8. Other Deep Architectures

CS 519 Deep Learning, Winter 2018

Fuxin Li

With materials from Zsolt Kira and Ian Goodfellow

A brief overview of other architectures

- Unsupervised Architectures
  - Deep Belief Networks
  - Autoencoders
  - GANs
- Temporal Architectures
  - Recurrent Neural Networks (RNN)
  - LSTM
- We will carefully cover those items later
  - Right now just a brief overview in case that you might be tempted to use them in your project

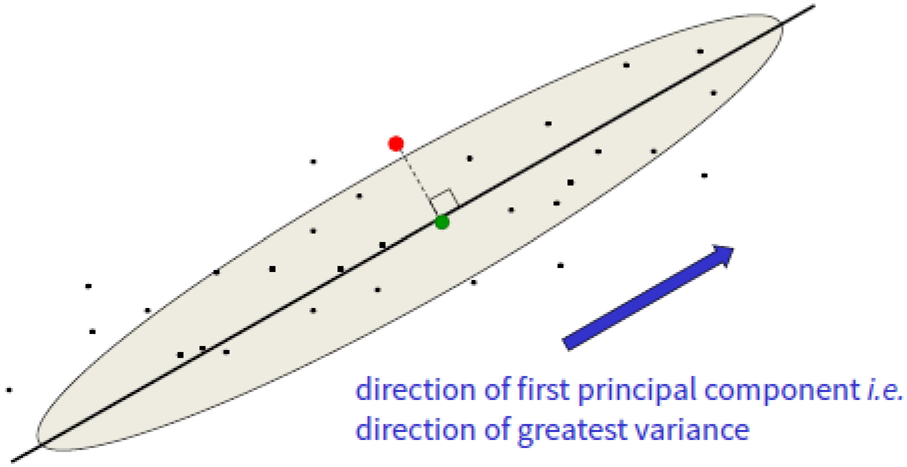
Unsupervised Deep Learning

- CNN is most successful with a lot of training examples

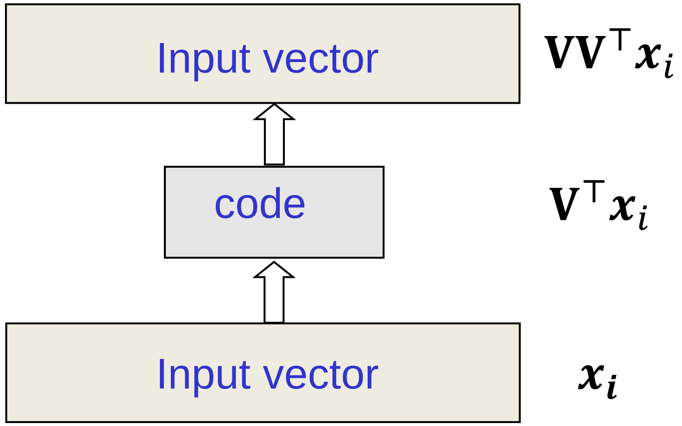
- What can we do if we do not have any training example?
    - Or have very few of them?

Remember PCA: Characteristics and Limitations

- PCA works well when the data is near a linear manifold in high-dimensional space
- Project the data onto this subspace spanned by principal components
- In dimensions orthogonal to the subspace the data has low variance



PCA as a "neural network"

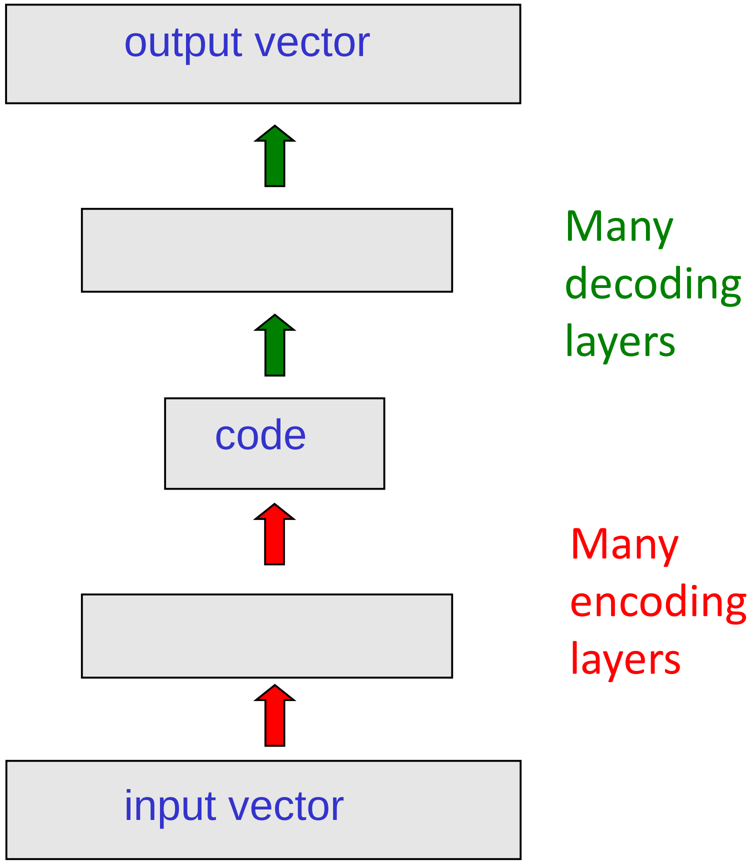


- PCA goal:
  - Minimize reconstruction error

$$\min_{\mathbf{V}} \sum_{i=1}^{n} \left(\boldsymbol{x}_i - \mathbf{V}\mathbf{V}^\top \boldsymbol{x}_i\right)^2$$

