AI in Built Environment DCP4300

Lec09: Computer Vision

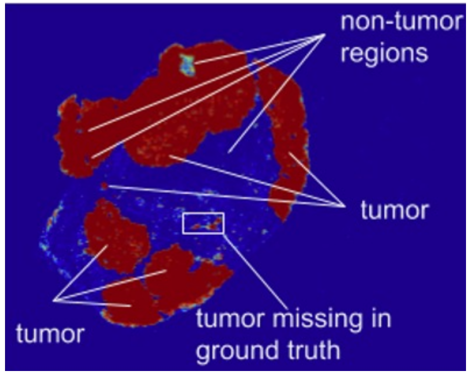
Part A

Dr. Chaofeng Wang
Jianhao Gao (TA)
University of Florida
College of Design Construction and Planning

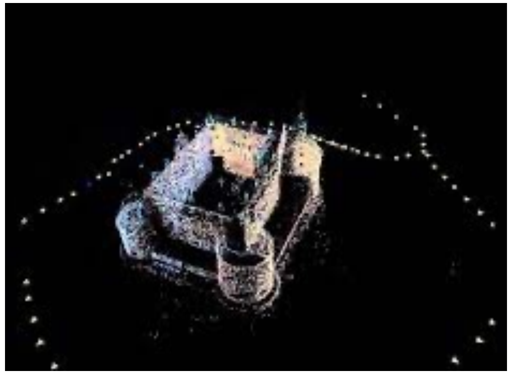
Applications of Computer Vision

Face recognition

Autonomous driving



Disease diagnosis



3D reconstruction

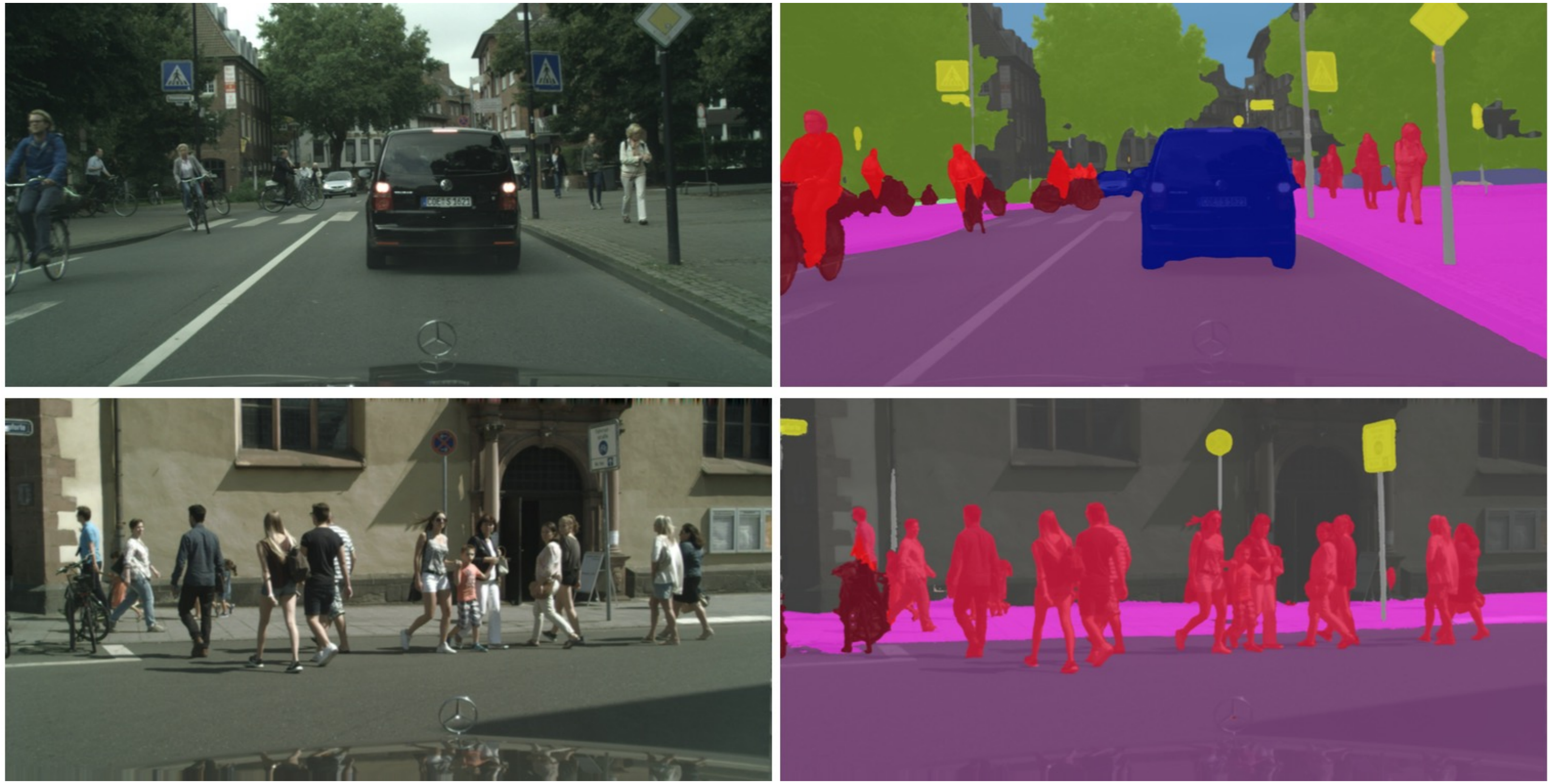
…

Applications of Computer Vision

Classification

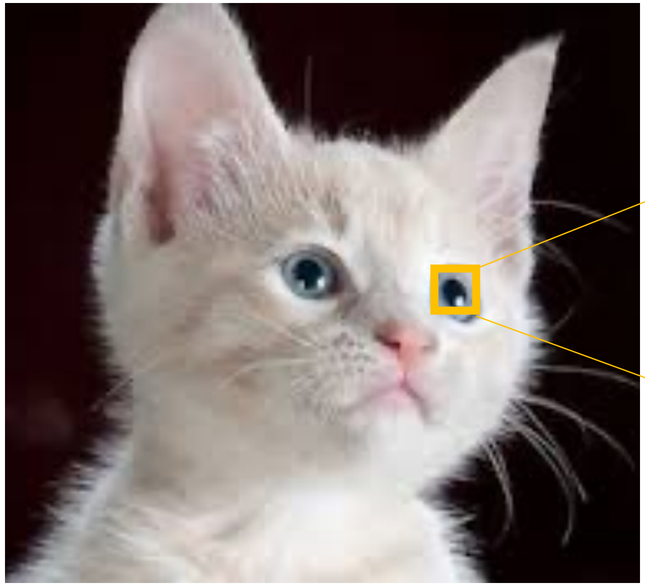
Applications of Computer Vision

Segmentation



Images

What we see



What the computer sees

$$\begin{bmatrix} 1 & 5 & 10 & 5 \\ 6 & 4 & 12 & 4 \\ 10 & 5 & 12 & 11 \\ 5 & 11 & 23 & 9 \end{bmatrix}$$

