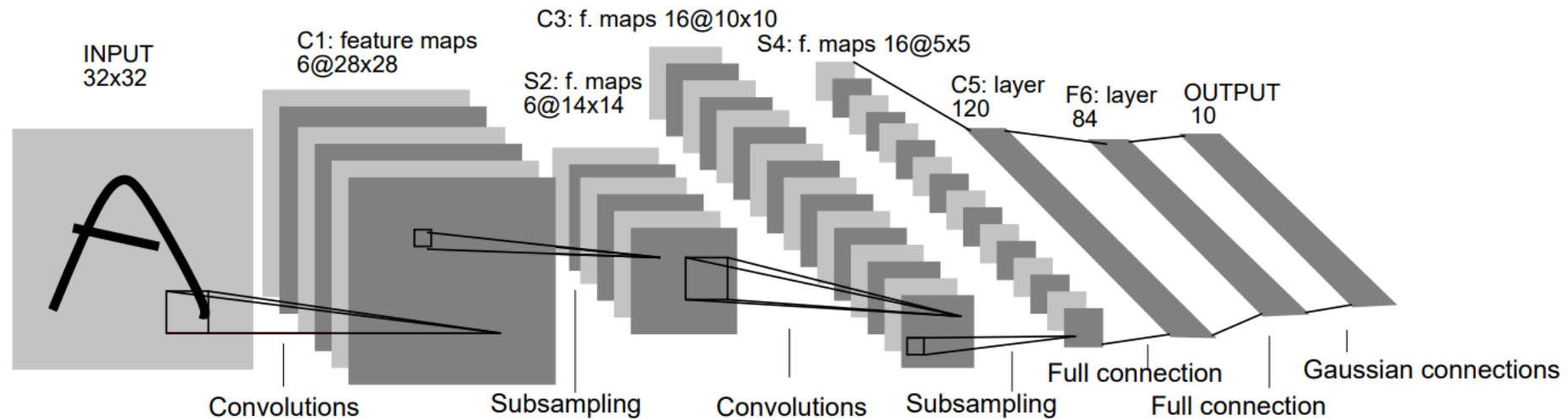# Deep Learning (CS 470, CS 570)
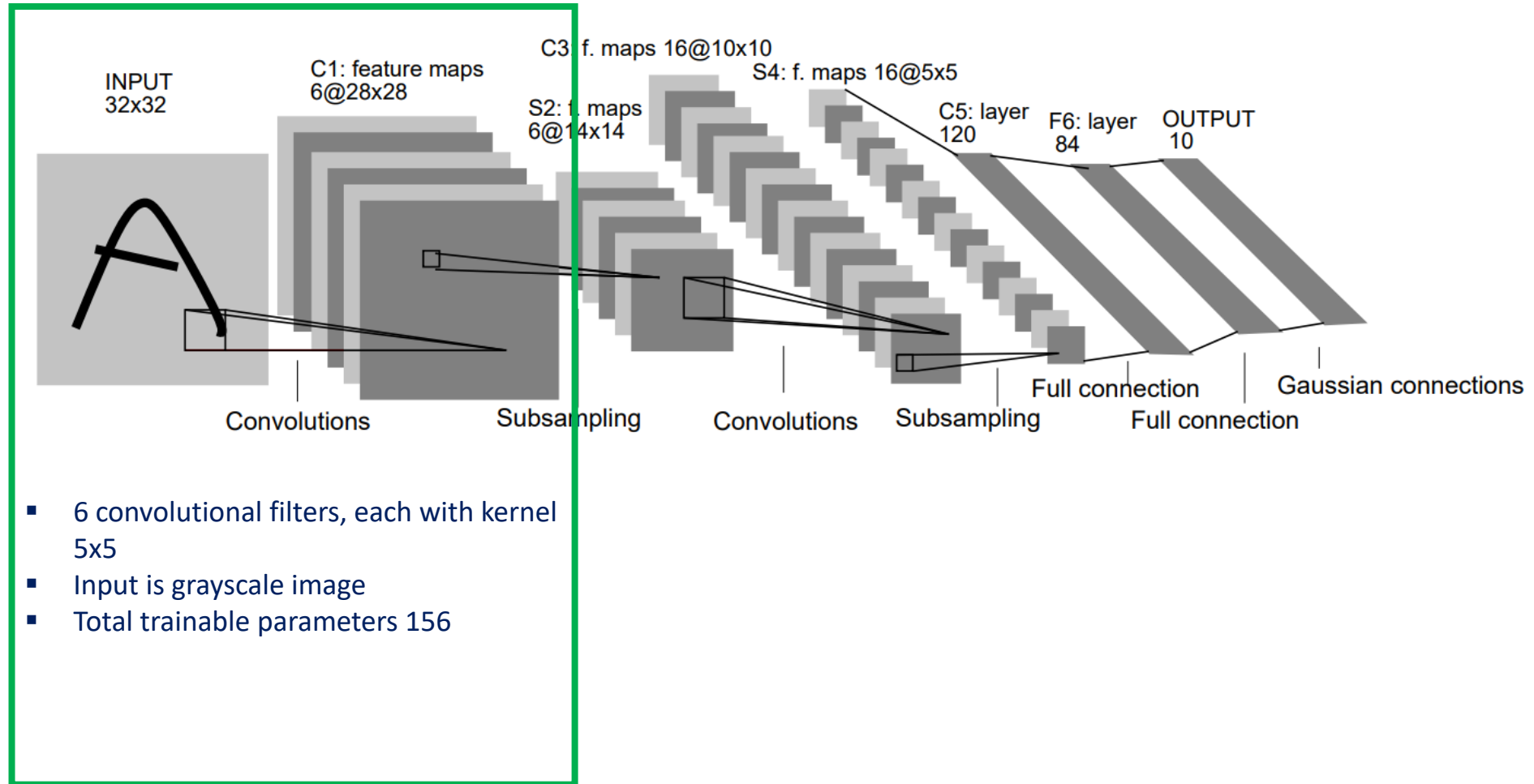
Module 4, Lecture 2: LeNet and AlexNet

LeNet

- LeNet introduced by Yann Lecun in 1998 for Optical Character Recognition (OCR)
- 3 convolutional layers, two pooling layers, and two fully connected layers
- Boundary pixels are ignored during convolution operations therefore no padding is used
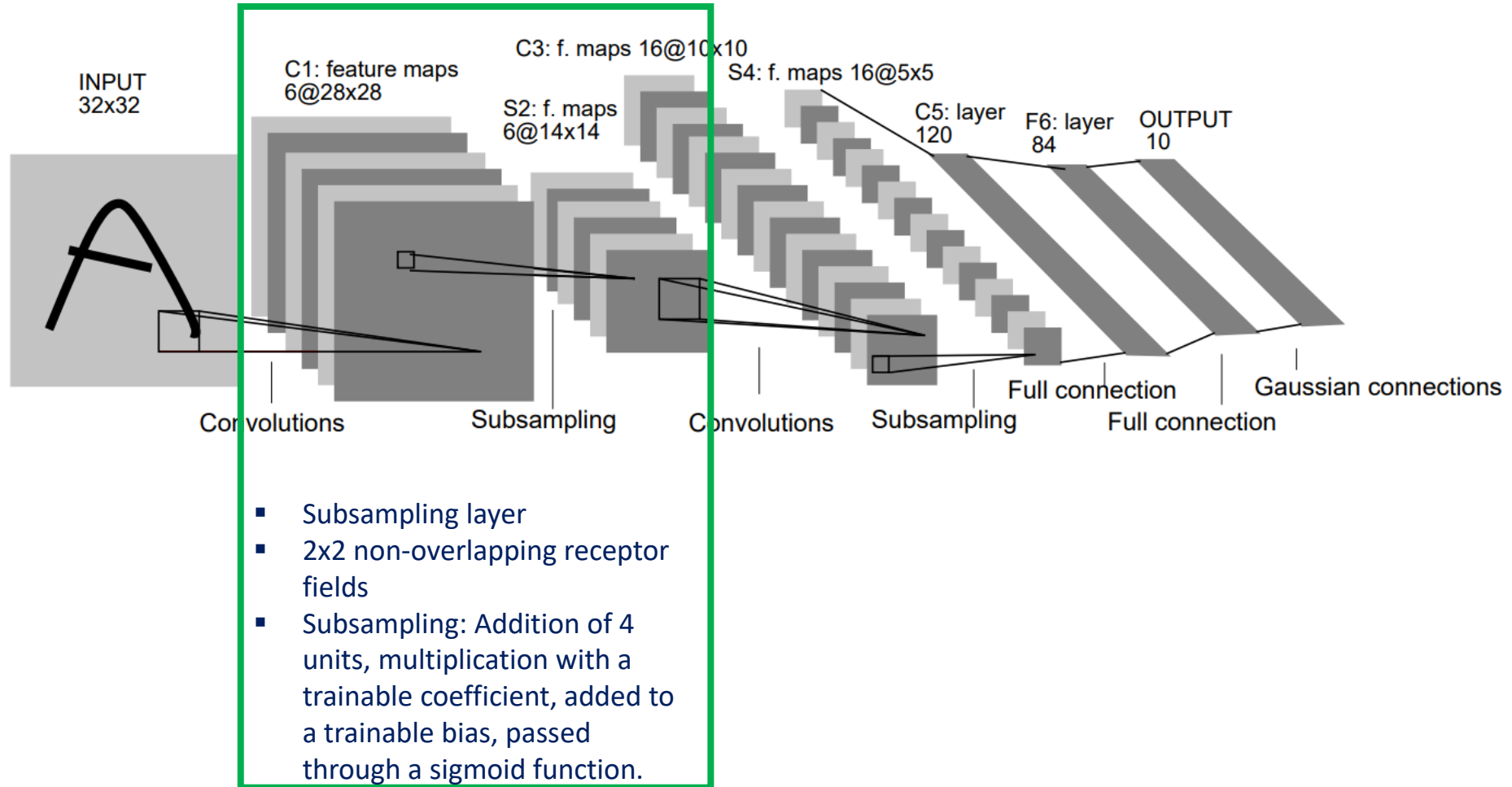
Database

Ref: "Gradient-Based Learning Applied to Document Recognition", Y. Lecun, L. Bottou, Y. Bangio, P Haffner.

