WALC 2024
Applied AI

Introduction to Convolutions

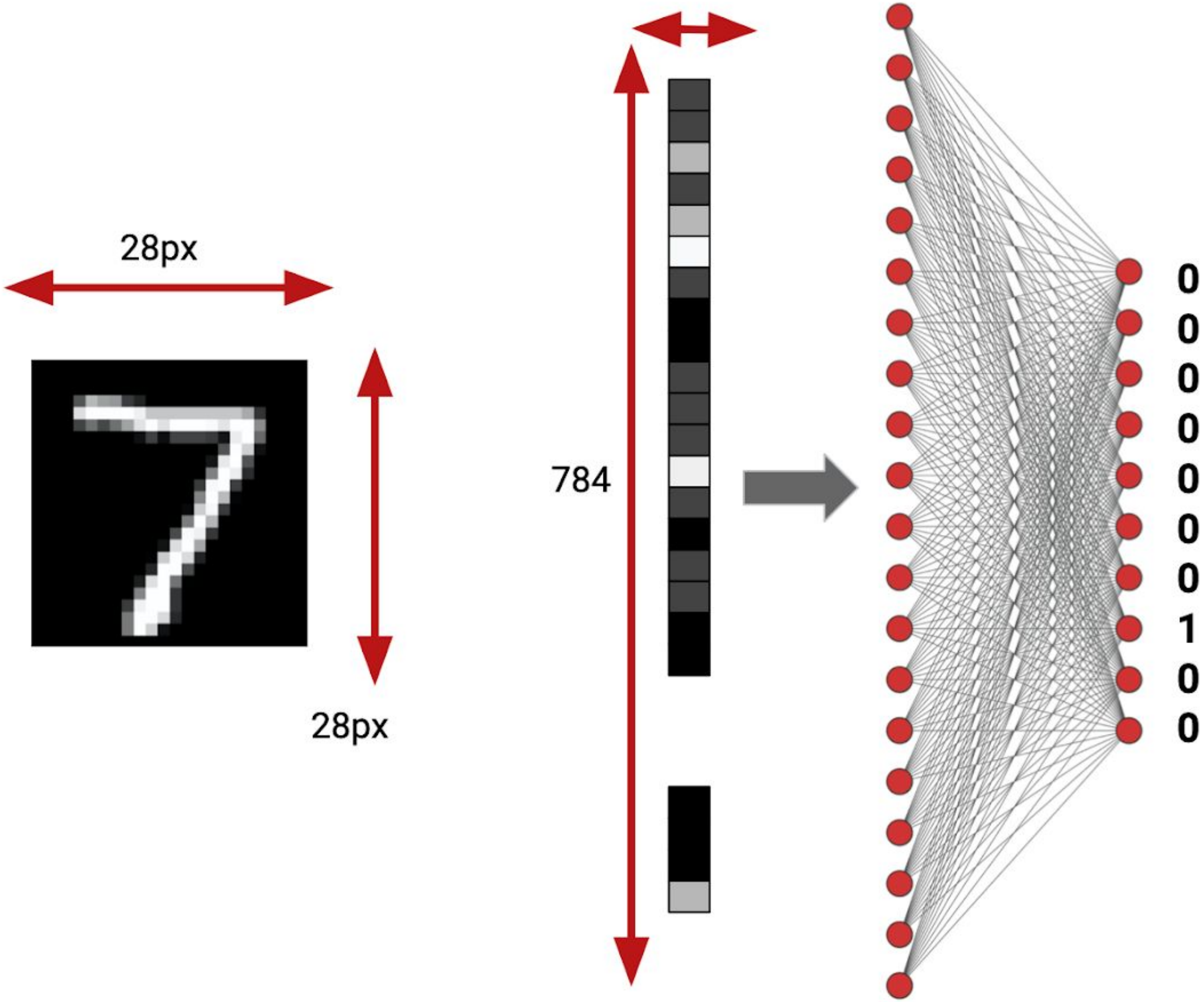
Prof. Marcelo J. Rovai
rovai@unifei.edu.br

UNIFEI - Federal University of Itajuba, Brazil
TinyML4D Academic Network Co-Chair

