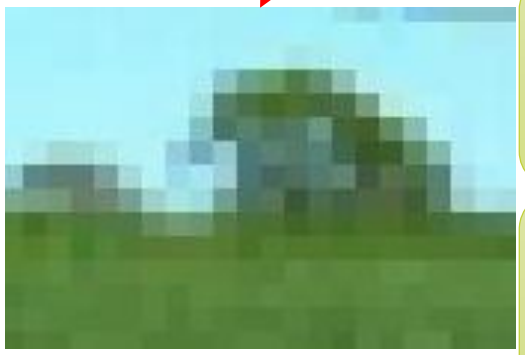
Colorization with Convolutional Neural Networks

Division of Computer Science University of Minnesota,
Morris Morris, Minnesota, USA 56267

17 Nov 2018

Yutaro Miyata

Introduction



Pixel: Smallest square that can be displayed on a screen.

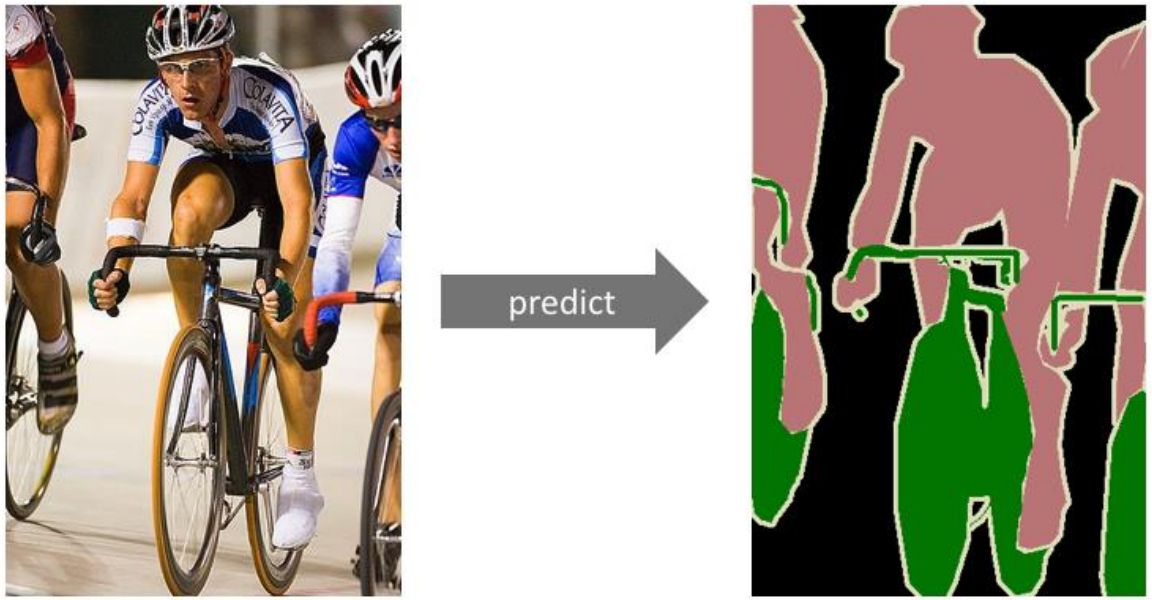
RGB: Color used in computer display. Combination of these can generate other colors.

Take Sample Color to find best matching

https://ieeexplore.ieee.org/document/8014766/references#references

Introduction

Semantic Segmentation



Person
Bicycle
Background

Classification

https://www.jeremyjordan.me/semantic-segmentation/

Outline

- Convolutional Neural Networks
- U-Net
- Generative Adversarial Network
- Image to Image Translation
- Global and Local Image Priors for Automatic Image Colorization
- Result

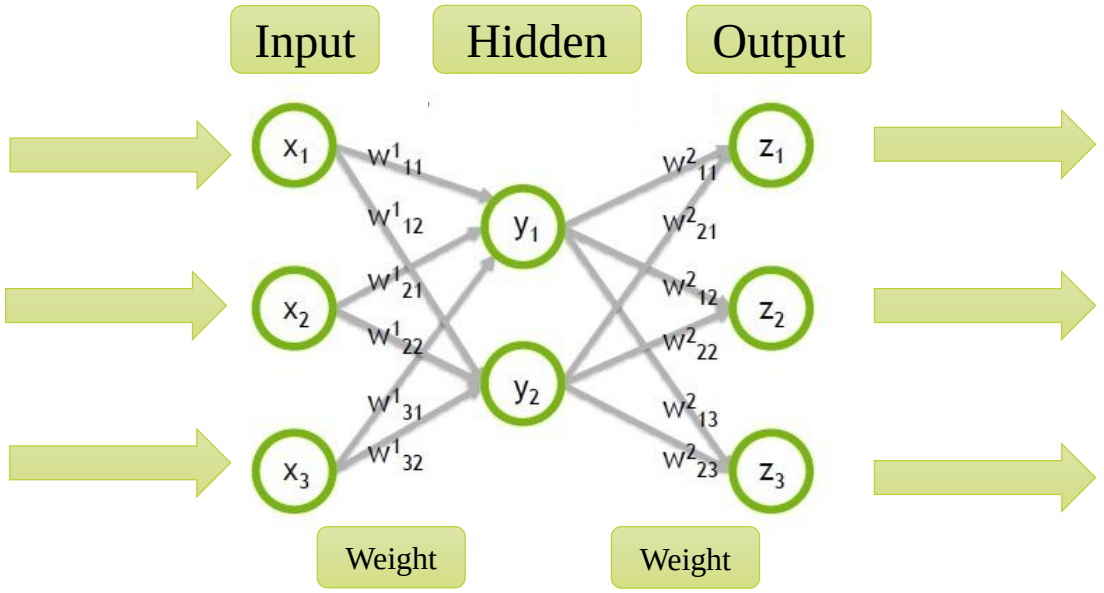
https://fstoppers.com/video/how-amazing-colorization-black-and-white-photos-are-done-5384

Artificial Neural Network

- What is an Artificial Neural Network?

