Deep learning for image processing

STAT 4710

November 21, 2023

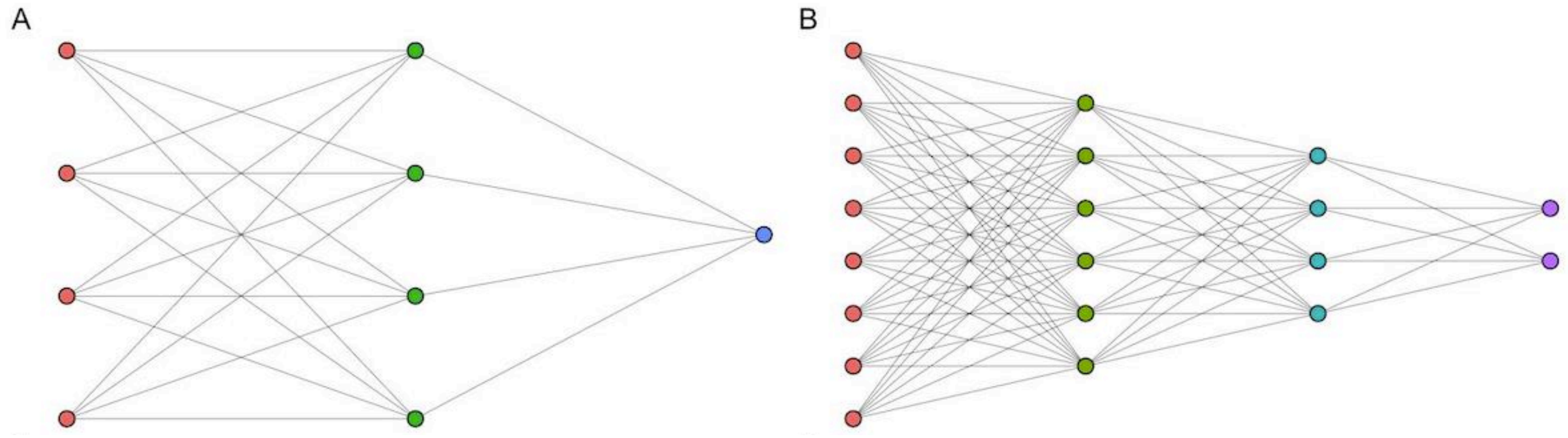
Where we are

- **Unit 1:** R for data mining
- **Unit 2:** Prediction fundamentals
- **Unit 3:** Regression-based methods
- **Unit 4:** Tree-based methods
- **Unit 5:** Deep learning
  - **Lecture 1:** Deep learning preliminaries
  - **Lecture 2:** Neural networks
  - $\boxed{\textbf{Lecture 3: } \text{Deep learning for images}}$
  - **Lecture 4:** Deep learning for text
  - **Lecture 5:** Unit review and quiz in class

Network architectures

Network architectures

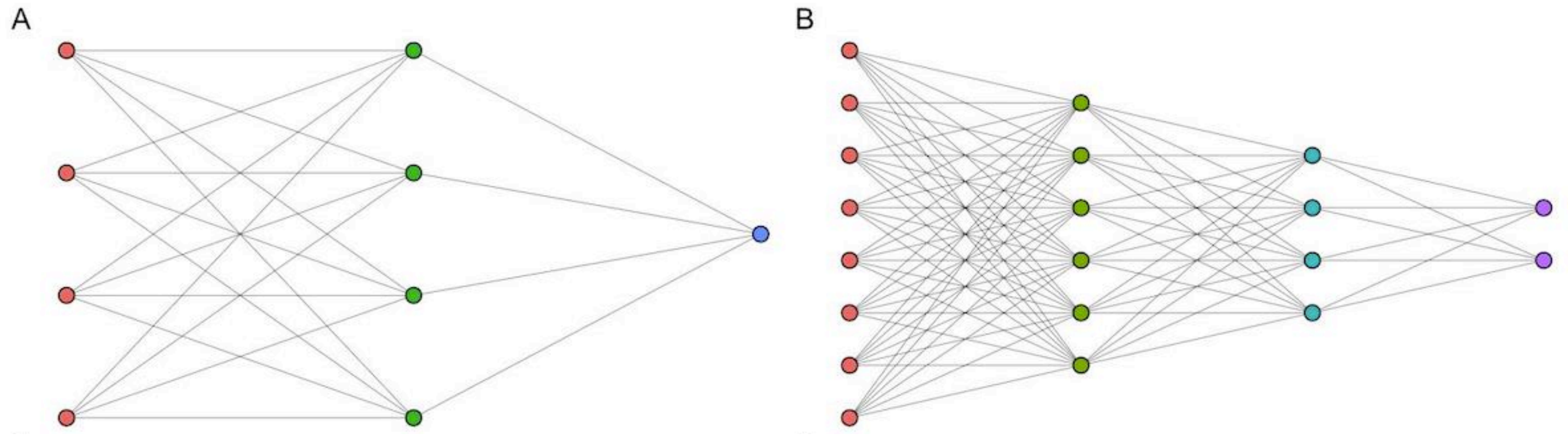
Fully connected architectures



https://community.rstudio.com/t/visualising-neural-network-architectures/41723

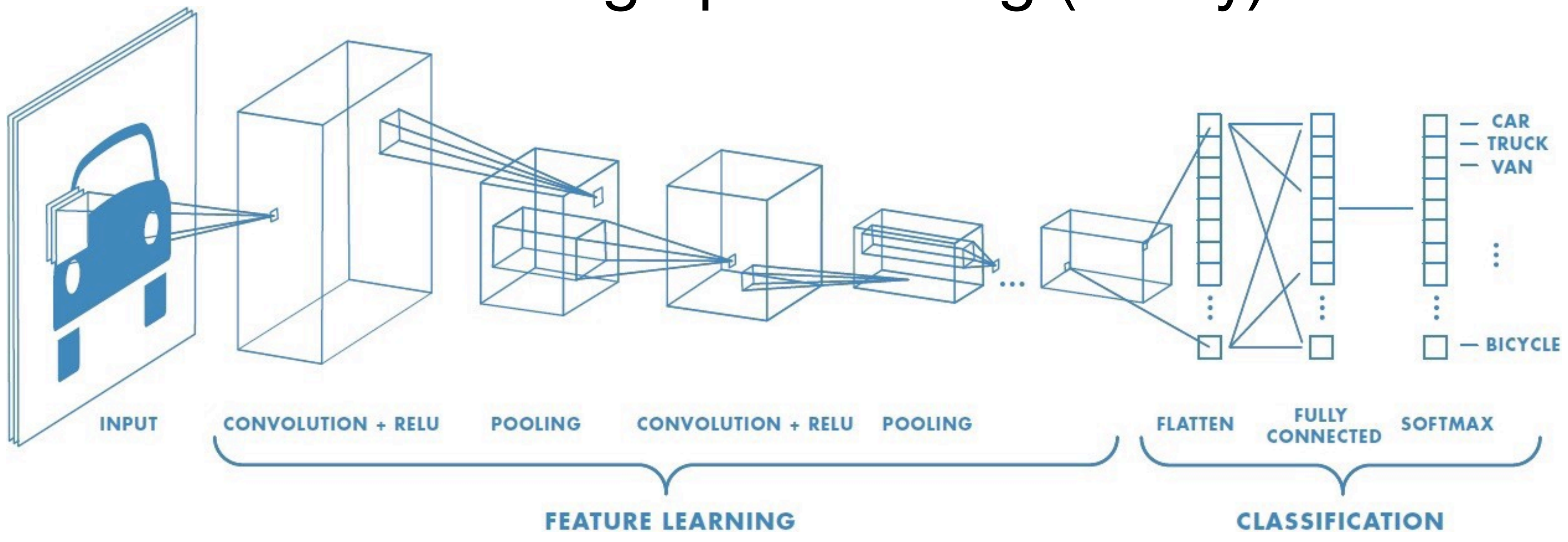
Network architectures

Fully connected architectures



https://community.rstudio.com/t/visualising-neural-network-architectures/41723

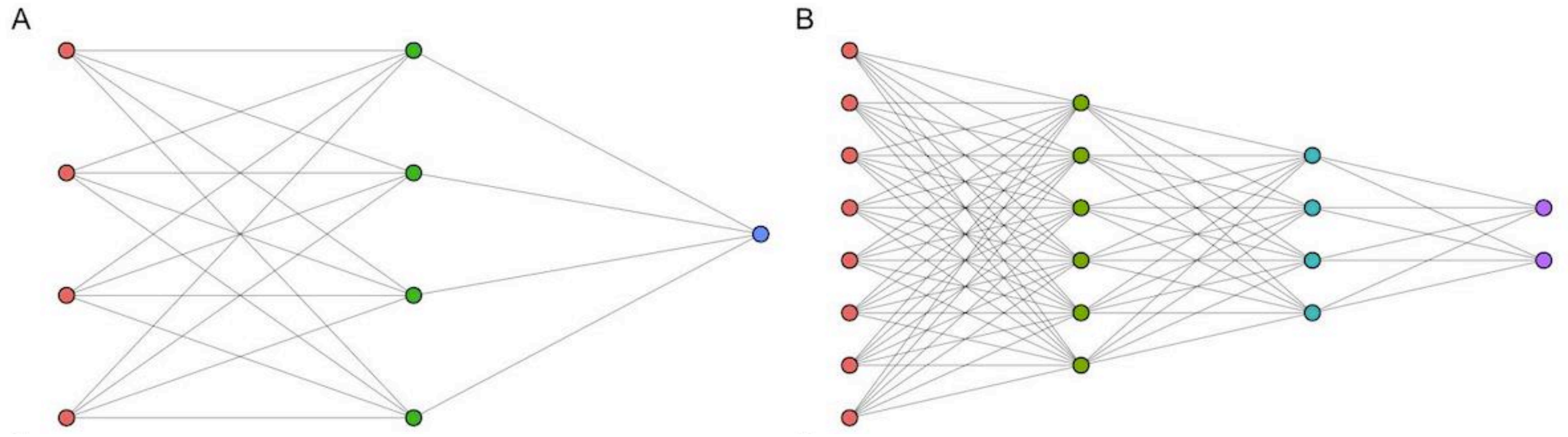
Convolutional neural network (CNN) architectures for image processing (today)



https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

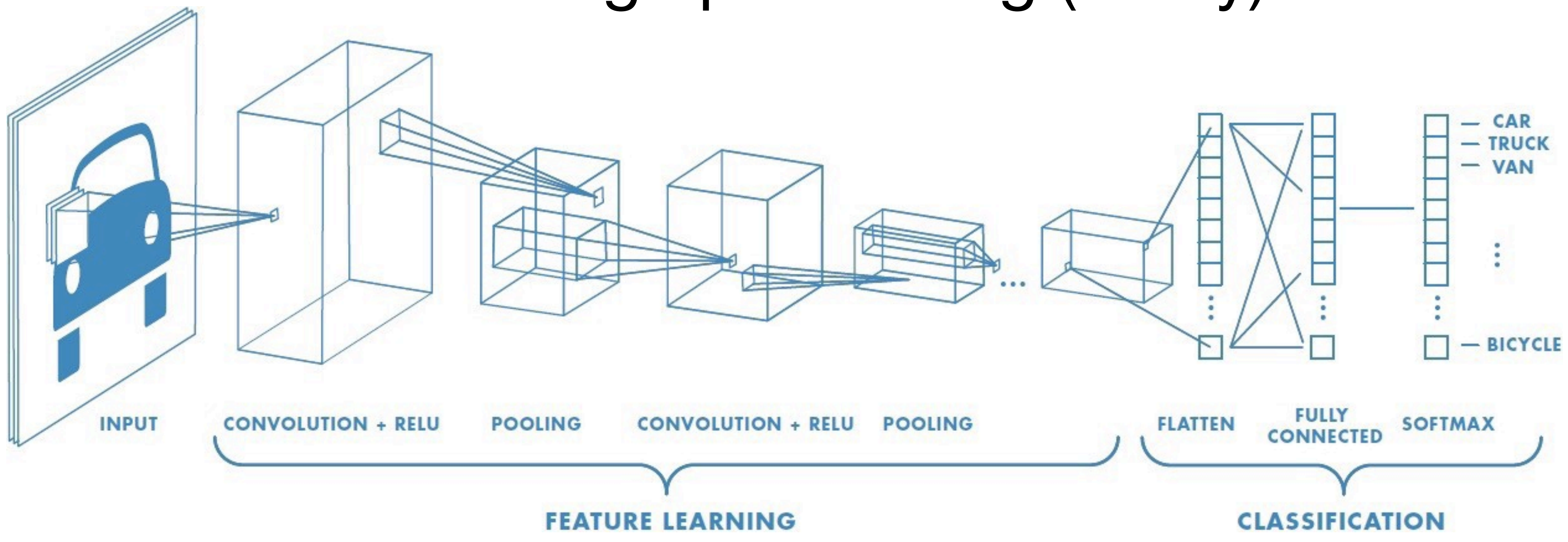
Network architectures

Fully connected architectures



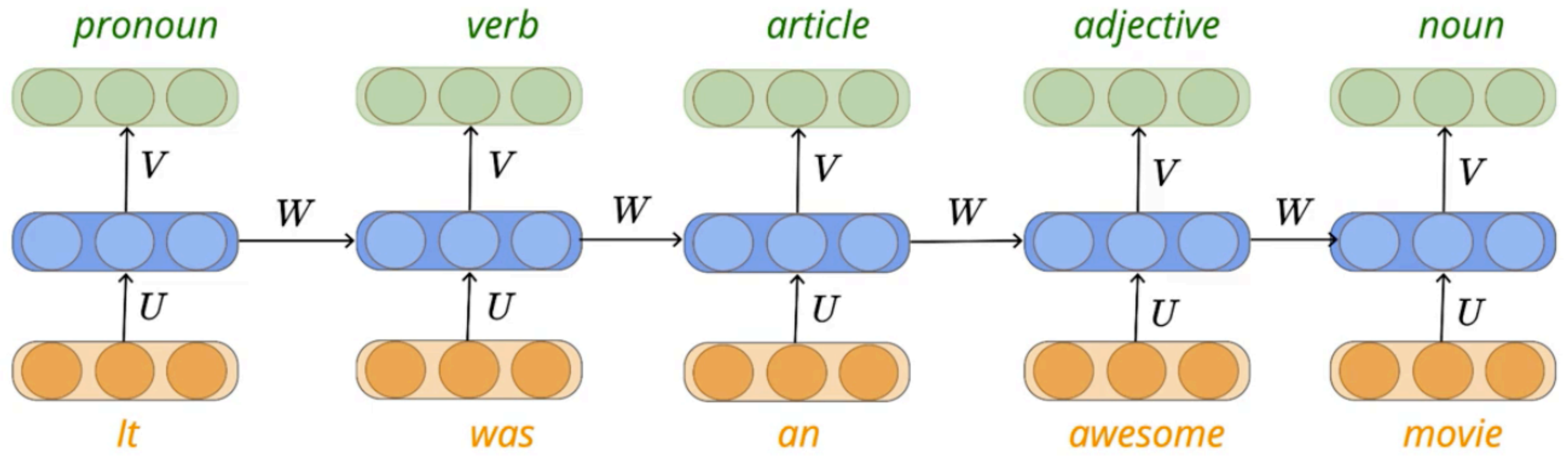
https://community.rstudio.com/t/visualising-neural-network-architectures/41723

Convolutional neural network (CNN) architectures for image processing (today)



https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

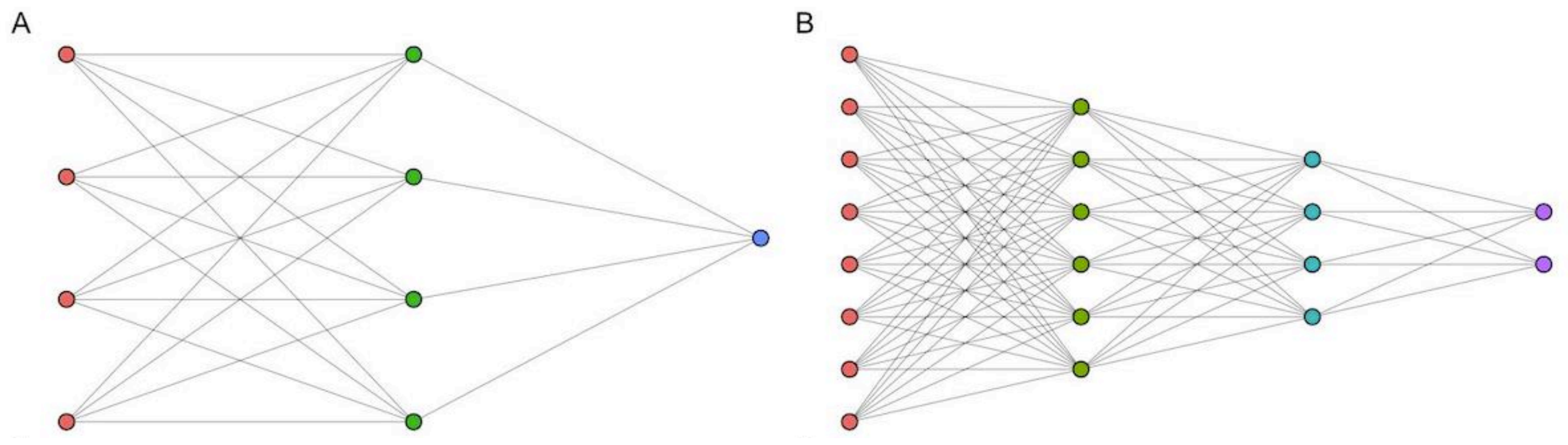
Recurrent neural network architectures for language processing (Thursday)



https://towardsdatascience.com/recurrent-neural-networks-rnn-explained-the-eli5-way-3956887e8b75

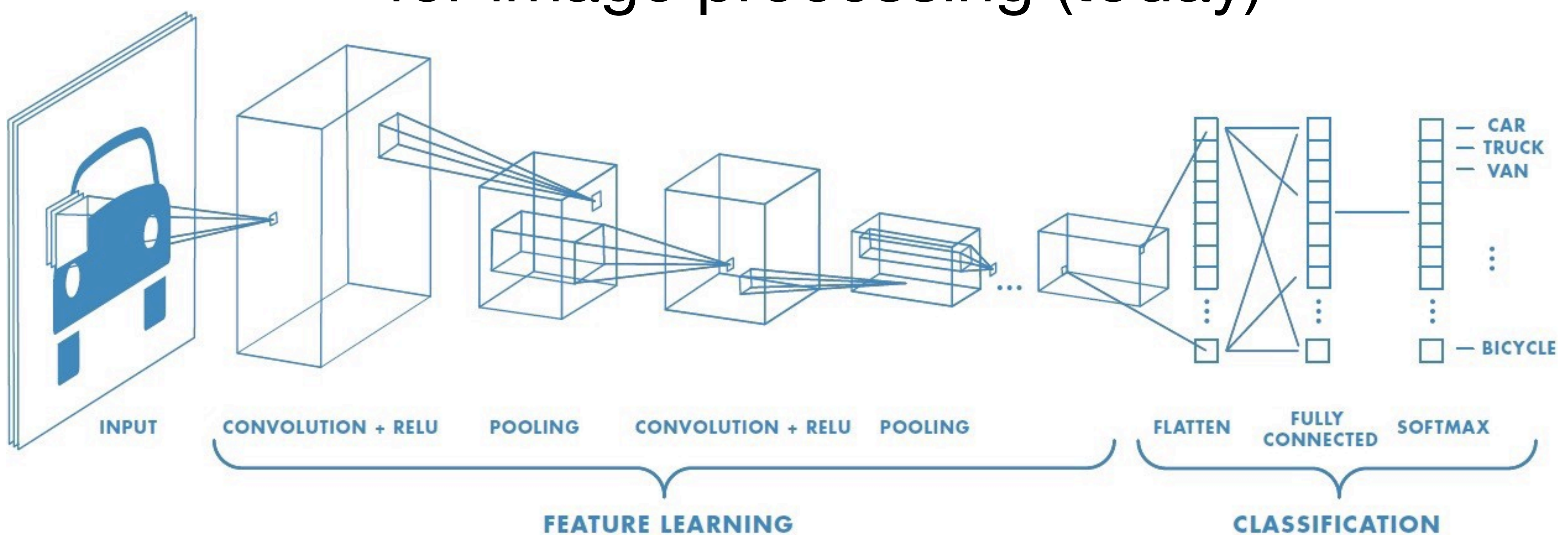
Network architectures

Fully connected architectures



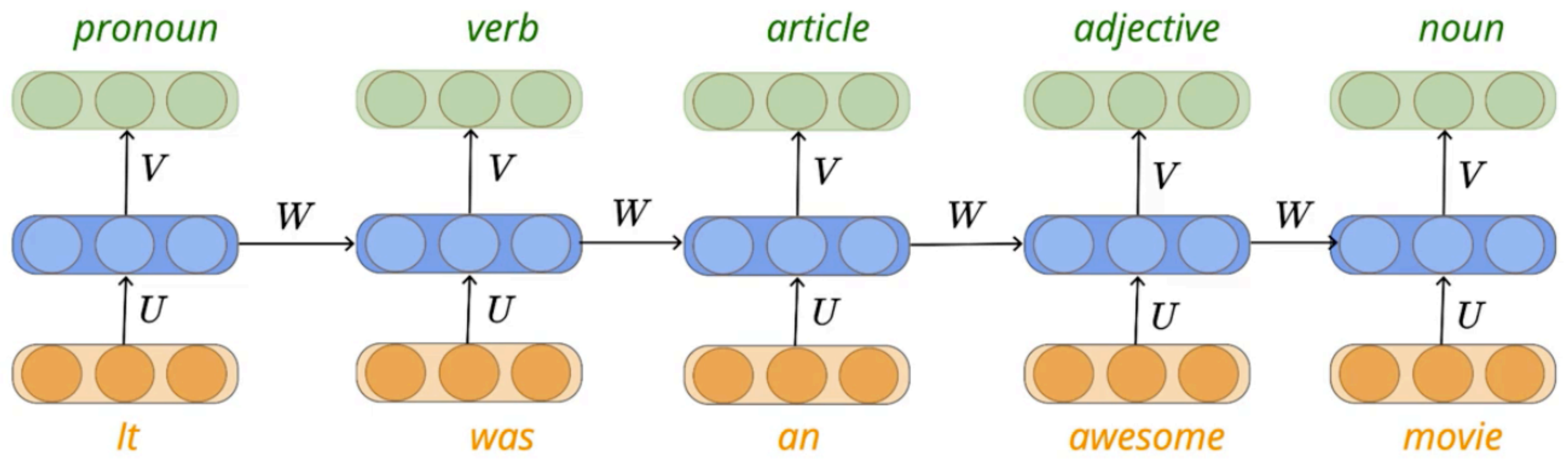
https://community.rstudio.com/t/visualising-neural-network-architectures/41723

Convolutional neural network (CNN) architectures for image processing (today)



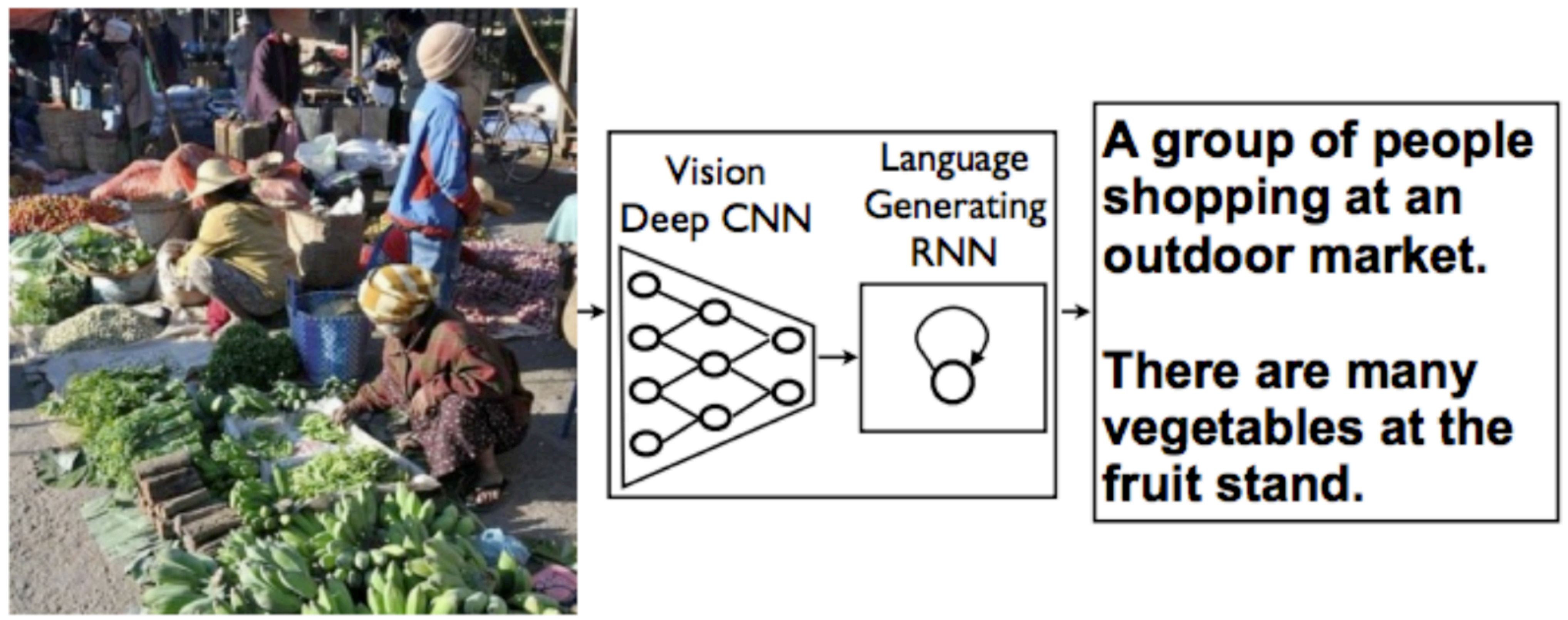
https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Recurrent neural network architectures for language processing (Thursday)



https://towardsdatascience.com/recurrent-neural-networks-rnn-explained-the-eli5-way-3956887e8b75

Architecture components are modular and can be composed, e.g. image captioning



https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781788398060/3/ch03lvl1sec22/what-is-caption-generation

Case study: Image classification

Case study: Image classification

Prototypical computer vision task:

Given an image, classify according to what object it depicts.

http://ai.stanford.edu/~olga/papers/iccv13-ILSVRCanalysis.pdf

Case study: Image classification

Prototypical computer vision task:

Given an image, classify according to what object it depicts.

Challenges:

http://ai.stanford.edu/~olga/papers/iccv13-ILSVRCanalysis.pdf

Case study: Image classification

Prototypical computer vision task:

Given an image, classify according to what object it depicts.

Challenges:

- Viewpoint variation

http://ai.stanford.edu/~olga/papers/iccv13-ILSVRCanalysis.pdf

Case study: Image classification

Prototypical computer vision task:

Given an image, classify according to what object it depicts.

Challenges:

- Viewpoint variation
- Illumination

http://ai.stanford.edu/~olga/papers/iccv13-ILSVRCanalysis.pdf

Case study: Image classification

Prototypical computer vision task:

Given an image, classify according to what object it depicts.

Challenges:

- Viewpoint variation
- Illumination
- Deformation

http://ai.stanford.edu/~olga/papers/iccv13-ILSVRCanalysis.pdf

Case study: Image classification

Prototypical computer vision task:

Given an image, classify according to what object it depicts.

Challenges:

- Viewpoint variation
- Illumination
- Deformation
- Occlusion

http://ai.stanford.edu/~olga/papers/iccv13-ILSVRCanalysis.pdf

Case study: Image classification

Prototypical computer vision task:

Given an image, classify according to what object it depicts.

Challenges:

- Viewpoint variation
- Illumination
- Deformation
- Occlusion
- Background clutter

http://ai.stanford.edu/~olga/papers/iccv13-ILSVRCanalysis.pdf

Case study: Image classification

Prototypical computer vision task:

Given an image, classify according to what object it depicts.

Challenges:

- Viewpoint variation
- Illumination
- Deformation
- Occlusion
- Background clutter
- Intraclass variation

http://ai.stanford.edu/~olga/papers/iccv13-ILSVRCanalysis.pdf

ImageNet

A large dataset for image classification

ImageNet

A large dataset for image classification

Assembled in 2009 by downloading lots of images from the web and crowdsourcing their labels.

https://medium.com/syncedreview/sensetime-trains-imagenet-alexnet-in-record-1-5-minutes-e944ab049b2c

ImageNet

A large dataset for image classification

Assembled in 2009 by downloading lots of images from the web and crowdsourcing their labels.

- Training set: 1.2 million images
- Test set: 100,000 images
- 1000 classes

https://medium.com/syncedreview/sensetime-trains-imagenet-alexnet-in-record-1-5-minutes-e944ab049b2c

ImageNet

A large dataset for image classification

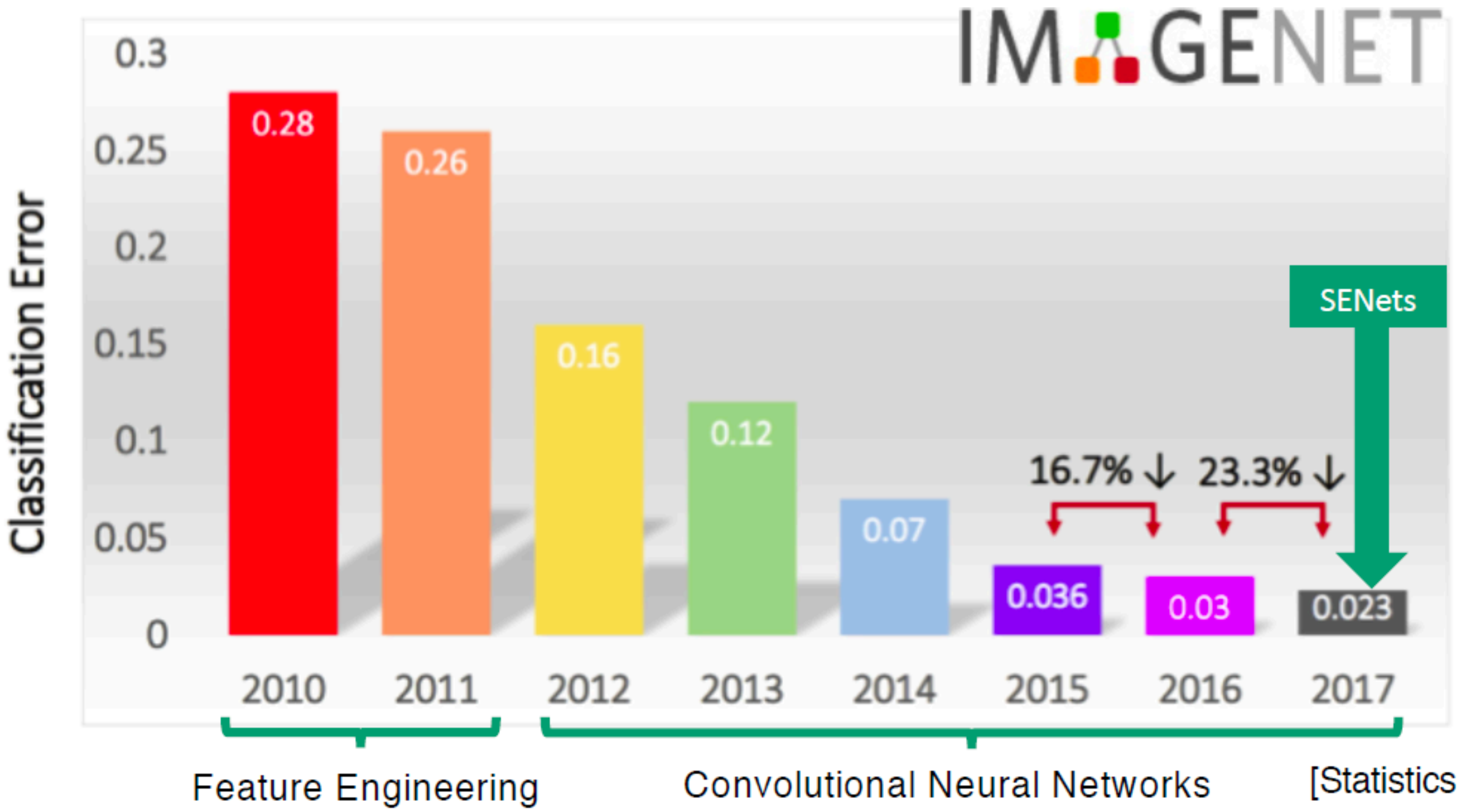
Assembled in 2009 by downloading lots of images from the web and crowdsourcing their labels.

- Training set: 1.2 million images
- Test set: 100,000 images
- 1000 classes

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) held annually between 2010-2017.