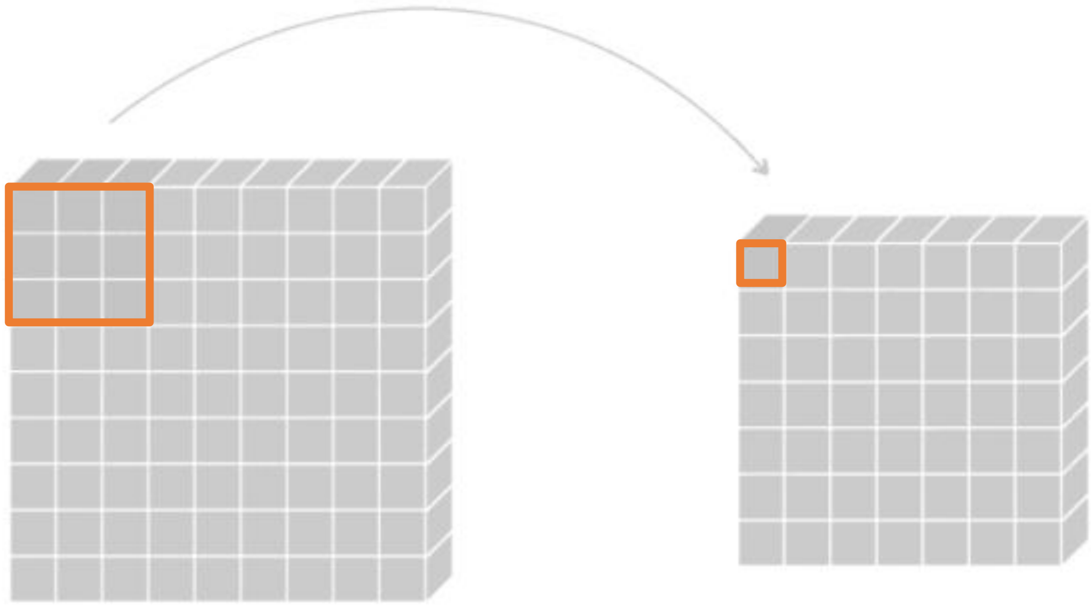
TINYML4D

Introducing Convolutions

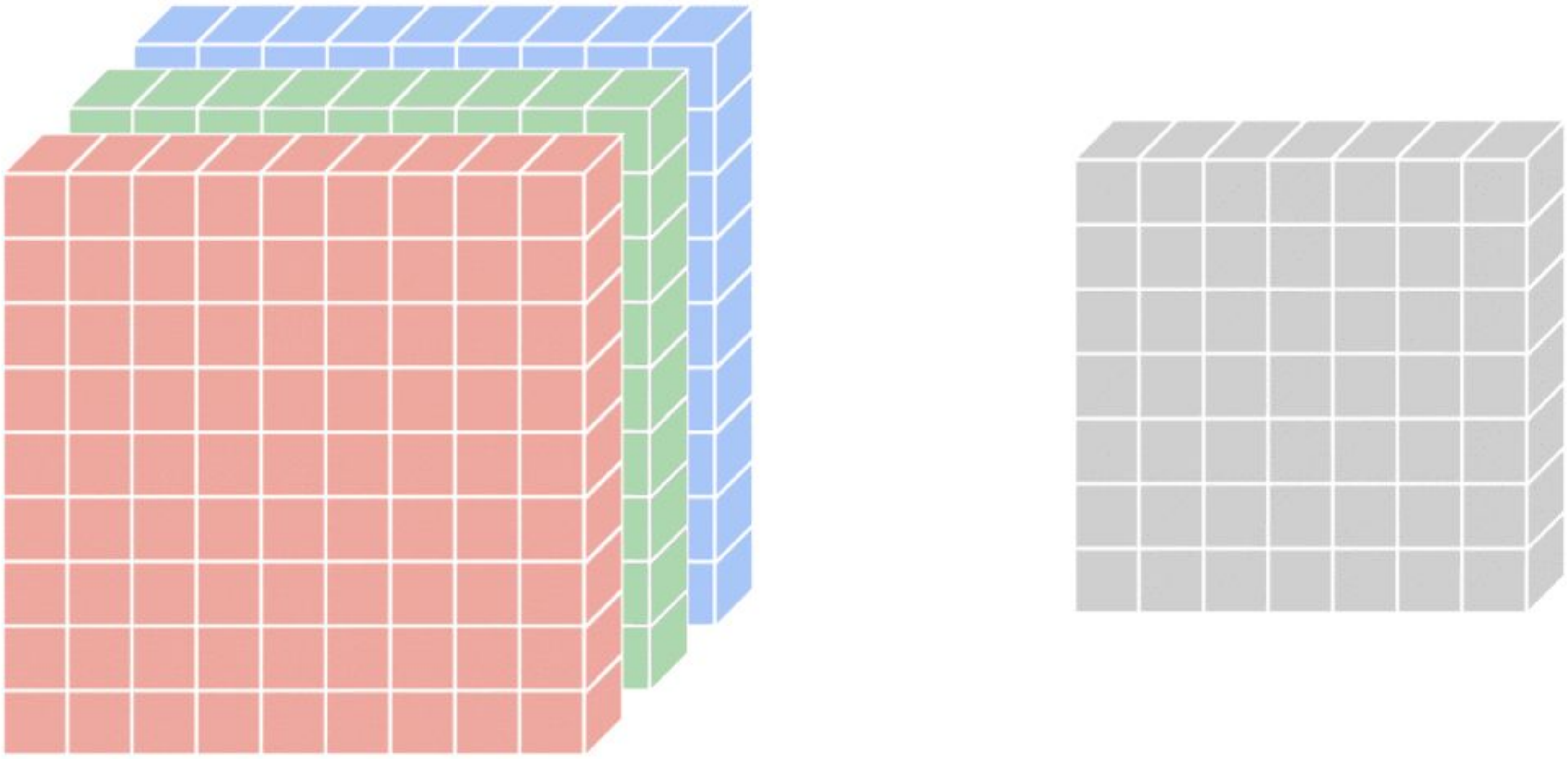
Beyond weights and biases…


28px
28px
784
0
0
0
0
0
0
0
1
0
0

Standard Convolution (1 Channel)

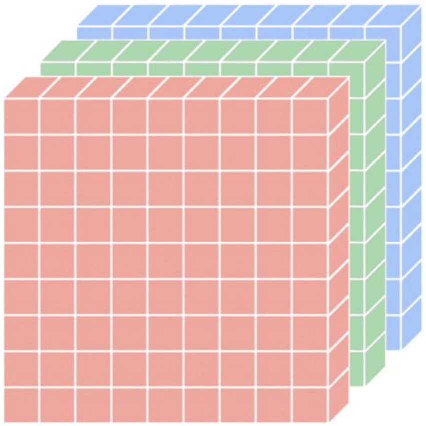


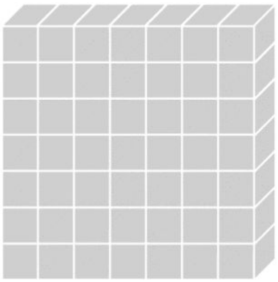
Standard Convolution ***(3 Channel**—e.g., RGB)*

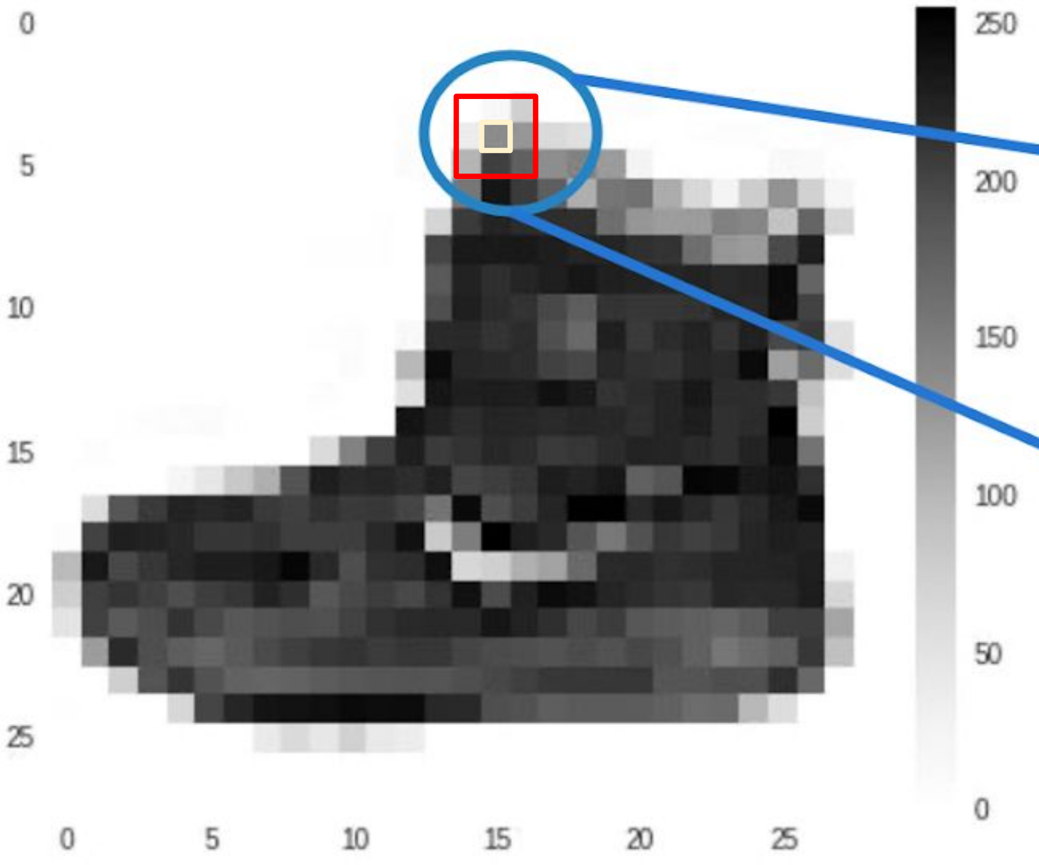


Standard Convolution ***(3 Channel**—e.g., RGB)*

- Input Feature Map
  - $8 \times 8 \times 3$
