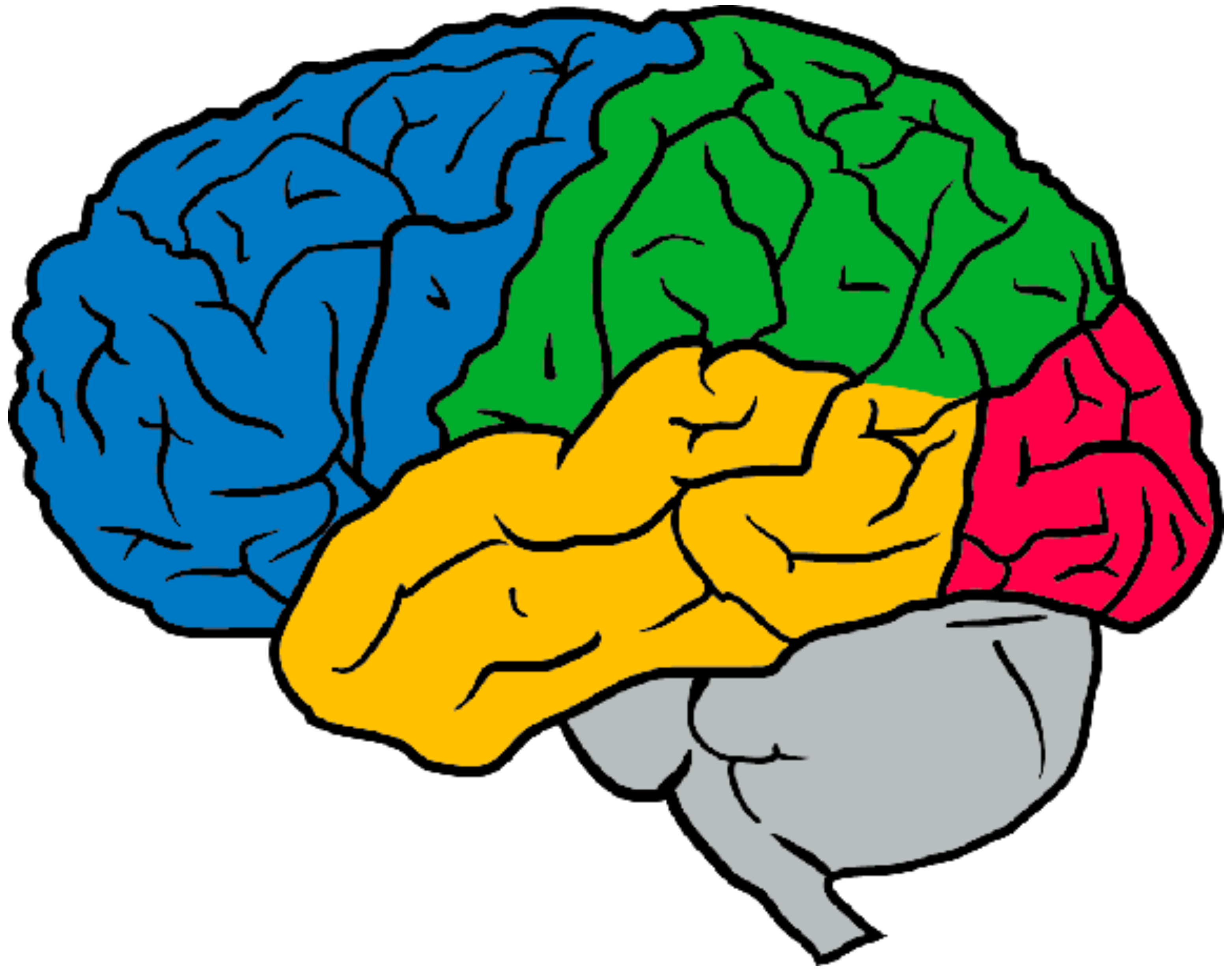
Generative Models I

Ian Goodfellow, Staff Research Scientist, Google Brain

MILA Deep Learning Summer School

Montréal, Québec 2017-06-27

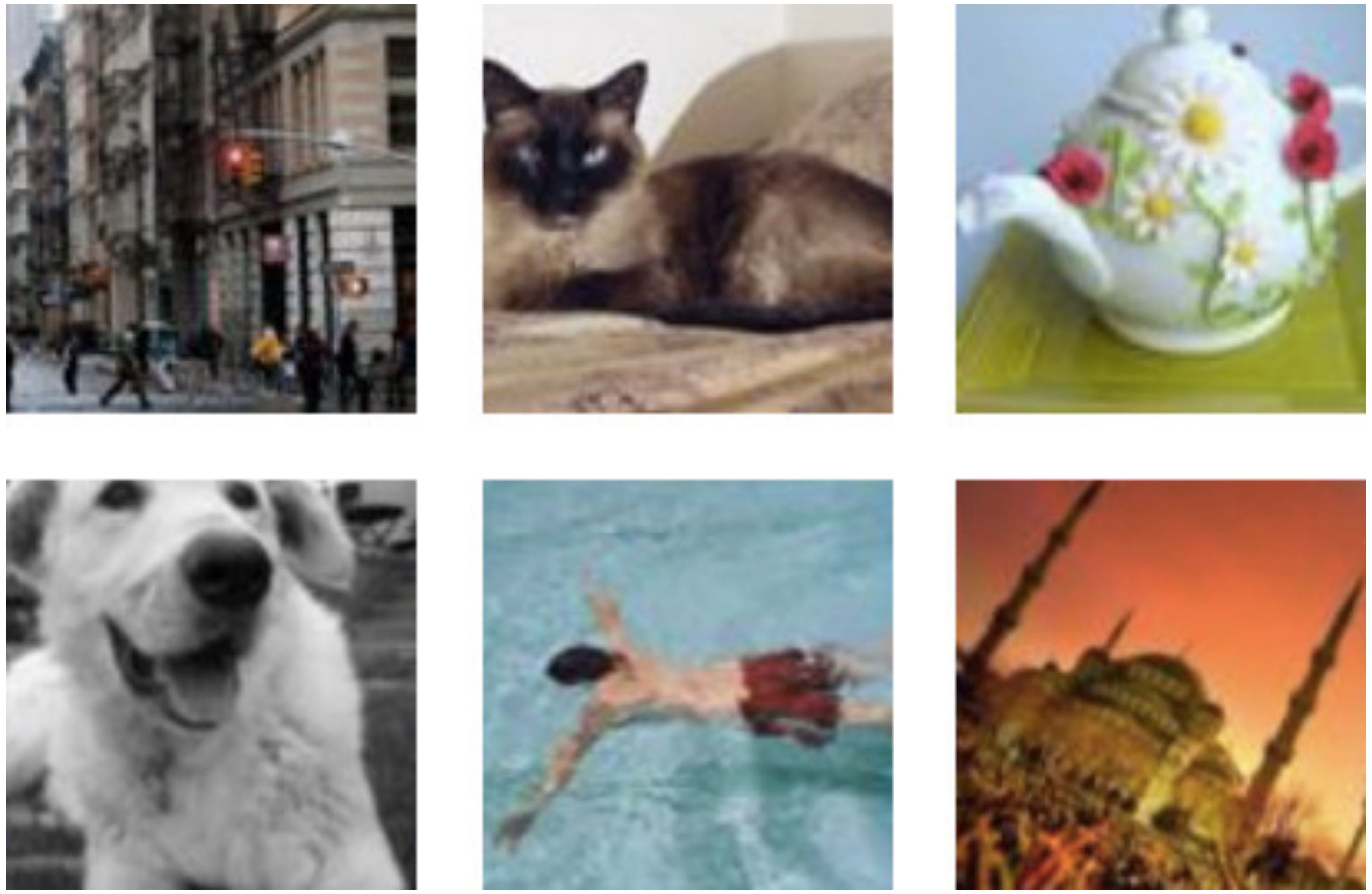


Density Estimation

(Goodfellow 2017)

Sample Generation

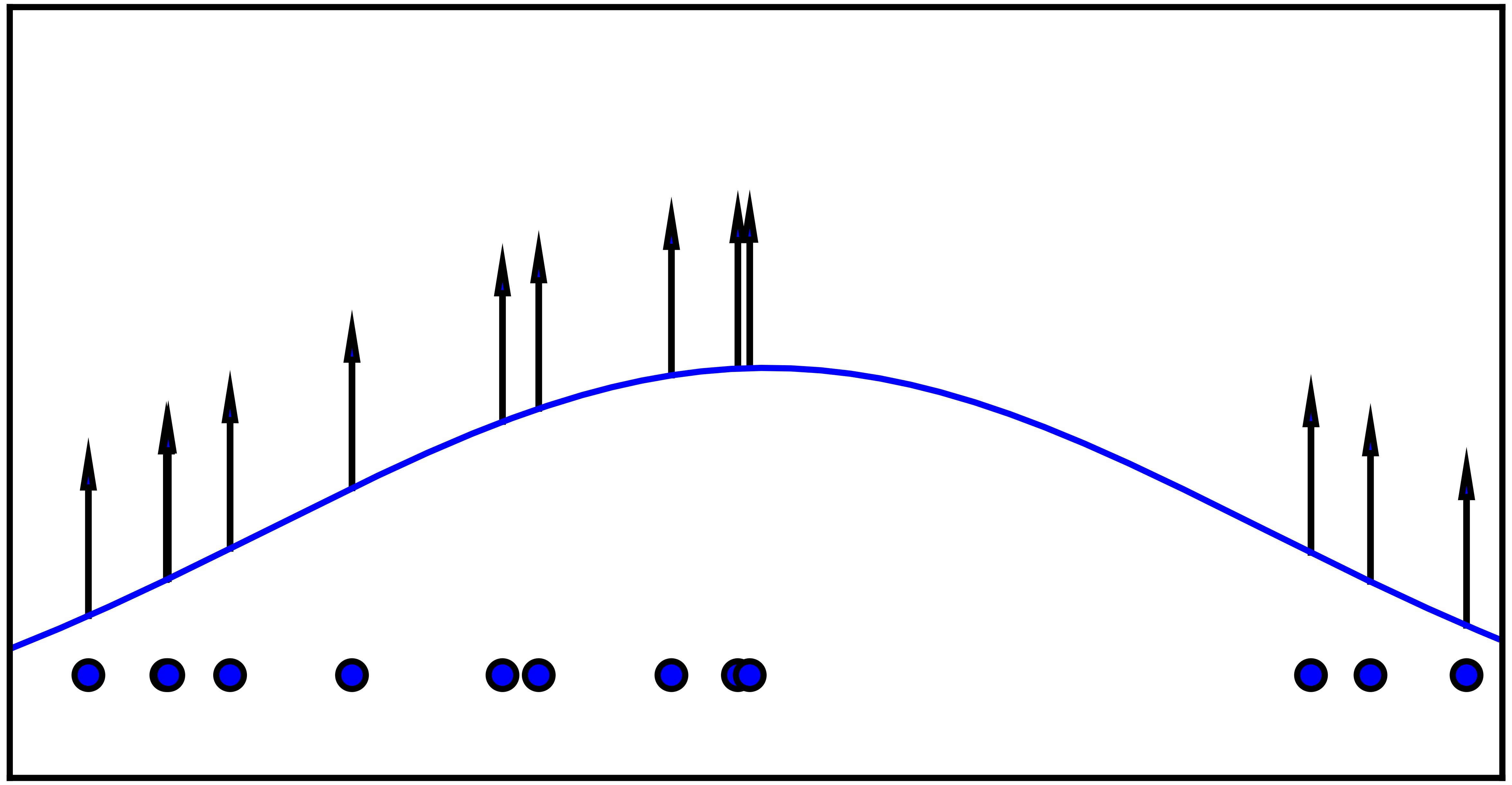
Training examples



Model samples

(Goodfellow 2017)

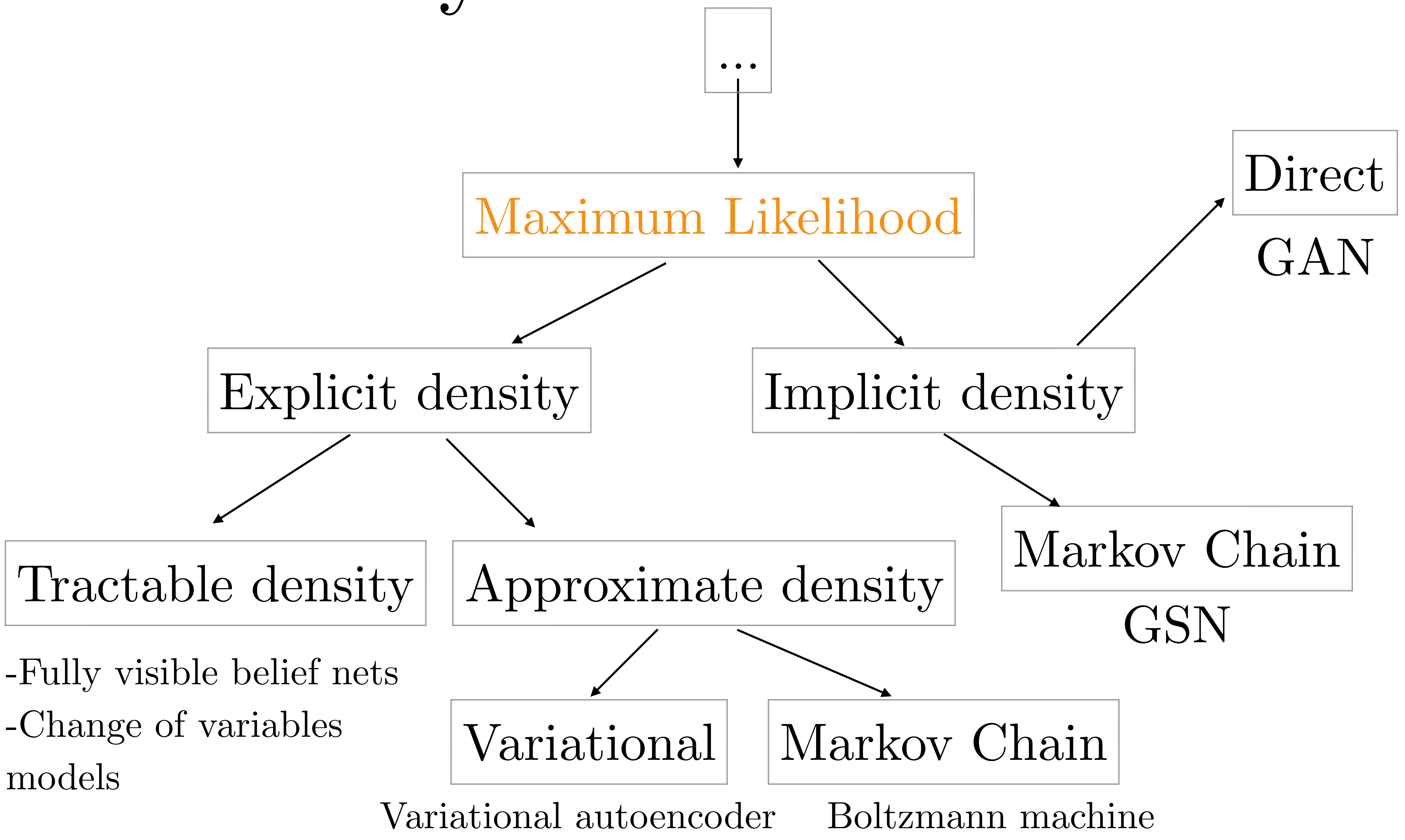
Maximum Likelihood



$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \mathbb{E}_{x \sim p_{\text{data}}} \log p_{\text{model}}(\boldsymbol{x} \mid \boldsymbol{\theta})$$

