# Deep Generative Models

Hao Dong

Peking University

# Introduction

- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- Challenges
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- Challenges
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

# What and Why

Speech

# What and Why

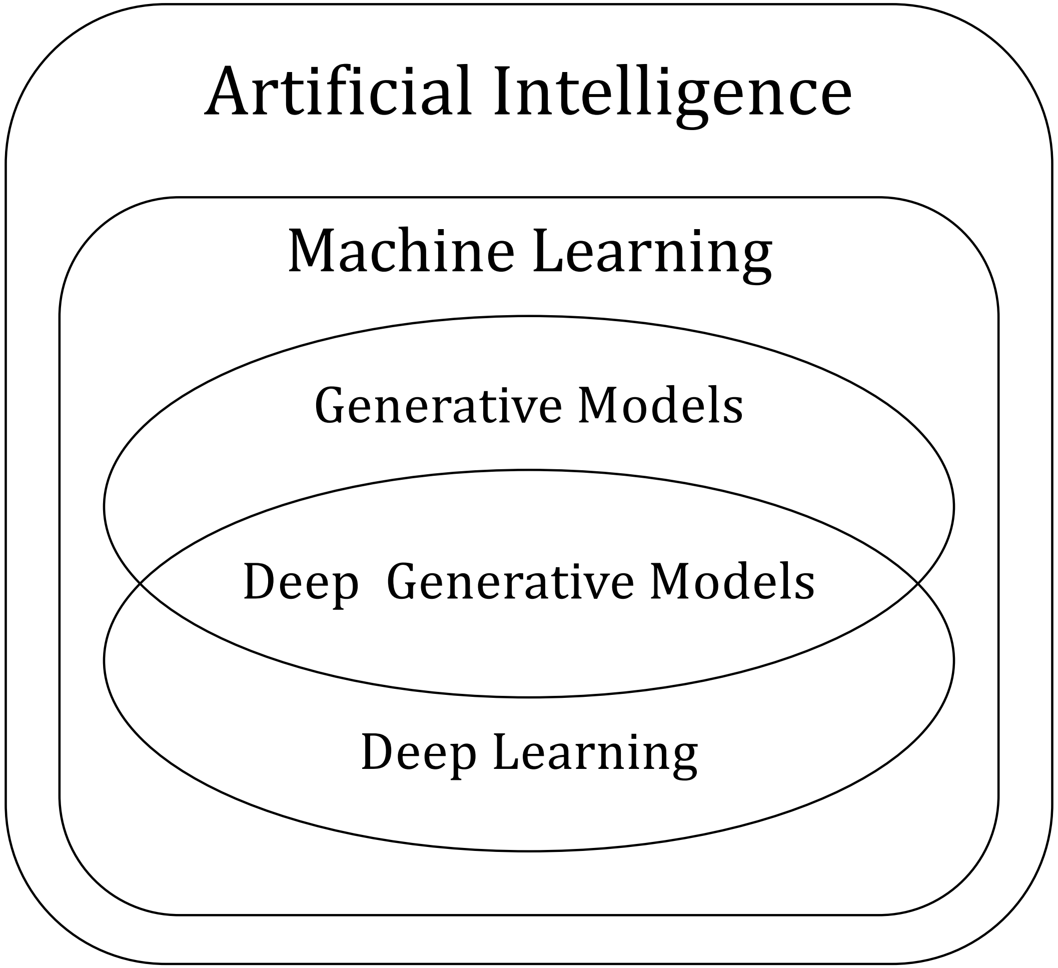
*"What I cannot create, I do not understand"*
--- Richard Feynman

Understand the complex and unstructured data

(image, text, speech, video …)

# Artificial Intelligence, Machine Learning, Deep Learning …

北京大學
PEKING UNIVERSITY



北京大學 PEKING UNIVERSITY

- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- Challenges
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

# Generative Models vs. Computer Graphics

- Generate data (e.g., image) in computer

**Computer Graphics**

Ball(color=yellow, position=(50, 100),…)
Ball(color=red, position=(30, 75),…)
Ball(color=blue, position=(30, 125),…)
…
…

**Computer Vision**


**"description"**

**"observation"**

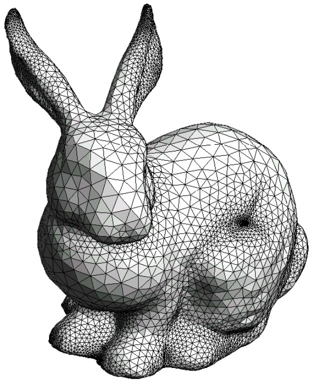
# Generative Models vs. Computer Graphics

- **Statistical Generative Models** are **data-driven** methods

**Computer Graphics** | **Statistical Generative Models**

**Prior Knowledge** | **Data**

Material, Physical Modeling, Lighting ….



# Generative Models vs. Computer Graphics

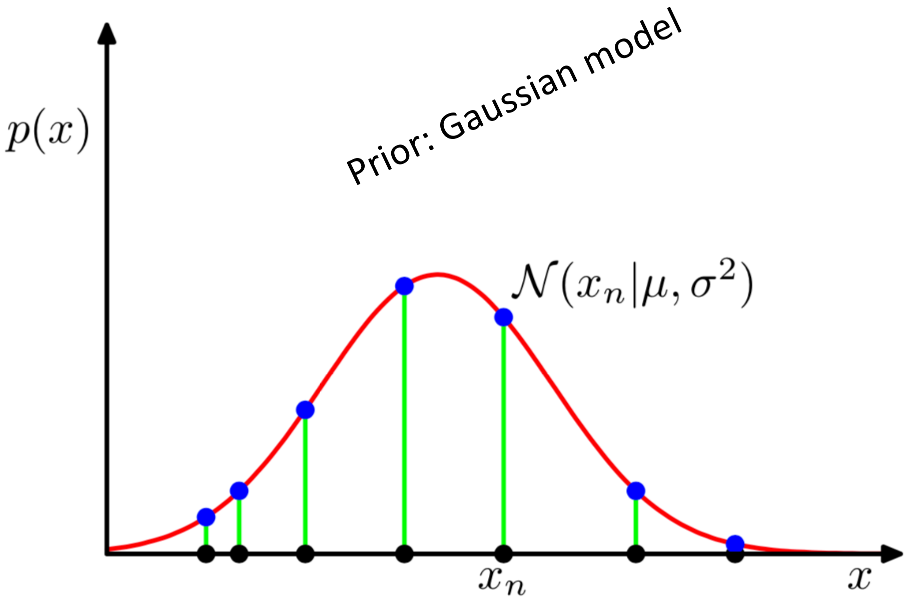
- **Computer Graphics**
  - Purely based on prior knowledge
  - Difficult to scale and generalize
  - Development is time-consuming

- **Machine Learning/Deep Learning**
  - Reduce the need of prior knowledge
  - Learn from data

- **Statistical/Deep Generative Models** still need some prior knowledge …
  - loss function, learning method, architecture, prior distribution (e.g., Gaussian)

