Introduction to generative models

Mihaela Rosca
@elaClaudia

DeepMind

Generative models are a core building block of intelligent systems.

What do intelligent systems need?

- Generate new data
- Imagine possible futures & have a model of the world
- Translate between data modalities
- Learn useful representations
- Completing missing data

Generate new data

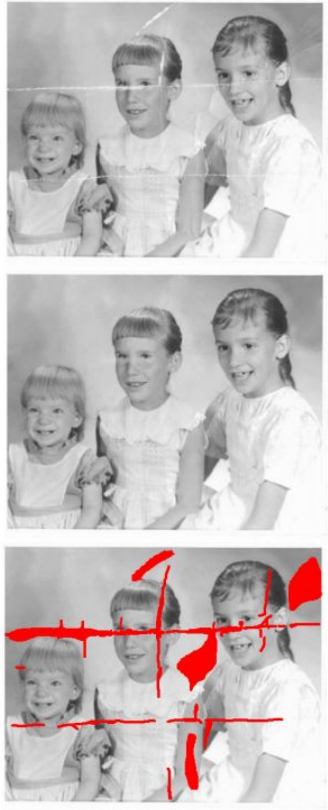
What do intelligent systems need?

Royour was I did too jest of this forget
That I must I should be report of tale
Decost we are bewarved:'d: yet my fearful scope
From whence the duty I may need their course,
Which thou wert sorry for my party was to show
Forthwith Edward for what stout King Richard death!
The queen hath cast froths me to said,
Is, and not to be framed, and let's with him.

https://arxiv.org/abs/1710.10196
https://arxiv.org/pdf/1609.03499.pdf
https://arxiv.org/abs/1308.0850

Complete missing data

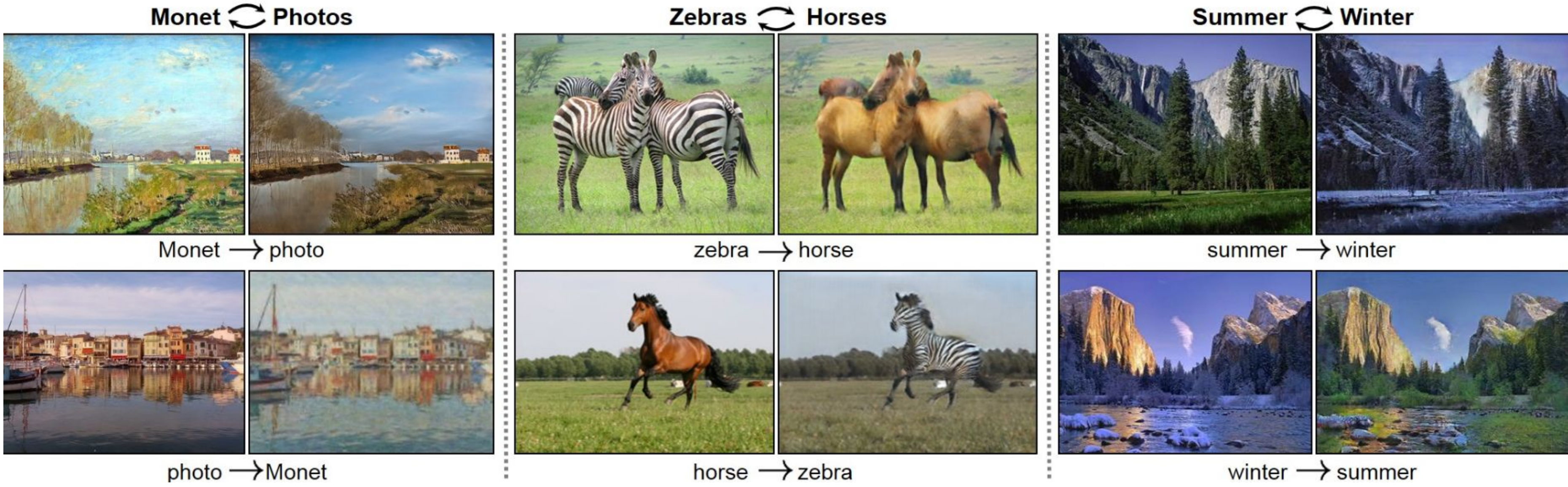
What do intelligent systems need?



