Science and Technology Facilities Council

Machine Learning for Science

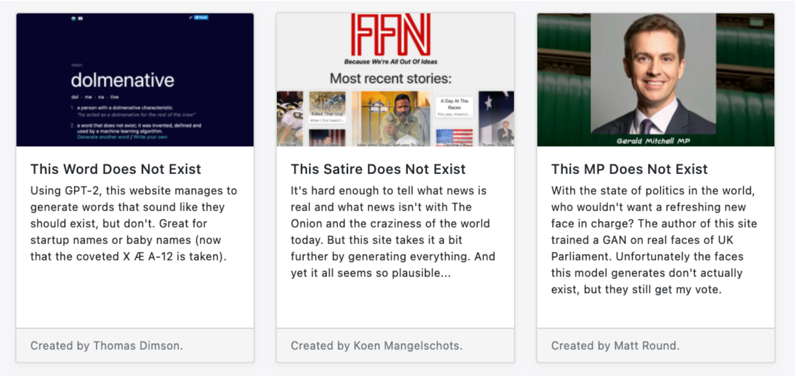
Generative Adversarial Networks

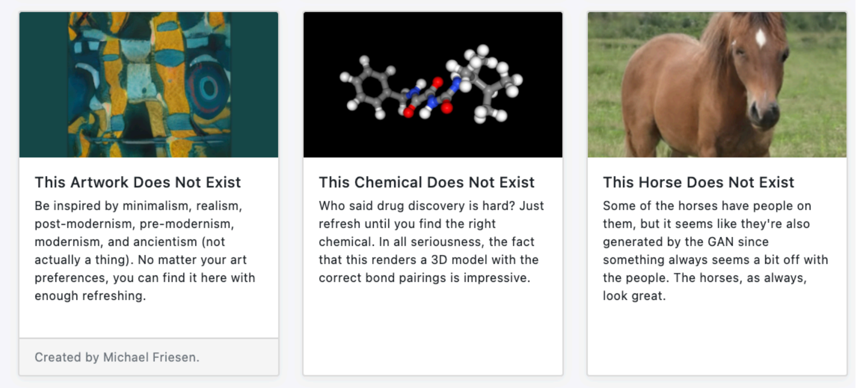
Kuangdai Leng

Scientific Machine Learning (SciML) Group
Rutherford Appleton Laboratory

Introduction

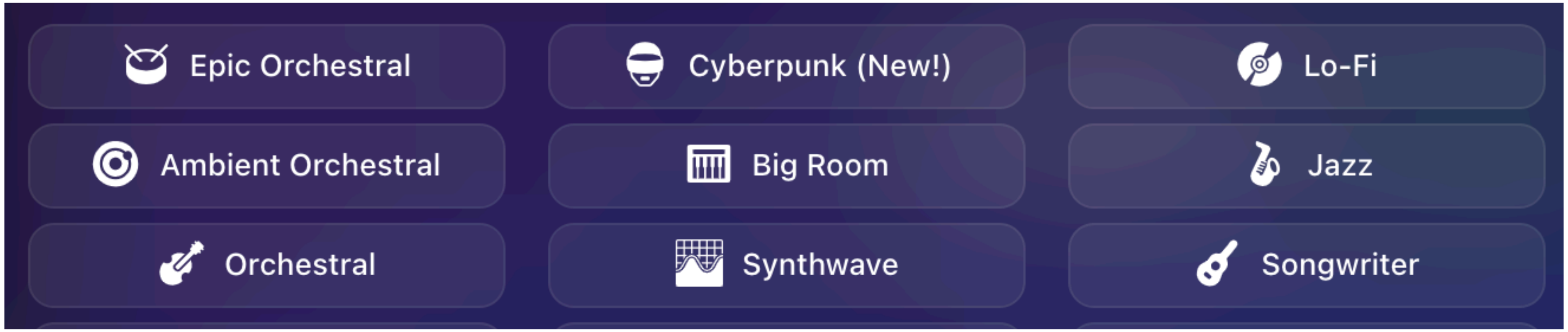
- Generative Adversarial Networks or GANs prove to be one of the most exciting ideas in deep learning
- GANs have been used to generate *diverse realistic comprehensive* contents





