



INTRODUCTION TO CONVOLUTIONAL NEURAL NETWORKS

Jakub Powierza

AGENDA

- Computer Vision - problems
- Architecture of the CNN
- What does happen inside?
- Popular architectures
- Popular datasets
- What's next?

COMPUTER VISION

“

Computer vision is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the **human visual system** can do.

~ *Wikipedia*

APPLICATIONS FOR COMPUTER VISION

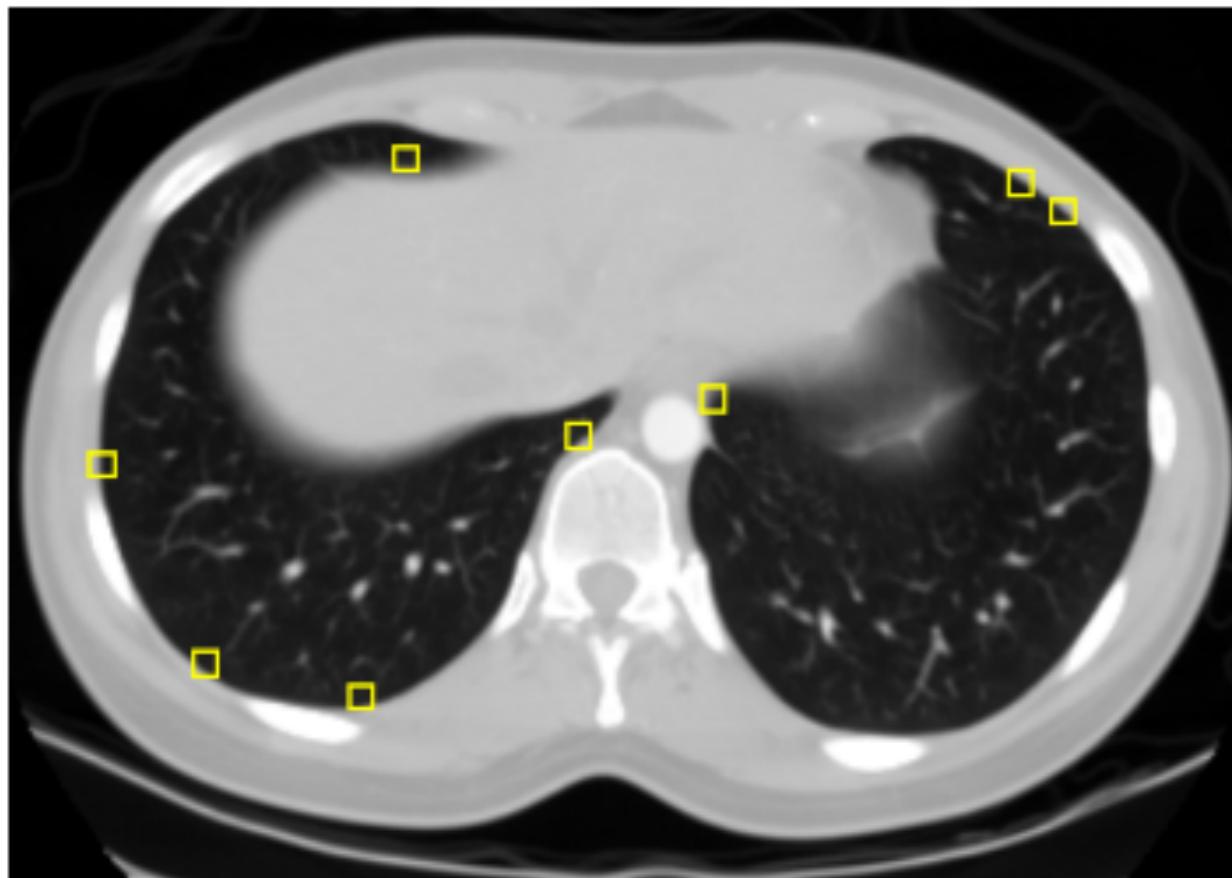
- agriculture
- augmented reality
- autonomous vehicles
- biometrics
- character recognition
- forensics
- quality inspection
- face recognition
- gesture analysis
- geoscience
- image restoration
- medical image analysis
- pollution monitoring
- process control
- remote sensing
- robotics
- security and surveillance
- transport
- ...



AUTONOMOUS VEHICLES

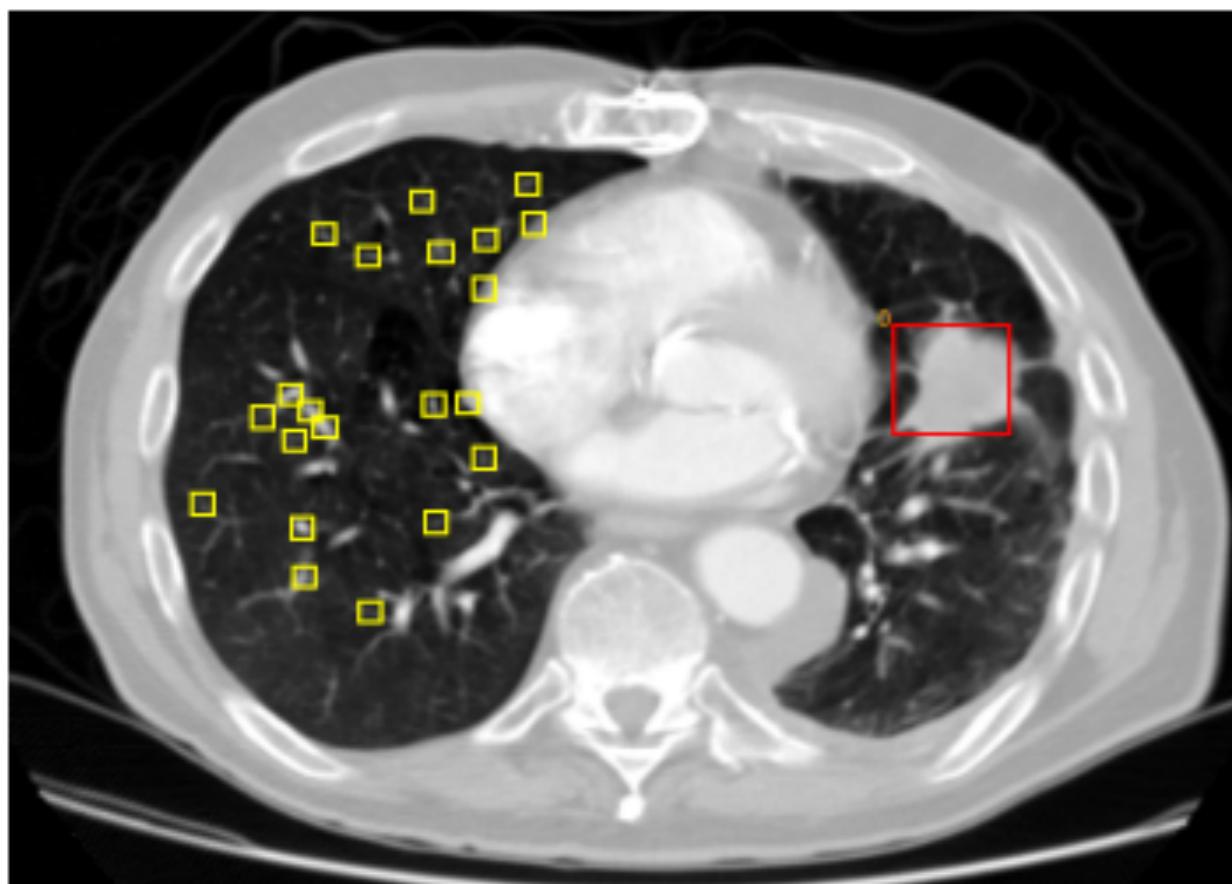
- Multiple sensors like:
 - Cameras,
 - Lidars,
 - Ultrasonic sensors,
- Big players like:
 - Intel & BMW,
 - Nvidia & Audi,
 - Google,
 - Tesla.



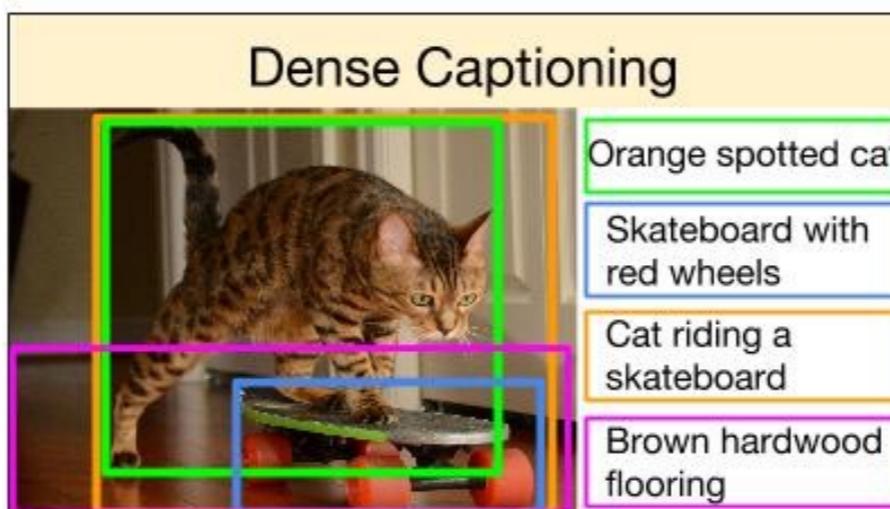


MEDICINE

- Assistance for doctors,
- Cancer detection (Kaggle's Data Science Bowl 2017),
- Healthcare automation (future?).



WHAT WE CAN DO?



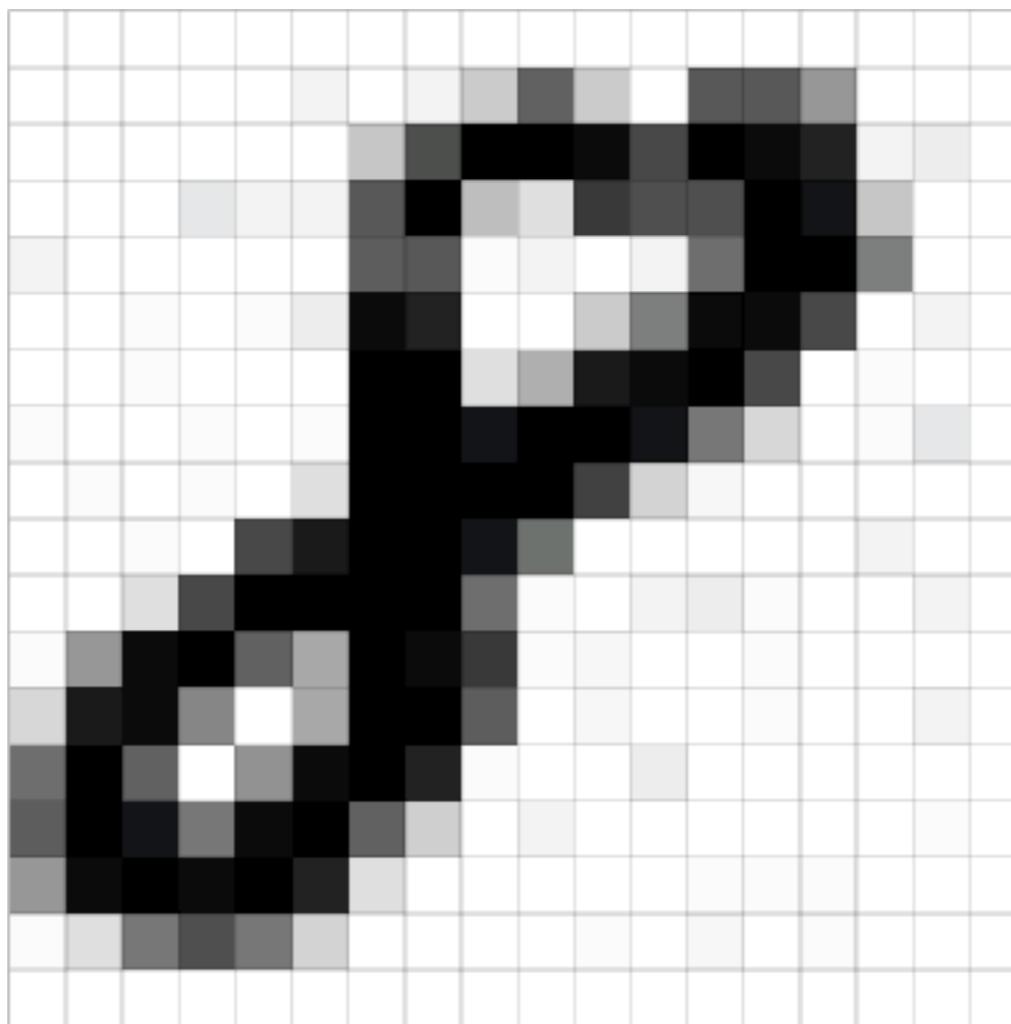
CNN ARCHITECTURE

What's inside?

