Deep Learning

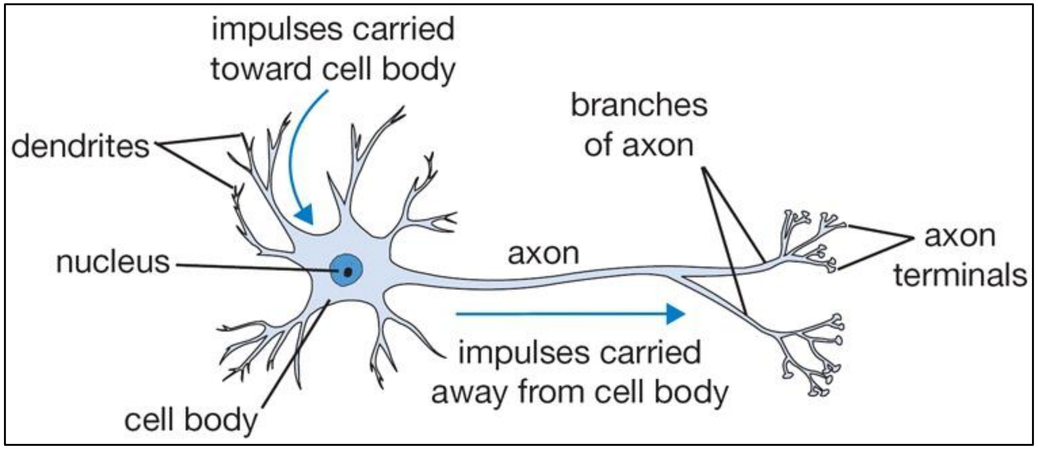
Onur Aydın

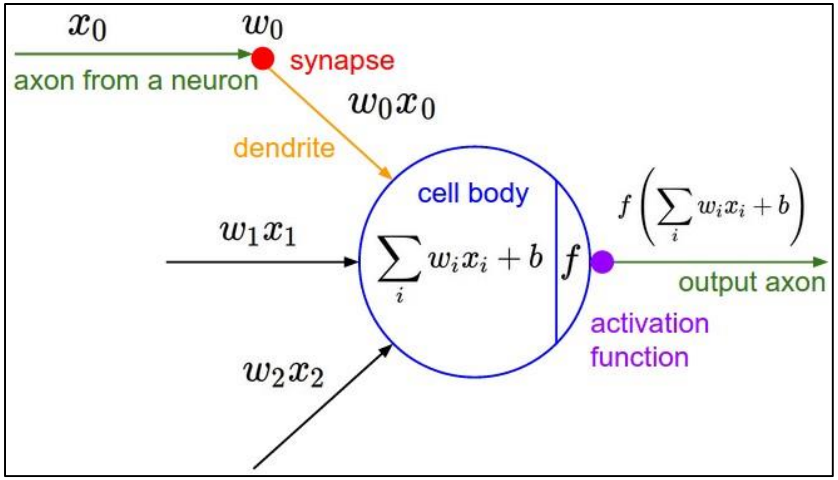
2017

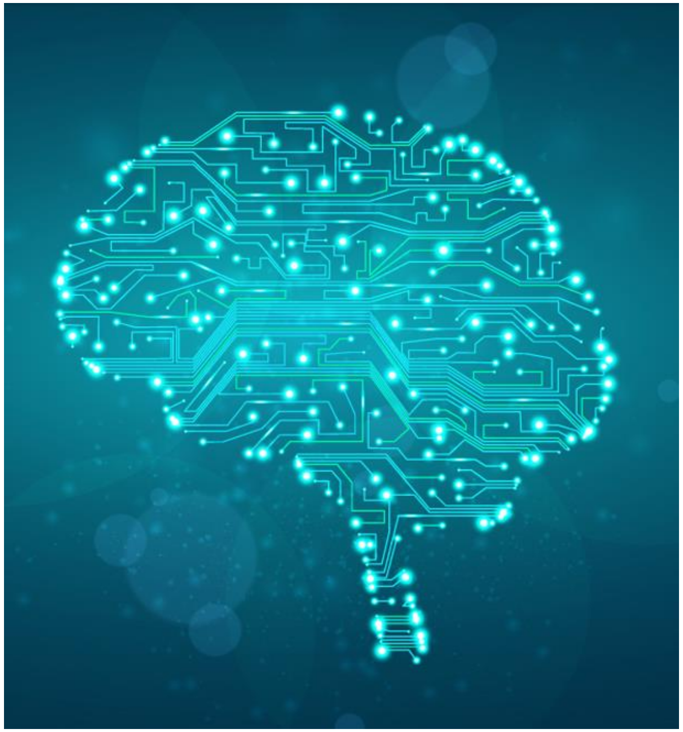
Outline

1. Introduction
2. **Feedforward Neural Networks**
3. Training and Optimization
4. **Convolutional Neural Networks**
5. Recurrent Neural Networks
6. **Deep Generative Models**
7. Advanced Topics in Deep Learning