A 256x256 RGB image is a 256x256x3 matrix

Images

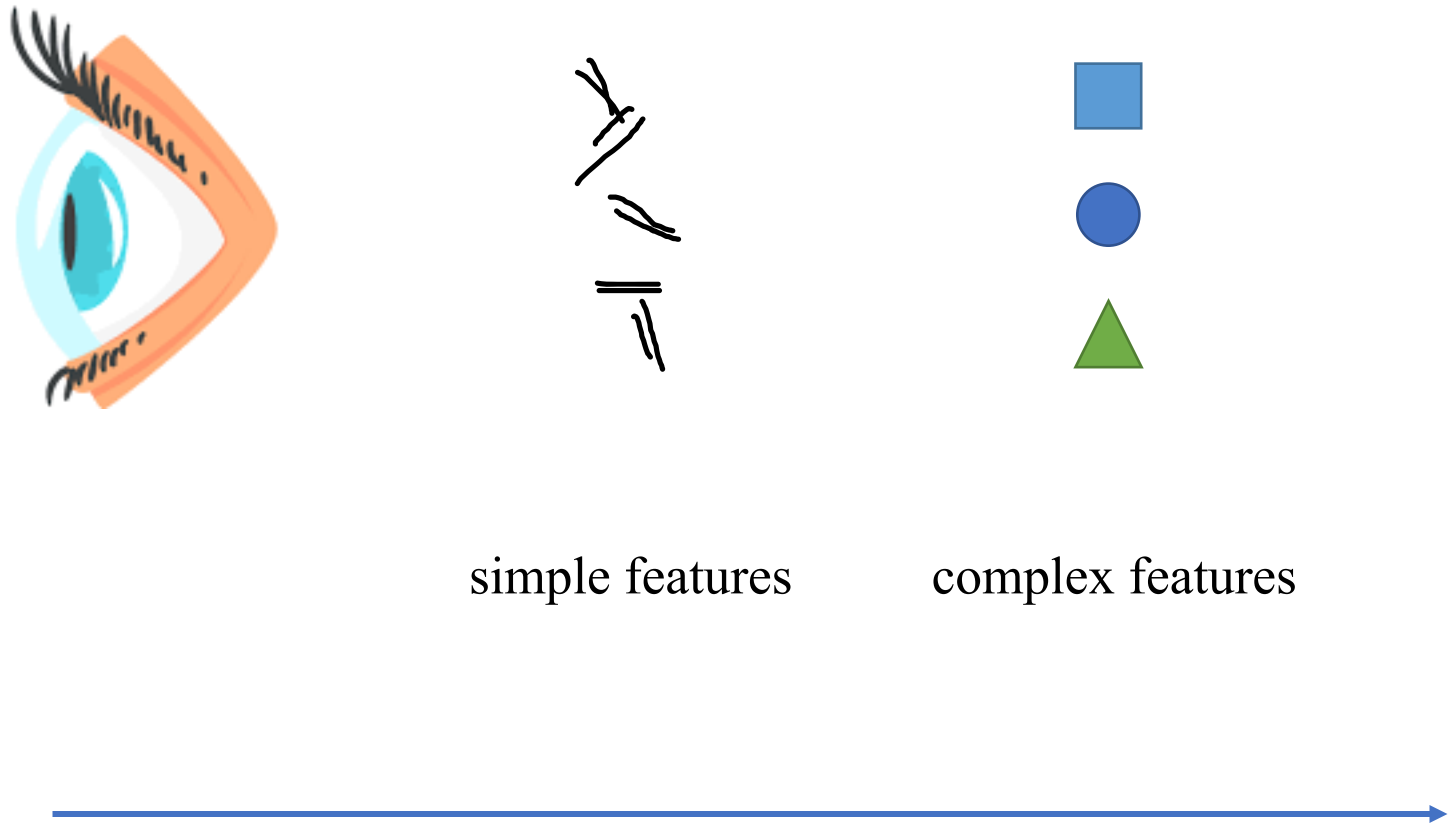
Visual Illusion: https://www.youtube.com/watch?v=9Gw23ayxY-I

Images

Vision

It started from the research on cat's vision system

Vision



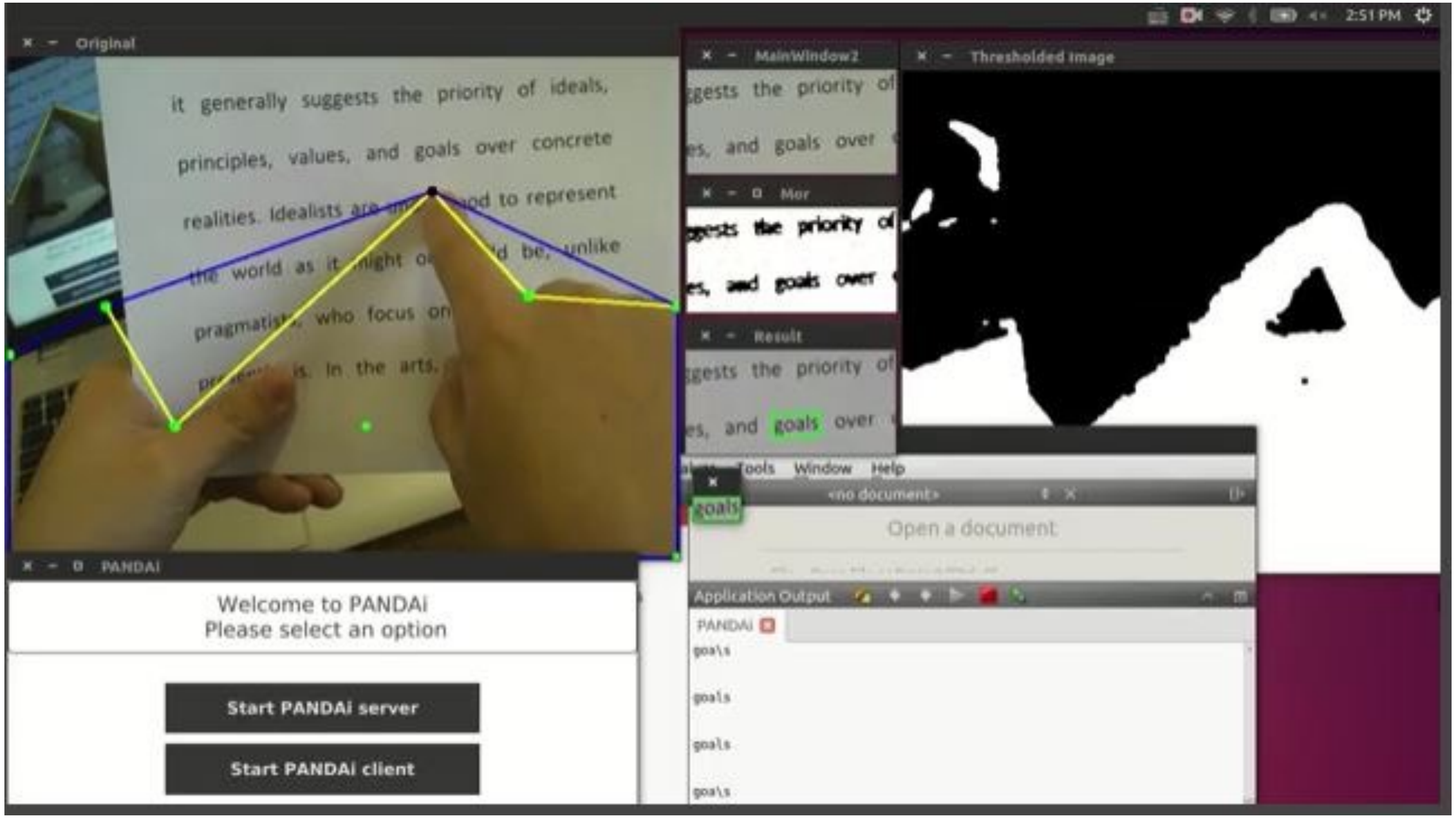
The early-stage computer vision

- Edge Detection
- Dilation, Erosion
- Perspective Transformation
- Cropping
- Scaling, Interpolations, And Re-Sizing
- Thresholding
- Sharpening
- Blurring
- Contours
- Line Detection
- Blob Detection
- …

