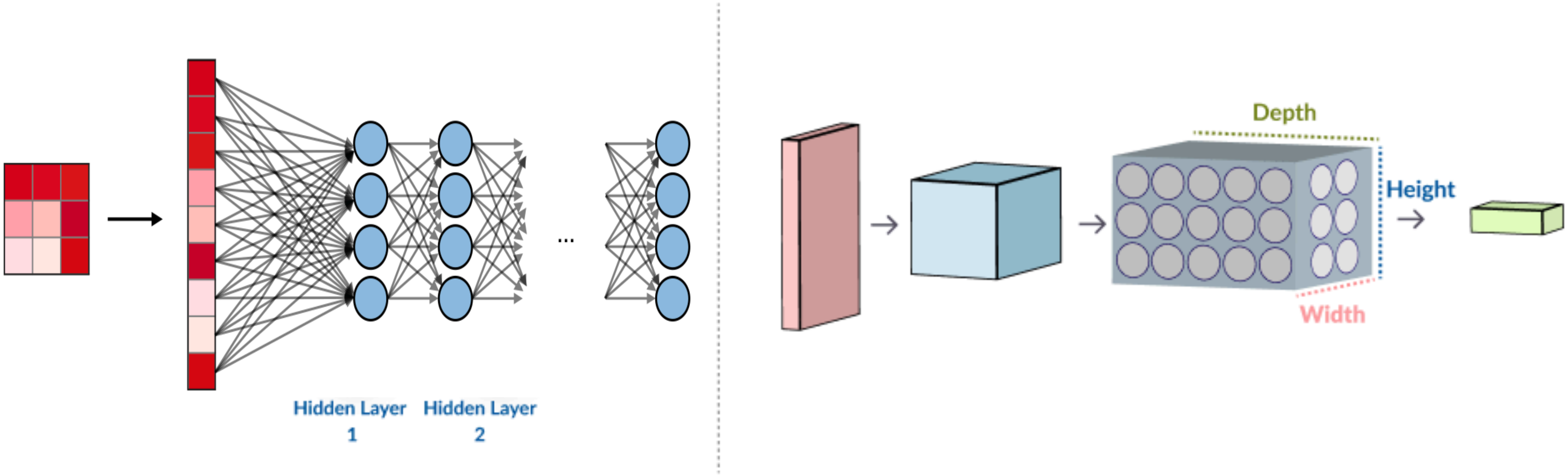
Convolution Neural Network

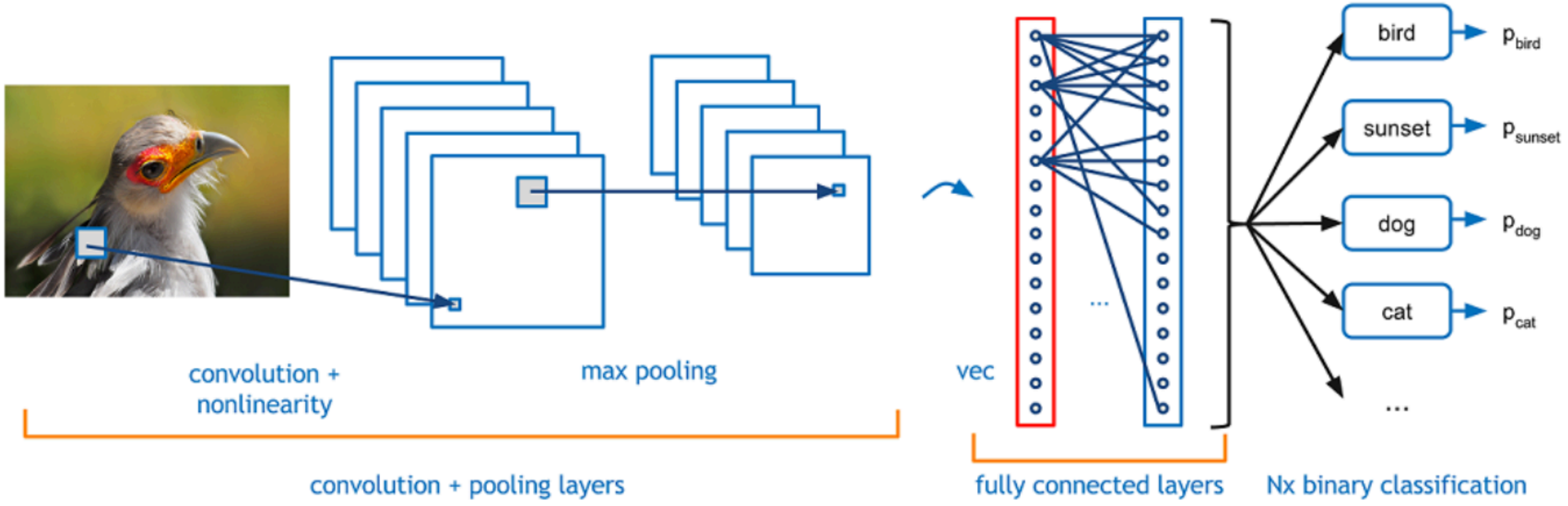
- Pramod Divakarmurthy

Traditional Neural Network
Vs
Convolution Neural Network

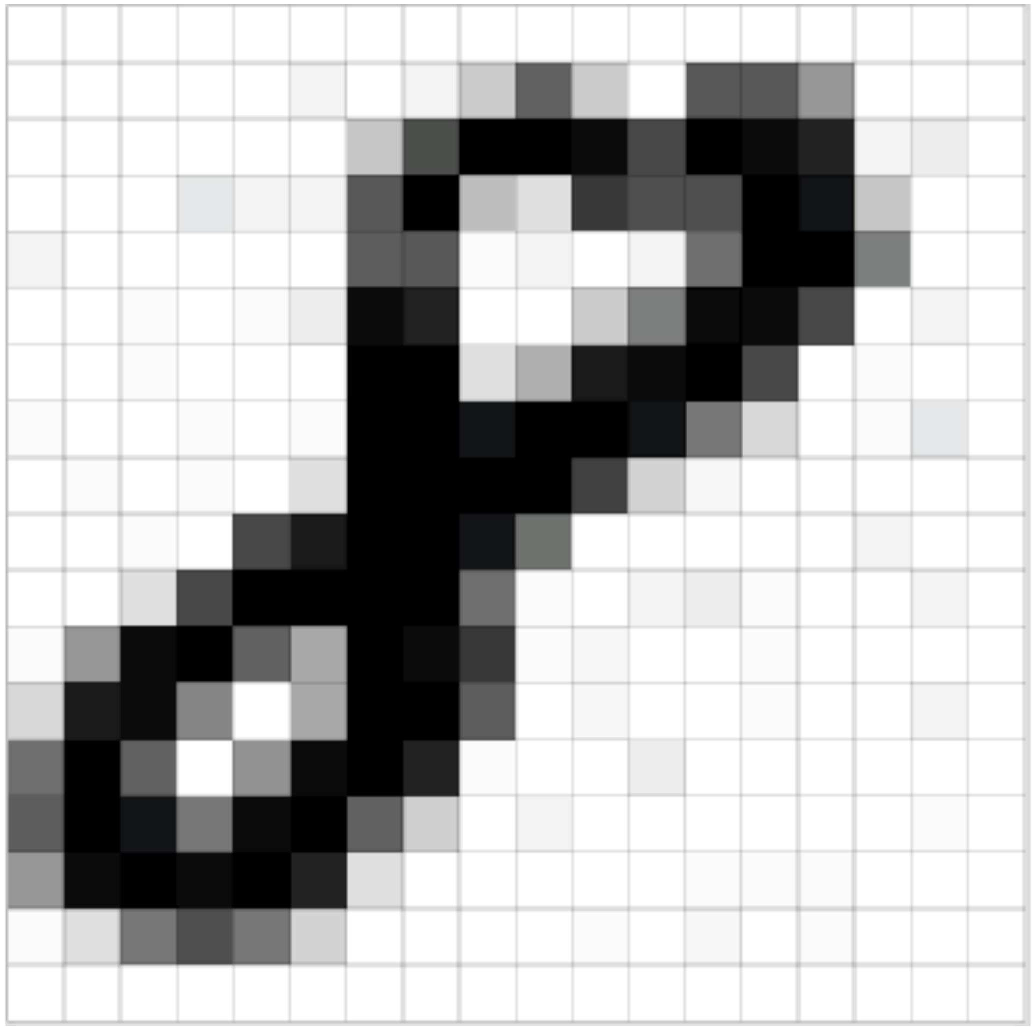


A traditional neural network structure (left) VS. 3-dimensional CNN structure (right)

