北京大學 PEKING UNIVERSITY

Visual Recognition – Part 2

Mu Yadong

Machine Intelligence Lab
Institute of Computer Science & Technology
Peking University

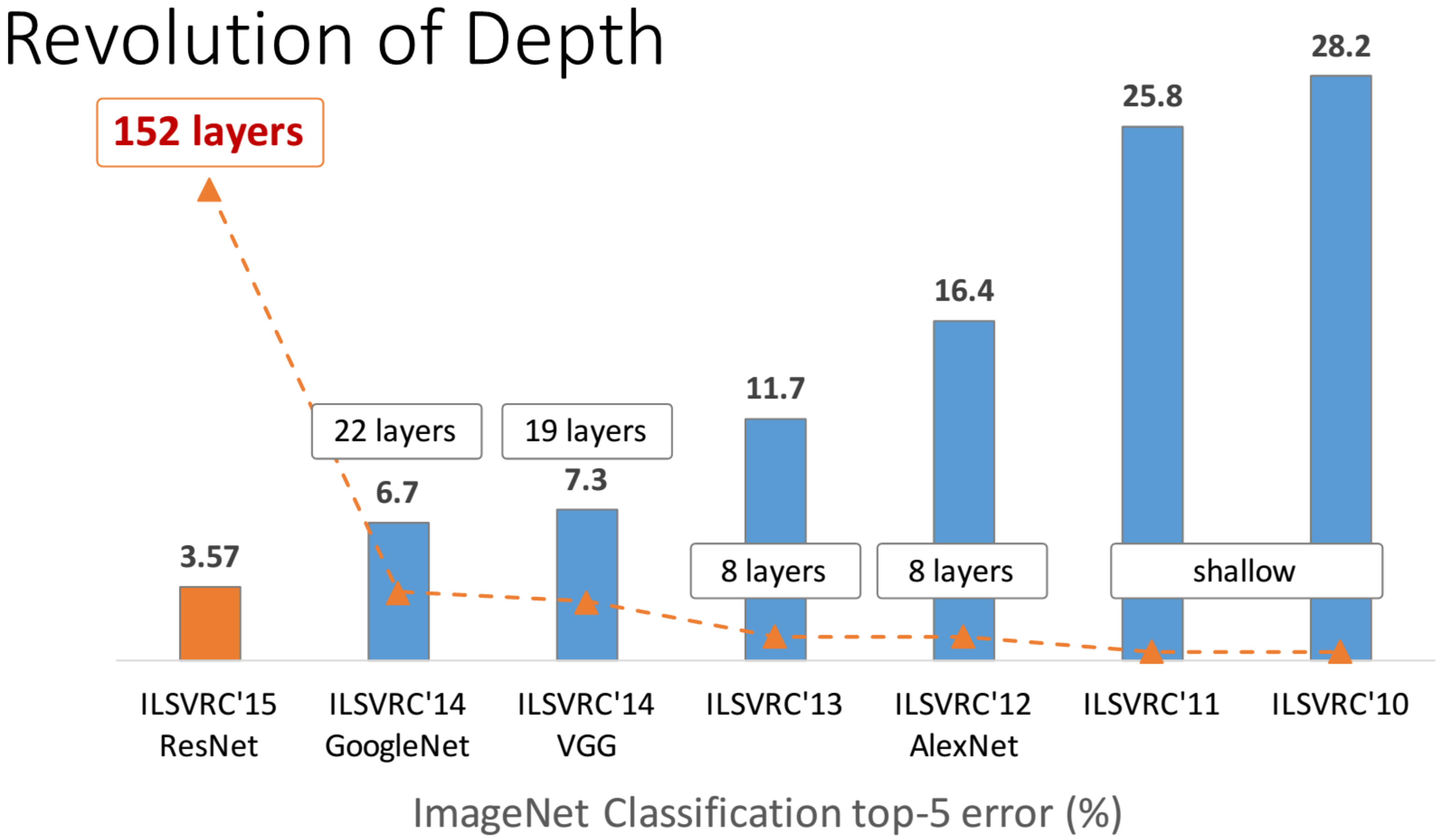
Several slides are adapted from related courses or tutorials.
Internal use only. Please do not distribute the slides.

