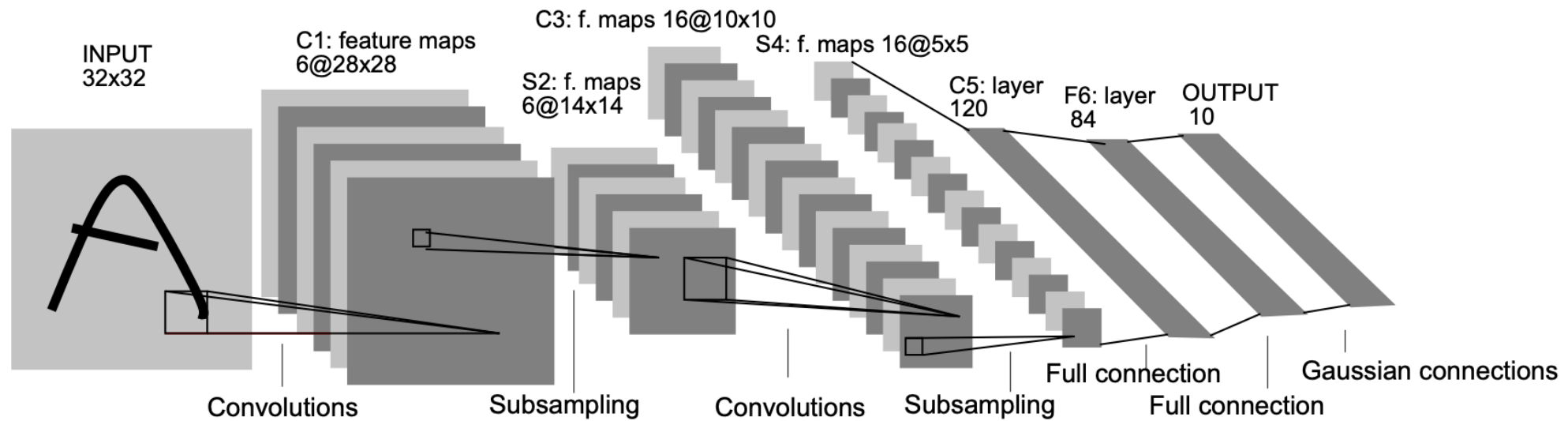


# Neural Architecture Search

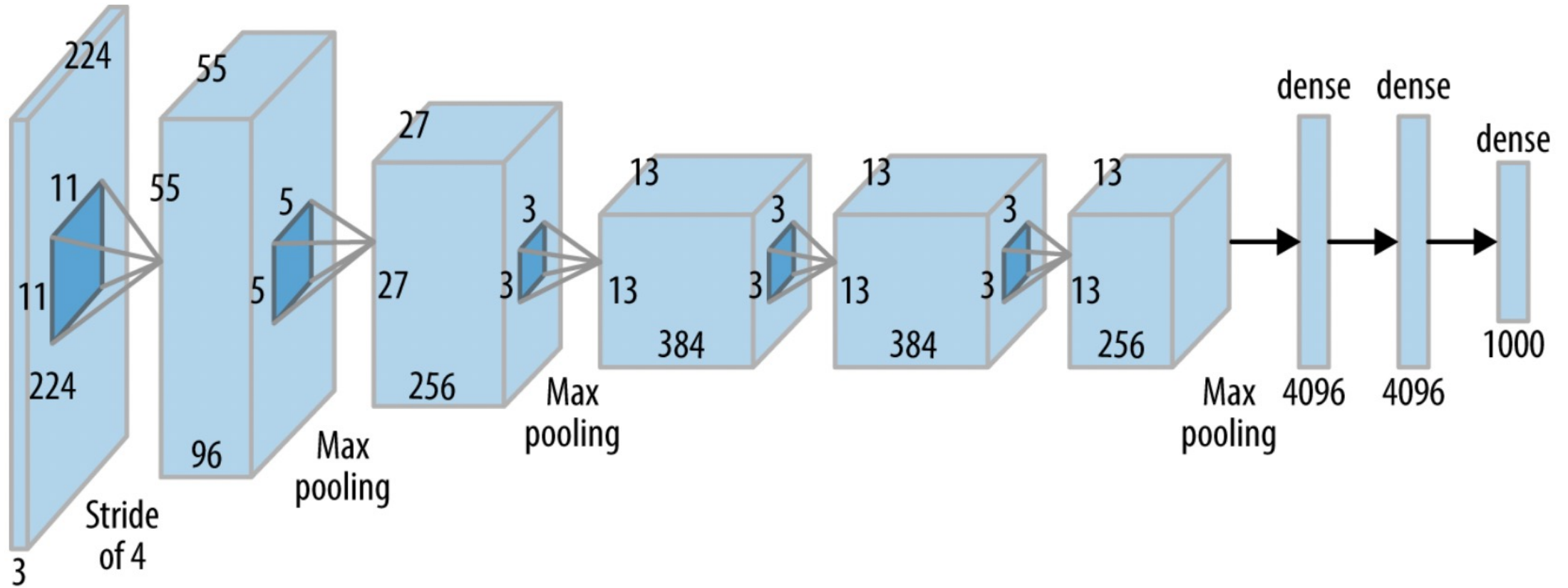
Neural Networks Design And Application

# Manually designed architectures

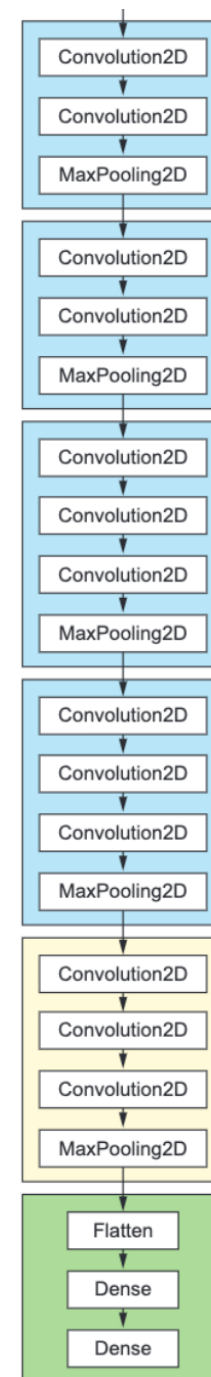
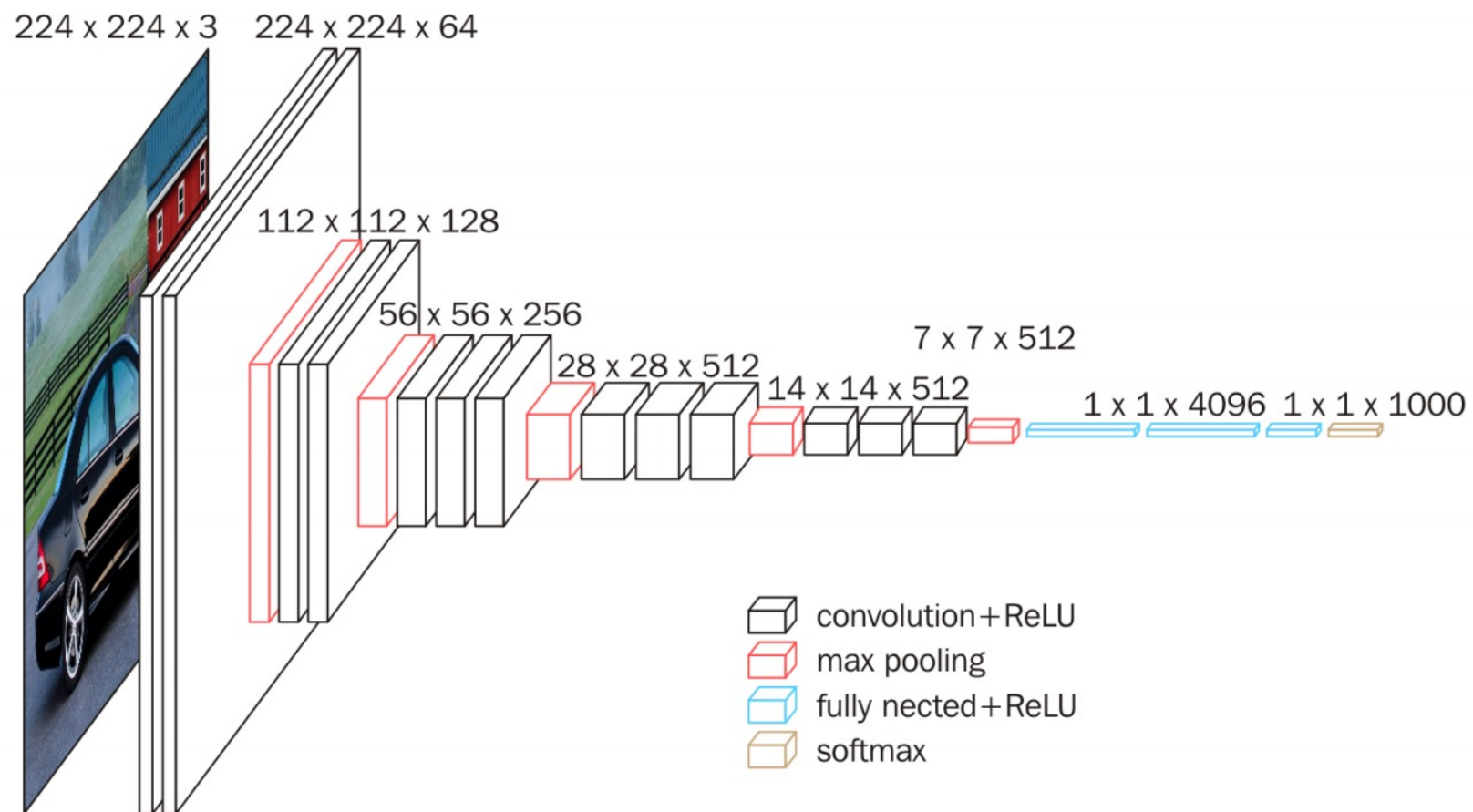


**Fig. 1.** Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

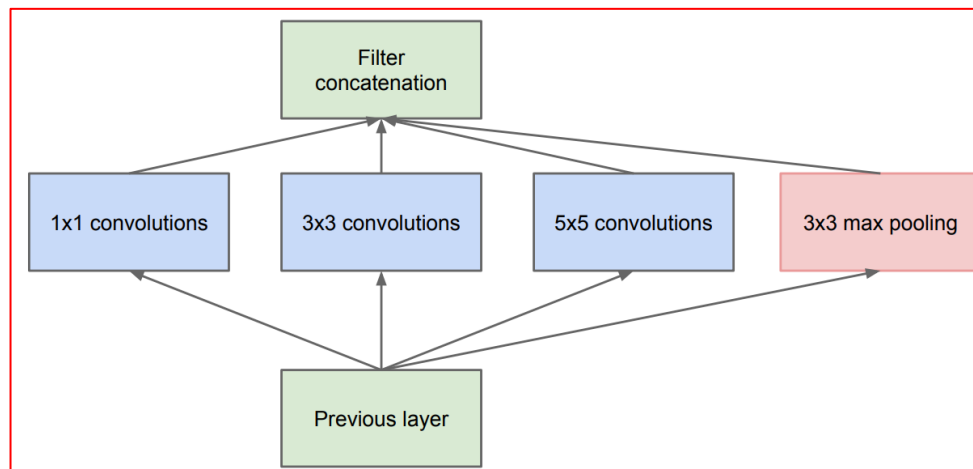
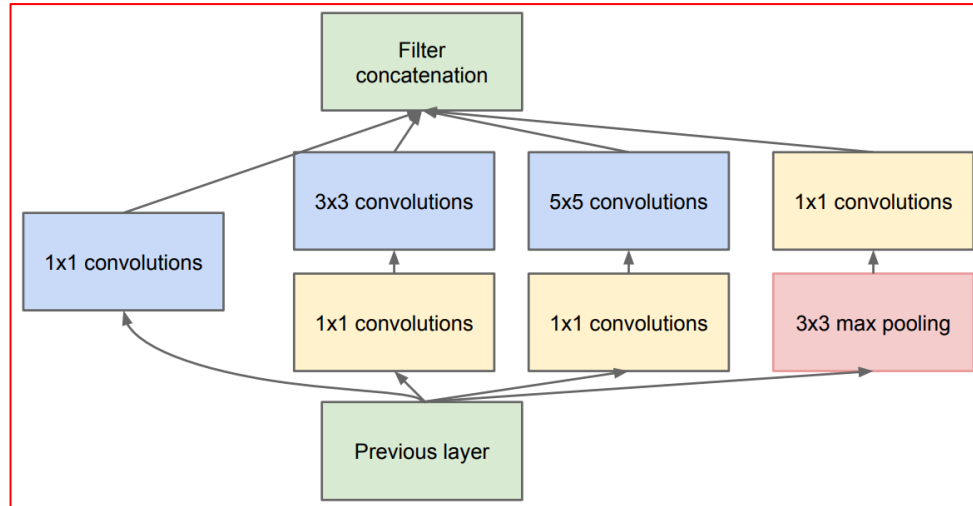
# AlexNet



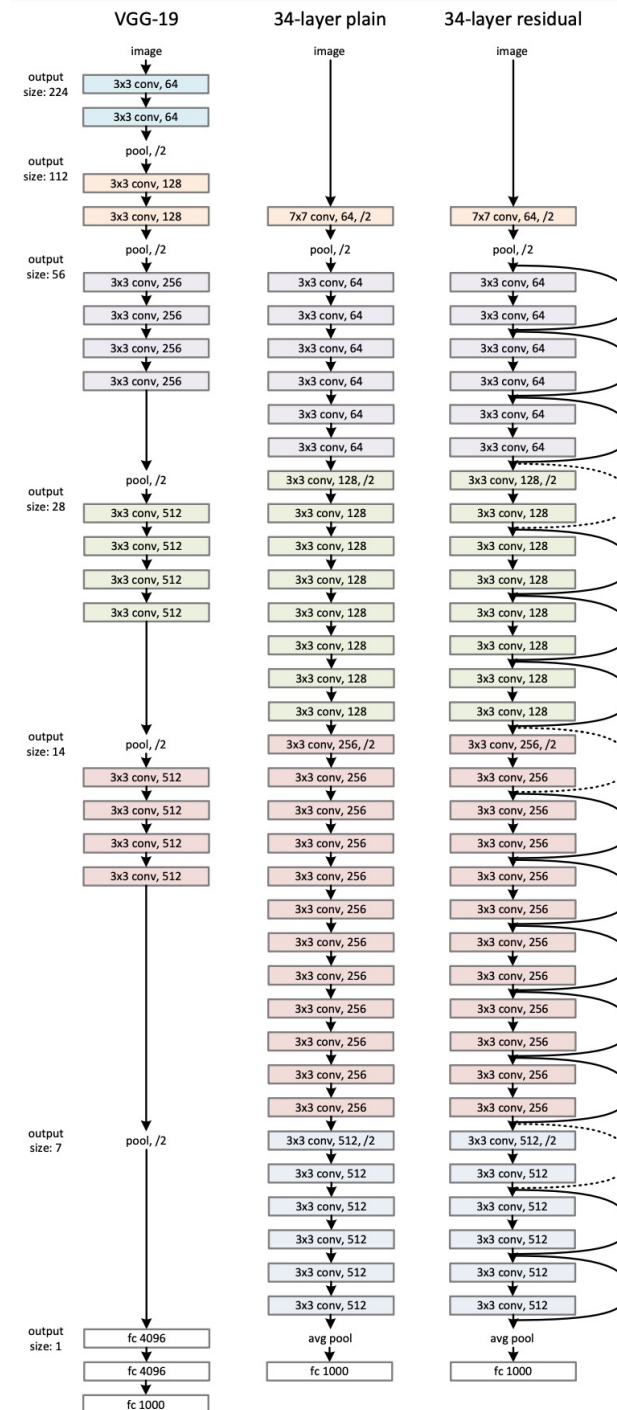
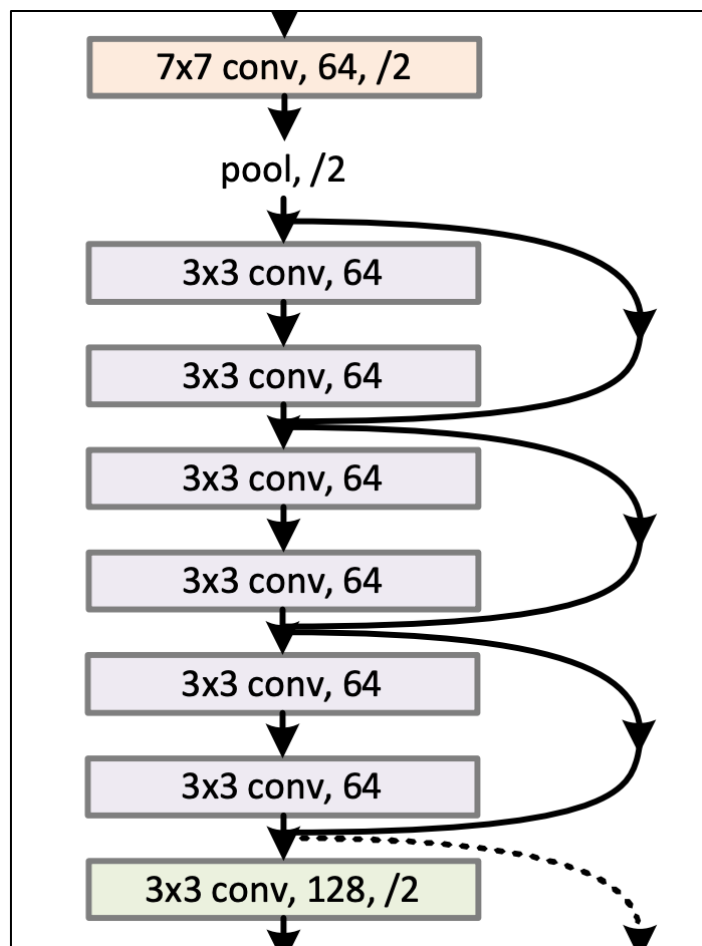
# VGG-16



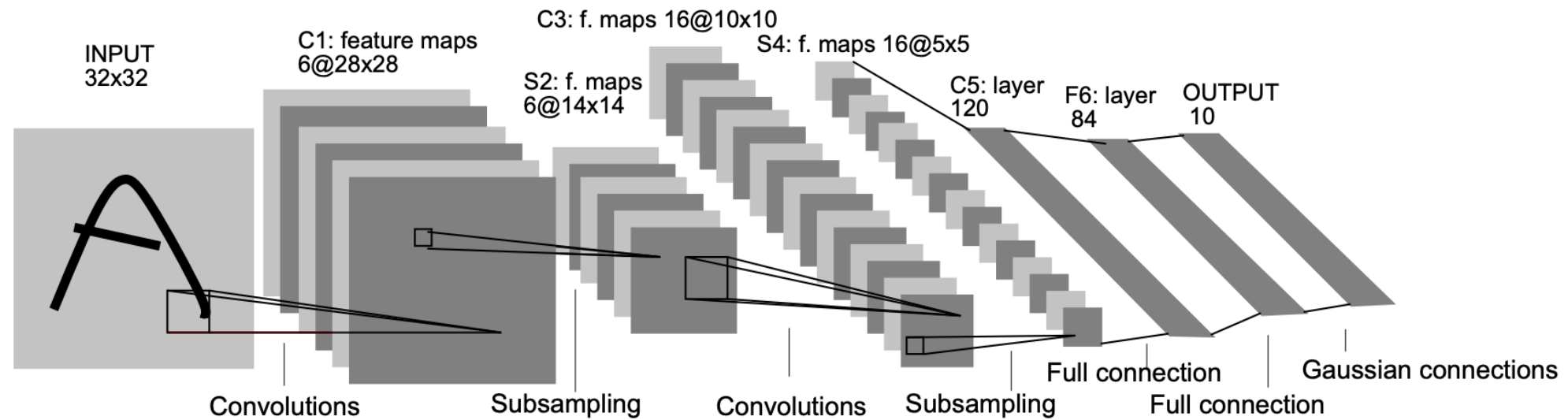
# Inception (GoogLeNet)