BW

Multilayer Perceptron

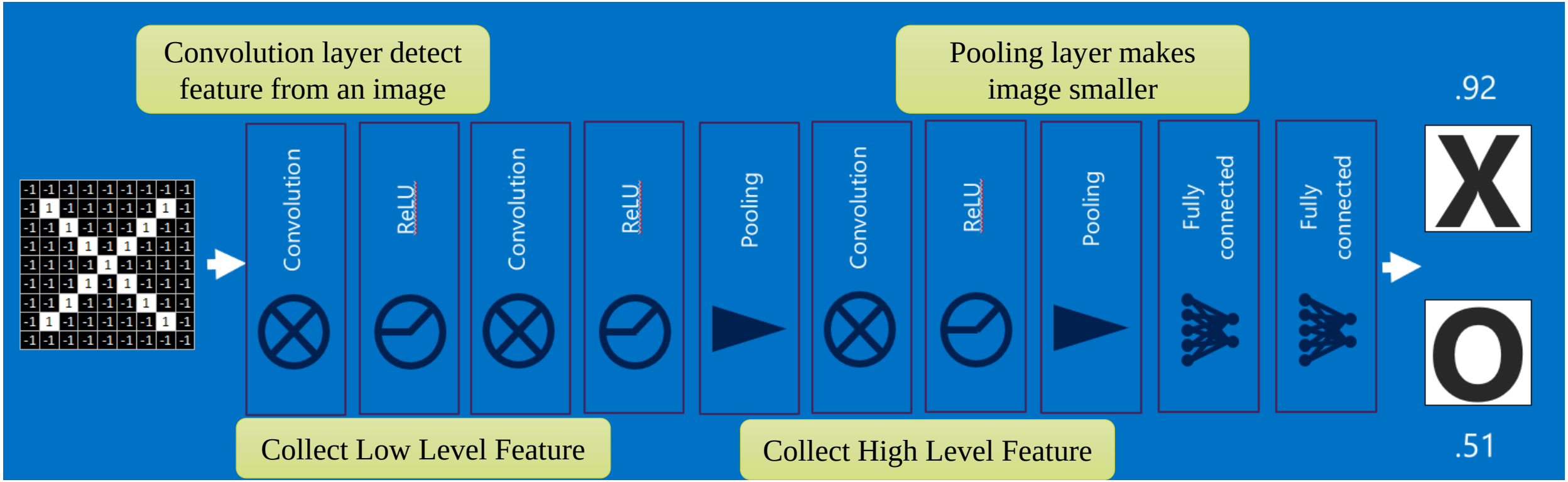


Color

https://www.sbbit.jp/article/cont1/33345

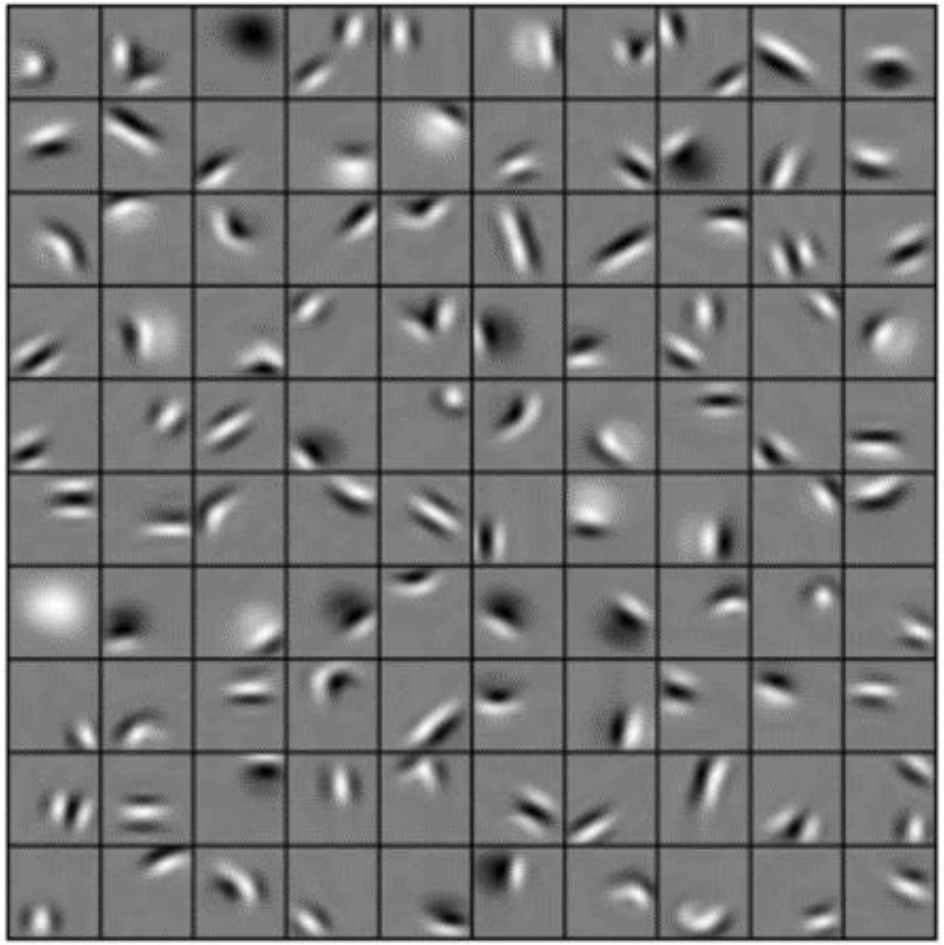
Convolutional Neural Networks

- Structure Of CNNs

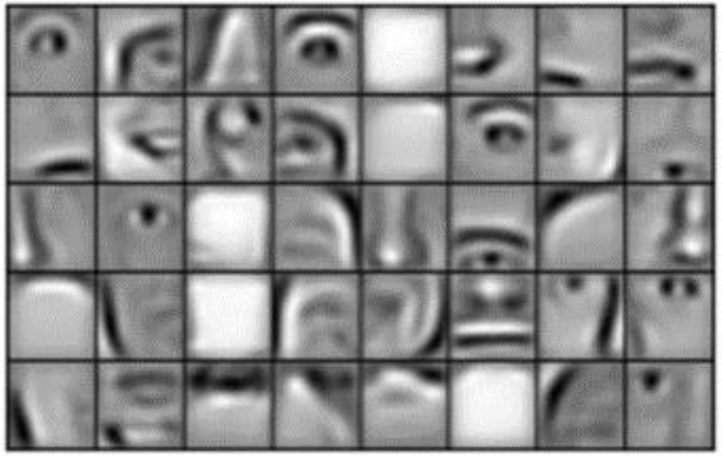


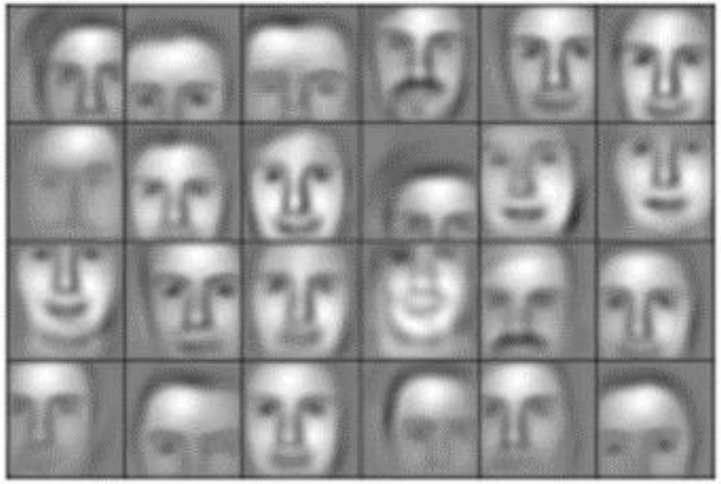
http://brohrer.github.io/how_convolutional_neural_networks_work.html

Low level feature



High level feature





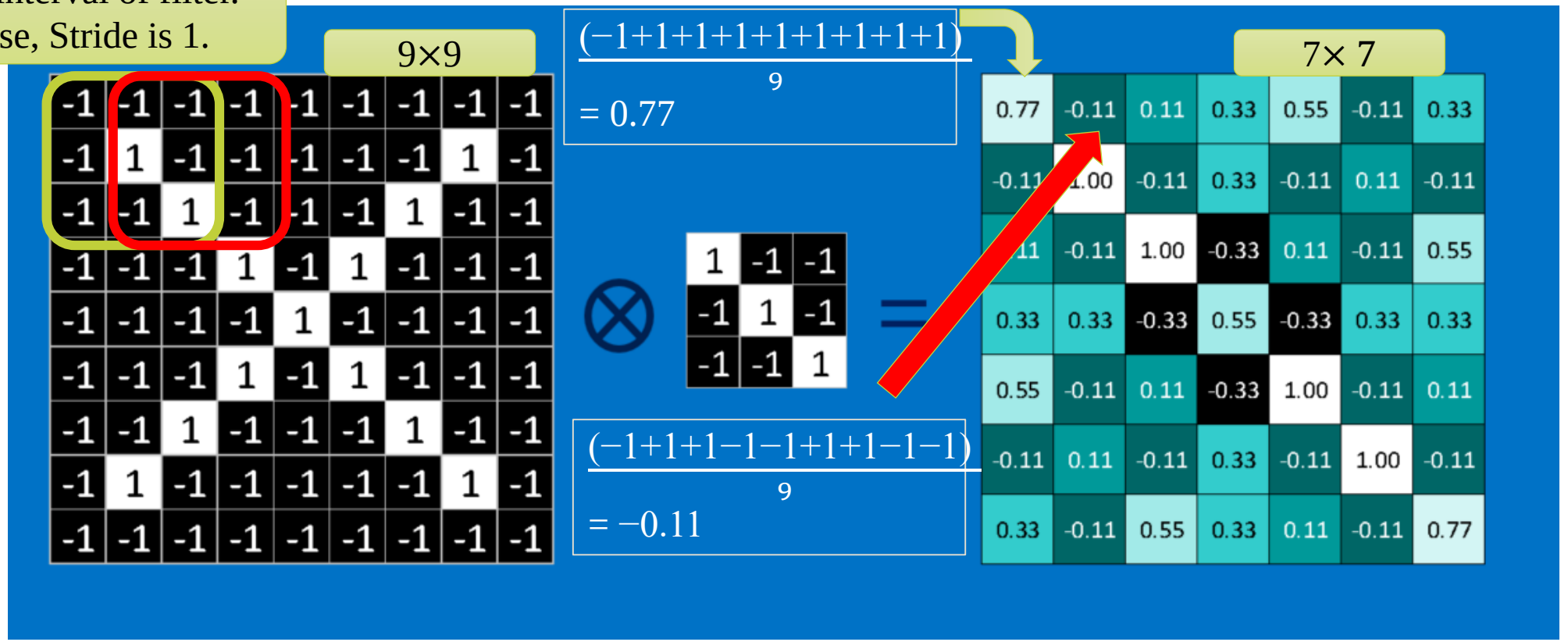
Feature: Pattern which network find from many data set.

Feature map: Certain combination of feature is found in an image.

http://brohrer.github.io/how_convolutional_neural_networks_work.html

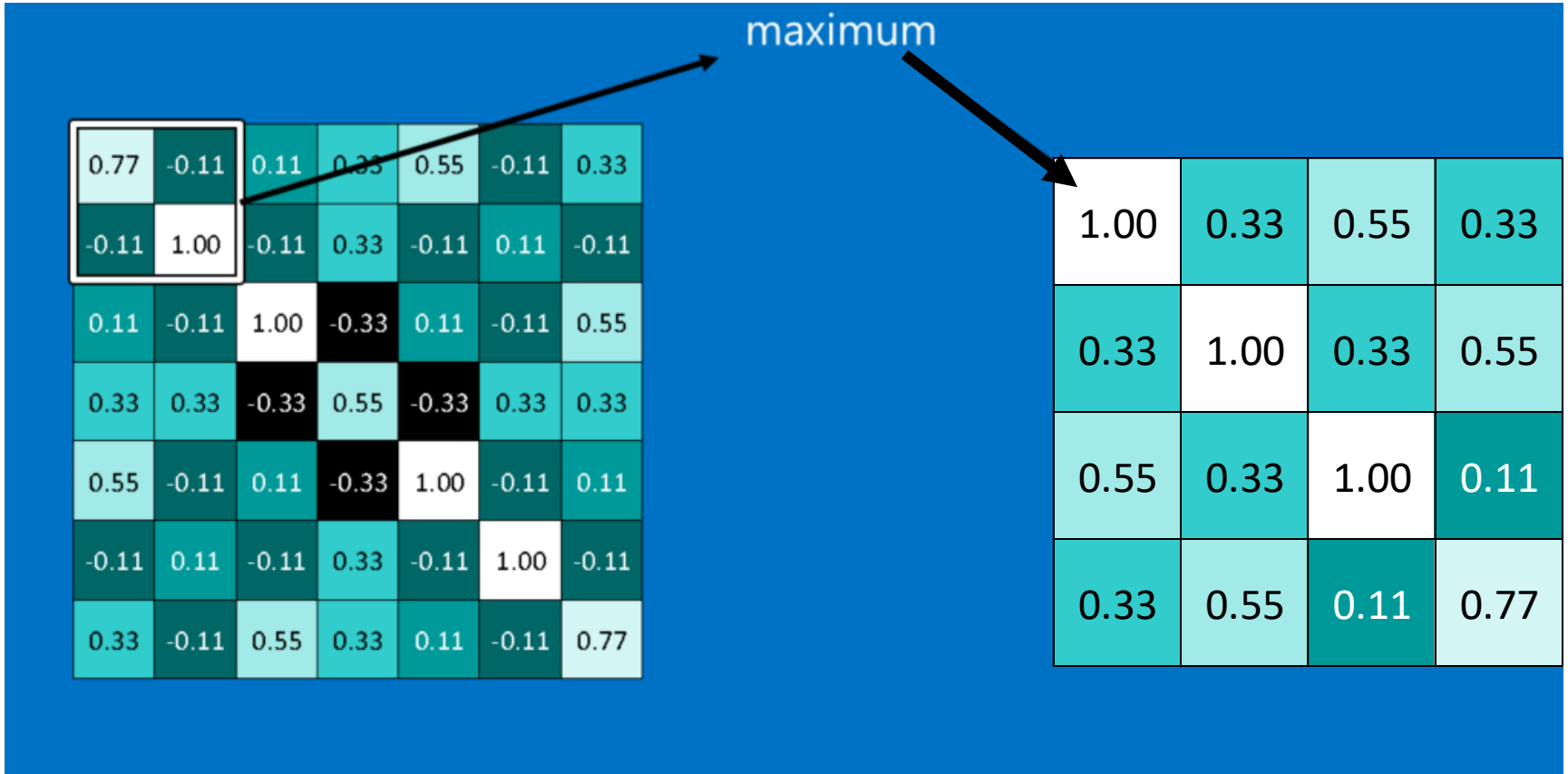
- Convolutional Layer
  - This layer detects many features from image. It applies many filters to image to get feature maps.

Stride is interval of filter.
In this case, Stride is 1.