## Generative Models vs. Computer Graphics

- **Statistical/Deep Generative Models**




- Given data samples
- Learn the probability distribution $p(x)$

So that

- It is generative because new data samples can be sampled from $p(x)$

$$x_{new} \sim p_x$$

# Generative Models vs. Computer Graphics

- **Statistical/Deep Generative Models**

The data distribution can be high-dimensional, like images

- Given data samples
- Learn the probability distribution $p(x)$

So that

- It is generative because new data samples can be sampled from $p(x)$

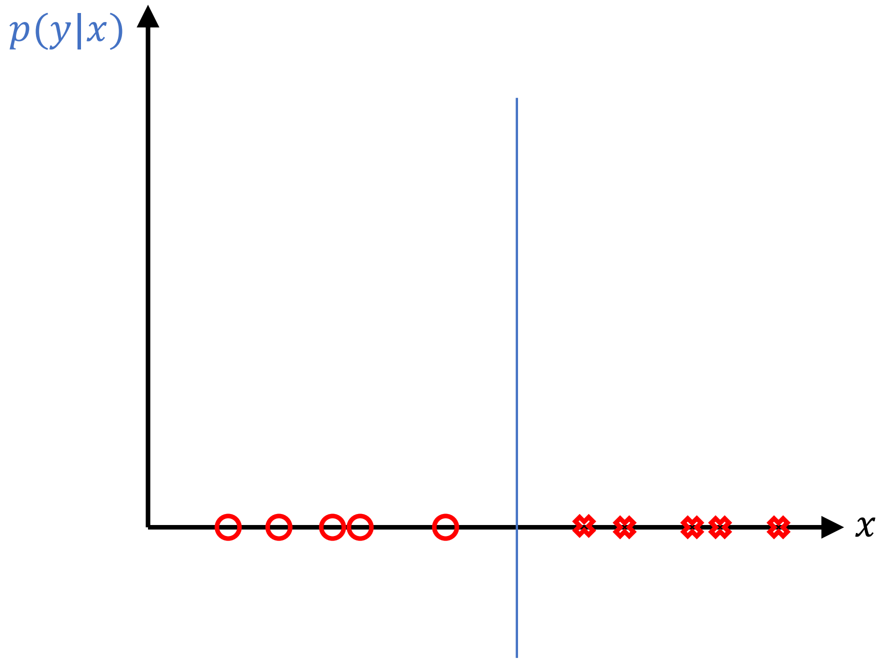
$$x_{new} \sim p_x$$

- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- Challenges
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

# Discriminative vs. Generative

**Discriminative models:** classify data

finding the **decision boundary** $P(Y|X)$

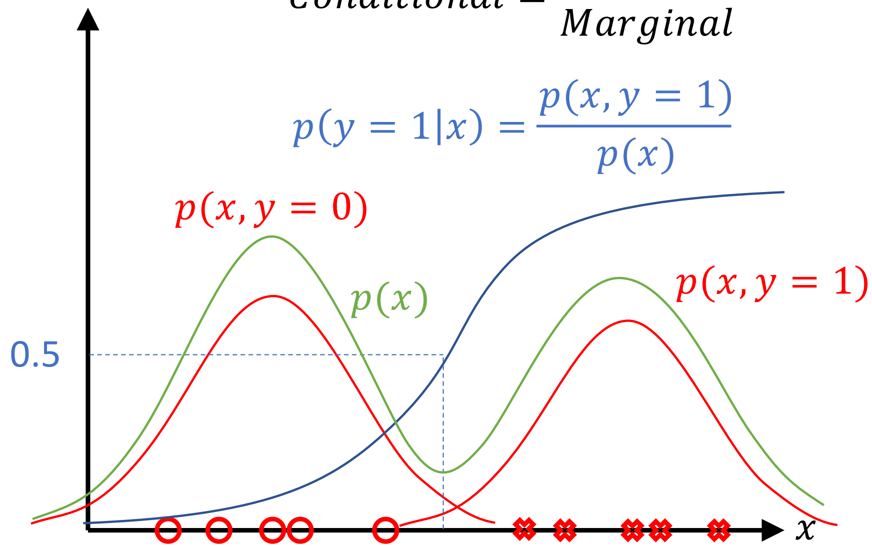



**Generative models:** generate data

finding **joint distribution** $P(Y,X)$

$$Conditional = \frac{Joint}{Marginal}$$

$$p(y=1|x) = \frac{p(x, y=1)}{p(x)}$$



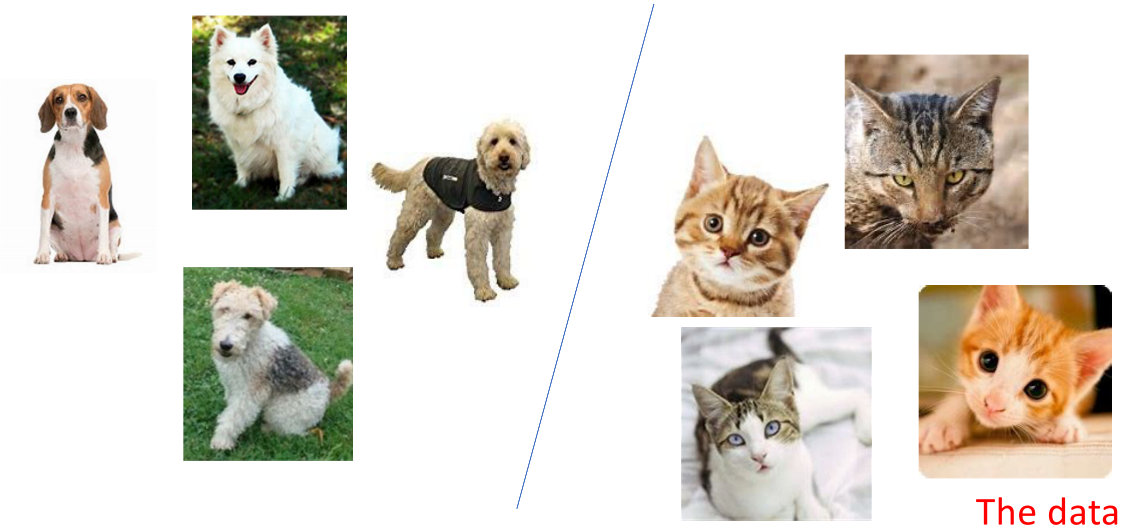

**Note: Generative models can perform both generative and discriminative tasks**

# Discriminative vs. Generative

**Discriminative models:** classify data

finding **conditional distribution** $P(Y|X)$

$P(Y = Cat|X =$  $) = 0.99$



Decision boundary

**Generative models:** generate data

finding **joint distribution** $P(Y,X)$

$Y = Cat, X =$ 

$Y = Dog, X =$ 

The data distribution can be high-dimensional, like images

## Discriminative vs. Generative

- Discriminative models do not model/learn the probability distribution of data $p(x)$
  and find the decision boundary directly to form $p(y|x)$

- Generative models need to
  first model/learn the probability distribution of data $p(x)$
  and the joint probability distribution $p(x, y)$
  and the estimated the conditional probability $p(y|x) = \frac{p(x,y)}{p(x)}$

- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- Challenges
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

# Selected Generative Applications

**We usually study generative models with:**

**image**

**text**

**speech**

**…**

**or their combinations**

## Selected Generative Applications

**Discriminative models**

$P(Y = Cat|X =$  $)$

**"Unconditional" generative models:** generate data from a prior distribution

$$P(X, Z) = P(X|Z)P(Z)$$

$P(X =$  $|Z = N(0,1))$

# Selected Generative Applications

**“Class” conditional generative models**

$P(X =$  $|Y = Cat)$

**“Text” conditional generative models**

$P(X =$  $|Y = \text{“}a\ flower\ with\ white\ petals\ and\ yellow\ stamen\text{”})$

**“Text-image” conditional generative models**

**Joint distribution**

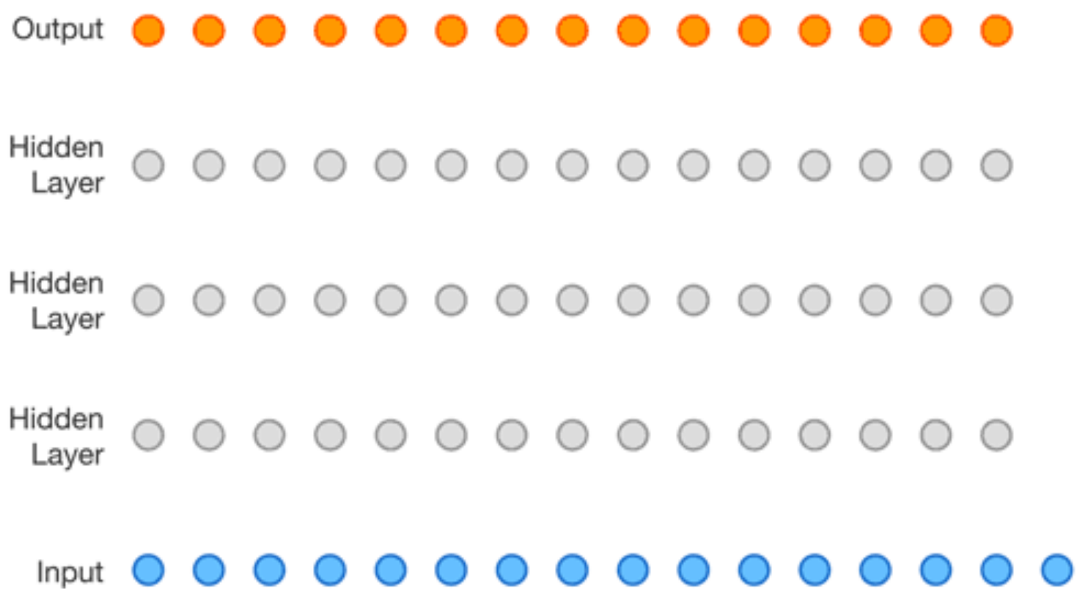
$P(X =$  $|Y_1 =$  $, Y_2 = \text{“}a\ yellow\ bird\ with\ grey\ wings\text{”})$