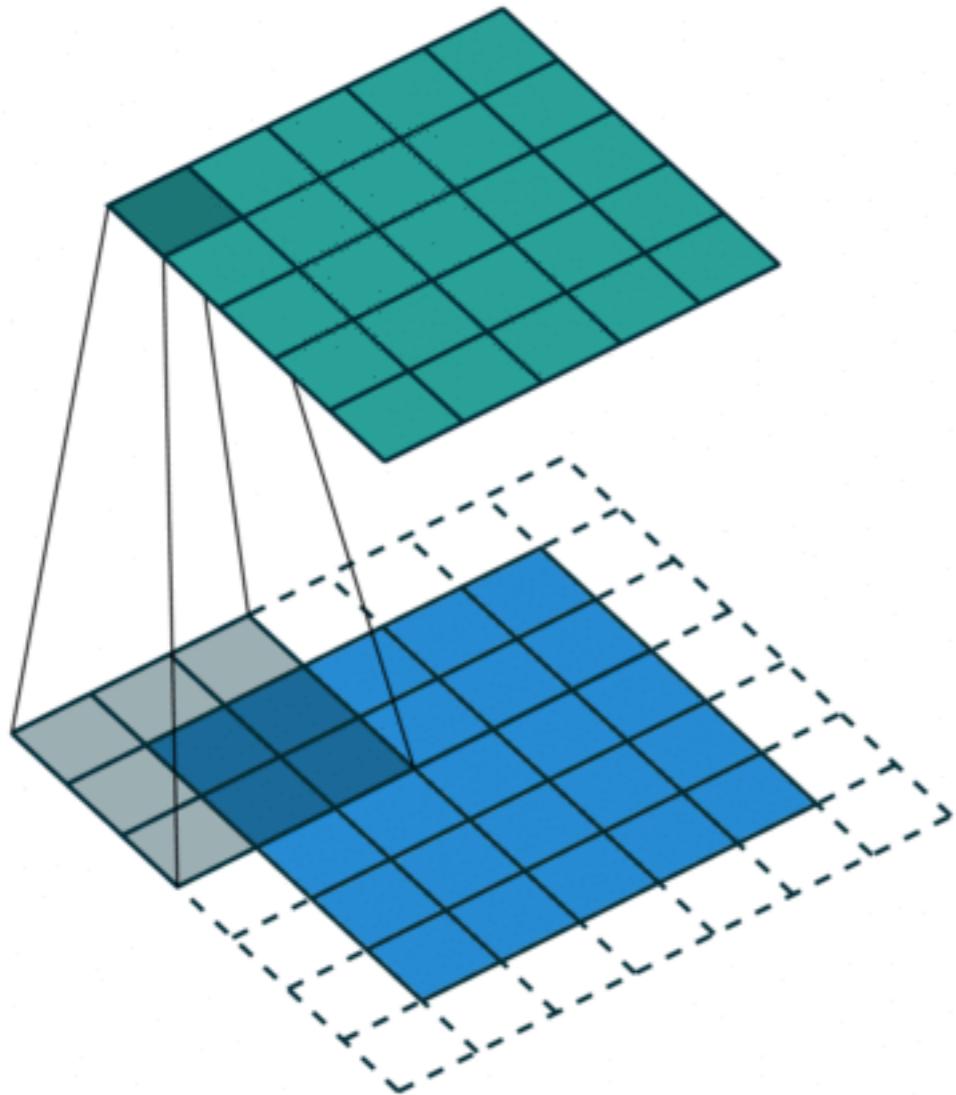
BASIC BLOCKS

- Input layer,
- Convolution layer,
- Activation layer,
- Pooling layer,
- Fully Connected layer.

WHAT IS OUR INPUT?



CONVOLUTION



- Creates “**feature maps**”,
- Apply **filters** on the image,
- Move such filter over the image and calculate **feature**,
- Follow the **stride** (how many fields it should “jump”),
- Is defined by **kernel size** (filter size),
- Can use **padding** for bigger receptive field.