Outline

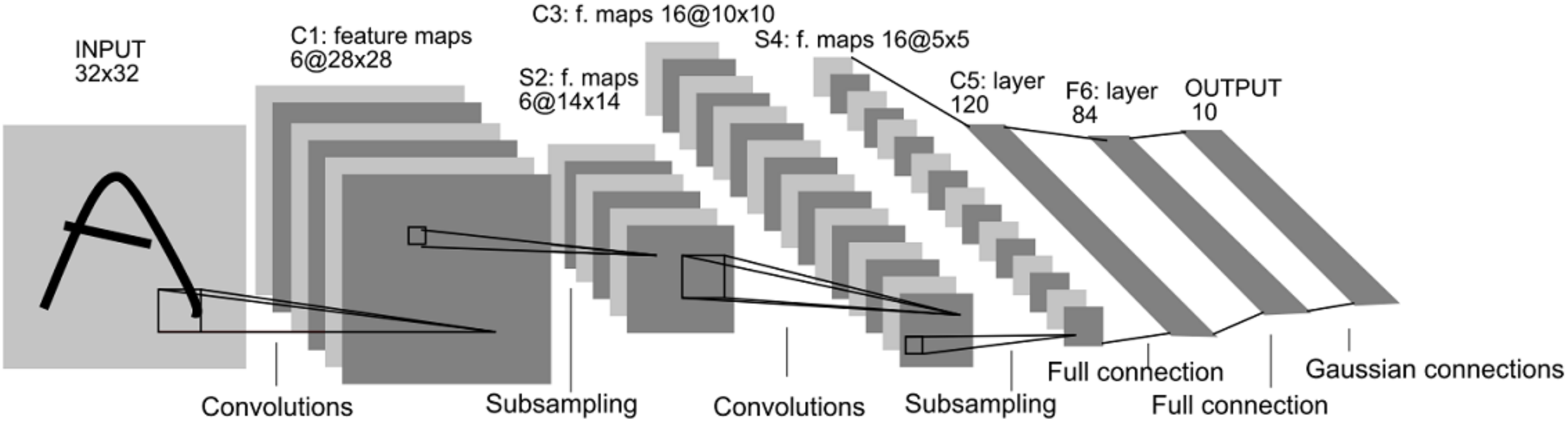
- **Why DL suddenly works? (AlexNet, 2012)**
- **Can it go deeper? (ResNet, 2015)**
- **Further extensions (DenseNet etc.)**