LeNet

INPUT
32x32

C1: feature maps
6@28x28

S2: f. maps
6@14x14

C3: f. maps 16@10x10

S4: f. maps 16@5x5

C5: layer
120

F6: layer
84

OUTPUT
10

Convolutions   Subsampling   Convolutions   Subsampling   Full connection   Gaussian connections

Full connection

- 6 convolutional filters, each with kernel 5x5
- Input is grayscale image
- Total trainable parameters 156

Ref: "Gradient-Based Learning Applied to Document Recognition", Y. Lecun, L. Bottou, Y. Bangio, P Haffner.

LeNet

- Subsampling layer
- 2x2 non-overlapping receptor fields
- Subsampling: Addition of 4 units, multiplication with a trainable coefficient, added to a trainable bias, passed through a sigmoid function.

Ref: "Gradient-Based Learning Applied to Document Recognition", Y. Lecun, L. Bottou, Y. Bangio, P Haffner.

LeNet

INPUT 32x32

C1: feature maps 6@28x28

S2: f. maps 6@14x14

C3: f. maps 16@10x10

S4: f. maps 16@5x5

C5: layer 120

F6: layer 84

OUTPUT 10

Convolutions   Subsampling   Convolutions   Subsampling   Full connection   Gaussian connections   Full connection

Total trainable parameters: 1516

Connection strategy

Input feature maps

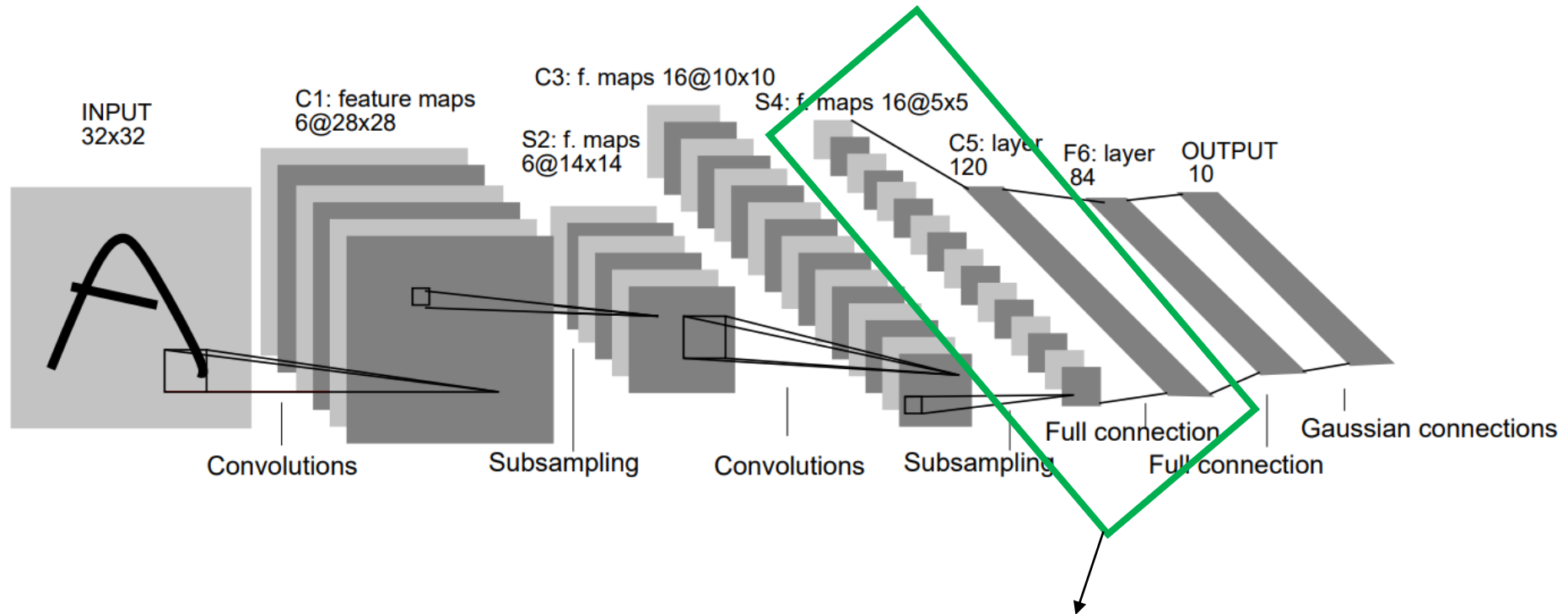|	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

Current layer convolutional filters

- 16 convolutional filters of size 5x5
- Each of the 16 filters are connected to a subset of input feature maps.
- The reason for subset selection
  - Less parameters to train
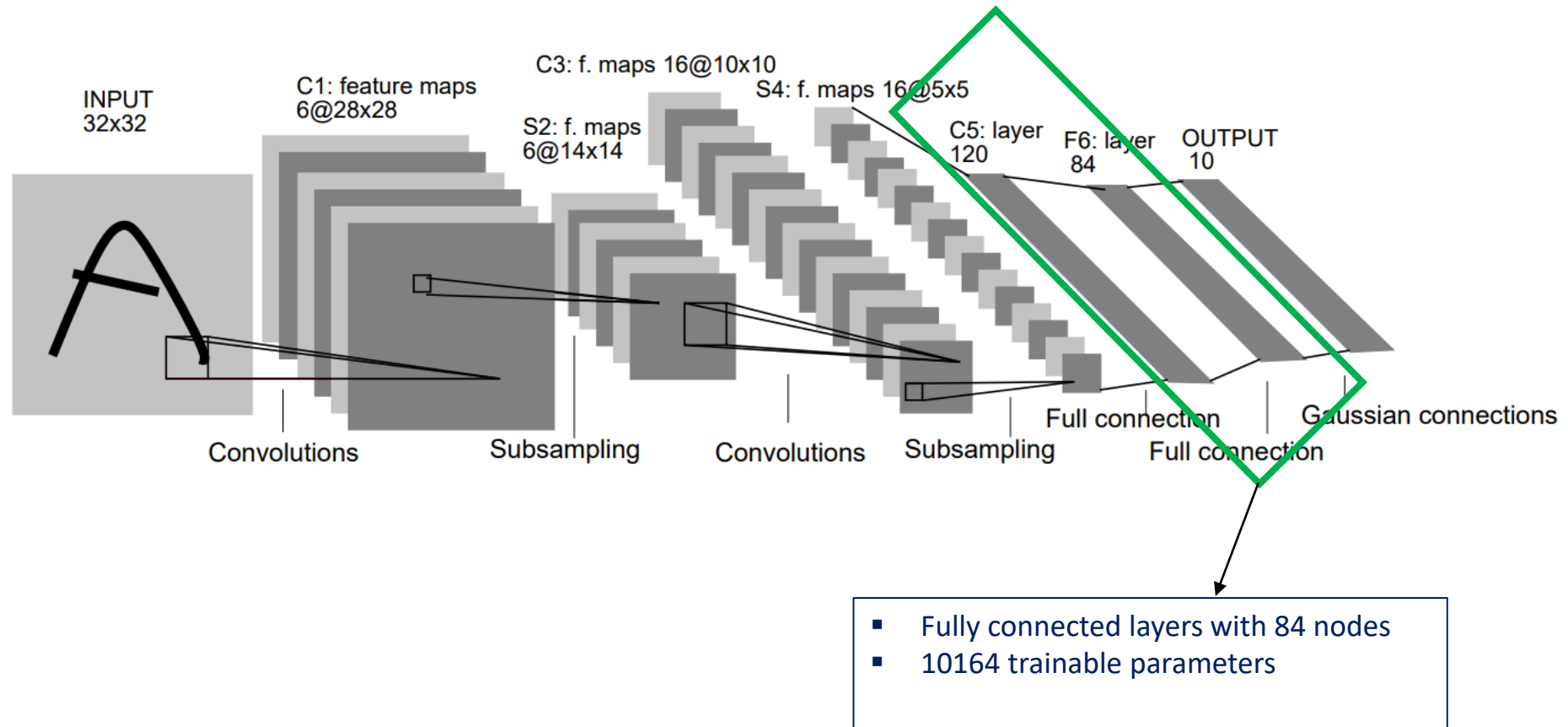 - Different filters are forced to extract different and (hopefully) complementary features