# ResNet

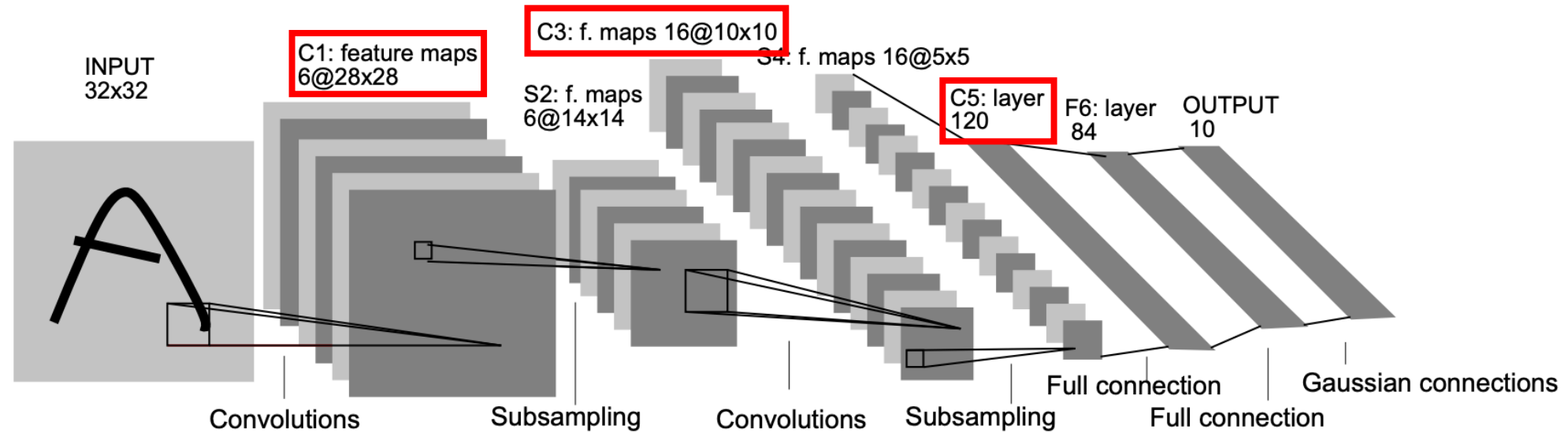


# LeNet-5



**Fig. 1.** Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

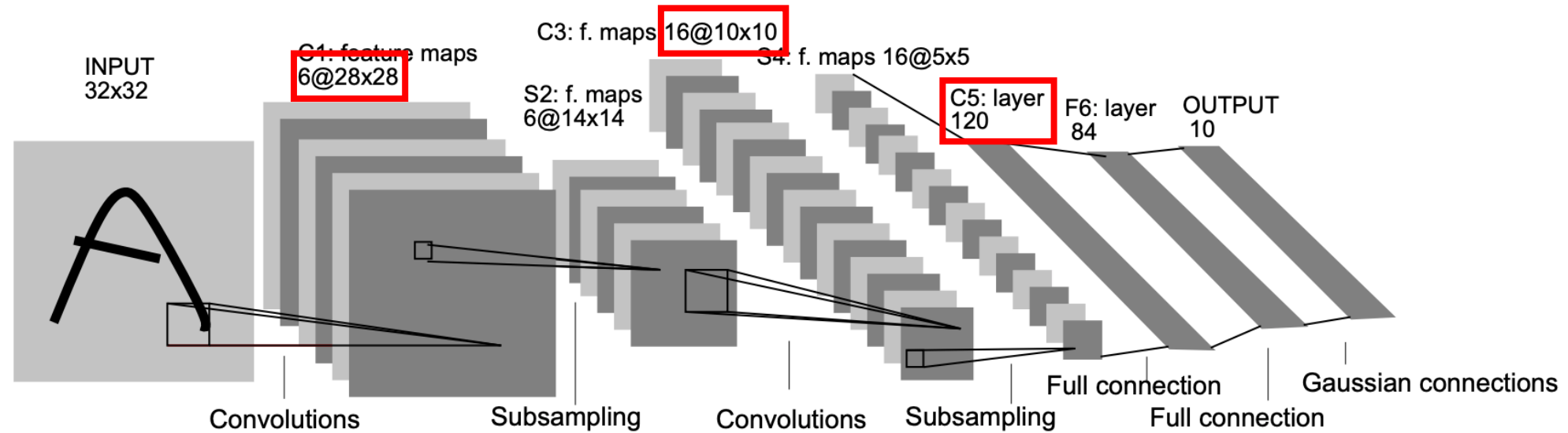
# LeNet-5



**Fig. 1.** Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

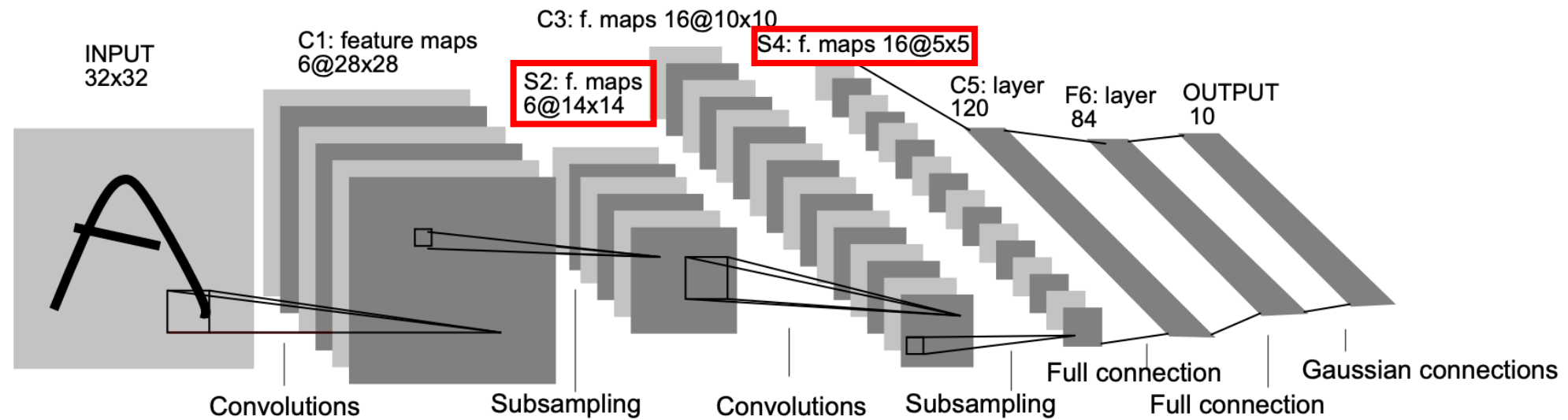


# LeNet-5



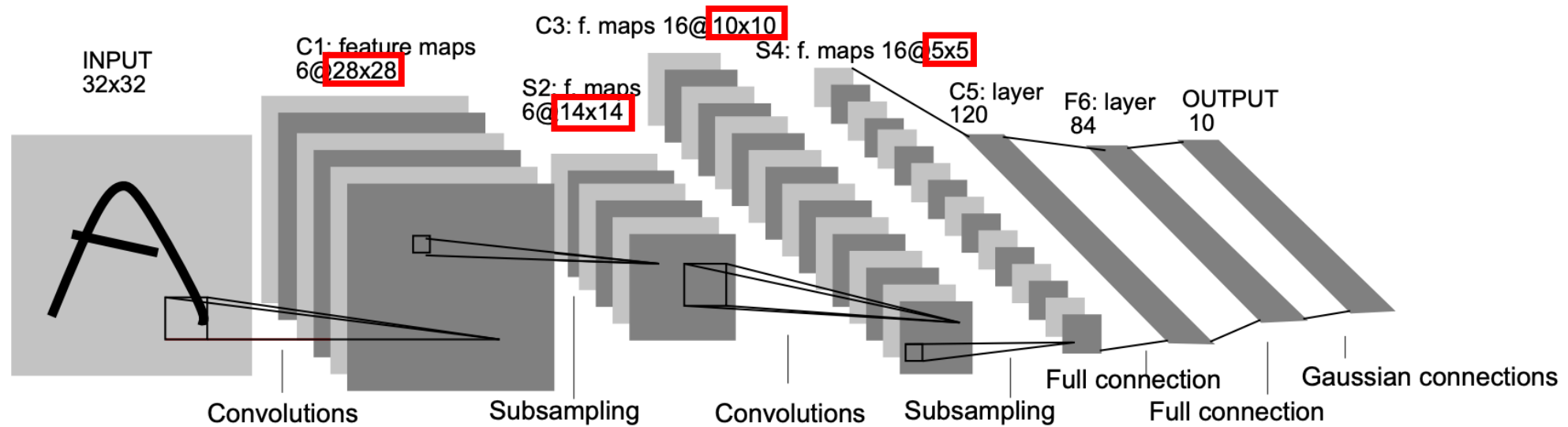
**Fig. 1.** Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

# LeNet-5



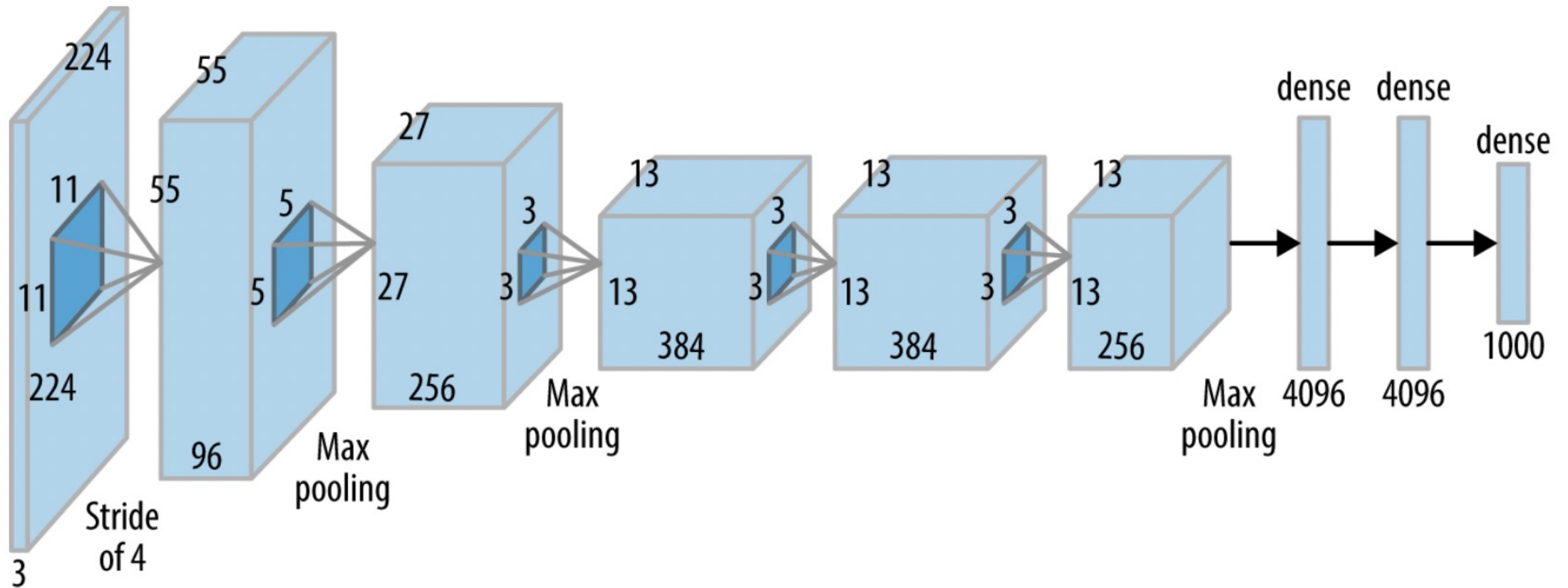
**Fig. 1.** Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

# LeNet-5

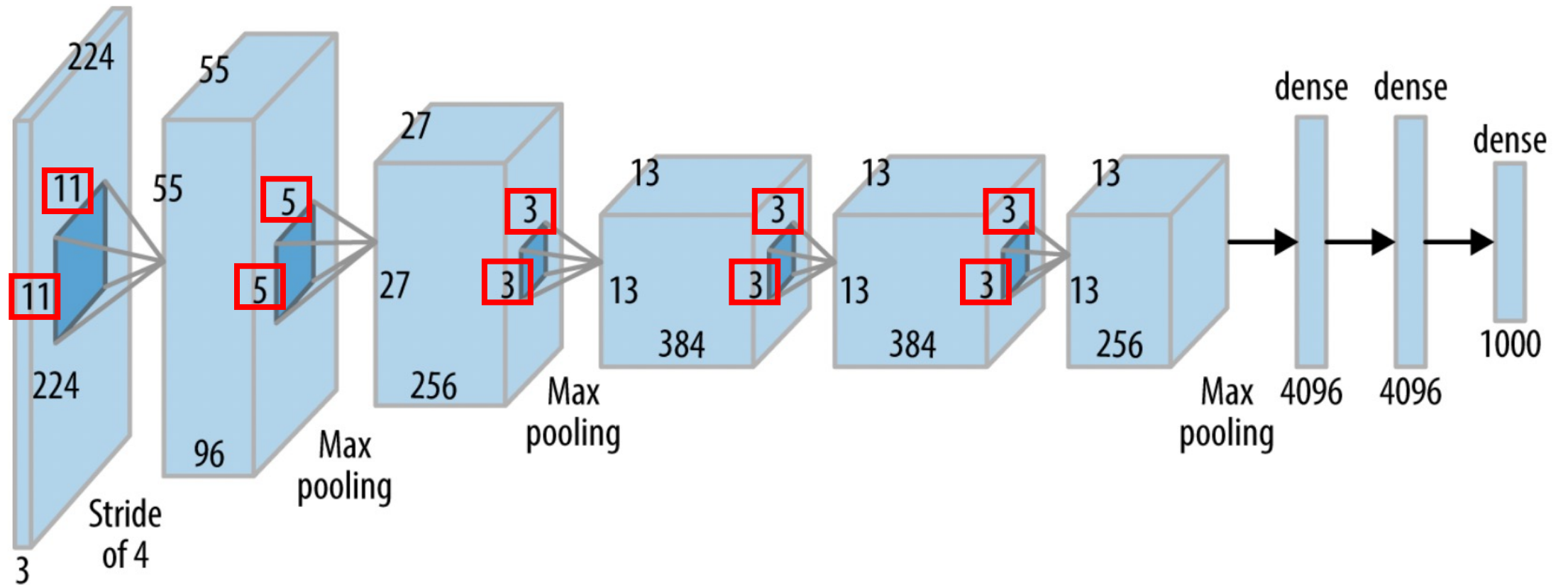


**Fig. 1.** Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

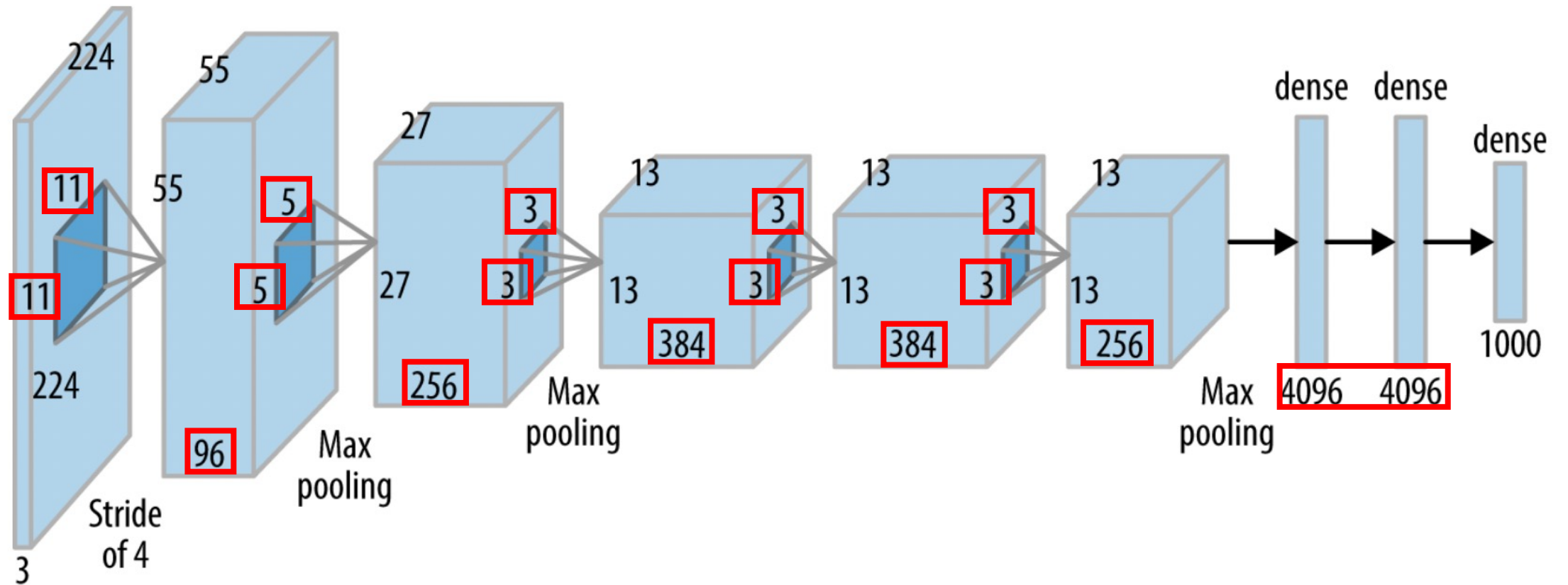
# AlexNet



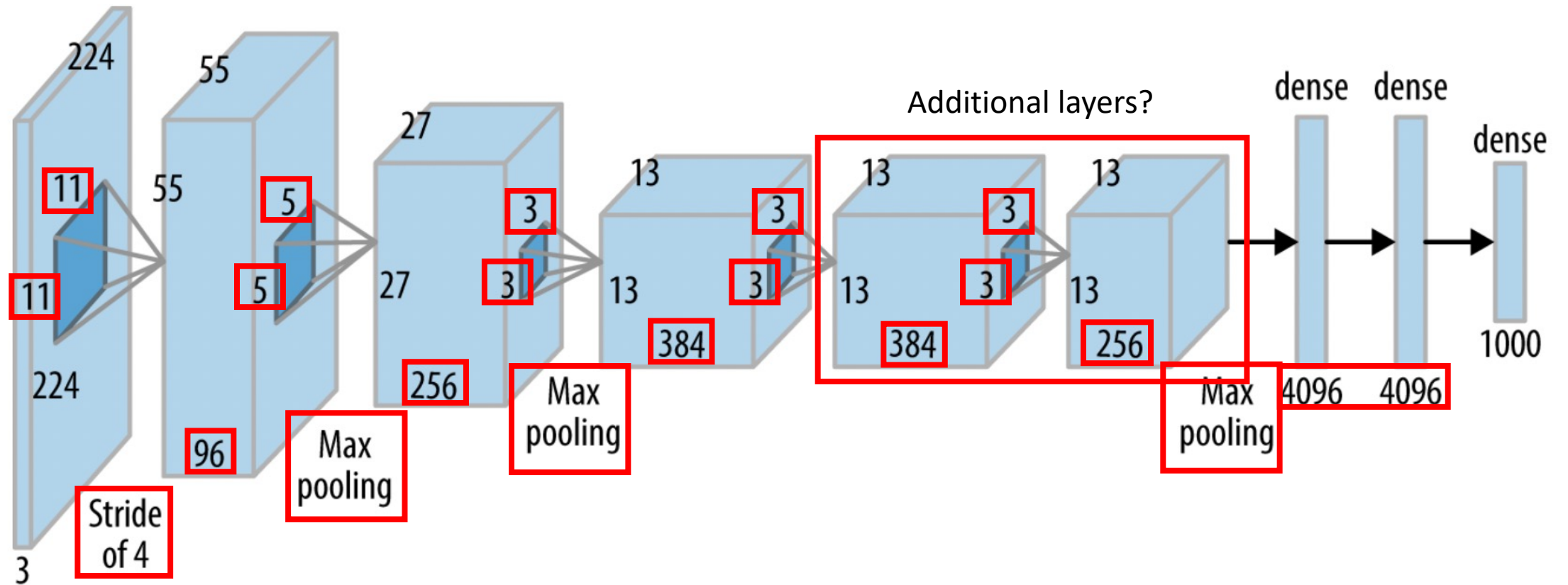
# AlexNet



# AlexNet

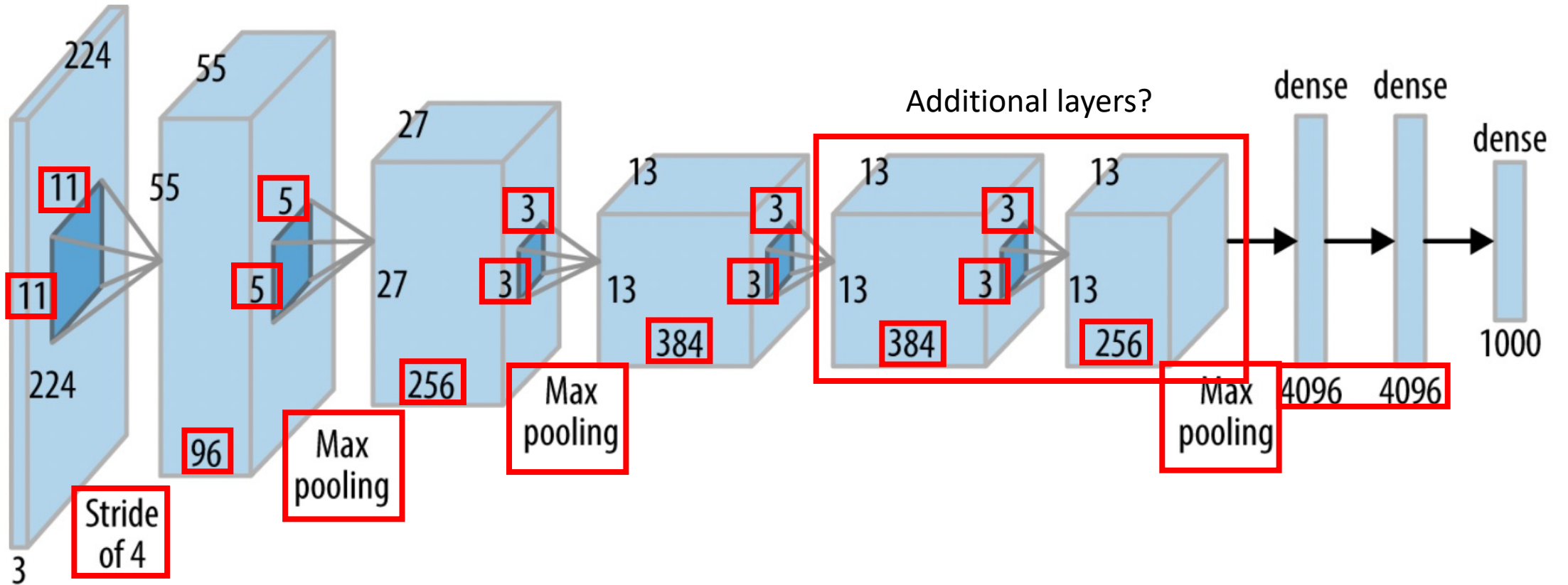


# AlexNet



# AlexNet

Q: what about other **settings**?





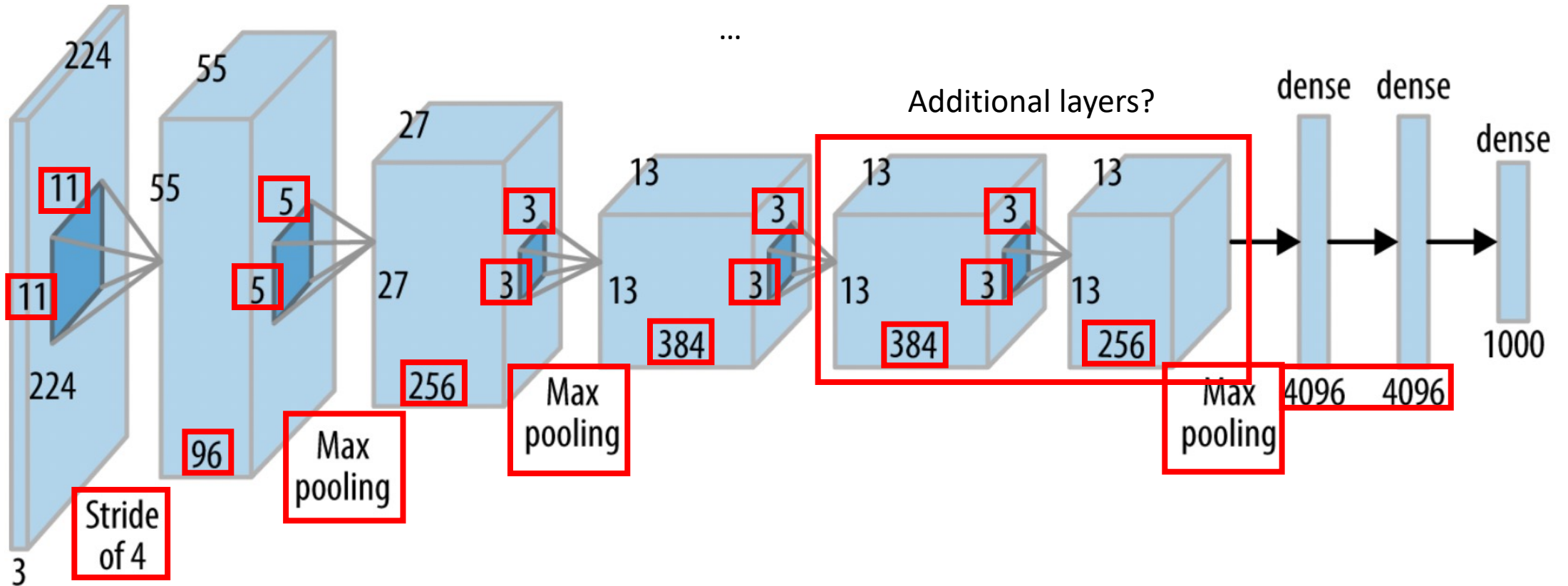
# AlexNet

Q: what about other settings?

## Size of filters

## # of filters

...