1 x1	1 x0	1 x1	0	0
0 x0	1 x1	1 x0	1	0
0 x1	0 x0	1 x1	1	1
0	0	1	1	0
0	1	1	0	0

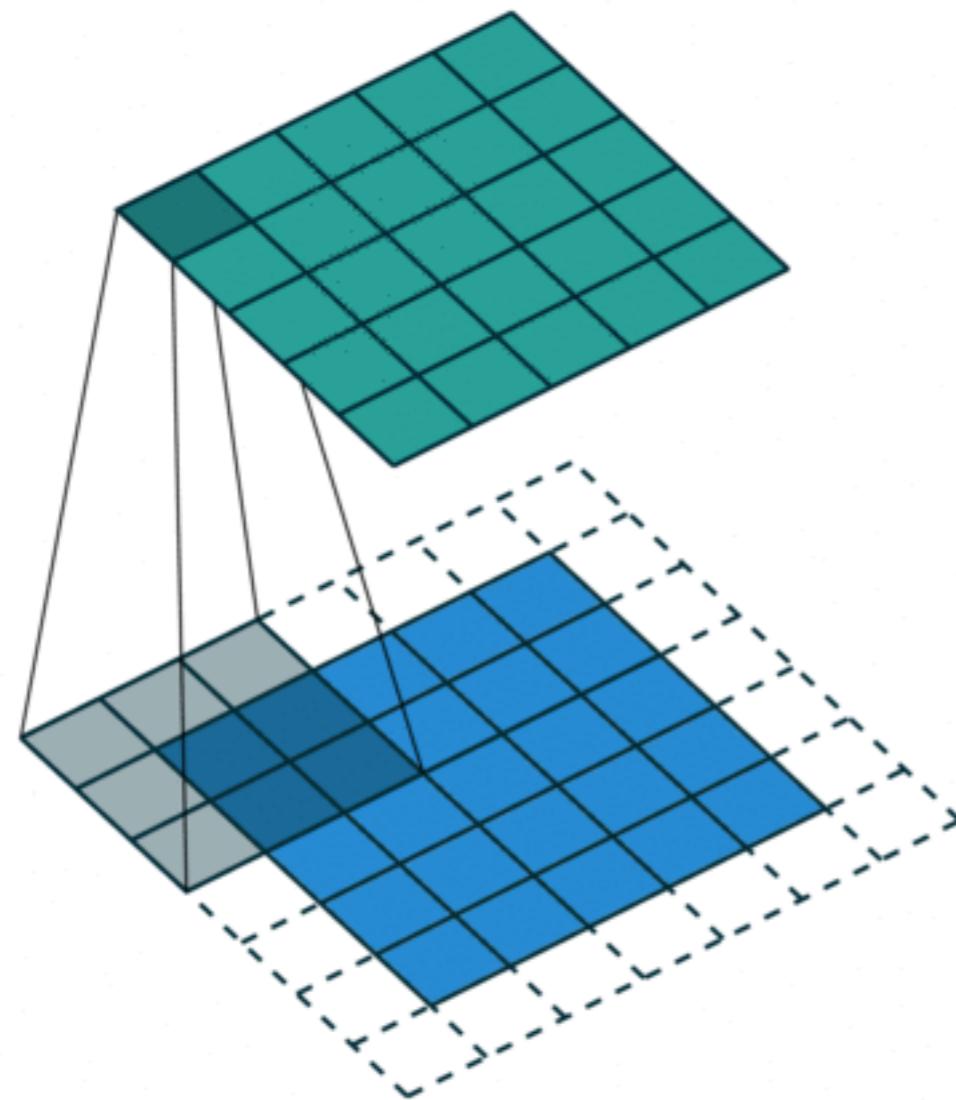
Image

4		

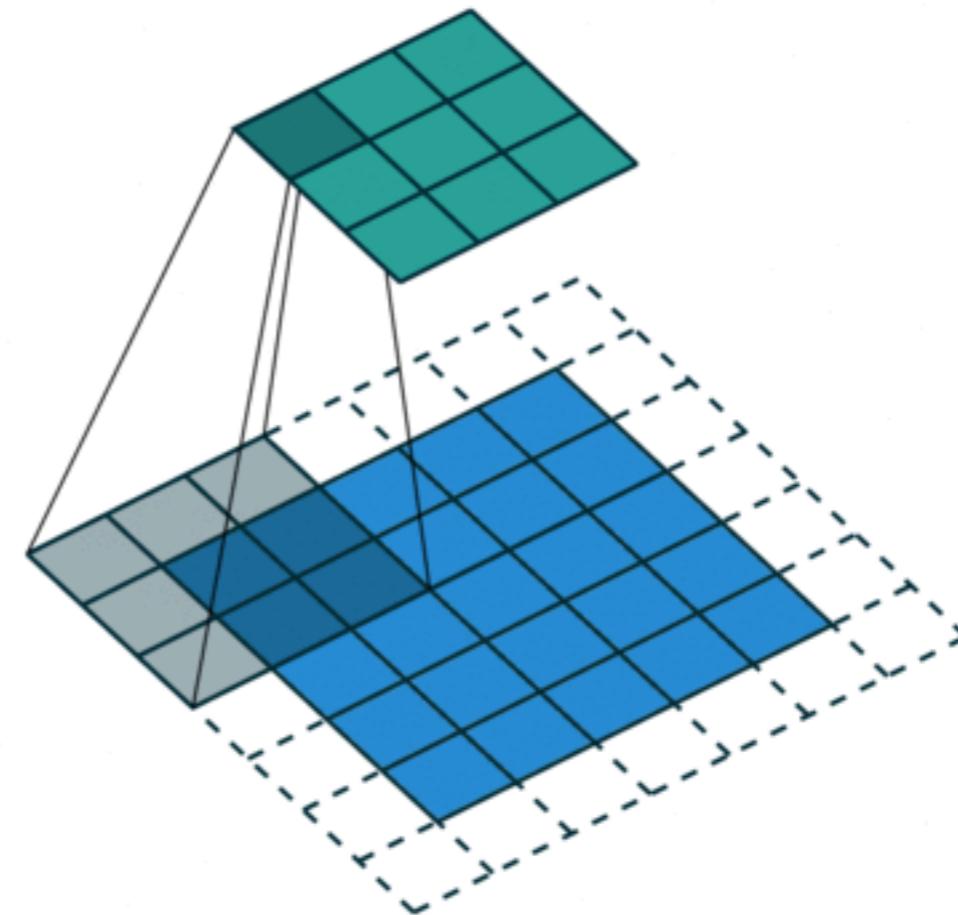
Convolved
Feature

STRIDE

Stride = 1

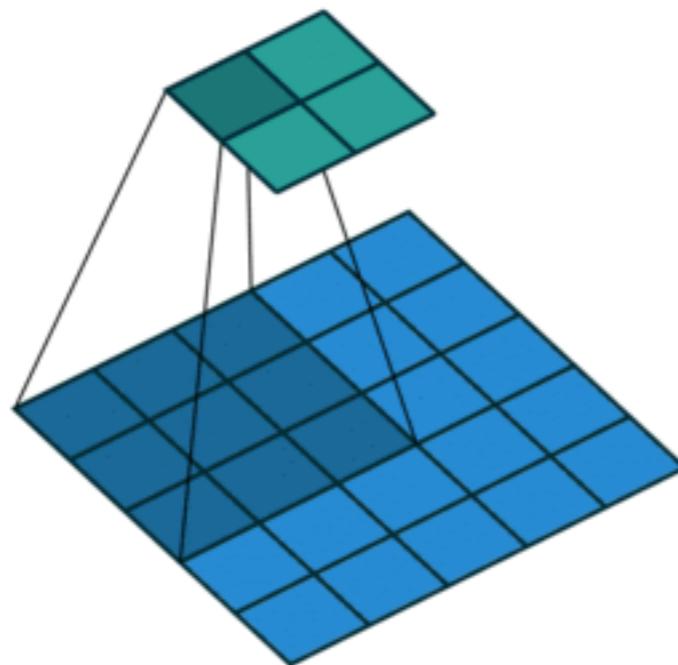


Stride = 2

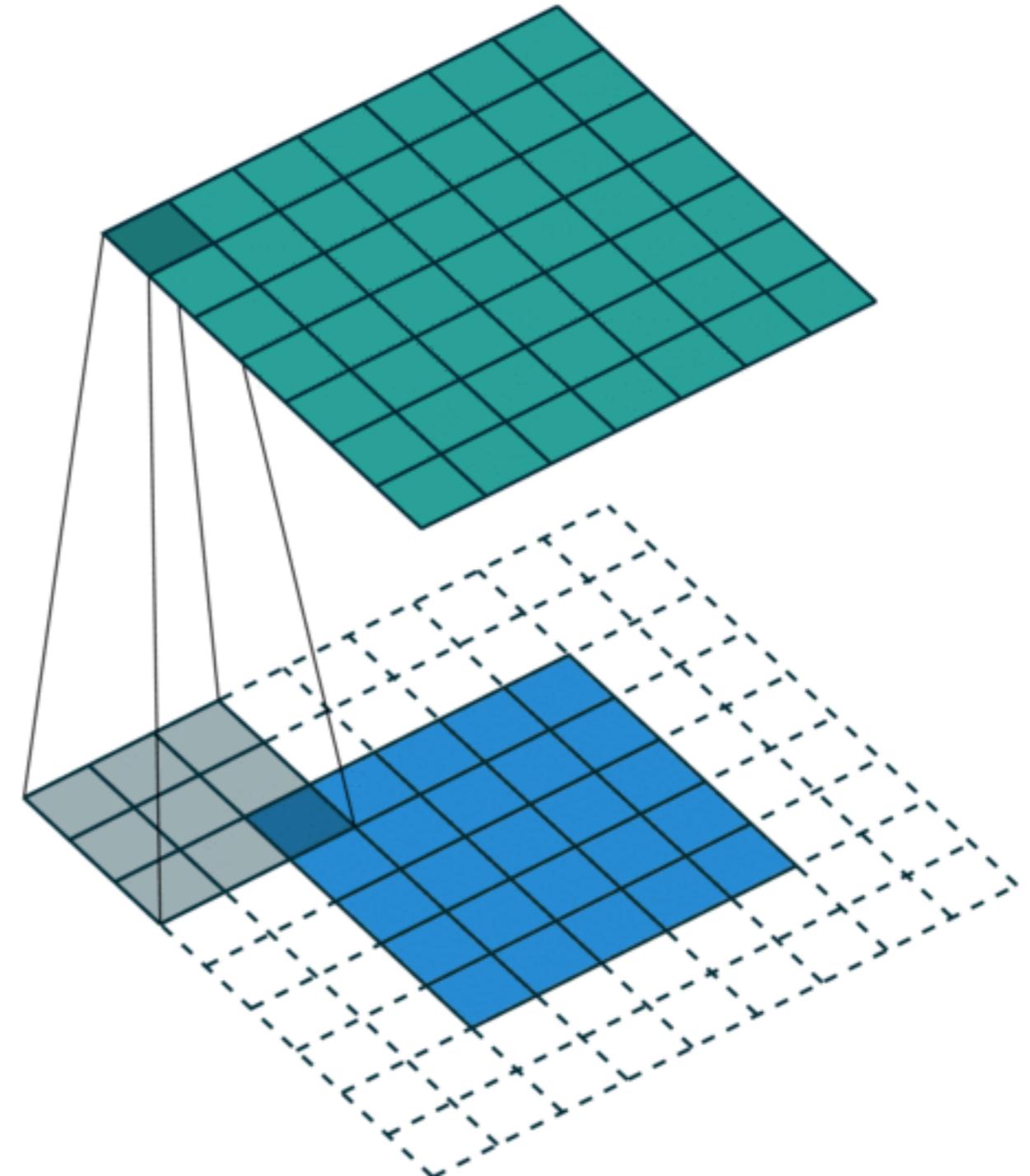


PADDING

Padding = 0

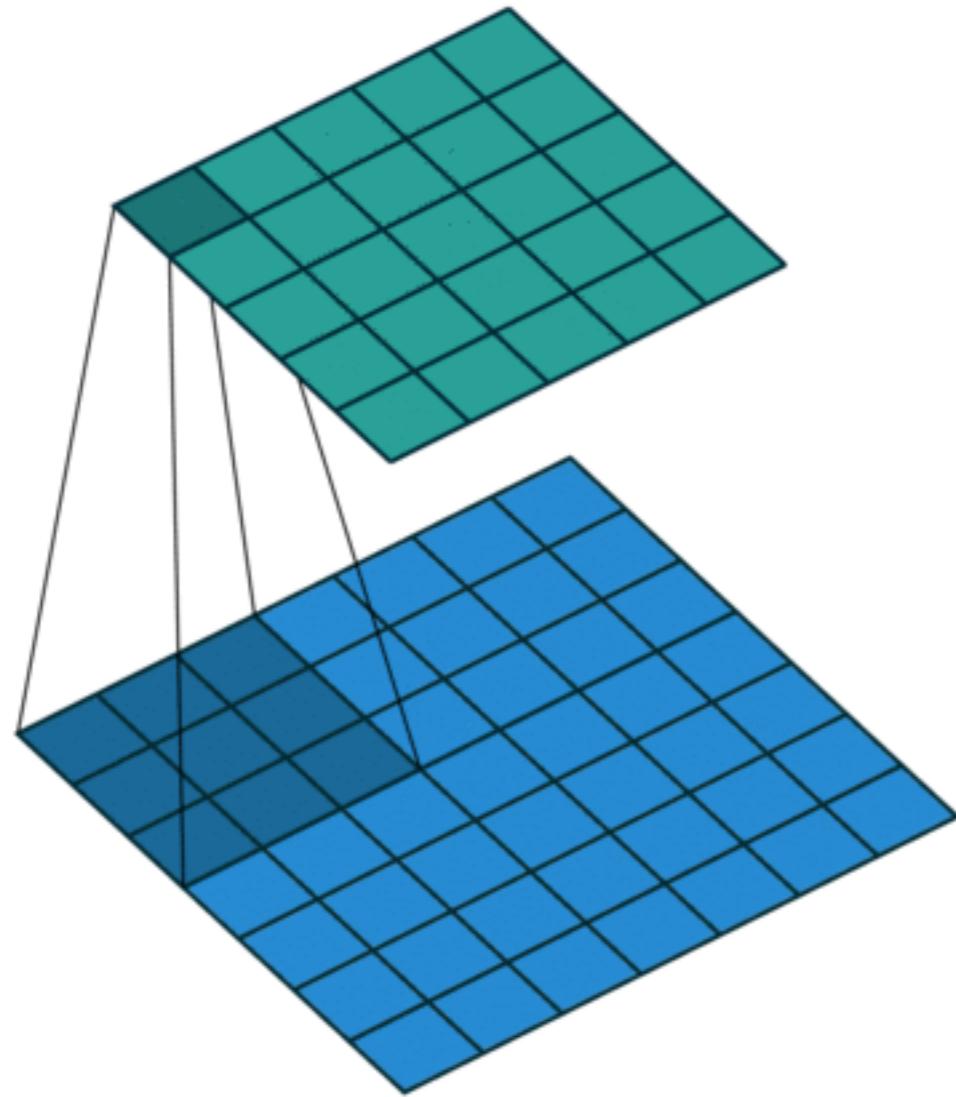


Padding = 2

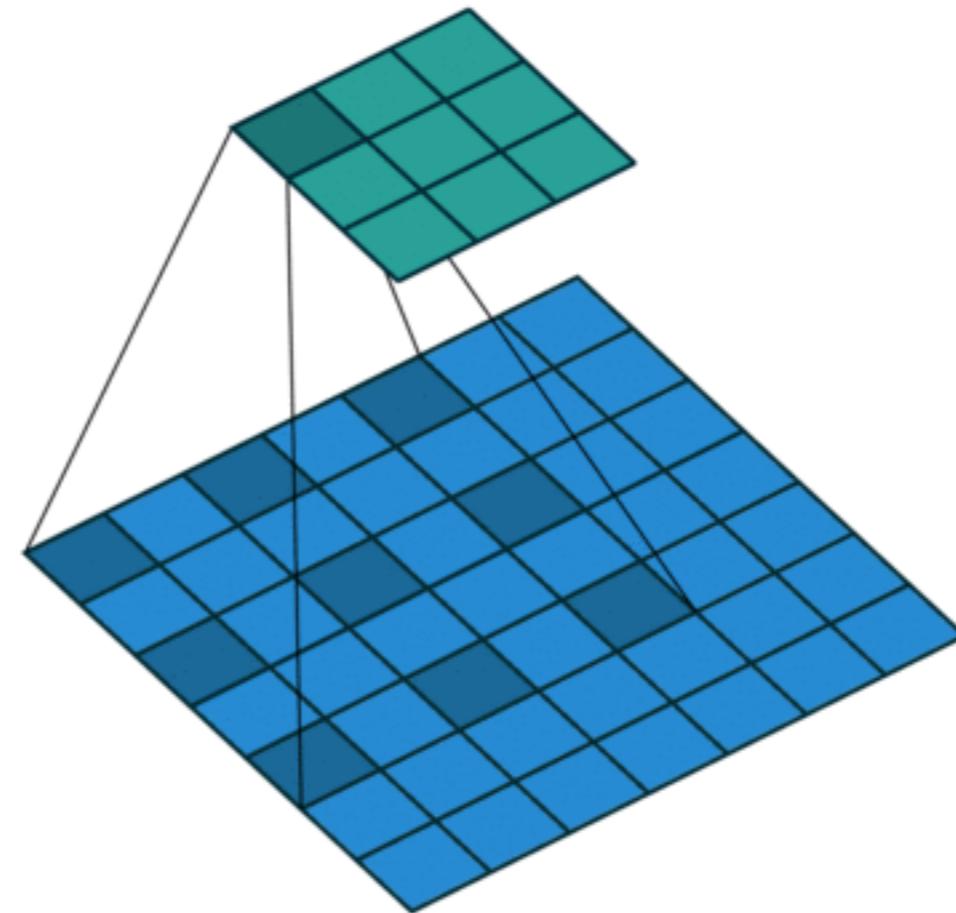


DILATION

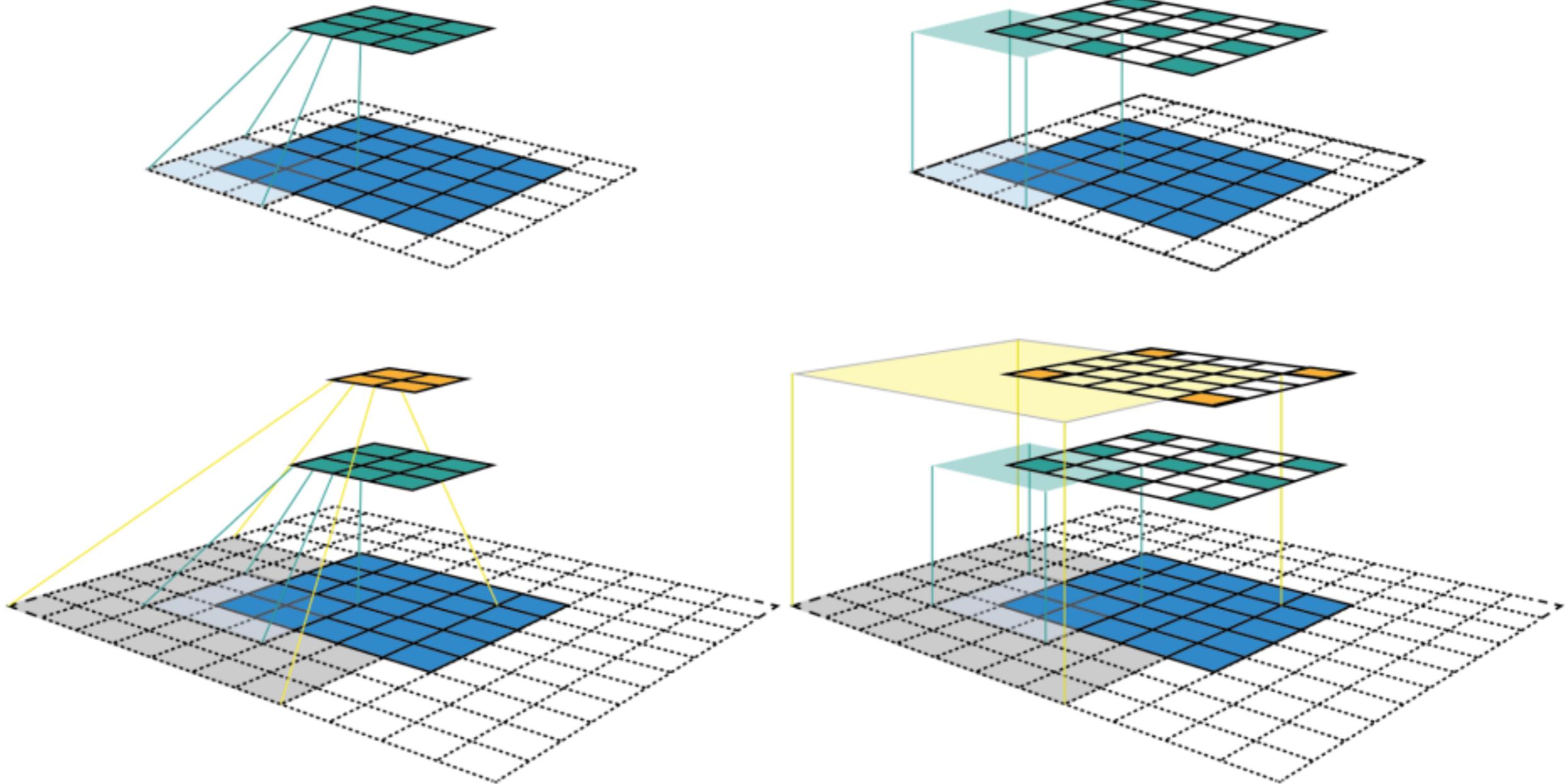
Dilation = 1



Dilation = 2



RECEPTIVE FIELD



EQUATION FOR NUMBER OF OUTPUT FEATURES

$$n_{out} = \left\lfloor \frac{n_{in} + 2p - k}{s} \right\rfloor + 1$$

n_{in} : number of input features

