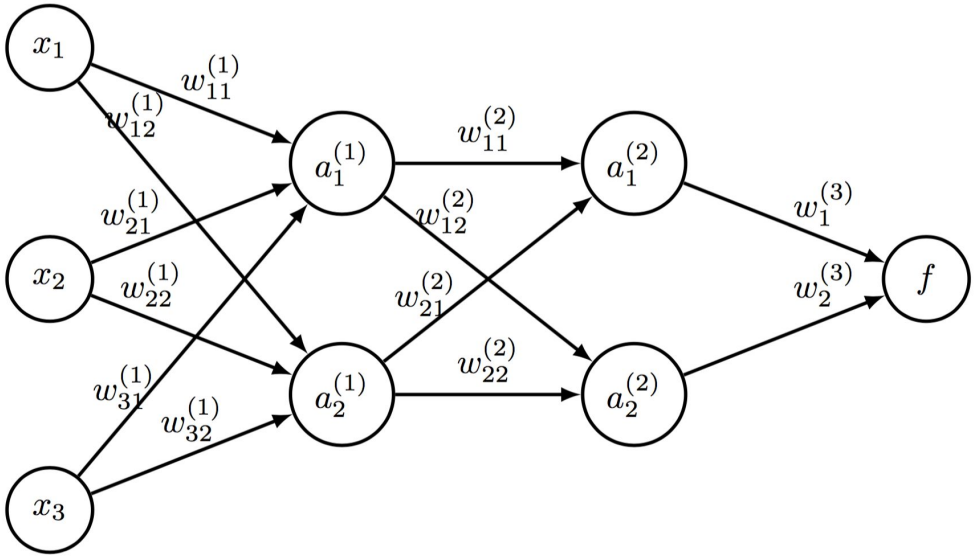
Neural Networks Tutorial (Part 2)

Kjong Lehmann

Solution HW4 Question 1

Xiaoming designed the following neural network to predict his grades in exams. He bases the predictions on his nervous degree (x_1), mood(x_2) and weather of the exam day (x_3). In hidden layers, he uses sigmoid function(x) = 1/(1 + e^x). In output layer, no activation function is used. As in many other regression tasks, he uses L_2 loss function: L = (y-f)^2



1. Forward pass, Compute Loss
2. Dropout with p=0.4. Expected loss given training
3. SGD update to weight w_21

Notable initialization issues

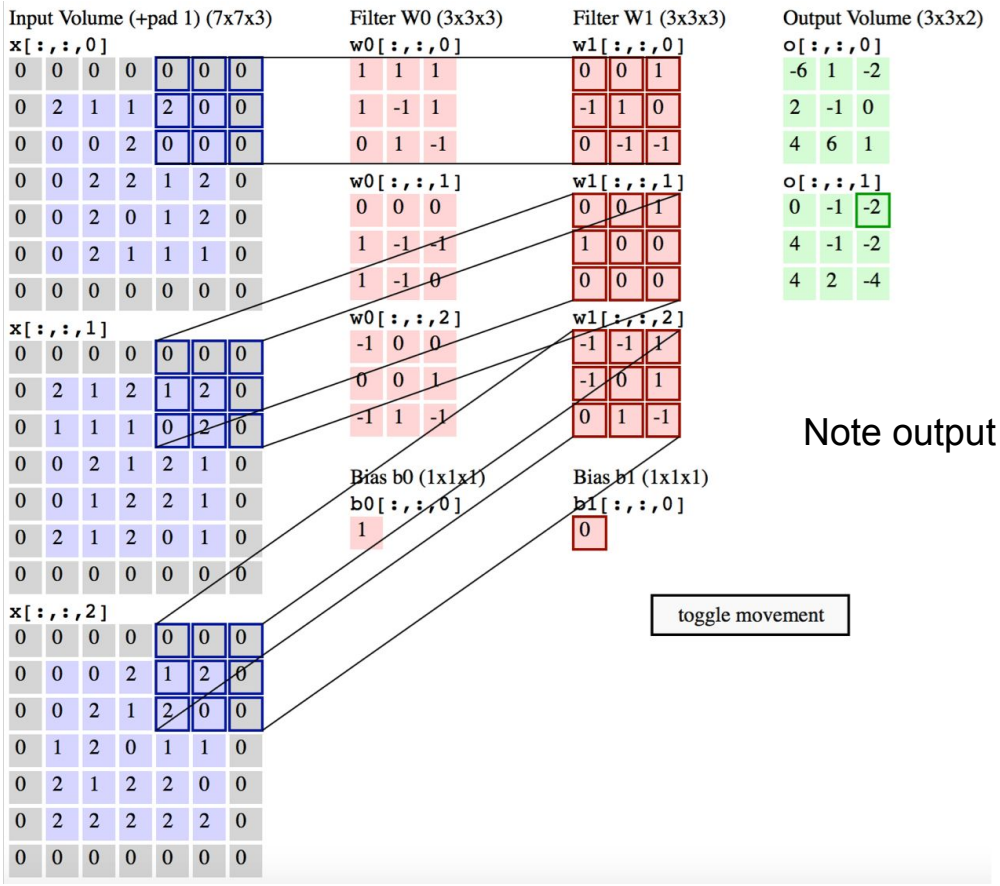
- Zero initialization
- Recurrent initializations
- Random Initialization with large spread
- Random Initialization with small spread