Generalize PCA to multi-layer nonlinear network

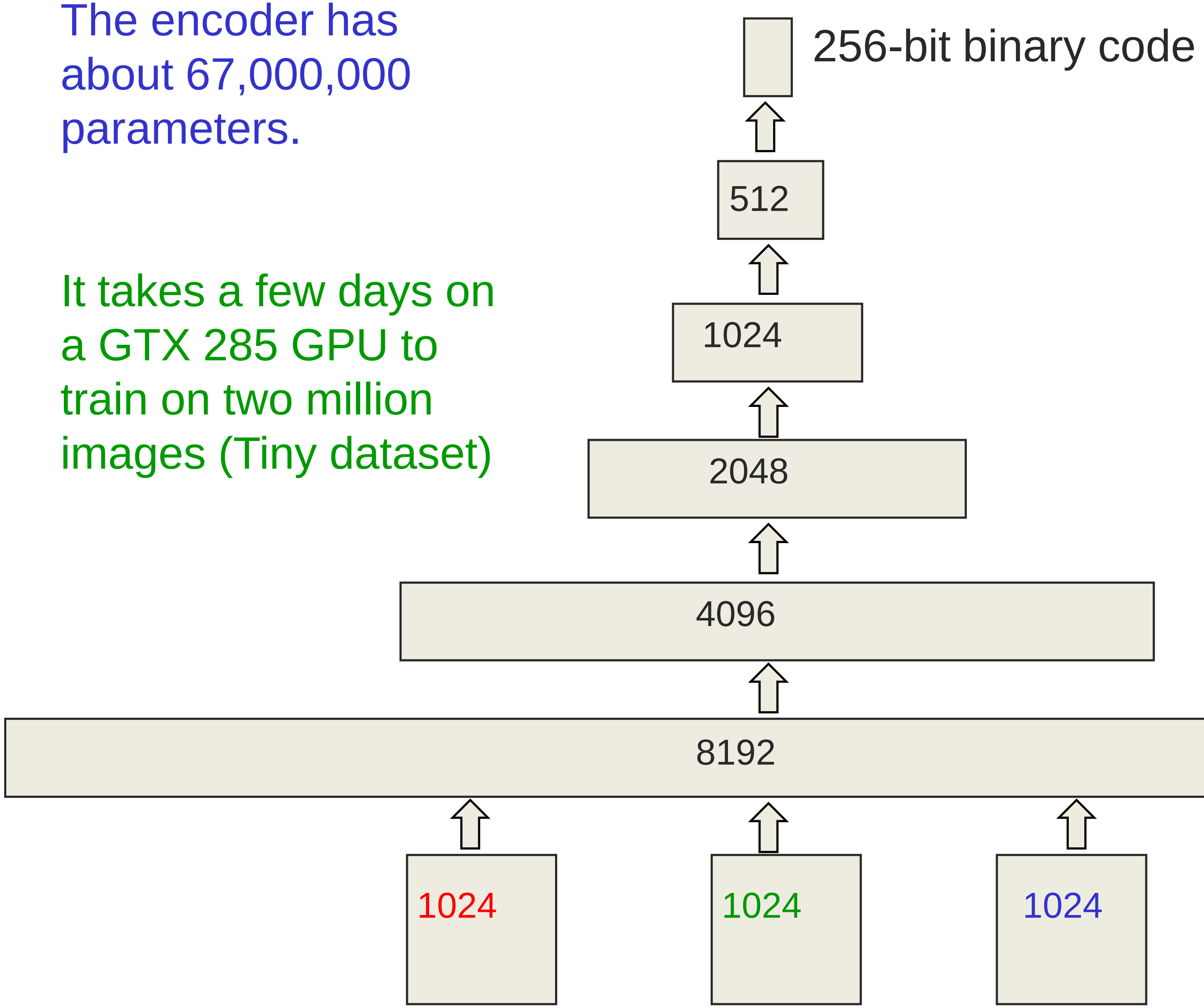
- Deep Autoencoder
  - Same as other NN (linear transform + nonlinearity + linear transform etc.)
  - Only difference is that after decoding, strive to reconstruct the original input
  - Can have convolutional/fully-connected/sparse versions



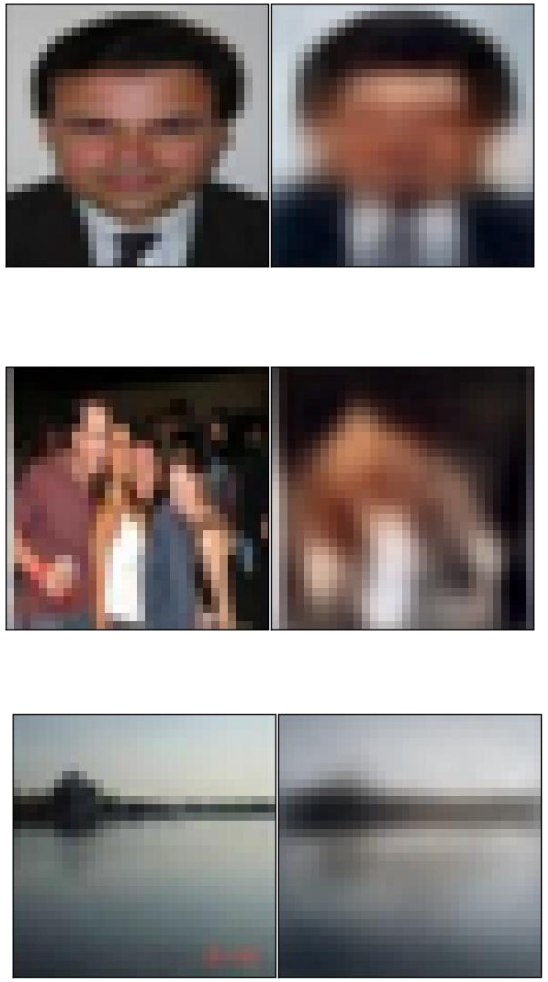
Krizhevsky’s deep autoencoder

The encoder has about 67,000,000 parameters.

