

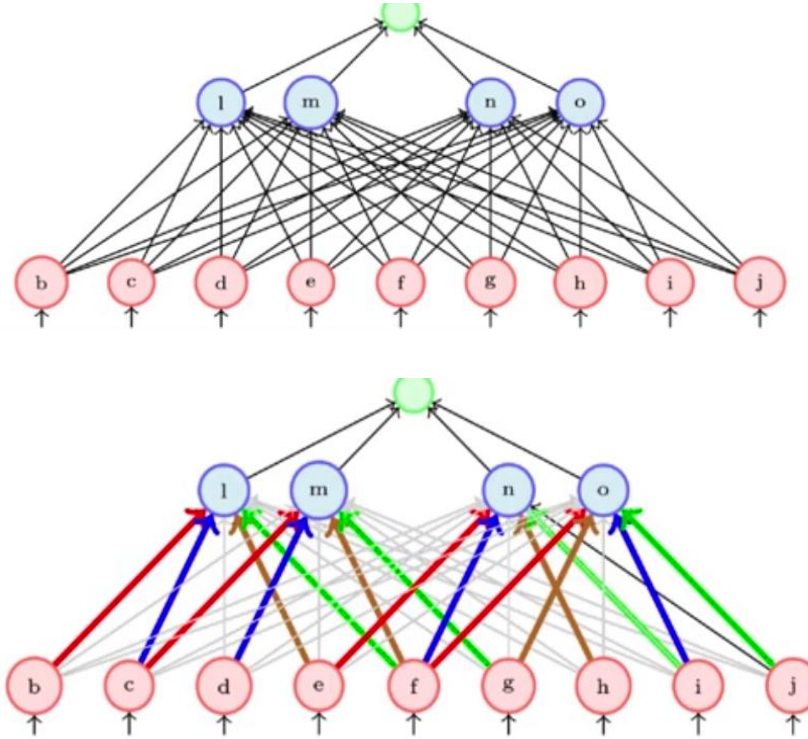
# Deep Learning : AlexNet and ZFNet



राष्ट्रीय प्रौद्योगिकी संस्थान सिक्किम  
NATIONAL INSTITUTE OF TECHNOLOGY SIKKIM

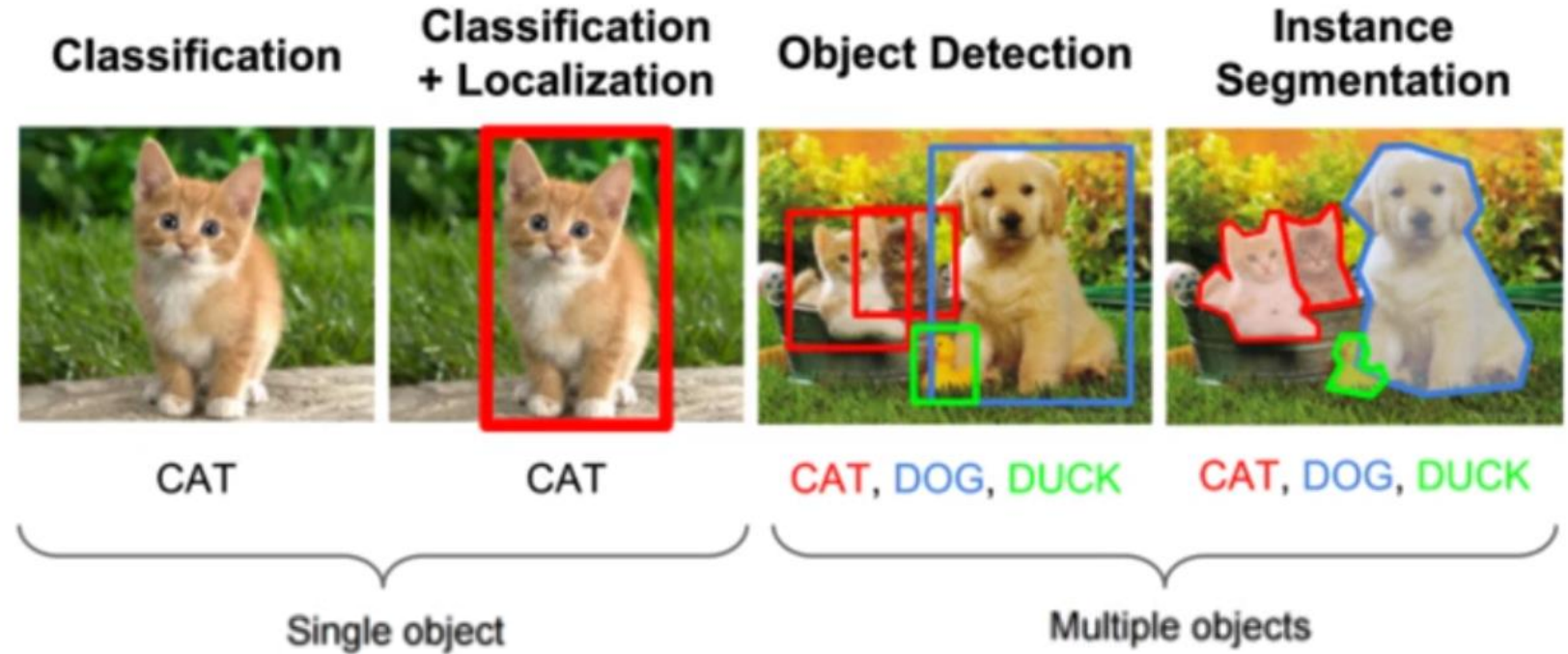
**Course Instructor:**  
Dr. Bam Bahadur Sinha  
*Assistant Professor*  
*Computer Science & Engineering*  
*National Institute of Technology*  
*Sikkim*

# How Do We Train A CNN Model?



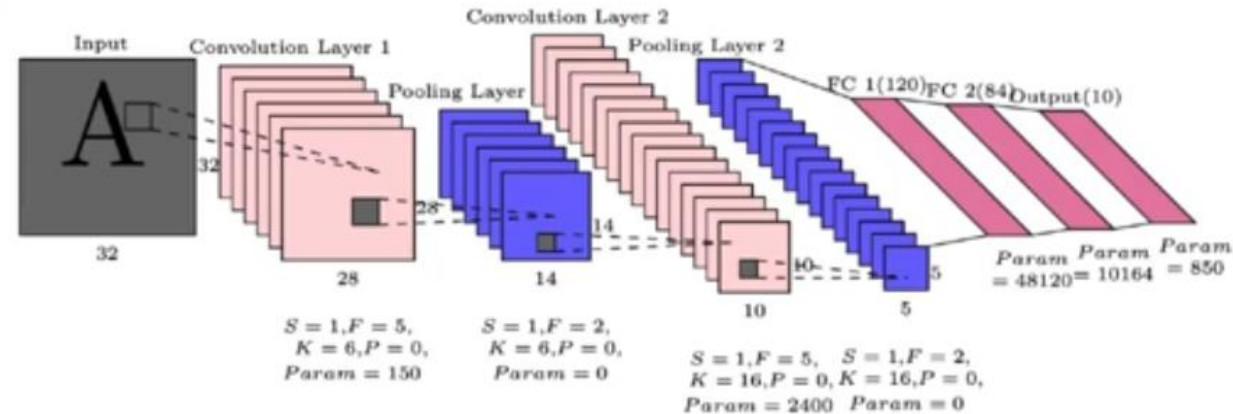
- A CNN can be implemented as a Feedforward Network
- Only a few weights (in color) are active
- The rest of weights (in grey) are zero/inactive

What kinds of tasks are CNNs used for?

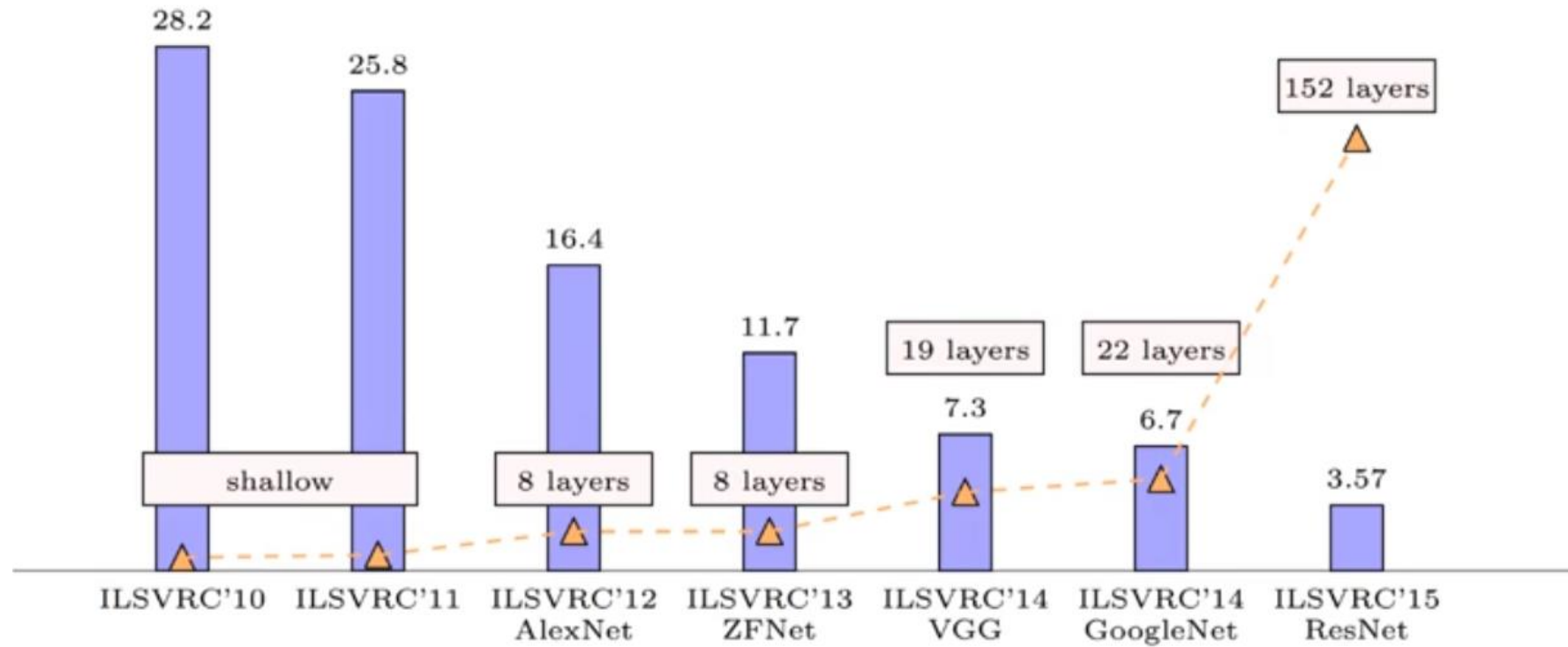


- ✓ Number of layers
- ✓ Number of filters in each layer
- ✓ Filter Size
- ✓ Max pooling

Or just use standard  
tried and tested  
architectures!