This X Does Not Exist

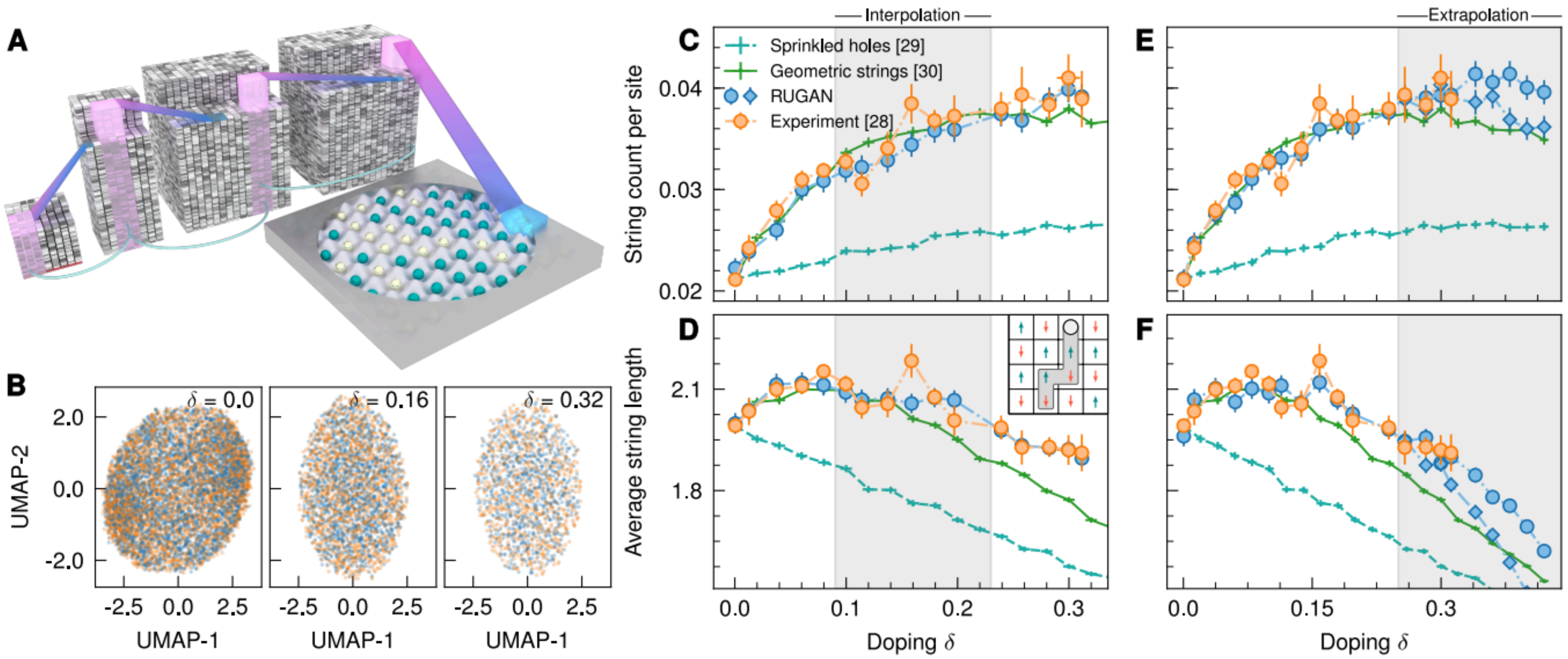
By Style-based GANs



AIVA

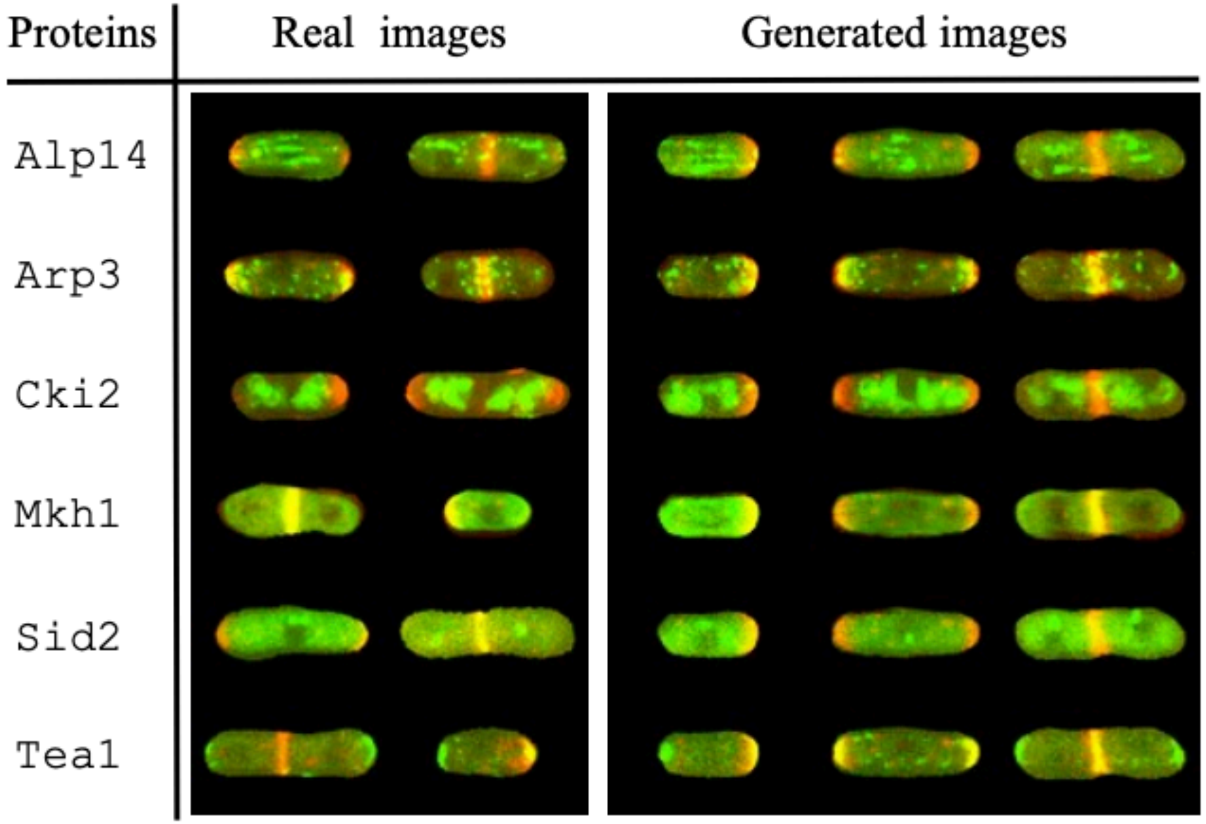
Scientific applications

- Regressive upscaling GANs to produce snapshots of a doped two-dimensional Fermi-Hubbard model (Casert et al., 2020)



Scientific applications

- GANs for Biological Image Synthesis of cells imaged by fluorescence microscopy (Osokin et al., 2017)



In this lecture

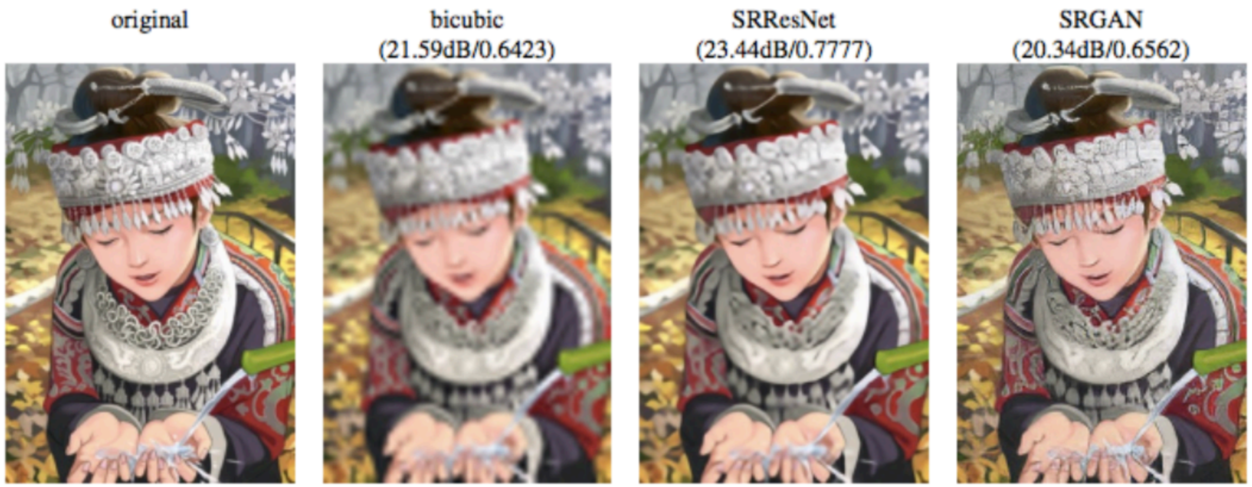
- Generative Models and Maximum Likelihood
- How GANs work and challenges
- GANs and VAEs: comparison and combination
- Summary
- Practicals: GANs for MNIST and Inelastic Neutron Scattering

Following the structure of

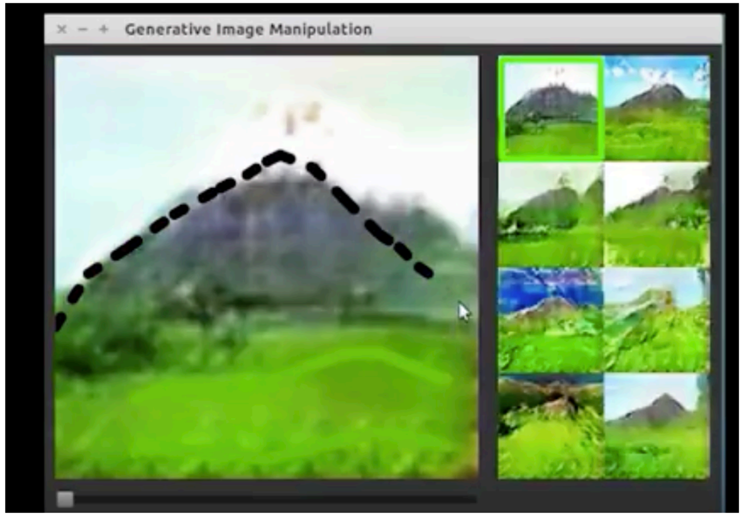
NIPS 2016 Tutorial: Generative Adversarial Networks, Ian Goodfellow, 2016

Why generative models?