LeNet
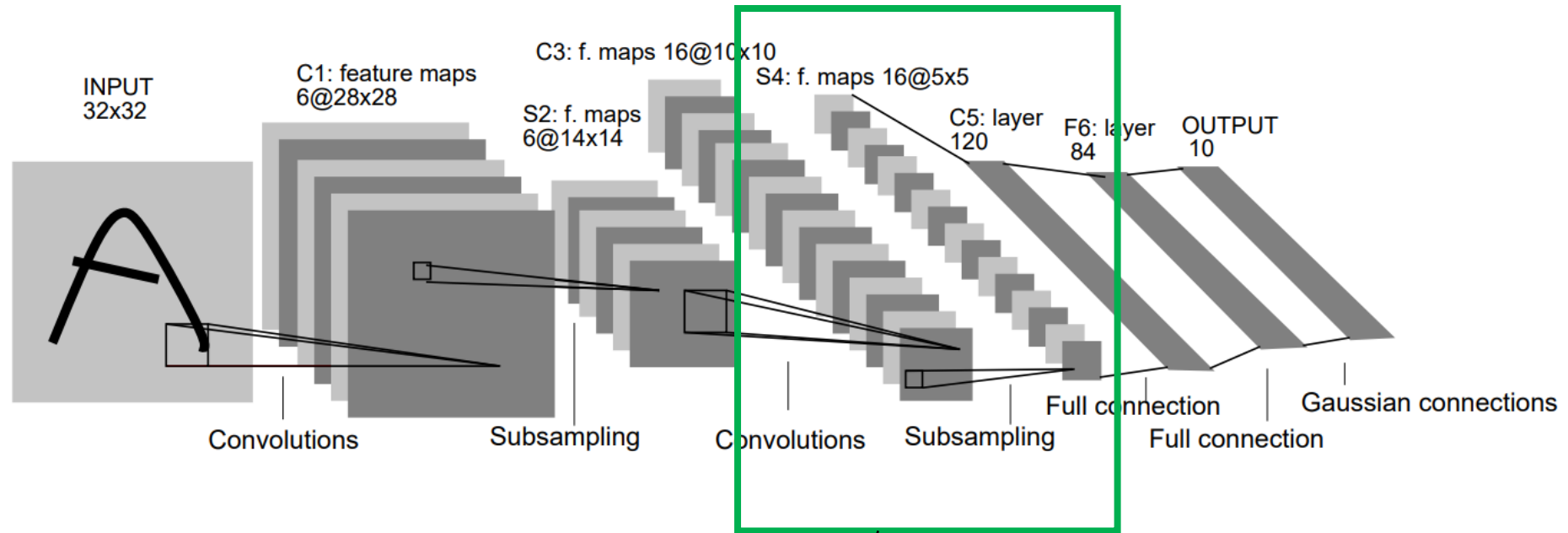
- Subsampling layer
- 2x2 non-overlapping receptor fields
- Subsampling: Addition of 4 units, multiplication with a trainable coefficient, added to a trainable bias, passed through a sigmoid function.

Ref: "Gradient-Based Learning Applied to Document Recognition", Y. Lecun, L. Bottou, Y. Bangio, P Haffner.
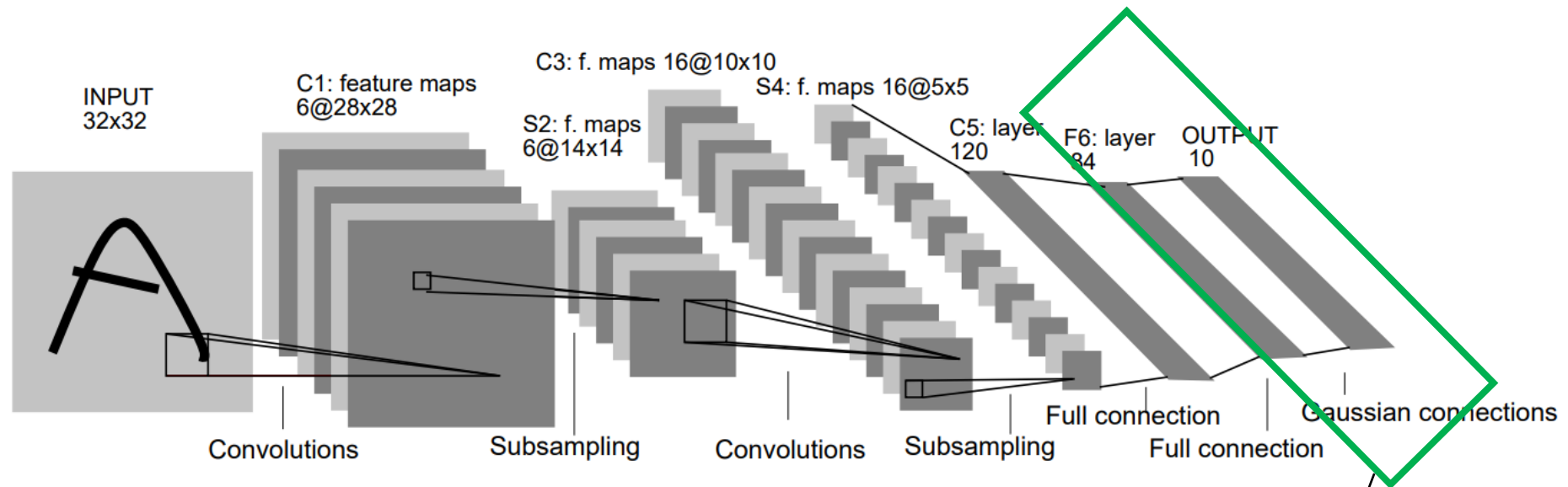
LeNet

- Fully connected convolutional layer
- Each convolutional filter is connected to all 400 features of previous layer
- 120 convolutional filters of size: 5x5x16
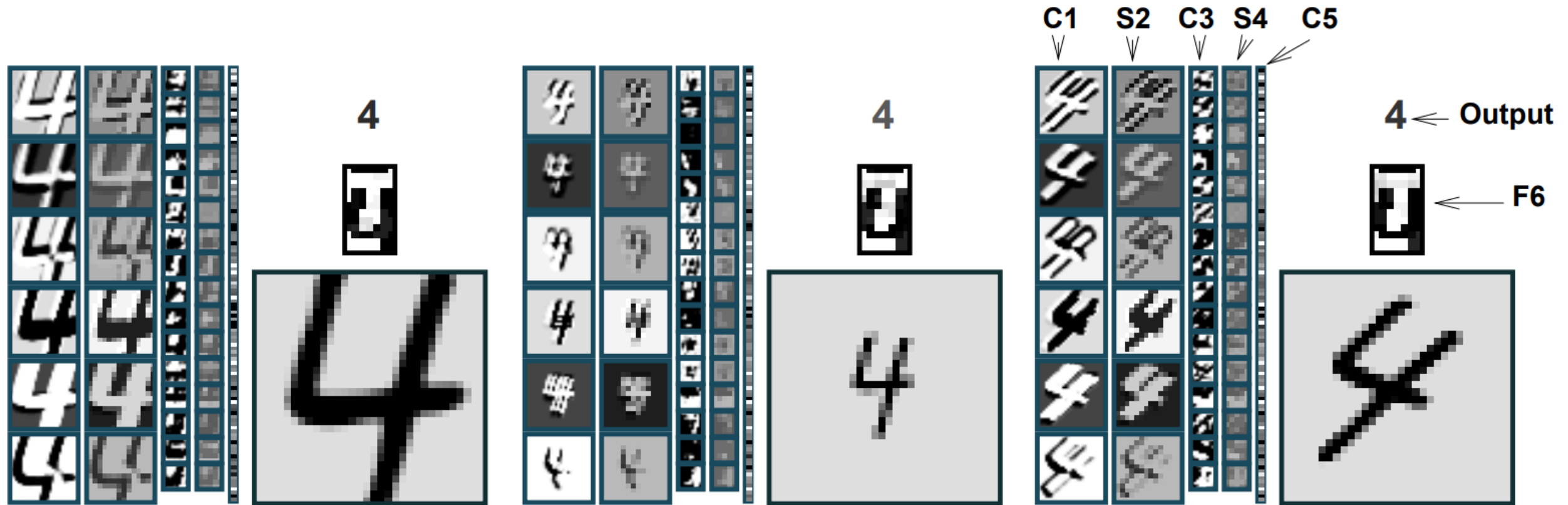- produce 120 feature maps of size 1x1
- Total trainable parameters: 48120

Ref: "Gradient-Based Learning Applied to Document Recognition", Y. Lecun, L. Bottou, Y. Bangio, P Haffner.

LeNet

INPUT
32x32

C1: feature maps
6@28x28

C3: f. maps 16@10x10

S2: f. maps
6@14x14

S4: f. maps 16@5x5

C5: layer
120

F6: layer
84

OUTPUT
10

Convolutions   Subsampling   Convolutions   Subsampling

Full connection   Gaussian connections

Full connection

- Fully connected layers with 84 nodes
- 10164 trainable parameters

Ref: "Gradient-Based Learning Applied to Document Recognition", Y. Lecun, L. Bottou, Y. Bangio, P Haffner.

LeNet

- Output layer with 10 nodes
- Connected to each node of previous layer

Ref: "Gradient-Based Learning Applied to Document Recognition", Y. Lecun, L. Bottou, Y. Bangio, P Haffner.

LeNet: Output

This example shows that LaNet can generate similar responses at F6 for same class (here digit '4') even when there are scale and rotational variations in the input images.