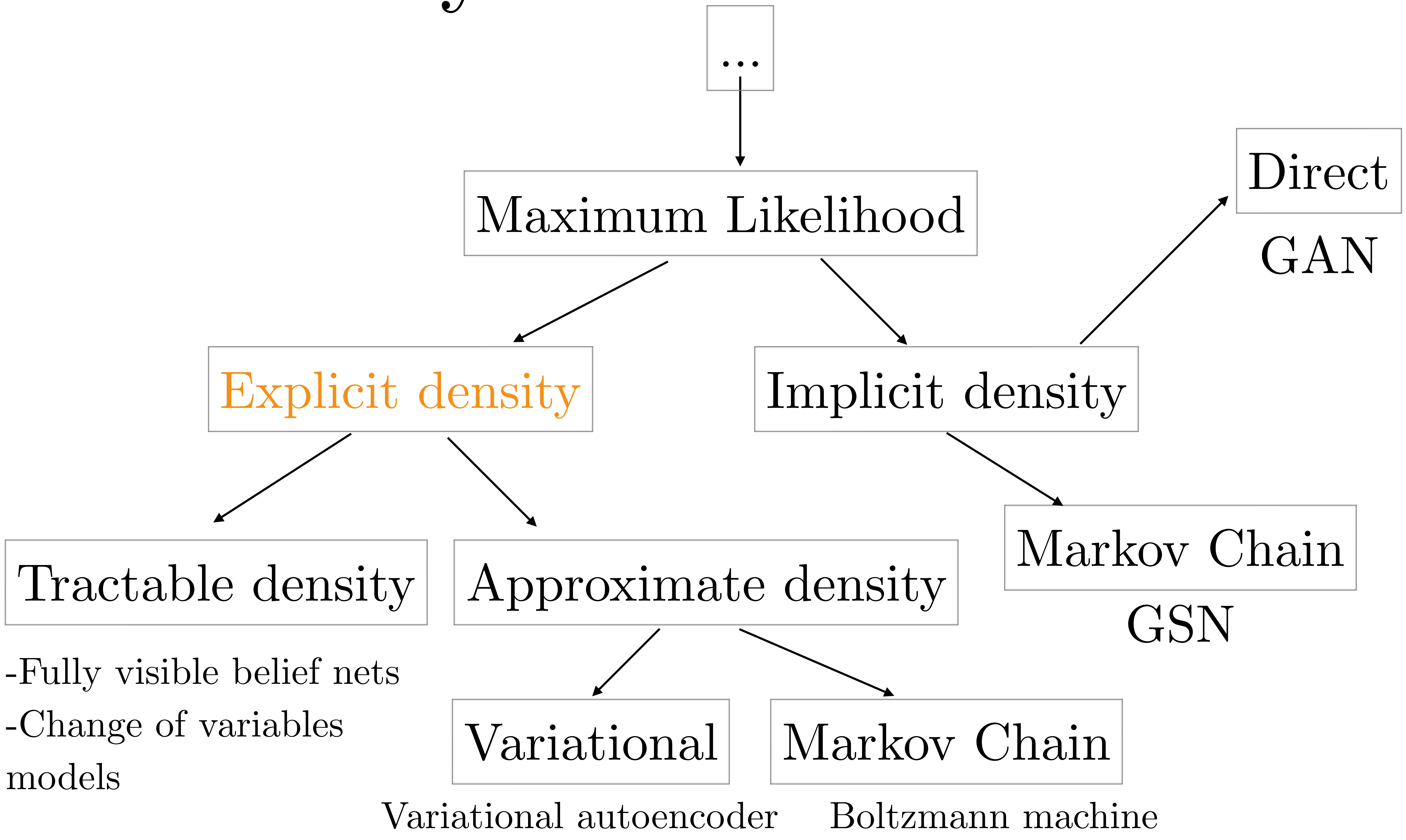
(Goodfellow 2017)

Taxonomy of Generative Models



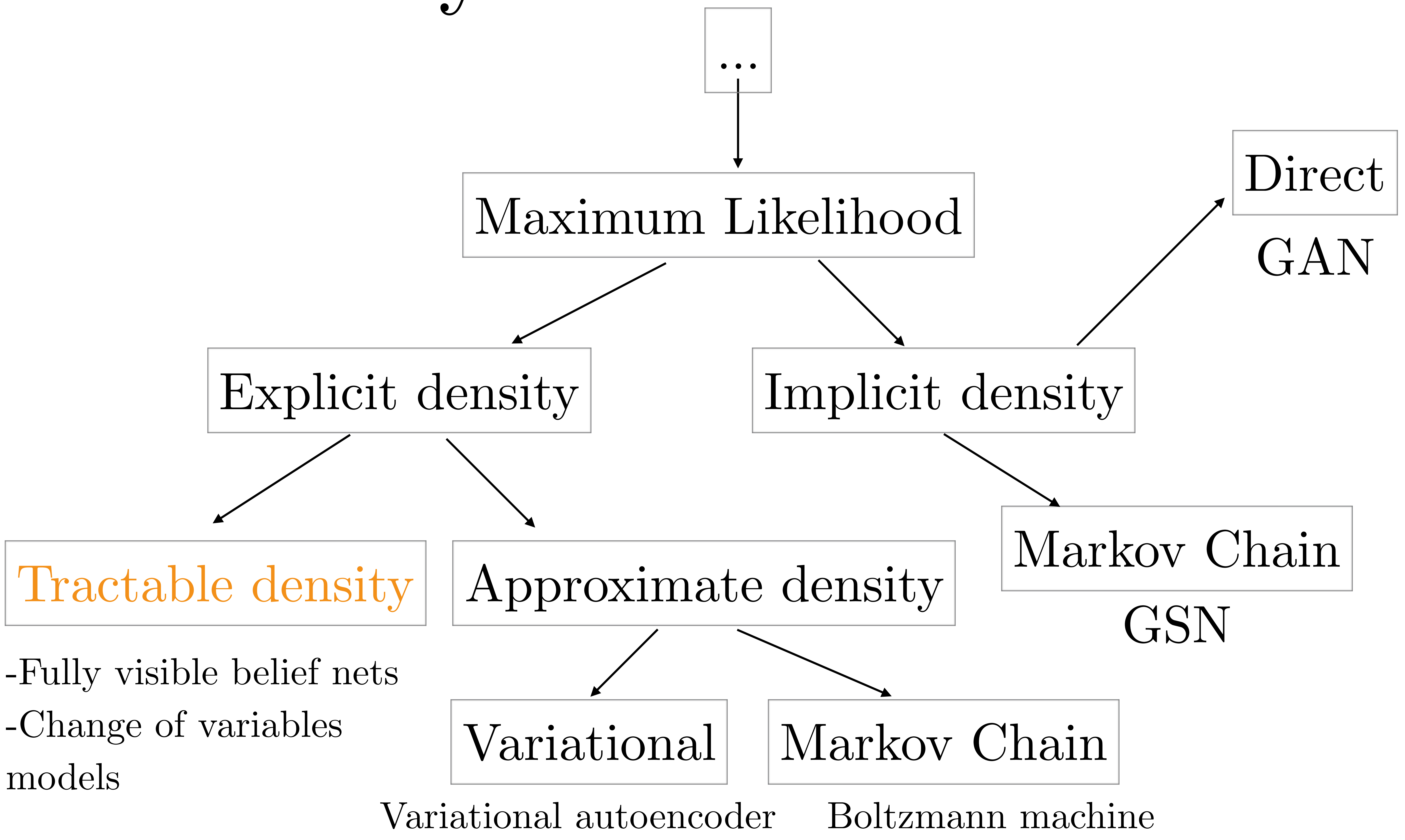
(Goodfellow 2017)

Taxonomy of Generative Models



(Goodfellow 2017)

Taxonomy of Generative Models

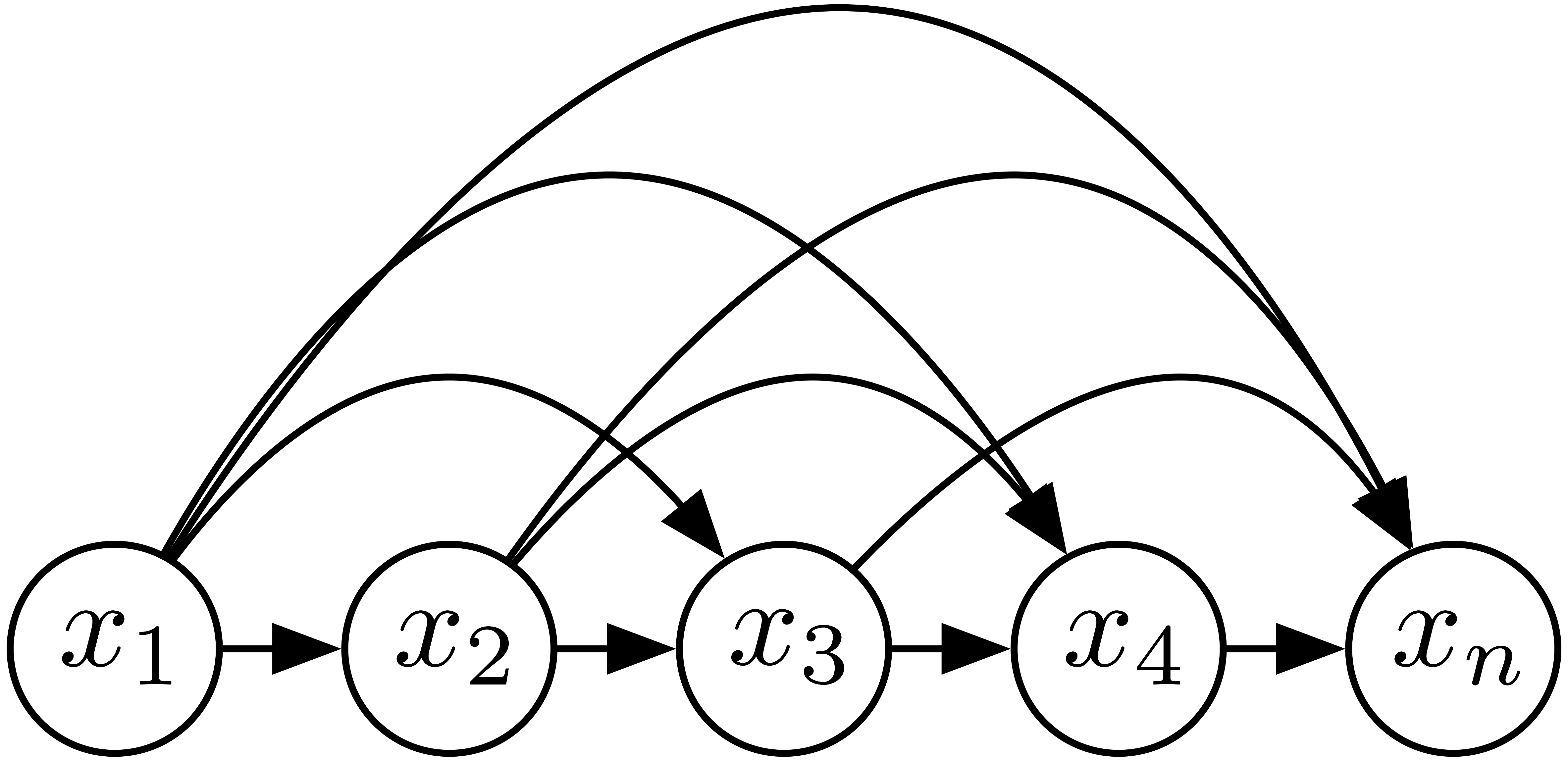


(Goodfellow 2017)

Fully Visible Belief Nets

- Explicit formula based on chain (Frey et al, 1996) rule:

$$p_{\text{model}}(\boldsymbol{x}) = p_{\text{model}}(x_1) \prod_{i=2}^{n} p_{\text{model}}(x_i \mid x_1, \ldots, x_{i-1})$$



(Goodfellow 2017)