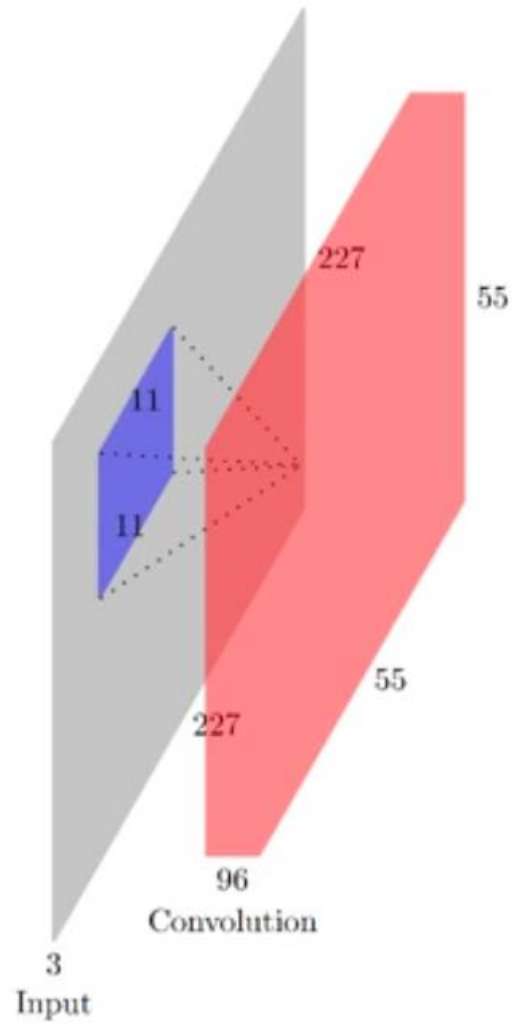


What are few decisions that needs to be taken?



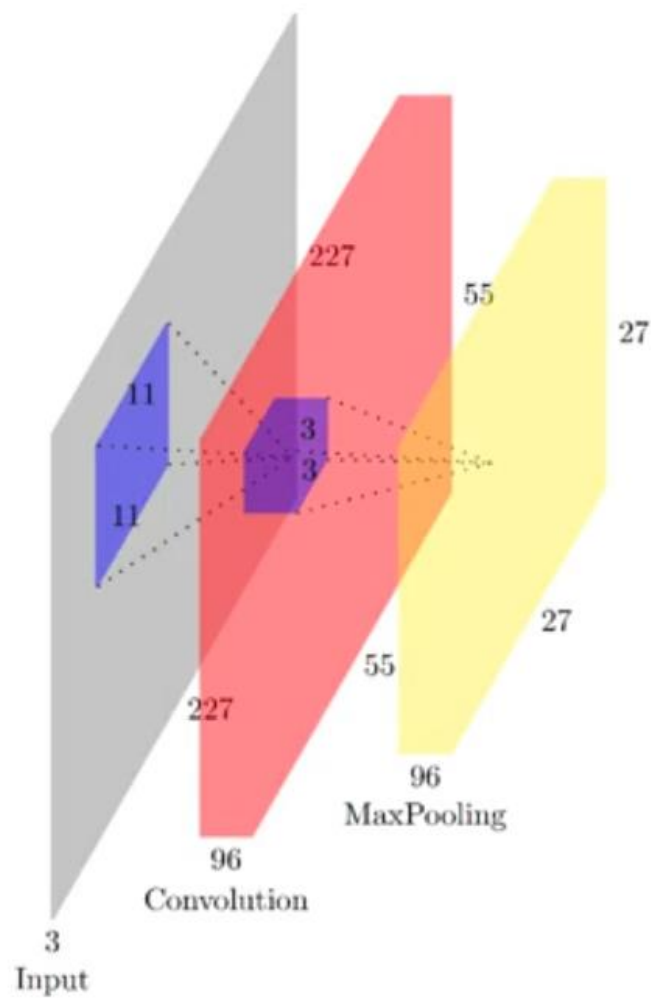
# ImageNet large Scale Visual Recognition Challenge

# AlexNet (2012)



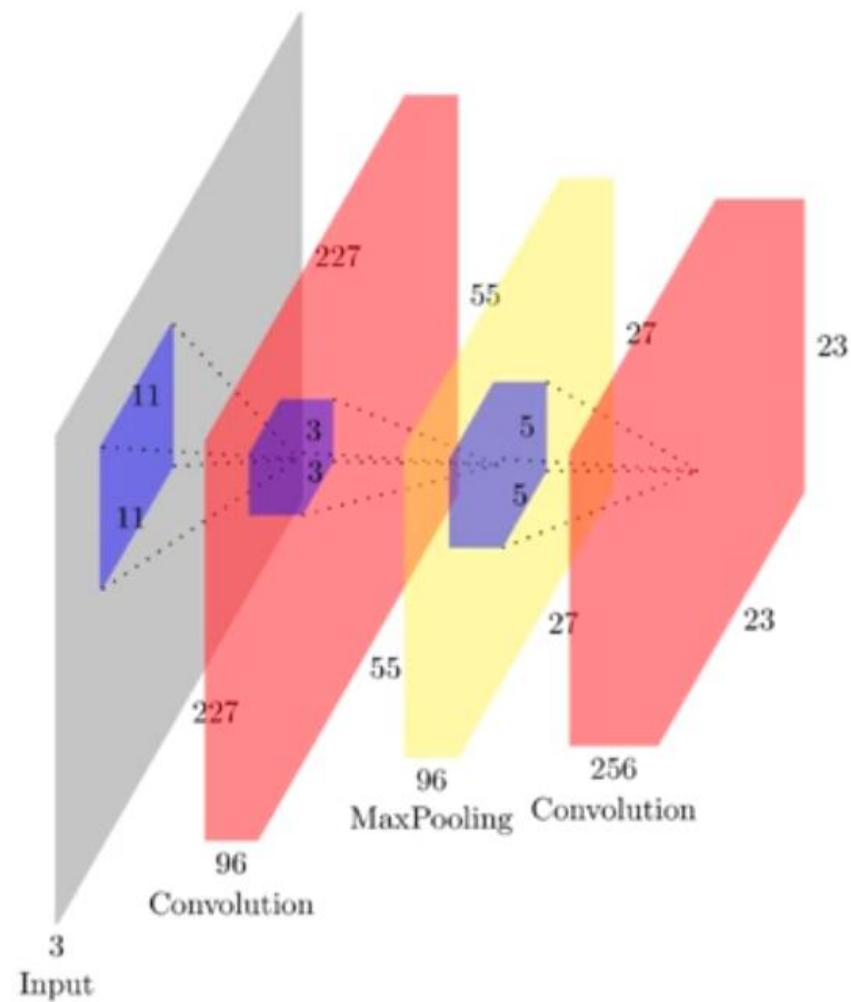
Input:  $227 \times 227 \times 3$   
Conv1:  $K = 96, F = 11$   
 $S = 4, P = 0$   
Output:  $W_2 = 55, H_2 = 55$   
Parameters:  $(11 \times 11 \times 3) \times 96 = 34K$

# AlexNet



Max Pool Input:  $55 \times 55 \times 96$   
 $F = 3, S = 2$   
Output:  $W_2 = 27, H_2 = 27$   
Parameters: 0

