Convolutional Neural Networks (CNN)

Rina BUOY

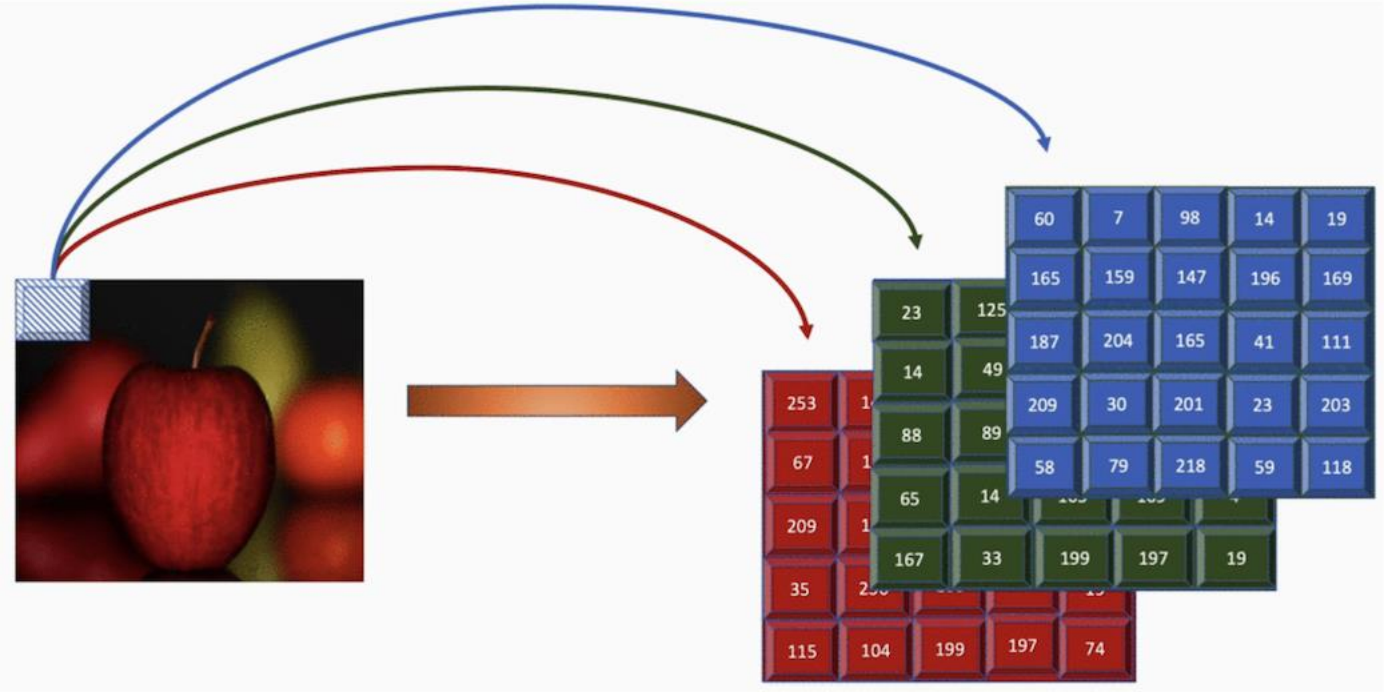
AMERICAN UNIVERSITY OF PHNOM PENH

STUDY LOCALLY. LIVE GLOBALLY.

https://developer.ibm.com/articles/introduction-to-convolutional-neural-networks/

Recap – CNN

RGB Image



60
7
98
14
19
165
159
147
196
169
187
204
165
41
111
209
30
201
23
203
58
79
218
59
118
23
125
14
49
88
89
65
14
167
33
199
197
19
253
67
209
35
115
104
199
197
74

Representing RGB Image

253	144	120	251	41
67	100	32	241	23
209	118	124	27	59
35	236	105	169	19
115	104	199	197	74

23	125	233	201	98
14	49	32	241	59
88	89	124	27	19
65	14	105	169	4
167	33	199	197	19

60	7	98	14	19
165	159	147	196	169
187	204	165	41	111
209	30	201	23	203
58	79	218	59	118

Convolution

- In CNN, convolutions are applied to extract the prominent features within the images.
- In summary, for an input image of size *(n, n)* and a filter of size *(m,m)*, the resulting output is of size *(n-m+1,(n-m+1)*.

$$\begin{bmatrix} 60 & 7 & 98 & 14 & 19 \\ 165 & 159 & 147 & 196 & 169 \\ 187 & 204 & 165 & 41 & 111 \\ 209 & 30 & 201 & 23 & 203 \\ 58 & 79 & 218 & 59 & 118 \end{bmatrix} * \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} =$$

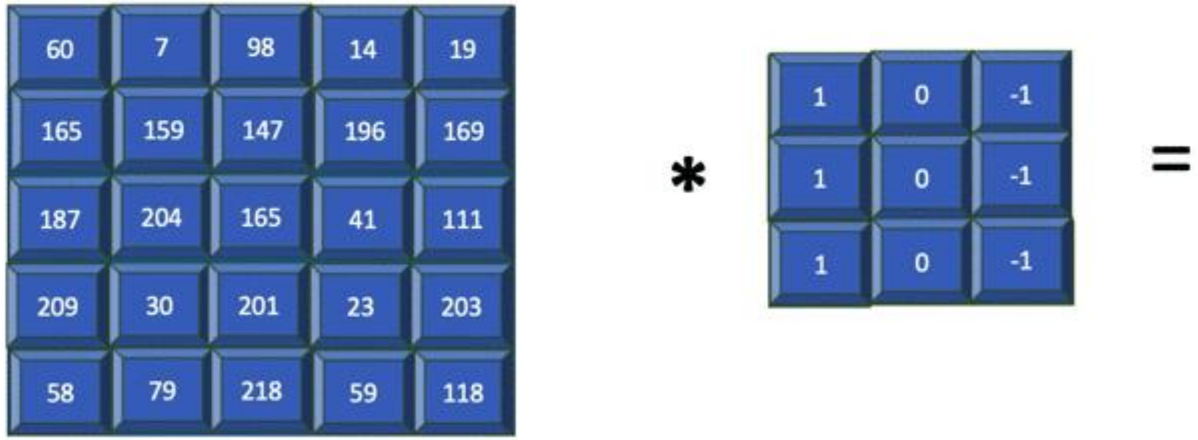
Stride

- Sliding of the filter over the input was done one step at a time. This is referred to as striding. The following example shows the same convolution, but strided with 2 steps.
- In summary, for an input image of size *(n,n)* and a filter of size *(m,m)* with stride=k, the resulting output will be of size *((n-m)/k+1),((n-m)/k+1)*.