  - Width $\times$ Height $\times$ Channels
- Kernel (*1 Filter*)
  - $3 \times 3 \times 3$







0	64	128
48	192	144
142	226	168

Current Pixel Value is 192

Consider neighbor Values

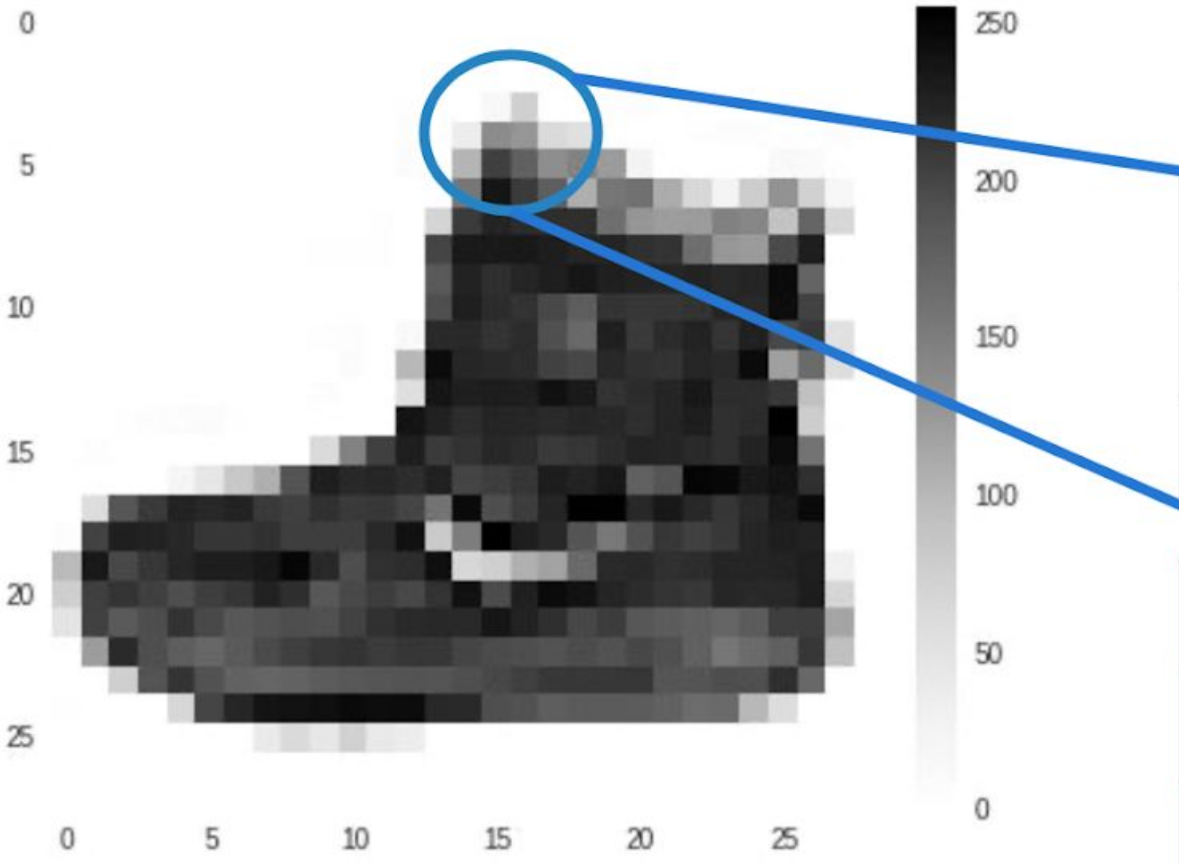
-1	0	-2
.5	4.5	-1.5
1.5	2	-3

Filter Definition

```
CURRENT_PIXEL_VALUE = 192
NEW_PIXEL_VALUE = (-1 * 0) + (0 * 64) + (-2 * 128) +
                  (.5 * 48) + (4.5 * 192) + (-1.5 * 144) +
                  (1.5 * 42) + (2 * 226) + (-3 * 168)
```

