Deep Learning

Unsupervised learning and Generative models

Alex Olson

Adapted from material by Charles Ollion & Olivier Grisel

Outline

Unsupervised learning

Outline

Unsupervised learning

Autoencoders

Outline

Unsupervised learning

Autoencoders

Generative Adversarial Networks

Unsupervised learning

Unsupervised learning

Generic goal of unsupervised learning is to **find underlying structrure** in data. Specific goals include:

- clustering: group similar observations together;
- reducing the dimensionality for visualization;
- building a better representation of data for a downstream supervised task;
- learning a likelihood function, e.g. to detect anomalies;
- generating new samples similar to past observations.

Unsupervised learning

For complex data (text, image, sound, ...), there is plenty of hidden latent structure we hope to capture:

- **Image data**: find low dimensional semantic representations, independent sources of variation;
- **Text data**: find fixed size, dense semantic representation of data.

Unsupervised learning

For complex data (text, image, sound, ...), there is plenty of hidden latent structure we hope to capture:

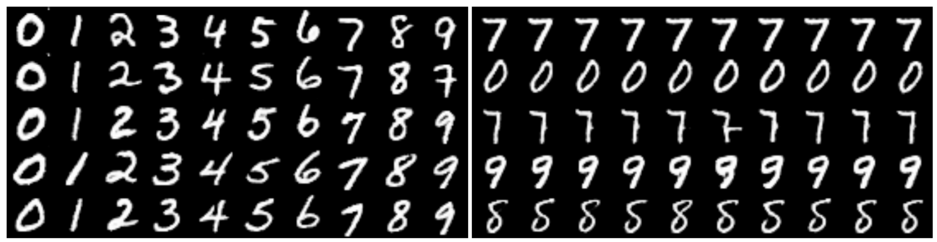
- **Image data**: find low dimensional semantic representations, independent sources of variation;
- **Text data**: find fixed size, dense semantic representation of data.

Latent space might be used to help build more efficient human labeling interfaces.

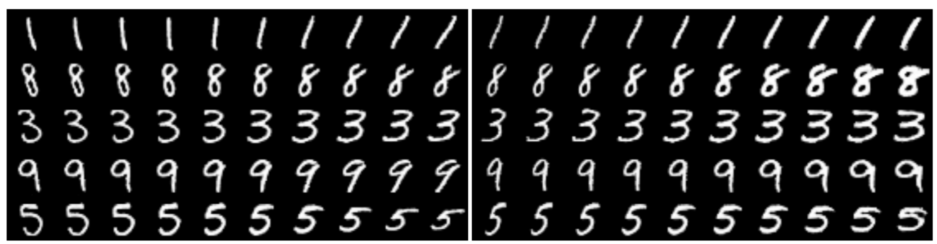
=> Goal: reduce labeling cost via active learning.

Goal of unsupervised learning

A low dimension space which captures all the **variations** of data and **disentangles** the different latent factors underlying the data.



(a) Varying c_1 on InfoGAN (Digit type) (b) Varying c_1 on regular GAN (No clear meaning)



(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation) (d) Varying c_3 from -2 to 2 on InfoGAN (Width)

Chen, Xi, et al. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. NIPS, 2016.

Self-supervised learning

find smart ways to **build supervision** without labels, exploiting domain knowledge and regularities

Self-supervised learning

find smart ways to **build supervision** without labels, exploiting domain knowledge and regularities

Use **text structure** to create supervision

- Word2Vec, BERT or GPT-1,2,3 (soon 4) language models

Self-supervised learning

find smart ways to **build supervision** without labels, exploiting domain knowledge and regularities

Use **text structure** to create supervision

- Word2Vec, BERT or GPT-1,2,3 (soon 4) language models

Can we do the same for other domains?

- **Image**: exploit spatial context of an object
- **Sound, video**: exploit temporal context

Self-supervised learning

find smart ways to **build supervision** without labels, exploiting domain knowledge and regularities

Use **text structure** to create supervision

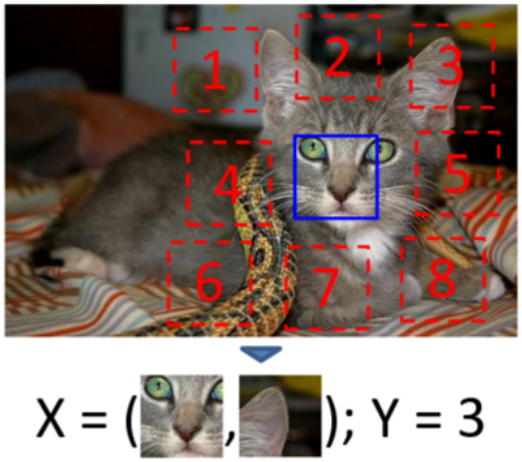
- Word2Vec, BERT or GPT-1,2,3 (soon 4) language models

Can we do the same for other domains?