Why do dropouts work?

Linear regression and dropouts:

See
https://nlp.stanford.edu/pubs/sidaw13fast.pdf

Convolutional Neural Network



Note output volume

Convolutional neural network

Consider a convolution Layer. The input consists of 6 feature maps of size 20 × 20. The output consists of 8 feature maps, and the filters are of size 5 × 5. The convolution is done with a stride of 2 and zero padding, so the output feature maps are of size 10×10.

Number weights in convolution layer:

Number of weights if fully connected layer but input/output dimension are the same:

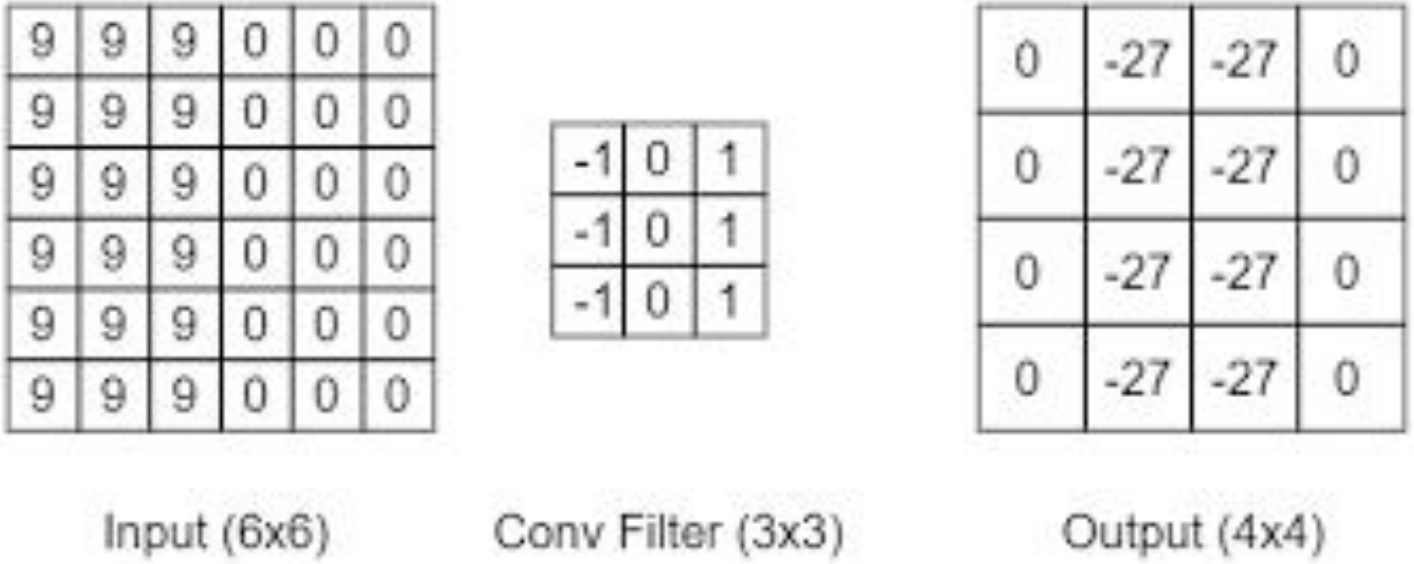
Convolutional neural network

There's one filter for each pair of an input and output feature map, and the filters are each 5×5. Therefore, the number of weights is 6×8×5×5 = 1200.