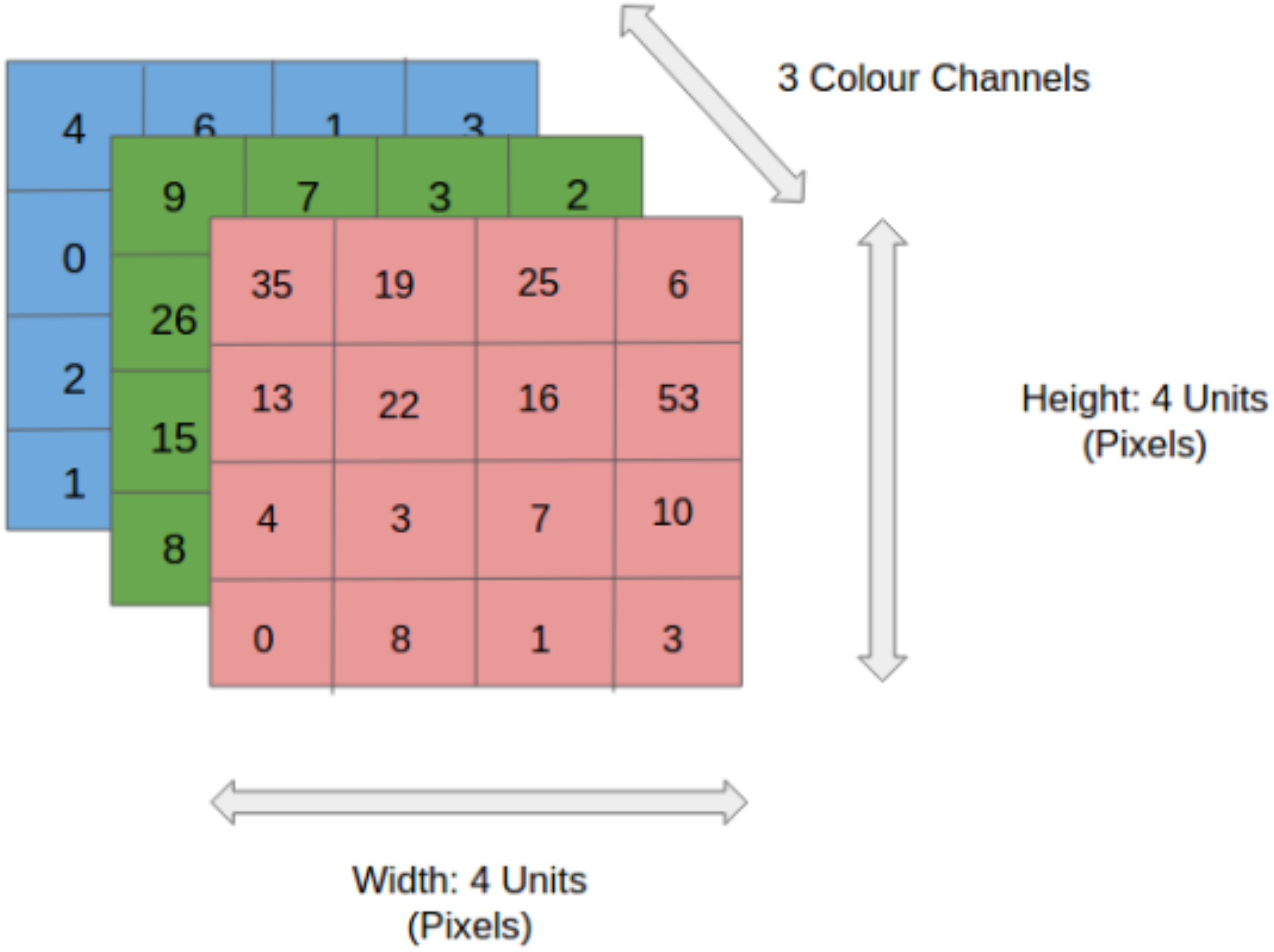
CNN And Layers



1. Input Layer - An Image is a matrix of pixel values



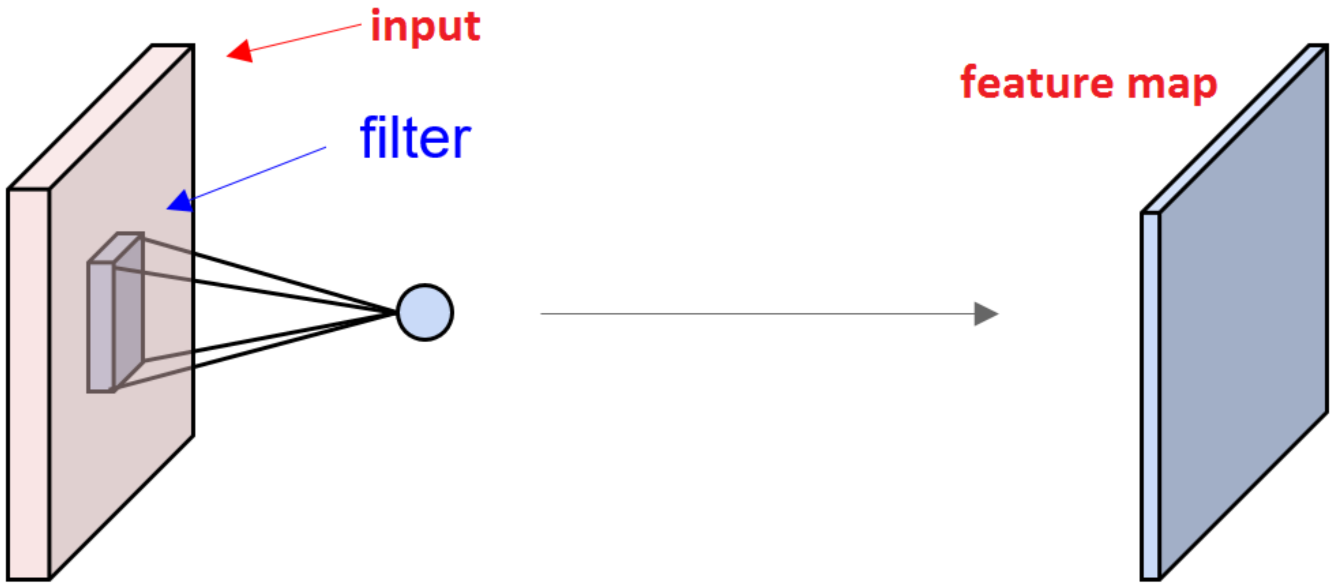
GrayScale

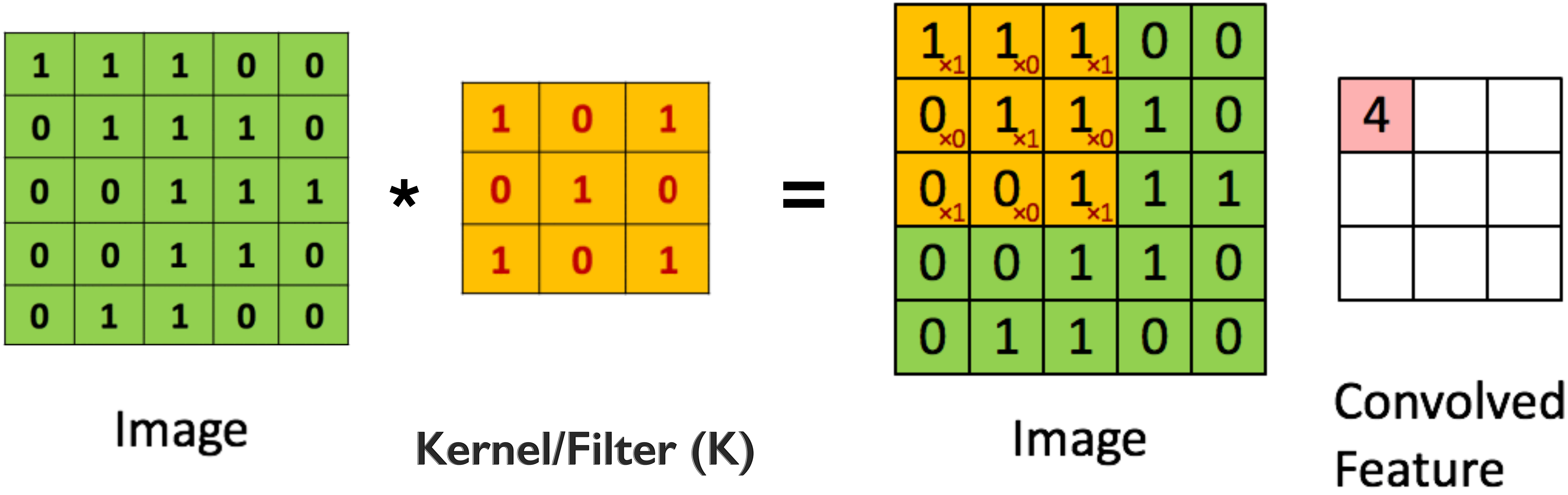


35	19	25	6
13	22	16	53
4	3	7	10
0	8	1	3

RGB

2. Convolution Layer — Convolution





1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Image

*

1	0	1
0	1	0
1	0	1

Kernel/Filter (K)

=

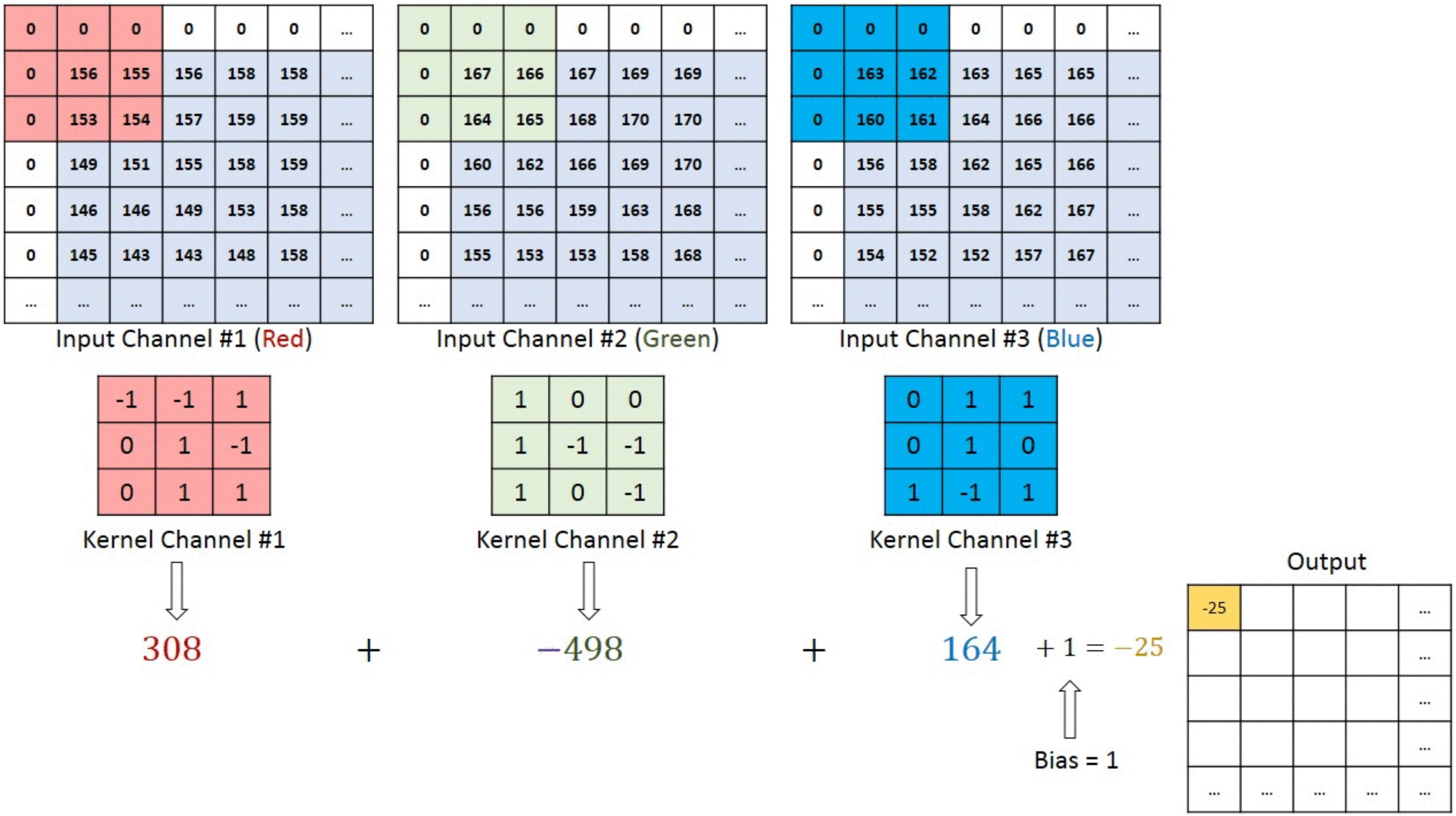
$1_{\times1}$	$1_{\times0}$	$1_{\times1}$	0	0
$0_{\times0}$	$1_{\times1}$	$1_{\times0}$	1	0
$0_{\times1}$	$0_{\times0}$	$1_{\times1}$	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved Feature

Convolution Operation on a MxNx3 image matrix with a 3x3x3 Kernel



0	0	0	0	0	0	...
0	156	155	156	158	158	...
0	153	154	157	159	159	...