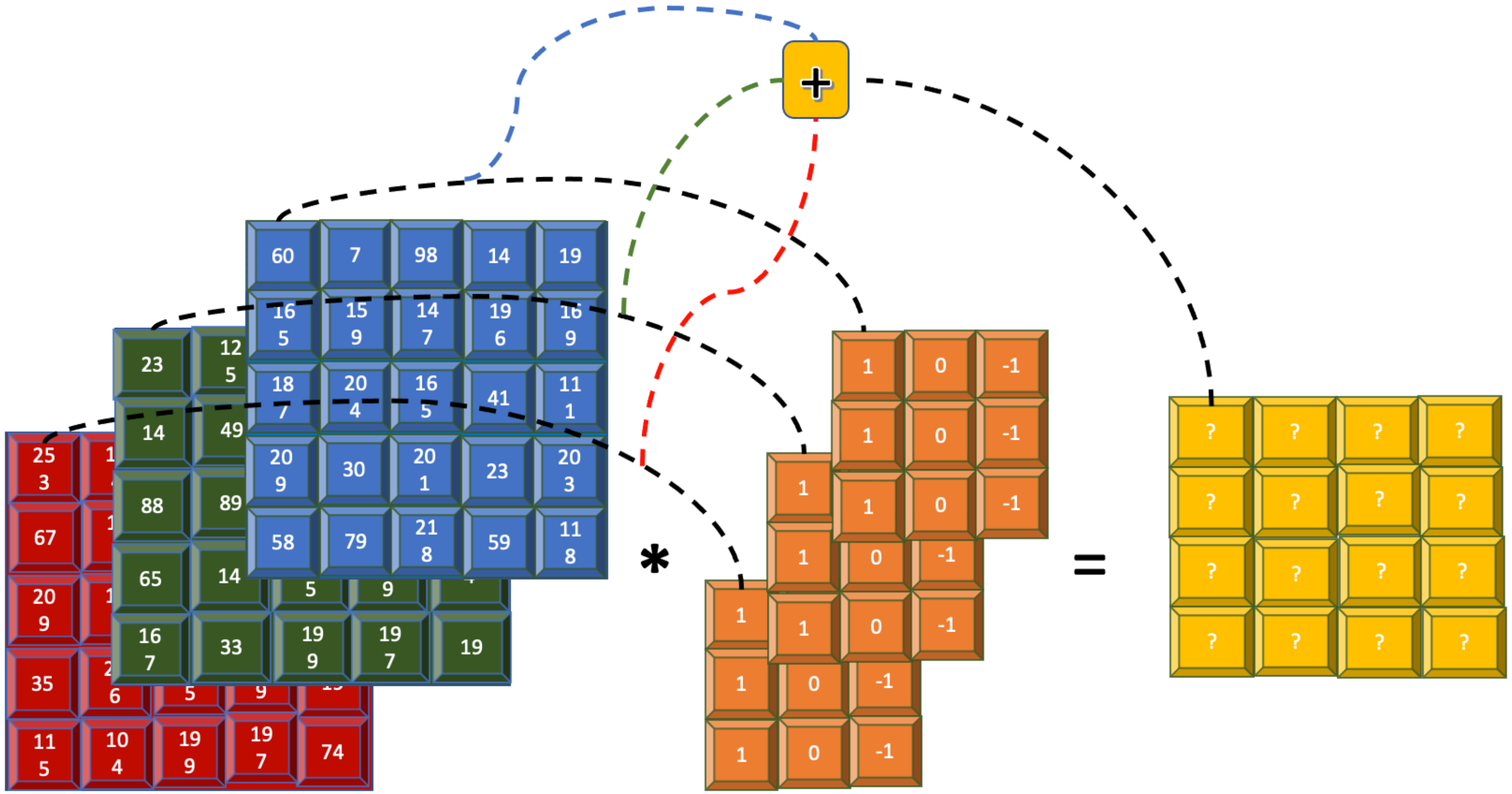
$$
\begin{bmatrix} 60 & 7 & 98 & 14 & 19 \\ 165 & 159 & 147 & 196 & 169 \\ 187 & 204 & 165 & 41 & 111 \\ 209 & 30 & 201 & 23 & 203 \\ 58 & 79 & 218 & 59 & 118 \end{bmatrix} * \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} =
$$

Padding

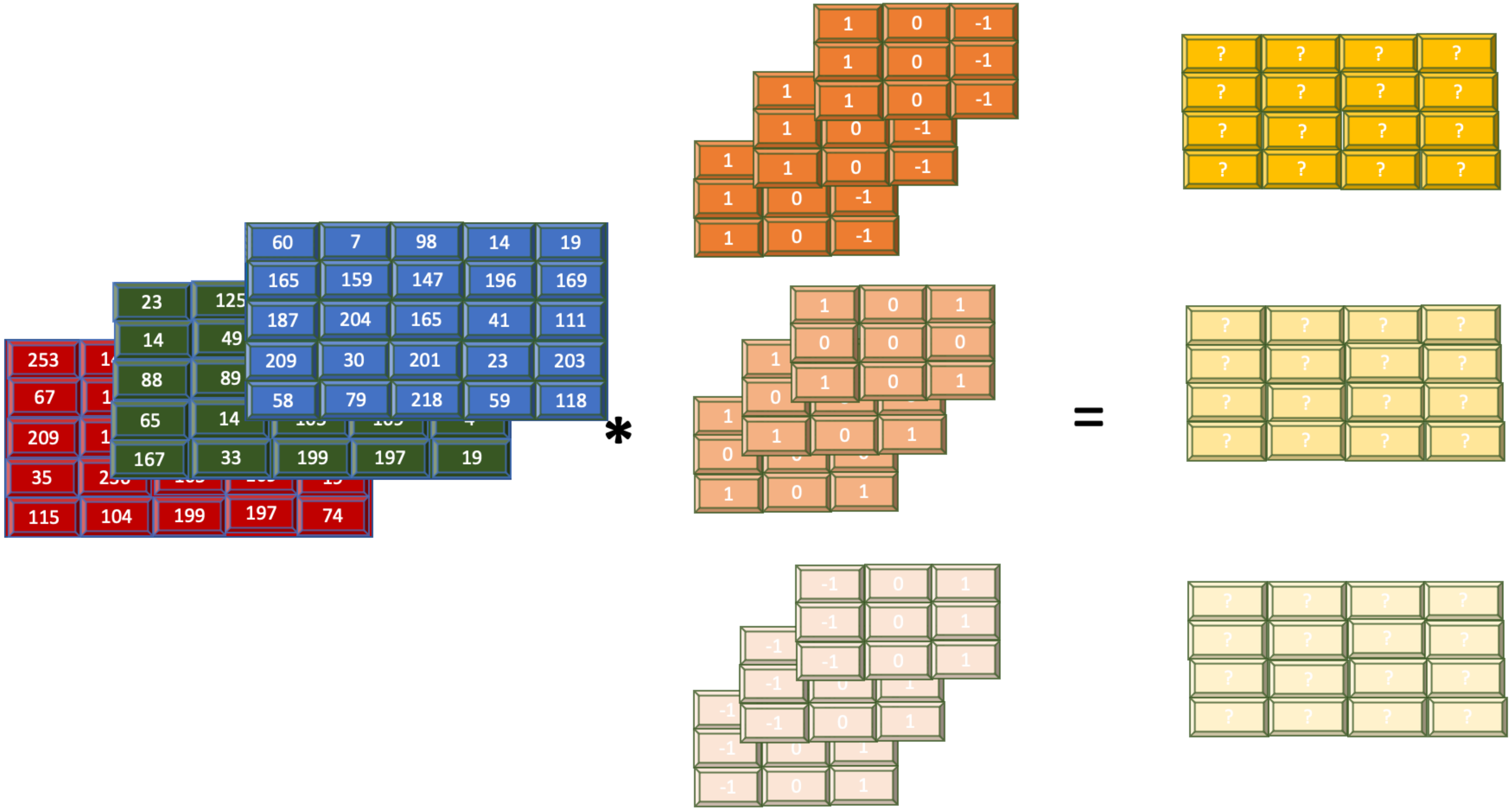
- Padding refers to the process of adding extra layers of zeros to the outer rows and columns of the input array.



How are convolutions applied over the RGB channels



More Filters



Pooling

- Pooling is the process of summarizing the features within a group of cells in the feature map. This summary of cells can be acquired by taking the maximum, minimum, or average within a group of cells.

-166	-20	-44	57	210
370	11	119	111	251
393	48	133	30	260
-313	-130	190	152	123
-109	-158	27	98	82