n_{out} : number of output features

k : convolution kernel size

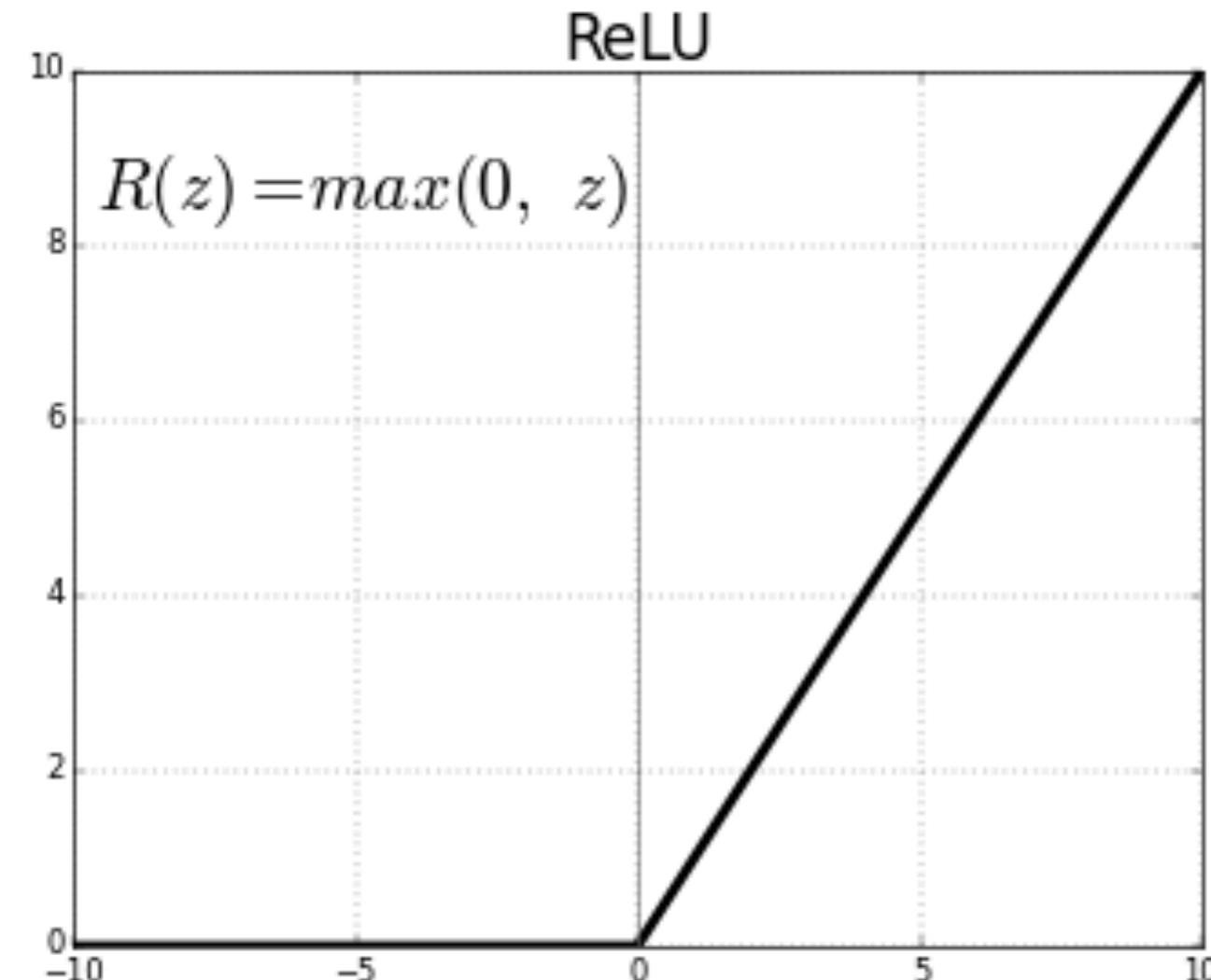
p : convolution padding size

s : convolution stride size



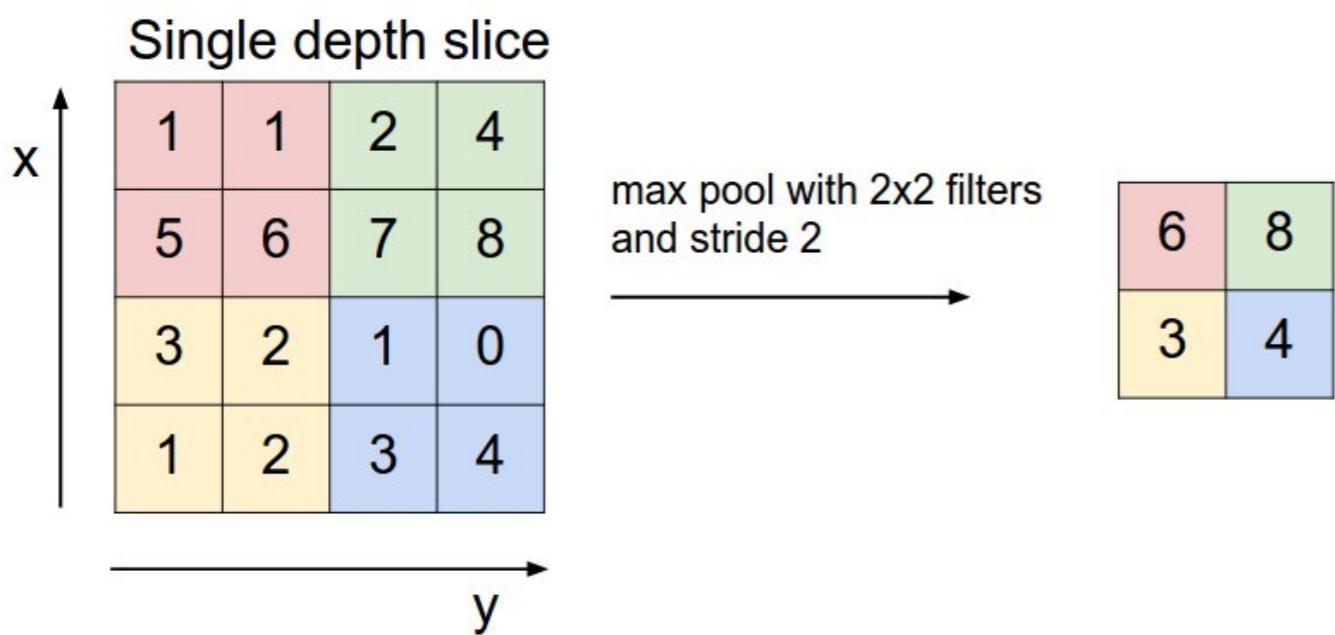
Input

ACTIVATION



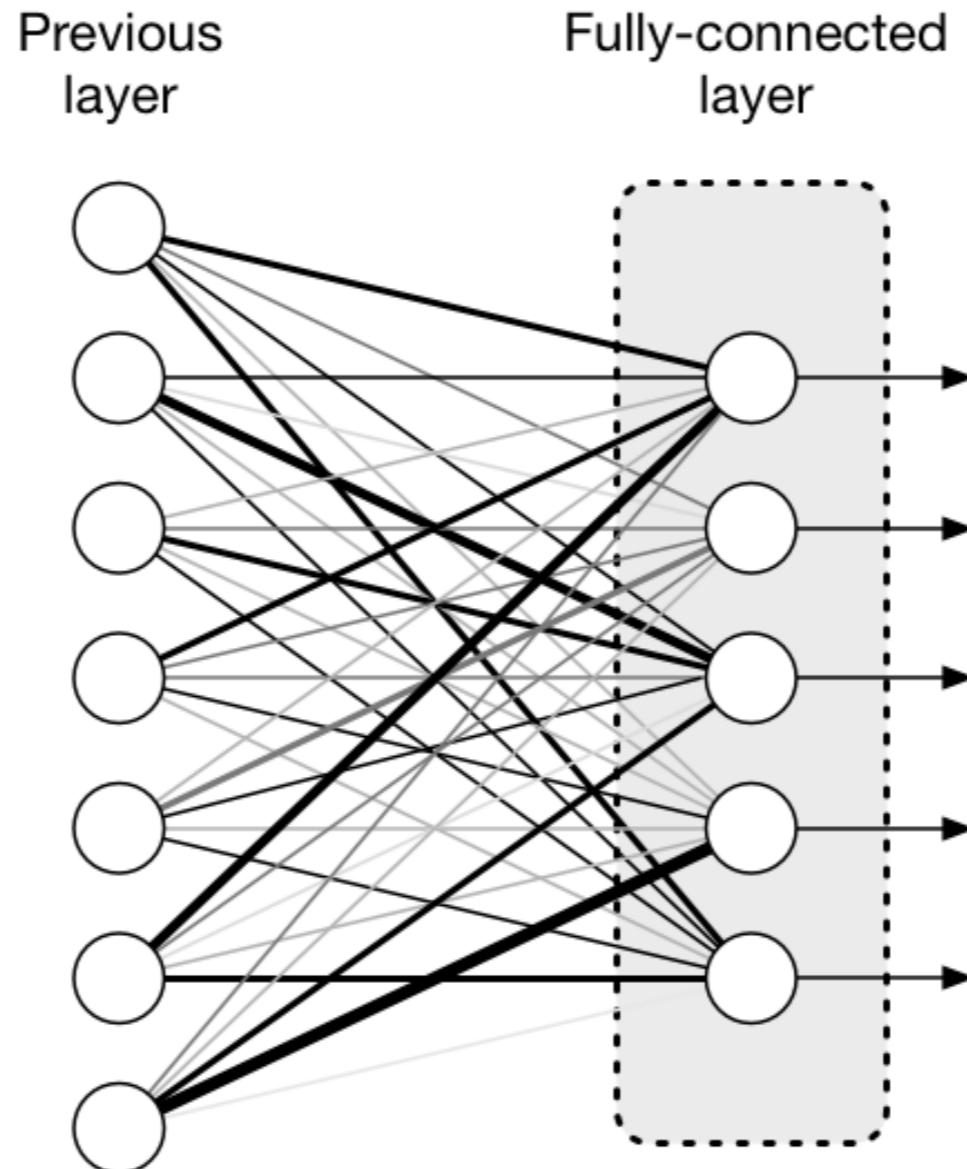
- ReLU is a default way to go,
- Some similar layers like:
 - Leaky ReLU,
 - Maxout,
- There are many other layers but have many problems (vanishing gradients, etc.).

POOLING

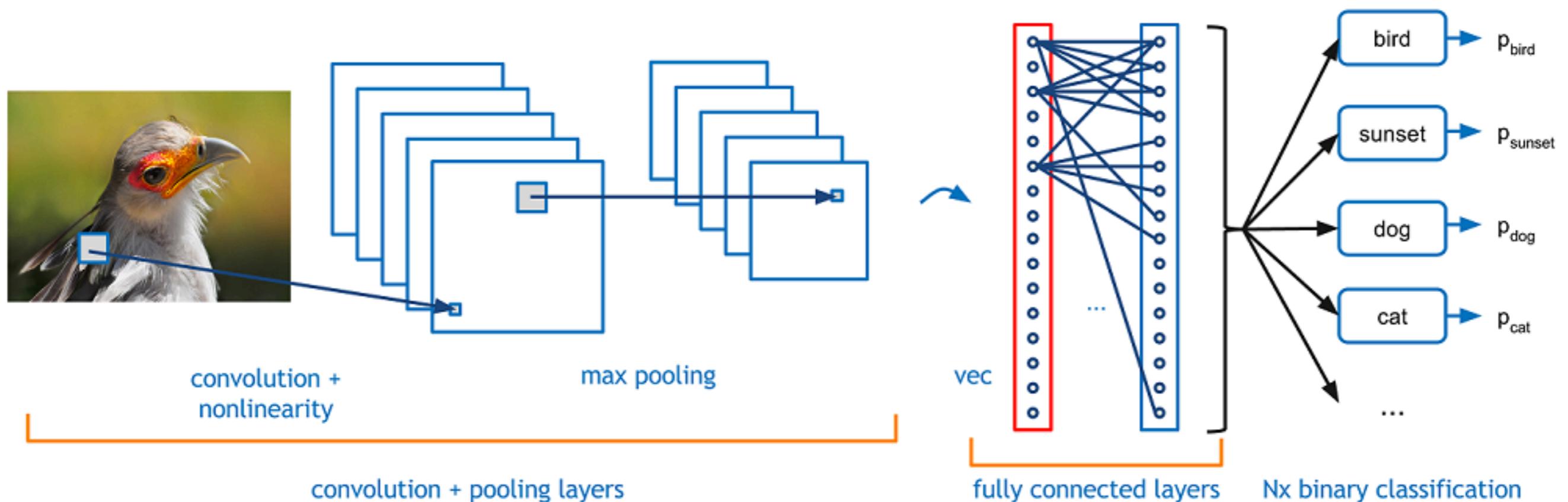


- Pooling reduces spatial space,
- Reduces amount of parameters,
- Reduces overfitting,
- A simple routing (during back propagation),
- Most common: MaxPooling,
- Also: AvgPooling, ...