0	149	151	155	158	159	...
0	146	146	149	153	158	...
0	145	143	143	148	158	...
...	...	...	...	...	...	...

Input Channel #1 (Red)

0	0	0	0	0	0	...
0	167	166	167	169	169	...
0	164	165	168	170	170	...
0	160	162	166	169	170	...
0	156	156	159	163	168	...
0	155	153	153	158	168	...
...	...	...	...	...	...	...

Input Channel #2 (Green)

0	0	0	0	0	0	...
0	163	162	163	165	165	...
0	160	161	164	166	166	...
0	156	158	162	165	166	...
0	155	155	158	162	167	...
0	154	152	152	157	167	...
...	...	...	...	...	...	...

Input Channel #3 (Blue)

-1	-1	1
0	1	-1
0	1	1

Kernel Channel #1

1	0	0
1	-1	-1
1	0	-1

Kernel Channel #2

0	1	1
0	1	0
1	-1	1

Kernel Channel #3

$$308 + -498 + 164 + 1 = -25$$

Bias = 1

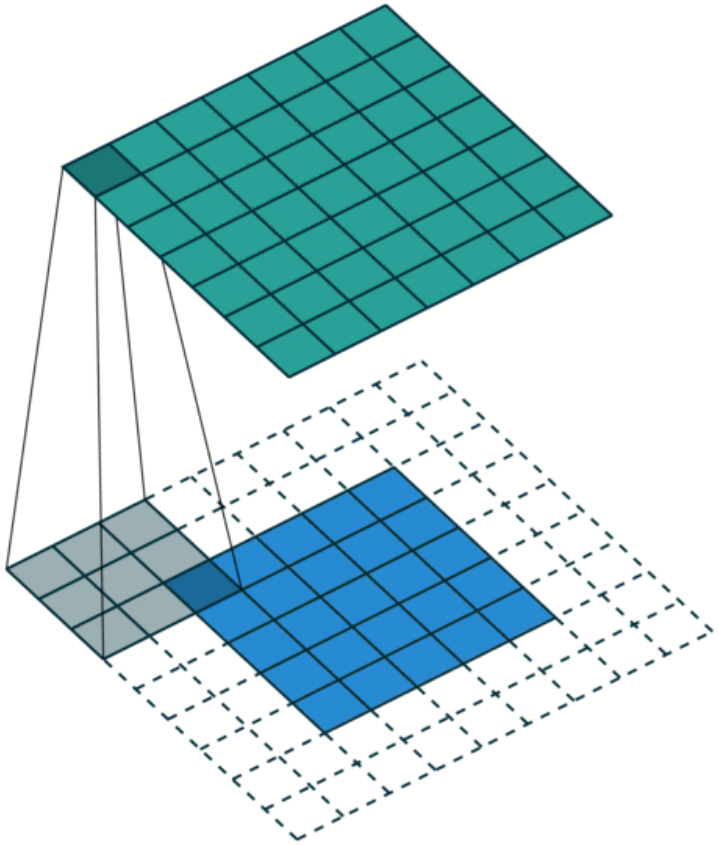
Output

-25				...
				...
				...
				...
...	...	...	...	...

