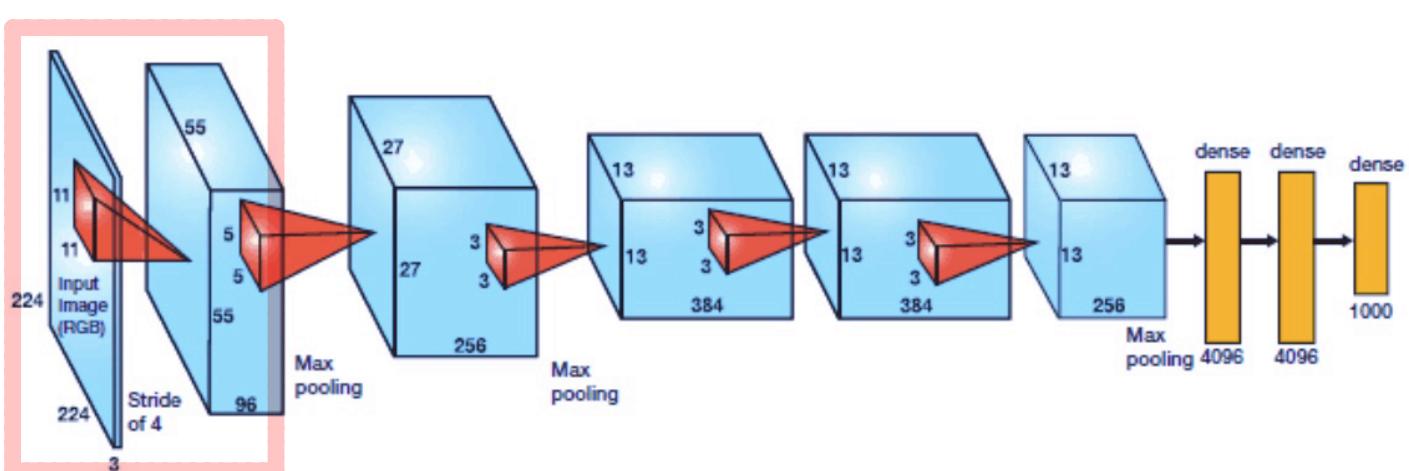


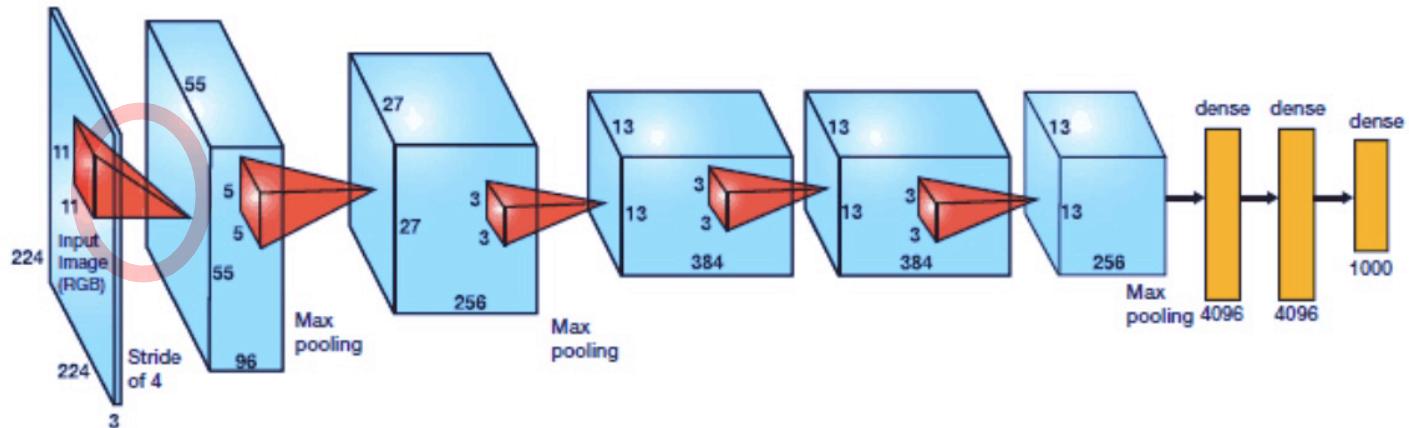
Alex Net



Convolution layer

- Input image size $224 \times 224 \times 3$
- Number of filters 96
- Filter size $11 \times 11 \times 3$
- Stride 4
- Output size $\text{remember } \{(N-F)/\text{stride}\} + 1$