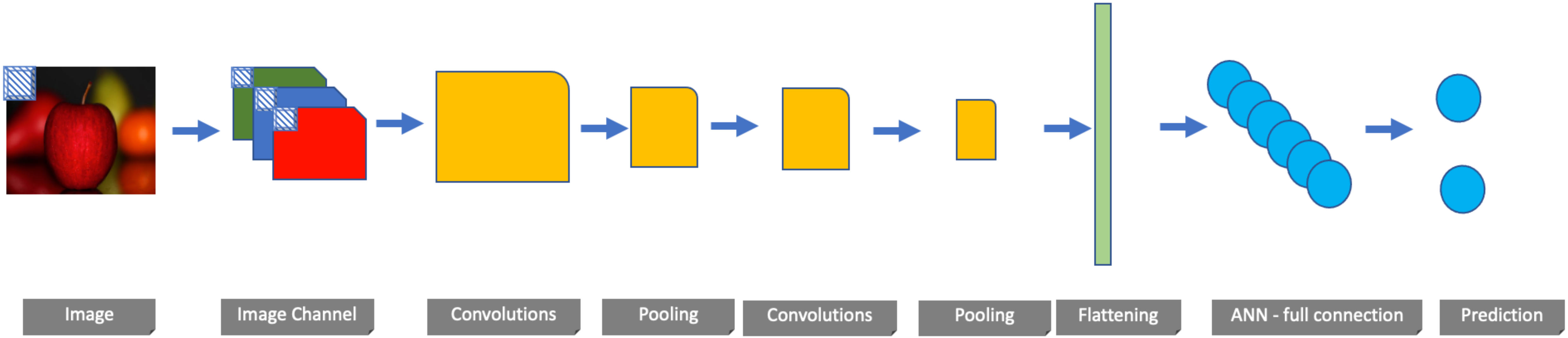
=

Flattening

11	119	119	251
393	119	119	260
393	190	190	260
-109	190	190	152

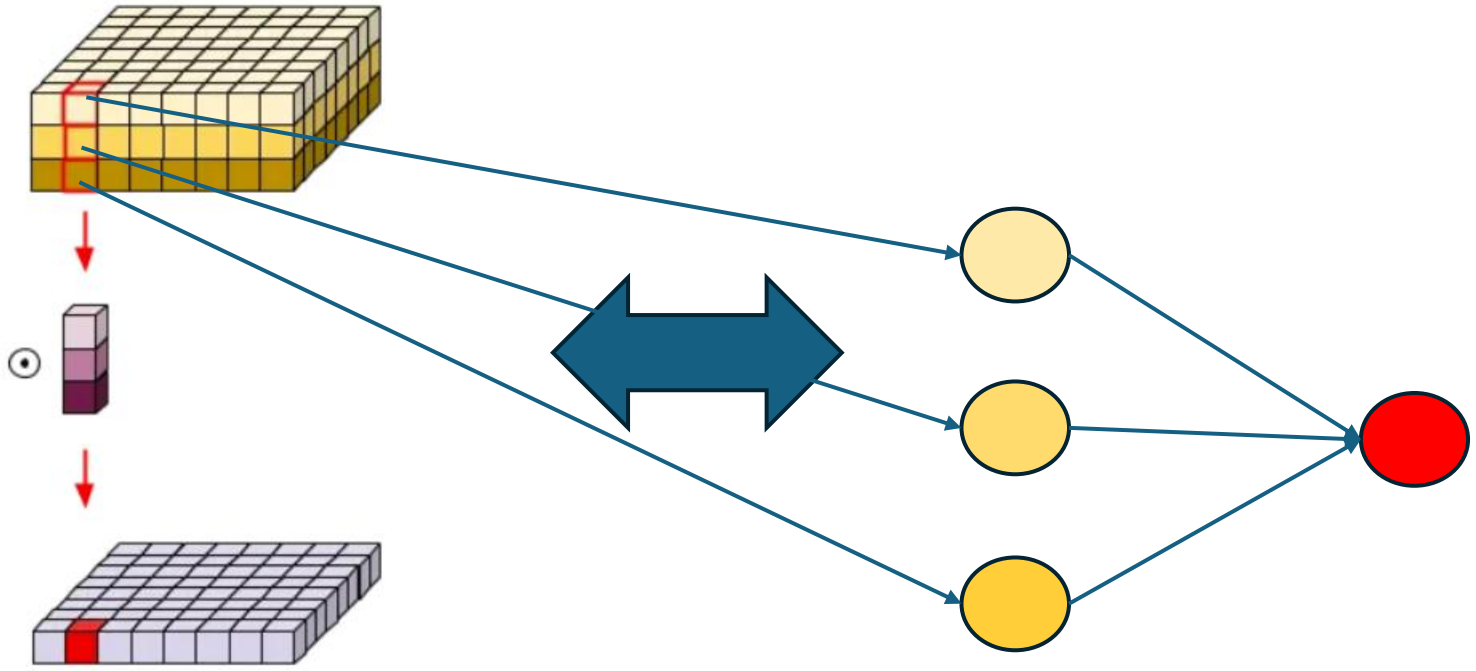
=

CNN Architecture

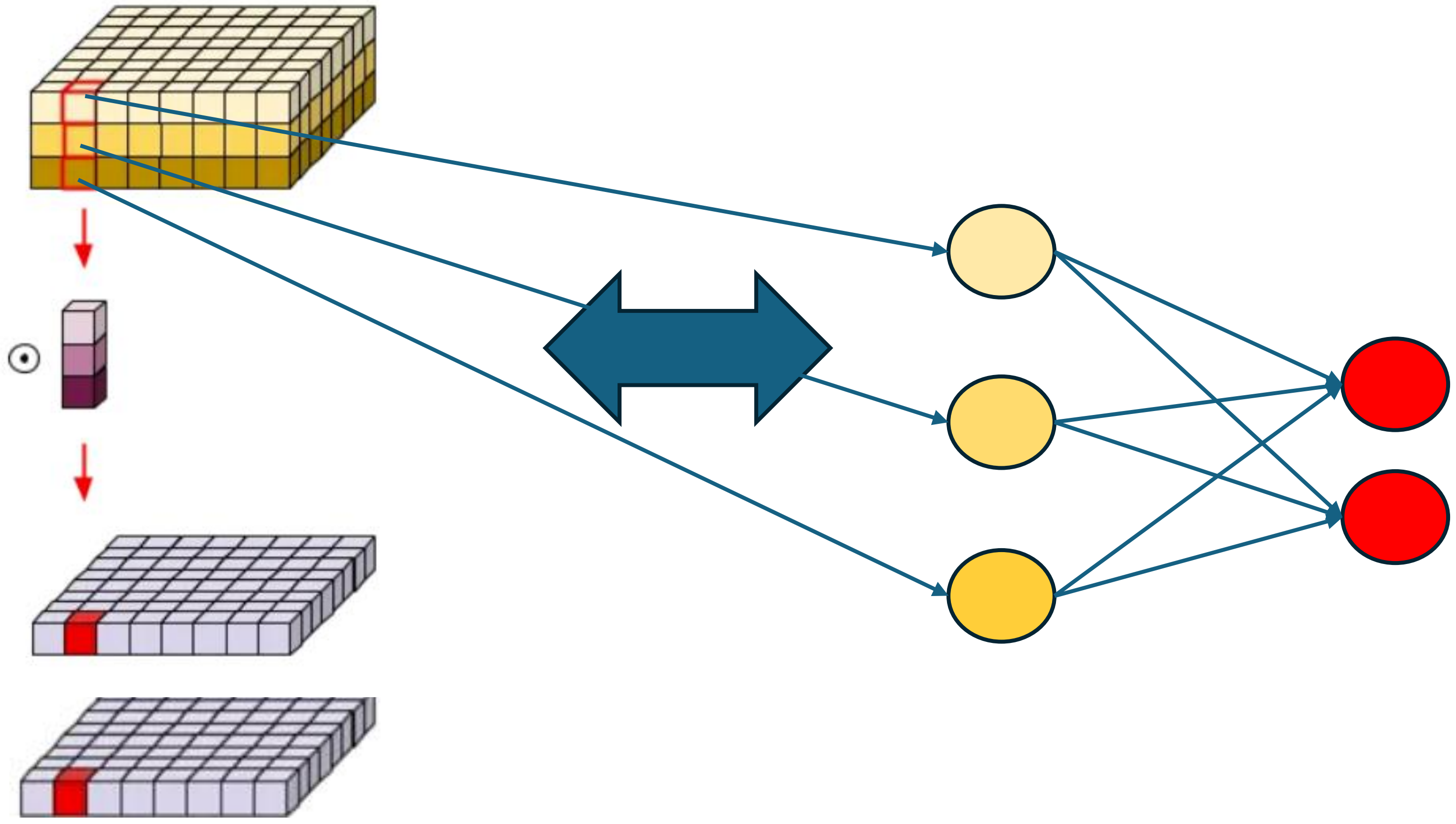


Depth-wise Convolution and Depth-wise Separable Convolution

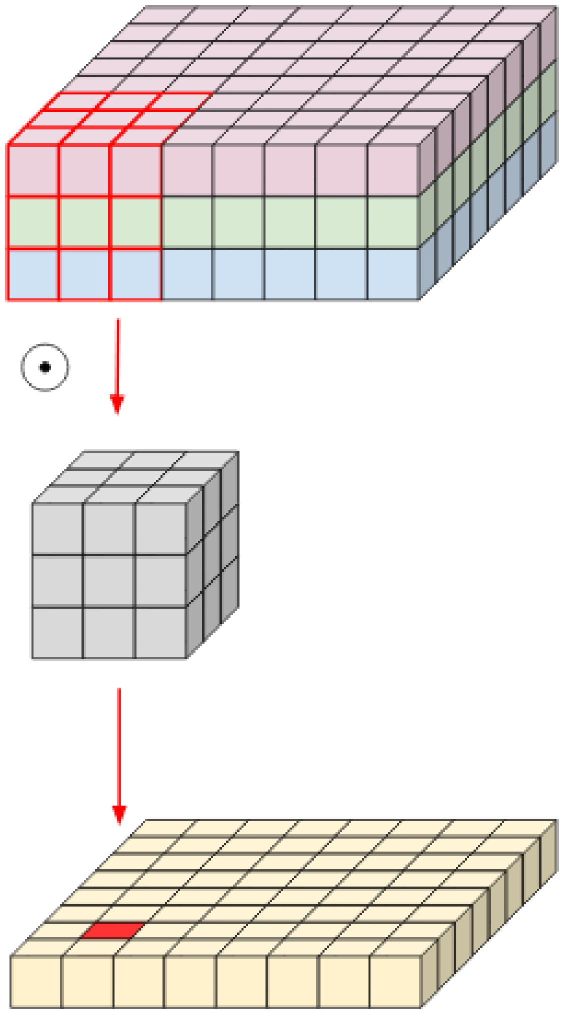
1x1 Convolution – Network in network