- High-dimensional probability distributions
- Realistic content generation (fake is good or even better, e.g., art)
- Data expansion and missing labels (semi-supervised)
- Reinforcement learning
- Inspiring human beings



SRGAN
Ledig et al. 2016



Interactive GAN (iGAN)
Zhu et al. 2016

Maximum Likelihood Estimate

- Maximum Likelihood

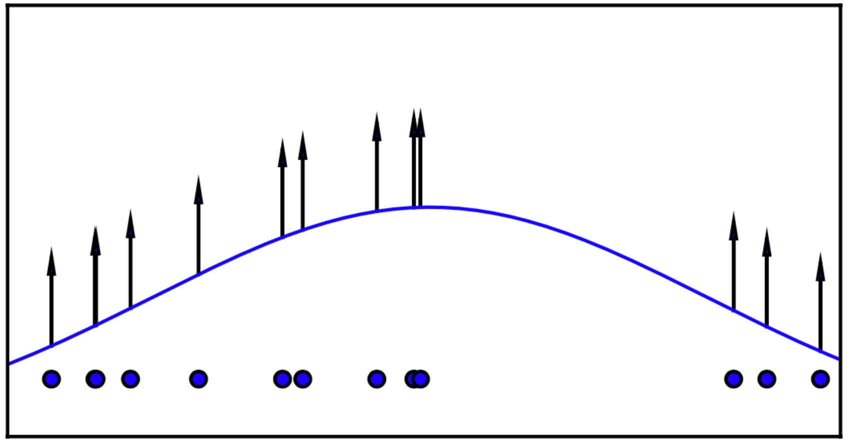
$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \prod_{i=1}^{m} p_{\text{model}}\left(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}\right)$$

$$= \arg\max_{\boldsymbol{\theta}} \log \prod_{i=1}^{m} p_{\text{model}}\left(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}\right)$$

$$= \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{m} \log p_{\text{model}}\left(\boldsymbol{x}^{(i)}; \boldsymbol{\theta}\right).$$

- Minimising KL-divergence between data and model distributions

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} D_{\text{KL}}\left(p_{\text{data}}(\boldsymbol{x}) \| p_{\text{model}}(\boldsymbol{x}; \boldsymbol{\theta})\right).$$