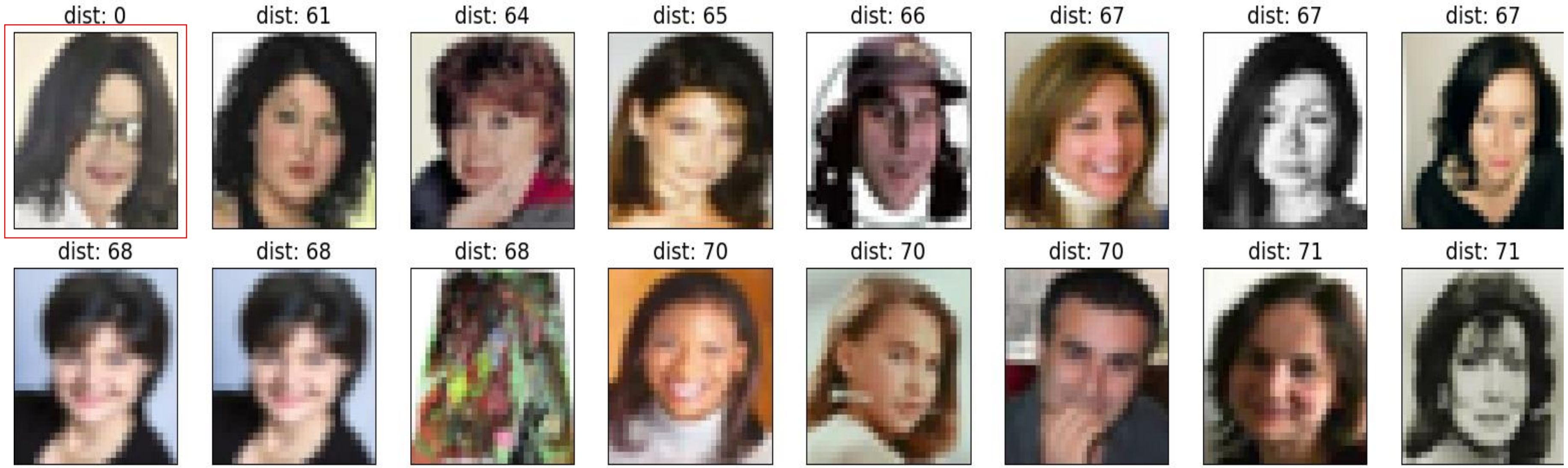
It takes a few days on a GTX 285 GPU to train on two million images (Tiny dataset)



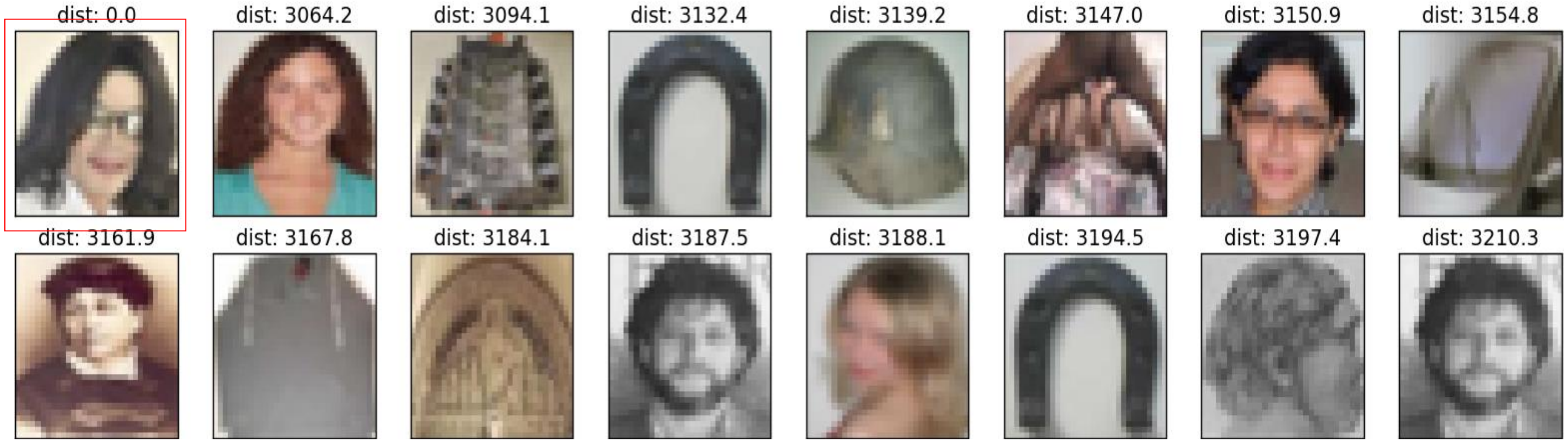
Reconstructions of 32x32 color images from 256-bit codes



retrieved using 256 bit codes



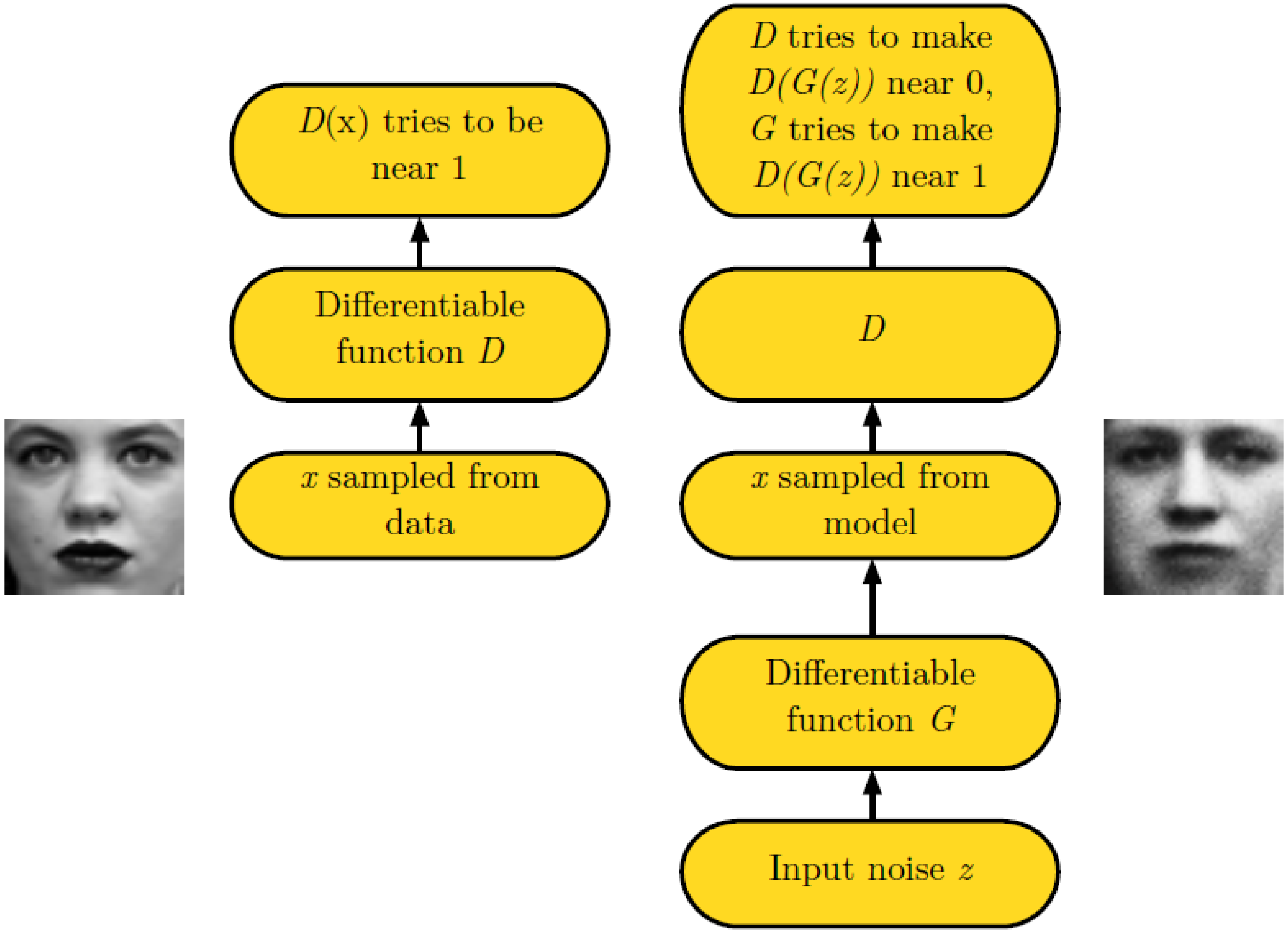
retrieved using Euclidean distance in pixel intensity space



