Grasping Deep Learning from Fundamentals to Applications

June 15, 2023

Lecture 2 – Convolutional Neural Networks (CNNs)

Instructors: **Yufei Huang**, PhD; **Arun Das**, PhD

University of Pittsburgh

nature > review articles > article

nature

Published: 27 May 2015

Deep learning

Yann LeCun, Yoshua Bengio & Geoffrey Hinton

Nature **521**, 436–444(2015) | Cite this article

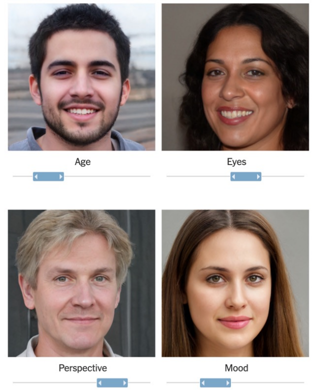
204k Accesses | **13997** Citations | **975** Altmetric | Metrics

Forbes

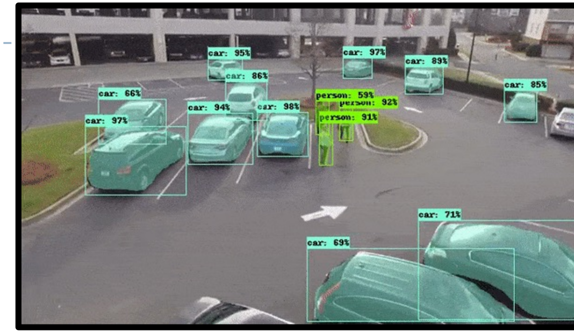
Oct 3, 2021, 07:34pm EDT | 53,450 views

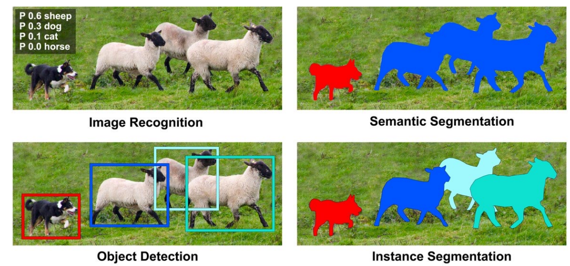
AlphaFold Is The Most Important Achievement In AI —Ever

https://www.forbes.com/sites/robtoews/2021/10/03/alphafold-is-the-most-important-achievement-in-ai-ever/



Deepfake Generation: Forbes Article





37% of tech organizations use AI!

VB VentureBeat

Uber's self-driving AI predicts the trajectories of pedestrians, vehicles, and cyclists

In a paper, Uber researchers describe an autonomous vehicle perception system that reasons about the behavior of pedestrians, vehicles, and ...

World Economic Forum

How AI and machine learning are helping to tackle COVID-19

Organizations have been quick to apply AI and machine-learning in the fight to curb the pandemic - and here are some of the most exciting ...

Test with your own text

This product was very bad!

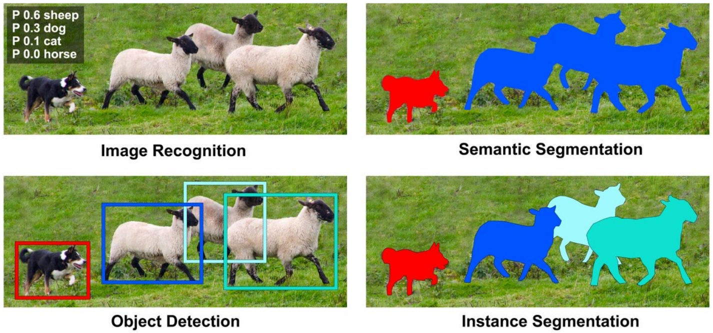
Classify Text

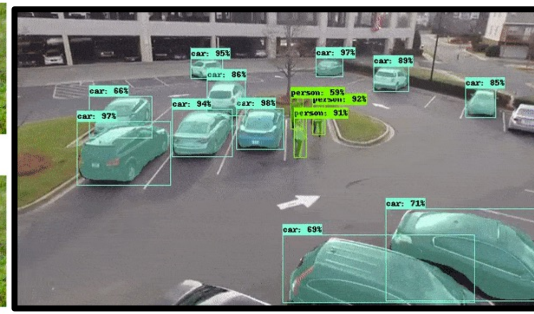
Results

TAG	CONFIDENCE
Negative	99.7%

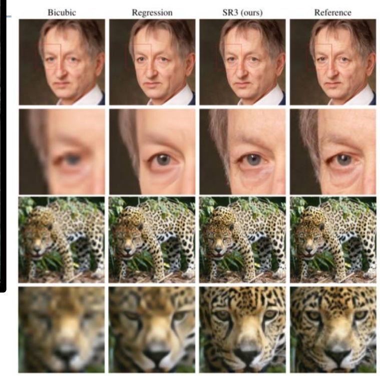
Sentiment Analysis of reviews.