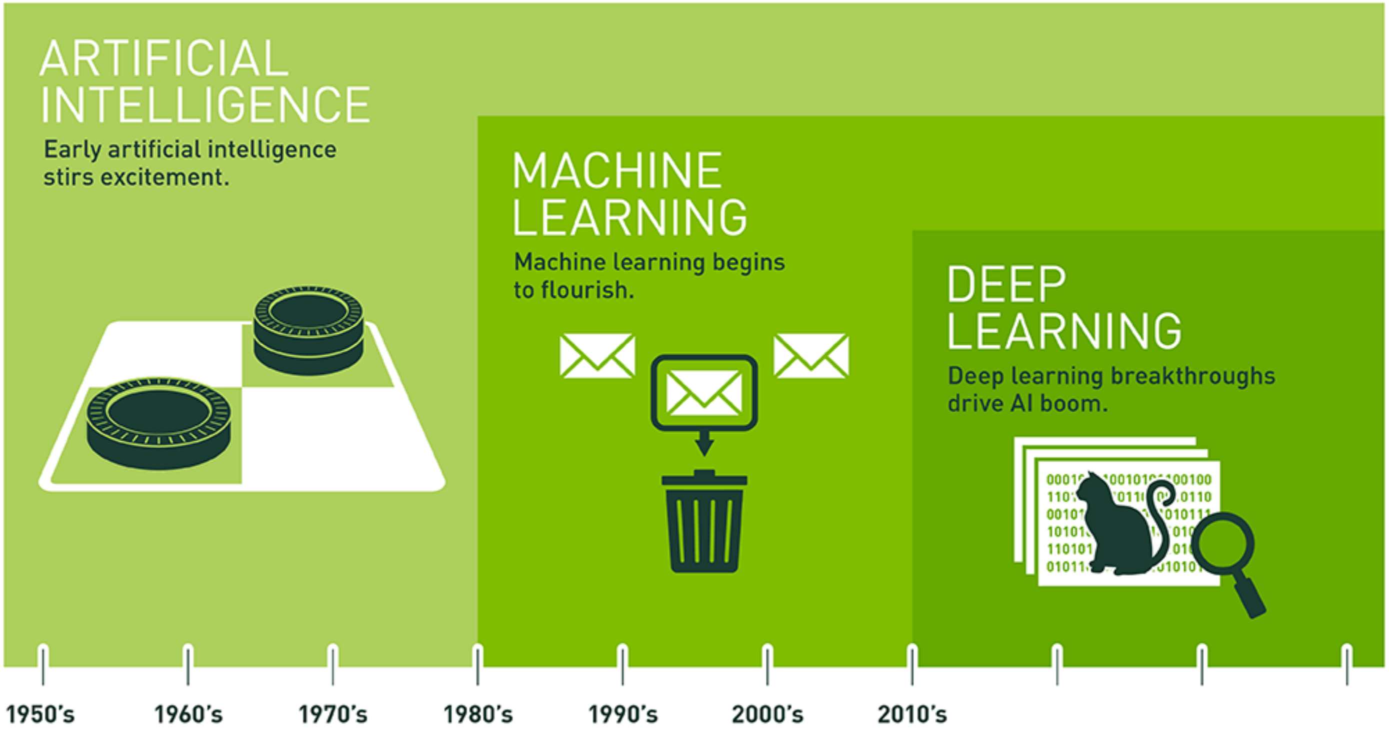
1. Introduction





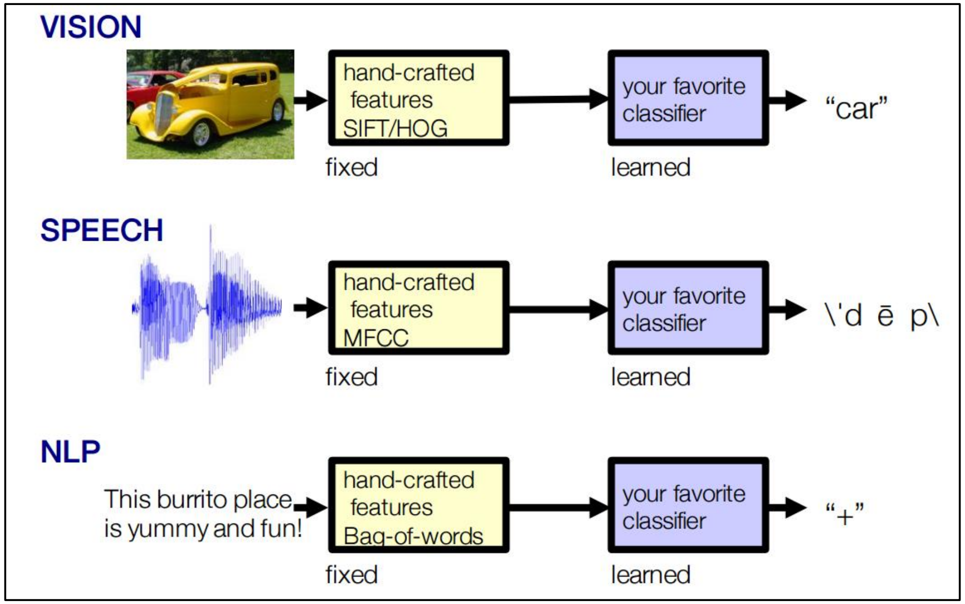


1. Introduction

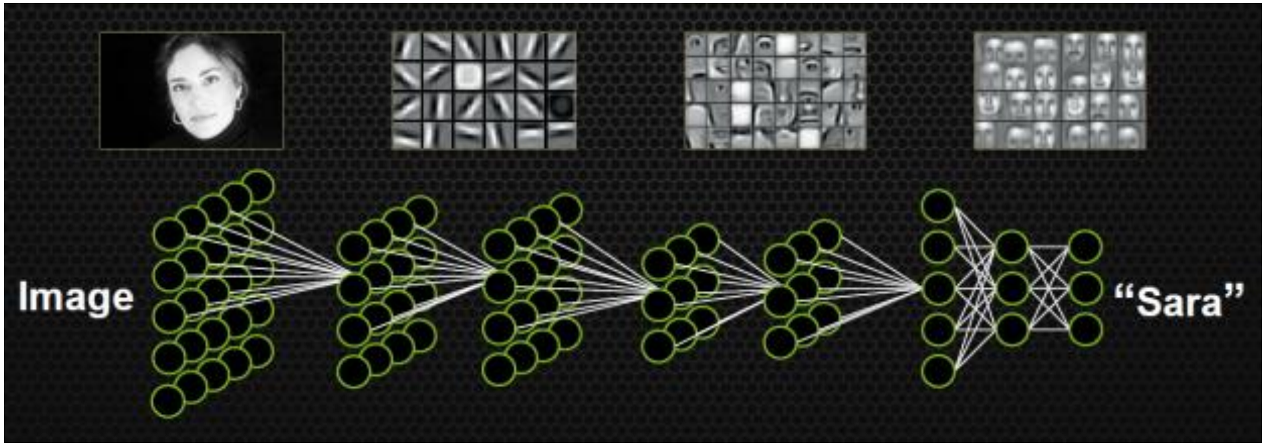


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

1. Introduction

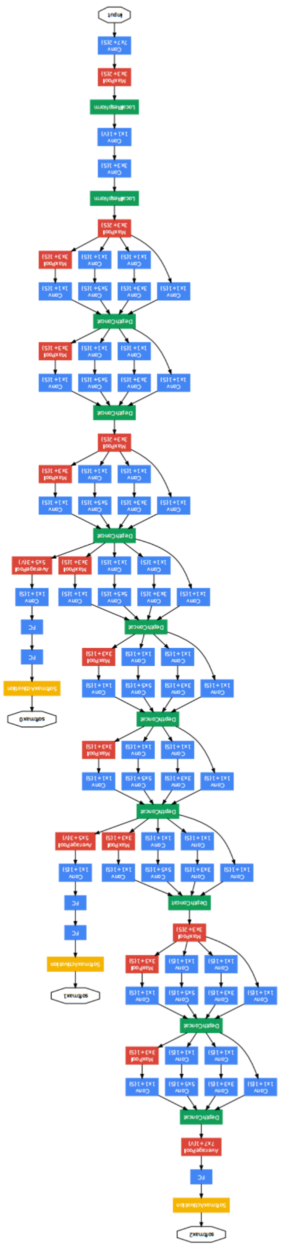


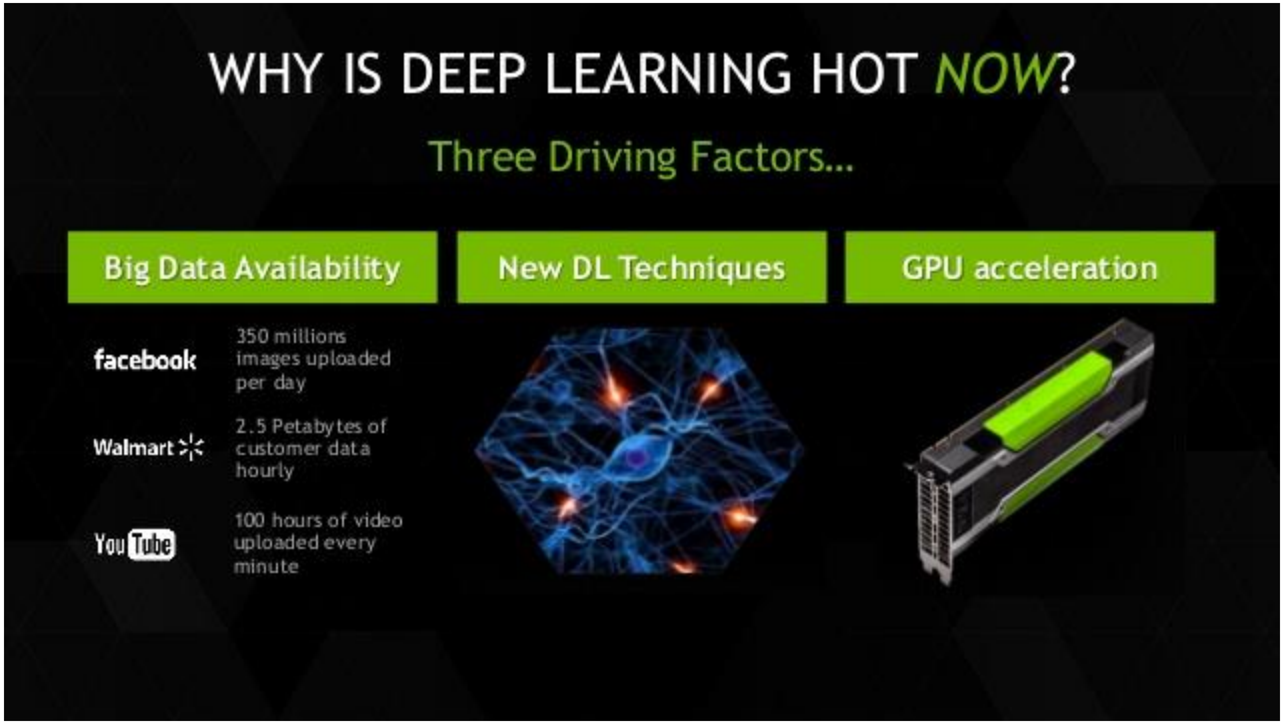
Traditional Machine Learning



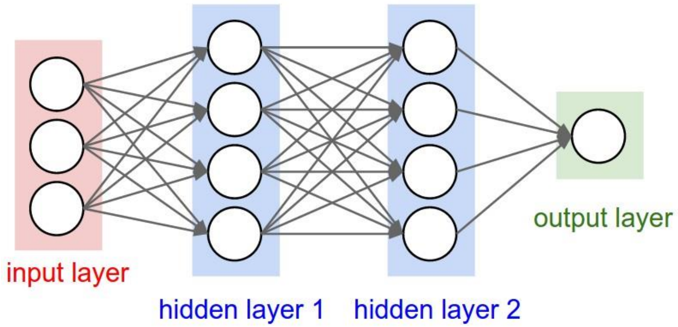
Deep Learning
(End-to-end Learning)

1. Introduction



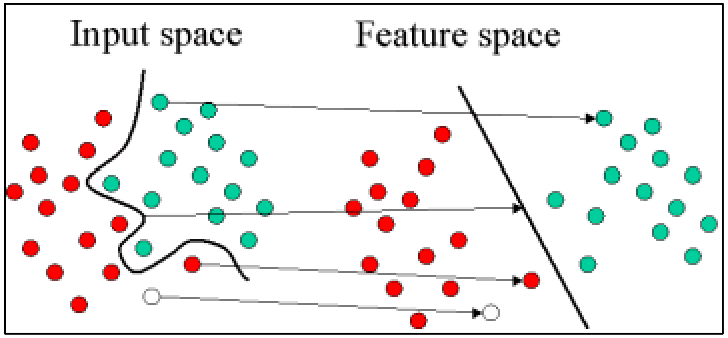


2. Feedforward Neural Networks



$$L(\theta, \hat{\theta})$$

$$\frac{\partial L}{\partial w}$$



- **NN with No Hidden Layers:**
  - Learns a Linear Function
  - Convex Problem

$$y = f(xW)$$

- **NN with Single Hidden Layer:**
  - Learns a Non-Linear Function
  - Non-Convex Problem

$$y = f_2(f_1(xW_1)W_2)$$

2. Feedforward Neural Networks

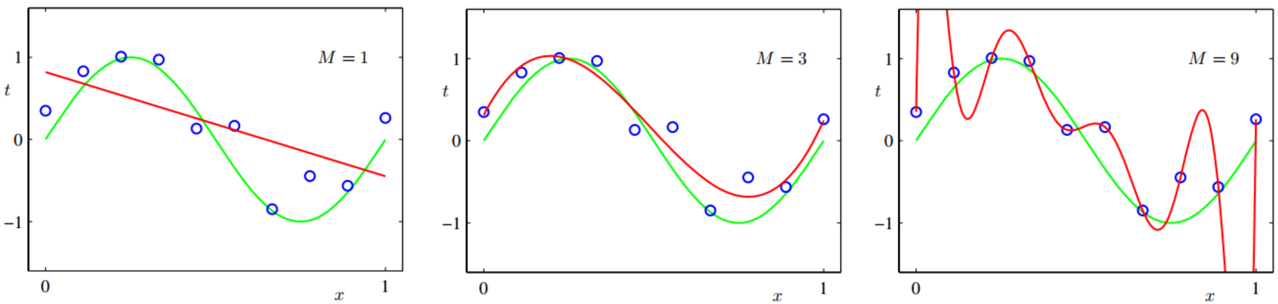
Universal Approximation Theorem

Theoretically,

'A feedforward network with a single hidden layer containing a finite number of neurons can approximate any Borel measurable function from one finite dimensional space to another with any desired non-zero amount of error.'

But in practice:

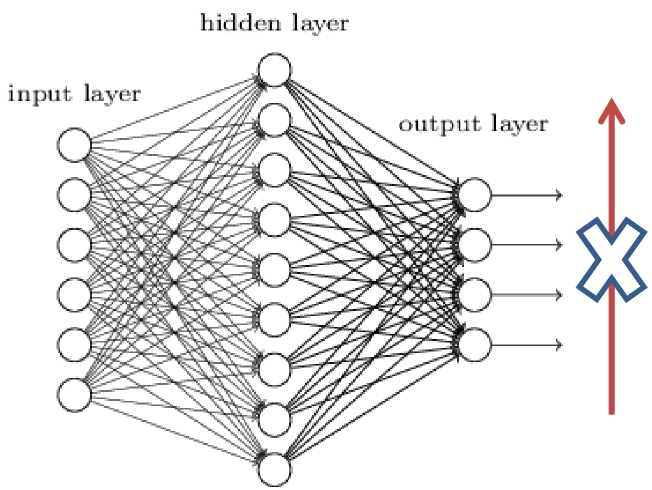
- ✓ It may require too many parameters
- ✓ The optimization algorithm may fail
- ✓ Finding a good generalization is more important