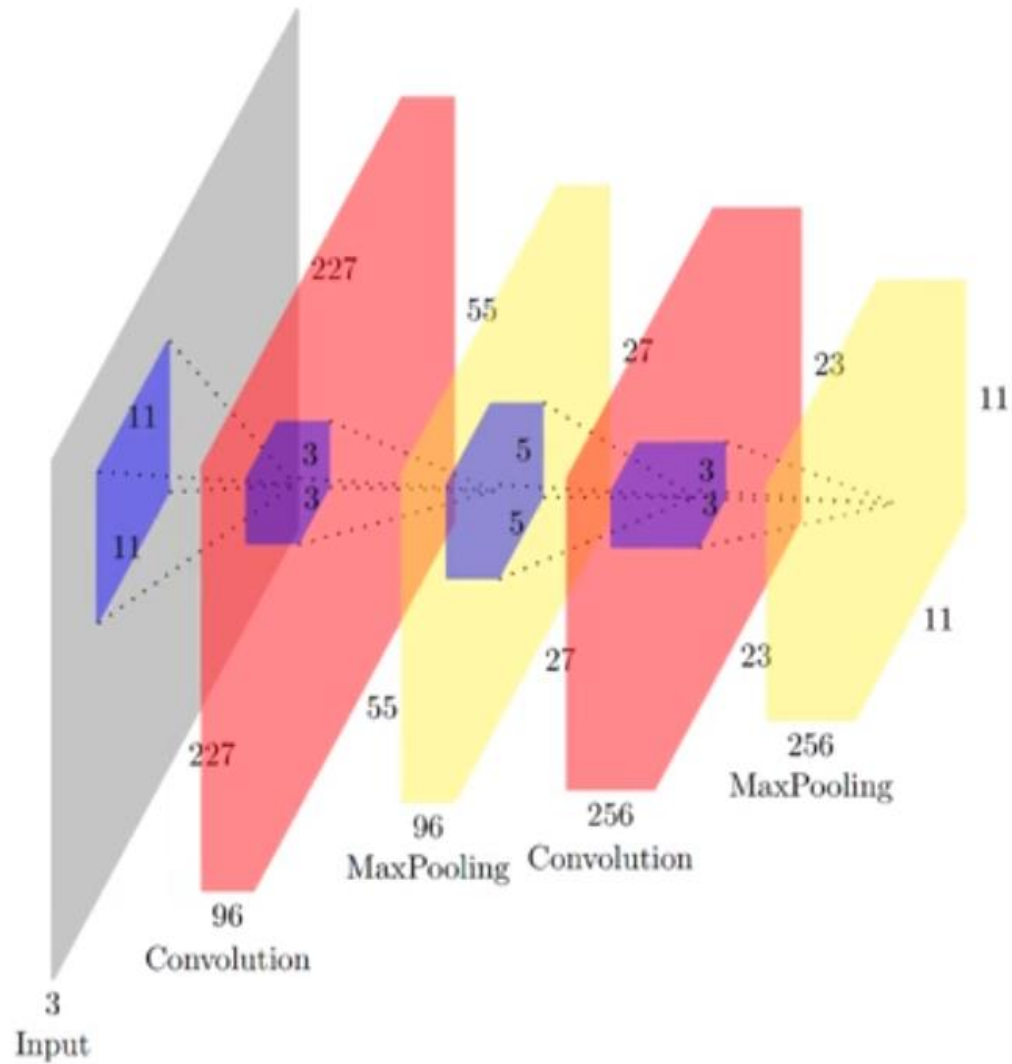
# AlexNet



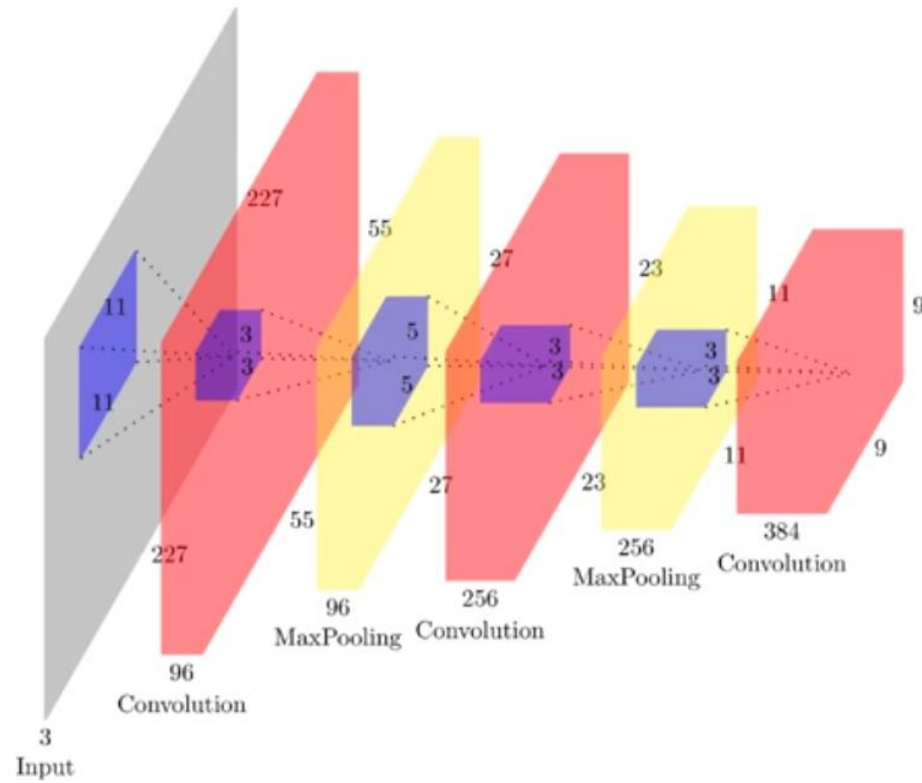
Input:  $27 \times 27 \times 96$   
Conv1:  $K = 256, F = 5$   
 $S = 1, P = 0$   
Output:  $W_2 = 23, H_2 = 23$   
Parameters:  $(5 \times 5 \times 96) \times 256 = 0.6M$



# AlexNet

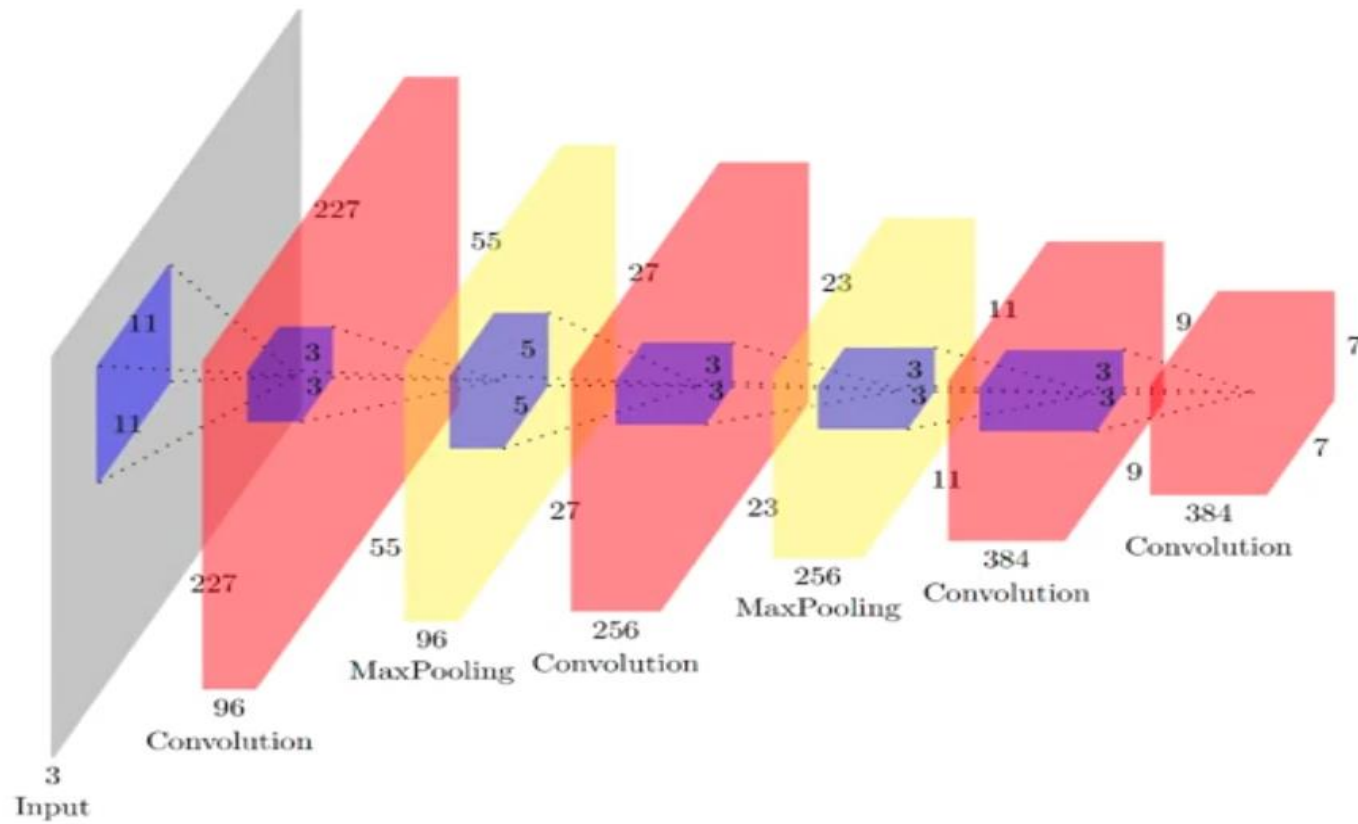


Max Pool Input:  $23 \times 23 \times 256$   
 $F = 3, S = 2$   
Output:  $W_2 = 11, H_2 = 11$   
Parameters: 0