# AlexNet

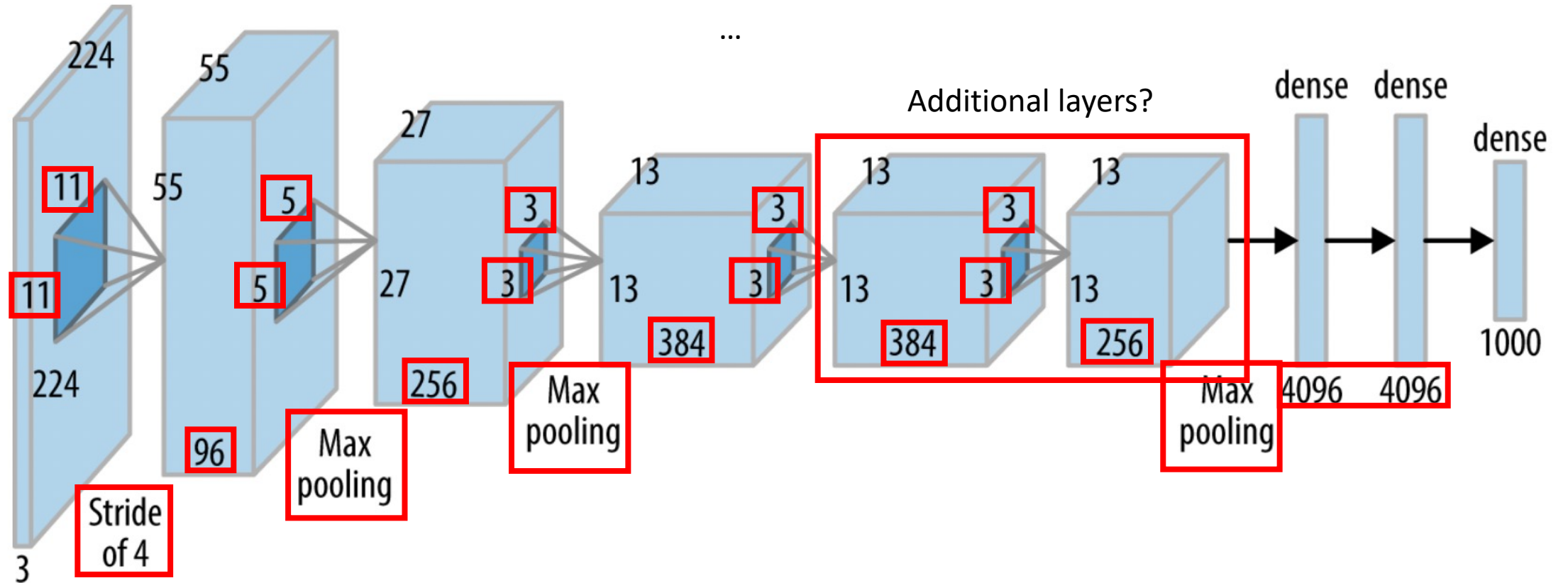
Q: what about other **settings**?

Size of filters

# of filters

...

Neural architecture search  
(NAS)



# A framework for NAS

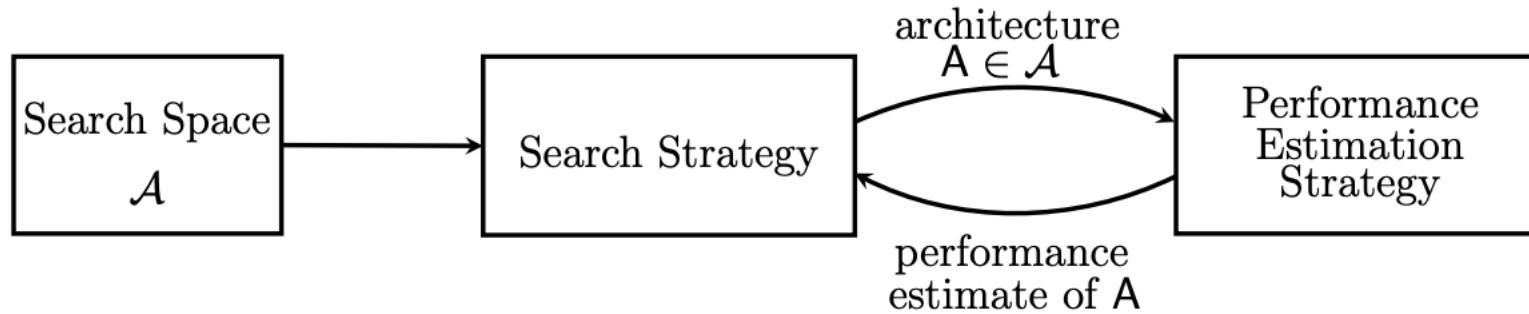
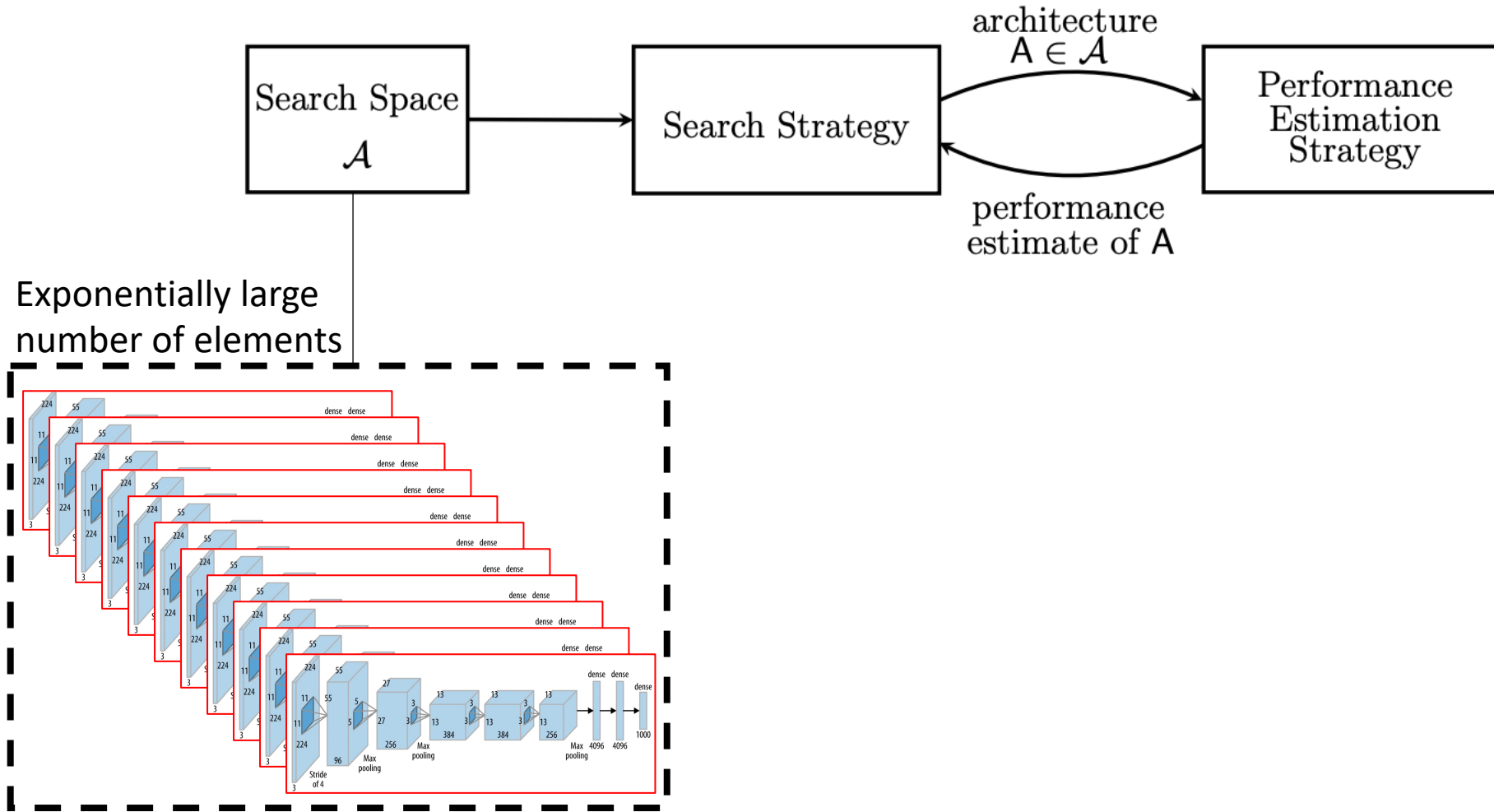


Image credit:

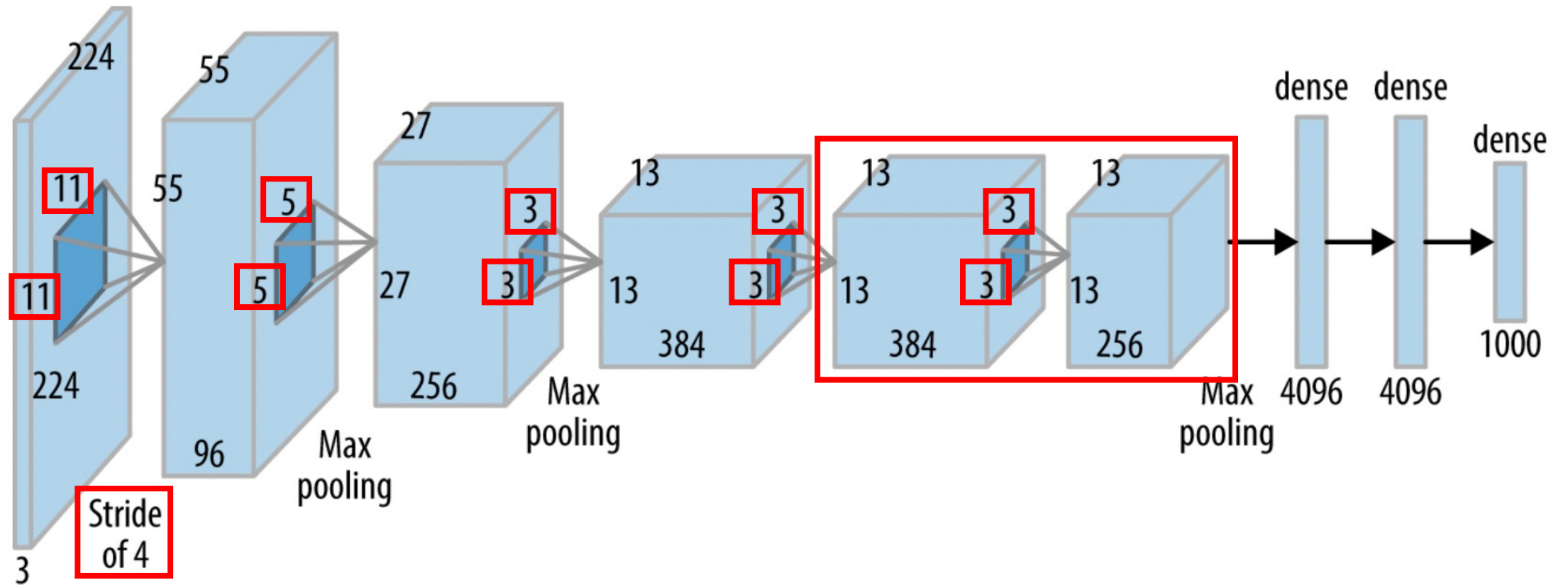
Elsken, Thomas, Jan Hendrik Metzen, and Frank Hutter. "Neural architecture search: A survey." *J. Mach. Learn. Res.* 20, no. 55 (2019): 1-21.

