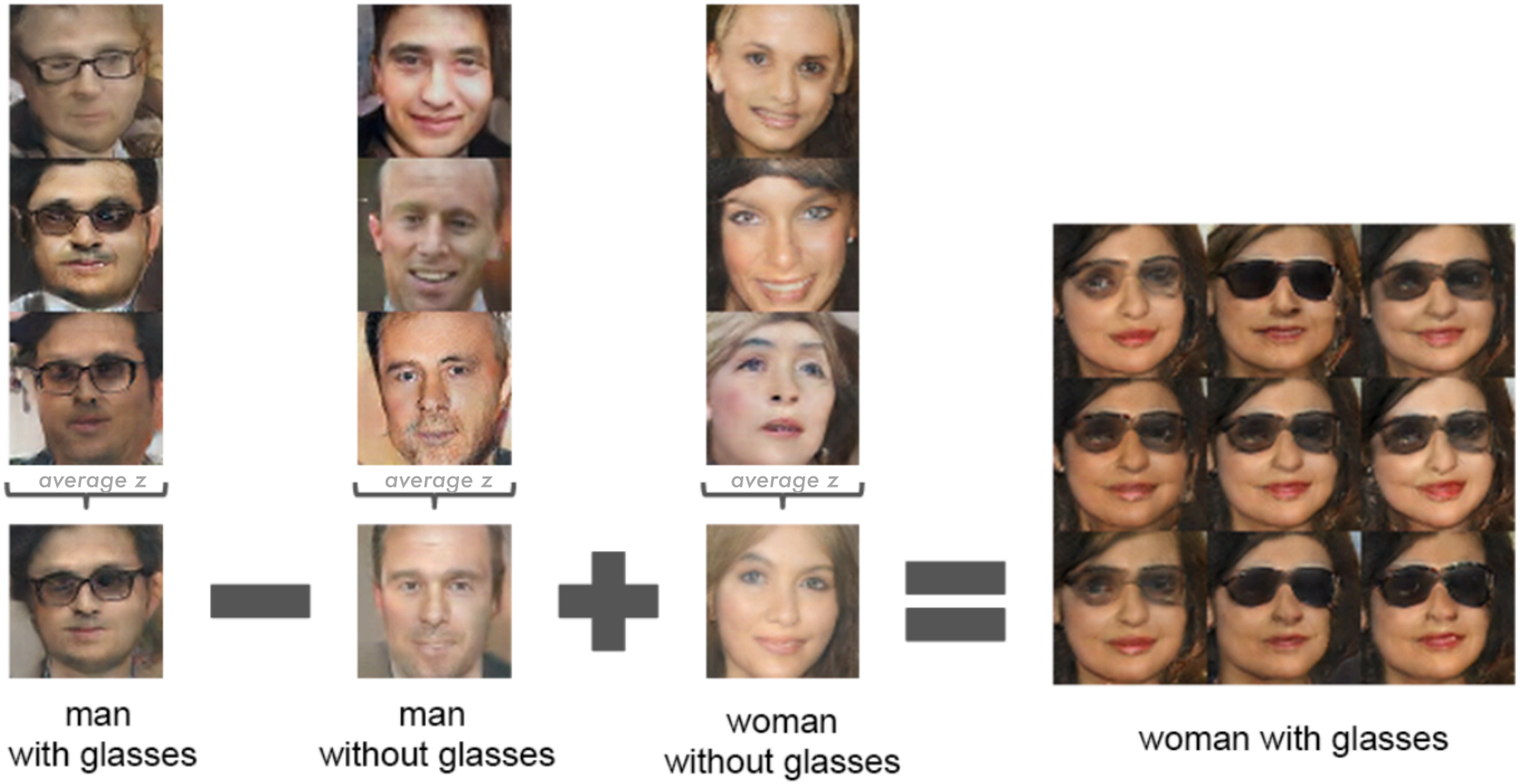


"Experiments working on only single samples per concept were unstable, but averaging the Z vector for ***three*** *examplars showed consistent and stable generations that semantically obeyed the arithmetic."*

A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, ICLR 2016

DCGAN results

- Vector arithmetic in the z space

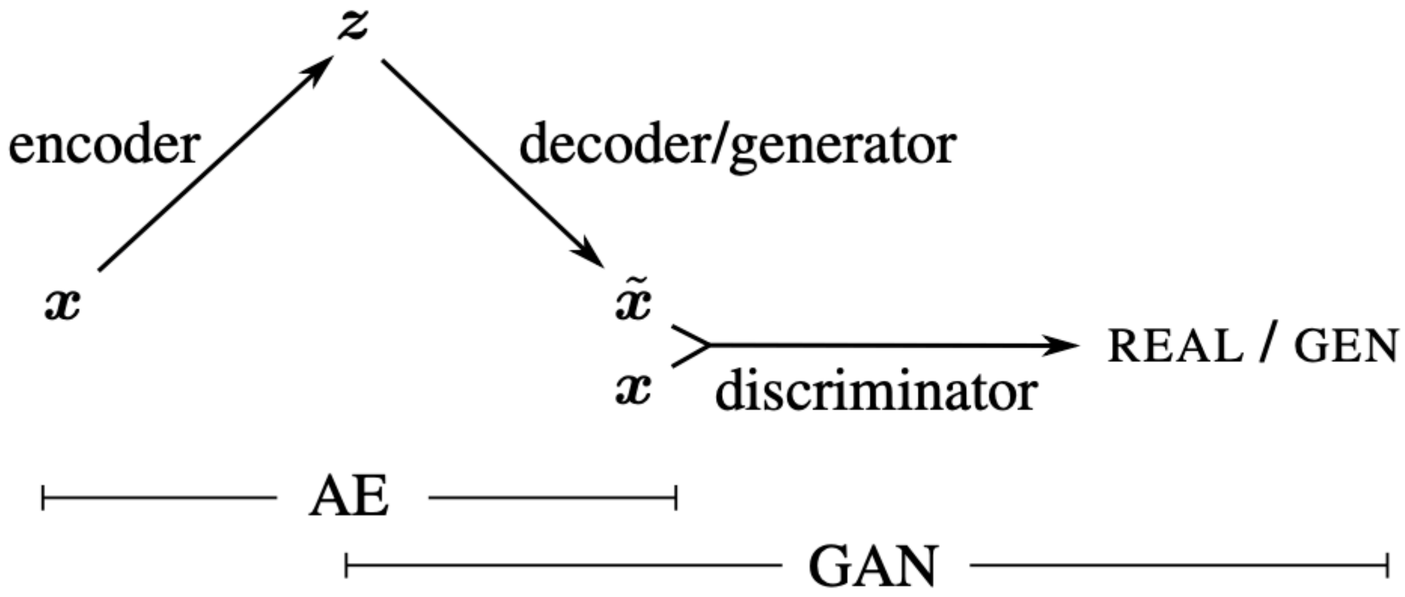


"Experiments working on only single samples per concept were unstable, but averaging the Z vector for three examplars showed consistent and stable generations that semantically obeyed the arithmetic."

A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, ICLR 2016

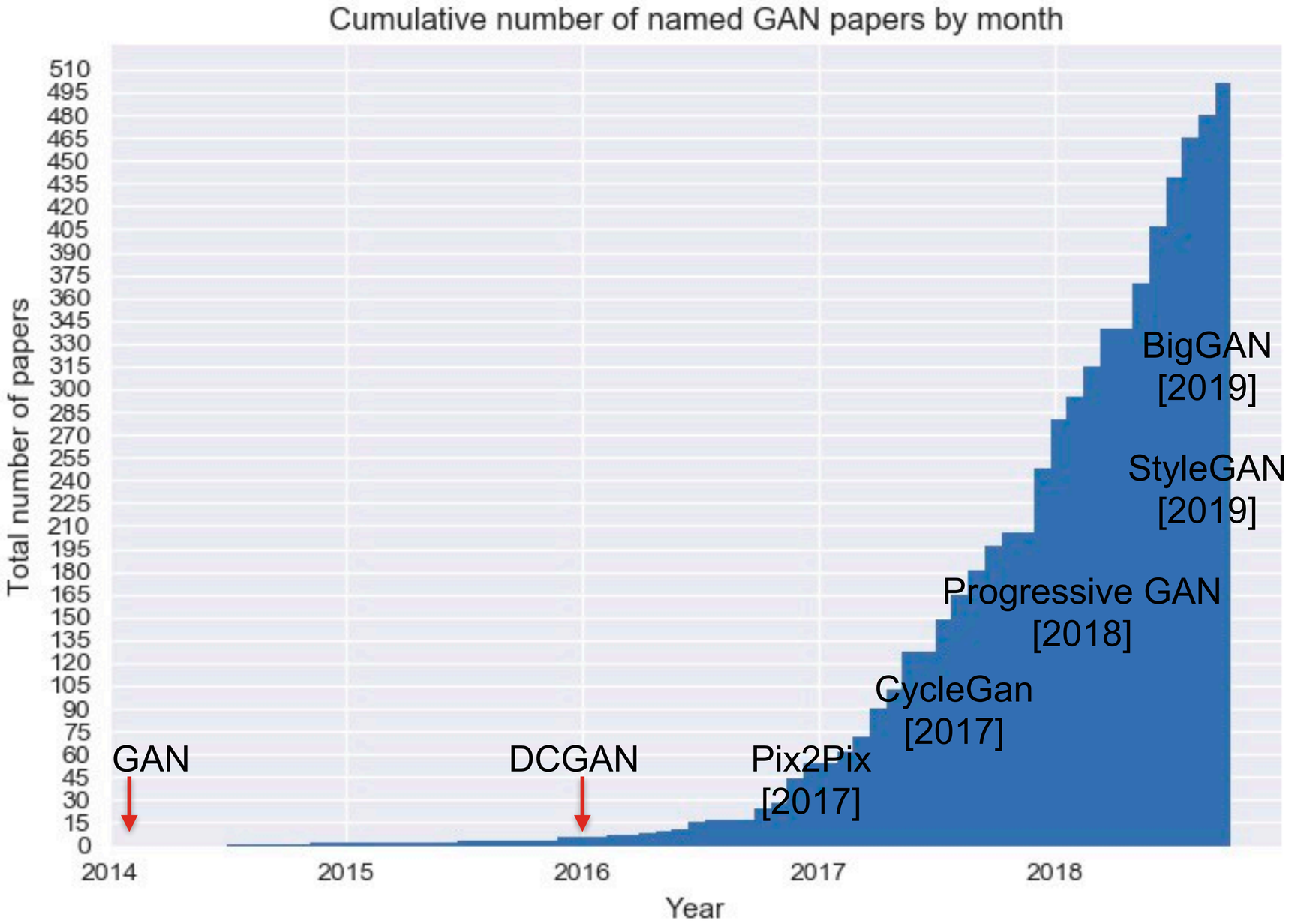
Hybrid approaches: e.g., Combining VAEs and GANs

- Define decoder probability model $p_\theta(x \mid z)$ not in terms of reconstruction errors in pixel space, but in terms of errors in discriminator feature space



A. Larsen, S. Sonderby, H. Larochelle, O. Winther, Autoencoding beyond pixels using a learned similarity metric, ICML 2016

Fast-forwarding a little...