http://math.univ-lyon1.fr/homes-www/masnou/fichiers/publications/survey.pdf
http://www.dtic.upf.edu/~mbertalmio/bertalmi.pdf

Translate between data modalities

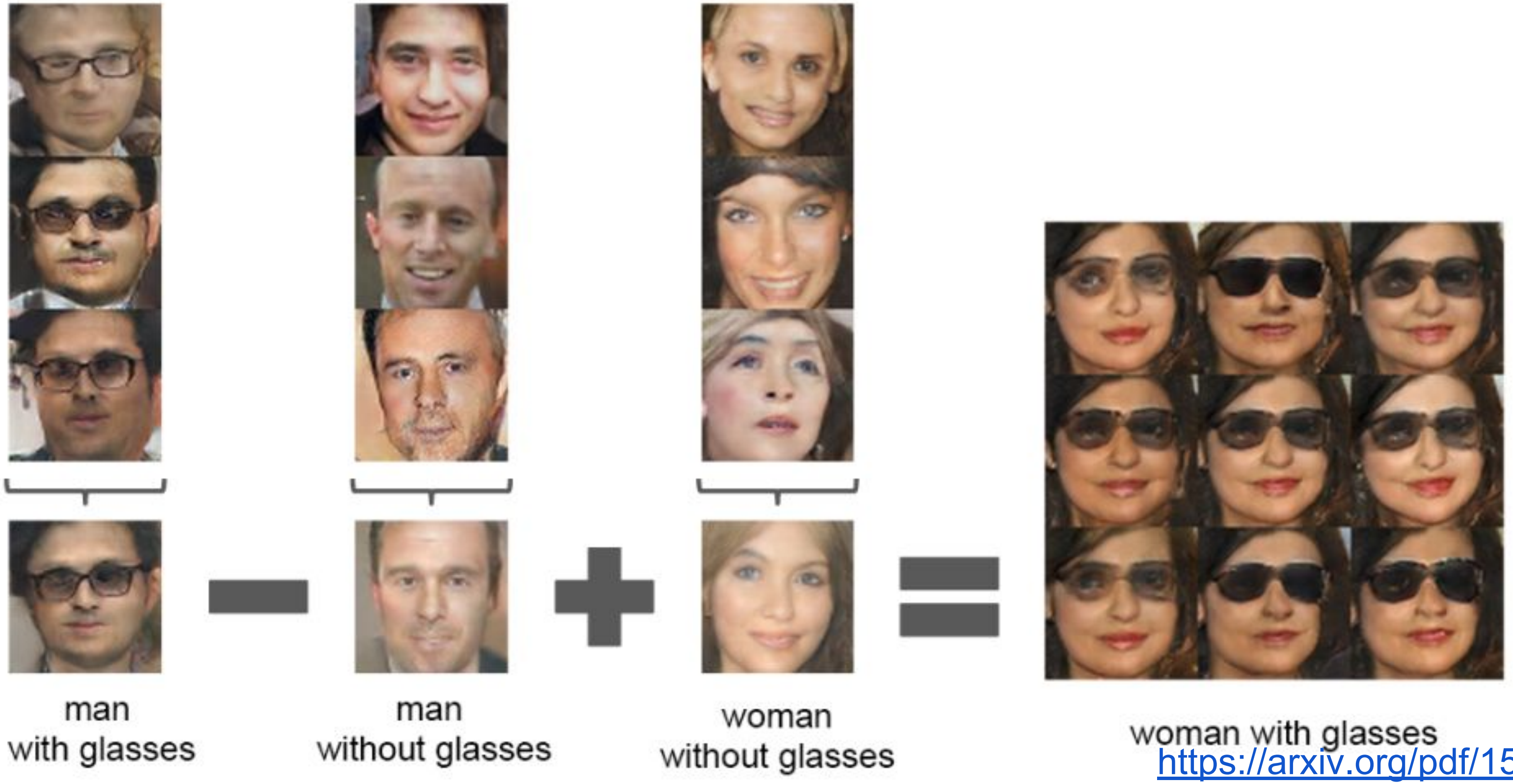
What do intelligent systems need?