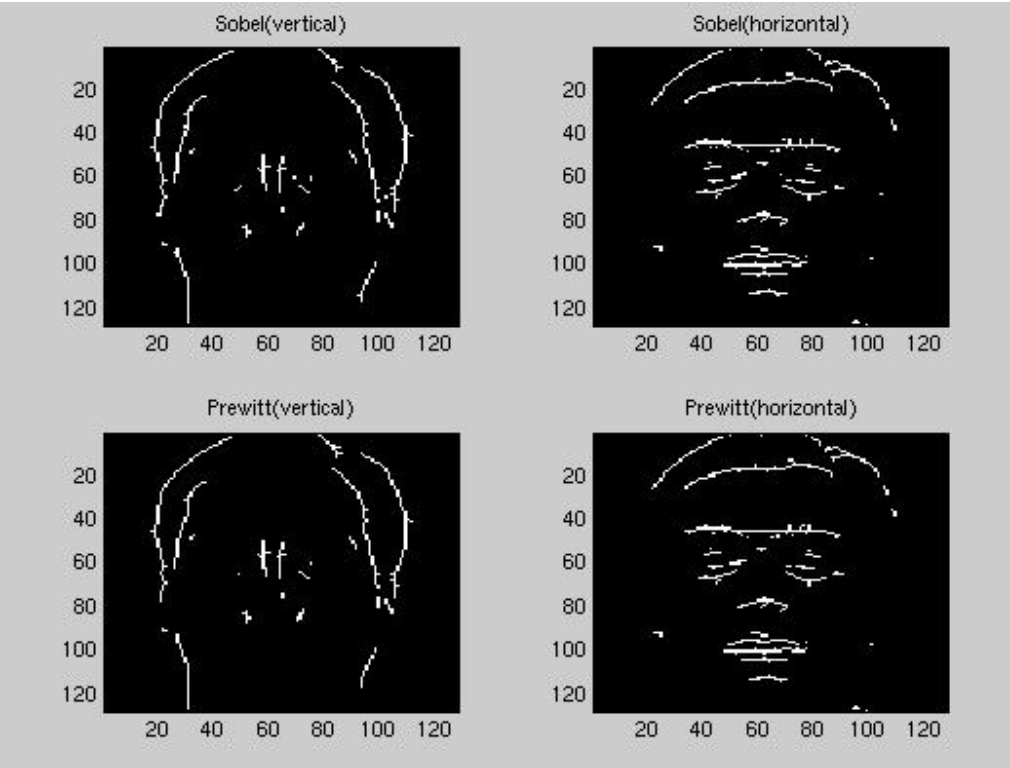
There are 20×20×6 units in the input layer and 10×10×8 units in the output layer, so the number of weights is 20×20×6×10×10×8 = 1,920,000.