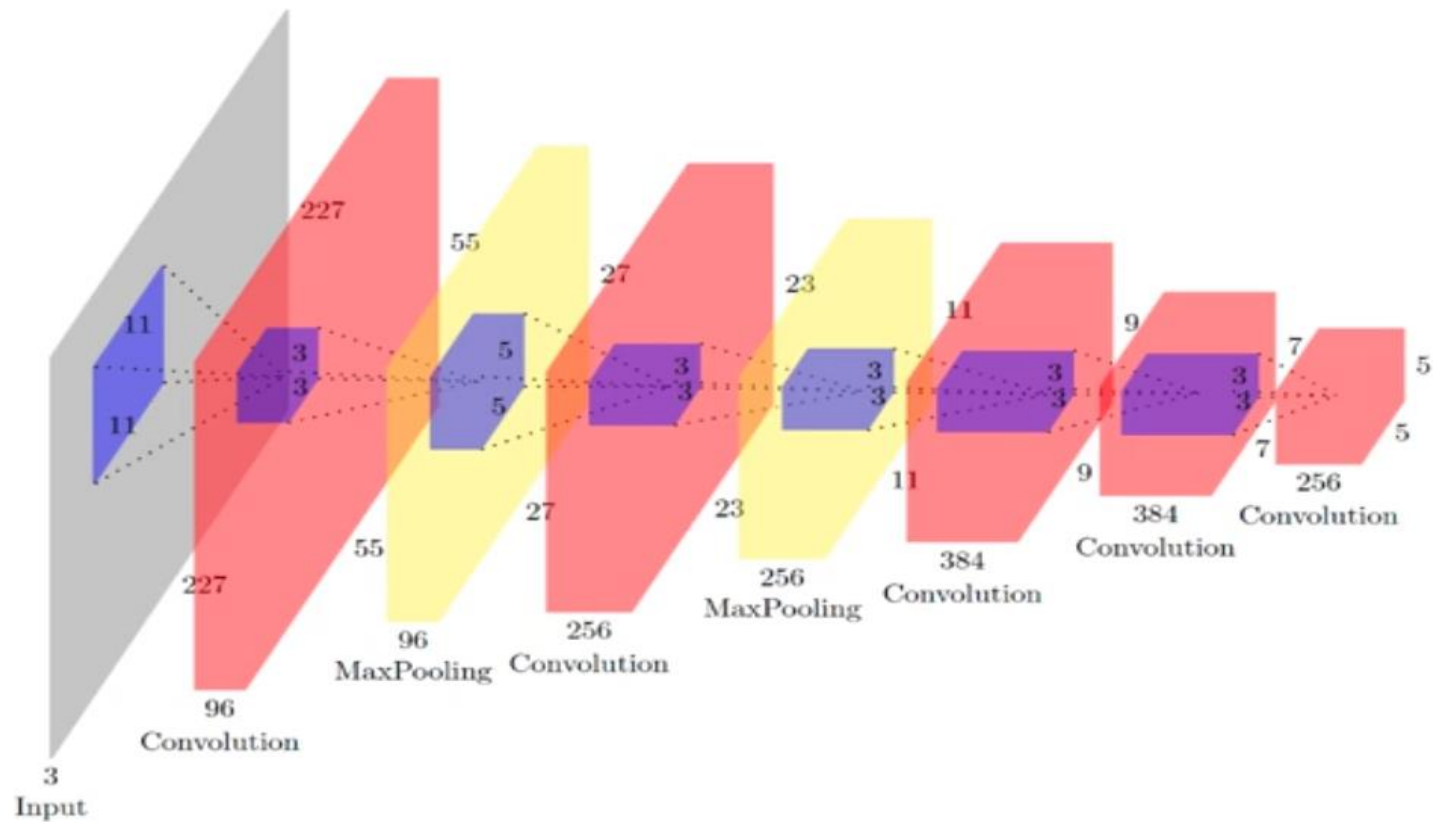
Input:  $11 \times 11 \times 256$   
 Conv1:  $K = 384, F = 3$   
 $S = 1, P = 0$   
 Output:  $W_2 = 9, H_2 = 9$   
 Parameters:  $(3 \times 3 \times 256) \times 384 = 0.8M$

# AlexNet



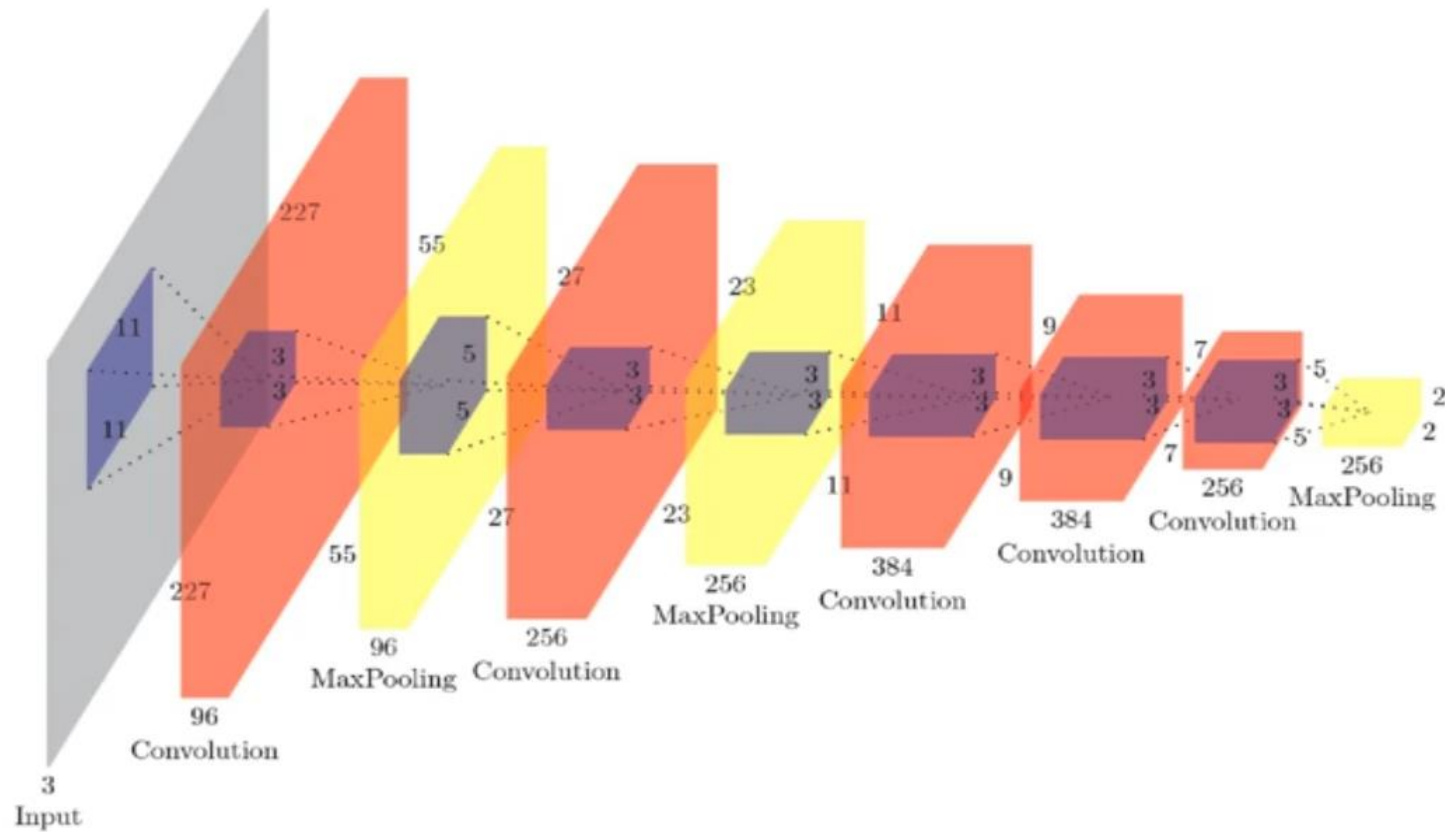
Input:  $9 \times 9 \times 384$   
 Conv1:  $K = 384, F = 3$   
 $S = 1, P = 0$   
 Output:  $W_2 = 7, H_2 = 7$   
 Parameters:  $(3 \times 3 \times 384) \times 384 = 1.327M$

# AlexNet



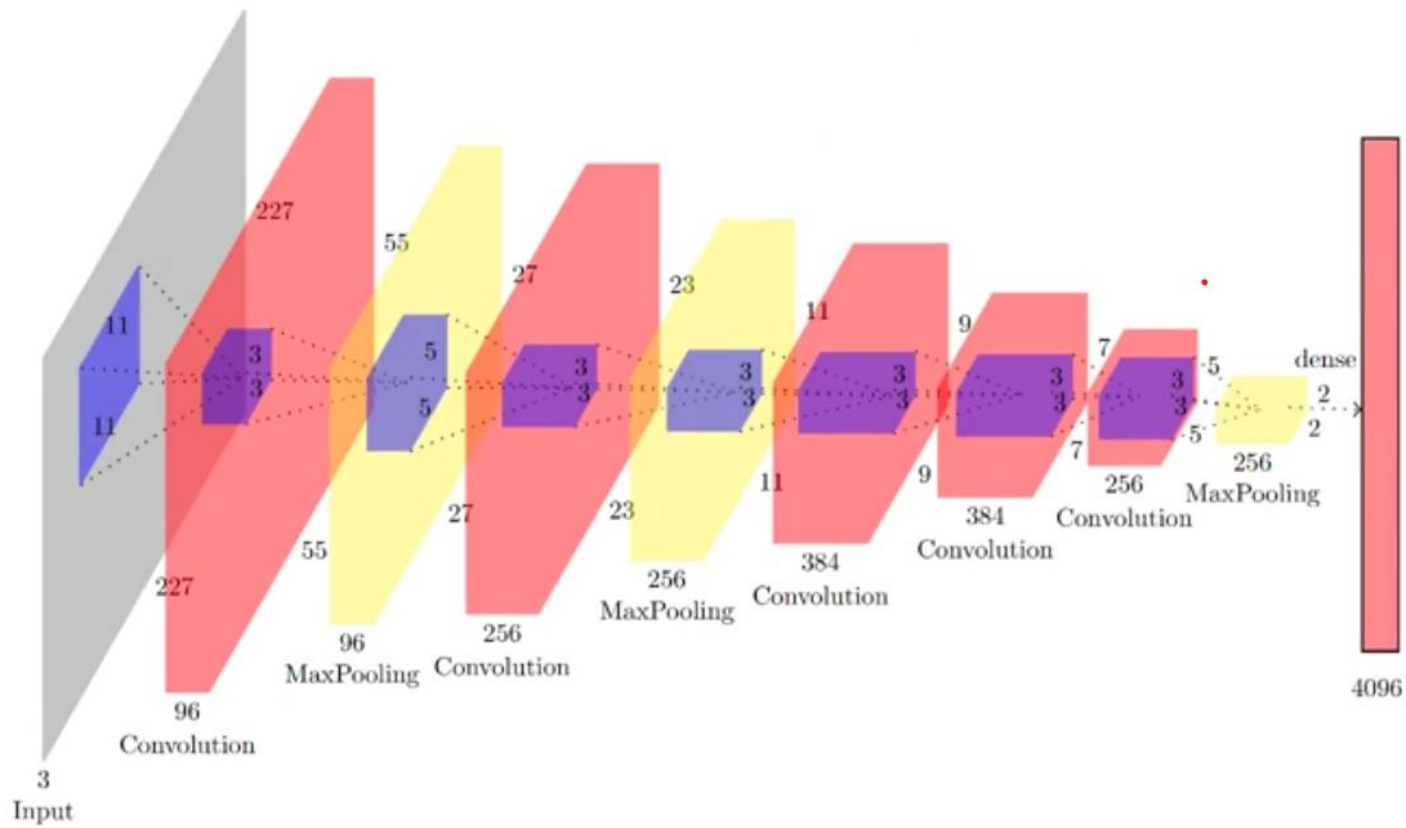
Input:  $7 \times 7 \times 384$   
 Conv1:  $K = 256, F = 3$   
 $S = 1, P = 0$   
 Output:  $W_2 = 5, H_2 = 5$   
 Parameters:  $(3 \times 3 \times 384) \times 256 = 0.8M$