Fully Visible Belief Nets

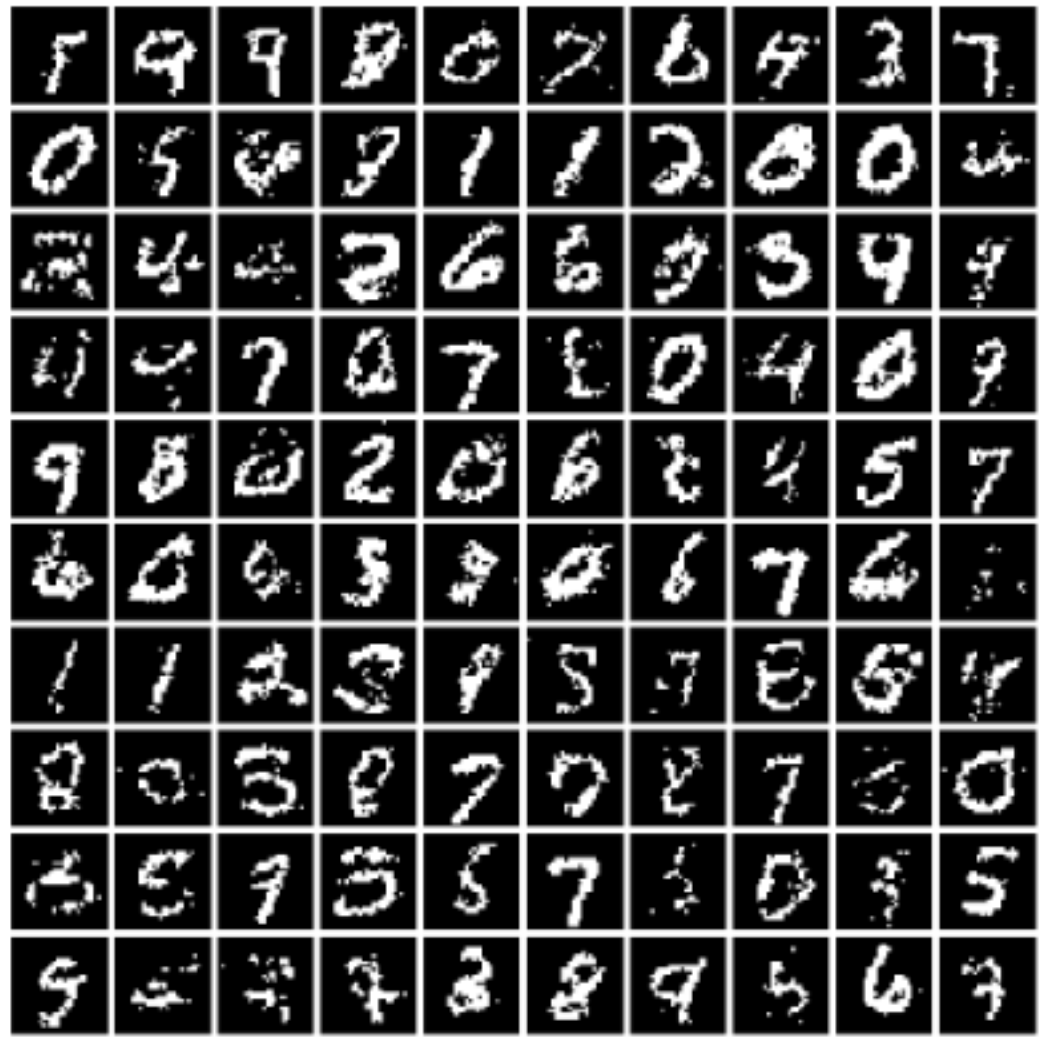
- Disadvantages:
  - O(n) non-parallelizable sample generation runtime
  - Generation not controlled by a latent code

(Goodfellow 2017)

Notable FVBNs

NADE

(Larochelle et al 2011)



MADE

(Germain et al 2016)

PixelCNN

(van den Ord et al 2016)

"Autoregressive models"

(Goodfellow 2017)

Change of Variables

$$y = g(x) \Rightarrow p_x(\boldsymbol{x}) = p_y(g(\boldsymbol{x})) \left| \det \left(\frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}} \right) \right|$$

e.g. Nonlinear ICA (Hyvärinen 1999)

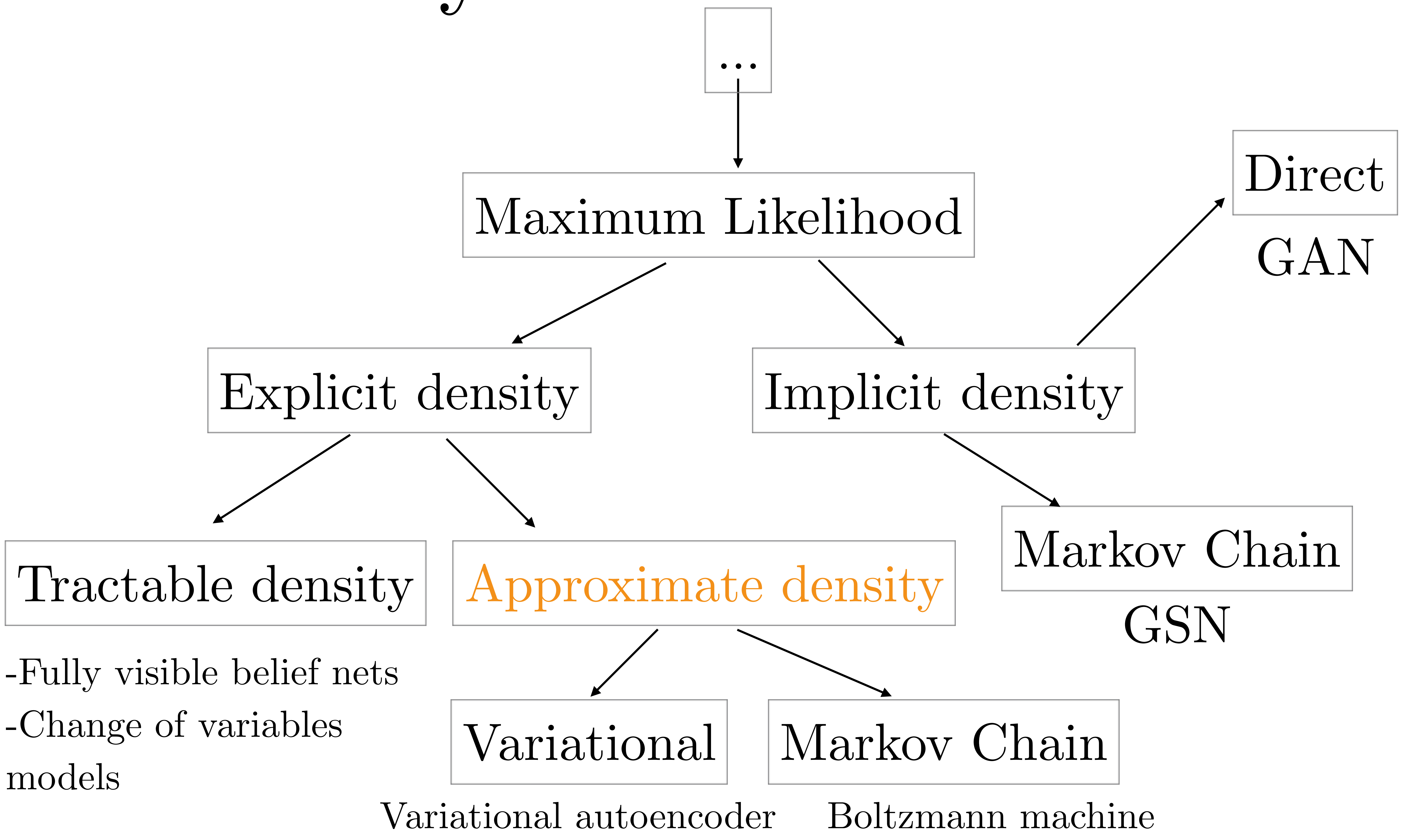
64x64 ImageNet Samples
Real NVP (Dinh et al 2016)

Disadvantages:

- Transformation must be invertible
- Latent dimension must match visible dimension

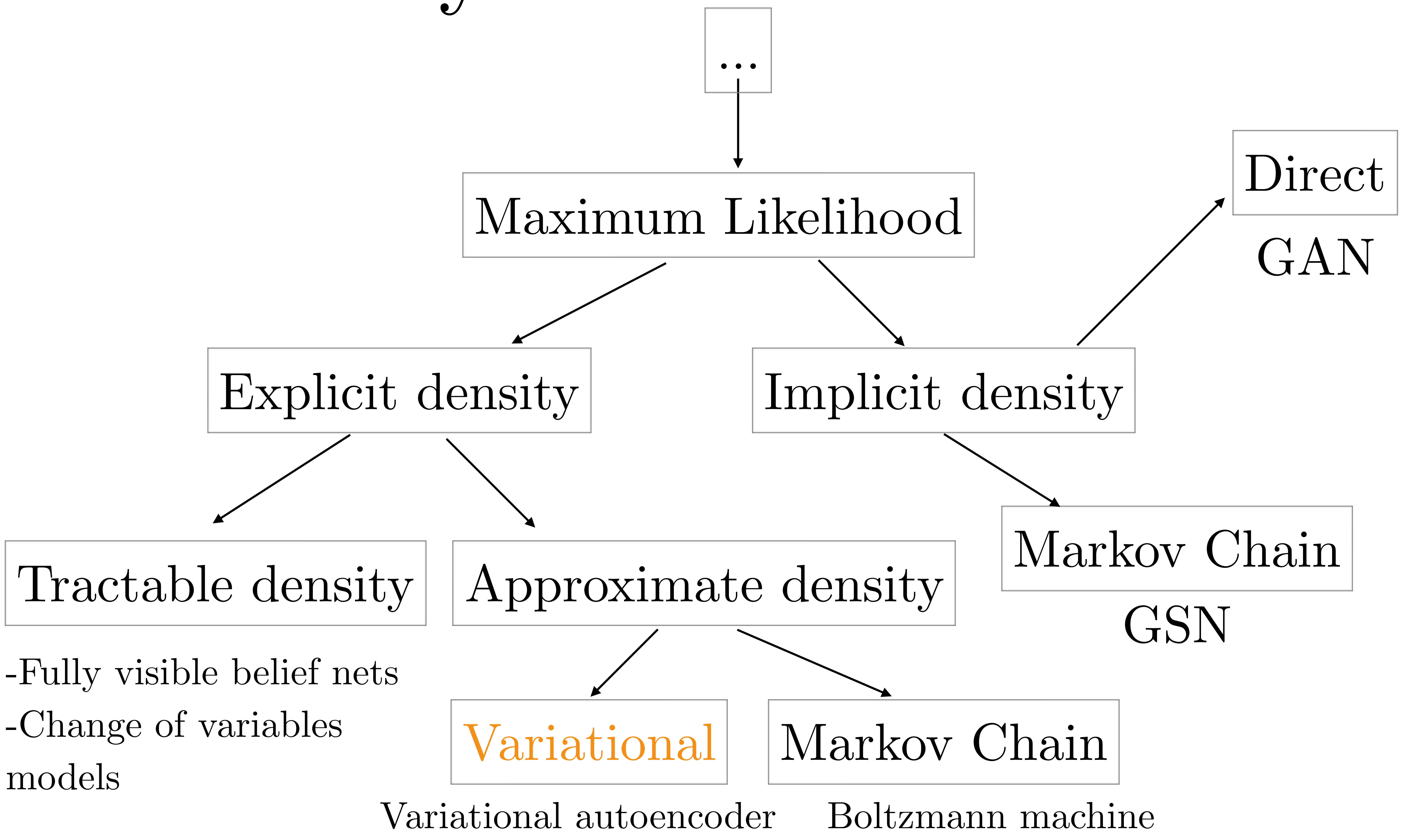
(Goodfellow 2017)

Taxonomy of Generative Models



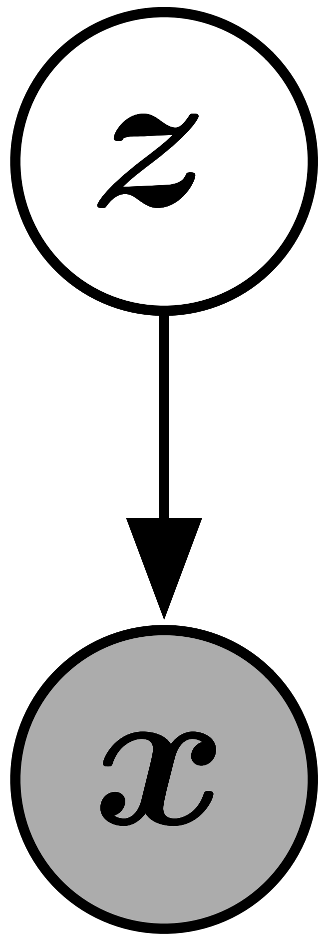
(Goodfellow 2017)

Taxonomy of Generative Models



(Goodfellow 2017)

Variational Learning



$$p_{\text{model}}(\boldsymbol{x}) = \int p_{\text{model}}(\boldsymbol{x}, \boldsymbol{z}) d\boldsymbol{z}$$

Latent variable models often have intractable density

Variational Bound

$$\log p(\boldsymbol{x}) \geq \log p(\boldsymbol{x}) - D_{\mathrm{KL}}\left(q(\boldsymbol{z}) \| p(\boldsymbol{z} \mid \boldsymbol{x})\right)$$
$$= \mathbb{E}_{\boldsymbol{z} \sim q} \log p(\boldsymbol{x}, \boldsymbol{z}) + H(q)$$

Variational inference: maximize with respect to q

Variational learning: maximize with respect to parameters of p

Variational Autoencoder

(Kingma and Welling 2013, Rezende et al 2014)

Define a neural network that predicts optimal q
Define $p(z \mid x)$ via another neural network