-256
672
11

427 New Pixel Value



Kernels = Filters

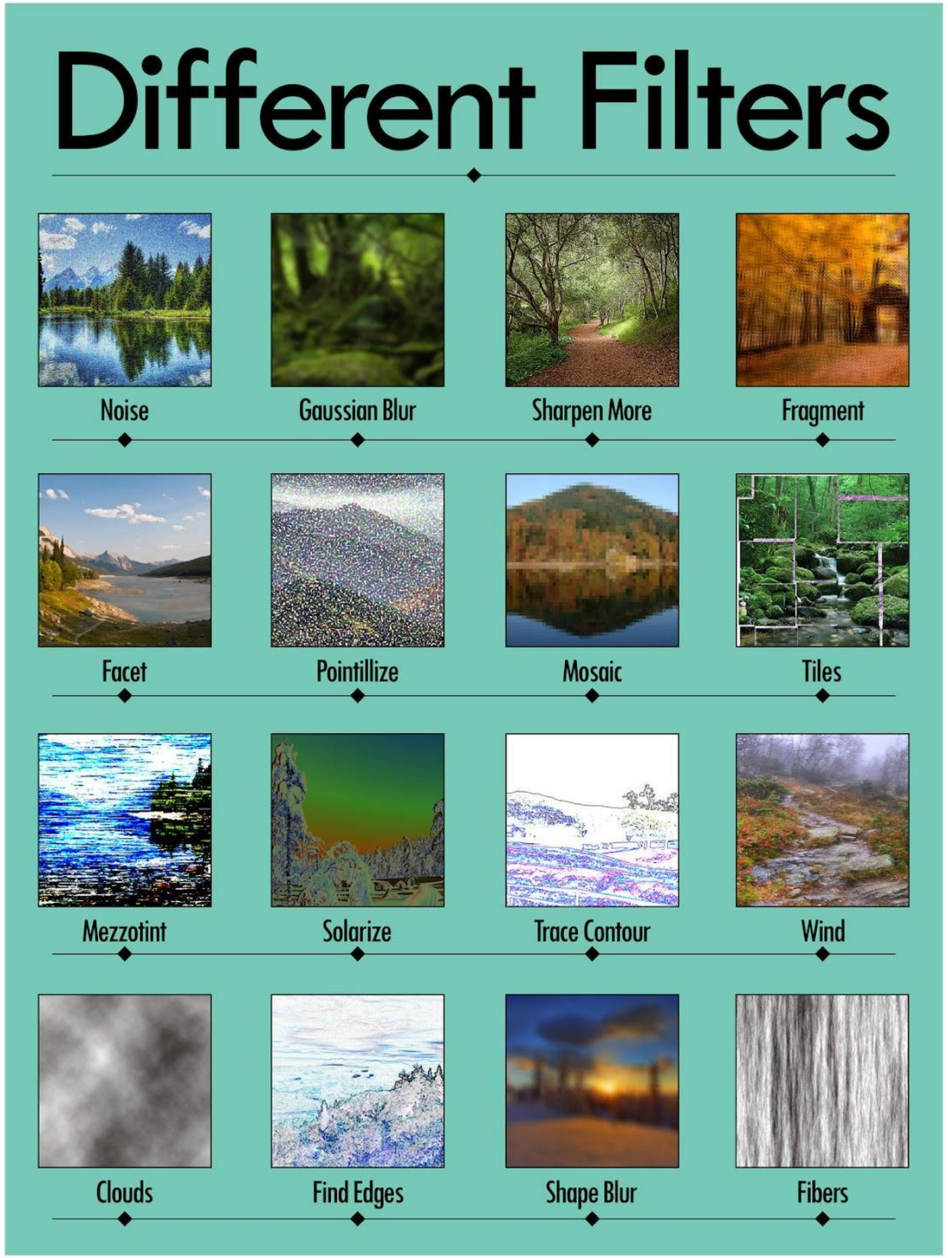


Image Kernels

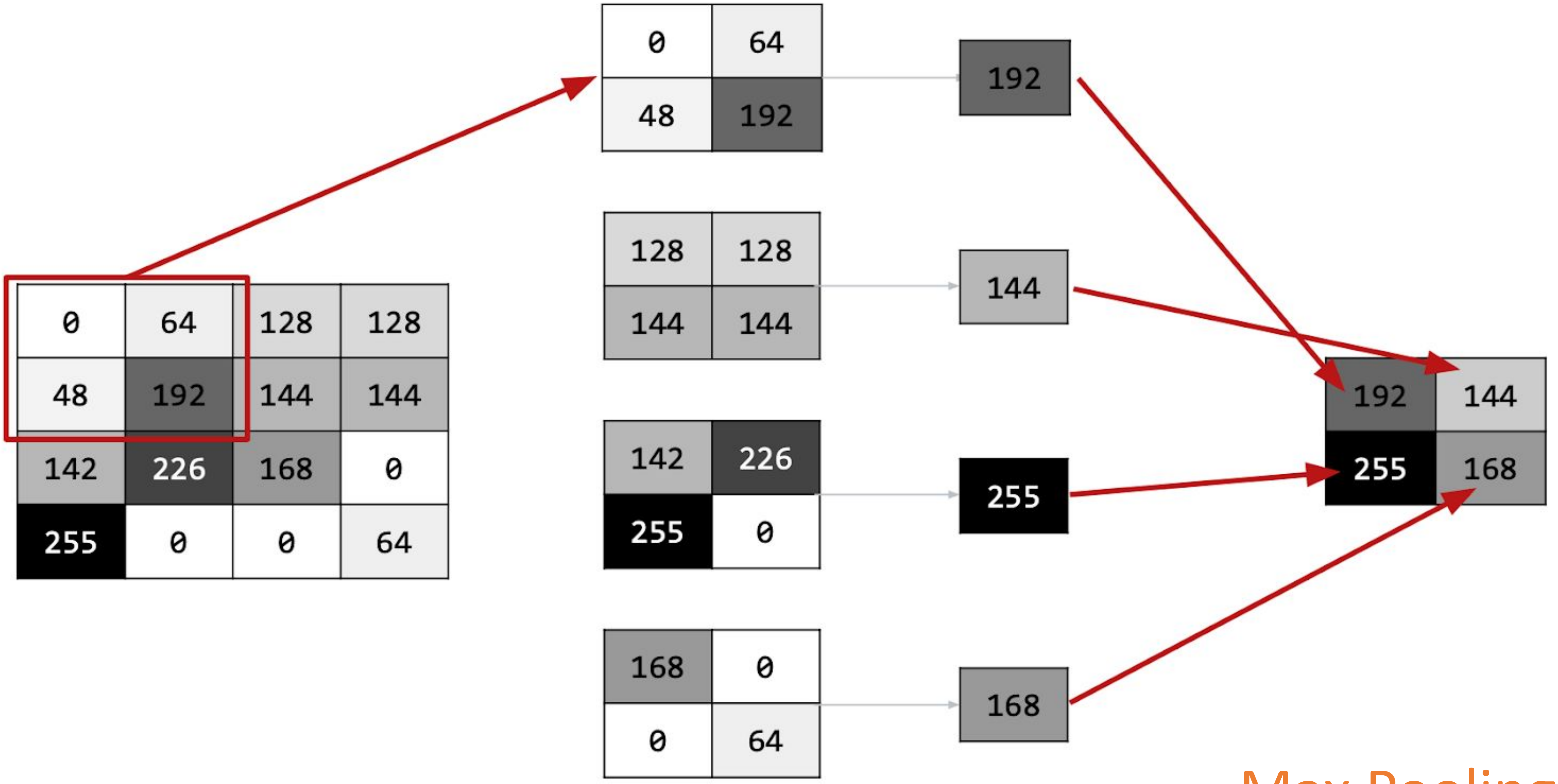
-1	0	1
-2	0	2
-1	0	1

custom

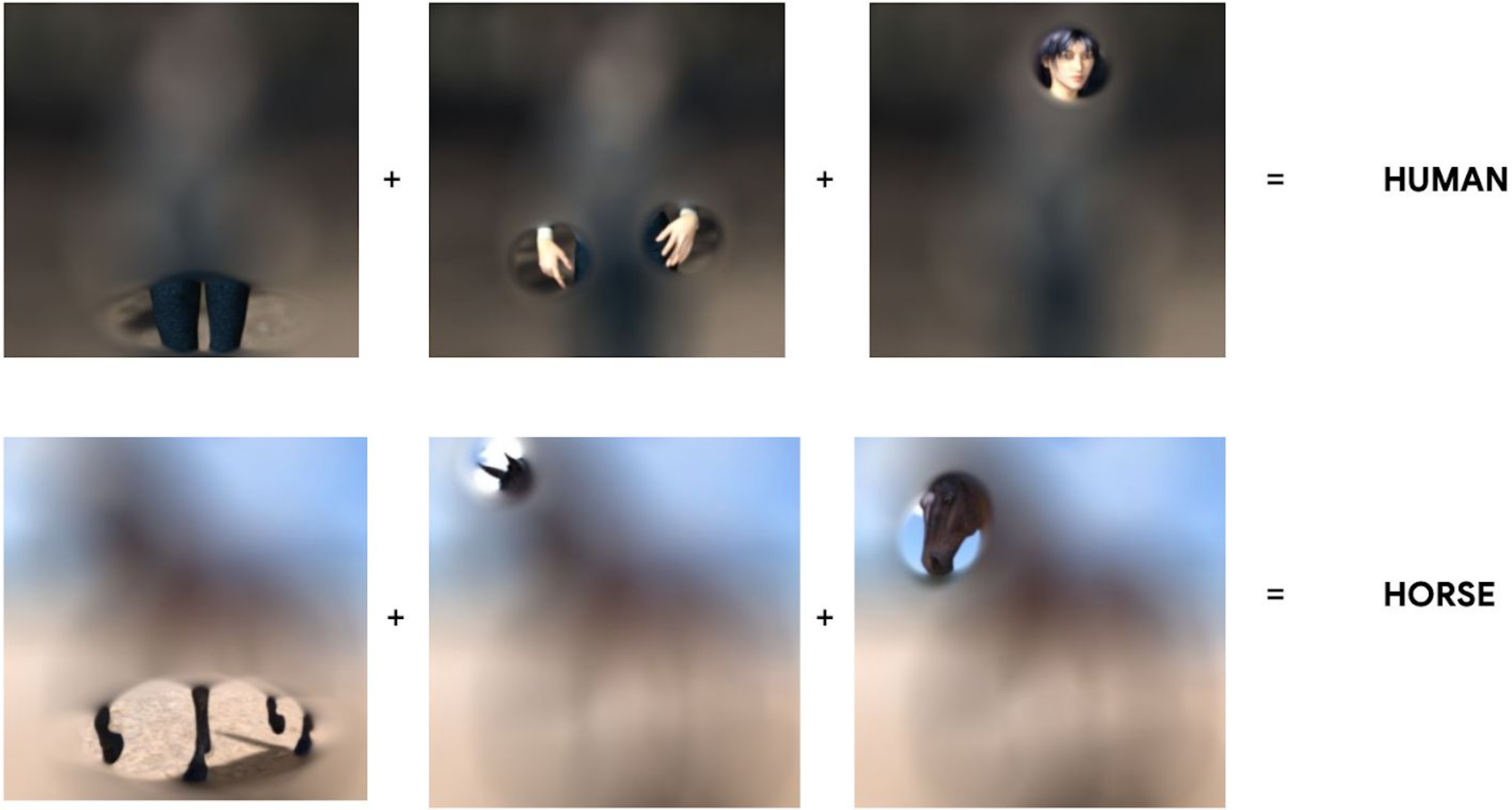
-1	-2	-1
0	0	0
1	2	1

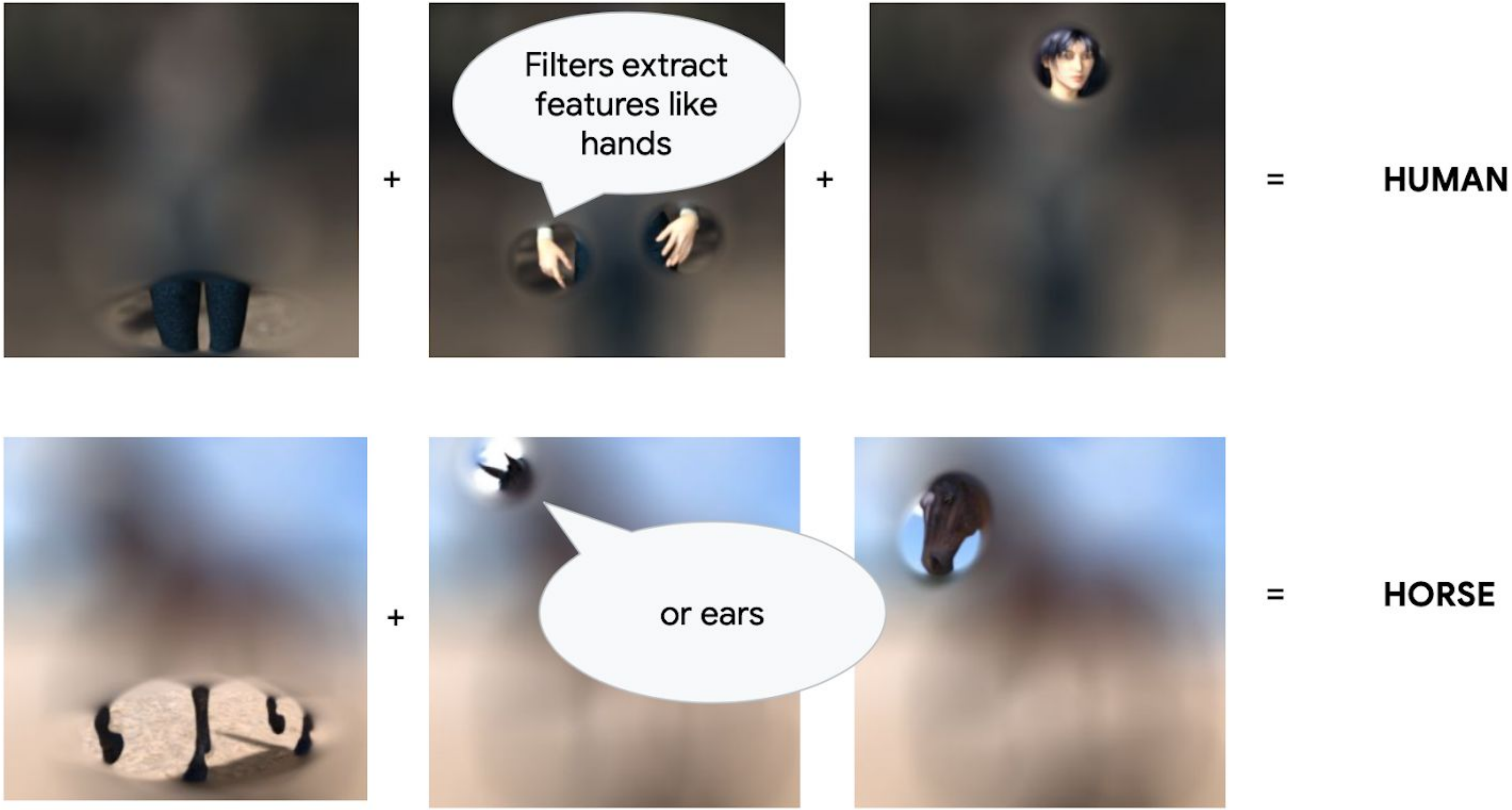
custom

https://setosa.io/ev/image-kernels/

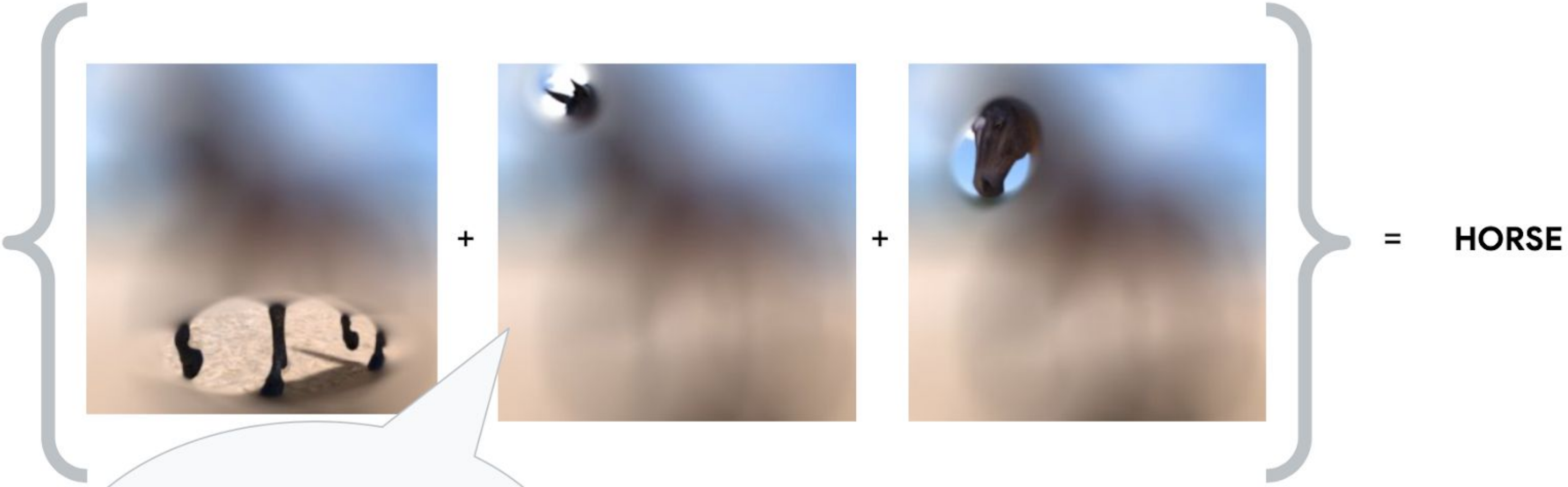

0 64 128 128
48 192 144 144