1x1 Convolution – Network in network



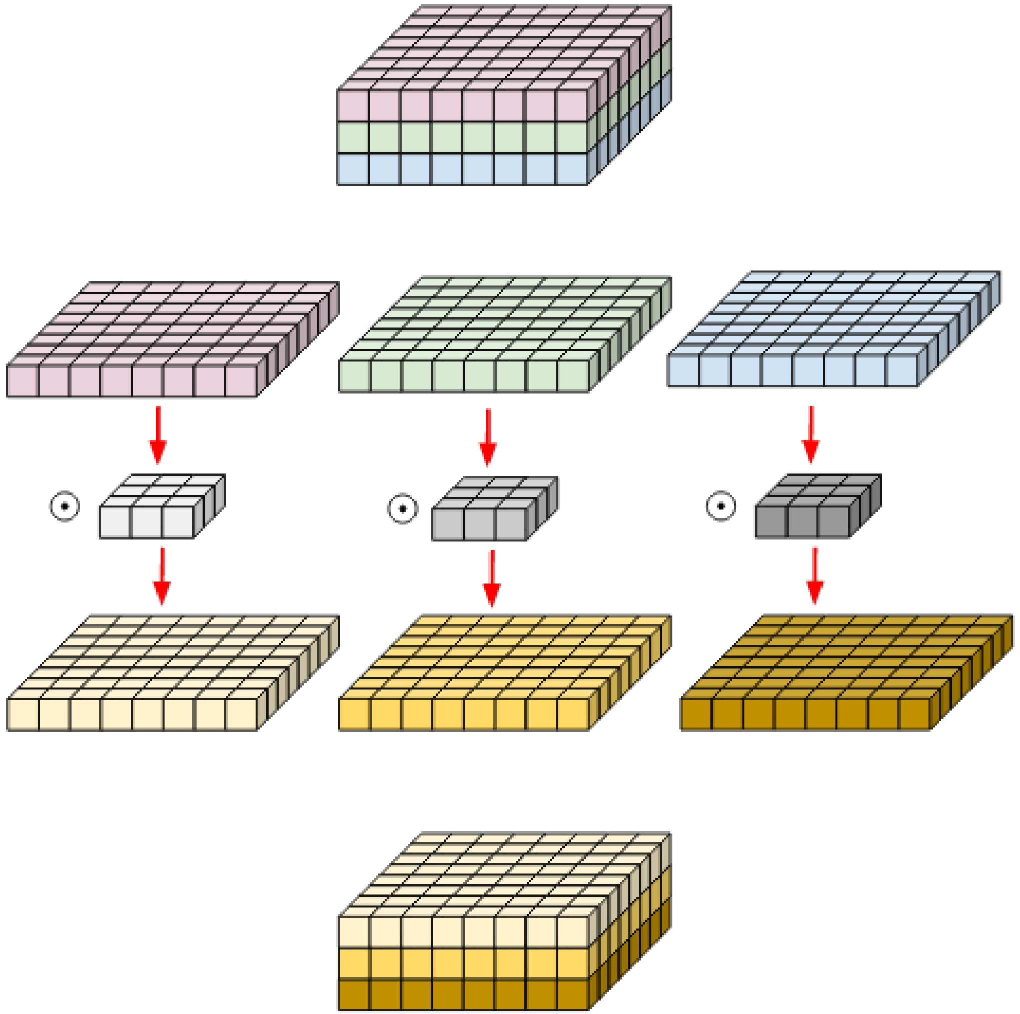
Depth-wise Convolution and Depth-wise Separable Convolution



Parameters / filter : ...?

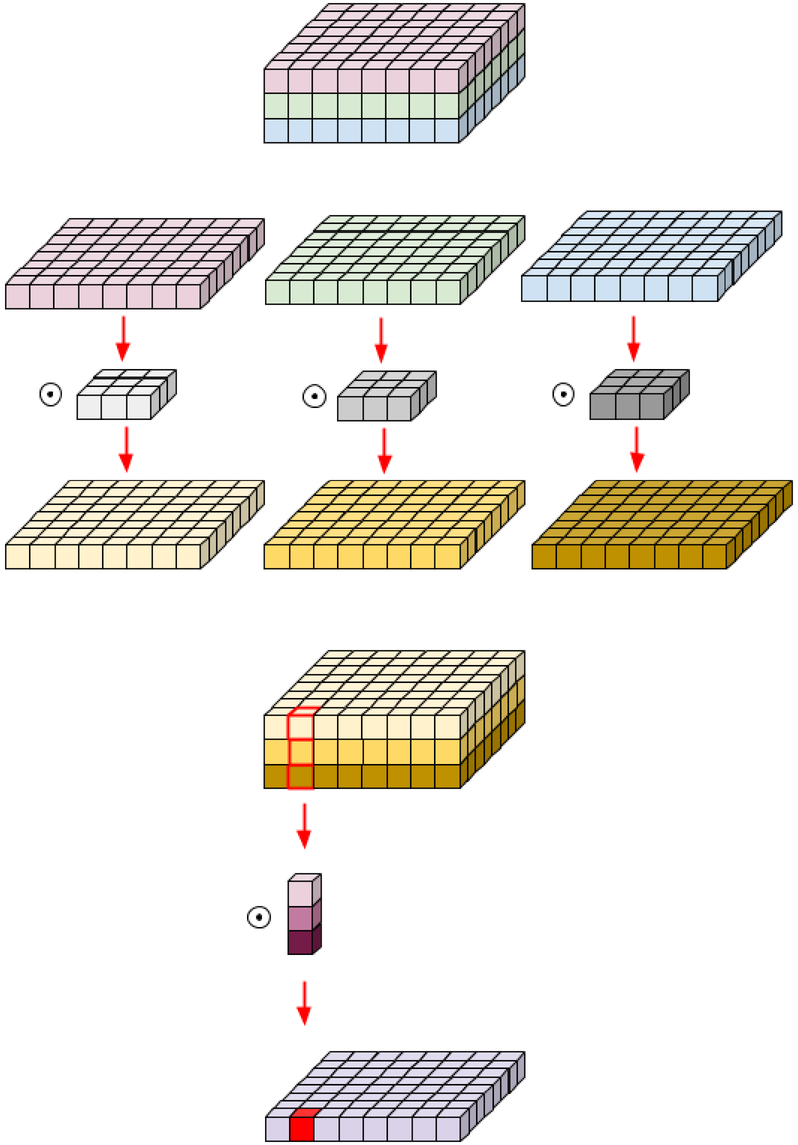
Parameters for n filters : ...?

Depth-wise Convolution



Parameters : ...?

Depth-wise Separable Convolution

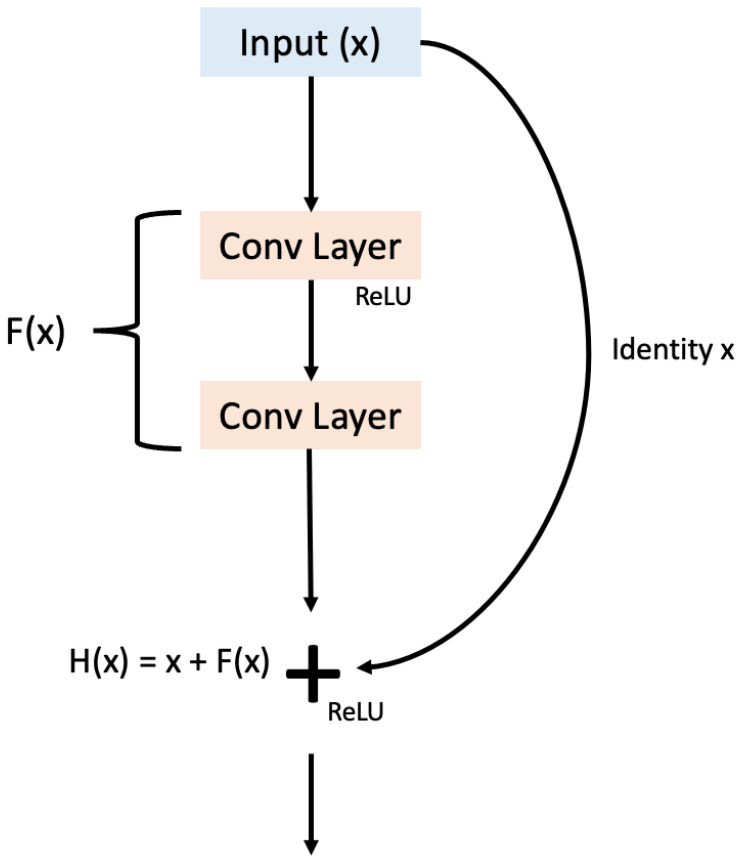


Parameters for n output maps : ...?