OpenCV

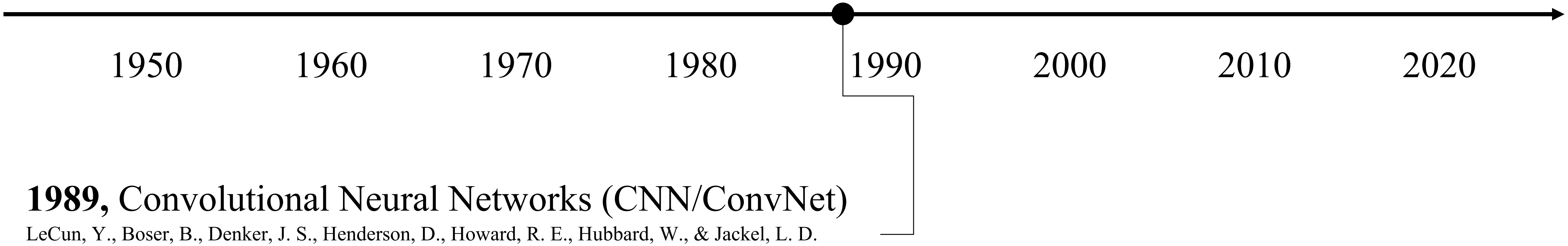
https://opencv.org/

The early-stage computer vision



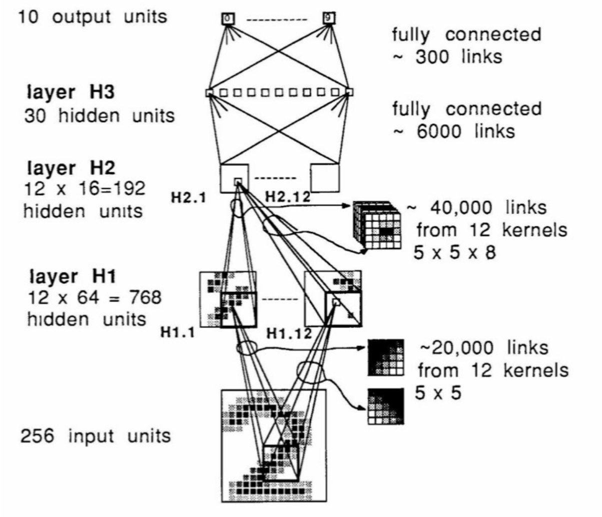
My first computer vision project

Key points in the history of computer vision



1989, Convolutional Neural Networks (CNN/ConvNet)

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4), 541-551.



Yann LeCun demonstrating LeNet 1, 1993

Key points in the history of computer vision

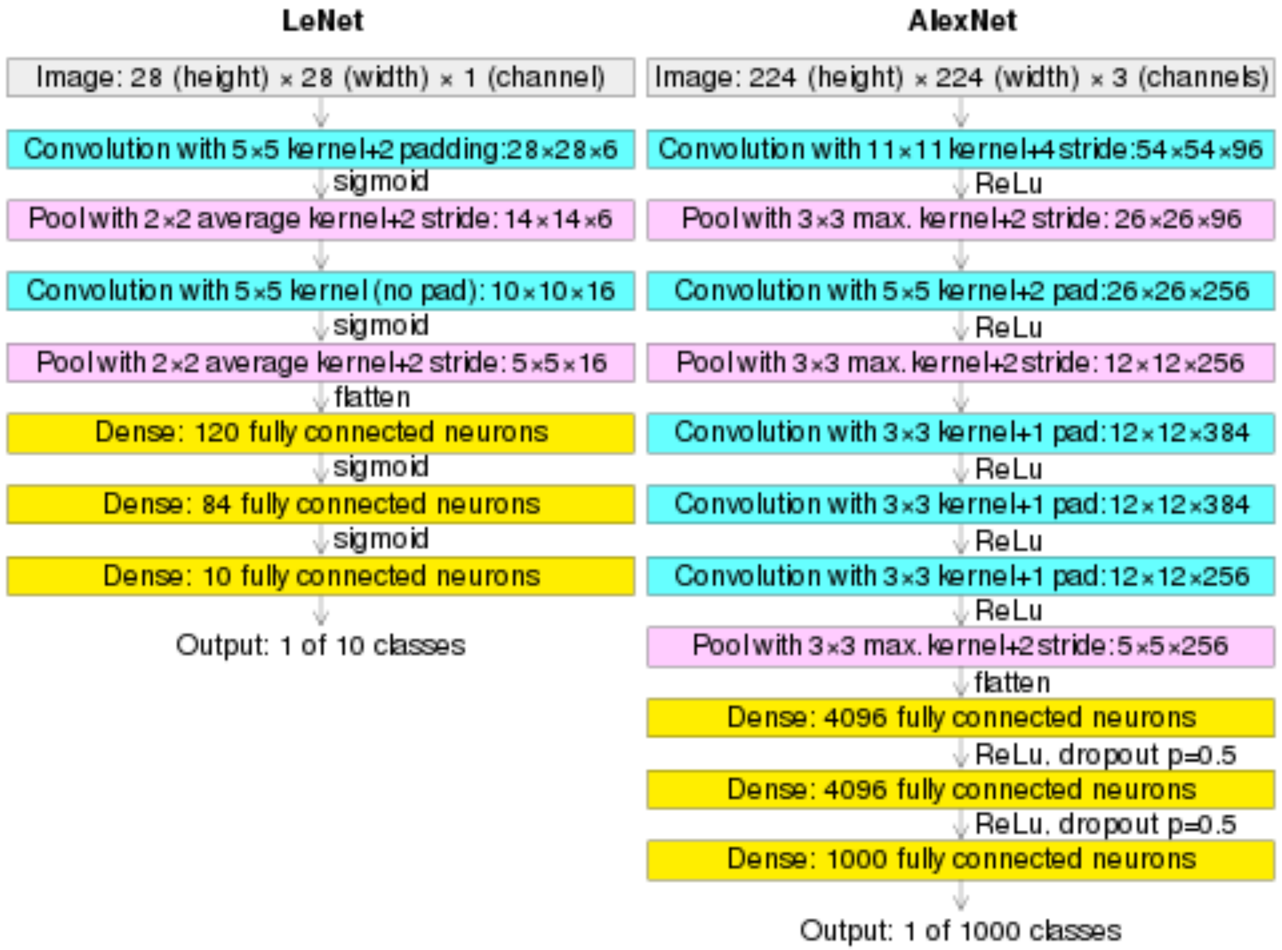
1950 1960 1970 1980 1990 2000 2010 2020

1989, Convolutional Neural Networks (CNN/ConvNet)

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4), 541-551.