Generative Adversarial Text to Image Synthesis. *S. Reed, Z. Akata et al. ICML. 2016.*
Semantic Image Synthesis via Adversarial Learning. *H. Dong, S. Yu et al. ICCV 2017.*

## Selected Generative Applications

### **Wavenet: Text to Speech**

$$P(X = speech|Y = sentence)$$




Parametric

Concatenative

**WaveNet**

Unconditional

Music



WaveNet: A Generative Model for Raw Audio. *A. Oord, S. Dieleman et al. arXiv. 2016.*

# Selected Generative Applications

## Image Super Resolution

$$P(High\ resolution\ image \mid Low\ resolution\ image)$$


Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *C. Ledig, L. Theis et al. CVPR 2017.*

# Selected Generative Applications

**Image Super Resolution**

$$P(High\ quality\ image\ |\ Low\ quality\ image)$$

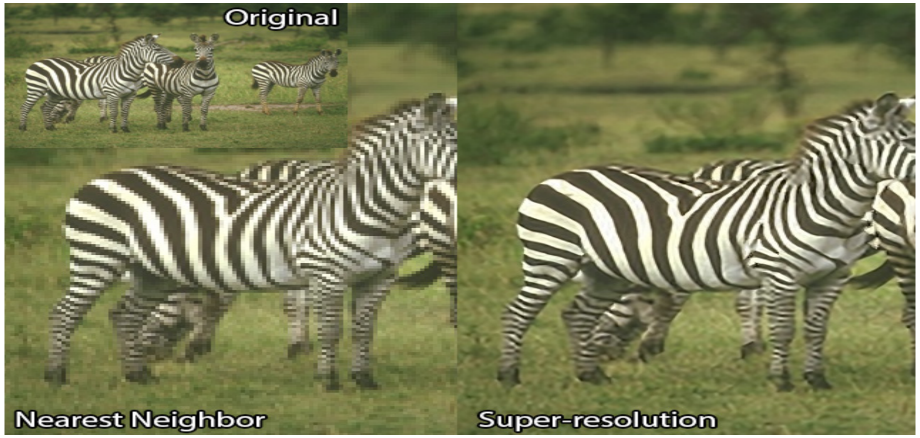
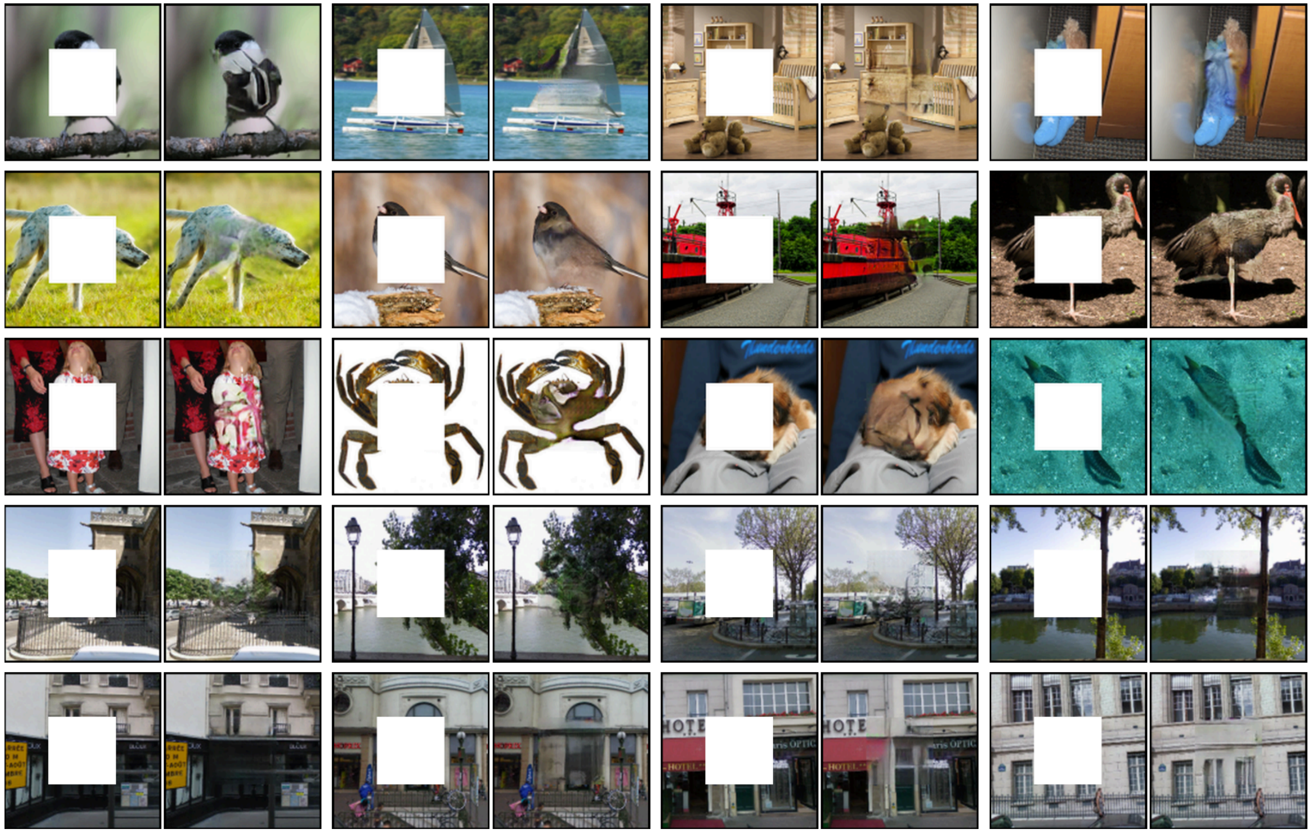




Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *C. Ledig, L. Theis et al. CVPR 2017.*