AlexNet

- Named after Alex Krizhevsky, proposed in 2012



LeNet-5

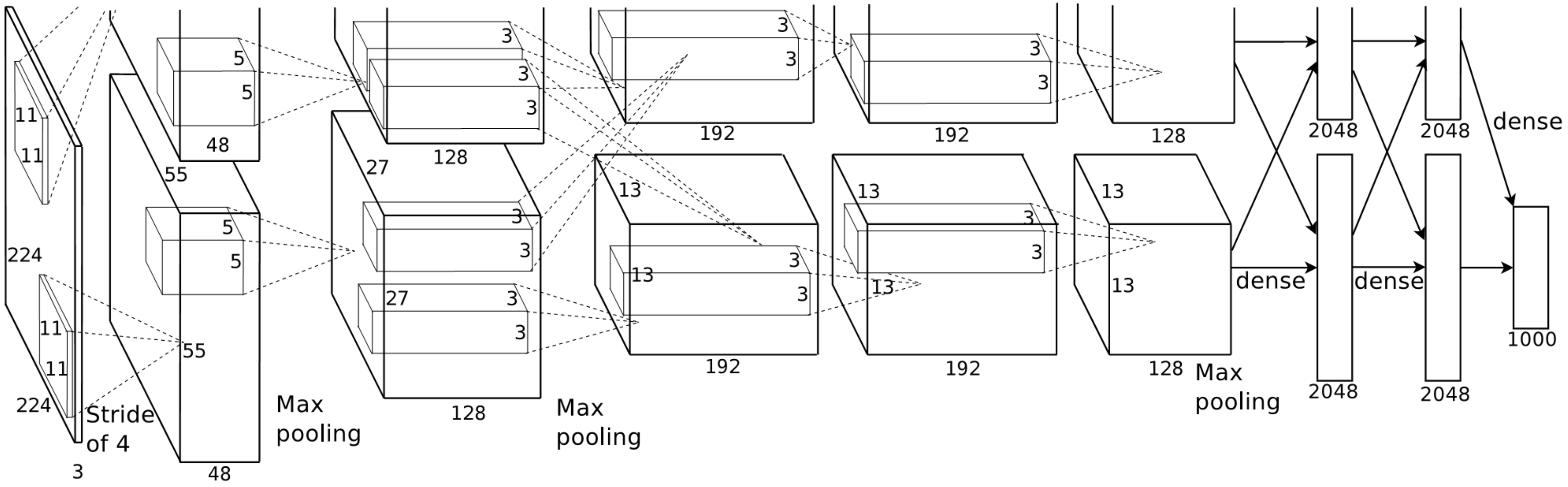


- Input: 32x32 pixel image. Largest character is 20x20 (All important info should be in the center of the receptive field of the highest level feature detectors)
- Cx: Convolutional layer
- Sx: Subsample layer
- Fx: Fully connected layer
- Black and White pixel values are normalized:
  E.g. White = -0.1, Black =1.175 (Mean of pixels = 0, Std of pixels =1)

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, november 1998.

AlexNet

- Much larger than LeNet-5
- Trained on two GTX 580 GPU
- Largest networks at its time
- Utilize multiple engineering tricks (dropout, ReLU)



Alex Krizhevsky et al., ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012

Why DL Suddenly works?

…It may be that the primary barriers to the success of neural networks were psychological (practitioners did not expect neural networks to work, so they did not make a serious effort to use neural networks)…

-- Goodfellow et al. “deep Learning”

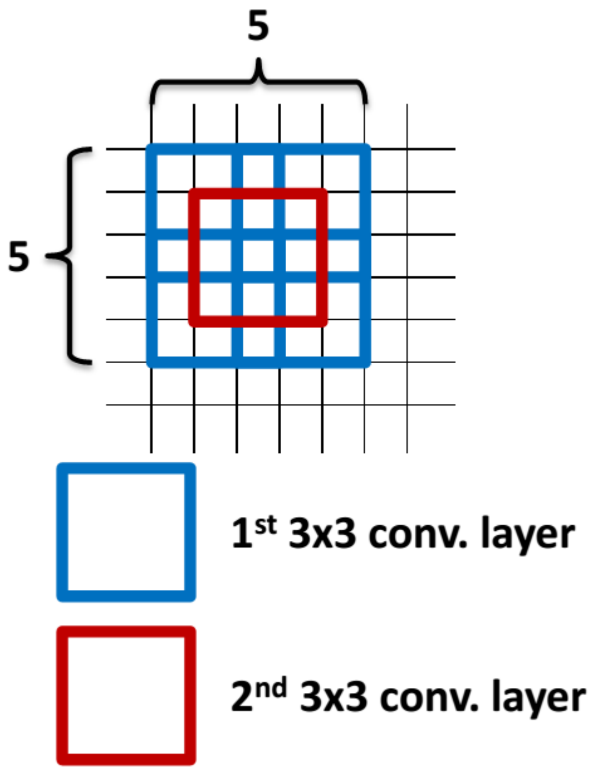
Why DL Suddenly works? – My Two Cents

- Emerging of big visual data
- GPU -> large network
- New engineering tricks (dropout, ReLU etc.)

VGG Net

Why 3x3 layers?

- Stacked conv. layers have a large receptive field
  - two 3x3 layers – 5x5 receptive field
  - three 3x3 layers – 7x7 receptive field
- More non-linearity
- Less parameters to learn
  - ~140M per net