Padding

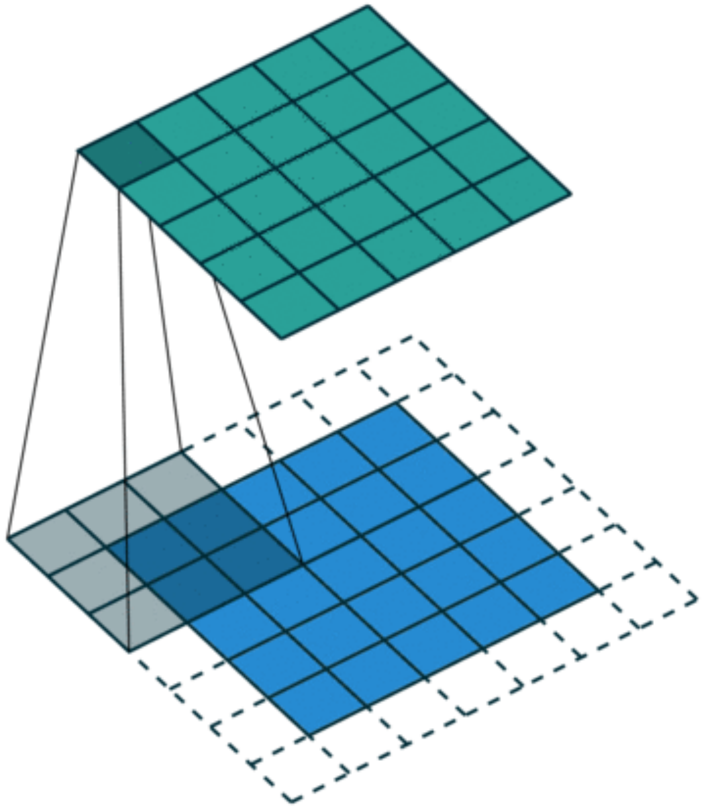
- Pixels in the middle are used more often than pixels on corners and edges
- Information on the borders of images are not preserved

Full Padding



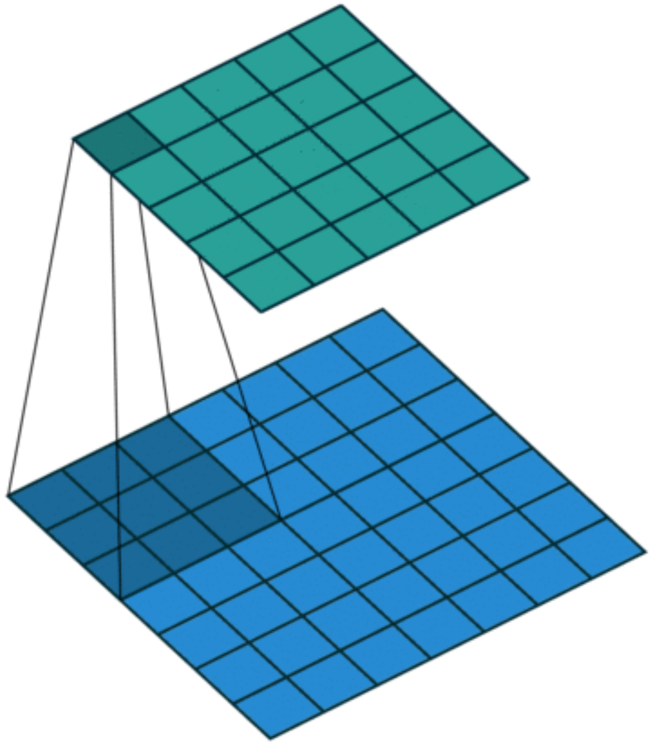
Output size
==
(m+k-1) * (m+k-1)