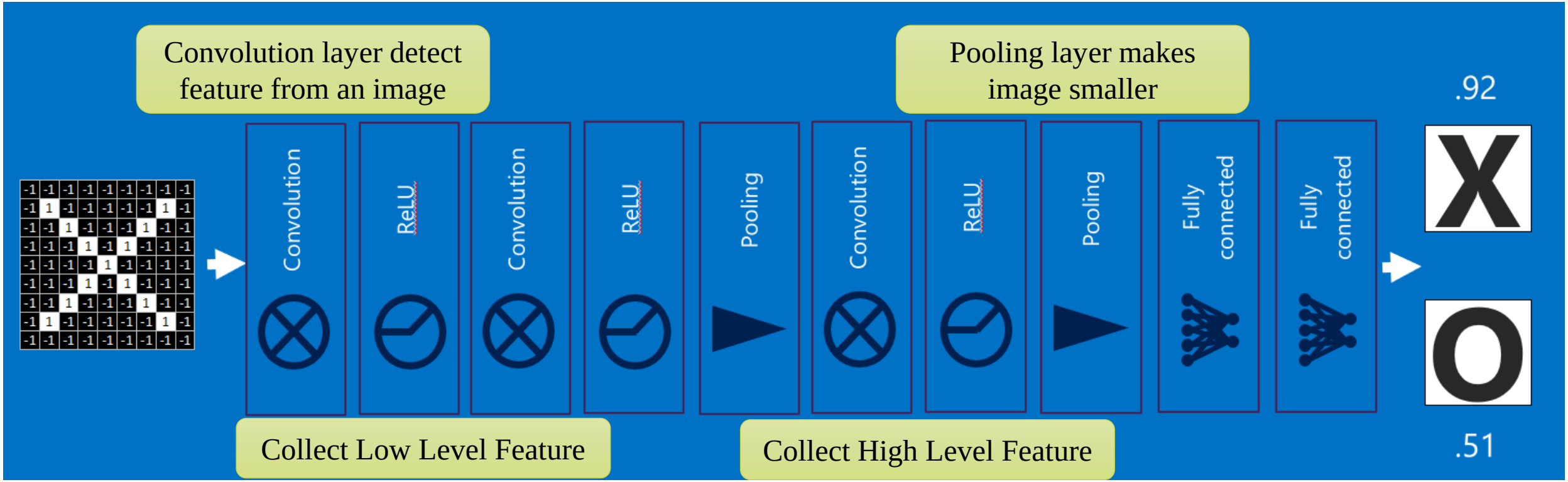

9×9×1
Filters
7× 7 × 3
⊗
=
⊗
=
⊗
=

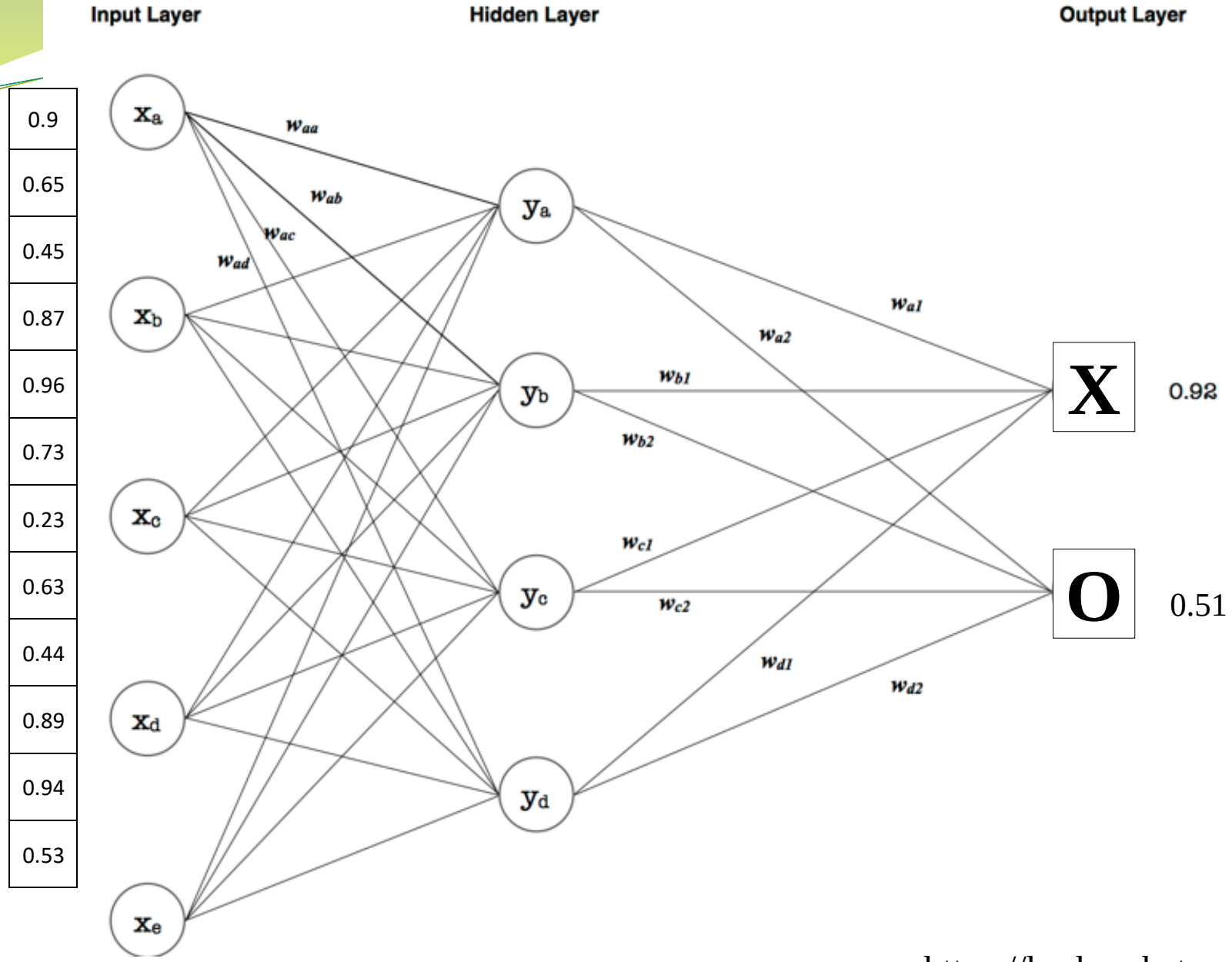
- Pooling Layer
  - This layer transform a manageable form to clarify features. After pooling process, an image size will be ¼, and a computational cost is reduced.



Convolutional Neural Networks

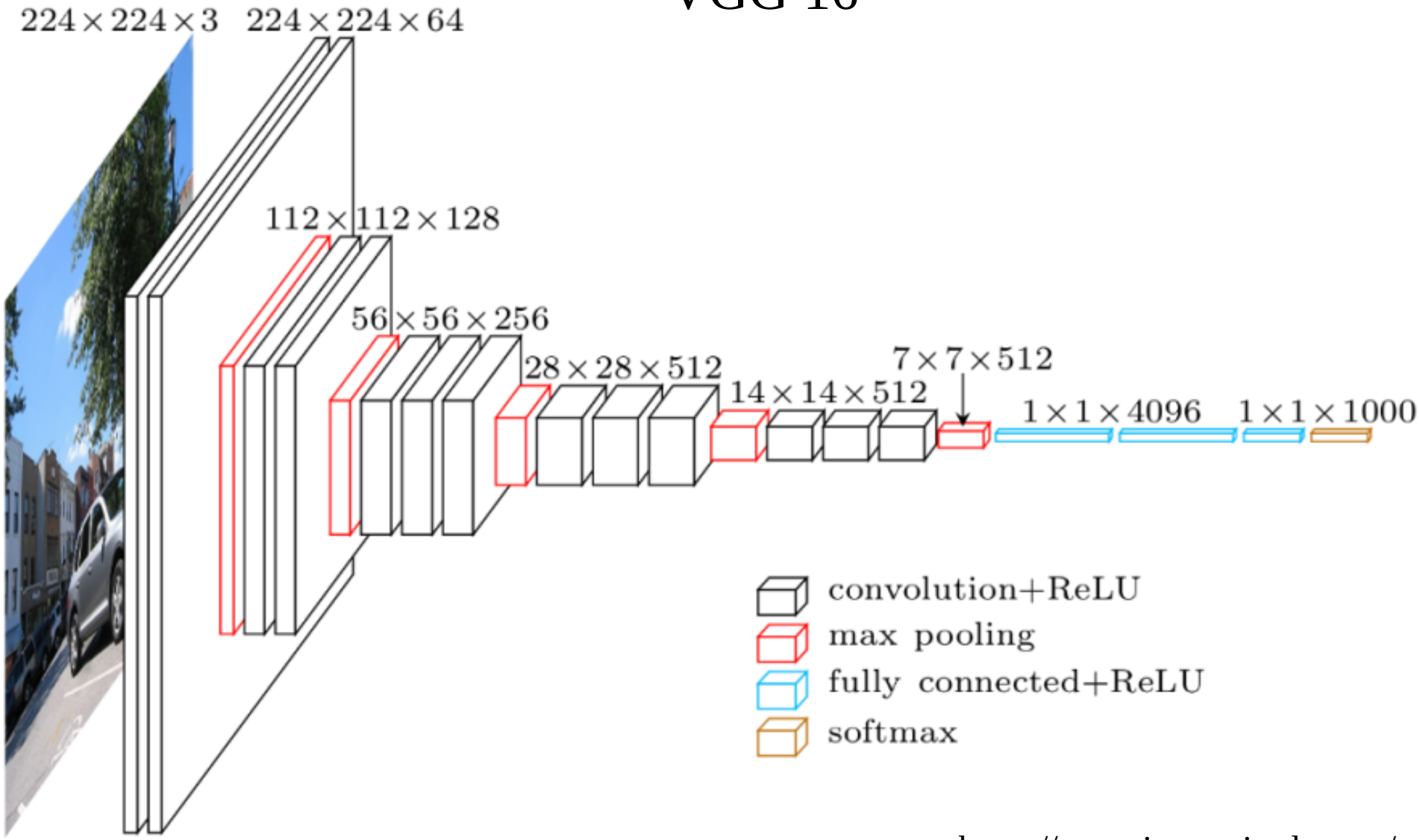
- Structure Of CNNs
  - CNNs consists of many convolution layer and pooling layer.





https://hashrocket.com/blog/posts/a-friendly-introduction-to-convolutional-neural-networks

