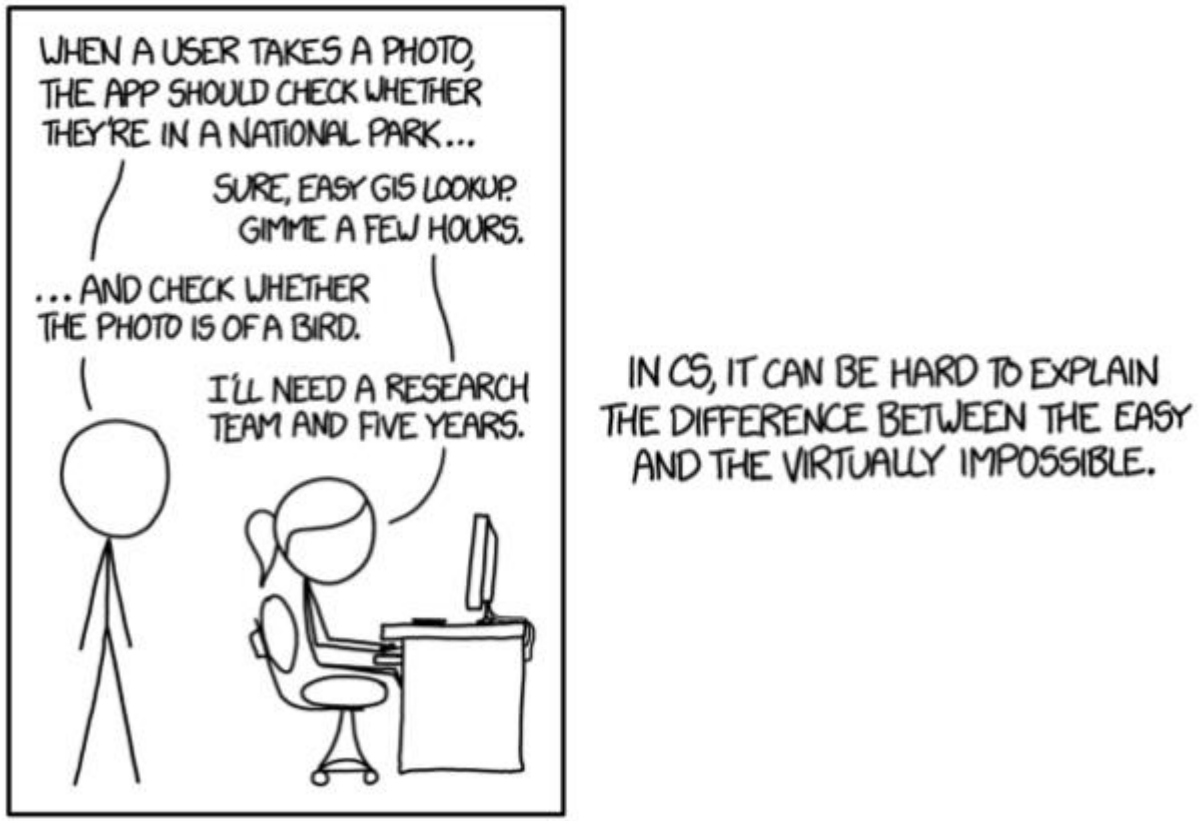
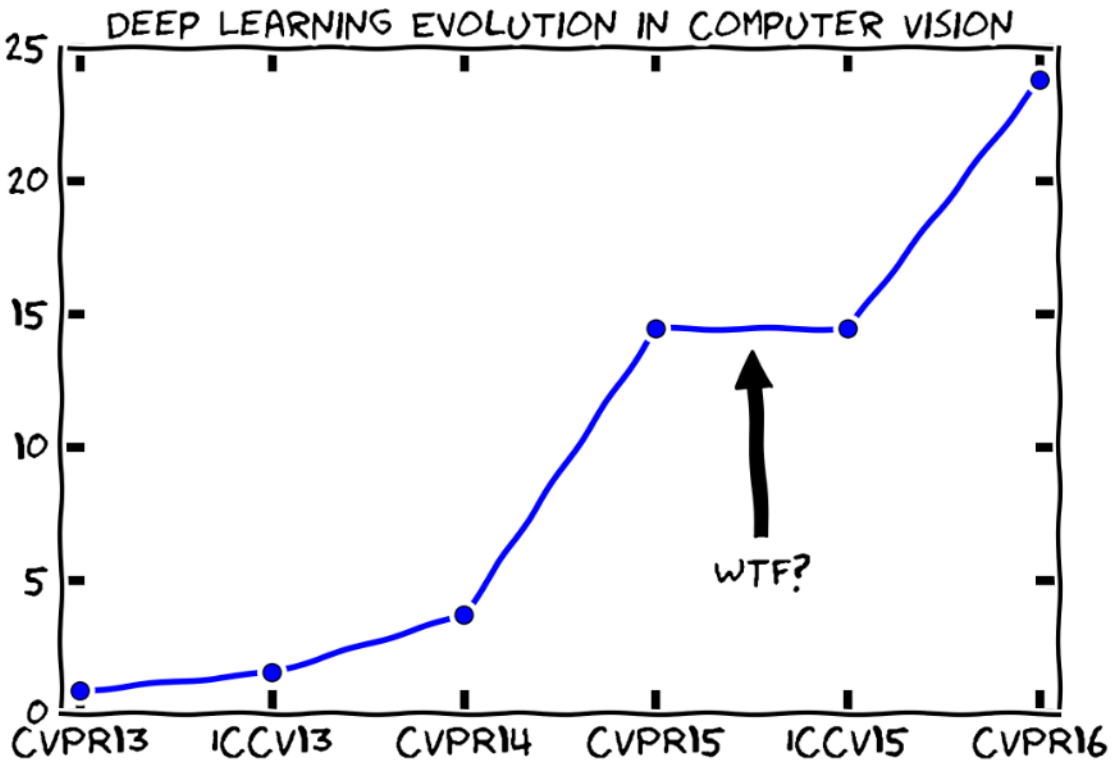
Image data