Same Padding



Output size
==
Image size

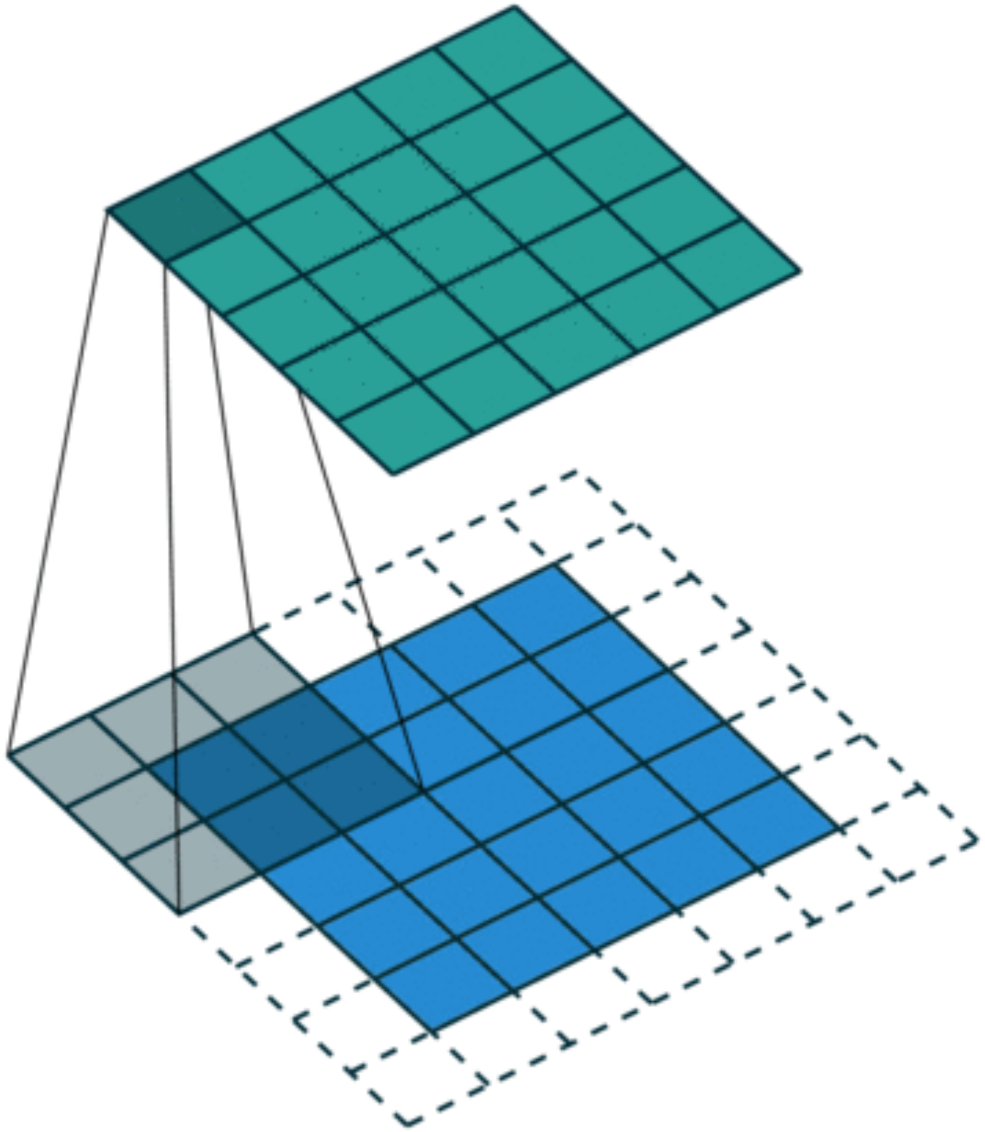
Valid Padding



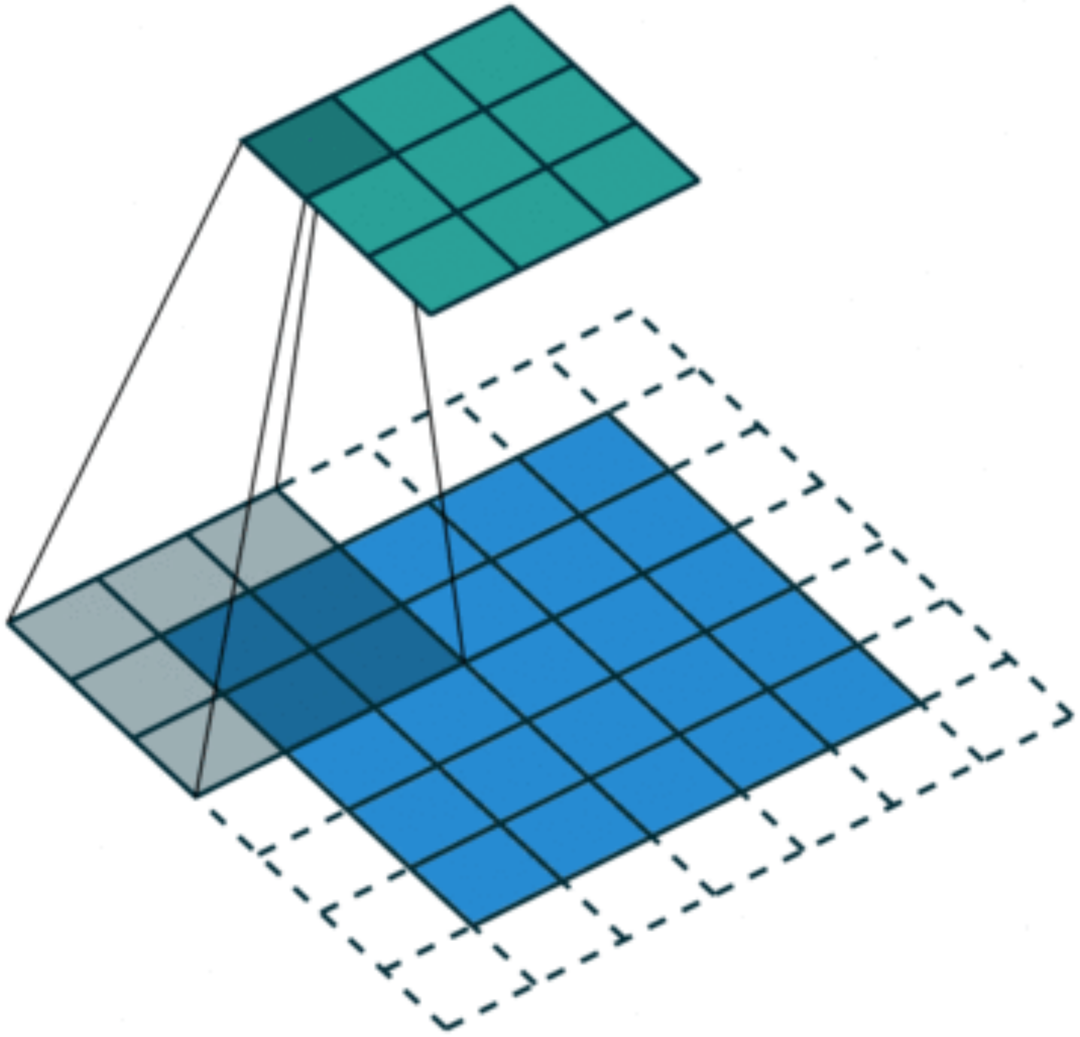
Output size
==
(m-k+1)* (m-k+1)

Stride

- The number of pixels by which we slide our filter matrix over the input matrix
- Having a larger stride will produce smaller feature maps.



Stride Length = 1 (Default)



Stride Length = 2

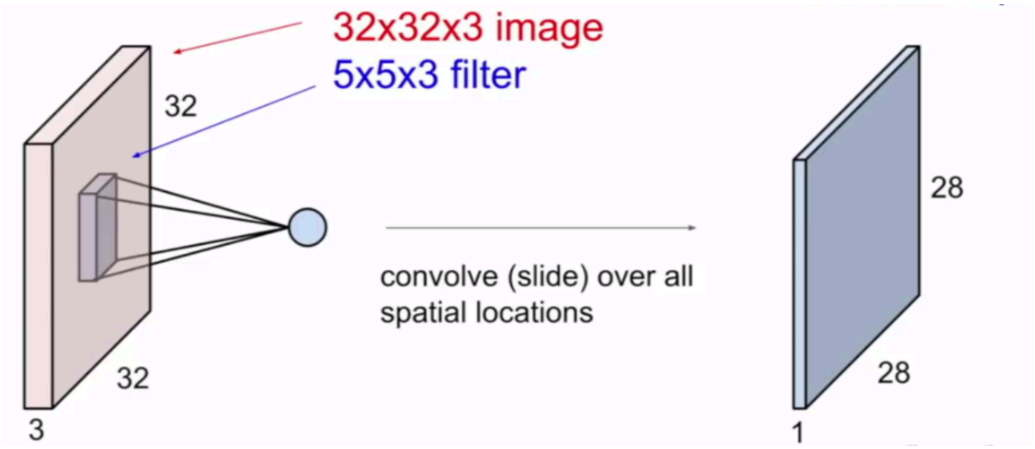
Size of Activation Map

There is a formula which is used in determining the dimension of the activation maps:

(N + 2P — F)/ S + 1

where

- N = Dimension of image (input) file
- P = Padding
- F = Dimension of filter
- S = Stride

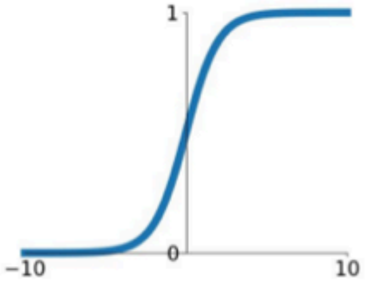


Activation Map Size = (32 + 0 - 5)/1 + 1
= 28

Activation Function

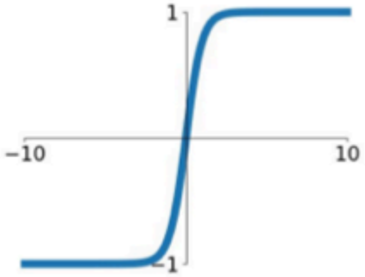
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



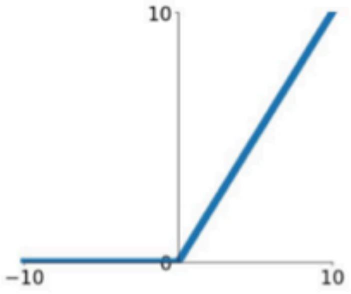
tanh

$\tanh(x)$



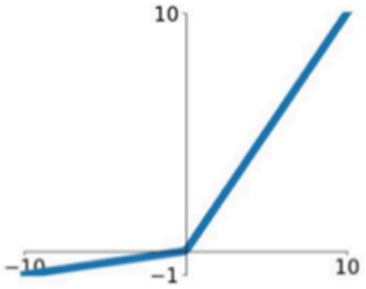
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