$$\Rightarrow \{(224-11)/4\}+1 = 54.25 + 1 = 55.25 ???$$



1. Convolution layer_1

- Input image size $224 \times 224 \times 3$

$$\rightarrow 227 \times 227 \times 3$$

$$\{(227-11)/4\}+1 = 55.$$

- Number of filters 96

- Filter size $11 \times 11 \times 3$

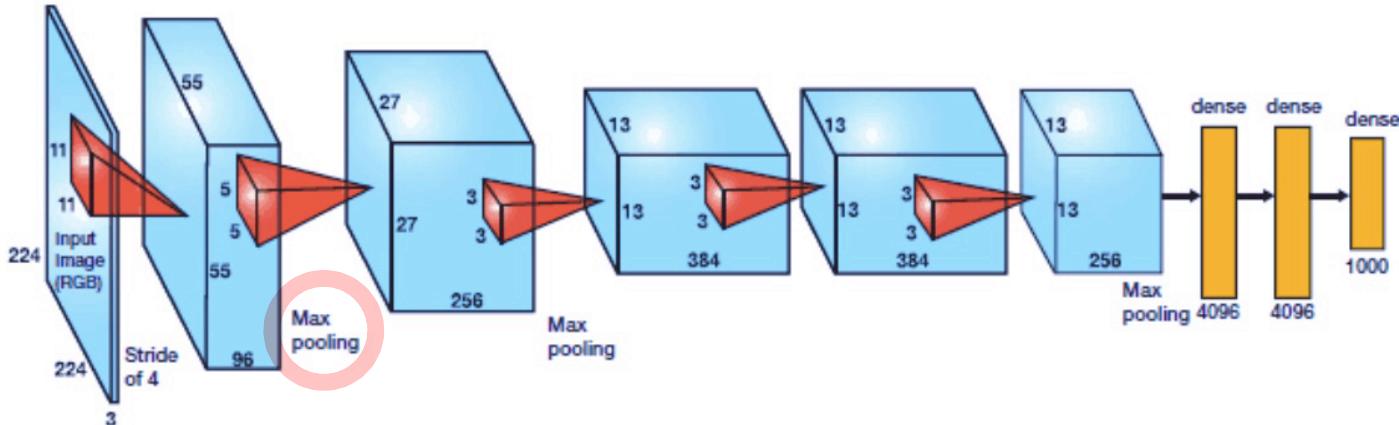
- Stride 4

- Output size remember $\{(N-F)/\text{stride}\}+1$



$$\frac{55 \times 55 \times 96}{96 \text{ per feature map.}}$$

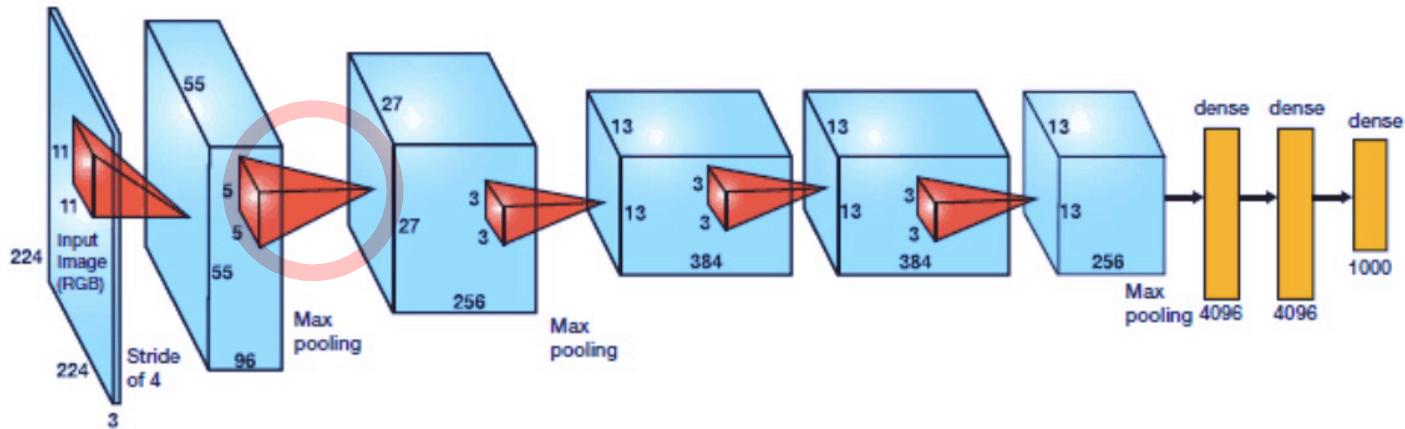
$$\Rightarrow \{(224-11)/4\}+1 = 54.25 + 1 = 55.25 ???$$