http://www.cs.toronto.edu/~rgrosse/courses/csc321_2017/exams/midterm_sol.pdf

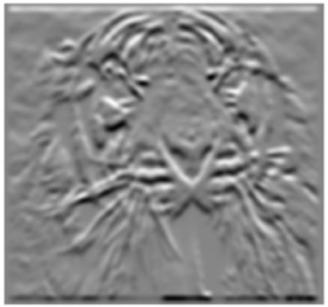
Filter detection



Different filters

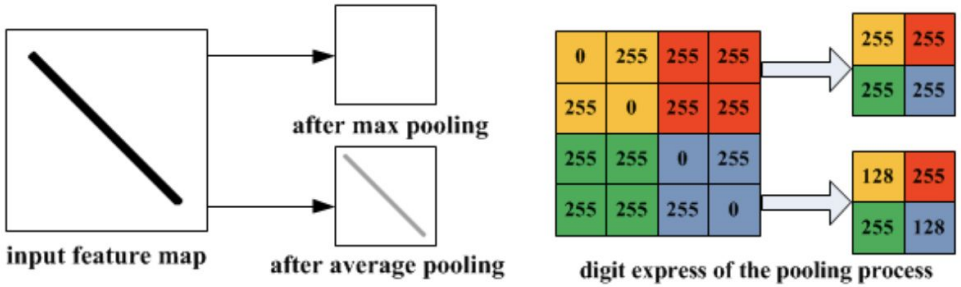


Convolutional Neural Network

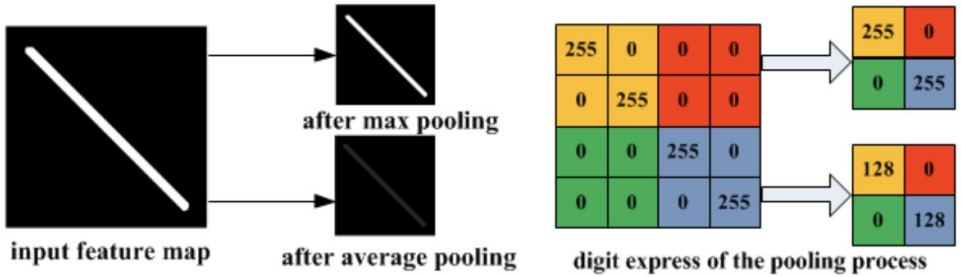


What are the filters used here?

Pooling



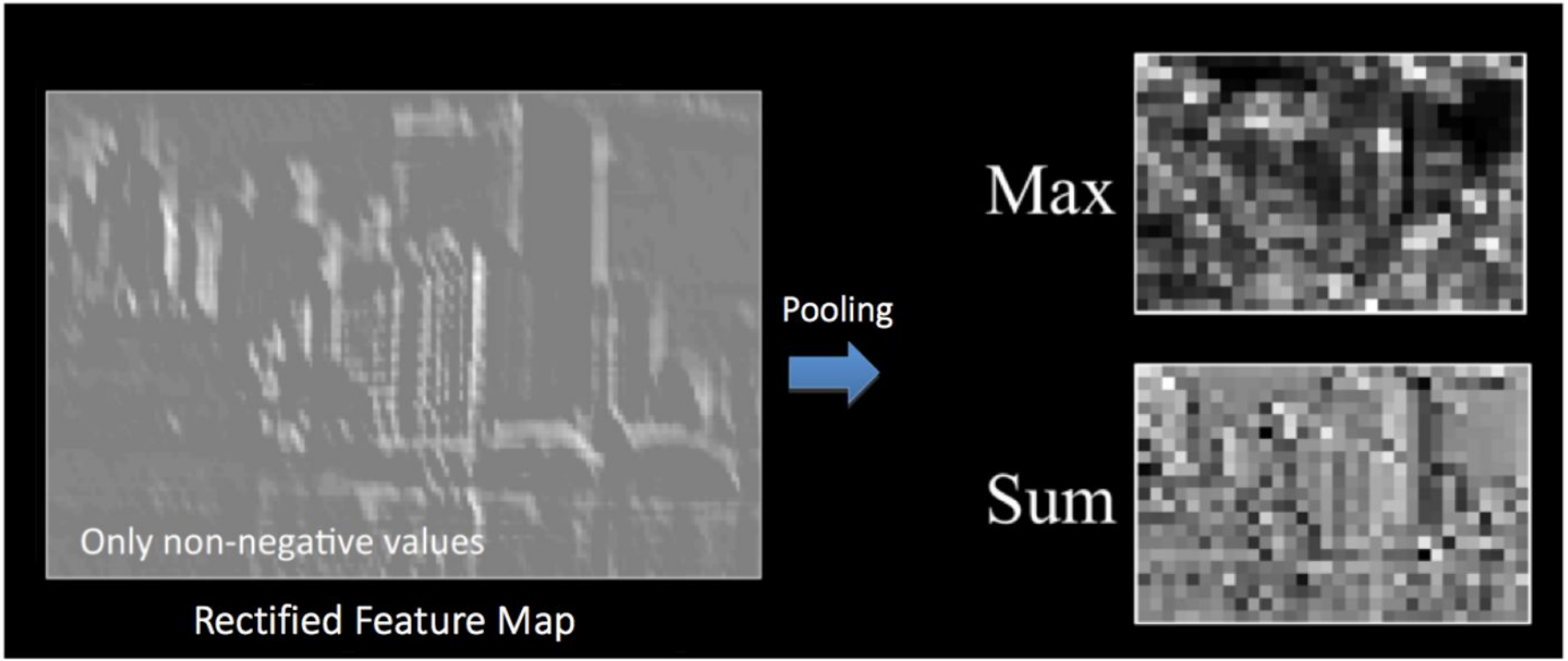
(a) Illustration of max pooling drawback