Residual Network

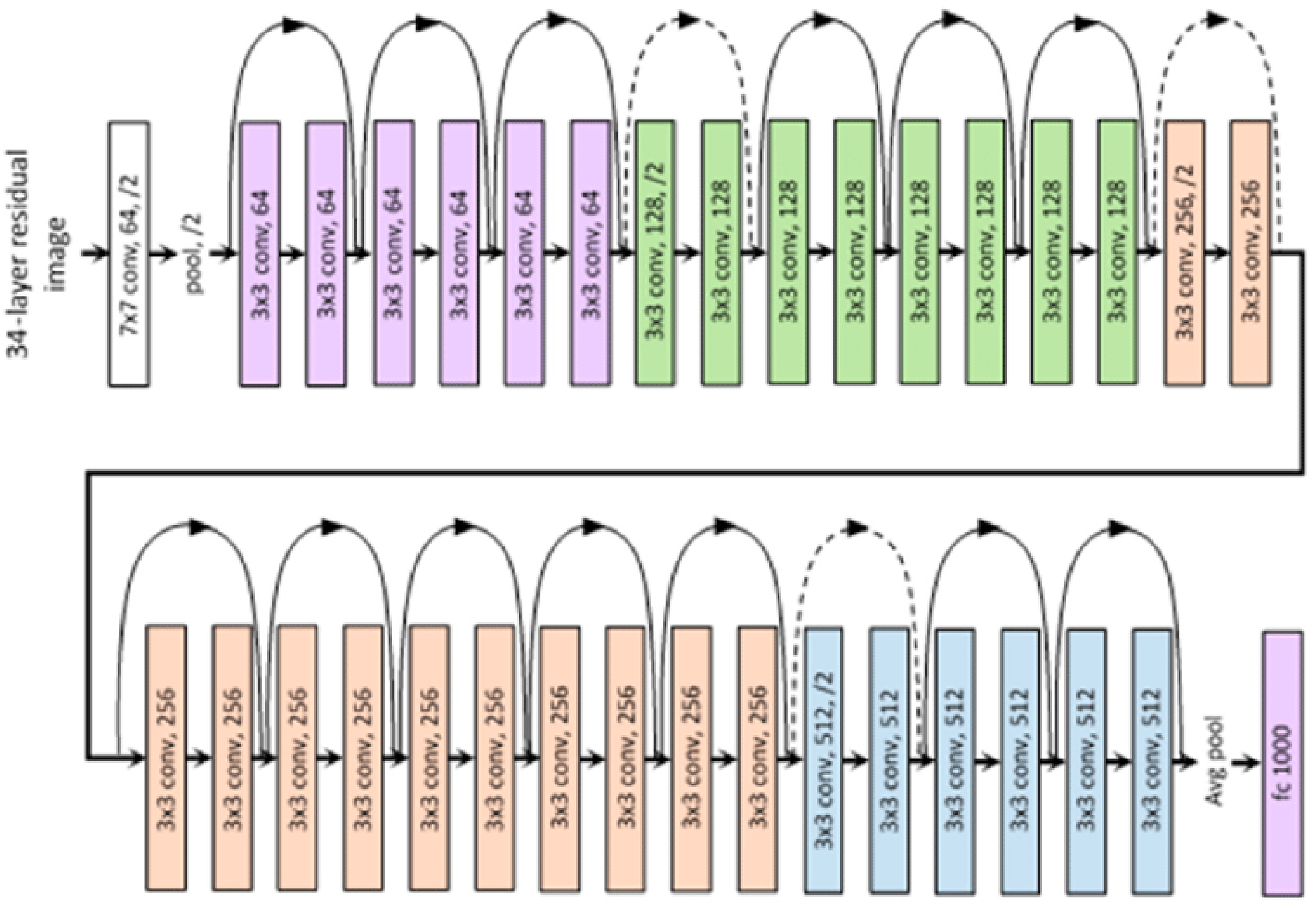
Challenges for Deep Network

- Challenges faced by Deep Neural Networks :
  - Vanishing/Exploding Gradient Problem:



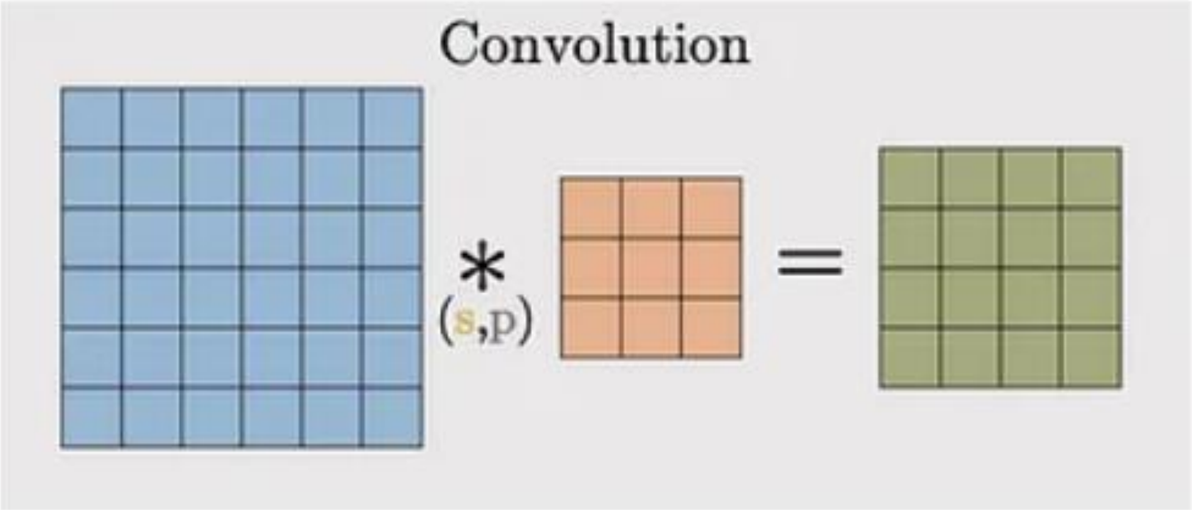
ResNet

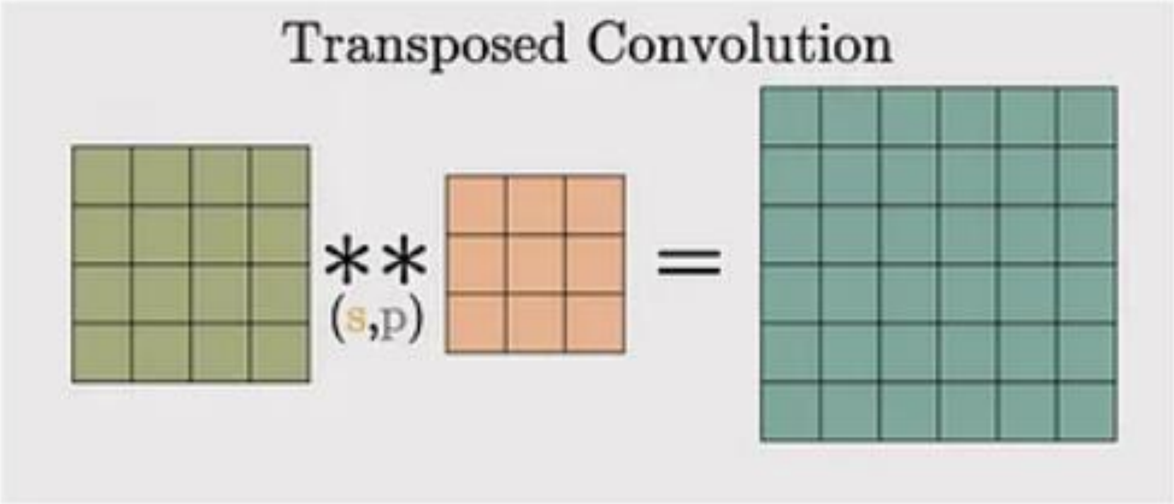
(a)