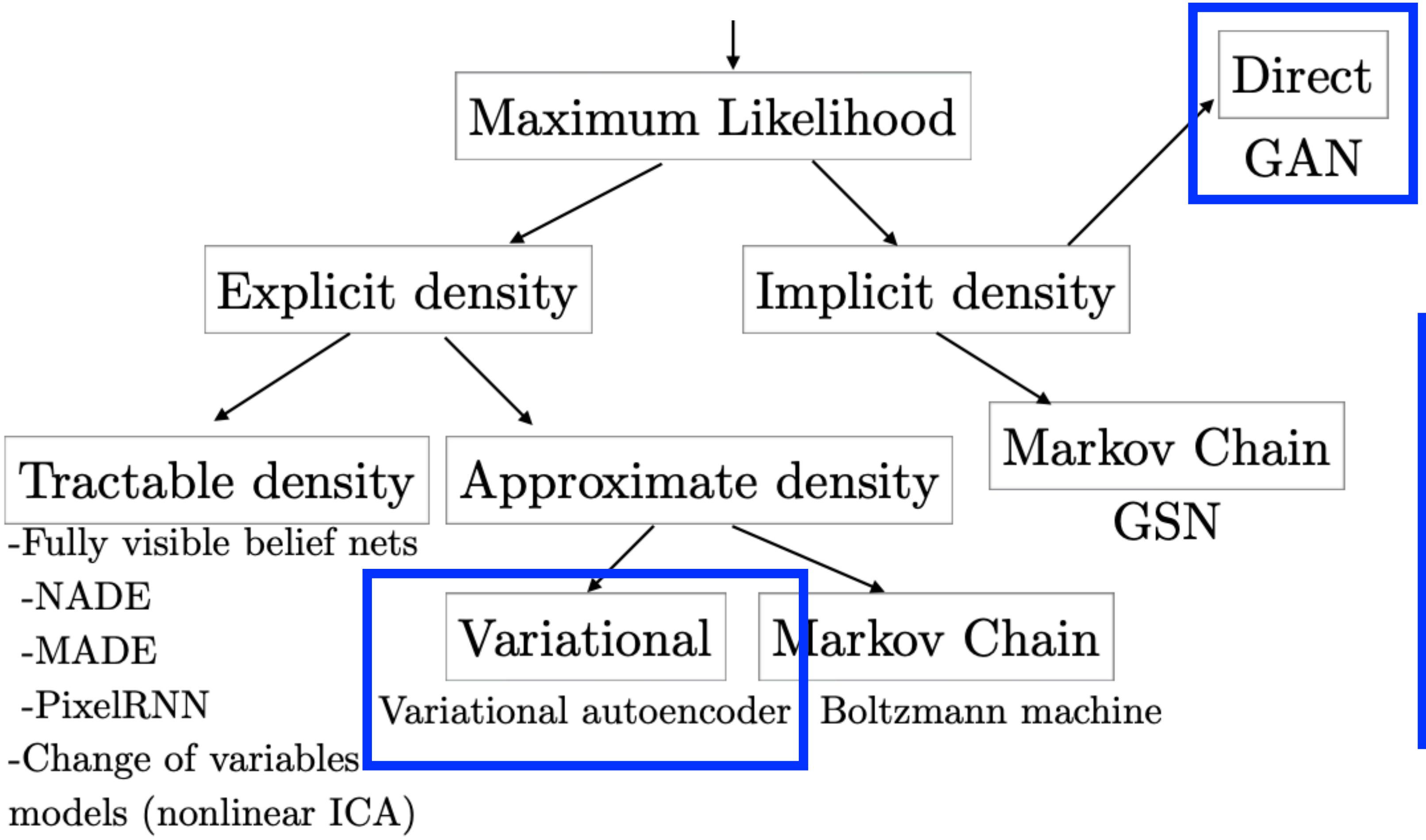


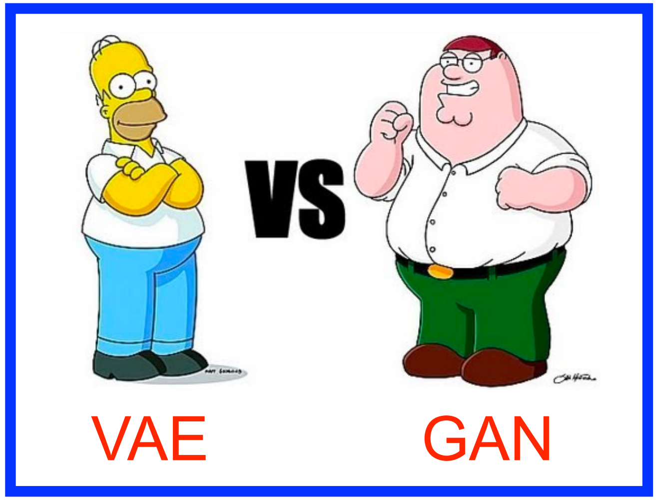
$p_{\text{data}}(x)$ is unknown

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

Maximum Likelihood Estimate

- Taxonomy of generative models based on Maximum likelihood



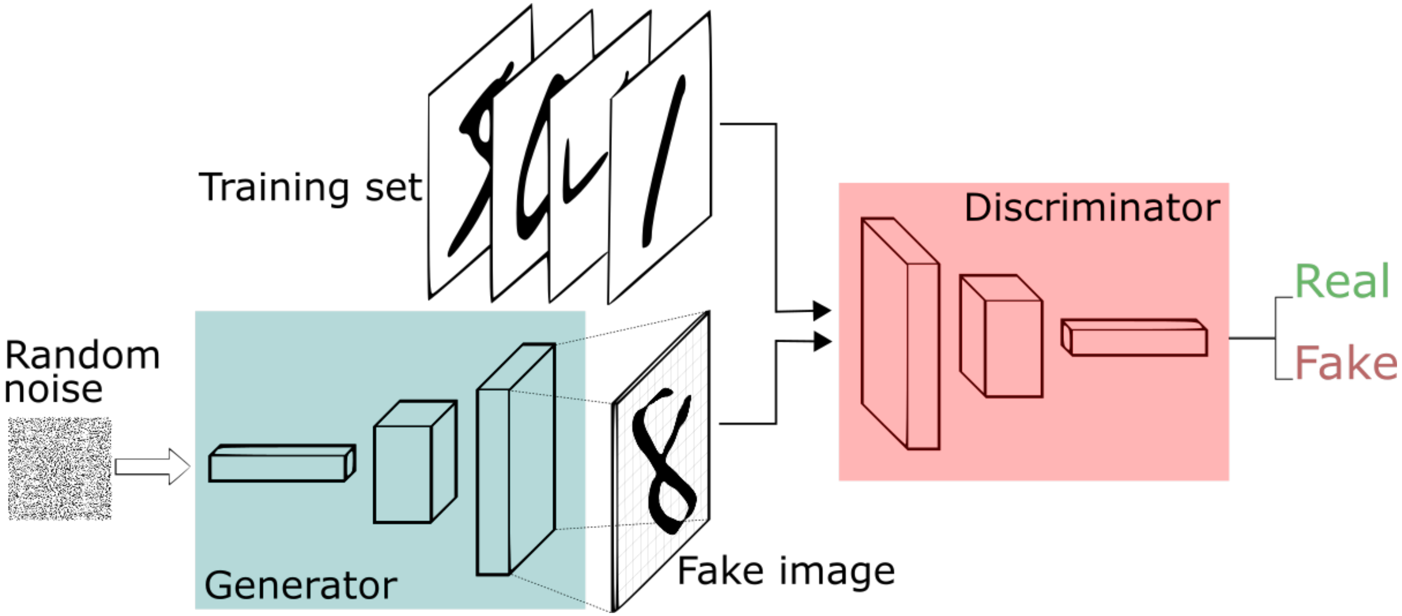


In this lecture

- Generative Models and Maximum Likelihood
- How GANs work and challenges
- GANs and VAEs: comparison and combination
- Summary
- Practicals: GANs for MNIST and Inelastic Neutron Scattering

How GANs work

- A GAN contains a generator and a discriminator competing in a two-player game
  - **Generator**: generates fake images from a latent sampling, $G(z) \rightarrow x'$, and tries to fool the discriminator that these images are real
  - **Discriminator**: tries to distinguish between fake and real images, $D(x) \rightarrow 0/1$



They can be regard as collaborators because they share information completely: the discriminator is more like a teacher than an adversary (Goodfellow, 2016)

How GANs work

- Discriminator Loss

$$J^{(D)}(\boldsymbol{\theta}^{(D)}, \boldsymbol{\theta}^{(G)}) = -\frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}} \log D(\boldsymbol{x}) - \frac{1}{2}\mathbb{E}_{\boldsymbol{z}} \log\left(1 - D\left(G(\boldsymbol{z})\right)\right).$$

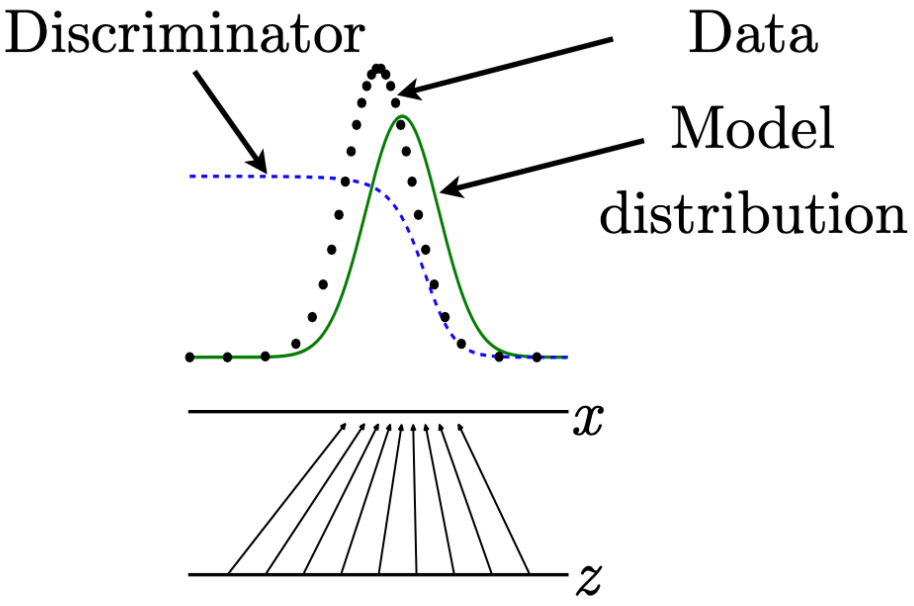
Disc decision for real images

Disc decision for fake images

- The ratio of densities

$$\frac{\delta}{\delta D(\boldsymbol{x})} J^{(D)} = 0.$$

$$D^*(\boldsymbol{x}) = \frac{p_{\text{data}}(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x}) + p_{\text{model}}(\boldsymbol{x})}.$$