- **Image**: exploit spatial context of an object
- **Sound, video**: exploit temporal context

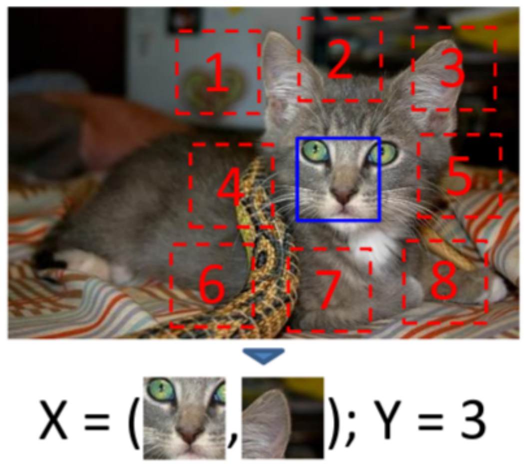
No direct **accuracy** measure: usually tested through a downstream task

Self-supervised learning



Doersch, Carl, Abhinav Gupta, and Alexei A. Efros. "Unsupervised visual representation learning by context prediction." ICCV 2015.

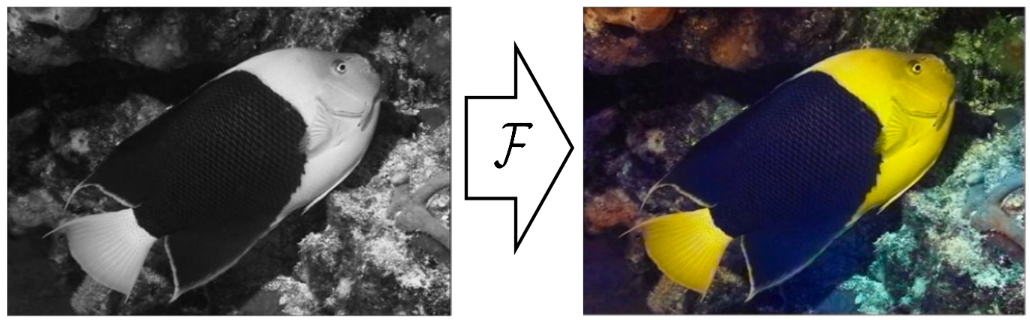
Self-supervised learning



- Predict patches arrangement in images: 8 class classifier
- Siamese architecture for the two patches + concat

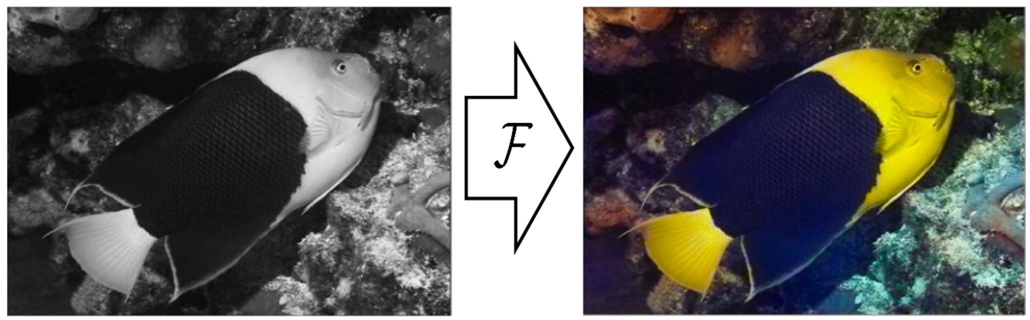
Doersch, Carl, Abhinav Gupta, and Alexei A. Efros. "Unsupervised visual representation learning by context prediction." ICCV 2015.