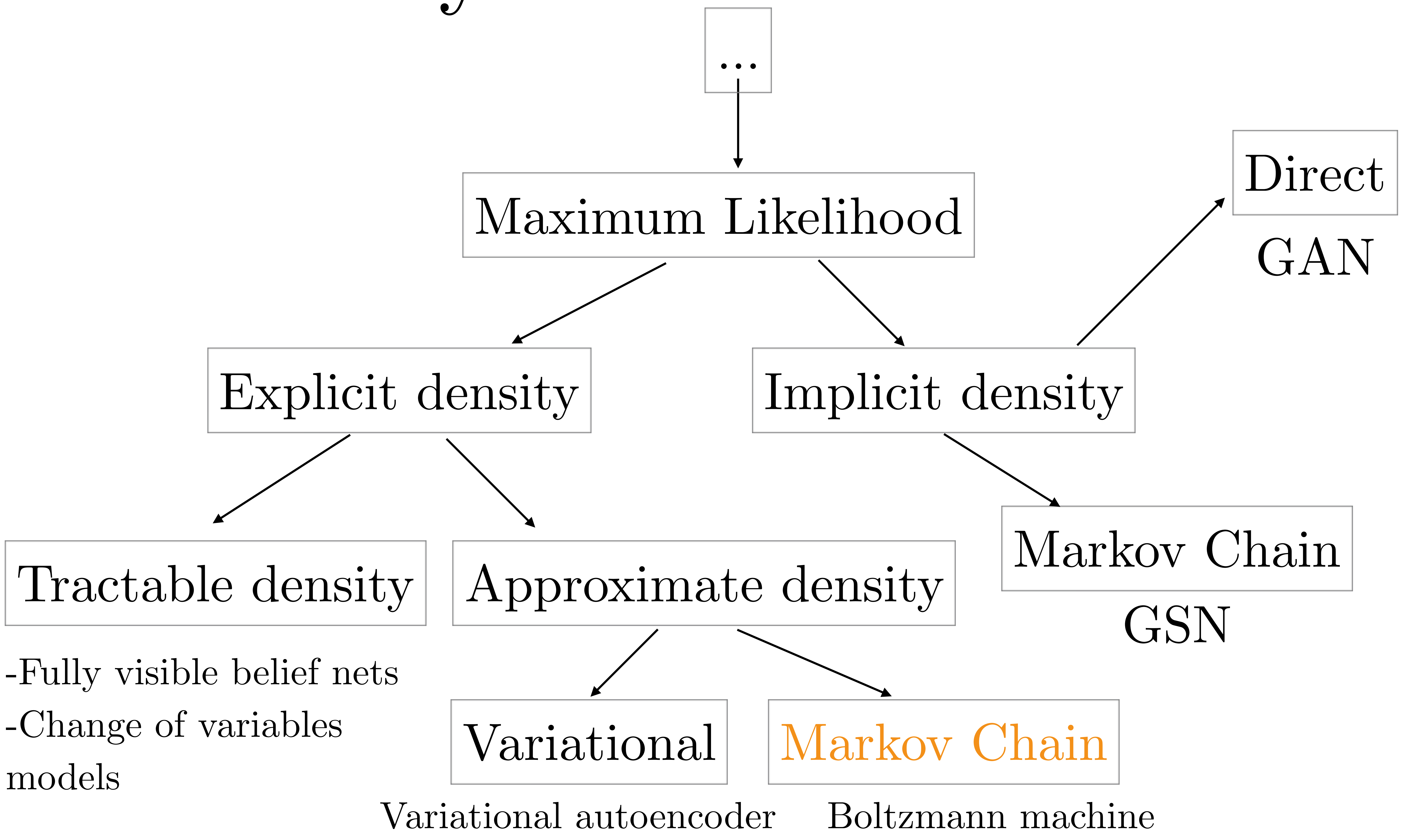
Whole model can be fit via maximization of a single objective function with gradient- based optimization

CIFAR-10 samples
(Kingma et al 2016)

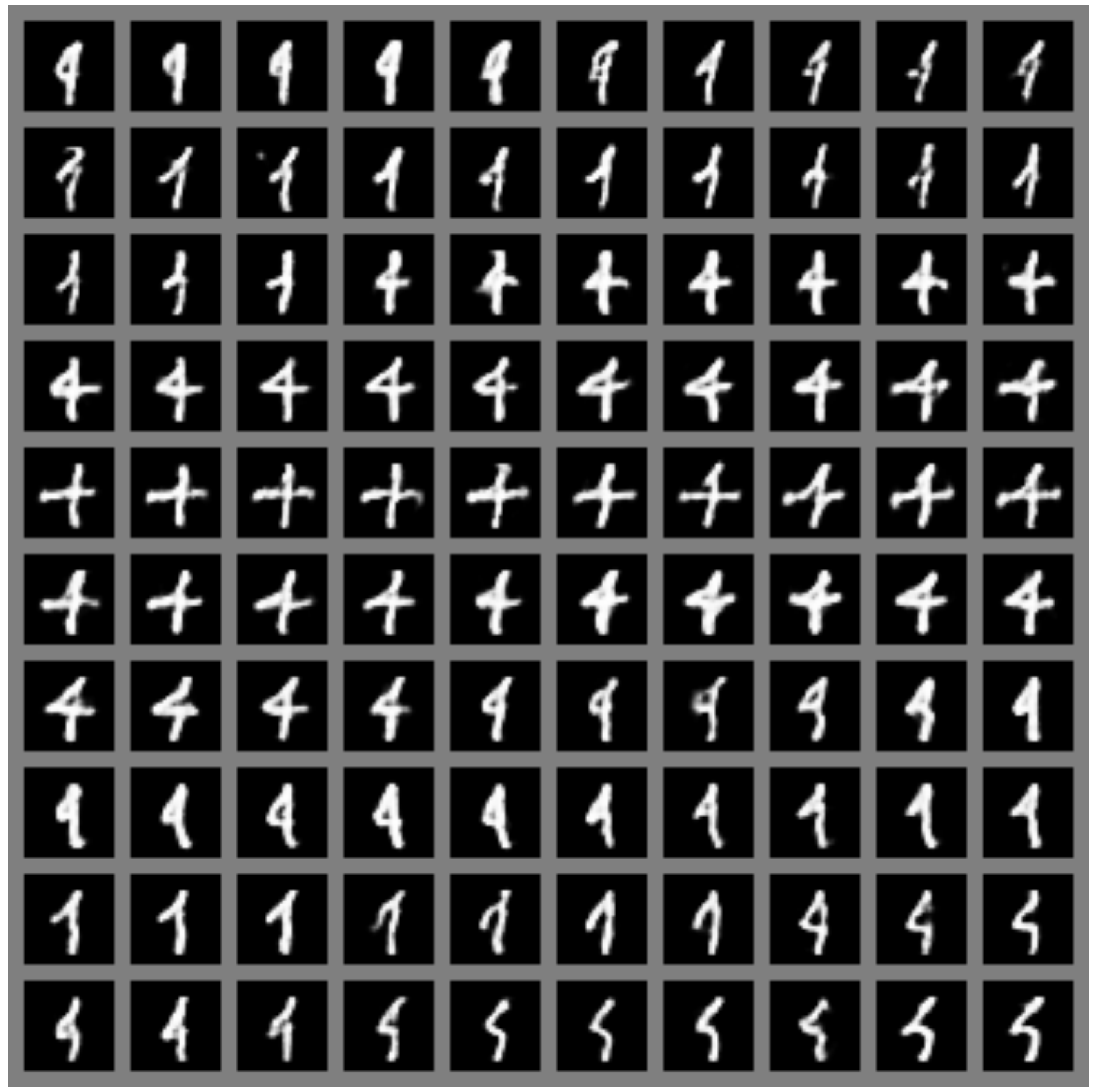
For more information...

- Max Welling will teach a lesson on variational inference

Taxonomy of Generative Models

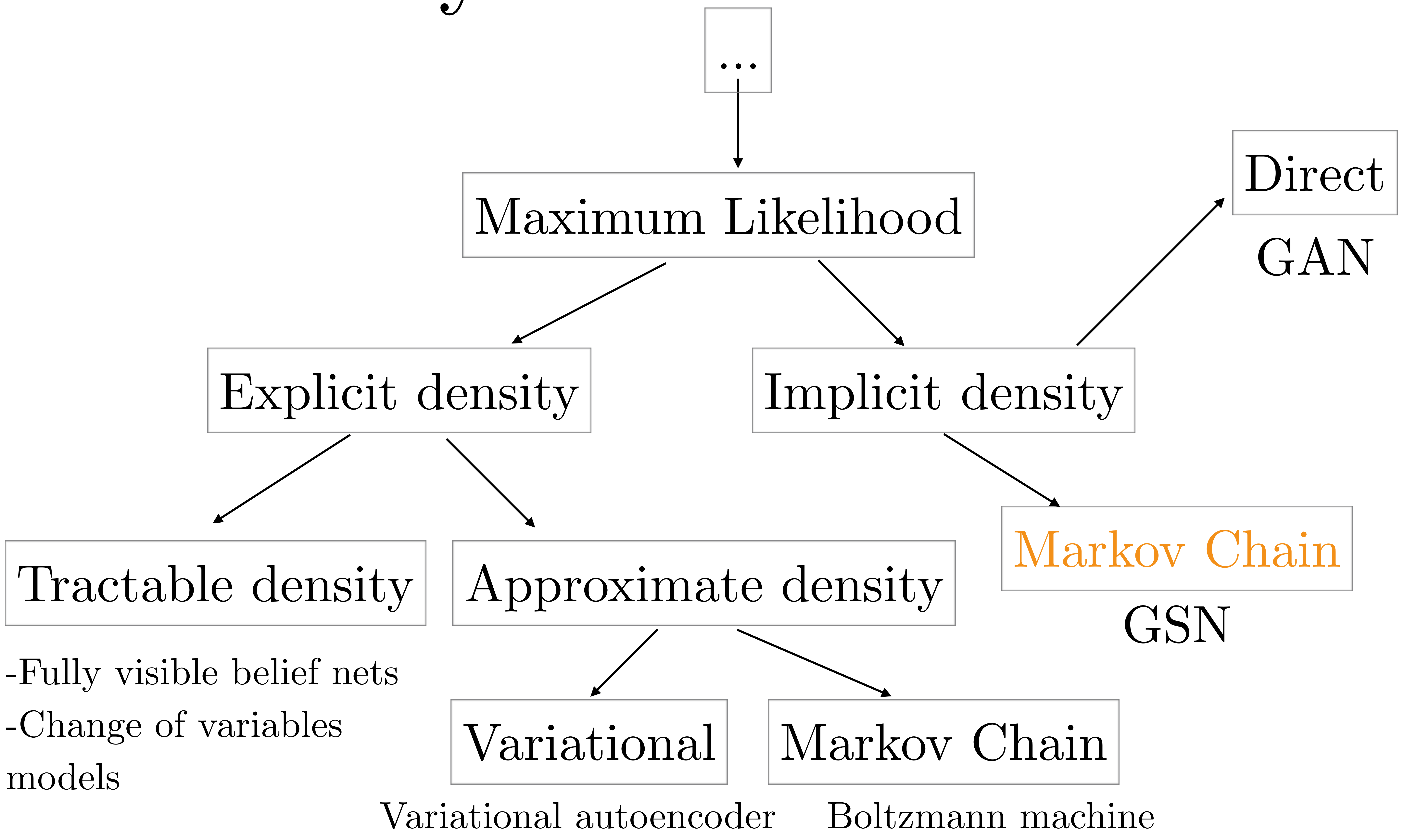


Deep Boltzmann Machines



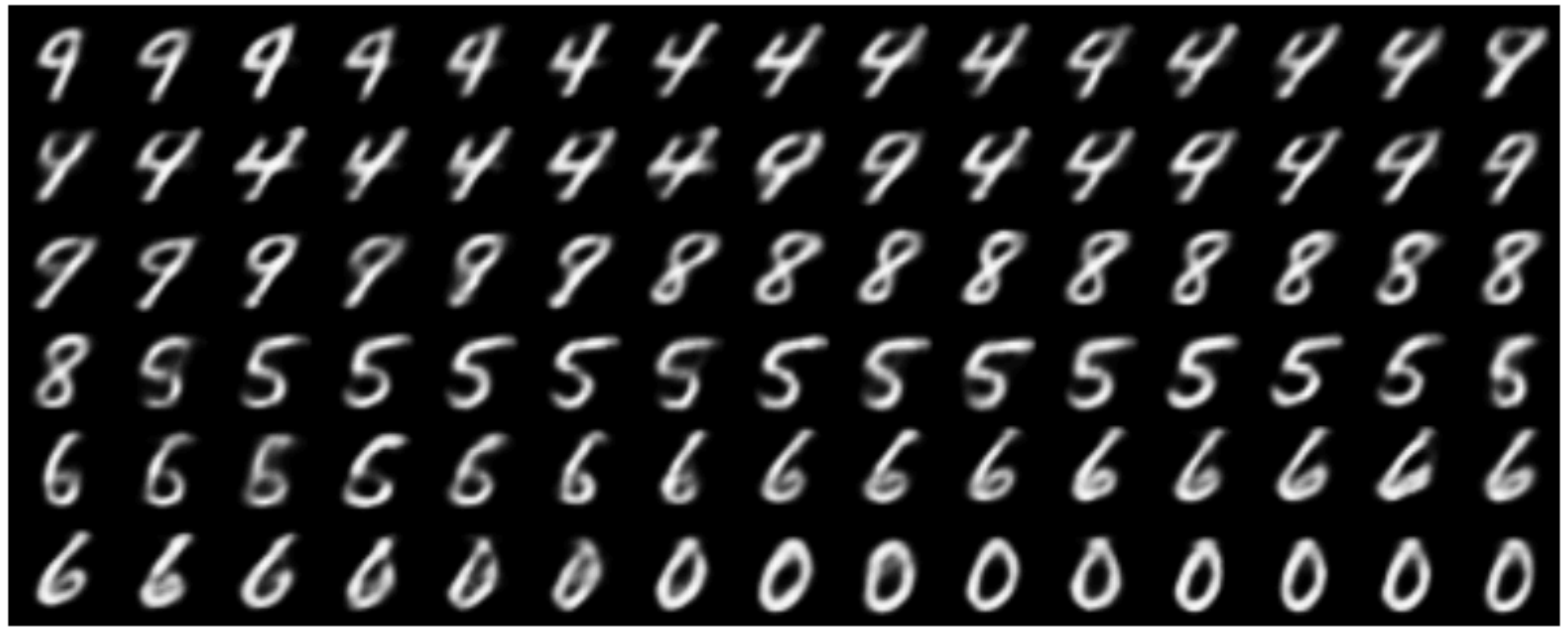
(Salakhutdinov and Hinton, 2009)