How GANs work

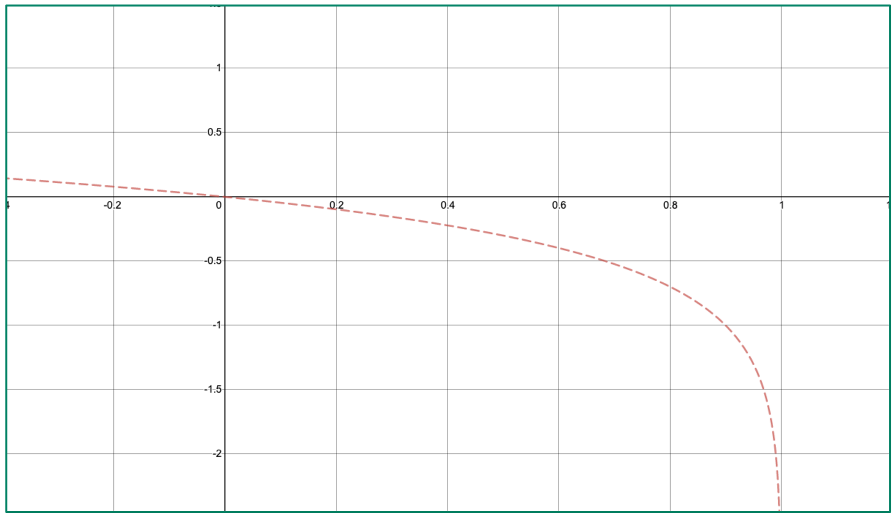
- Generator Loss: zero-sum game (simplest case)

$$J^{(G)} = -J^{(D)}.$$

- The minimax of a value function

$$\boldsymbol{\theta}^{(G)*} = \arg\min_{\boldsymbol{\theta}^{(G)}} \max_{\boldsymbol{\theta}^{(D)}} V\left(\boldsymbol{\theta}^{(D)}, \boldsymbol{\theta}^{(G)}\right).$$

where $V\left(\boldsymbol{\theta}^{(D)}, \boldsymbol{\theta}^{(G)}\right) = -J^{(D)}\left(\boldsymbol{\theta}^{(D)}, \boldsymbol{\theta}^{(G)}\right).$



- Minimization in an outer loop
- Maximization in an inner loop
- Useful for theoretical analysis
- Impractical because of vanishing gradient of generator ($\log(1-x)$ is flat near the fake end)

How GANs work

- Generator Loss: heuristic game

$$J^{(G)} = -\frac{1}{2}\mathbb{E}_{\boldsymbol{z}} \log D(G(\boldsymbol{z}))$$

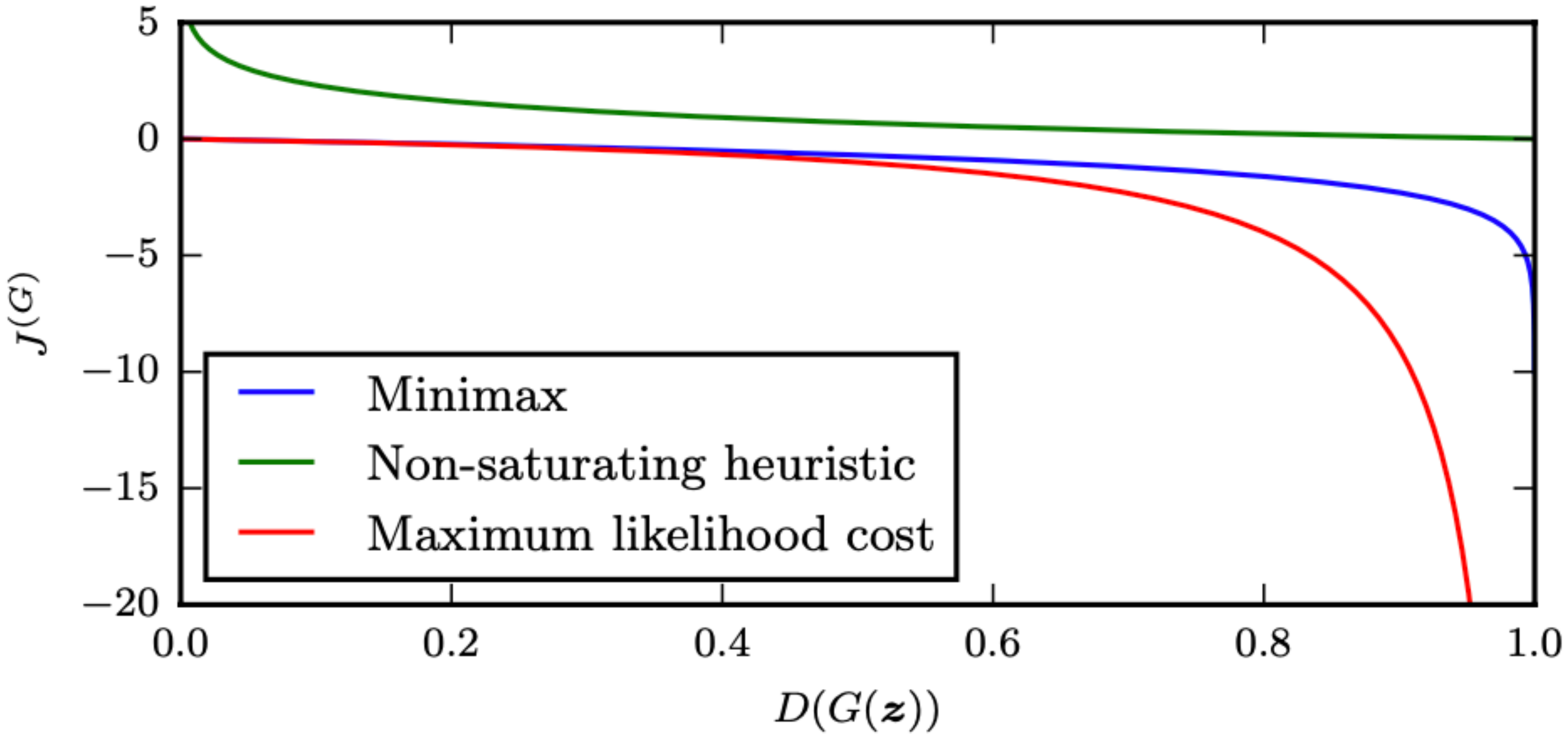
  - In the minimax game, the generator minimizes the log-probability of the discriminator being correct
  - In this game, the generator maximizes the log-probability of the discriminator being mistaken

- Generator Loss: maximum likelihood

$$J^{(G)} = -\frac{1}{2}\mathbb{E}_{z} \exp\left(\sigma^{-1}\left(D(G(\boldsymbol{z}))\right)\right)$$