University of Pittsburgh

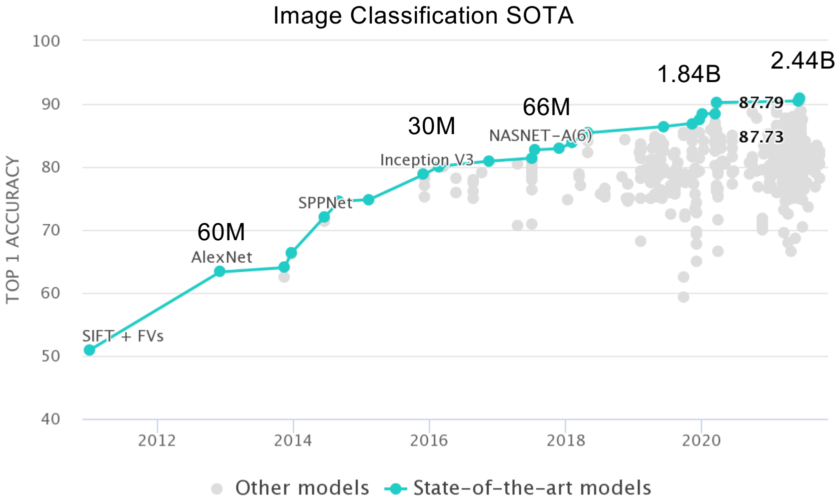




Super Resolution Upscaling



https://arxiv.org/abs/2104.07636

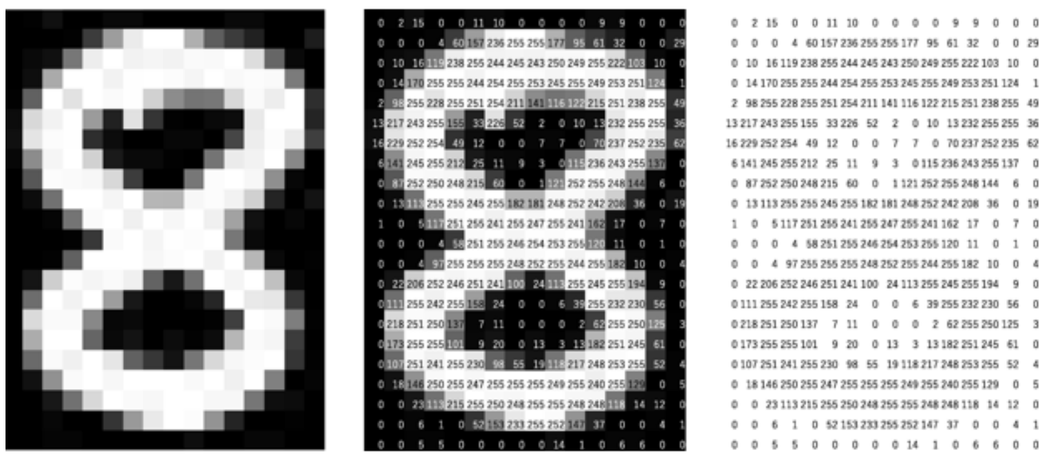


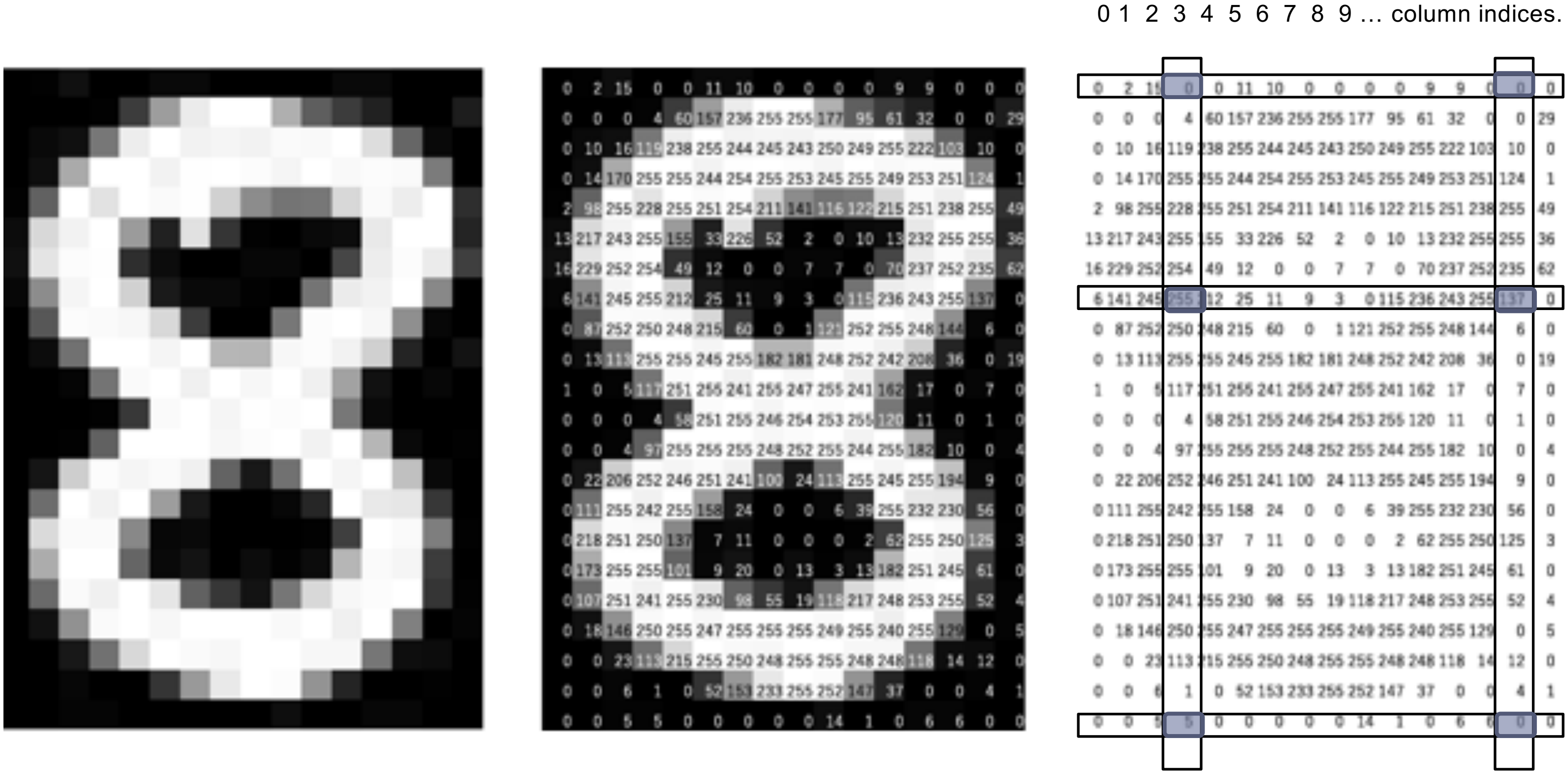
https://thispersondoesnotexist.com

University of Pittsburgh

The AI needs to see!

- Human vision is a complex phenomenon starting with the light rays entering through the cornea of the eye and the visual cortex making sense of the various signals it receives.
- However, computers speak only numbers. Hence, images are represented as numbers, usually in intensities ranging from 0 to 255.




0 1 2 3 4 5 6 7 8 9 … column indices.
0 2 15 0 0 11 10 0 0 0 0 9 9 0 0 0
0 0 0 4 60 157 236 255 255 177 95 61 32 0 0 29
0 10 16 119 238 255 244 245 243 250 249 255 222 103 10 0
0 14 170 255 255 244 254 255 253 245 255 249 253 251 124 1
2 98 255 228 255 251 254 211 141 116 122 215 251 238 255 49
13 217 243 255 155 33 226 52 2 0 10 13 232 255 255 36
16 229 252 254 49 12 0 0 7 7 0 70 237 252 235 62
6 141 245 255 212 25 11 9 3 0 115 236 243 255 137 0
0 87 252 250 248 215 60 0 1 121 252 255 248 144 6 0
0 13 113 255 255 245 255 182 181 248 252 242 208 36 0 19
1 0 5 117 251 255 241 255 247 255 241 162 17 0 7 0
0 0 0 4 58 251 255 246 254 253 255 120 11 0 1 0
0 0 4 97 255 255 255 248 252 255 244 255 182 10 0 4
0 22 206 252 246 251 241 100 24 113 255 245 255 194 9 0
0 111 255 242 255 158 24 0 0 6 39 255 232 230 56 0
0 218 251 250 137 7 11 0 0 0 2 62 255 250 125 3
0 173 255 255 101 9 20 0 13 3 13 182 251 245 61 0
0 107 251 241 255 230 98 55 19 118 217 248 253 255 52 4
0 18 146 250 255 247 255 255 255 249 255 240 255 129 0 5
0 0 23 113 215 255 250 248 255 255 248 248 118 14 12 0
0 0 6 1 0 52 153 233 255 252 147 37 0 0 4 1
0 0 5 5 0 0 0 0 0 14 1 0 6 6 0 0

Challenges with learning images

Sun flower?

Problems:

- High dimensional input
  - 150 x 150 pixels x 3 (RGB) = 67,500
- 2D correlations
- Operational invariance
  - Scale, translation, etc

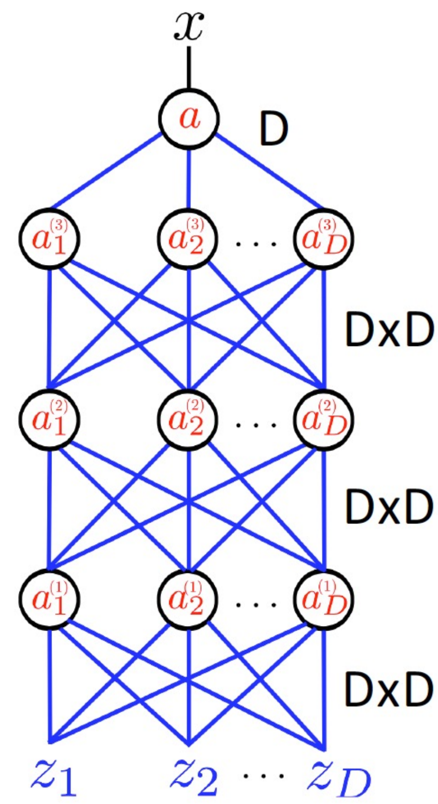
Very hard to train with DNN!

Number of parameters = 3 x (D x D) + D

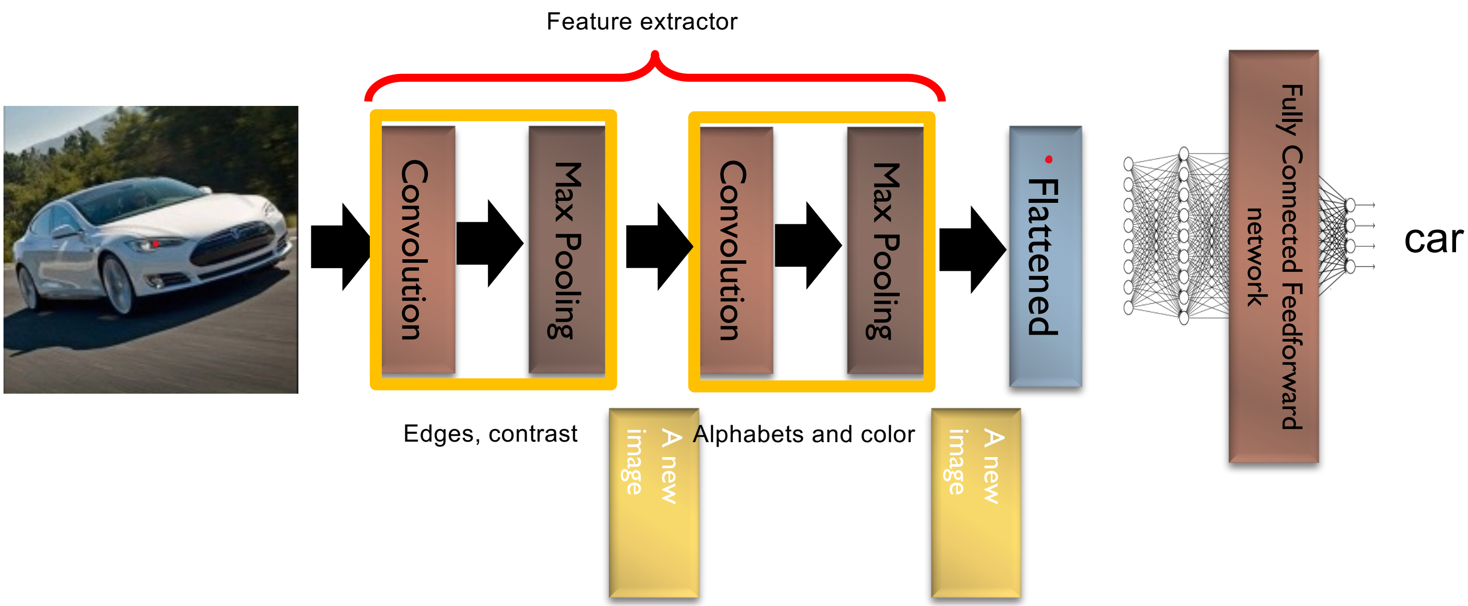
To feed images to FCN (DNN), we can flatten the images.

For a 32x32 image, D=1024.