FULLY CONNECTED LAYER



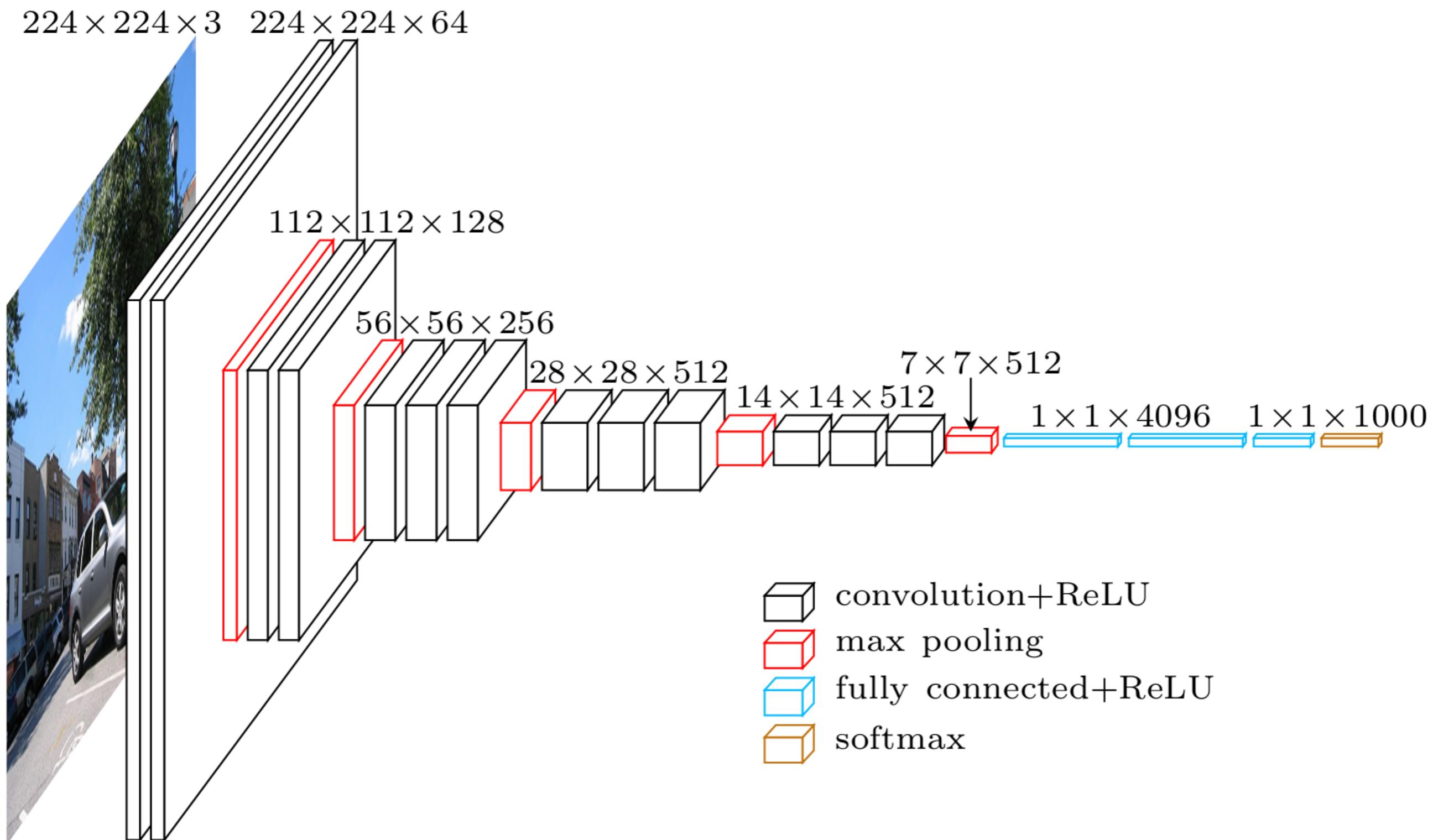
FULLY CONNECTED LAYER



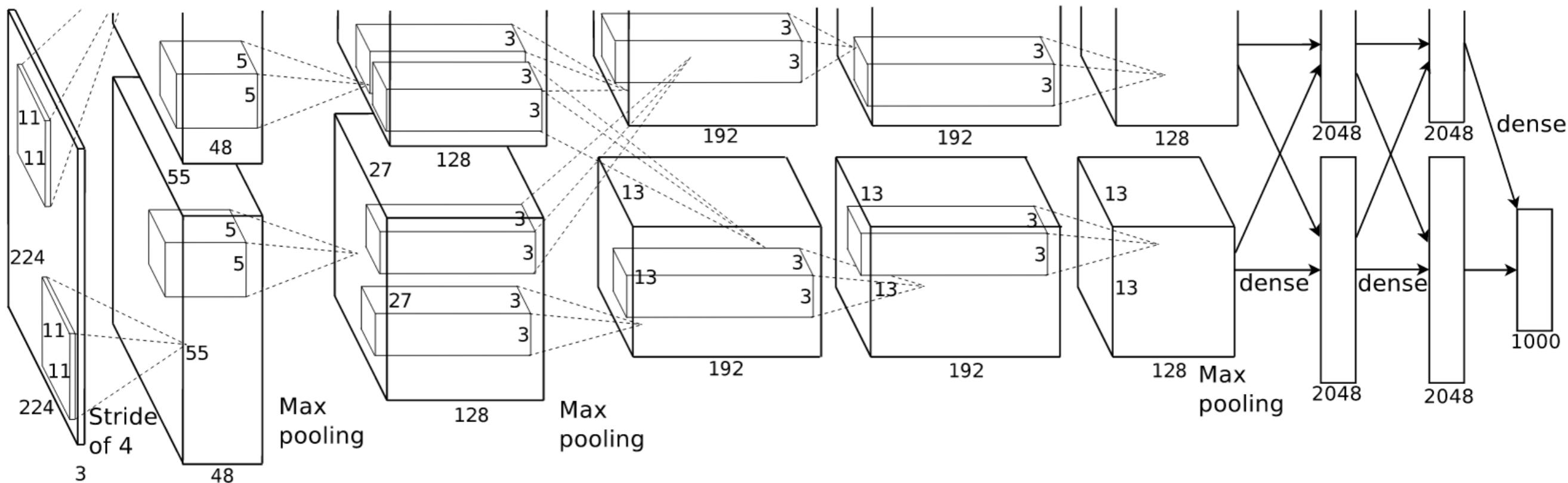
POPULAR ARCHITECTURES

...that might be helpful!