How GANs work

- Generator Loss comparisons



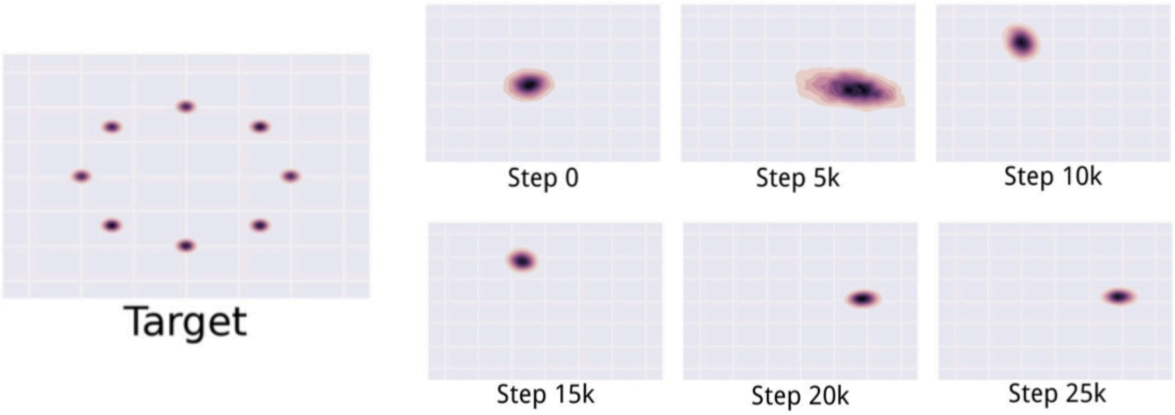
- Variance distribution
- Variance reduction

Challenges: Nash Equilibrium

- G and D are competing; if one stops evolve, both will stop (vanishing gradient)
  - What if G is much stronger than D?
  - What if D is much stronger than G?
- Balancing G and D is not straightforward because it depends on $p_{\text{data}}(x)$
- What we can do:
  - Loss parameterisation
  - Model architecture
  - Asynchronous training (one exercise in the notebook)

Challenges: Mode Collapse

- Mode collapse is the most significant challenge for GANs
- The generator decodes every z into one or one type of x

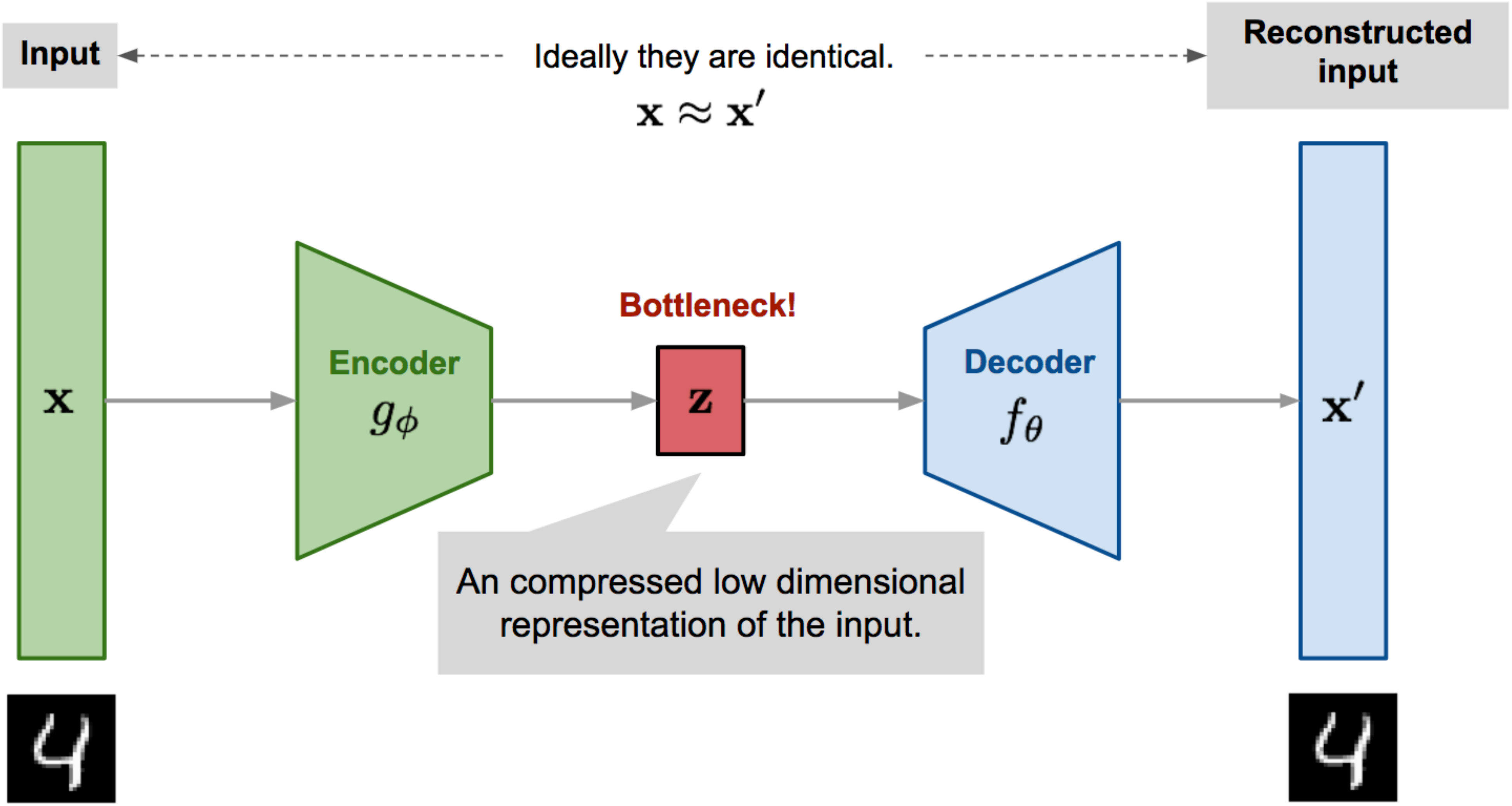


- This is an unsolved problem yet with some solutions
  - Mini-batch features (Salimans et al., 2016)
  - Unrolled GANs (Metz et al., 2016)
  - Wasserstein loss (Arjovsky et al., 2017)
    - Discriminator loss: $D(x) - D(G(z))$; Generator loss: $D(G(z))$