Taxonomy of Generative Models



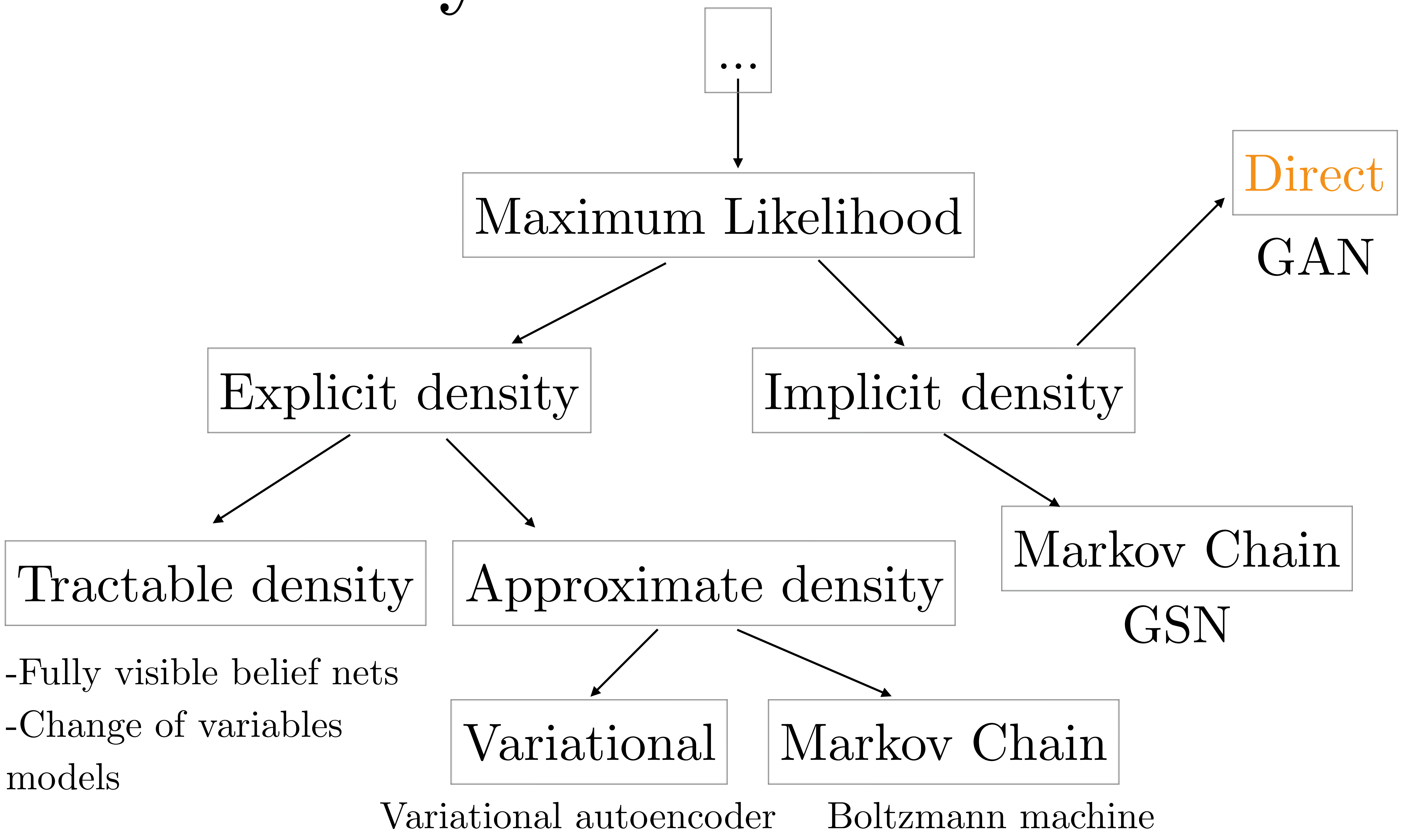
(Goodfellow 2017)

Generative Stochastic Networks



(Bengio et. al, 2013)

Taxonomy of Generative Models



(Goodfellow 2017)

Generative Adversarial Networks

$D(\mathrm{x})$ tries to be near 1

↑

Differentiable function D

↑

x sampled from data

D tries to make $D(G(z))$ near 0, G tries to make $D(G(z))$ near 1

↑

D

↑

x sampled from model

↑

Differentiable function G

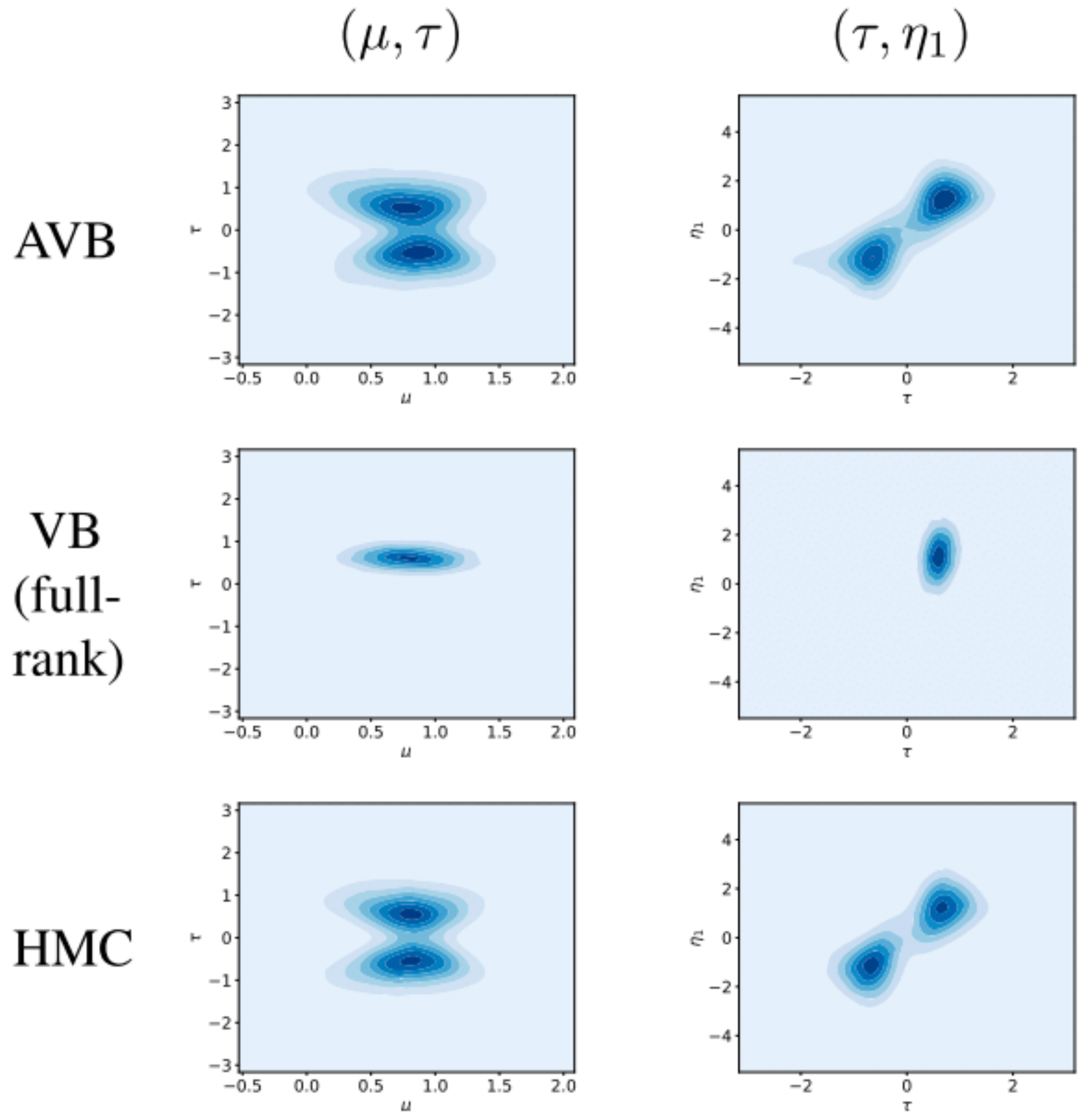
↑

Input noise z

(Goodfellow et al., 2014)

(Goodfellow 2017)

Combining VAEs and GANs: Adversarial Variational Bayes



Related:
-Adversarial autoencoders
-Adversarially learned inference
-BiGANs

(Mescheder et al, 2017)

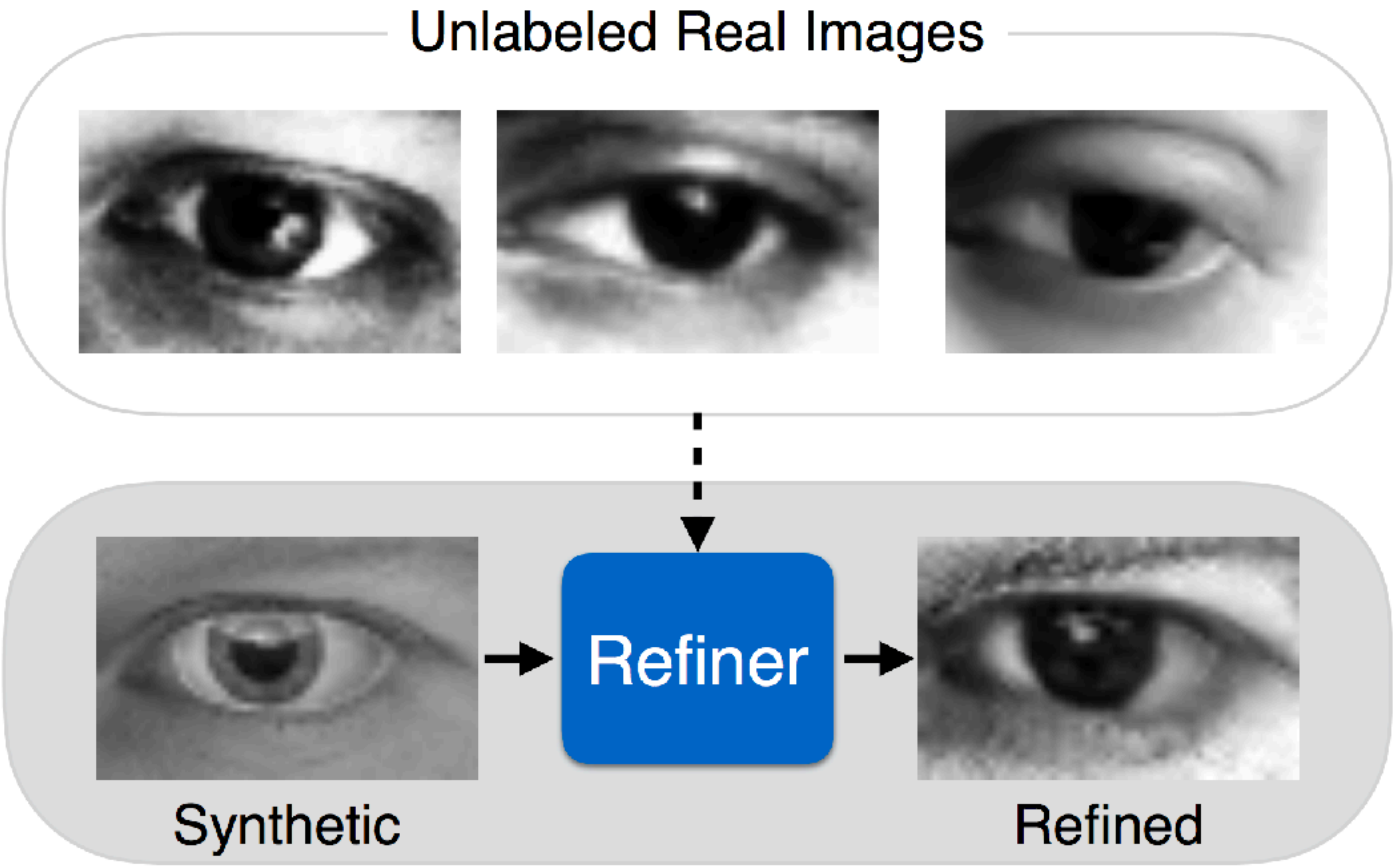
What can you do with generative models?

- Simulated environments and training data
- Missing data
  - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Learn useful embeddings

TEACHING AID

Apple's first research paper tries to solve a problem facing every company working on AI

Generative models for simulated training data



(Shrivastava et al., 2016)

What can you do with generative models?

- Simulated environments and training data
- Missing data
    - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Learn useful embeddings

What is in this image?

(Yeh et al., 2016)

Generative modeling reveals a face

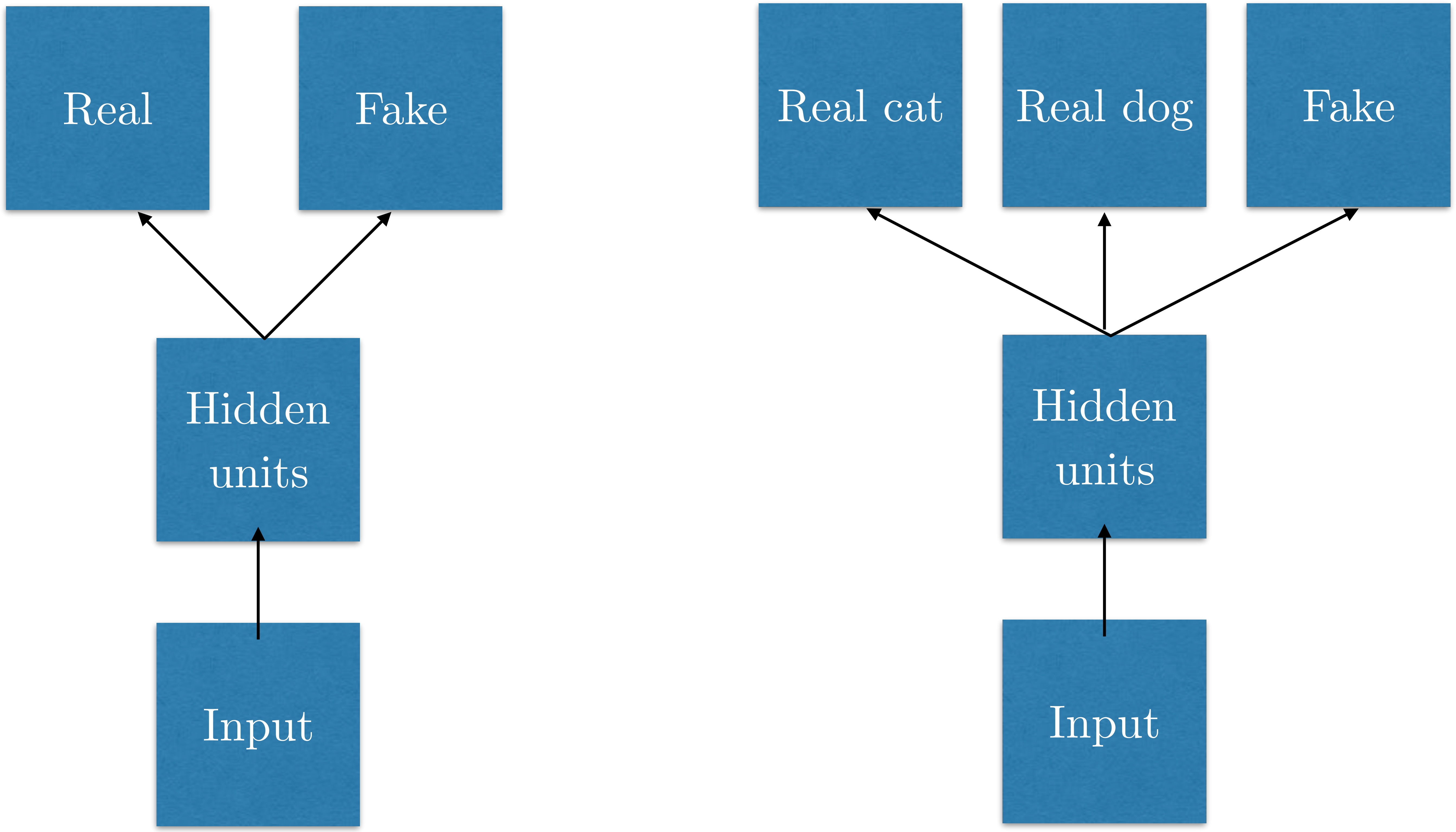
(Yeh et al., 2016)

What can you do with generative models?