34-layer residual
image
7x7 conv, 64, /2
pool, /2
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 128
3x3 conv, 256, /2
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
3x3 conv, 512, /2
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
Avg pool
fc 1000

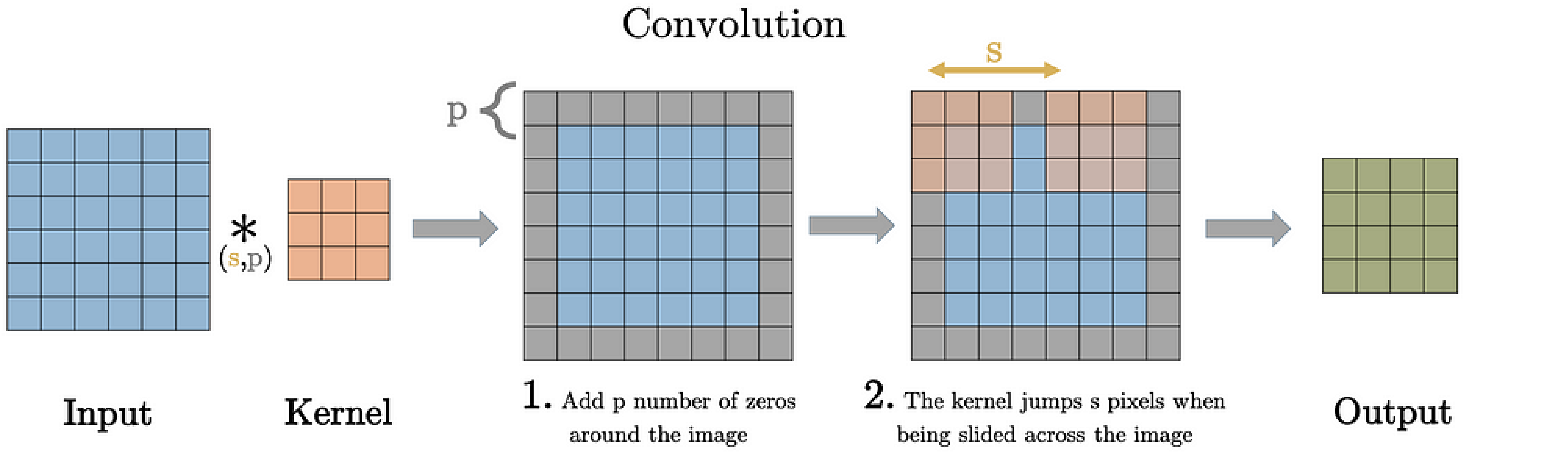
(b)

Transposed

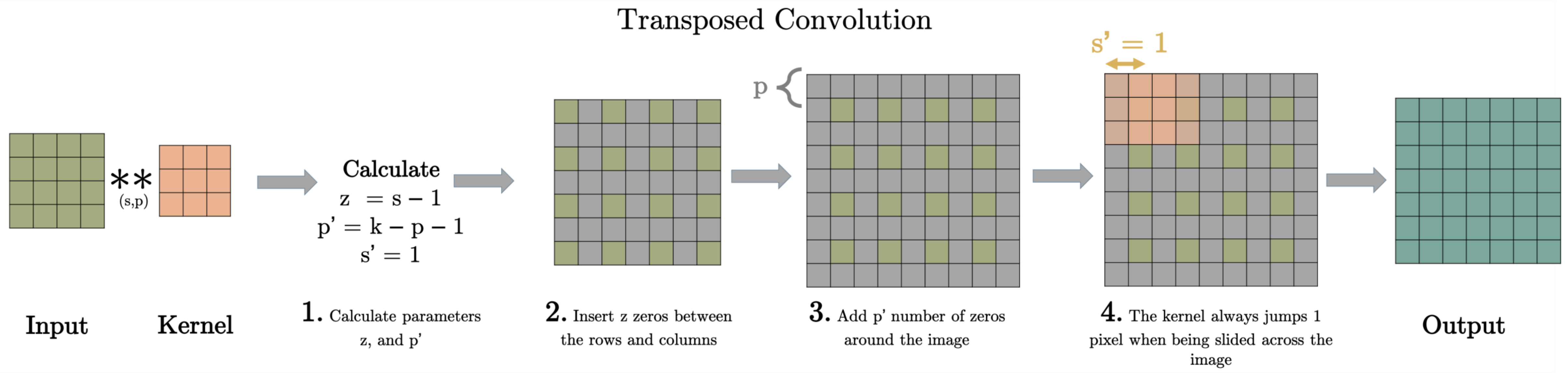

Convolution
(s,p)


Transposed Convolution
(s,p)

Convolution



Transposed Convolution



Autoencoders

Components of Autoencoder

Autoencoders have two main components, an encoder, and a decoder.

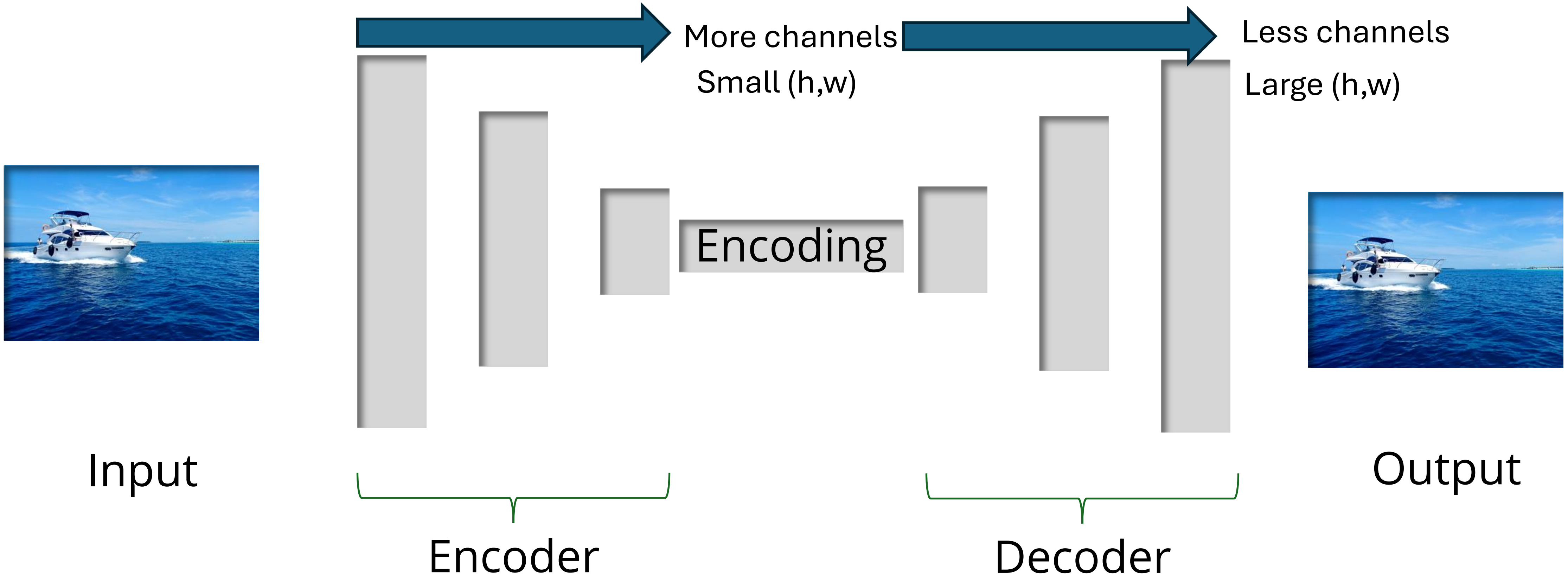


Photo by Alina Kacharho on Unsplash

What is an Autoencoder?

It is an artificial neural network that learns efficient coding of unlabeled data.

It learns and refines encoding by attempting to regenerate the input from the encoding.

It learns a representation (or encoding) for a dataset.

This is dimensionality reduction and can be used to ignore insignificant data such as noise.

You can use it for feature detection, anomaly detection, and denoising images.

It can generate new data that is similar to the input data.

Denoising Autoencoder

In this example, you will regenerate the input image. You can use the same network to learn how to denoise images.