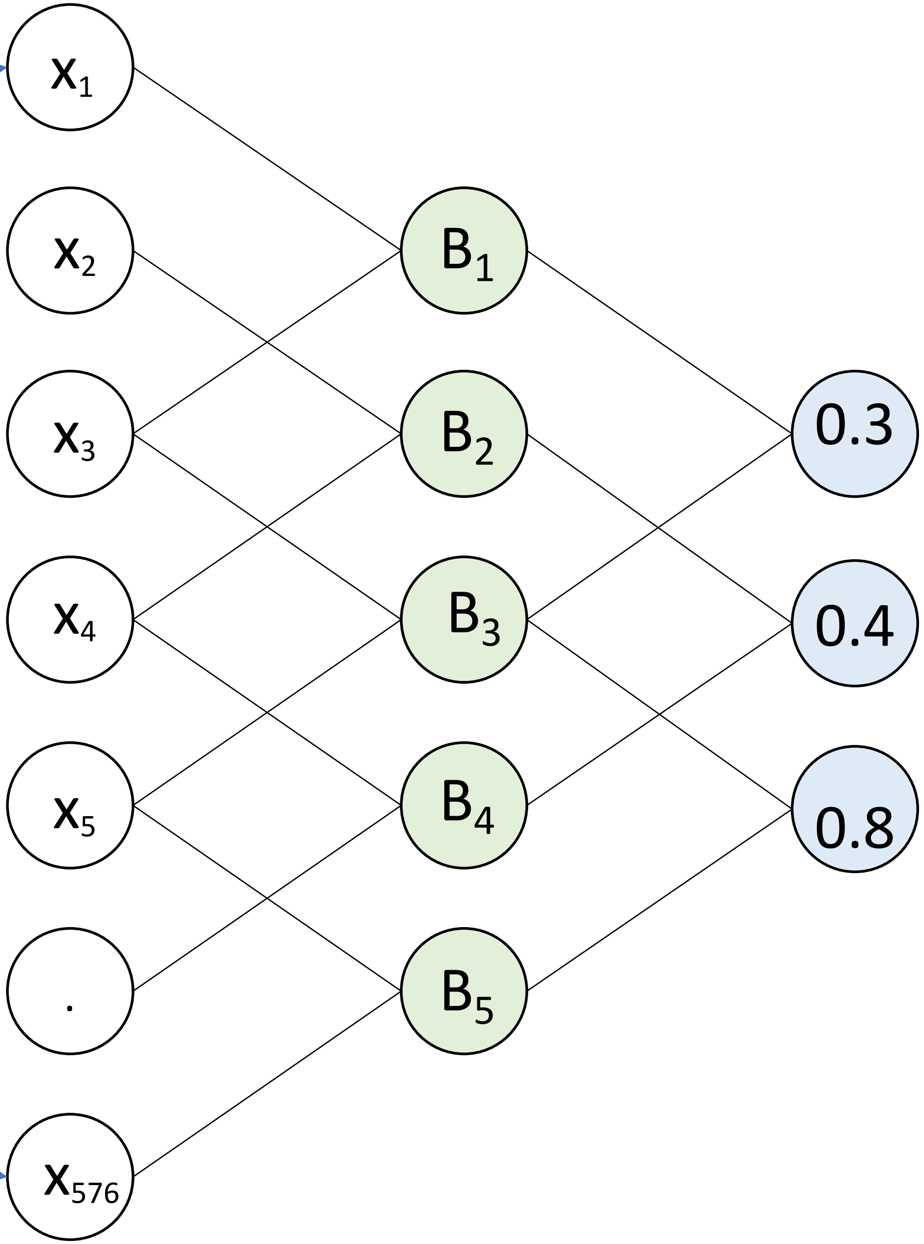
2012, AlexNet**,** ImageNet Large Scale Visual Recognition Challenge
Deep network + GPU + Data (ImageNet)

Key points in the history of computer vision



https://en.wikipedia.org/wiki/AlexNet

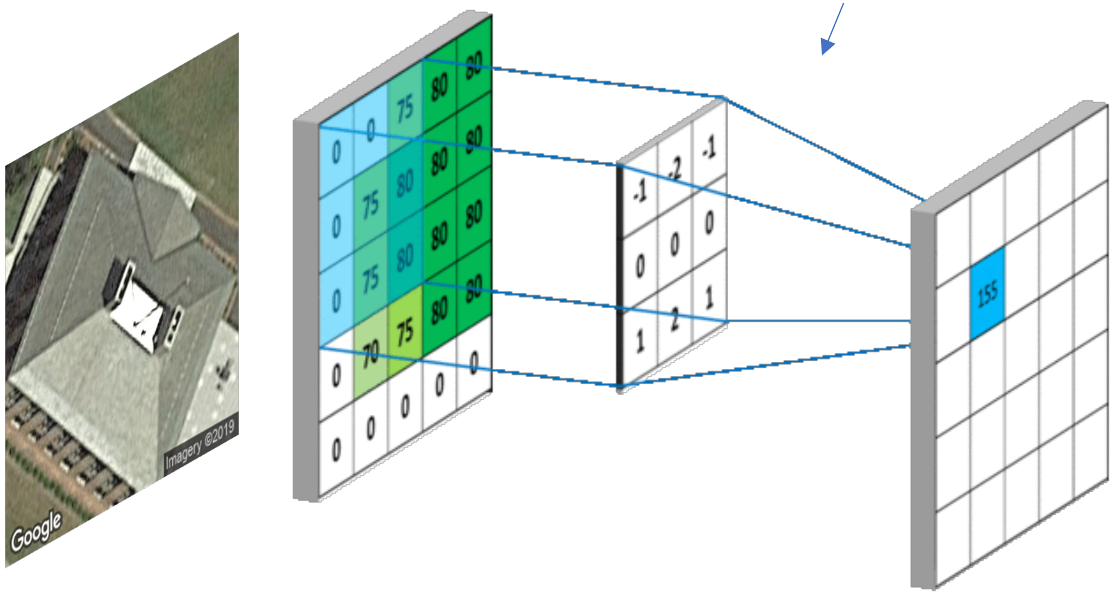
A 24x24 image can be expanded as a vector $[x_1, x_2, \ldots x_{576}]$



A brute way…

Convolutional layer

Kernel (filter) matrix



Image

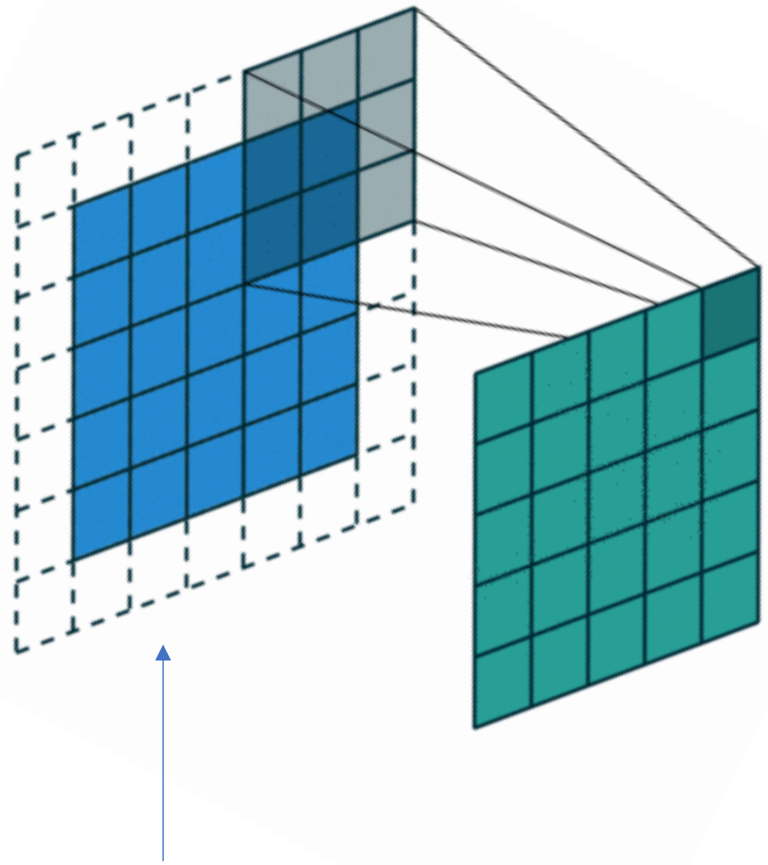
Convolved feature

A 2D convolution operation

Input: 5x5x1
Kernel: 3x3
Stride: 1 (size of the ‘slide’)
Padding: 0
Output: 3x3x1

Purpose: extracting features.