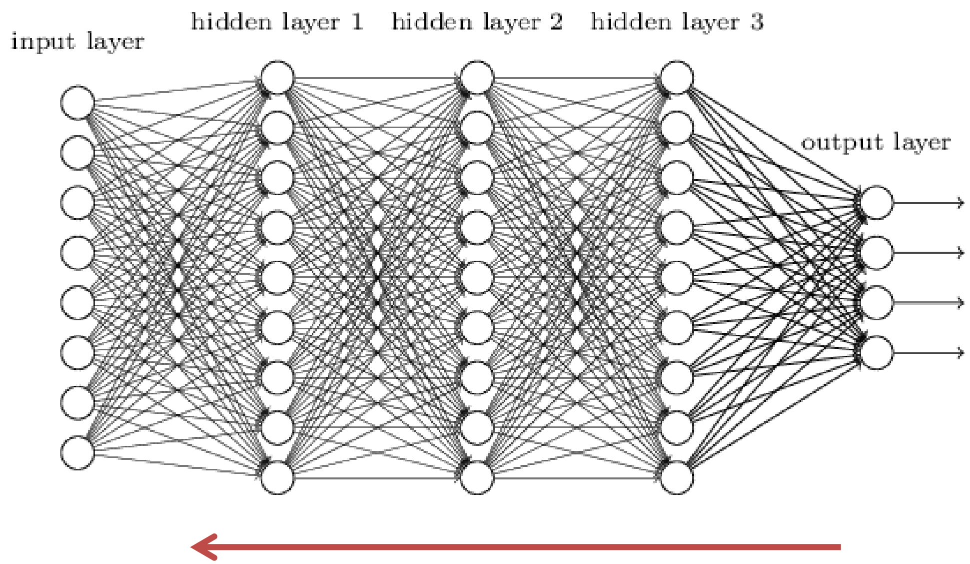
2. Feedforward Neural Networks

The universal approximation theorem

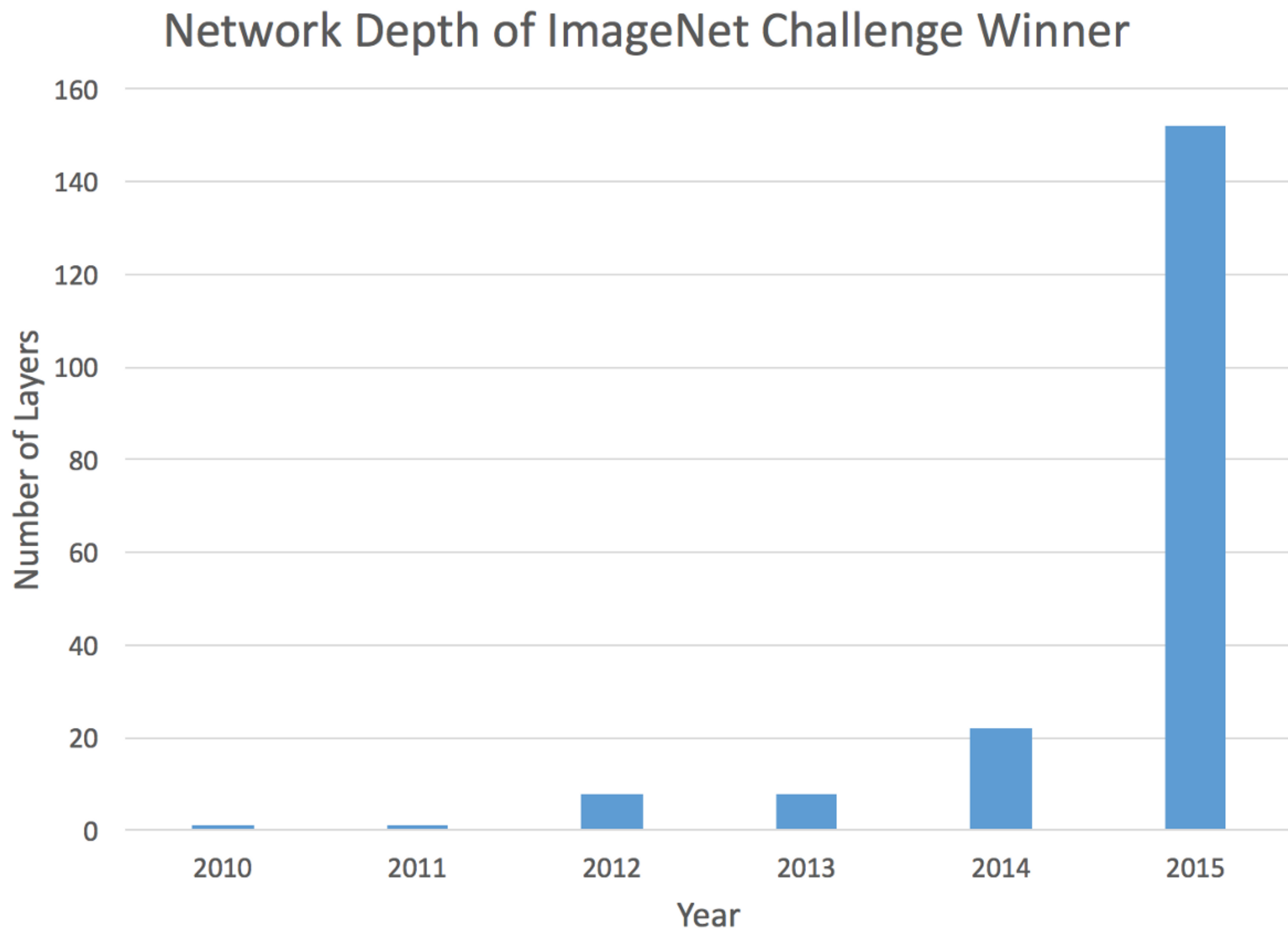
"Non-deep" feedforward
neural network



Deep neural network

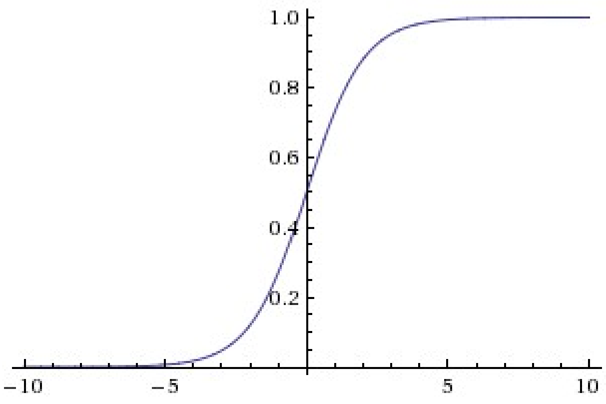


2. Feedforward Neural Networks

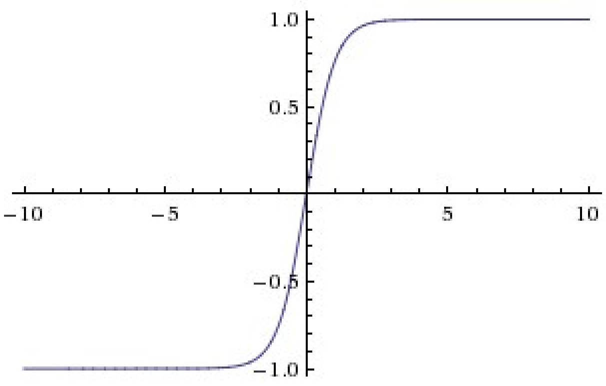


3. Training and Optimization

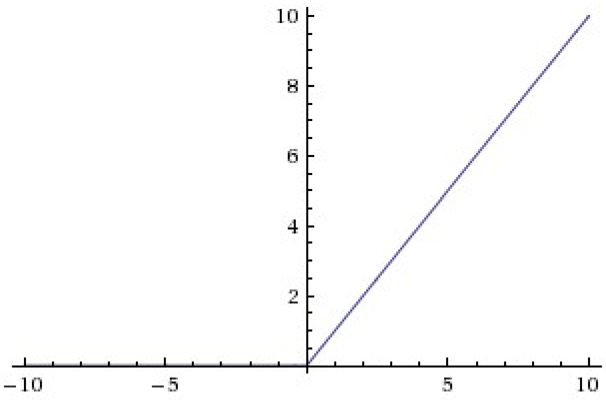
Non-linear Activation Functions



Sigmoid



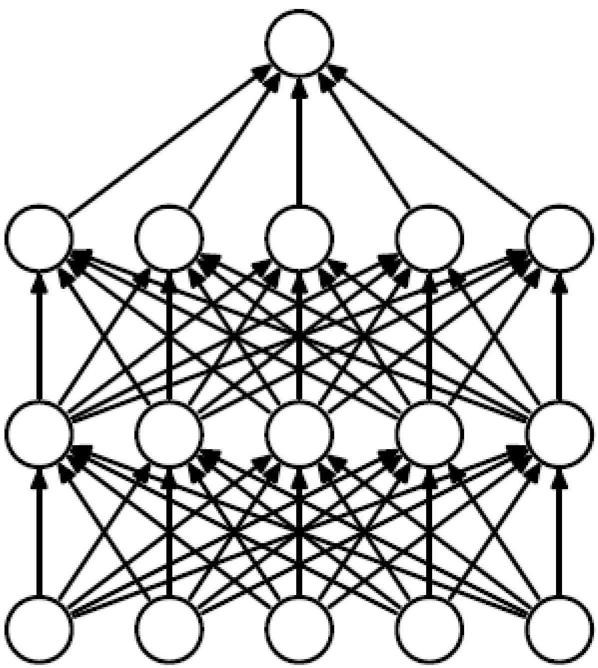
Hyperbolic Tangent (tanh)



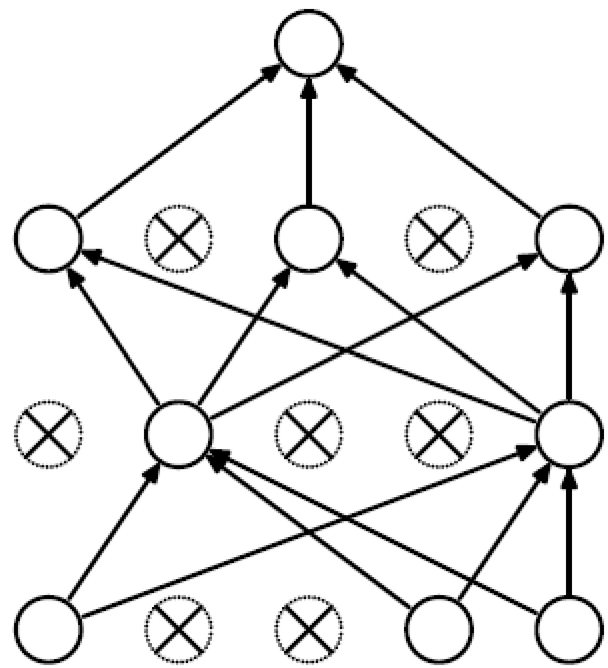
Rectified Linear Unit (RELU)

3. Training and Optimization

Dropout



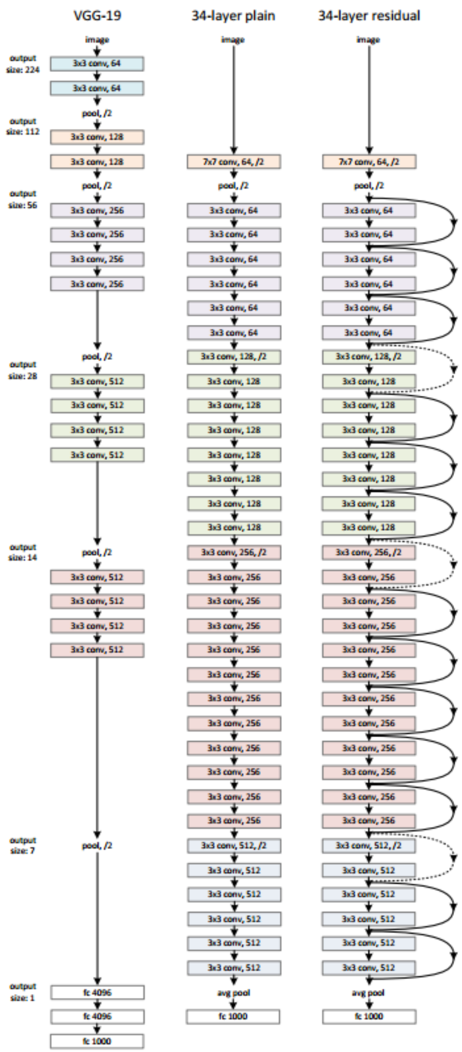
(a) Standard Neural Net

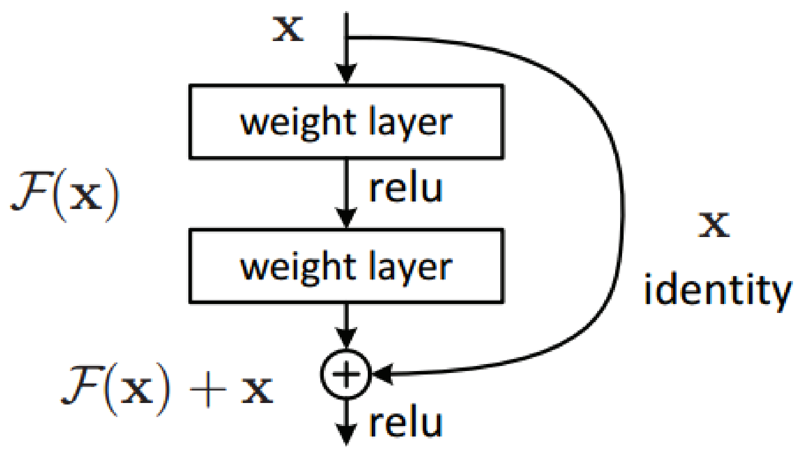


(b) After applying dropout.