CycleGAN: https://arxiv.org/abs/1703.10593

Learn useful representations

What do intelligent systems need?



https://arxiv.org/pdf/1511.06434.pdf

Learn useful representations

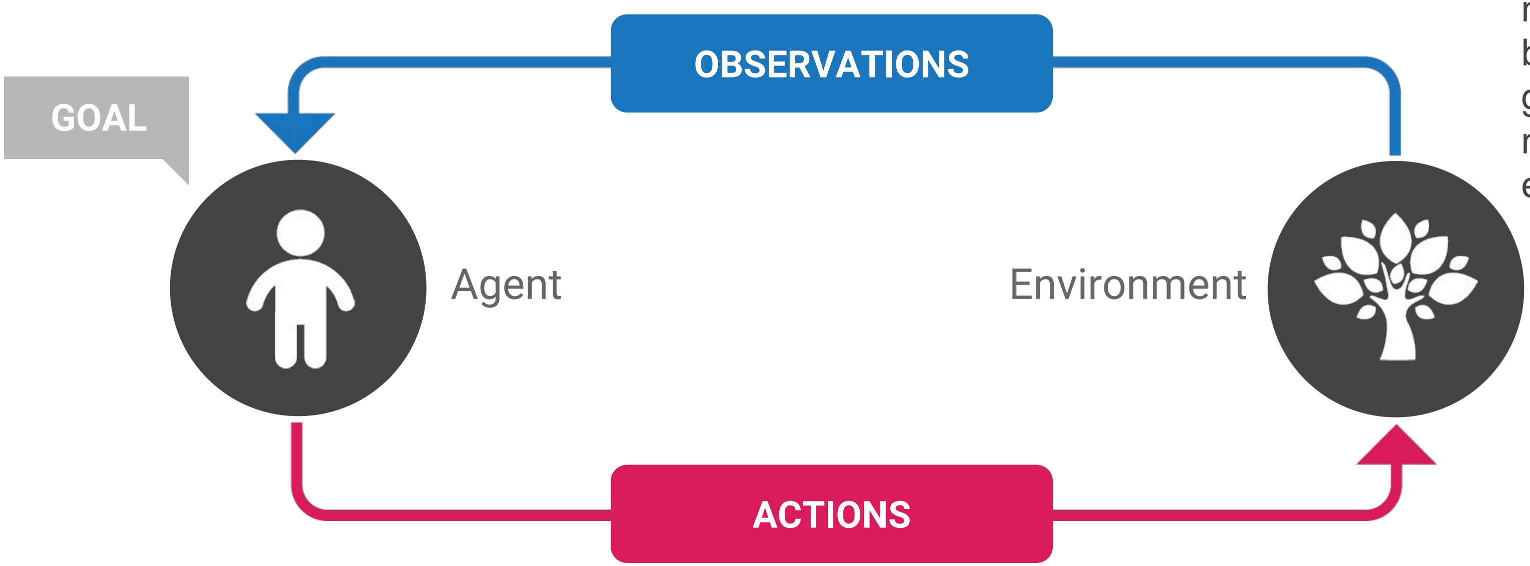
What do intelligent systems need?

Leverage learned structure for better classification performance on labelled data.

DeepMind

Have a model of the world

What do intelligent systems need?



Generative models can be used to generate a model of the environment.

How do generative models allow us to build intelligent systems?