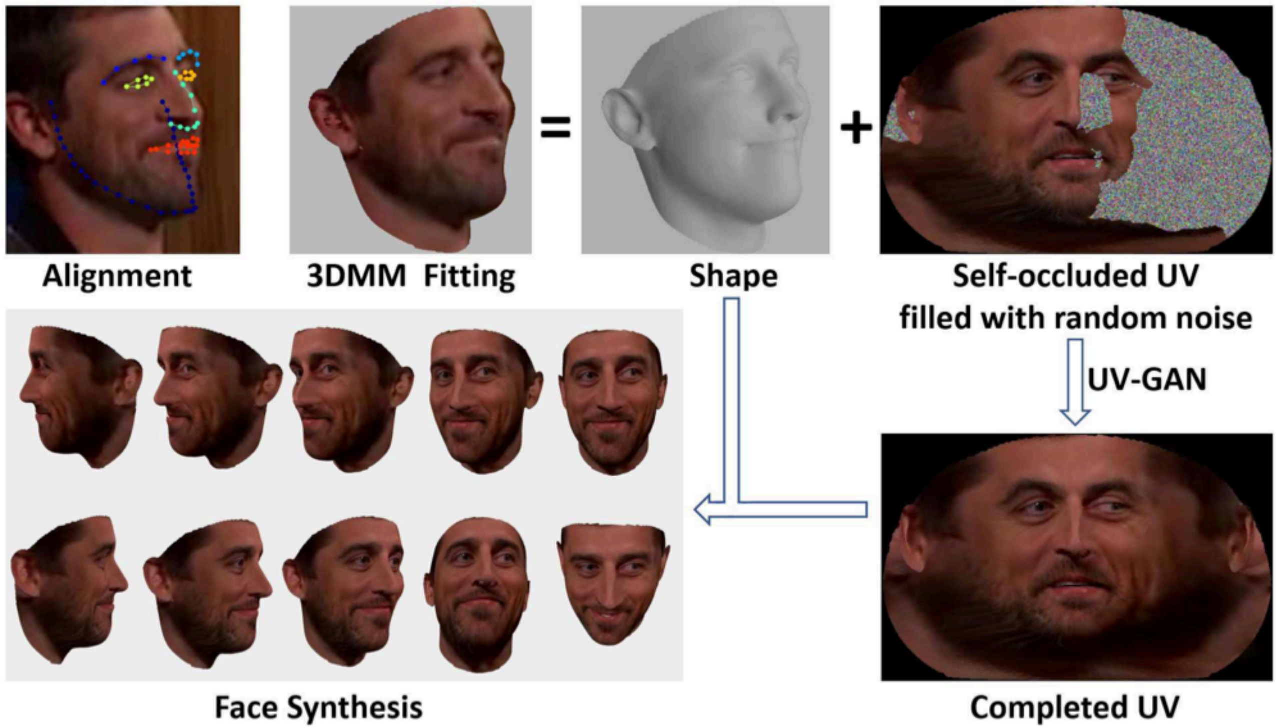
## Selected Generative Applications

- **Image inpainting**



Context Encoders: Feature Learning by Inpainting. *D. Pathak, J. Donahue. CVPR. 2017*

- **2D→3D via Image Inpainting**

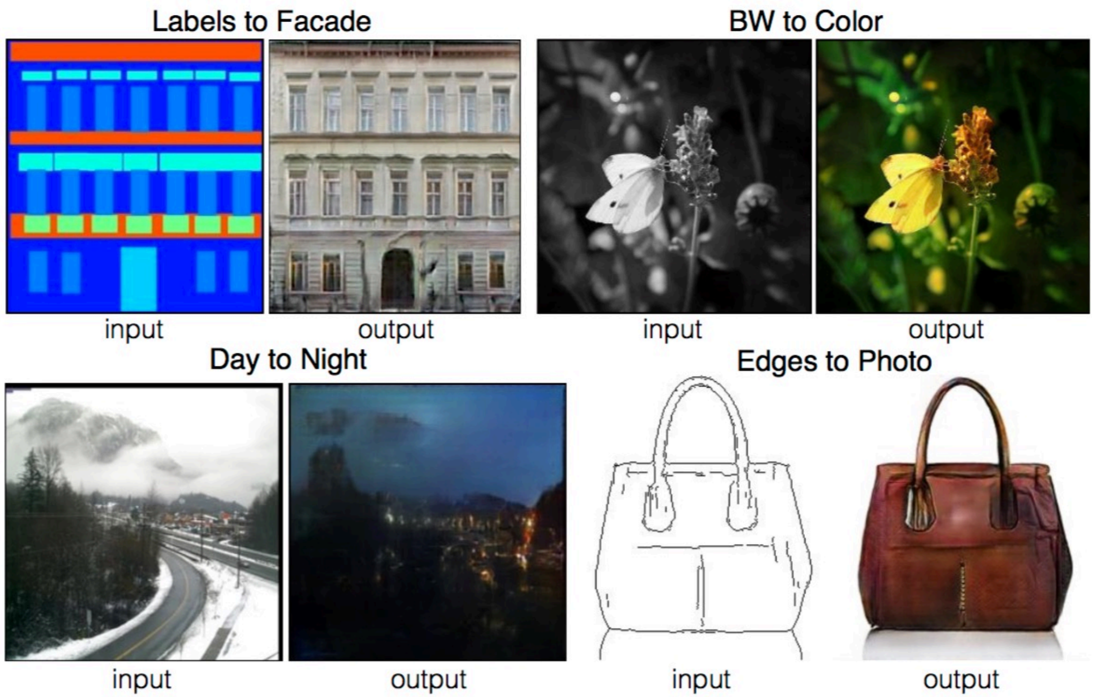



UV-GAN: Adversarial Facial UV Map Completion for Pose-invariant Face Recognition.
*J. Deng, S. Cheng et al. CVPR. 2018.*

# Selected Generative Applications

## Image-to-Image Translation

$$P(image\ from\ domain\ B \mid image\ from\ domain\ A)$$



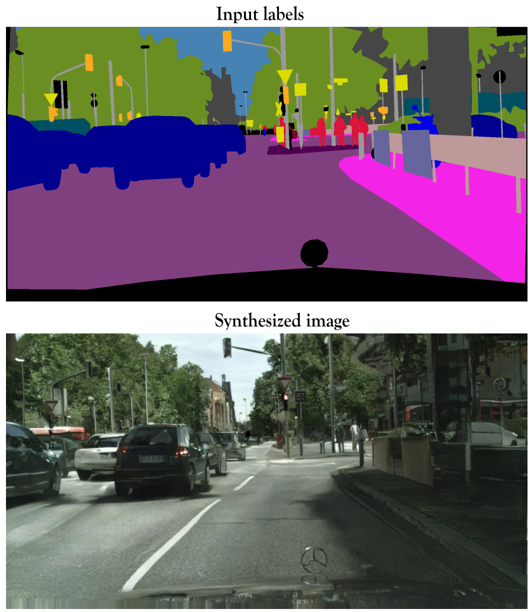



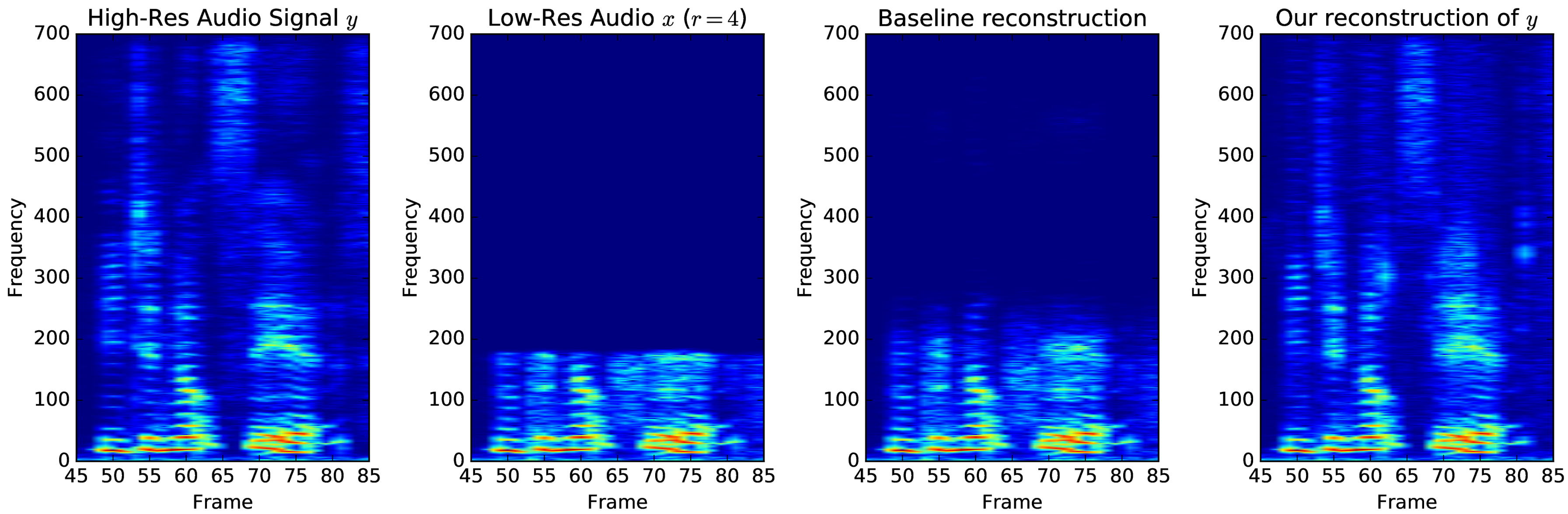


Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *C. Ledig, L. Theis et al. CVPR 2017.*