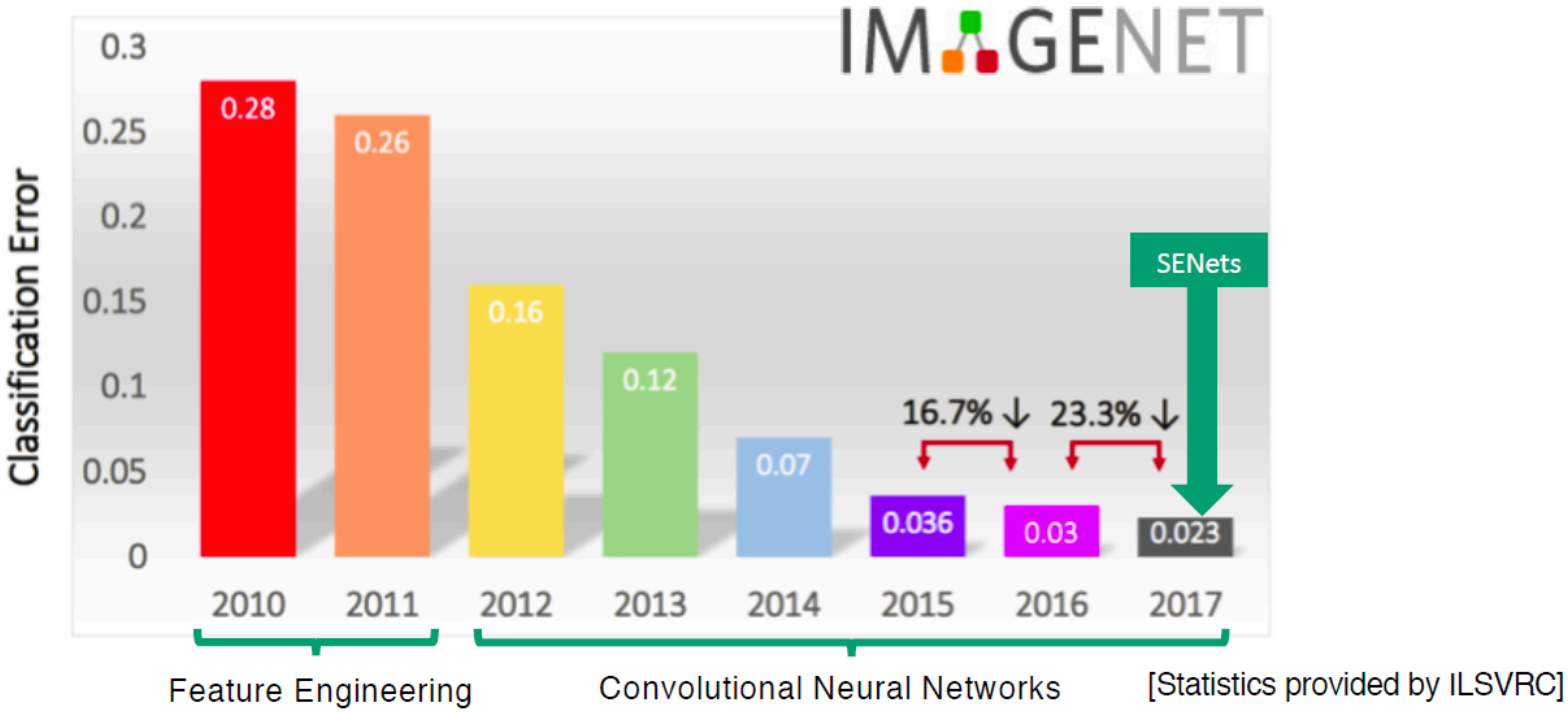
https://medium.com/syncedreview/sensetime-trains-imagenet-alexnet-in-record-1-5-minutes-e944ab049b2c

ILSVRC results over the years



[Statistics provided by ILSVRC]

ILSVRC results over the years

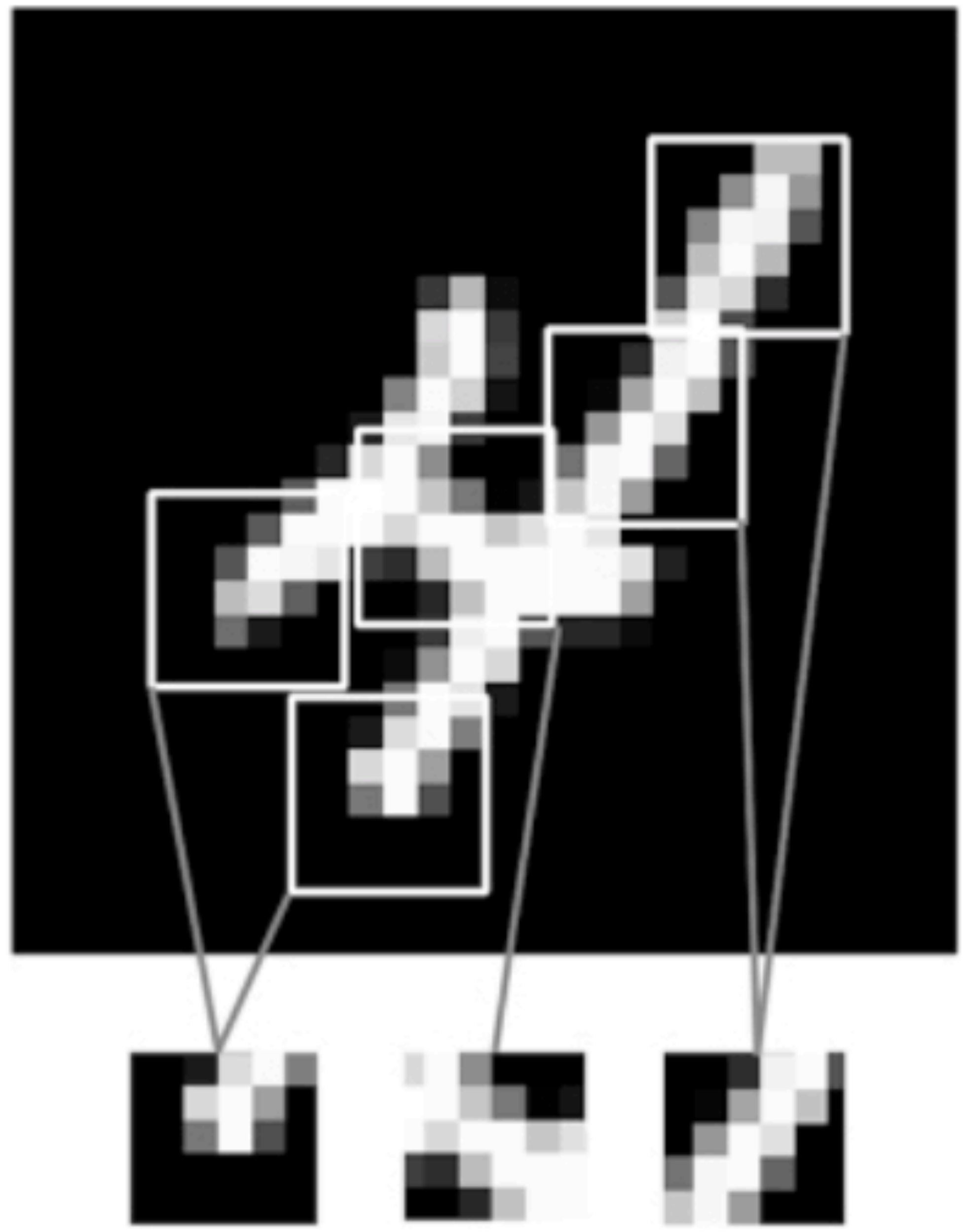


Convolutional neural networks (CNNs) have dominated since 2012.

CNNs are built on image-specific properties

CNNs are built on image-specific properties

Figure 5.1. Images can be broken into local patterns such as edges, textures, and so on.



Images from *Deep Learning with R* (Chapter 5)

CNNs are built on image-specific properties

Figure 5.1. Images can be broken into local patterns such as edges, textures, and so on.

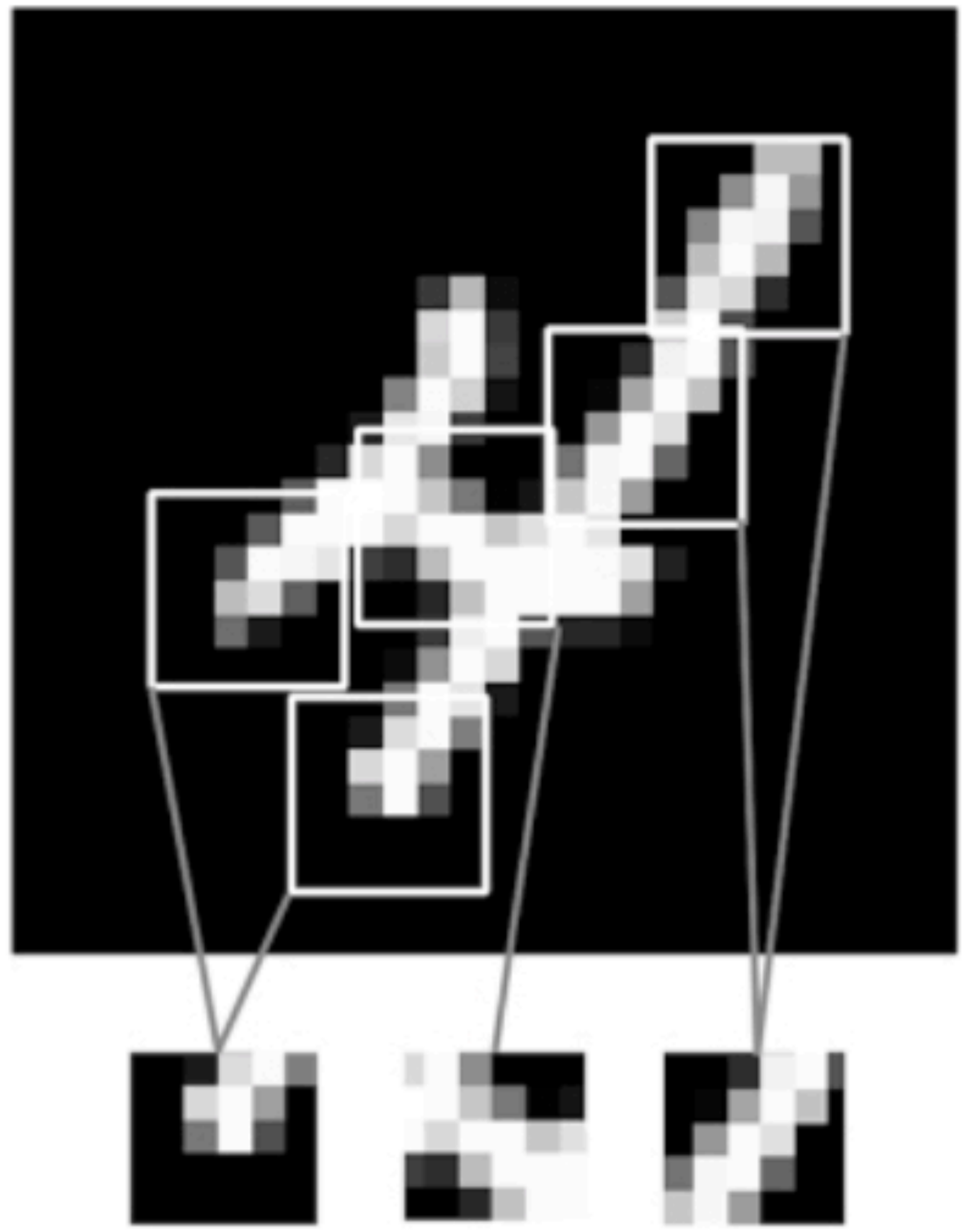
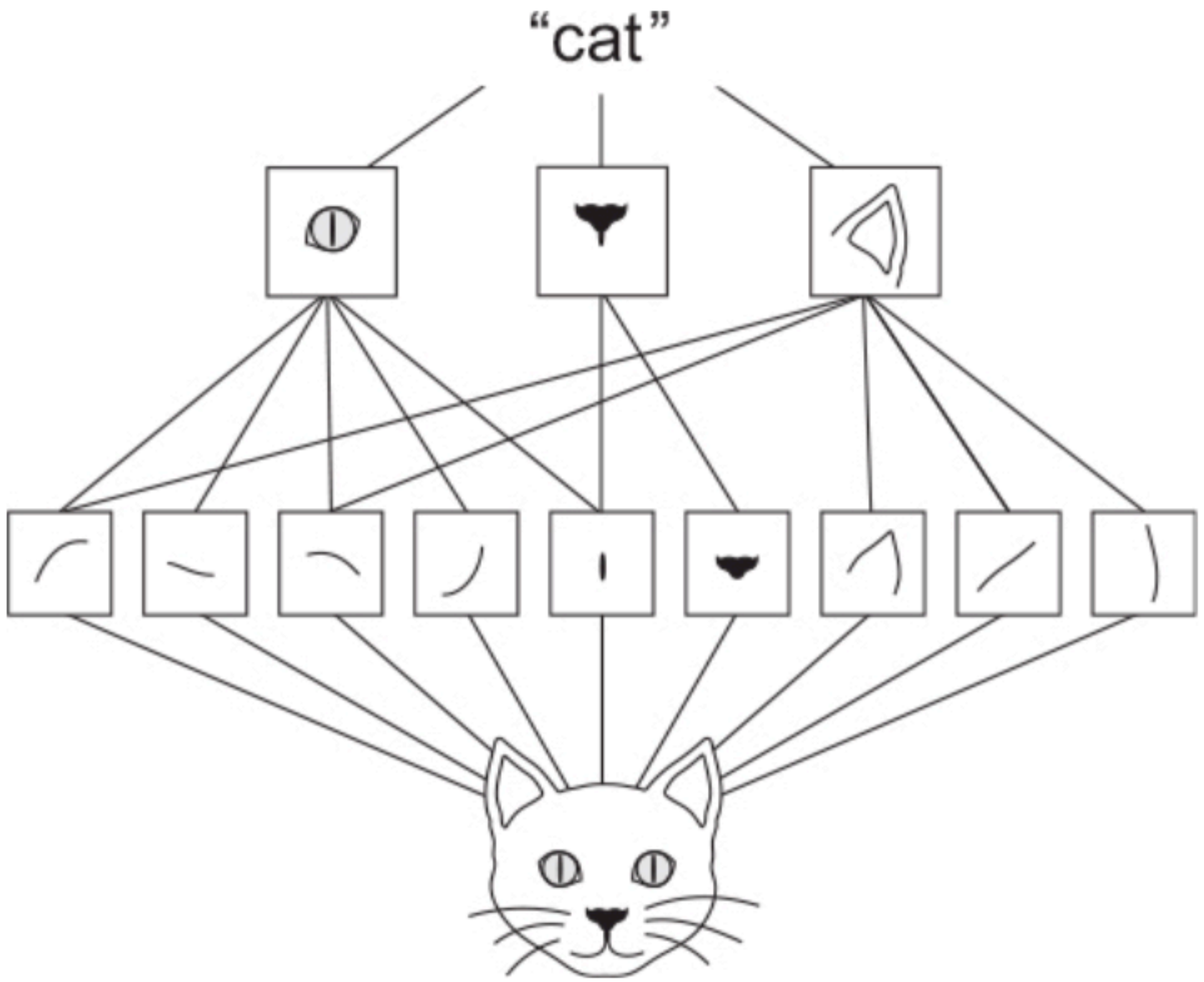


Figure 5.2. The visual world forms a spatial hierarchy of visual modules: hyperlocal edges combine into local objects such as eyes or ears, which combine into high-level concepts such as "cat."



Images from *Deep Learning with R* (Chapter 5)

CNNs are built on image-specific properties

Figure 5.1. Images can be broken into local patterns such as edges, textures, and so on.

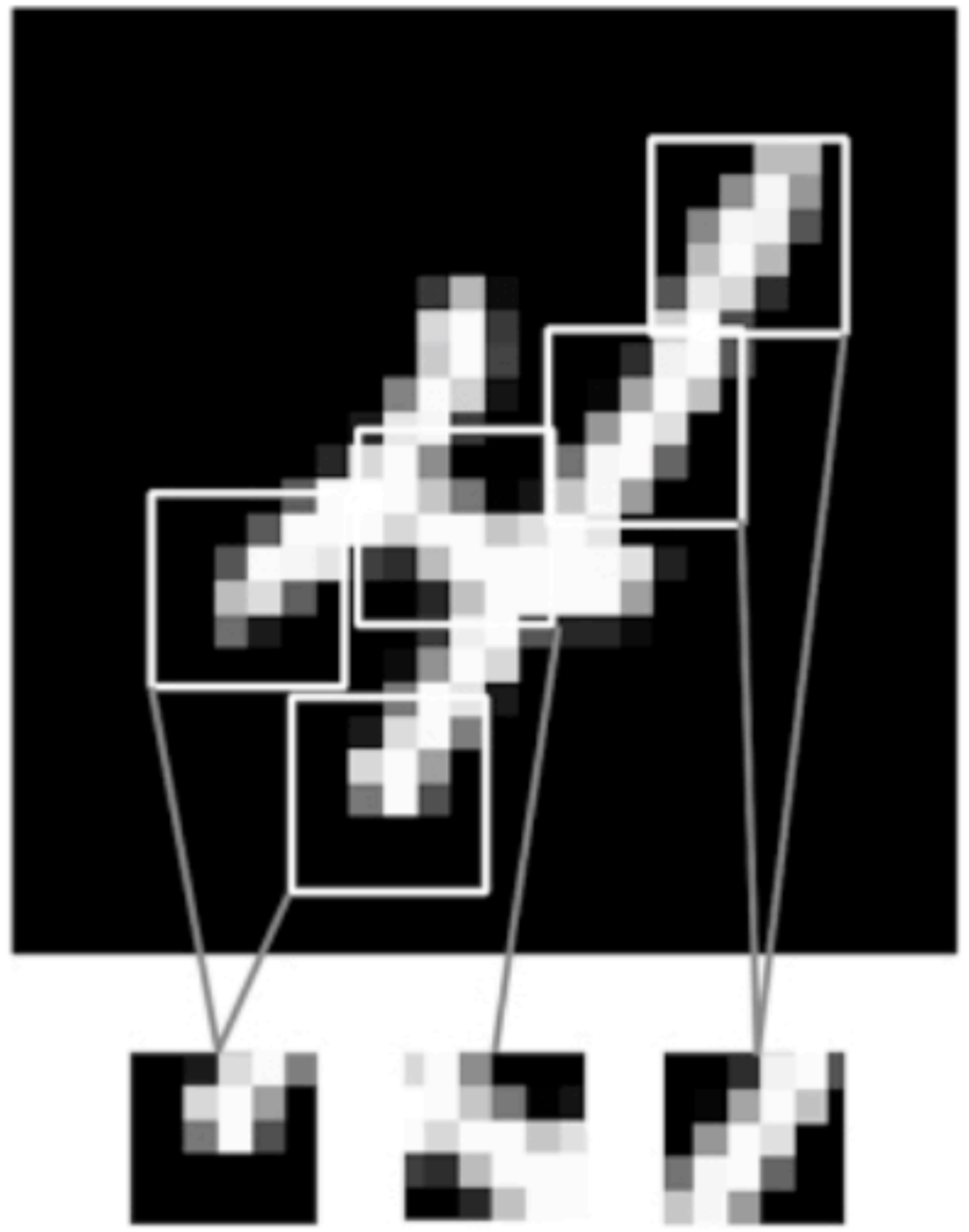
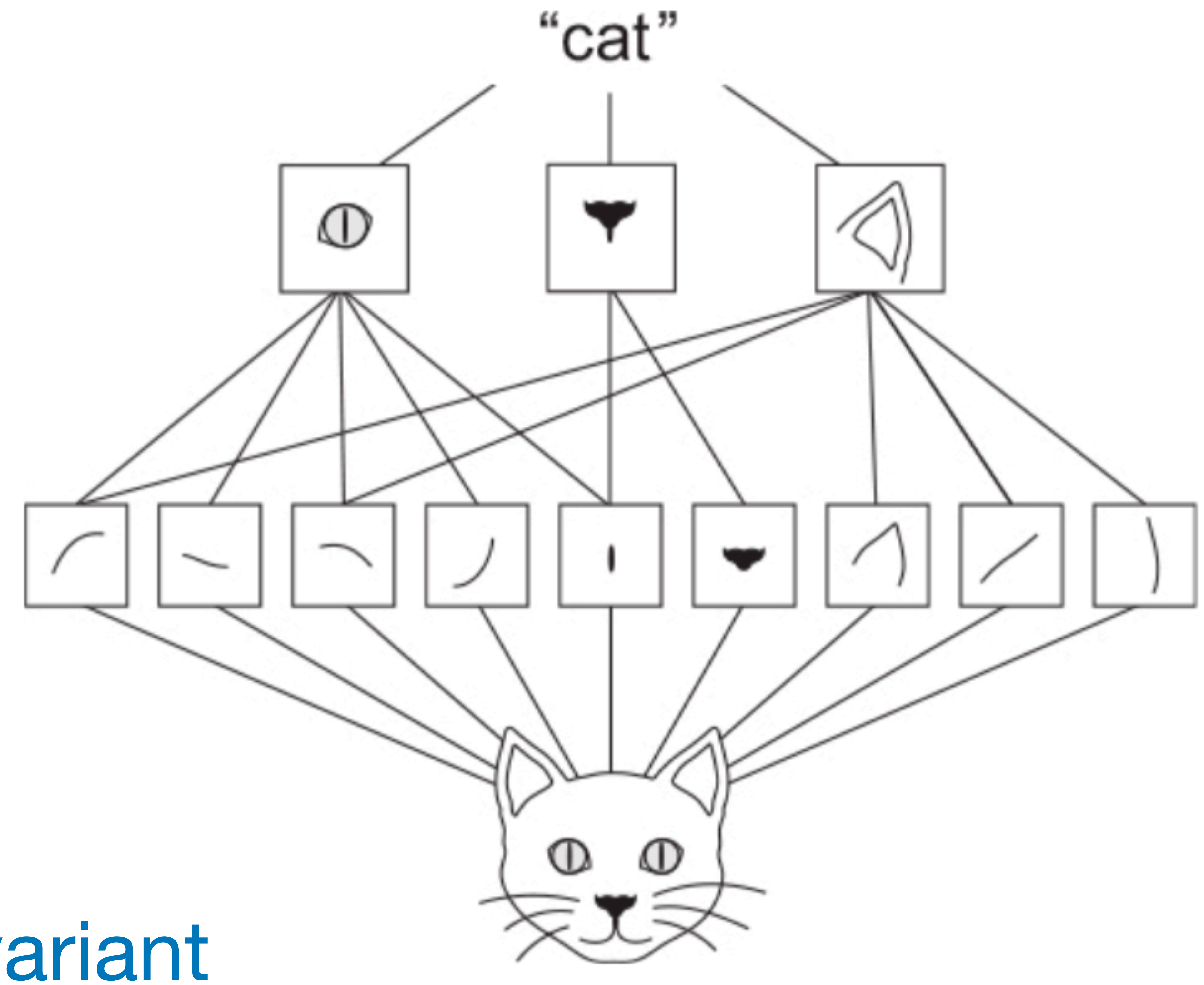


Figure 5.2. The visual world forms a spatial hierarchy of visual modules: hyperlocal edges combine into local objects such as eyes or ears, which combine into high-level concepts such as "cat."



Patterns are

- Local
- Translation-invariant
- Hierarchical

Images from *Deep Learning with R* (Chapter 5)

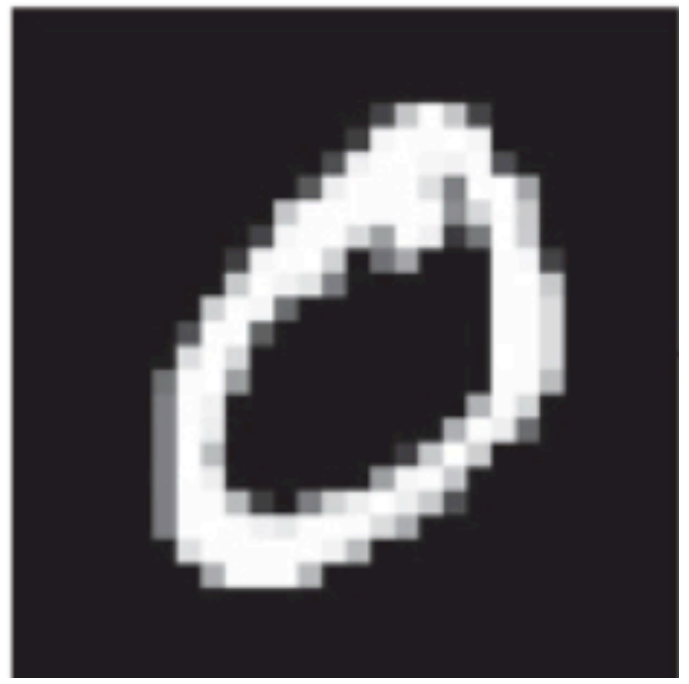
Convolution: Searching for patterns

Filter (3x3)

(pattern)

0	1	2
2	2	0
0	1	2

Input image



3_{0}	3_{1}	2_{2}	1	0
0_{2}	0_{2}	1_{0}	3	1
3_{0}	1_{1}	2_{2}	2	3
2	0	0	2	2
2	0	0	0	1

Activation map

(presence of pattern)

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

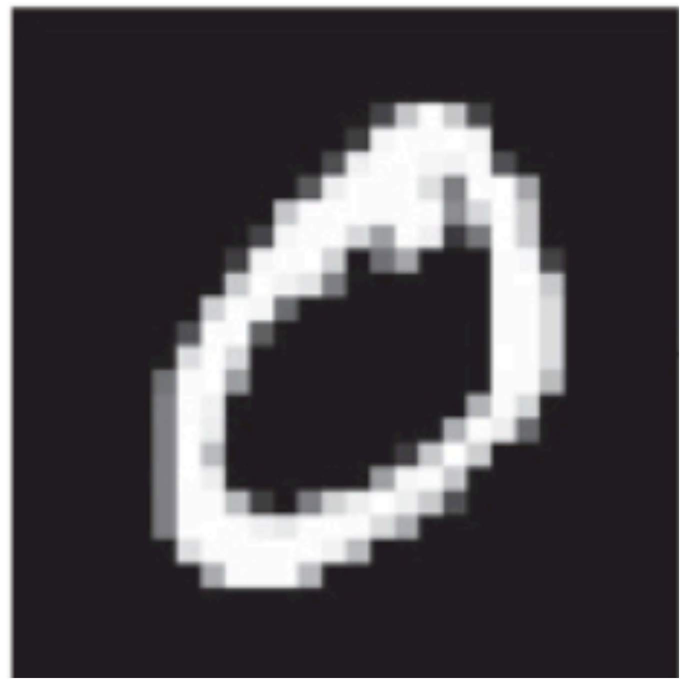
Convolution: Searching for patterns

Filter (3x3)

(pattern)

0	1	2
2	2	0
0	1	2

Input image



3_0	3_1	2_2	1	0
0_2	0_2	1_0	3	1
3_0	1_1	2_2	2	3
2	0	0	2	2
2	0	0	0	1

Activation map

(presence of pattern)

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

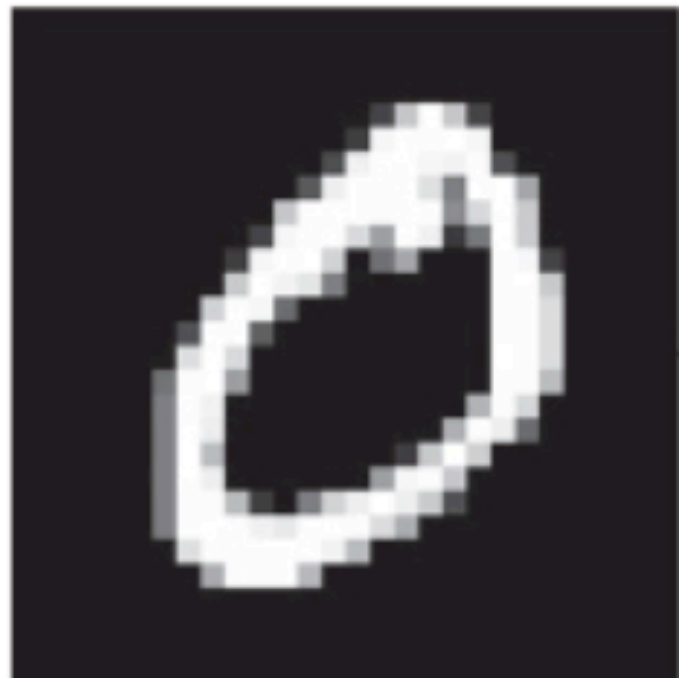
Convolution: Searching for patterns

Filter (3x3)

(pattern)

0	1	2
2	2	0
0	1	2

Input image



3_0	3_1	2_2	1	0
0_2	0_2	1_0	3	1
3_0	1_1	2_2	2	3
2	0	0	2	2
2	0	0	0	1

Activation map

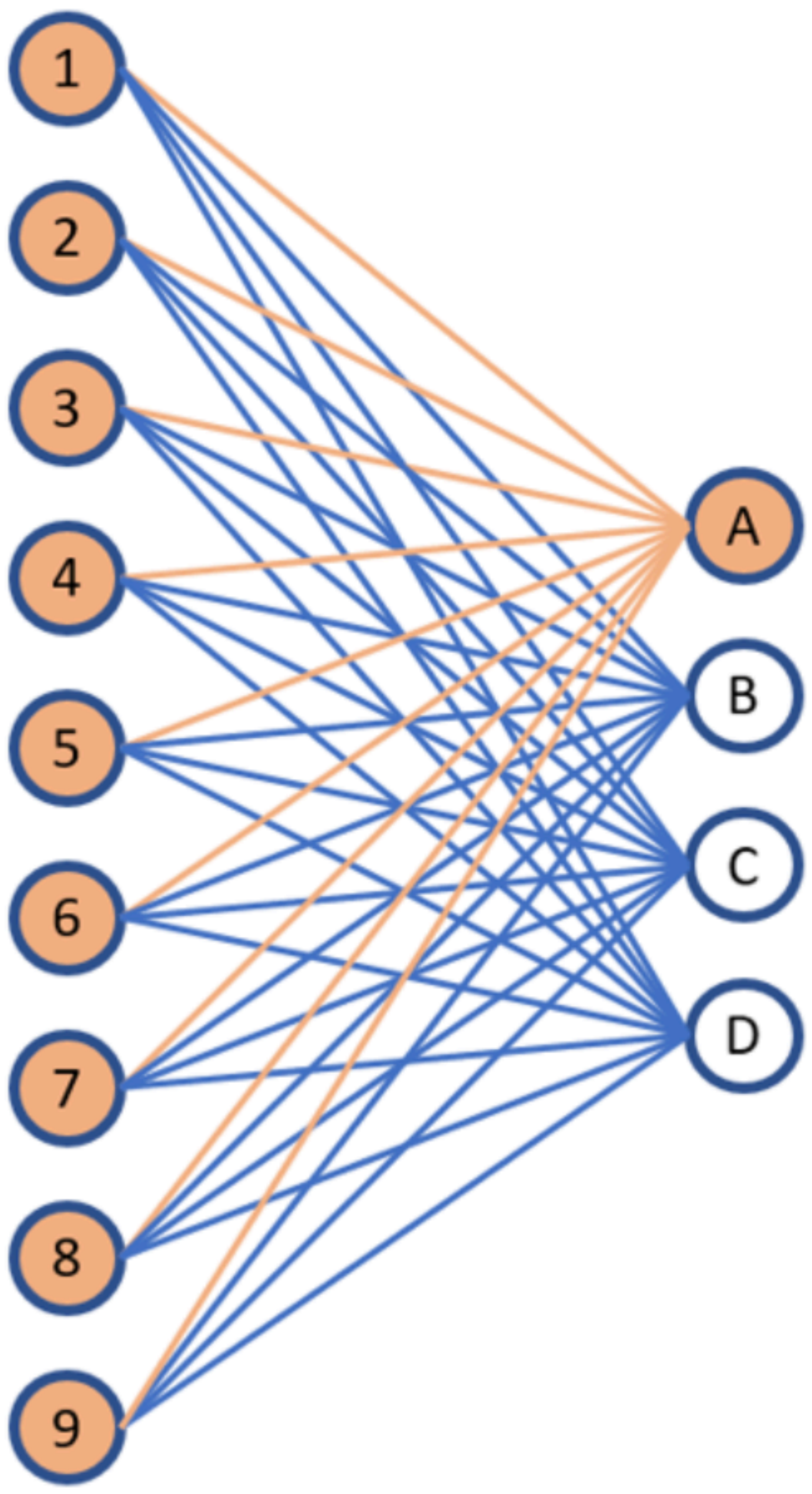
(presence of pattern)

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

We want to use many filters, each sensitive to a different kind of pattern.

Convolutional layer versus fully-connected layer

A convolutional layer can be visualized similarly to a fully-connected layer.