<https://arxiv.org/pdf/1808.05377.pdf>

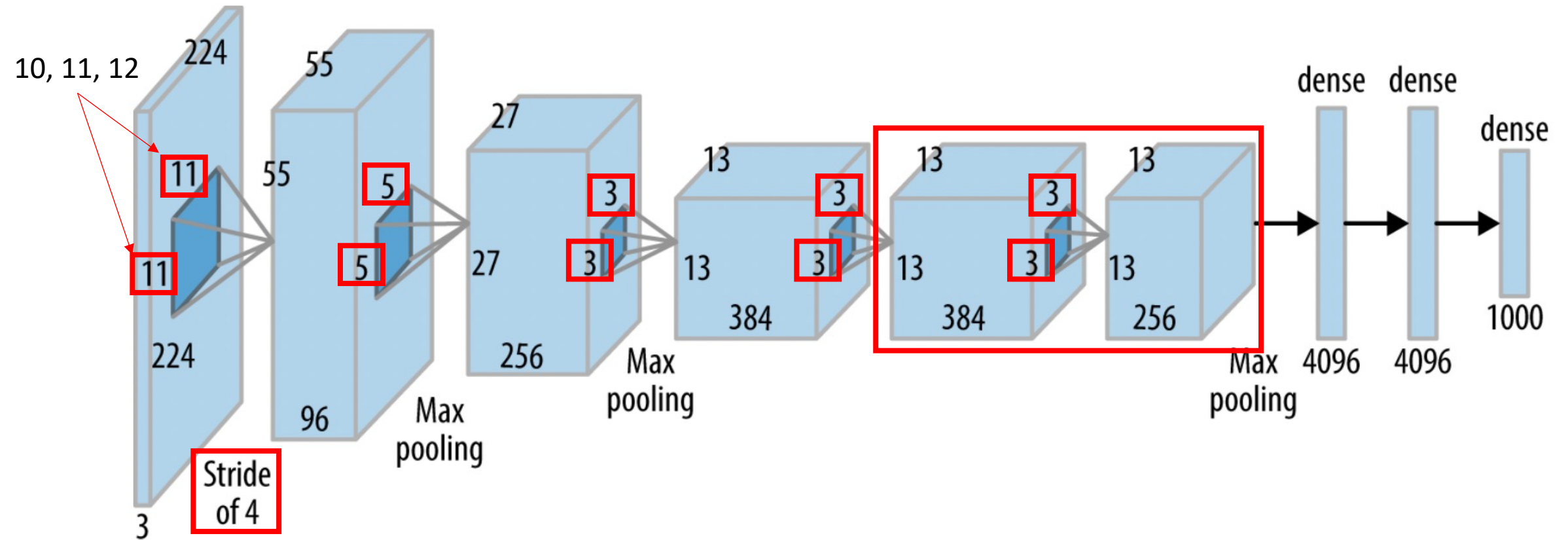
# A framework for NAS



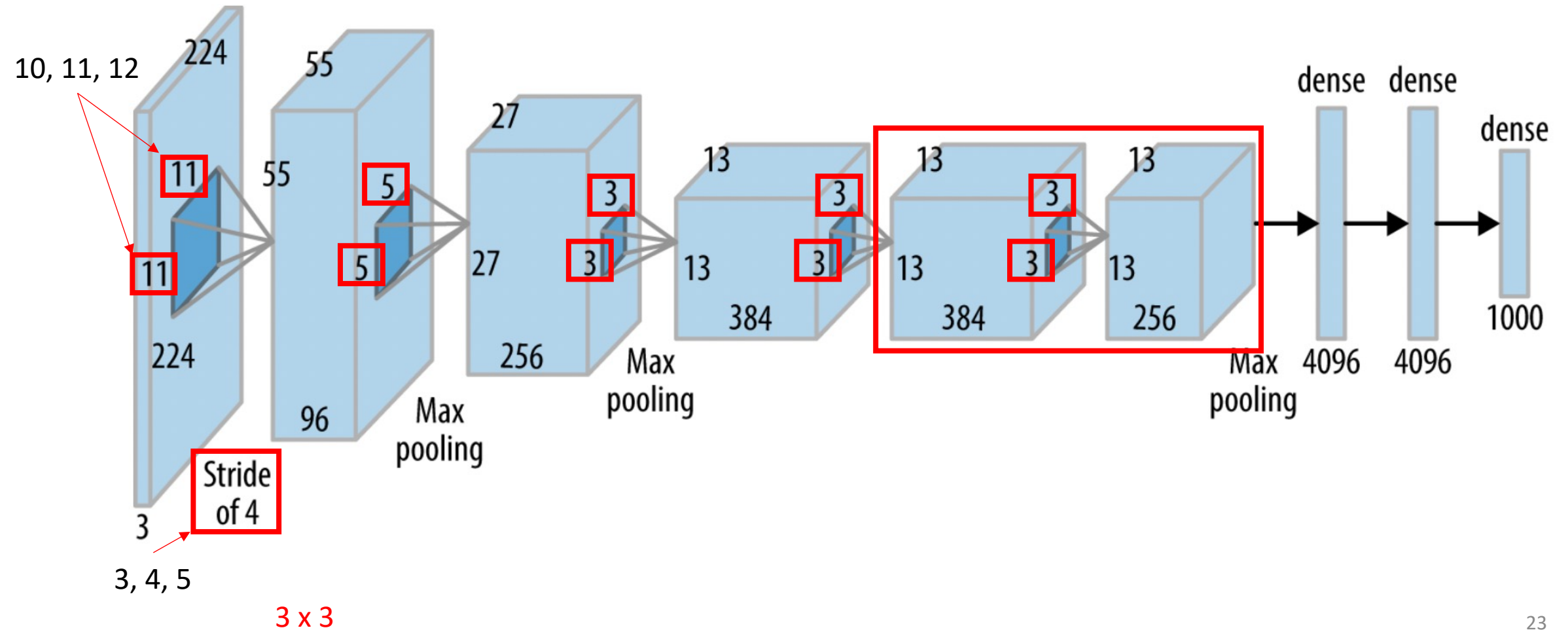
# AlexNet



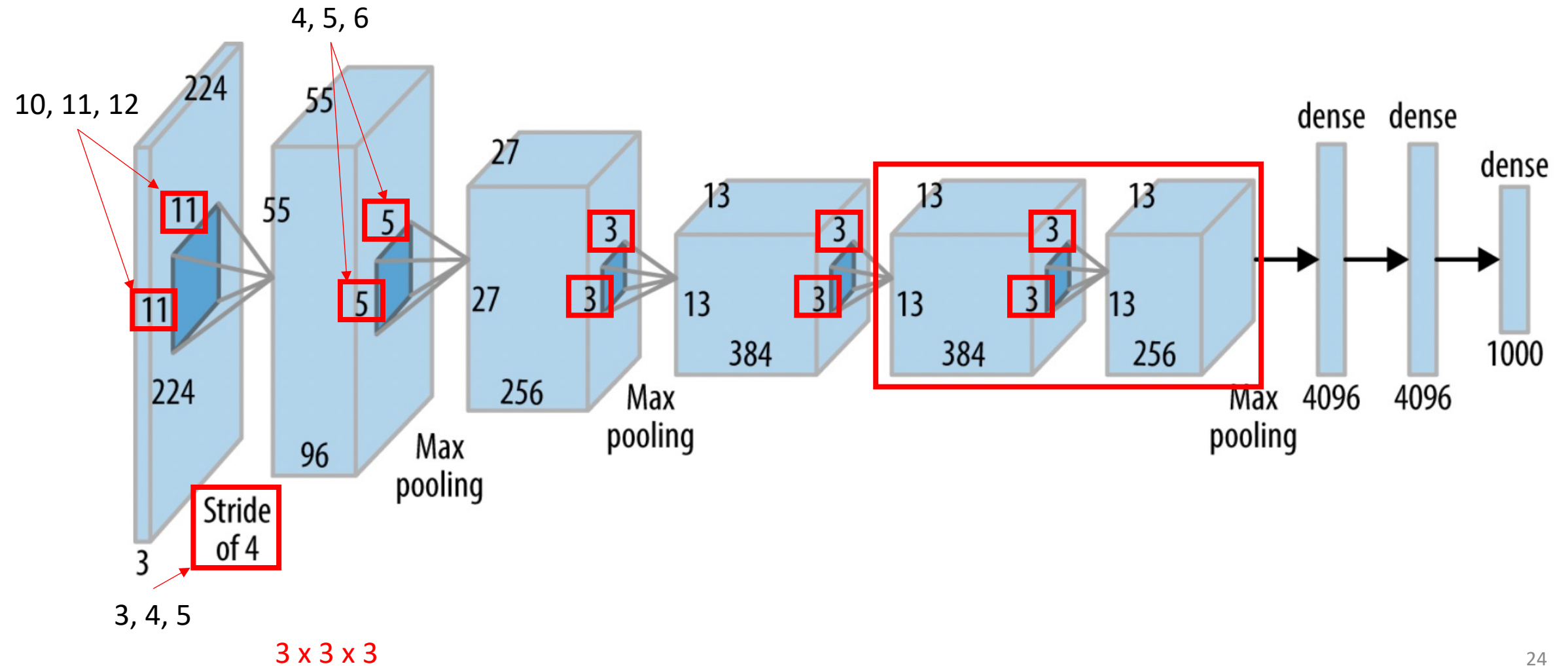
# AlexNet



# AlexNet

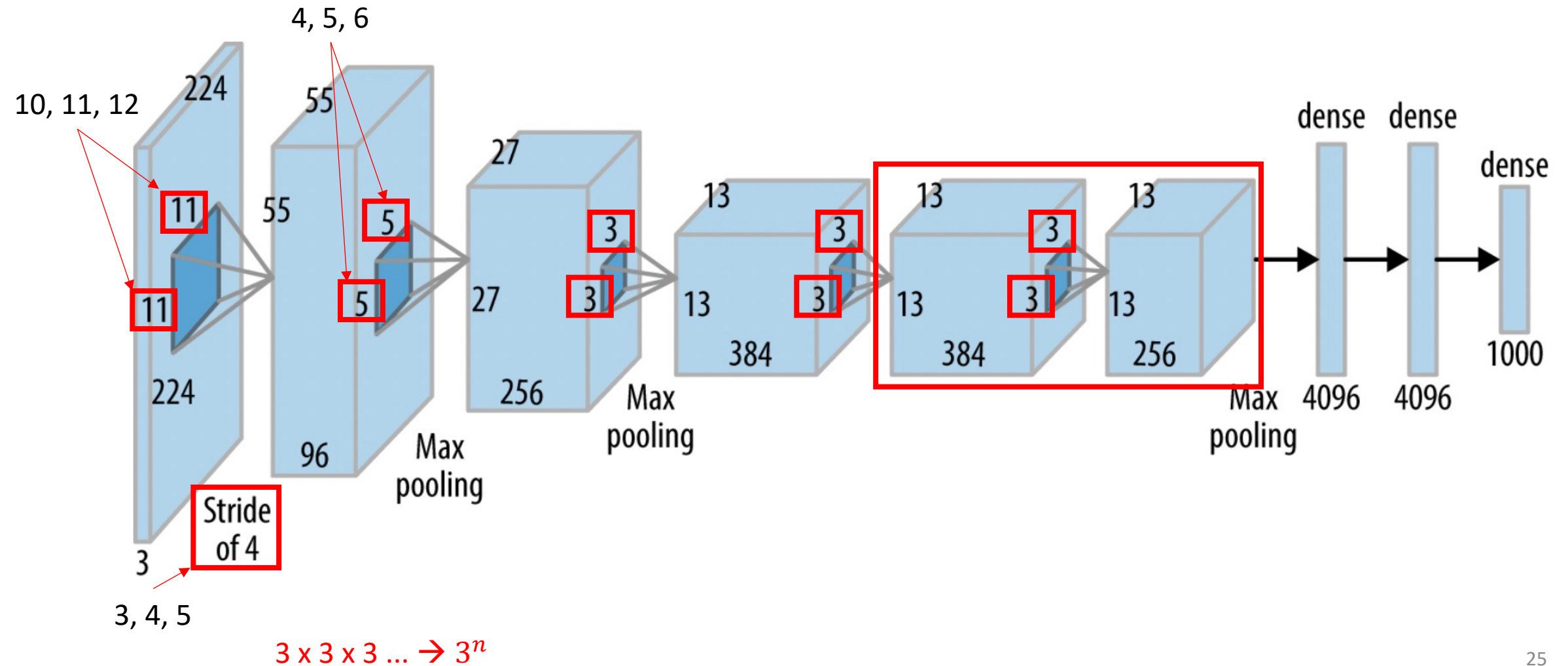


# AlexNet

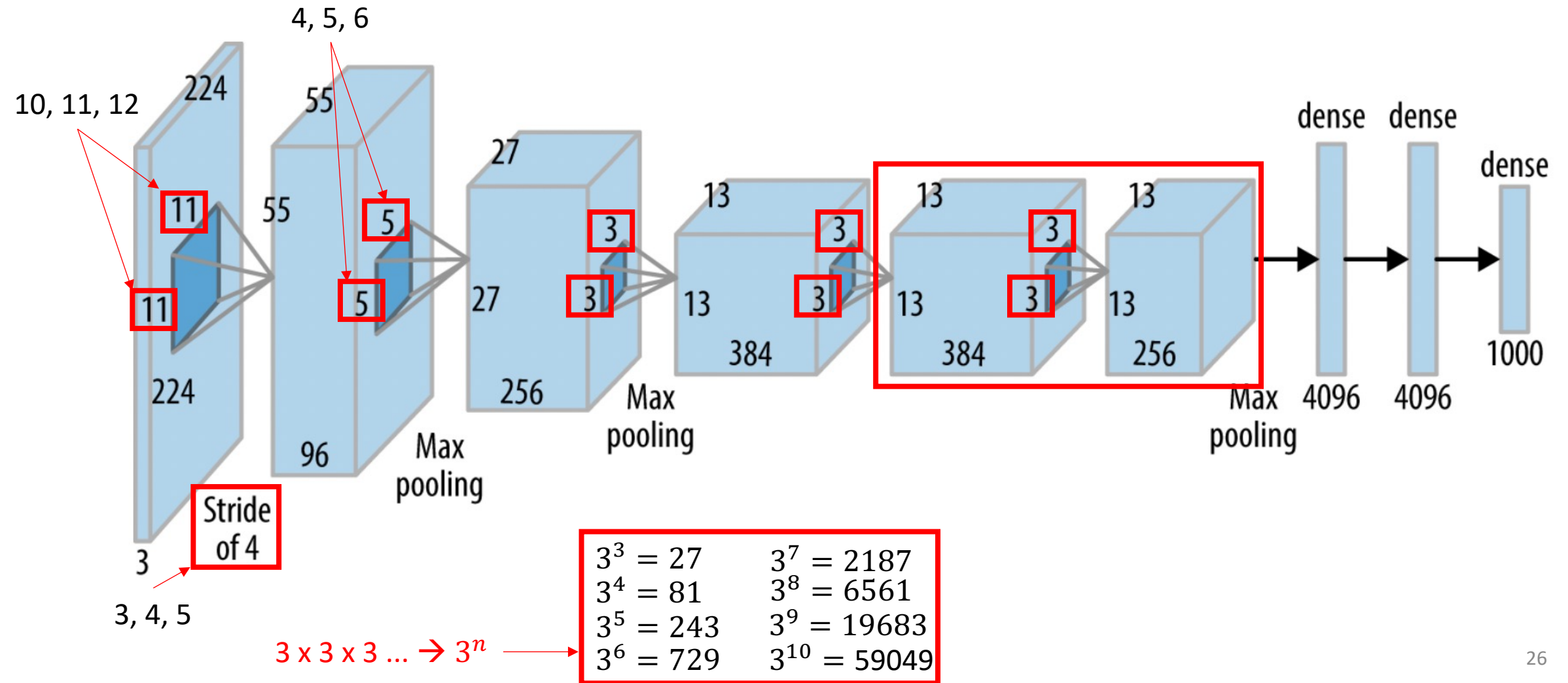




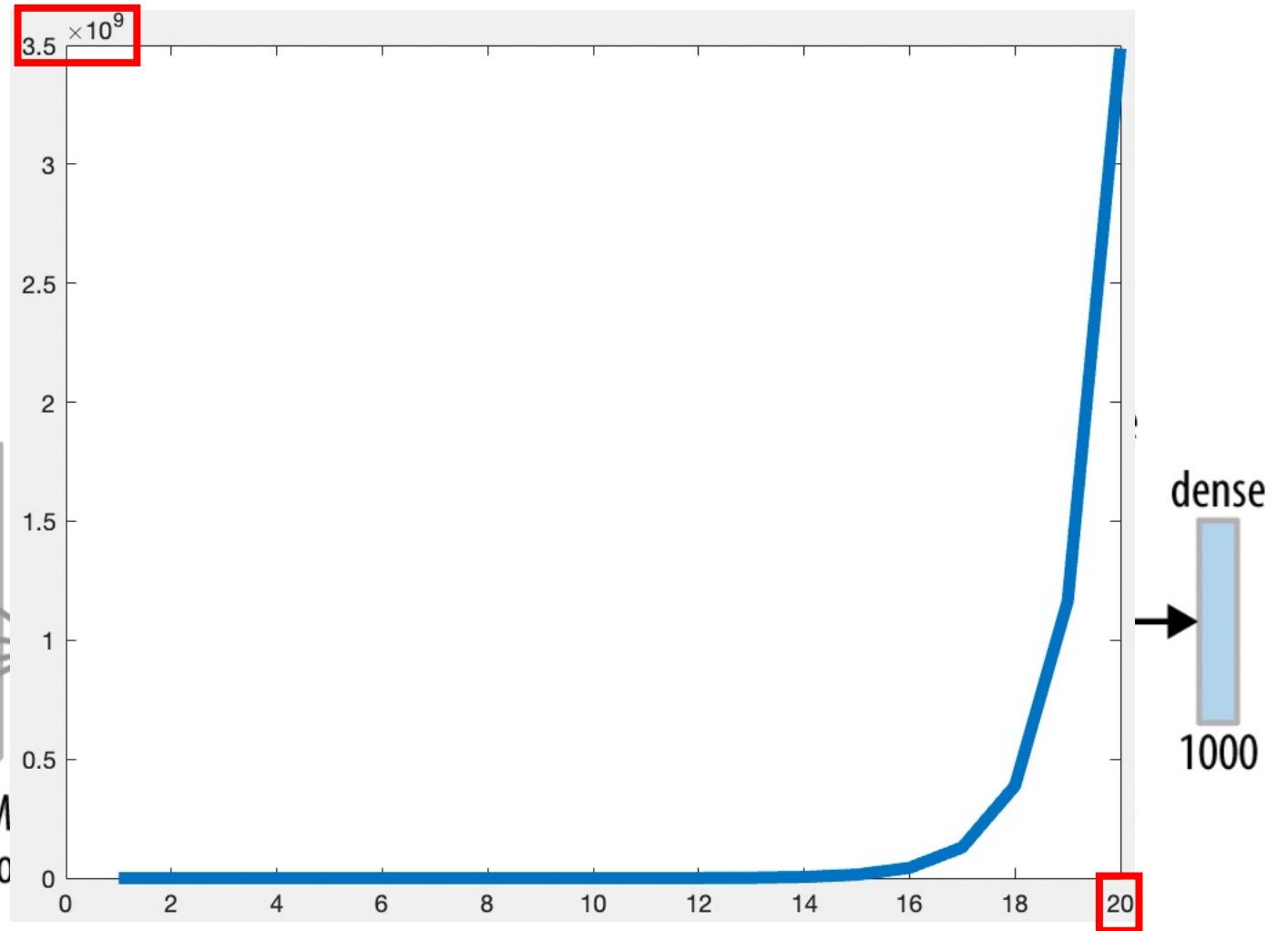
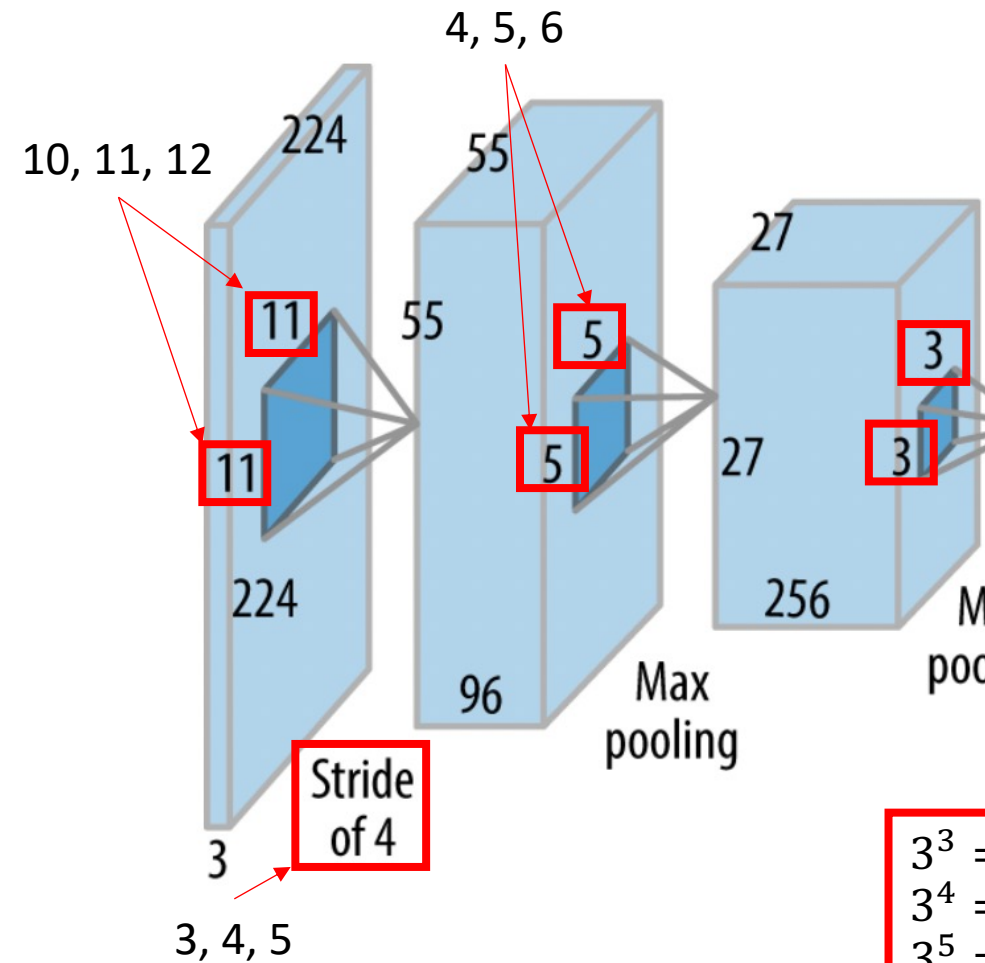
# AlexNet



# AlexNet



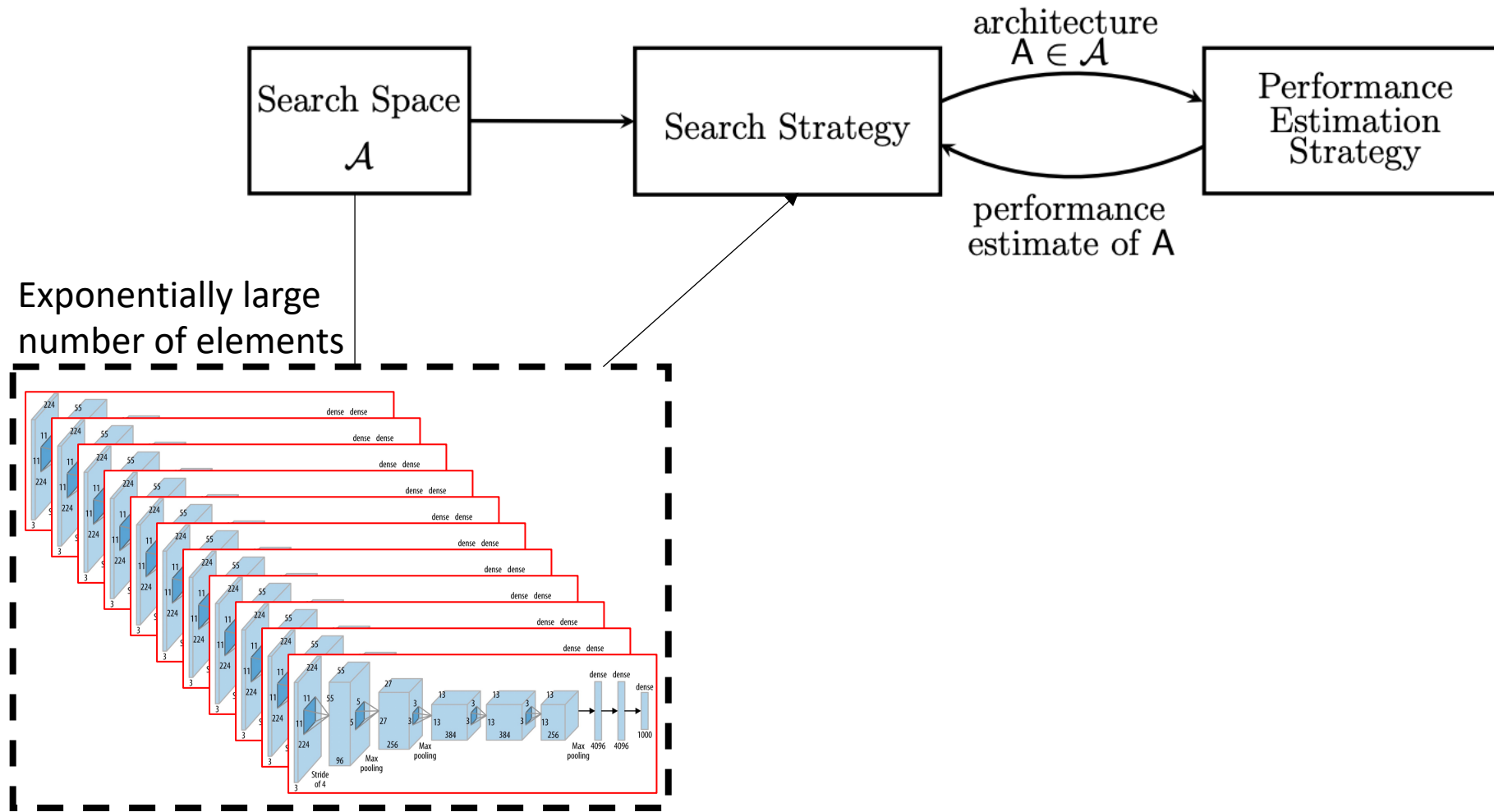
# AlexNet



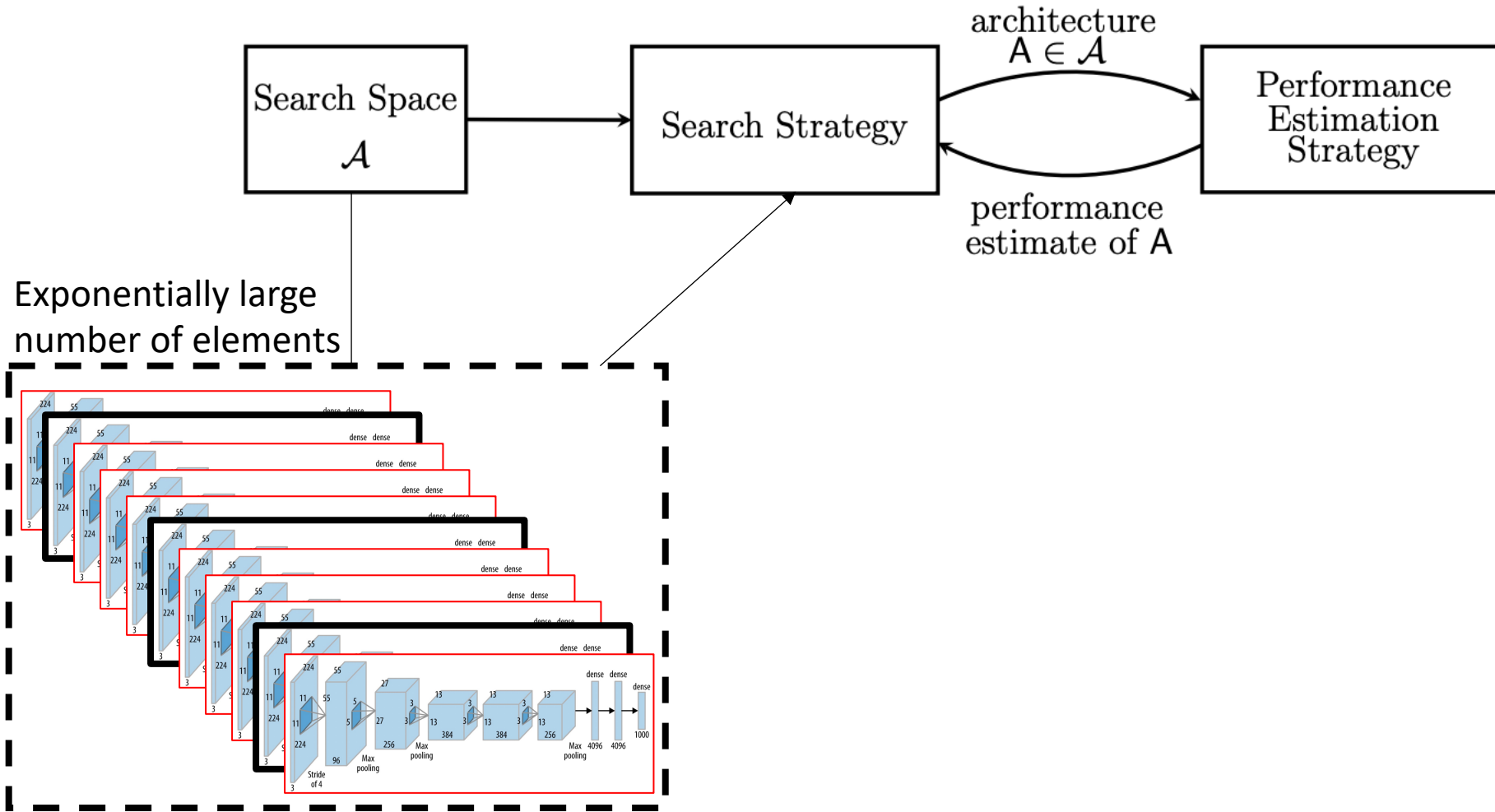
$3 \times 3 \times 3 \dots \rightarrow 3^n$

$3^3 = 27$	$3^7 = 2187$
$3^4 = 81$	$3^8 = 6561$
$3^5 = 243$	$3^9 = 19683$
$3^6 = 729$	$3^{10} = 59049$