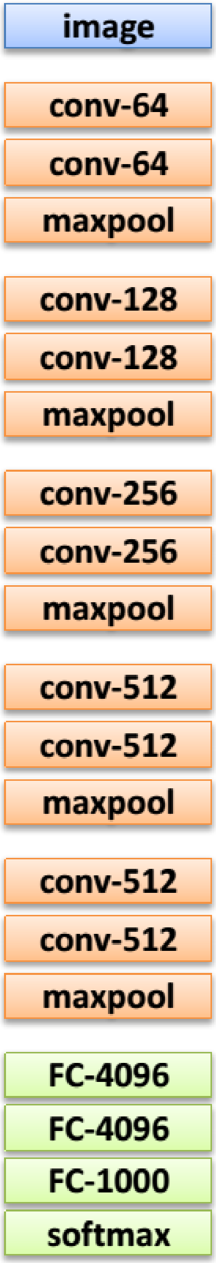
Network Design

Key design choices:

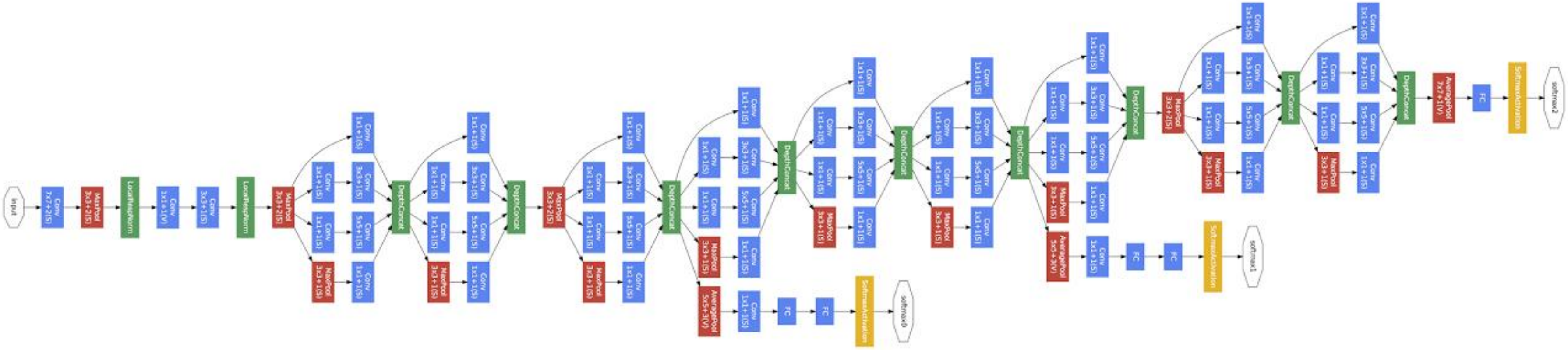
- 3x3 conv. kernels – very small
- conv. stride 1 – no loss of information

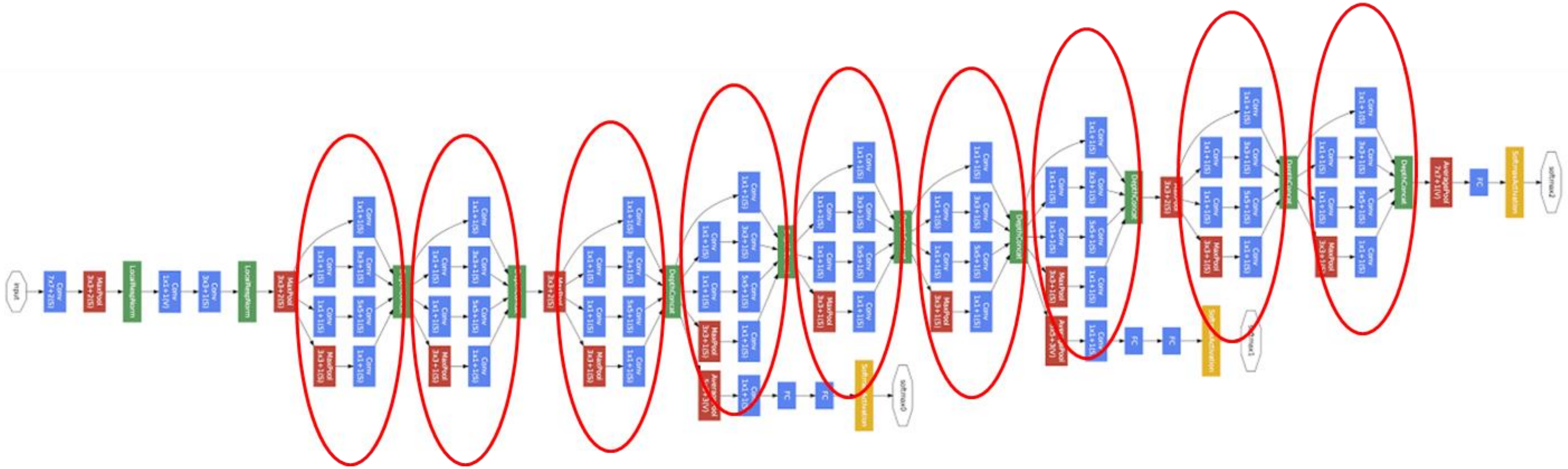
Other details:

- Rectification (ReLU) non-linearity
- 5 max-pool layers (x2 reduction)
- no normalisation
- 3 fully-connected (FC) layers



GoogLeNet



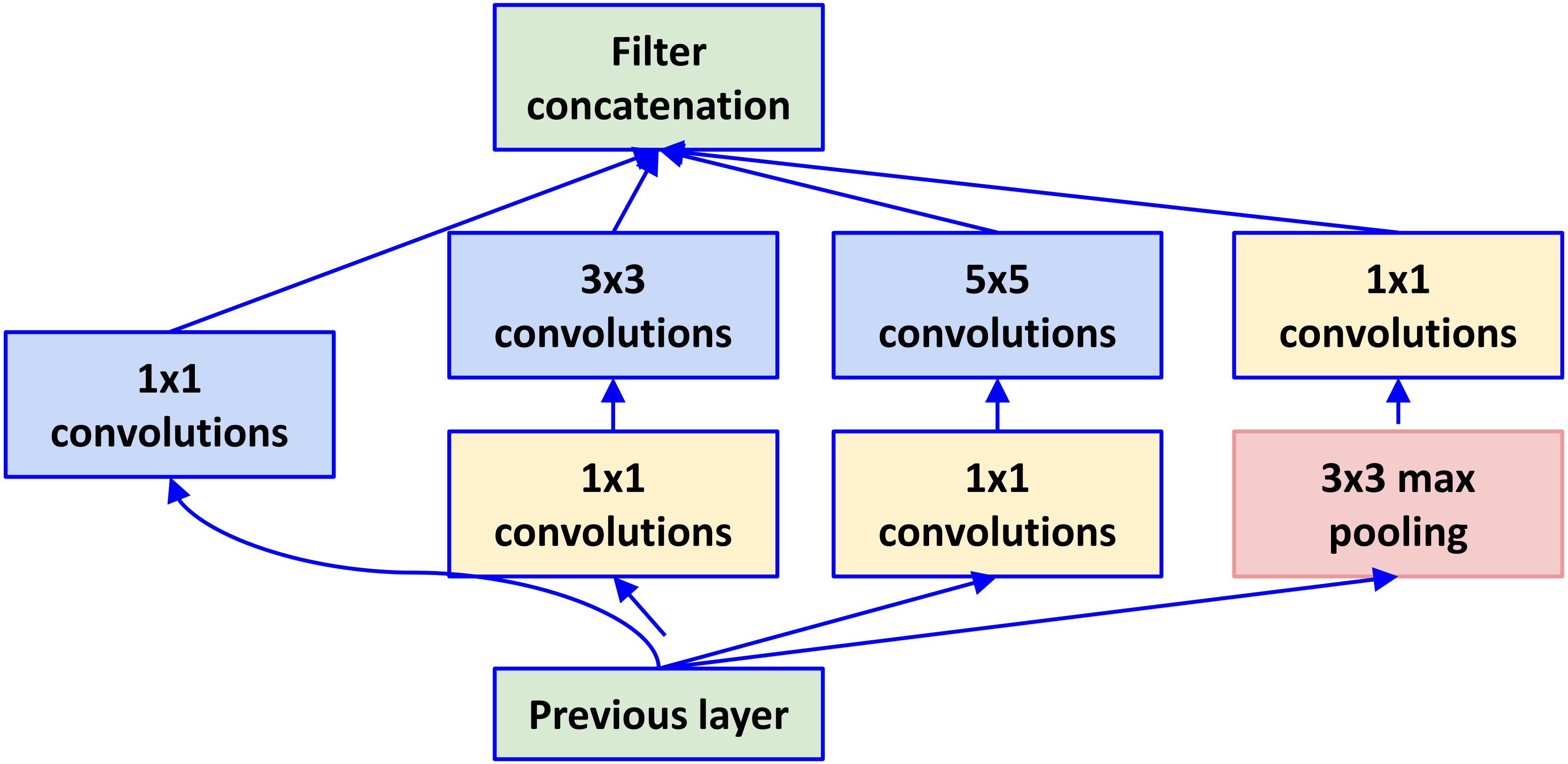


Inception

Network in a network in a network...