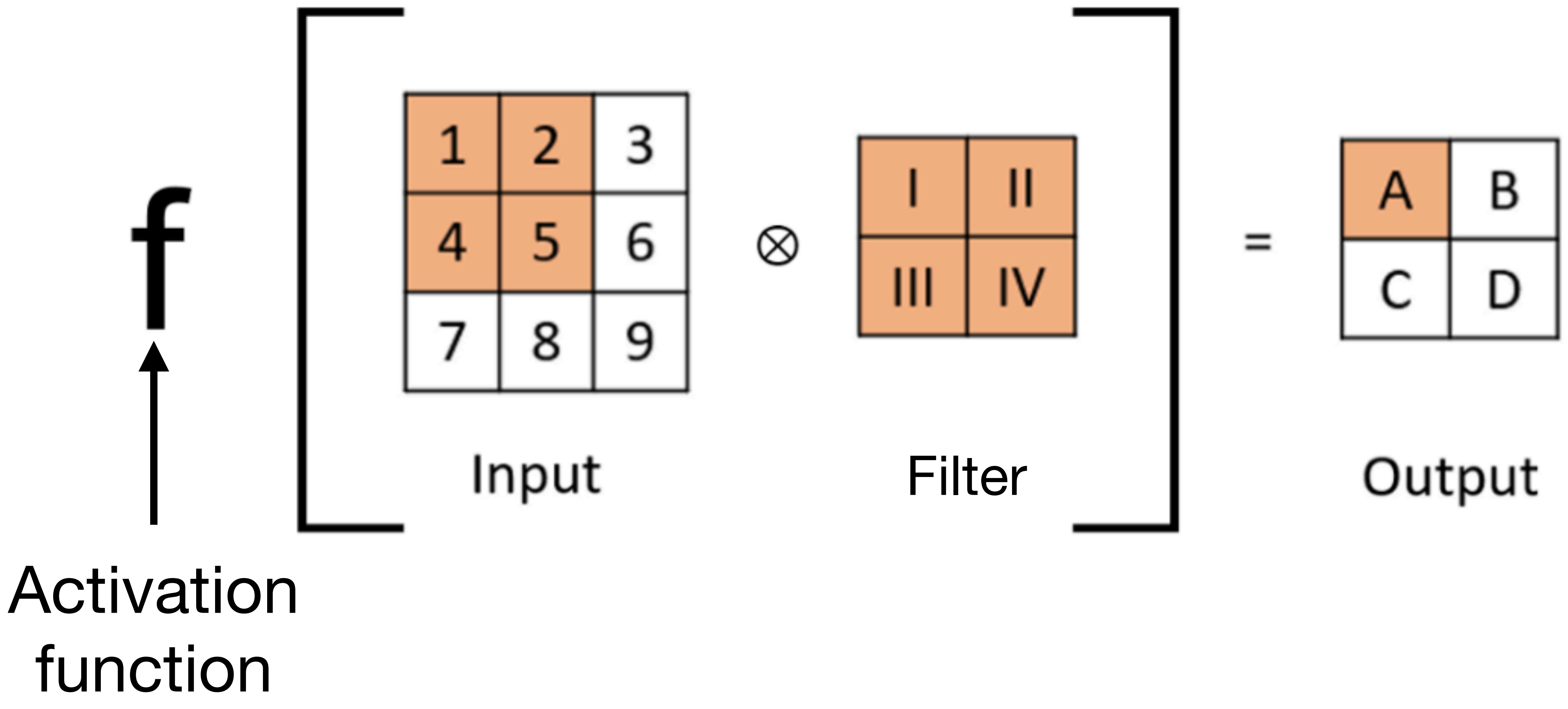


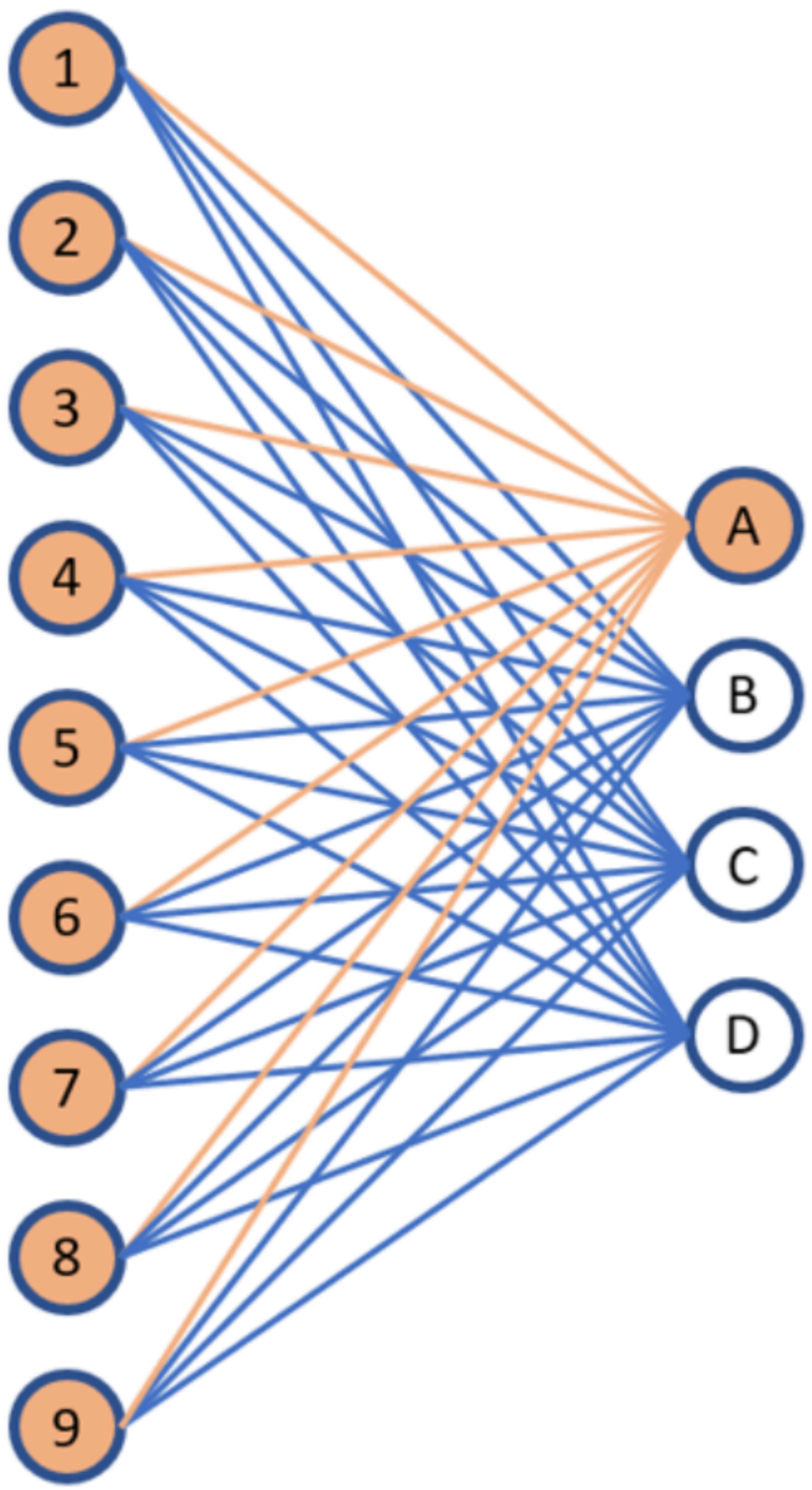
Fully-connected layer

Images from: https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b

Convolutional layer versus fully-connected layer

A convolutional layer can be visualized similarly to a fully-connected layer.



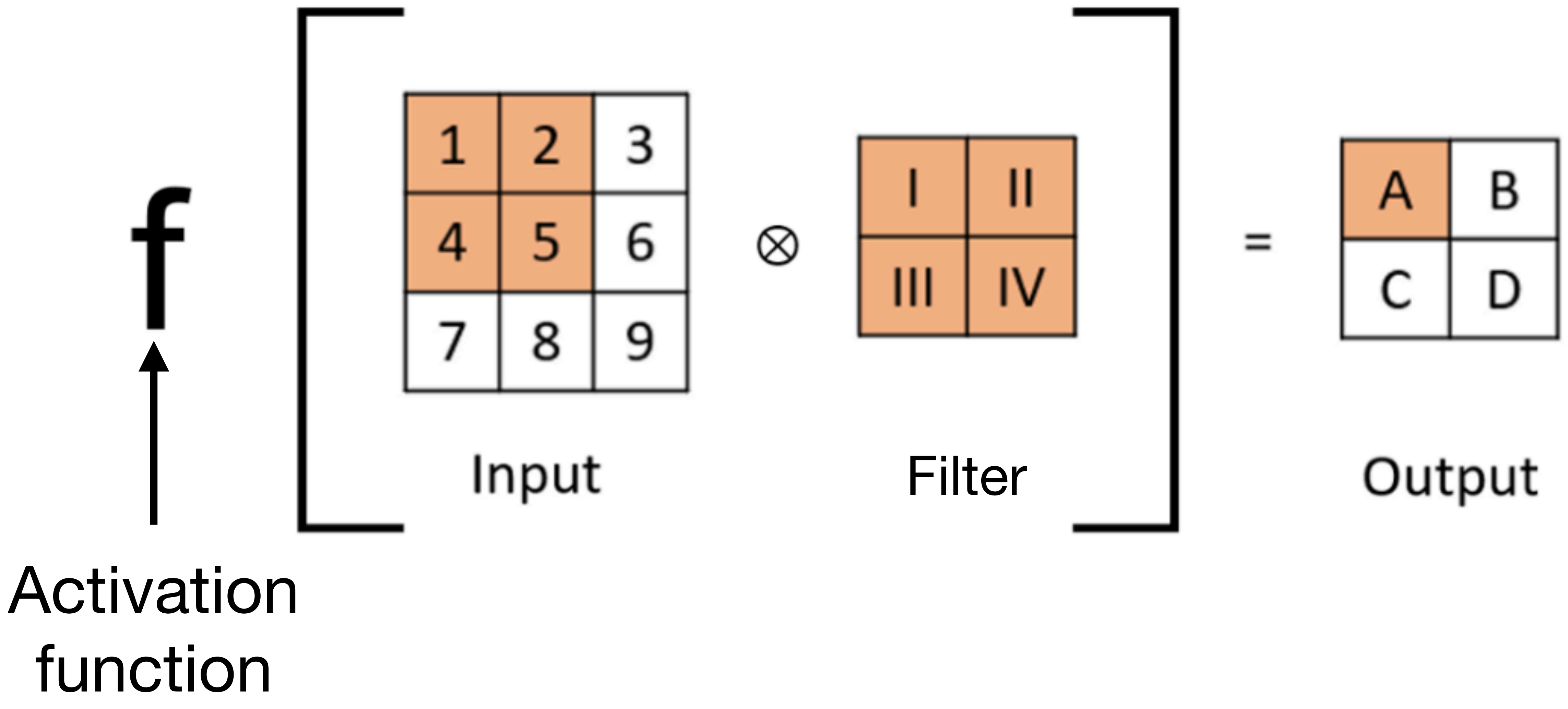


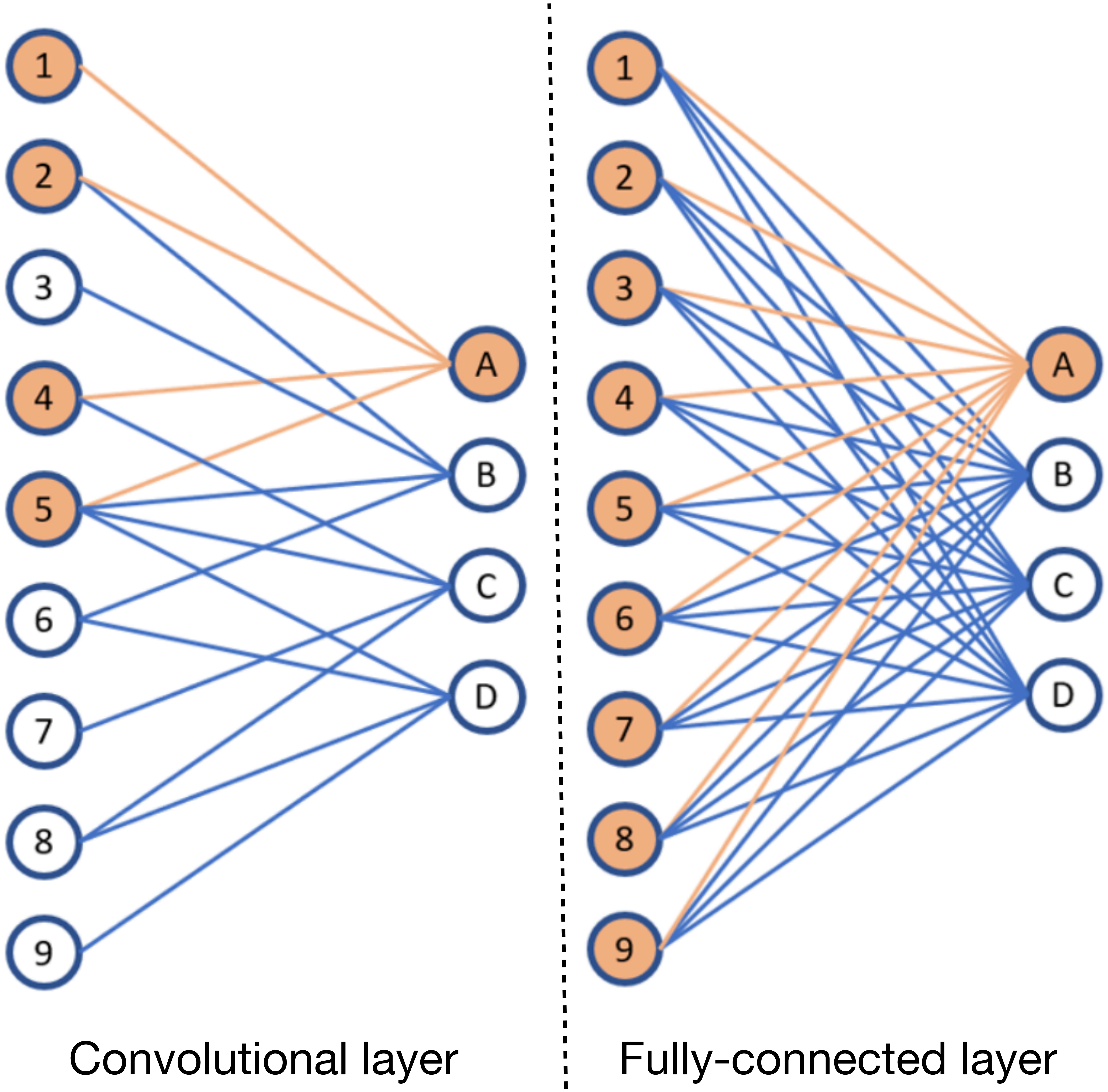
Fully-connected layer

Images from: https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b

Convolutional layer versus fully-connected layer

A convolutional layer can be visualized similarly to a fully-connected layer.

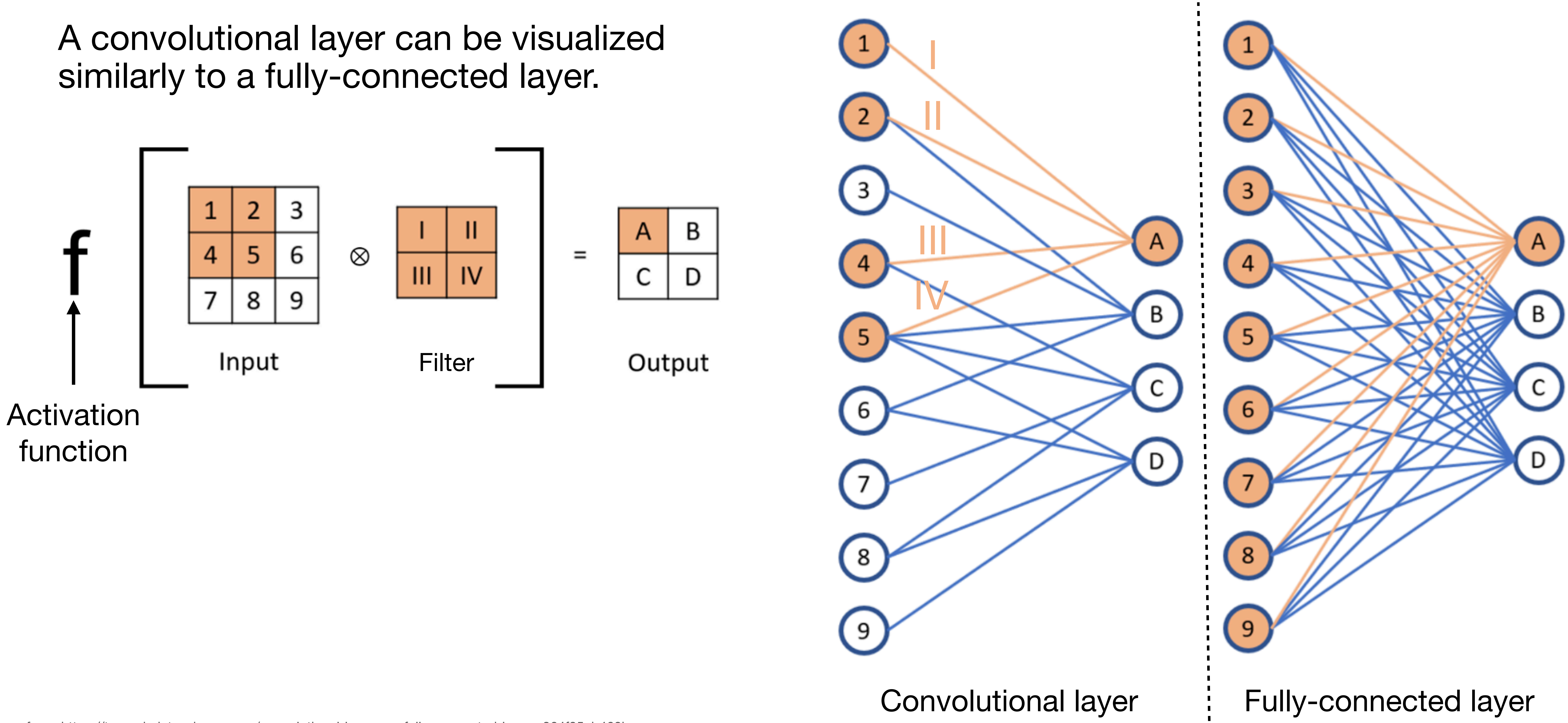




Images from: https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b

Convolutional layer versus fully-connected layer

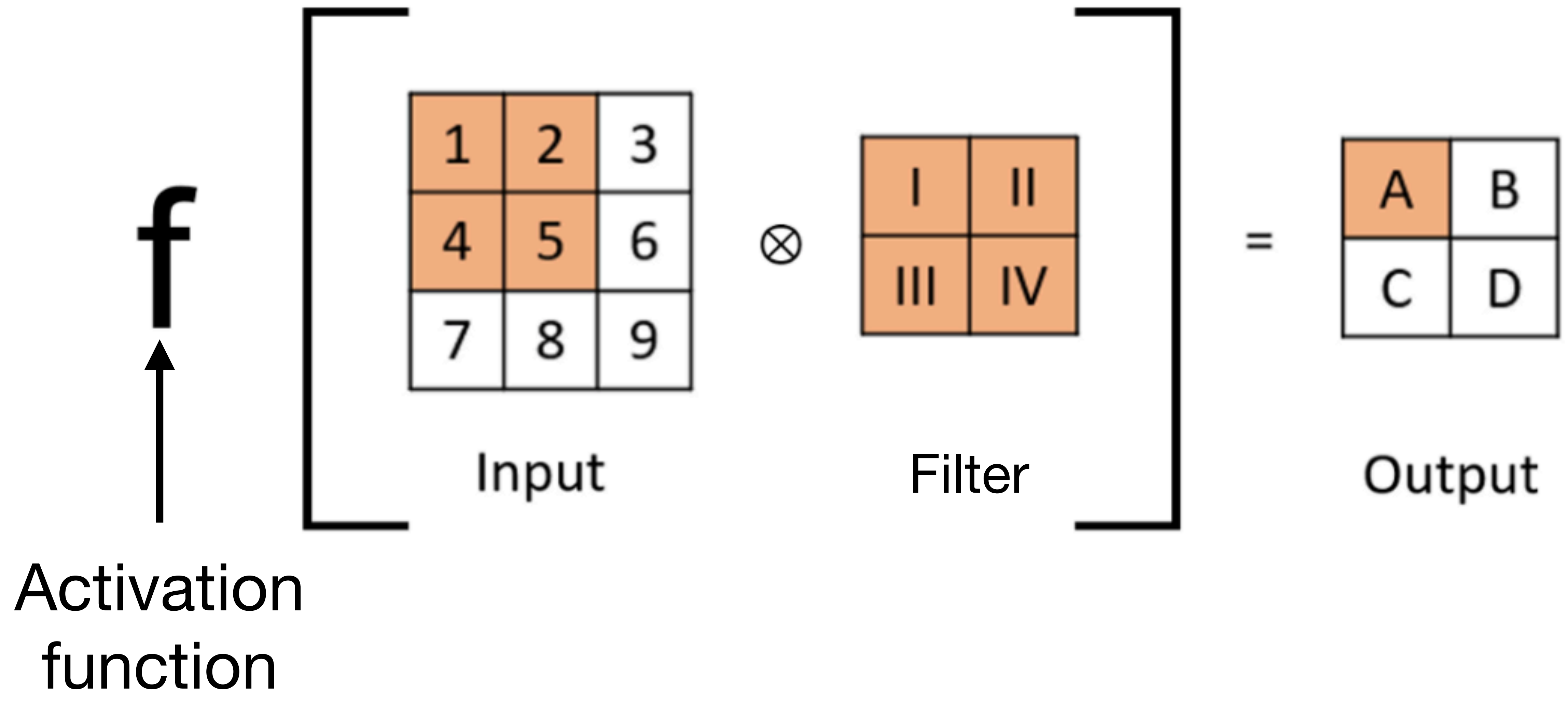
A convolutional layer can be visualized similarly to a fully-connected layer.



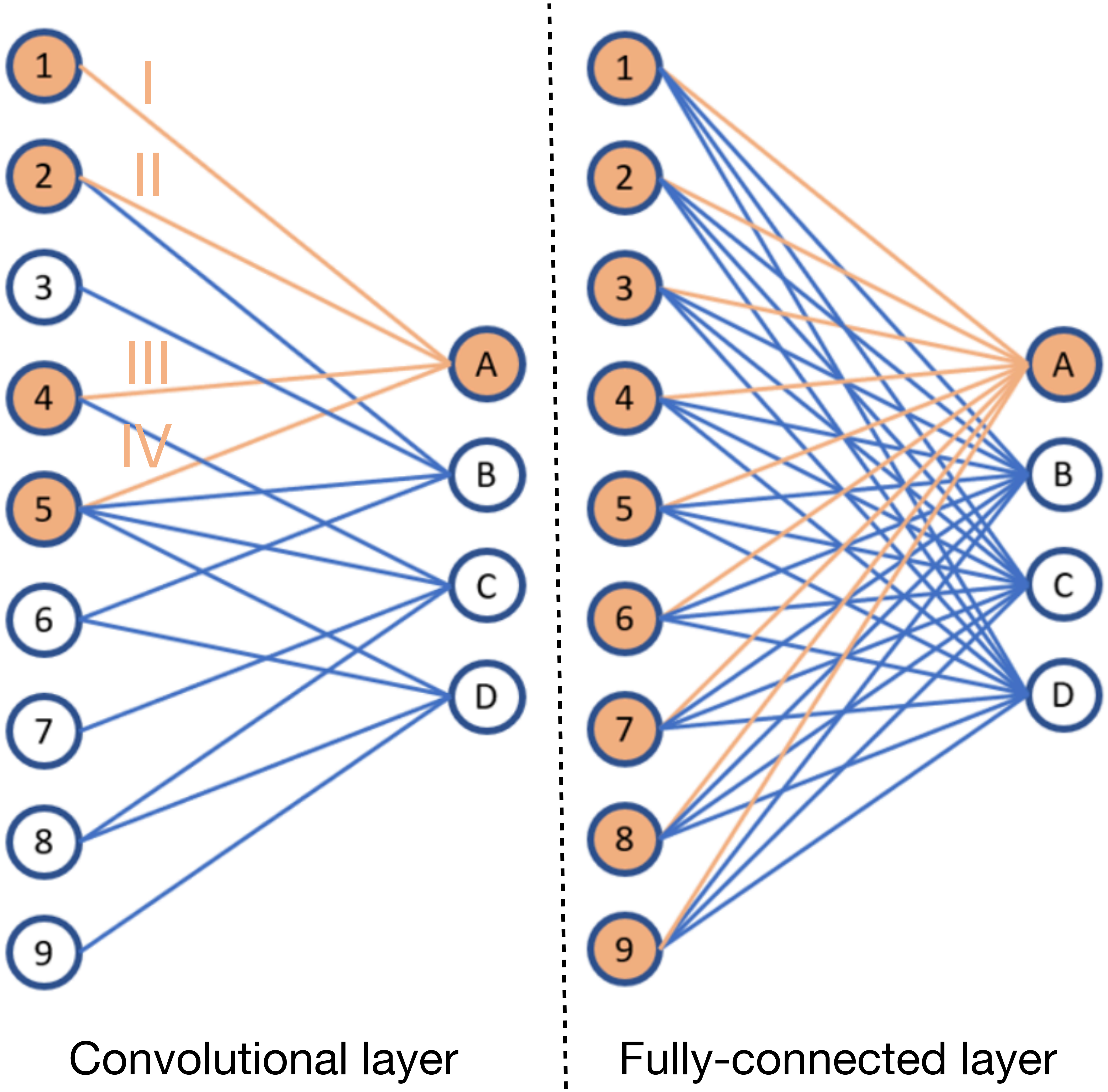
Images from: https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b

Convolutional layer versus fully-connected layer

A convolutional layer can be visualized similarly to a fully-connected layer.



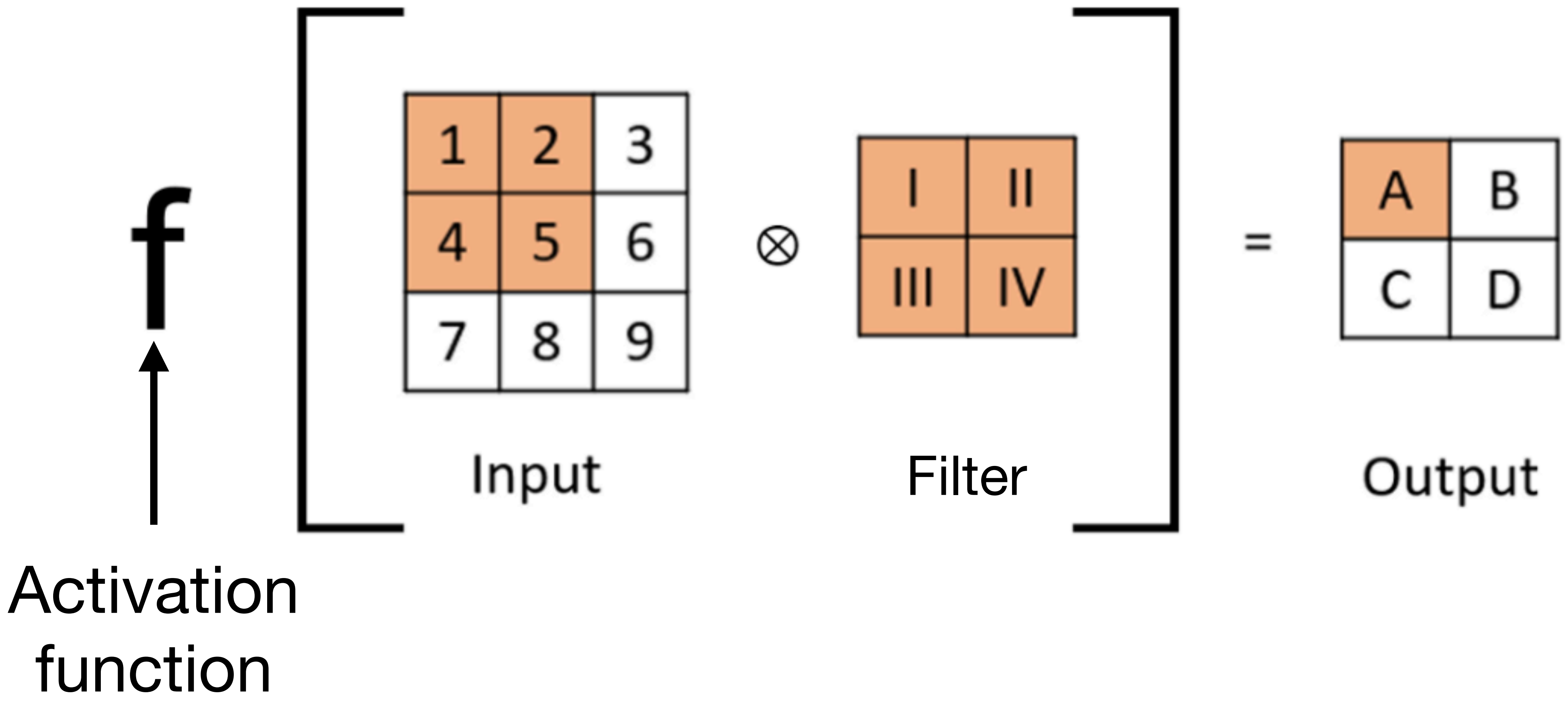
In a convolutional layer:



Images from: https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b

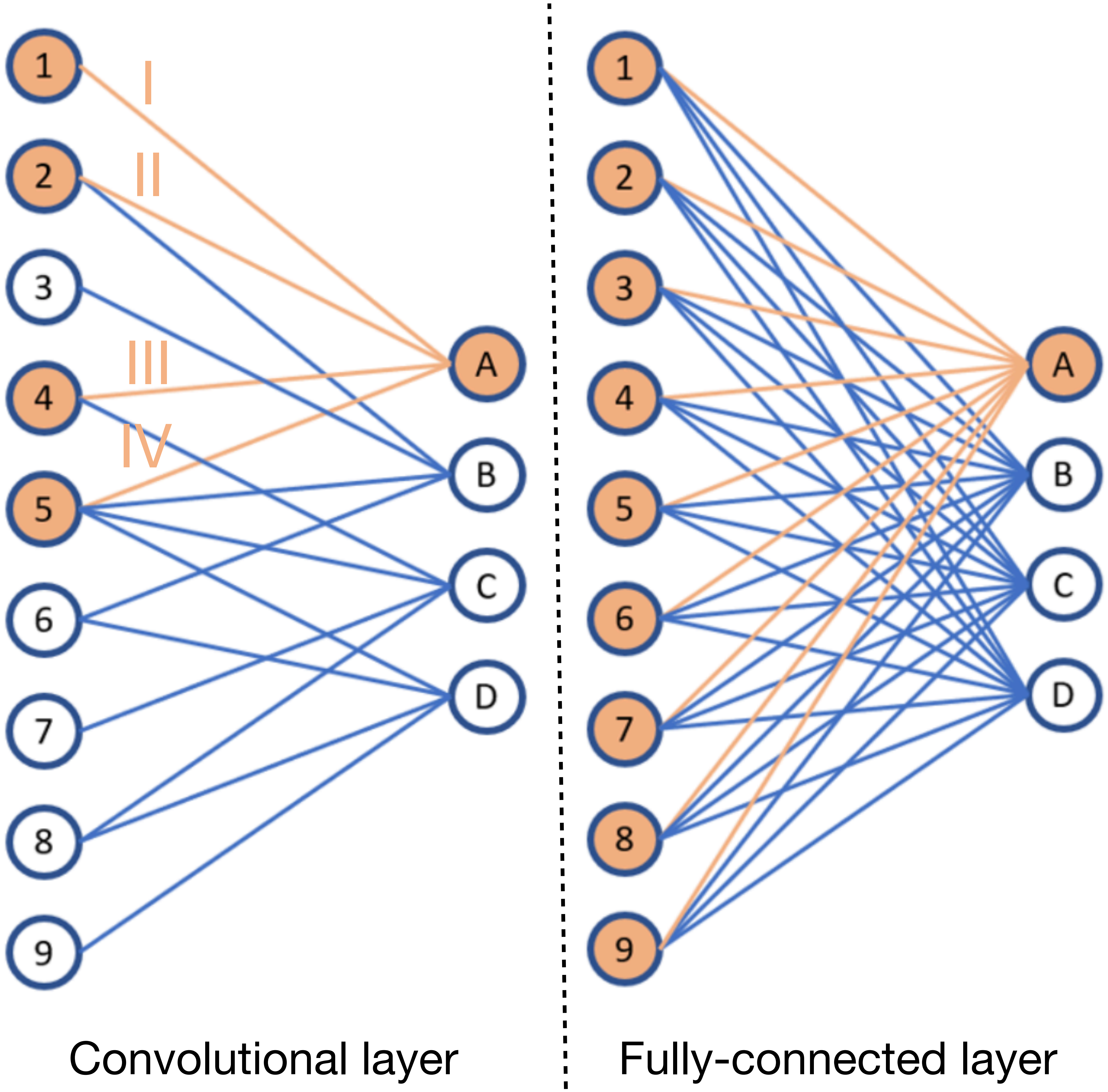
Convolutional layer versus fully-connected layer

A convolutional layer can be visualized similarly to a fully-connected layer.



In a convolutional layer:

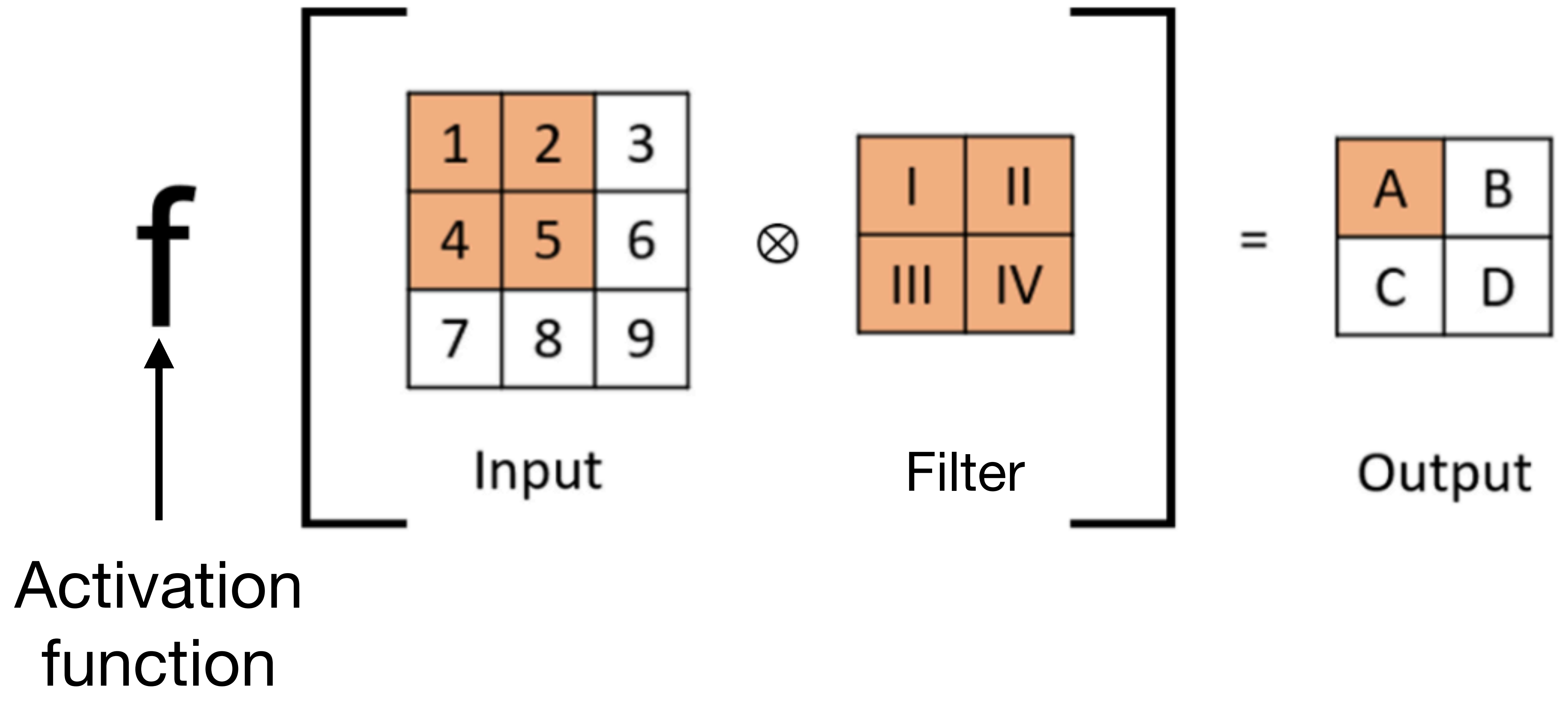
- Not all node pairs are connected with edges



Images from: https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b

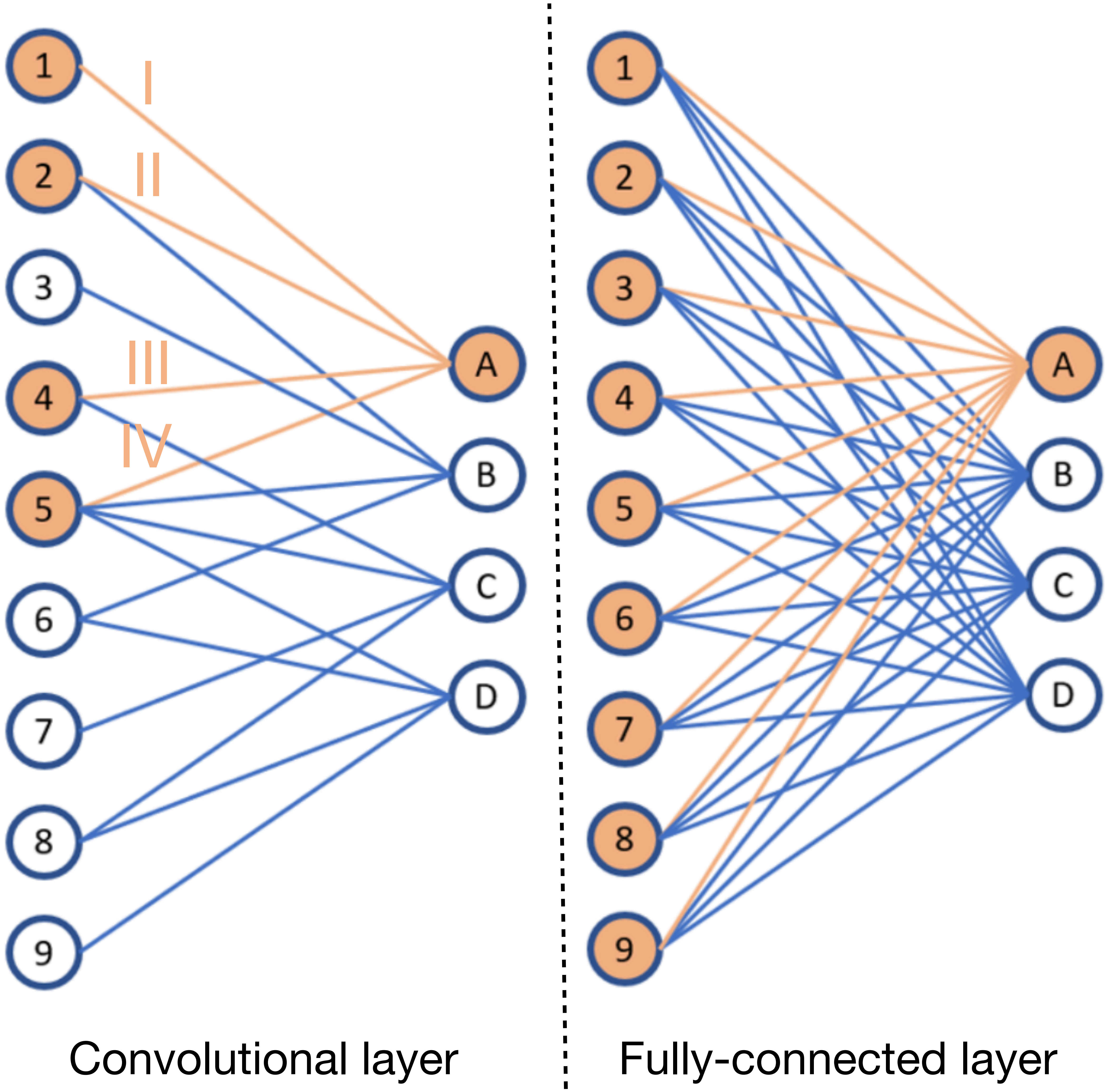
Convolutional layer versus fully-connected layer

A convolutional layer can be visualized similarly to a fully-connected layer.