142 226 168 0
255 0 0 64
0 64
48 192
128 128
144 144
142 226
255 0
168 0
0 64
192
144
255
168
192 144
255 168

Max Pooling


+
+
=
HUMAN
+
+
=
HORSE

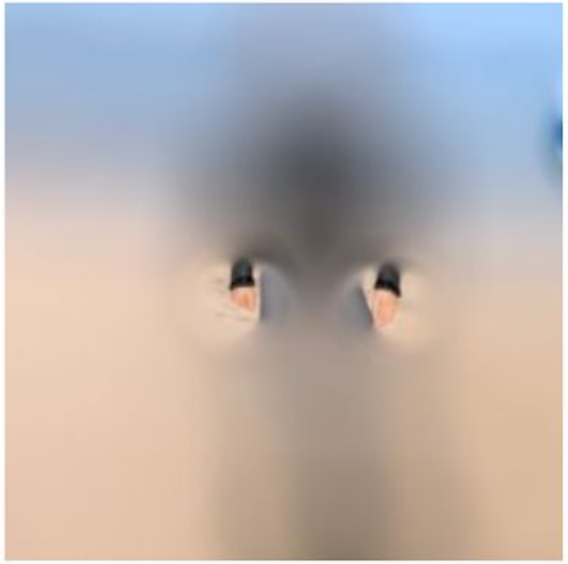


Filters extract features like hands
+
+
= HUMAN
+
or ears
= HORSE

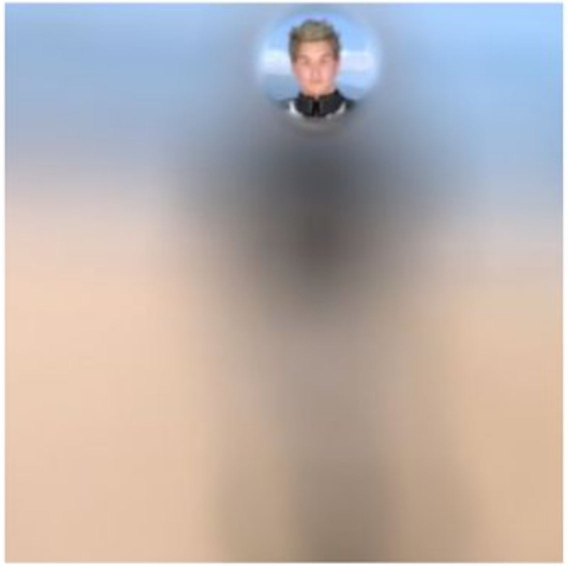


Filters can then be combined with **labels** to make a **prediction** of the image contents...