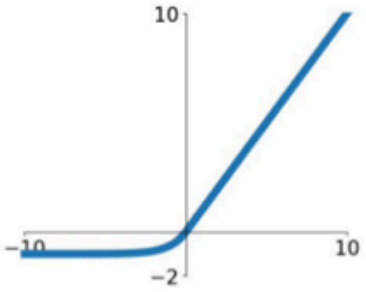


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

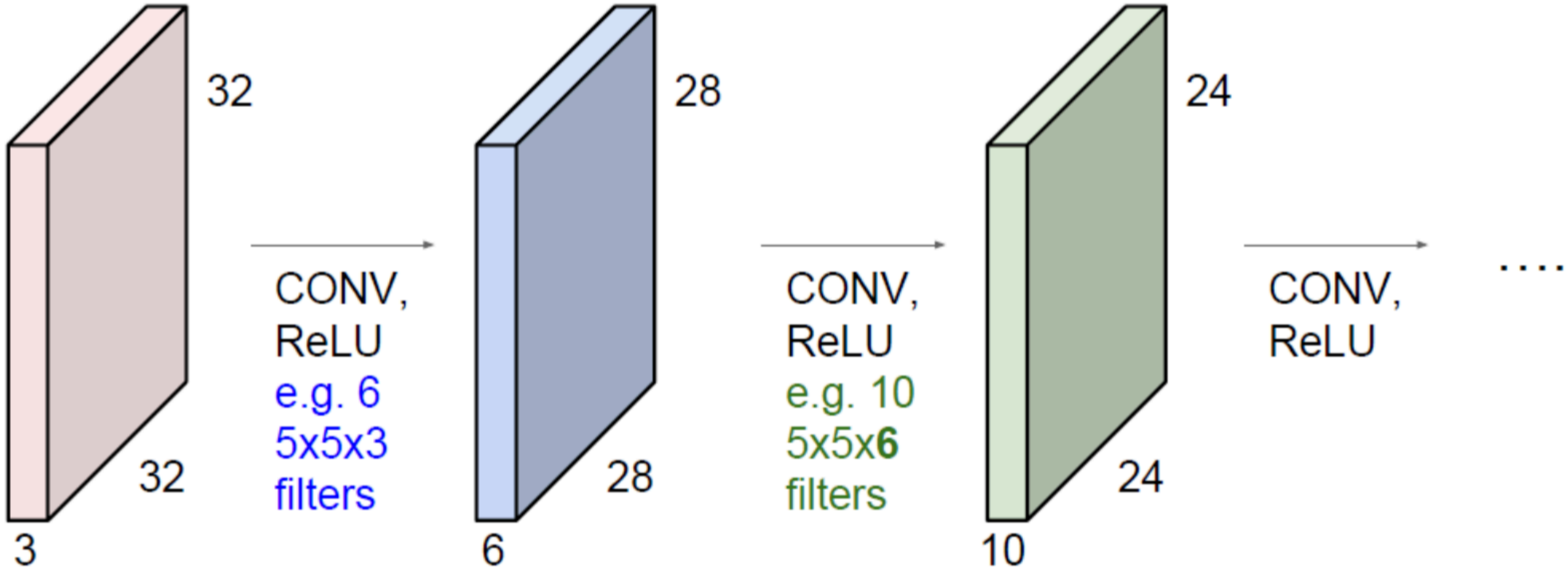
ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



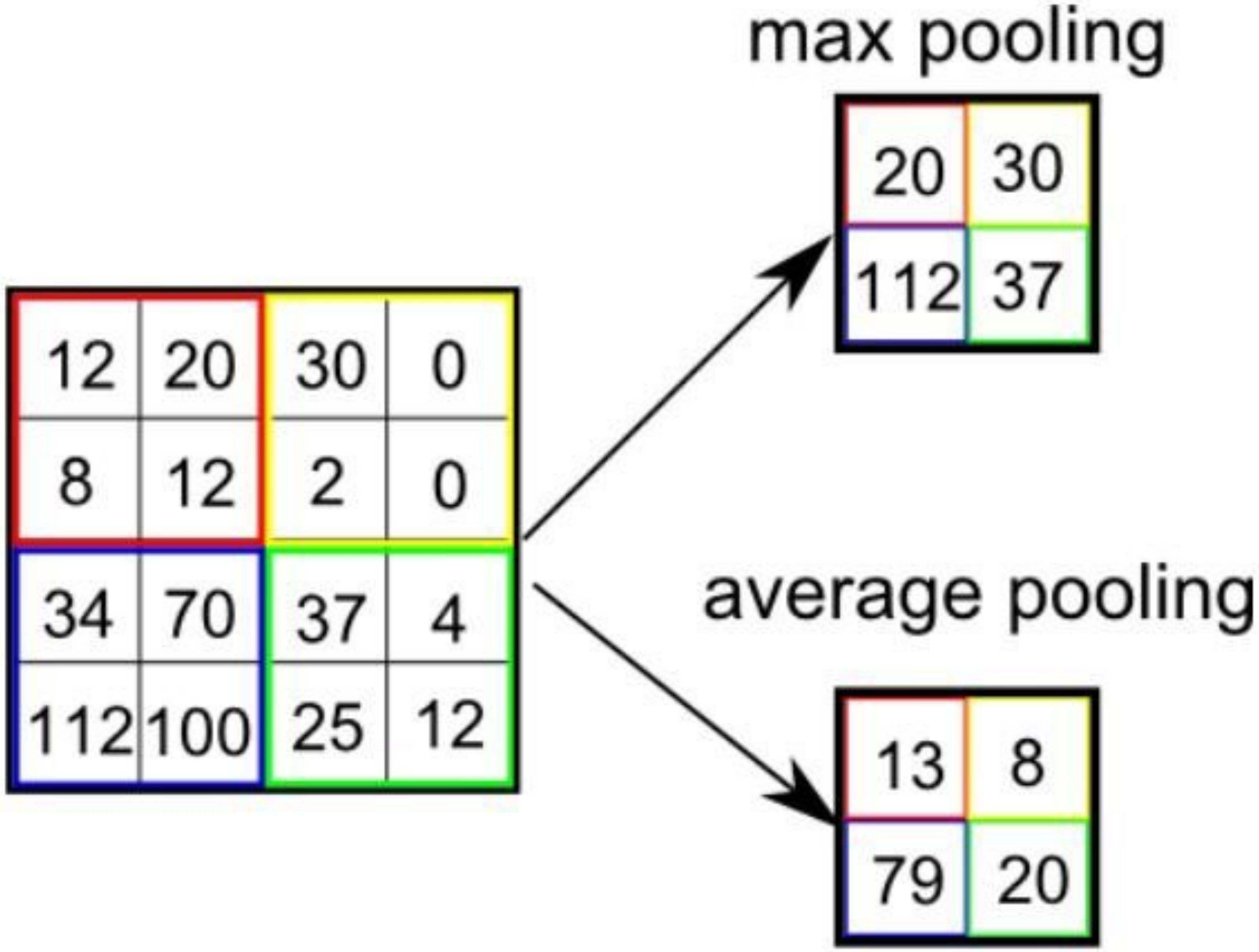
Stacking Convolution Layers

ConvNet is a sequence of convolutional layer, interspersed with activation function



3. Pooling Layer

- To progressively reduce the spatial size (downsampling) of the Convolved Feature
- Reduces the number of parameters and computational power
- **Extracting dominant features** which are rotational and positional invariant