AlexNet

- Even after LeNet, deep learning was not popular because:
  - Lack of computational power didn't allow the network to be deep
  - Shallow architecture was not appropriate for high resolution image classification
- After winning the ImageNet image classification competition, AlexNet become a game changer.
- It is named after Alex Krizhevsky
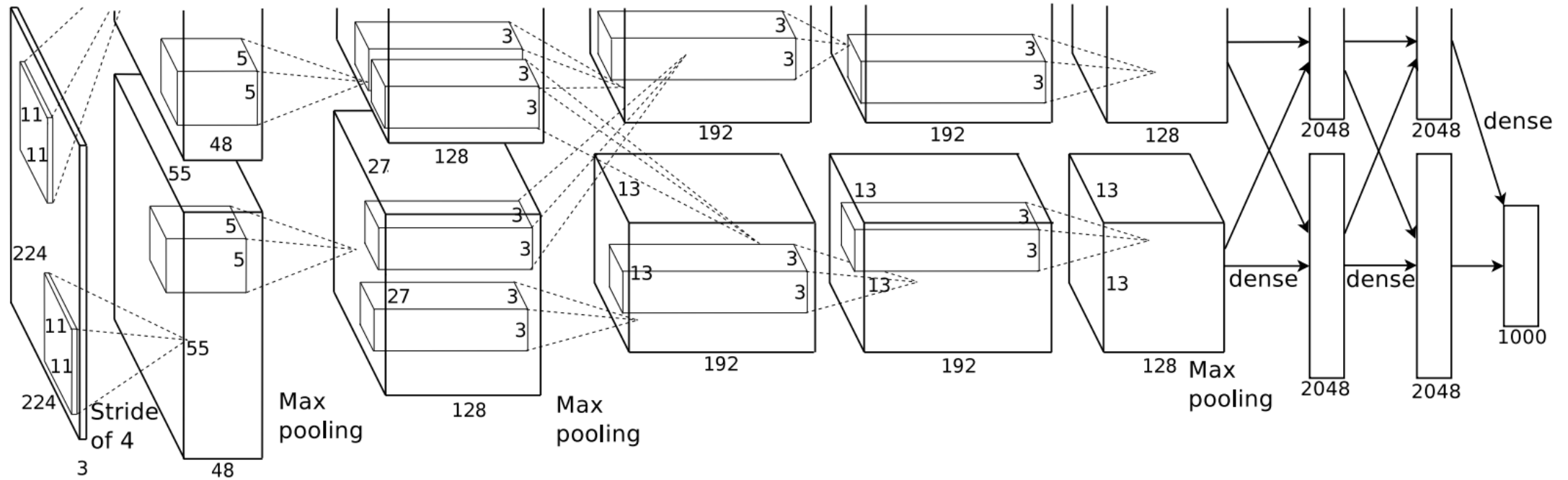- First CNN which used GPU for fast computation

ImageNet dataset
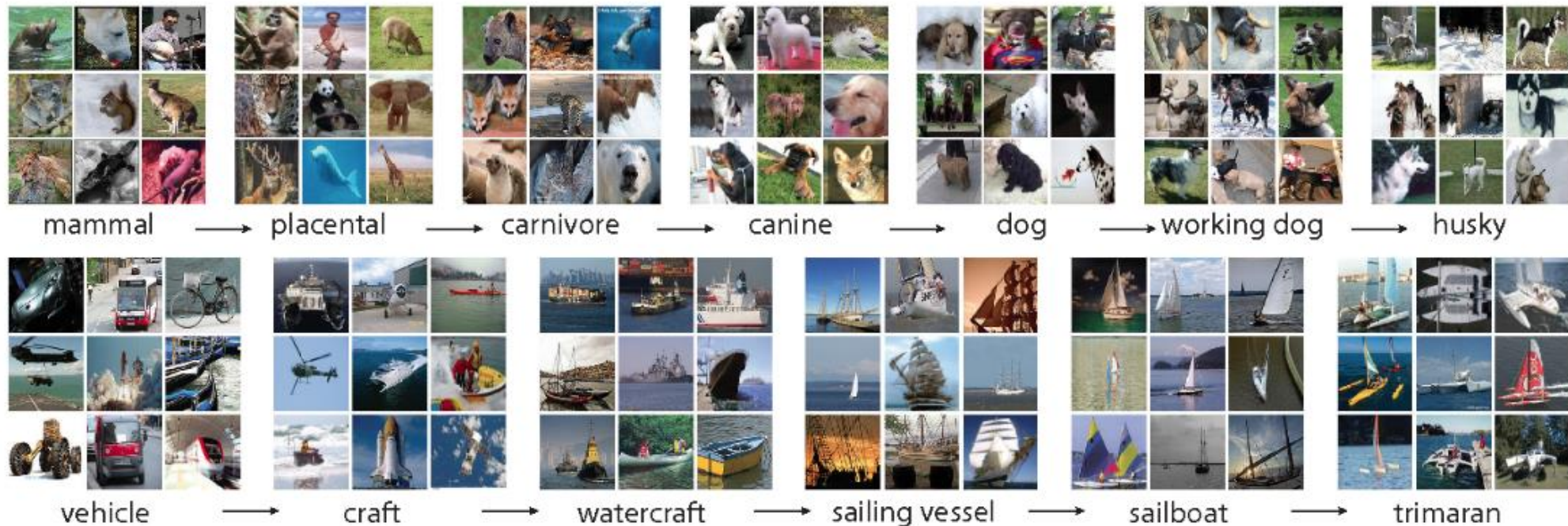
Fig: "ImageNet: A large-scale hierarchical image database", Jia Deng et al.
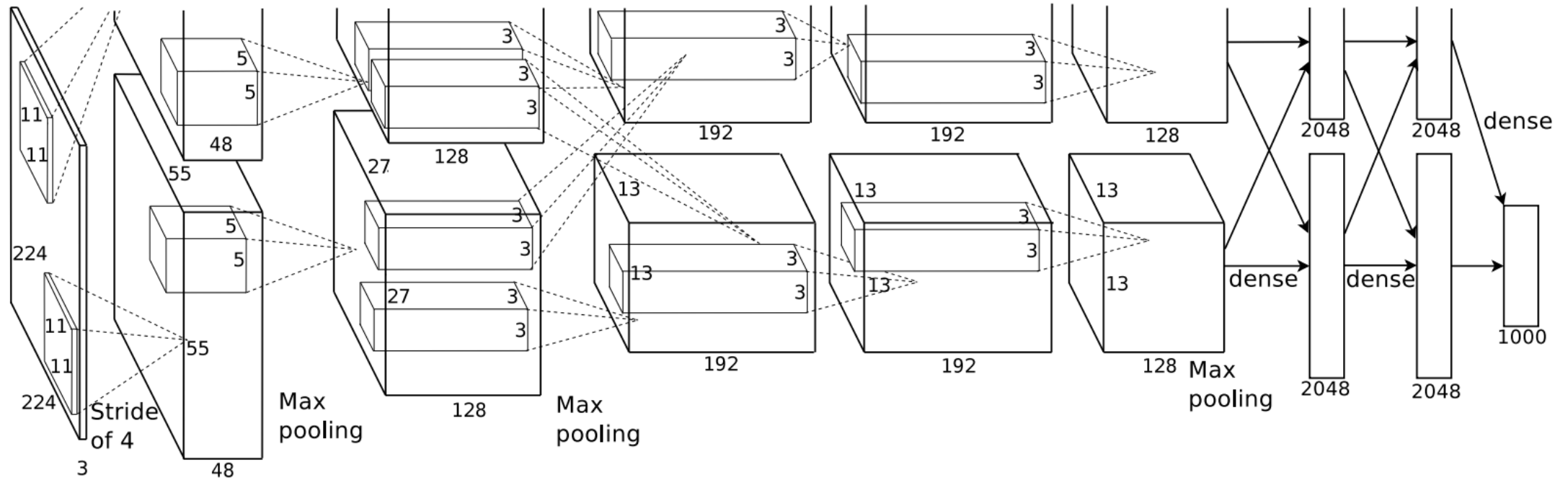
AlexNet

Database: ImageNet

- 15 million high resolution images
- 22000 different classes
- Manually labeled

http://www.image-net.org/

Database used: subset of ImageNet

- 1000 different classes
- Around 1000 images per class
- 1.2 million high resolution training images
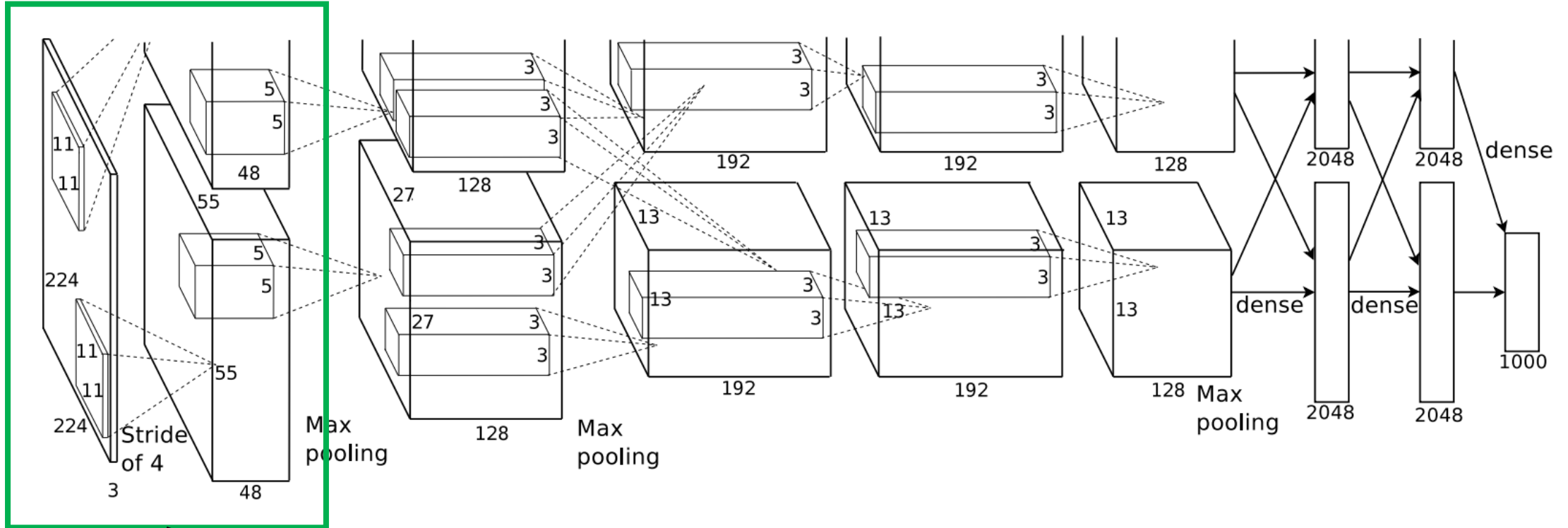- 50,000 validation images
- 150,000 testing images