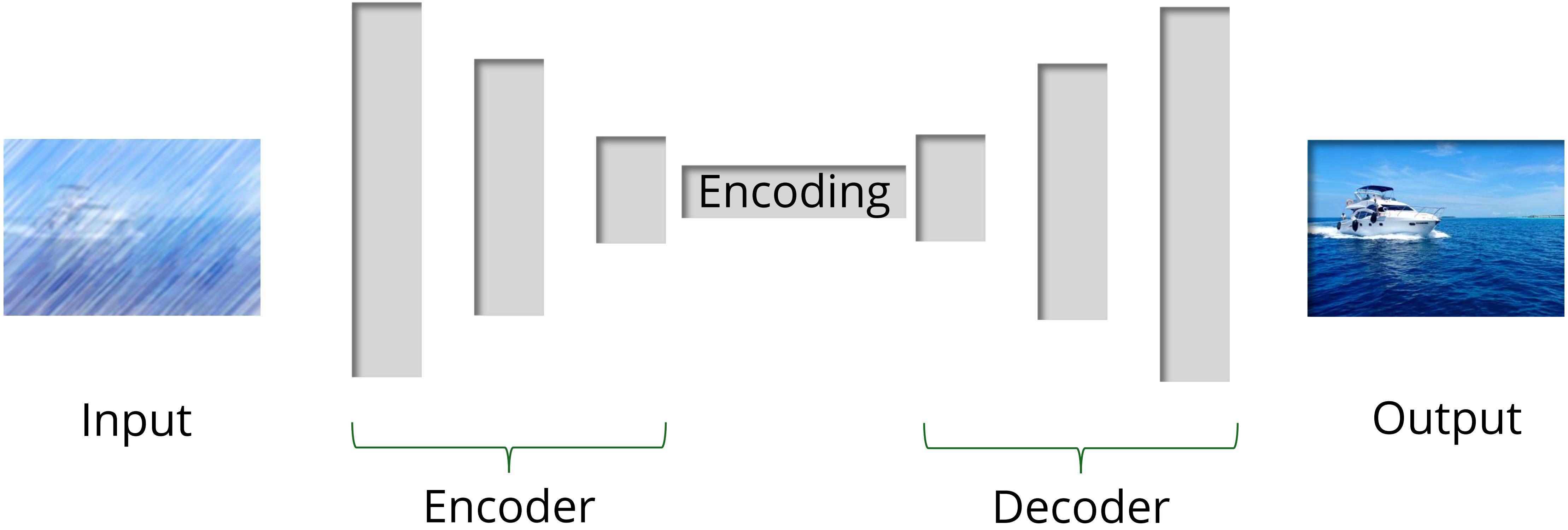


Photo by Alina Kacharho on Unsplash

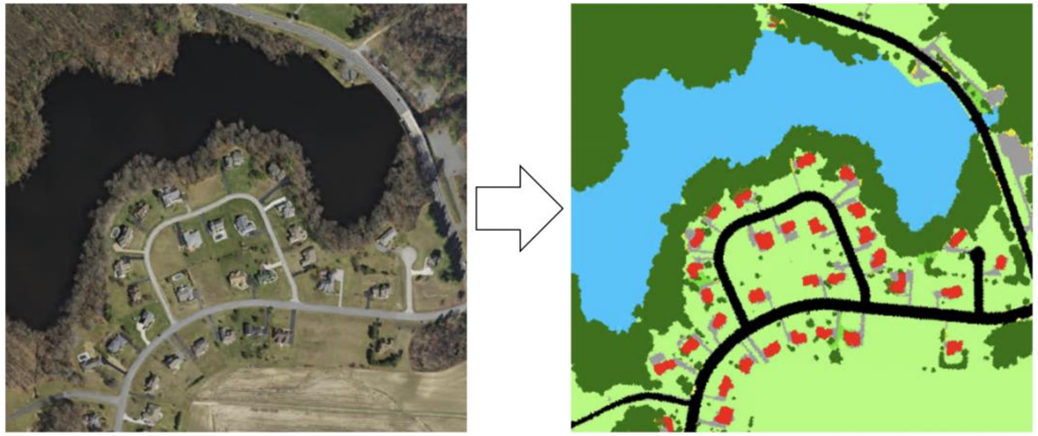
What is U-Net?

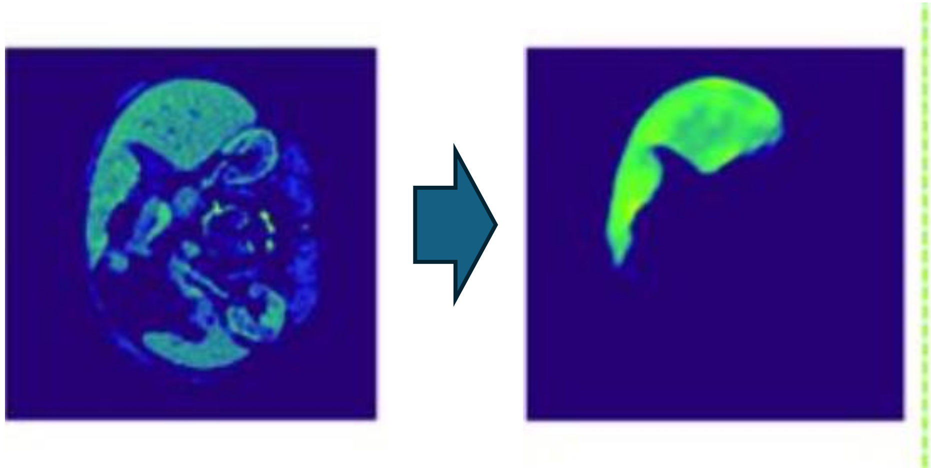
It is a model based on the autoencoder approach.

It was developed to help with biomedical image segmentation.

It is fully convolutional, so there are no dense (fully connected) layers.

It was designed to train on limited samples of data using augmentation.

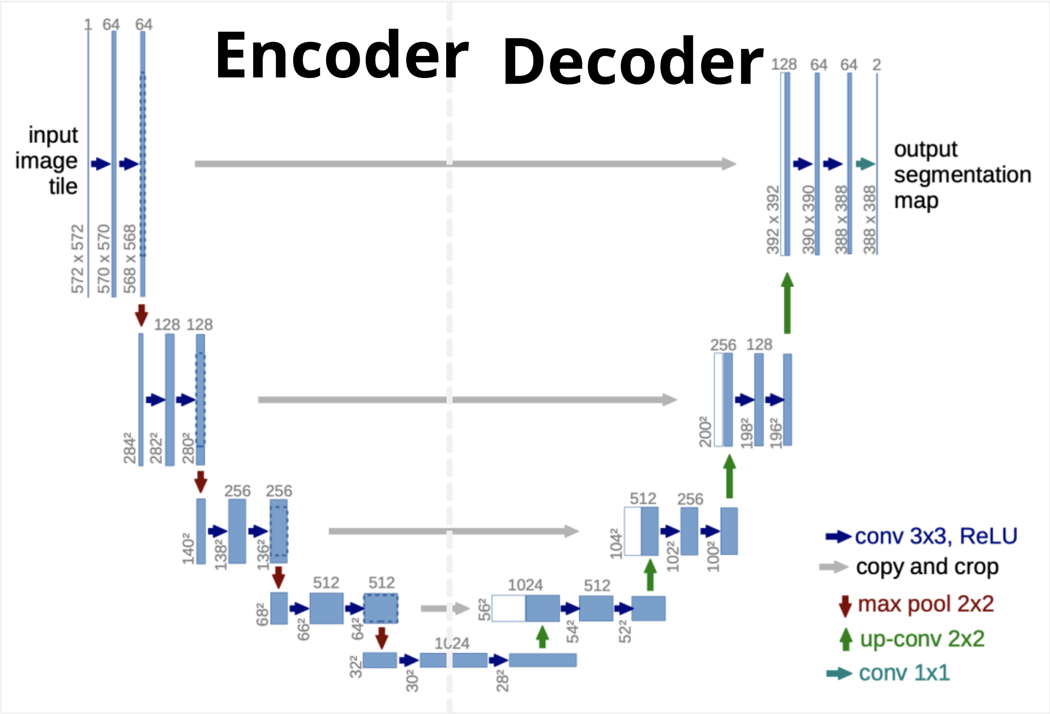




What does U-Net Give?

In a classification task, the main focus is on ”what”, the class label. But the output semantic segmentation is more than just the class label or bounding box parameters. The output is a complete high-resolution image in which all the pixels are classified.

The decoder part of the network recovers the ”where”, and the skip connections from the encoder provide more precise locations.