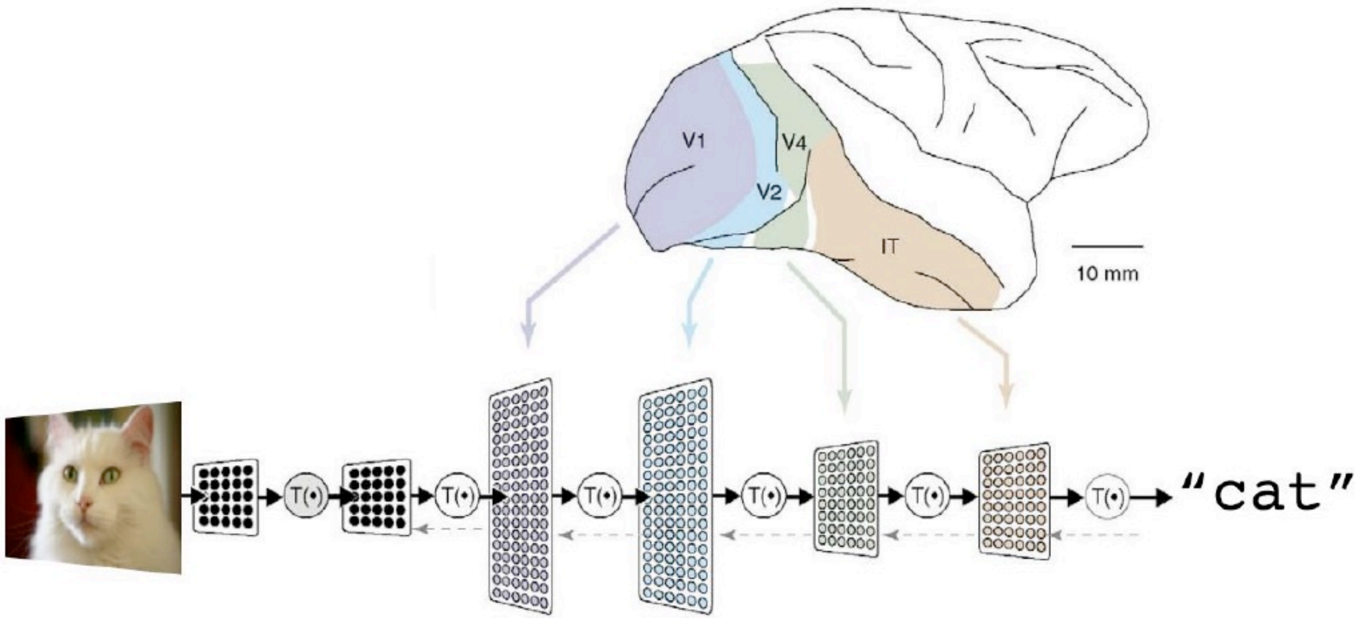
Number of parameters = 3 x (1024 x 1024) + 1024 = ~ 3×10^6



Convolutional neural networks (CNNs)

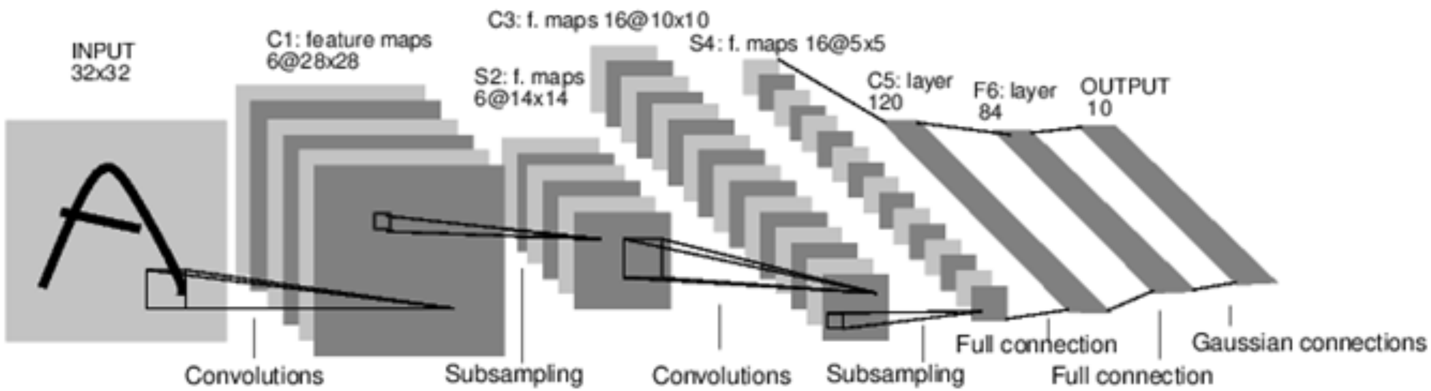


Hierarchical Architecture of the mammalian visual cortex



- Ventral (recognition) pathway in the visual cortex has multiple stages
  Retina - LGN - V1 - V2 - V4 - PIT - AIT ....
- It's hierarchical
- There is local processing

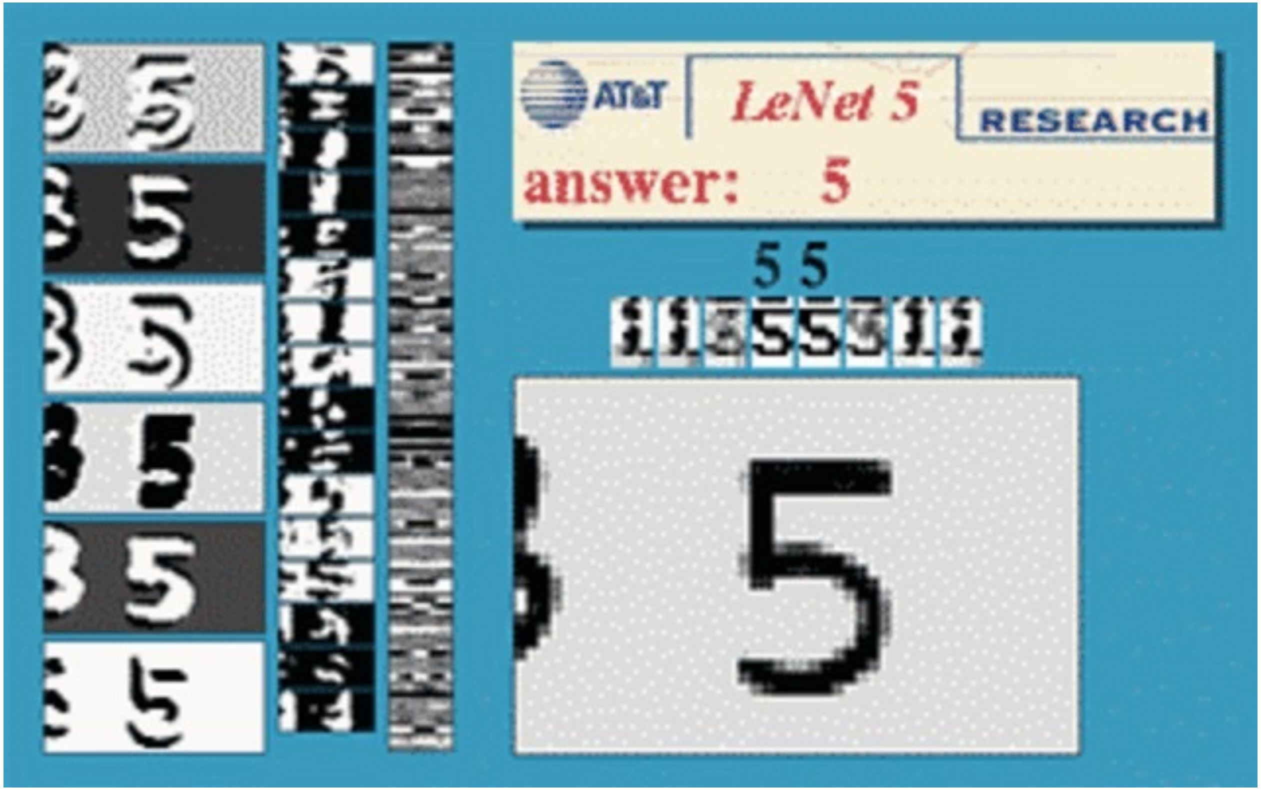
LeNet (1989)



A Full Convolutional Neural Network (LeNet)

LeNet1 Demo from 1993

- Running on a 486 PC with an AT&T DSP32C add-on board (20 Mflops!)



AT&T
LeNet 5
RESEARCH
answer: 5
5 5


AT&T
LeNet 5
RESEARCH
answer: 384
3 3 8 8 4 4

Why CNN now? A: ImageNet and GPU

- The ImageNet dataset [Fei-Fei et al. 2012]
  - 1.5 million training samples
  - 1000 categories

- NVIDIA Graphical Processing Units (GPU)
  - Capable of 1 trillion operations/second

Sea lion

Flute

Strawberry

Backpac
k

Racket

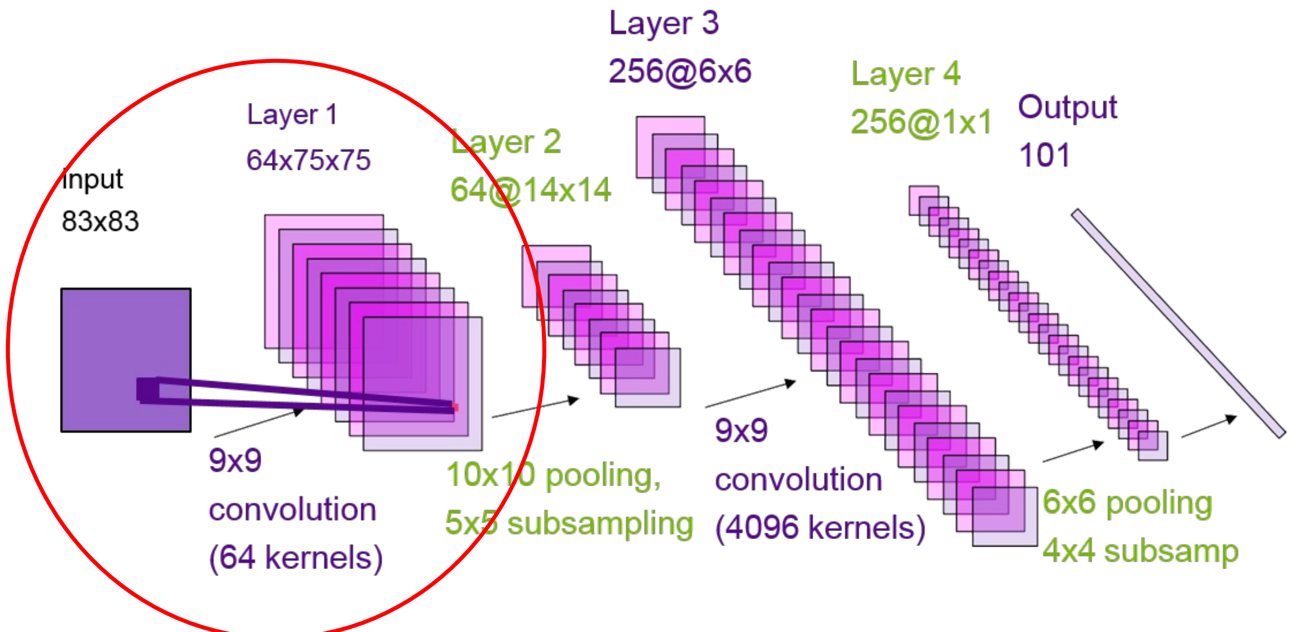
ImageNet large-scale visual recognition challenge (ILSVRC)

– The ImageNet dataset
  - 1.5 million training samples of size 224x224x3
  - 1000 fine-grained categories (breeds of dogs....)