Convolutional layer



Padded

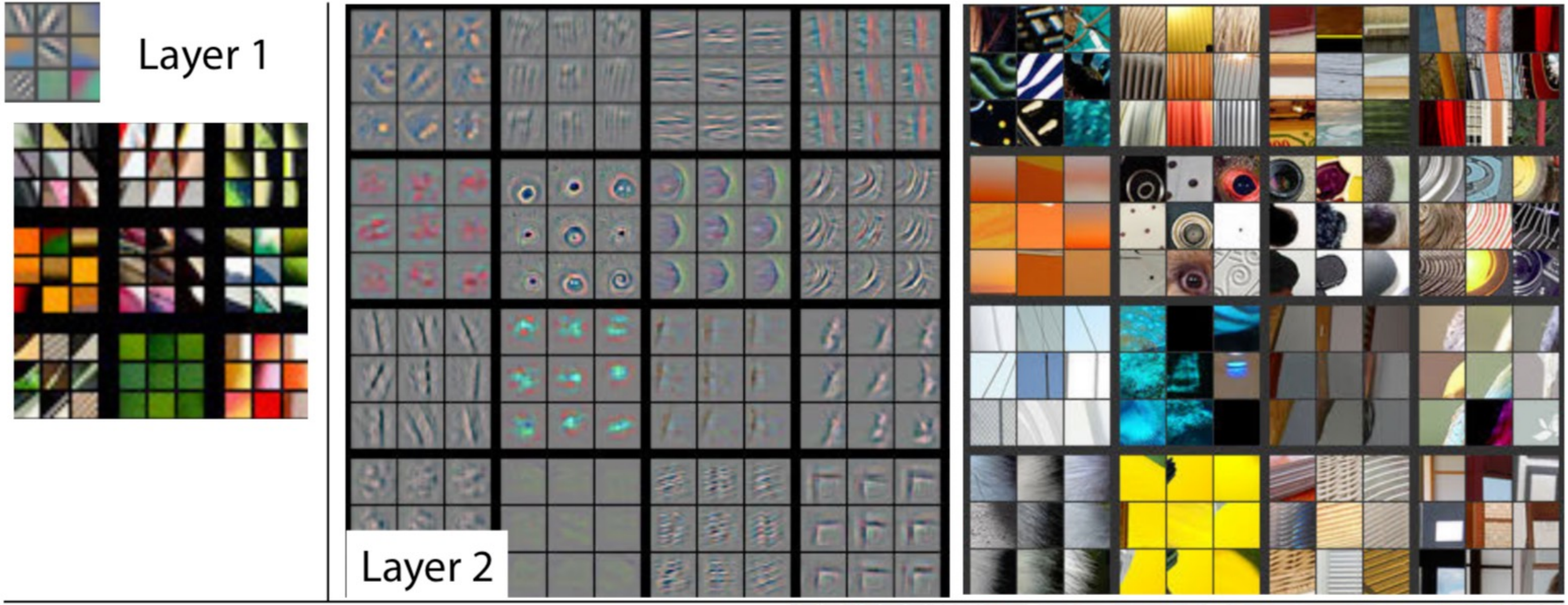
https://github.com/vdumoulin/conv_arithmetic

Input: 5x5x1
Kernel: 3x3
Stride: 1 (size of the ‘slide’)
Padding: 1
Output: 5x5x1

Purpose: extracting features.

Learn features hierarchically

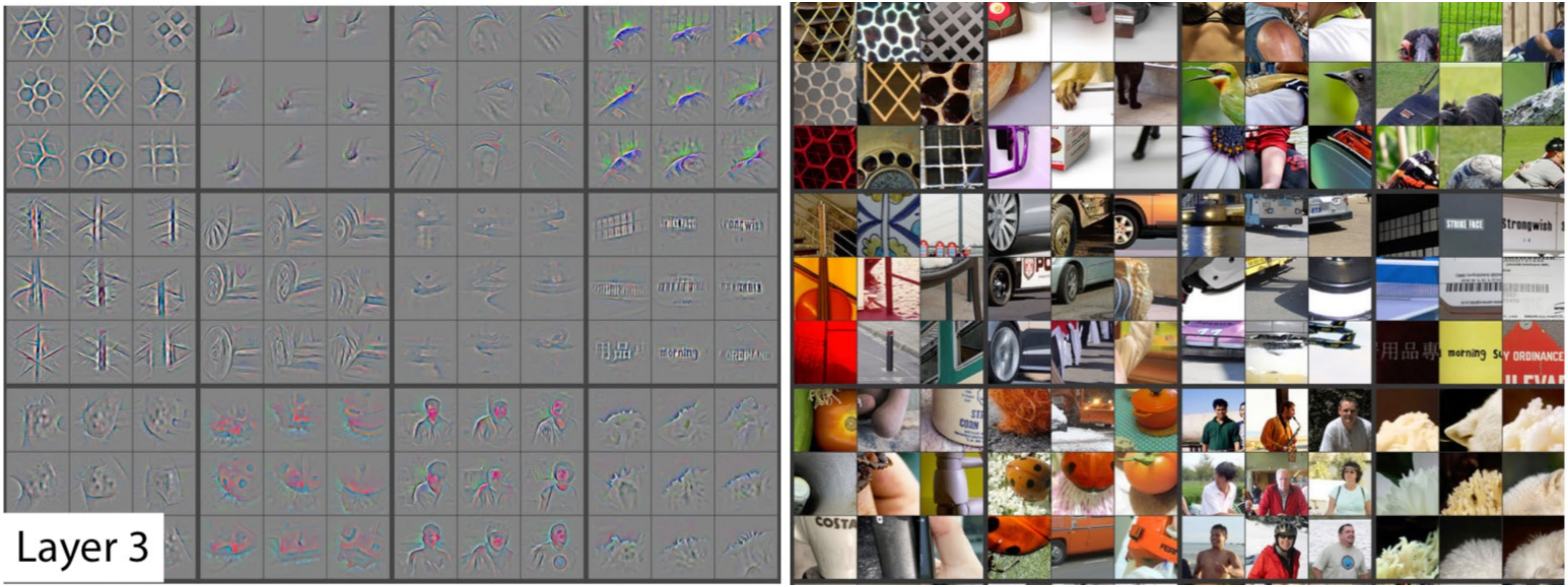
The first layers of CNN detect general features: Edges, Corners, Circles, Blobs colors, …



Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." European conference on computer vision. Springer, Cham, 2014.