In a convolutional layer:

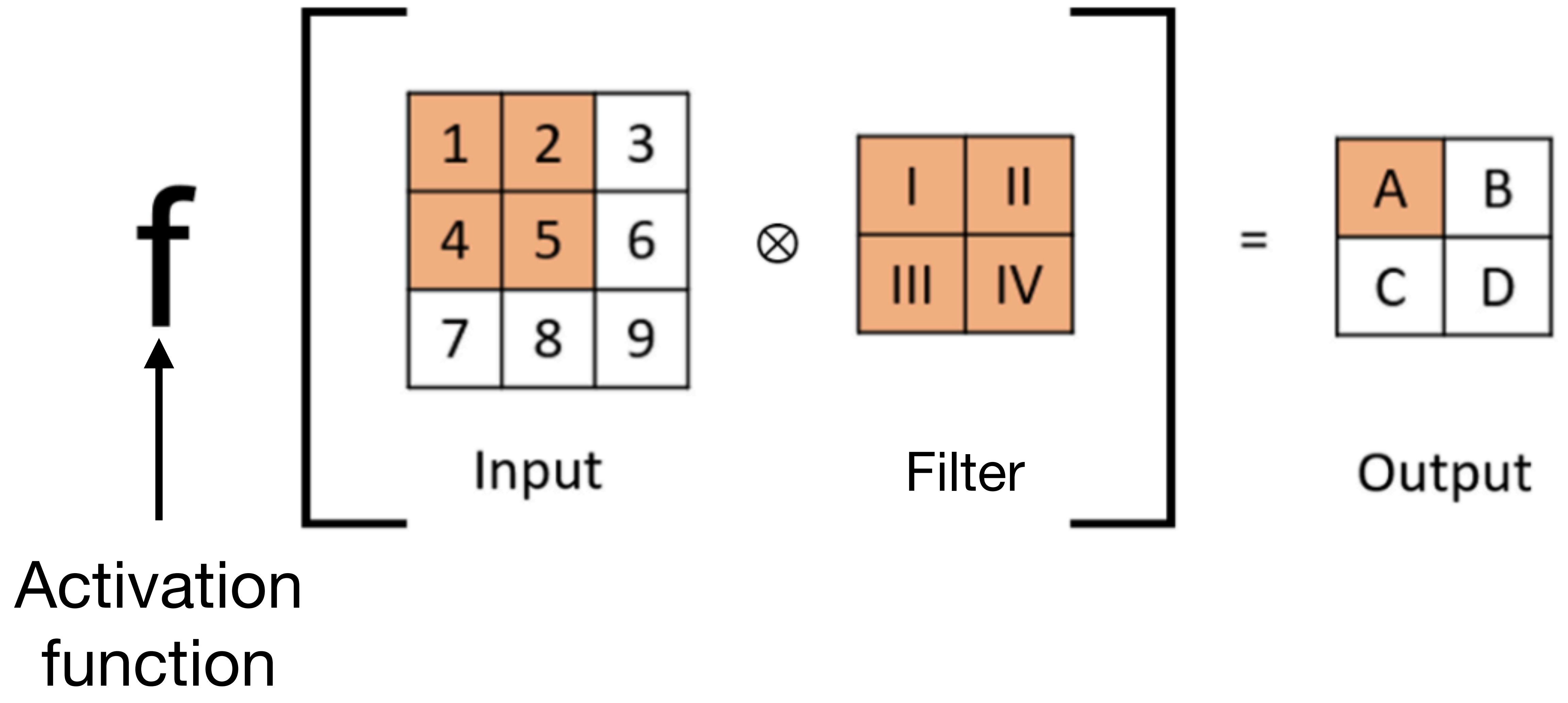
- Not all node pairs are connected with edges
- Weights (from filter) reused across edges



Images from: https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b

Convolutional layer versus fully-connected layer

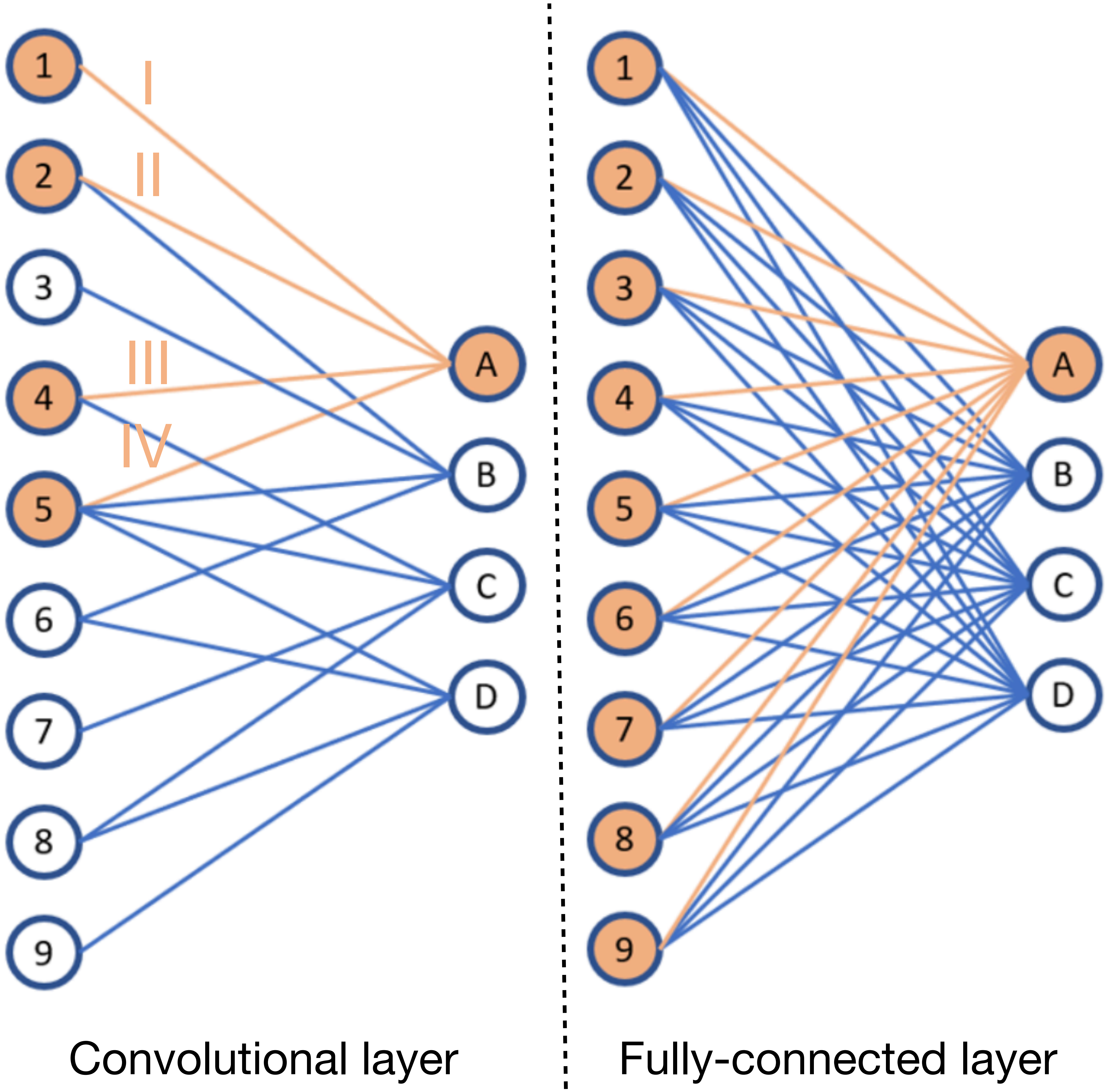
A convolutional layer can be visualized similarly to a fully-connected layer.



In a convolutional layer:

- Not all node pairs are connected with edges
- Weights (from filter) reused across edges

Consequence: Conv layers have fewer parameters!

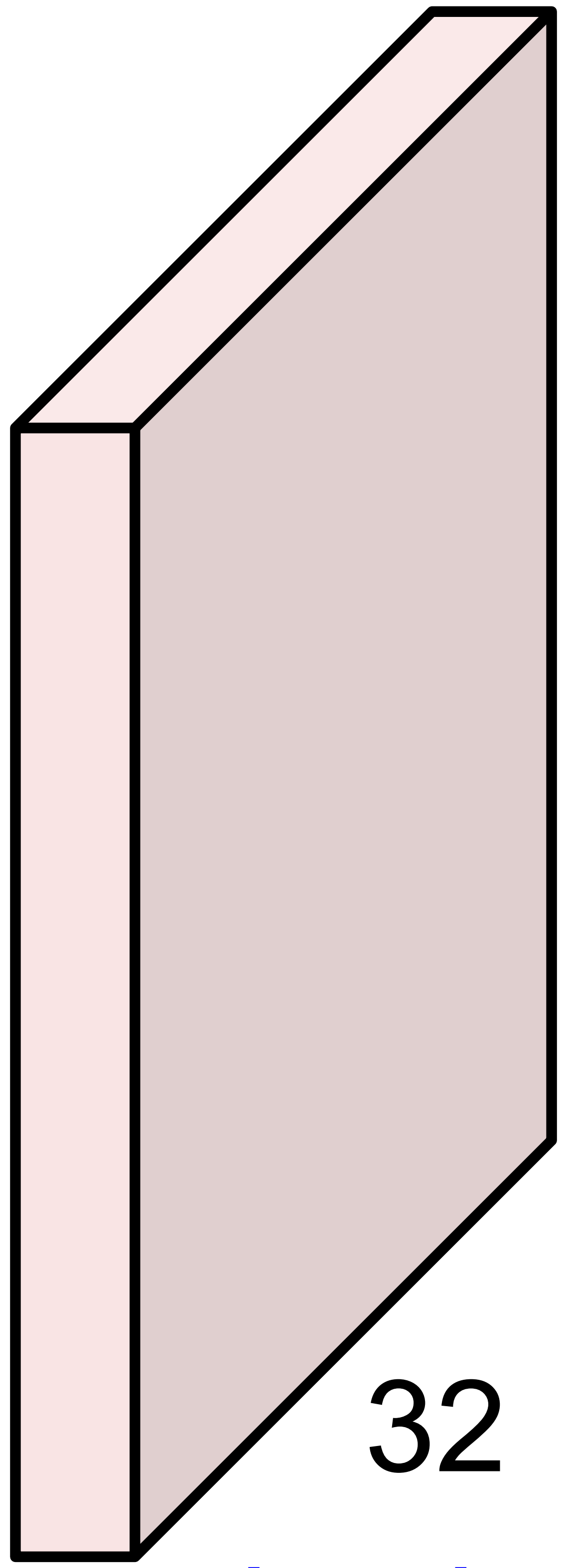


Images from: https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b

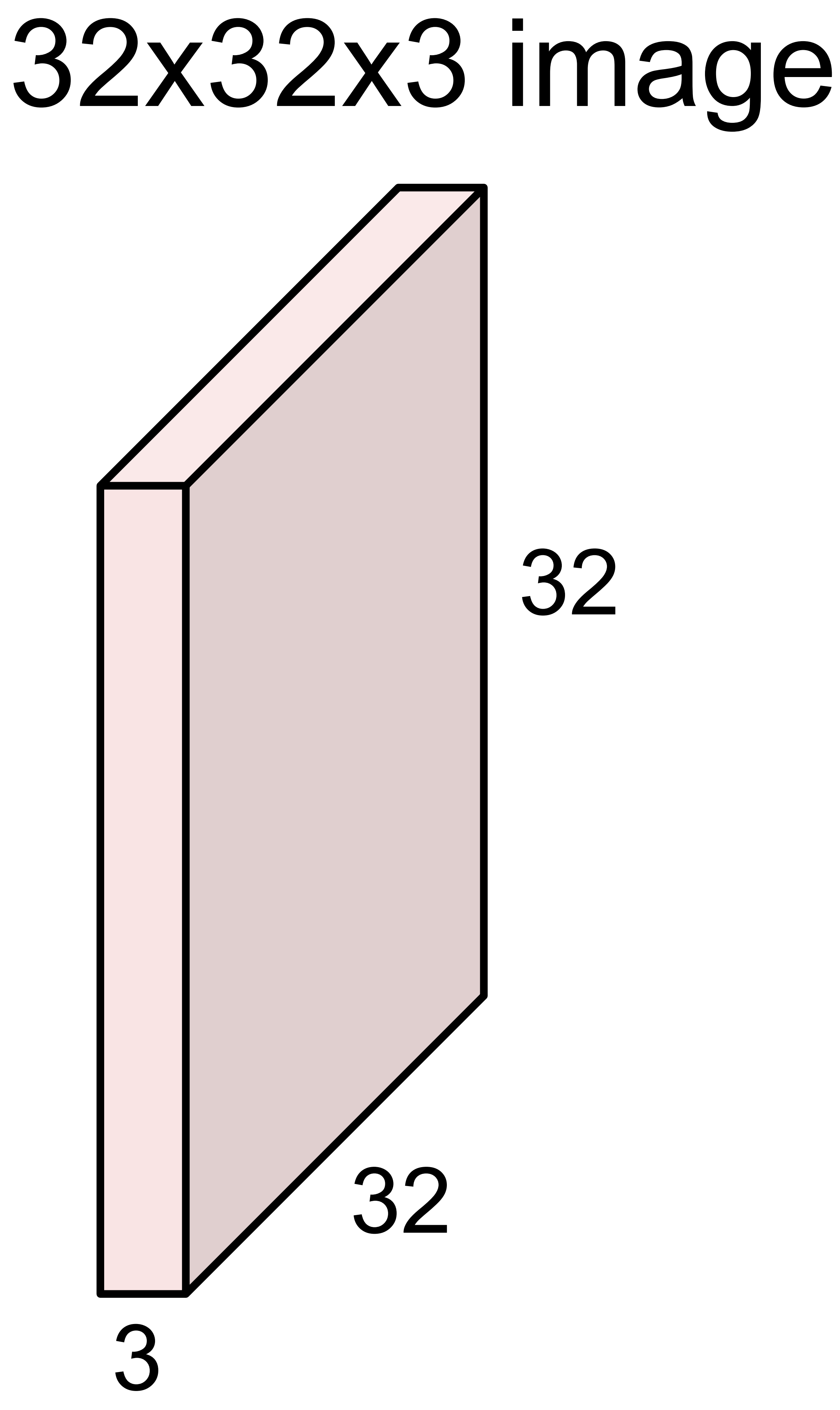
Convolution Layer

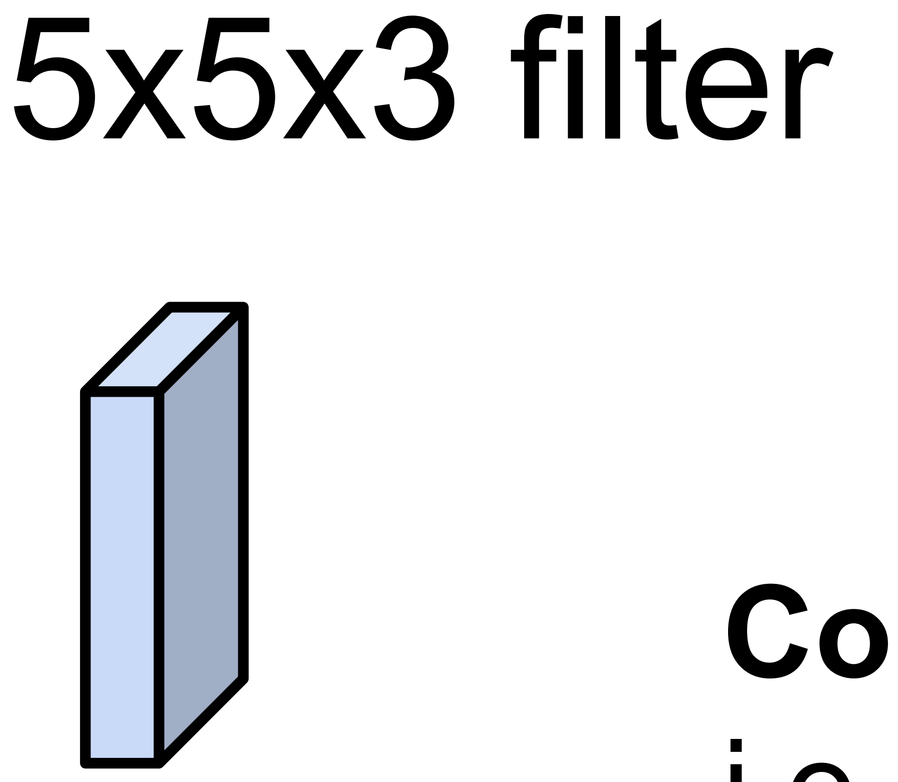
Note: This slide and several following ones are borrowed from Stanford's CS231n.

32x32x3 image (images typically have red, green, and blue channels.)



Convolution Layer



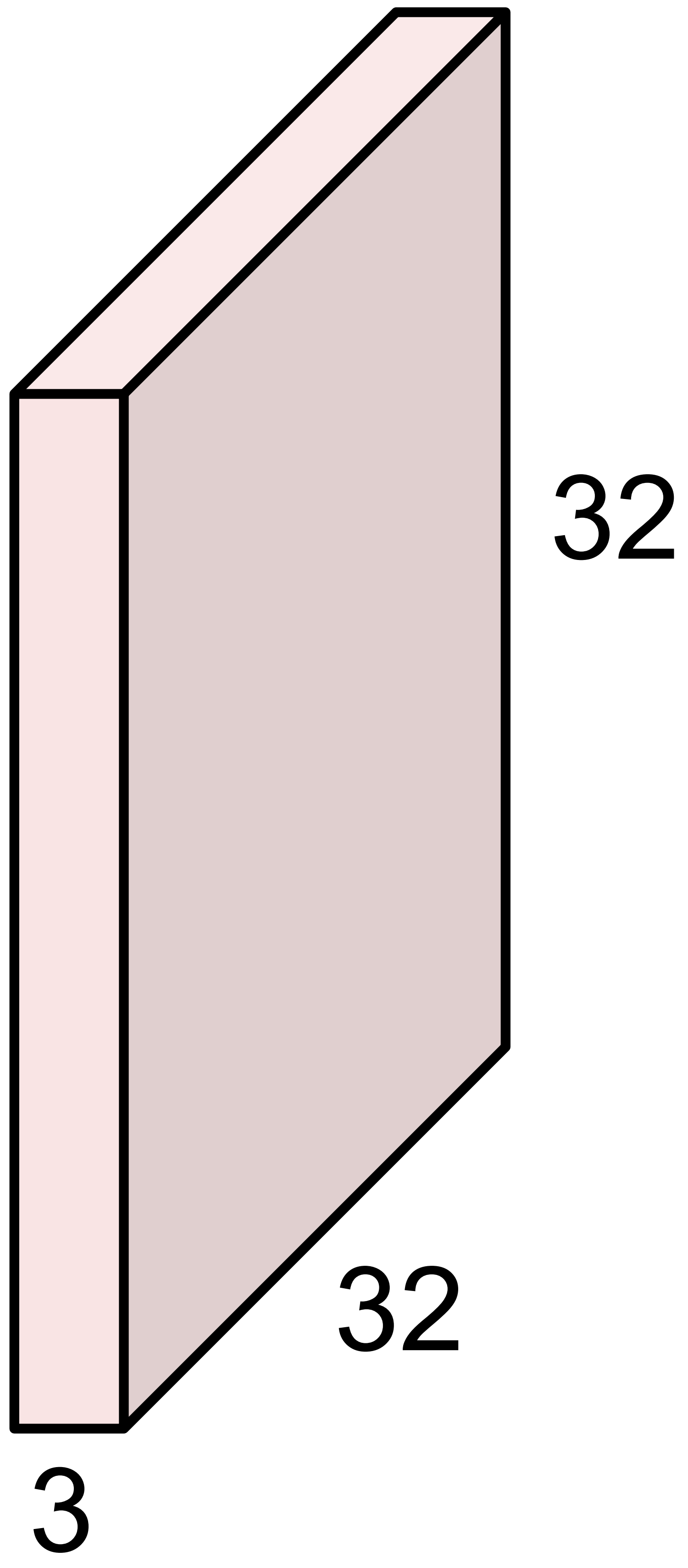


Convolve the filter with the image i.e. "slide over the image spatially, computing dot products"

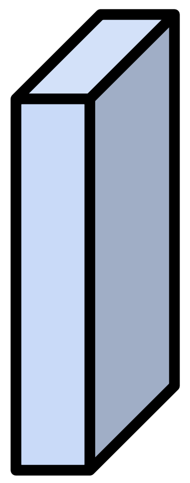
Convolution Layer

Filters always extend the full depth of the input volume

32x32x3 image

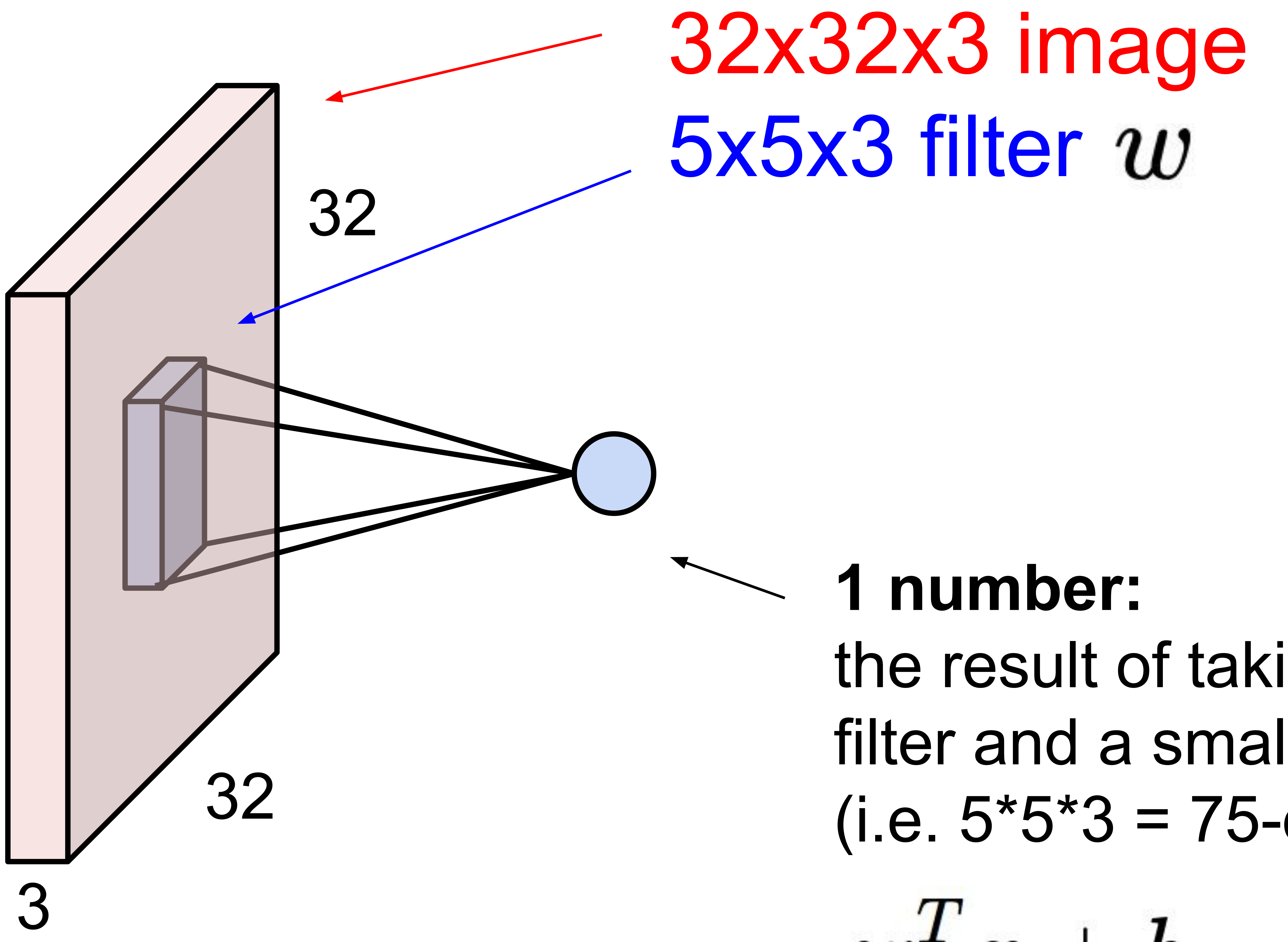


5x5x3 filter



Convolve the filter with the image i.e. “slide over the image spatially, computing dot products”

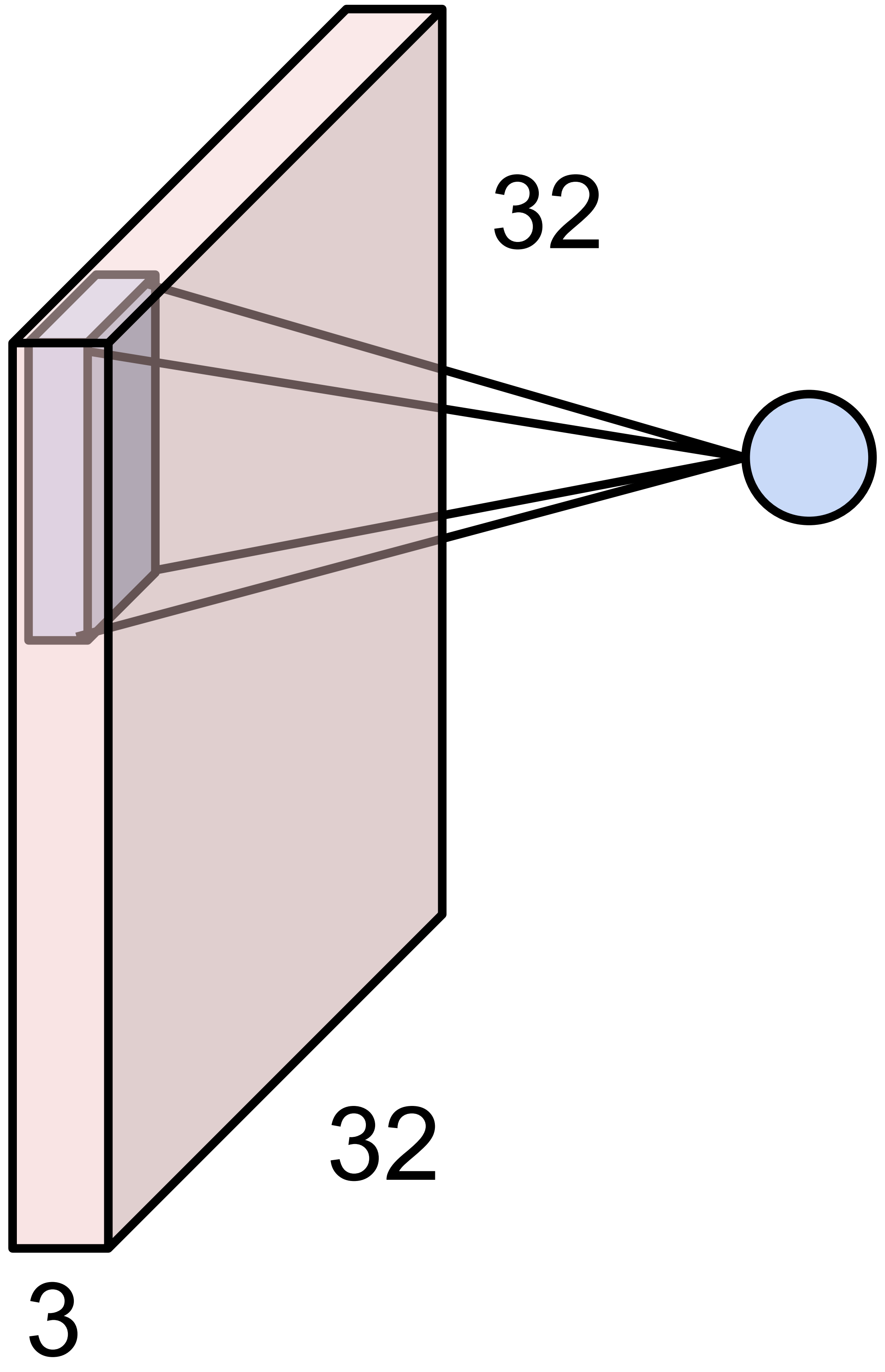
Convolution Layer



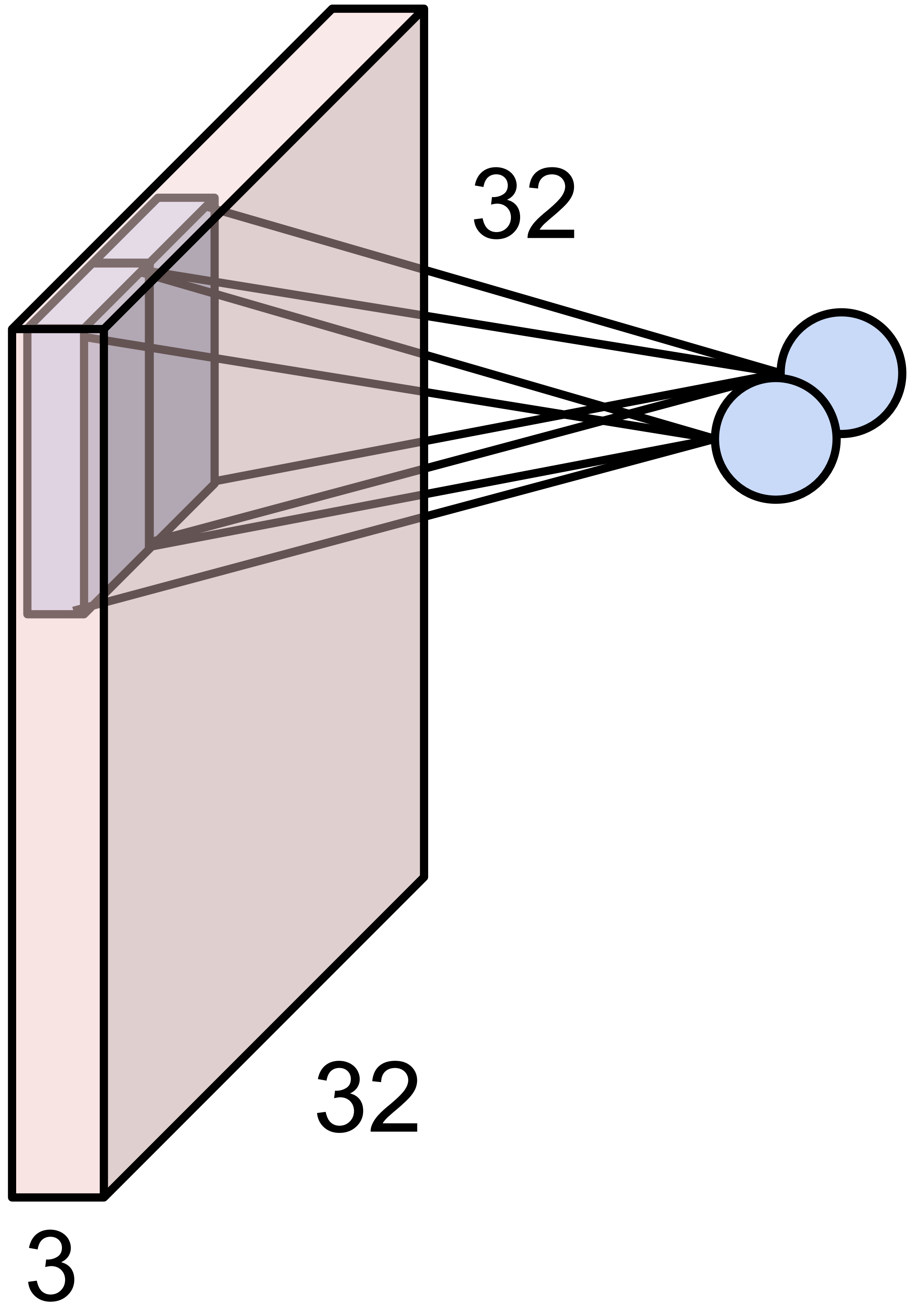
1 number:
the result of taking a dot product between the filter and a small 5x5x3 chunk of the image
(i.e. 5*5*3 = 75-dimensional dot product + bias)