"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton
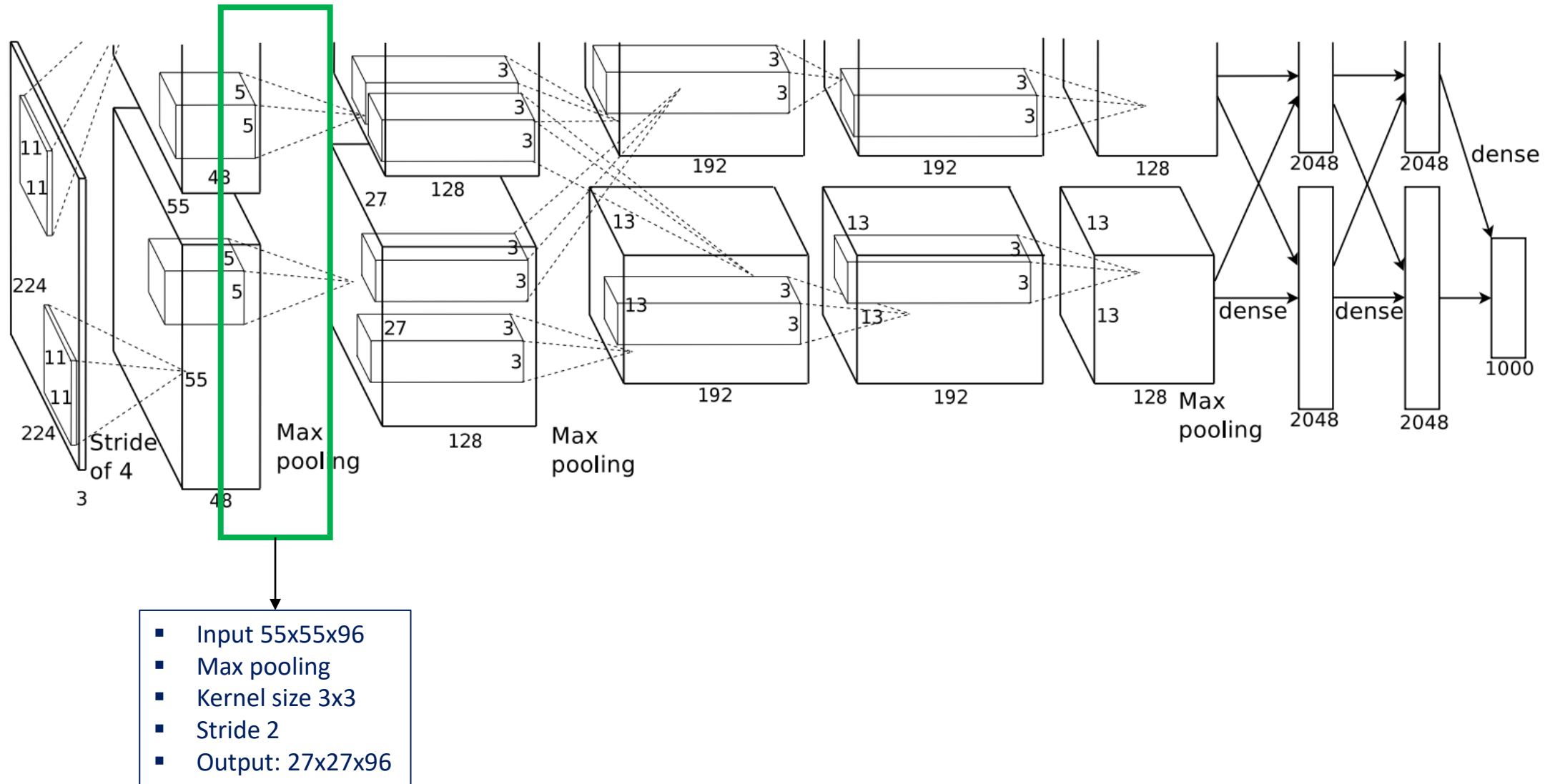
AlexNet

Network Architecture:
- 8 trainable layers
- 5 convolutional layers, 3 fully connected layers
- Each convolutional layer consist of convolutional filters and a non-linear activation function ReLU
- 3 maxpooling layers for subsampling
- Overall 60 million parameters

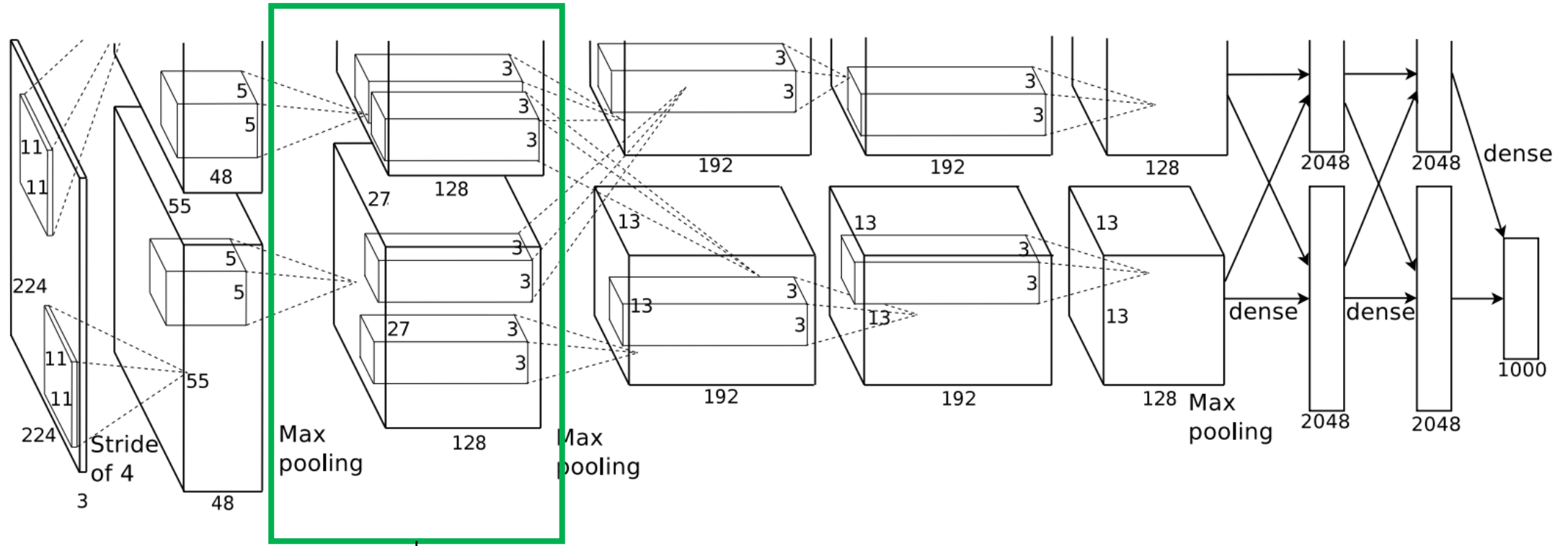"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton

AlexNet

- Input images 224x224x3 color images
- 96 convolutional filters of size 11x11x3
- Stride 4, Padding 2
- Output 55x55x96