Source

Progress in GANs

- **Progressive GAN, StyleGAN, StyleGan2 (higher quality)**

T. Karras, T. Aila, S. Laine, J. Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. ICLR 2018

T. Karras, S. Laine, T. Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. CVPR 2019

T. Karras et al. Analyzing and Improving the Image Quality of StyleGAN. CVPR 2020

- **GAN Dissection (interpretability)**

D. Bau et al. GAN Dissection: Visualizing and understanding generative adversarial networks. ICLR 2019

- **BigGan (class-conditioned)**

A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, ICLR 2019

- **Pix2Pix, CycleGan (image-conditioned)**

P. Isola, J.-Y. Zhu, T. Zhou, A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, CVPR 2017

J.-Y. Zhu, T. Park, P. Isola, A. Efros, Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, ICCV 2017

…

Progress in GANs: Faces

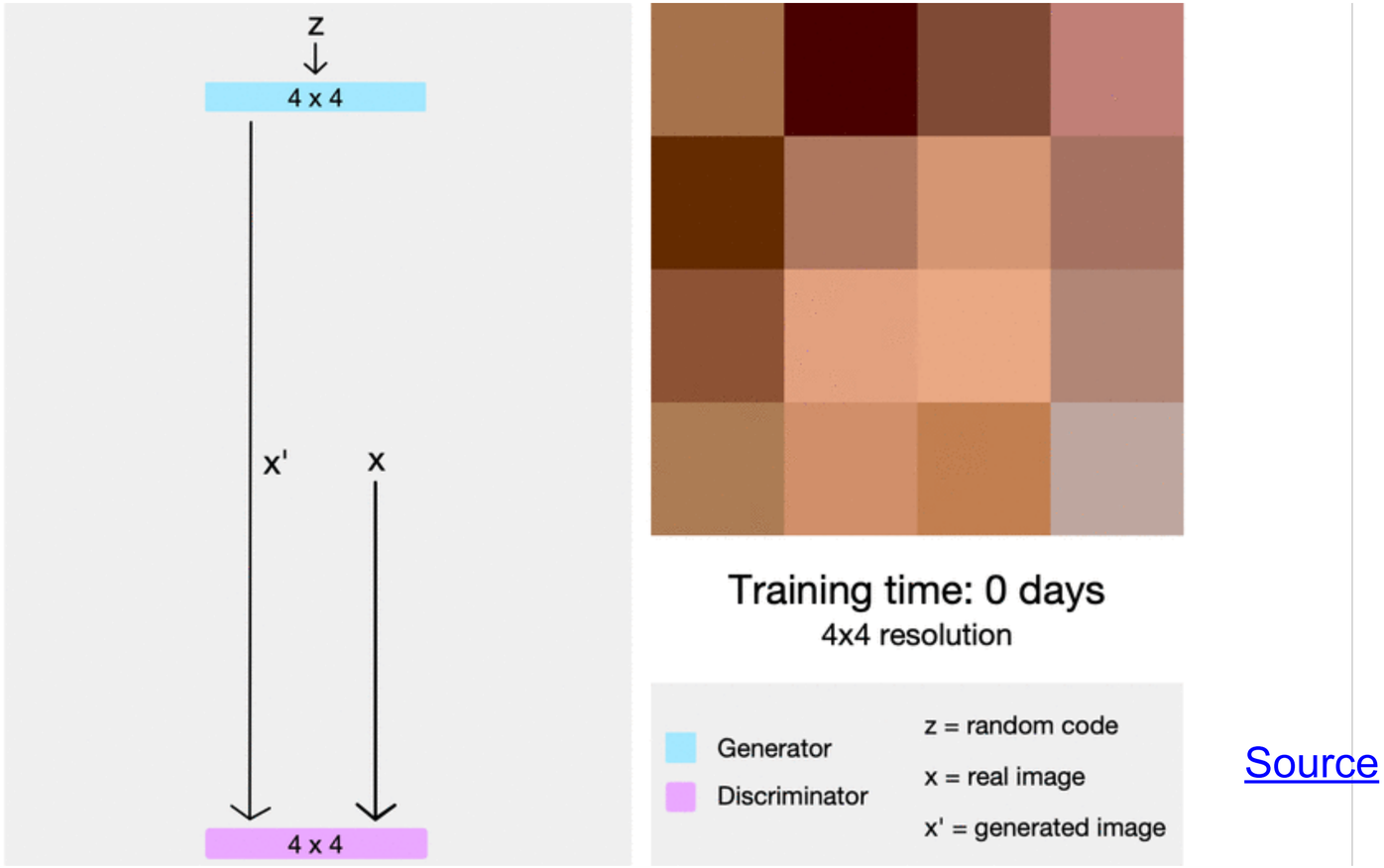
Progressive GANs

Realistic face images up to 1024 x 1024 resolution

T. Karras, T. Aila, S. Laine, J. Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. ICLR 2018

Progressive GANs

- Key idea: train lower-resolution models, gradually add layers corresponding to higher-resolution outputs



Source

T. Karras, T. Aila, S. Laine, J. Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. ICLR 2018

Progressive GANs: Results

256 x 256 results for LSUN categories

"A separate network was trained for each category using identical parameters."

StyleGAN: Results

Built on top of Progressive GAN

T. Karras, S. Laine, T. Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. CVPR 2019

GAN Dissection

D. Bau et al. GAN Dissection: Visualizing and understanding generative adversarial networks. ICLR 2019

GAN Dissection

GAN dissection allows us to ask:

1. Does the network learn internal neurons that match meaningful concepts?
2. Do these sets of neurons merely correlate with objects, or does the GAN use those neurons to reason about objects?
3. Can causal neurons be manipulated to improve the output of a GAN?

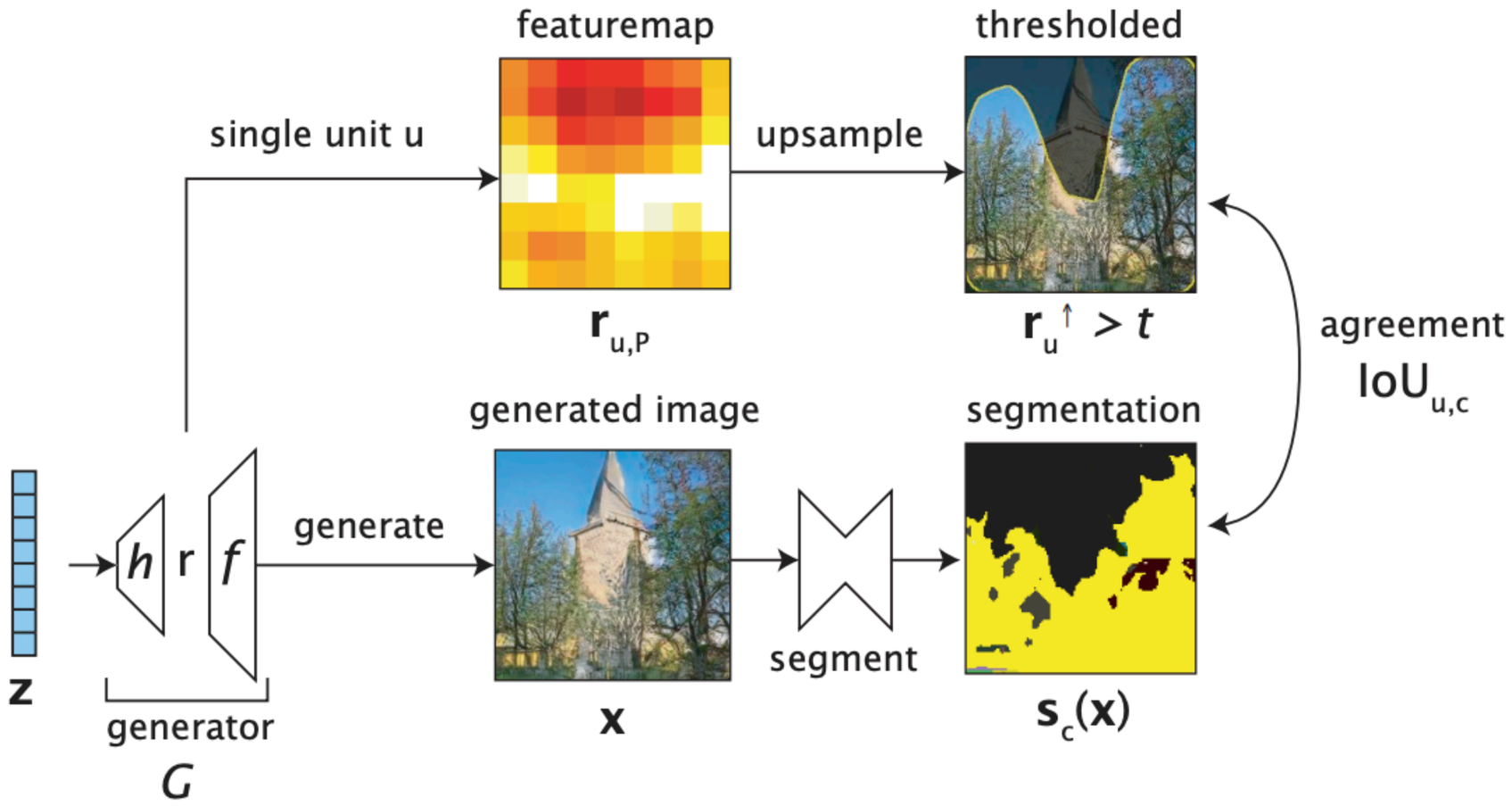
D. Bau et al. GAN Dissection: Visualizing and understanding generative adversarial networks. ICLR 2019