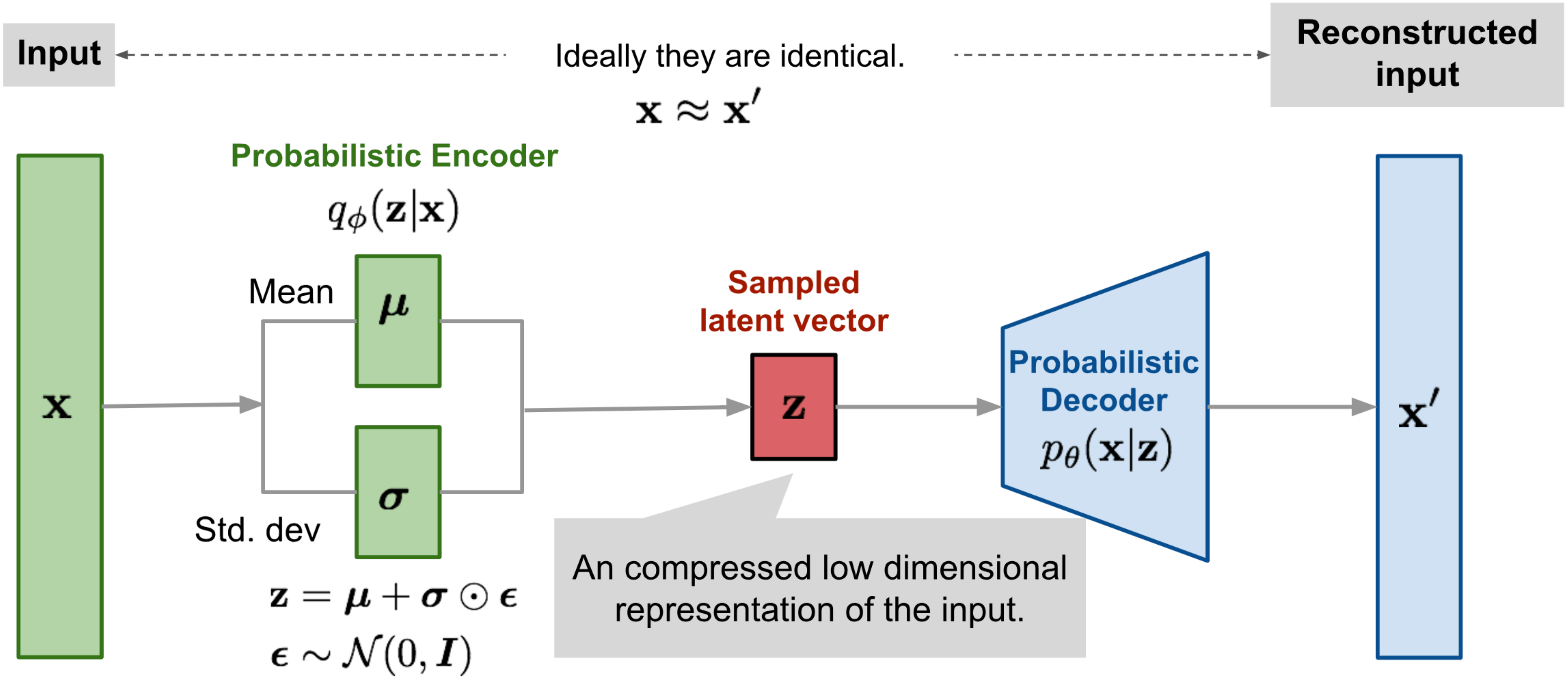
In this lecture

- Generative Models and Maximum Likelihood
- How GANs work and challenges
- GANs and VAEs: comparison and combination
- Summary
- Practicals: GANs for MNIST and Inelastic Neutron Scattering

Autoencoder (AE) and Variational Autoencoder (VAE)



Autoencoder (AE) and Variational Autoencoder (VAE)



Variational inference

- True generative process (e.g., PDE for a physics problem)

$$z \quad \Rightarrow \quad p(x|z) \quad \Rightarrow \quad x$$

- Given x (image data), we want to infer $p(z|x)$, the *posterior probability*

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} = \frac{p(x|z)p(z)}{\int p(x|z')p(z')\,dz'}$$ intractable (MC)

- Variational inference: use some distribution $q(z|x)$ to approximate $p(z|x)$

$$q(z|x) \sim p(z|x) \quad \Leftarrow \quad \min \mathrm{KL}(q(z|x), p(z|x))$$

Variational inference

- Kullback–Leibler divergence: “distance” between two distributions

$$\mathrm{KL}(q,p) = -\,\mathrm{E}_q \log \frac{p}{q} \quad \Rightarrow \quad \mathrm{KL}(q,p) \geq 0 \quad \mathrm{KL}(q,p) \neq \mathrm{KL}(p,q)$$

Don’t have to use KL divergence (β-VAE)

- Variational identity

$$\mathrm{KL}(q(z\,|\,x), p(z\,|\,x)) = \log p(x) - \left(-\mathrm{KL}(q(z\,|\,x), p(z)) + \mathrm{E}_{q(z|x)} \log p(x\,|\,z)\right)$$

- Evidence lower bound

$$\log p(x) \geq \underbrace{-\,\mathrm{KL}(q(z\,|\,x), p(z)) + \mathrm{E}_{q(z|x)} \log p(x\,|\,z)}_{\textbf{ELBO}}$$

Variational Autoencoder

- Evidence lower bound

$$\log p(x) \geq -\,\mathrm{KL}(q(z\,|\,x), p(z)) + \mathrm{E}_{q(z|x)} \log p(x\,|\,z)$$

$$\Downarrow \quad \textbf{VAE}$$

$$\log p_\theta(x) \geq -\mathrm{KL}(q_\phi(z \mid x), p(z)) \\ + \mathrm{E}_{q_\phi(z|x)} \log p_\theta(x \mid z).$$

- θ: Generative model or the decoder (θ looks like g)
- ϕ: Inference model or the encoder (ϕ has an I in the middle)

Variational Autoencoder

- Maximum likelihood training

$$\mathrm{E}_{p_{\mathcal{D}}(x)} \log p_\theta(x),$$

Data

$$\Downarrow$$

Approx. posterior

Prior

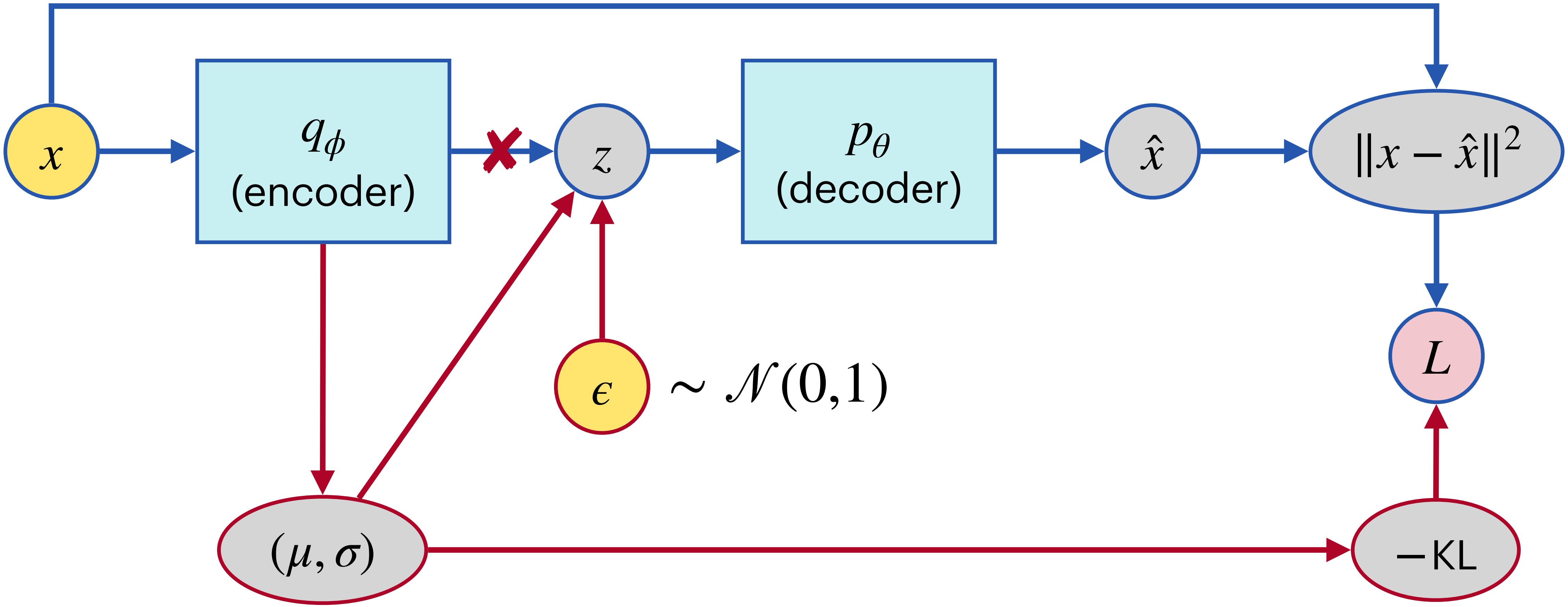
$$\max_\theta \max_\phi \mathrm{E}_{p_{\mathcal{D}}(x)} \Big[-\mathrm{KL}(q_\phi(z \mid x), p(z)) + \mathrm{E}_{q_\phi(z|x)} \log p_\theta(x \mid z) \Big].$$