(b) Illustration of average pooling drawback

DOI: 10.1007/978-3-319-11740-9_34

Pooling



http://mlss.tuebingen.mpg.de/2015/slides/fergus/Fergus_1.pdf

Convolutional neural network

Alice and Bob implemented two neural networks for recognizing handwritten digits from 16 × 16 grayscale images. Each network has a single hidden layer, and makes predictions using a softmax output layer with 10 units, one for each digit class.

-) Alice's network is a convolutional net. The hidden layer consists of three 16 × 16 convolutional feature maps, each with filters of size 5 × 5, and uses the logistic nonlinearity. All of the hidden units are connected to all of the output units.

-) Bob's network is a fully connected network with no weight sharing. The hidden layer consists of 768 logistic units (the same number of units as in Alice's convolutional layer).

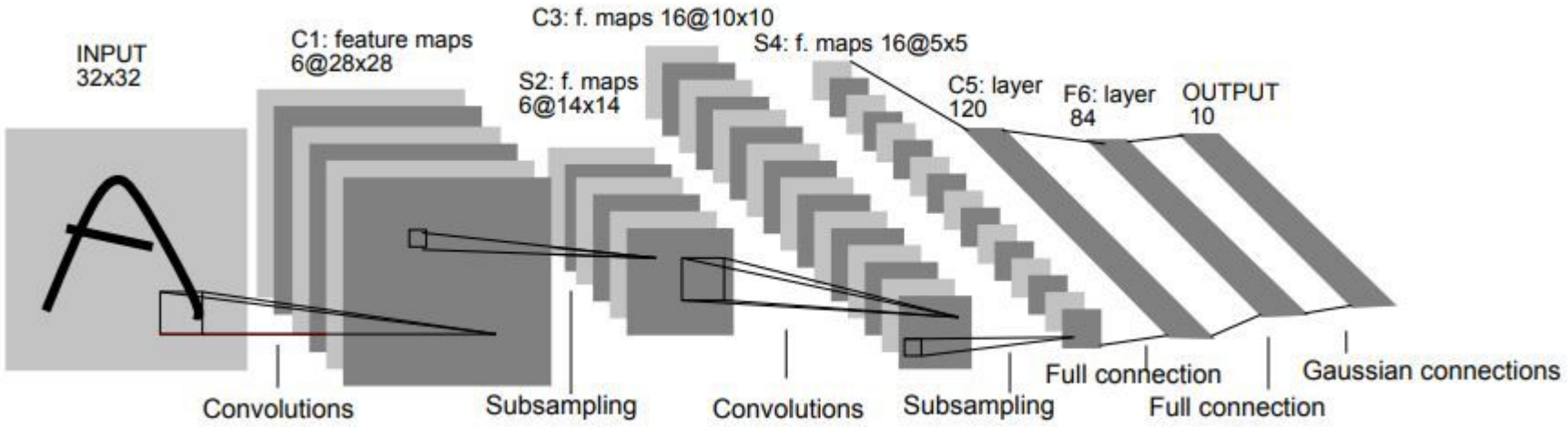
Convolutional neural network

The inputs to the convolution layer are a linear function of the images. In Bob's network, the hidden units can compute any linear function of the images; by contrast, Alice's convolutional layer is more restricted because of weight sharing and local connectivity. The advantage of Bob's network is that it is more powerful, i.e. it can compute any function Alice's network can compute, plus some additional functions. Advantages of Alice's network include:

(a) It has fewer parameters, so it is less likely to overfit.
(b) It has fewer connections, so it requires fewer arithmetic operations to compute the activations or the weight gradients.

http://www.cs.toronto.edu/~rgrosse/courses/csc321_2017/practice_tests/midterm2015_2_solutions.pdf