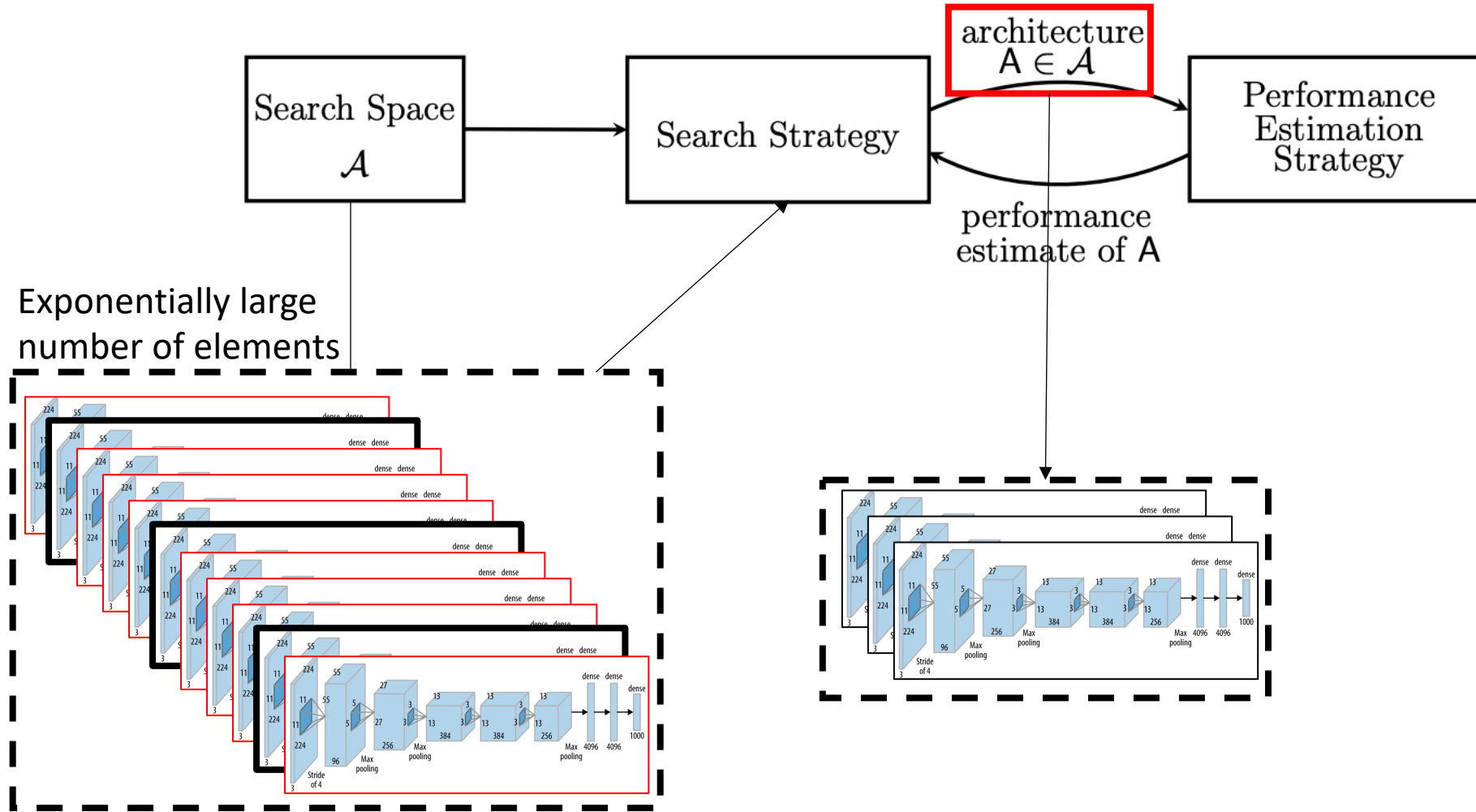
# A framework for NAS



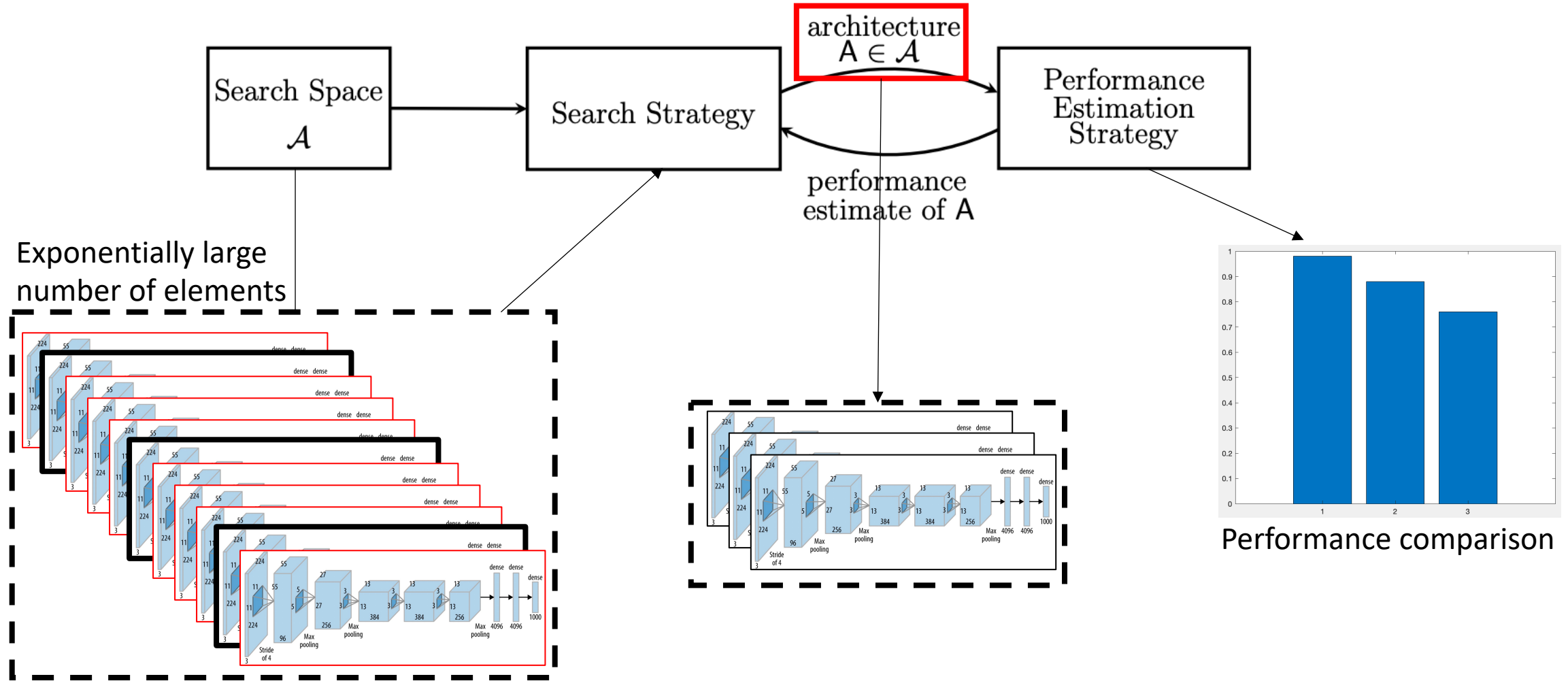
# A framework for NAS



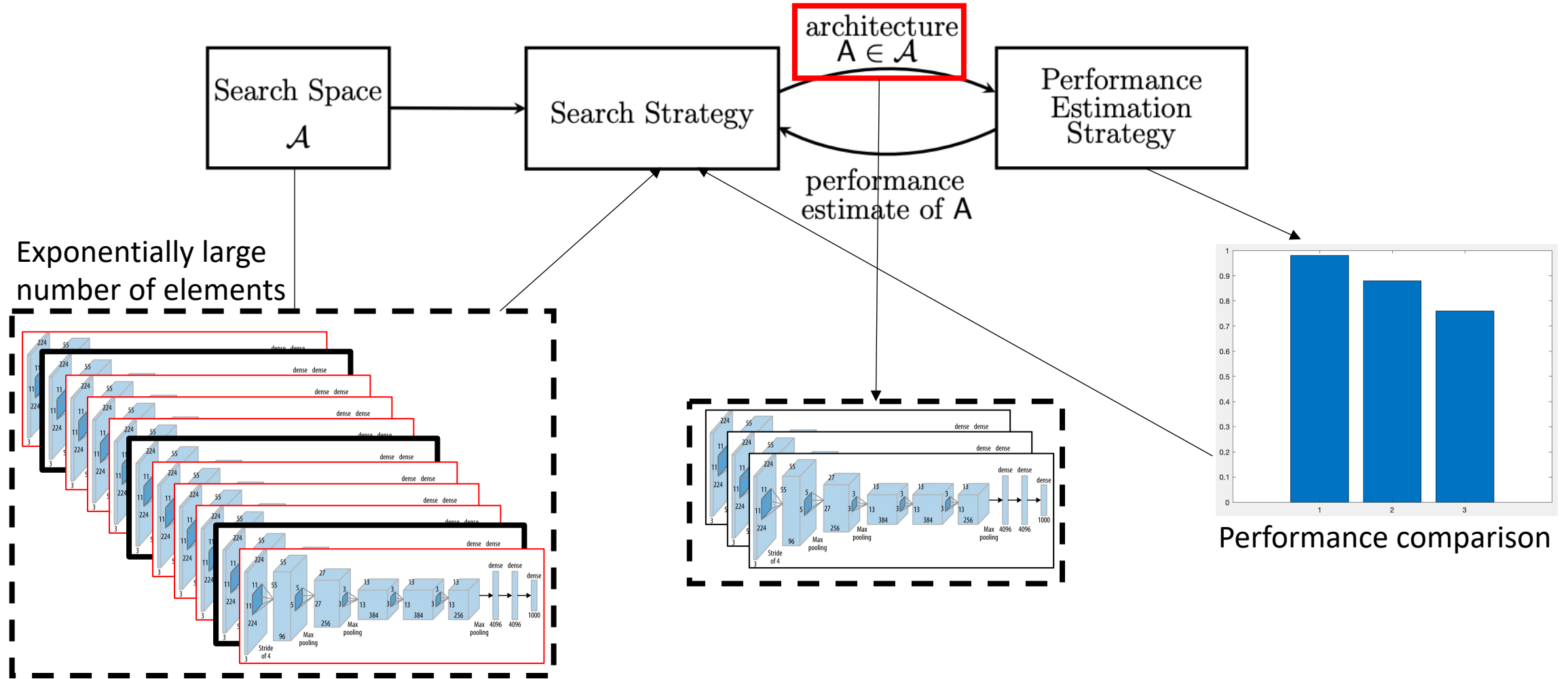
# A framework for NAS



# A framework for NAS

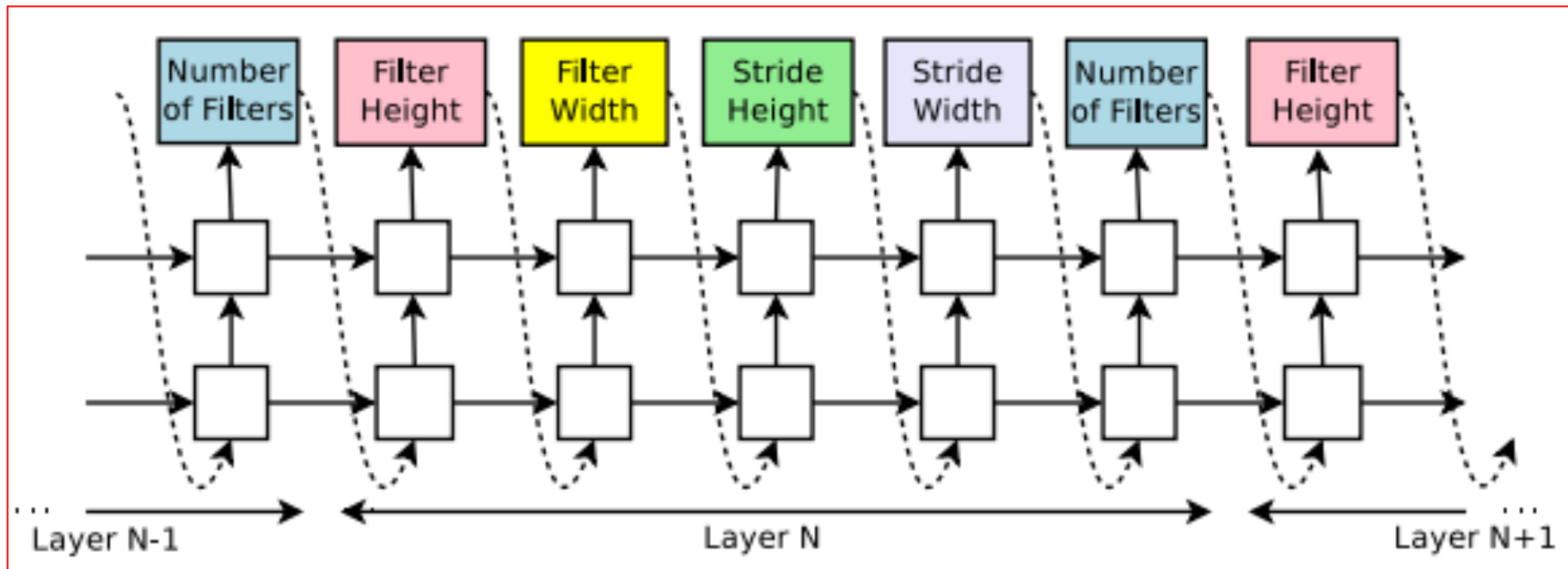


# A framework for NAS



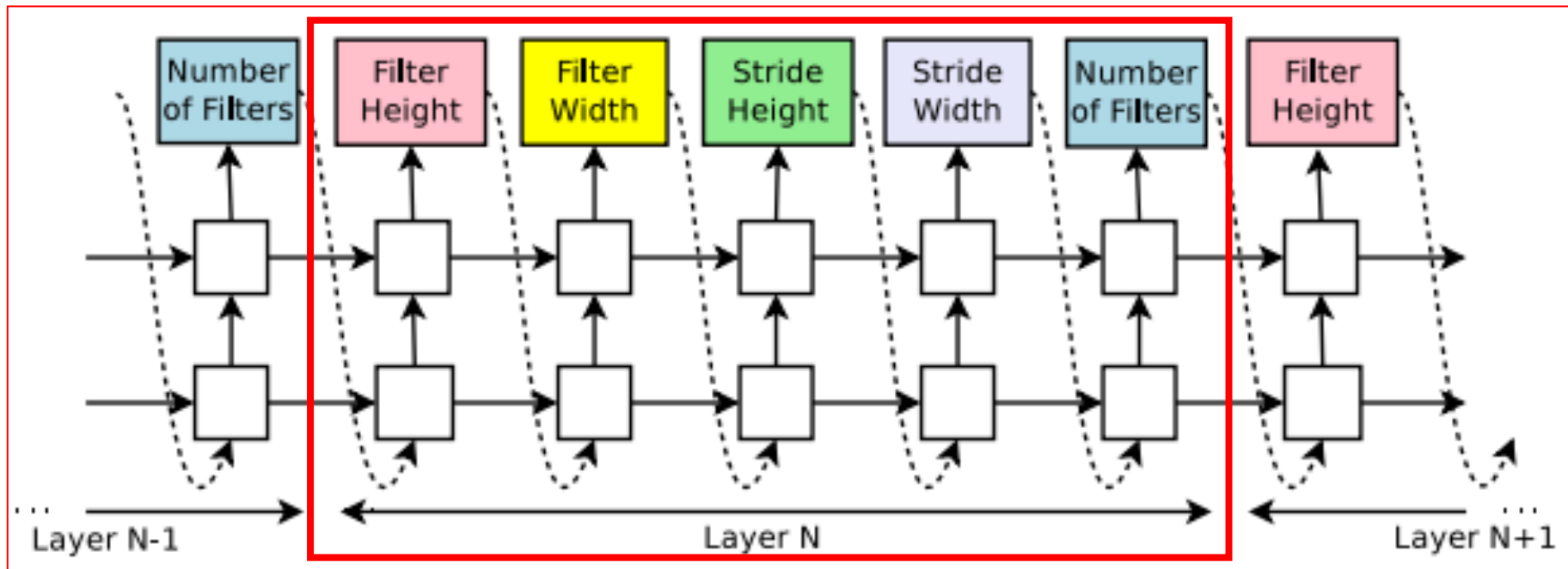


# Defining search space



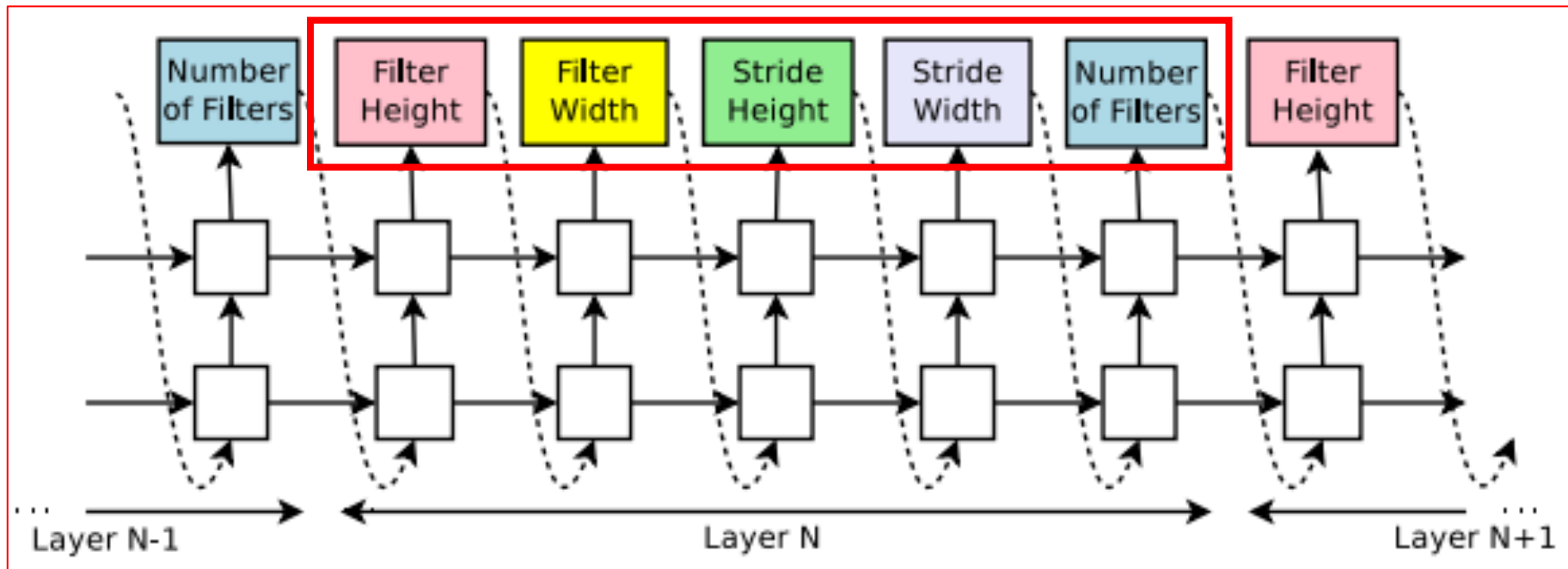
A RNN network

# Defining search space



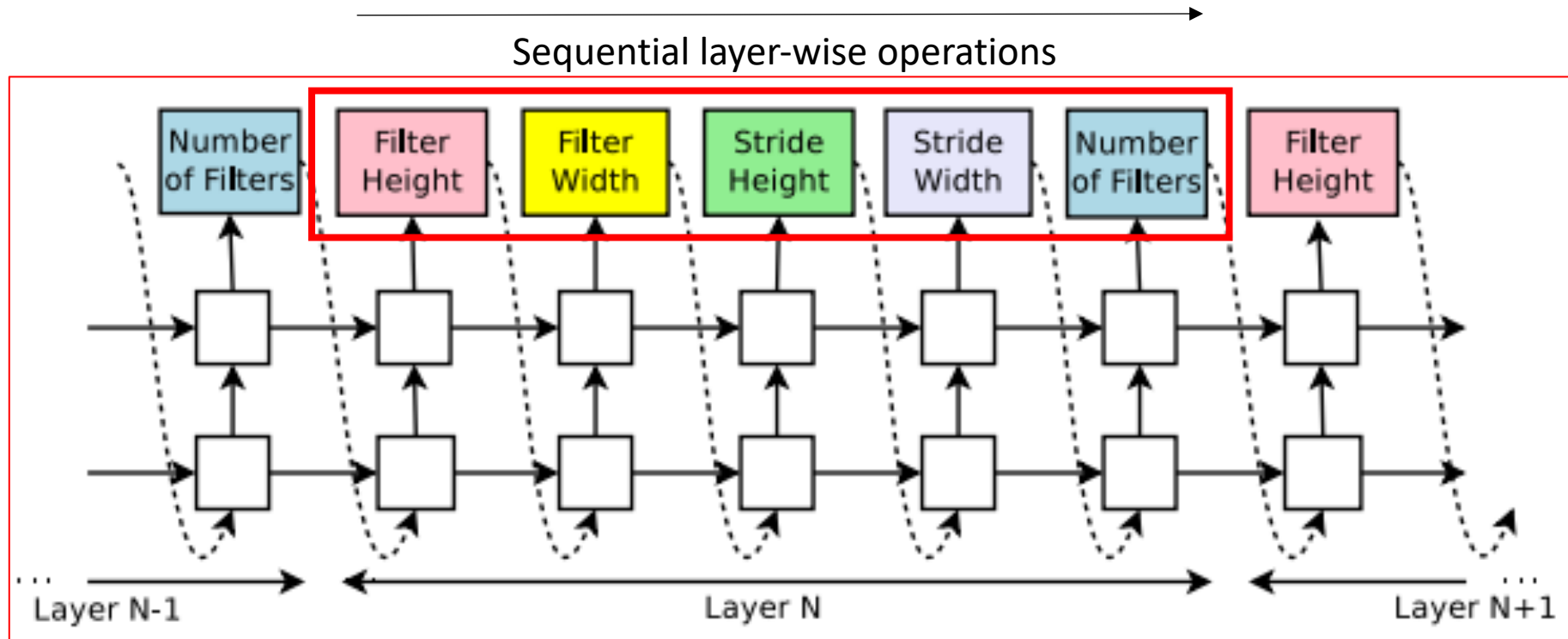
A RNN network

# Defining search space



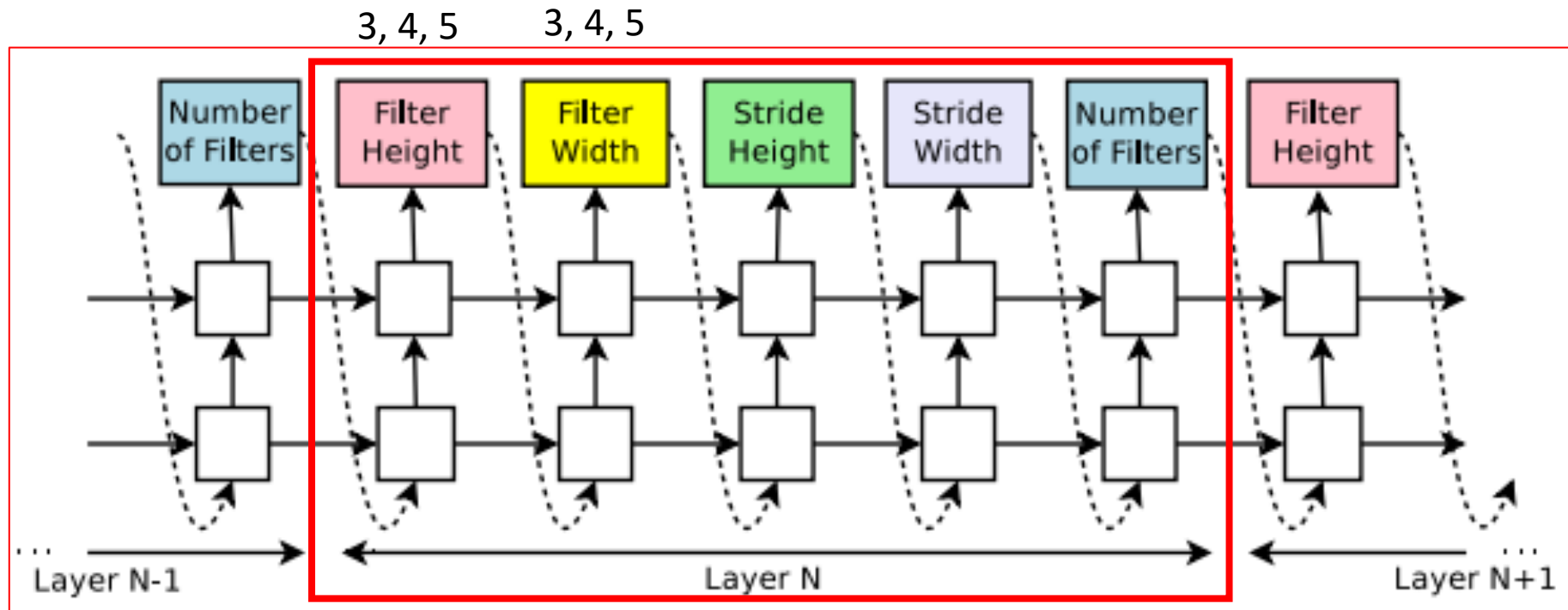
A RNN network

# Defining search space



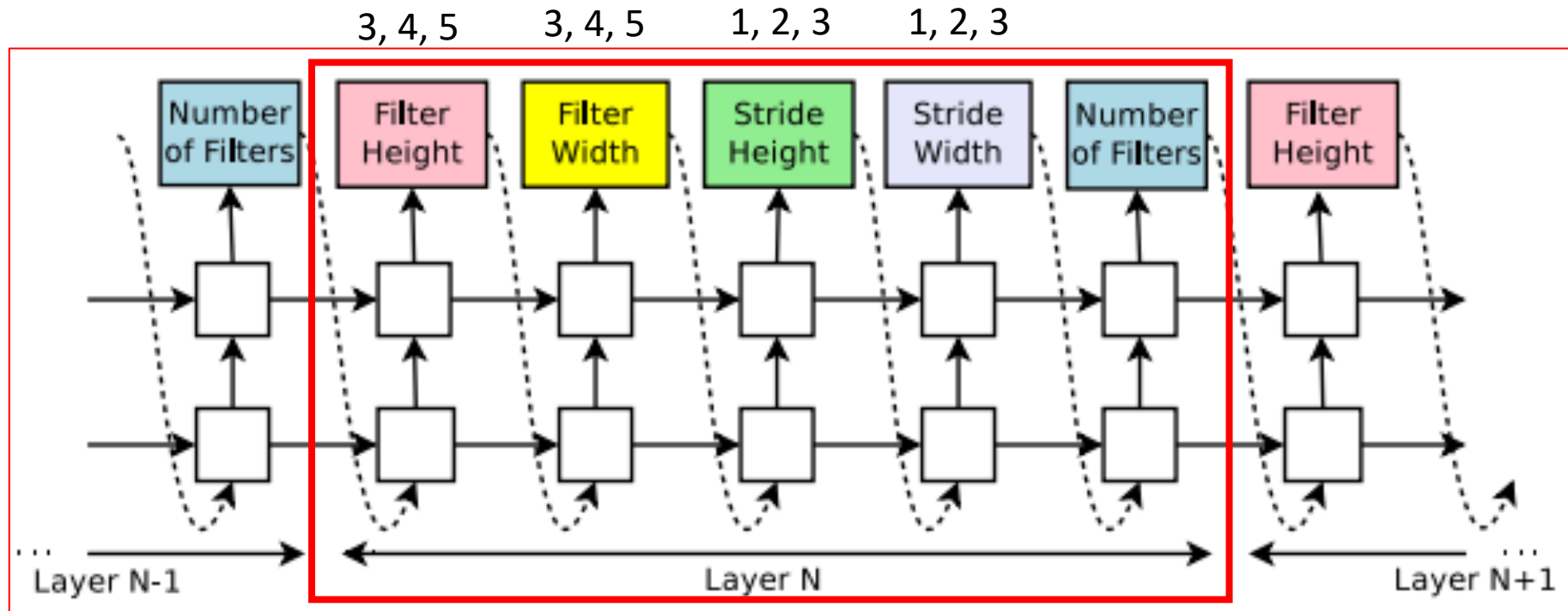
A RNN network

# Defining search space



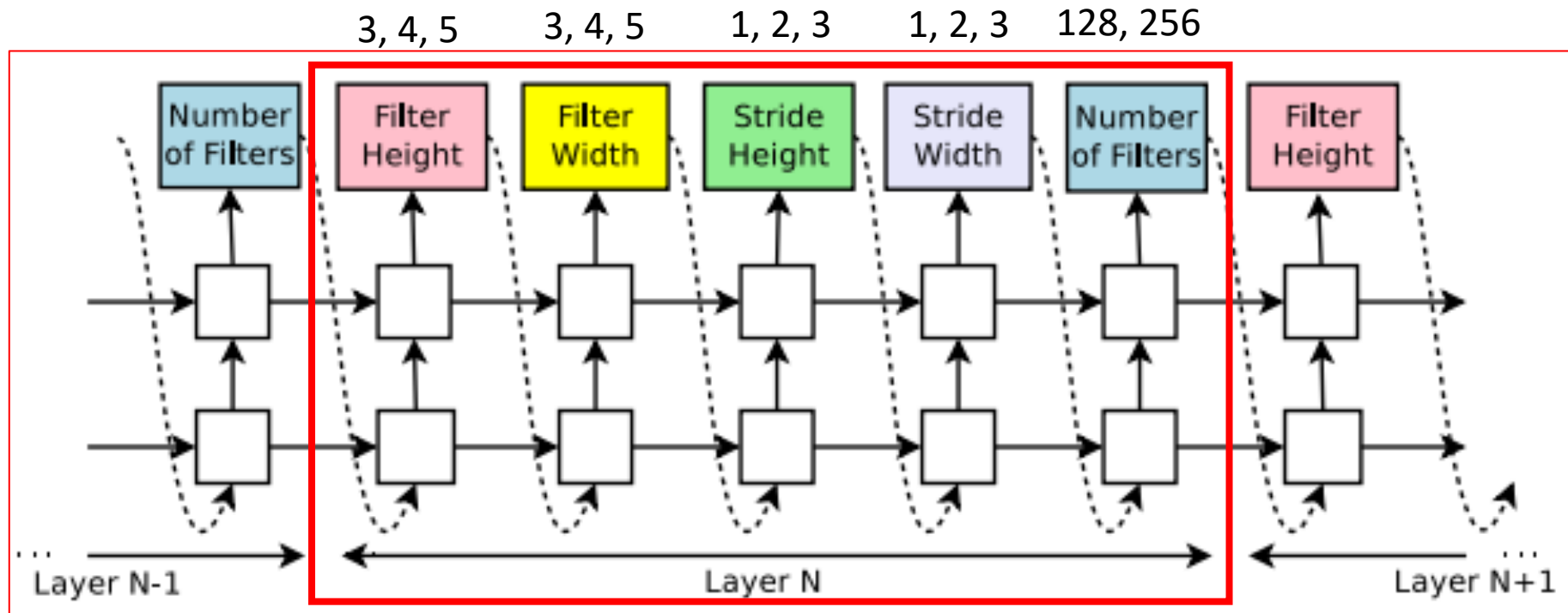
A RNN network

# Defining search space



A RNN network

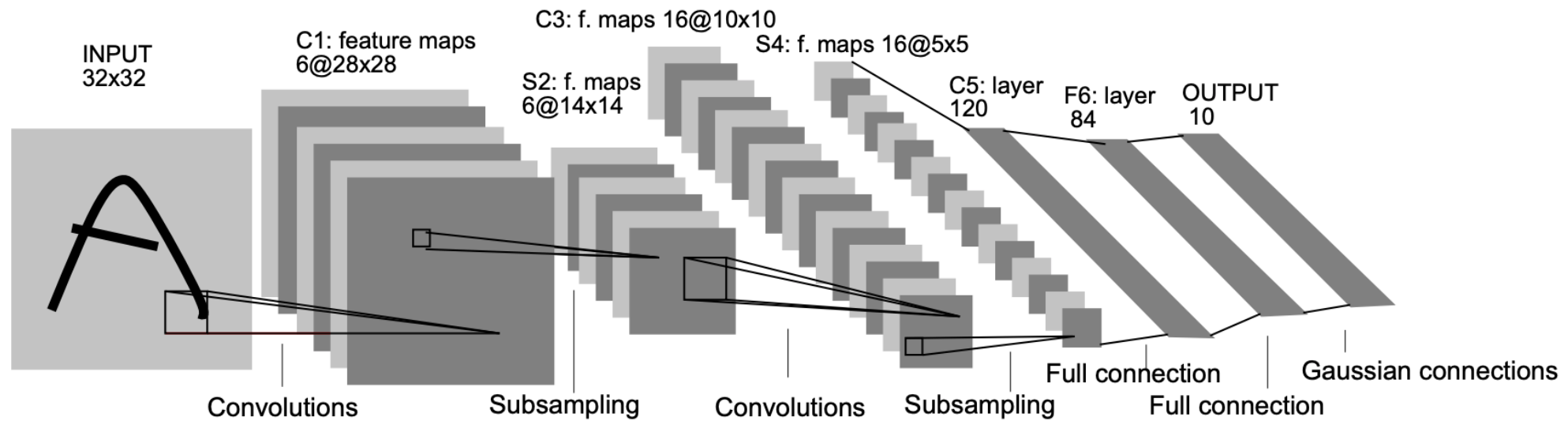
# Defining search space



A RNN network