# AlexNet



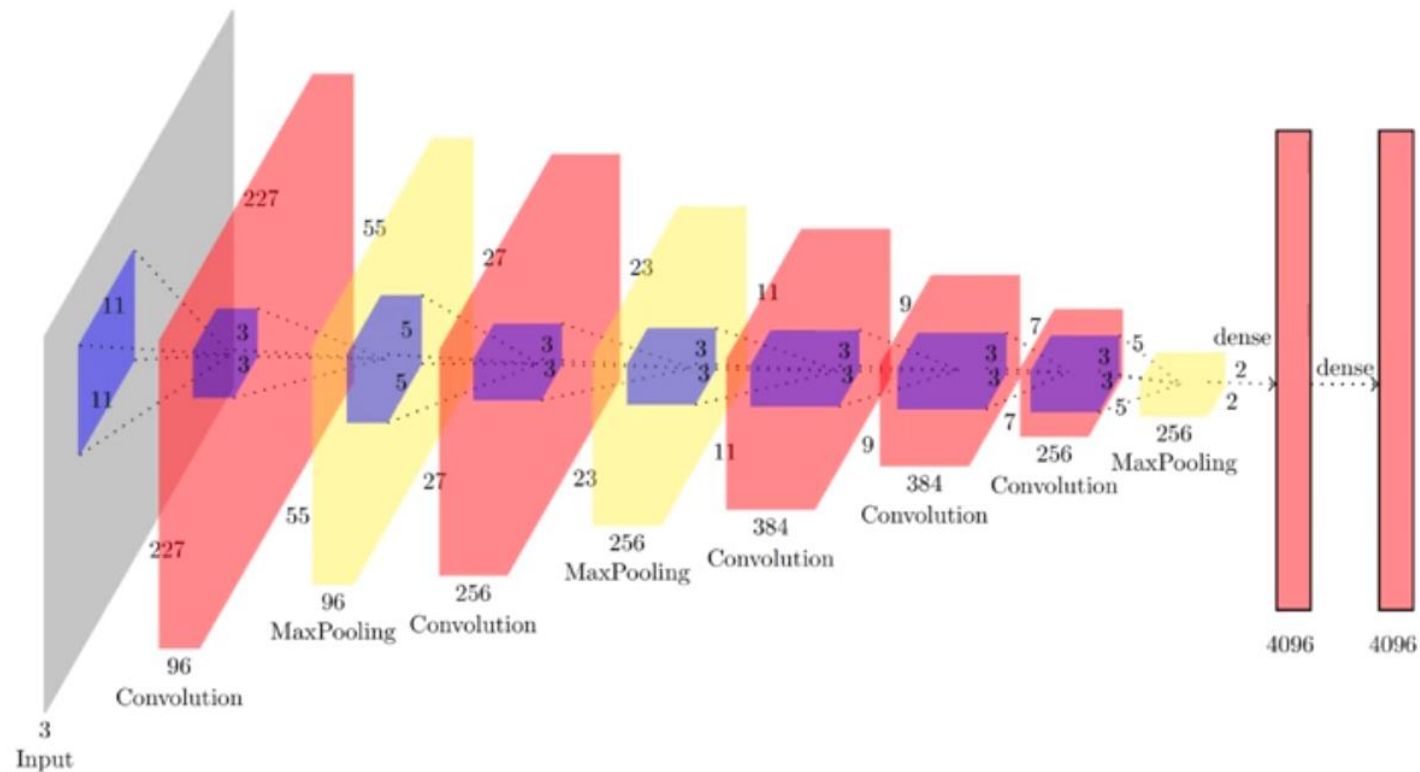
Max Pool Input:  $5 \times 5 \times 256$   
 $F = 3, S = 2$   
 Output:  $W_2 = 2, H_2 = 2$   
 Parameters: 0

# AlexNet



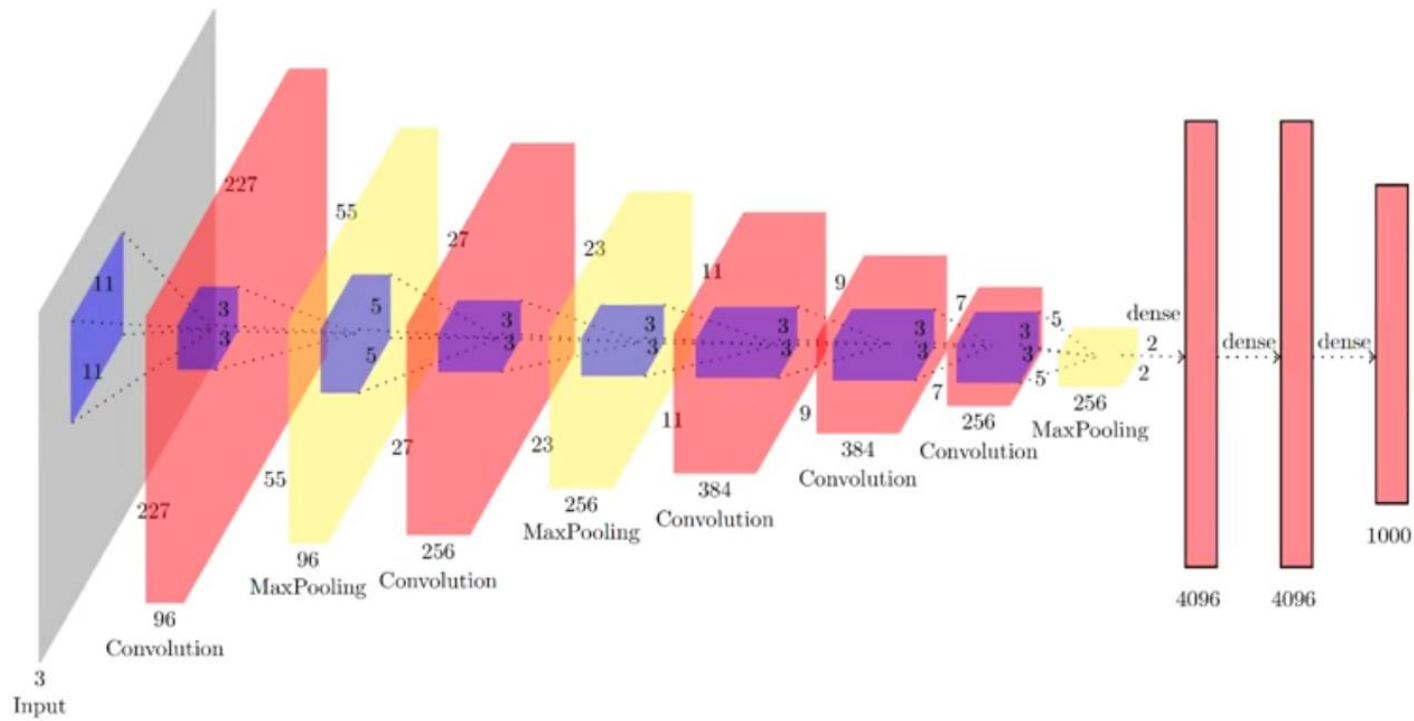
FC1  
Parameters:  $(2 \times 2 \times 256) \times 4096 = 4M$