CNN ingredients

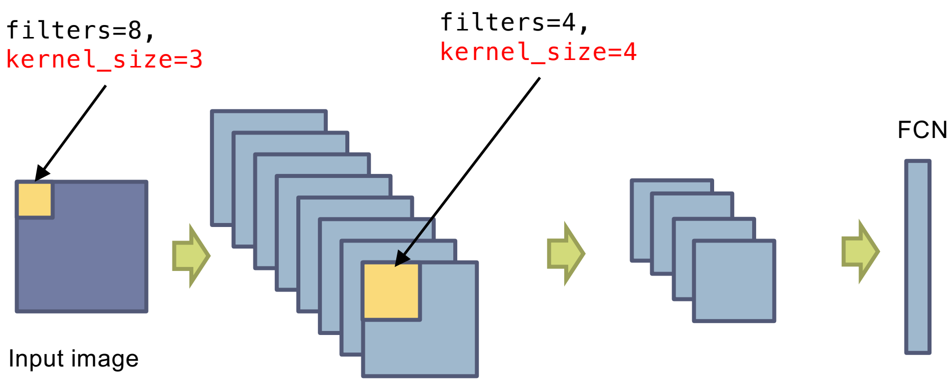
- Convolutional filters
  - local connectivity
  - parameter sharing

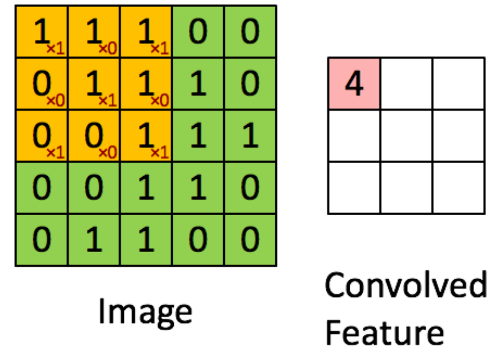
- Pooling/subsampling hidden units



University of Pittsburgh

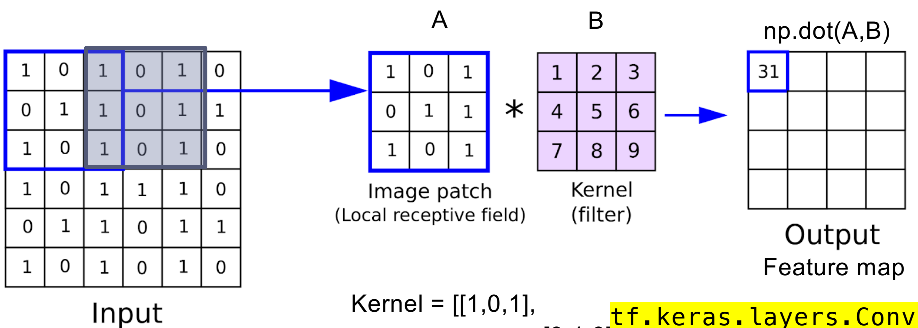
Convolution filters





8 feature maps.
Size of feature map -> parameters we set for the kernel.

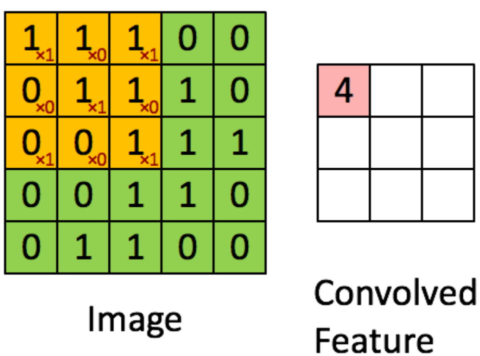
Elementwise
Multiply and Add

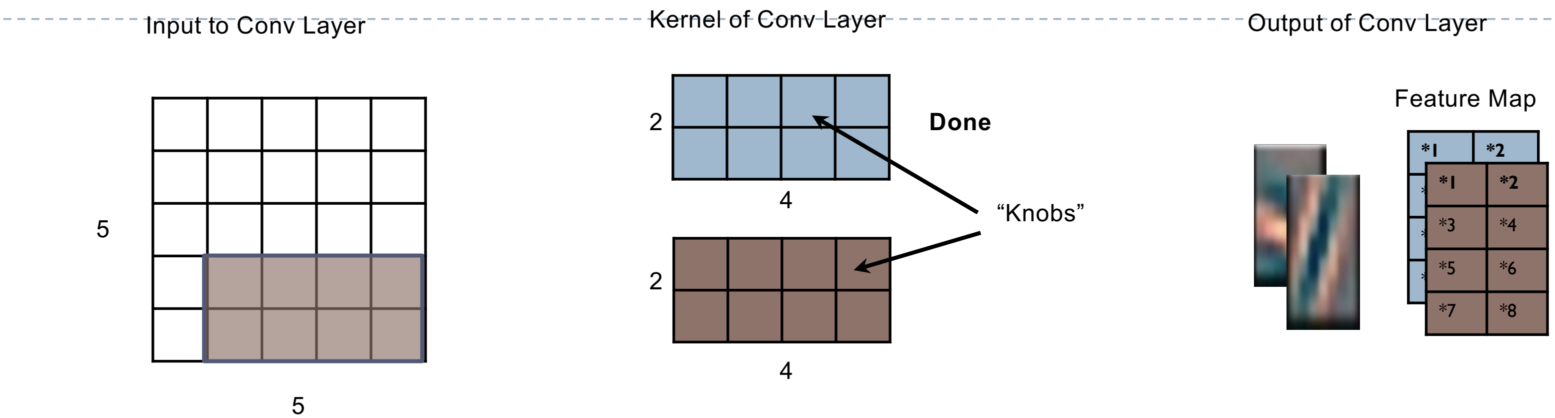


(height, width, channels)

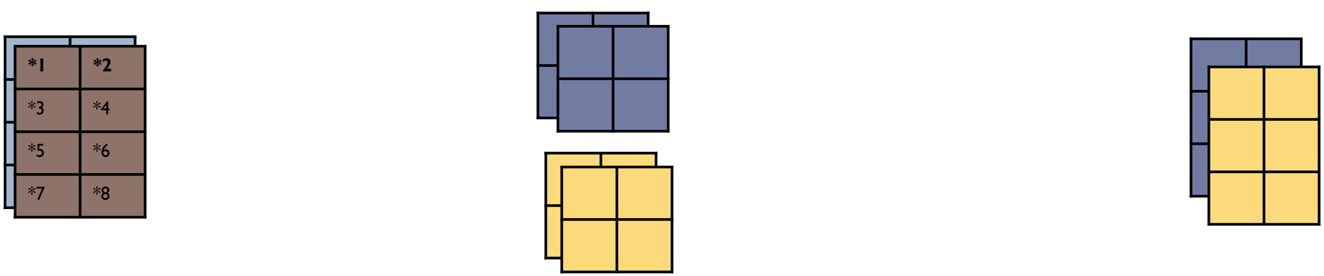
Kernel = [[1,0,1],
[0,1,0]
[1,0,1]]

```
tf.keras.layers.Conv2D(filters=1, kernel_size=(4,2),
                padding='same', activation='relu’,
input_shape=(5,5,1)),
```



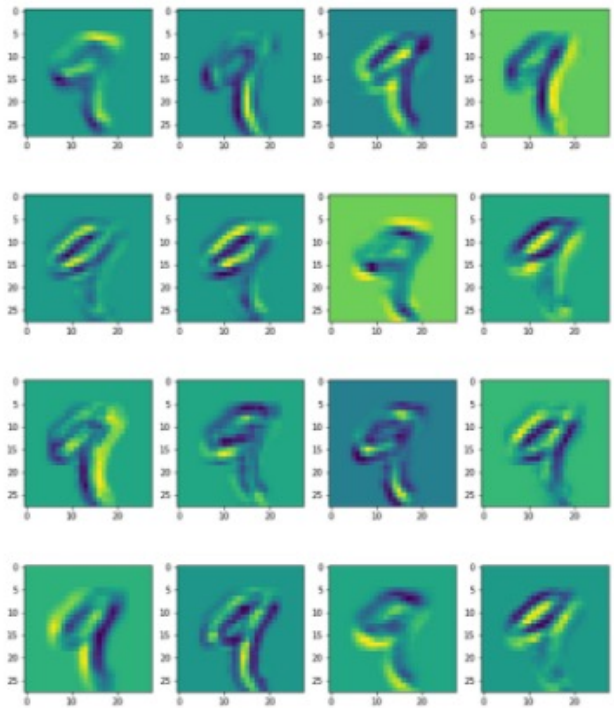



```
tf.keras.layers.Conv2D(filters=2, kernel_size=(4,2),
                       padding='same', activation='relu’,
input_shape=(5,5,1)),
```