retrieved using 256 bit codes

retrieved using Euclidean distance in pixel intensity space

Generative Adversarial Networks



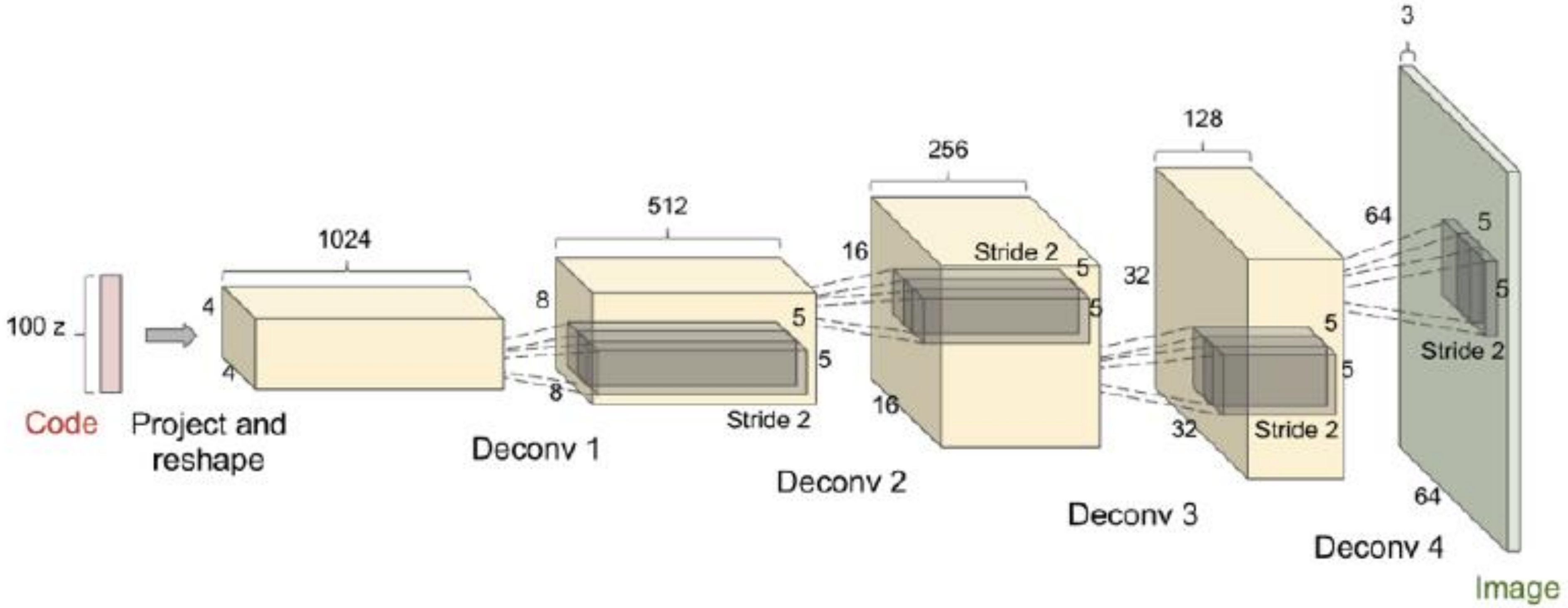
Generative Adversarial Networks

- Cost for the discriminator:
  - Standard cross-entropy loss, with everything from p_{data} label 1, and everything from z label 0

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

- Cost for the generator:
  - Try to generate examples to “fool” the discriminator

DCGAN



100 z
Code
Project and reshape
1024
4
4
512
8
8
5
5
Stride 2
Deconv 1
256
16
16
Stride 2
5
5
Deconv 2
128
32
32
5
5
Stride 2
Deconv 3
3
64
64
5
5
Stride 2
Deconv 4
Image