GAN Dissection

D. Bau et al. GAN Dissection: Visualizing and understanding generative adversarial networks. ICLR 2019

GAN Dissection

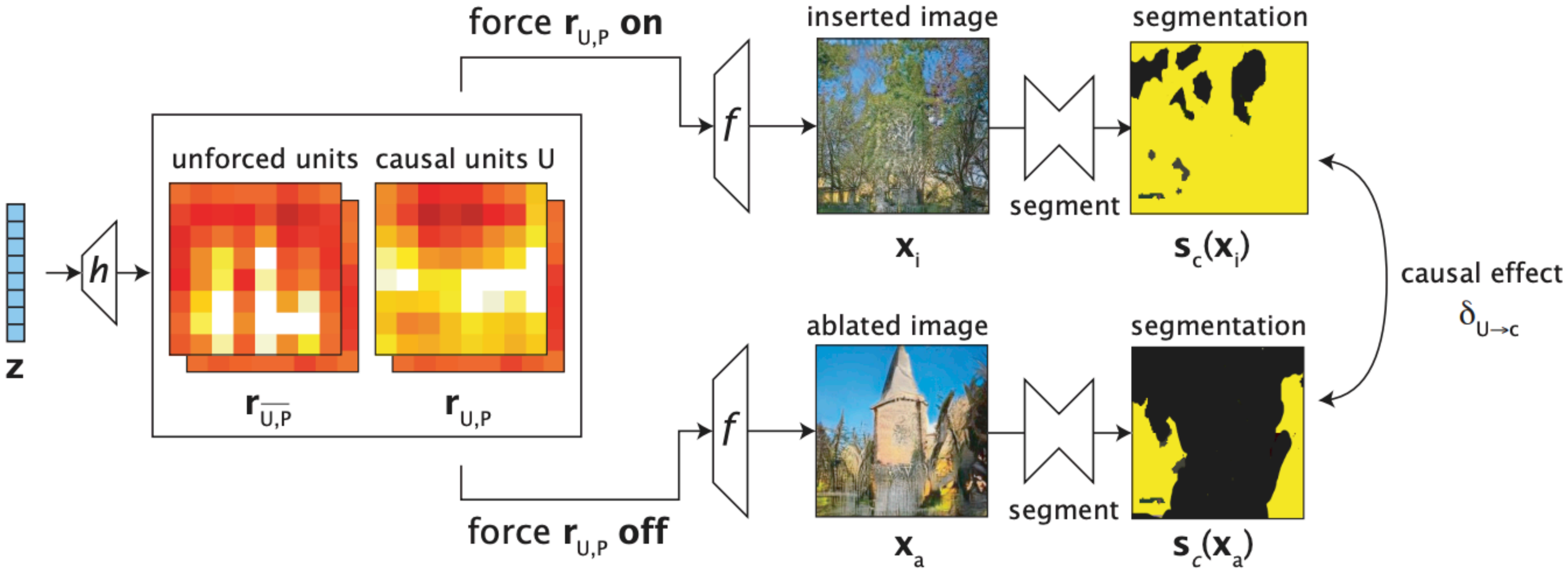
- Dissection: measure agreement between a unit and a concept



D. Bau et al. GAN Dissection: Visualizing and understanding generative adversarial networks. ICLR 2019

GAN Dissection

- Intervention: measure the causal effect of a set of units and a concept

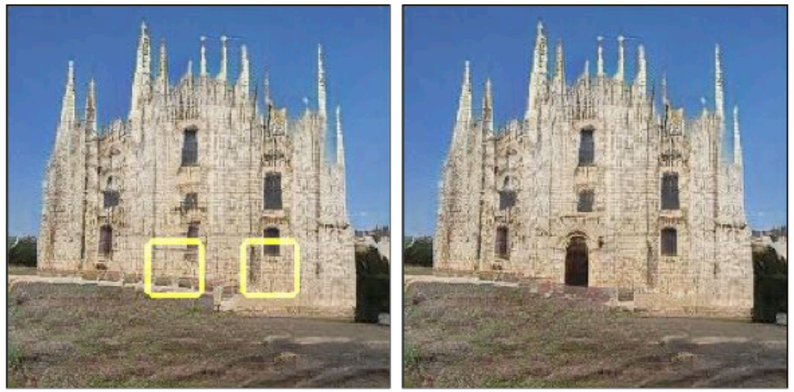


D. Bau et al. [GAN Dissection: Visualizing and understanding generative adversarial networks](). ICLR 2019

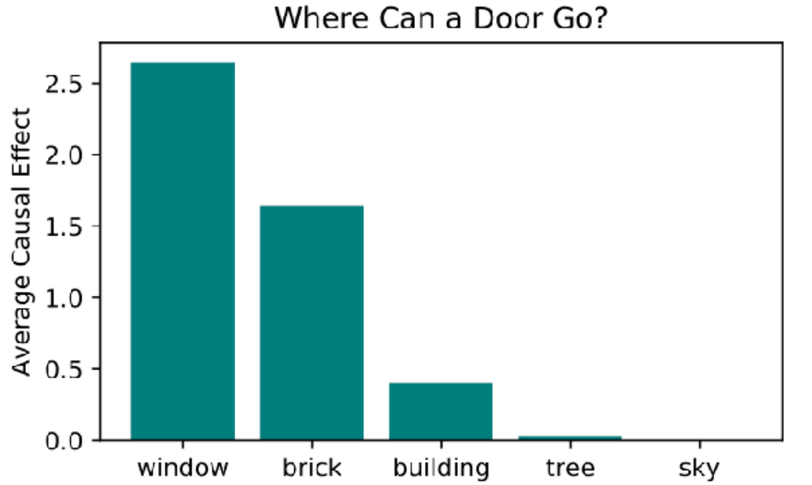
GAN Dissection

*"The network also understands when it can and cannot compose objects. For example, turning on neurons for a door in the proper location of a building will **add a door**. But doing the same in the sky or on a tree will typically have no effect. This structure can be quantified."*

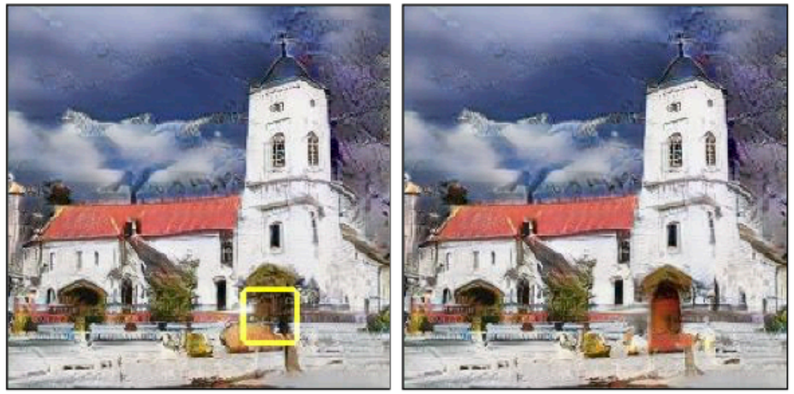
(a)



(b)



(c)

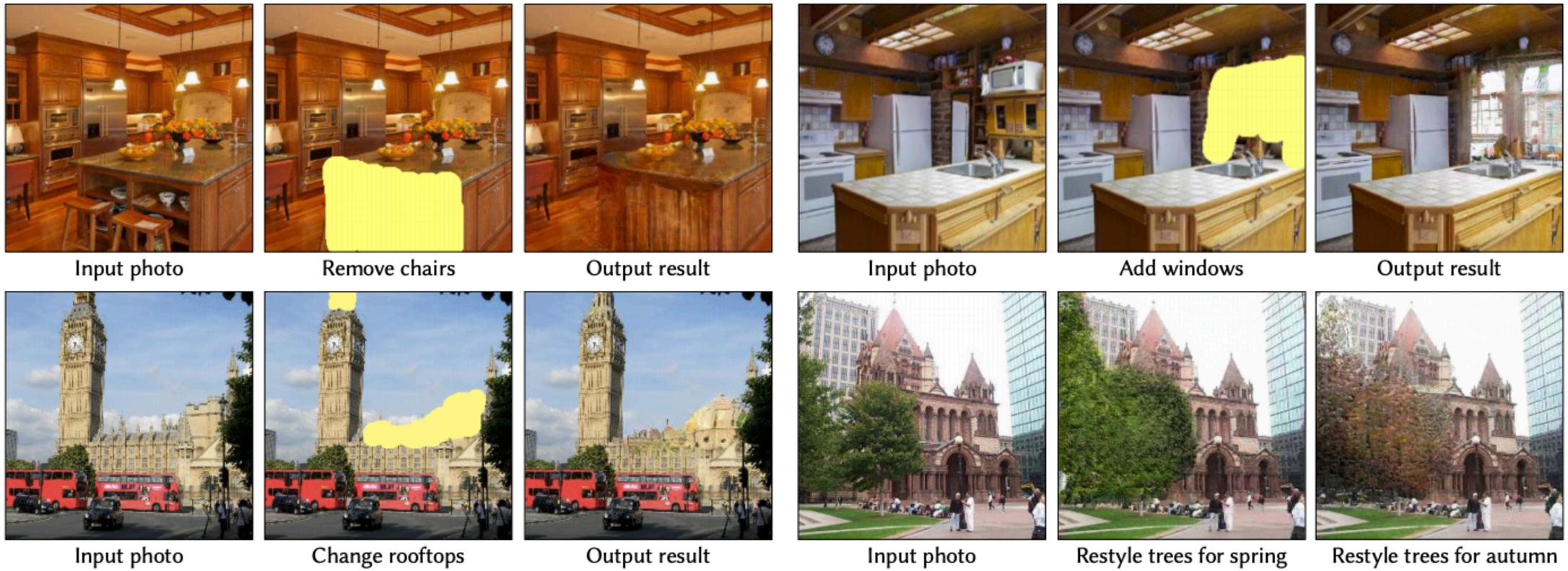


(d)

(e)

D. Bau et al. GAN Dissection: Visualizing and understanding generative adversarial networks. ICLR 2019

GANPaint demo



https://ganpaint-demo.vizhub.ai/

D. Bau et al. GAN Dissection: Visualizing and understanding generative adversarial networks. ICLR 2019

Parenthesis:
(Conditional generation

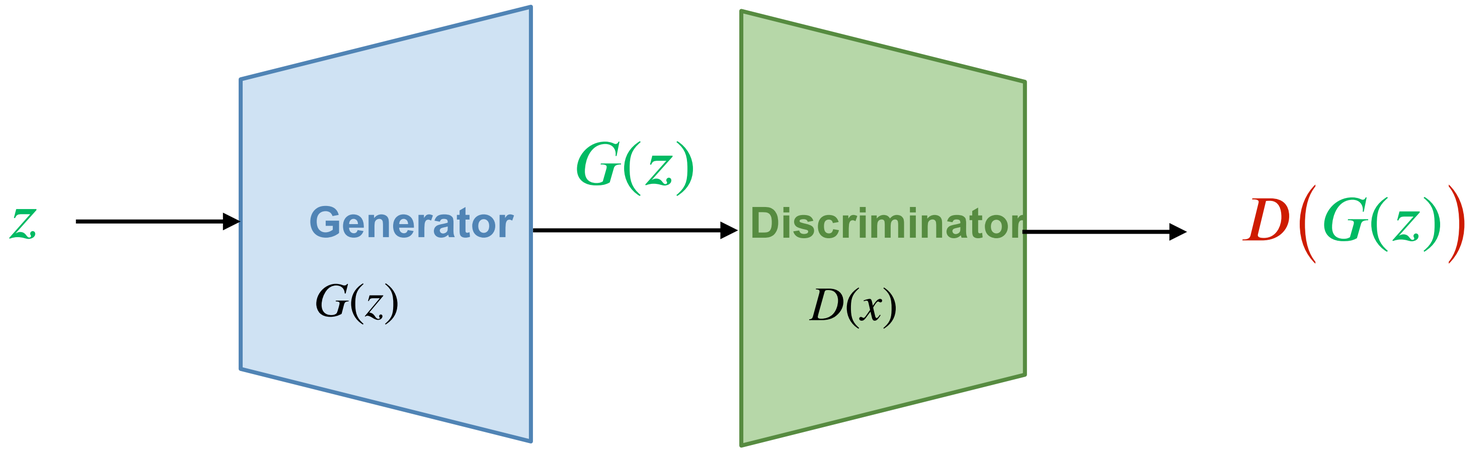
Conditional generation

- One may want to *control* the generation, e.g., instead of a random image sample, generating for a given:
  - category label (class conditioning),
  - natural language description (text conditioning),
  - ...