2. Max pooling - I

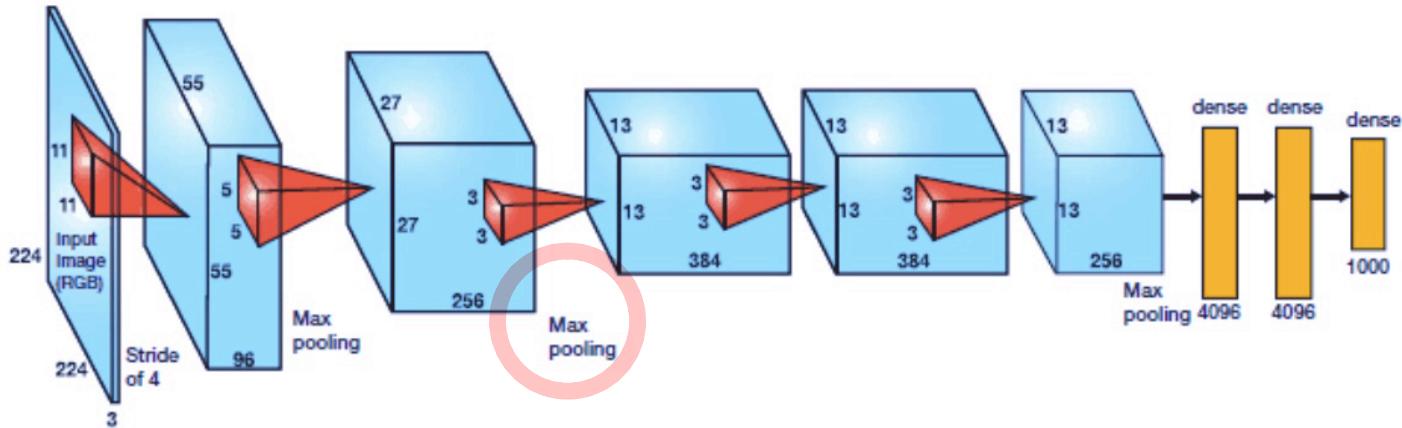
- Input image size $55 \times 55 \times 96$

- Pooling (Overlapping) 3×3
- Stride 2
- Output size $\lceil \frac{55-3}{2} \rceil + 1 = 27$, $27 \times 27 \times 96$



3. Convolution layer-2

- Input image size $27 \times 27 \times 96$ $27 - 5 + 4$
- Number of filters 256 $26 \div 1$
- Filter size $5 \times 5 \times 96$ 27
- Stride 1
- Padding 2
- Output size $(27 - 5 + 1) \div 1 = 27$, $27 \times 27 \times 256$