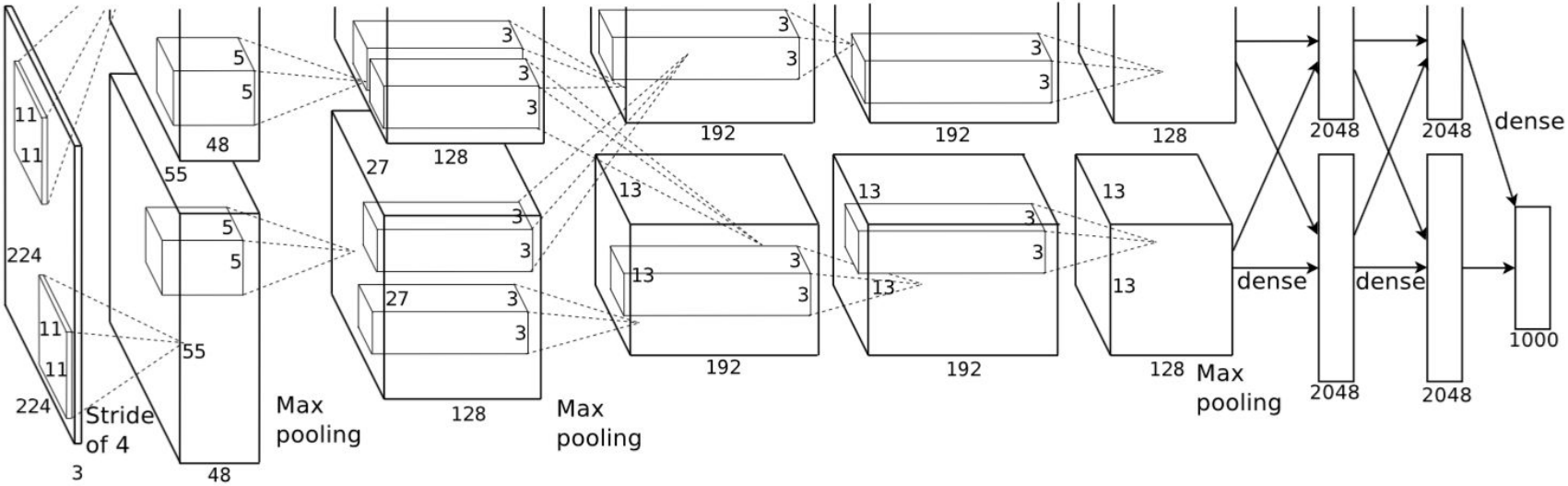
LeNet-5 (Digit recognition)



Lenet

- MNIST dataset (Digits)
- Tanh/Sig functions
- First 7-level convolutional neural network (1998)
- Convolution,Pooling,Convolution,Pooling plus two FC Layers
- 60k parameters

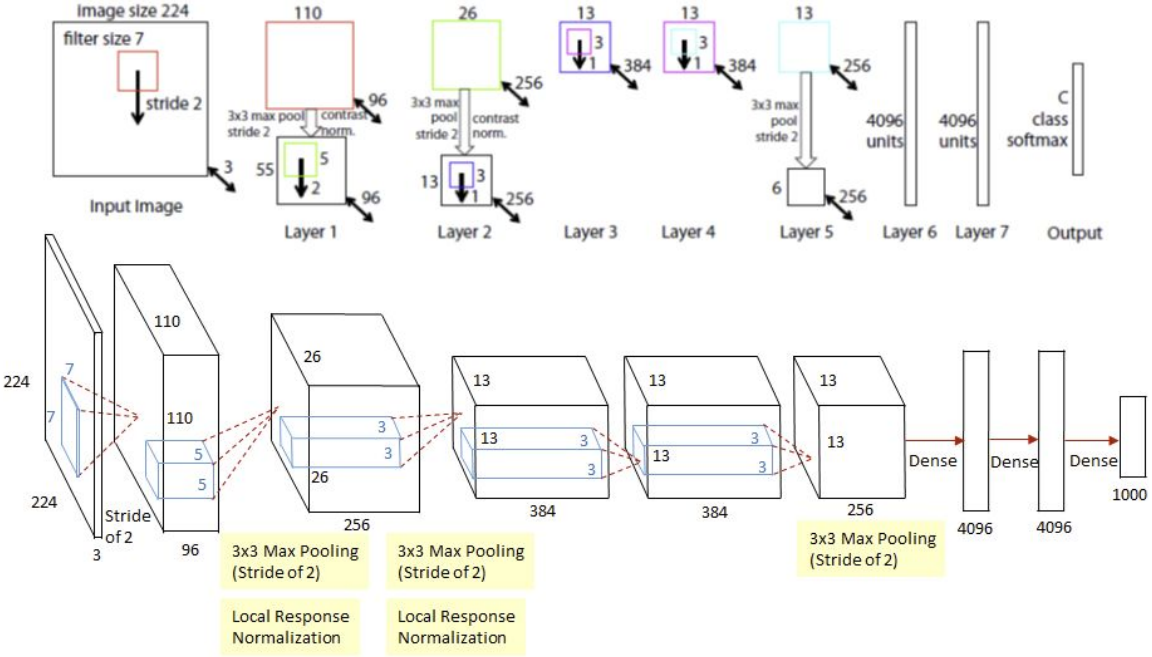
AlexNet (2012)