Samples of DCGAN-generated images

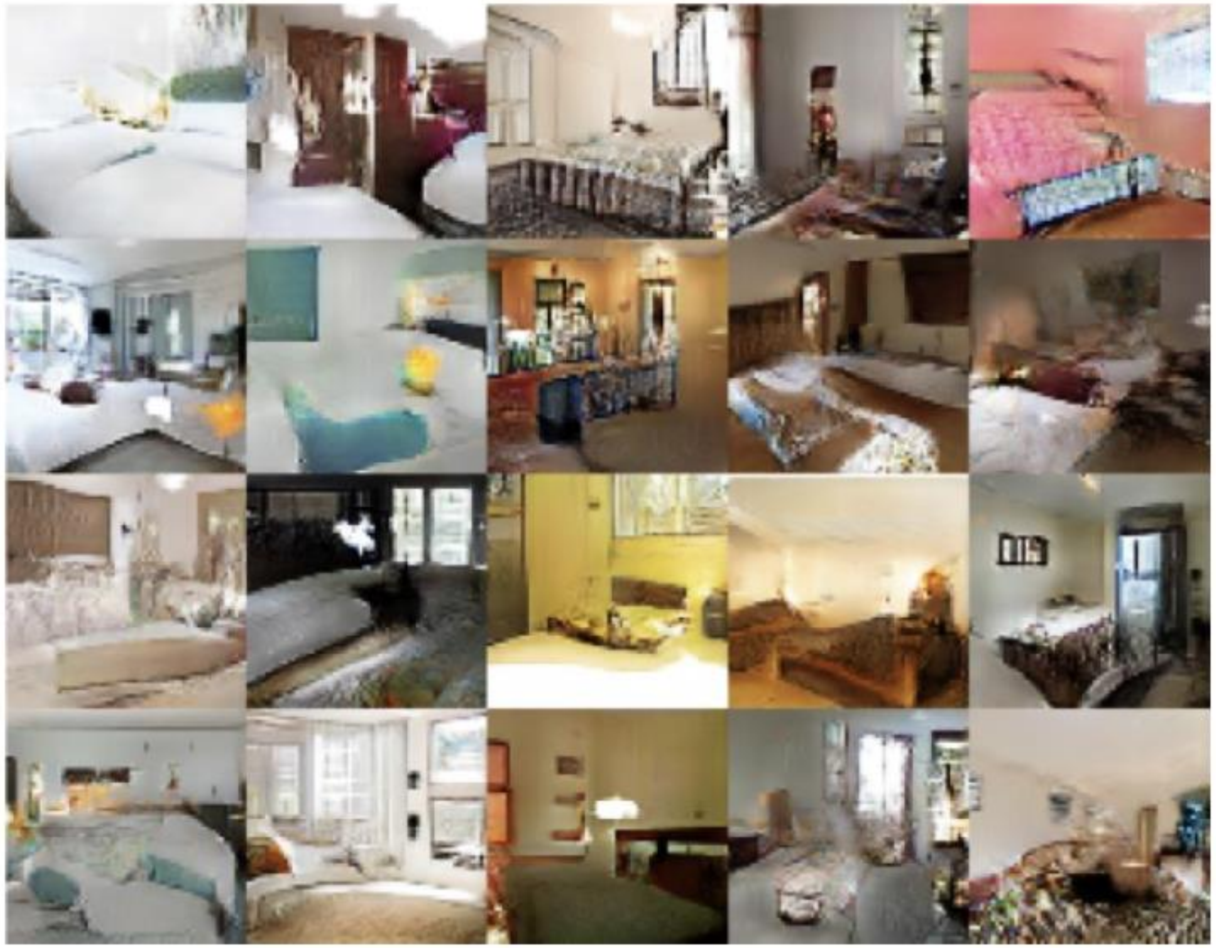
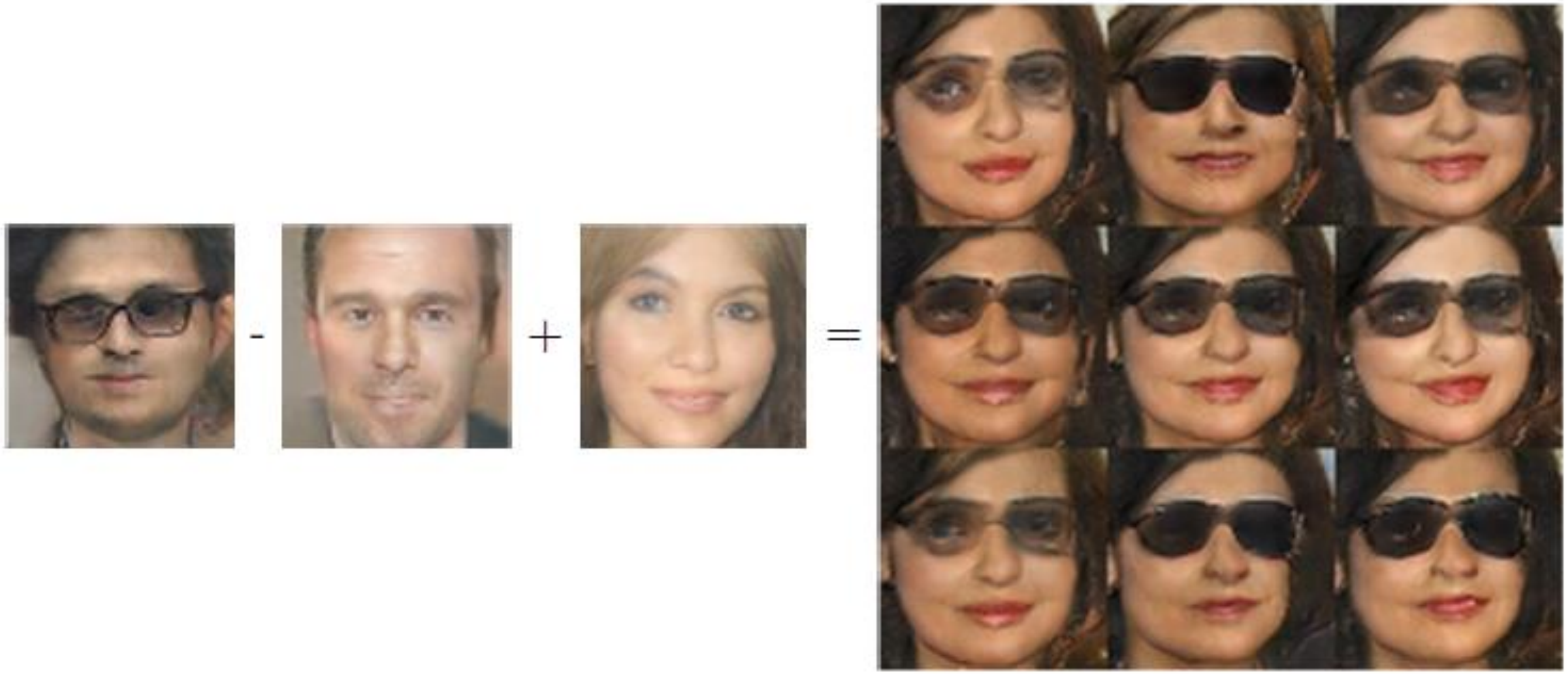


Figure 18: Samples of images of bedrooms generated by a DCGAN trained on the LSUN dataset.