- Simply add the condition as input.

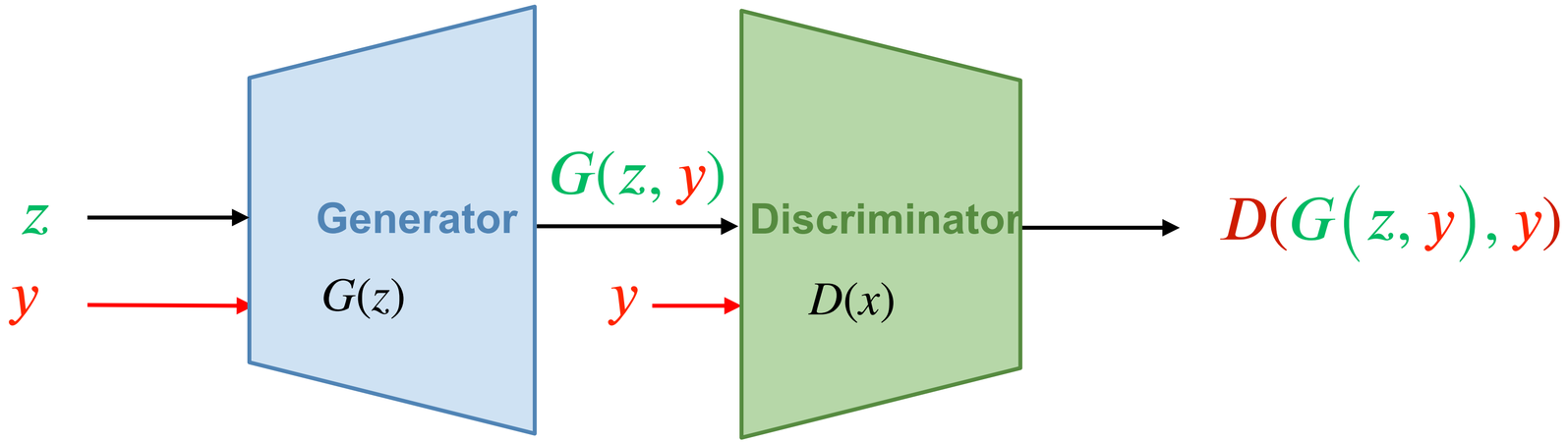
Conditional GANs:

- To condition the generation of samples on discrete side information (e.g., label) y, we need to add y as an input to **both generator and discriminator**



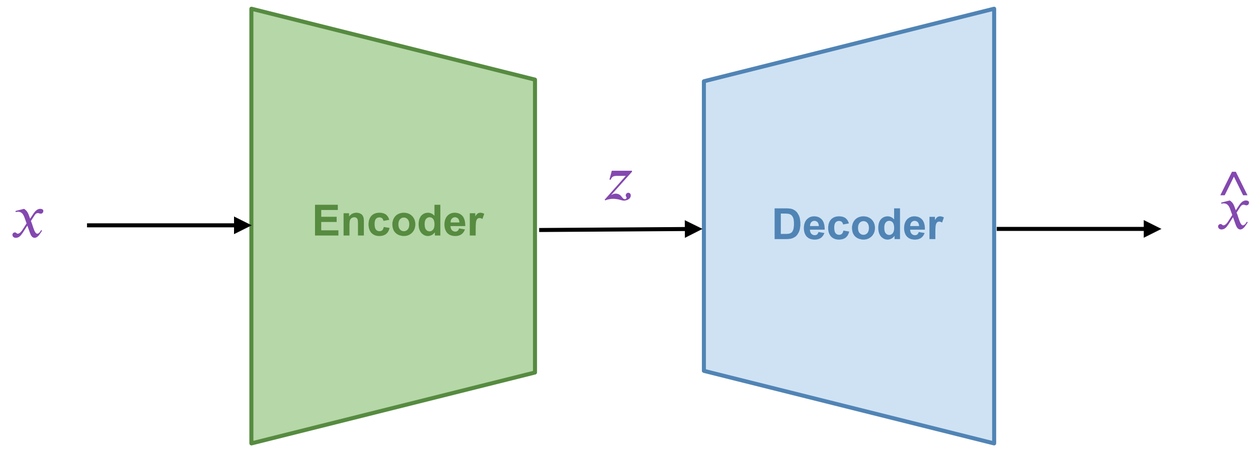
Conditional GANs:

- To condition the generation of samples on discrete side information (e.g., label) y, we need to add y as an input to **both generator and discriminator**



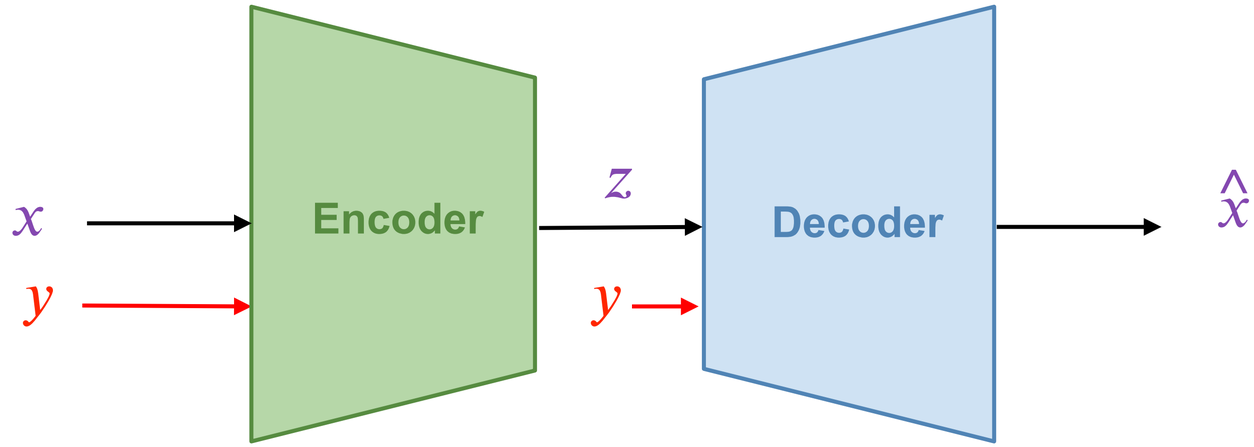
Conditional VAEs:

- To condition the generation of samples on discrete side information (e.g., label) y, we need to add y as an input to **both encoder and decoder**



Conditional VAEs:

- To condition the generation of samples on discrete side information (e.g., label) y, we need to add y as an input to **both encoder and decoder**



Parenthesis Closed:
Conditional generation)

BigGAN

- Class-conditional generation of ImageNet images up to 512 x 512 resolution

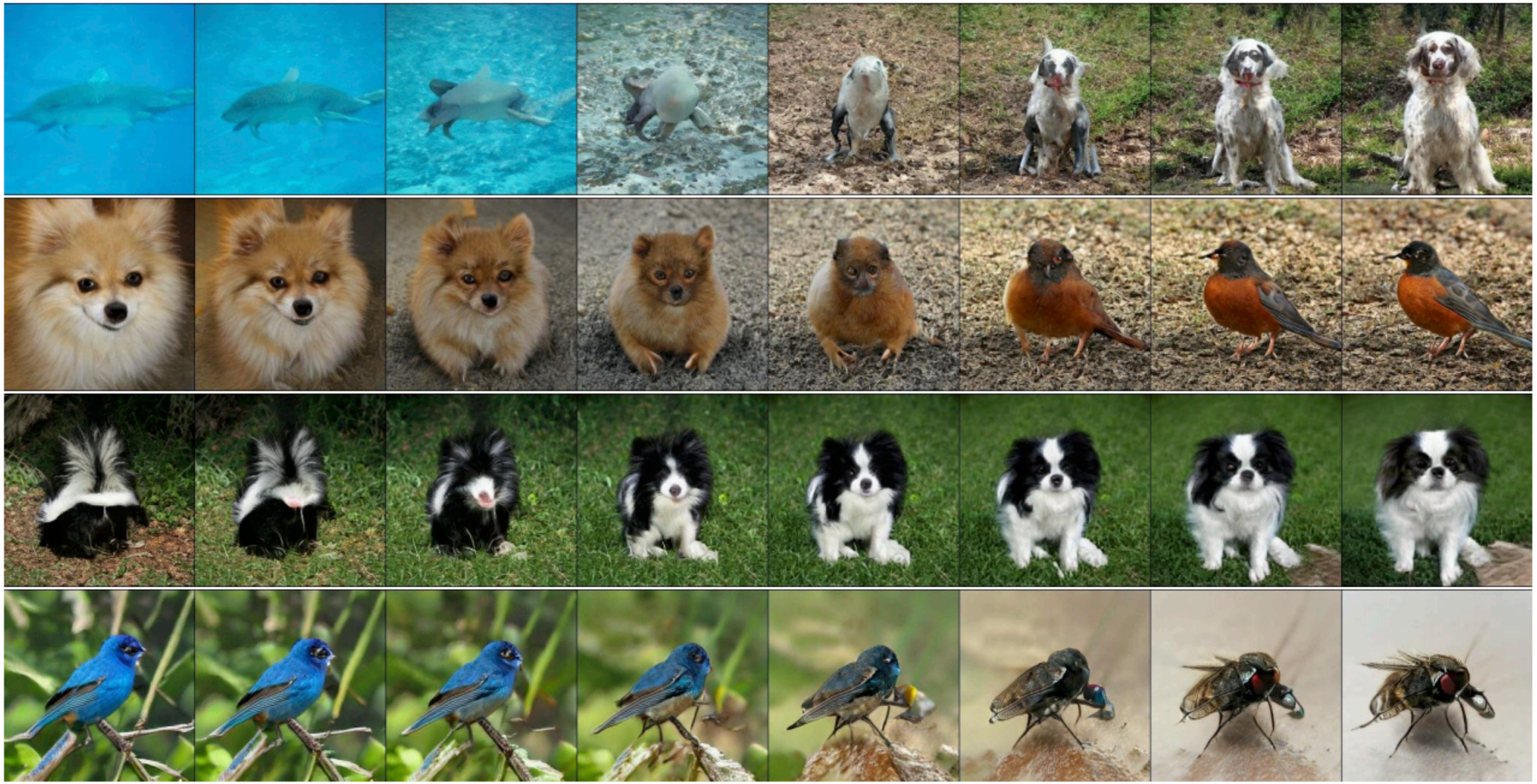
A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, ICLR 2019