4. Max_Pooling_2

- Input image size : 224x224x3

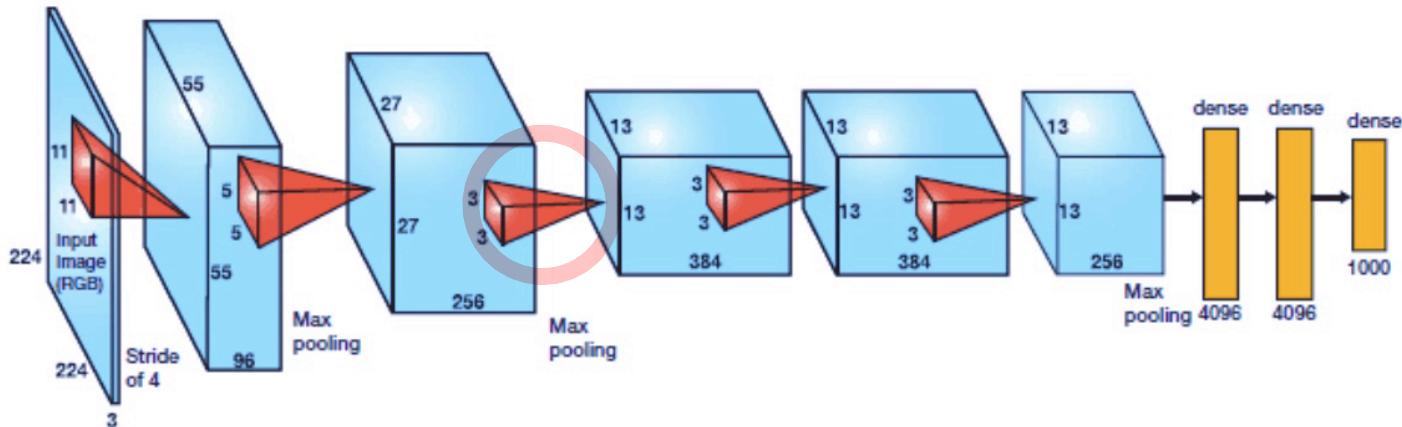
- Pooling (overlapping) : 3×3

- Stride : 2

- Output size : $(\frac{224 - 3 + 2}{2} + 1) = 112$

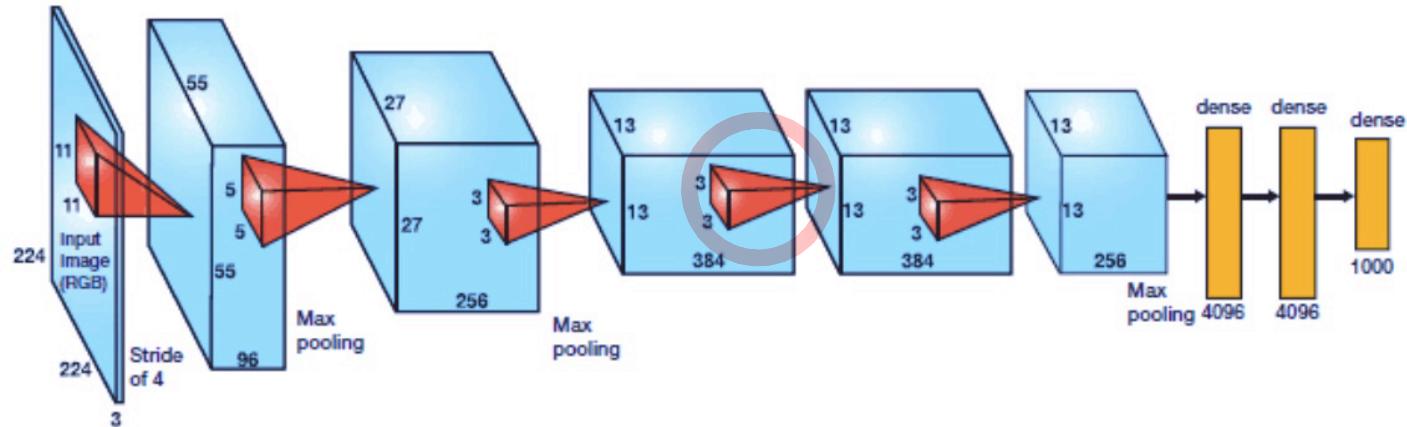
\Rightarrow

112x112x3



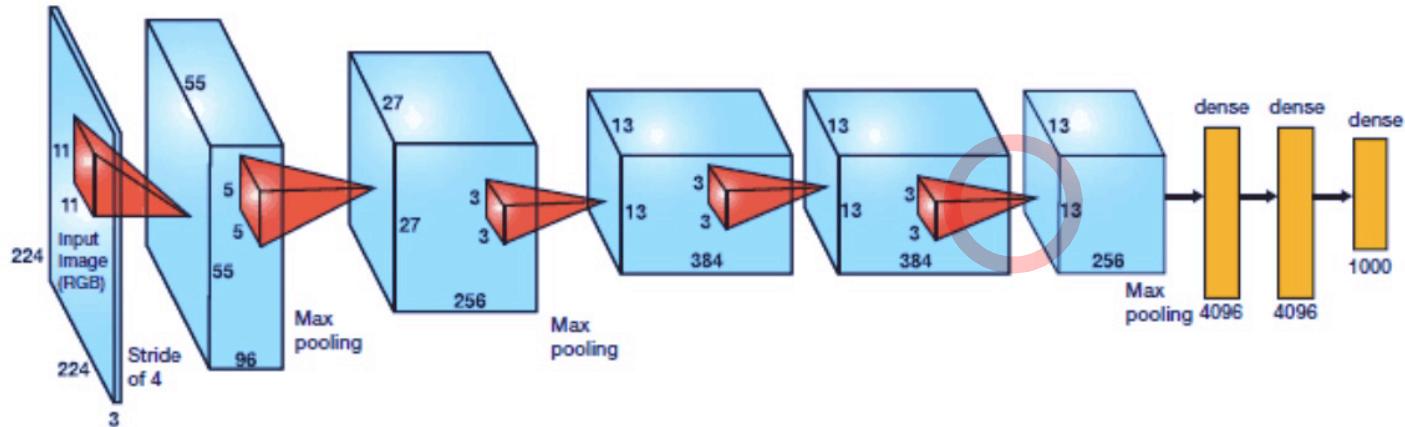
5. Convolution layer -3

- Input Image size [REDACTED]
- Number of filters [REDACTED]
- Filter size 3×3
- Stride 1
- Padding 1
- Output size $\{(3+2\cdot1)-2\}/1 + 1 = 12, (12 \times 12) \times 256$