Learn features hierarchically

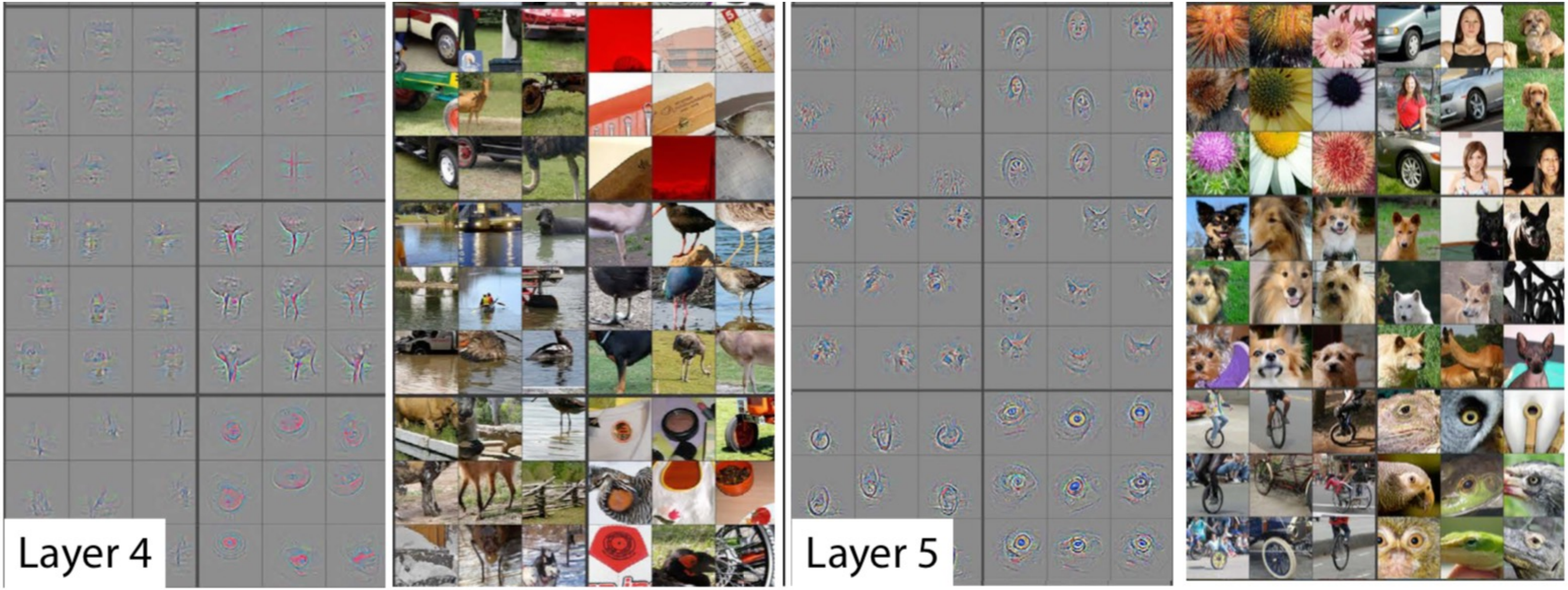
As it goes deeper into the CNN, it starts to detect more concrete things such as eyes, faces, and full objects.



Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." European conference on computer vision. Springer, Cham, 2014.

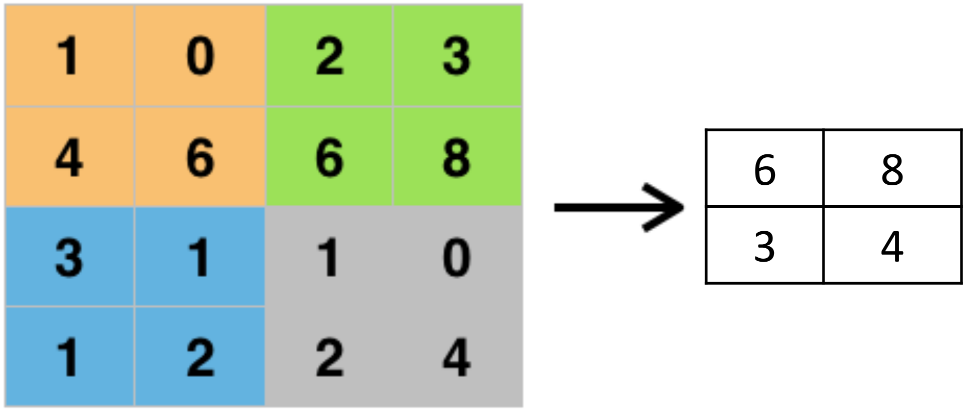
Learn features hierarchically

More concrete things …



Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." European conference on computer vision. Springer, Cham, 2014.