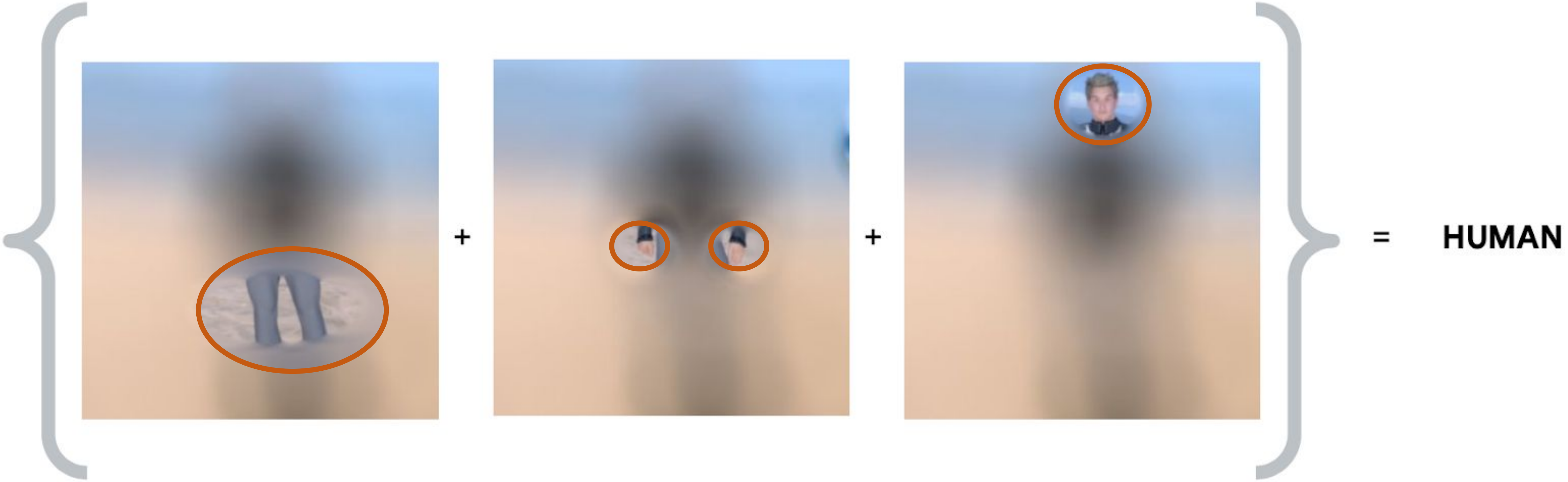
+

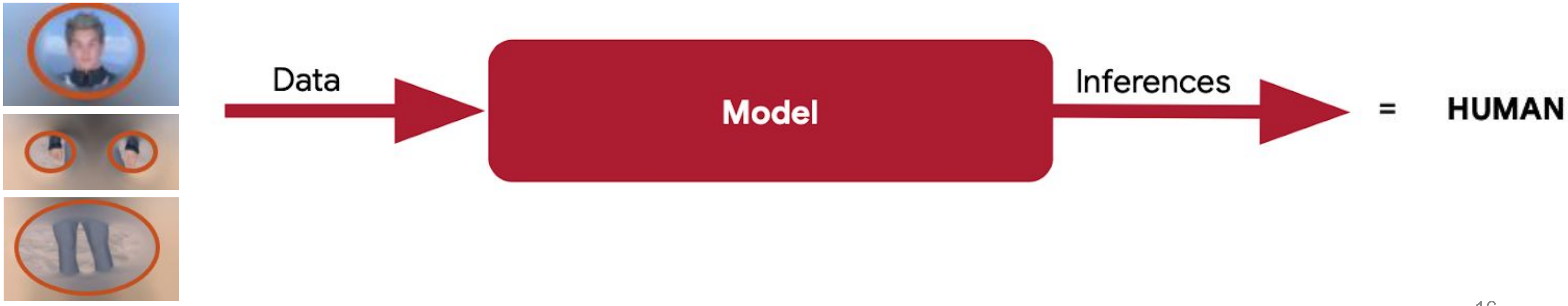


+

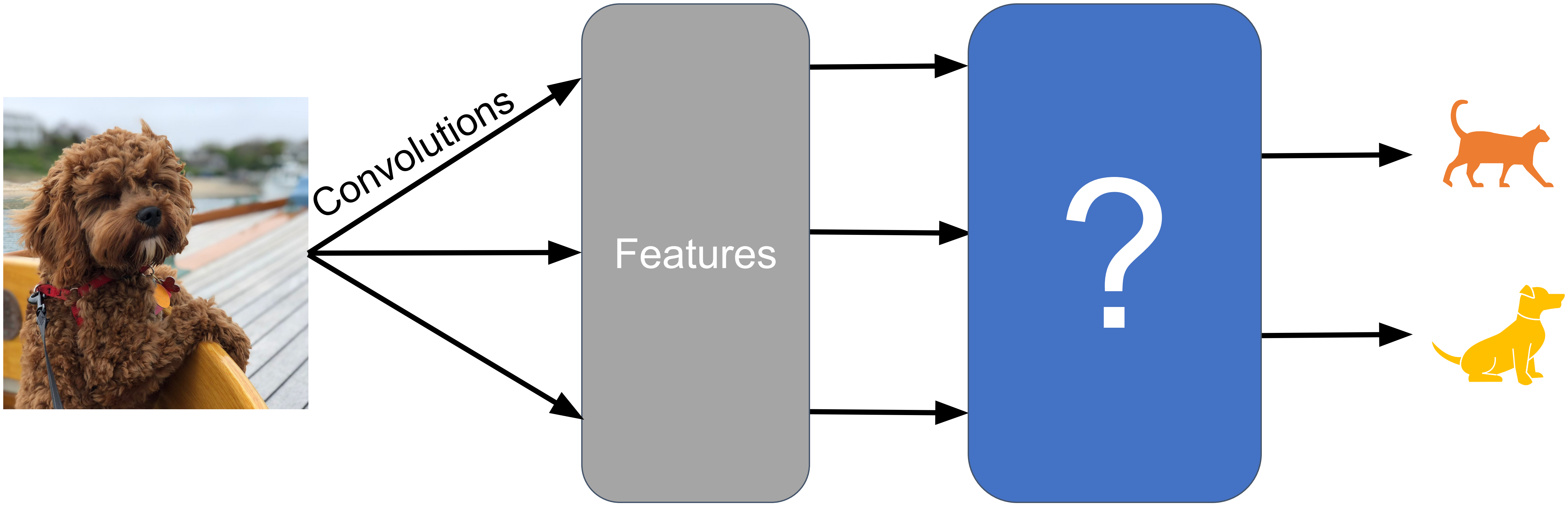


= **HUMAN**

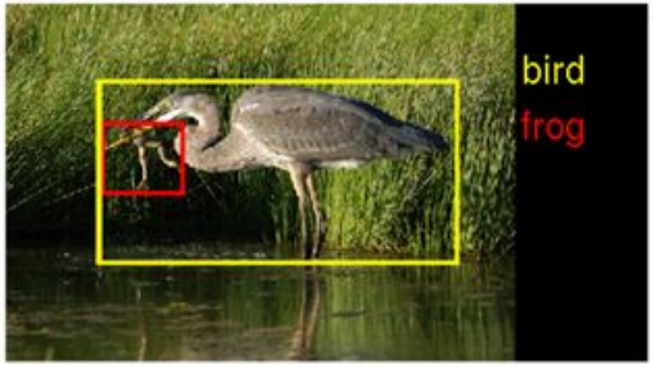

+
+
=
HUMAN

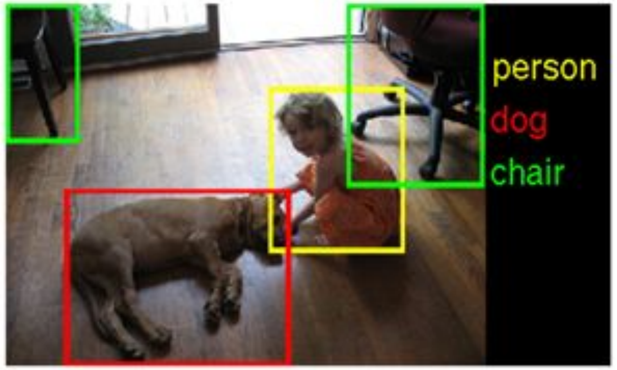

Data
Model
Inferences
=
HUMAN

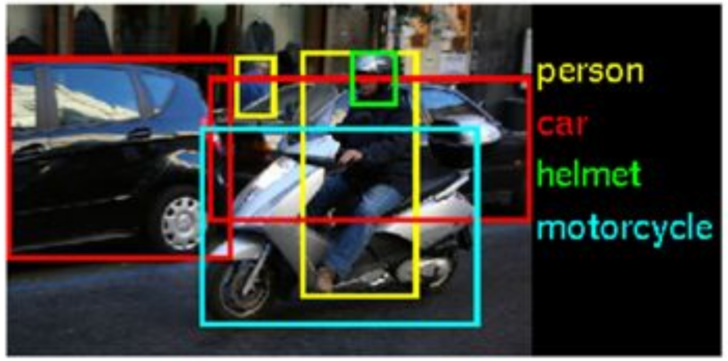
How might we combine these features to **classify an object**?



The ImageNet Challenge and the Birth of CNNs







The ImageNet Challenge provided 1.2 million examples of 1,000 **labeled** items and challenged algorithms to learn from the data and then was tested on another 100,000 images

18

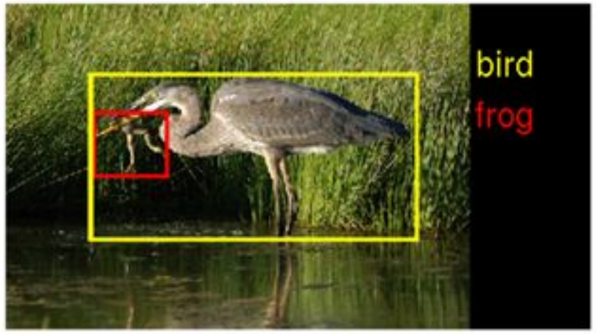
The ImageNet Challenge and the Birth of CNNs

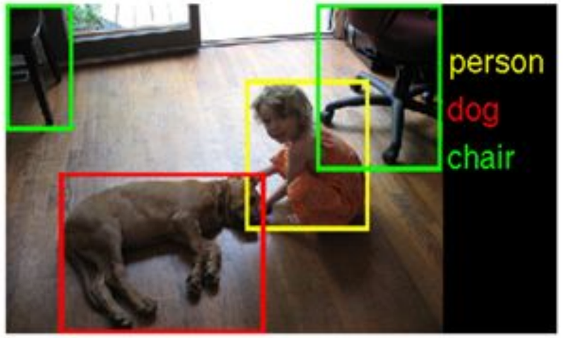
Traditional Machine Learning Flow

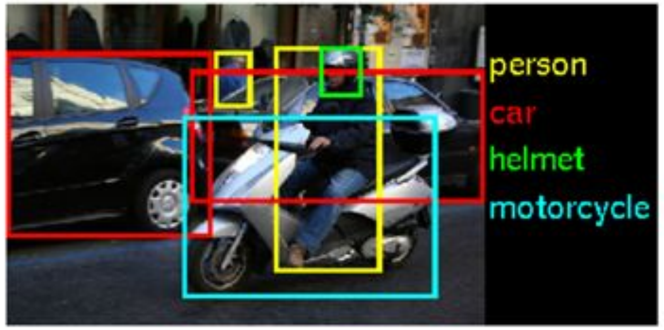
Vertical Lines, Horizontal Lines, Changes in Color, Changes in Focus, etc.

Regression, Clustering, etc.