- Simulated environments and training data
- Missing data
  - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Learn useful embeddings

(Goodfellow 2017)

Supervised Discriminator



(Odena 2016, Salimans et al 2016)

Semi-Supervised Classification

MNIST (Permutation Invariant)

Model	Number of incorrectly predicted test examples for a given number of labeled samples			
	20	50	100	200
DGN [21]			333 ± 14	
Virtual Adversarial [22]			212	
CatGAN [14]			191 ± 10	
Skip Deep Generative Model [23]			132 ± 7	
Ladder network [24]			106 ± 37	
Auxiliary Deep Generative Model [23]			96 ± 2	
Our model	1677 ± 452	221 ± 136	93 ± 6.5	90 ± 4.2
Ensemble of 10 of our models	1134 ± 445	142 ± 96	86 ± 5.6	81 ± 4.3

(Salimans et al 2016)

Semi-Supervised Classification

CIFAR-10

Model	Test error rate for a given number of labeled samples			
	1000	2000	4000	8000
Ladder network [24]			20.40 ± 0.47	
CatGAN [14]			19.58 ± 0.46	
Our model	21.83 ± 2.01	19.61 ± 2.09	18.63 ± 2.32	17.72 ± 1.82
Ensemble of 10 of our models	19.22 ± 0.54	17.25 ± 0.66	15.59 ± 0.47	14.87 ± 0.89

SVHN

Model	Percentage of incorrectly predicted test examples for a given number of labeled samples		
	500	1000	2000
DGN [21]		36.02 ± 0.10	
Virtual Adversarial [22]		24.63	
Auxiliary Deep Generative Model [23]		22.86	
Skip Deep Generative Model [23]		16.61 ± 0.24	
Our model	18.44 ± 4.8	8.11 ± 1.3	6.16 ± 0.58
Ensemble of 10 of our models		5.88 ± 1.0	

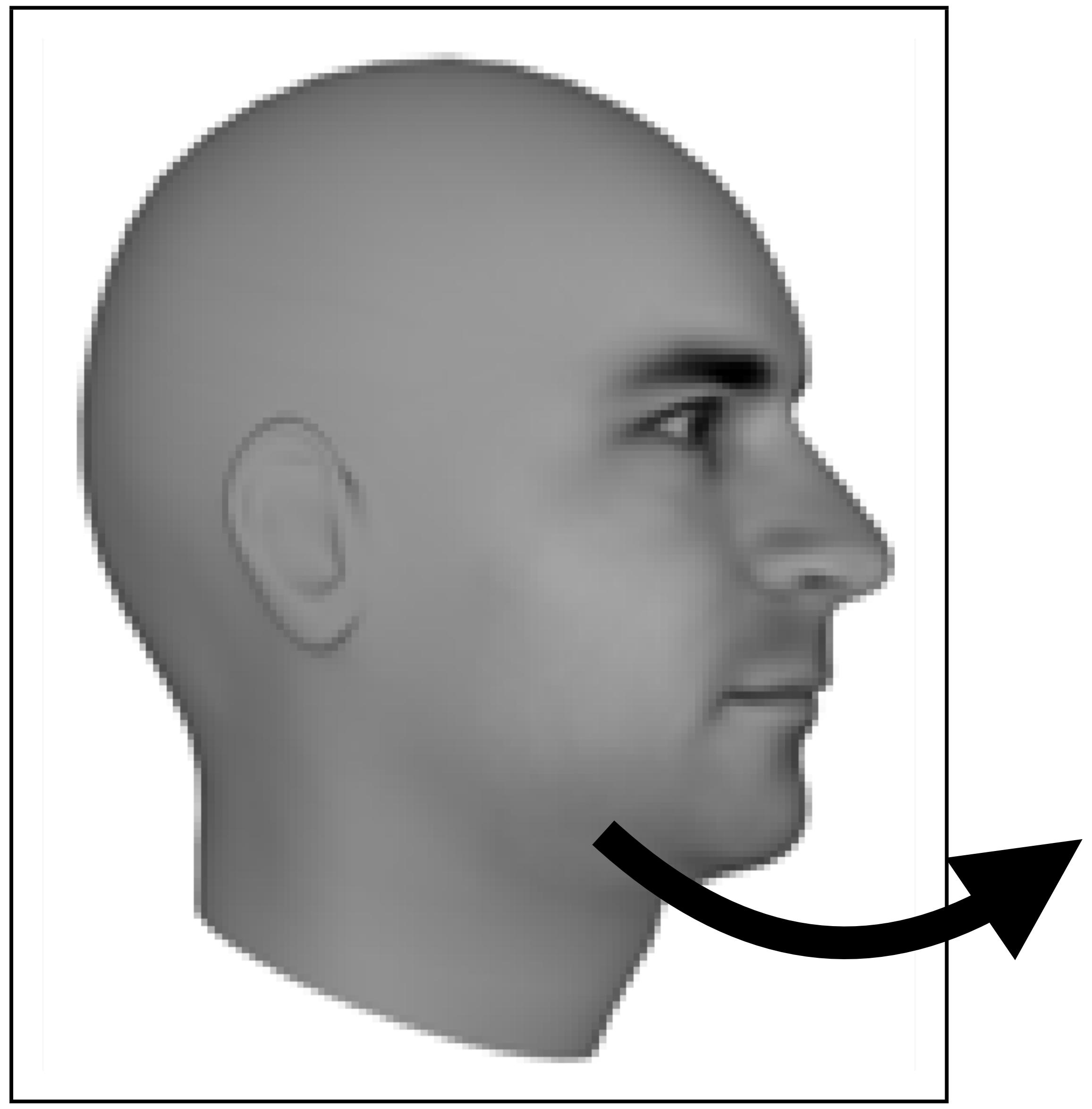
(Salimans et al 2016)

What can you do with generative models?

- Simulated environments and training data
- Missing data
  - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Learn useful embeddings

Next Video Frame Prediction

Ground Truth

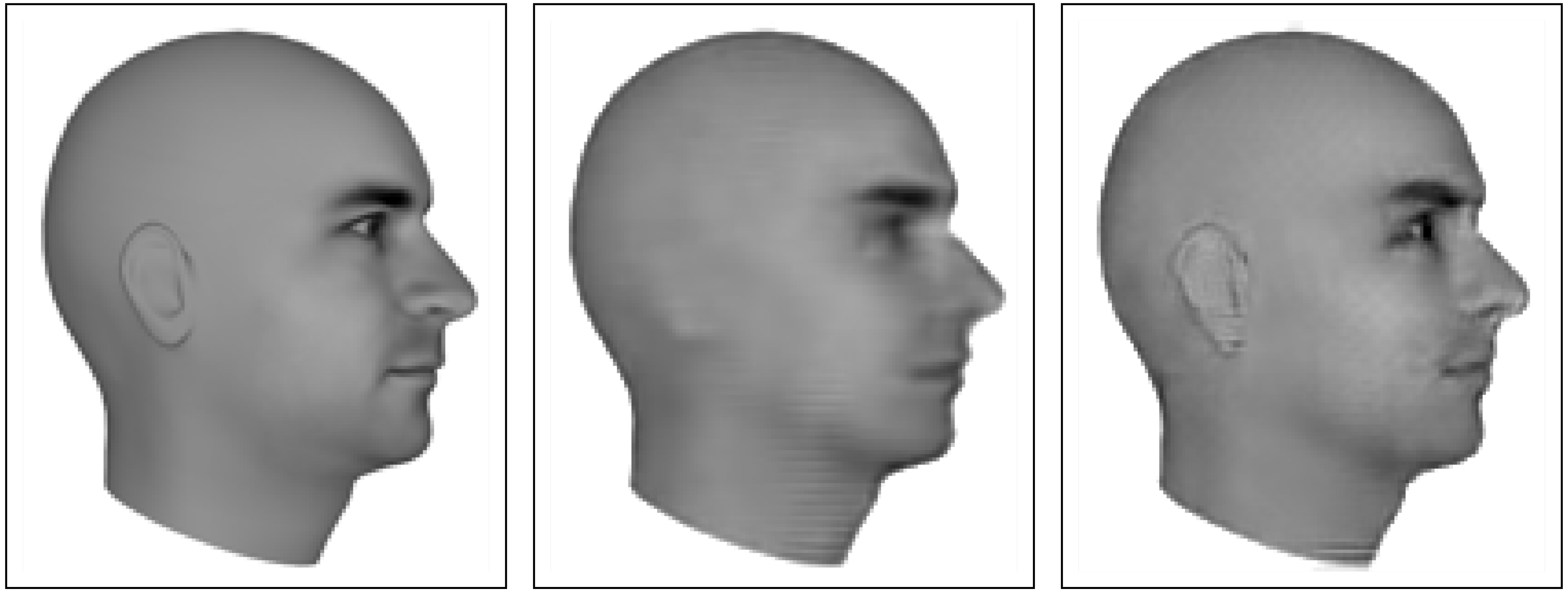


What happens next?

(Lotter et al 2016)

Next Video Frame Prediction

Ground Truth | MSE | Adversarial

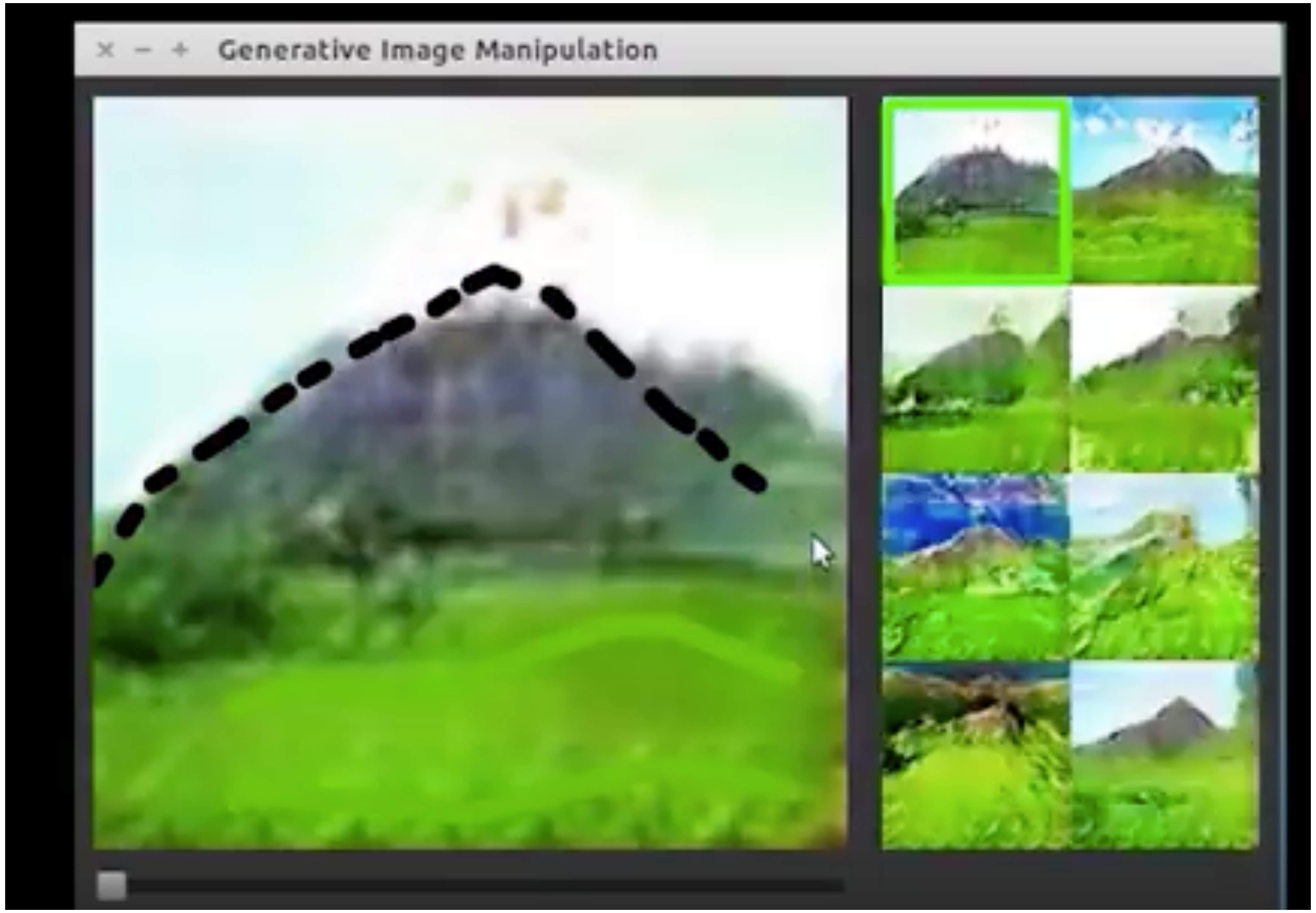


(Lotter et al 2016)

What can you do with generative models?