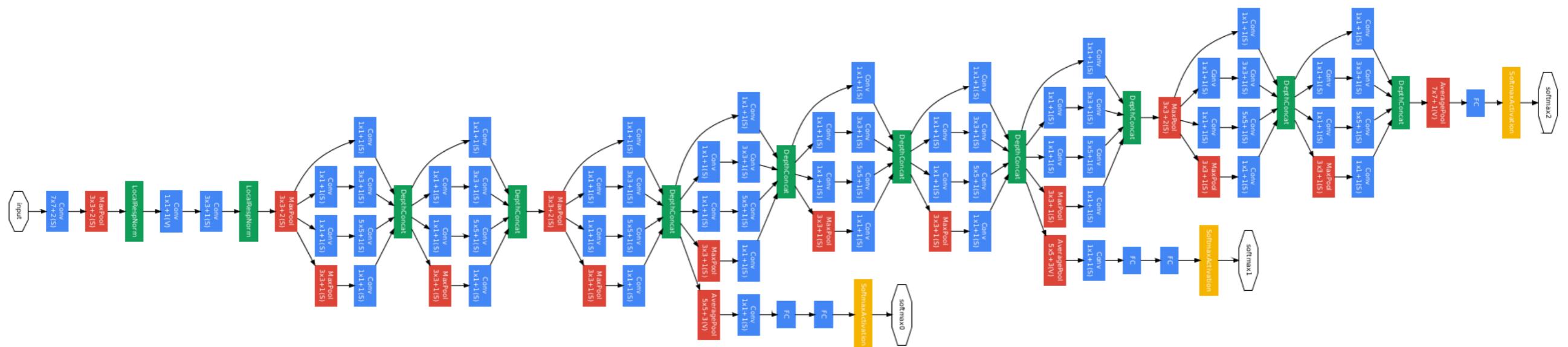
VGG



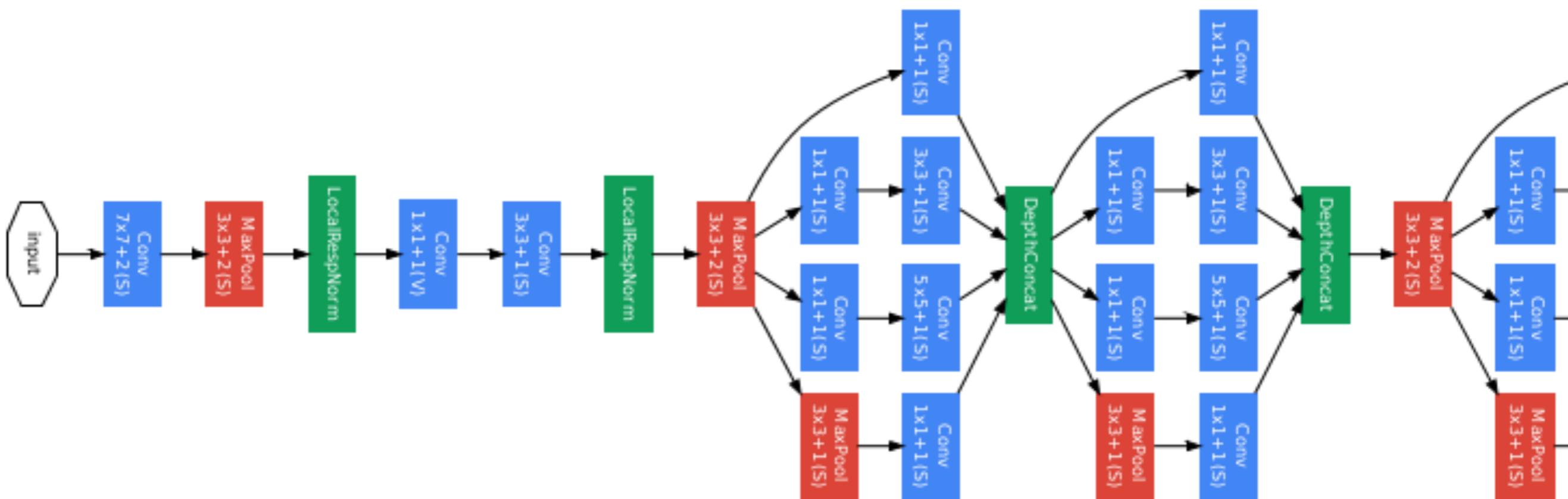
ALEXNET



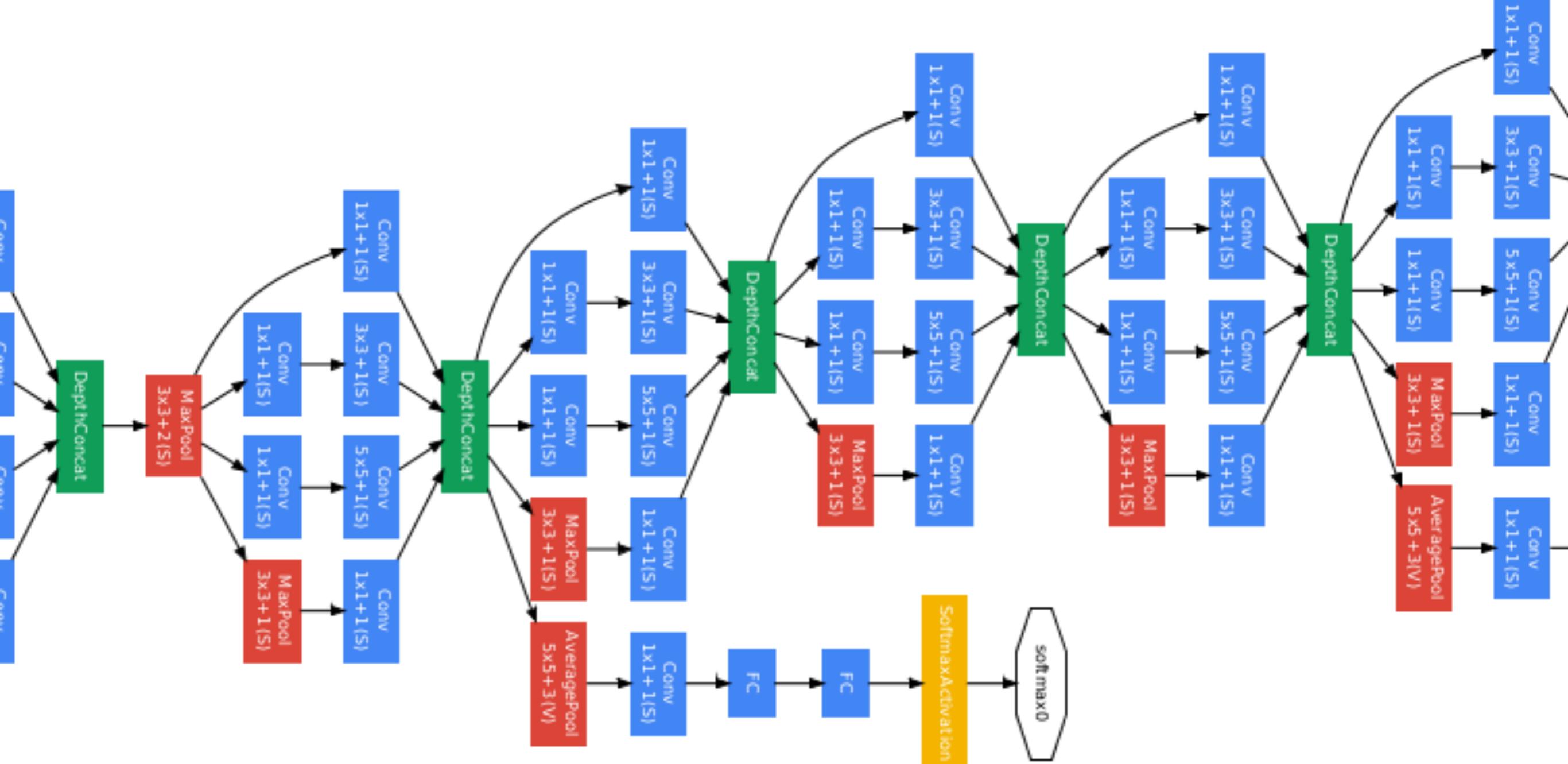
GOOGLENET



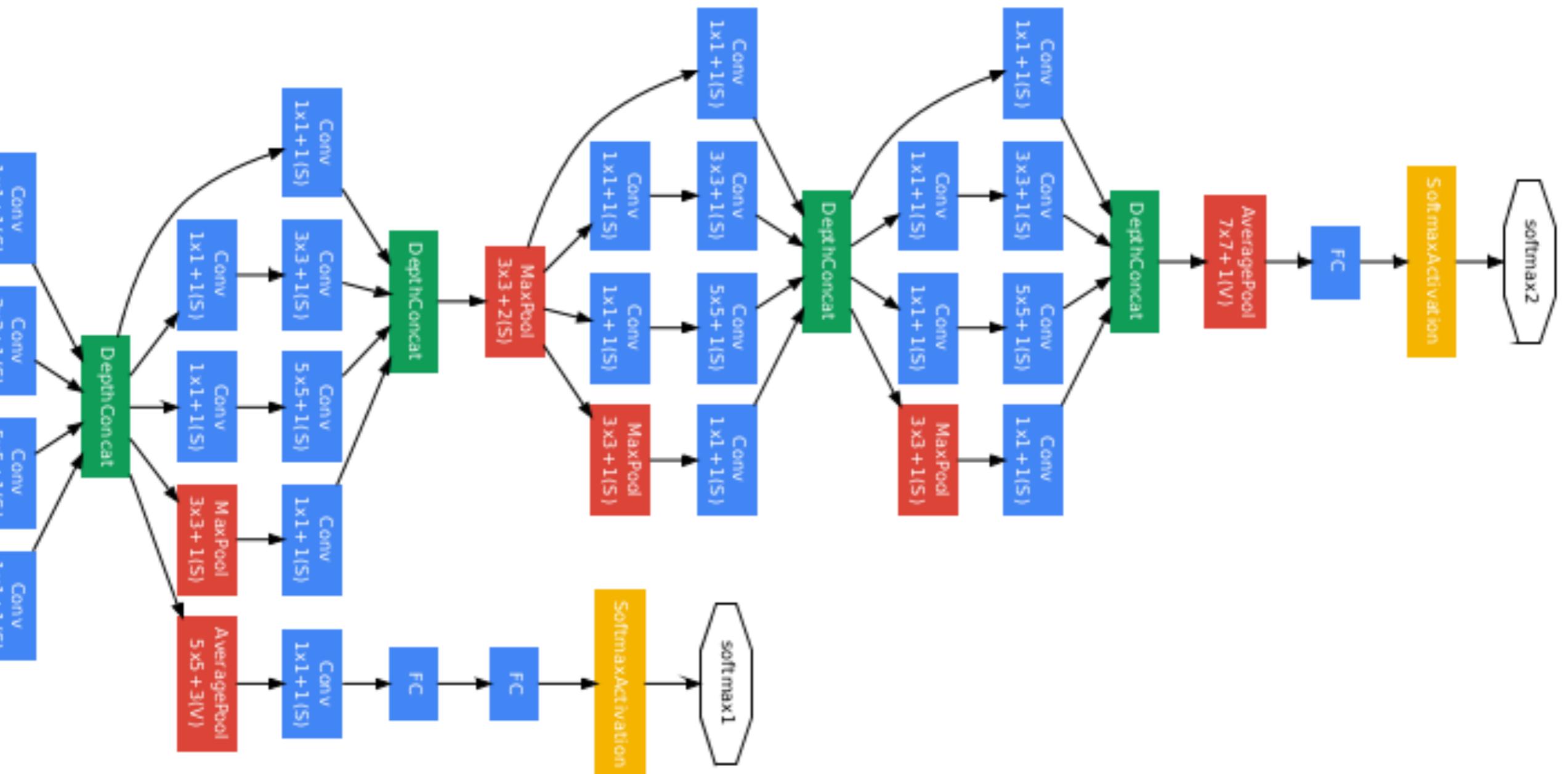
GOOGLENET



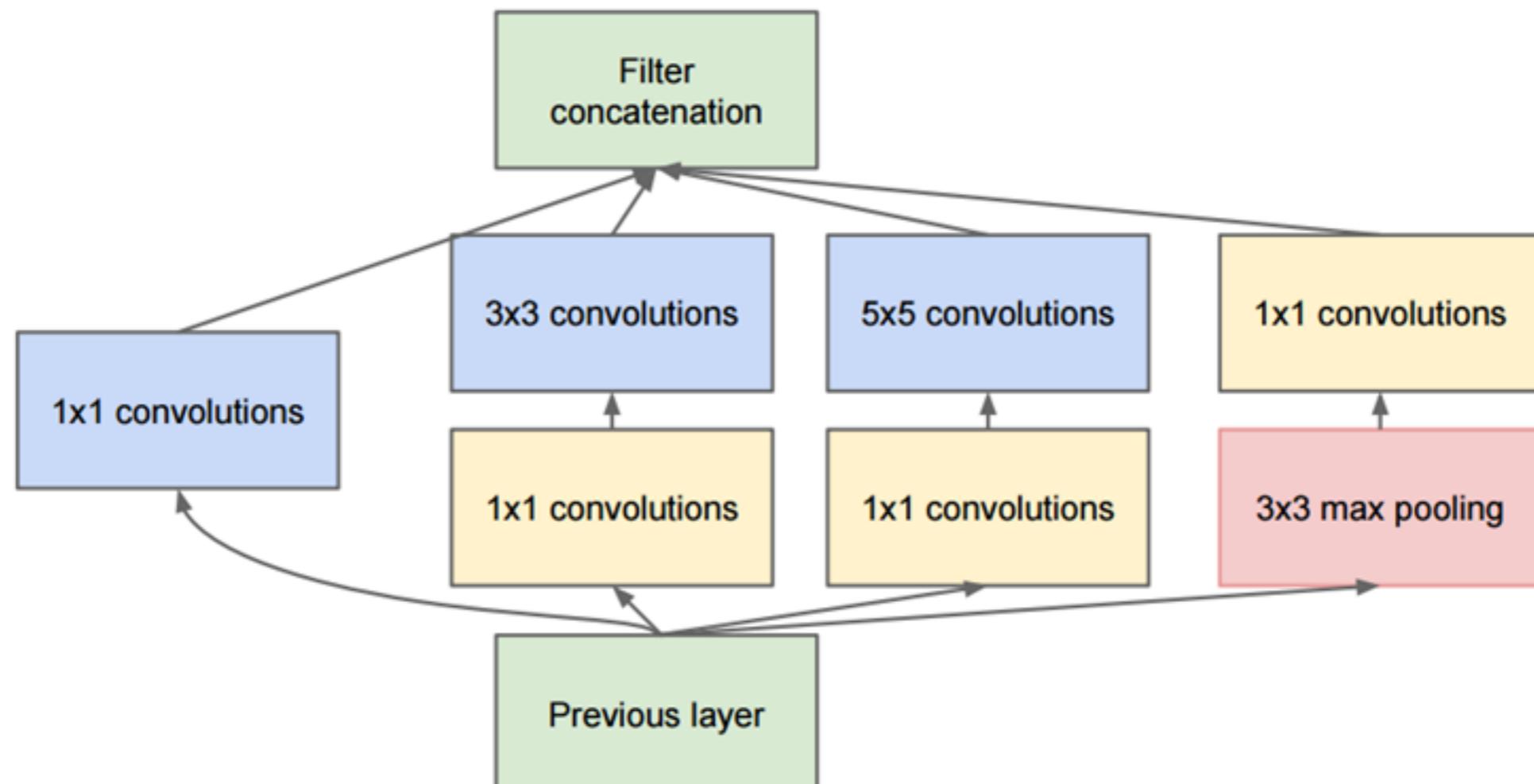
GOOGLENET



GOOGLENET

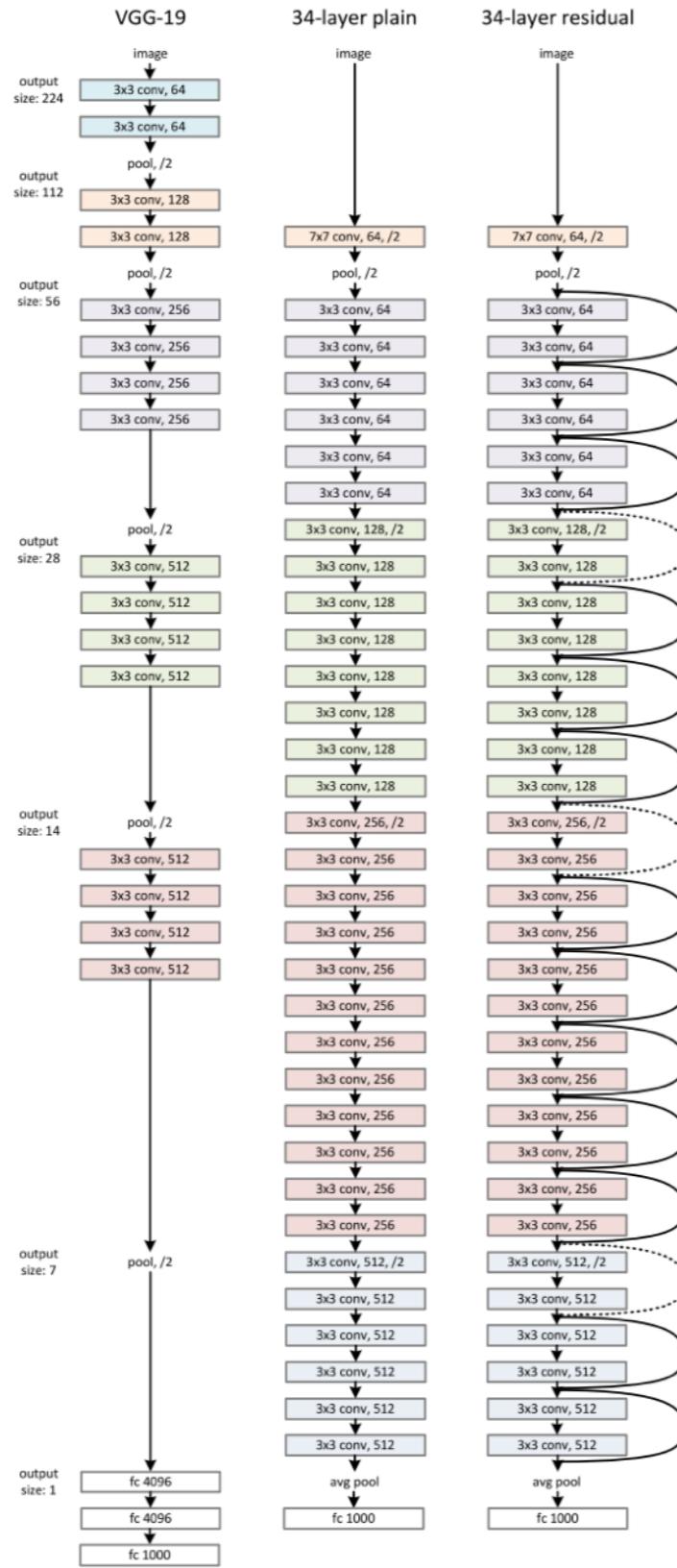


GOOGLENET

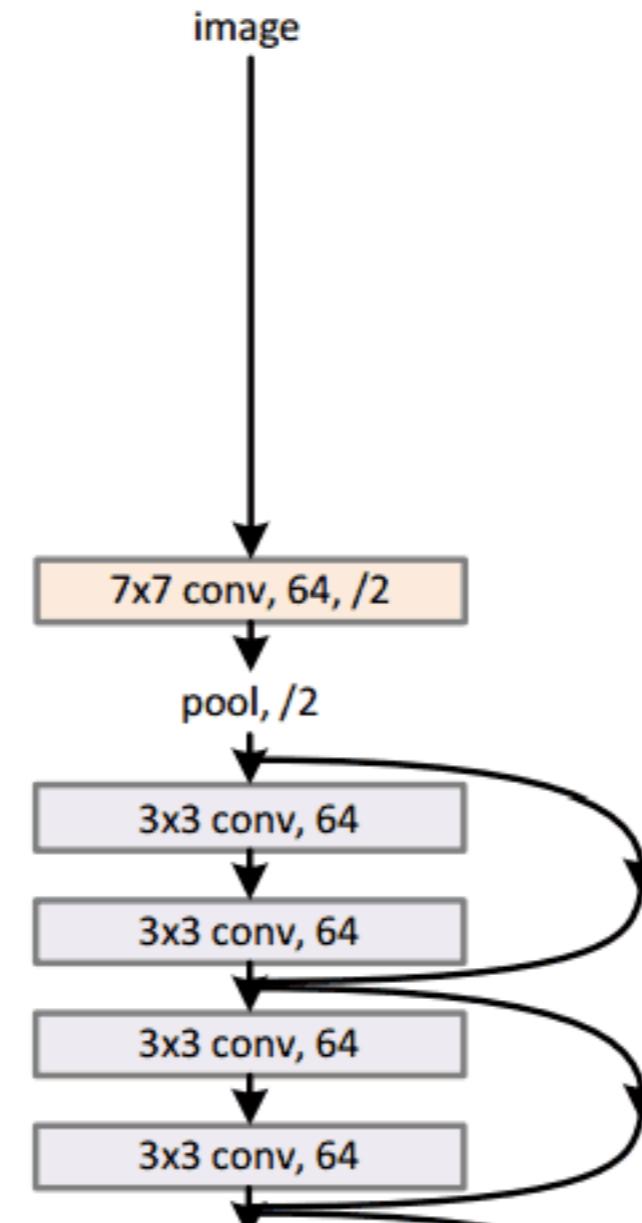


Full Inception module

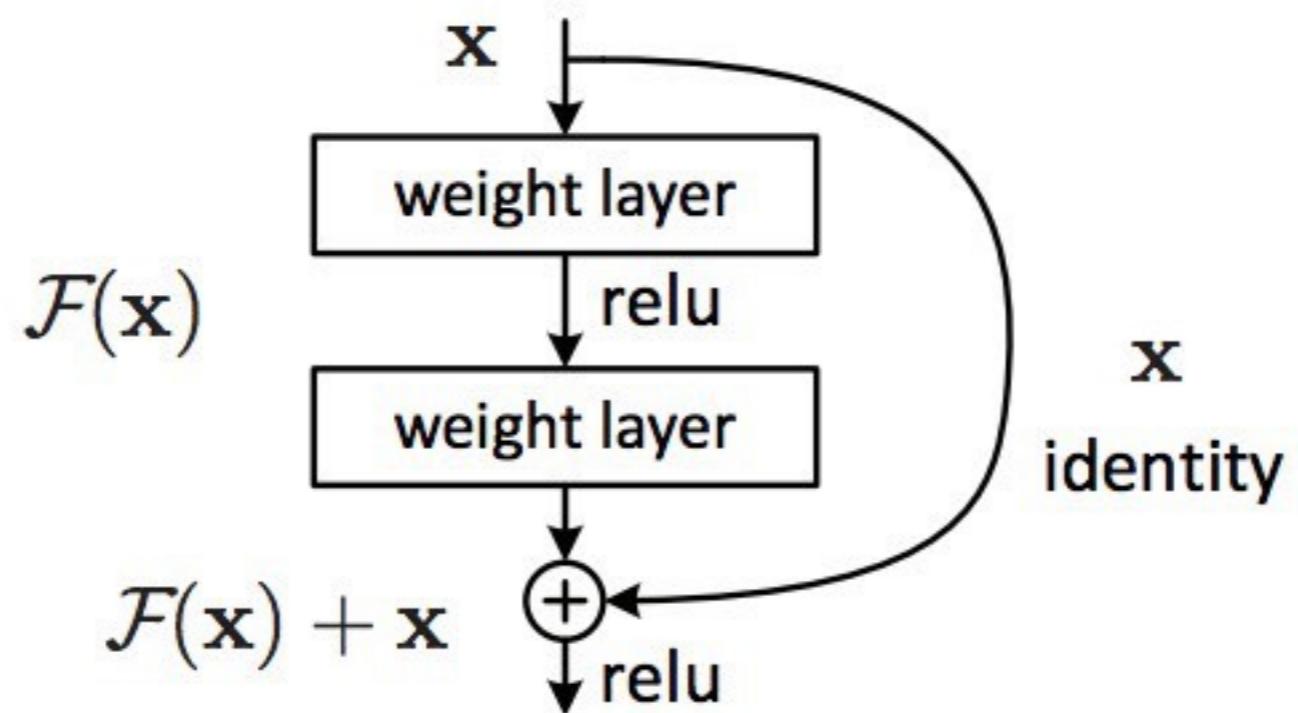
RESNET



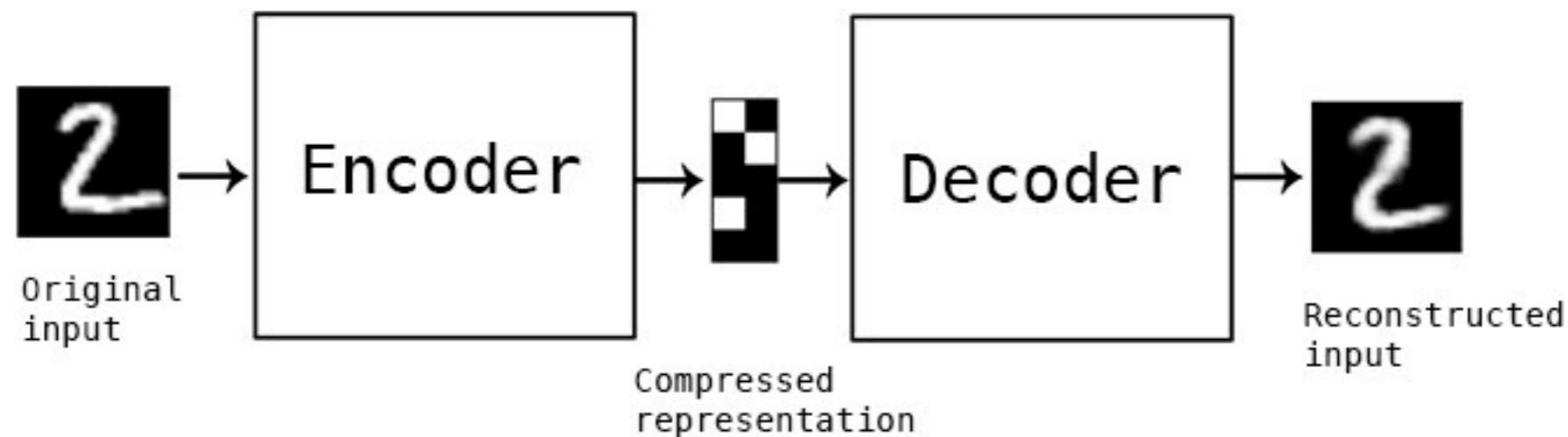