3. Training and Optimization

Residual Connections

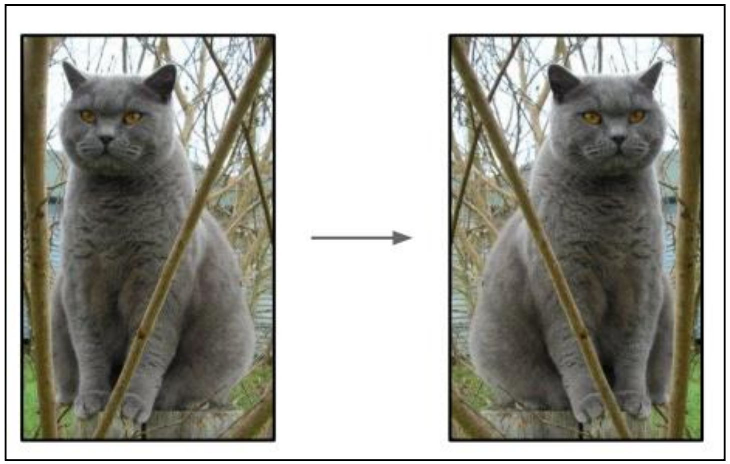




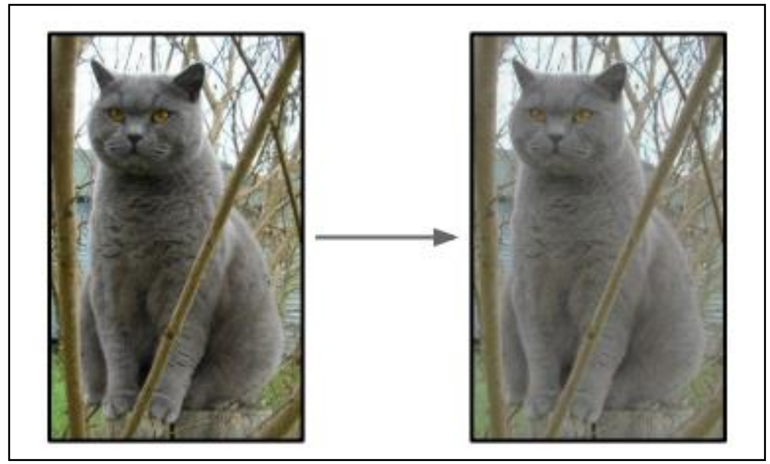
3. Training and Optimization

Data Augmentation

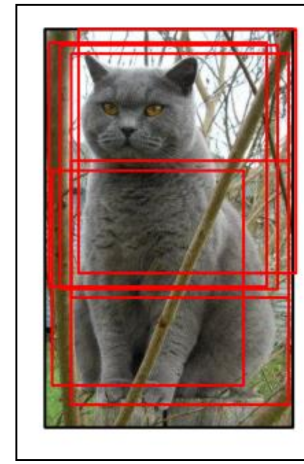
- ✓ Artificially increase dataset size



Horizontal Flips



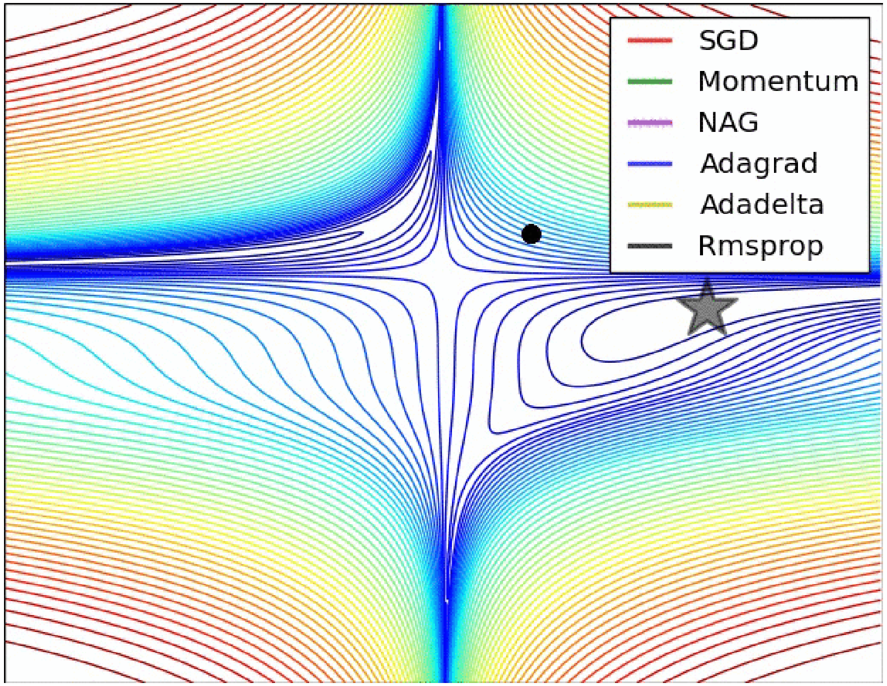
Change Contrast

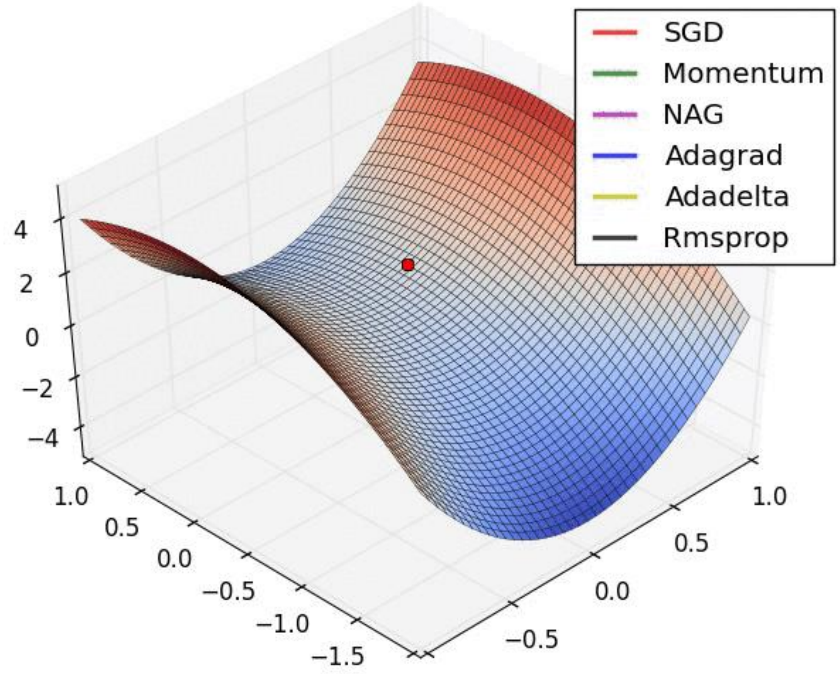


Random Crops and Scales

3. Training and Optimization

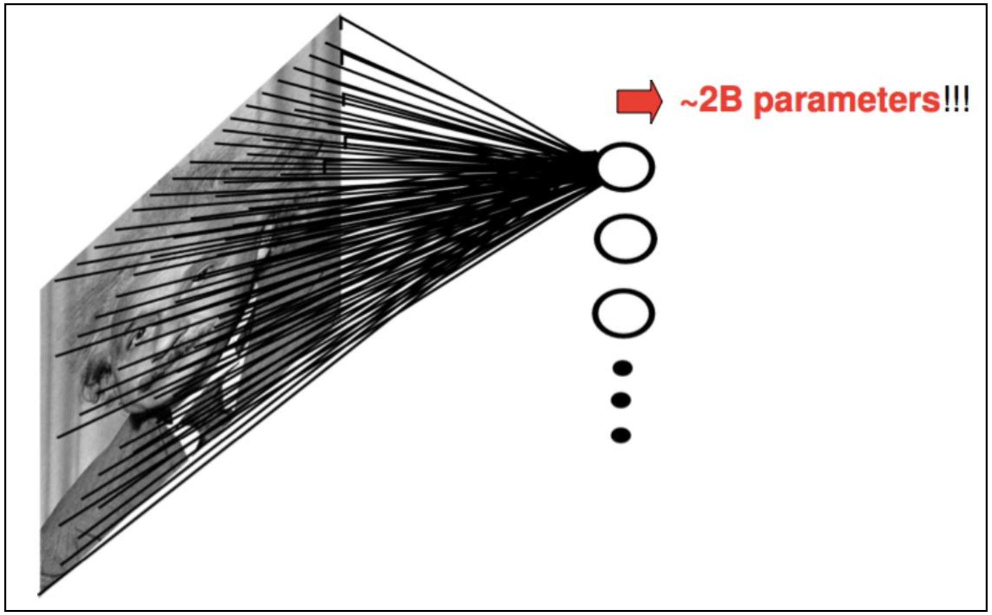
Optimization



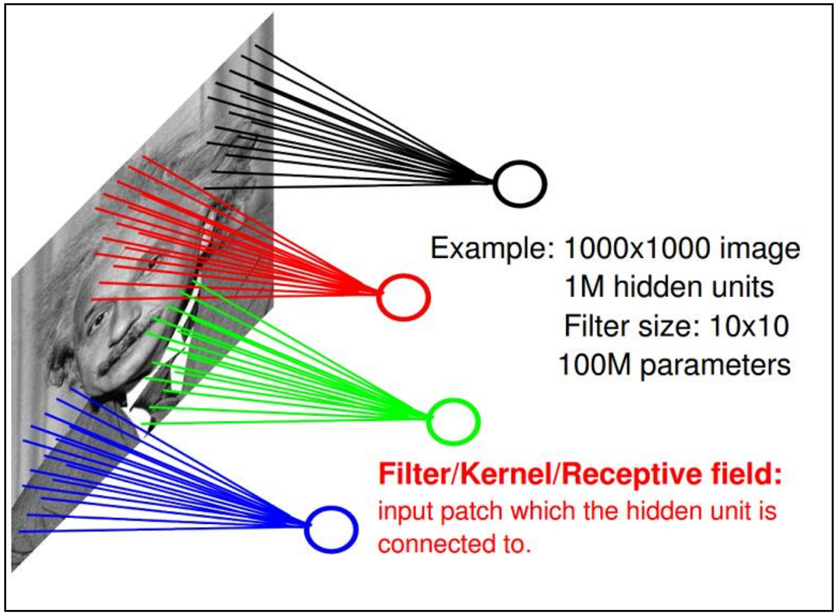


4. Convolutional Neural Networks

Feedforward Network



Convolutional Network

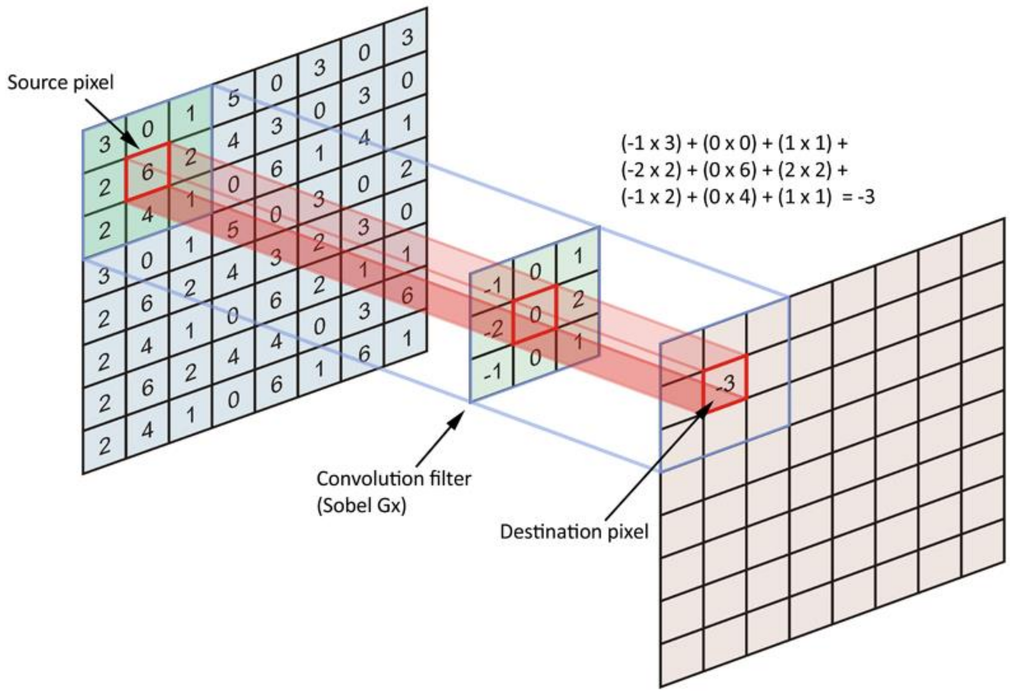


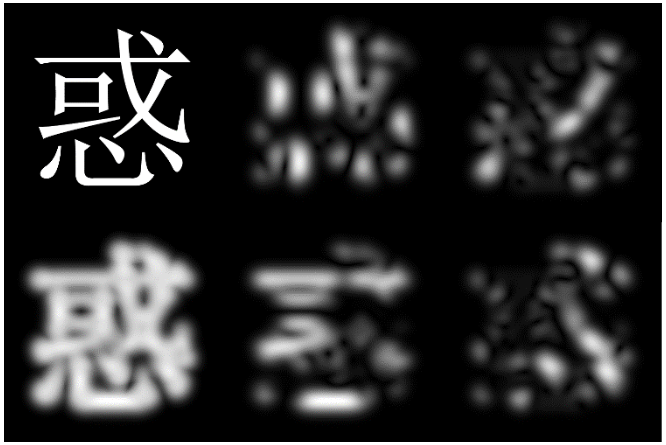
Matrix Multiplication

Convolution Operation

4. Convolutional Neural Networks

The Convolution Operation

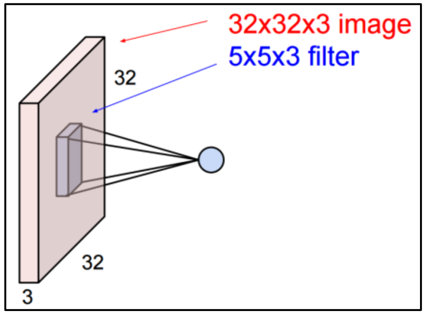




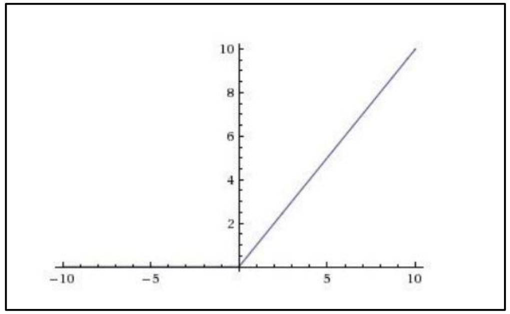
$$f[x] * g[x] = \sum_{k=-\infty}^{\infty} f[k] \cdot g[x-k]$$

$$f[x,y] * g[x,y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1,n_2] \cdot g[x-n_1, y-n_2]$$

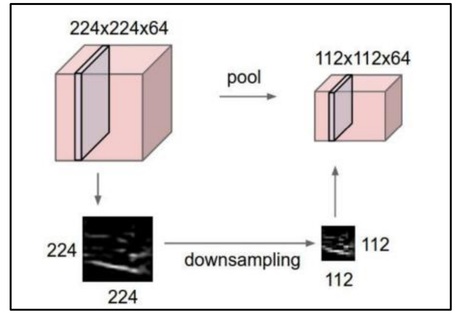
4. Convolutional Neural Networks



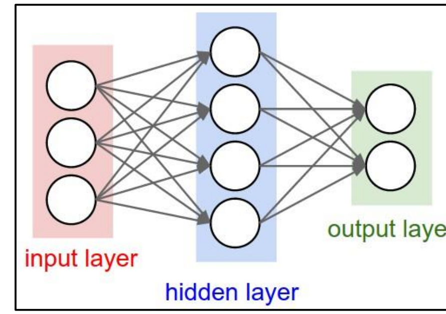
CONV



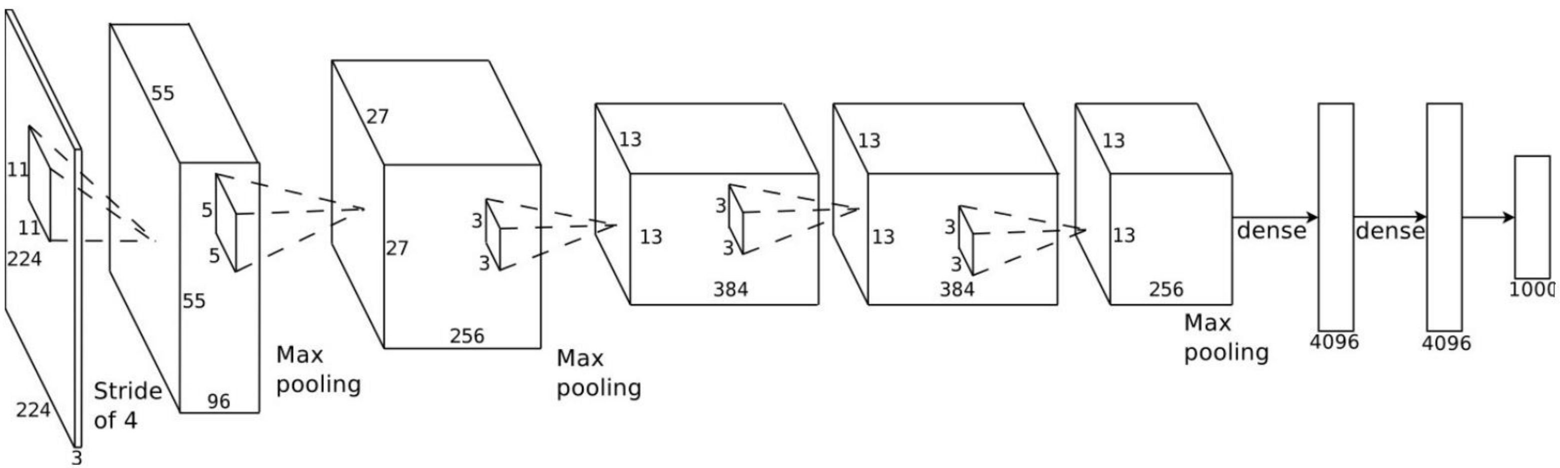
RELU



POOL

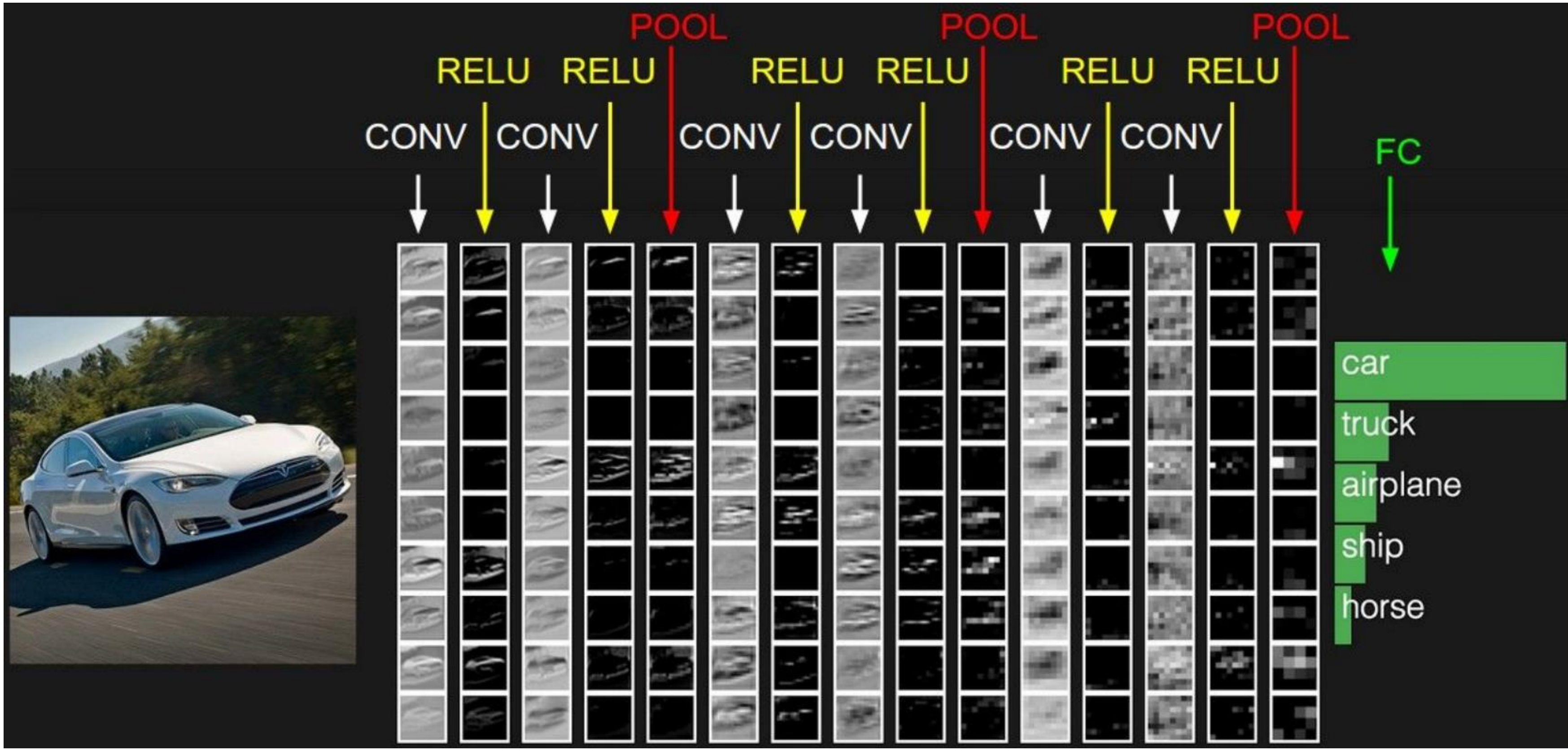


FC



AlexNet – [Krizhevsky et al. 2012]