- VGG 16



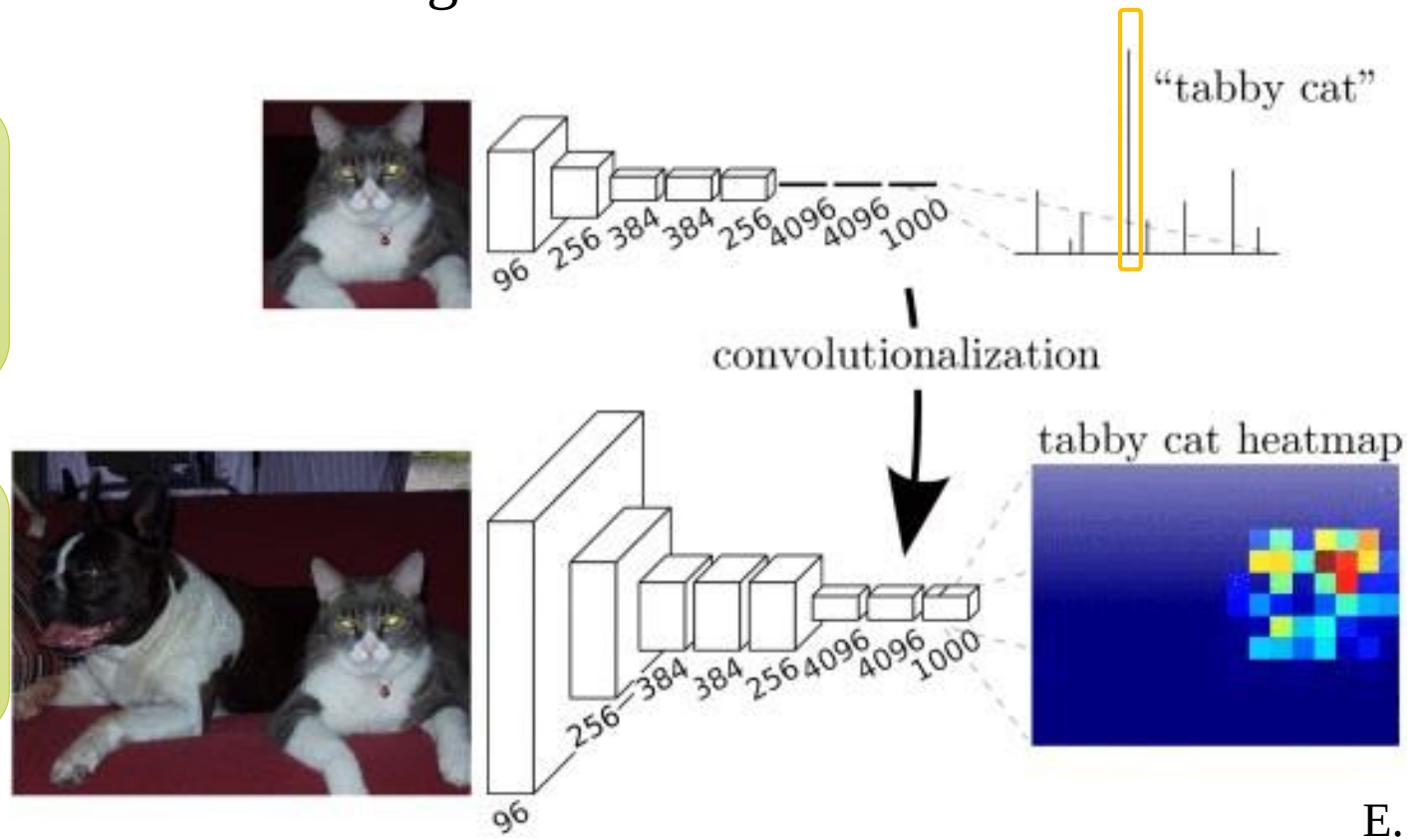
https://www.jeremyjordan.me/convnet-architectures/

U-Net

- What is U-Net ?
  - U-Net is one method of Semantic Segmentation.

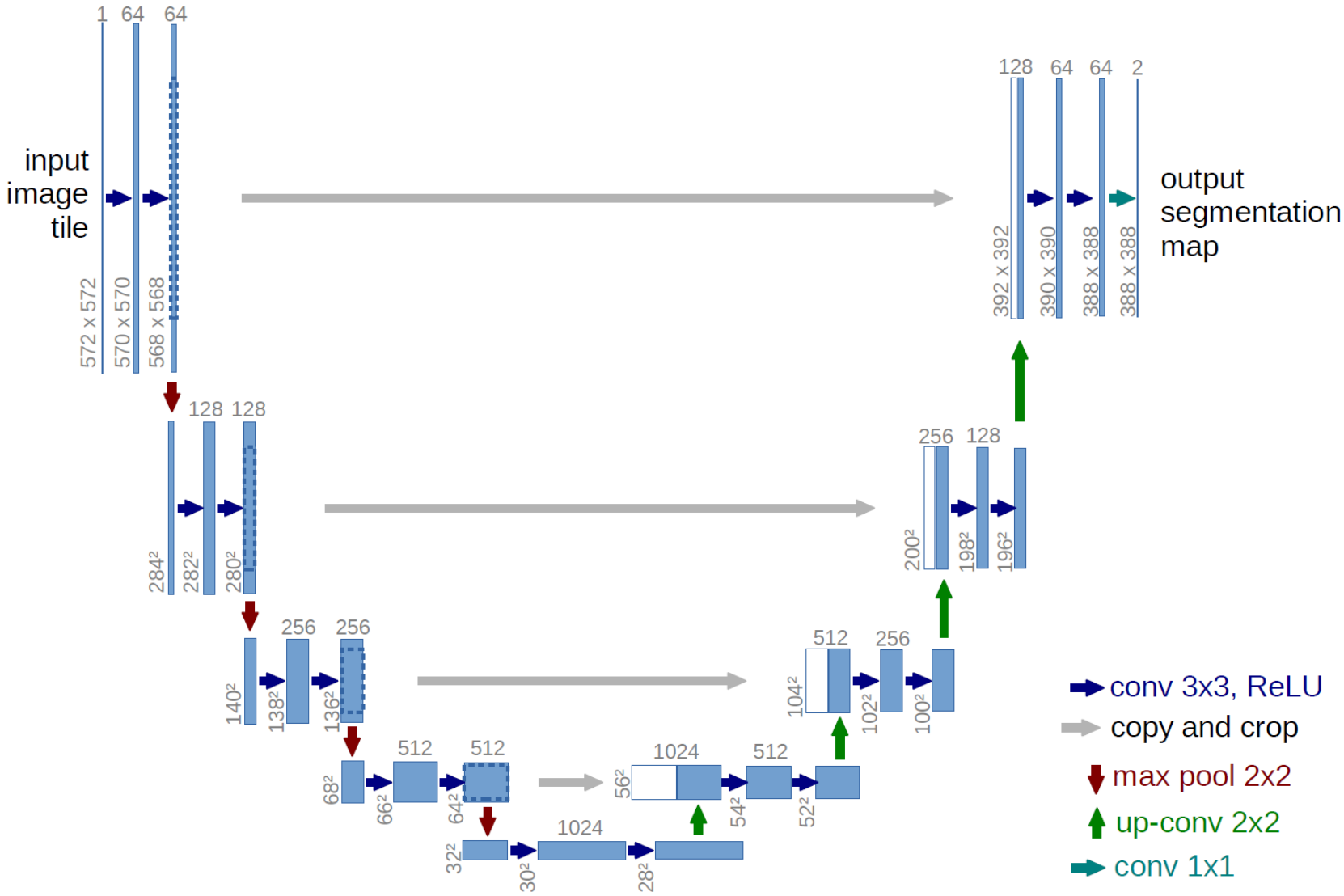
CNNs:
One dimensional vector

U-Net:
Two dimensional map
(H × W)