Pooling layer

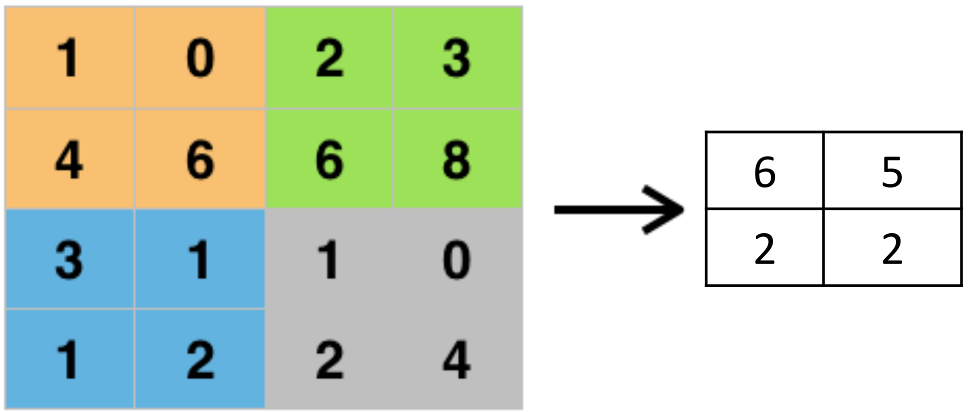


Max pooling

Max pooling
Average pooling
L2-norm polling
…

Purpose: extracting dominant feature and reduce dimensionality

Pooling layer

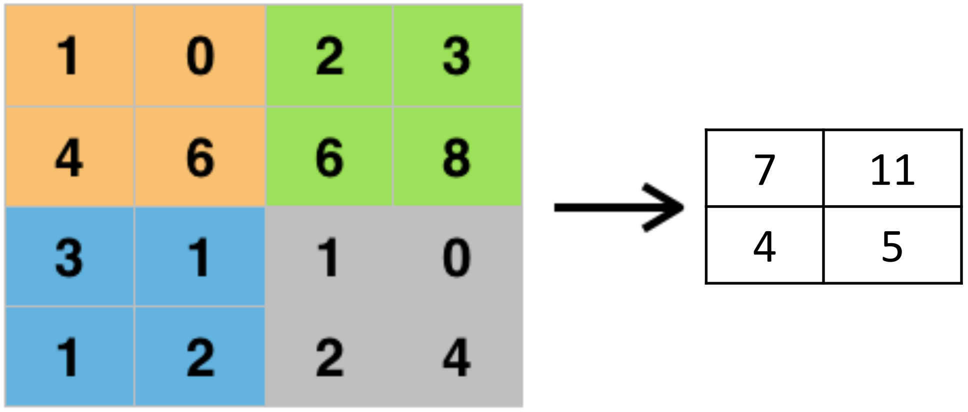


Average pooling

Max pooling
Average pooling
L2-norm pooling
…

Purpose: extracting dominant feature and reduce dimensionality

Pooling layer

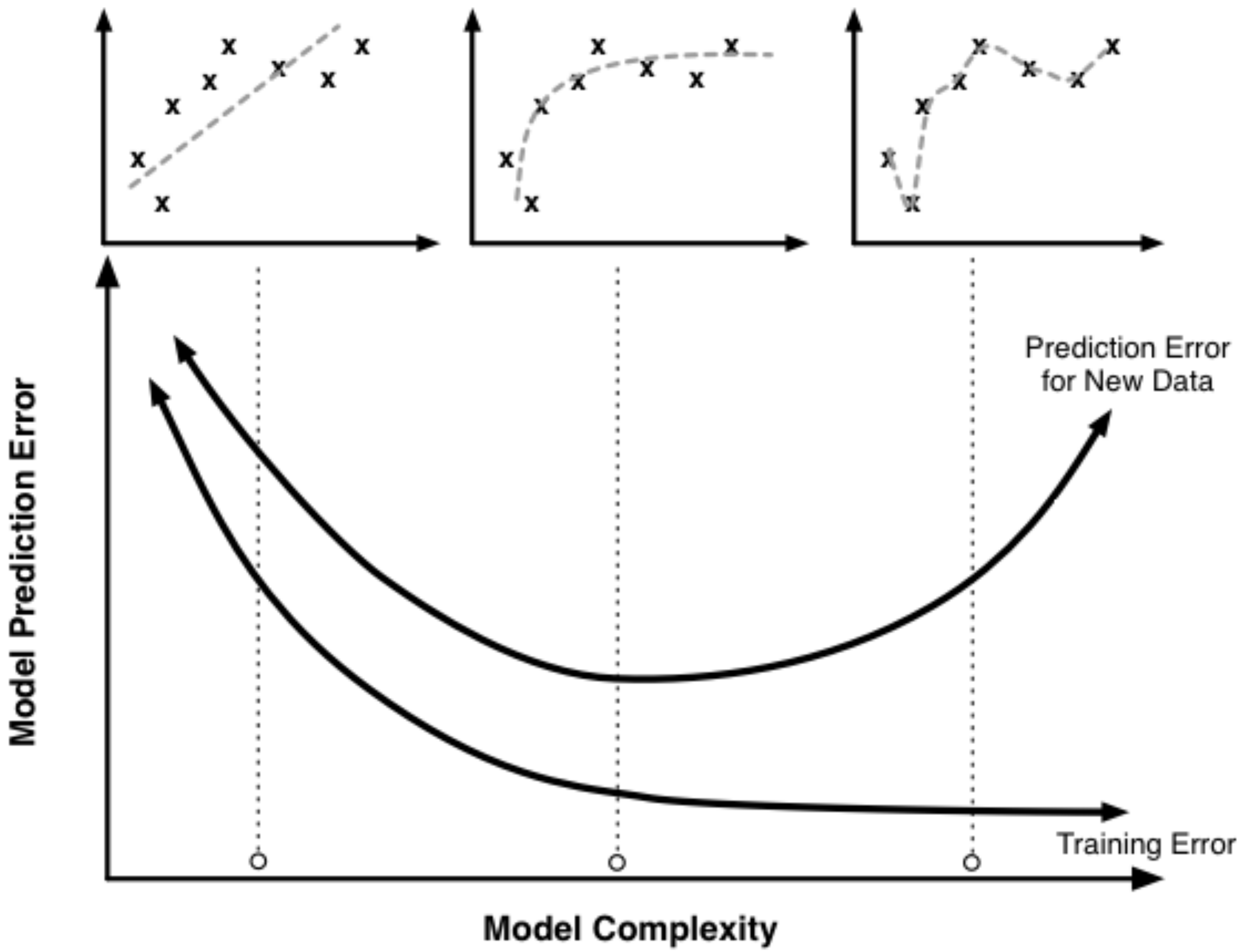


L2-norm pooling

Max pooling
Average pooling
L2-norm pooling
…

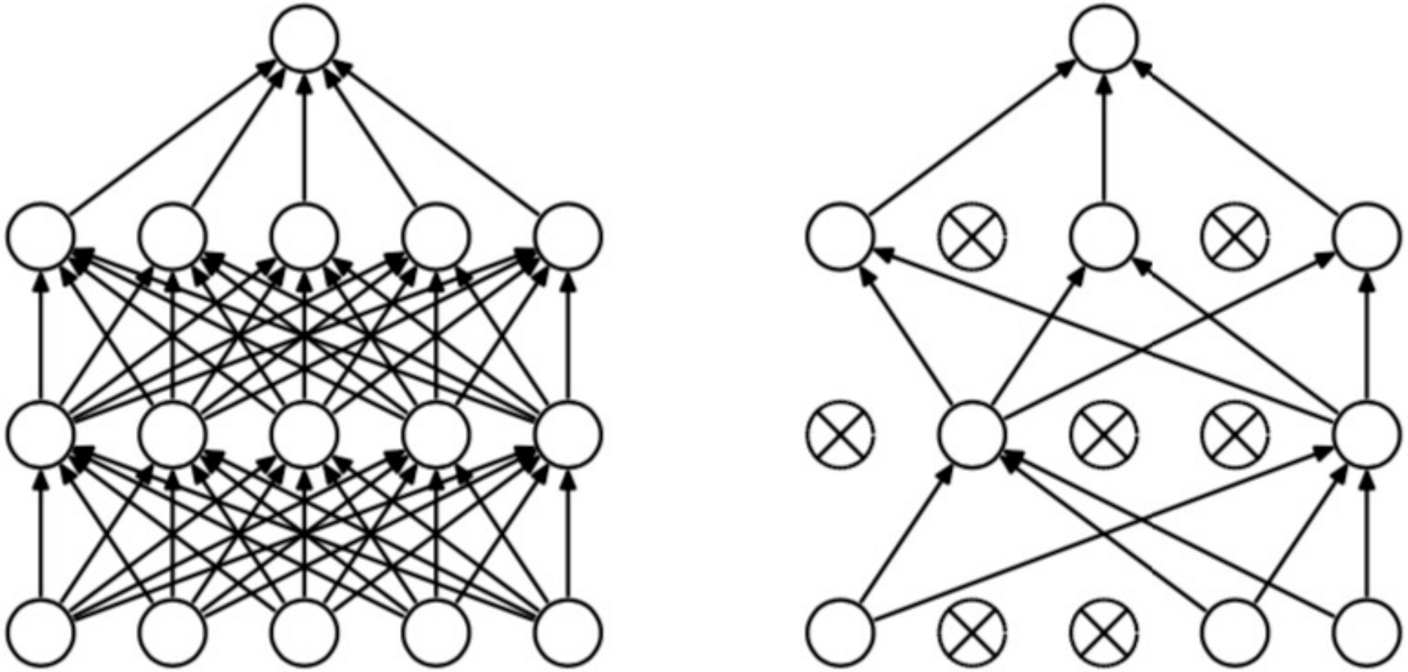
Purpose: extracting dominant feature and reduce dimensionality

What is a good model?