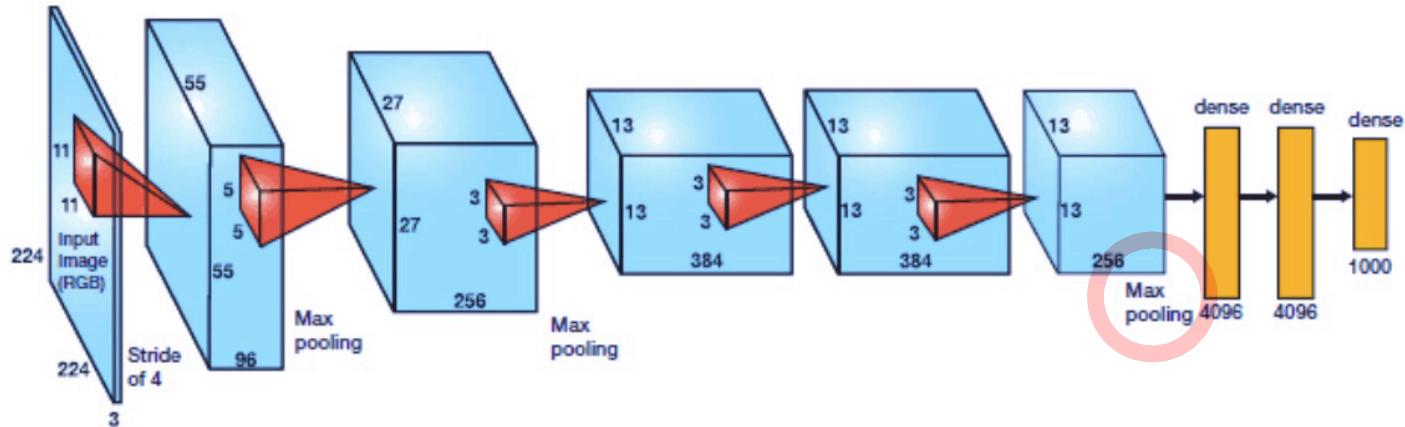
6. Convolution layer - 4

- Input image size 224x224x3
- Number of filters 384
- Filter size 3x3
- Stride 1
- Padding 0
- Output size 13x13x256



7. Convolution layer - 5

- Input image size 224x224x3
- Number of filters 256
- Filter size 3x3
- Stride 1
- Padding 1
- Output size 13x13x256