The ImageNet Challenge and the Birth of CNNs

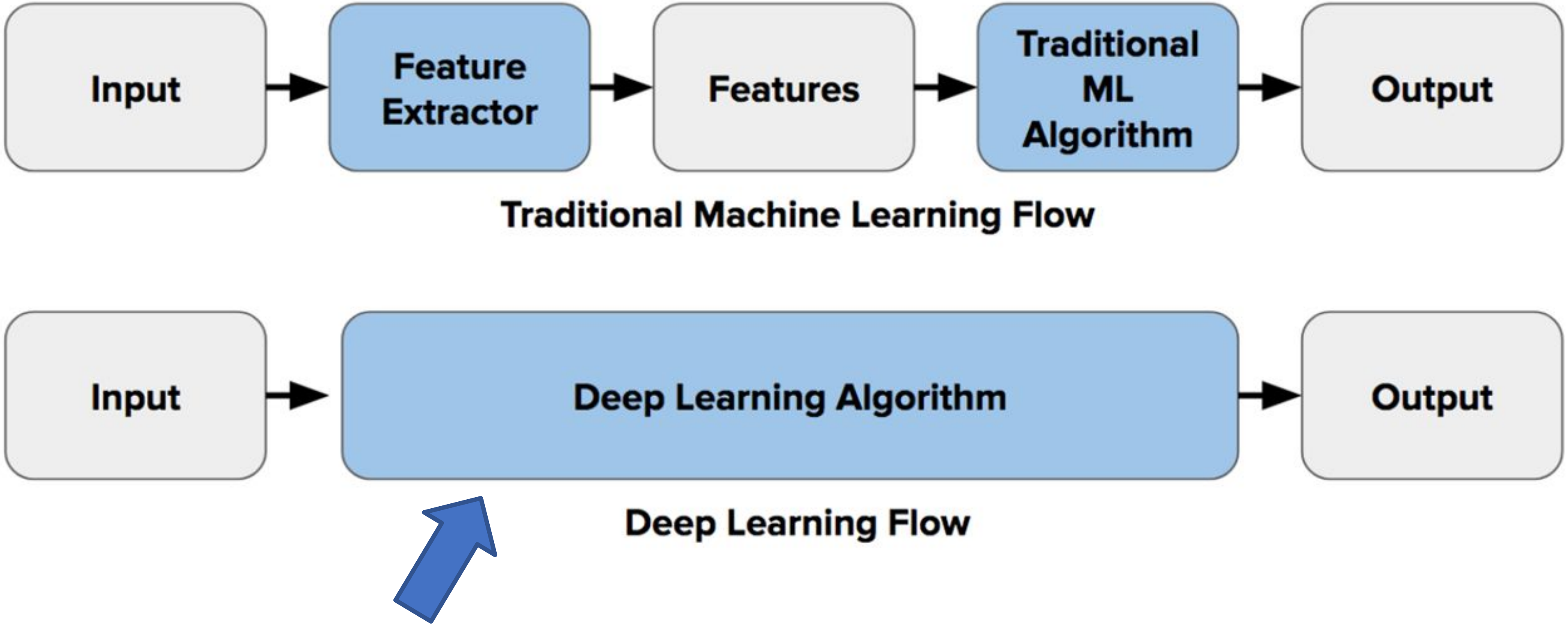






- In 2010, teams had a 75-50% error
- In 2011, teams had 75-25% error
- In 2012 still, no team had less than 25% error barrier, except AlexNet at 15% (Top-5)

The ImageNet Challenge and the Birth of CNNs

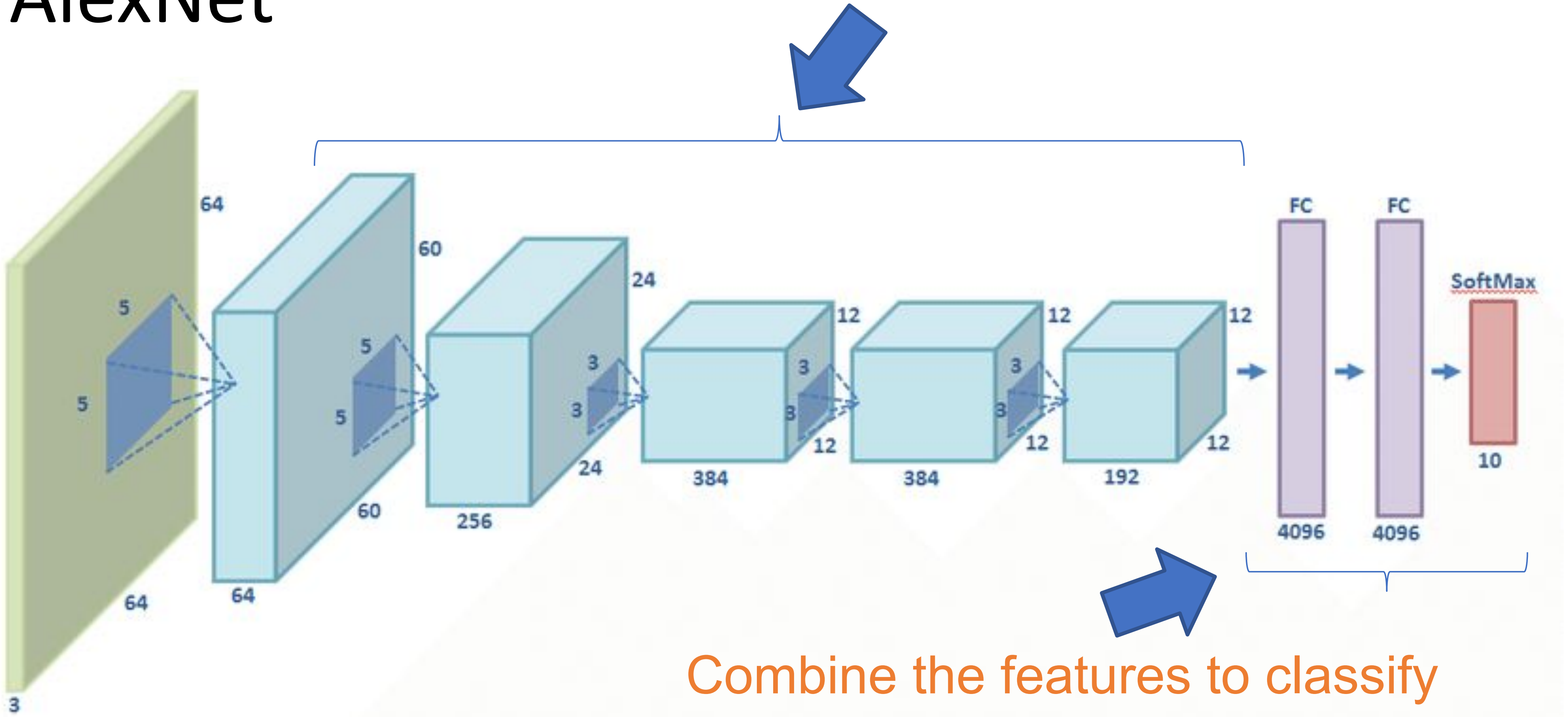


Let the computer figure out its features and how to combine them!

[AlexNet Paper]

AlexNet

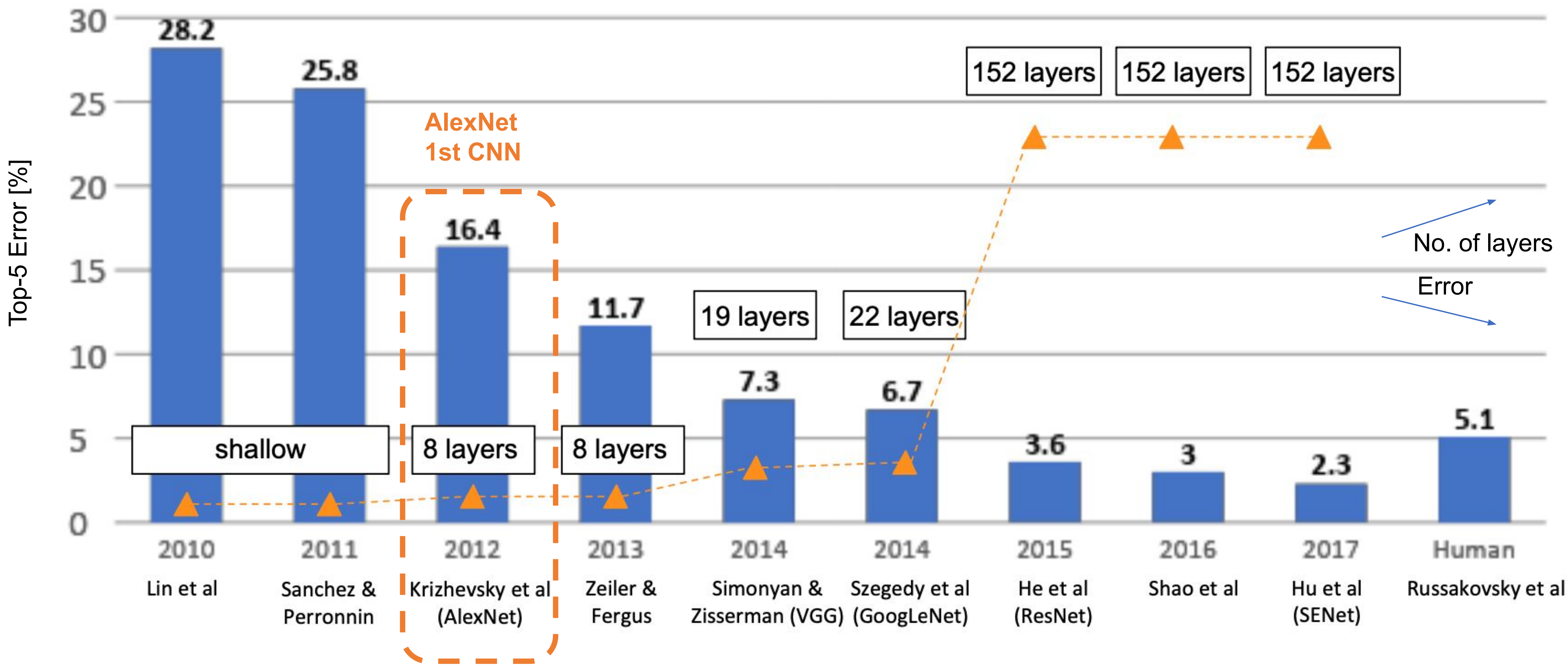
Use convolutions to find features and summarize them into higher-level features



Combine the features to classify the various objects in the dataset.