Convolution
Pooling
Softmax
Other

Inception module



Filter concatenation

1x1 convolutions

3x3 convolutions

5x5 convolutions

1x1 convolutions

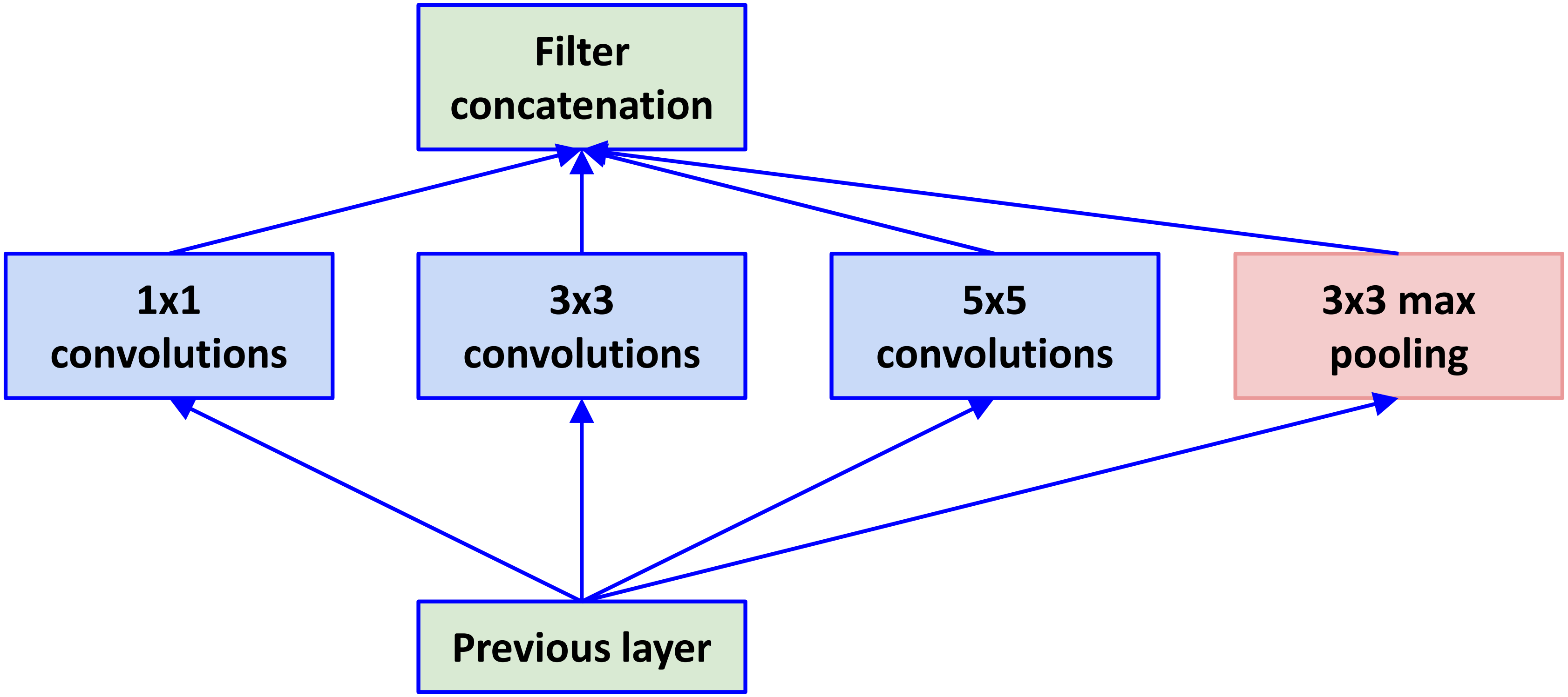
1x1 convolutions

1x1 convolutions

3x3 max pooling

Previous layer

Naive idea (does not work!)