34-layer residual



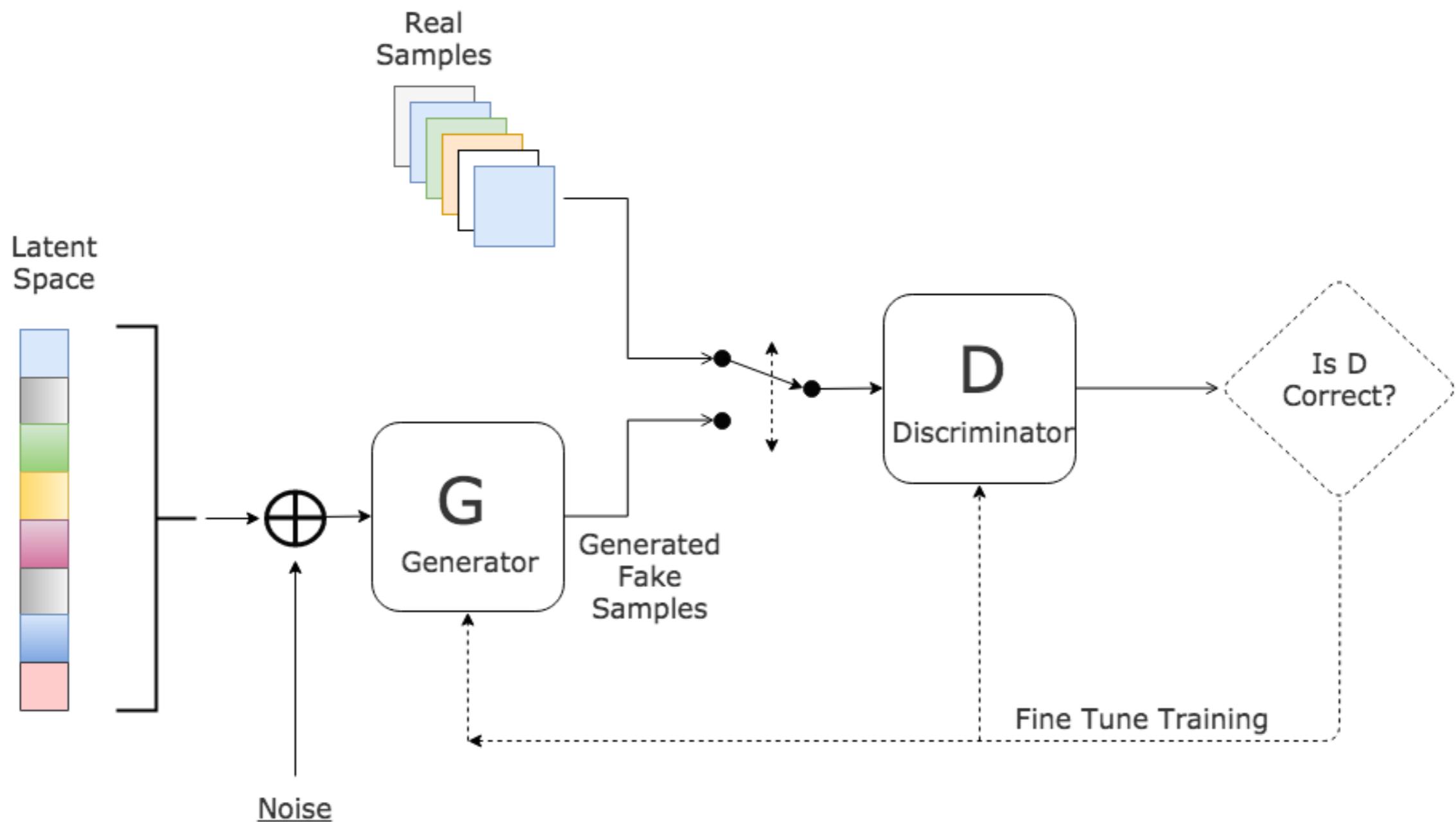
RESNET



AUTOENCODERS

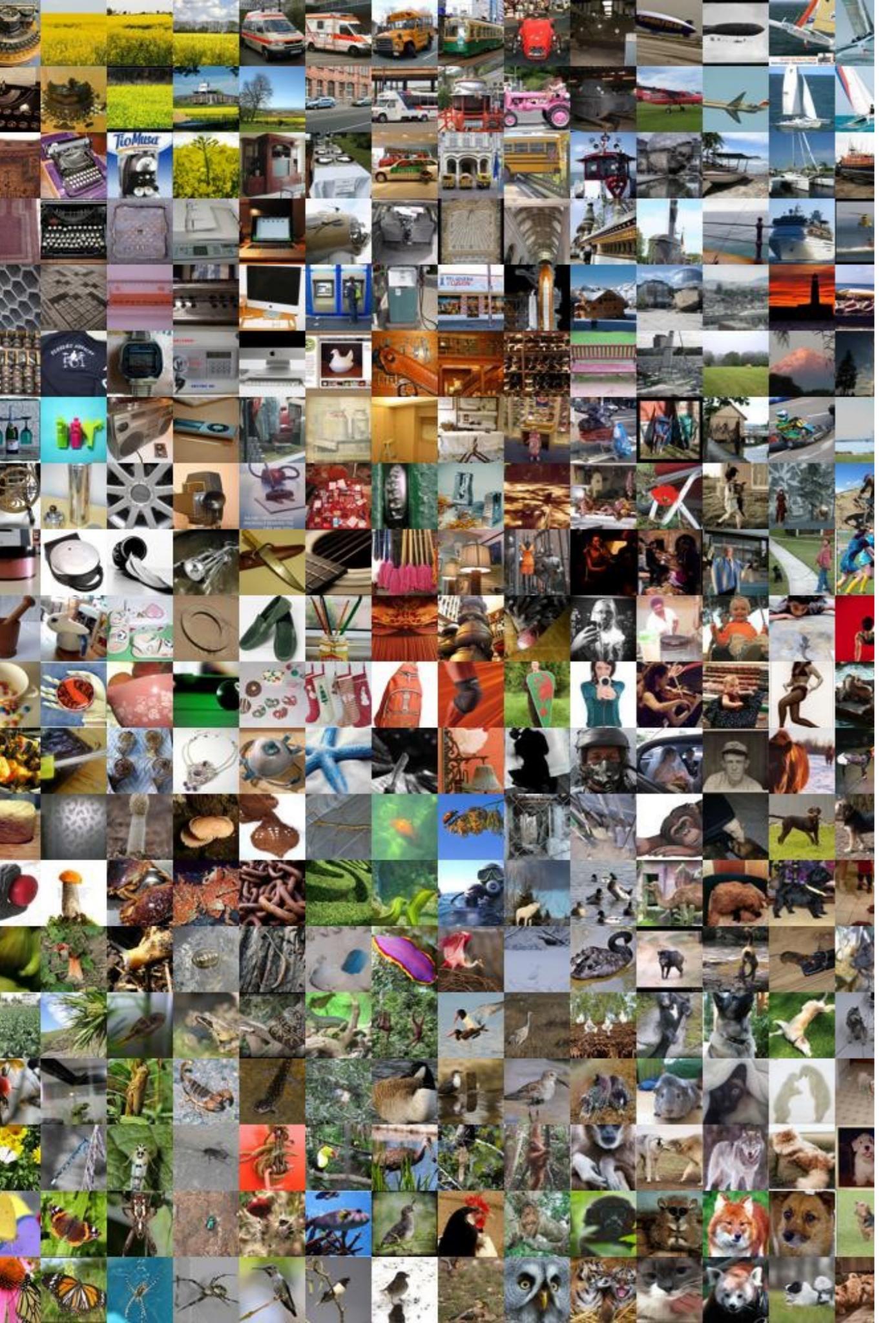


Generative Adversarial Network



POPULAR DATASETS

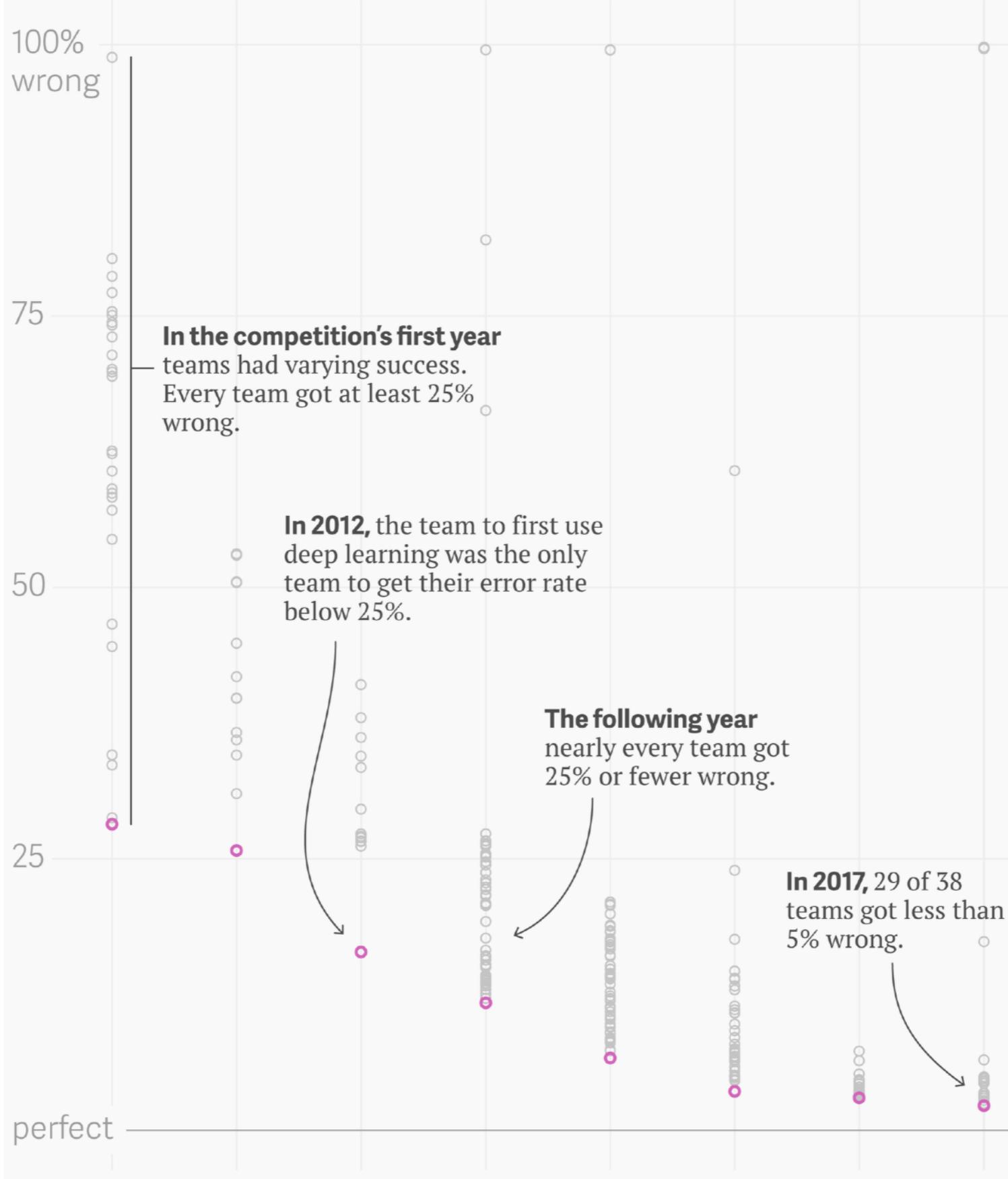
...to train your CNNs!



IMAGENET

- 10 millions hand-labelled images,
- 1 million with bounding boxes,
- Labels based on WordNet (hierarchical dictionary).

ImageNet Large Scale Visual Recognition Challenge results



CIFAR10