The goal of generative models

Learn a model of the true underlying data distribution $p^*(x)$ from samples

$$x_1, x_2 \dots x_n$$

Generative models learning principles

The goal of generative models

Find p_θ to minimize the distance between p_θ and p^*

Finding p_θ

Choices in generative models

- Model of p_θ
    - you can leverage prior knowledge of the problem
        - what kind of data do you have?
        - what kind of process generated the data?

- The learning principle used to minimize the distance between p_θ and p^*

Generative model algorithm = learning principle + model

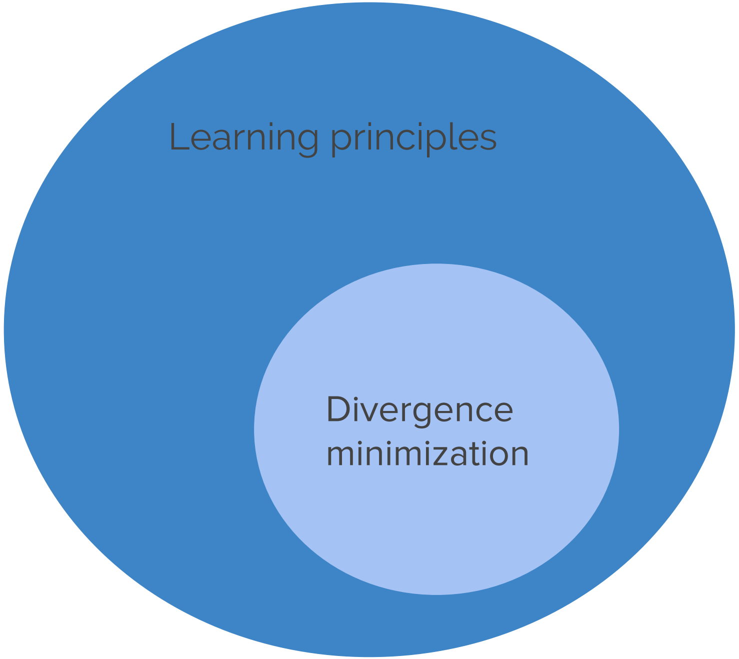
Learning principle

	Moment matching	Maximum likelihood	…	Optimal transport
Autoregressive				
Implicit		Algorithm		
…				
Encoder-decoder				

Model of p_θ

Learning principles

Divergence minimization as a learning principle



Other learning principles:
moment matching
optimal transport

Divergences minimized

Divergence minimization for generative model learning

Aim: Minimize a divergence between p_θ and p^*

Divergences minimized

Common divergence choices

- KL divergence (most common)
    - results in *maximum likelihood* learning
- Reverse KL divergence
- Jensen Shannon