BigGAN: Results

- Interpolation between class c with noise z held constant:

BigGAN: Results

- Interpolation between c, z pairs:



BigGAN: Results

- Difficult classes:

- Human bodies are still difficult today in 2024, although getting better

Progress in GANs

- **Progressive GAN, StyleGAN, StyleGan2** (higher quality)

T. Karras, T. Aila, S. Laine, J. Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation. ICLR 2018

T. Karras, S. Laine, T. Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. CVPR 2019

T. Karras et al. Analyzing and Improving the Image Quality of StyleGAN. CVPR 2020

- **GAN Dissection** (interpretability)

D. Bau et al. GAN Dissection: Visualizing and understanding generative adversarial networks. ICLR 2019

- **BigGan** (class-conditioned)

A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, ICLR 2019

- **Pix2Pix, CycleGan** (image-conditioned)

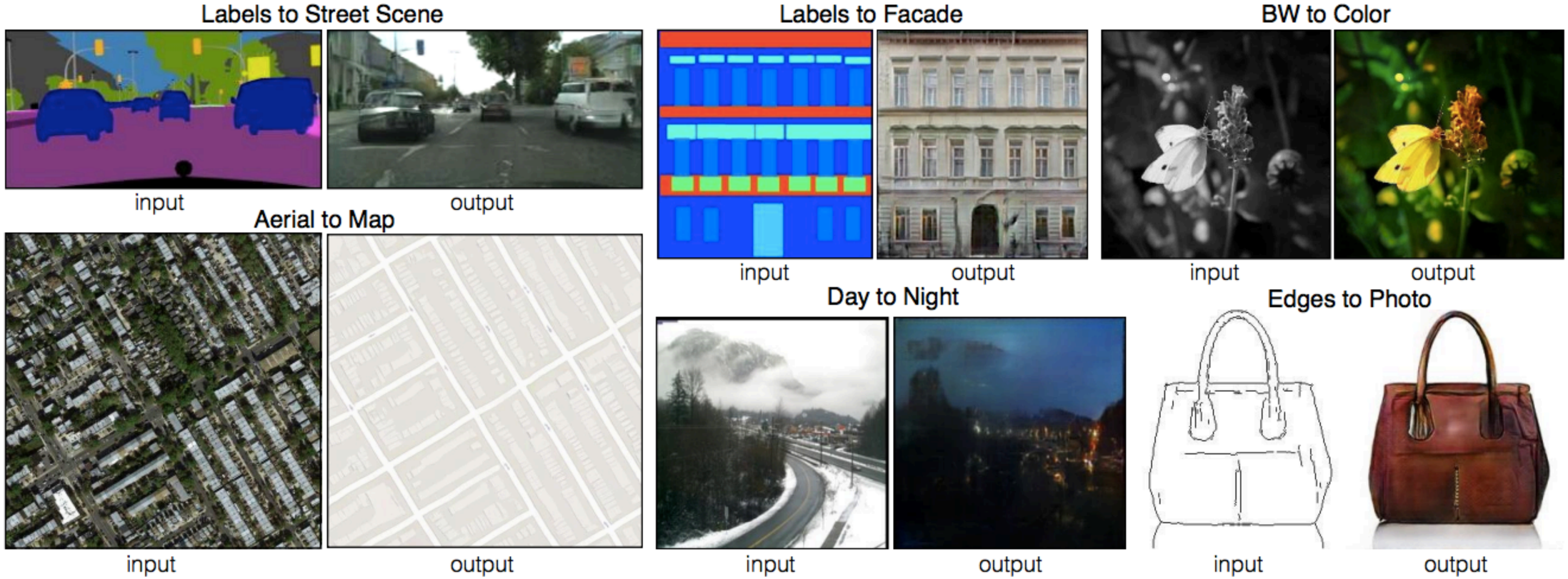
P. Isola, J.-Y. Zhu, T. Zhou, A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, CVPR 2017

J.-Y. Zhu, T. Park, P. Isola, A. Efros, Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, ICCV 2017

…

Paired image-to-image translation

- Deterministic



P. Isola, J.-Y. Zhu, T. Zhou, A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, CVPR 2017

Image-to-image translation

- Produce modified image y conditioned on input image x (note change of notation)
- Generator receives x as input
- Discriminator receives an x, y pair and has to decide whether it is real or fake

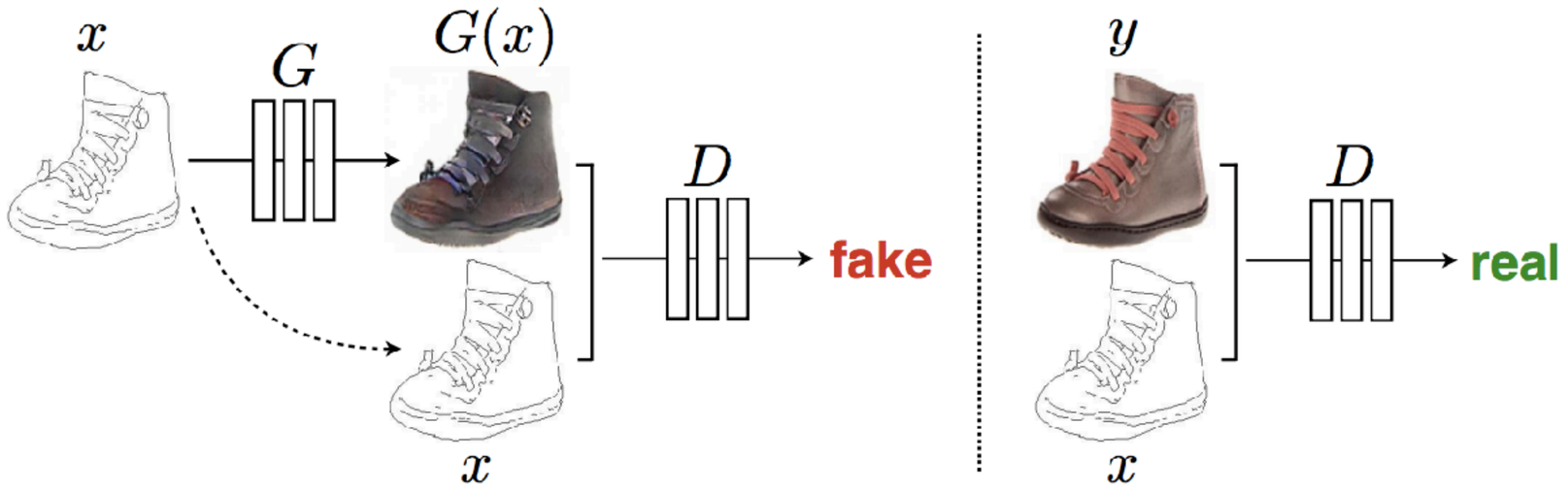
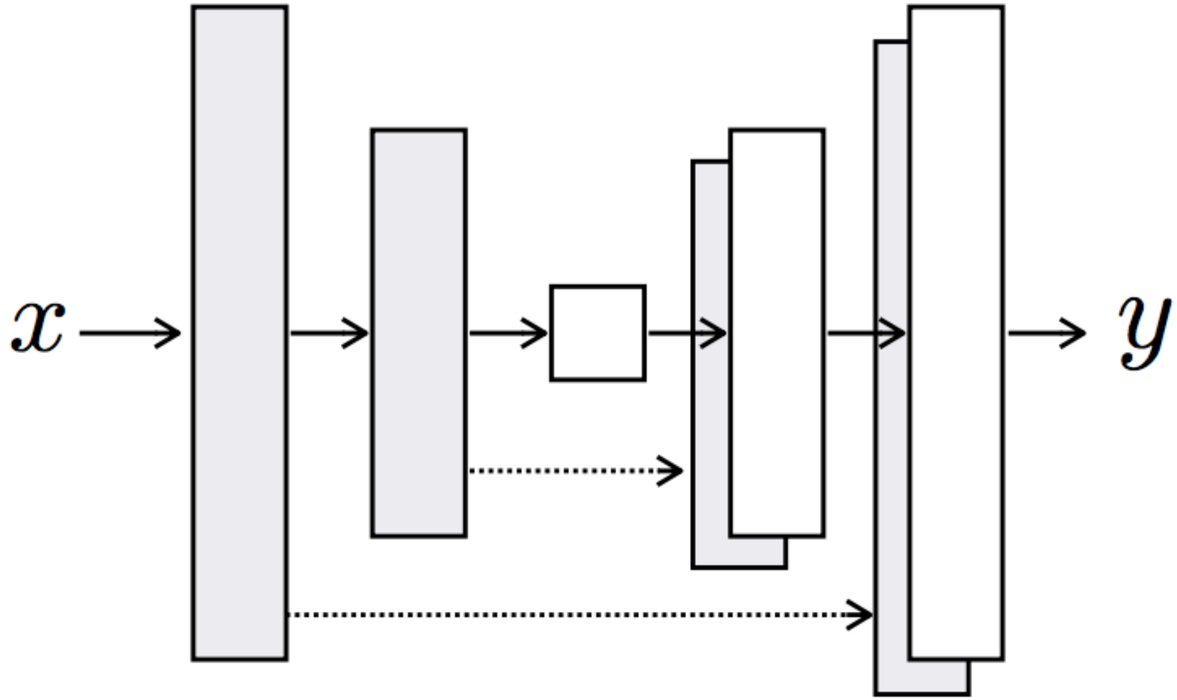