## Selected Generative Applications

### Audio Super Resolution

$$P(High\ resolution\ signal\ |\ Low\ resolution\ signal)$$




Audio Super Resolution Using Neural Networks. *V. Kuleshov, Z. Enam et al. ICLR Workshop 2017.*

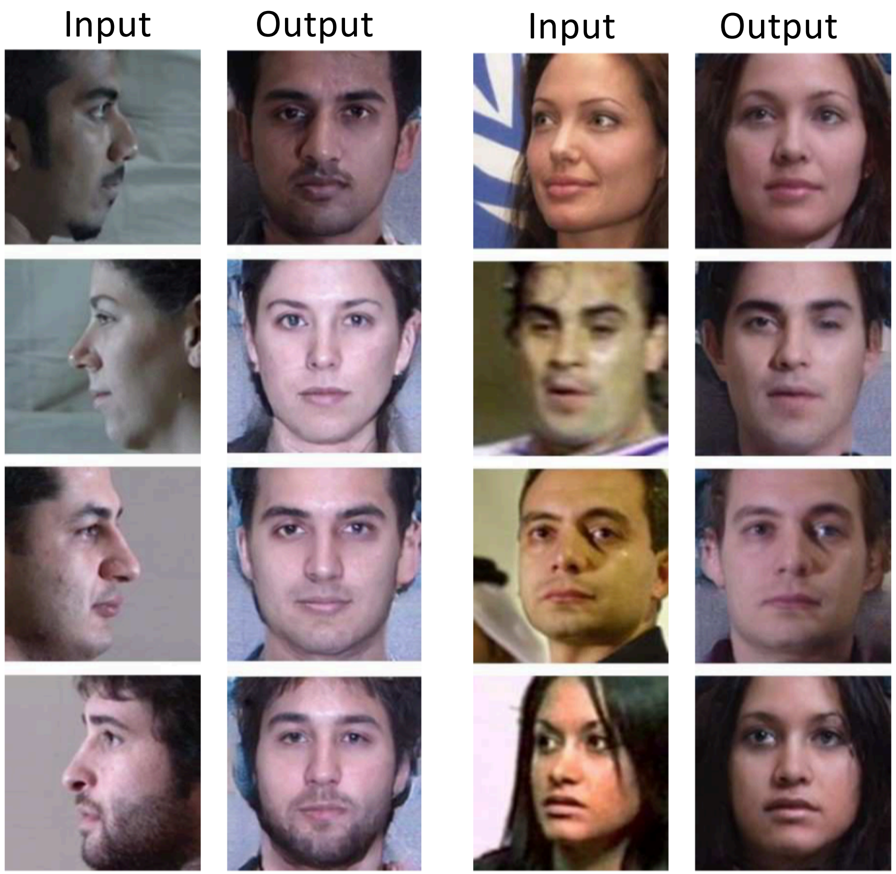
## Selected Generative Applications

**DeepFake**

$$P(me \mid you)$$

Selected Generative Applications

- **Face Rotation**

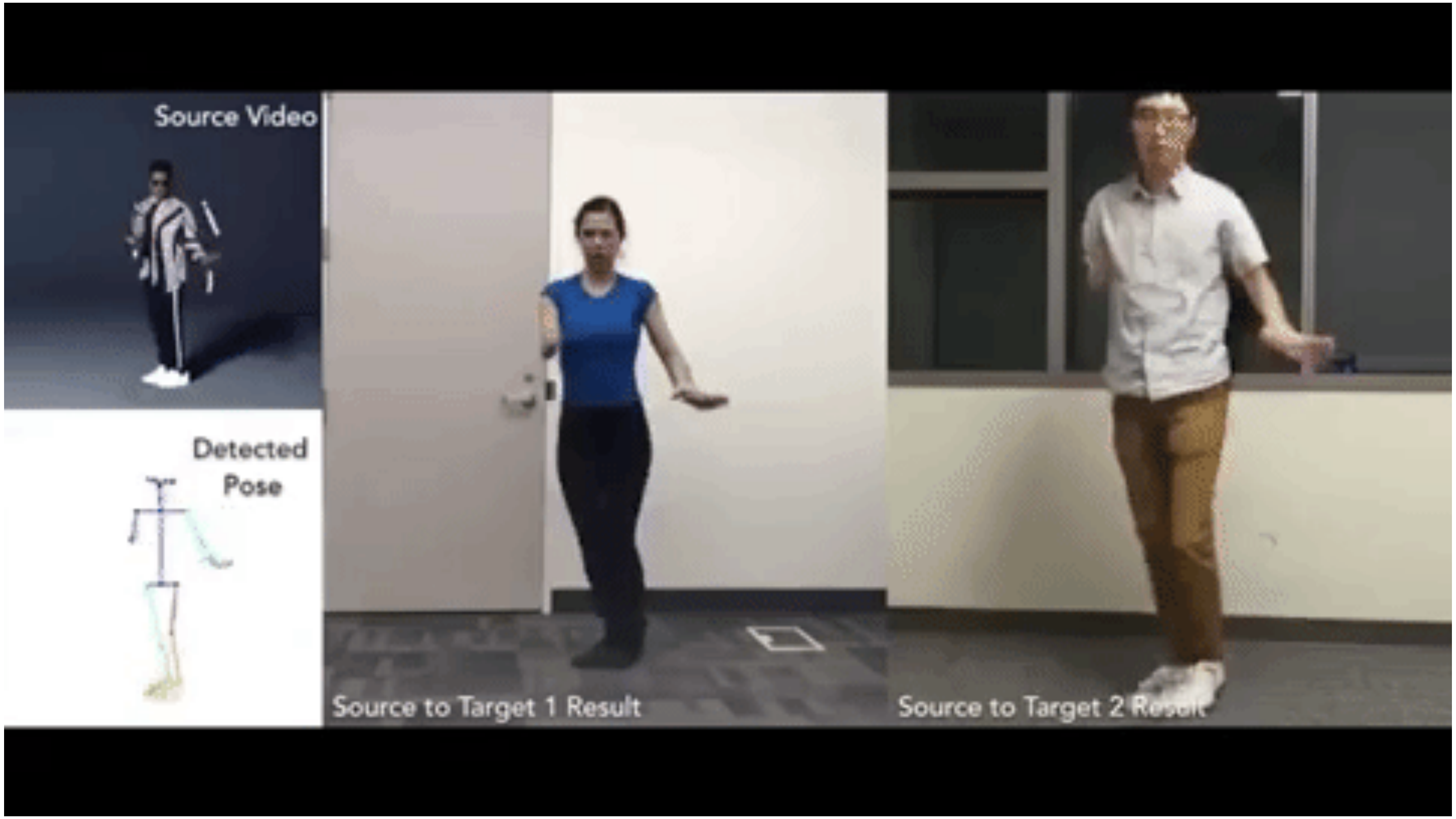



Pose-Guided Photorealistic Face Rotation. *Y. Hu, X. Wu et al. CVPR. 2018*

# Selected Generative Applications

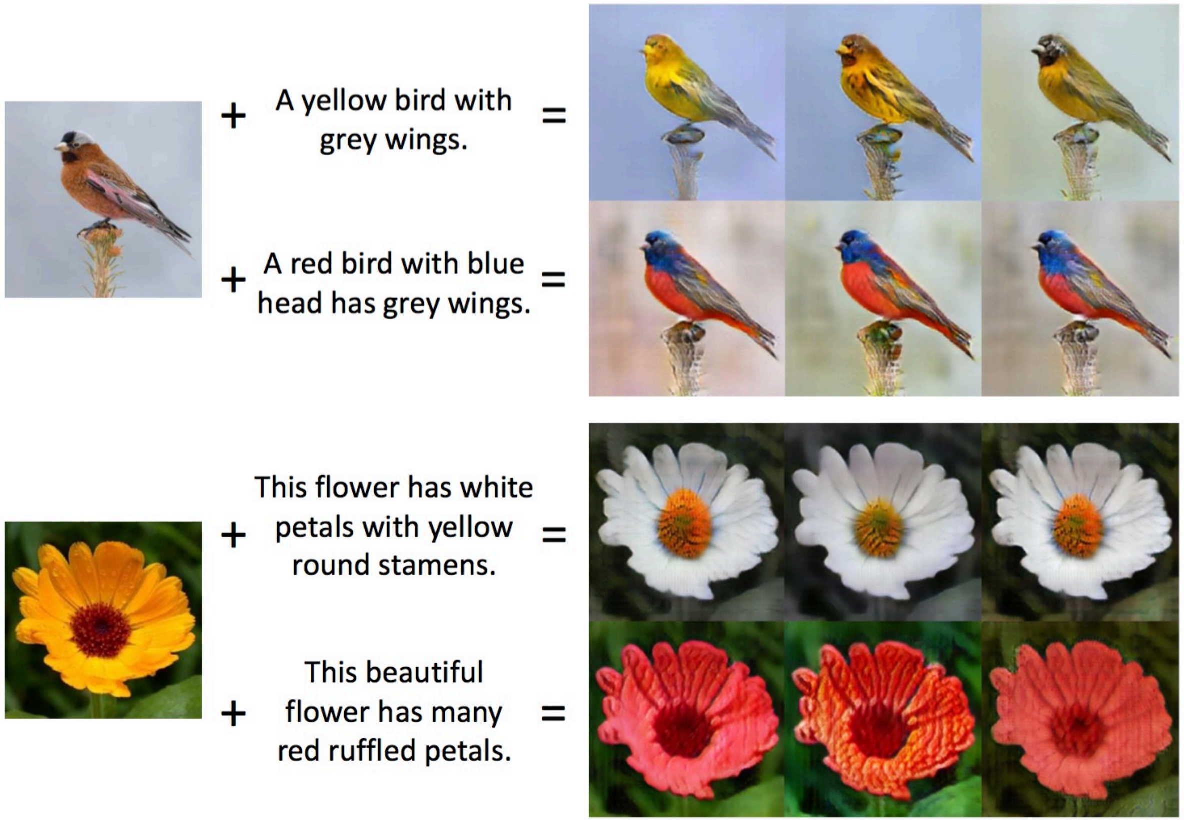
**Everybody Dance Now**

$$P(my\ dance \mid your\ dance)$$



# Selected Generative Applications

**Combine Image and Sentence: <u>Two Conditions</u>**






Semantic Image Synthesis via Adversarial Learning. *H. Dong, S. Yu et al. ICCV 2017.*

## Selected Generative Applications

- **2D Video to 3D shape**

Learning 3D Human Dynamics from Video. *A. Kanazawa, J. Zhang et al*. CVPR, 2019

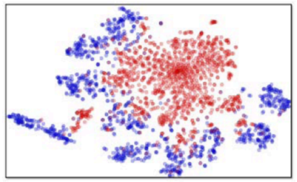
- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- Challenges
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

Selected Advanced Topics

## Domain Adaptation: Model the distribution

Source: Labelled

Target: Unlabelled



$$S(\mathbf{f}) = \{G_f(\mathbf{x};\theta_f)\,|\,\mathbf{x}{\sim}S(\mathbf{x})\}$$