4. Convolutional Neural Networks



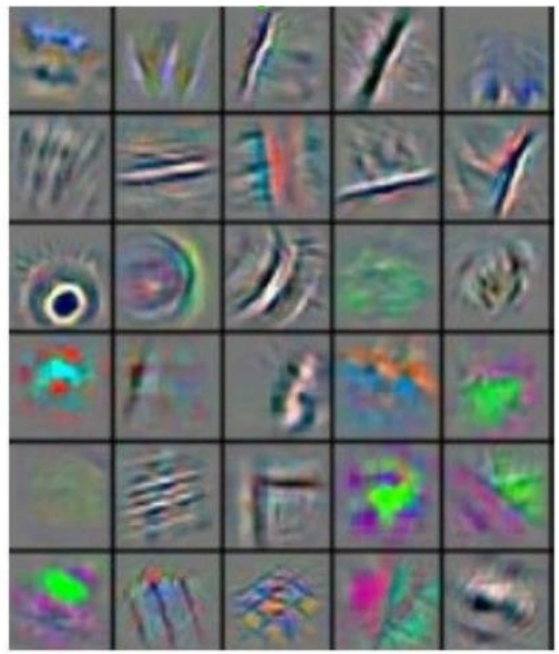
Feature Extraction

Classification

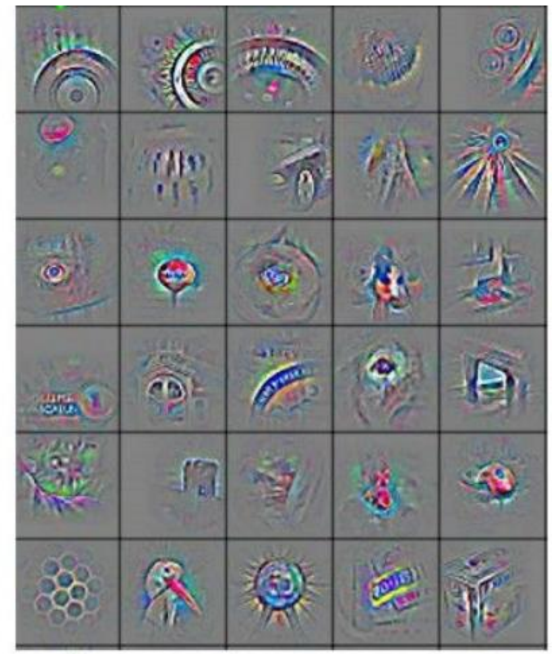
4. Convolutional Neural Networks

Filters learned by CNN

Low level Structures



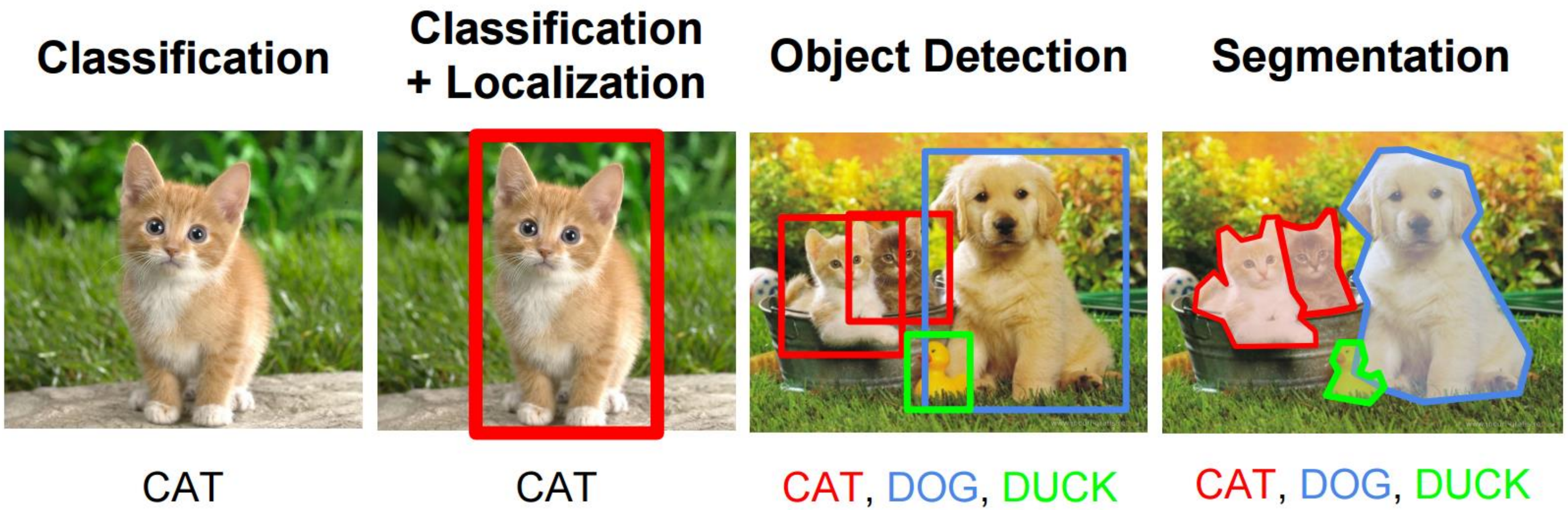
More Complex Structures



Highly Complex, Abstract Concepts

4. Convolutional Neural Networks

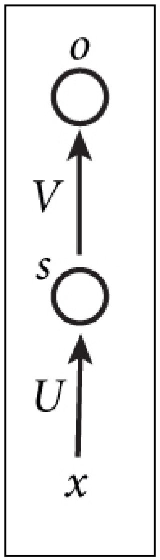
CNN Architectures for Other Computer Vision Problems:



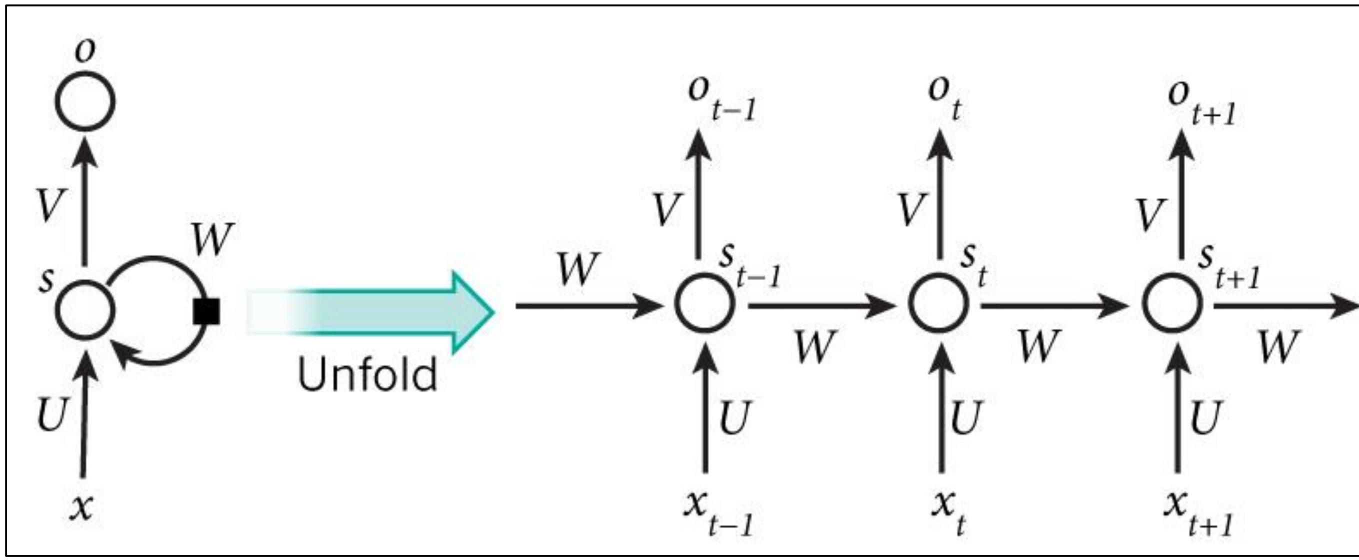
Classification: LeNet, AlexNet, GoogLeNet, VGGNet
Object Detection: R-CNN, Faster R-CNN
Segmentation: FCN, DeconvNet, SegNet
Image Super Resolution: SRCNN, EnhanceNet, SRGAN

5. Recurrent Neural Networks

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$



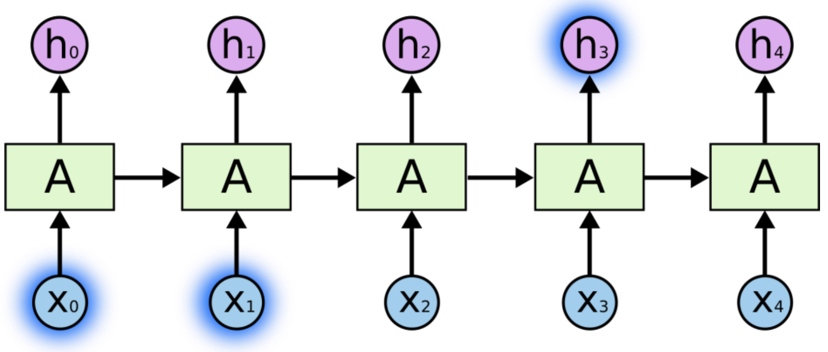
Feedforward Networks



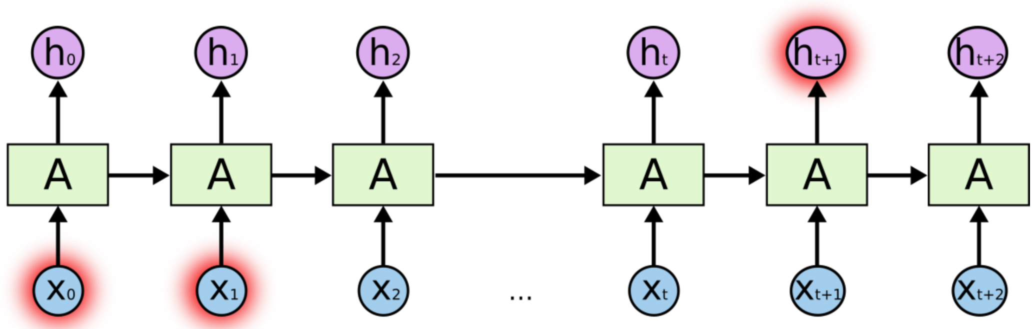
Recurrent Networks

5. Recurrent Neural Networks

The Problem of Long-Term Dependencies



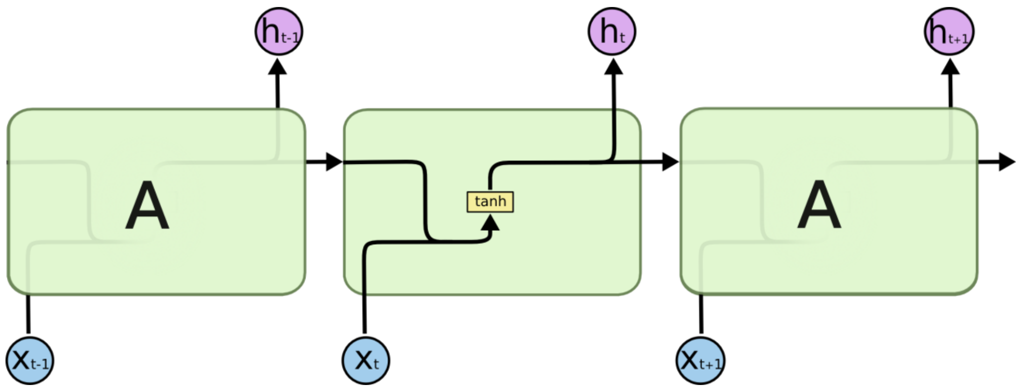
'the clouds are in the *sky*'