"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton

AlexNet

- Input 55x55x96
- Max pooling
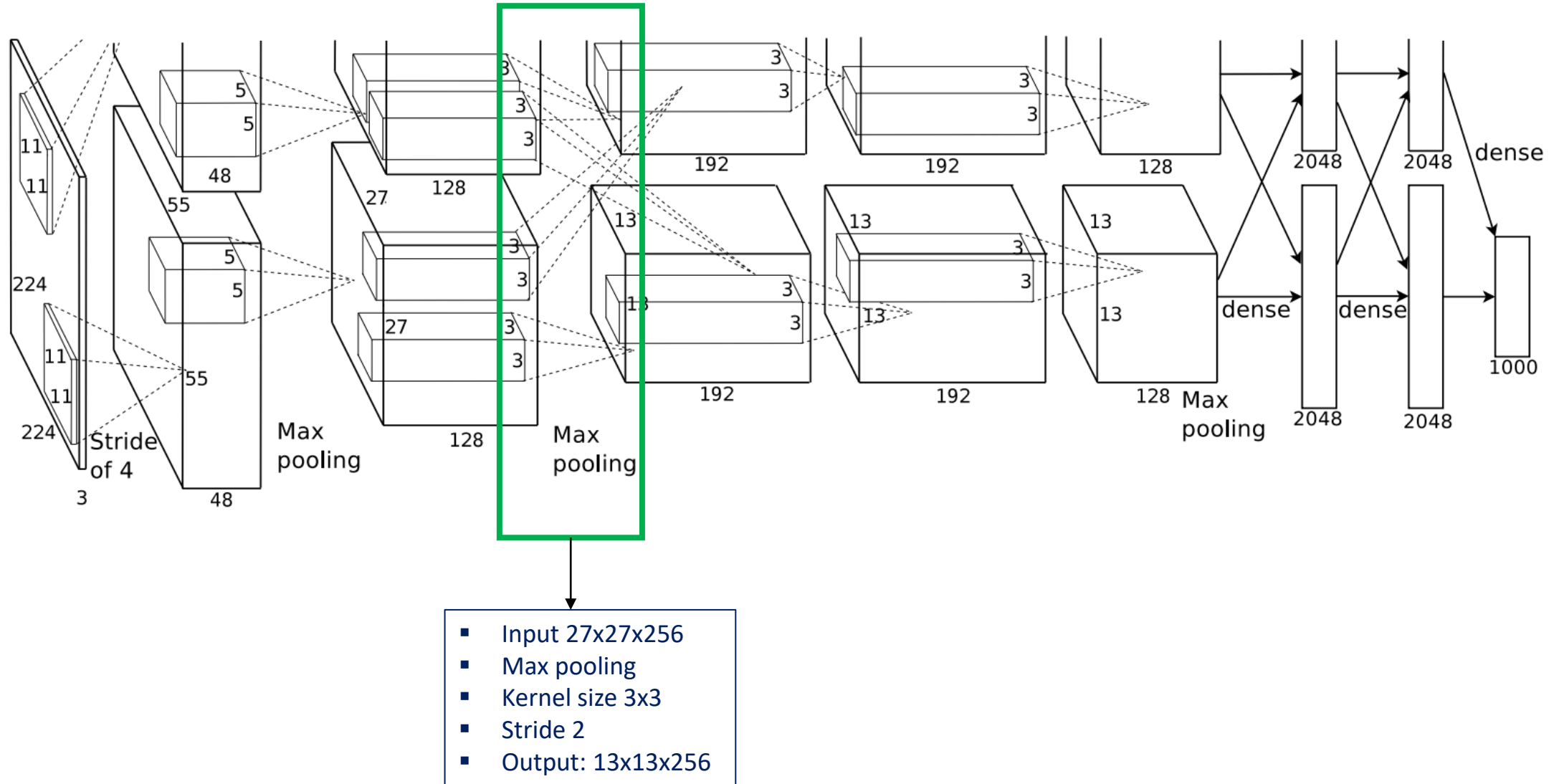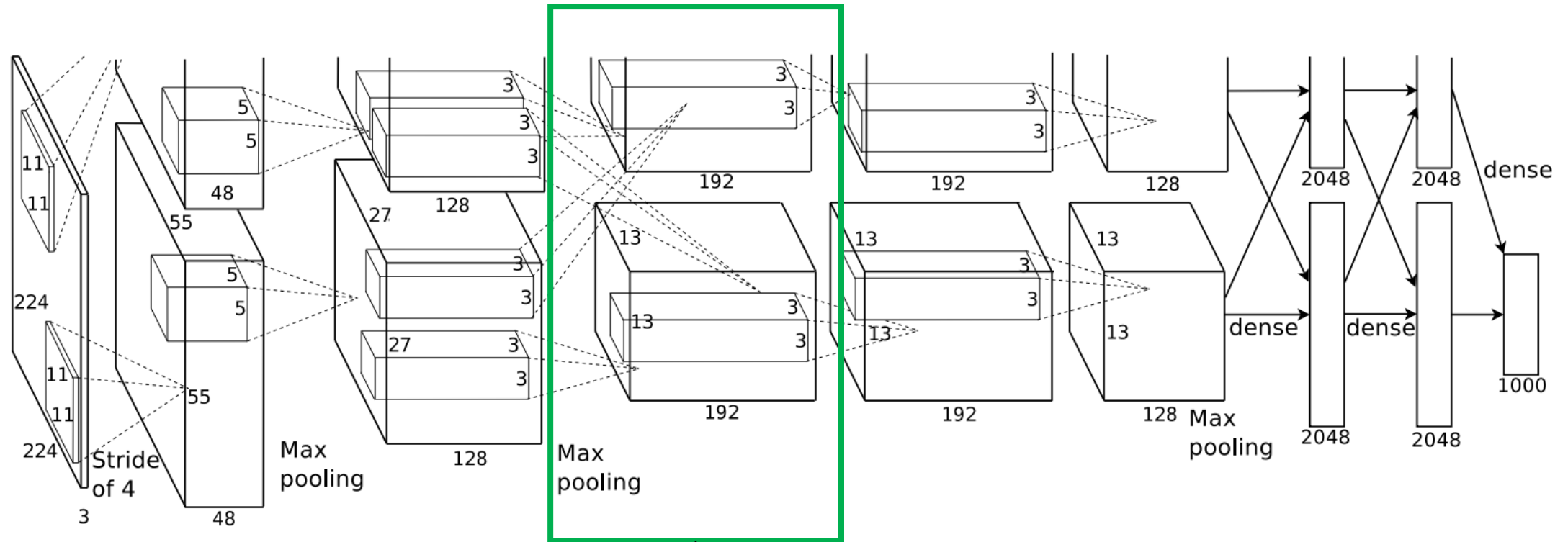- Kernel size 3x3
- Stride 2
- Output: 27x27x96

"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton

AlexNet

- Input 27x27x96
- 256 filters of size 5x5x48
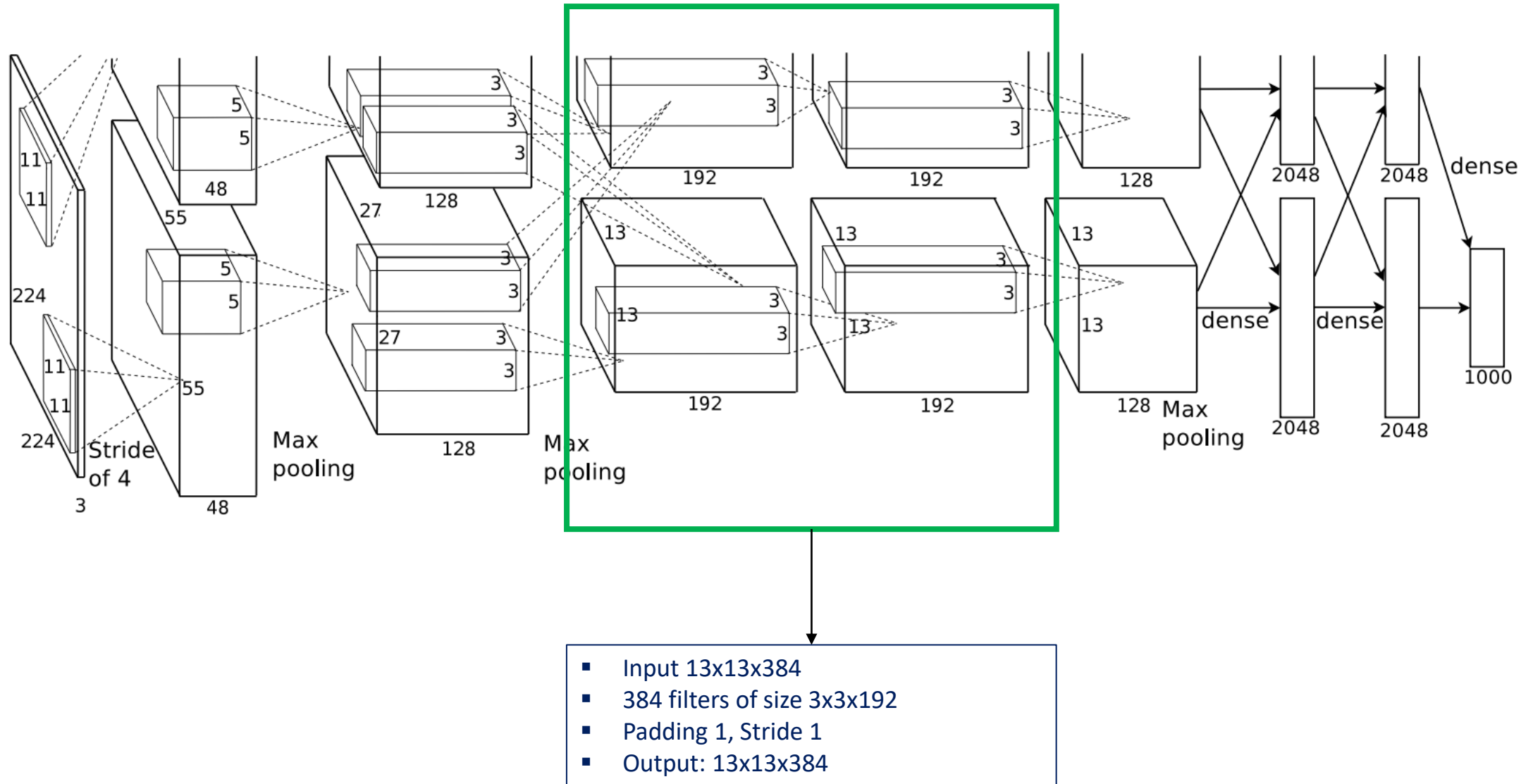- Padding 2, Stride 1
- Output: 27x27x256

"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton

AlexNet

- Input 27x27x256
- Max pooling
- Kernel size 3x3
- Stride 2
- Output: 13x13x256

"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton

AlexNet

- Input 13x13x256
- 384 filters of size 3x3x256
- Padding 1, Stride 1
- Output: 13x13x384

"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton

AlexNet

- Input 13x13x384
- 384 filters of size 3x3x192
- Padding 1, Stride 1
- Output: 13x13x384