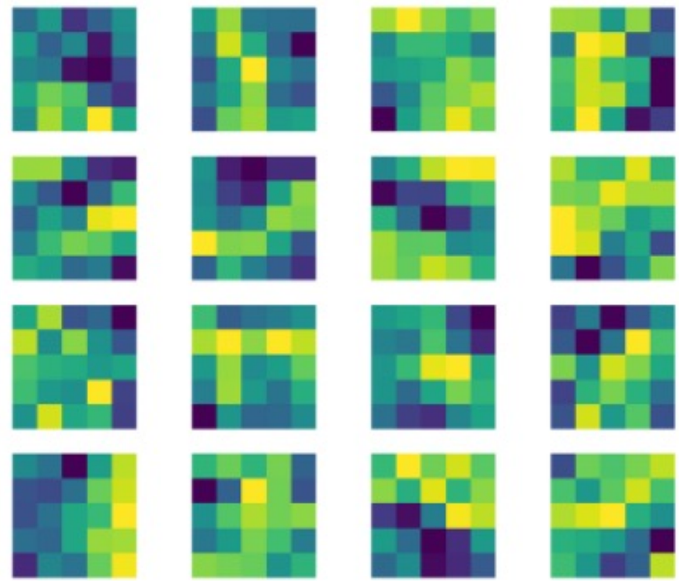



```
tf.keras.layers.Conv2D(filters=2, kernel_size=2,
                       padding='same', activation='relu’)
```

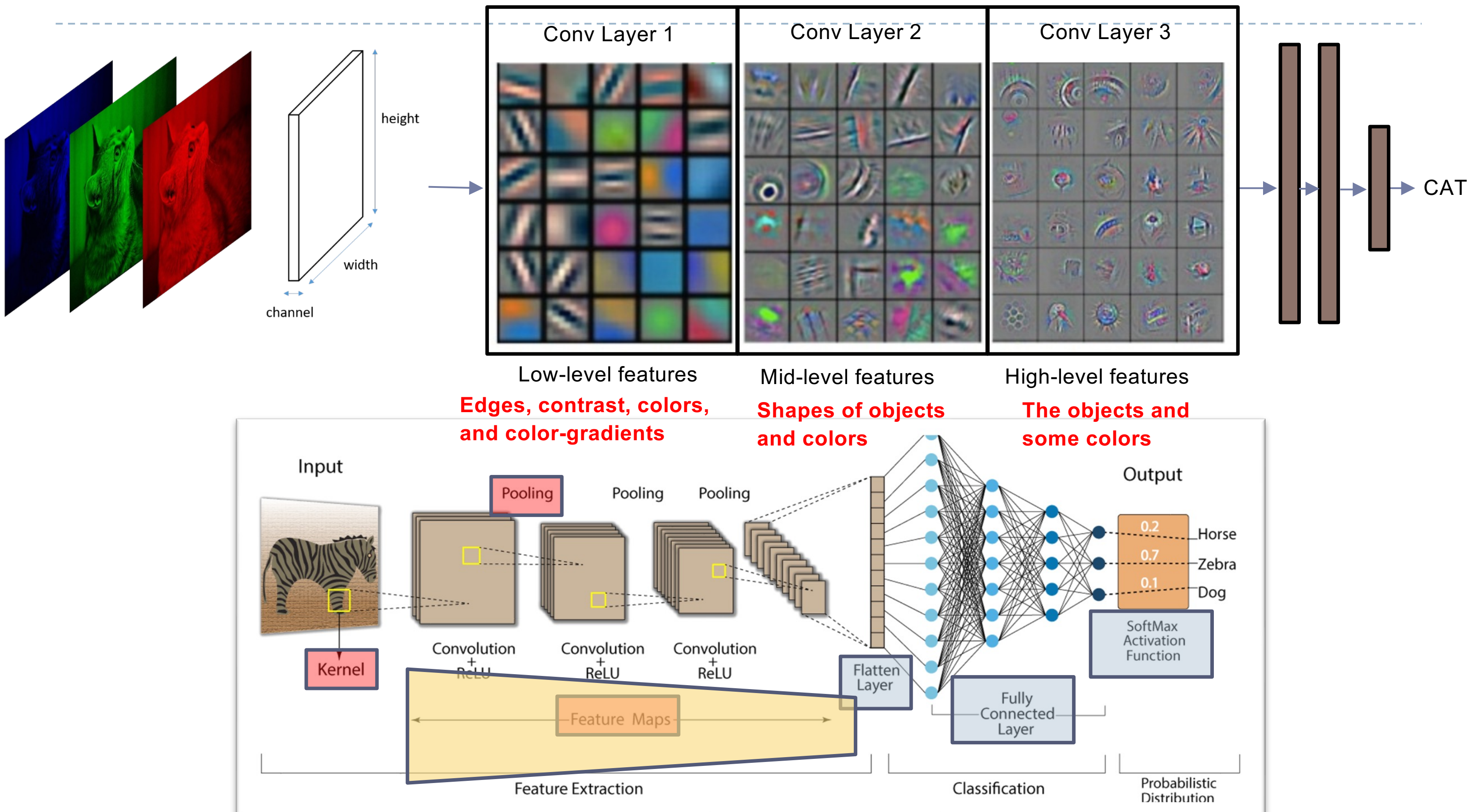
Feature map



Filters

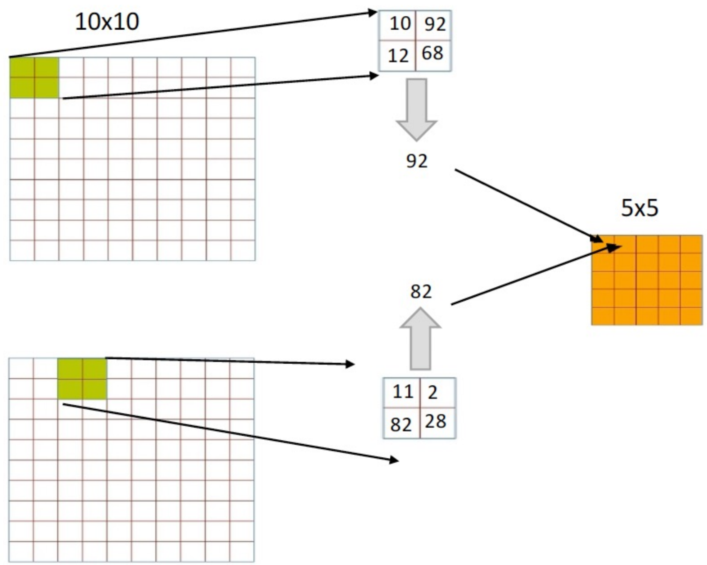


```
filters=16,
kernel_size=5
```



https://www.analyticsvidhya.com/blog/2021/05/20-questions-to-test-your-skills-on-cnn-convolutional-neural-netw

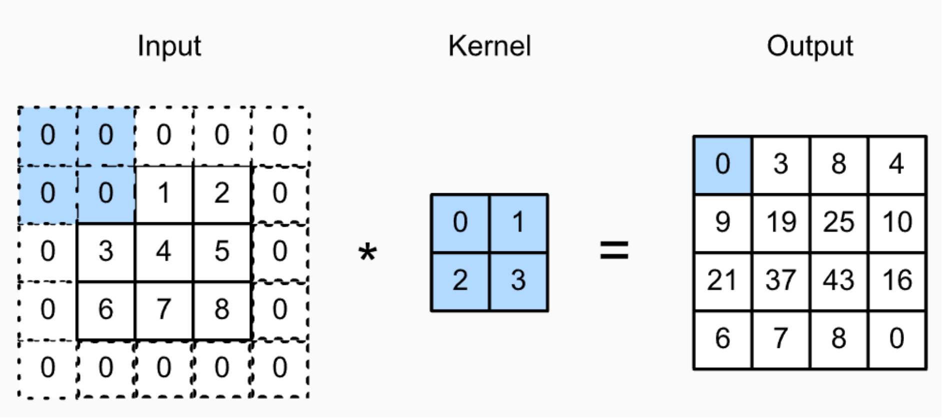
Pooling



Example of Max Pooling.

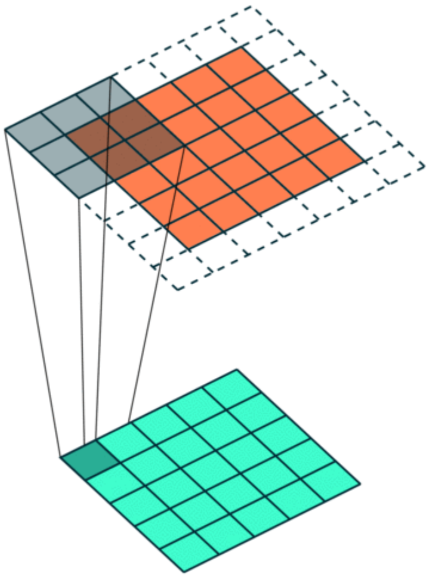
University of Pittsburgh

Padding

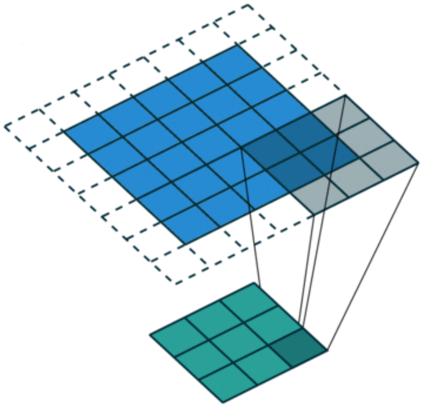


https://classic.d2l.ai/chapter_convolutional-neural-networks/padding-and-strides.html