12	20	30	0
8	12	2	0
34	70	37	4
112	100	25	12

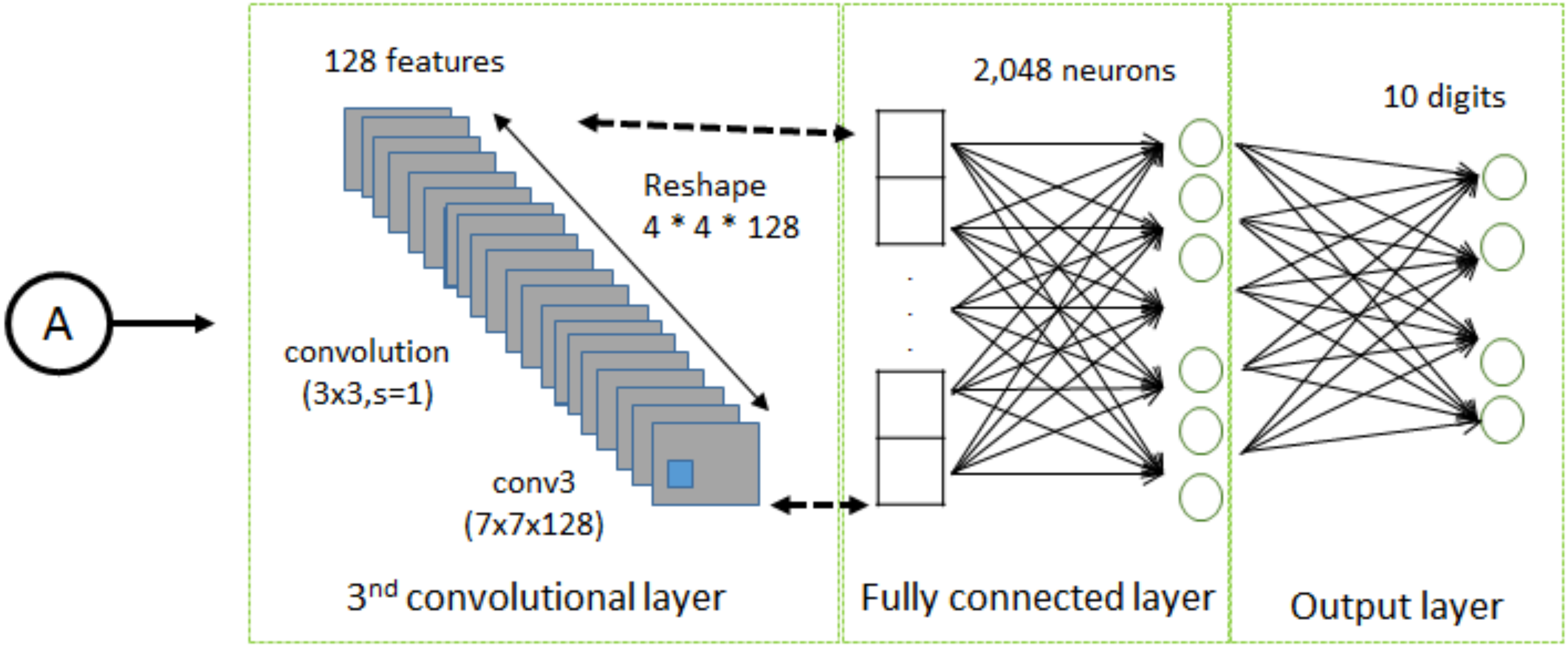
max pooling

20	30
112	37

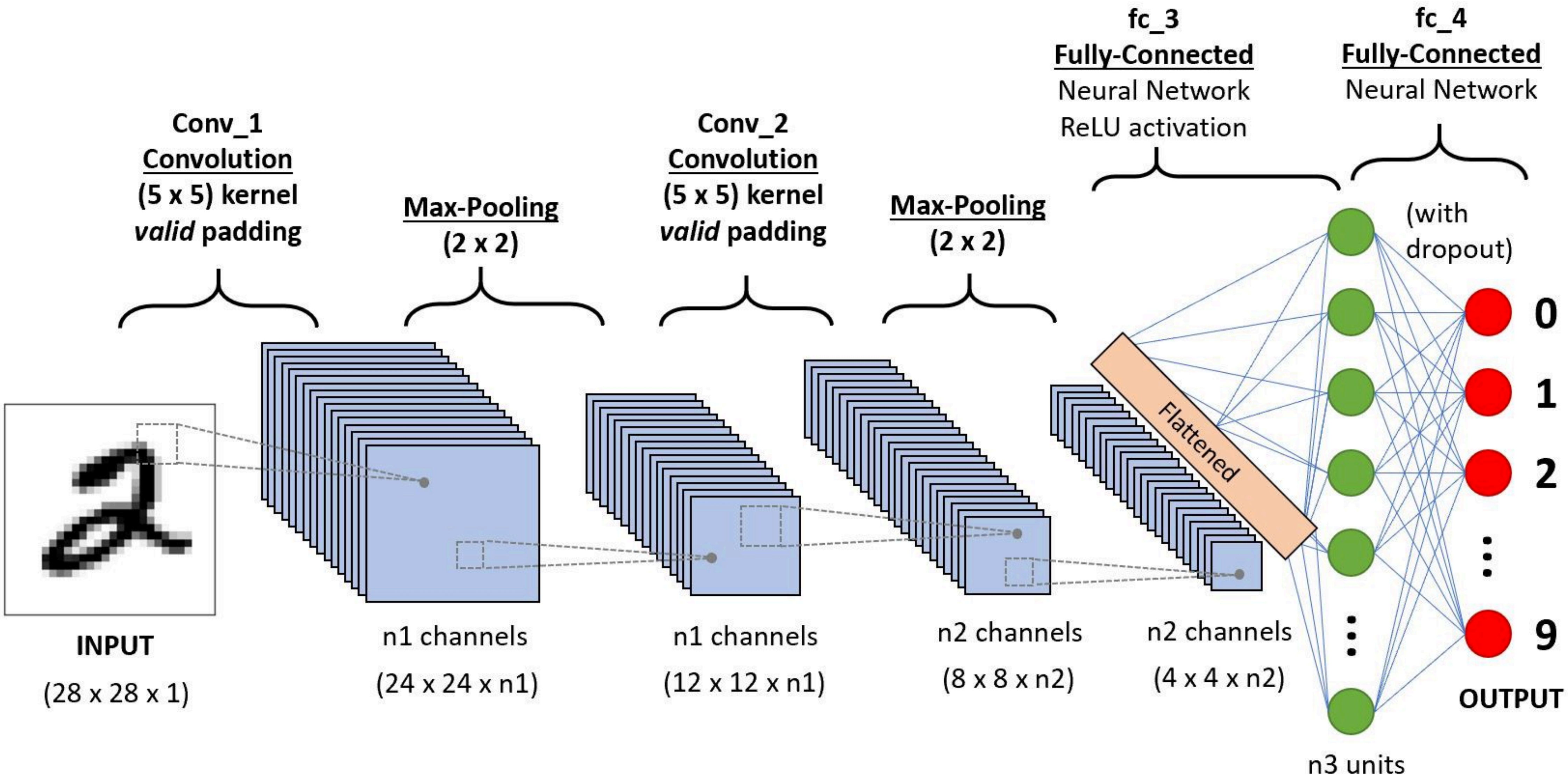
average pooling

13	8
79	20

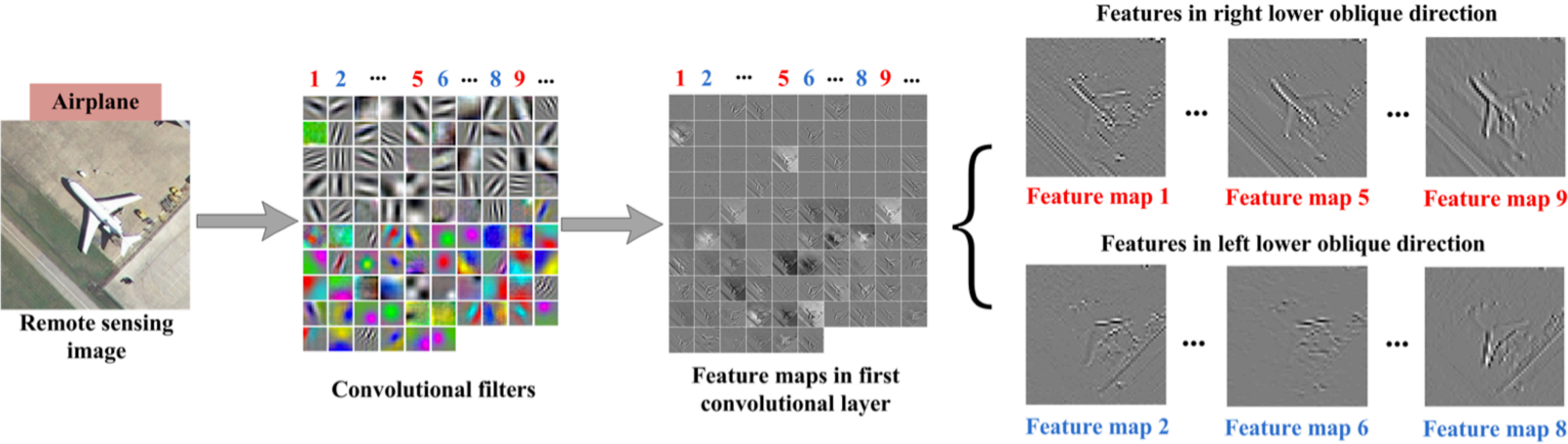
4. Fully Connected Layer (FC Layer)