Self-supervised learning



Zhang et al. "Colorful Image Colorization" ECCV 2016

Self-supervised learning



- Given RGB images, generate their grayscale version
- Train a network to predict pixels color given grayscale image

Zhang et al. "Colorful Image Colorization" ECCV 2016

Self-supervised learning

Dosovitskiy et al. "Exemplar Networks" 2014

Self-supervised learning

- Heavy augmentation of the images
- Network must predict that augmented images are similar, and another random image dissimilar

Dosovitskiy et al. "Exemplar Networks" 2014

Self-supervised learning

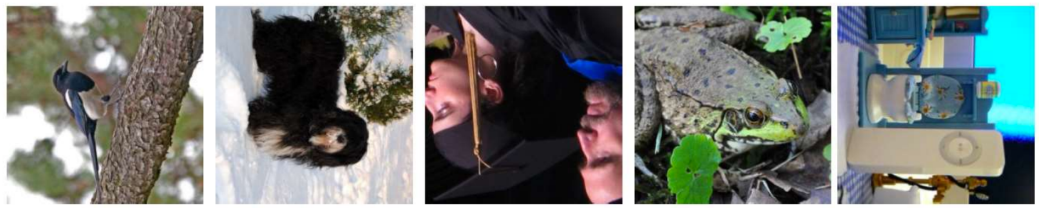


Figure 1: Images rotated by random multiples of 90 degrees (e.g., 0, 90, 180, or 270 degrees). The core intuition of our self-supervised feature learning approach is that if someone is not aware of the concepts of the objects depicted in the images, he cannot recognize the rotation that was applied to them.

Spyros Gidaris, Praveer Singh, Nikos Komodakis. "Unsupervised representation learning by predicting image rotations, " ICLR 2018

Self-supervised learning

90° rotation | 270° rotation | 180° rotation | 0° rotation | 270° rotation

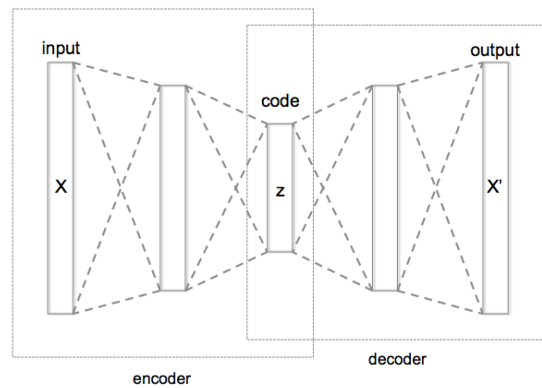
Figure 1: Images rotated by random multiples of 90 degrees (e.g., 0, 90, 180, or 270 degrees). The core intuition of our self-supervised feature learning approach is that if someone is not aware of the concepts of the objects depicted in the images, he cannot recognize the rotation that was applied to them.

- Generate 4 versions of the image, rotated by 0°, 90°, 180°, and 270°
- Network must predict the angle

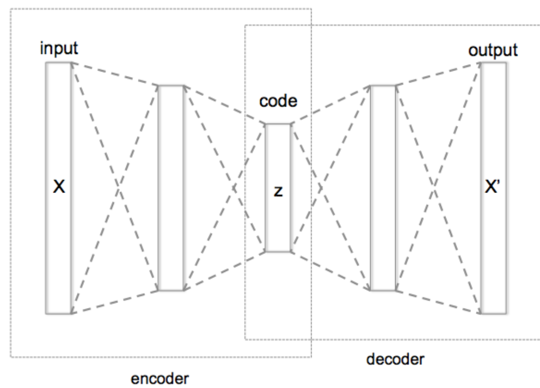
Spyros Gidaris, Praveer Singh, Nikos Komodakis. "Unsupervised representation learning by predicting image rotations, " ICLR 2018

Autoencoders

Autoencoder



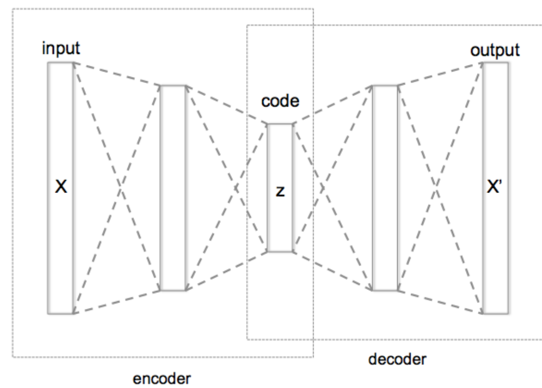
Autoencoder



Supervision : reconstruction loss of the input, usually:

$$l(x, f(x)) = ||f(x) - x||_2^2$$

Autoencoder

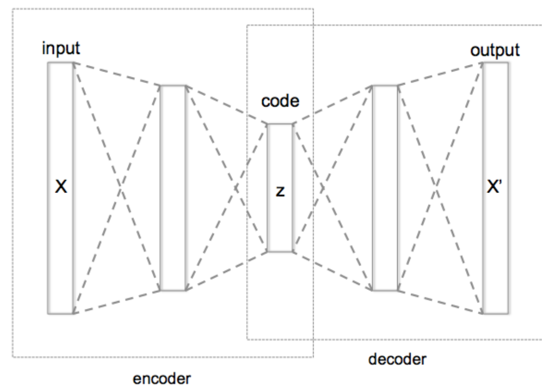


Supervision : reconstruction loss of the input, usually:

$$l(x, f(x)) = ||f(x) - x||_2^2$$

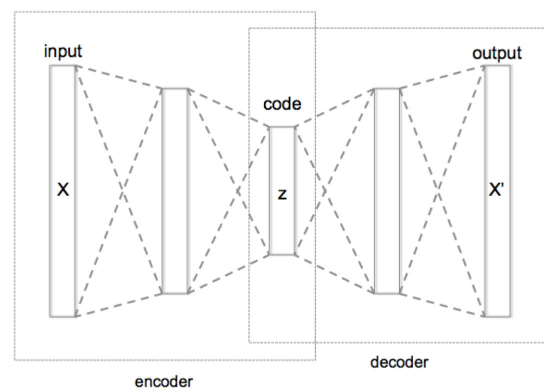
Binary crossentropy is also used

Autoencoder



Keeping the **latent code z** low-dimensional forces the network to learn a "smart" compression of the data, not just an identity function

Autoencoder



Keeping the **latent code z** low-dimensional forces the network to learn a "smart" compression of the data, not just an identity function

Encoder and decoder can have arbitrary architecture (CNNs, RNNs...)

Sparse/Denoising Autoencoder

Adding a sparsity constraint on activations:

$$||encoder(x)||_1 \sim \rho, \rho = 0.05$$

Learns sparse features, easily interpretable

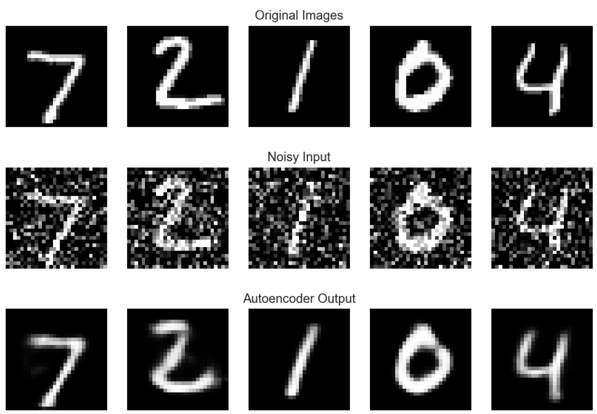
Sparse/Denoising Autoencoder

Adding a sparsity constraint on activations:

$$||encoder(x)||_1 \sim \rho, \rho = 0.05$$

Learns sparse features, easily interpretable

Denoising Autoencoder: train features for robustness to noise.



Uses and limitations

After **pre-training** use the latent code $\mathbf{z}$ as input to a classifier instead of $\mathbf{x}$

Semi-supervised learning simultaneous learning of the latent code (on a large, unlabeled dataset) and the classifier (on a smaller, labeled dataset)

Uses and limitations

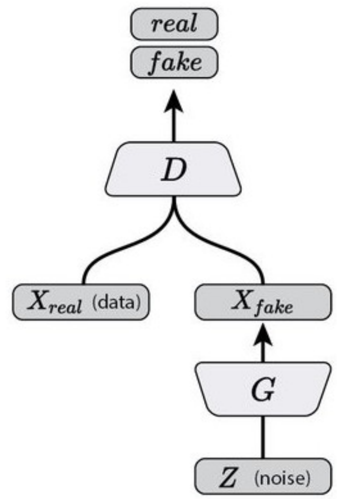
After **pre-training** use the latent code $\mathbf{z}$ as input to a classifier instead of $\mathbf{x}$

Semi-supervised learning simultaneous learning of the latent code (on a large, unlabeled dataset) and the classifier (on a smaller, labeled dataset)

Other use: Use decoder $D(x)$ as a **Generative model**: generate samples from random noise

Generative Adversarial Networks

Generative Adversarial Networks



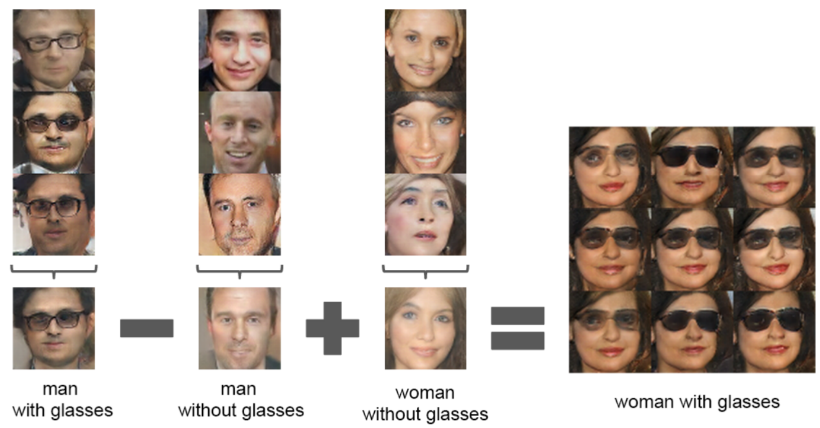
Alternate training of a **generative network** G and a **discrimininative network** D

Goodfellow, Ian, et al. Generative adversarial nets. NIPS 2014.