Strides



Strides = 1

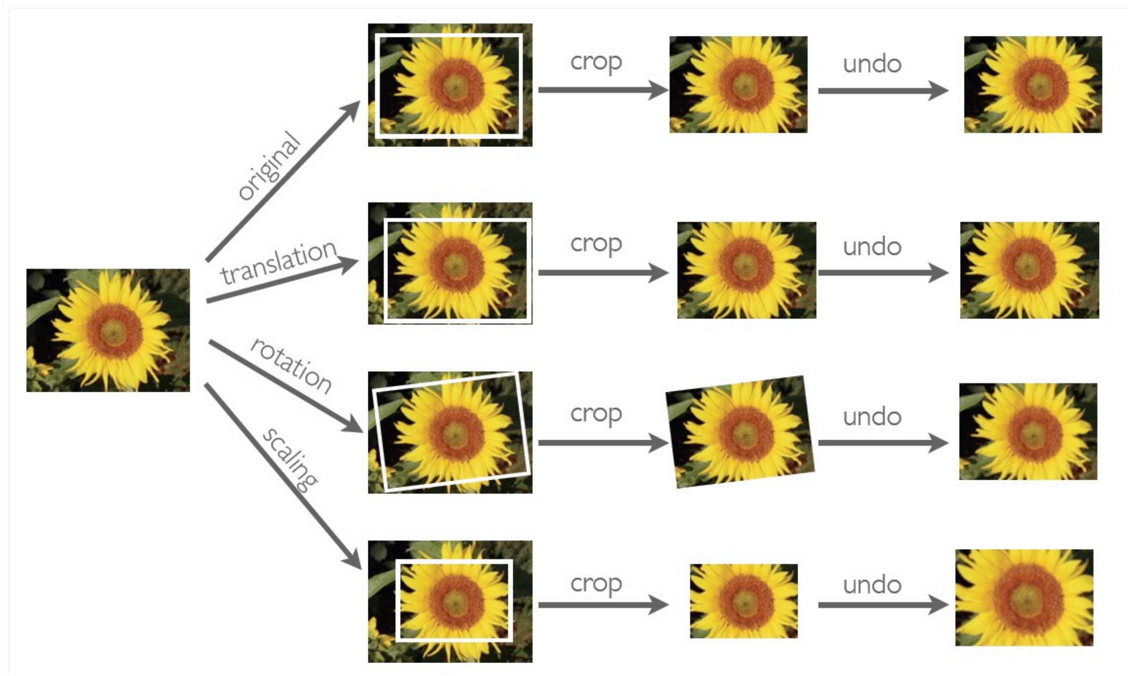


Strides = 2

Stride is how much we move the kernels forward at each step during the convolution operation. When the stride is 1 then we move the filters one pixel at a time. When the stride is 2 then the filters jump 2 pixels at a time as we slide them around. This will produce smaller output volumes spatially.

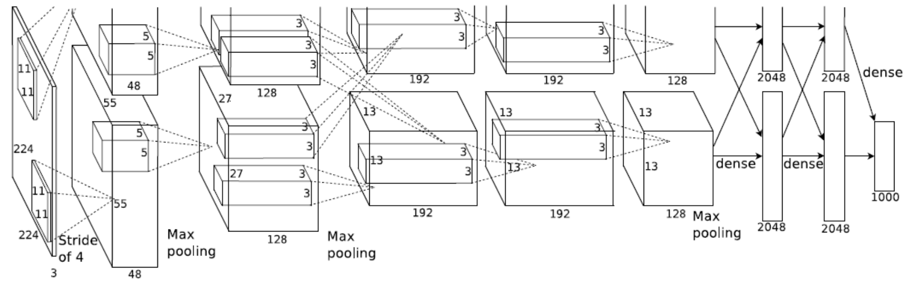
Data augmentation

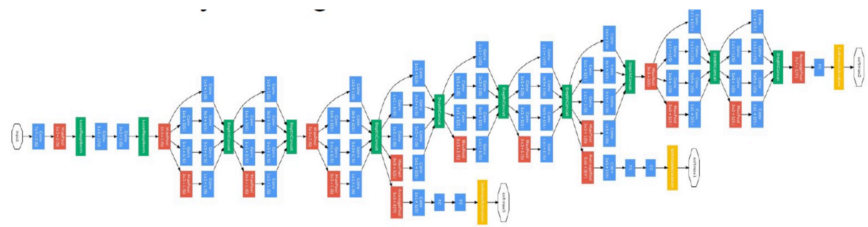
- Goal: introduce scale and rotational invariance
- How? Generate artificial images

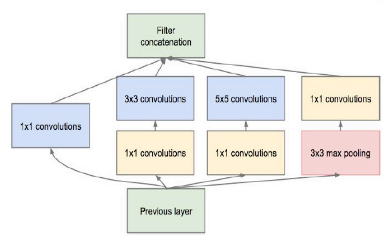


Different CNNs

- AlexNet
- VGGNet
- Inception model
- ResNet
- …







Inception module

ILSVRC 2014 winner (6.7% top 5 error)

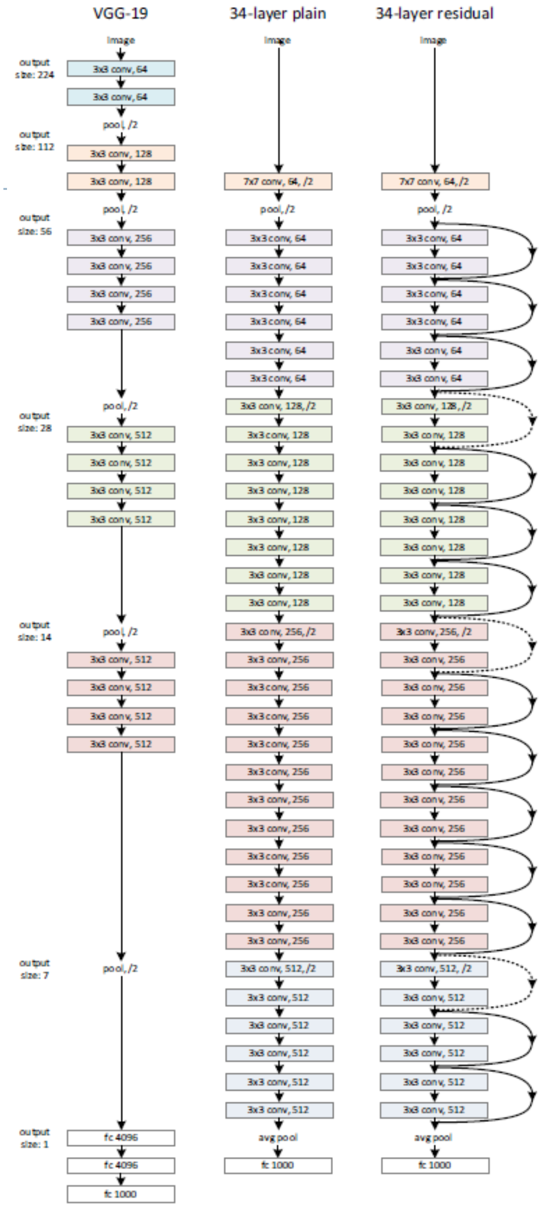
ResNet (He et al, 2015)

ILSVRC 2015 winner (3.6% top 5 error)

- **1st places in all five main tracks**
  - ImageNet Classification: *"Ultra-deep"* (quote Yann) 152-layer nets
  - ImageNet Detection: 16% better than 2nd
  - ImageNet Localization: 27% better than 2nd
  - COCO Detection: 11% better than 2nd
  - COCO Segmentation: 12% better than 2nd

152 layers!!!

25.5M parameters



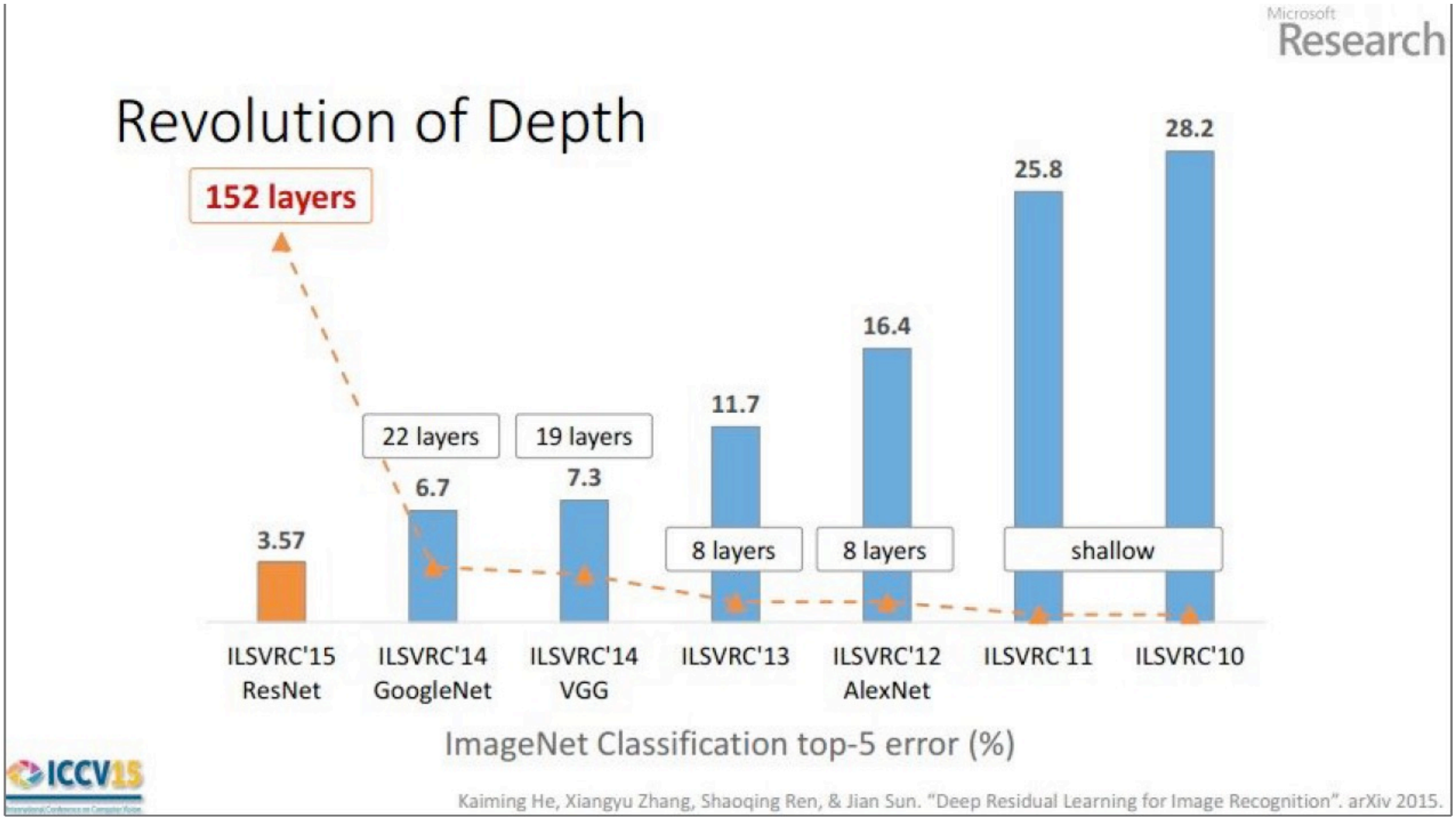
method	top-1 err.	top-5 err.
VGG [41] (ILSVRC'14)	-	8.43†
GoogLeNet [44] (ILSVRC'14)	-	7.89
VGG [41] (v5)	24.4	7.1
PReLU-net [13]	21.59	5.71
BN-inception [16]	21.99	5.81
ResNet-34 B	21.84	5.71
ResNet-34 C	21.53	5.60
ResNet-50	20.74	5.25
ResNet-101	19.87	4.60
ResNet-152	**19.38**	**4.49**

Table 4. Error rates (%) of **single-model** results on the ImageNet validation set (except † reported on the test set).