Image-to-image translation

- Generator architecture: U-Net



- Note: no z used as input, transformation is basically deterministic

Image-to-image translation

- Generator architecture: U-Net

Effect of adding skip connections to the generator

without skip connections

with skip connections

Image-to-image translation

- Generator loss: GAN loss plus L1 reconstruction penalty

$$G^* = \text{argmin}_G \text{max}_D \mathcal{L}_{GAN}(G, D) + \lambda \sum_{\cdot} \left\| y_i - G(x_i) \right\|_1$$

Image-to-image translation: Results

- Translating between maps and aerial photos

Image-to-image translation: Results

- Semantic labels to scenes

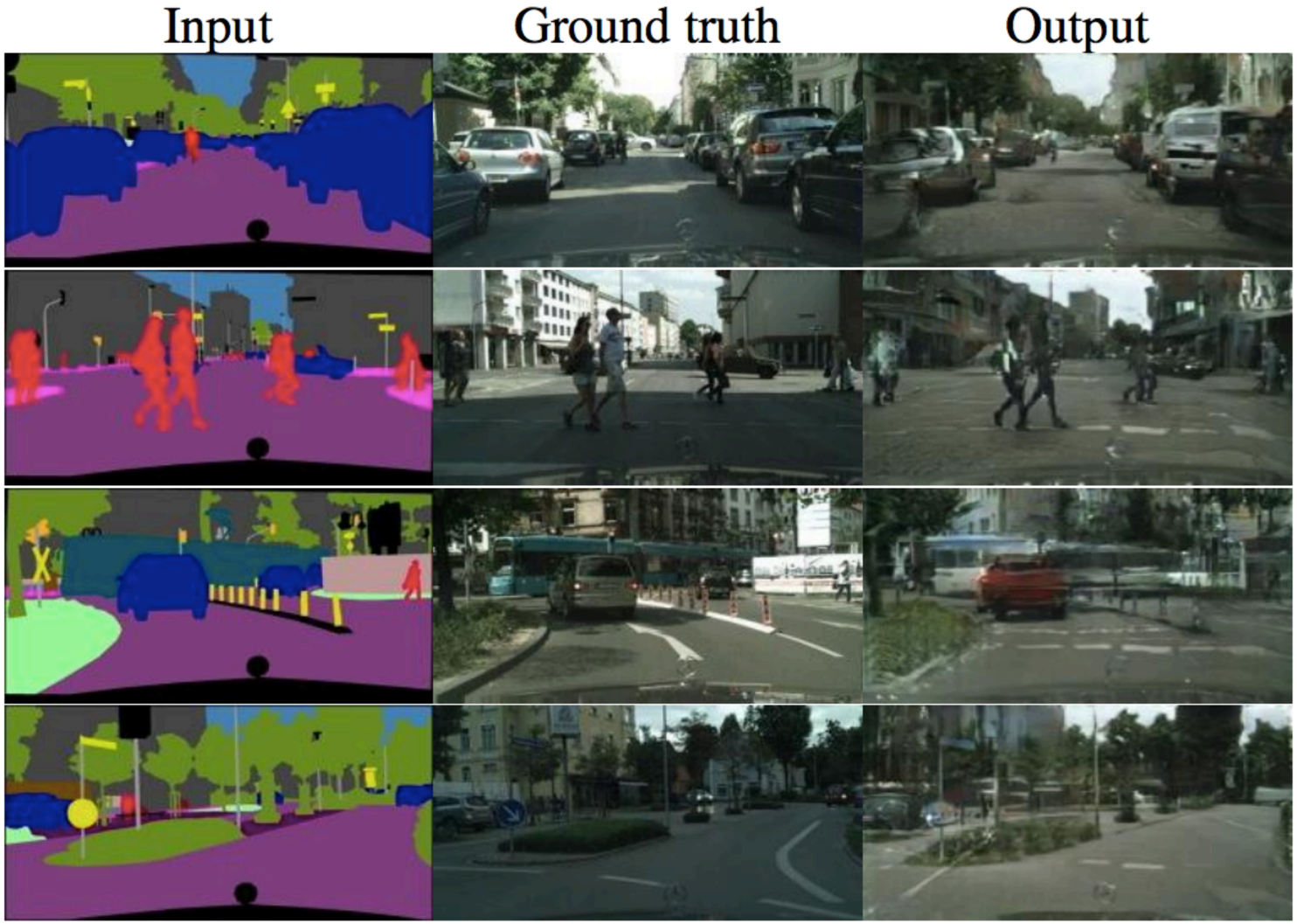


Image-to-image translation: Results