- Simulated environments and training data
- Missing data
    - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Learn useful embeddings

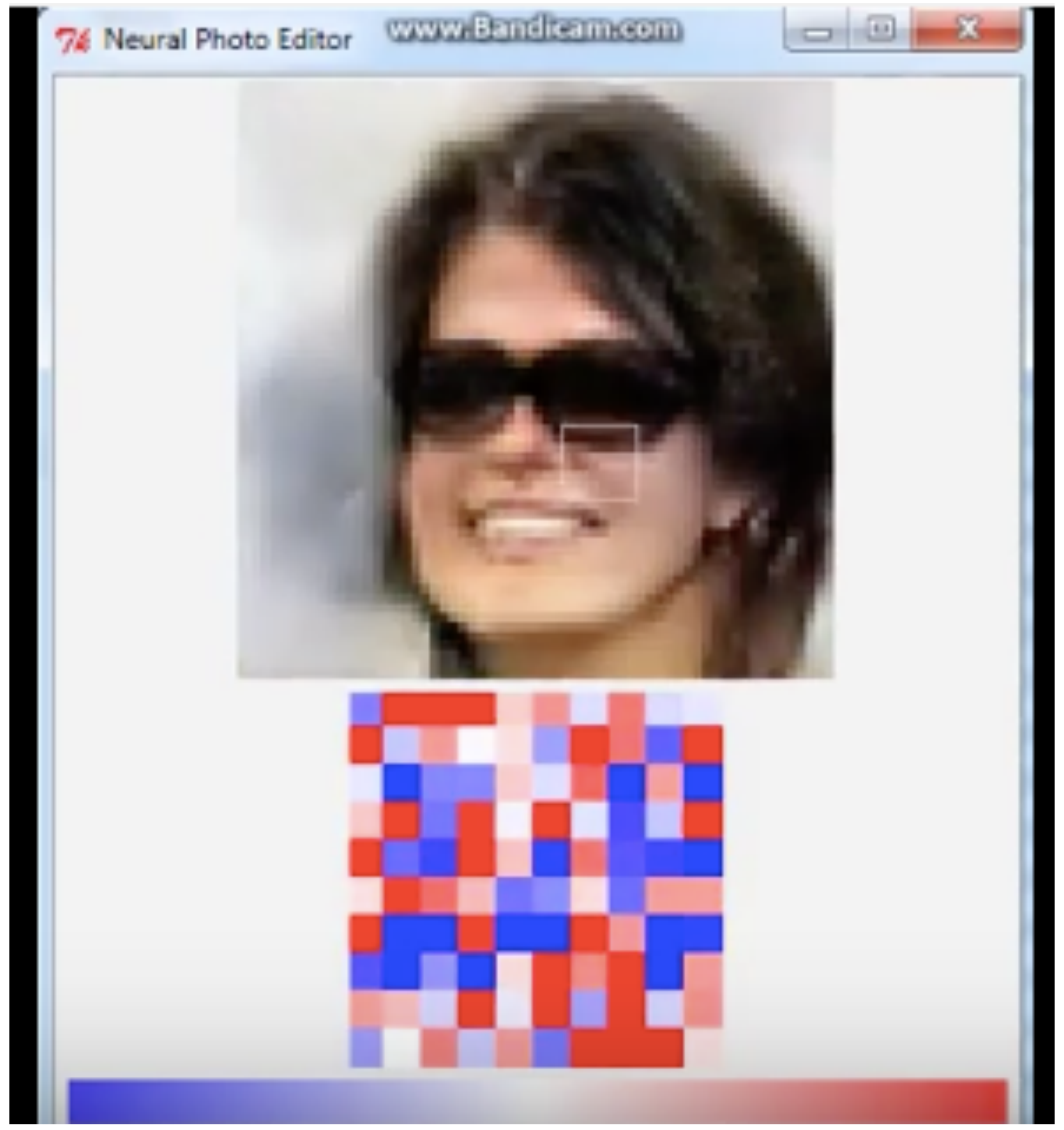
iGAN



youtube

(Zhu et al., 2016)

Introspective Adversarial Networks



youtube

(Brock et al., 2016)

Image to Image Translation

(Isola et al., 2016)

Unsupervised Image-to-Image Translation

Day to night

(Liu et al., 2017)

CycleGAN

(Zhu et al., 2017)

Text-to-Image Synthesis

This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face

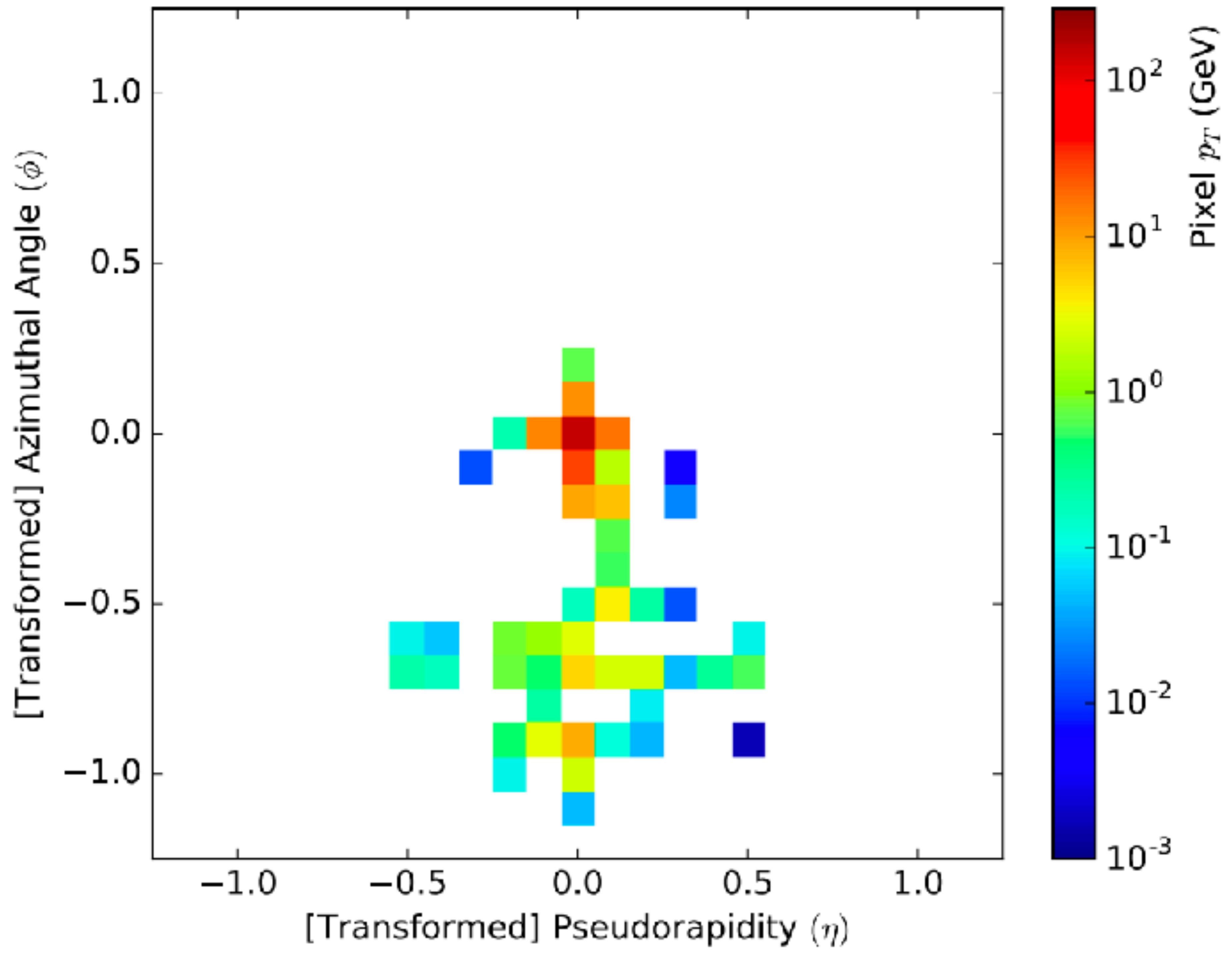
(Zhang et al., 2016)

What can you do with generative models?

- Simulated environments and training data
- Missing data
  - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Learn useful embeddings

Simulating particle physics

Save millions of dollars of CPU time by predicting outcomes of explicit simulations

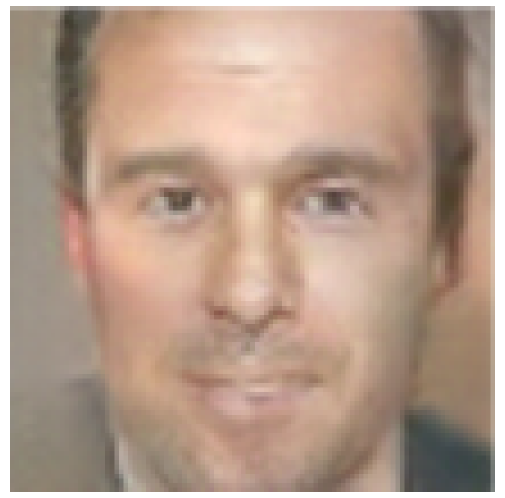
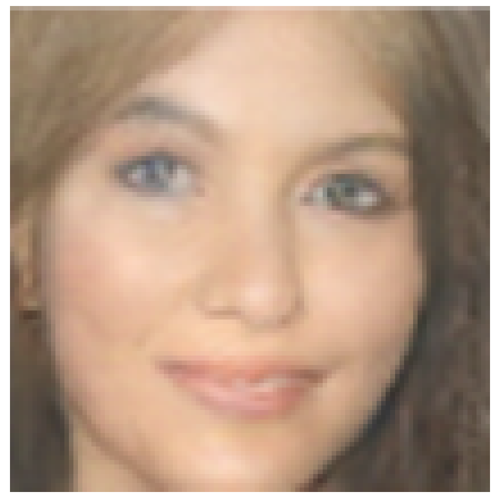


(de Oliveira et al., 2017)

What can you do with GANs?

- Simulated environments and training data
- Missing data
    - Semi-supervised learning
- Multiple correct answers
- Realistic generation tasks
- Simulation by prediction
- Learn useful embeddings

Vector Space Arithmetic

 -  +  = 

Man with glasses

Man

Woman

Woman with Glasses

(Radford et al, 2015)

Learning interpretable latent codes / controlling the generation process

(a) Azimuth (pose)

(b) Elevation

(c) Lighting

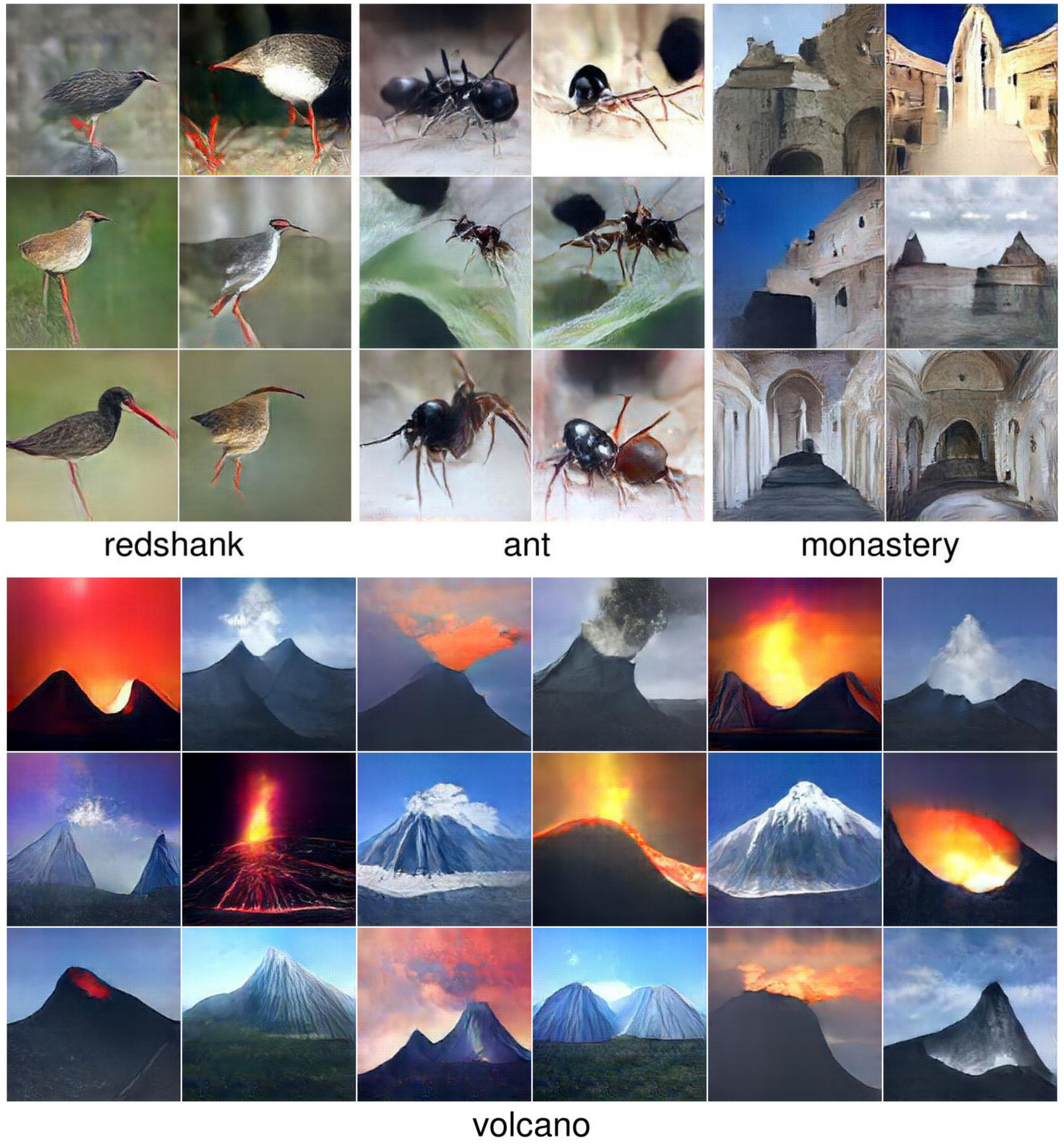
(d) Wide or Narrow

InfoGAN (Chen et al 2016)