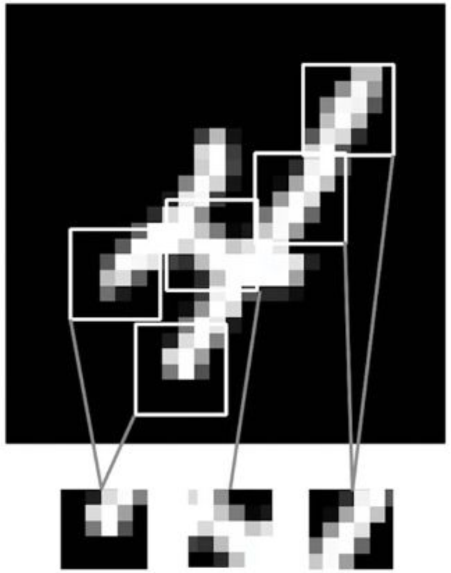
22

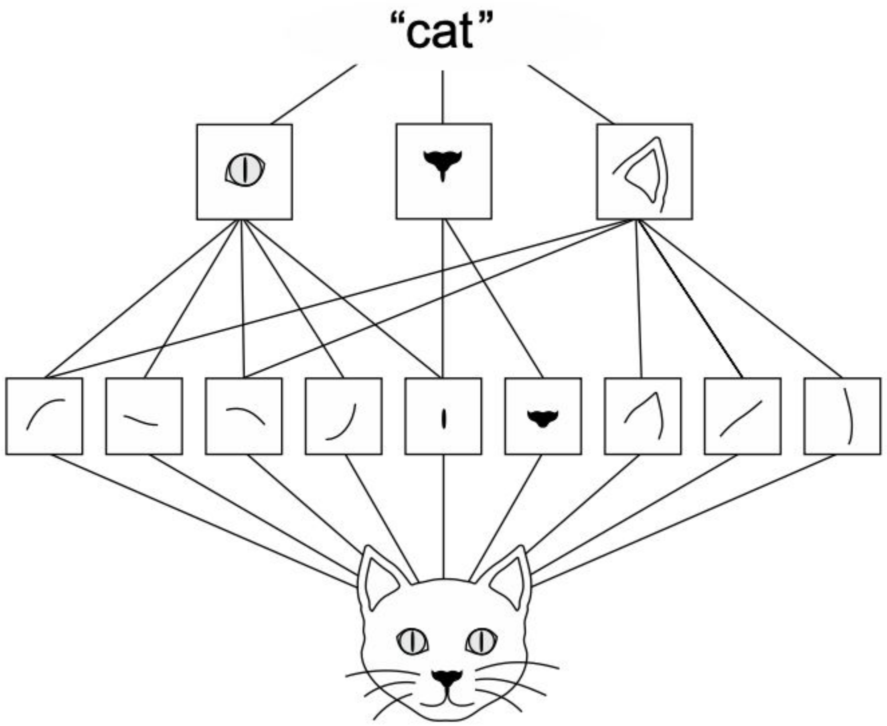
ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



The convolution operation

The fundamental difference between a densely connected layer and a convolution layer is this: **Dense layers learn global patterns** in their input feature space (for example, for an MNIST digit, patterns involving all pixels), whereas **convolution layers learn local patterns**—in the case of images, patterns found in small 2D windows of the inputs In the previous example, these windows were all 3 × 3.





They can learn spatial hierarchies of patterns. A first convolution layer will learn small local patterns such as edges, a second convolution layer will learn larger patterns made of the features of the first layers, and so on.

"Deep Learning with Python" by François Chollet

Image Classification with CNN

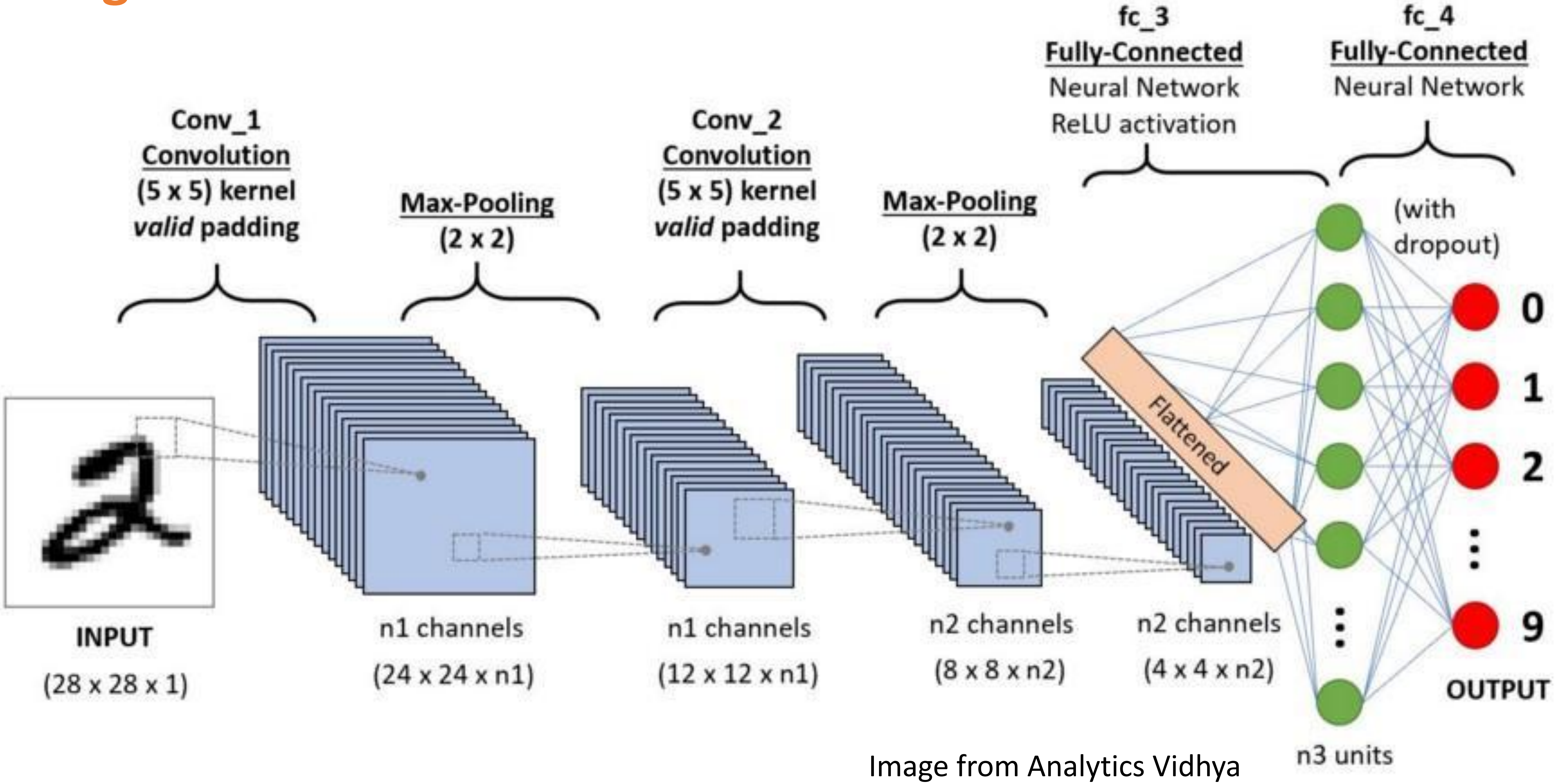


Image from Analytics Vidhya

Exploring CNN

ConvNetJS MNIST demo
https://cs.stanford.edu/people/karpathy/convnetjs/demo/mnist.html

ConvNetJS CIFAR-10 demo
https://cs.stanford.edu/people/karpathy/convnetjs/demo/cifar10.html

CNN Explainer
https://poloclub.github.io/cnn-explainer/

Questions?