"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton
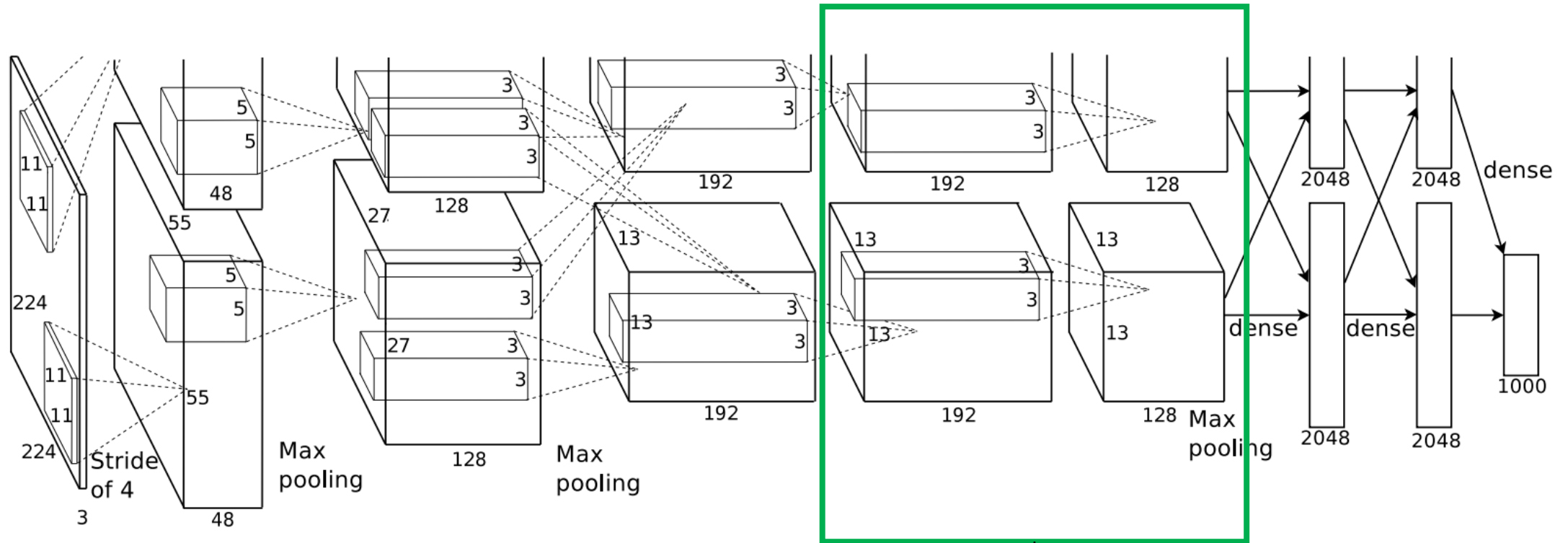
AlexNet

- Input 13x13x384
- 256 filters of size 3x3x192
- Padding 1, Stride 1
- Output: 13x13x256

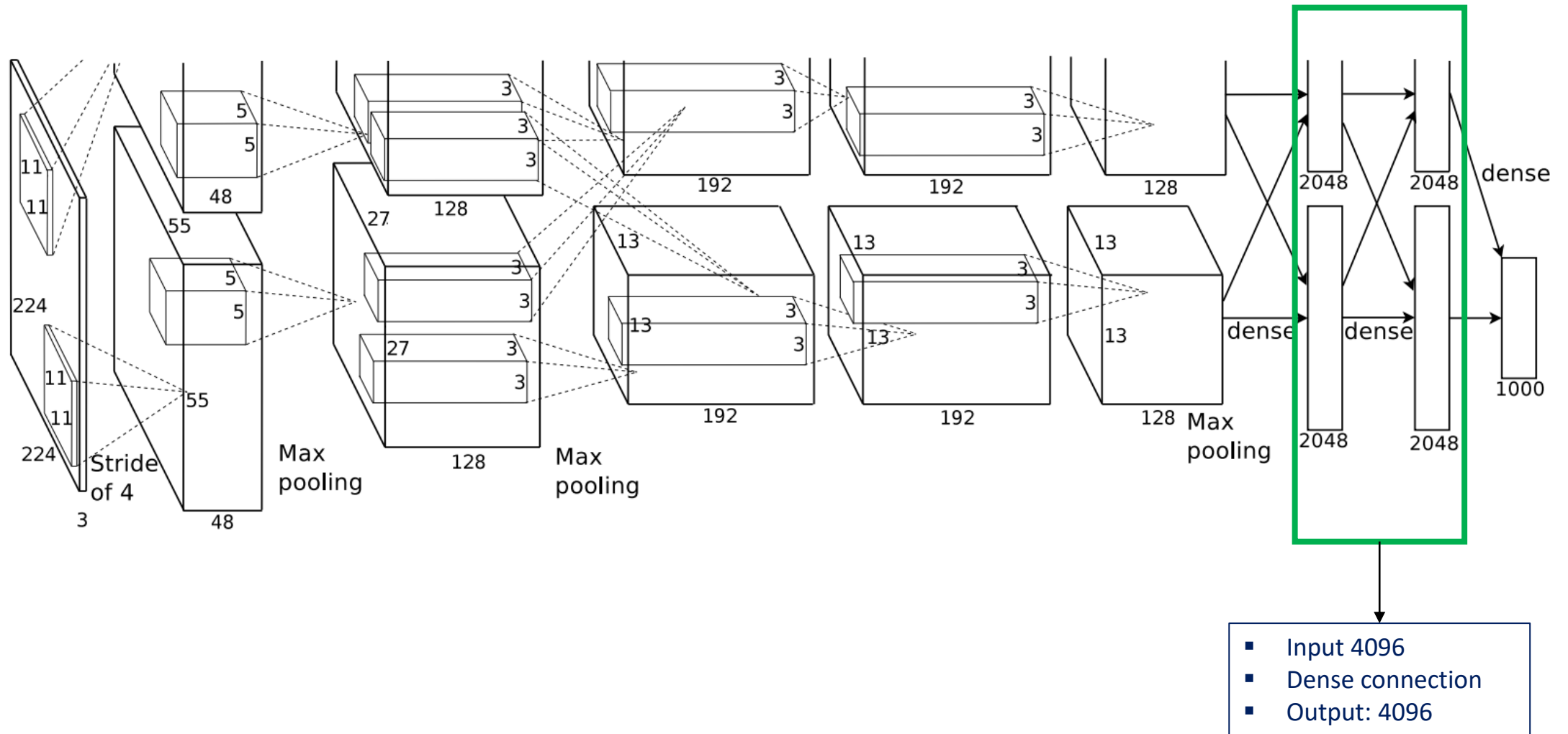"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton
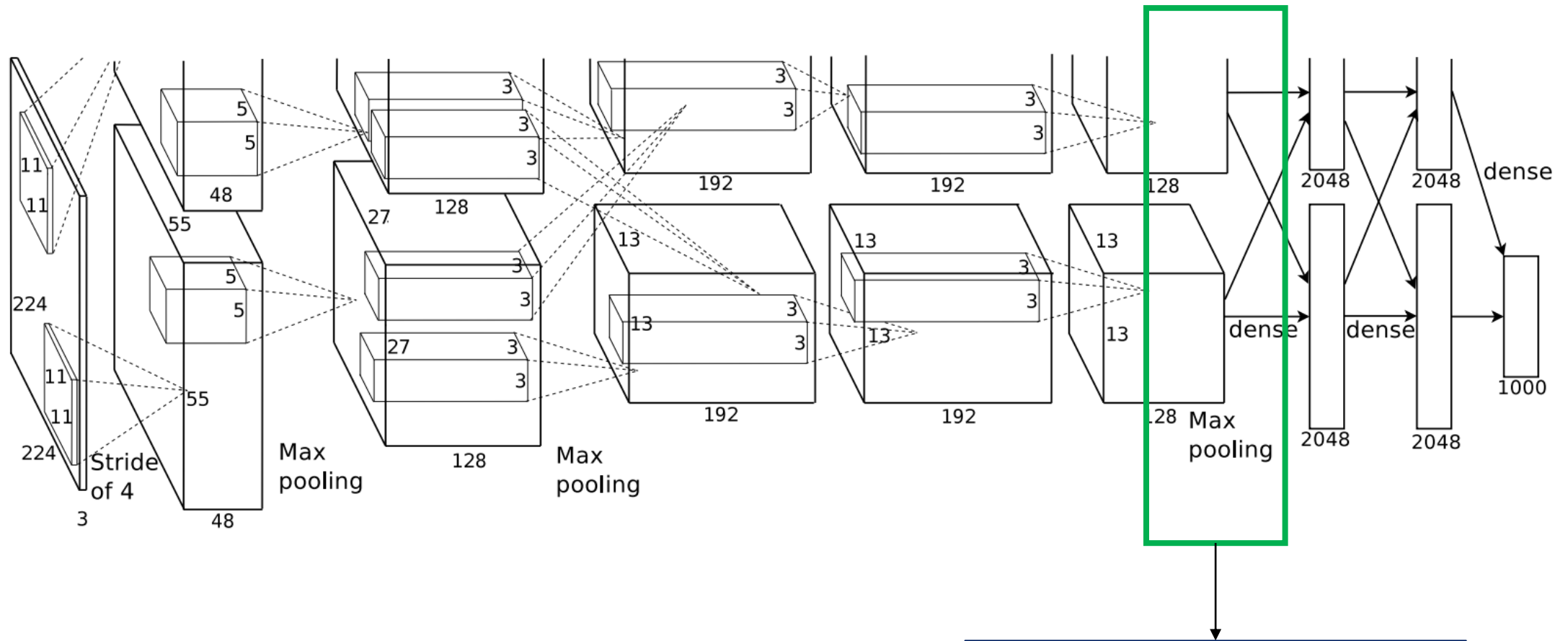
AlexNet