KL divergence

Maximum likelihood

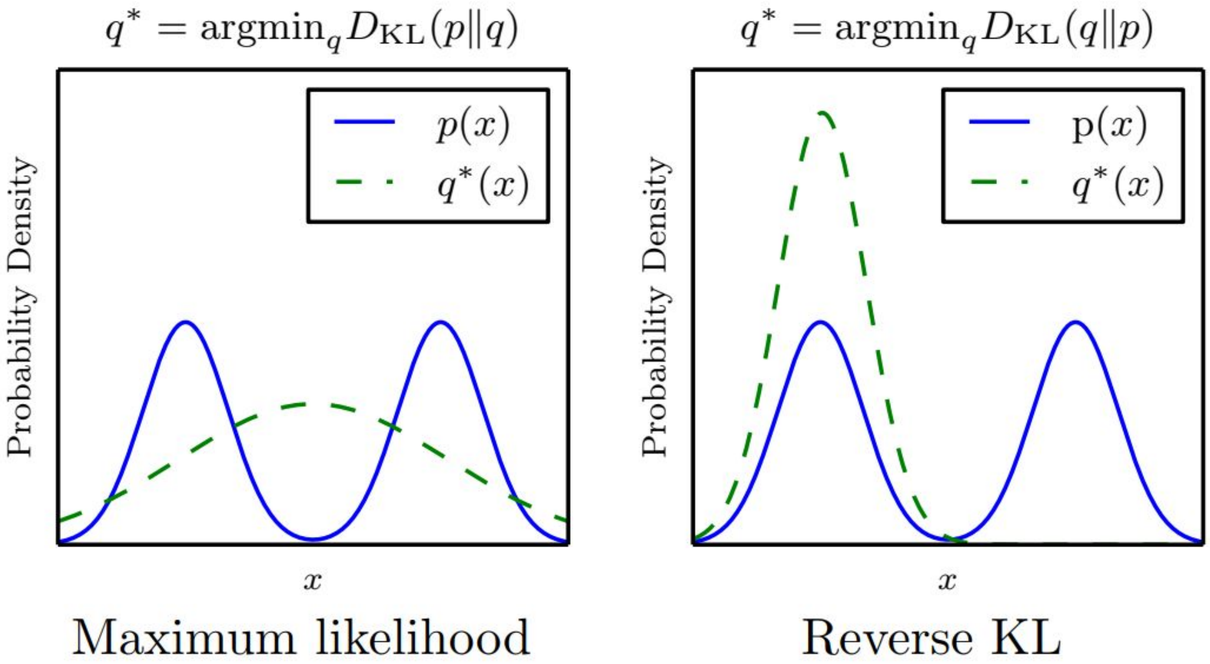
Minimizing the KL divergence between p_θ and p^* ⇒ maximum likelihood learning:

$$\text{argmax}_\theta \, \mathcal{E}_{x \sim p^*} \log p_\theta(x)$$

Intuition: find the model which gives highest likelihood to the data.

Trade-offs for a fixed model choice

KL vs Reverse KL model fit



https://arxiv.org/pdf/1701.00160.pdf

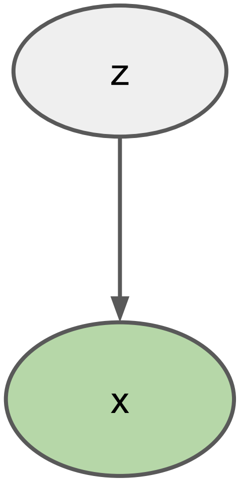
Model choices

Model choices

Graphical structure

Directly model $p_\theta(x)$

Leverage underlying data structure in generative process.



z = latents
x = observed

Model choices

Embedding priors in your model

Leverage data knowledge:

- convolutional models for images
- recurrent models for text and sound
- appropriate priors for latent variables

Algorithms

PixelCNN, PixelRNN, WaveNet

Autoregressive maximum likelihood models

Principle: maximum likelihood

Model: Autoregressive model with no latent variables

$$p_\theta(x) = \prod_i p_\theta(x_i | x_1, x_2 \ldots x_{i-1})$$

Model $p_\theta(x_i | x_1, x_2 \ldots x_{i-1})$ using a recurrent neural network or convolutional masking.

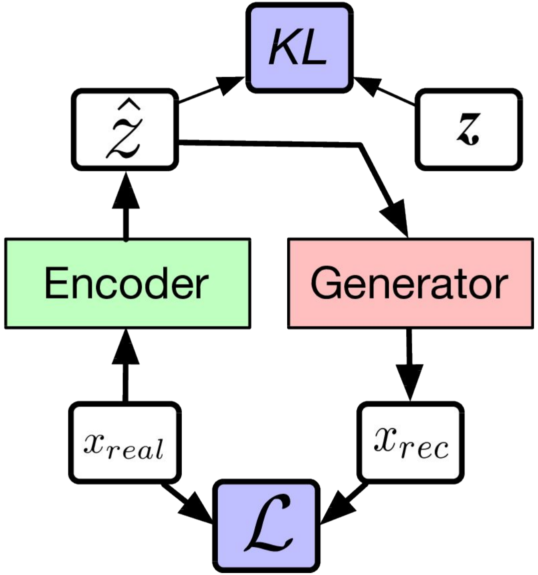
https://arxiv.org/pdf/1601.06759.pdf
https://arxiv.org/abs/1606.05328
https://deepmind.com/blog/wavenet-generative-model-raw-audio/