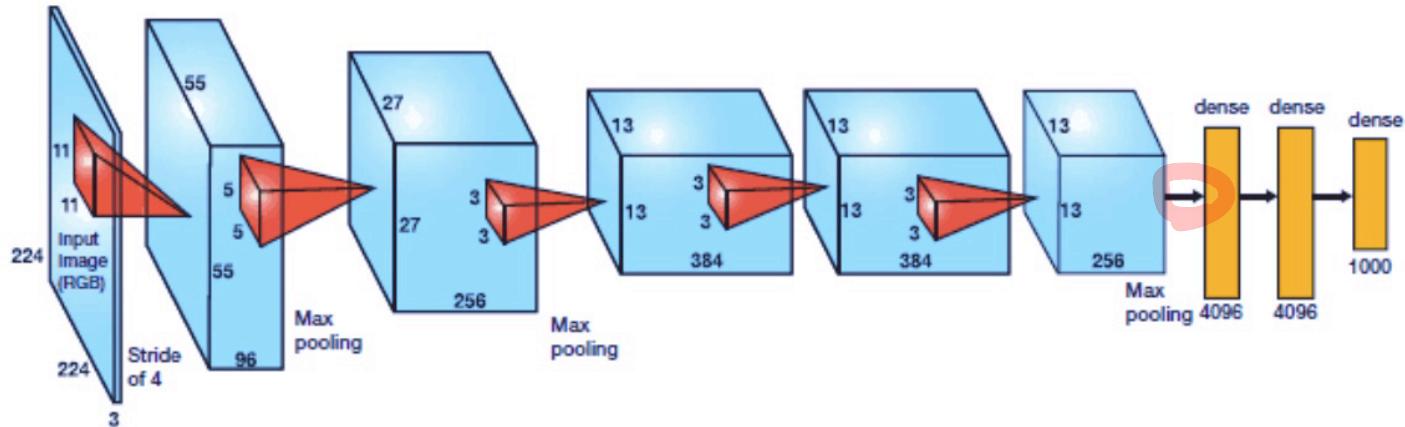
8. Pooling layer_3

Input image size $13 \times 13 \times 256$

• Pooling (overlapping) 3×3

• Stride 2

• Output size $6 \times 6 \times 6$

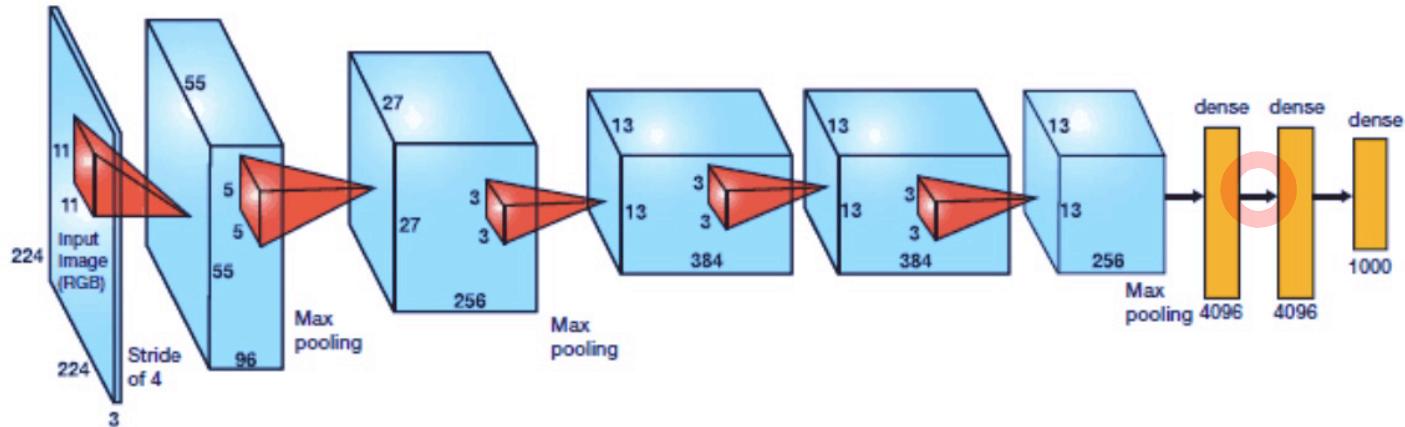


9. FC1

- Input image size:

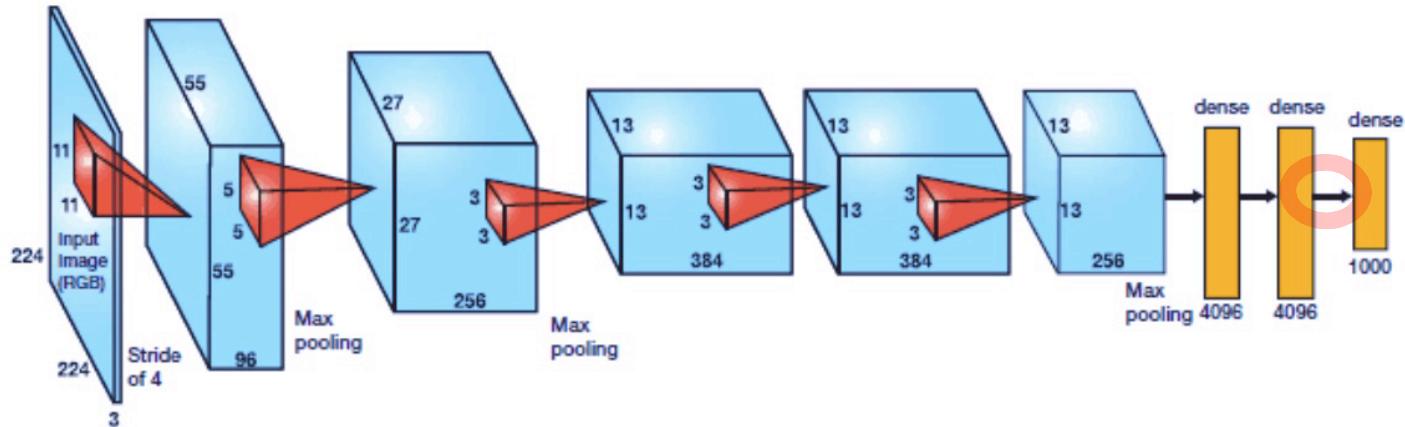
↳ transformed into a vector ($N \times 1$)

- output dimension : $(1, 4096) \times (4096, 1000)$



10. FC2

- Input image size : $(N, 4096)$
- Output dimension : $(N, 4096) \times (4096, 4096)$



11. FC3

- Input Image size : $(N, 4096)$
- output dimension : $(N, 4096) \times (\overbrace{4096, \text{num_classes}}^{\hookrightarrow 313 \text{ label } +})$