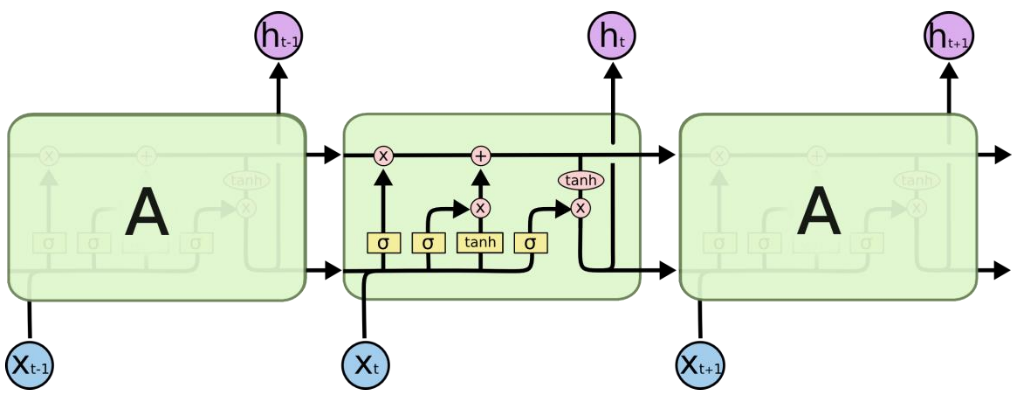
'I grew up in France… I speak fluent *French*'

5. Recurrent Neural Networks

Long Short Term Memory Networks (LSTM)



RNN



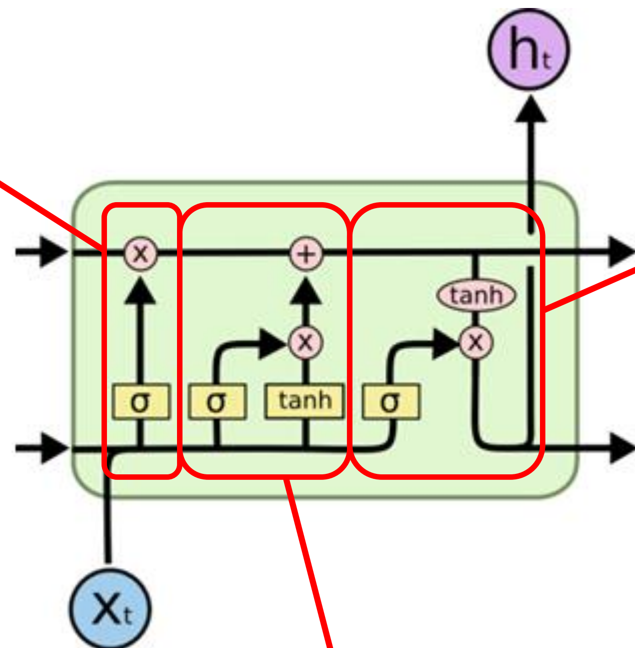
LSTM

LSTM with fewer parameters = Gated Recurrent Unit (GRU)

5. Recurrent Neural Networks

Long Short Term Memory Networks (LSTM)

Forget irrelevant parts of previous state



Output certain parts of cell state

Selectively update cell state values

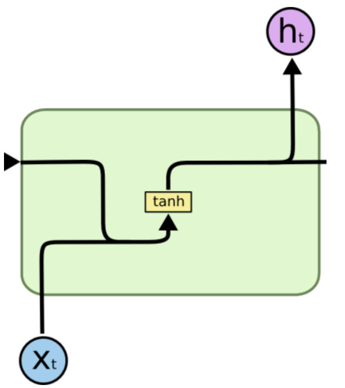
5. Recurrent Neural Networks

Long Short Term Memory Networks (LSTM)

RNN:

$$\boxed{h_t^l = \tanh W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix}}$$

$h \in \mathbb{R}^n$. $\qquad W^l \; [n \times 2n]$



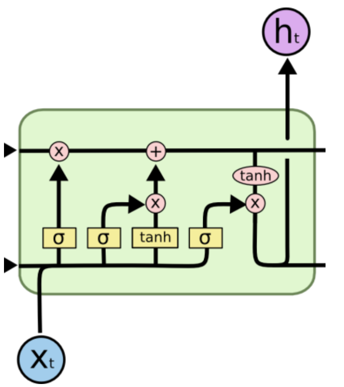
LSTM:

$W^l \; [4n \times 2n]$

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \text{sigm} \\ \text{sigm} \\ \text{sigm} \\ \tanh \end{pmatrix} W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix}$$

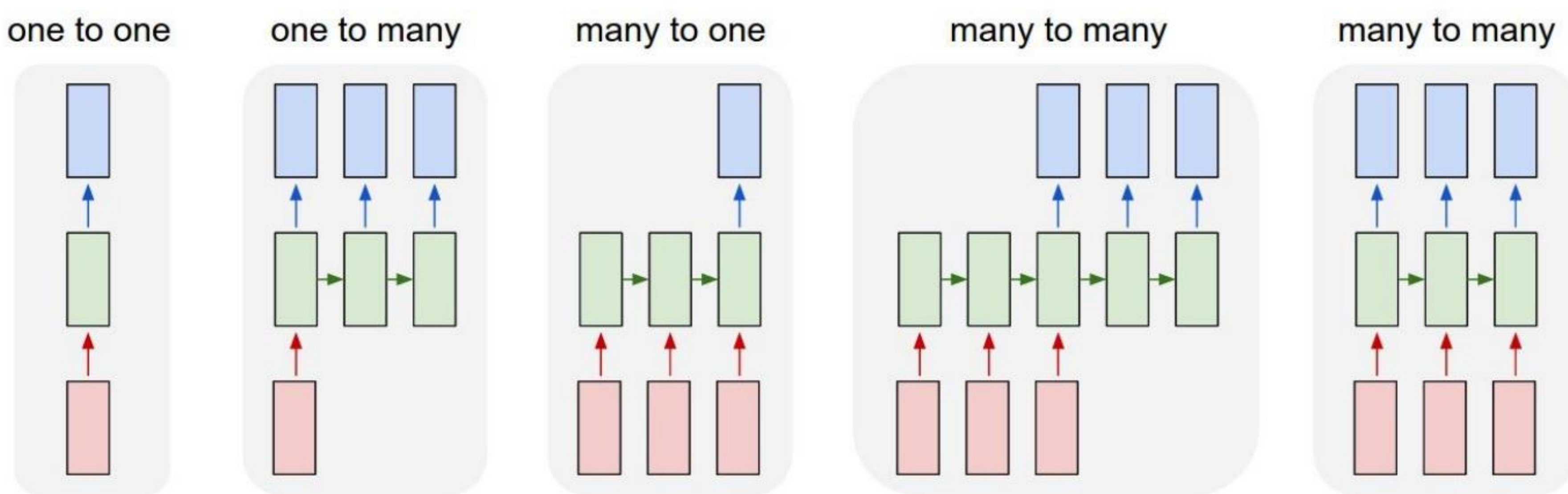
$$c_t^l = f \odot c_{t-1}^l + i \odot g$$

$$h_t^l = o \odot \tanh(c_t^l)$$



5. Recurrent Neural Networks

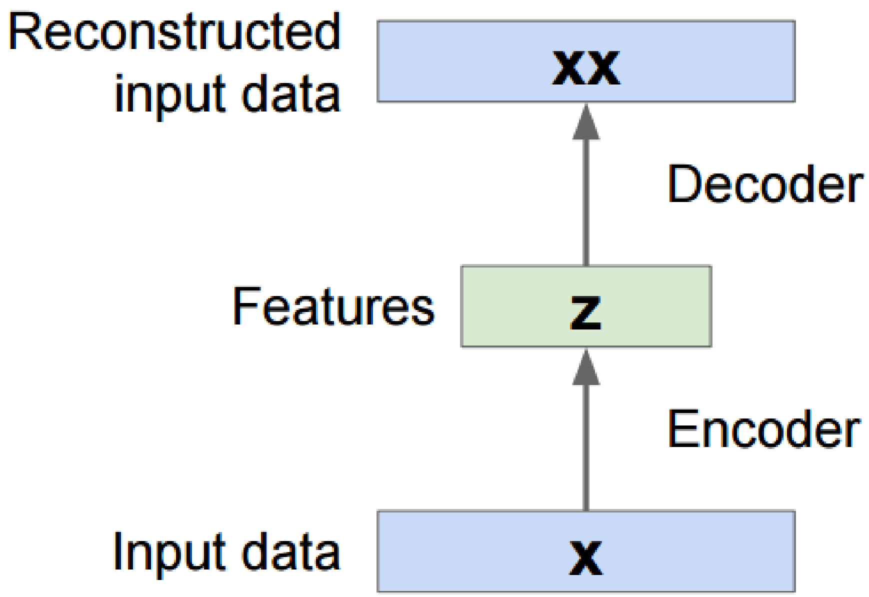
Applications and Network Architectures



Feedforward Networks | Image Captioning | Sentiment Classification | Machine Translation | Video Classification

6. Deep Generative Models

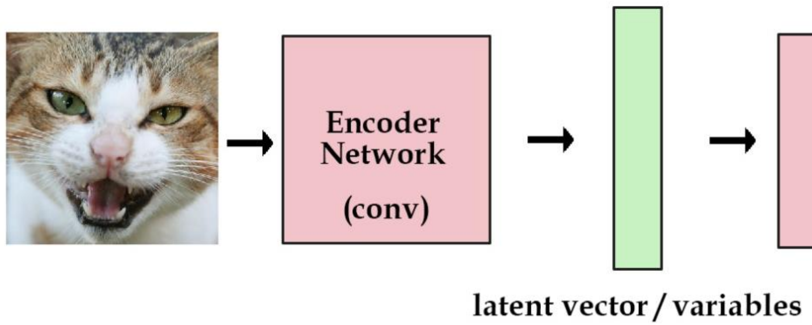
Autoencoders

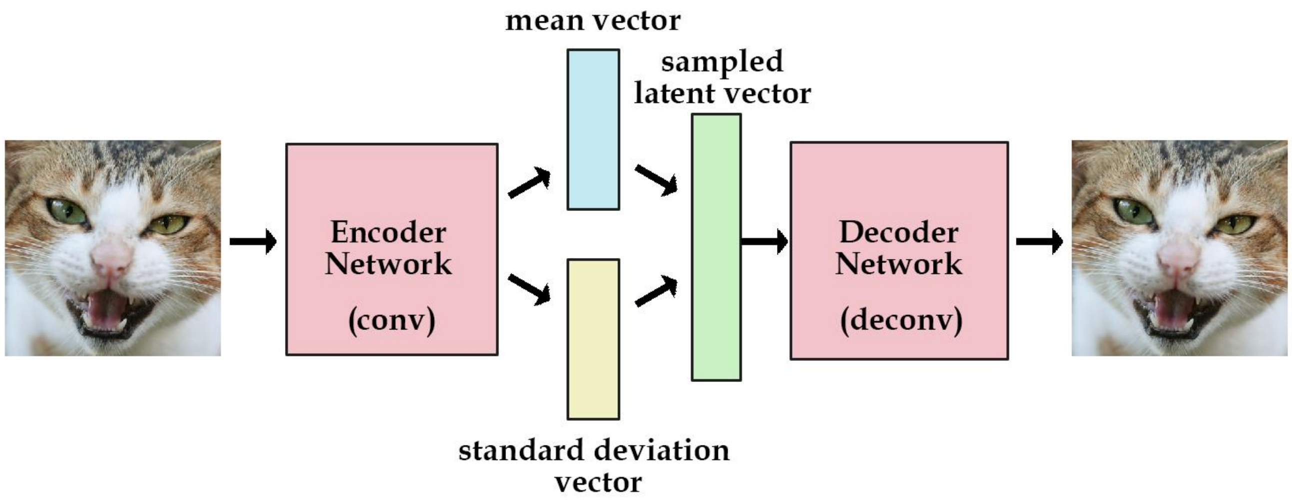


- ✓ Minimizes reconstruction error (e.g. MSE or KL divergence)
  - ✓ z < x - Dimension Reduction
  - ✓ z > x - Denoising Autoencoder
- ✓ Linear Autoencoder = PCA