- Scenes to semantic labels

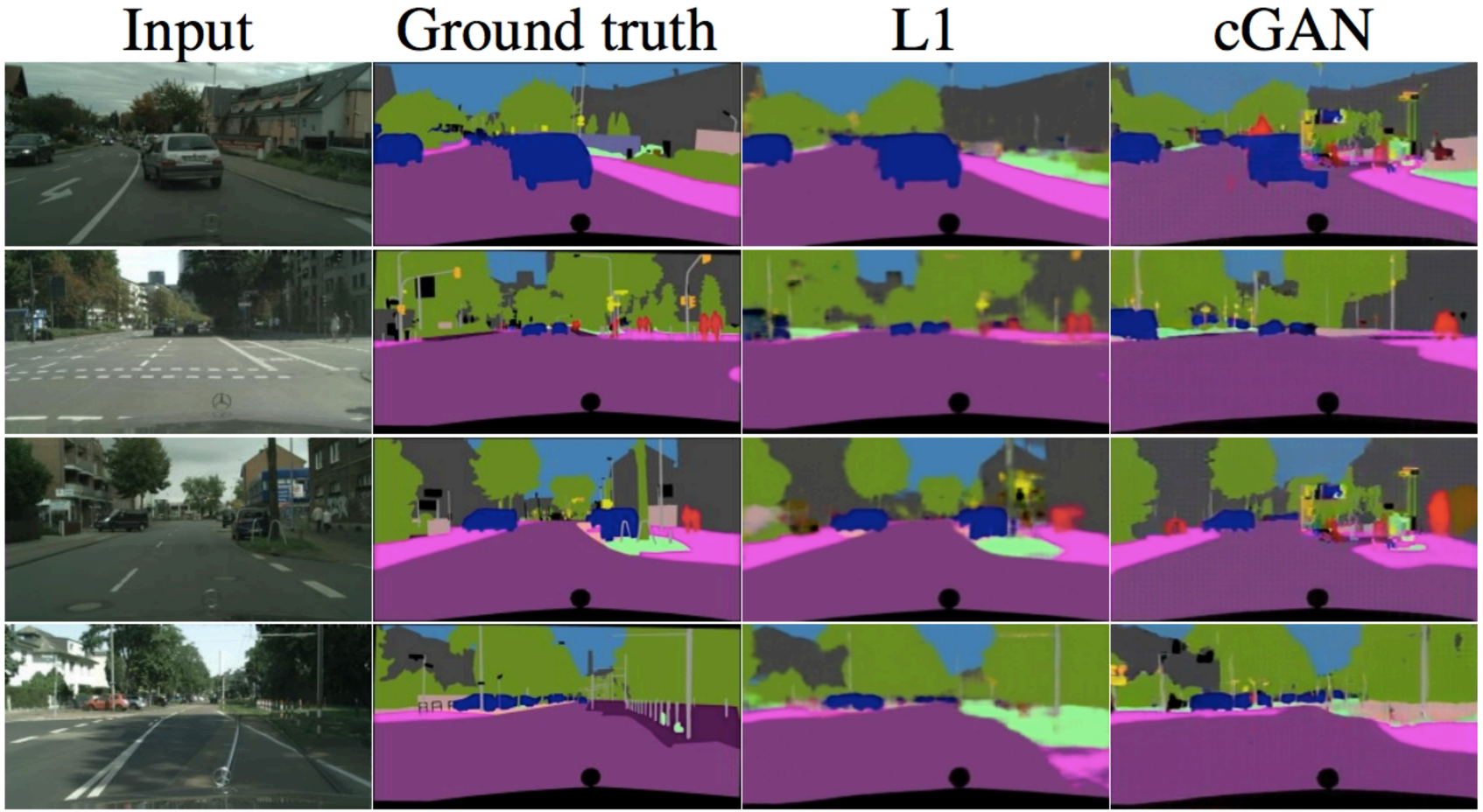


Image-to-image translation: Results

- Semantic labels to facades

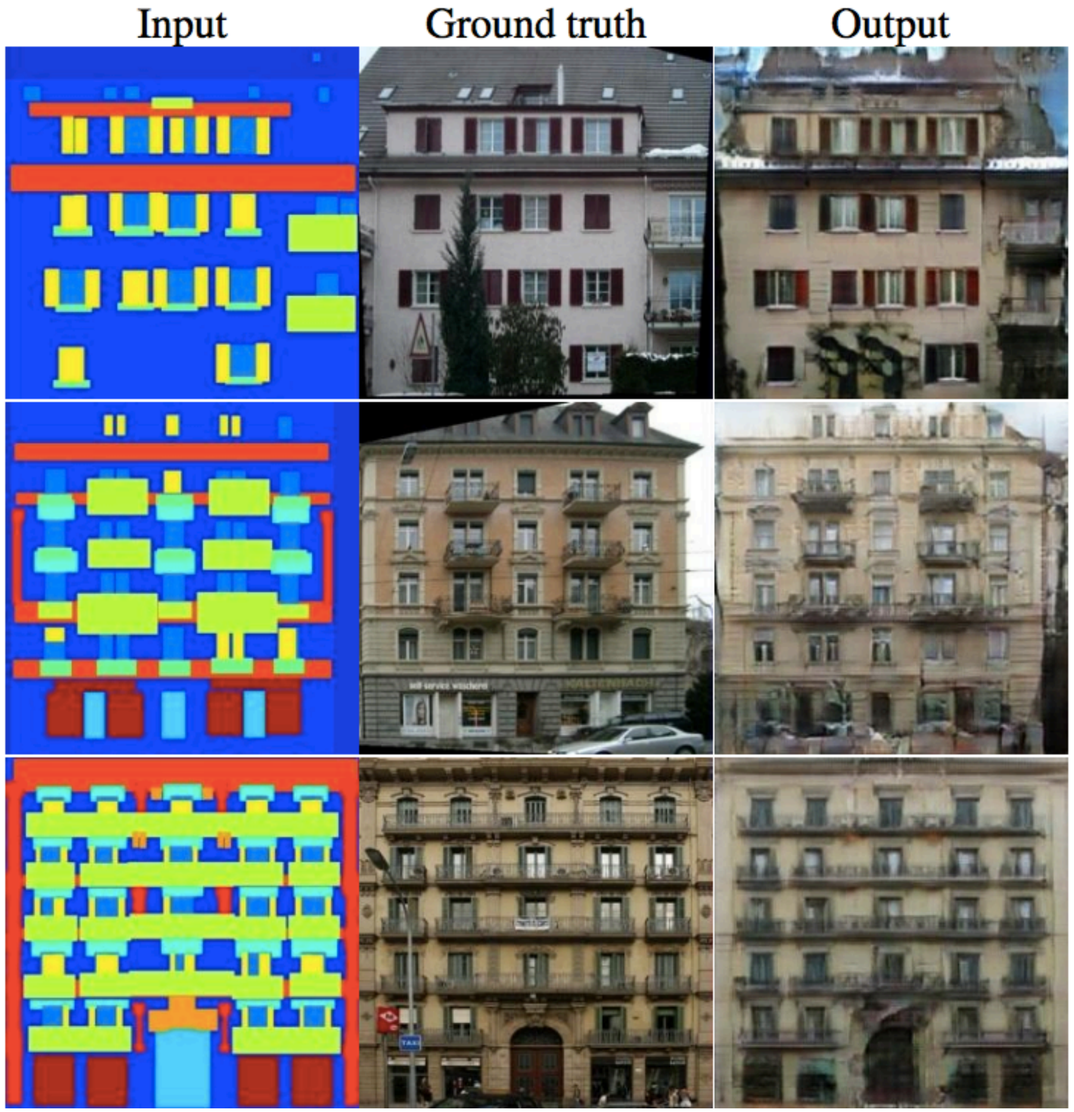


Image-to-image translation: Results

- Day to night

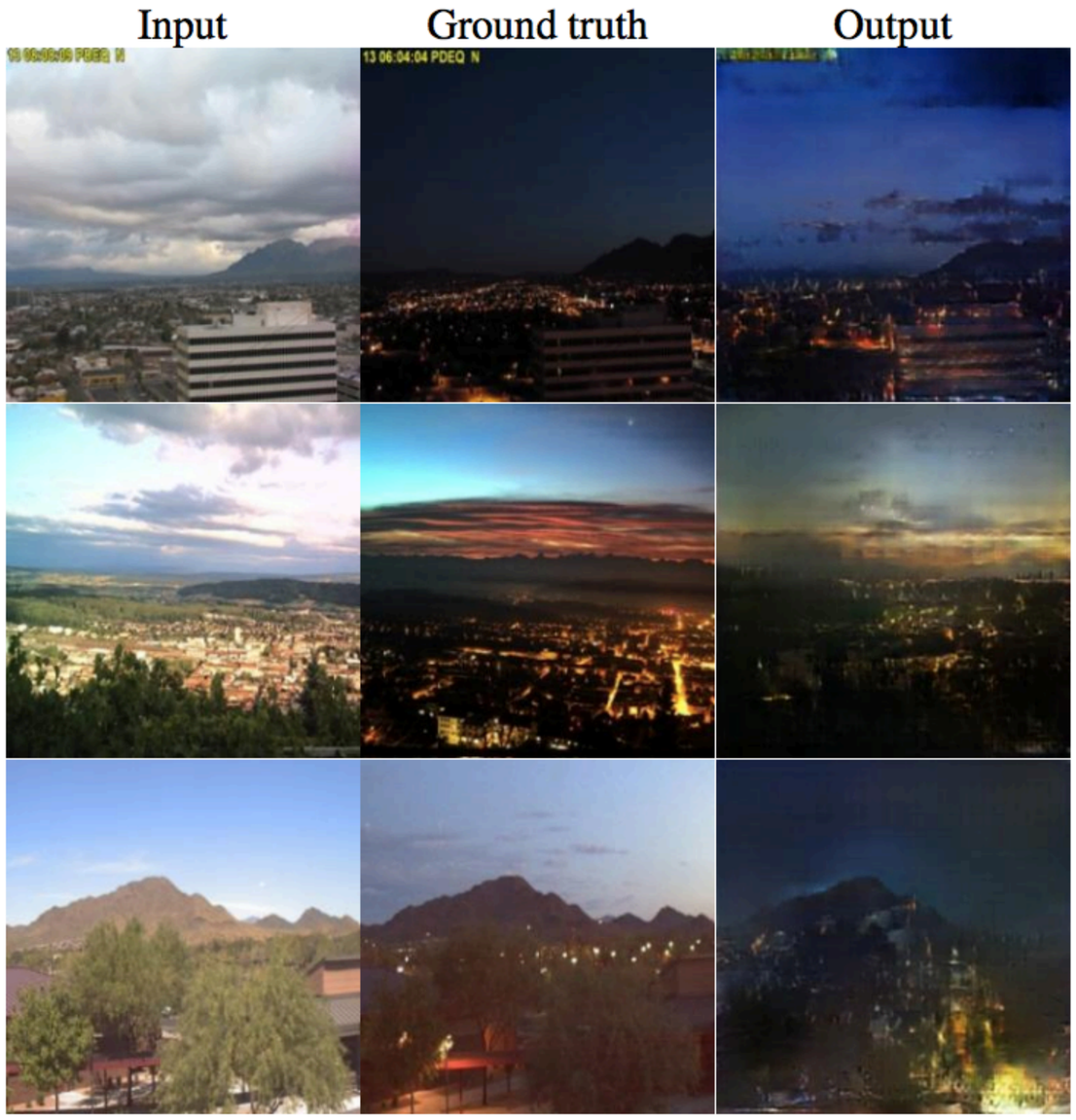
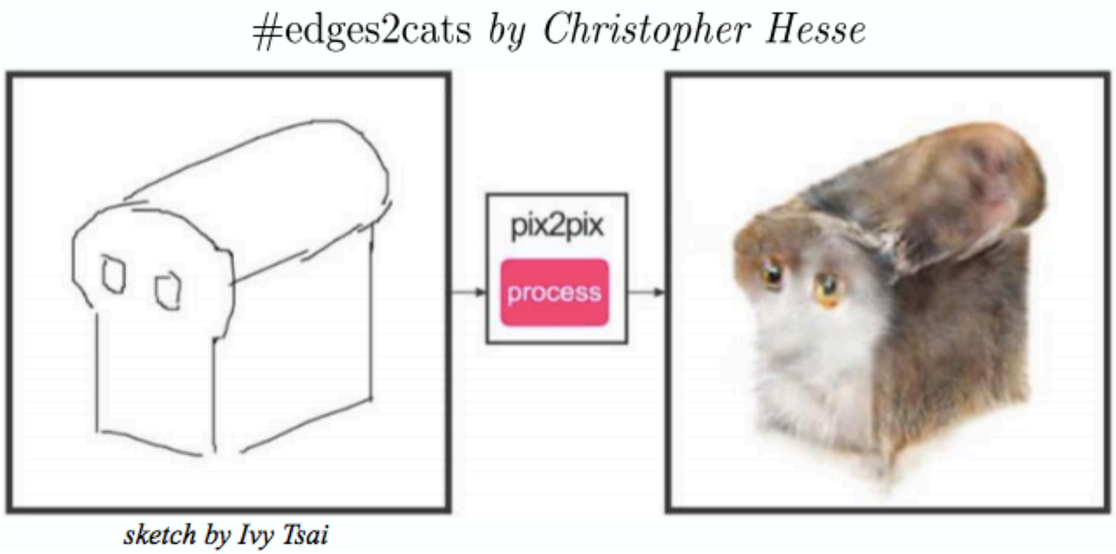


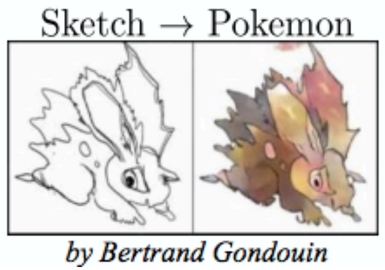
Image-to-image translation: Results

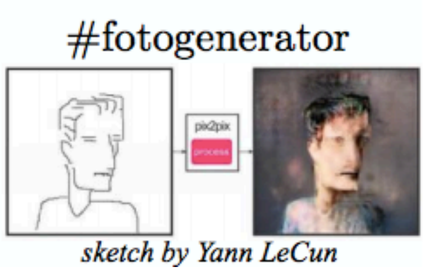
- Edges to photos

Image-to-image translation: Results

- [pix2pix demo]





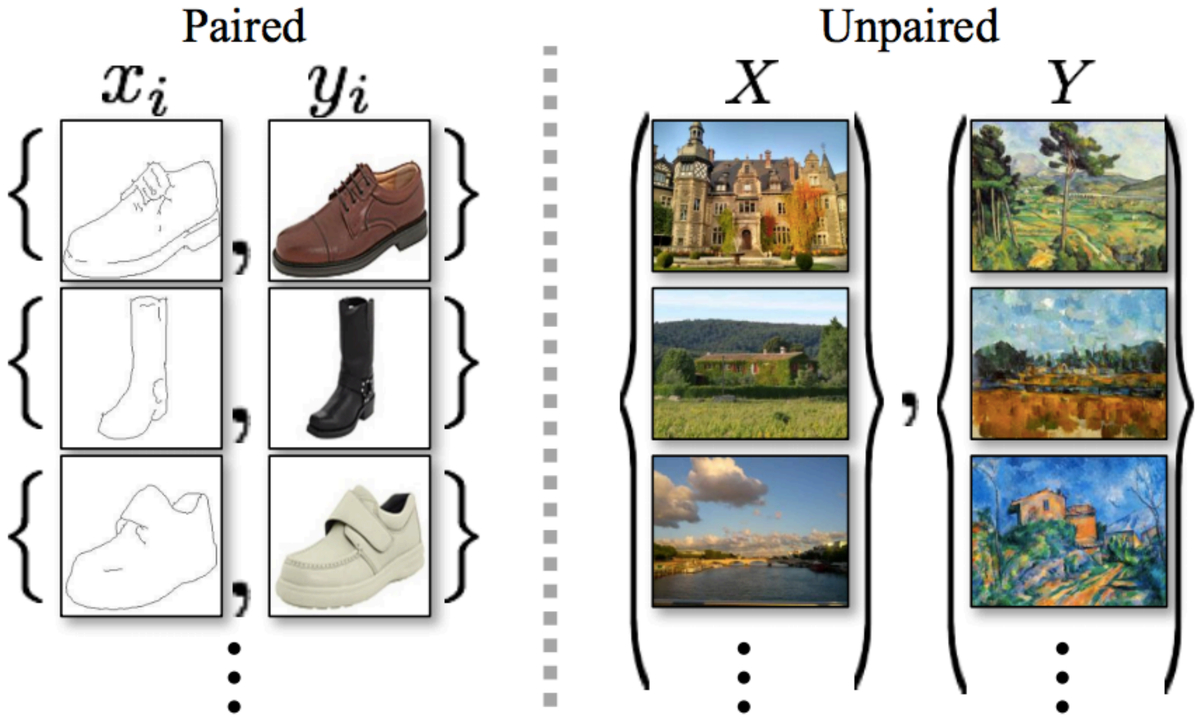


Pix2pix: Limitations

- Visual quality could be improved
- Requires $x,\ y$ pairs for training
- Does not model conditional distribution $P(y|x)$, returns a single mode instead