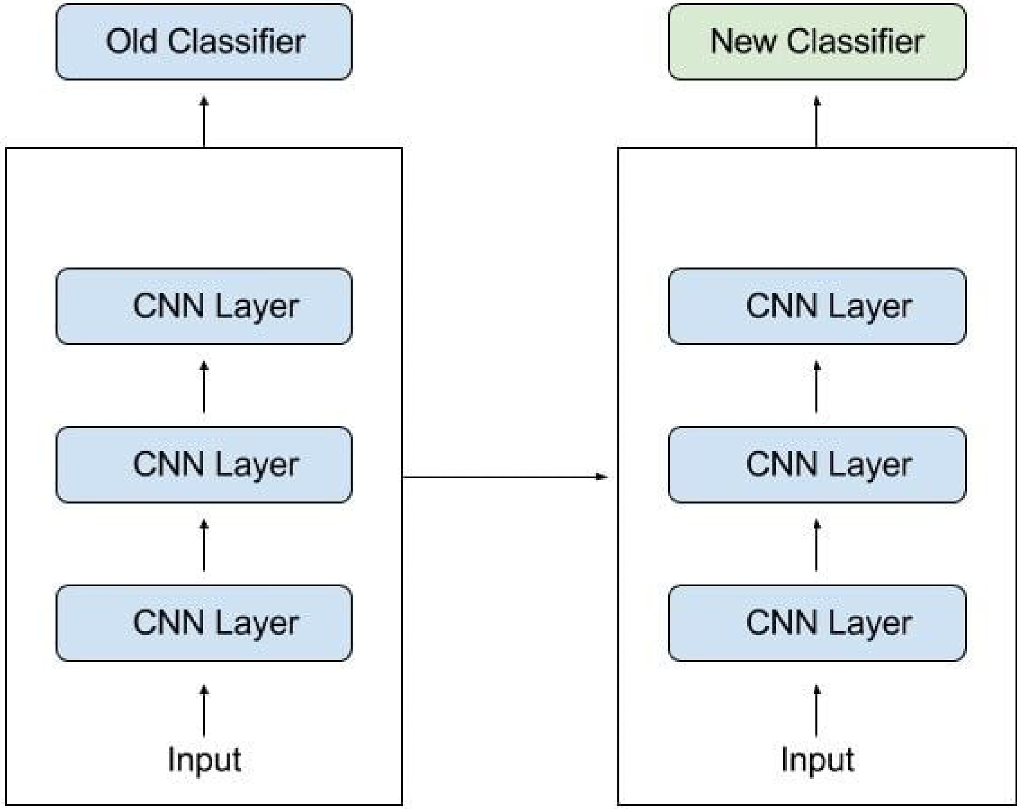


Source and copyright: https://arxiv.org/pdf/1505.04597.pdf

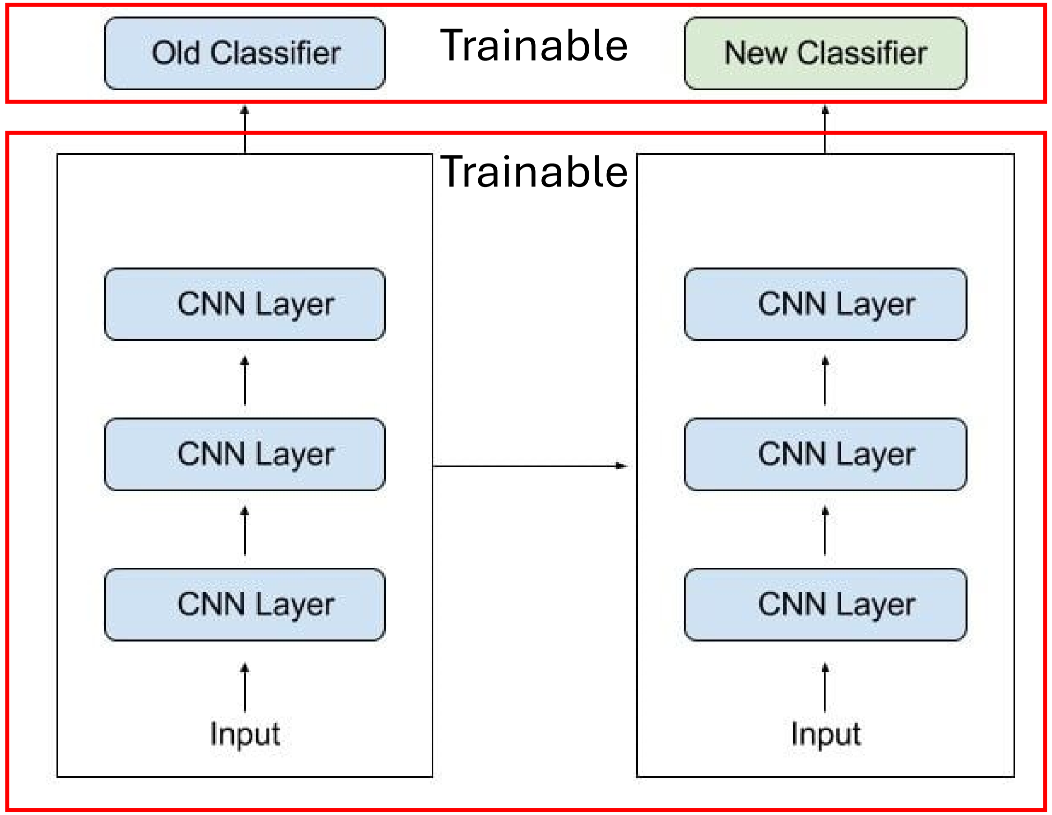
Transfer Learning

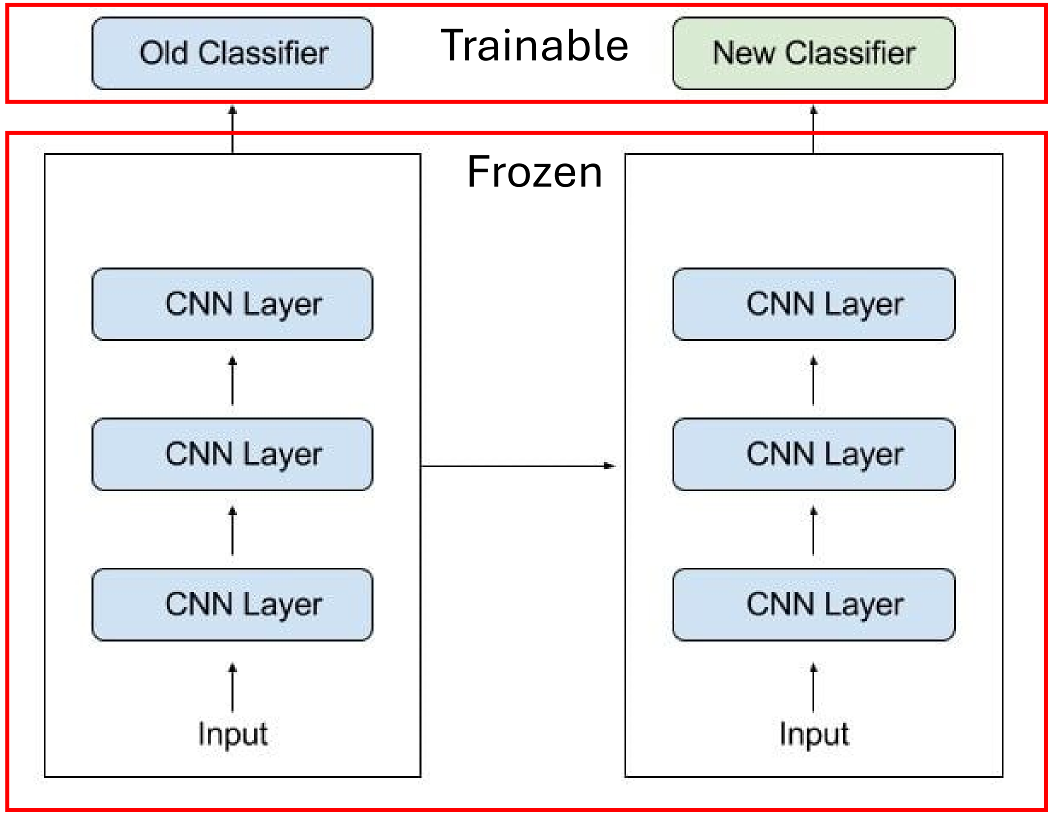
Transfer Learning

- Knowledge of an already trained machine learning model is applied to a different but related problem.
- For example, if you trained a simple classifier to predict whether an image contains a backpack, you could use the knowledge that the model gained during its training to recognize other objects like sunglasses.



Transfer Learning





Thank you