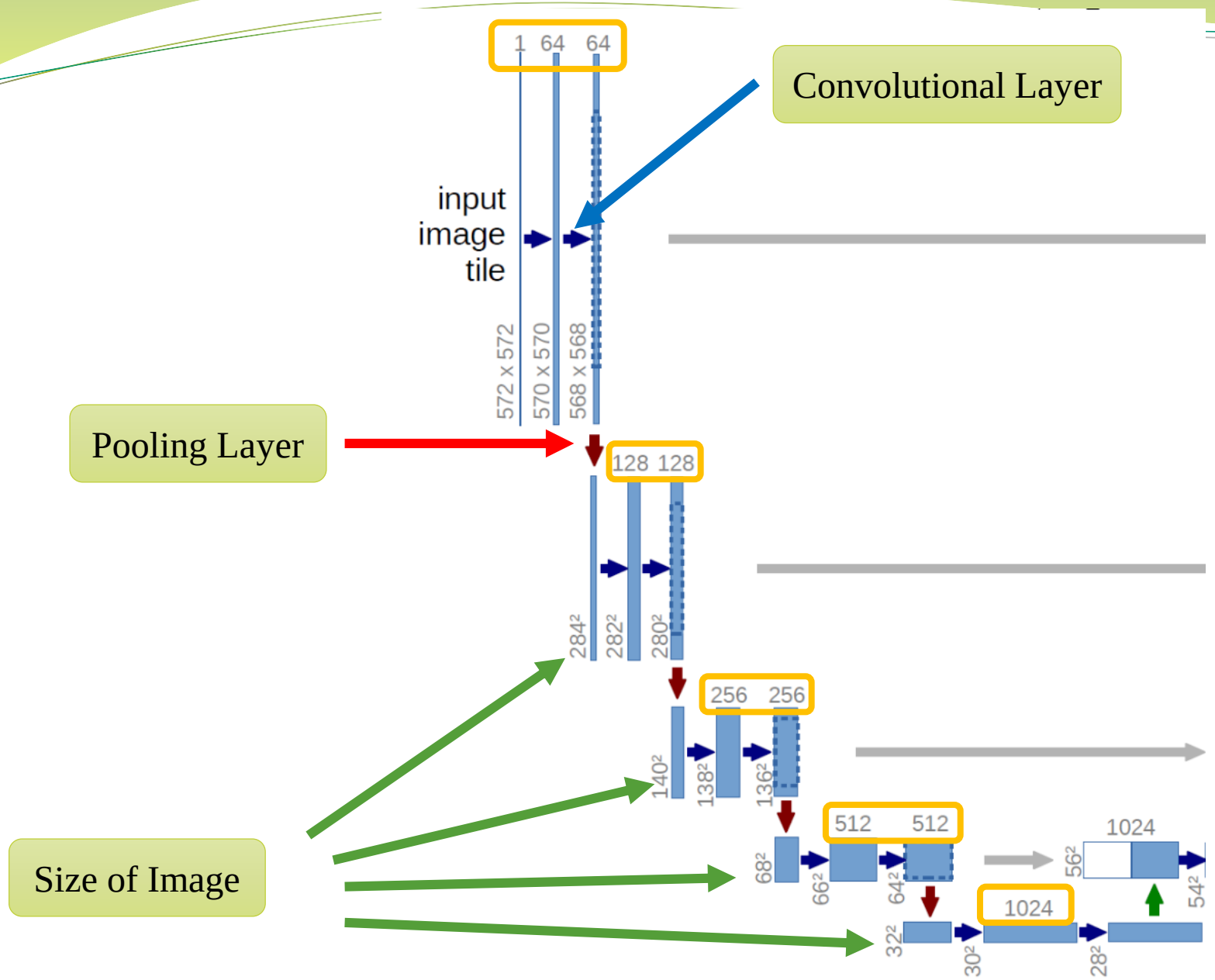


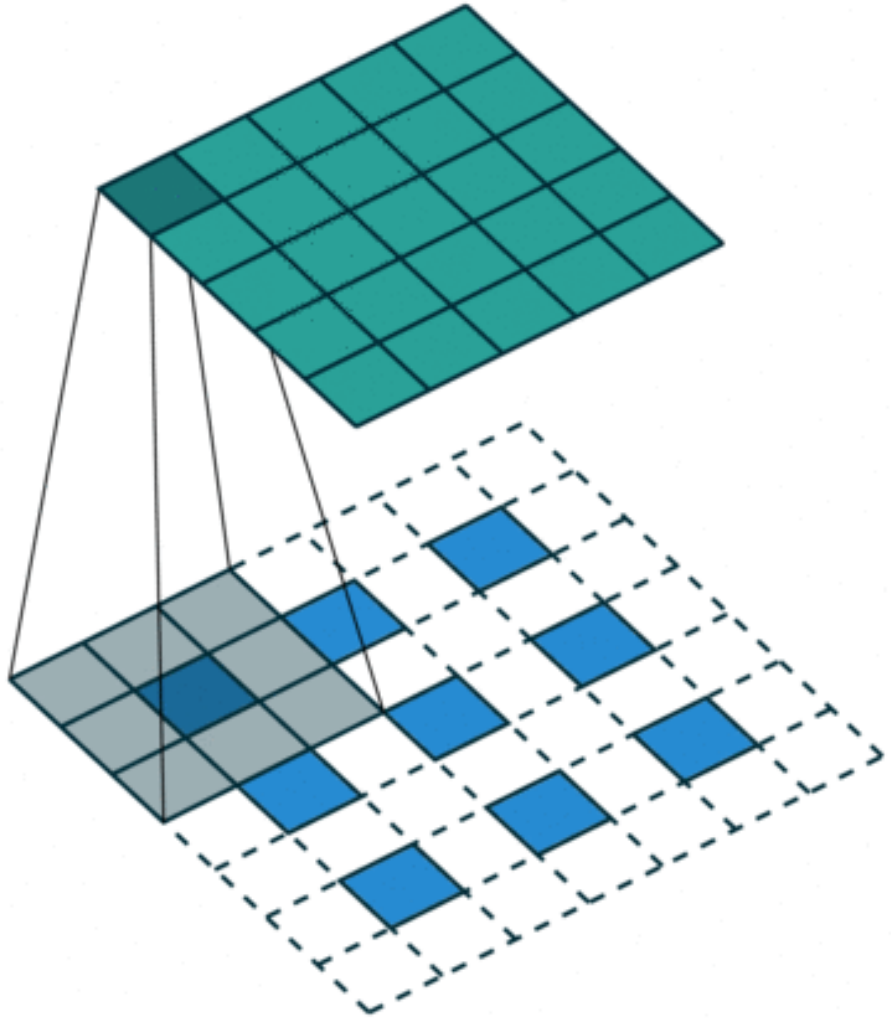
E. Shelhamer 2016

- Structure Of U-Net

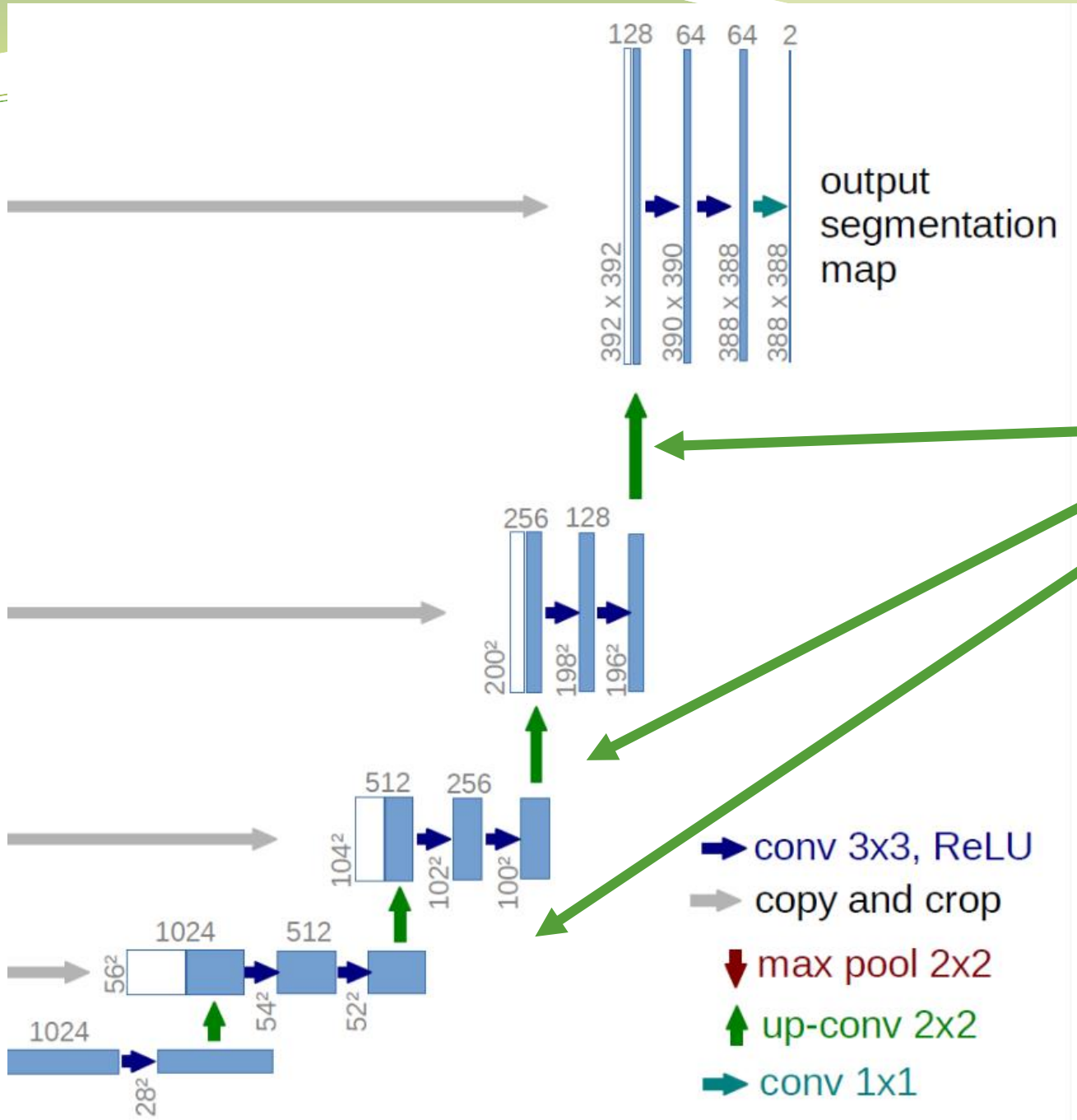


O. Ronneberger 2015

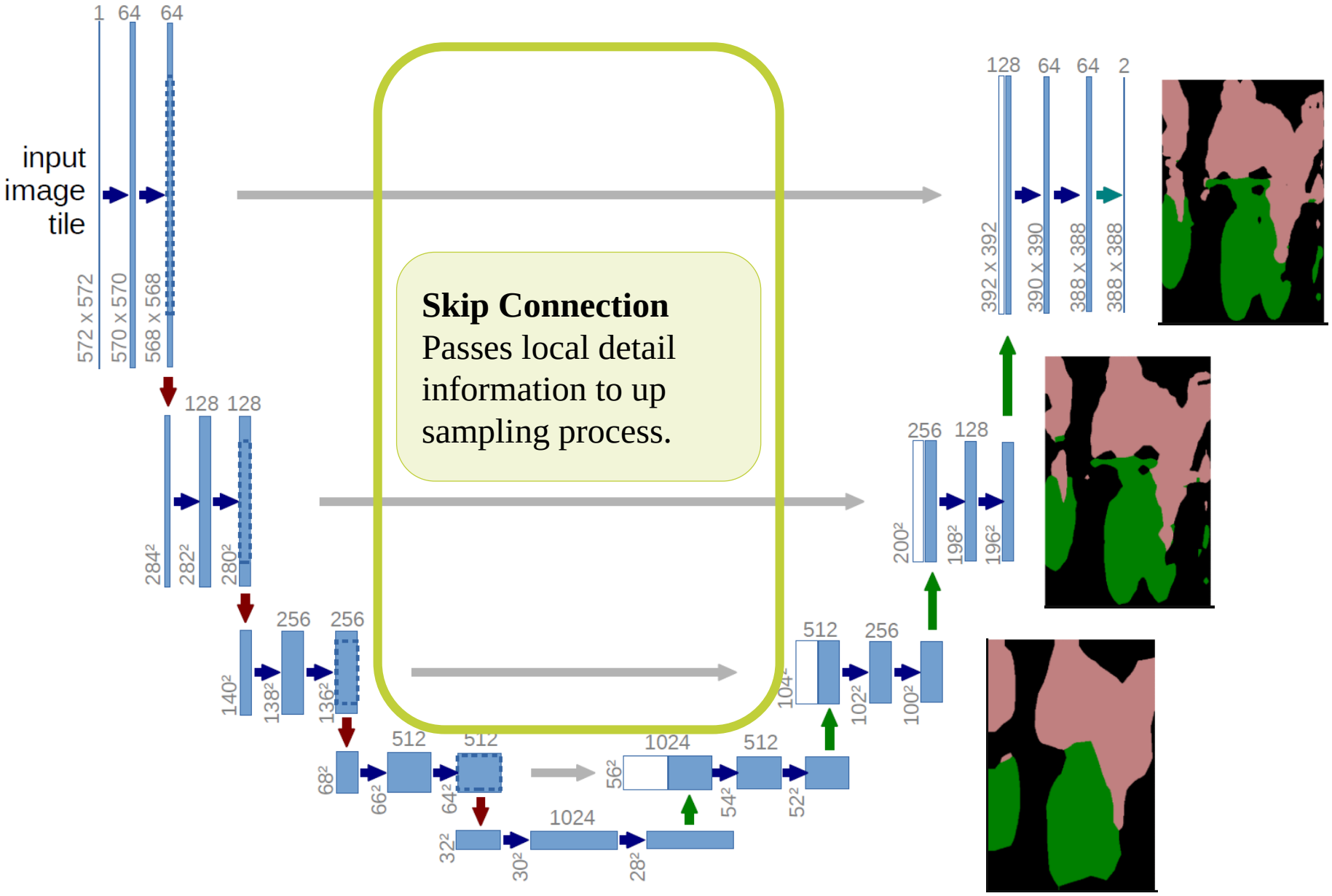

Convolutional Layer
Pooling Layer
Size of Image
input image tile
1 64 64
572 x 572
570 x 570
568 x 568
128 128
284²
282²
280²
256 256
140²
138²
136²
512 512
68²
66²
64²
1024
56²
54²
1024
32²
30²
28²



https://towardsdatascience.com/types-of-convolutions-in-deep-learning-717013397f4d



Up Sampling:
Expands feature map because it became coarse due to Pooling Layer


input image tile
1 64 64
572 x 572
570 x 570
568 x 568
128 128
284²
282²
280²
256 256
140²
138²
136²
512 512
68²
66²
64²
1024
32²
30²
28²
512
56²
54²
52²
512 256
104²
102²
100²
256 128
200²
198²
196²
128 64 64 2
392 x 392
390 x 390
388 x 388
388 x 388
Skip Connection
Passes local detail information to up sampling process.