# LeNet-5

Q: The definition of sequential layer-wise operations works for LeNet-like networks?

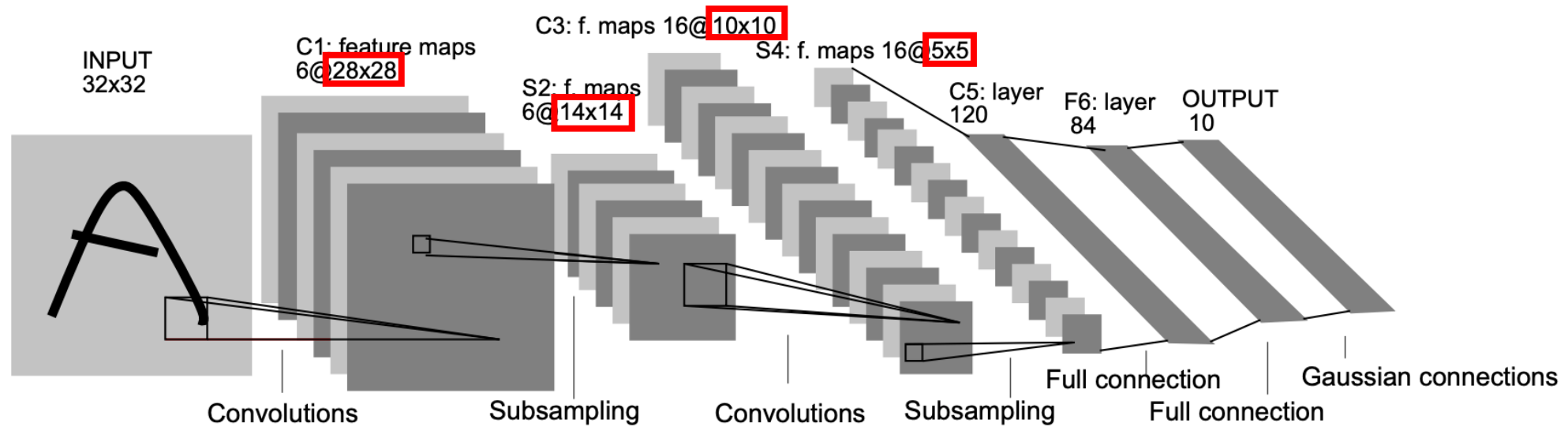


**Fig. 1.** Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.



# LeNet-5

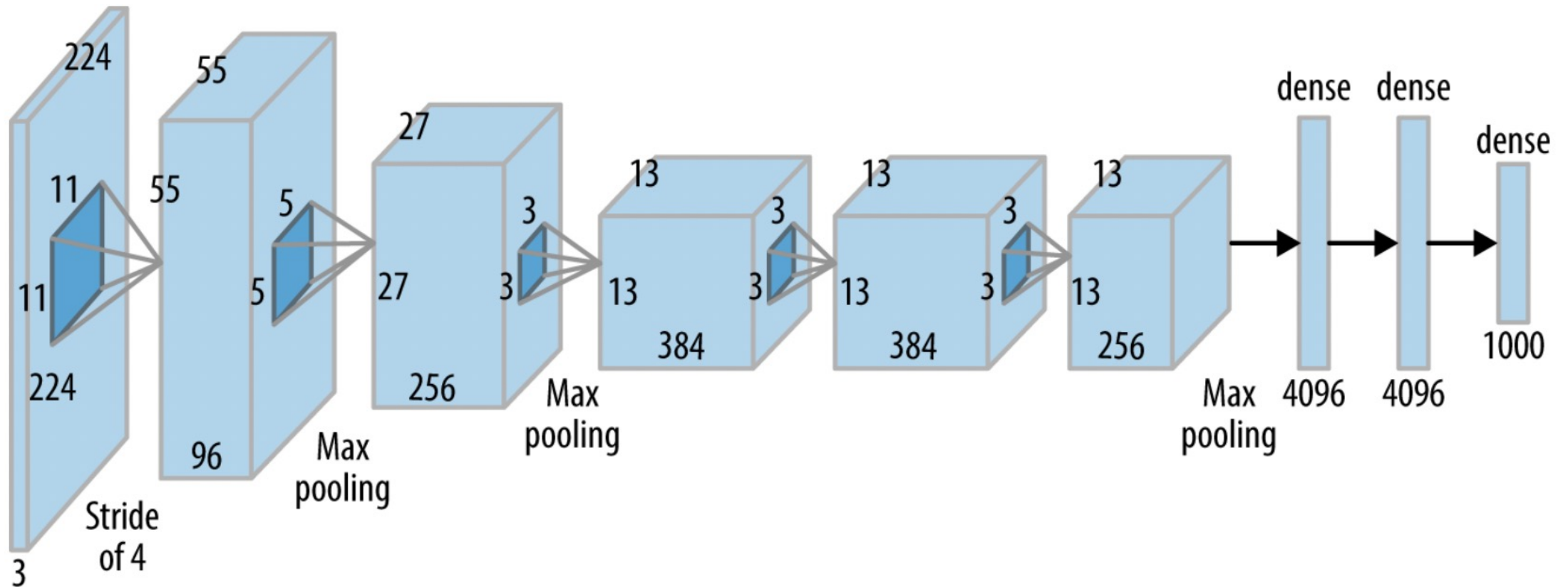
Q: The definition of sequential layer-wise operations works for LeNet-like networks?



**Fig. 1.** Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

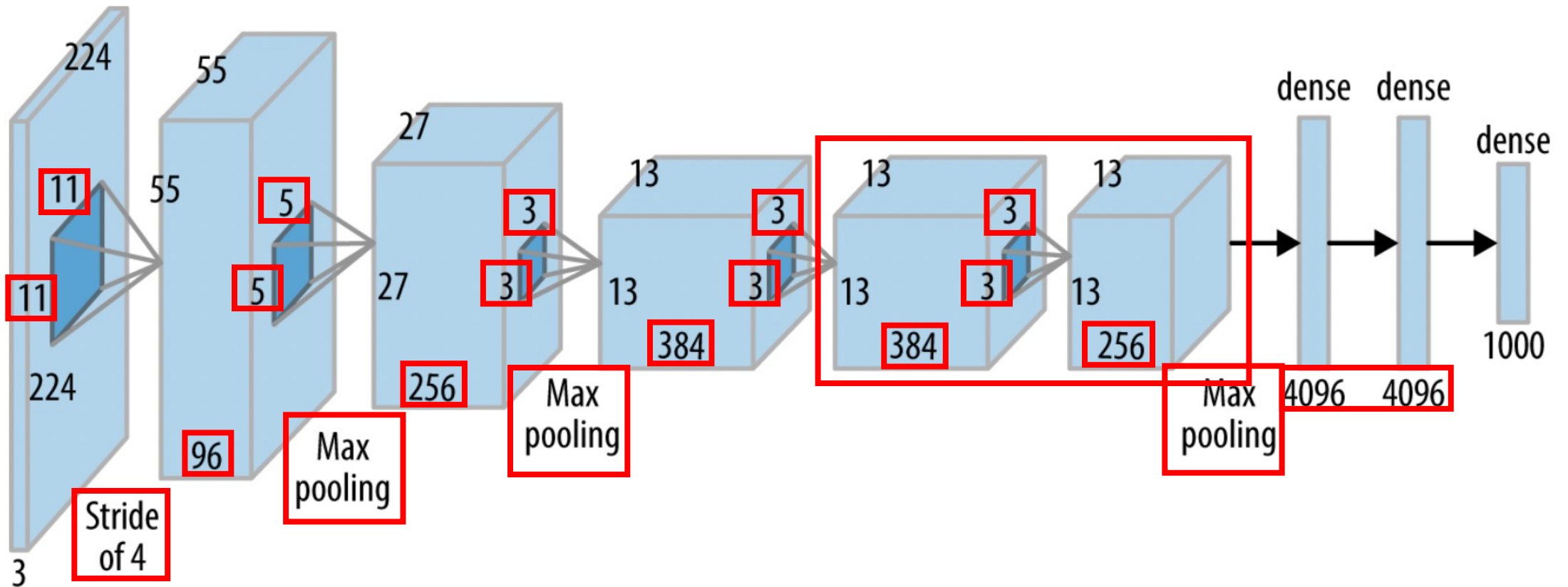
# AlexNet

Q: The definition of sequential layer-wise operations works for AlexNet-like networks?

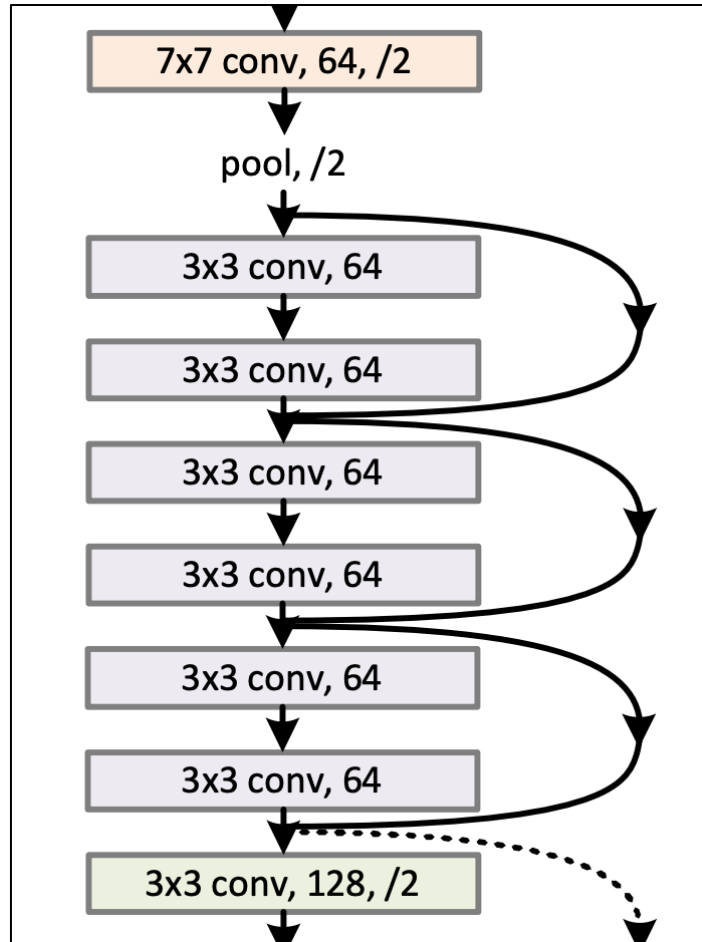


# AlexNet

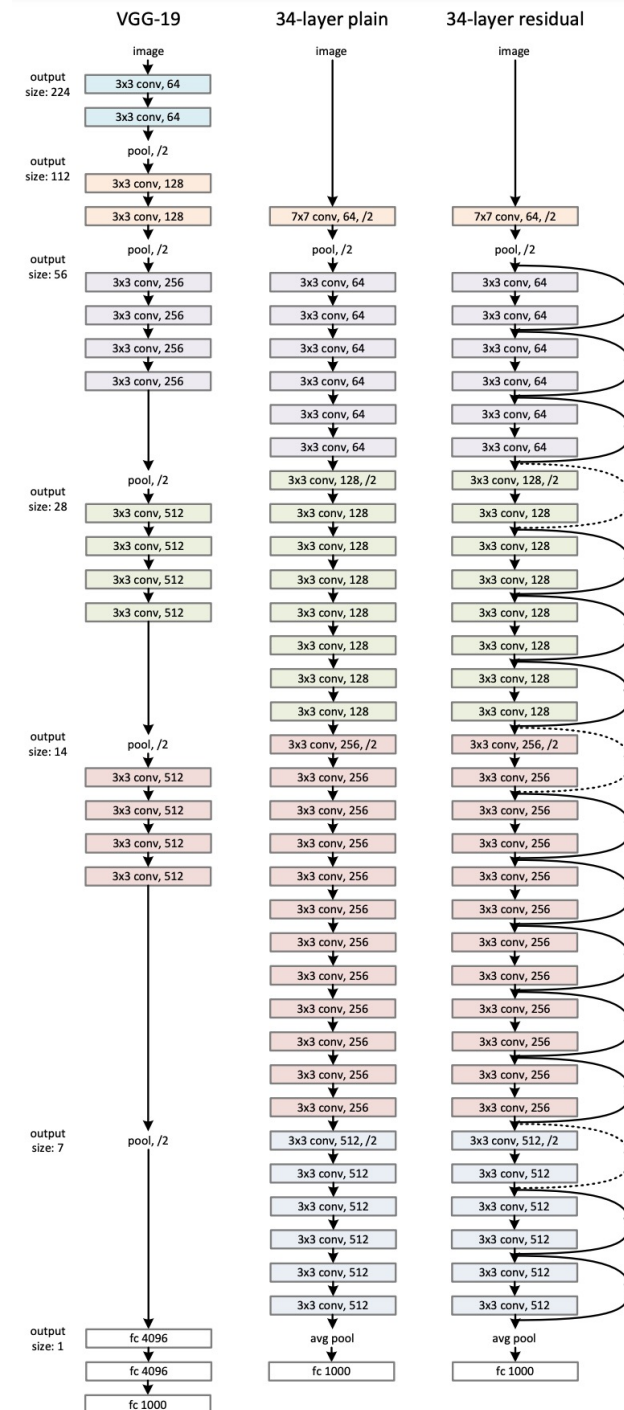
Q: The definition of sequential layer-wise operations works for AlexNet-like networks?



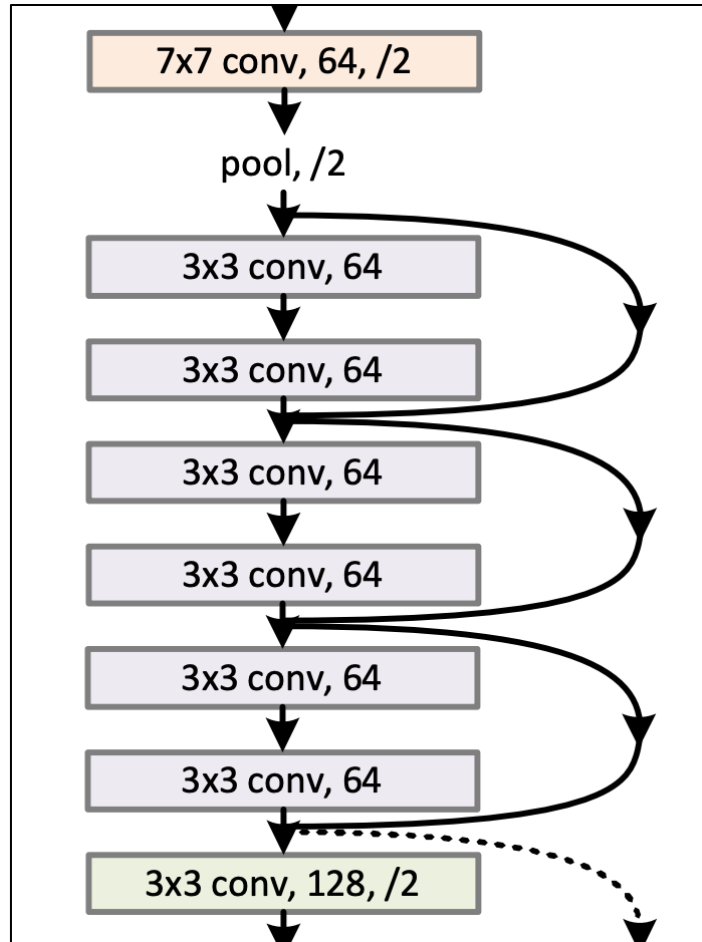
# ResNet



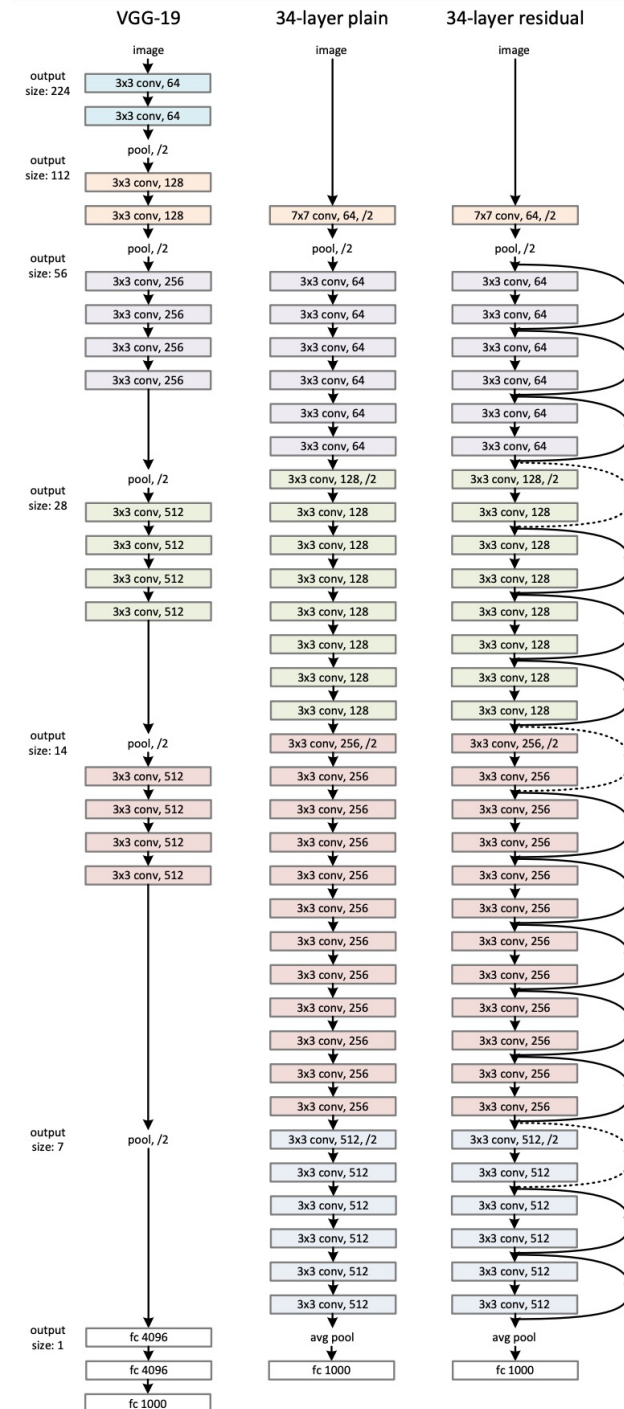
**Q:** The definition of sequential layer-wise operations works for ResNet-like networks?