Variational autoencoders

Latent approximate maximum likelihood models

Principle: (approximate) maximum likelihood

Model: Encoder-decoder model with latent variables



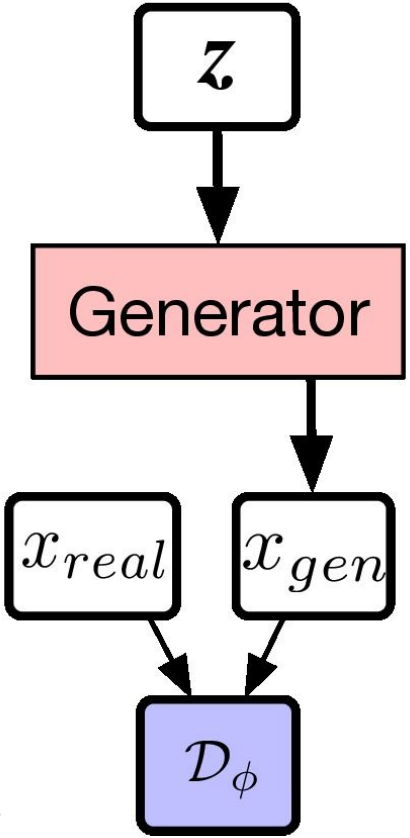
$$\log p_{\boldsymbol{\theta}}(\mathbf{x}) = \log \int p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z} \geq \mathbb{E}_{q_\eta(\mathbf{z}|\mathbf{x})}[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})] - \mathrm{KL}[q_\eta(\mathbf{z}|\mathbf{x})||p(\mathbf{z})]$$

https://arxiv.org/abs/1312.6114

Generative adversarial networks

Latent variable models without maximum likelihood

- Minimize distance between p_θ and p^*
  - Jensen Shannon
  - Earth movers
- How they model p_θ
  - Implicitly: you do not have access to p_θ, but you can sample from it
  - Latent models