Unpaired image-to-image translation

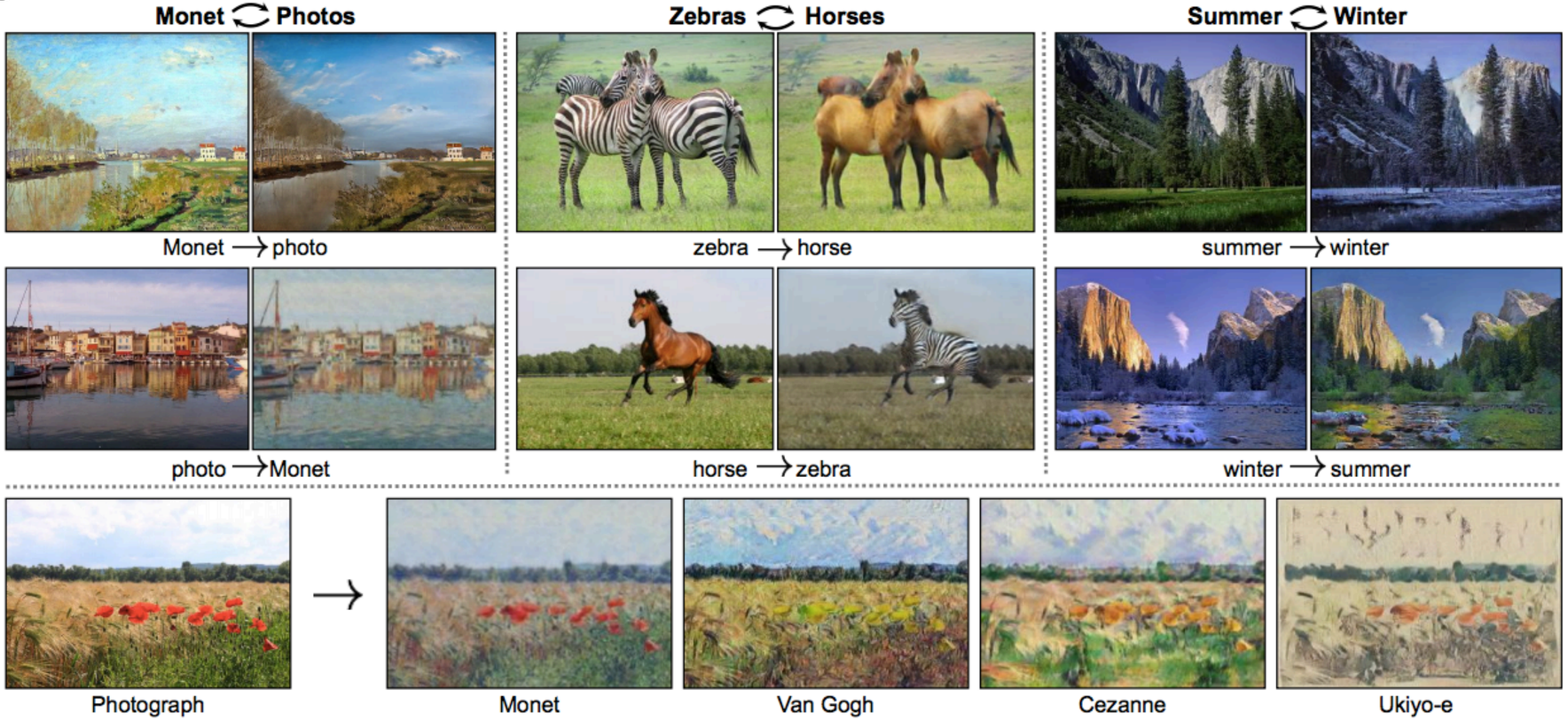
- Given two unordered image collections X and Y, learn to "translate" an image from one into the other and vice versa



J.-Y. Zhu, T. Park, P. Isola, A. Efros, Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, ICCV 2017

Unpaired image-to-image translation

- Given two unordered image collections X and Y, learn to "translate" an image from one into the other and vice versa

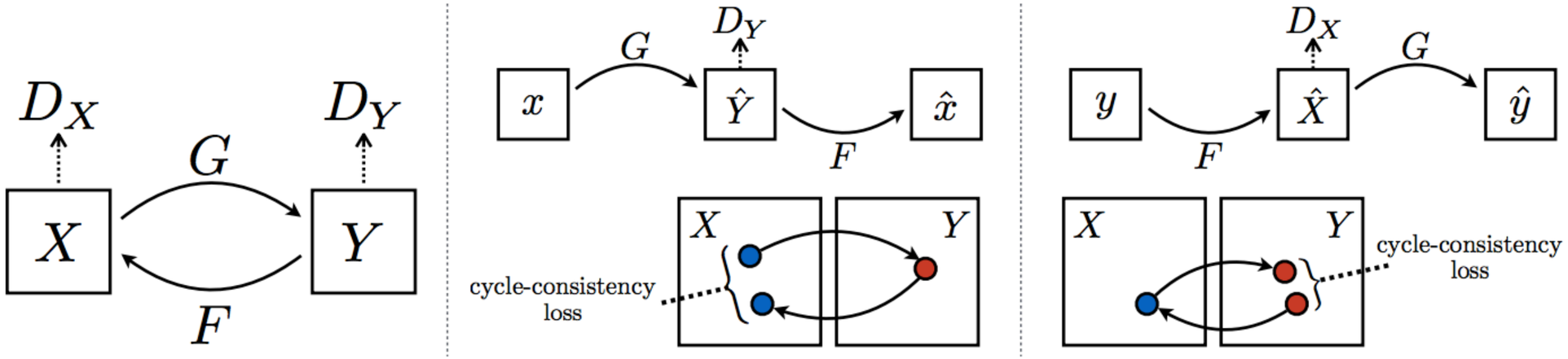


J.-Y. Zhu, T. Park, P. Isola, A. Efros, Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, ICCV 2017

CycleGAN

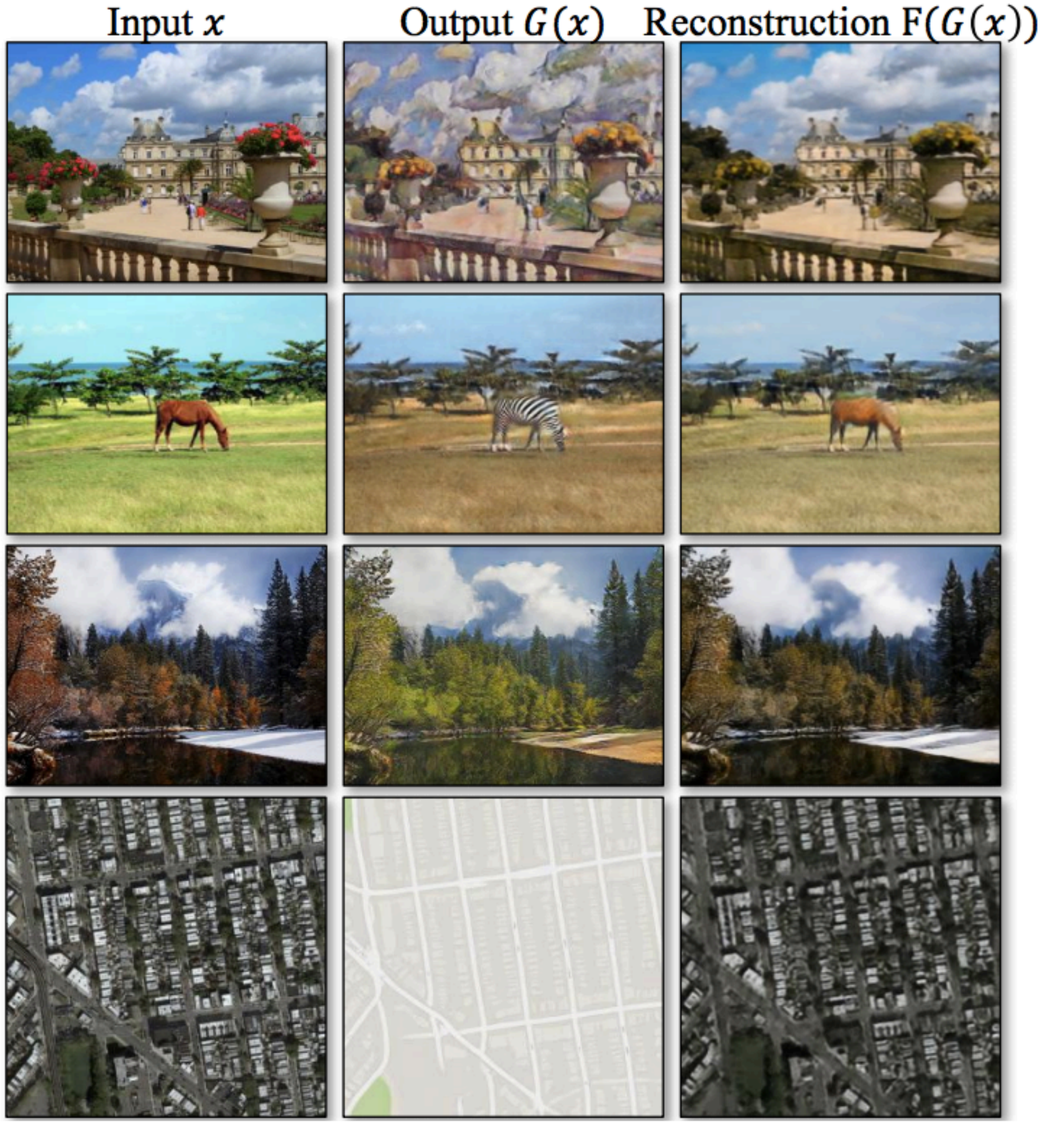
- Given: domains X and Y

- Train two generators F and G and two discriminators D_X and D_Y
  - G translates from X to Y, F translates from Y to X
  - D_X recognizes images from X, D_Y from Y
  - *Cycle consistency*: we want $F(G(x)) \approx x$ and $G(F(y)) \approx y$



CycleGAN

- Illustration of cycle consistency:



CycleGAN: Results

- Translation between maps and aerial photos