Plug and Play Generative Networks

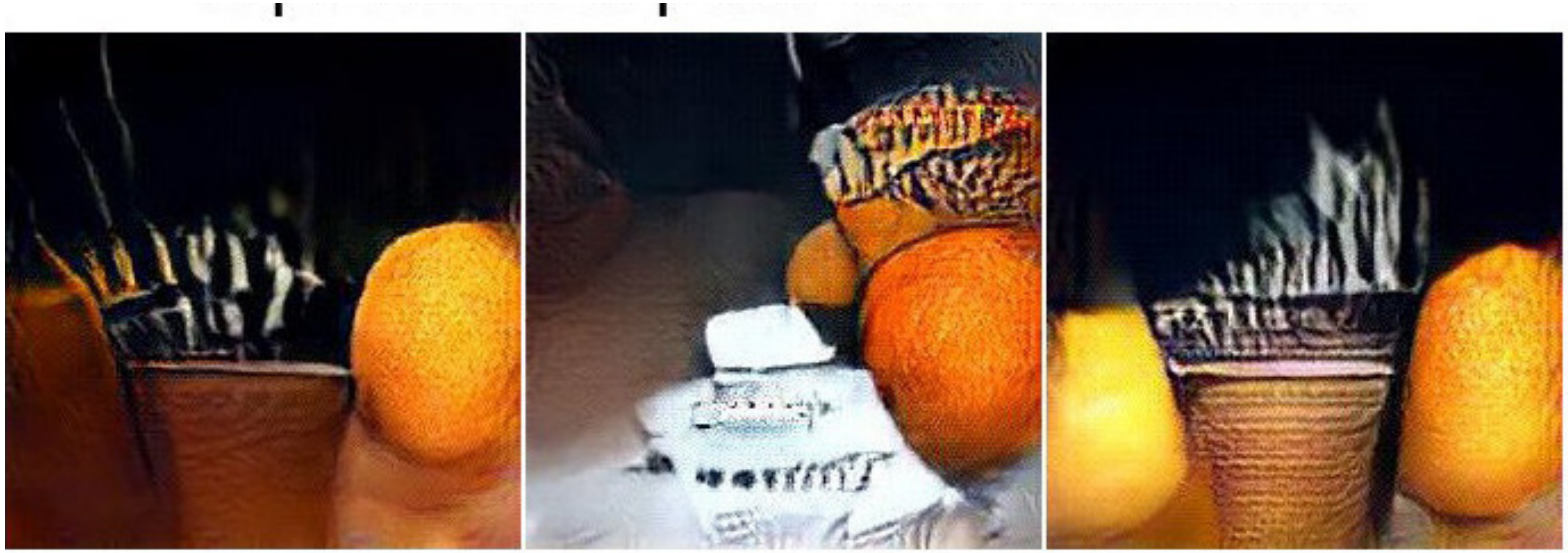
- New state of the art generative model (Nguyen et al 2016)
- Generates 227x227 realistic images from all ImageNet classes
- Combines adversarial training, moment matching, denoising autoencoders, and Langevin sampling

PPGN Samples



(Nguyen et al 2016)

PPGN for caption to image



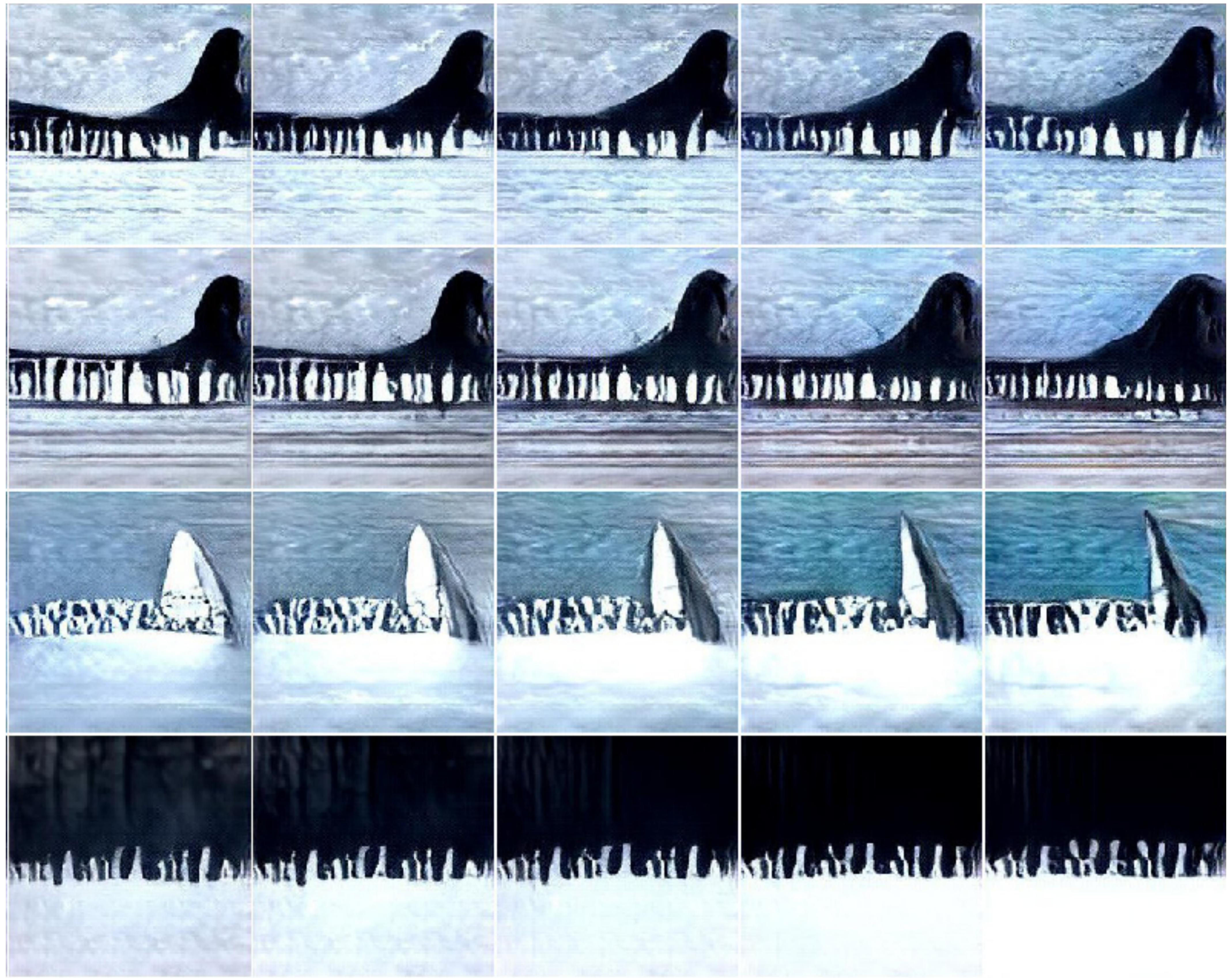
oranges on a table next to a liquor bottle

(Nguyen et al 2016)

Basic idea

- Langevin sampling repeatedly adds noise and gradient of log $p(x,y)$ to generate samples (Markov chain)
- Denoising autoencoders estimate the required gradient
- Use a special denoising autoencoder that has been trained with multiple losses, including a GAN loss, to obtain best results

Sampling without class gradient



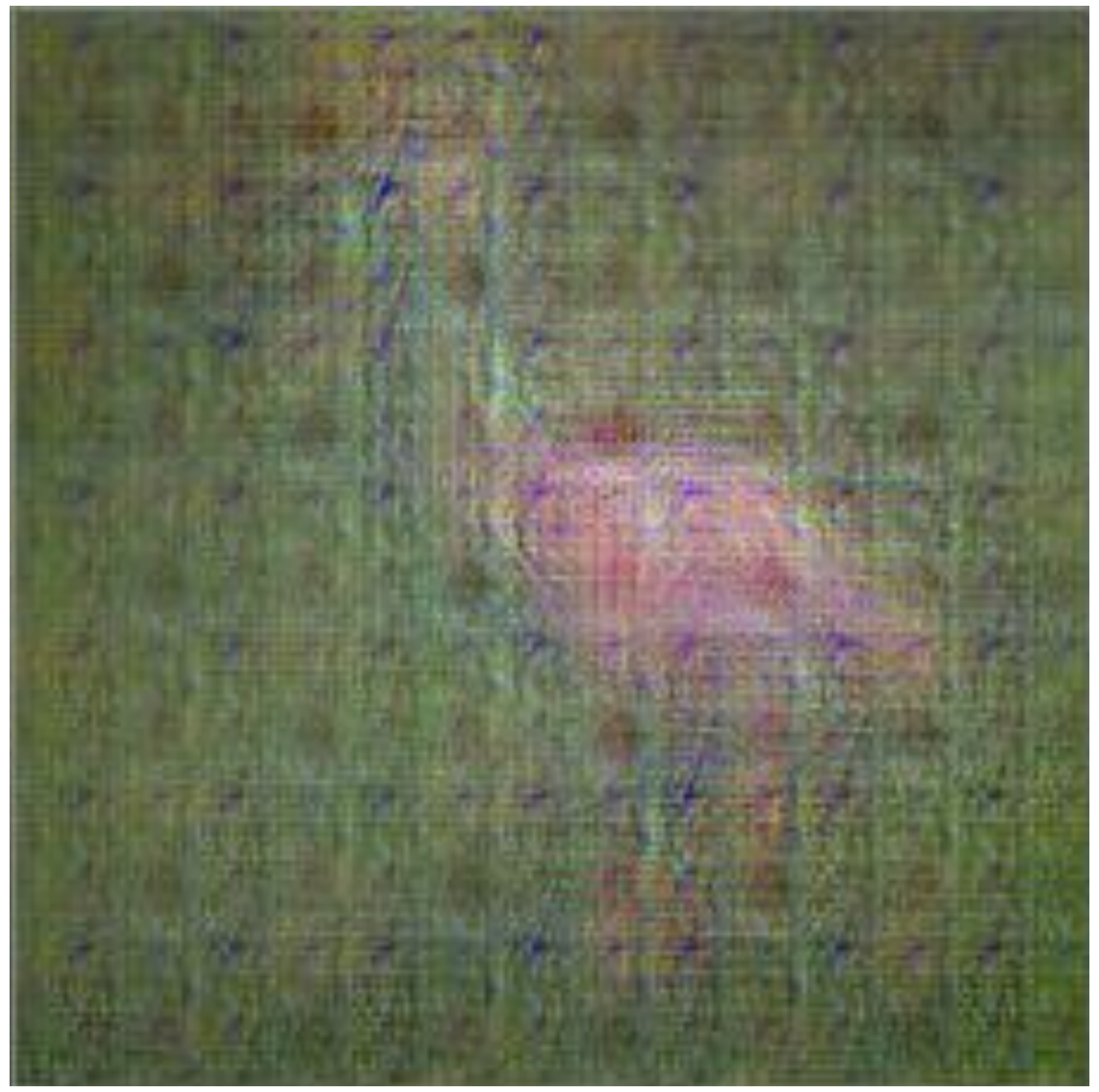
epsilon1 = 0, epsilon2 = 1e-5

(Nguyen et al 2016)

GAN loss is a key ingredient

Raw data

Reconstruction by PPGN



Reconstruction by PPGN without GAN

Images from Nguyen et al 2016

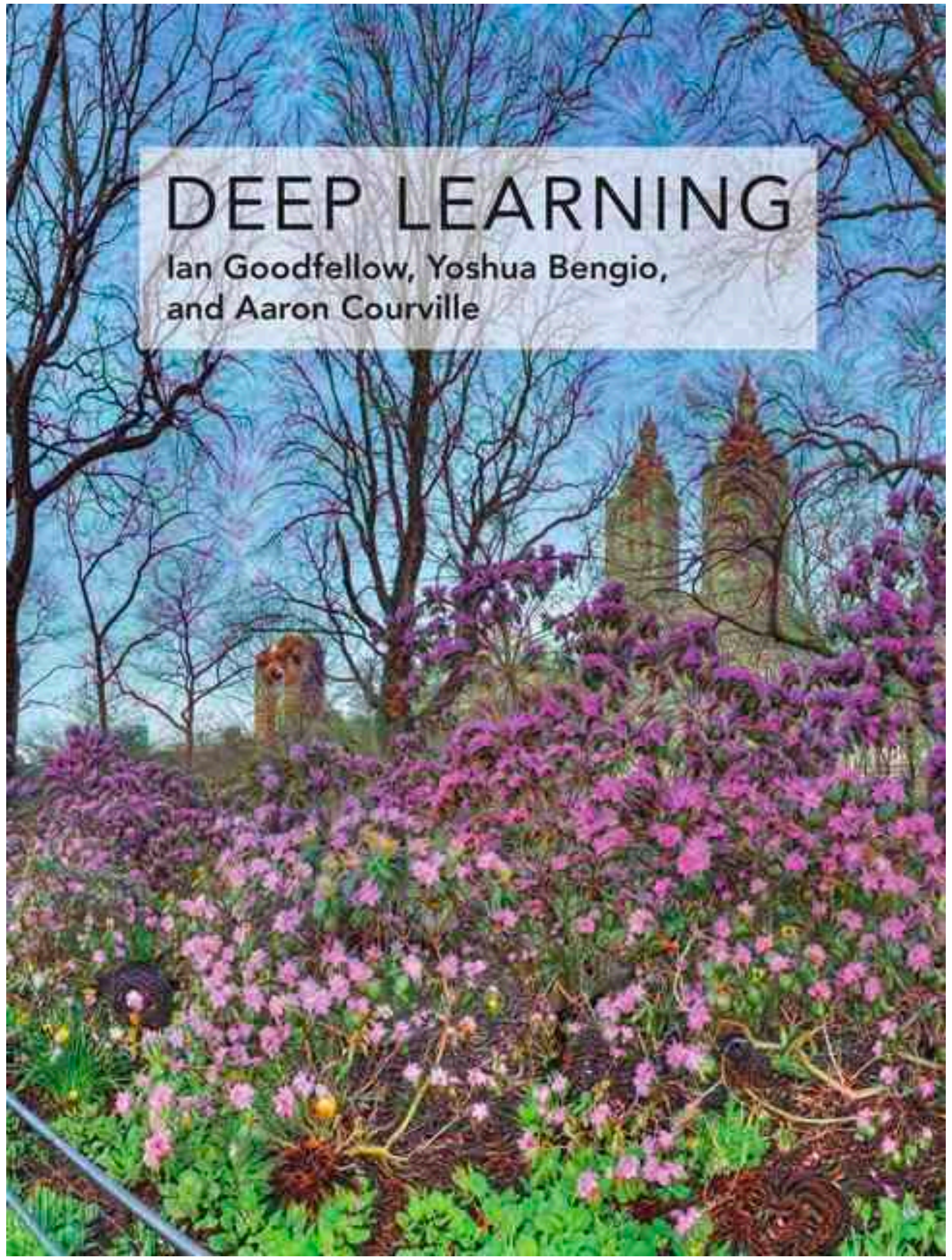
First observed by Dosovitskiy et al 2016

(Goodfellow 2017)

To be continued...

- Generative Models II will be taught by Aaron Courville

For more information...



www.deeplearningbook.org