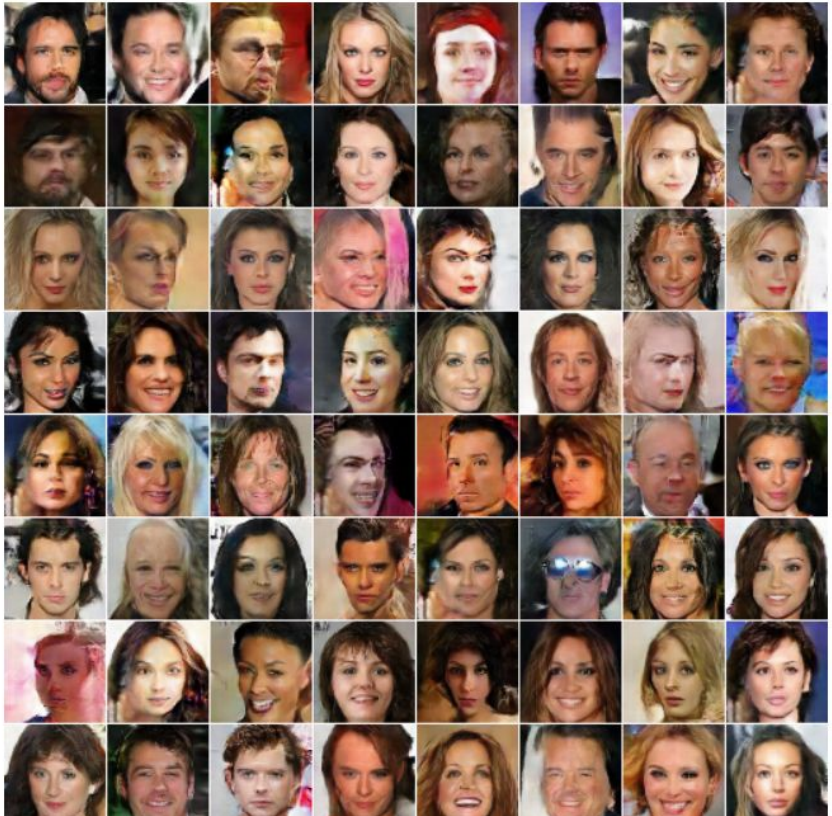


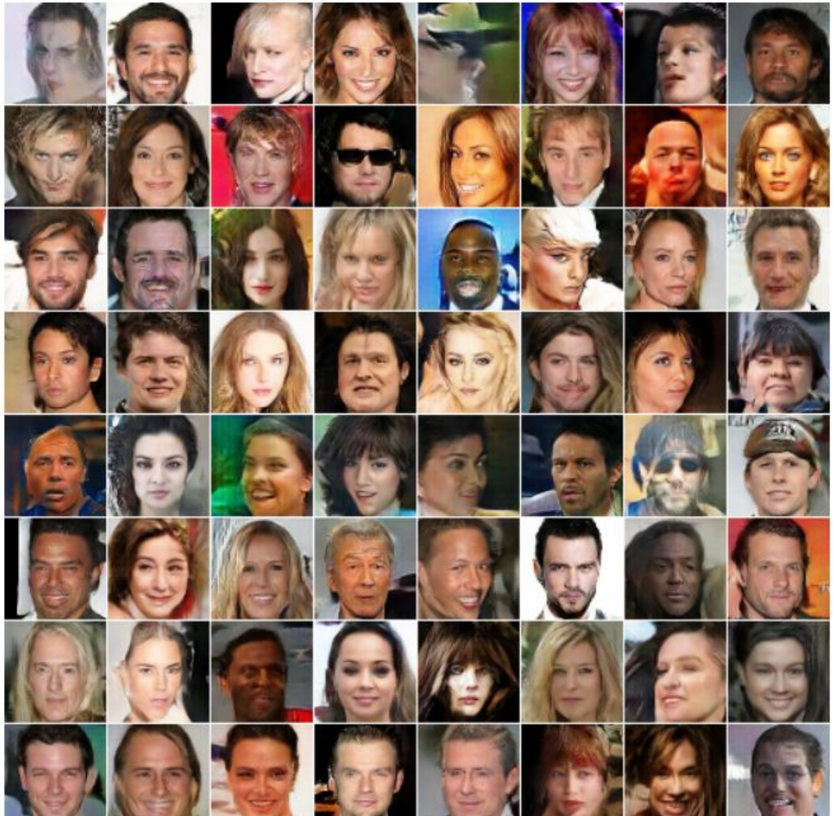
https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

Evaluating generative models

Evaluation of generative models

How can we compare generative models?





Evaluation of generative models

Use application specific metrics

No evaluation metric is able to capture all desired properties.

Evaluate performance based on the end goal:

- semi supervised learning: classification accuracy
- reinforcement learning: total agent reward
- data generation (eg: text to speech): human (user) evaluation via beta testing

When to think about generative models

Using generative models

Use them when...

- learning with scarce labelled data
- estimating uncertainty
- eliminating outliers
- completing missing data
- generating data
- building useful (disentangled) representations

Using generative models

What model to use?

- need to learn underlying structure?

  $\Rightarrow$ use an inference model

- need only samples?

  $\Rightarrow$ consider GANs

- need a likelihood p(x)

  $\Rightarrow$ use maximum likelihood models

  $\Rightarrow$ use autoregressive models if sampling cost is not an issue

THANK YOU

Credits

Shakir Mohamed, Balaji Lakshminarayanan

Additional Credits