# AlexNet



FC1  
Parameters:  $4096 \times 4096 = 16M$

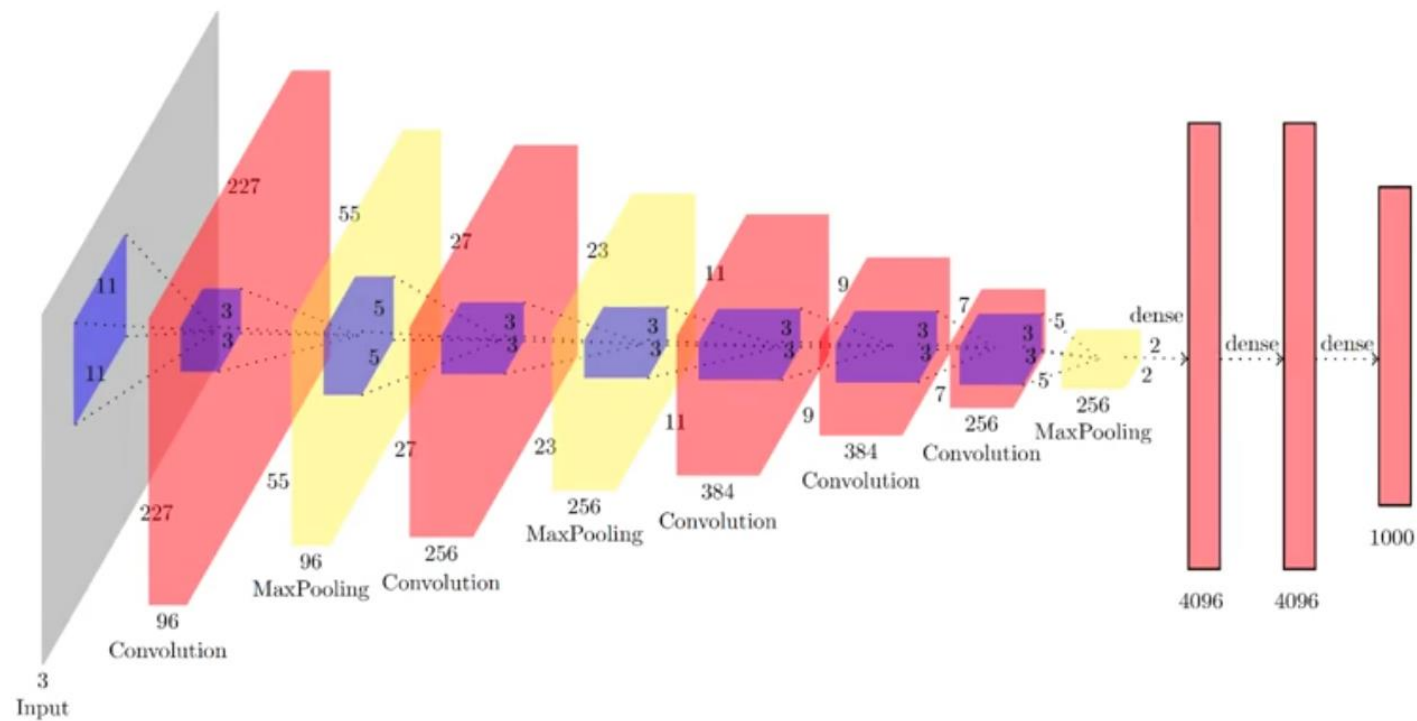
# AlexNet



FC1  
Parameters:  $4096 \times 1000 = 4M$

# AlexNet

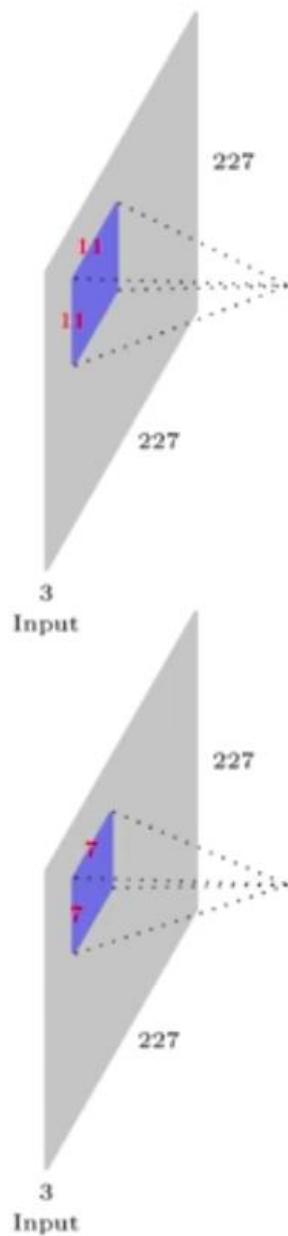




Total Parameters: 27.55M

# AlexNet

# AlexNet (8 layers) vs ZFNet (8 layers)



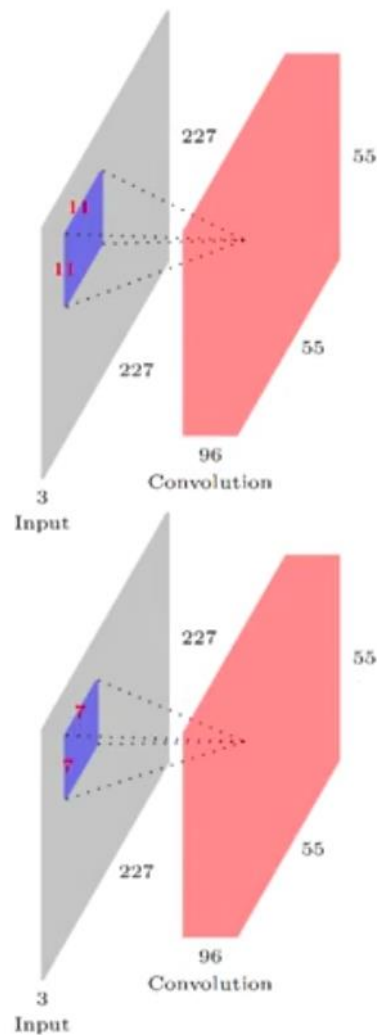
$P = 0$

$S = 4$

Layer1:  $F = 11 \rightarrow 7$   
Difference in Parameters  
 $((11^2 - 7^2) \times 3) \times 96 = 20.7K$

# AlexNet Vs ZFNet

## LAYER 1



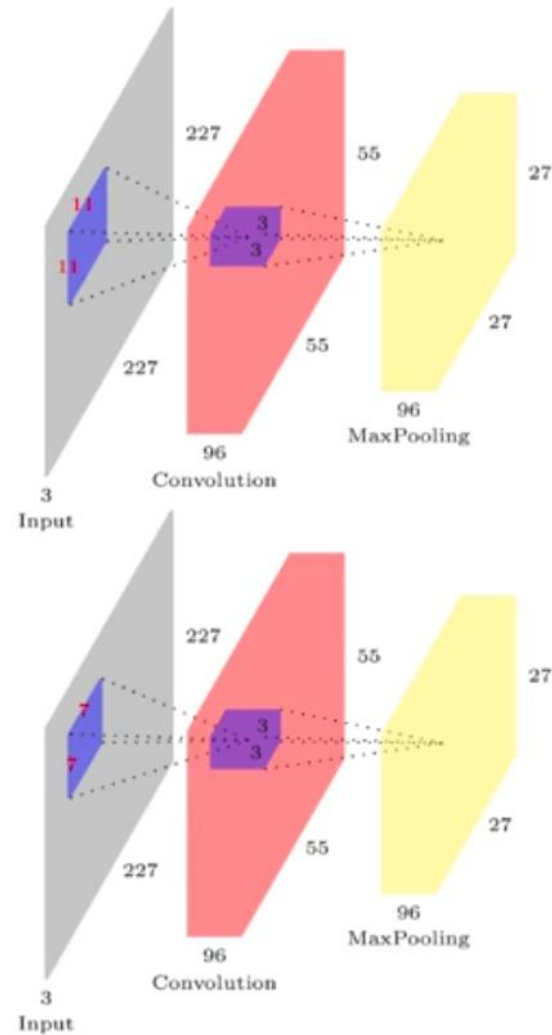
$P = 0$

$S = 4$

Layer1:  $F = 11 \rightarrow 7$   
Difference in Parameters  
 $((11^2 - 7^2) \times 3) \times 96 = 20.7K$