Generative Adversarial Network

- Structure of GAN
  - The Generative Adversarial Network is generator (G) and discriminator (D) model.

D: Discriminator

Determine whether it is generated or real

Generator continues to improve the accuracy of generated image until it cannot be determined.

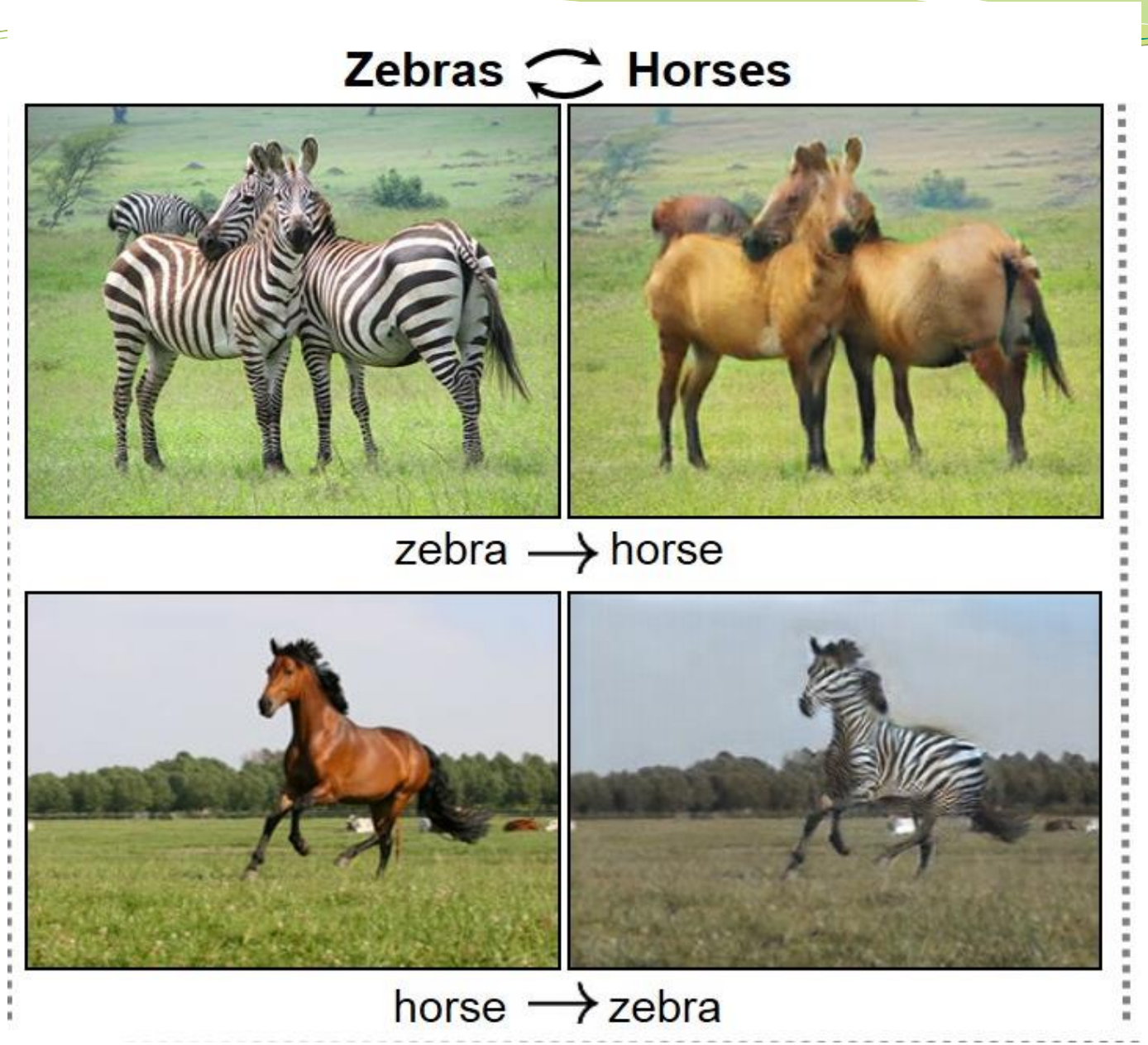
X: Training real Data

G: Generator

Z: Random Noise
Or image

Trying to generate imitated image which nearly a real image

https://medium.com/@devnag/generative-adversarial-networks-gans-in-50-lines-of-code-pytorch-e81b79659e3f



https://junyanz.github.io/CycleGAN/

Take real image as input

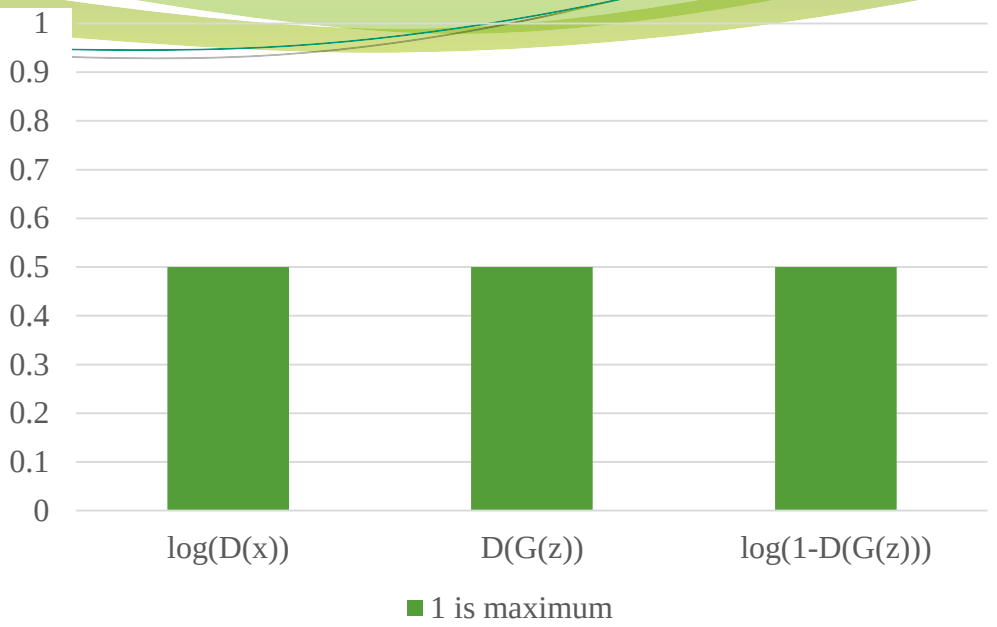
Trying to be bigger(1)

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)]$$
$$+ E_{z\sim P_z(z)}[\log(1 - D(G(z)))].$$

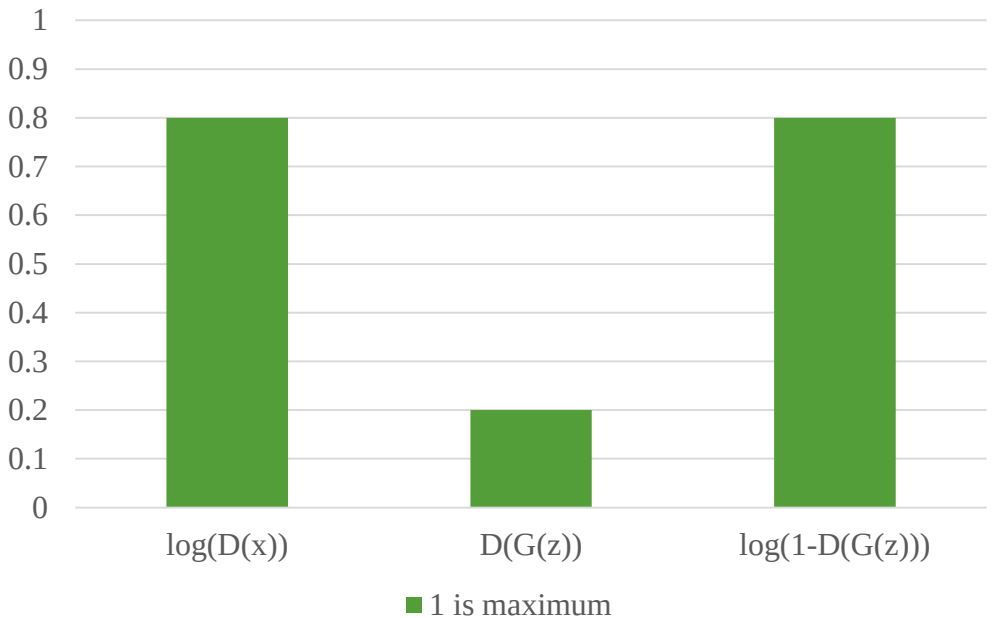
Take noise as input

Trying to be smaller (0)