$$T(\mathbf{f}) = \{G_f(\mathbf{x};\theta_f)\,|\,\mathbf{x}{\sim}T(\mathbf{x})\}$$

**Domain shift among sources and target**

**Domain adaptation needed!**

Domain-Adversarial Training of Neural Networks. *Y. Ganin, H. Ajakan et al. JMLR. 2016*

# Selected Advanced Topics

## Adversarial Attack

Fig. 4: An example of digital dodging. Left: An image of actor Owen Wilson, correctly classified by VGG143 with probability 1.00. Right: Dodging against VGG143 using AGN's output (probability assigned to the correct class: $< 0.01$).

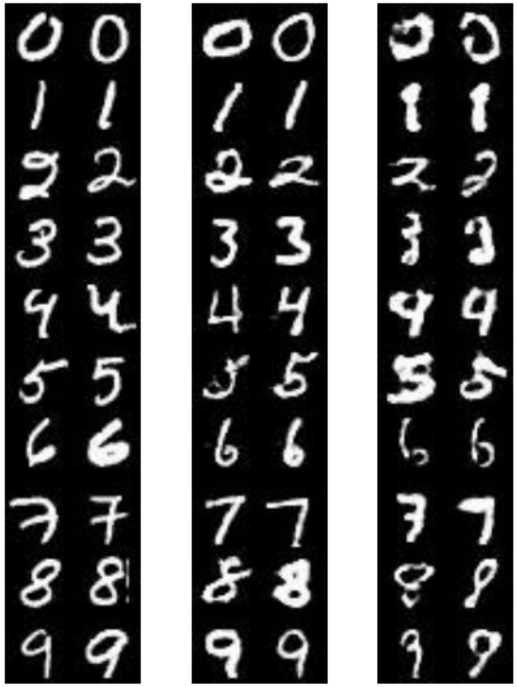


Fig. 9: An illustrations of attacks generated via AGNs. Left: A random sample of digits from MNIST. Middle: Digits generated by the pretrained generator. Right: Digits generated via AGNs that are misclassified by the digit-recognition DNN.

Sharif M, Bhagavatula S, Bauer L, et al. Adversarial generative nets: Neural network attacks on state-of-the-art face recognition[J]

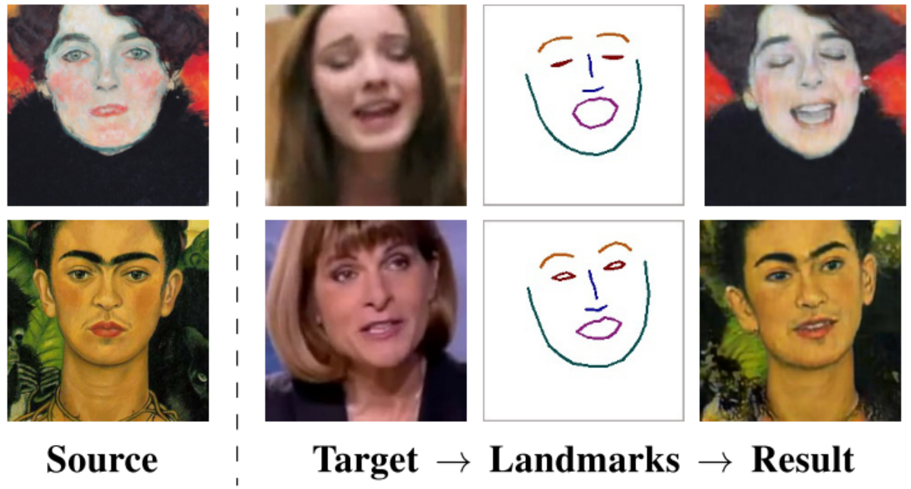
## Meta Learning

face landmark tracks


extracted face landmark tracks from a different video sequence of the same person

source frame

target frame




The results are conditioned on the landmarks taken from the target frame, while the source frame is an example from the training set.