$$w^T x + b$$

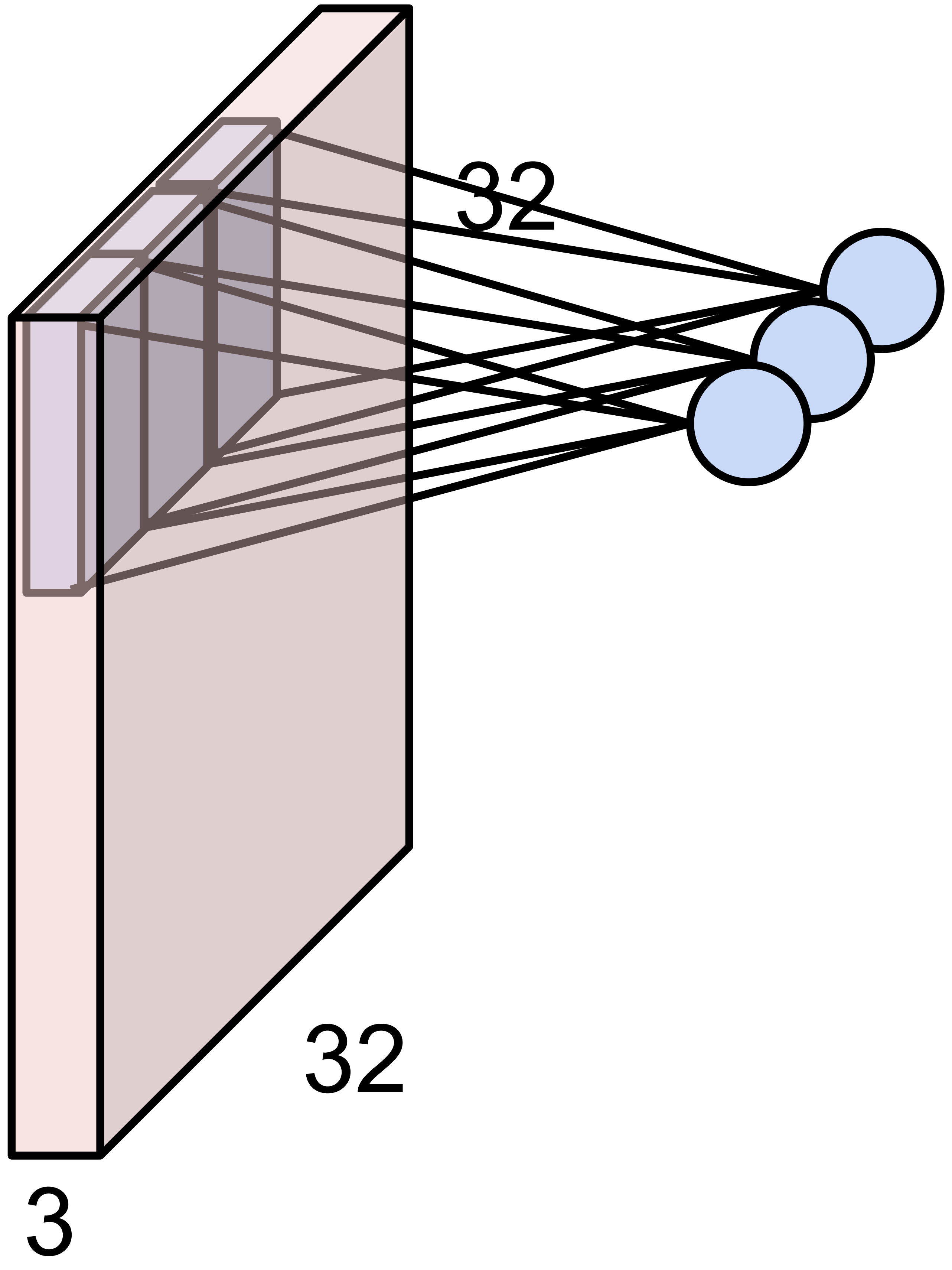
Convolution Layer



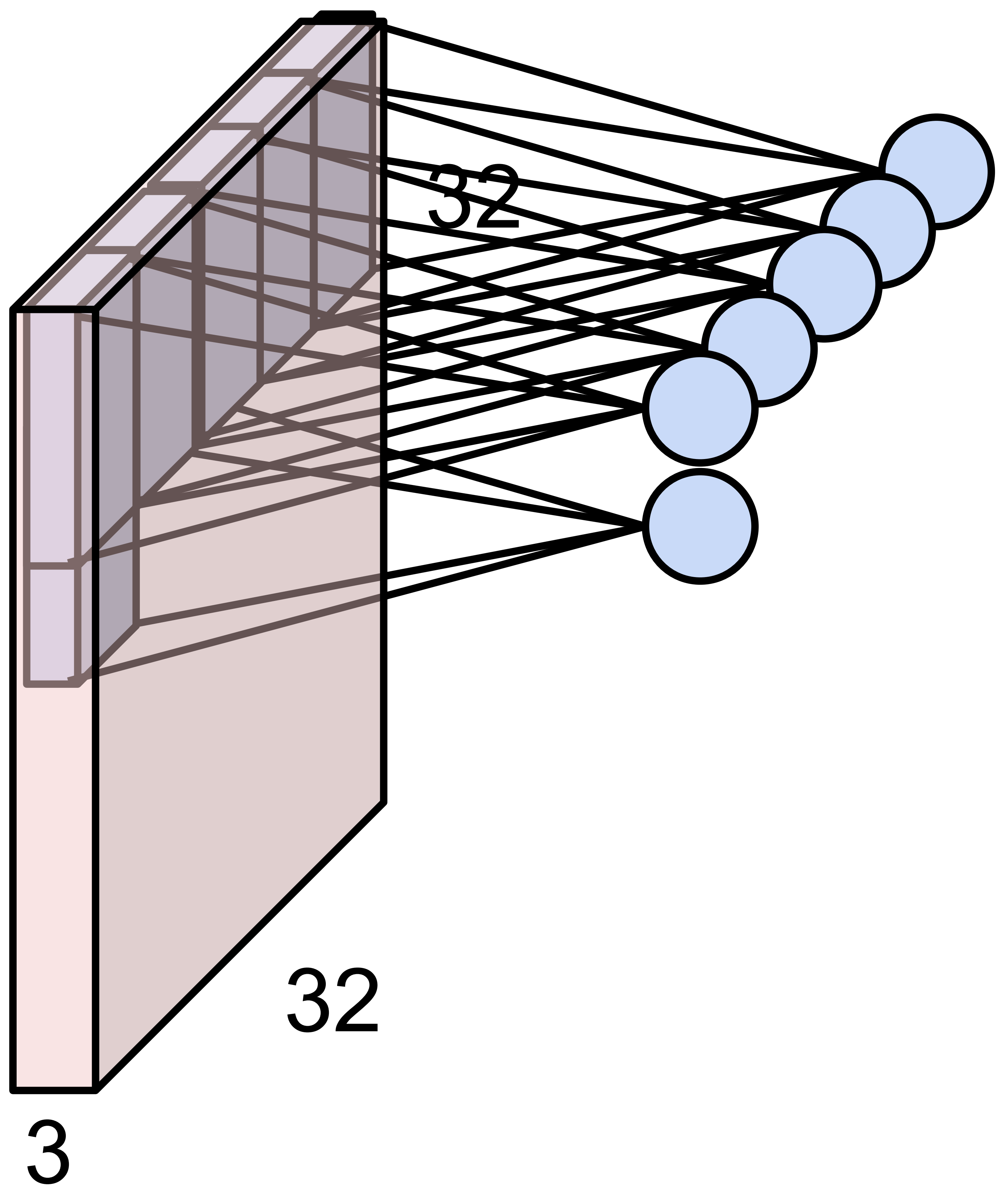
Convolution Layer



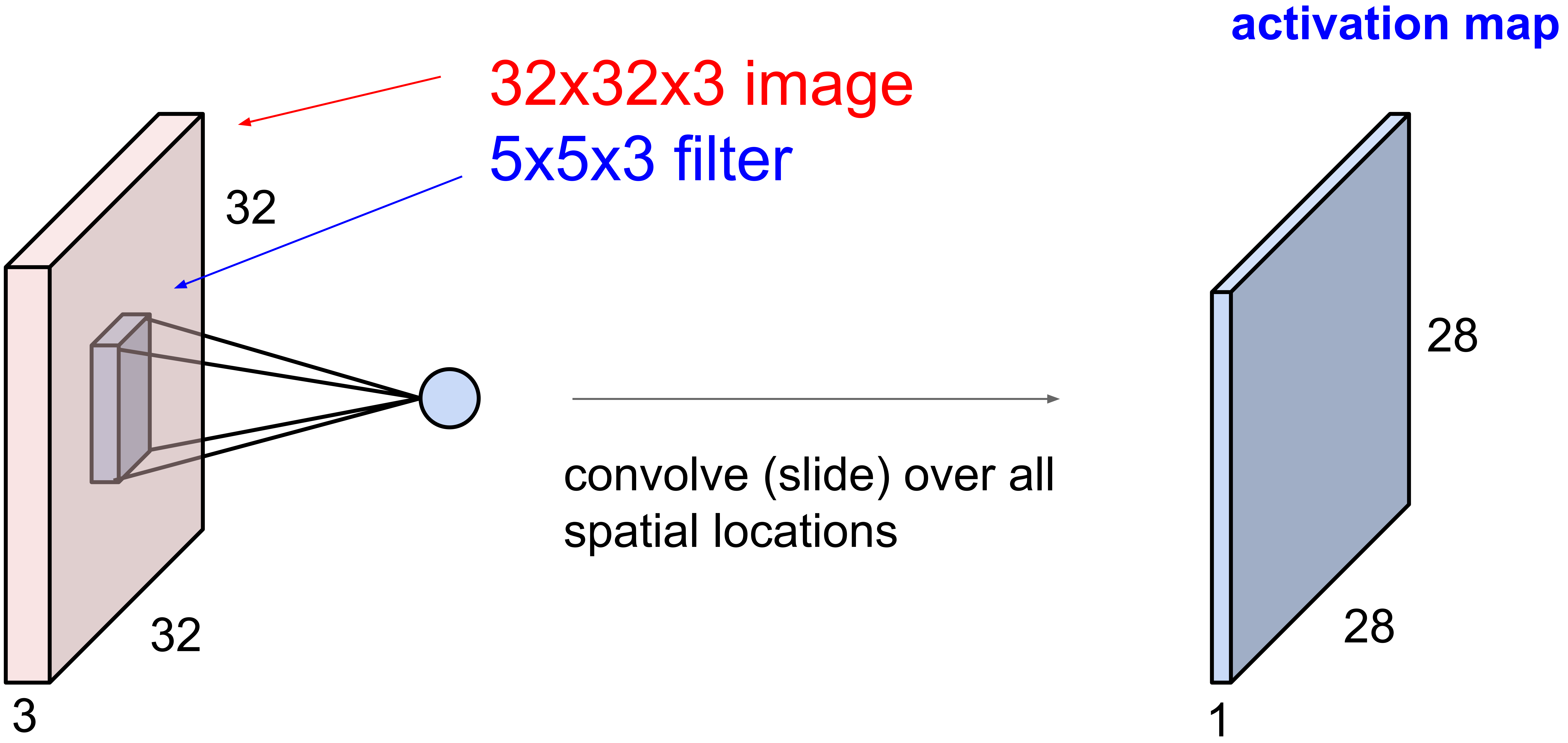
Convolution Layer



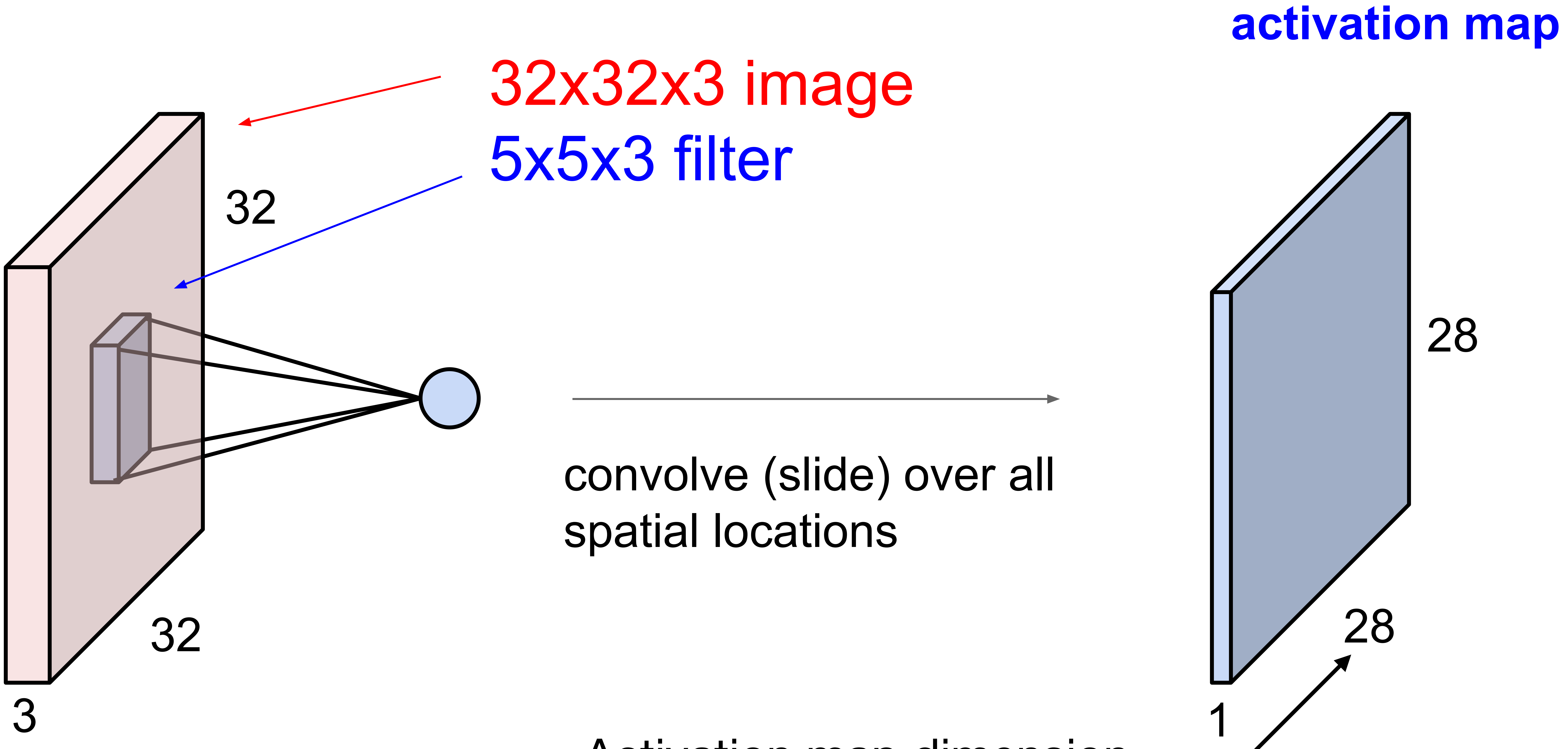
Convolution Layer



Convolution Layer



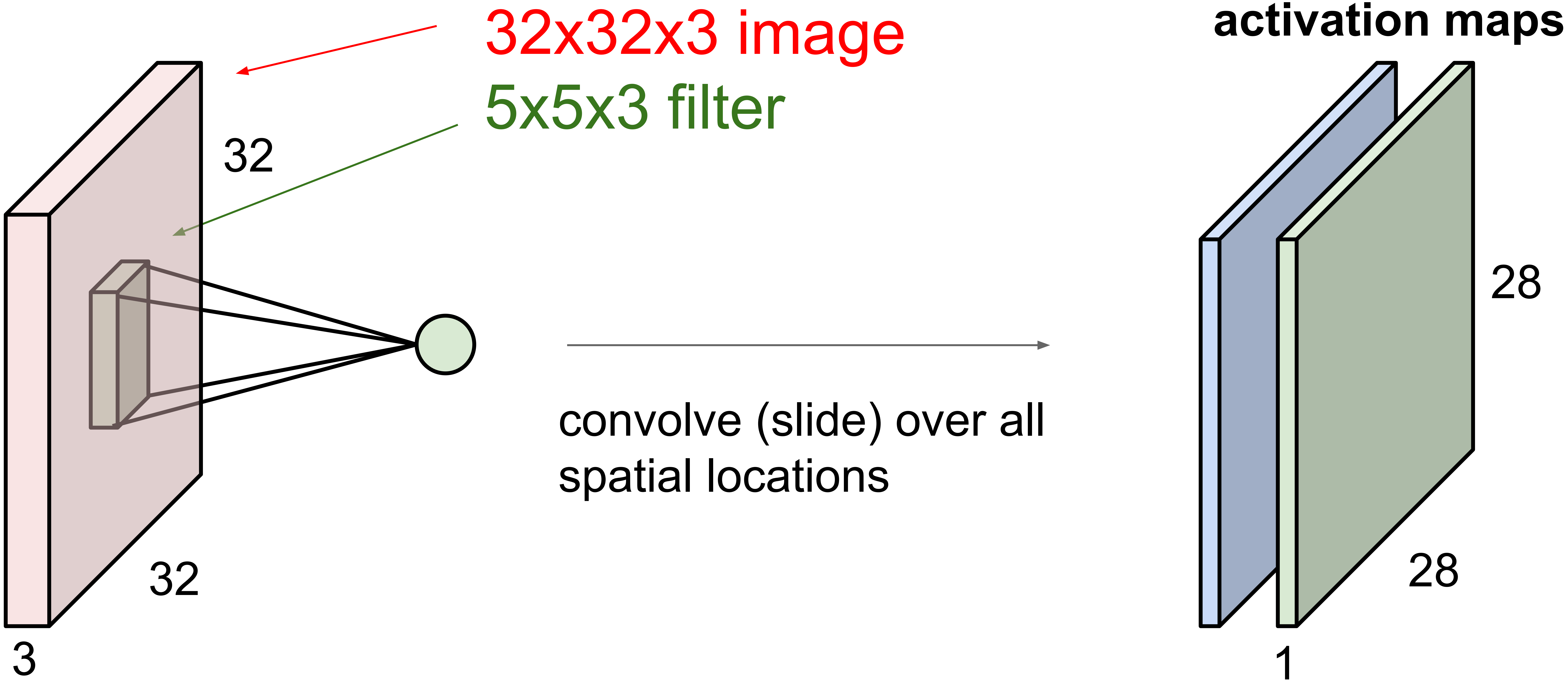
Convolution Layer



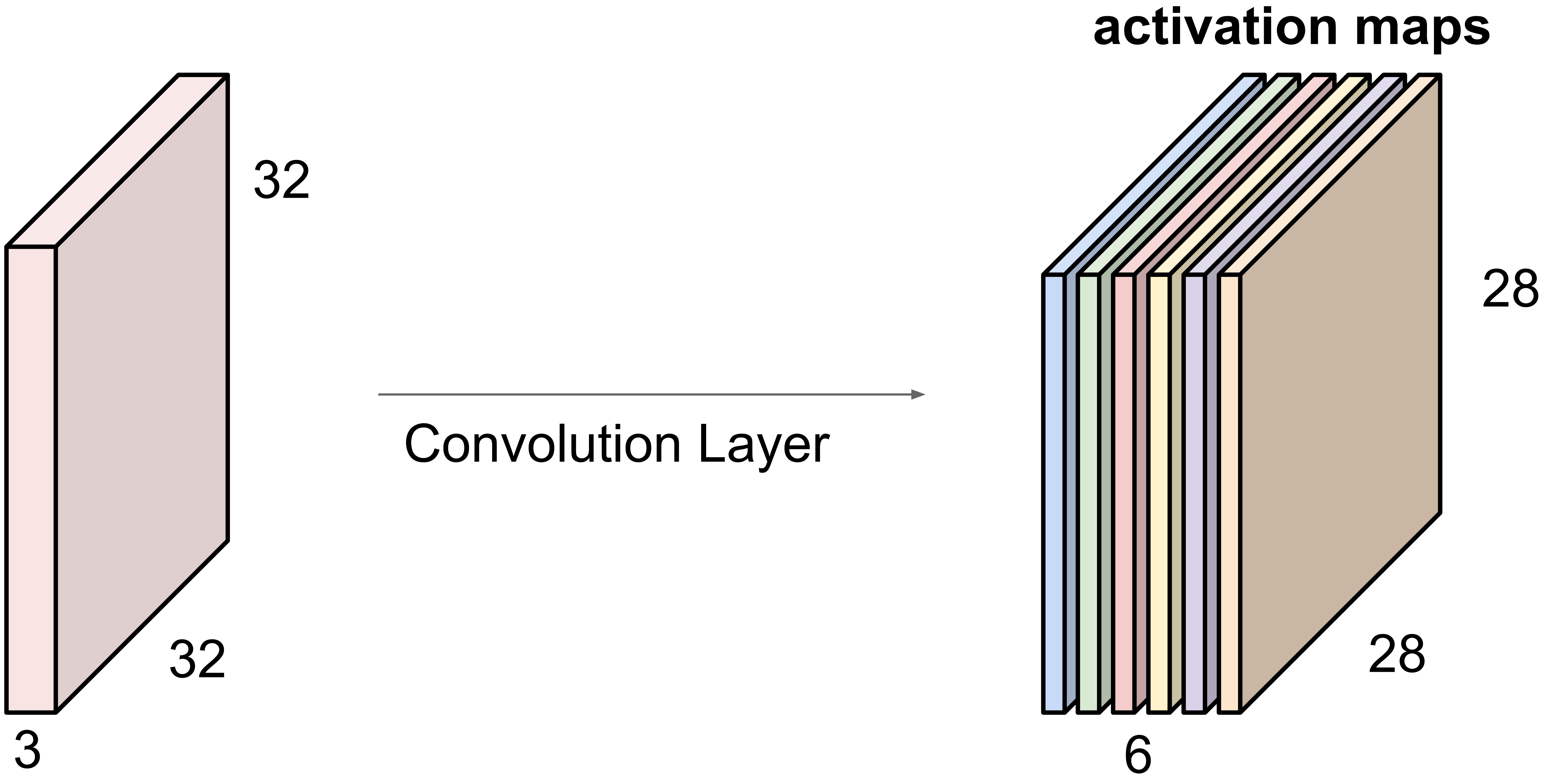
Activation map dimension =
Input image dimension - Filter dimension + 1

Convolution Layer

consider a second, green filter

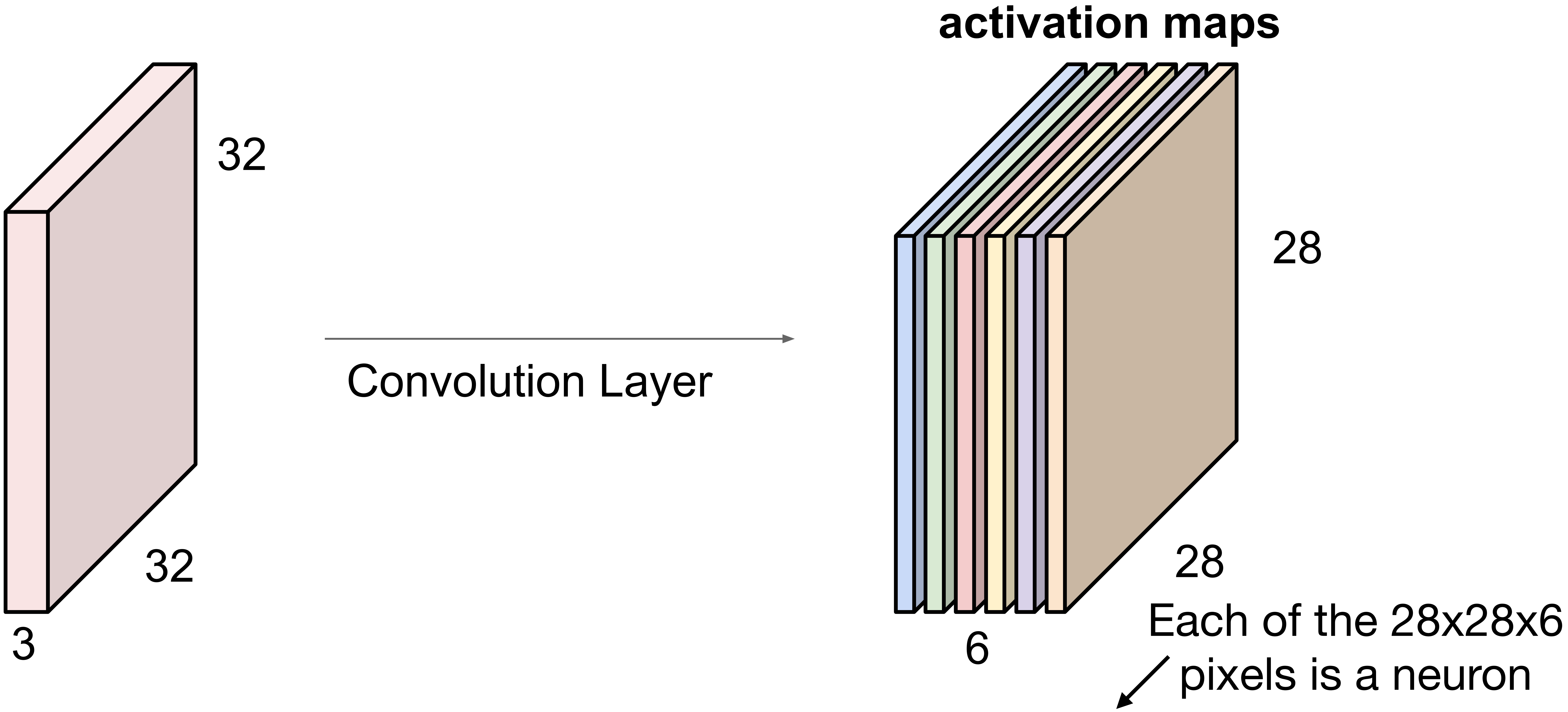


For example, if we had 6 5x5 filters, we'll get 6 separate activation maps:



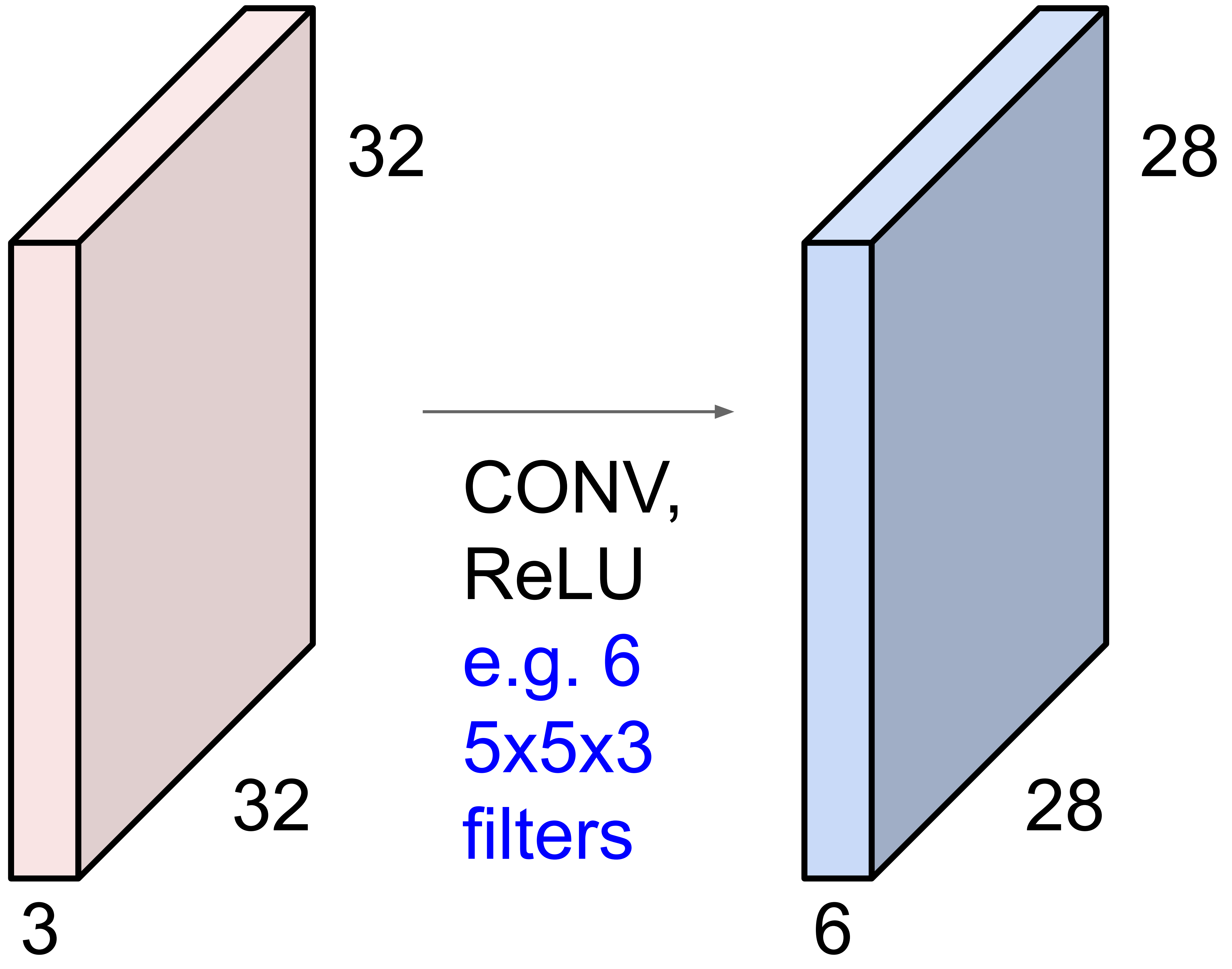
We stack these up to get a "new image" of size 28x28x6!

For example, if we had 6 5x5 filters, we'll get 6 separate activation maps:

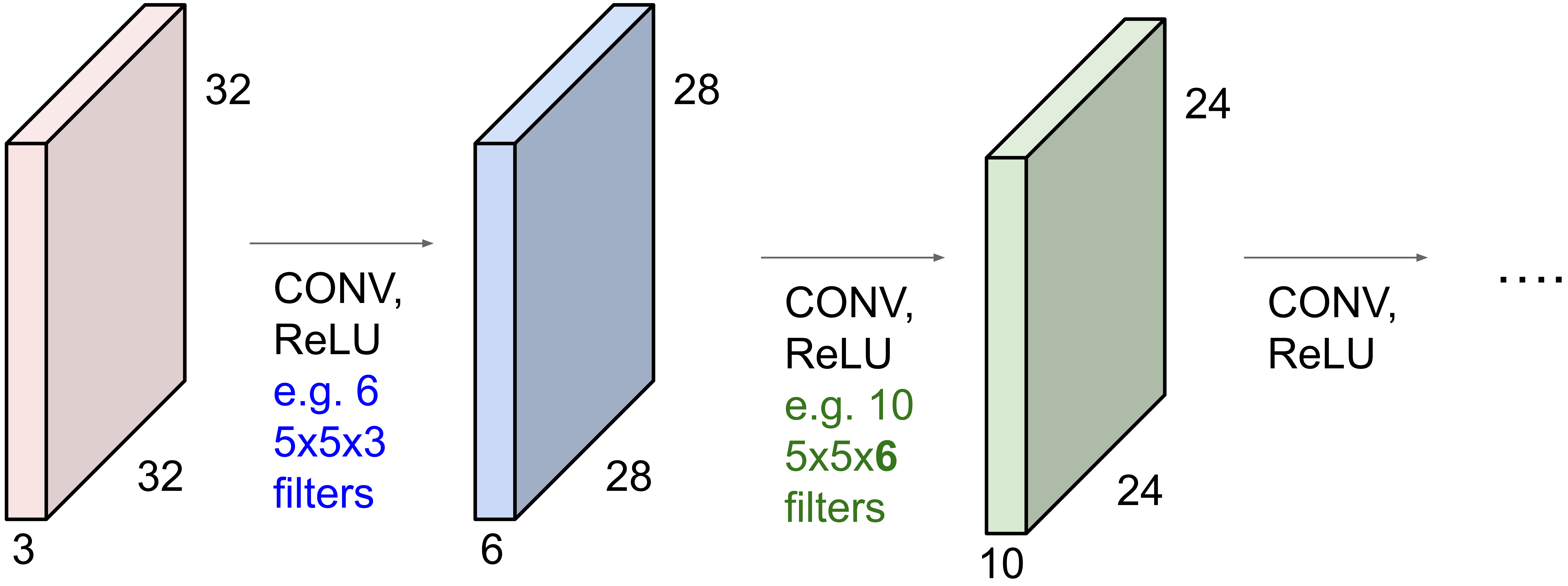


We stack these up to get a "new image" of size 28x28x6!