University of Pittsburgh

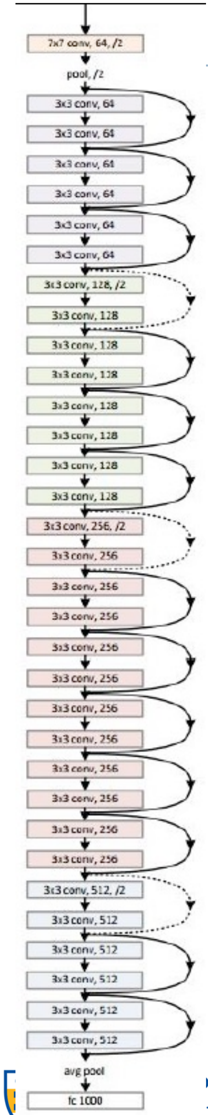
ResNet (He et al, 2015)

ILSVRC 2015 winner (3.6% top 5 error)



University of Pittsburgh

ResNet (He et al, 2015)

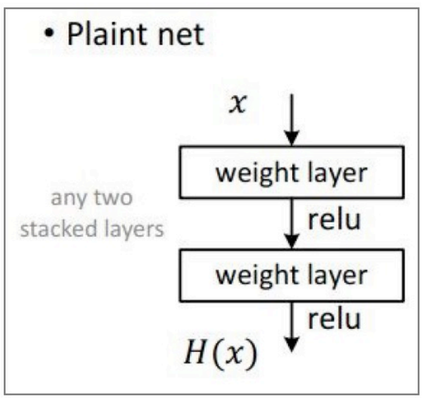


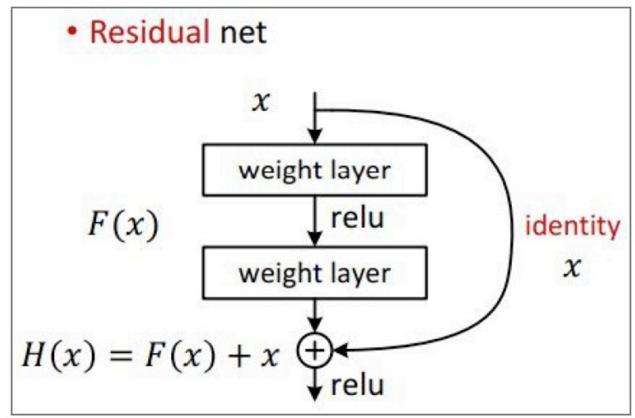
Deep Residual Learning for Image Recognition

Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun

Microsoft Research

{kahe, v-xiangz, v-shren, jiansun}@microsoft.com

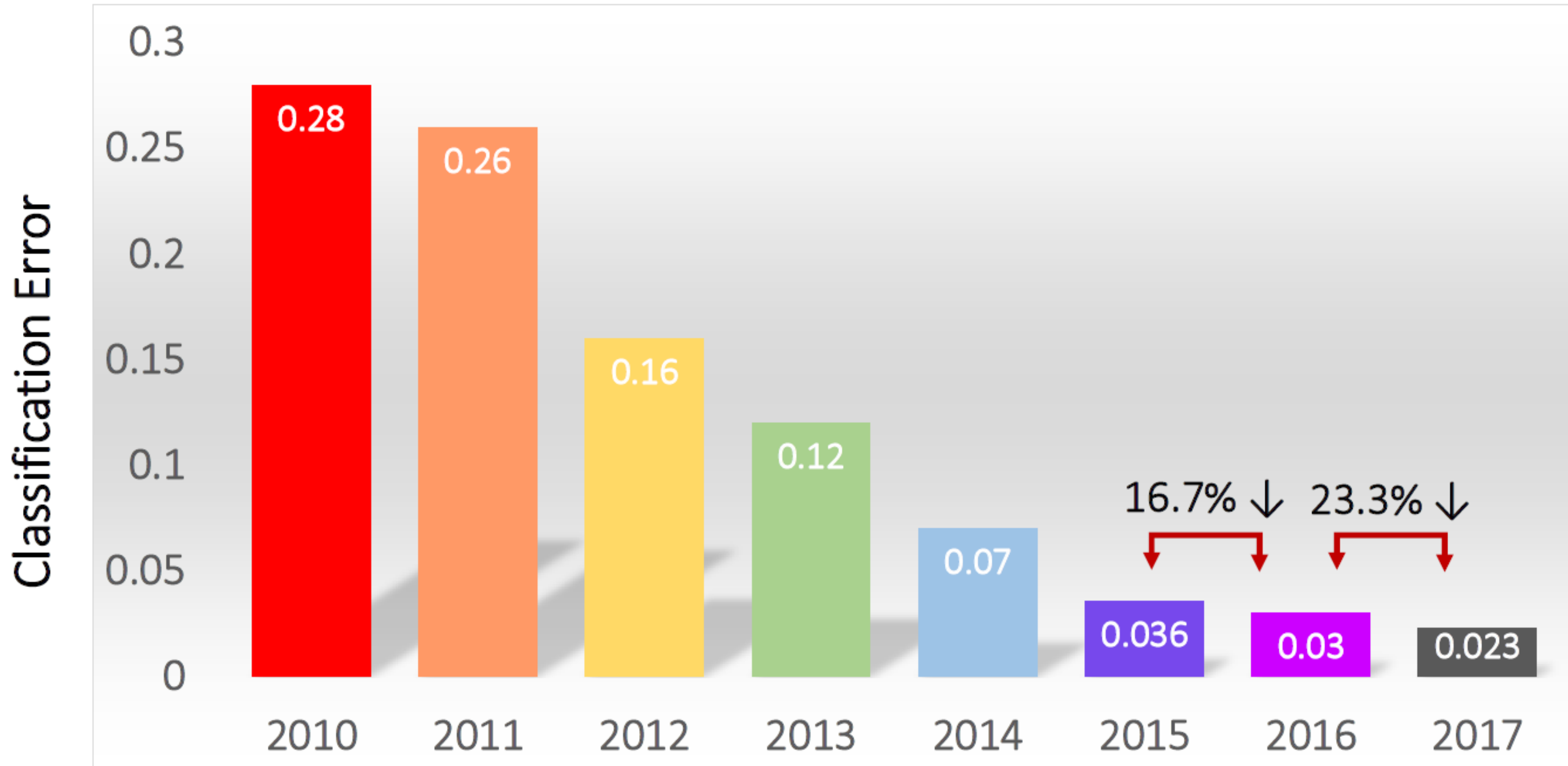




Why does it work?

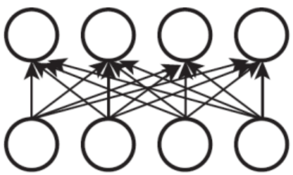
- The “identity” path preserve the gradient!