6. Deep Generative Models

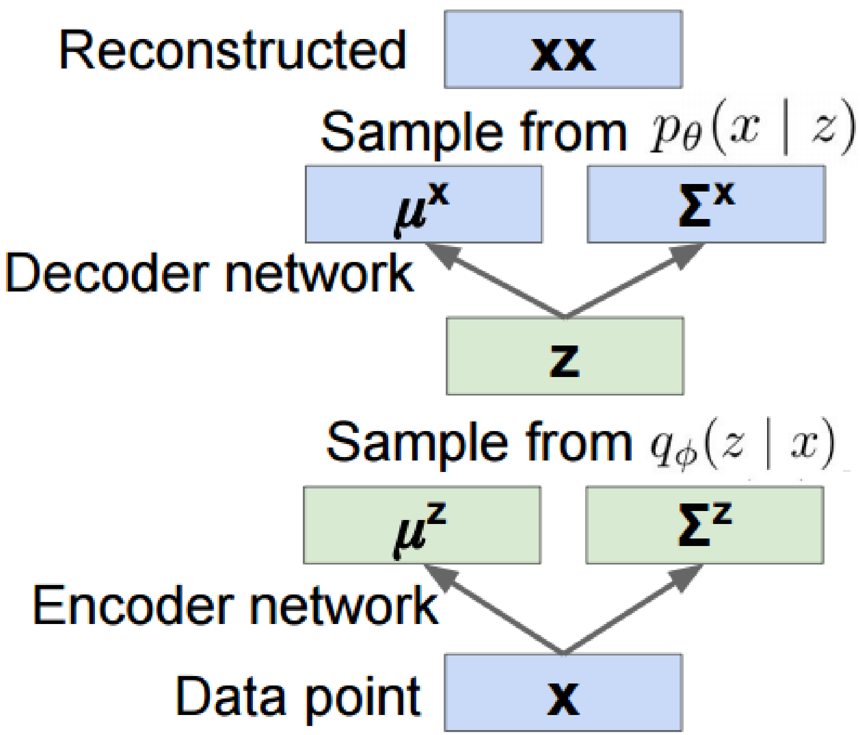
Variational Autoencoders





6. Deep Generative Models

Variational Autoencoders



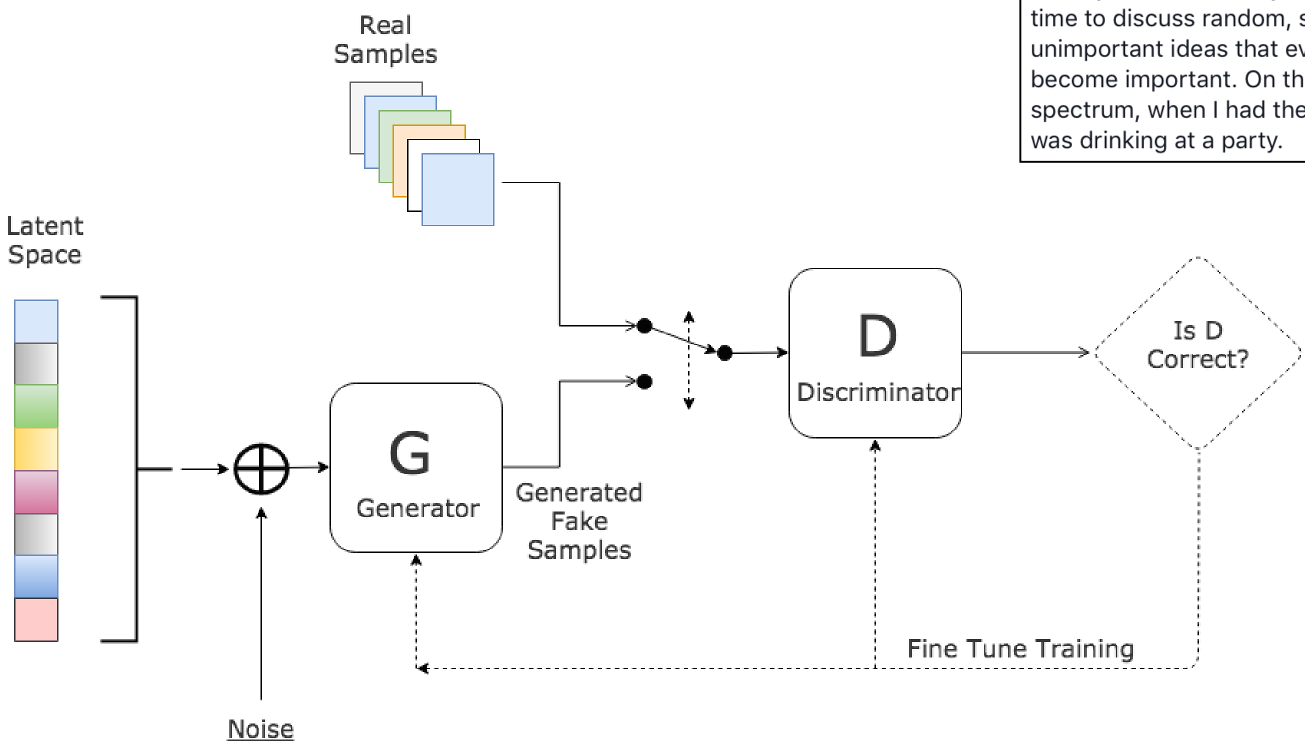
- ✓ Force network to generate latent vectors that follow unit Gaussian distribution

- ✓ **Generative Loss (MSE):** How accurately the network reconstructed inputs
- ✓ **Latent Loss (KL Divergence):** How closely the latent variable match a unit Gaussian
- ✓ Loss = latent loss + generative loss

6. Deep Generative Models

> **Ian Goodfellow** shared **Stevie Jeung**'s post.
> 3 September 2016 ·
>
> I think the least productive I've ever been was also the time I worked the most hours, as a masters student at Stanford. Almost nothing I did in that time got published. Sleep deprivation makes you less intelligent, and when you're constantly busy you don't have time to discuss random, seemingly unimportant ideas that eventually click and become important. On the other end of the spectrum, when I had the idea for GANs, I was drinking at a party.

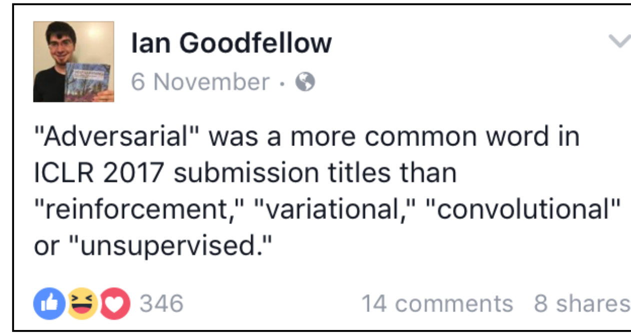
Generative Adversarial Networks (GAN)



Generator Network (G) generates artificial samples

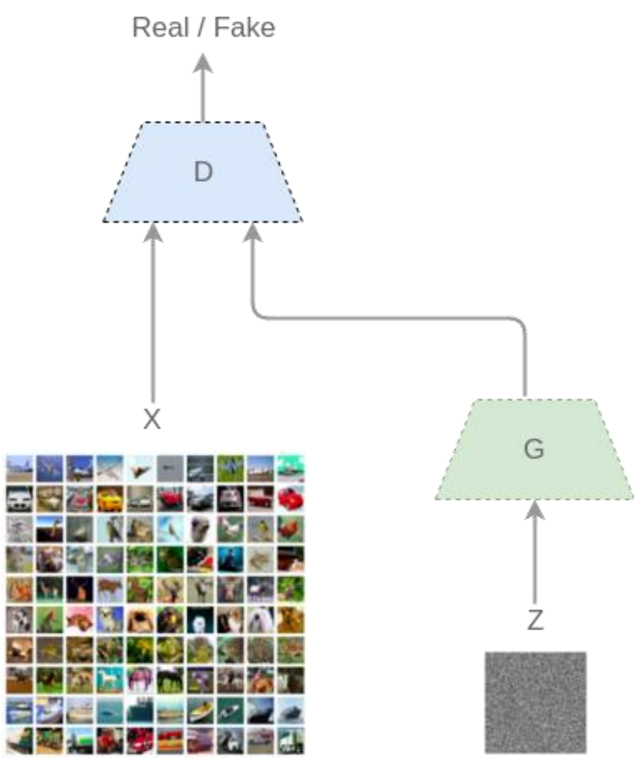
Discriminator Network (D) discriminates real samples from artificial samples

6. Deep Generative Models



Generative Adversarial Networks (GAN)

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$



Both networks are trained simultaneously until the artificial samples are indistinguishable from real samples

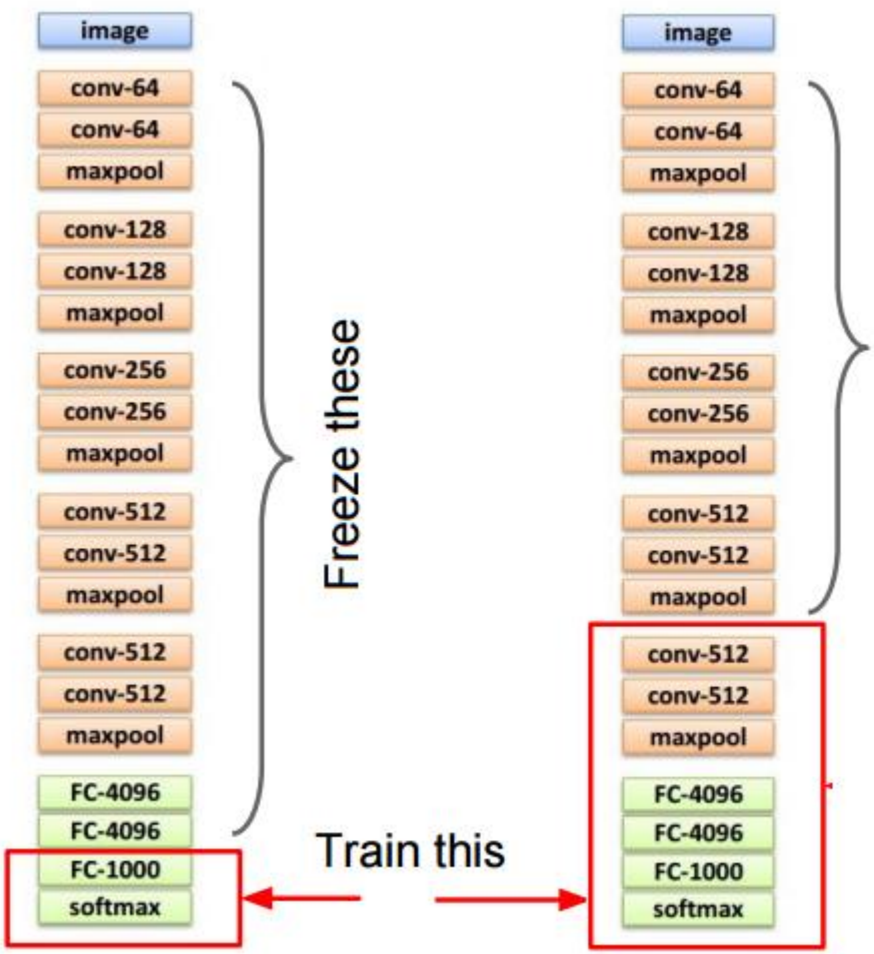
6. Deep Generative Models

Deep Convolutional Generative Adversarial Networks (DCGAN)