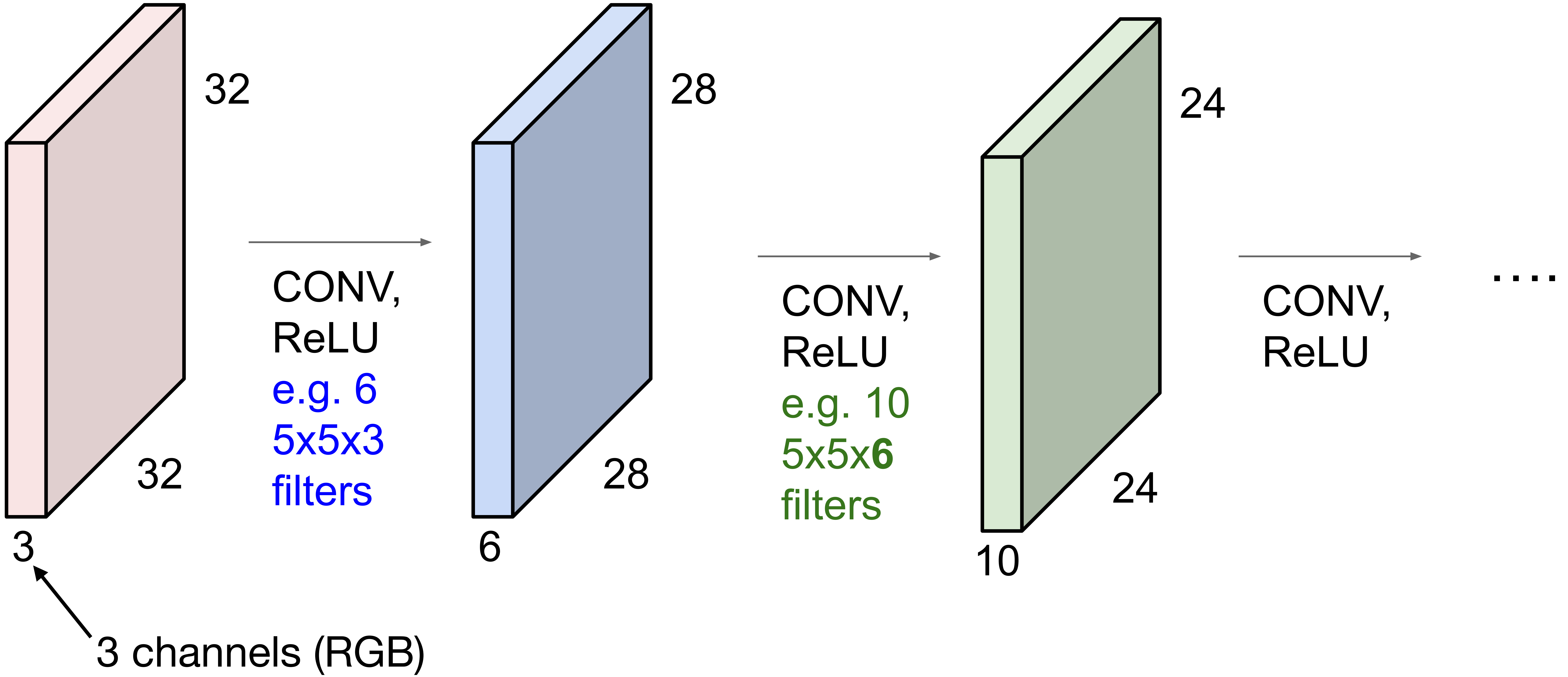
Preview: ConvNet is a sequence of Convolution Layers, interspersed with activation functions



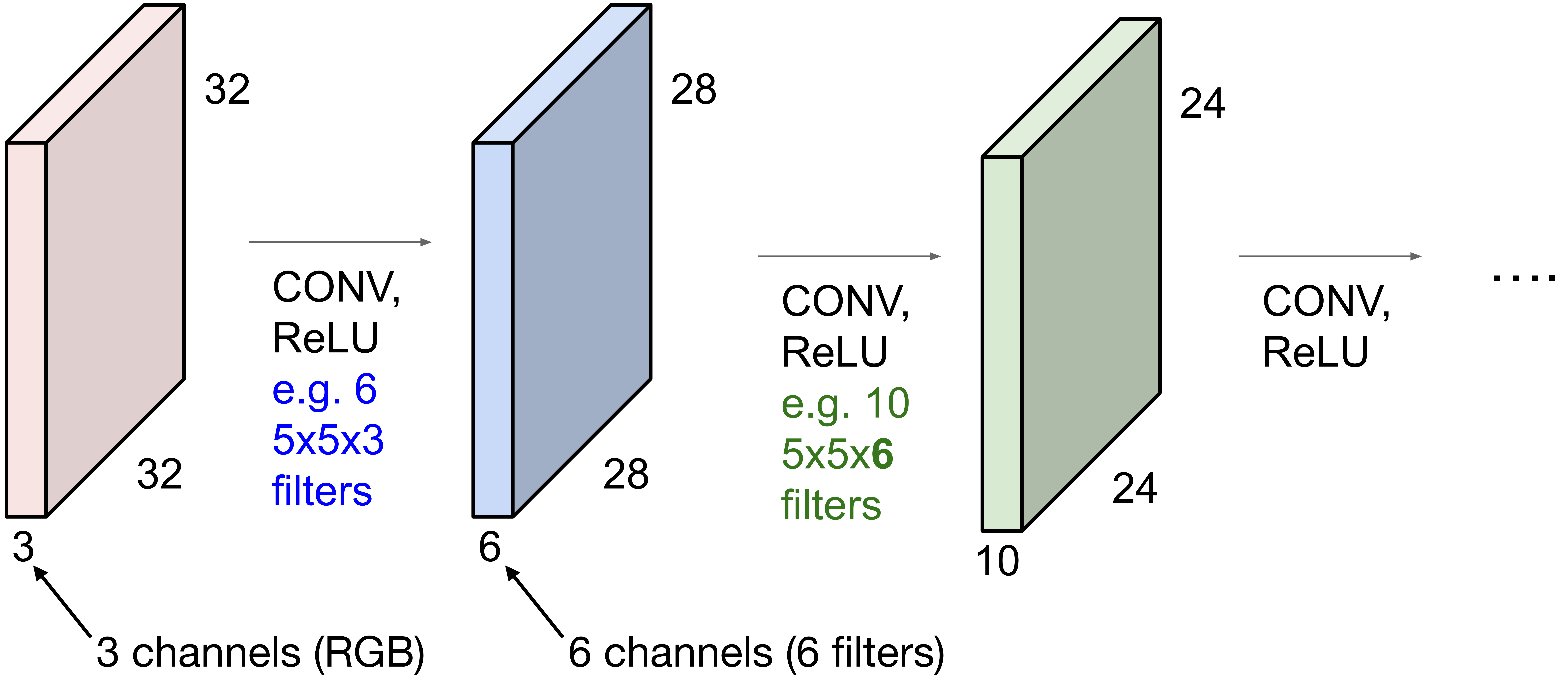
Preview: ConvNet is a sequence of Convolution Layers, interspersed with activation functions



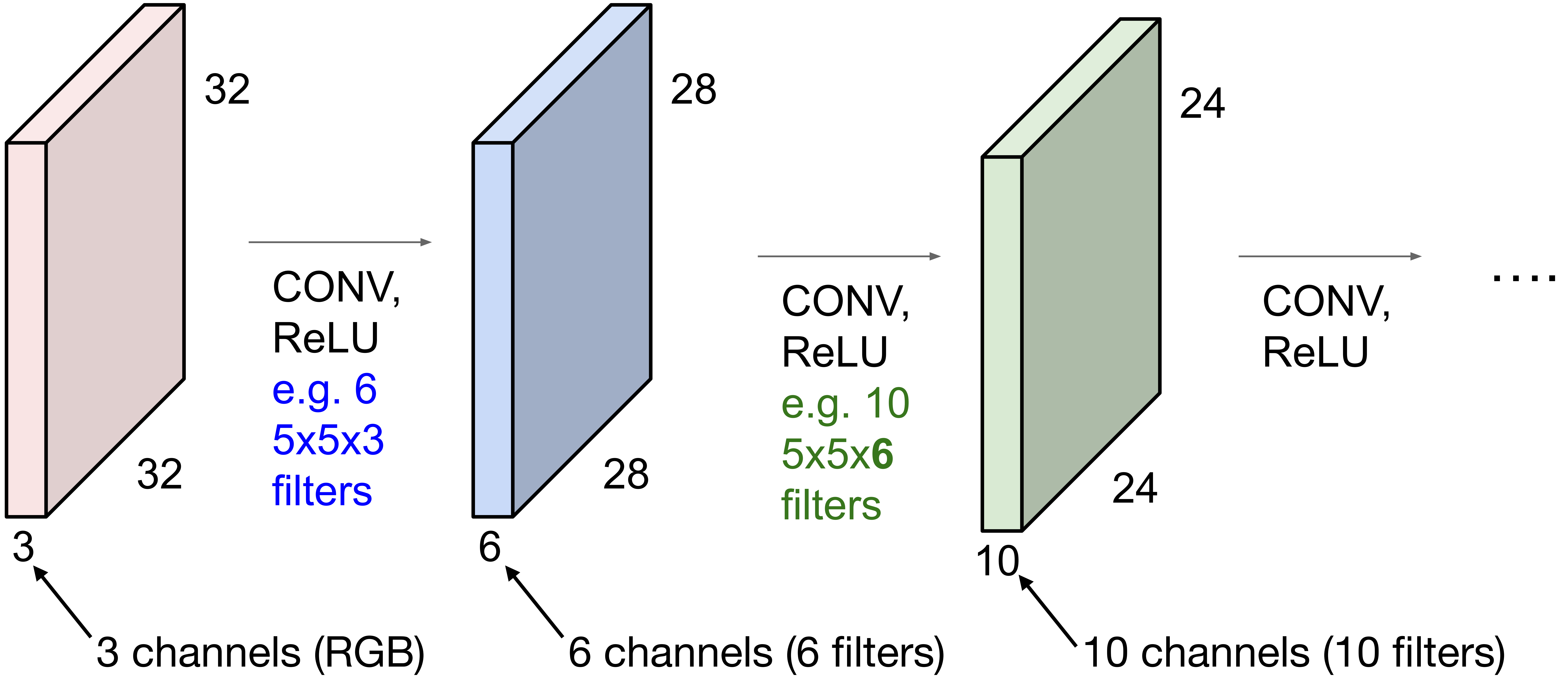
Preview: ConvNet is a sequence of Convolution Layers, interspersed with activation functions



Preview: ConvNet is a sequence of Convolution Layers, interspersed with activation functions

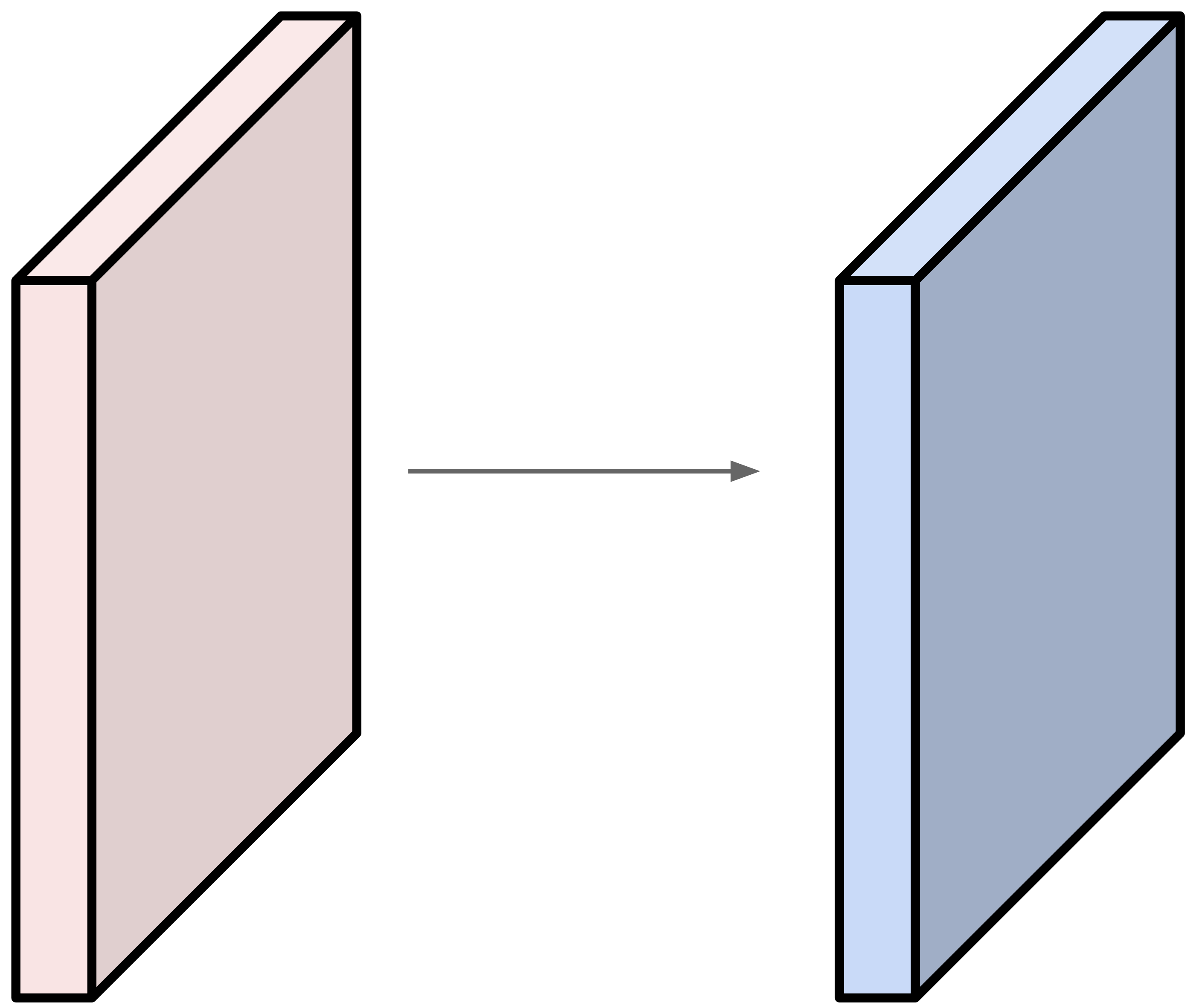


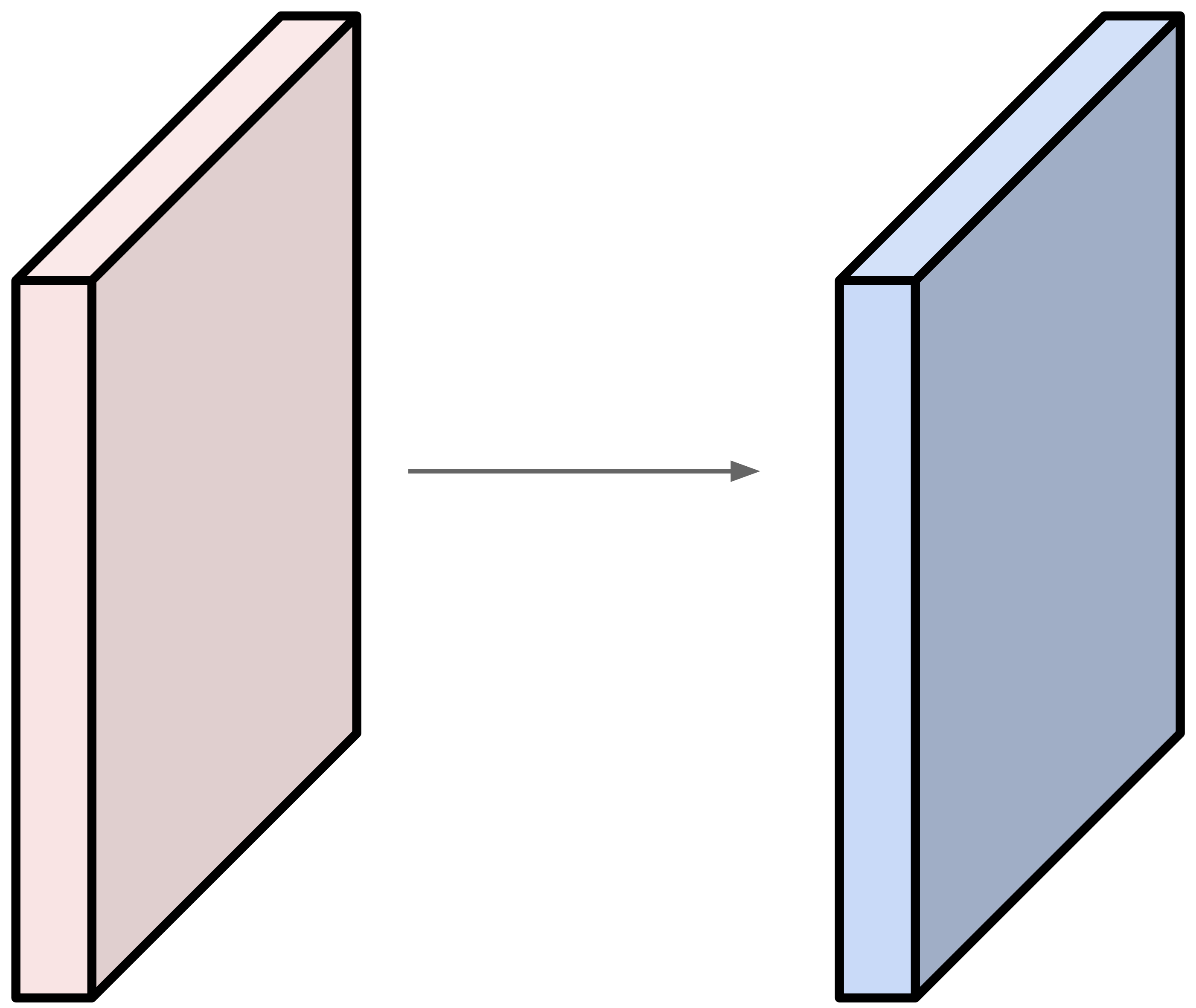
Preview: ConvNet is a sequence of Convolution Layers, interspersed with activation functions



Input volume: **32x32x3**
10 5x5 filters

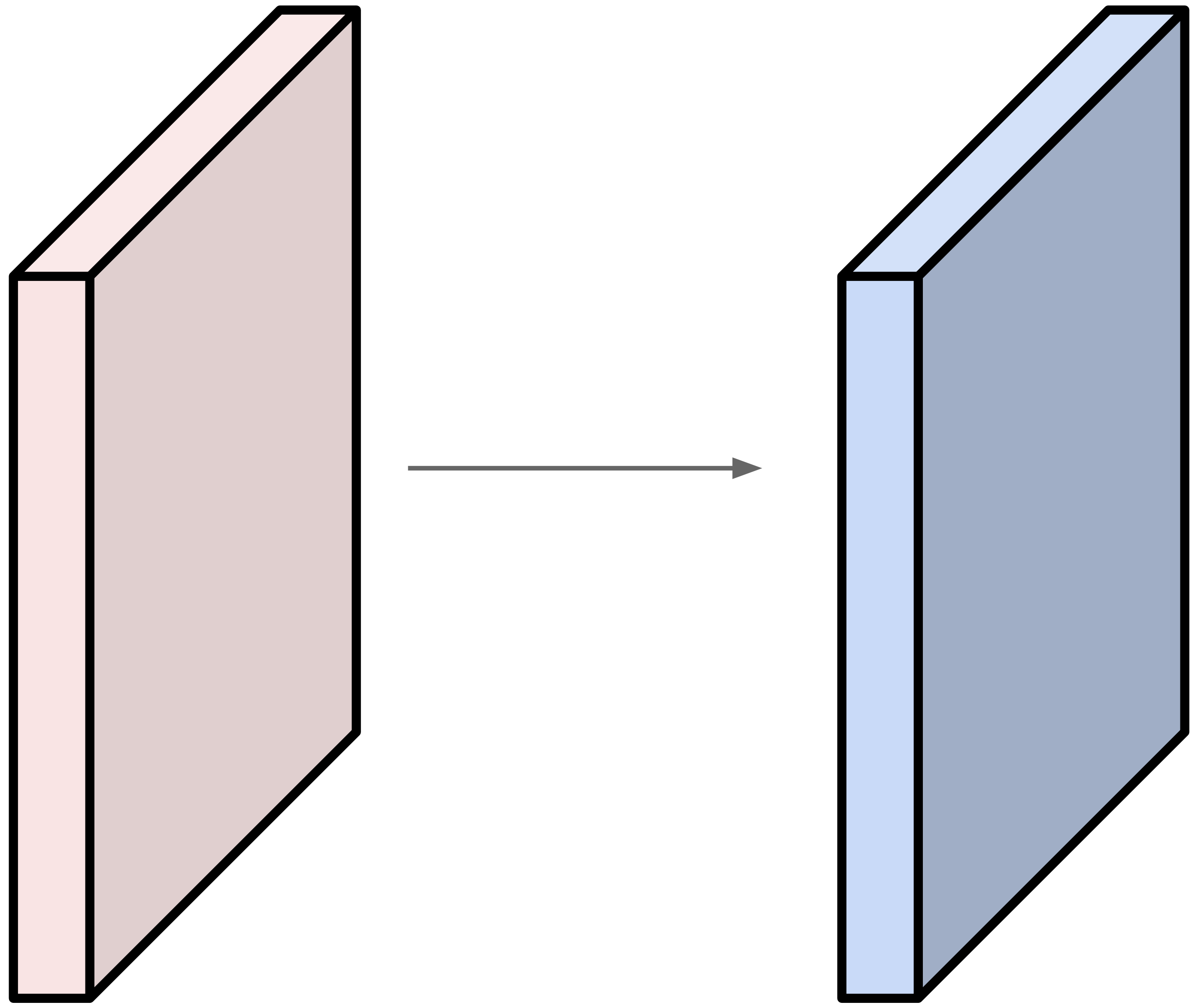
Number of parameters in this layer?





Input volume: **32x32x3**
10 5x5 filters

Number of parameters in this layer?
each filter has 5*5*3 + 1 = 76 params (+1 for bias)
=> 76*10 = **760**



Input volume: 32x32x3

10 5x5 filters

Number of parameters in this layer?

each filter has 5*5*3 + 1 = 76 params (+1 for bias)

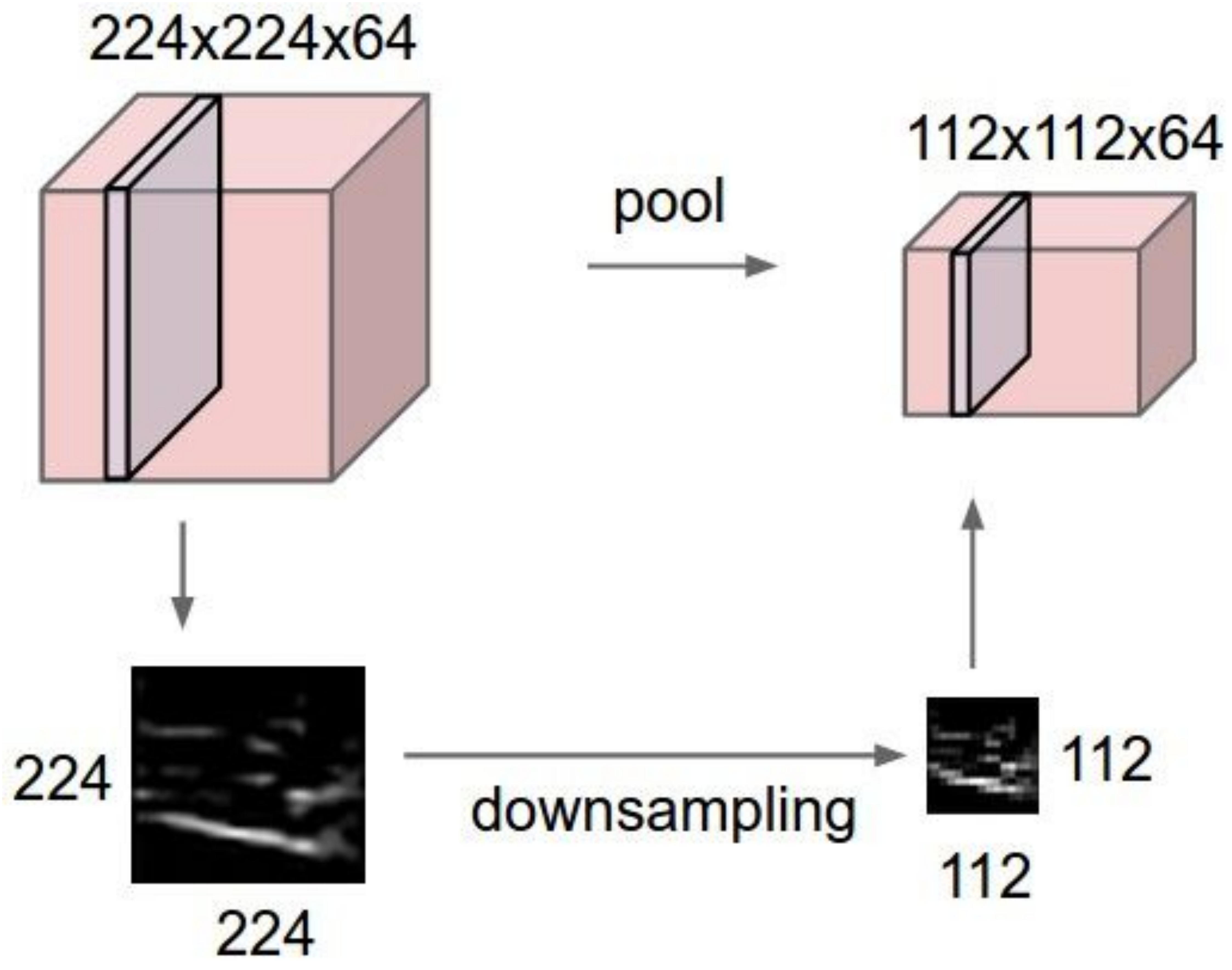
=> 76*10 = **760**

In general, parameters in conv layer =

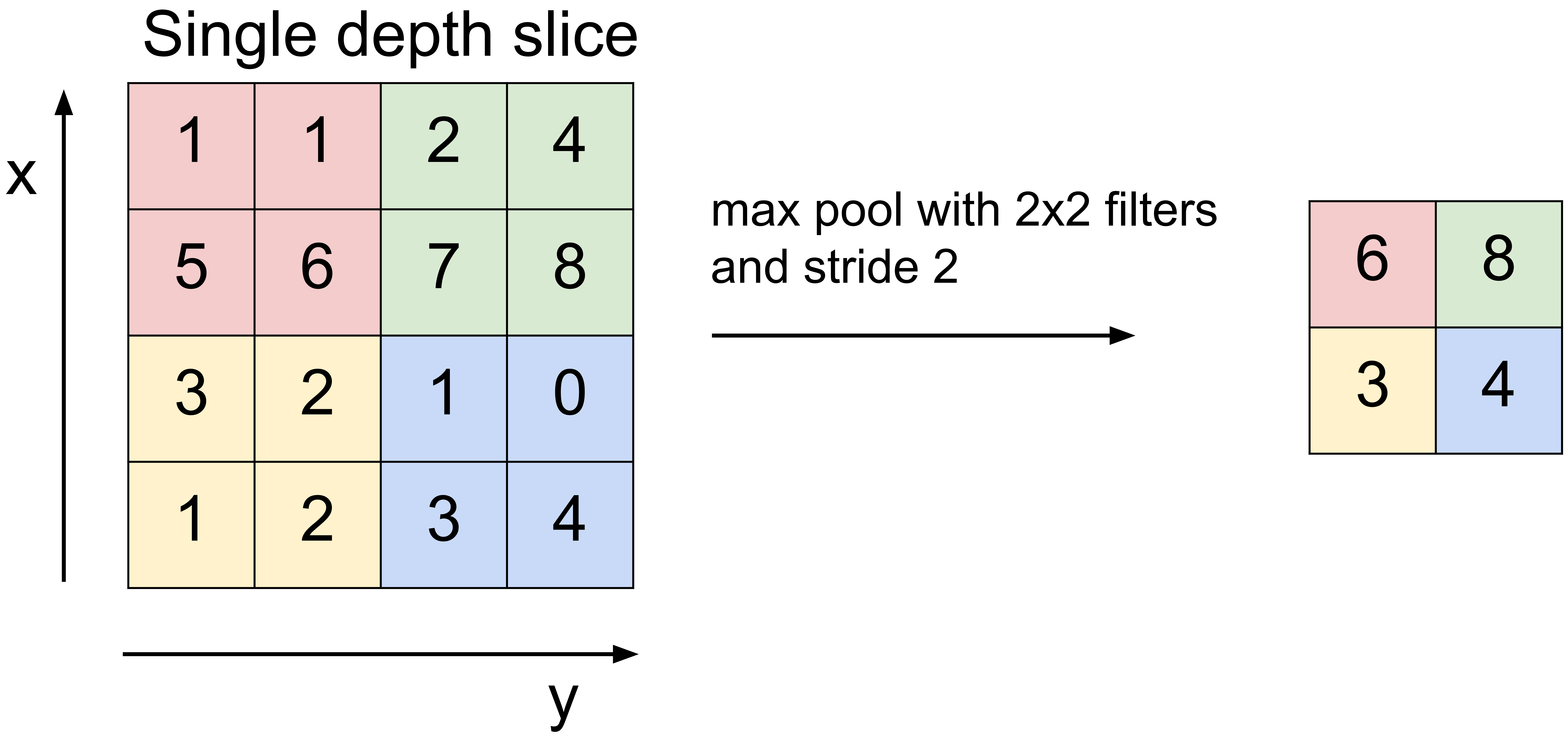
(filter width x filter height x input channels + 1) x number of filters.