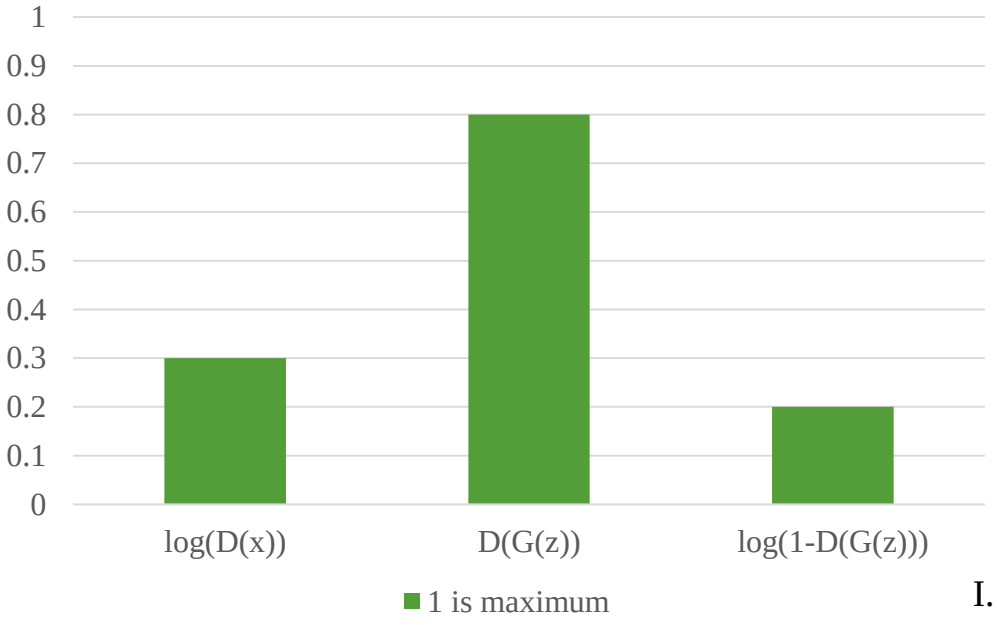
Desired result



D is well trained



D is not well trained



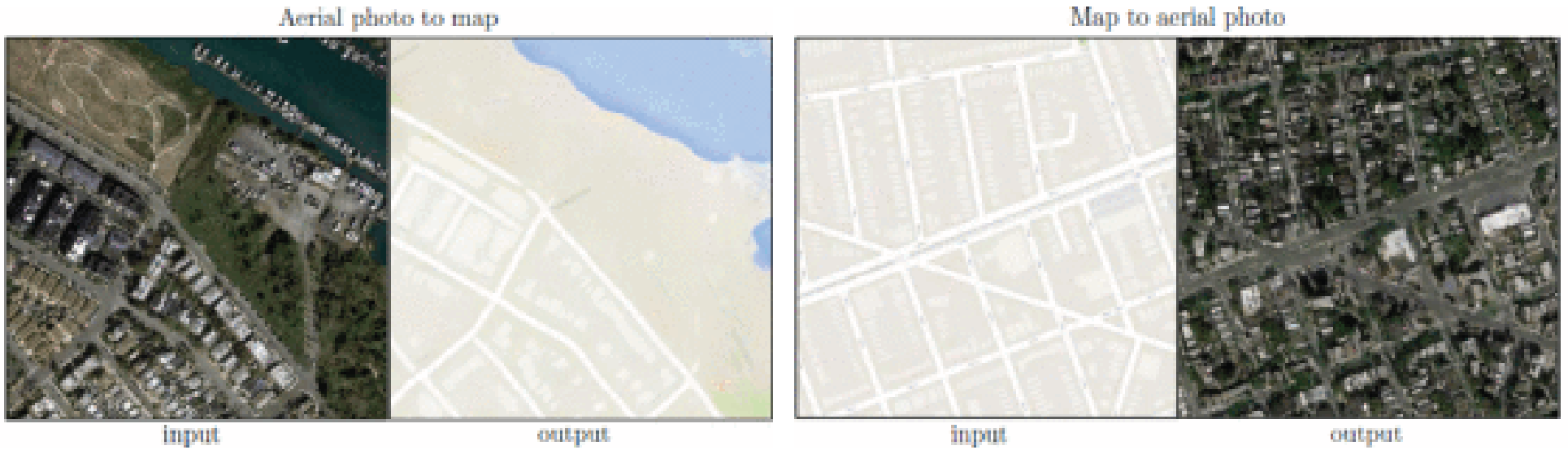
I. J. Goodfellow 2014

Image to Image Translation

- Prove effectiveness of cGAN and U-Net
- The authors adopt one methods.
  - Amazon Mechanical Turk: Show a generated image and real image to human and evaluate it. Measure the percentage which tester says real.

	Turker	**Turker**
	Photo to Map	Map to Photo
L1	2.8%±1.0%	0.8%±0.3%
L1 + cGAN	6.1%±1.3%	18.9%±2.5%

L1: This is traditional method




Ground truth
L1
cGAN
L1 + cGAN

Effectiveness on Colorization

Method	Turkers Labeled real
Zhang et al 2016	27.8%±2.7%
L1 + cGAN	22.5%±1.6%

Another method of colorization which specifically engineered to do well on colorization.