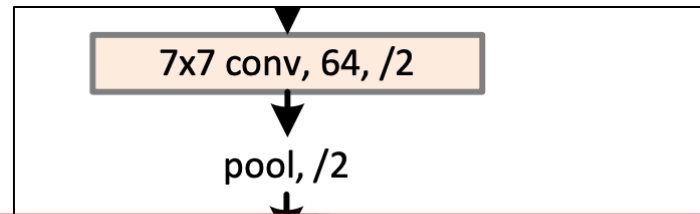
# ResNet



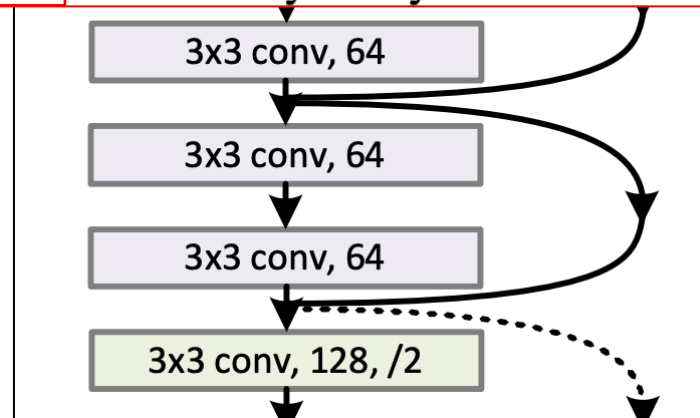
**Q:** The definition of sequential layer-wise operations works for ResNet-like networks? **Yes, but the scalability?**



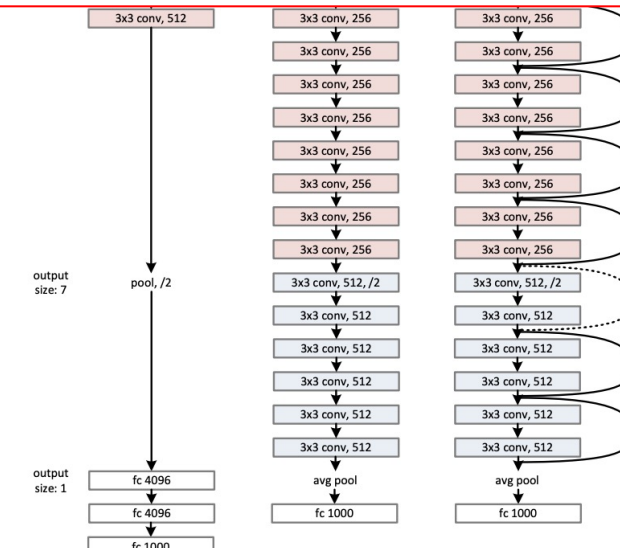
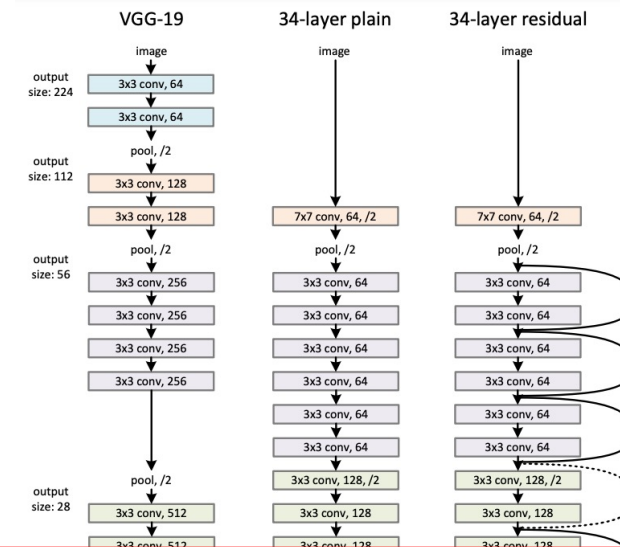
# ResNet



**Training details:** The controller RNN is a two-layer LSTM with 35 hidden units on each layer. It is trained with the ADAM optimizer (Kingma & Ba, 2015) with a learning rate of 0.0006. The weights of the controller are initialized uniformly between -0.08 and 0.08. For the distributed training, we set the number of parameter server shards  $S$  to 20, the number of controller replicas  $K$  to 100 and the number of child replicas  $m$  to 8, which means there are 800 networks being trained on 800 GPUs concurrently at any time.

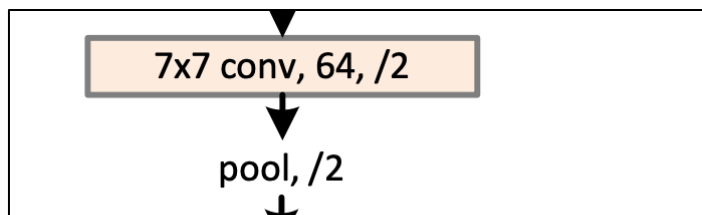


Q: The definition of sequential layer-wise operations works for ResNet-like networks? **Yes, but the scalability?**



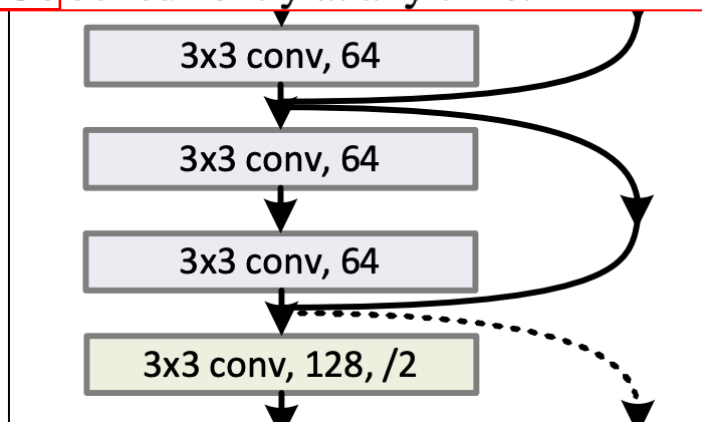


# ResNet

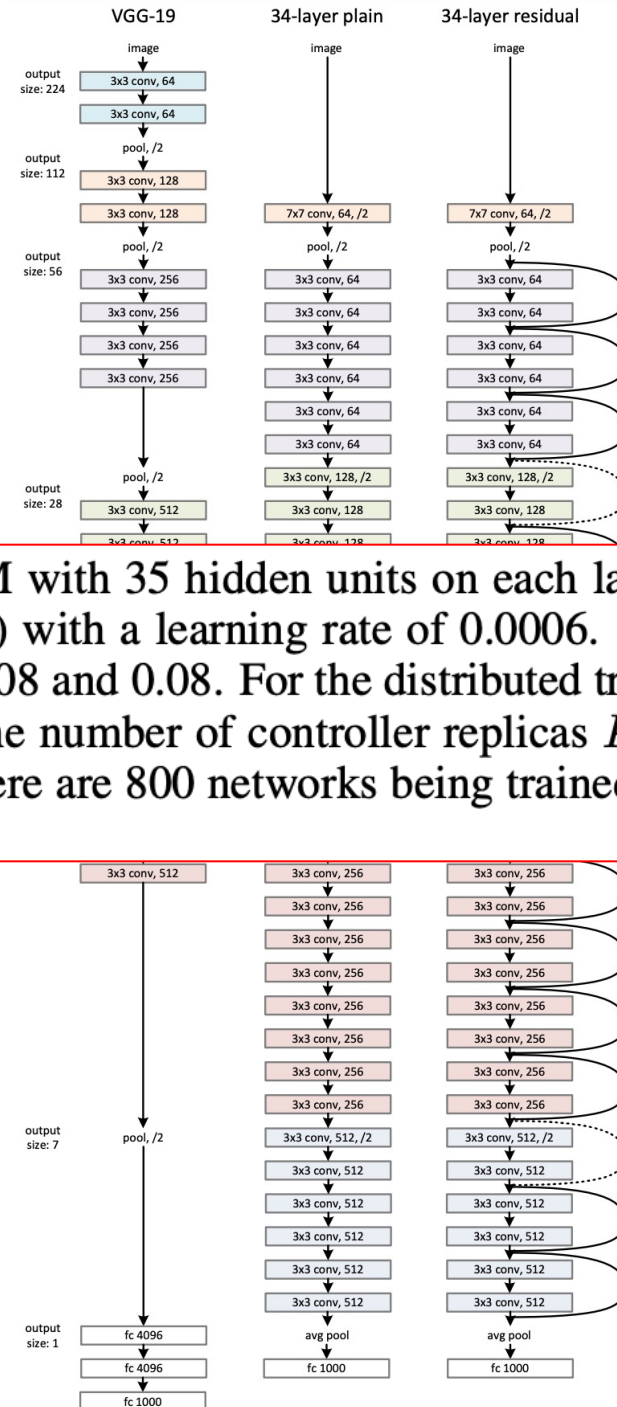


**Training details:** The controller RNN is a two-layer LSTM with 35 hidden units on each layer. It is trained with the ADAM optimizer (Kingma & Ba, 2015) with a learning rate of 0.0006. The weights of the controller are initialized uniformly between -0.08 and 0.08. For the distributed training, we set the number of parameter server shards  $S$  to 20, the number of controller replicas  $K$  to 100 and the number of child replicas  $m$  to 8, which means there are 800 networks being trained on 800 GPUs concurrently at any time.

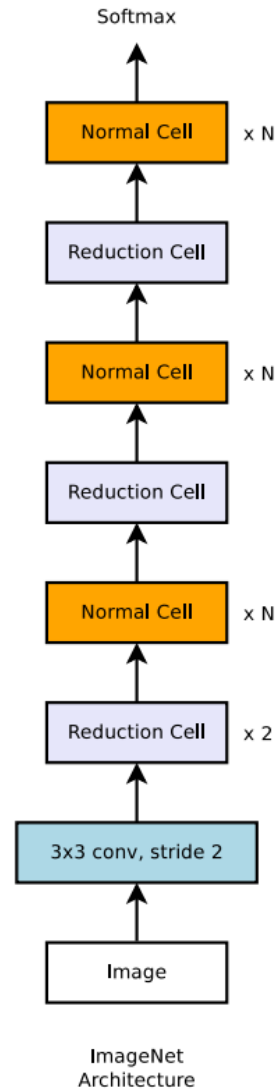
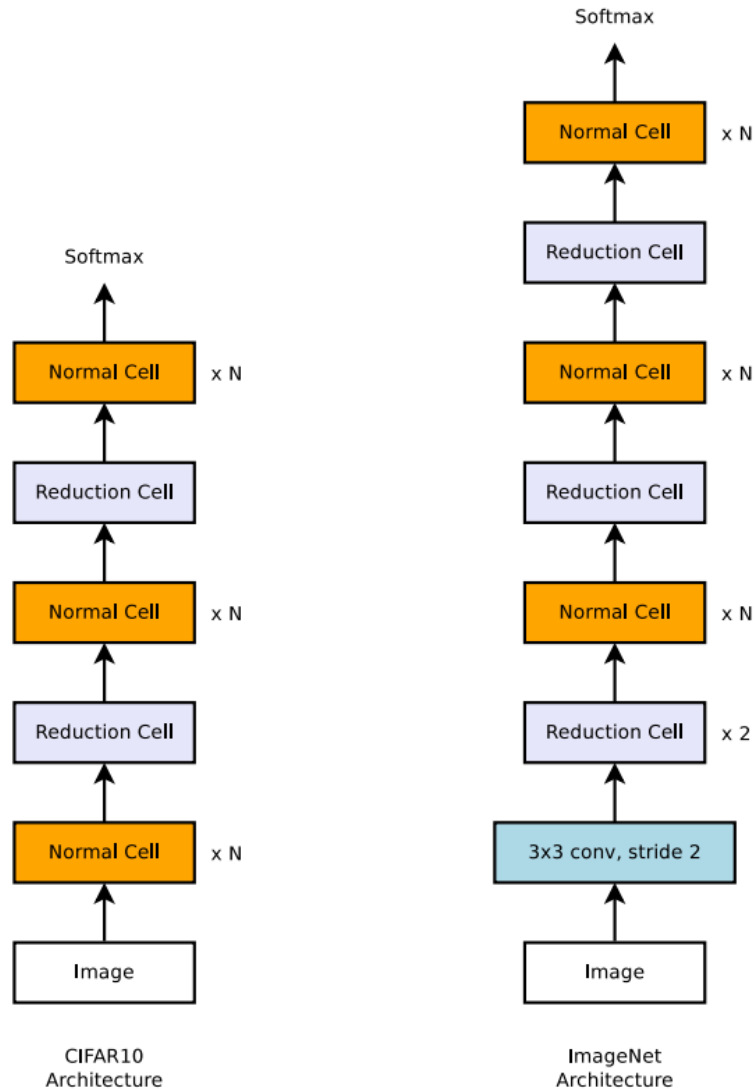
28 days



Q: The definition of sequential layer-wise operations works for ResNet-like networks? **Yes, but the scalability?**



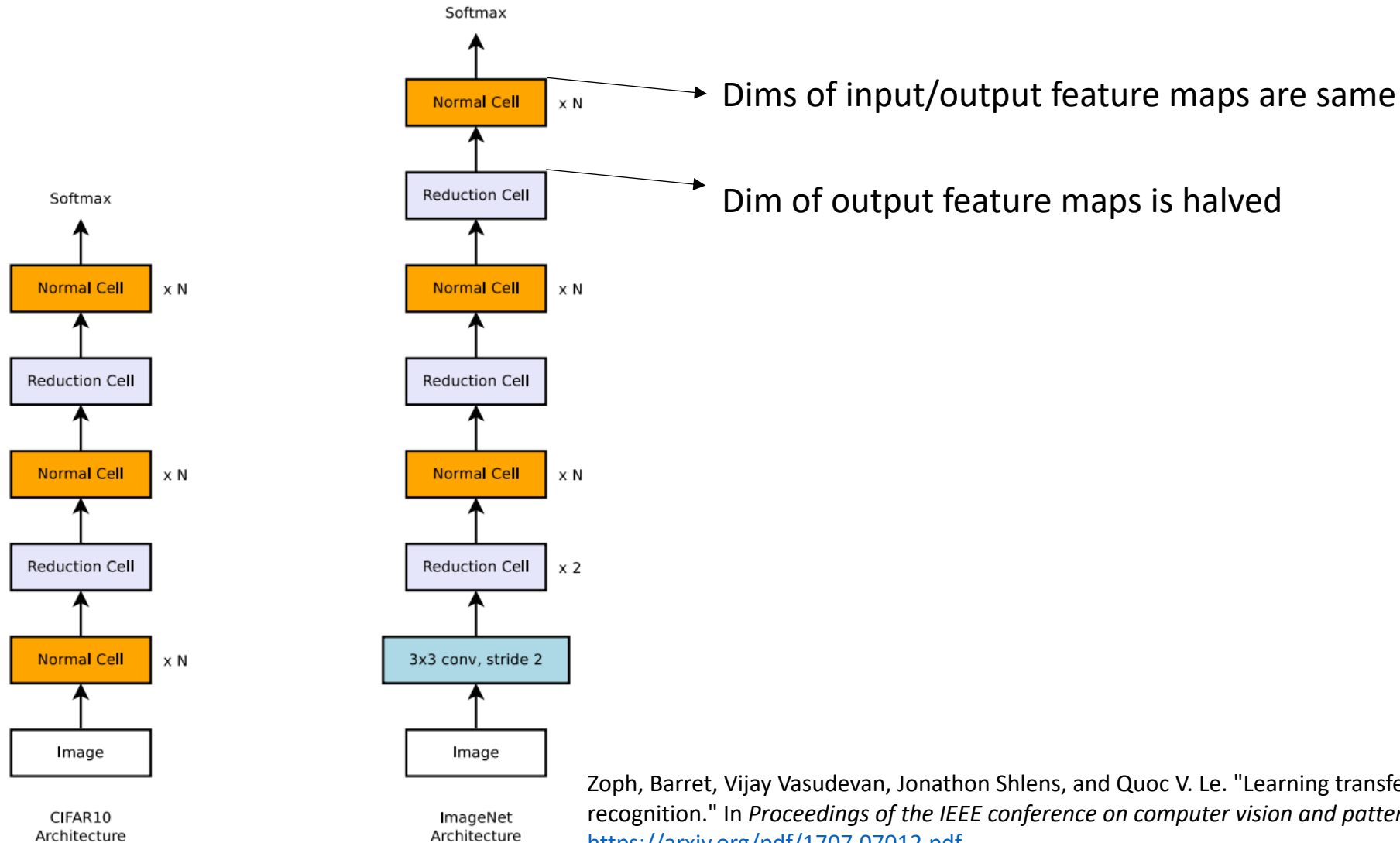
# Defining search space: Cell-based



Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. "Learning transferable architectures for scalable image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710. 2018.  
<https://arxiv.org/pdf/1707.07012.pdf>



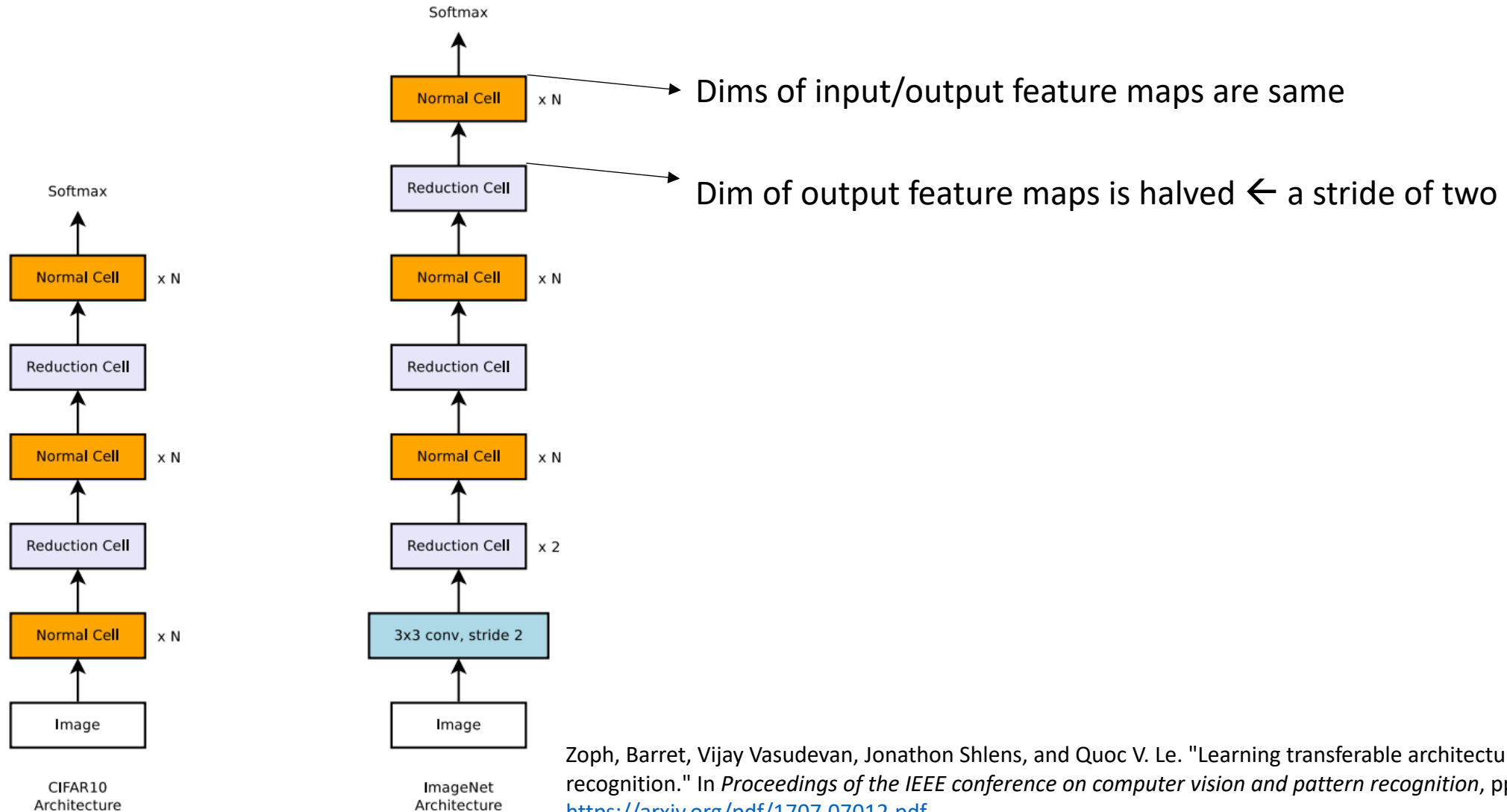
# Defining search space: Cell-based



Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. "Learning transferable architectures for scalable image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710. 2018.

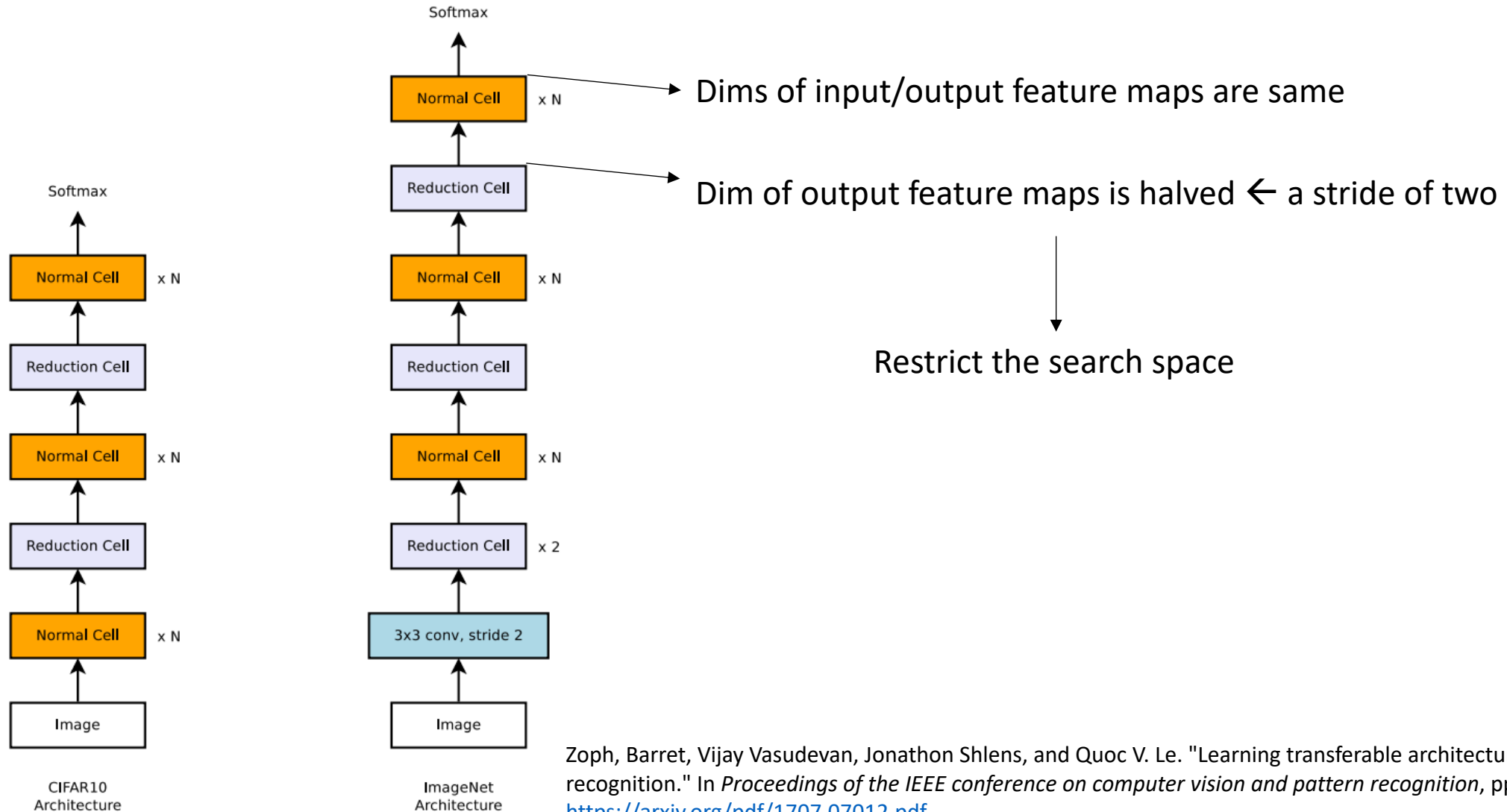
<https://arxiv.org/pdf/1707.07012.pdf>

# Defining search space: Cell-based

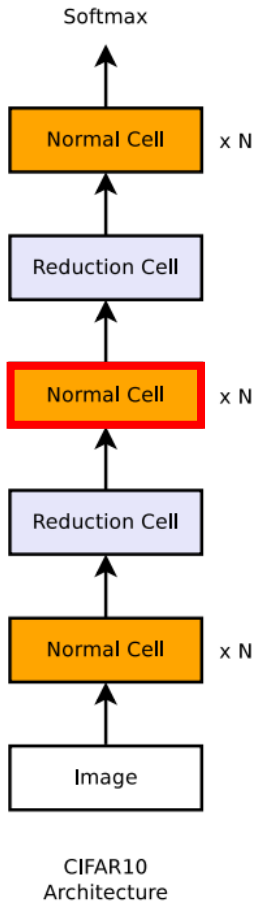


Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. "Learning transferable architectures for scalable image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710. 2018.  
<https://arxiv.org/pdf/1707.07012.pdf>