DCGAN representations



Text-to-Image with GANs

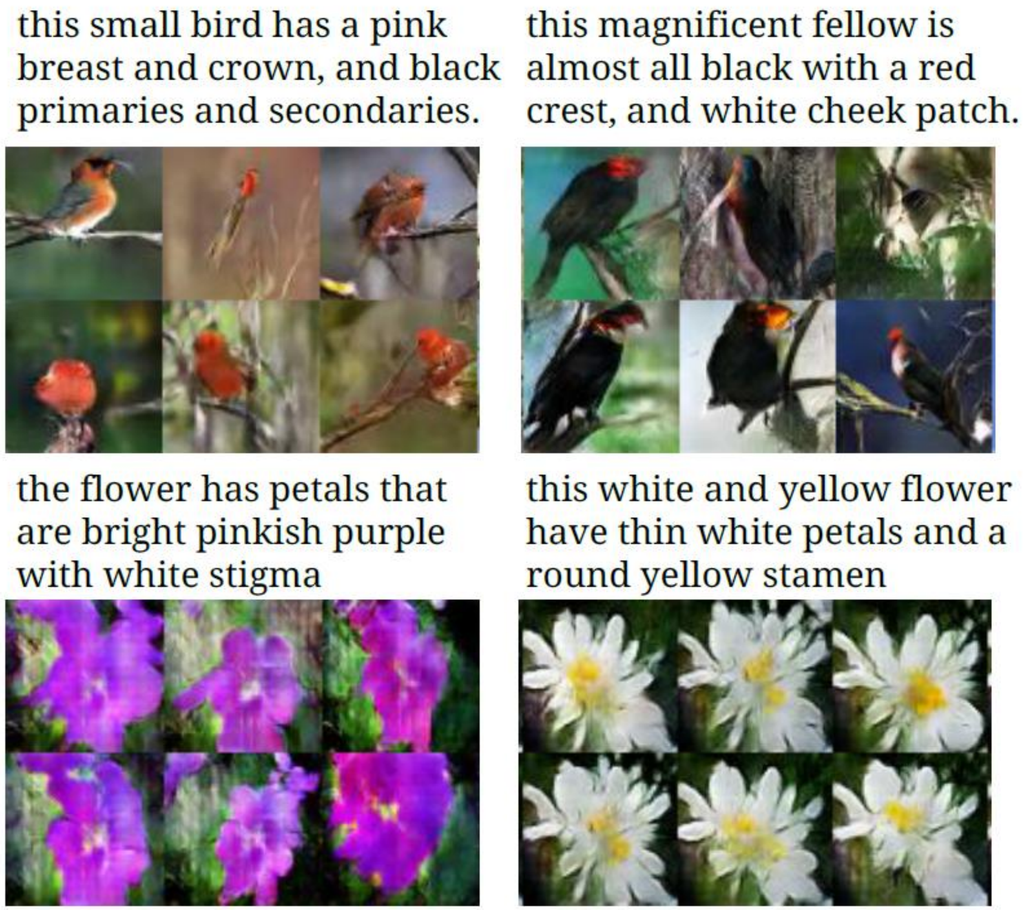
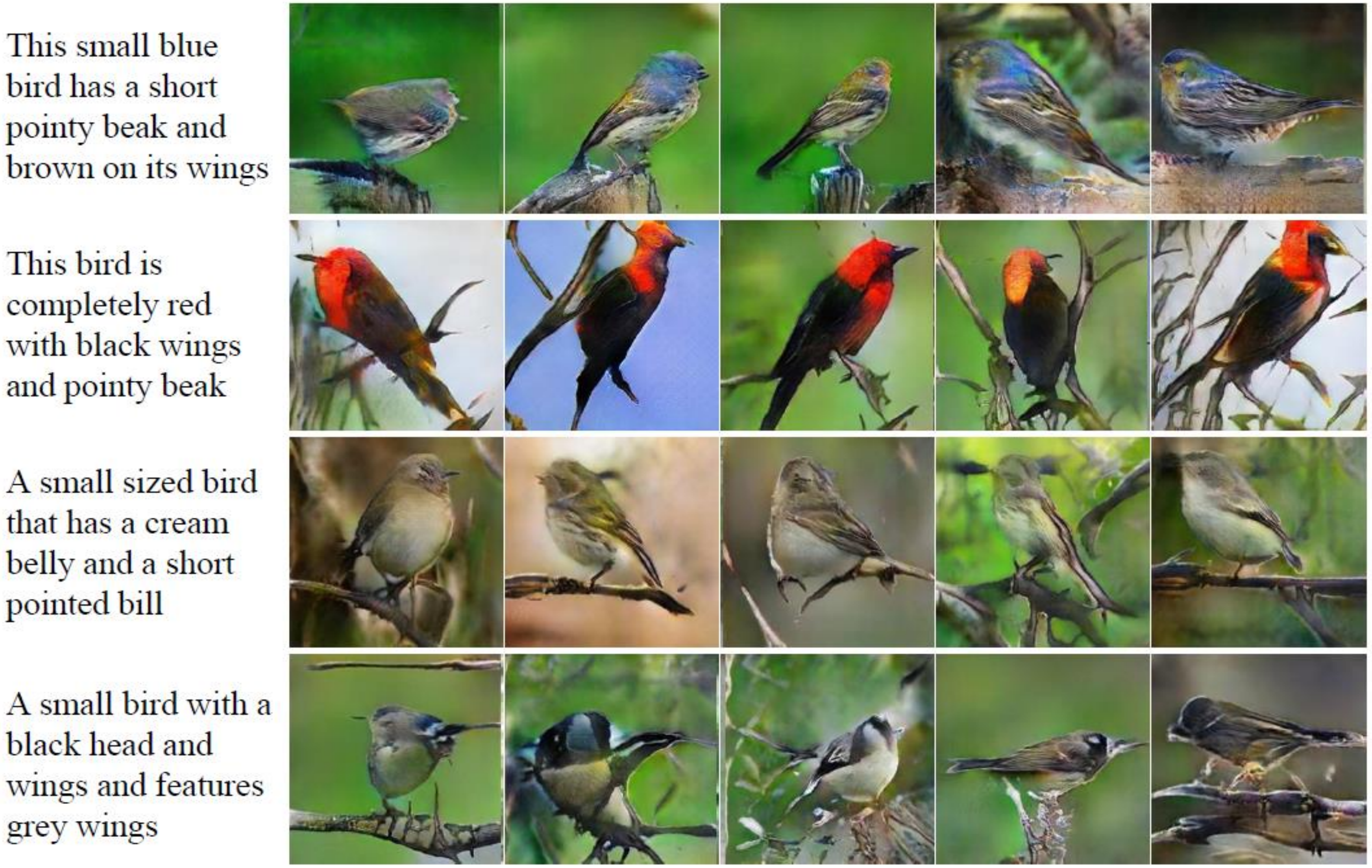


Figure 23: Text-to-image synthesis with GANs. Image reproduced from Reed *et al.* (2016b).