airplane



automobile



bird



cat



deer



dog



frog



horse



ship



truck



- 60.000 images for training,
- 6000 images for testing,
- 10 classes,
- Also: Version with 100 classes (CIFAR100).



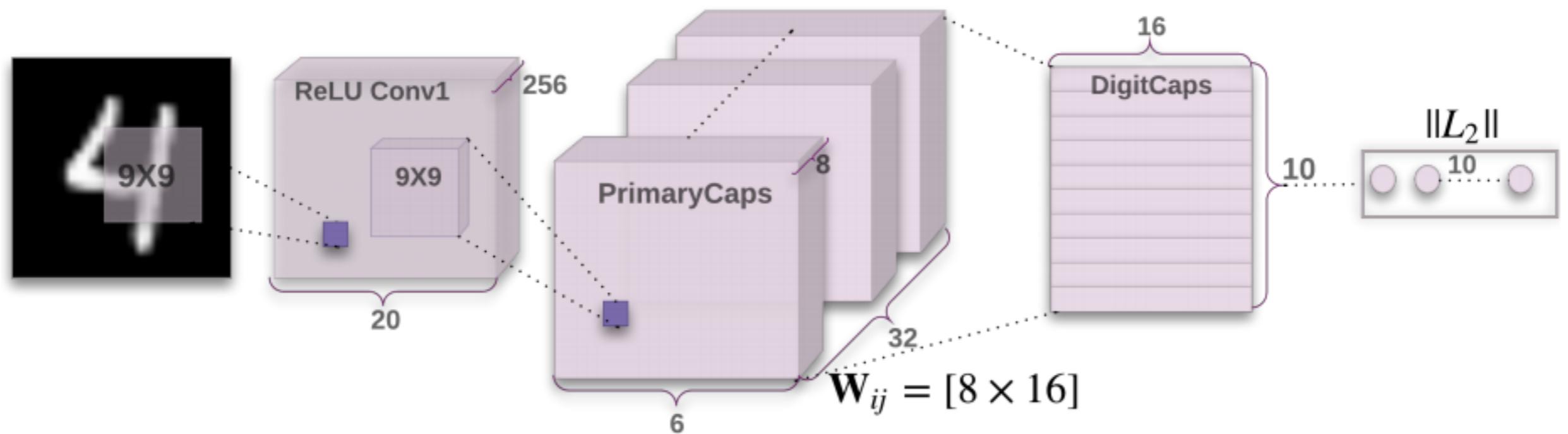
KAGGLE

- Competitions,
- 100+ Datasets,
- Community,
- Many code examples,
- Many CNNs challenges!

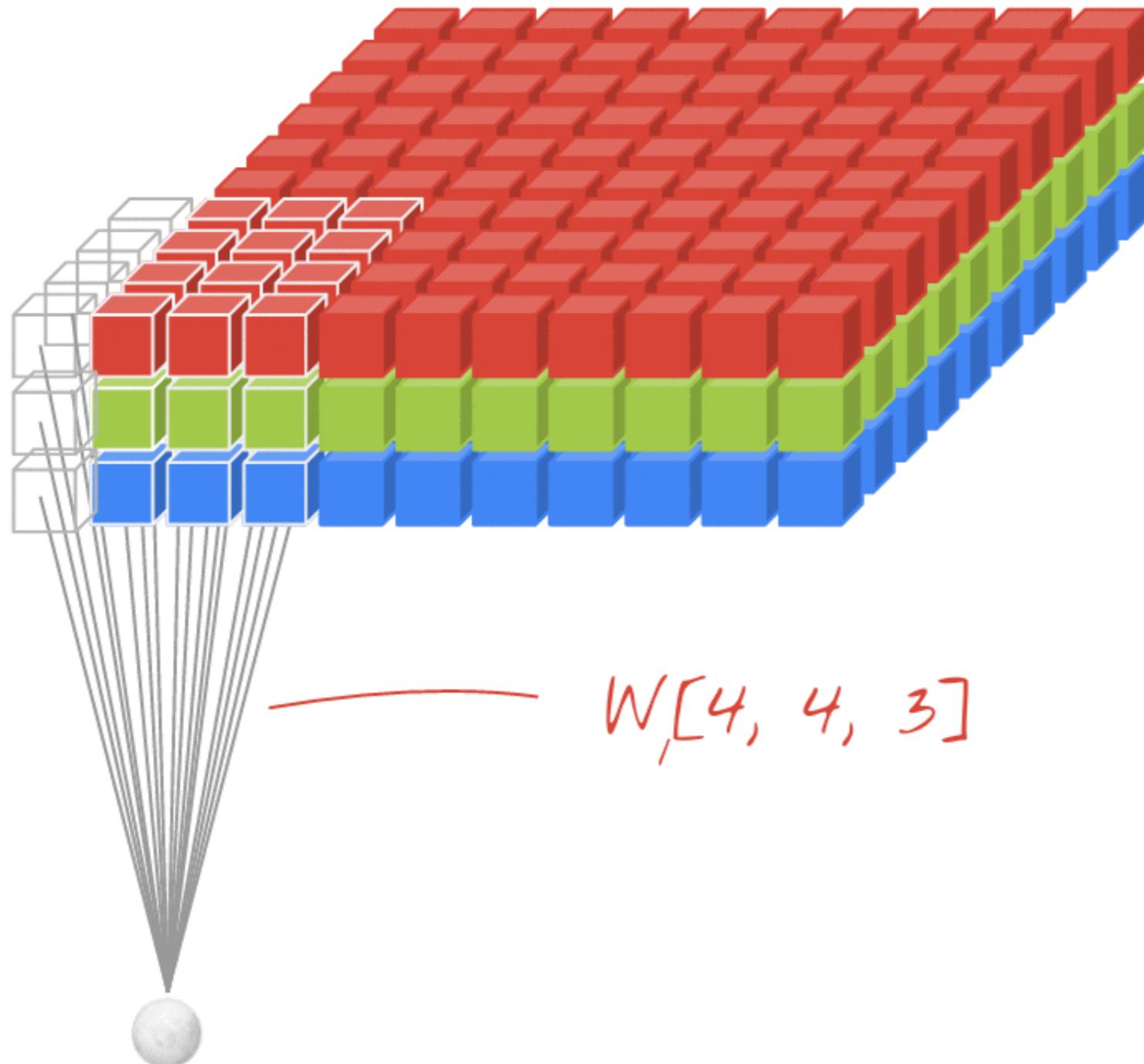
WHAT'S NEXT?

Future?

CAPSNET



3D CONVOLUTIONS



$W[4, 4, 3]$

THANKS!

Q&A