# AlexNet Vs ZFNet

## LAYER 2

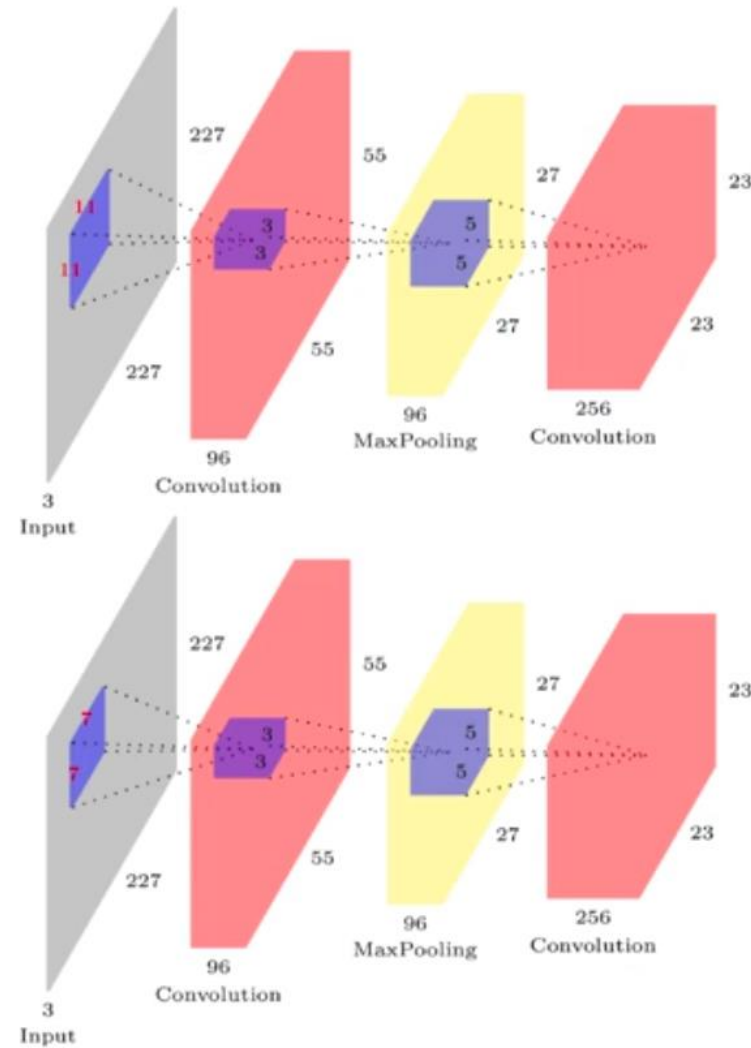


$P = 0$   
 $S = 2$

Layer2: No difference

# AlexNet Vs ZFNet

## LAYER 3

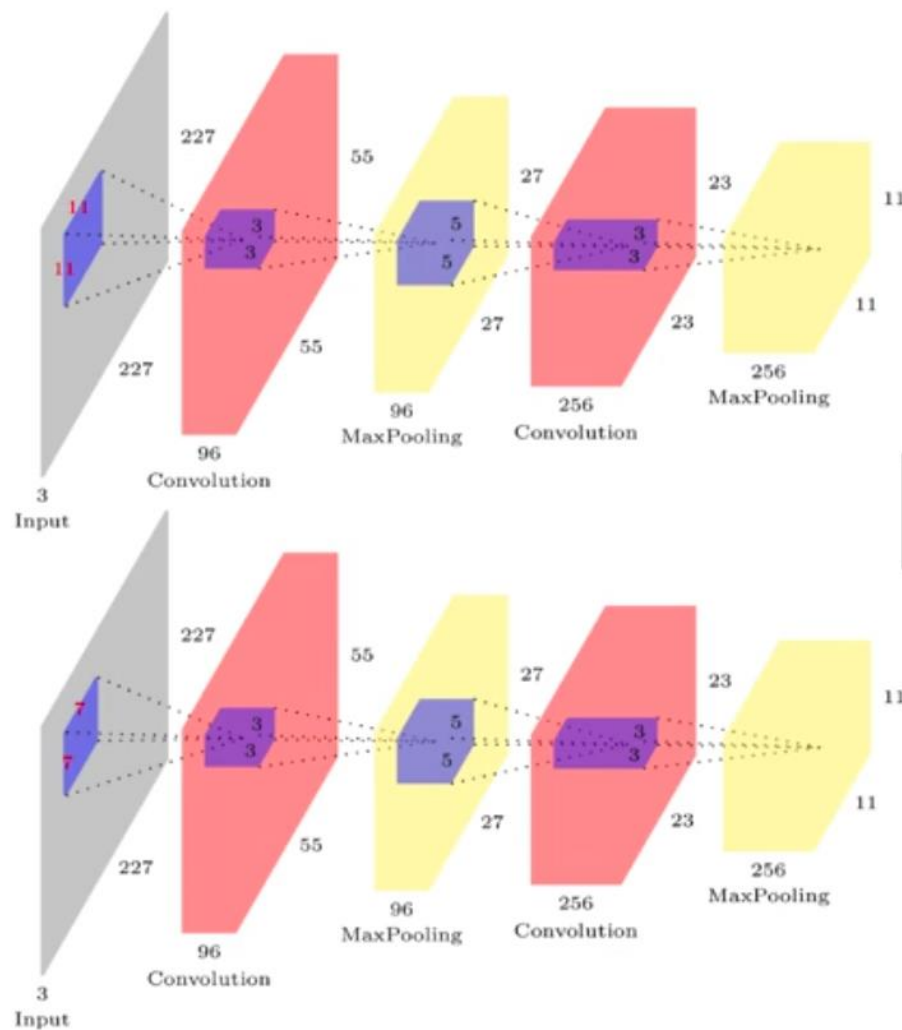


$P = 0$   
 $S = 1$

Layer3: No difference

# AlexNet Vs ZFNet

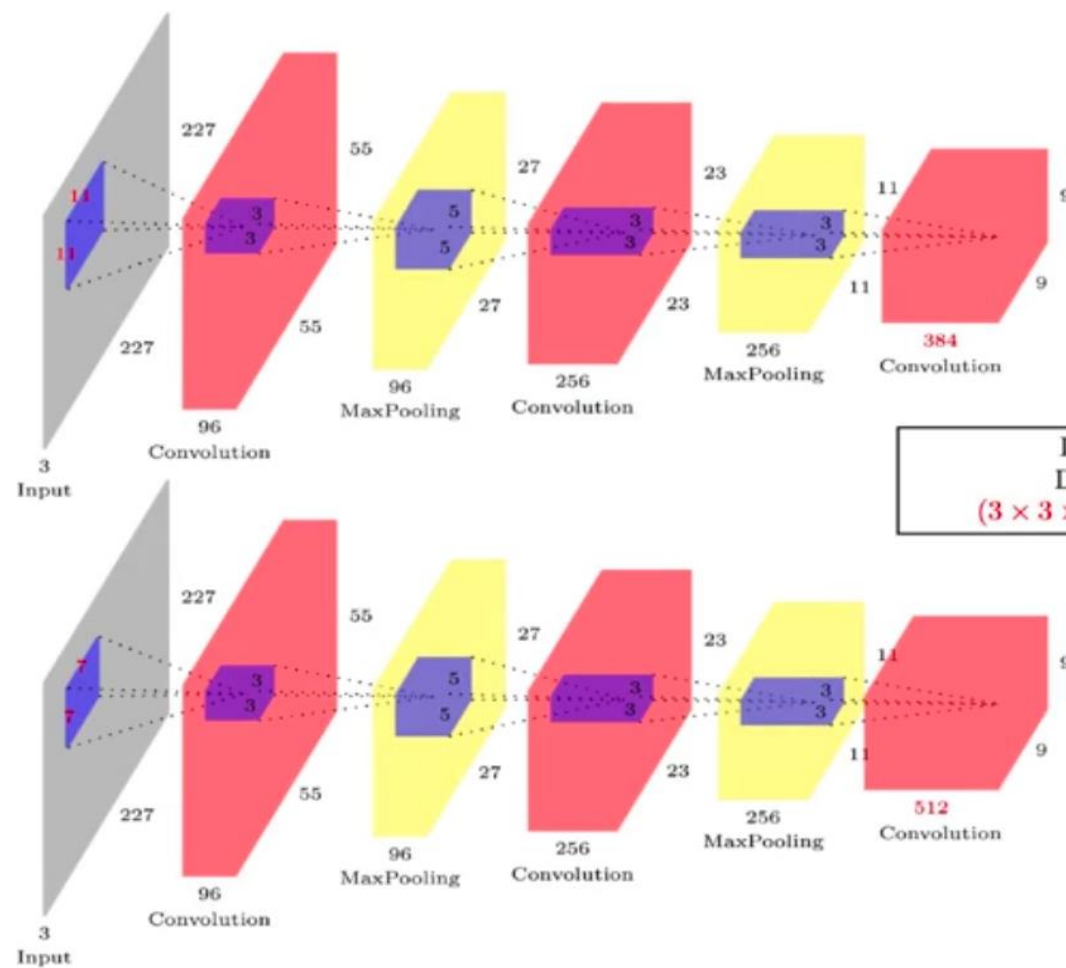
## LAYER 4



P = 0  
S = 2

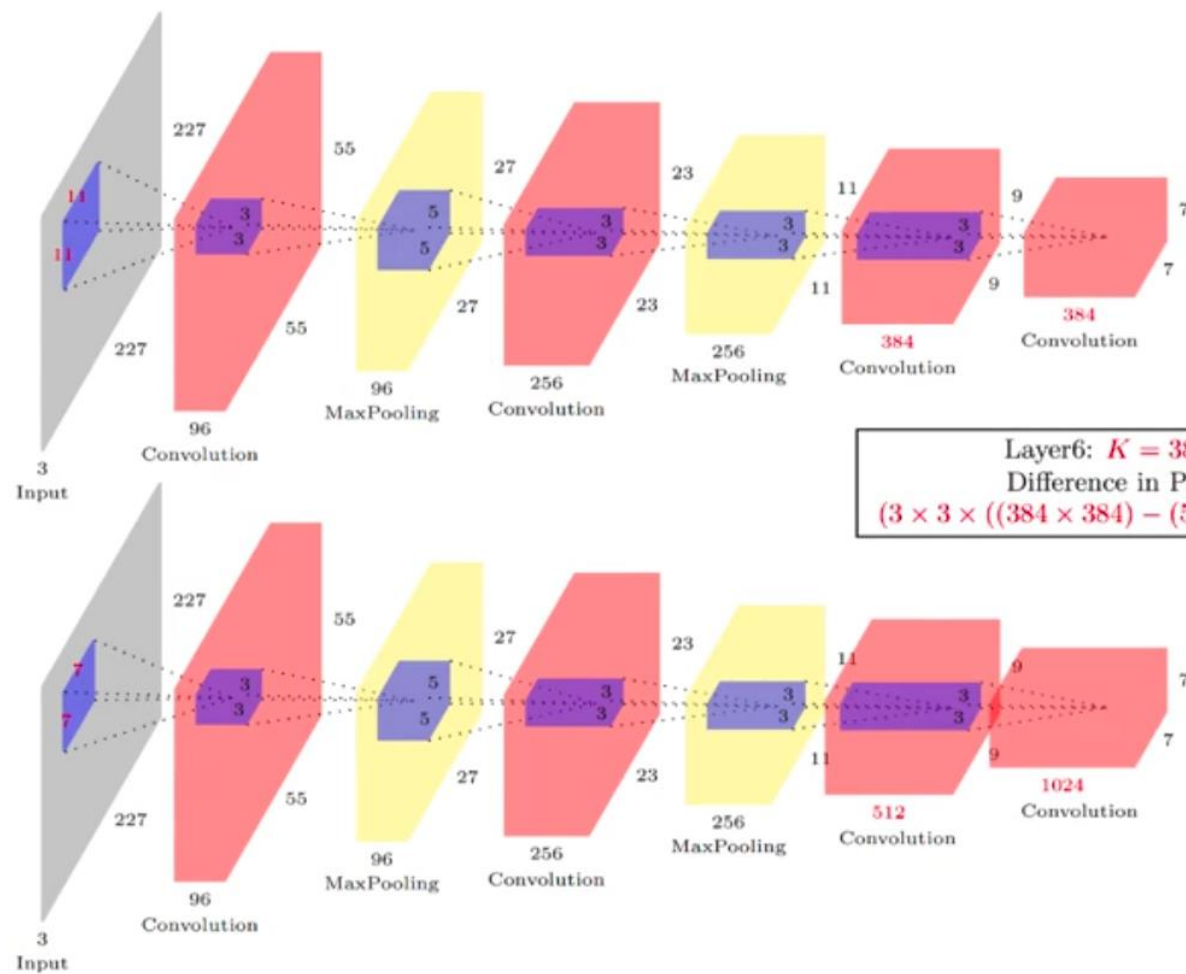