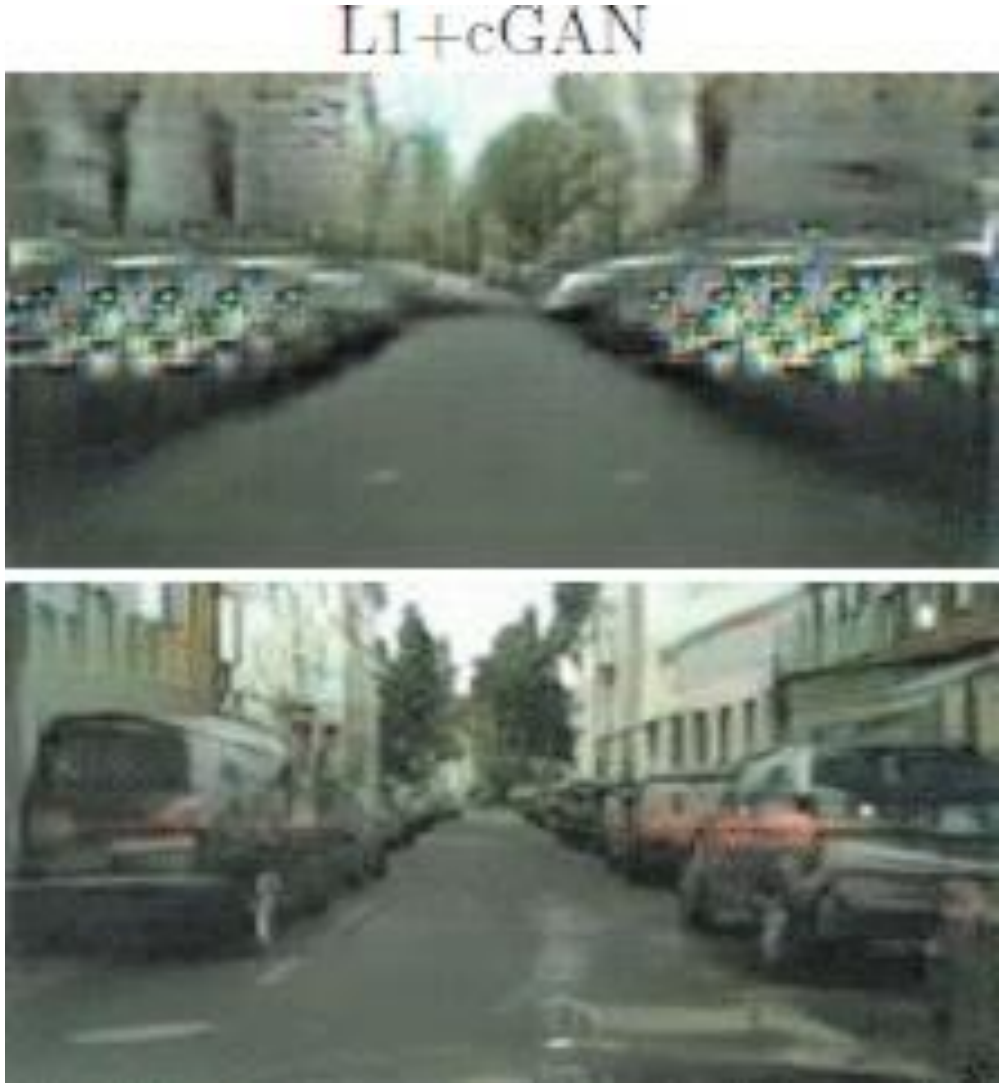
Sketches to Bags
Sketches to Shoes

sketches→shoes sketches→bag

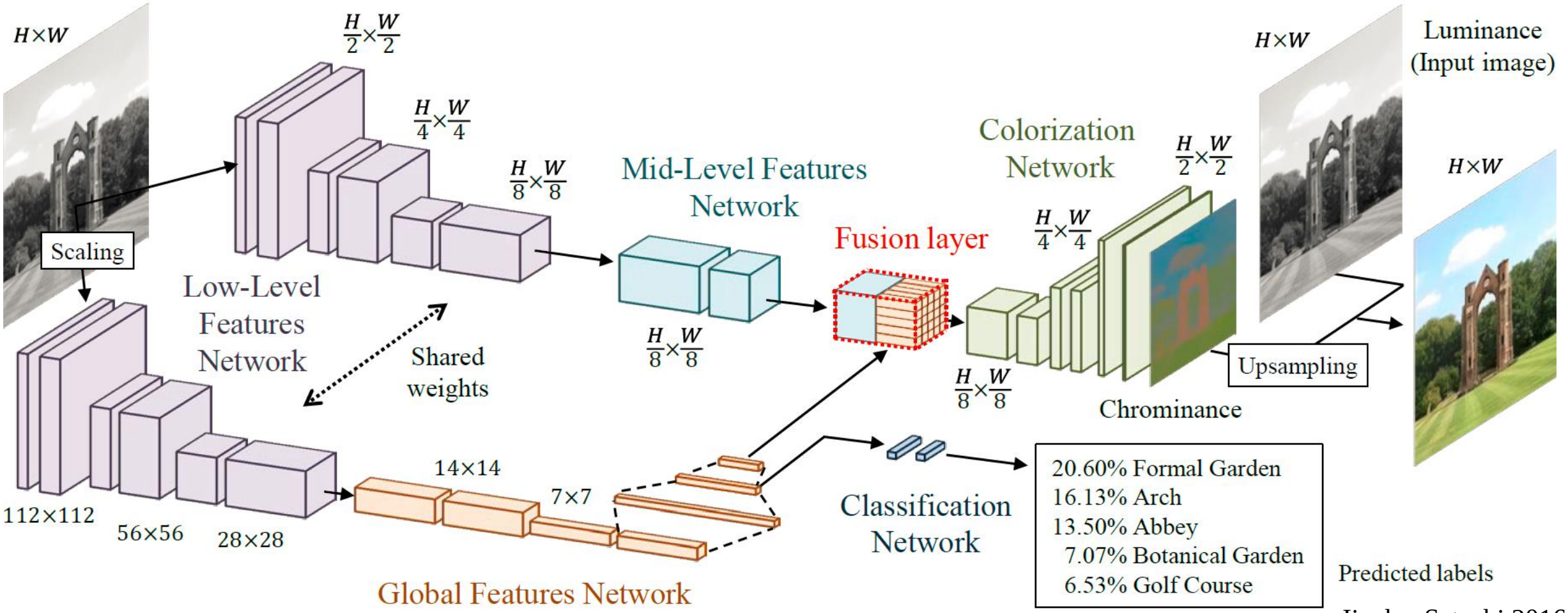
Effectiveness of U-Net

Traditional Encoder Decoder Network

U-Net



Global and Local Image Priors for Automatic Image Colorization



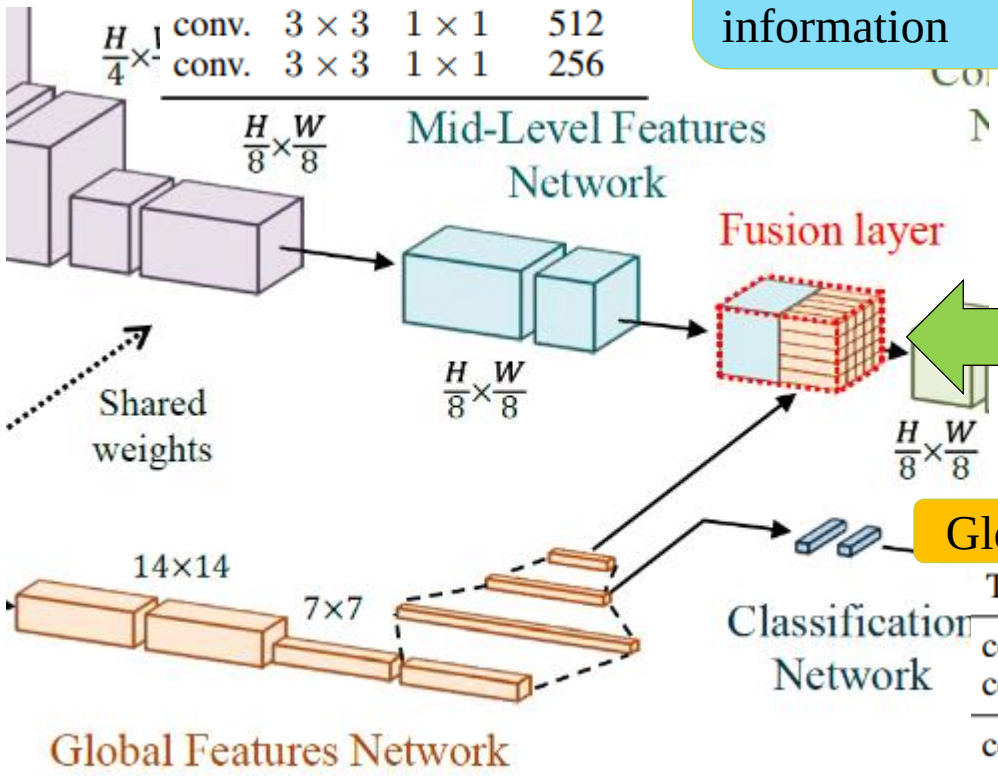
Iizuka, Satoshi 2016

Middle and Global Level Feature Network

Middle Level Feature Network

It consists of 2 convolution layer to collect middle level feature information

Type	Kernel	Stride	Outputs
conv.	3×3	1×1	512
conv.	3×3	1×1	256



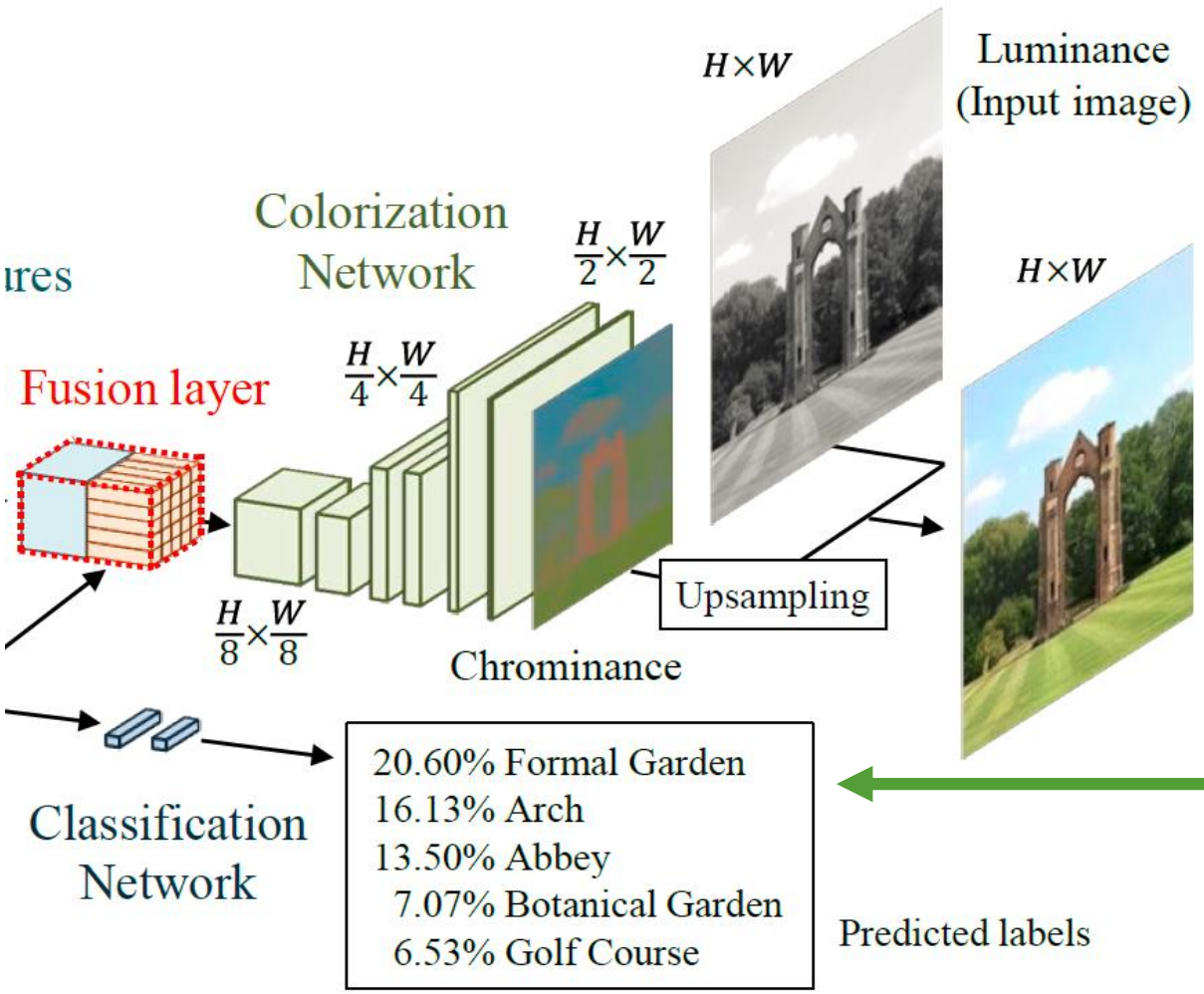
Fusion Layer:
Combine global feature(256) and mid level feature(256)

Global Feature Network

Type	Kernel	Stride	Outputs
conv.	3×3	2×2	512
conv.	3×3	1×1	512
conv.	3×3	2×2	512
conv.	3×3	1×1	512
FC	-	-	1024
FC	-	-	512
FC	-	-	256

It consists of 4 convolution layer and 3 fully connected layer to compute a 256 dimensional vector representation of image

Colorization

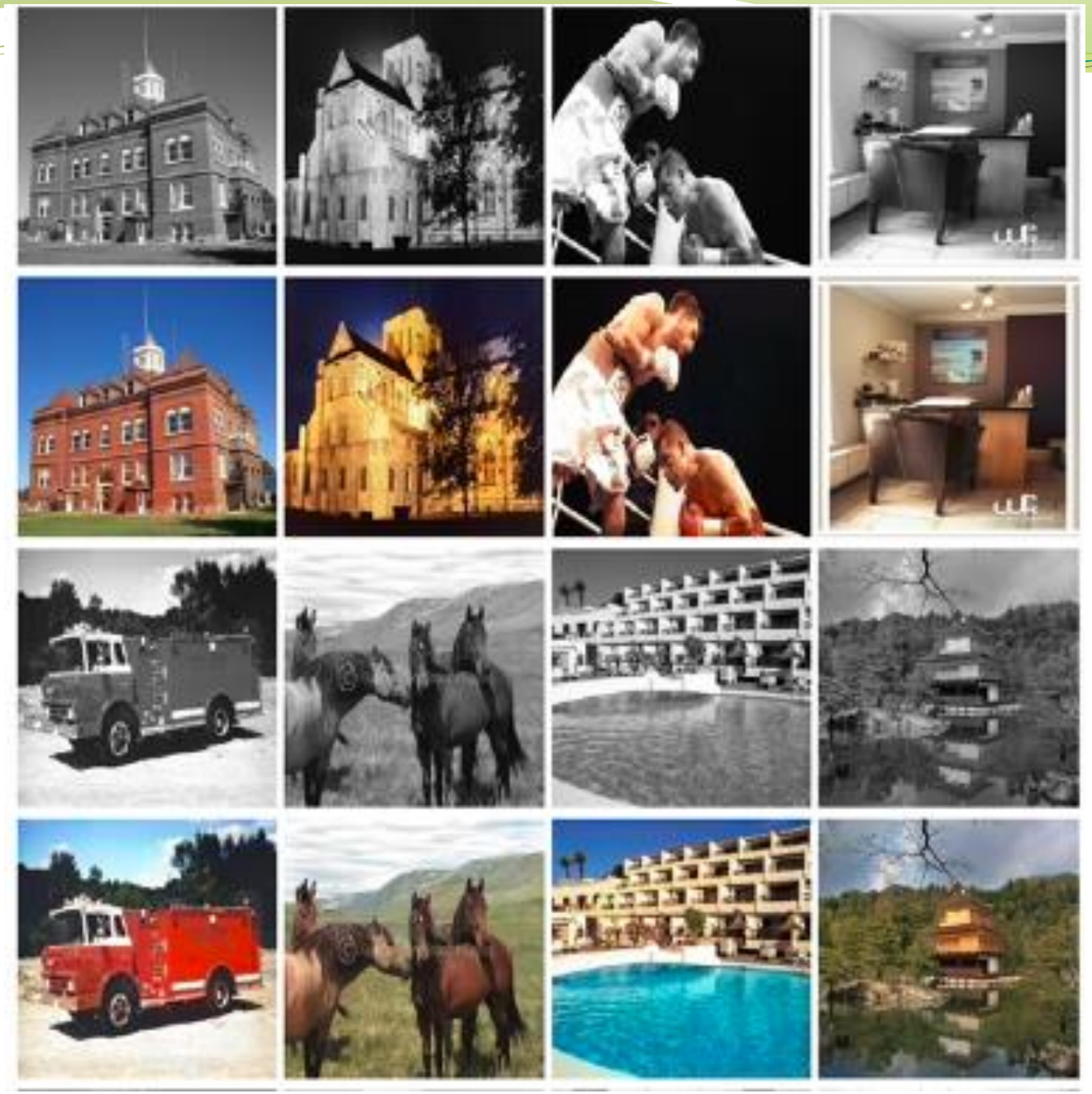


Colorization Network

Type	Kernel	Stride	Outputs
fusion	-	-	256
conv.	3 × 3	1 × 1	128
upsample	-	-	128
conv.	3 × 3	1 × 1	64
conv.	3 × 3	1 × 1	64
upsample	-	-	64
conv.	3 × 3	1 × 1	32
output	3 × 3	1 × 1	2

It consists of convolution layer and up sampling layer to expands feature map until twice as wide and tall.

Computer learns color and label at same time to understand global context.
It guide the training of the global image feature.



Result

- Each method has a field of excellence. For example, global local colorization is specialized for black and white picture colorization. Image to image translation method can be widely used in various fields.
- GAN will be used in various fields because of versatility.

Question?