Reconstruction error

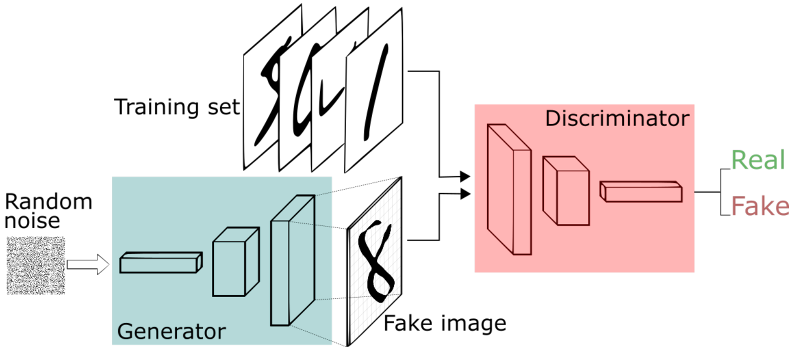
$$\boxed{\begin{aligned} q_\phi(z \,|\, x) &= \mathcal{N}(\mu, \sigma^2) \\ p(z) &= \mathcal{N}(0,1) \end{aligned}}$$

Variational Autoencoder

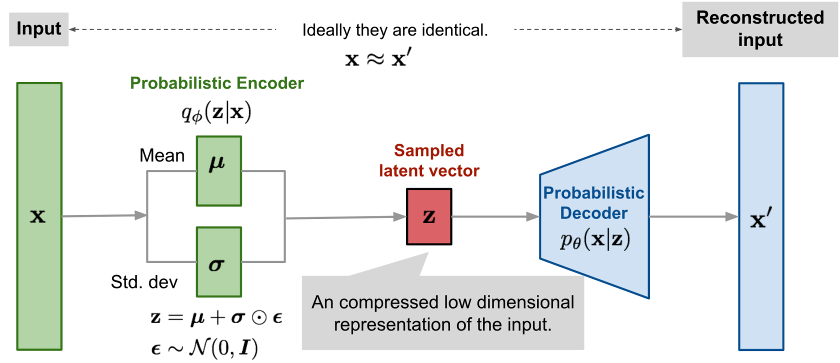
- Architecture



GANs and VAEs

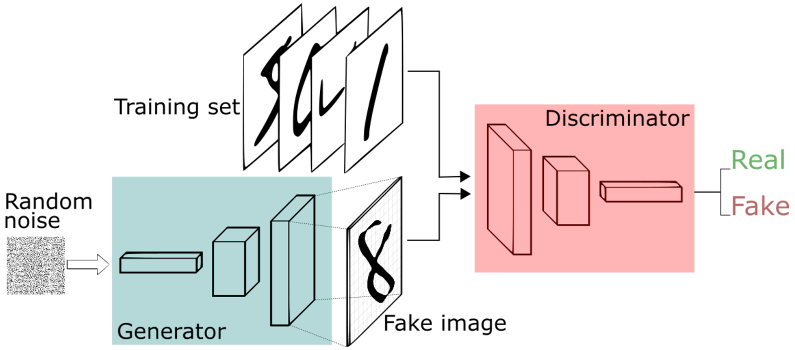


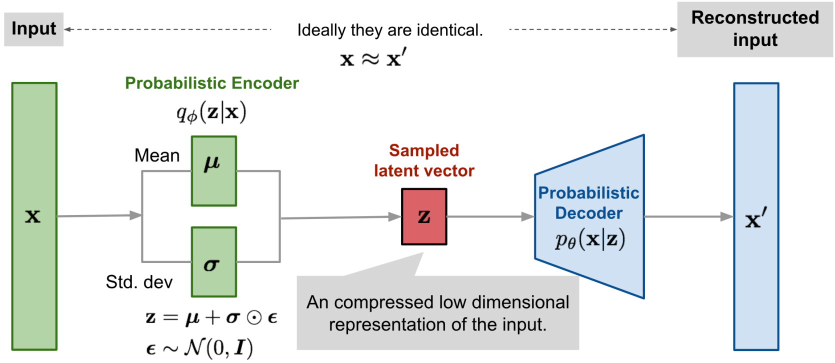
VS



	Sharp	Content generation	Vague
icGAN, infoGAN	No	Inference model	Yes
	Hard	Feature control and PCA	Easy
	Hard	Designing & training	Easy
	Implicit	Maximum likelihood	Explicit
	Indirect	Generator sees data	Direct

GANs and VAEs





- Notable efforts to combine GANs and VAEs
  - **Simple merge** (Larsen et al., 2015): VAE's decoder as GAN's generator
  - **Adversarial AE** (Makhzani et al., 2015): VAE's latent (+ seed) as input of GAN's generator
  - **Bidirectional GANs** (Donahue et al., 2016): Both VAE's encoder and decoder provide (feature, data) pairs as the input of GAN
  - **Adversarial Variational Bayes** (Lars et al., 2017): Use a GAN to compute VAE's variational loss without assuming a posterior form (most theoretically elegant)

In this lecture

- Generative Models and Maximum Likelihood
- How GANs work and challenges
- GANs and VAEs: comparison and combination
- Summary
- Practicals: GANs for MNIST and Inelastic Neutron Scattering

Summary

- GANs are powerful to generate realistic contents of nearly all types
- The generator and discriminator compete in a two-player game
- The choice of loss functions is a key issue
- GANs are difficult to train (instability)
  - Vanishing gradient
  - Mode collapse
- Lake of an inference model (important to science) is a major drawback
- Combing VAEs and GANs is a promising research direction

In this lecture

- Generative Models and Maximum Likelihood
- How GANs work and challenges
- GANs and VAEs: comparison and combination
- Summary
- Practicals: GANs for MNIST and Inelastic Neutron Scattering

Science and Technology Facilities Council

Thank you

Science and Technology Facilities Council

@STFC_matters

Science and Technology Facilities Council