AlexNet (2012)

- From 26% error to about 15%
- ImageNet data (about 15mill images)
- ReLU
- data augmentation
- Dropout
- SGD with momentum
- 60 Million Parameters

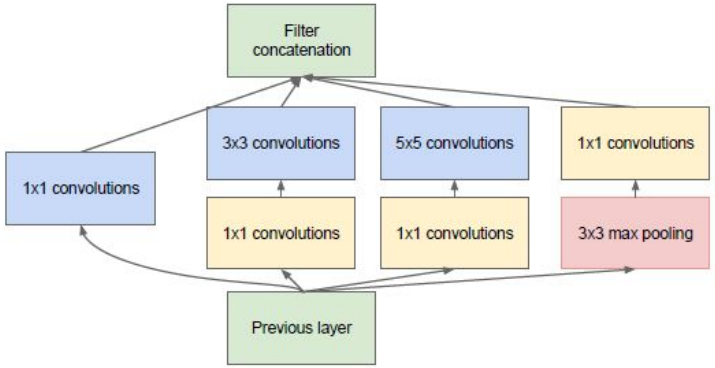
ZFNet (Class example)



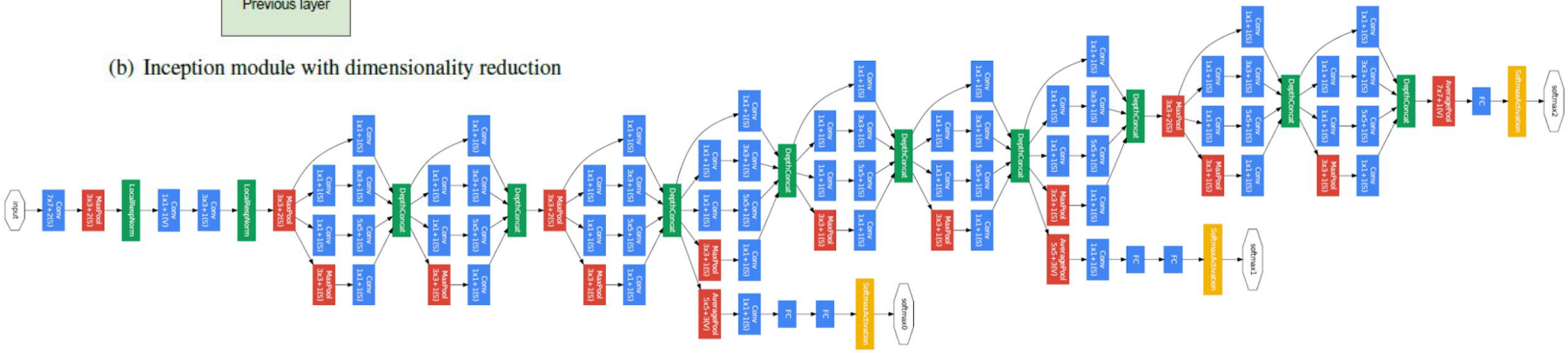
ZFNet (Class example)

- Similar to AlexNet (some mods: Smaller Filter etc)
- Achieved better accuracy on less training examples

GoogleNet



(b) Inception module with dimensionality reduction



GoogleNet

- 4 million parameter
- No fully connected layers
- Average pooling