Zakharov E, Shysheya A, Burkov E, et al. Few-shot adversarial learning of realistic neural talking head models[C]//Proceedings of the IEEE International Conference on Computer Vision

Selected Advanced Topics

## Imitation Learning

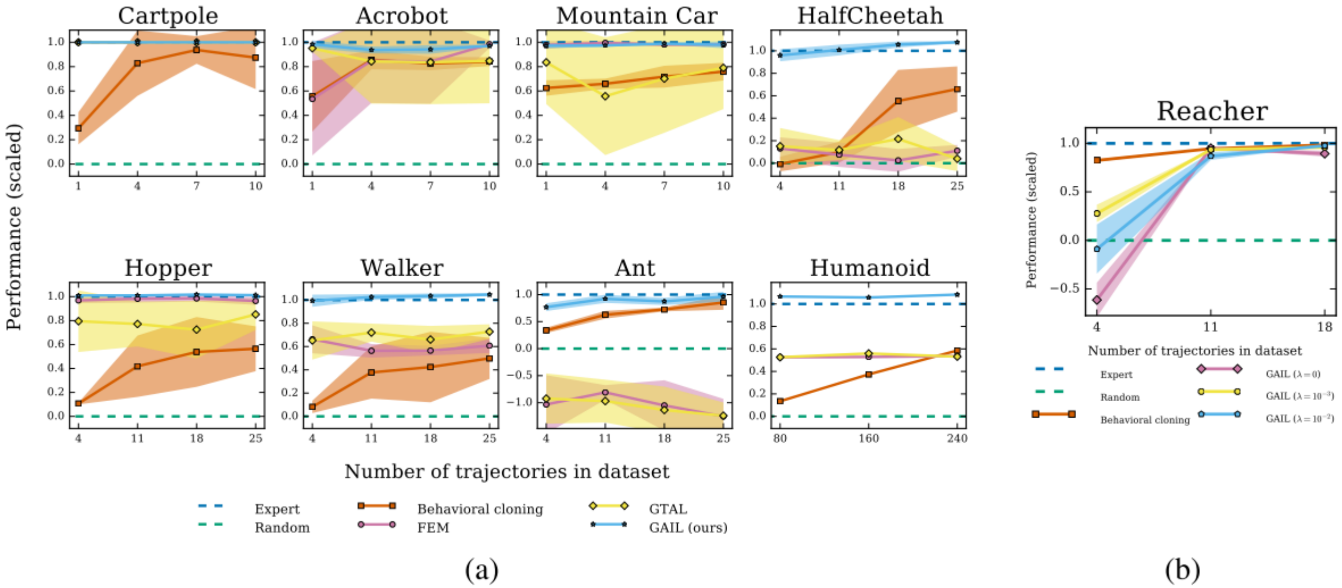
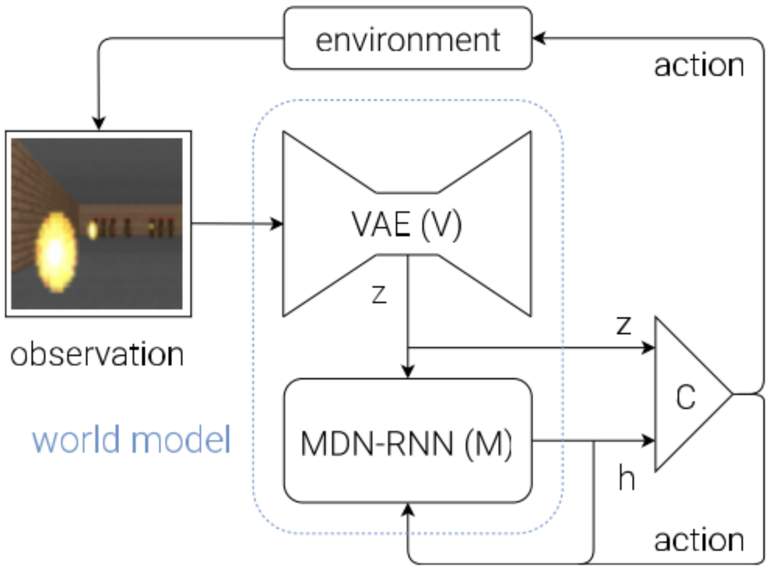


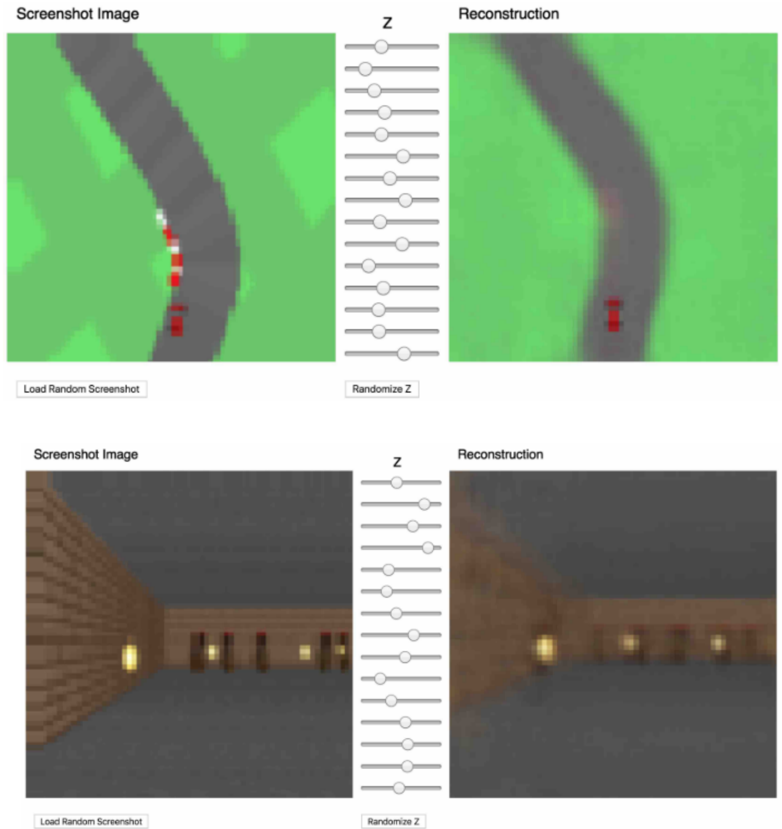

Figure 1: (a) Performance of learned policies. The $y$-axis is negative cost, scaled so that the expert achieves 1 and a random policy achieves 0. (b) Causal entropy regularization $\lambda$ on Reacher. Except for Humanoid, shading indicates standard deviation over 5-7 reruns.

Ho J, Ermon S. Generative adversarial imitation learning[C]//Advances in neural information processing systems.

Selected Advanced Topics

## Reinforcement Learning




Ha D, Schmidhuber J. World models[J]. arXiv preprint arXiv:1803.10122, 2018.

# Selected Advanced Topics

Deep Generative Models relate to all of the following topics:

- Unsupervised Learning
- Semi-supervised Learning
- Weakly-supervised Learning
- Dual Learning
- Self-supervised Learning
- Self-augmented Learning
- …
- …
- …

- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- **Challenges**
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

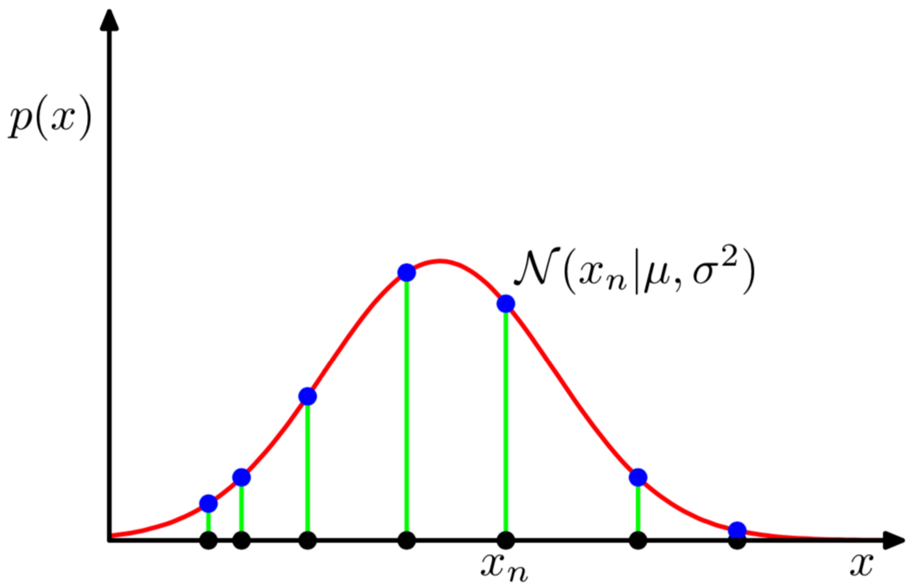
# Challenges

- **Representation ability**

For 1-D data $x$ , the probability distribution $p(x)$ is simple, e.g., Gaussian?