Text-to-Image with GANs



Problems

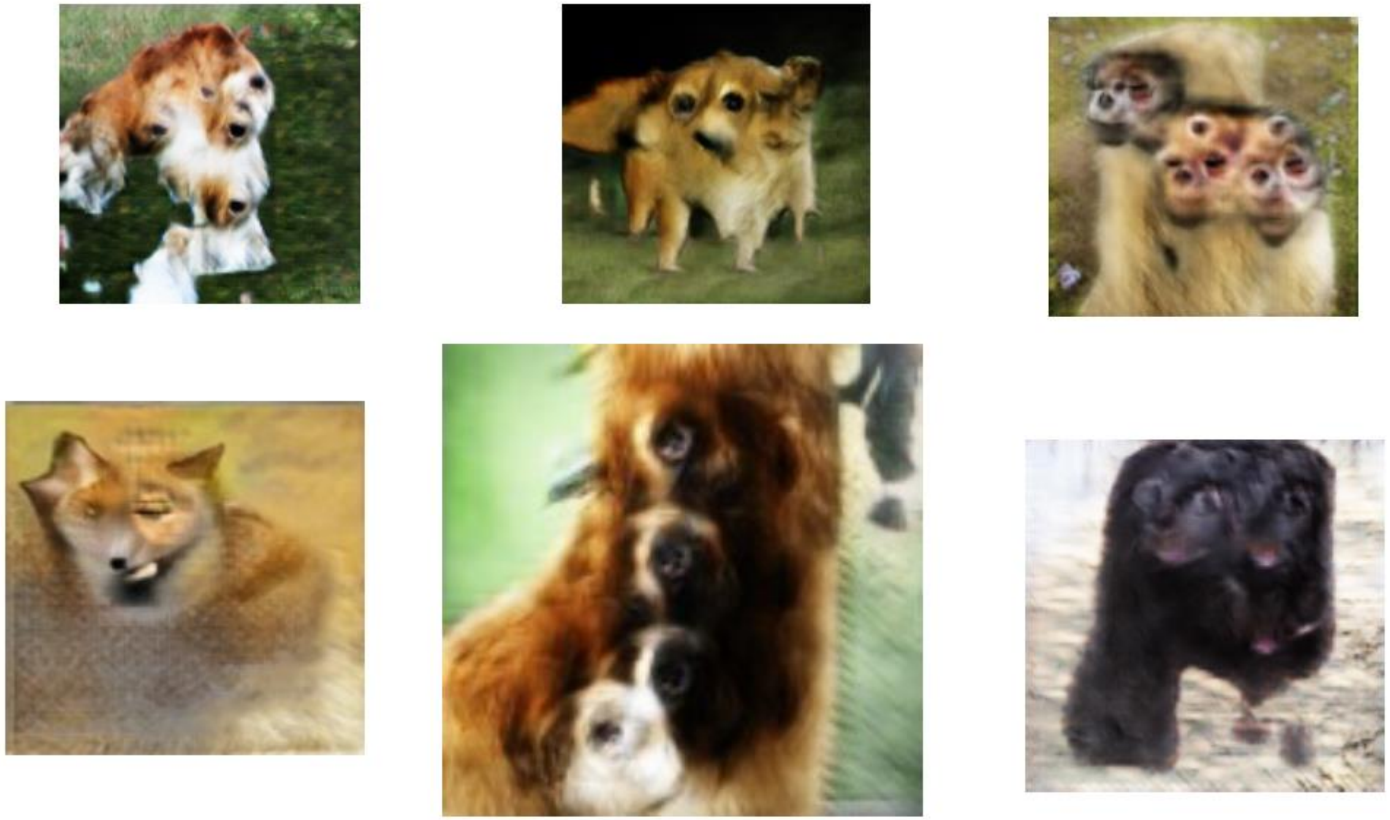


Figure 29: GANs on 128 × 128 ImageNet seem to have trouble with counting, often generating animals with the wrong number of body parts.

Problems

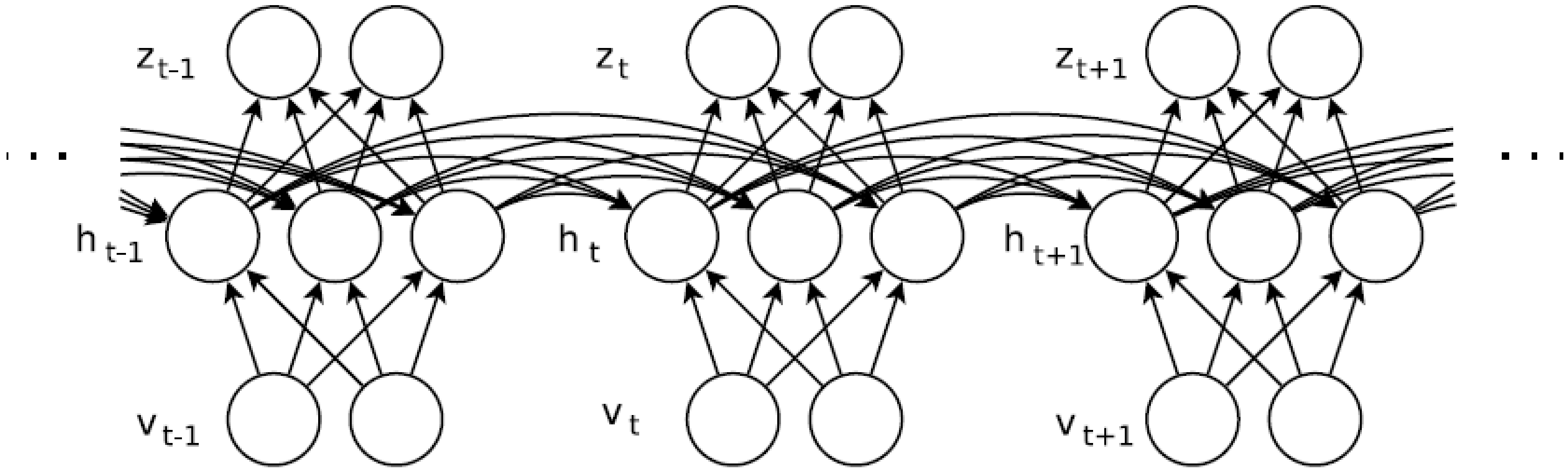
Figure 31: GANs on 128 × 128 ImageNet seem to have trouble coordinating global structure, for example, drawing "Fallout Cow," an animal that has both quadrupedal and bipedal structure.