ResNet

- See He Kaiming’s ICML tutorial

DenseNet

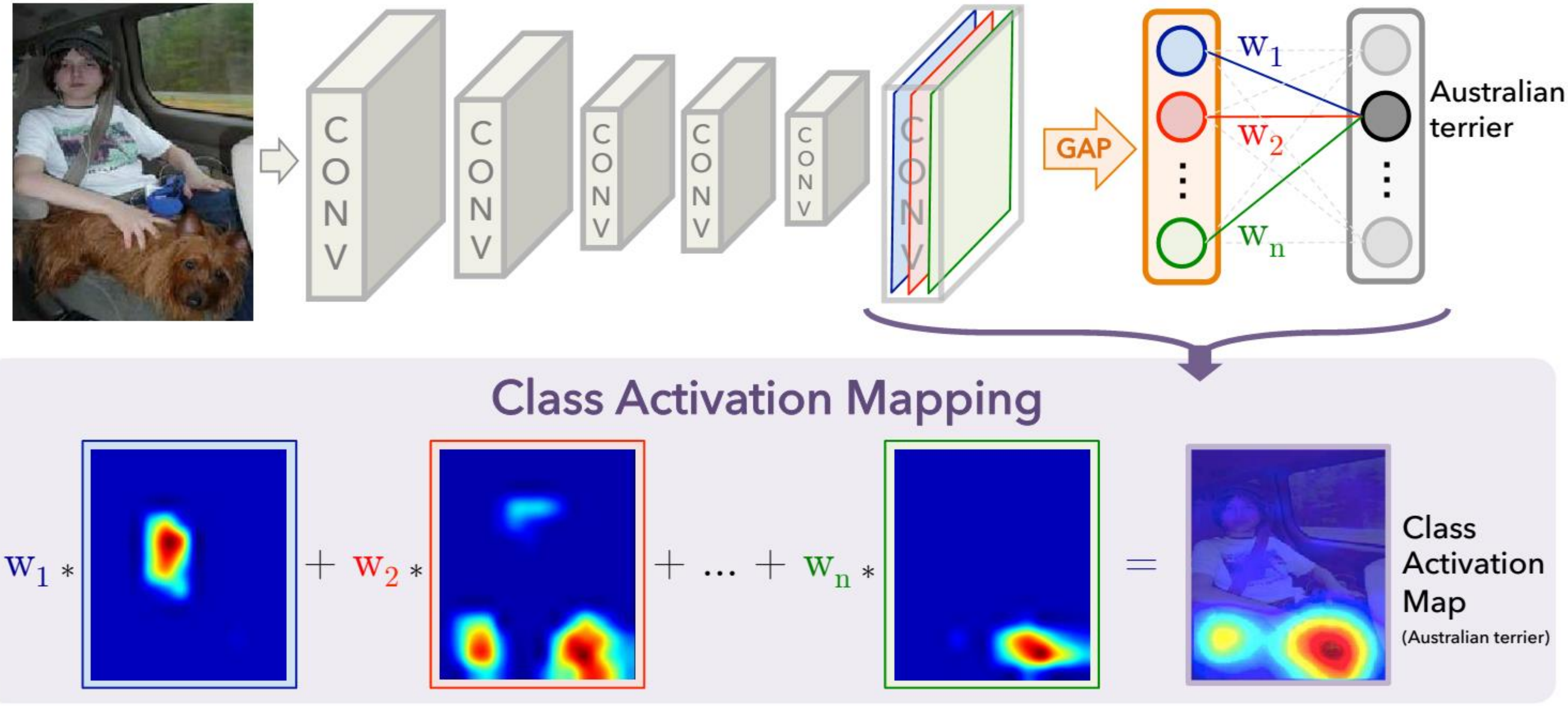
- See DenseNet’s CVPR slides

Dual Path Network

- See DualPathNet’s CVPR slides

Class Activation Map (CAM)

- Global Average Pooling



Class Activation Map (CAM)

- Examples