Pooling layer

- makes the representations smaller and more manageable
- operates over each activation map independently:

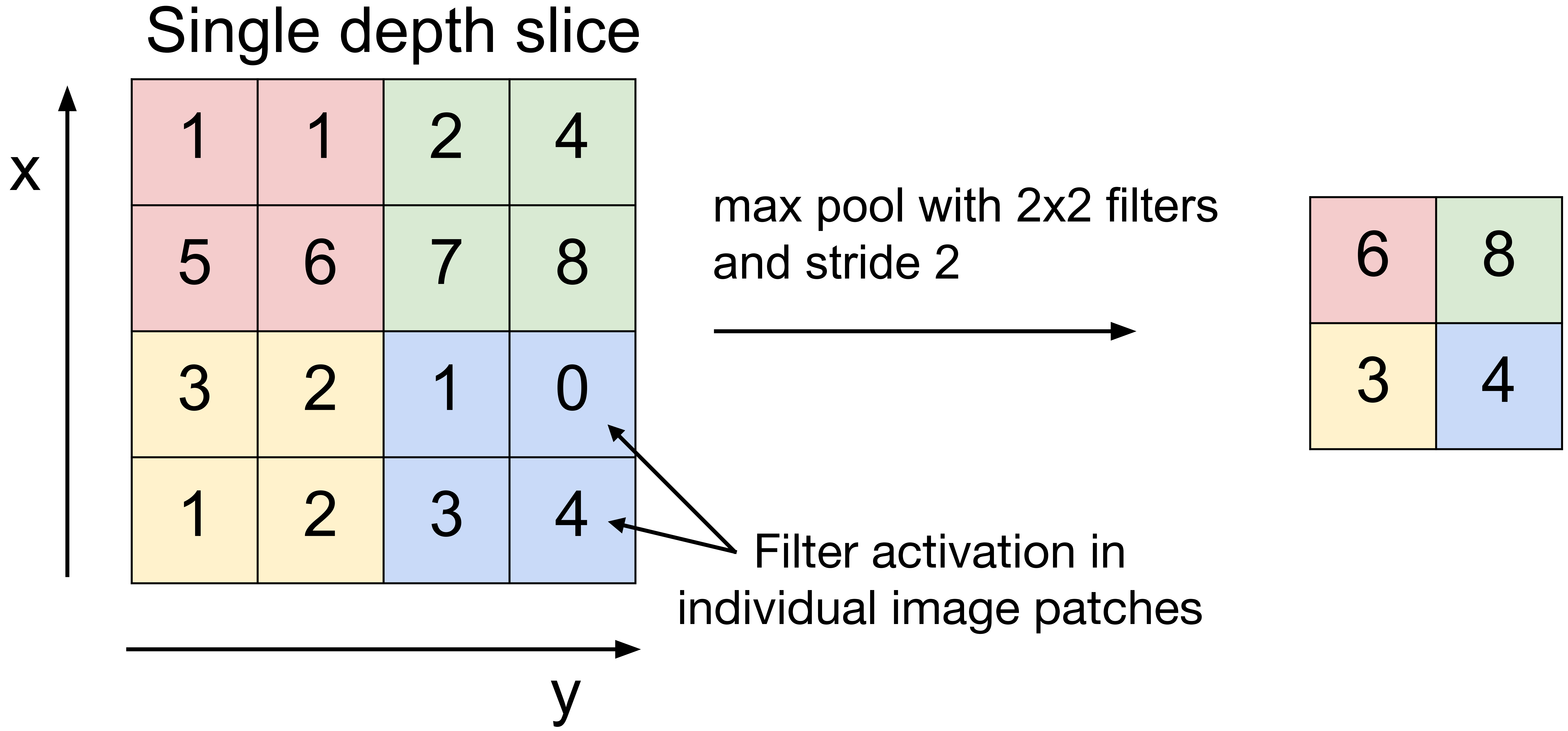


MAX POOLING

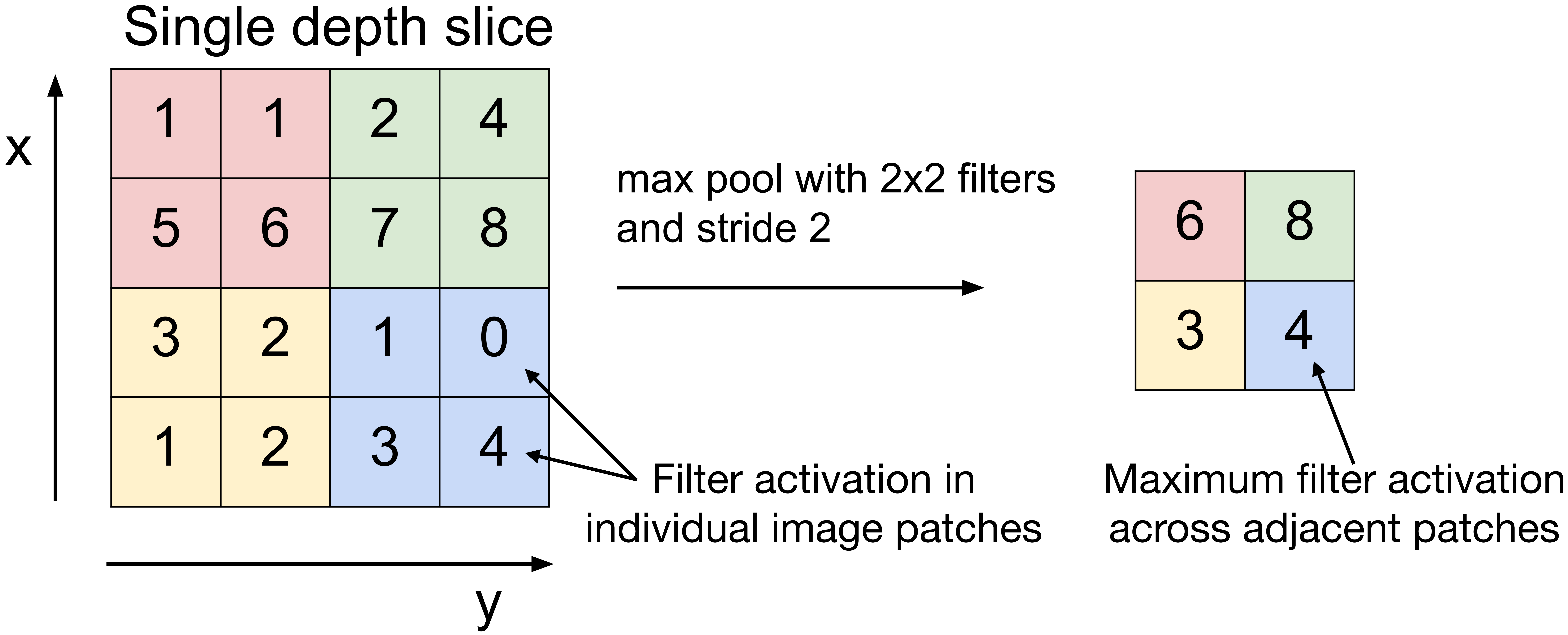


Single depth slice
x
y
1 1 2 4
5 6 7 8
3 2 1 0
1 2 3 4
max pool with 2x2 filters
and stride 2
6 8
3 4

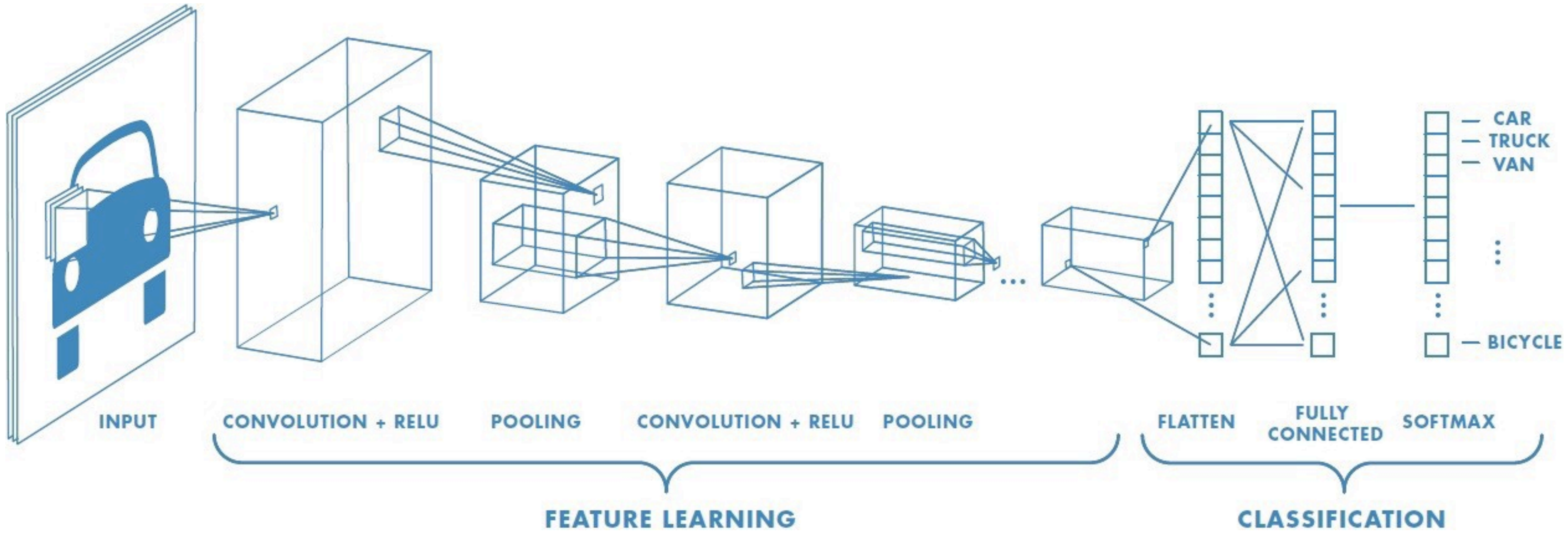
MAX POOLING


Single depth slice
x
y
1 1 2 4
5 6 7 8
3 2 1 0
1 2 3 4
max pool with 2x2 filters
and stride 2
6 8
3 4
Filter activation in
individual image patches

MAX POOLING

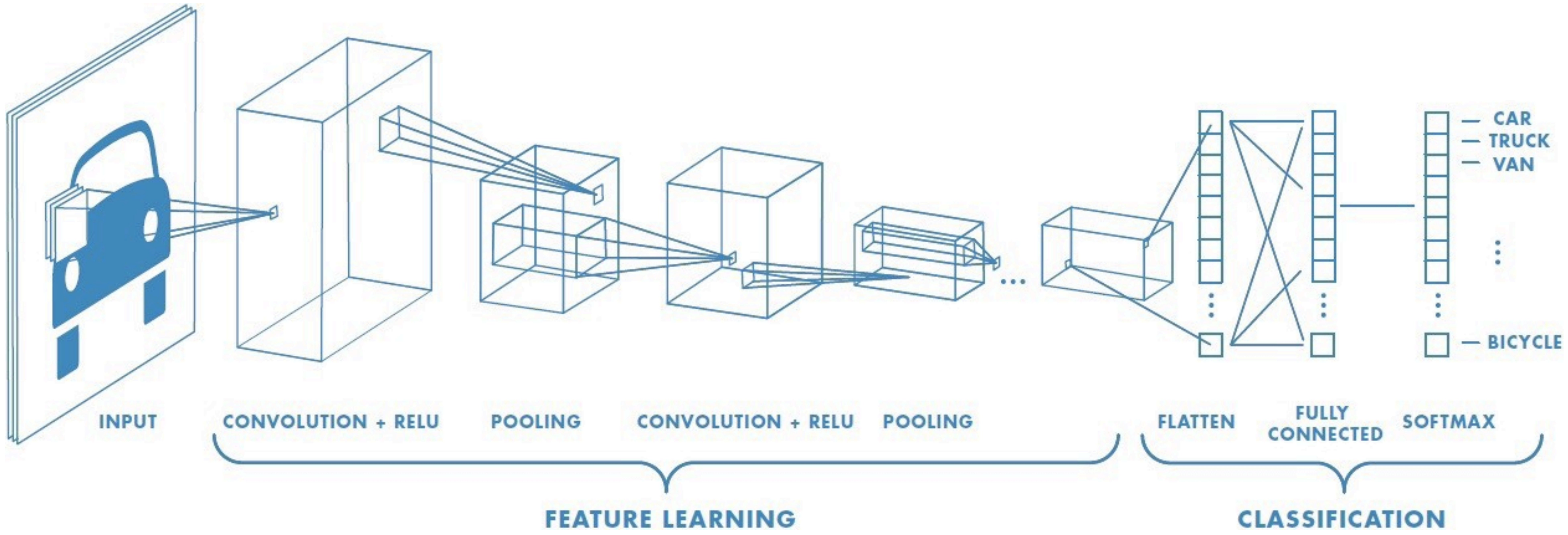

Single depth slice
x
y
1 1 2 4
5 6 7 8
3 2 1 0
1 2 3 4
max pool with 2x2 filters
and stride 2
6 8
3 4
Filter activation in
individual image patches
Maximum filter activation
across adjacent patches

Convolutional neural networks



https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

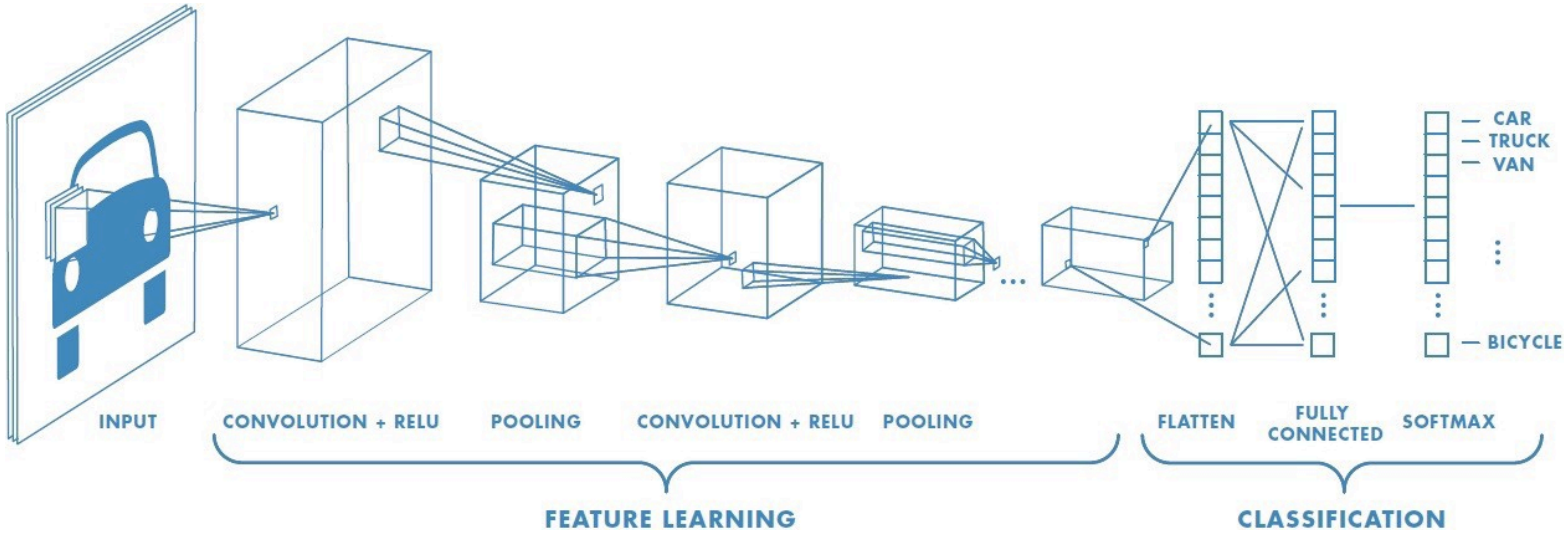
Convolutional neural networks



https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

A CNN stacks together several alternating convolution and pooling layers, followed by a fully connected layer and a softmax output.

Convolutional neural networks



https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

A CNN stacks together several alternating convolution and pooling layers, followed by a fully connected layer and a softmax output.

Filters, weights in fully connected layer, and biases learned by optimizing cross-entropy loss via stochastic gradient descent.

Interpreting the filters learned by a CNN

Interpreting the filters learned by a CNN

Use neural network for binary classification, e.g. faces versus not faces, cars versus not cars, etc.

Interpreting the filters learned by a CNN

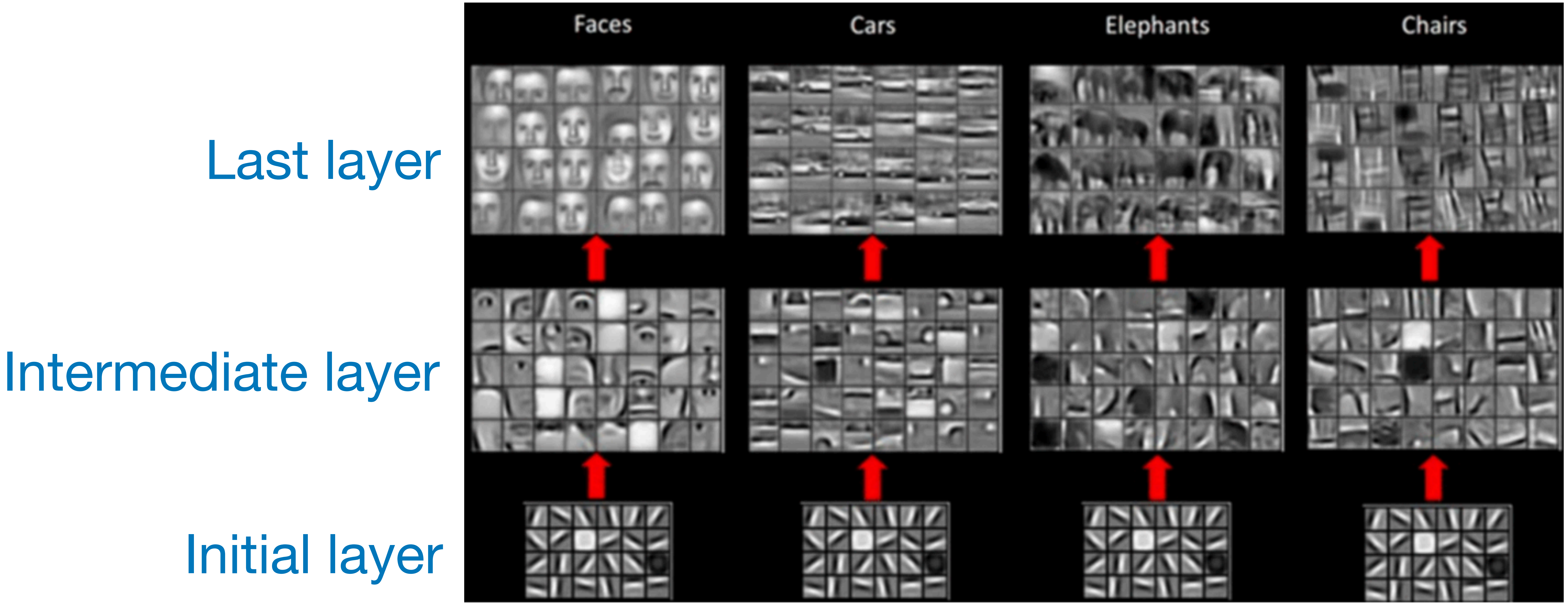
Use neural network for binary classification, e.g. faces versus not faces, cars versus not cars, etc.

For each neuron at each layer, find input image that activates it most strongly.

Interpreting the filters learned by a CNN

Use neural network for binary classification, e.g. faces versus not faces, cars versus not cars, etc.

For each neuron at each layer, find input image that activates it most strongly.



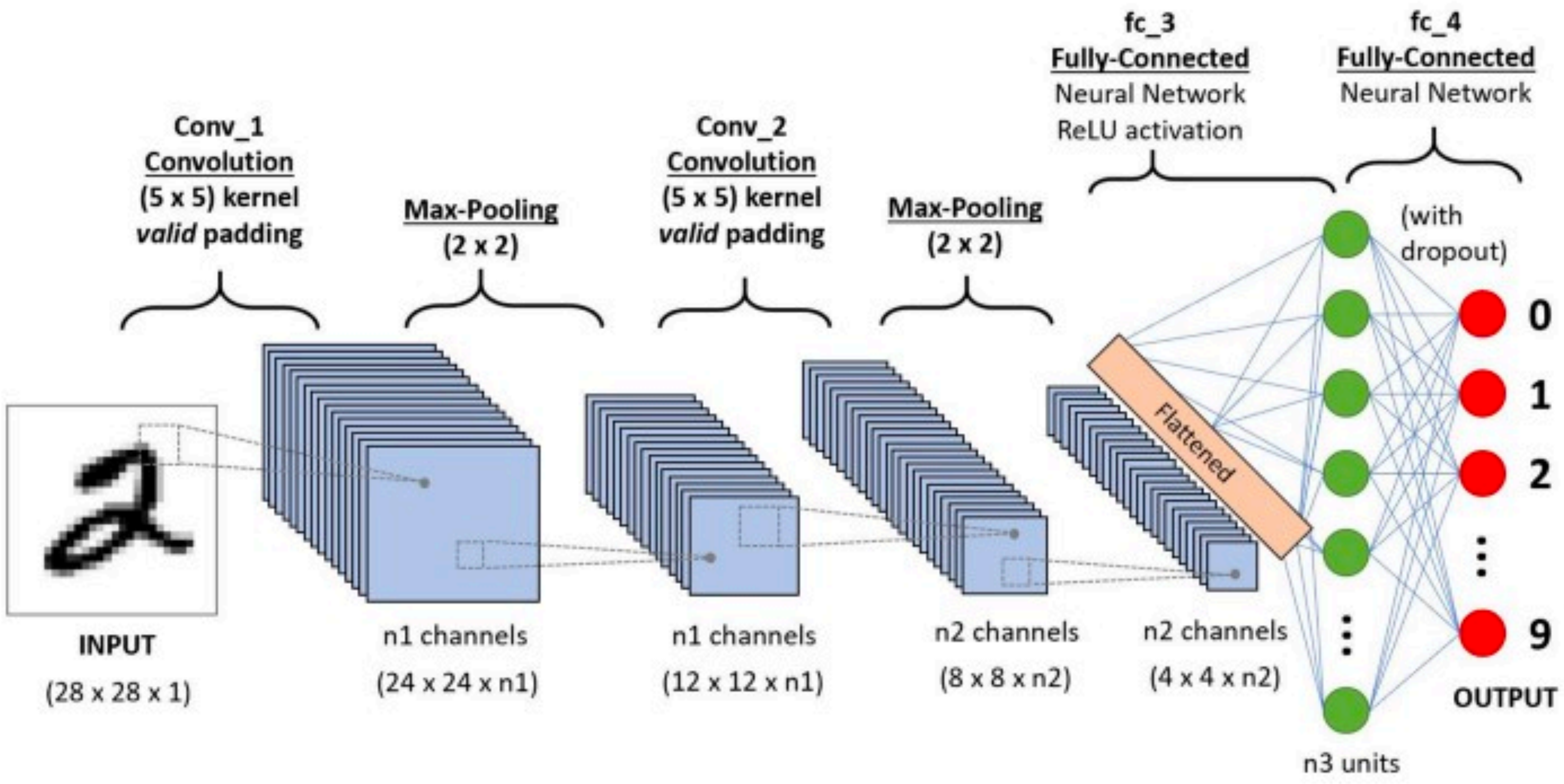
The original CNN architecture

The original CNN architecture

LeNet architecture for hand-written
digit recognition (1989).

The original CNN architecture

LeNet architecture for hand-written digit recognition (1989).



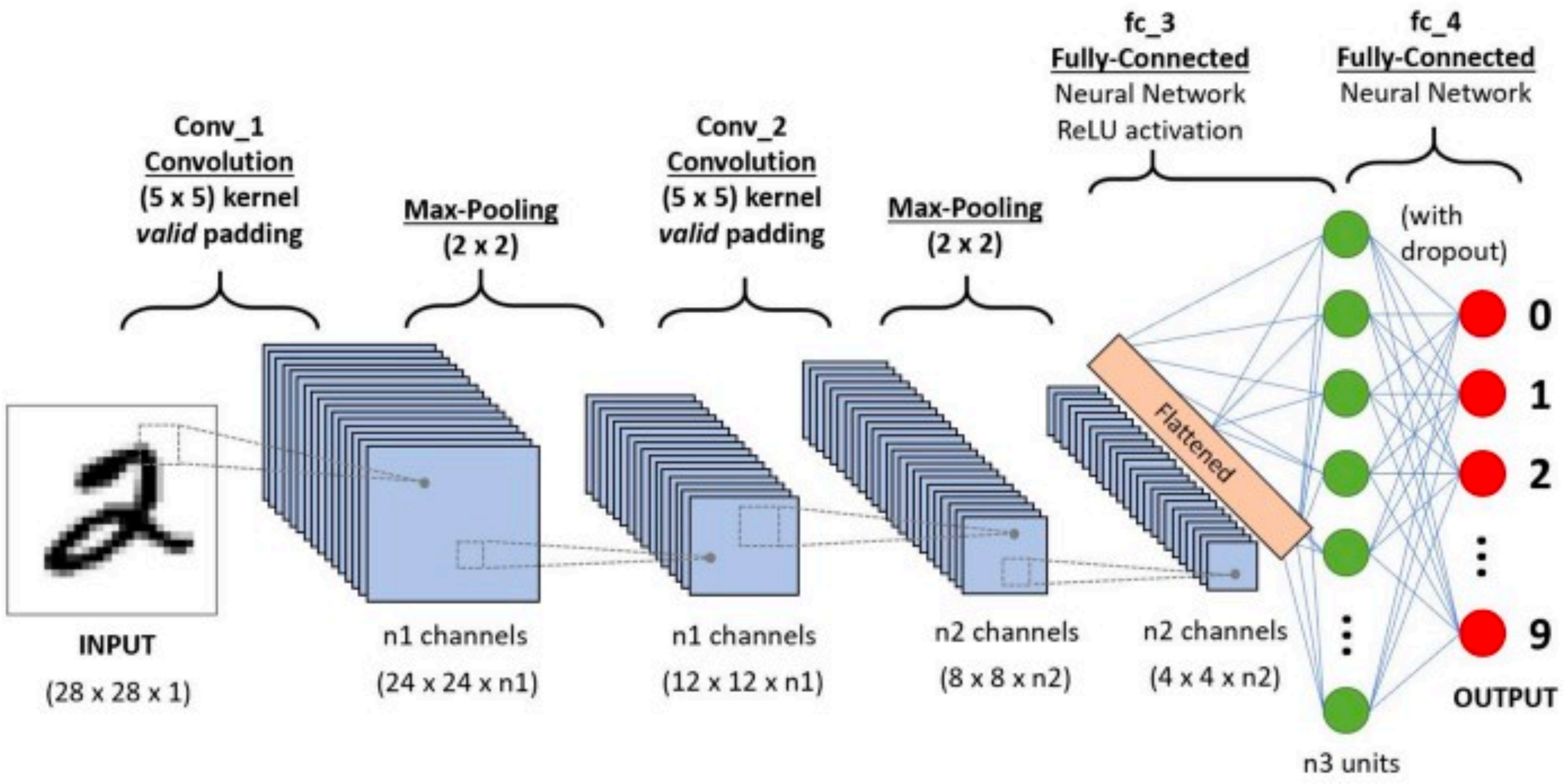
https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Yann LeCun

The original CNN architecture

LeNet architecture for hand-written digit recognition (1989).

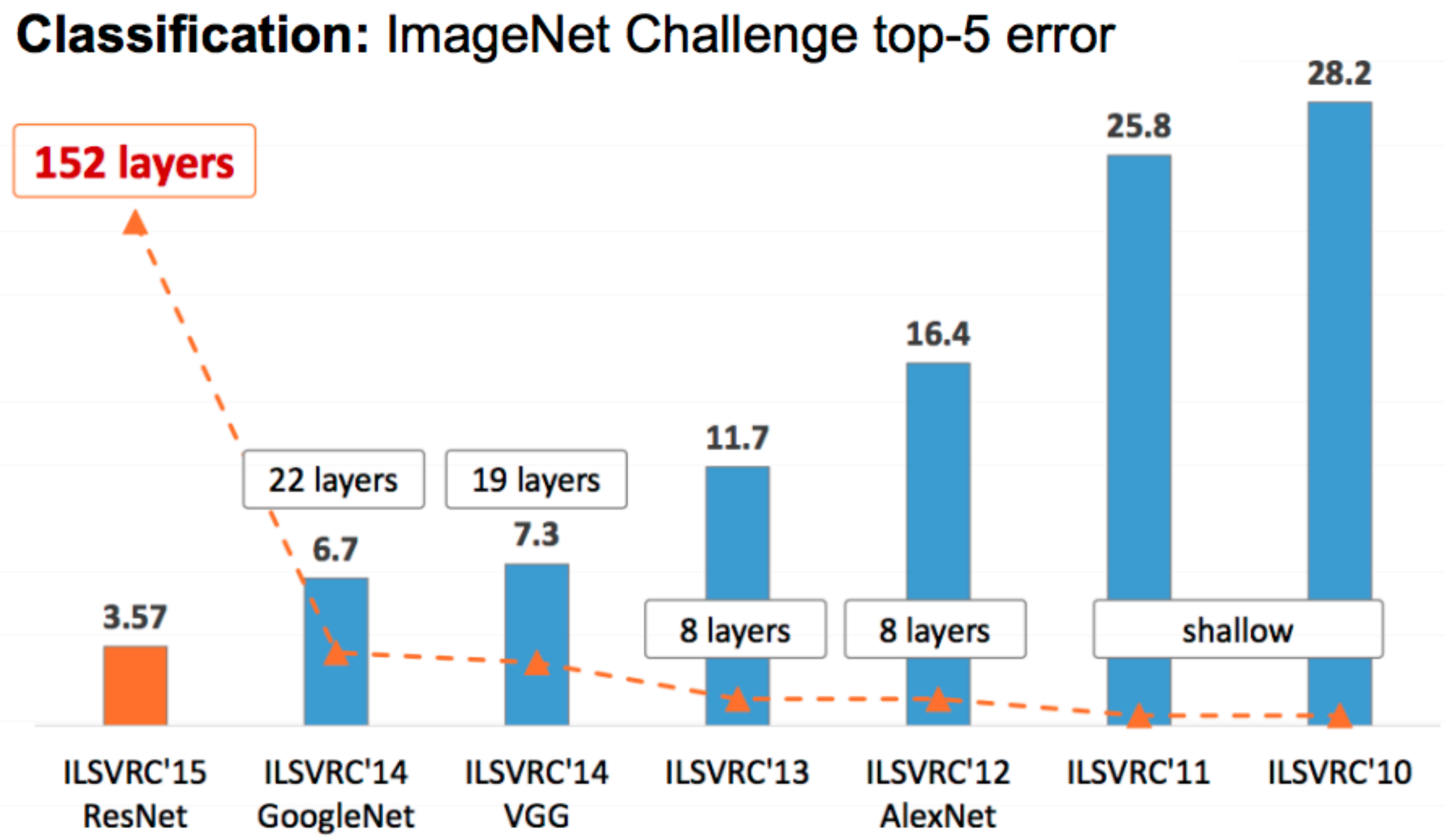
Idea existed decades ago but data and computing power only became available in 2010s.



https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Yann LeCun

Modern CNN architectures

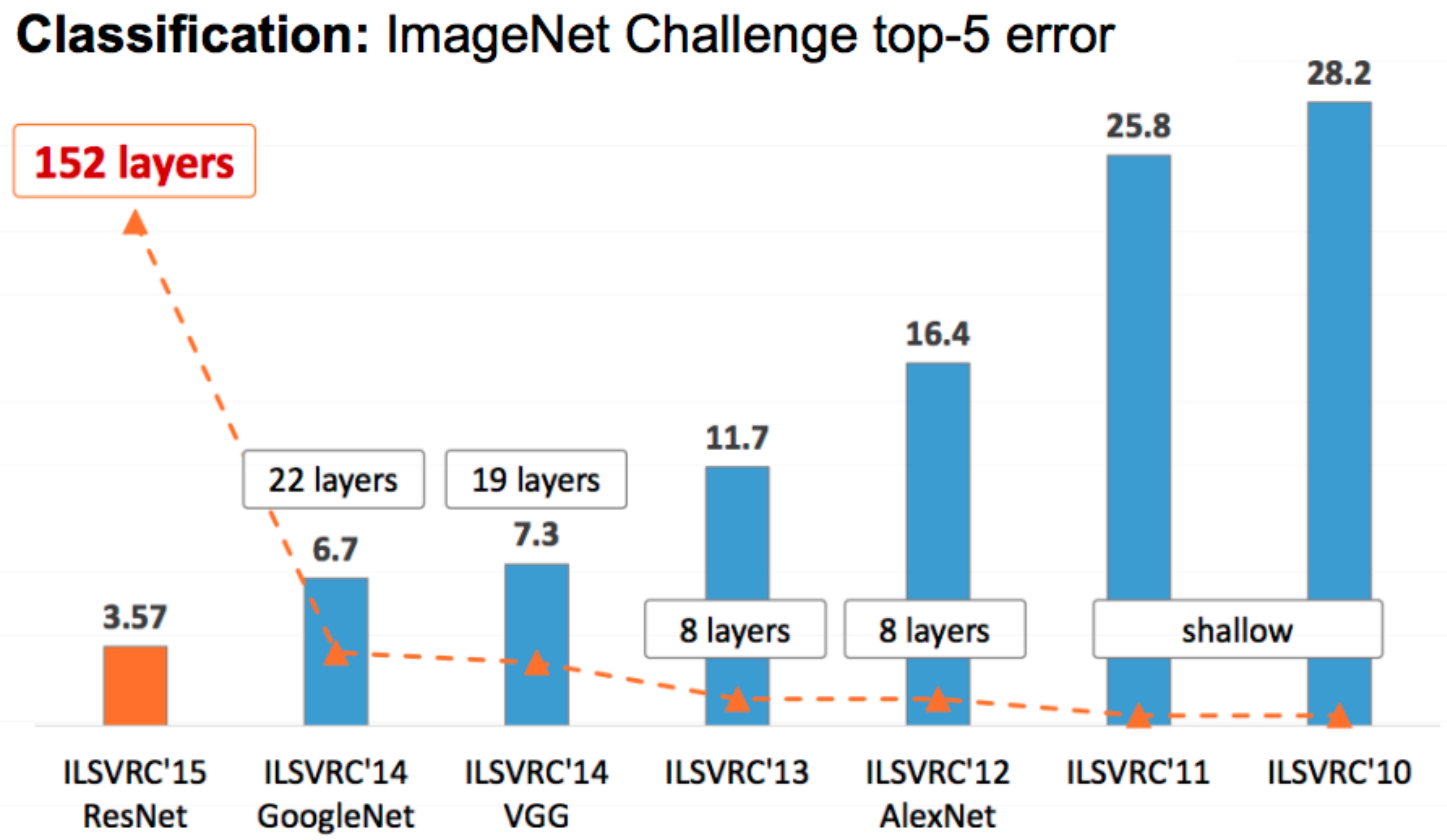


https://medium.com/@RaghavPrabhu/cnn-architectures-lenet-alexnet-vgg-googlenet-and-resnet-7c81c017b848

Model	Size (M)	Top-1/top-5 error (%)	# layers	Model description
AlexNet	238	41.00/18.00	8	5 conv + 3 fc layers
VGG-16	540	28.07/9.33	16	13 conv + 3 fc layers
VGG-19	560	27.30/9.00	19	16 conv + 3 fc layers
GoogleNet	40	29.81/10.04	22	21 conv + 1 fc layers
ResNet-50	100	22.85/6.71	50	49 conv + 1 fc layers
ResNet-152	235	21.43/3.57	152	151 conv + 1 fc layers

https://www.researchgate.net/figure/The-comparison-of-different-CNN-architectures-on-model-size-classification-error-rate_tbl1_320199404

Modern CNN architectures



https://medium.com/@RaghavPrabhu/cnn-architectures-lenet-alexnet-vgg-googlenet-and-resnet-7c81c017b848

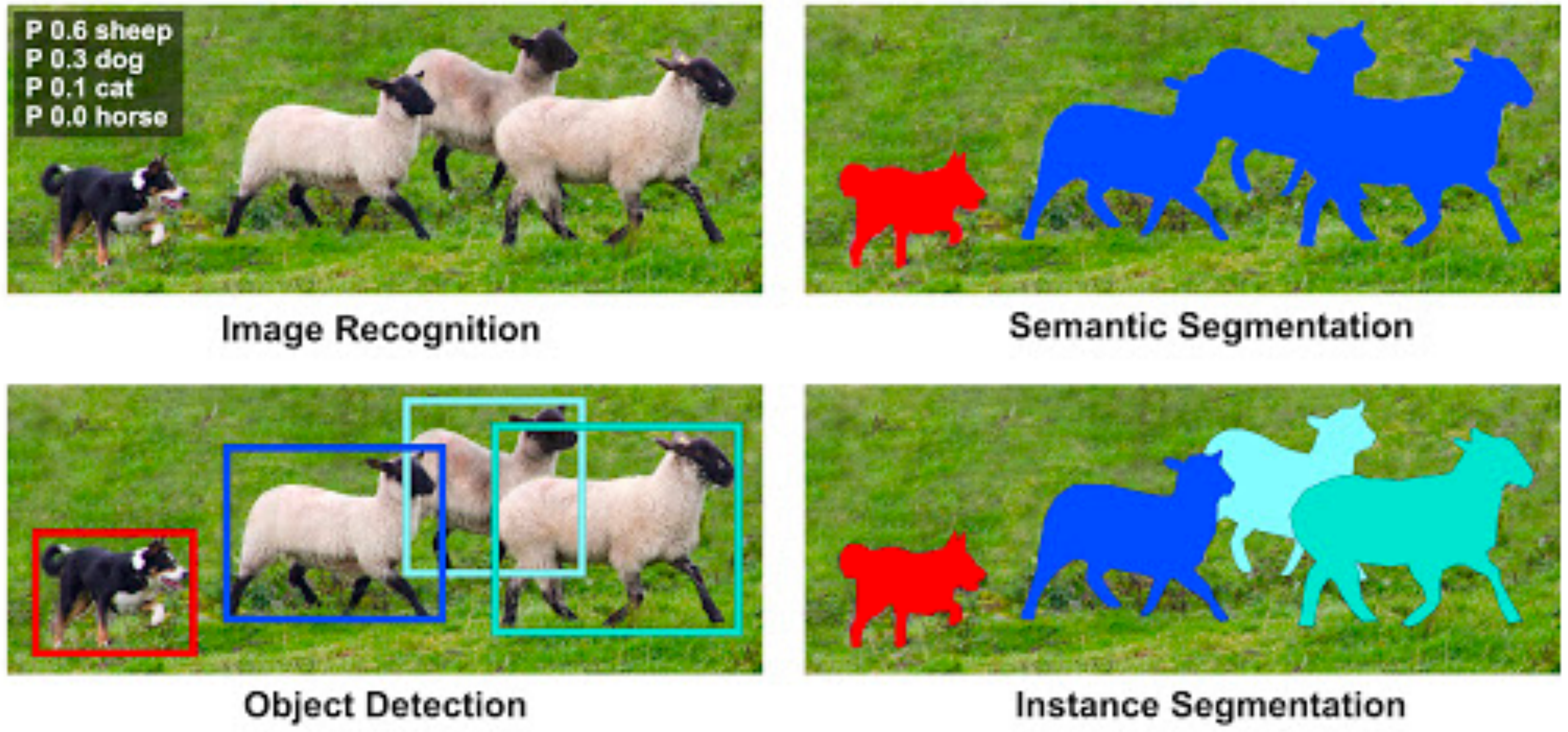
Model	Size (M)	Top-1/top-5 error (%)	# layers	Model description
AlexNet	238	41.00/18.00	8	5 conv + 3 fc layers
VGG-16	540	28.07/9.33	16	13 conv + 3 fc layers
VGG-19	560	27.30/9.00	19	16 conv + 3 fc layers
GoogleNet	40	29.81/10.04	22	21 conv + 1 fc layers
ResNet-50	100	22.85/6.71	50	49 conv + 1 fc layers
ResNet-152	235	21.43/3.57	152	151 conv + 1 fc layers

https://www.researchgate.net/figure/The-comparison-of-different-CNN-architectures-on-model-size-classification-error-rate_tbl1_320199404

CNNs are getting progressively deeper with time.