iGAN

https://www.youtube.com/watch?v=9c4z6YsBGQ0

Recurrent Neural Networks (RNNs)

- Temporal, Sequences
- Tied weights
- Some additional variants: Recursive Autoencoders, Long Short-Term Memory (LSTM)



Machine Translation

- Have to look at the entire sentence (or, many sentences)

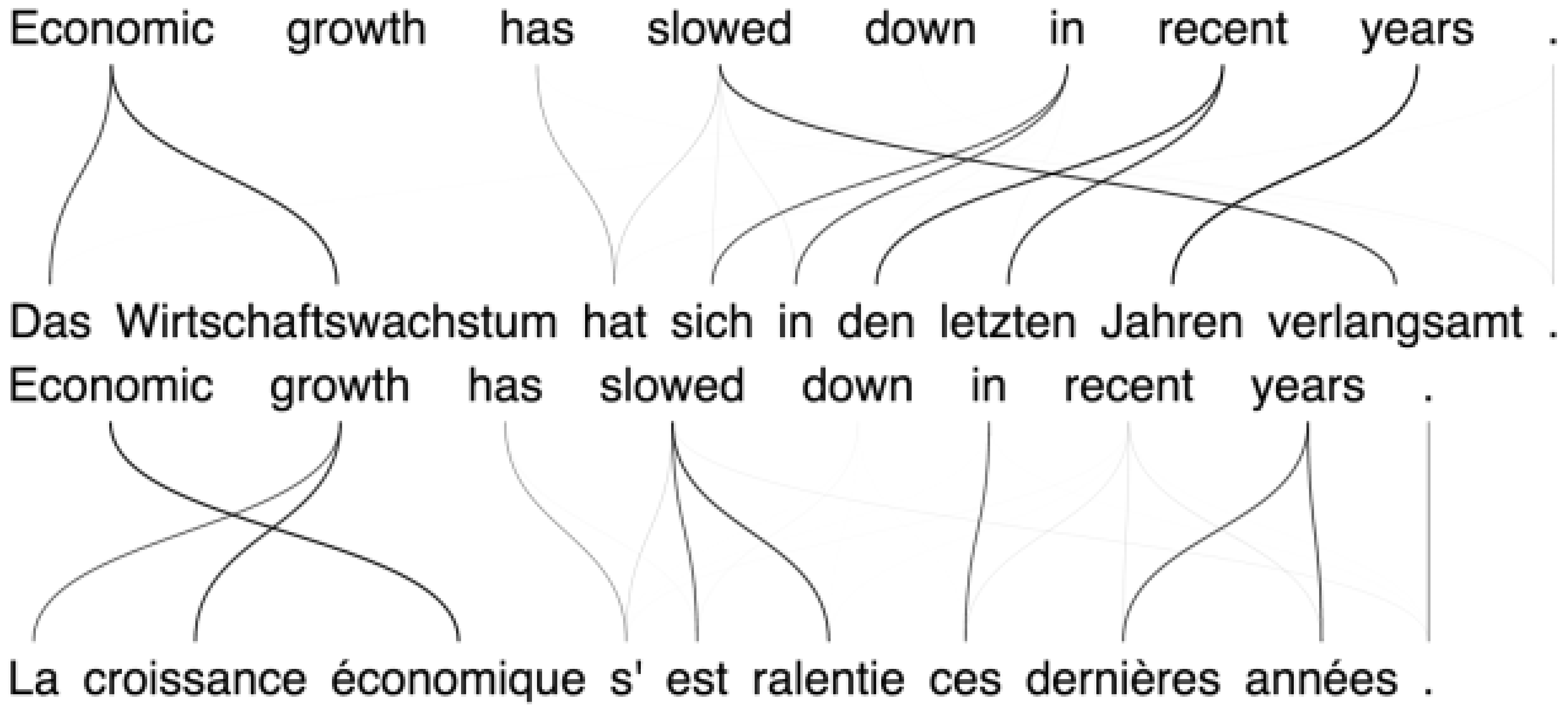


Image Captioning

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

"girl in pink dress is jumping in air."

"black and white dog jumps over bar."

"young girl in pink shirt is swinging on swing."

"man in blue wetsuit is surfing on wave."

Restricted Boltzmann Machines

- Generative version of the encoder
- Binary-valued hidden variables
  - Define probabilities such as $P(h_i|X)$ and $P(x_i|H)$
  - You can generate samples of observed variables from hidden
- Think as an extension of probabilistic PCA
- Only if you are into generative models (PGM class)
- Unsupervised pre-training method to train it (Hinton, Salakhutdinov 2006)
- Convolutional and fully connected version available
- Doesn’t perform very well..

Fooling a deep network (Szegedy et al. 2013)

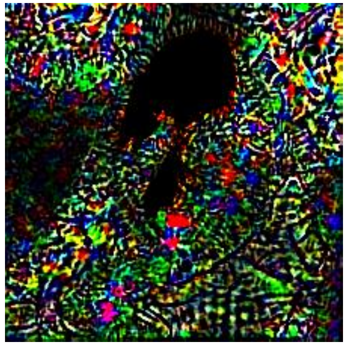
- Optimizing a delta from the image to maximize a class prediction $f_c(x)$

$$\max_{\Delta I} f_c(I + \Delta I) - \lambda||\Delta I||^2$$

I

Giant Panda (99.32% confidence)

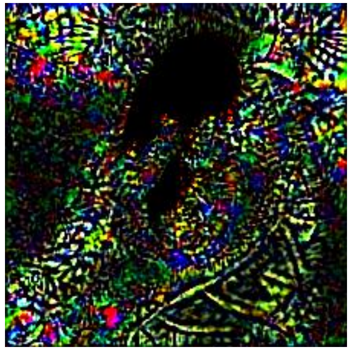
$+0.03$ ΔI



$=$

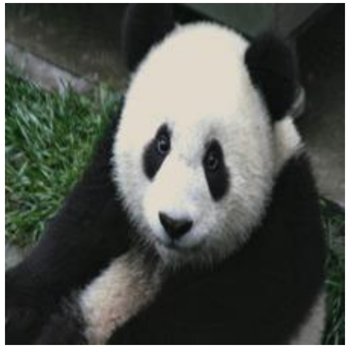
Shark (93.89% confidence)

$+0.03$ ΔI



$=$

Goldfish (95.15% confidence)



(Szegedy et al. 2013, Goodfellow et al. 2014, Nguyen et al. 2015)