For high-dimensional data $\mathbf{x} = (x_1, x_2, \dots, x_n)$, e.g., $n$ pixels

how do we learn the joint distribution $p(x_1, x_2, \dots, x_n)$?




## Challenges

- **Learning method**

If we can **represent** the $p(x)$, the next question:

how do we **measure** and **minimize** the distance
between the estimated distribution $p(x)$ and the real distribution $p_{data}$?

If we use a parametric model (e.g., Gaussian) to represent $p(x)$,
it can be an optimization problem:

$$\min_{\theta \in \mathcal{M}} \mathcal{L}(p_{data}, p_\theta(x))$$

where the parameter $\theta$ is from the model $\mathcal{M}$

# Challenges

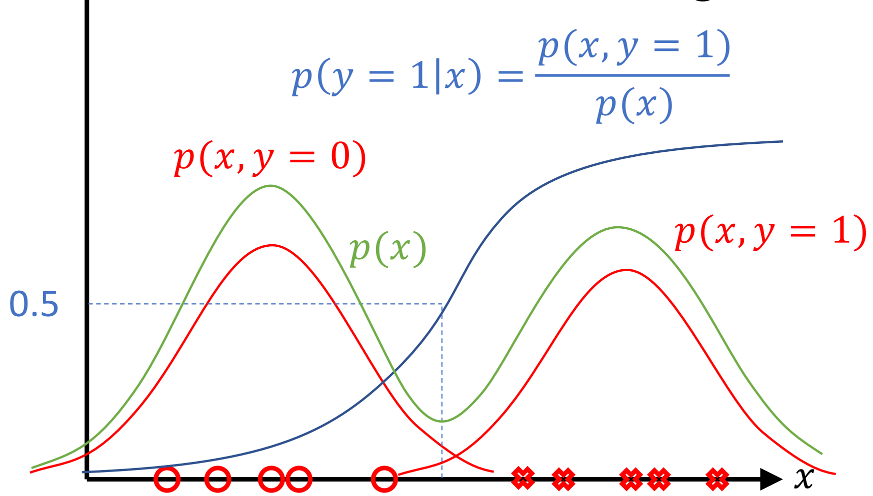
- **Inference**

If we can represent the $p(x)$ and successfully learn it, we now can:

1. Generative task (sampling): $\mathbf{x}_{new} \sim p(\mathrm{x})$
2. Density estimation: $p(\mathrm{x})$ high if $\mathbf{x}$ looks like a real data sample

the final question: how do we perform discriminative task?

i.e., invert the generative process




- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- Challenges
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

## Syllabus

- Week 1: Introduction (Today)
- Week 2: Autoregressive Models
- Week 3: Variational Autoencoders
- Week 4: Normalizing Flow Models
- Week 5: Generative Adversarial Networks
- Week 6: Practice

Foundation

- Week 7: Evaluation of Generative Models
- Week 8: Energy-based Models
- Week 9: Discreteness in Latent Variables
- Week 10: Challenges of Generative Models
- Week 11: Applications of Generative Models
- Week 12: Generative Model Variants

Research

Might changed later …

- Week 13-14: Paper Reading
- Week 15-16: Project Presentation

Practice

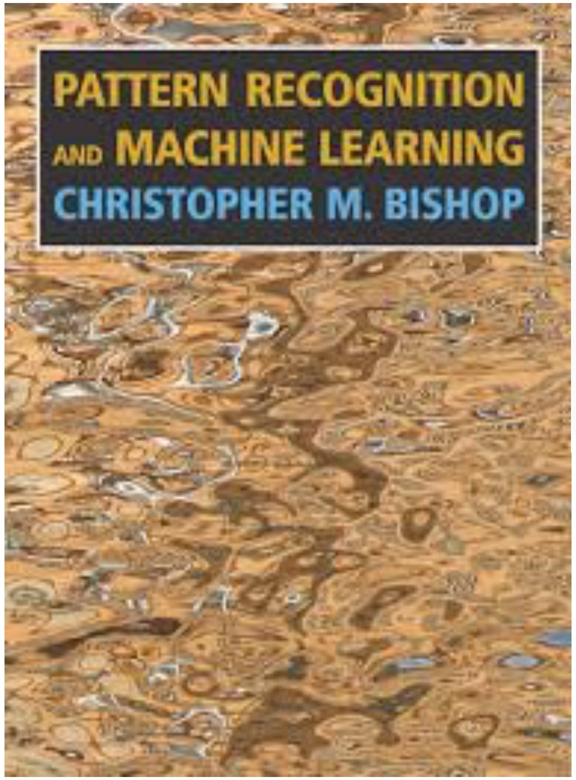
- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- Challenges
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

# Prerequisites

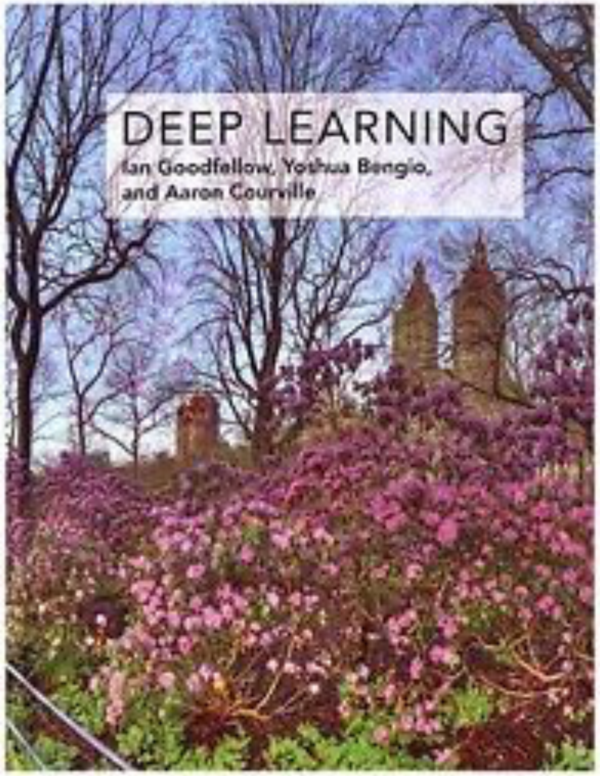
- Basic knowledge of probabilities
  - Bayes rule, chain rule, probability distribution …
- Basic knowledge of machine learning/deep learning
  - “Machine Learning”, “Pattern Recognition and Machine Learning”
  - “Computer Vision”, “Natural Language Processing” …
- Basic programming language
  - Python

- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- Challenges
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

# Logistics




**Free Download**




**Free Download**

# Logistics

**Deep Generative Models**
Stefano Ermon, Aditya Grover
https://deepgenerativemodels.github.io

**Deep Generative Models**
Rajesh Ranganath
https://cs.nyu.edu/courses/spring18/CSCI-GA.3033-022/

- What and Why
- Generative Models vs. Computer Graphics
- Discriminative vs. Generative
- Selected Generative Applications
- Selected Advanced Topics
- Challenges
- Syllabus
- Prerequisites
- Logistics
- Grading Policies

# Grading Policies

- **Paper Reading 40%**
  - Understanding (Q/A) 20%
  - Presentation 20%
- **Course Project 50%**
  - Proposal 10%
  - Open source quality 15%
  - Report 15%
- **Others 20%**
  - Discussion
  - Attendance

- 1~2 students/group
- Topic: application or theory
- Open source: Github repository
- 4 Pages Report
  - Motivation
  - Introduction
  - Related Work
  - Method
  - Evaluation
  - Conclusion

# Thanks