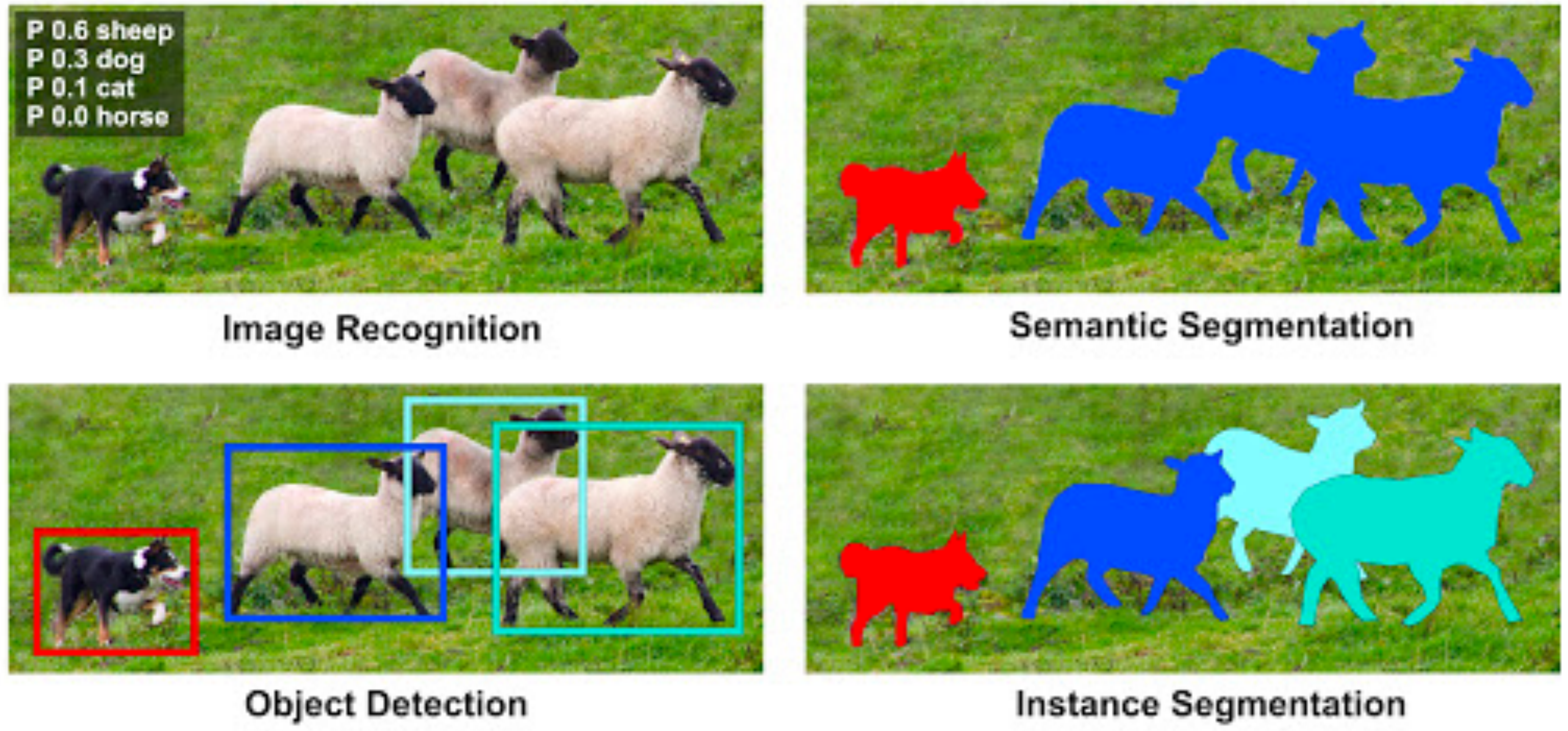
Other applications of CNNs

Other applications of CNNs



http://manipulation.csail.mit.edu/segmentation.html

Other applications of CNNs



http://manipulation.csail.mit.edu/segmentation.html

content image | **style image** | **generated image**

Ancient city of Persepolis

+

The Starry Night (Van Gogh)

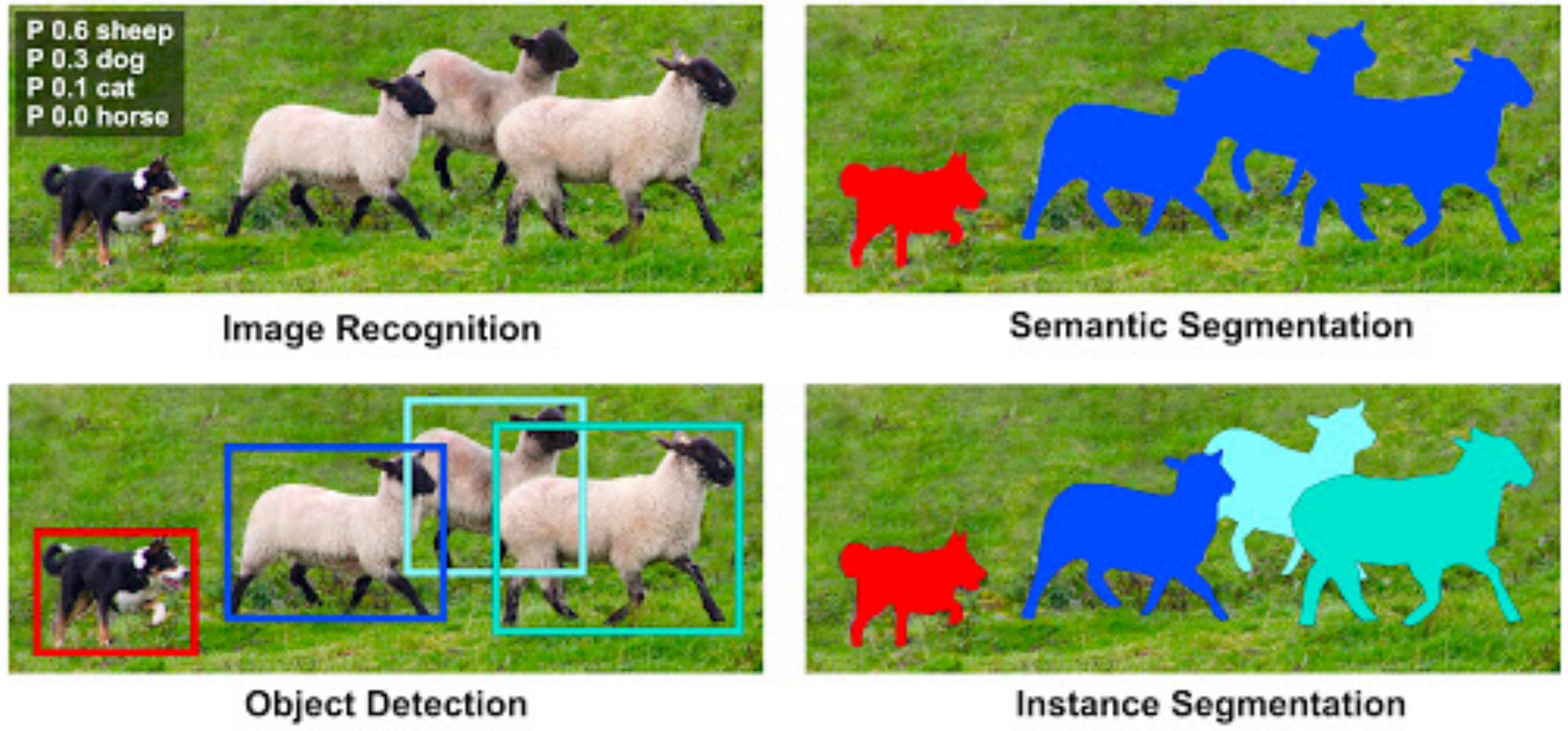
=

Persepolis
in Van Gogh style

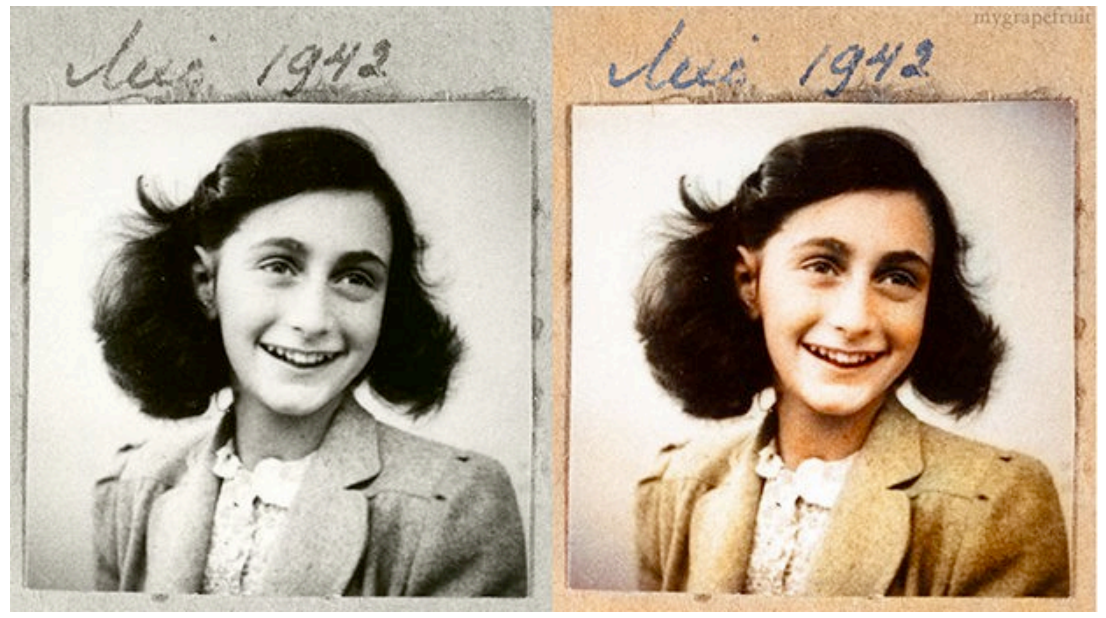
https://towardsdatascience.com/neural-style-transfer-on-real-time-video-with-full-implementable-code-ac2dbc0e9822

Style Transfer

Other applications of CNNs



http://manipulation.csail.mit.edu/segmentation.html



https://medium.com/@nabilliban14/chain-colorization-using-cdcgans-and-cnn-from-learned-deep-prior-1405dab48df3

Image colorization

content image **style image** **generated image**

 +  = 

Ancient city of Persepolis

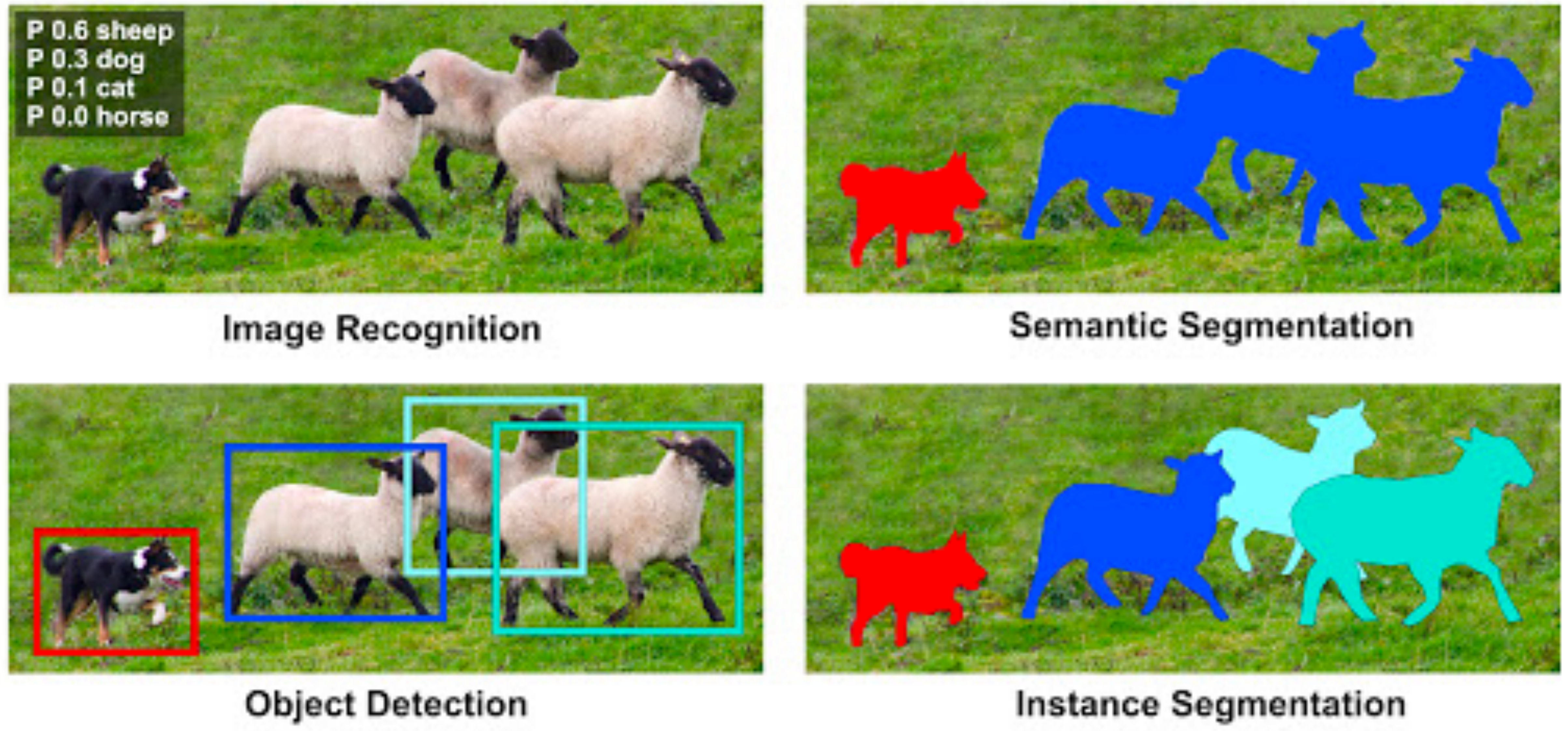
The Starry Night (Van Gogh)

Persepolis
in Van Gogh style

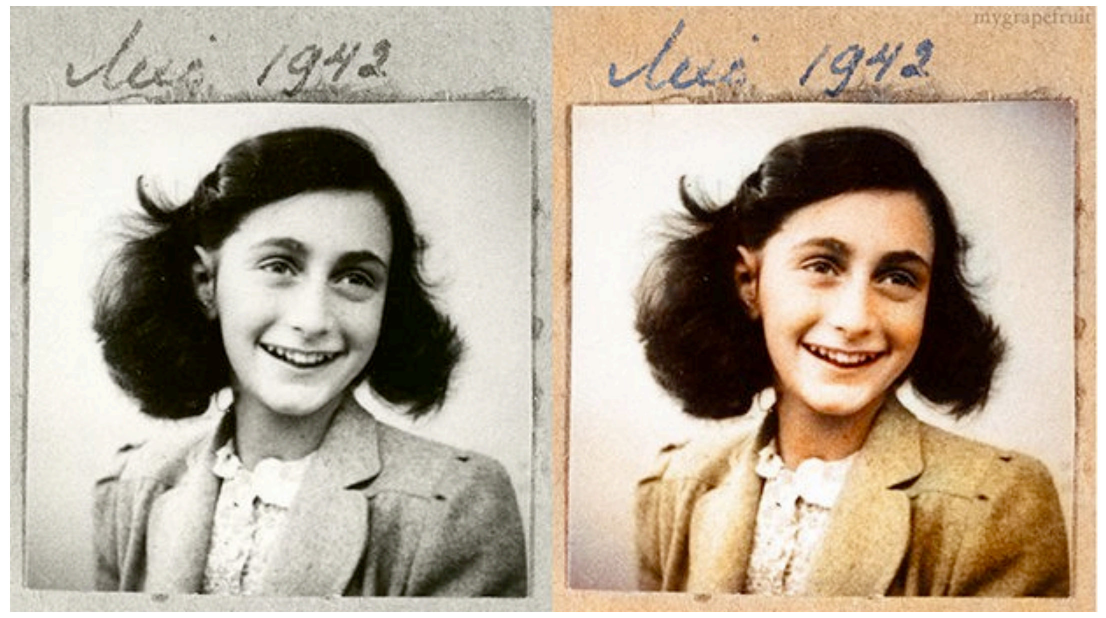
https://towardsdatascience.com/neural-style-transfer-on-real-time-video-with-full-implementable-code-ac2dbc0e9822

Style Transfer

Other applications of CNNs

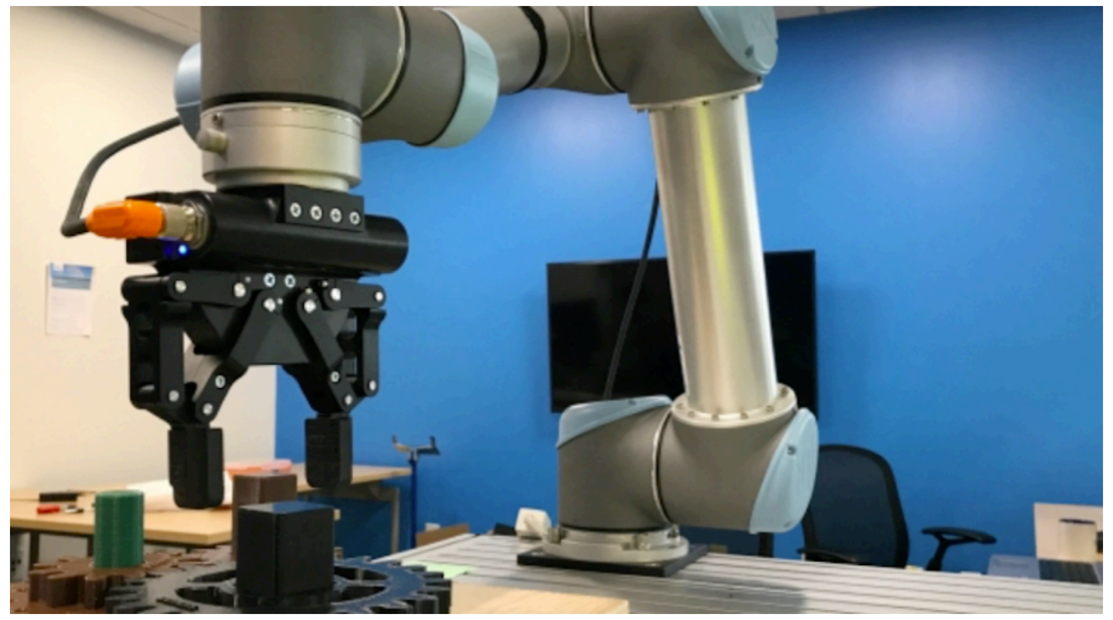


http://manipulation.csail.mit.edu/segmentation.html



https://medium.com/@nabilliban14/chain-colorization-using-cdcgans-and-cnn-from-learned-deep-prior-1405dab48df3

Image colorization



https://www.therobotreport.com/reinforcement-learning-industrial-robotics/

Reinforcement learning

content image **style image** **generated image**

 +  = 

Ancient city of Persepolis

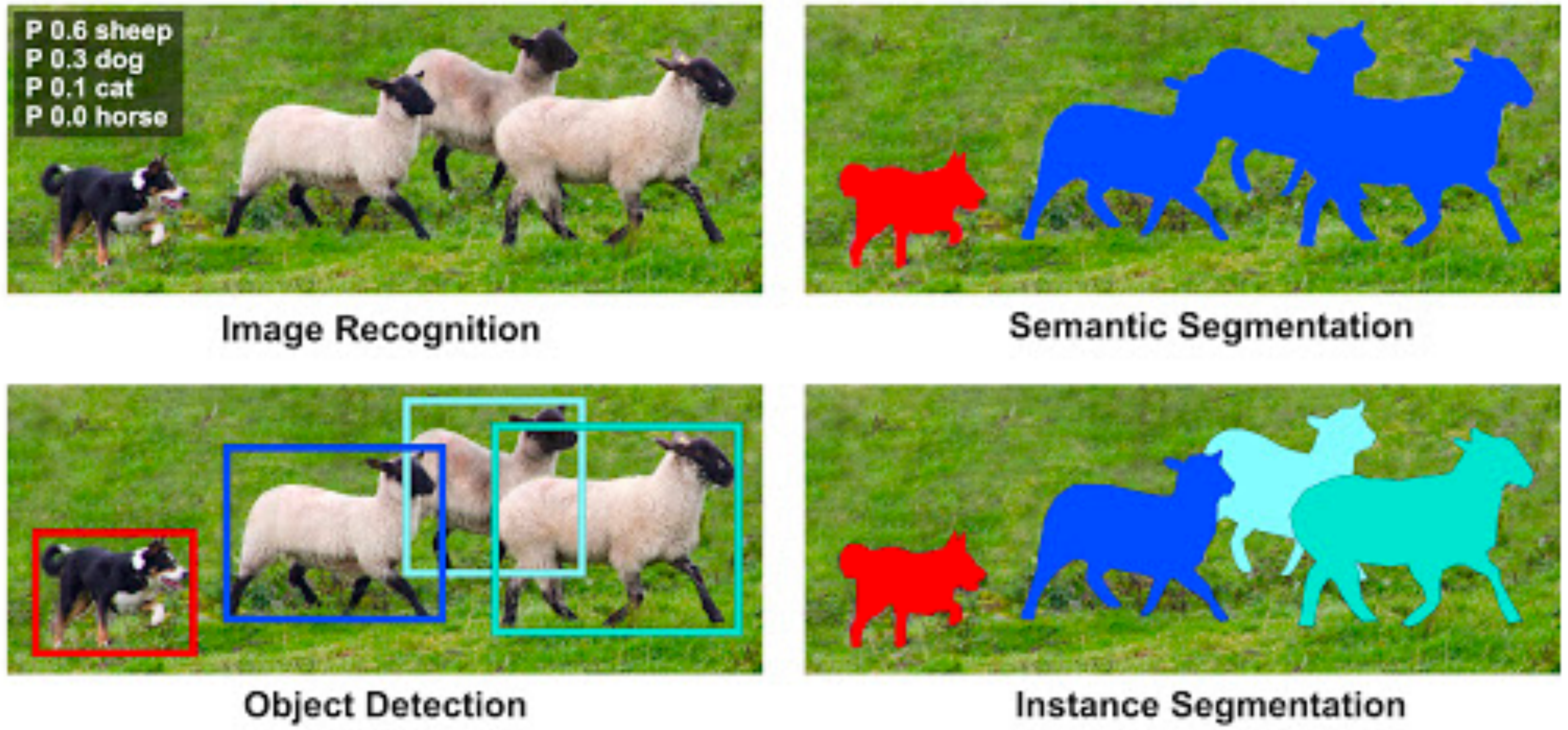
The Starry Night (Van Gogh)

Persepolis in Van Gogh style

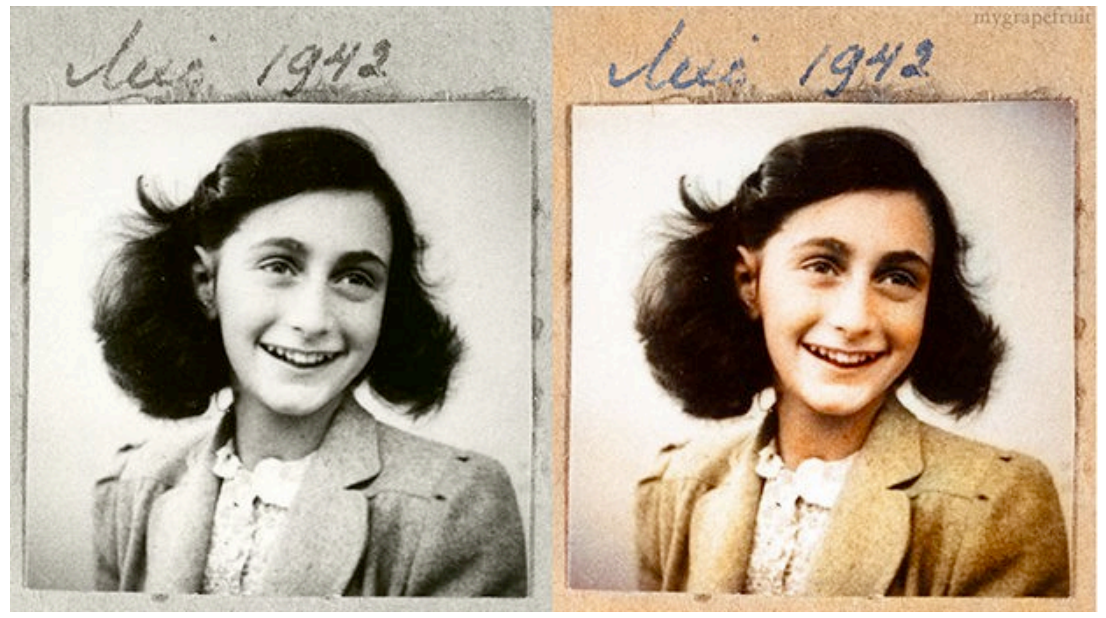
https://towardsdatascience.com/neural-style-transfer-on-real-time-video-with-full-implementable-code-ac2dbc0e9822

Style Transfer

Other applications of CNNs

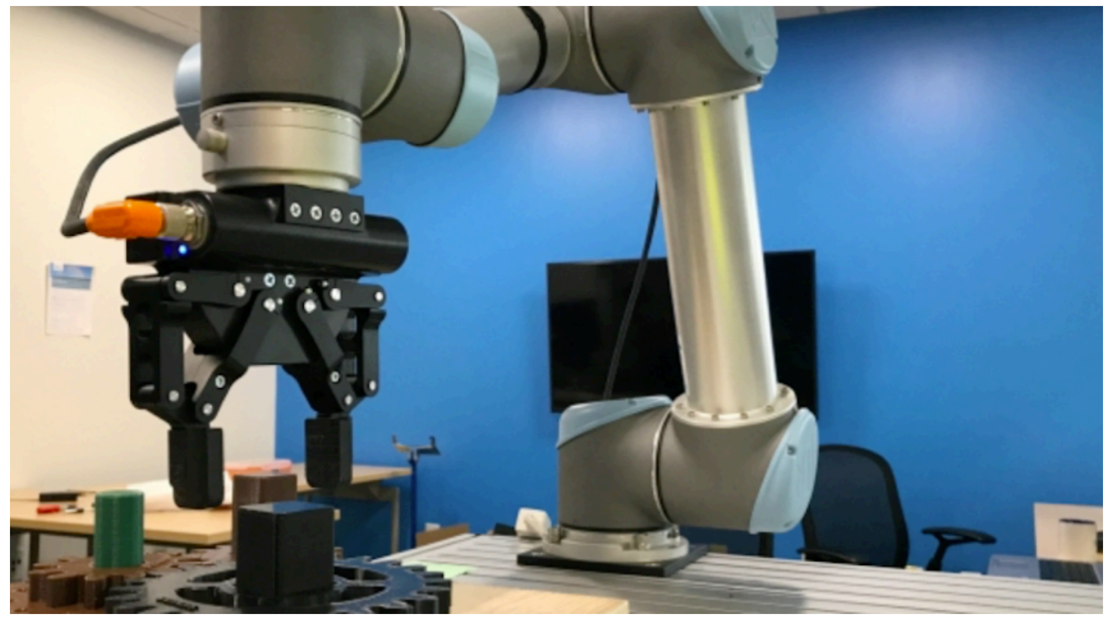


http://manipulation.csail.mit.edu/segmentation.html



https://medium.com/@nabilliban14/chain-colorization-using-cdcgans-and-cnn-from-learned-deep-prior-1405dab48df3

Image colorization



https://www.therobotreport.com/reinforcement-learning-industrial-robotics/

Reinforcement learning

content image **style image** **generated image**

 +  = 

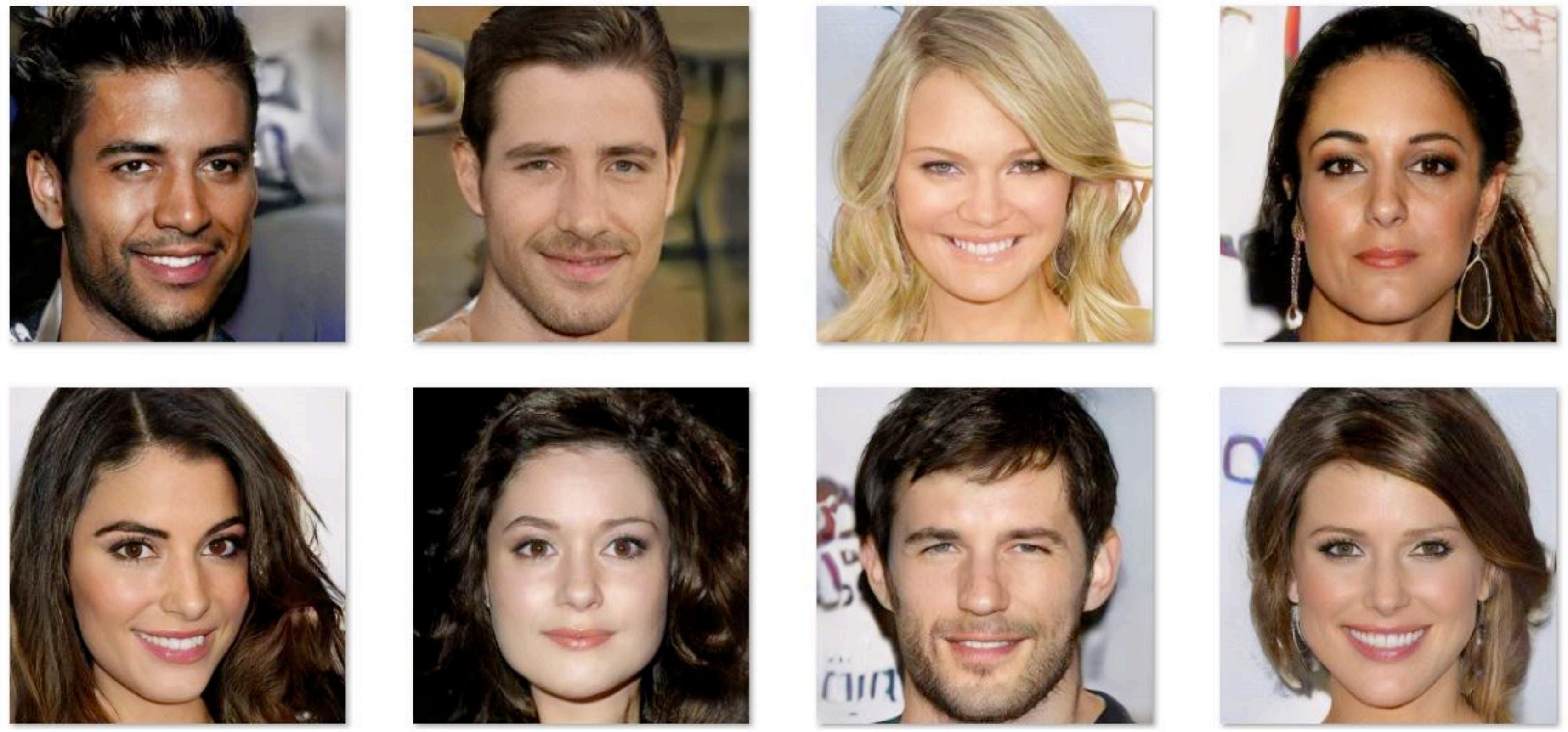
Ancient city of Persepolis

The Starry Night (Van Gogh)

Persepolis
in Van Gogh style

https://towardsdatascience.com/neural-style-transfer-on-real-time-video-with-full-implementable-code-ac2dbc0e9822

Style Transfer



https://medium.datadriveninvestor.com/artificial-intelligence-gans-can-create-fake-celebrity-faces-44fe80d419f7

Generative models

Summary

Summary

- Image classification is a prototypical task in image processing.

Summary

- Image classification is a prototypical task in image processing.
- A convolution sweeps over an image, looking for a specific pattern.

Summary

- Image classification is a prototypical task in image processing.
- A convolution sweeps over an image, looking for a specific pattern.
- Convolutions are local (applied to image patches) and translation-invariant (applied equally to patches across the whole image).

Summary

- Image classification is a prototypical task in image processing.
- A convolution sweeps over an image, looking for a specific pattern.
- Convolutions are local (applied to image patches) and translation-invariant (applied equally to patches across the whole image).
- Convolutional neural networks are a specialized architecture for image processing, consisting of alternating convolutional and pooling layers (feature learning), following by final fully connected layer (classification).

Summary

- Image classification is a prototypical task in image processing.
- A convolution sweeps over an image, looking for a specific pattern.
- Convolutions are local (applied to image patches) and translation-invariant (applied equally to patches across the whole image).
- Convolutional neural networks are a specialized architecture for image processing, consisting of alternating convolutional and pooling layers (feature learning), following by final fully connected layer (classification).
- People have built increasingly deep CNNs, which have performed increasingly well. Image classification problem is essentially solved.

Summary

- Image classification is a prototypical task in image processing.
- A convolution sweeps over an image, looking for a specific pattern.
- Convolutions are local (applied to image patches) and translation-invariant (applied equally to patches across the whole image).
- Convolutional neural networks are a specialized architecture for image processing, consisting of alternating convolutional and pooling layers (feature learning), following by final fully connected layer (classification).
- People have built increasingly deep CNNs, which have performed increasingly well. Image classification problem is essentially solved.
- Many other image processing tasks can be addressed with CNNs.