Artificially generated bedroom images

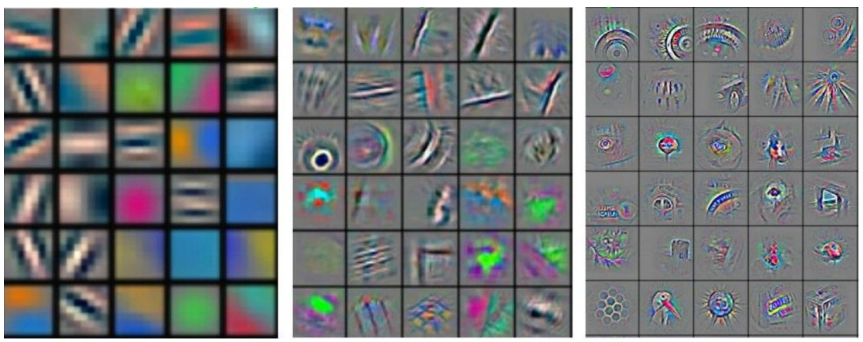
7. Advanced Topics in Deep Learning

1. Transfer Learning and Domain Adaptation



Small Dataset

Medium Dataset



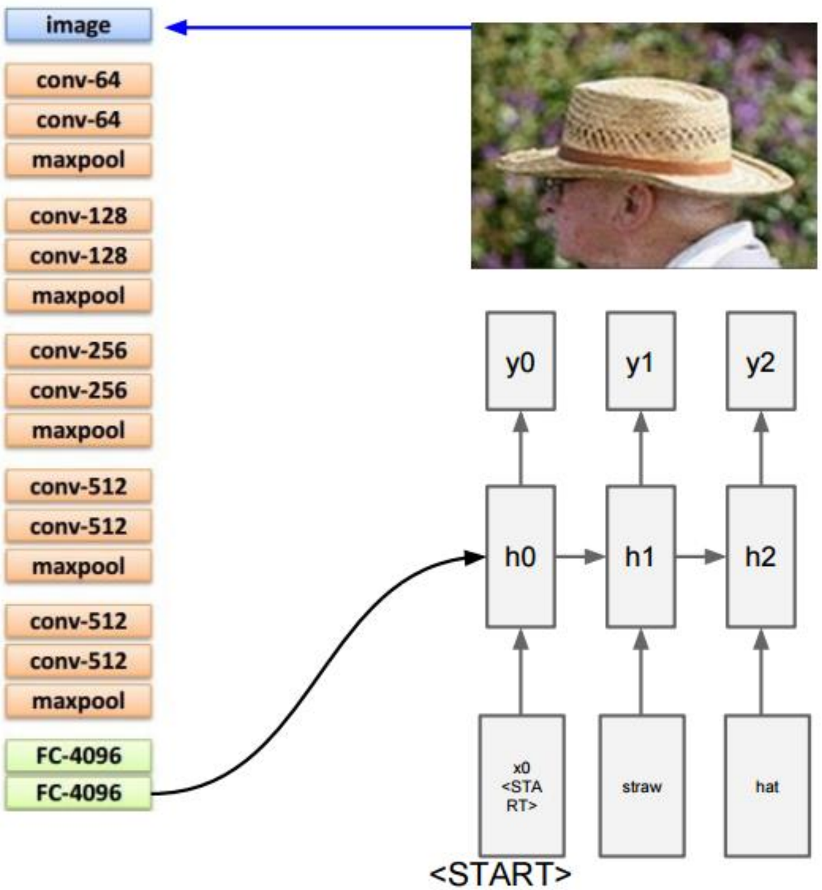
Low level Structures

More Complex Structures

Highly Complex, Abstract Concepts

7. Advanced Topics in Deep Learning

2. Image Captioning

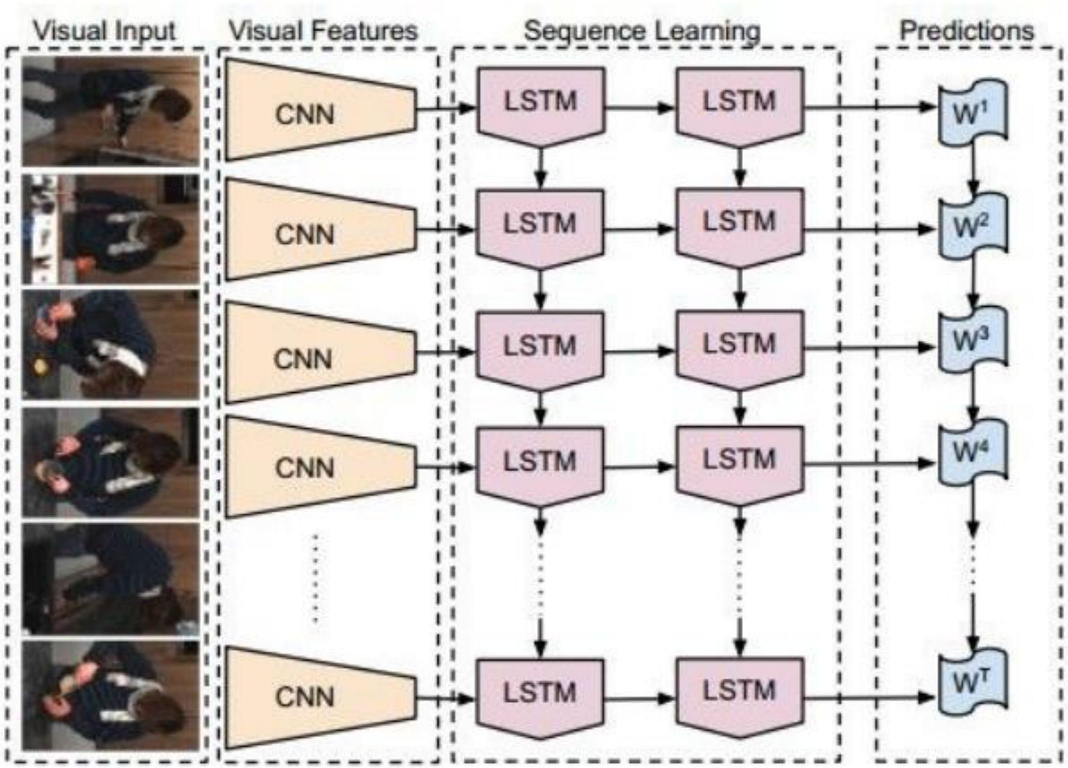


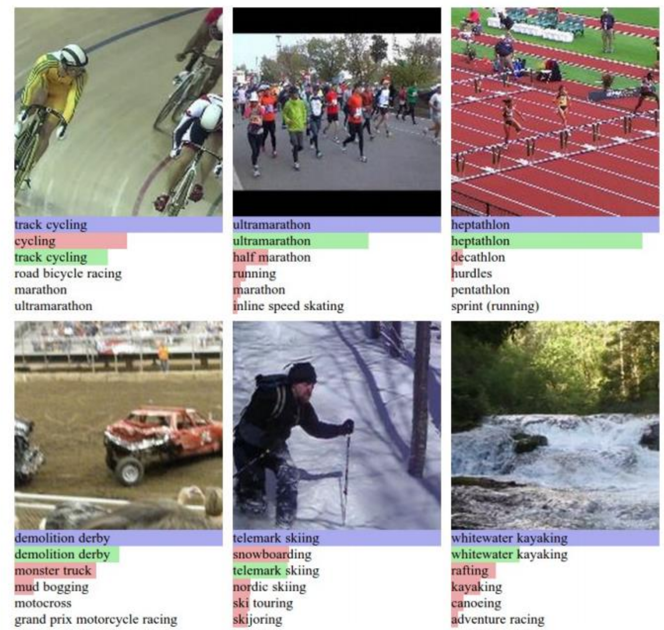
"man in black shirt is playing guitar."

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

7. Advanced Topics in Deep Learning

3. Video Classification





7. Advanced Topics in Deep Learning

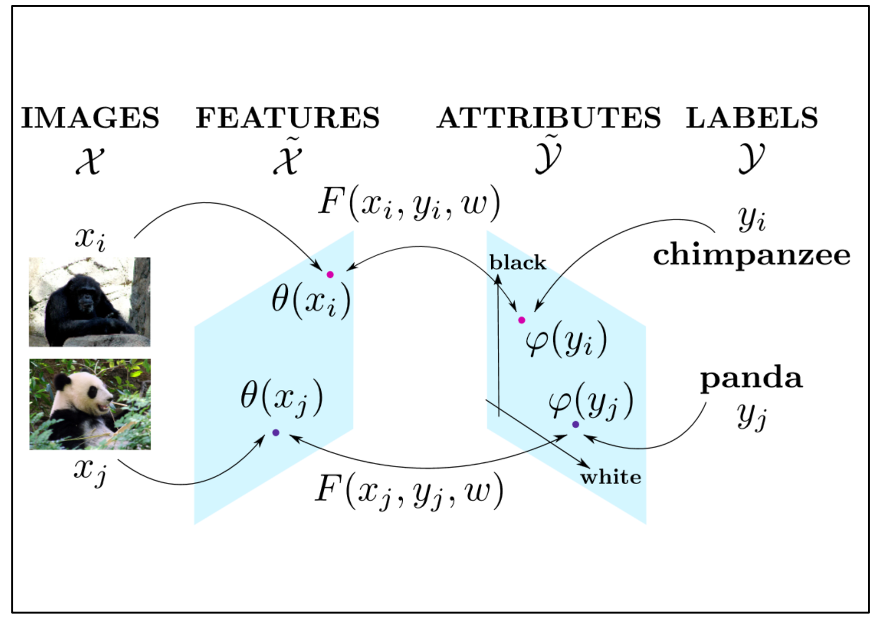
3. Zero-shot Learning, One-shot Learning, Few-shot Learning

- ✓ Red
- ✓ Fruit
- ✓ Eatable
- ✓ …

- ✓ Red
- ✓ Shoe
- ✓ Not Eatable
- ✓ …

- ✓ Yellow
- ✓ Fruit
- ✓ Eatable
- ✓ …

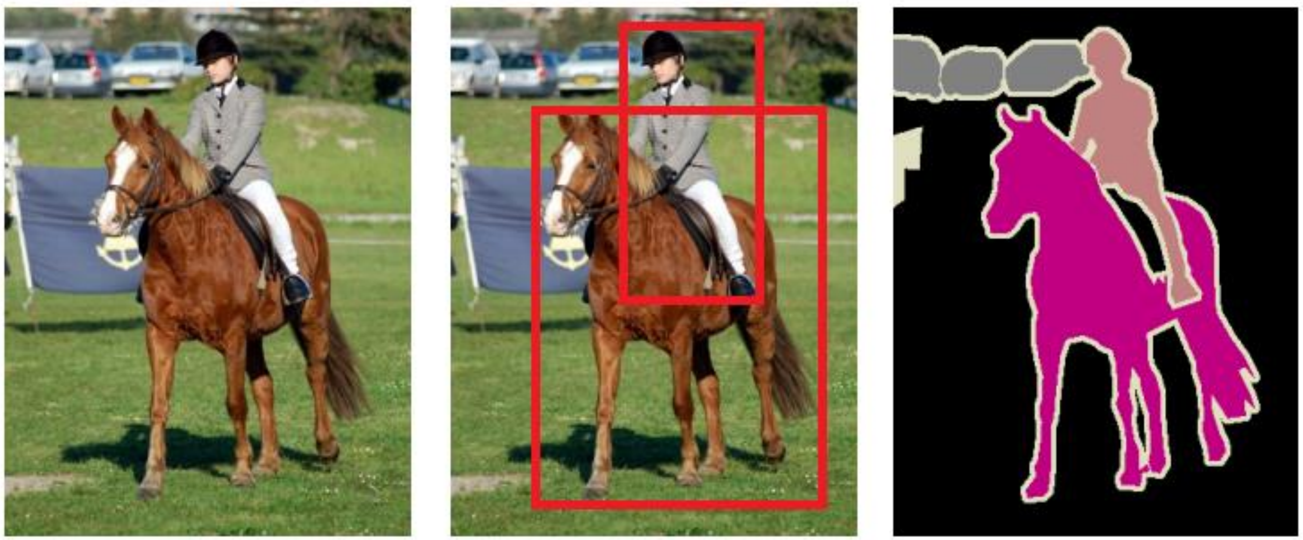
- ✓ Yellow
- ✓ Shoe
- ✓ Not Eatable
- ✓ …



7. Advanced Topics in Deep Learning

4. Weakly Supervised Learning

Image | Bounding Box | Ground Truth | Label

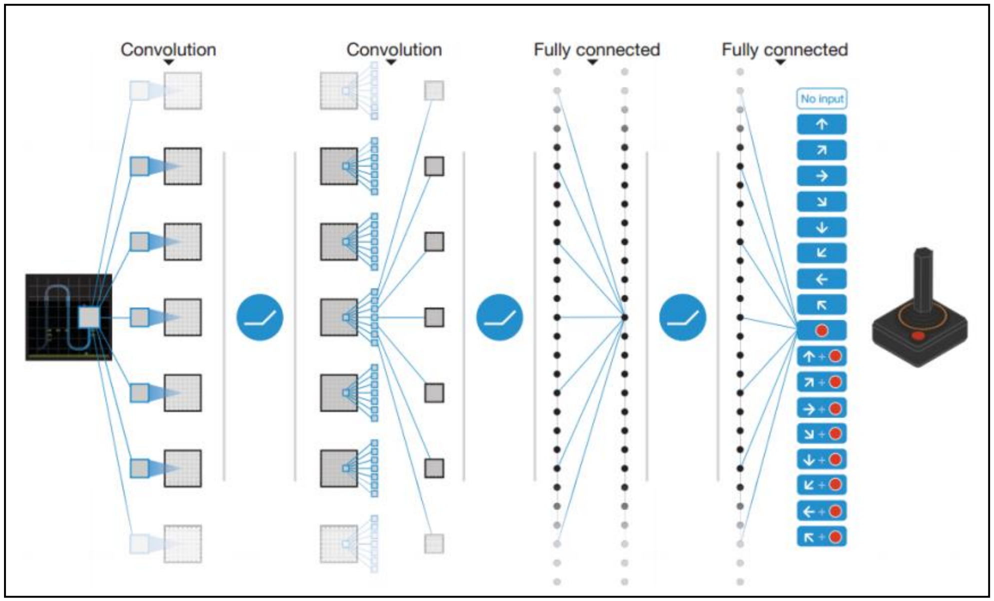


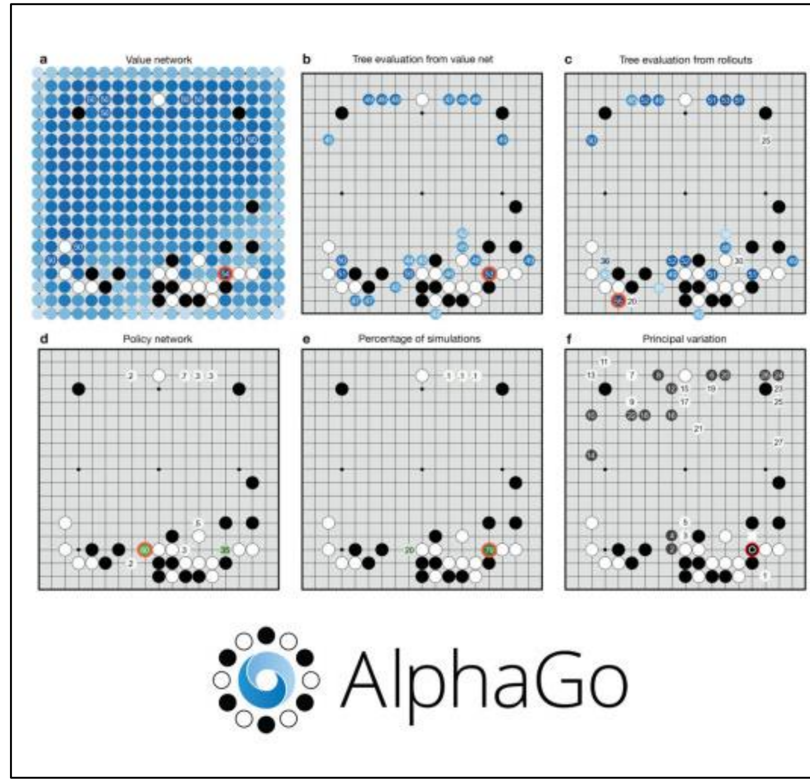
Label

- ✓ Horse
- ✓ Human
- ✓ …

7. Advanced Topics in Deep Learning

5. Deep Reinforcement Learning





References

Deep Learning

- ✓ I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016. www.deeplearningbook.org
- ✓ LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 436–444 (2015)
- ✓ Bilkent University CS 559 Deep Learning Course: www.cs.bilkent.edu.tr/~gcinbis/courses/Spring17/CS559/index.html

Convolutional Neural Networks

- ✓ Stanford University CS231n Convolutional Neural Networks for Visual Recogniton Course: cs231n.stanford.edu
- ✓ J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, and G. Wang, "Recent advances in convolutional neural networks," arXiv preprint arXiv:1512.07108, 2015.
- ✓ A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, 2012.
- ✓ B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a backpropagation network," in NIPS. Citeseer, 1990.
- ✓ C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. CoRR, abs/1409.4842, 2014.
- ✓ M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in ECCV, 2014.
- ✓ J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, 2015.
- ✓ H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In Proc. Int. Conf. Comp. Vis., 2015.

Recurrent Neural Networks

- ✓ Stanford University CS224d Deep Learning for Natural Language Processing Course: cs224d.stanford.edu
- ✓ Understanding LSTM Networks Blog Post: colah.github.io/posts/2015-08-Understanding-LSTMs

Generative Models

- ✓ Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.
- ✓ Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." *arXiv preprint arXiv:1511.06434* (2015).