GANs

- D tries to find out which example are generated or real
- G tries to fool D into thinking its generated examples are real

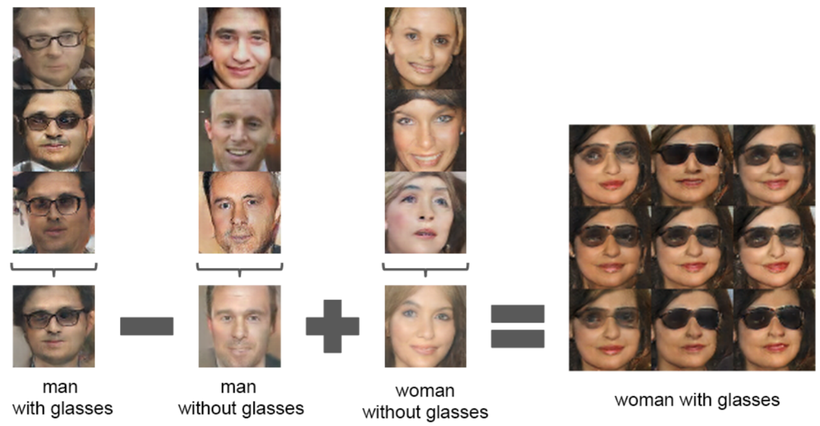
DC-GAN



- Generator generates high quality images

Radford, Alec, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. 2015.

DC-GAN



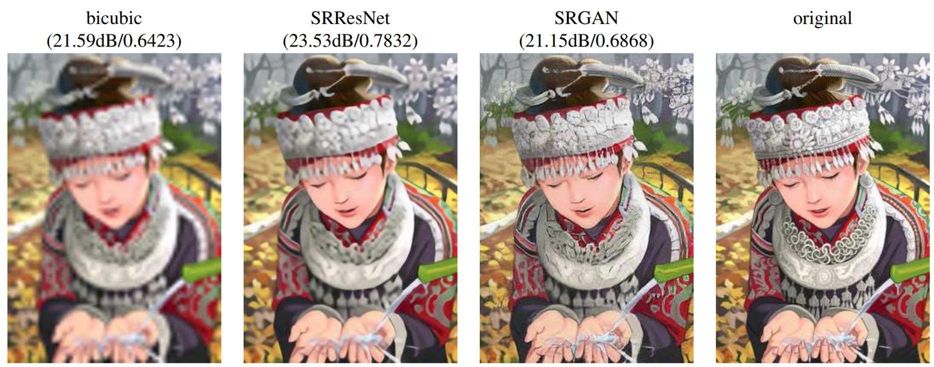
- Generator generates high quality images
- Latent space has some local linar properties (vector arithmetic like with Word2Vec)

Radford, Alec, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. 2015.

Style GANs

A Style-Based Generator Architecture for Generative Adversarial Networks by Tero Karras, Samuli Laine, Timo Aila, 2018, and later versions

Super Resolution



"Perceptual" loss = combining pixel-wise loss mse-like loss with GAN loss

Ledig, Christian, et al. Photo-realistic single image super-resolution using a generative adversarial network. CVPR 2016.

Takeaways

(Reconstruction) Autoencoders

- have no direct probabilistic interpretation;
- are not designed to generate useful samples;
- encoder defines a useful latent representation.

Takeaways

GANs

- likelihood-free generative models;
- high quality samples from high dimensional distributions;
- discriminator not meant be used as encoder

Takeaways

Adversarial training is useful beyond generative models:

- domain adaptation;
- learning representations blind to sensitive attributes;
- defend against malicious inputs (adversarial examples);
- regularization by training on adversarial examples.

Takeaways

Adversarial training is useful beyond generative models:

- domain adaptation;
- learning representations blind to sensitive attributes;
- defend against malicious inputs (adversarial examples);
- regularization by training on adversarial examples.

Quality of samples depends a lot on the architectures of sub-networks.

Next: Lab 10!