- Input 13x13x256
- Max pooling
- Kernel size 3x3
- Stride 2
- Output: 6x6x256
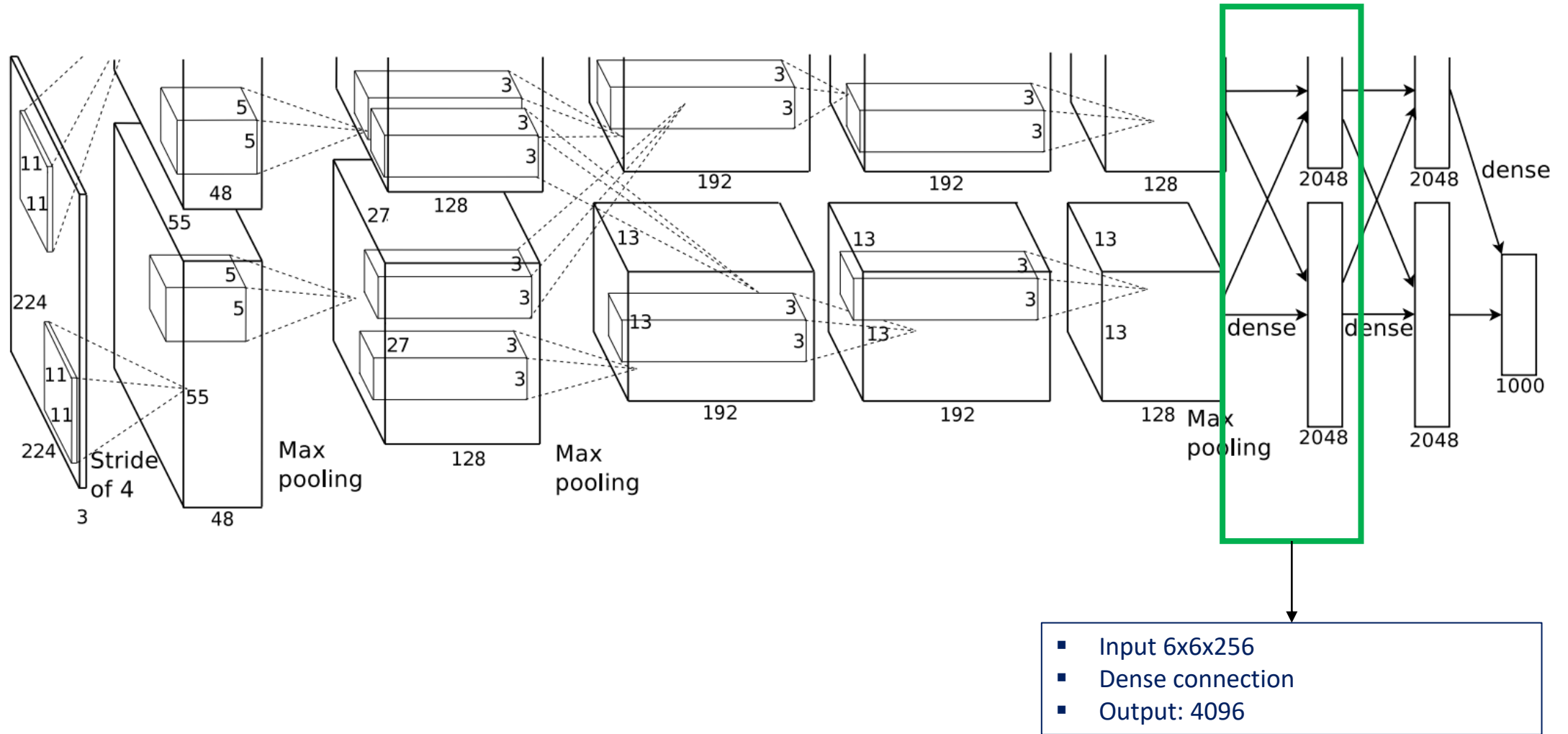
"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton

AlexNet

- Input 6x6x256
- Dense connection
- Output: 4096

"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton

AlexNet

- Input 4096
- Dense connection
- Output: 4096

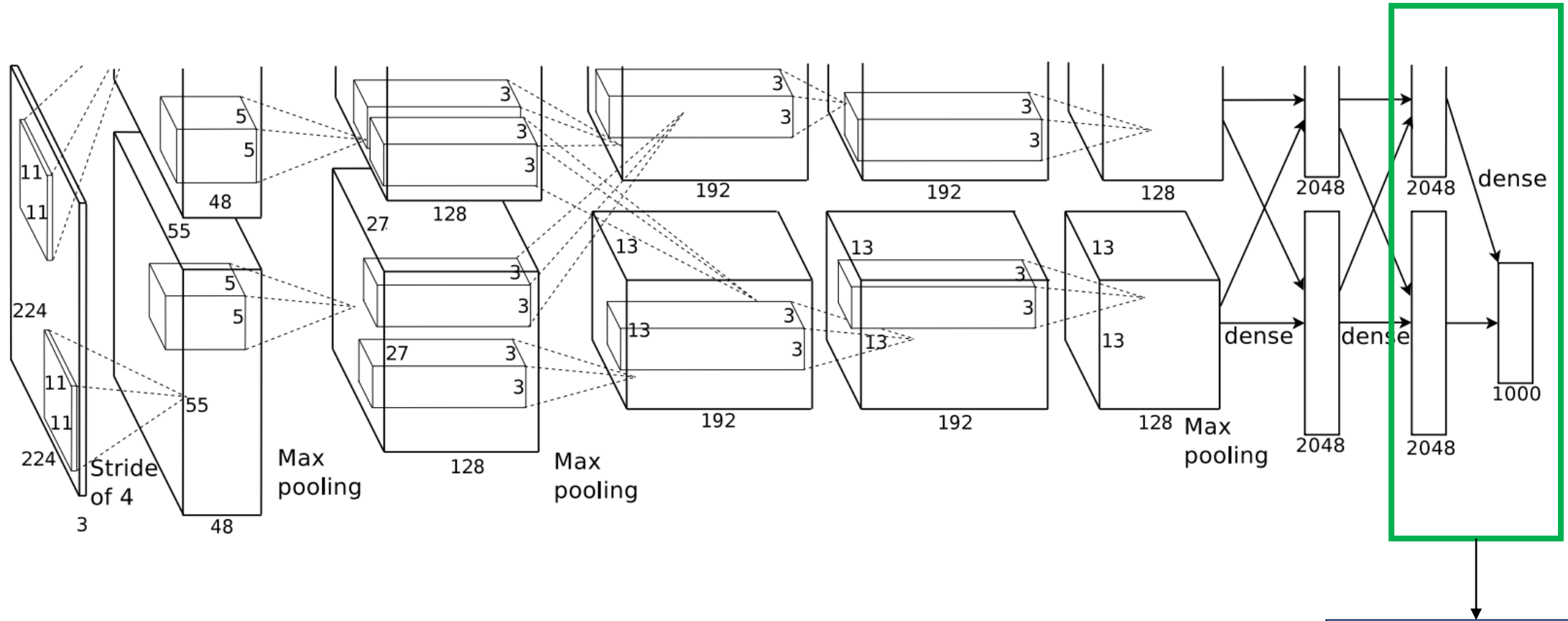"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton
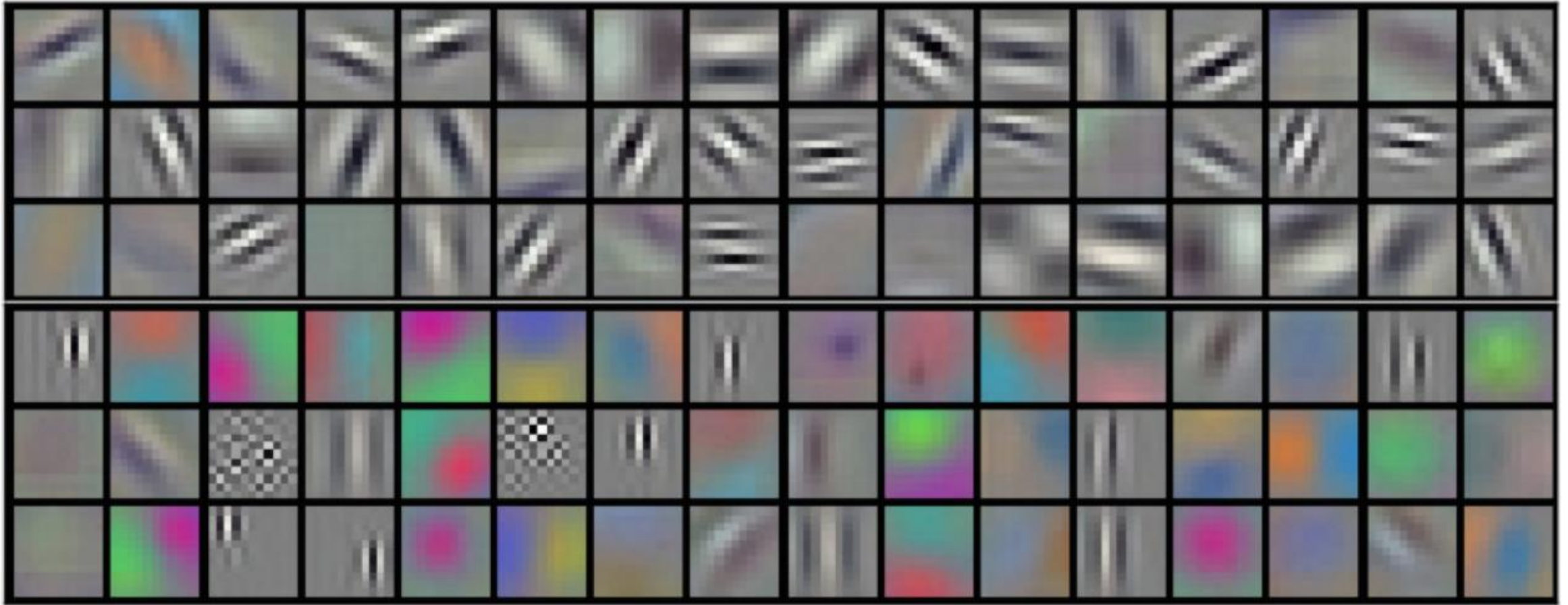
AlexNet

- In total 62.3 million parameters to train
- Conv. Layers: 3.7 million(6%), FC layers: 58.6 million (94%)
- RELU nonlinearity after each convolutional layer

- Input 4096
- Dense connection
- Output: 1000

"ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton

AlexNet

96 learned filters in convolutional layer 1

Fig ref: "ImageNet Classification with Deep Convolutional Neural Networks", A. Krizhevsky, I. Sutskever, G. E. Hinton

Additional Readings

LeNet
AlexNet