Layer4: No difference

# AlexNet Vs ZFNet LAYER 5



P = 0  
S = 1

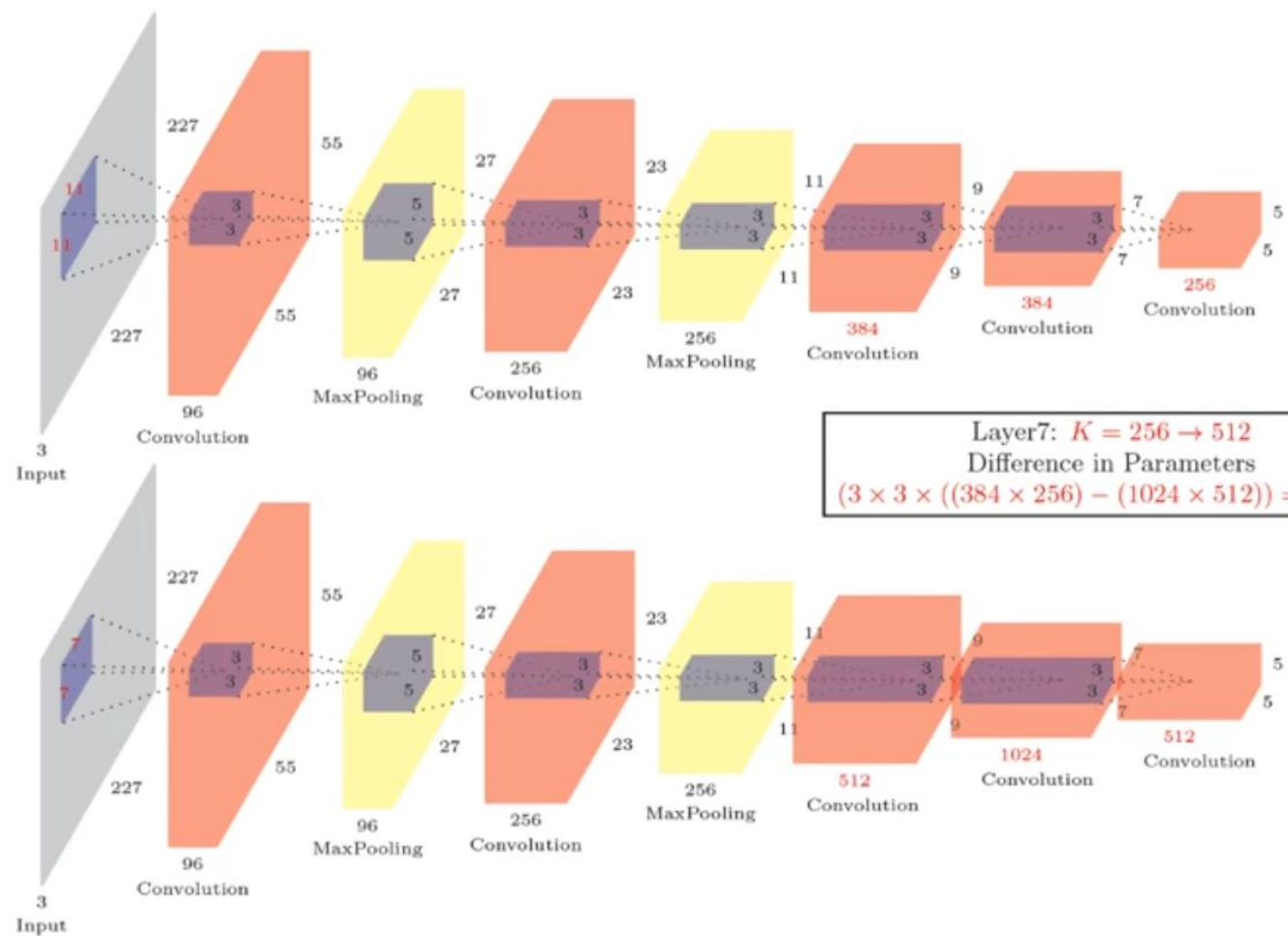
# AlexNet Vs ZFNet LAYER 6



P = 0  
S = 1



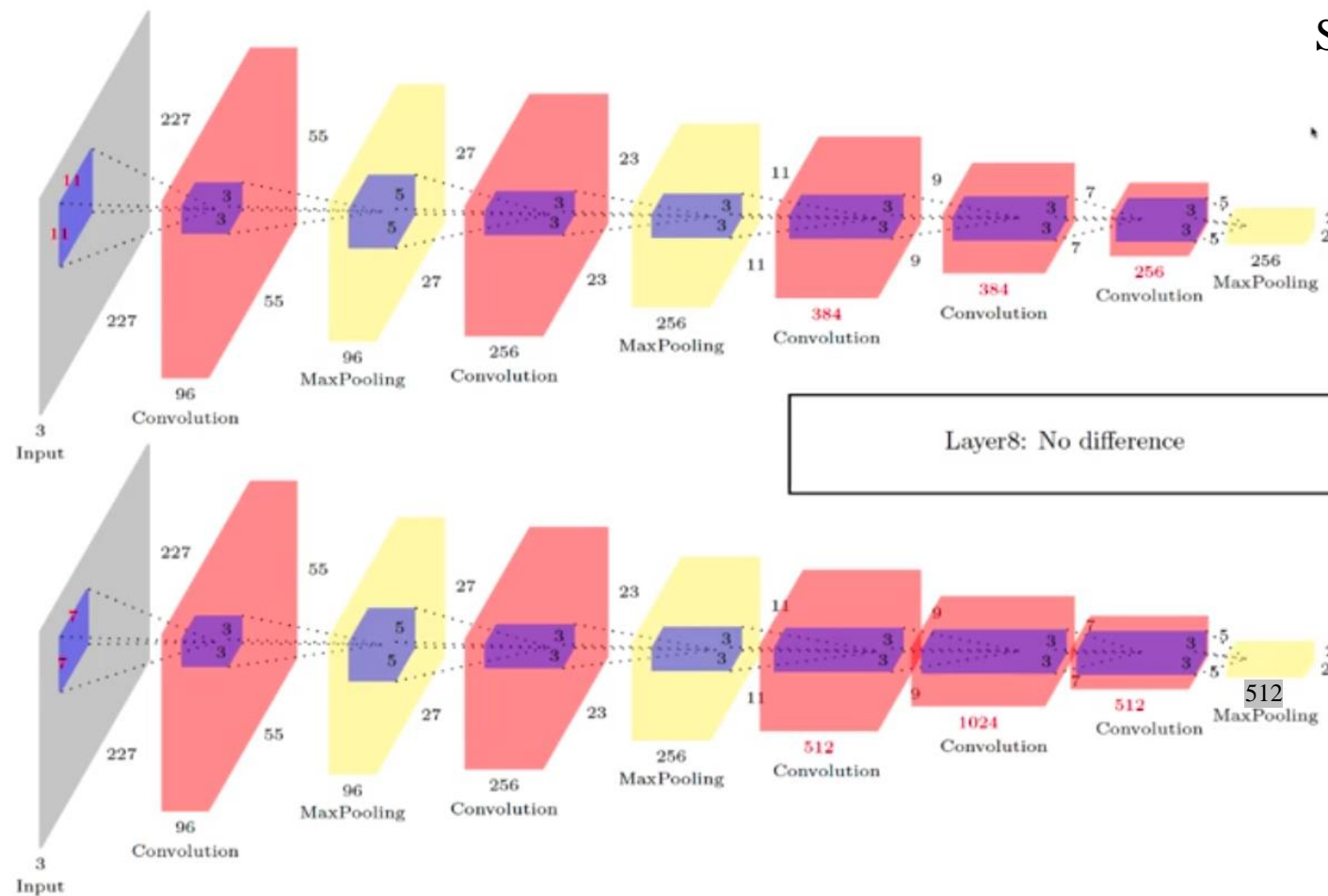
# AlexNet Vs ZFNet LAYER 7



P = 0  
S = 1

# AlexNet Vs ZFNet LAYER 8

$P = 0$   
 $S = 2$



# AlexNet Vs ZFNet

## Fully Connected Layers