underfit optimal overfit

Dropout layer



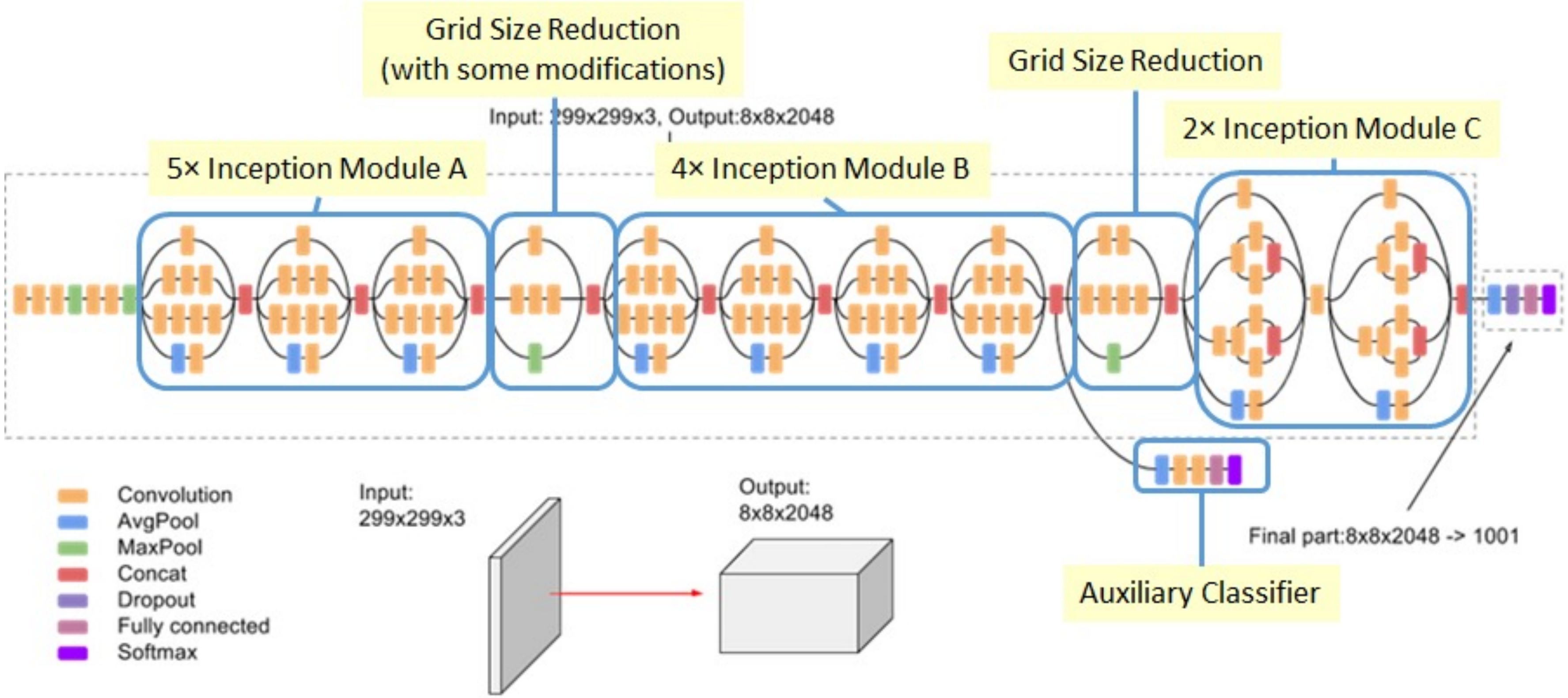
(a) Standard Neural Net (b) After applying dropout.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958. https://www.cs.toronto.edu/~hinton/absps/JMLRdropout.pdf

Purpose: fight overfitting

Popular deep CNN architectures

- AlexNet
- VGGNet
- GoogLeNet
- Microsfot ResNet
- Google Inception
- …



Example: Inception v3, https://arxiv.org/abs/1512.00567

Build one from scratch (ResNet)
https://www.analyticsvidhya.com/blog/2021/08/how-to-code-your-resnet-from-scratch-in-tensorflow/

Transfer Learning

The weights in a pretrained neural network is the leaned knowledge.

So a deep CNN trained on a large dataset contains knowledge (weights) that can be used to understand basic features in any given new image. This is the concept of transfer learning.

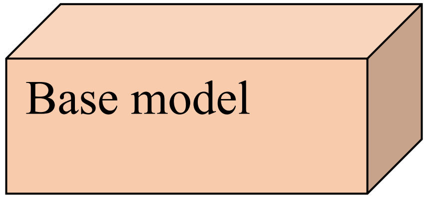
To do transfer learning, we

- Freeze the first layers of the pretrained neural network. These are the layers that detect general features that are common across all domains.

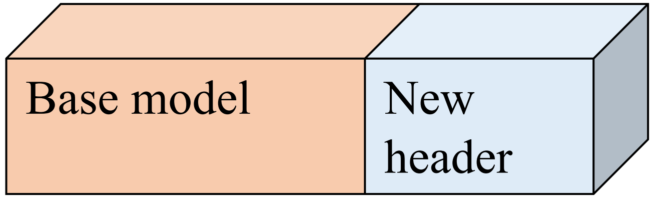
- Then we finetune the deeper layers with our own training data and add new layers to classify new categories included in our training dataset.

Transfer Learning: Fine tuning

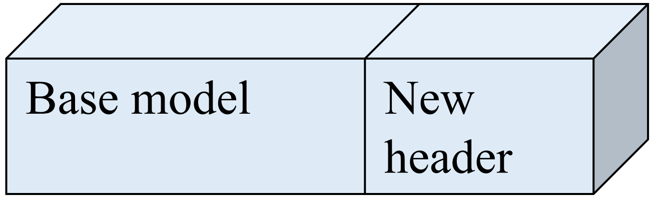
Take a pre-trained model (with learned weights) as base model



Add a header and train with the base model's weights frozen

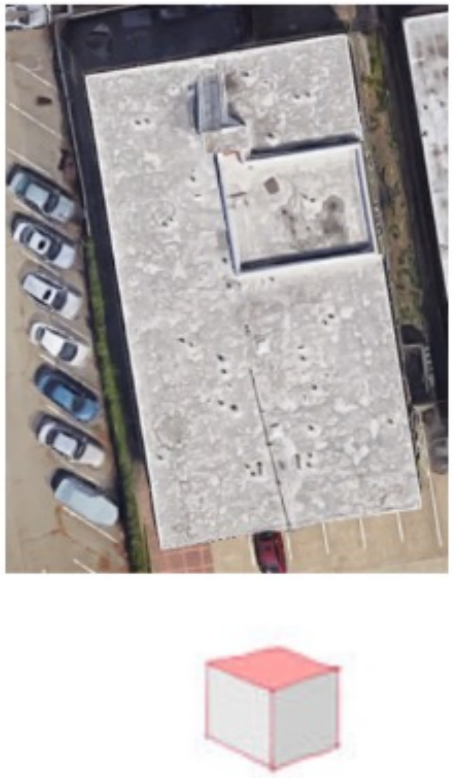


Unfreeze the base model and train



Demo: Classification of satellite images

Will do this demo in a Jupyter notebook on Google Colab:
https://colab.research.google.com/drive/1EKUEZEVgtWTfDTdR1grtAorMq3ZGLxmR?usp=sharing



Flat

Gabled

Hipped