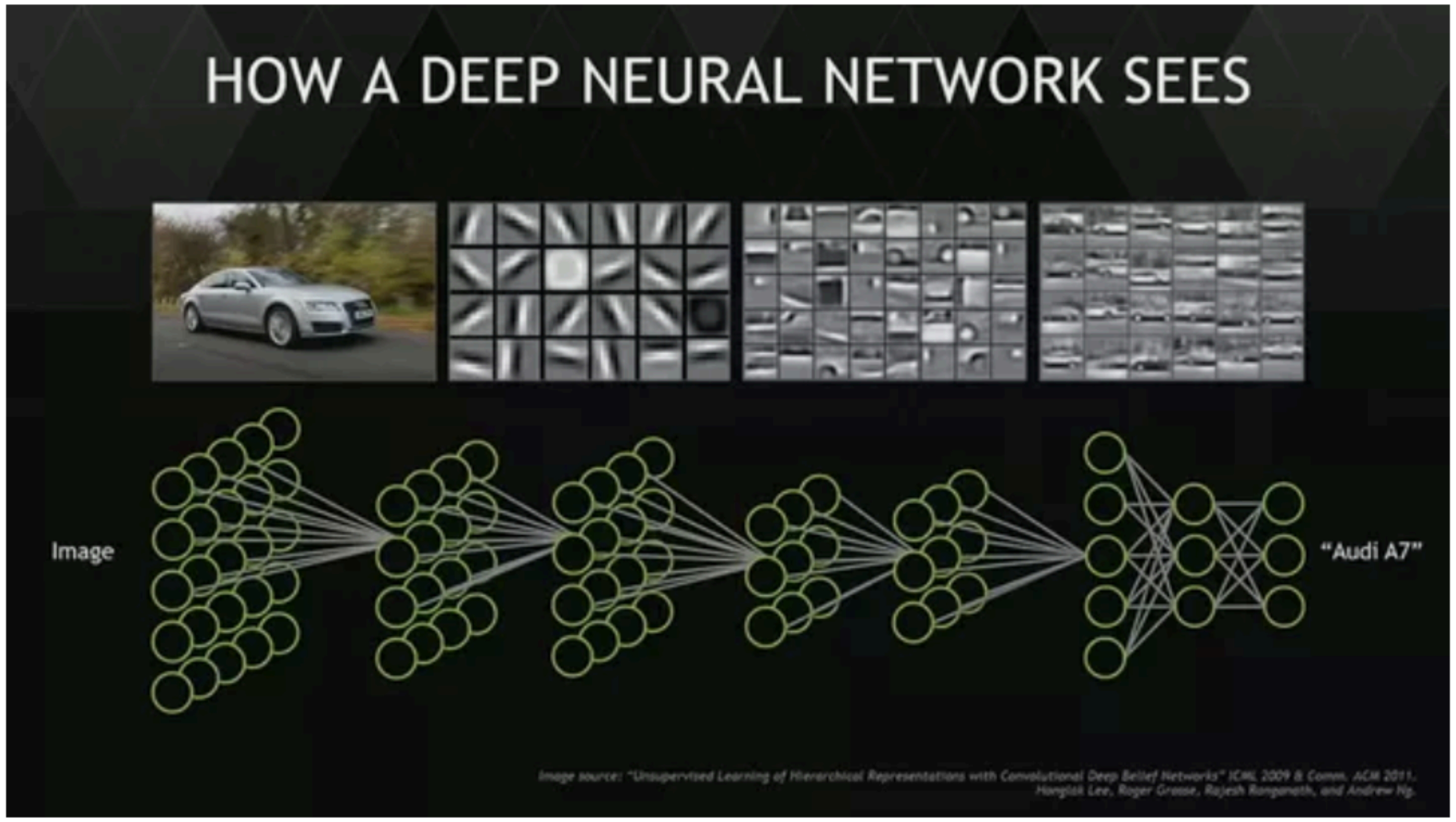


CNN Sequence To Classify Handwritten Digits



Visualizing Filters & Feature Maps

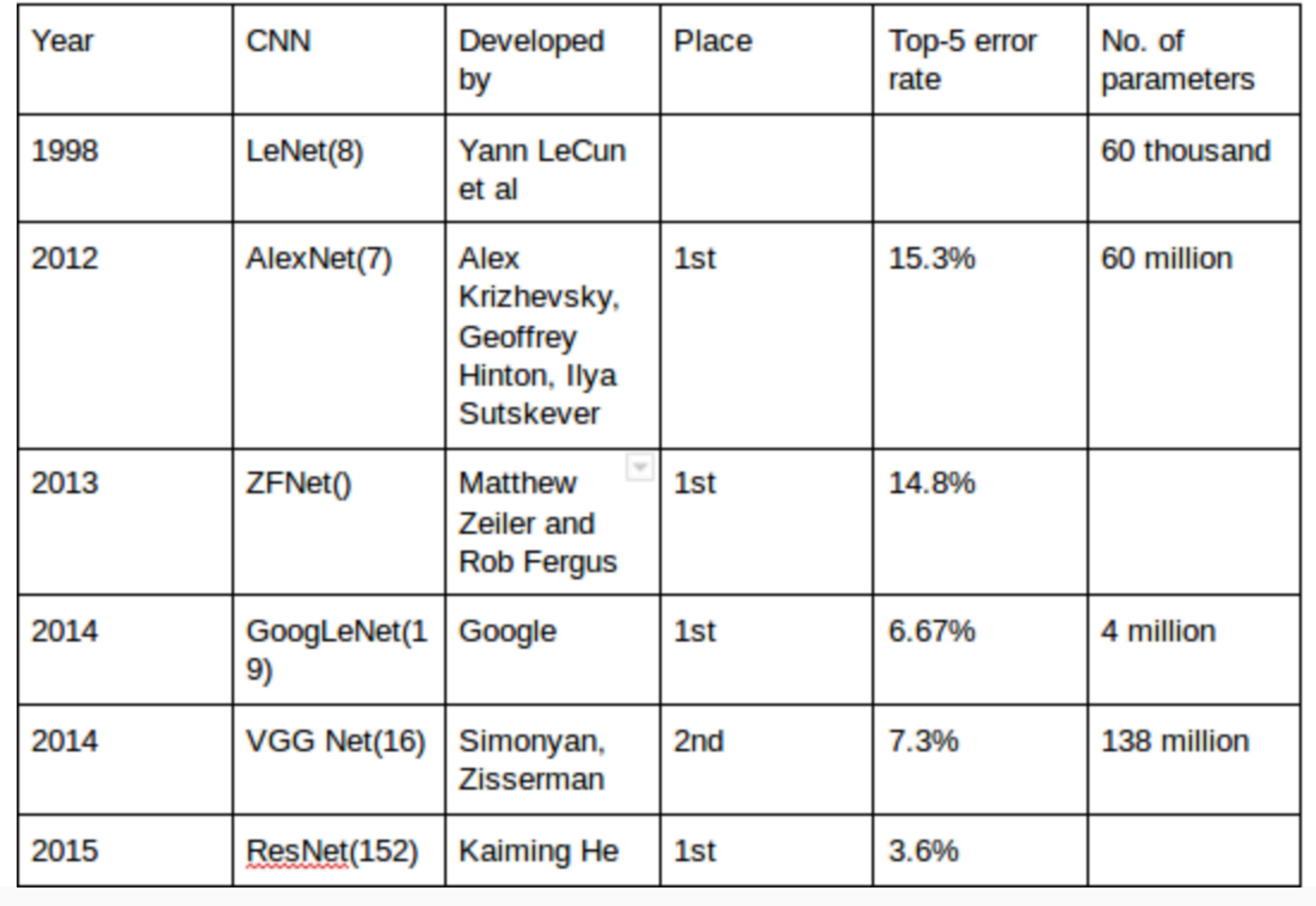




HOW A DEEP NEURAL NETWORK SEES

Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.

Popular Architectures



Year	CNN	Developed by	Place	Top-5 error rate	No. of parameters
1998	LeNet(8)	Yann LeCun et al			60 thousand
2012	AlexNet(7)	Alex Krizhevsky, Geoffrey Hinton, Ilya Sutskever	1st	15.3%	60 million
2013	ZFNet()	Matthew Zeiler and Rob Fergus	1st	14.8%	
2014	GoogLeNet(19)	Google	1st	6.67%	4 million
2014	VGG Net(16)	Simonyan, Zisserman	2nd	7.3%	138 million
2015	ResNet(152)	Kaiming He	1st	3.6%	