Results of 2017

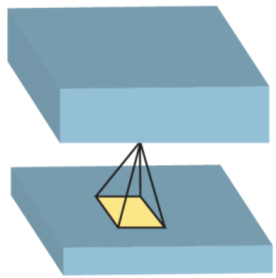


Deep learning modules

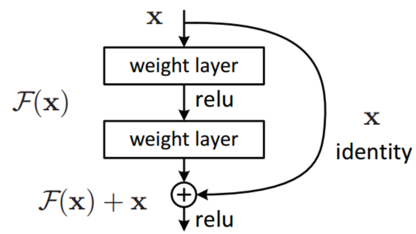
Dense layers



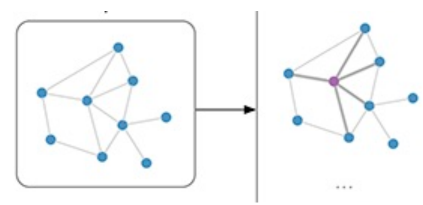
Convolutional layer



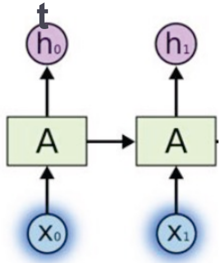
Residual layer



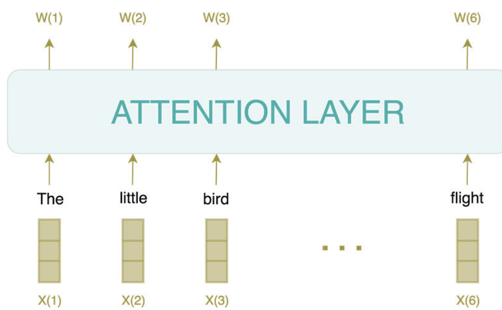
Graph convolution



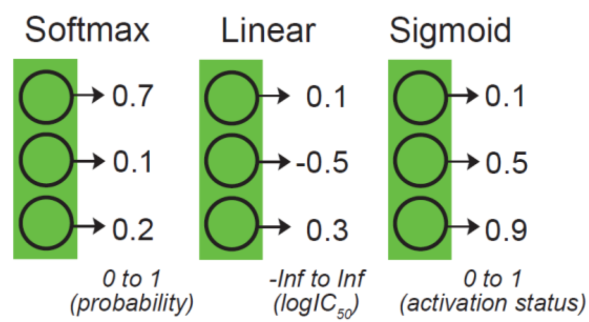
Recurrent



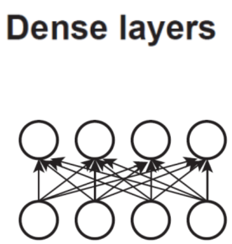
Attention layer

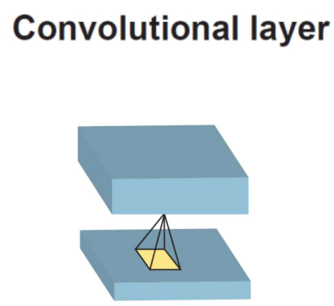


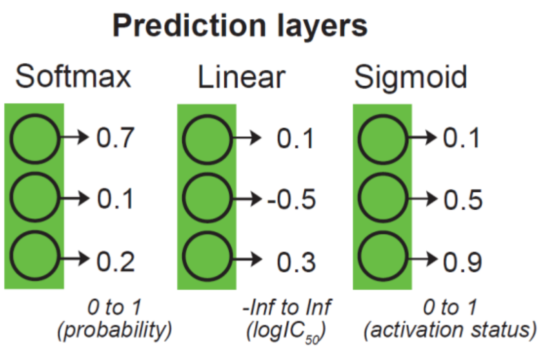
Prediction layers

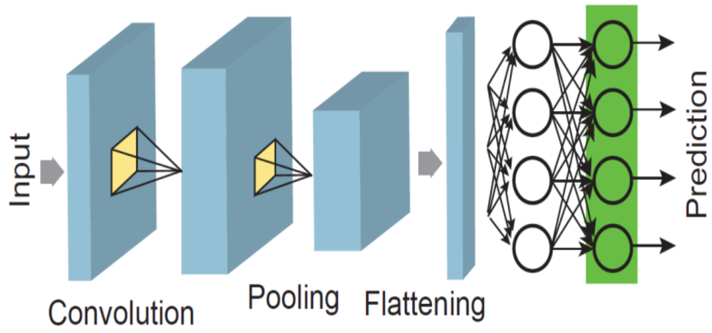


Building a convolution neural network (CNN)



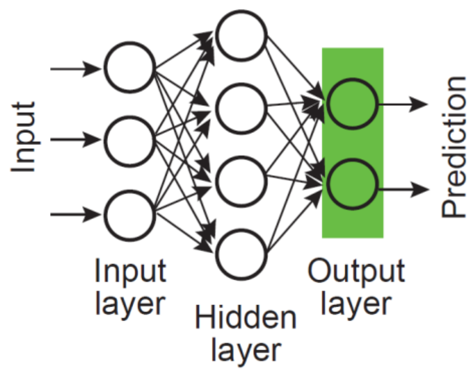




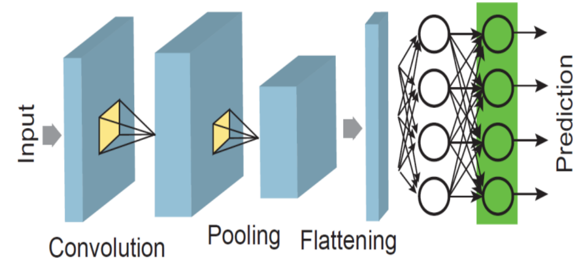


Supervised deep learning models

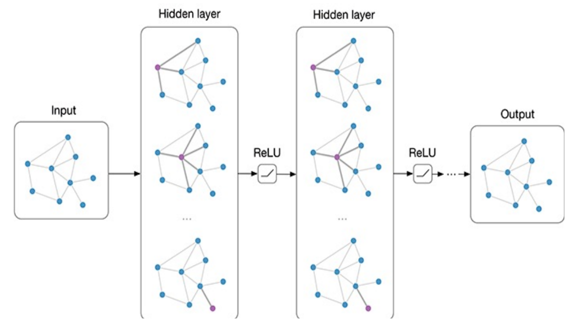
DNN



Convolutional neural networks (CNN)

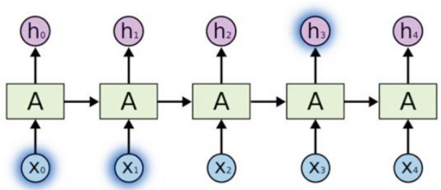


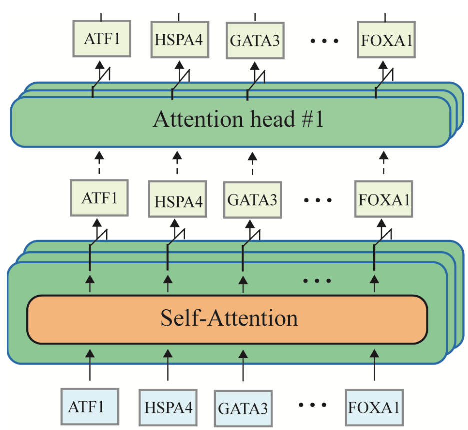
Graph CNN



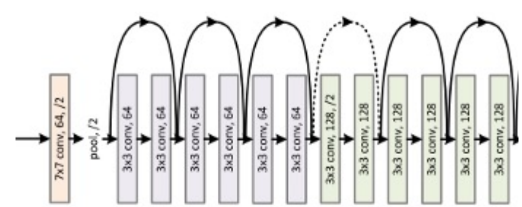
Transformer

RNN (LSTM, GRU)



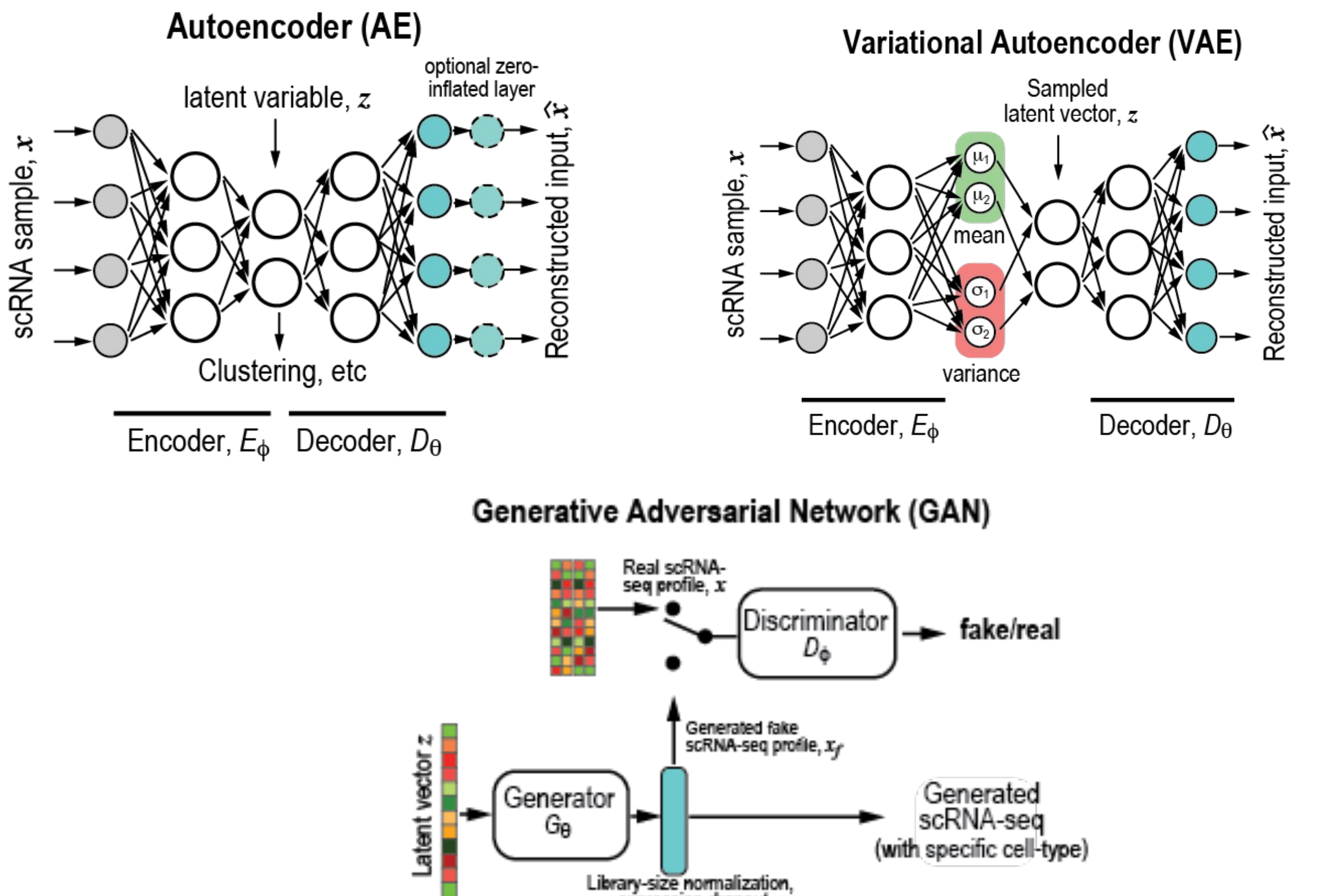


ResNet



University of Pittsburgh

Unsupervised deep learning models



University of Pittsburgh