# Defining search space: Cell-based

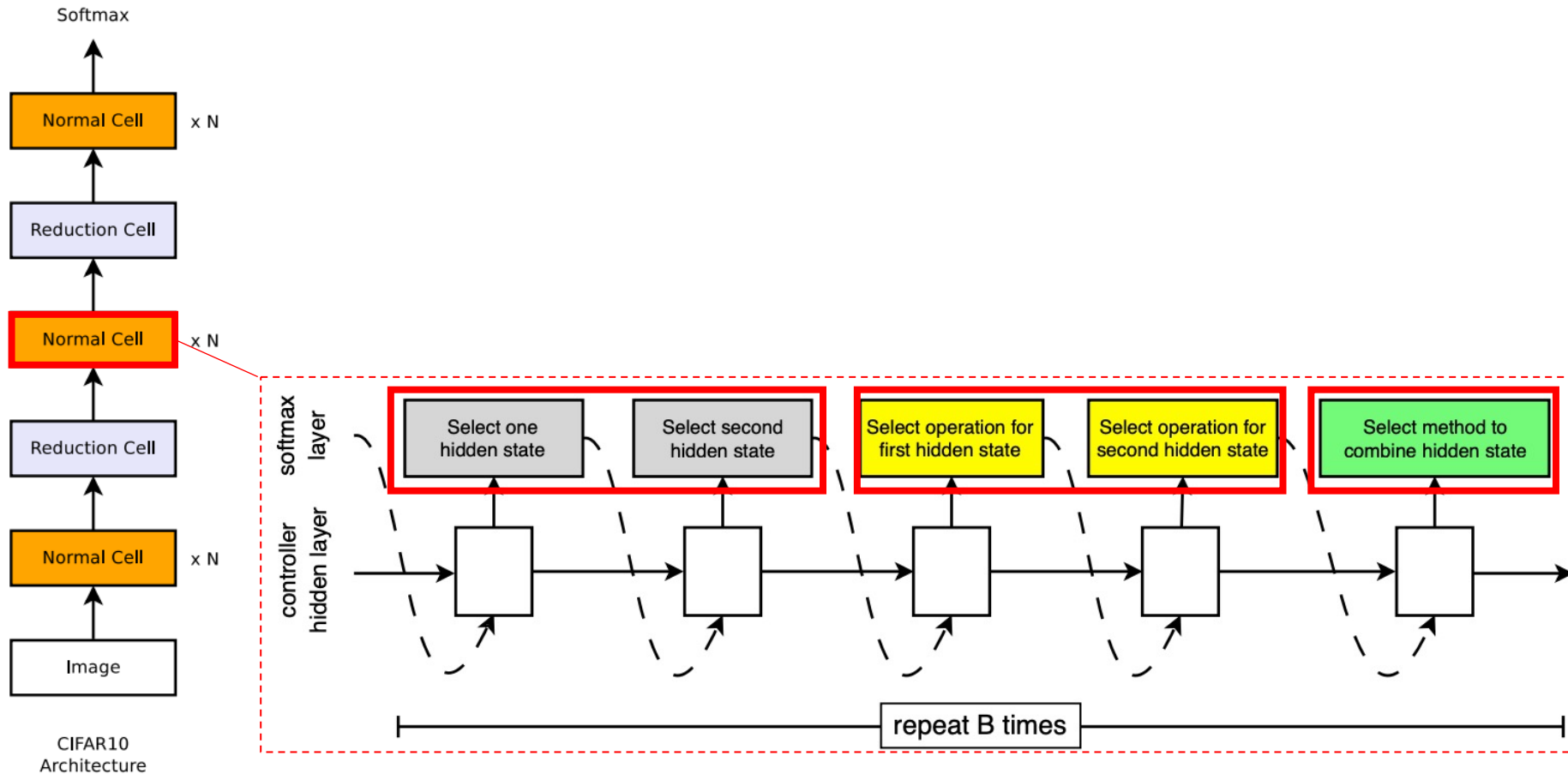


# Defining search space: Cell-based



Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. "Learning transferable architectures for scalable image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710. 2018.  
<https://arxiv.org/pdf/1707.07012.pdf>

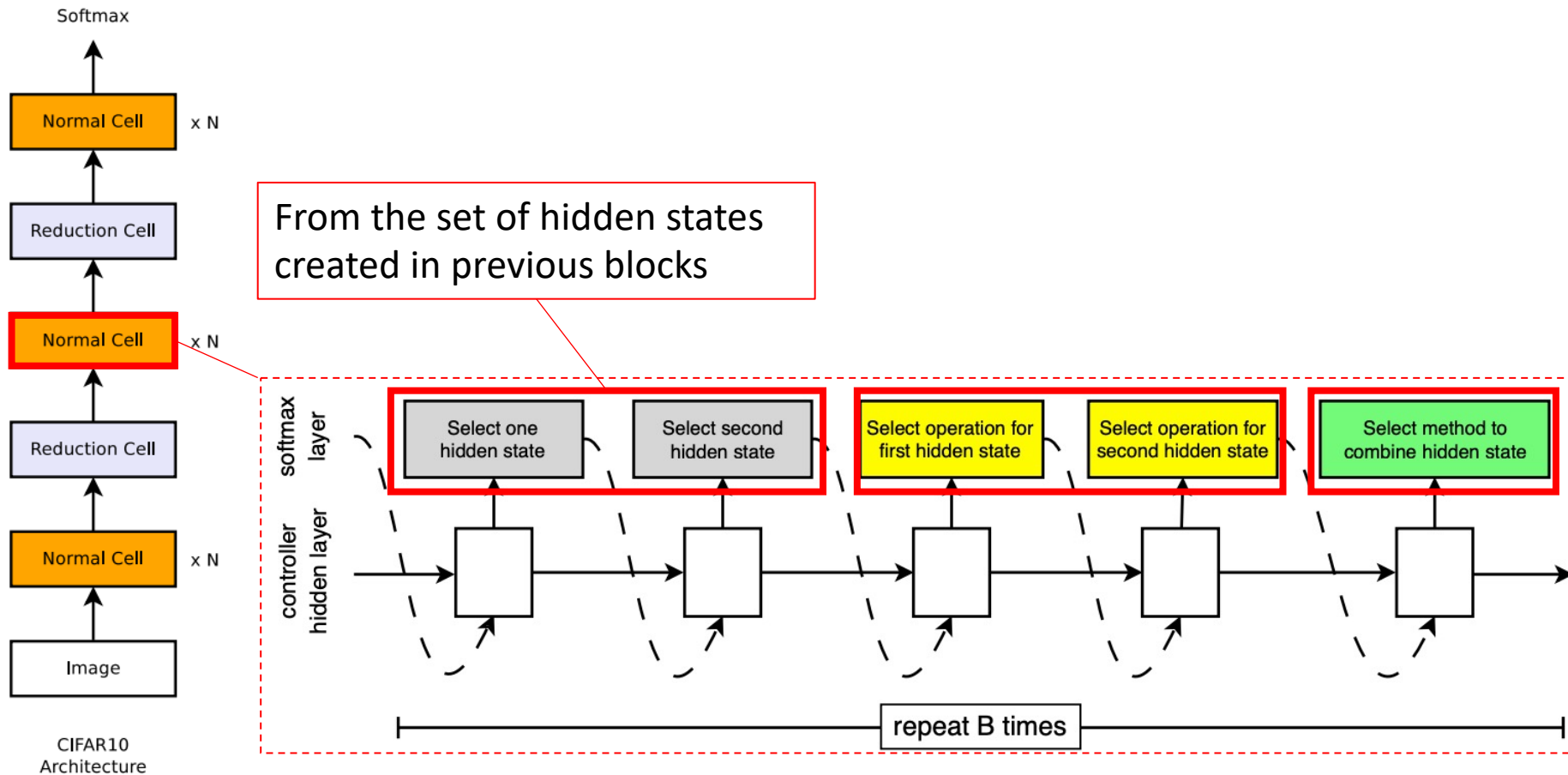
# Defining search space: Cell-based



Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. "Learning transferable architectures for scalable image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710. 2018.

<https://arxiv.org/pdf/1707.07012.pdf>

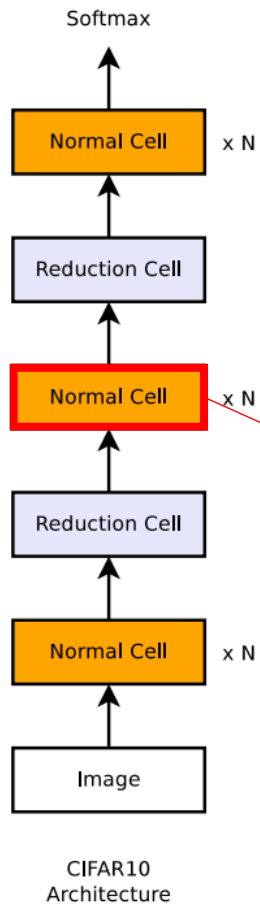
# Defining search space: Cell-based



Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. "Learning transferable architectures for scalable image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710. 2018.

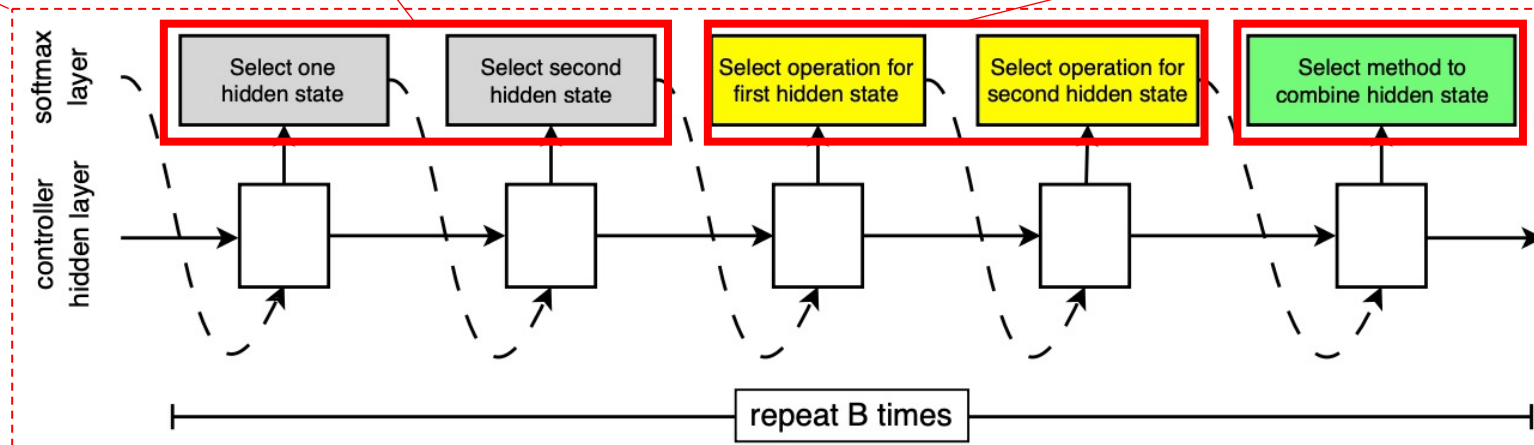
<https://arxiv.org/pdf/1707.07012.pdf>

# Defining search space: Cell-based



From the set of hidden states created in previous blocks

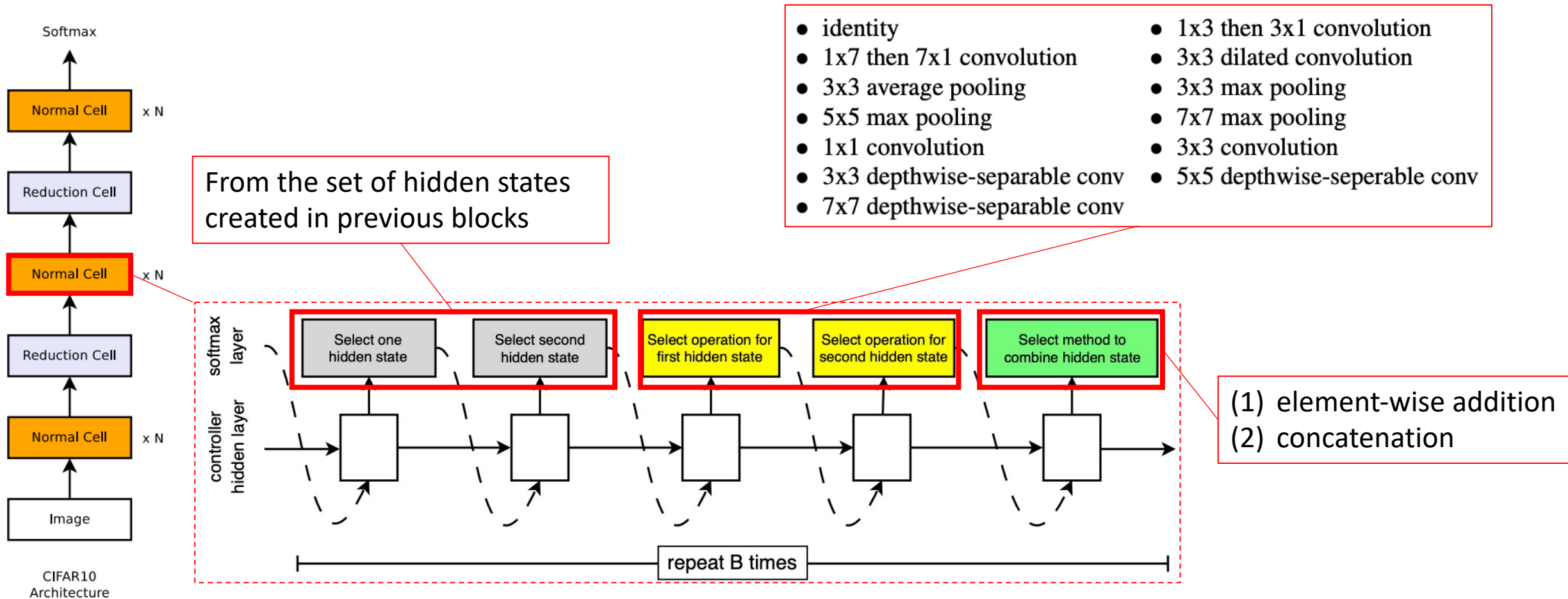
- identity
- 1x7 then 7x1 convolution
- 3x3 average pooling
- 5x5 max pooling
- 1x1 convolution
- 3x3 depthwise-separable conv
- 7x7 depthwise-separable conv
- 1x3 then 3x1 convolution
- 3x3 dilated convolution
- 3x3 max pooling
- 7x7 max pooling
- 3x3 convolution
- 5x5 depthwise-seperable conv



Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. "Learning transferable architectures for scalable image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710. 2018.

<https://arxiv.org/pdf/1707.07012.pdf>

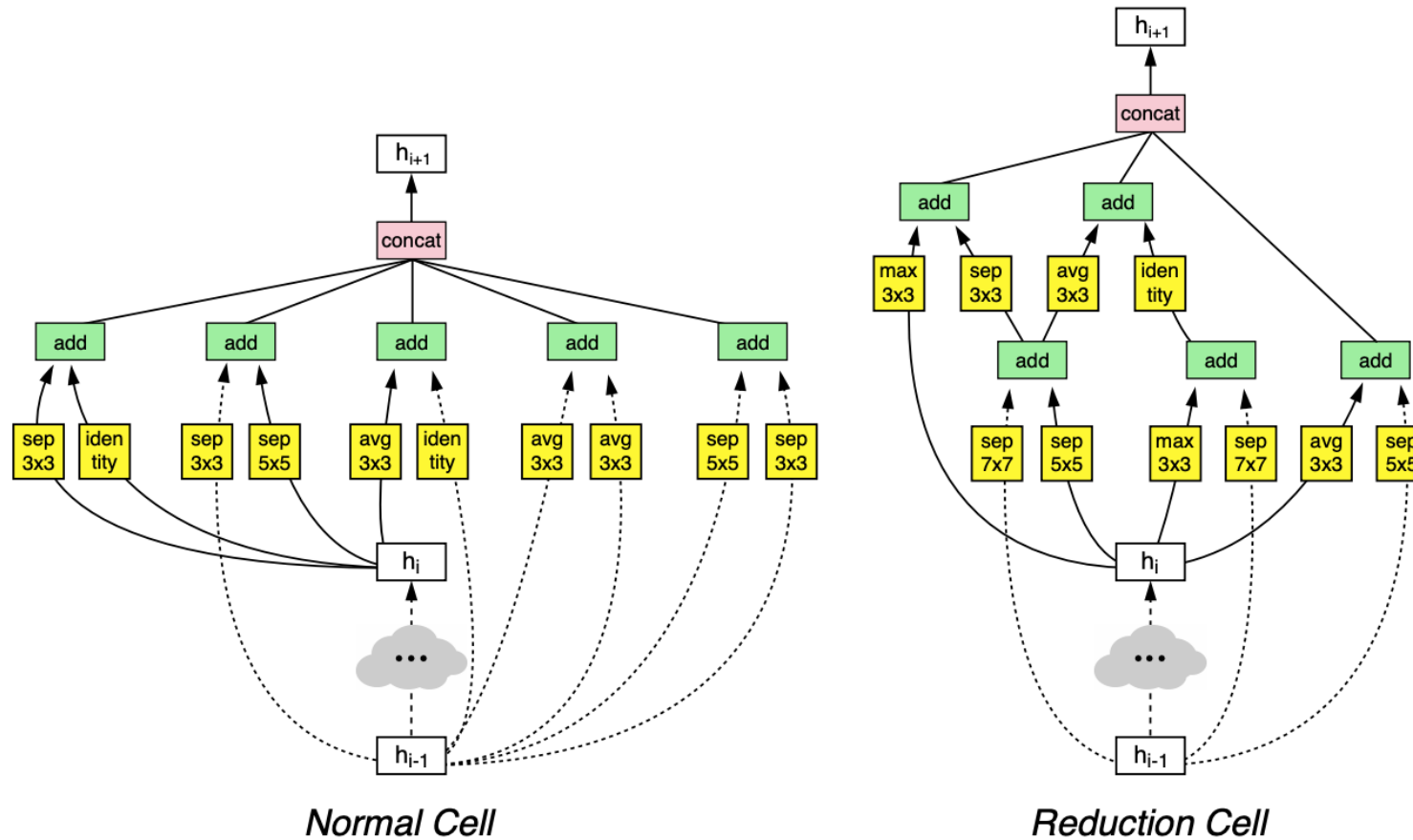
# Defining search space: Cell-based



Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. "Learning transferable architectures for scalable image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710. 2018.  
<https://arxiv.org/pdf/1707.07012.pdf>

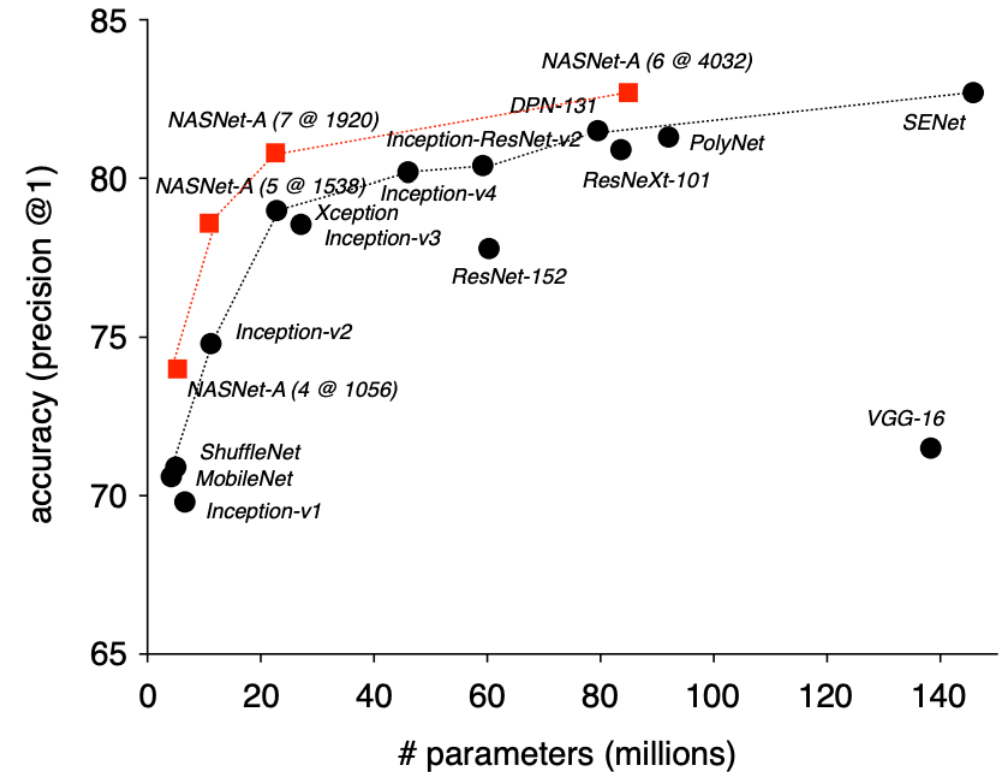
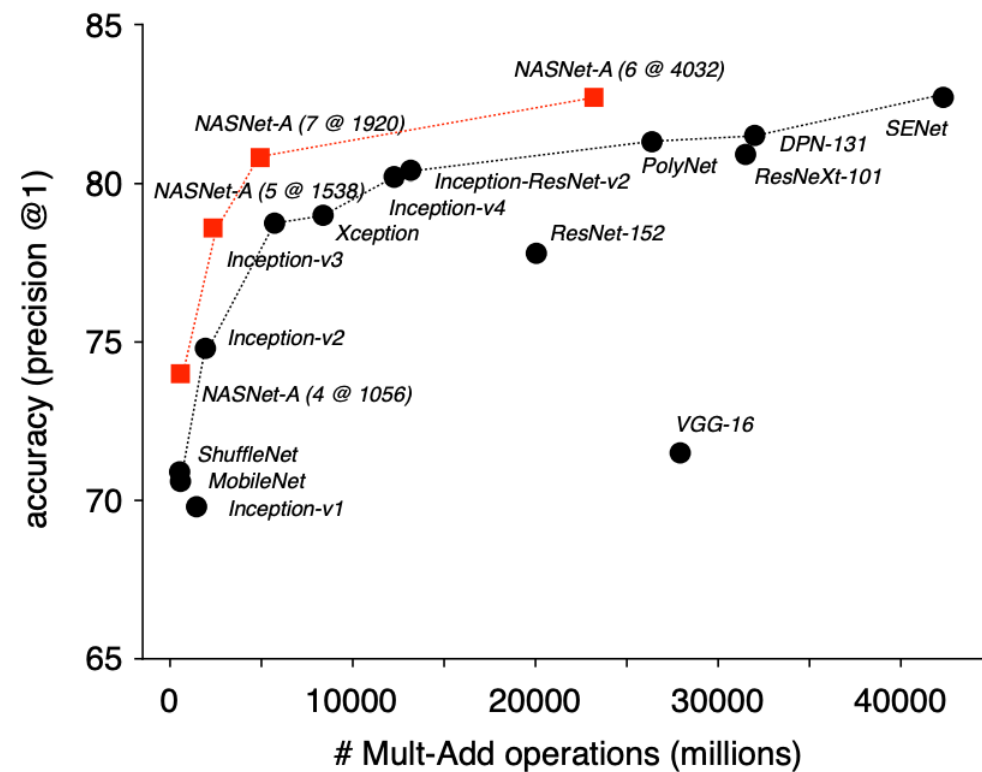


# Defining search space: Cell-based



Best cells on CIFAR-10 with  $B = 5$

# Performance of NAS models



Zoph, Barret, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. "Learning transferable architectures for scalable image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8697-8710. 2018.

<https://arxiv.org/pdf/1707.07012.pdf>