Xkcd, 2015

Image data

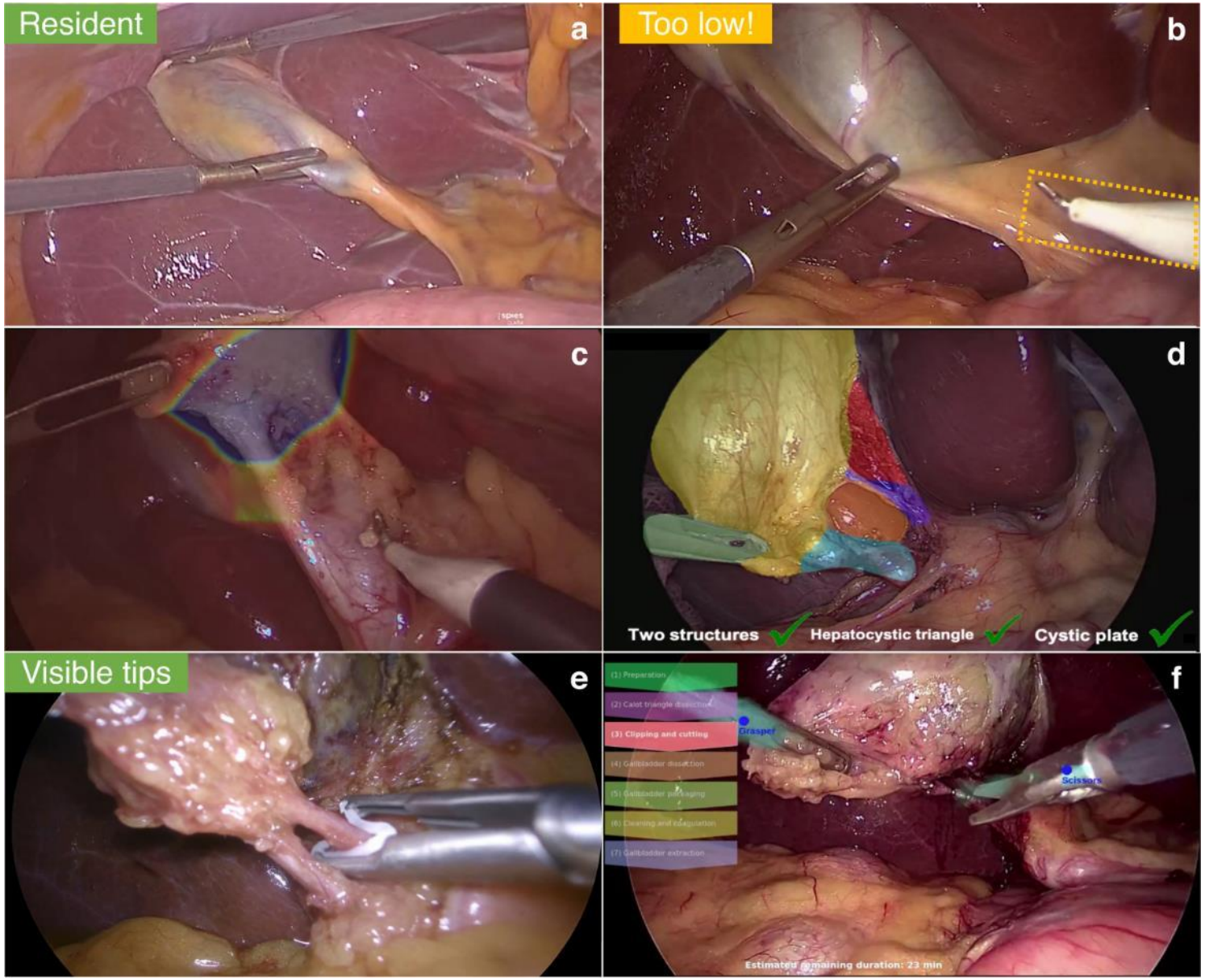


Xkcd

Computer vision

Derive meaningful information from digital images, videos and other visual inputs.

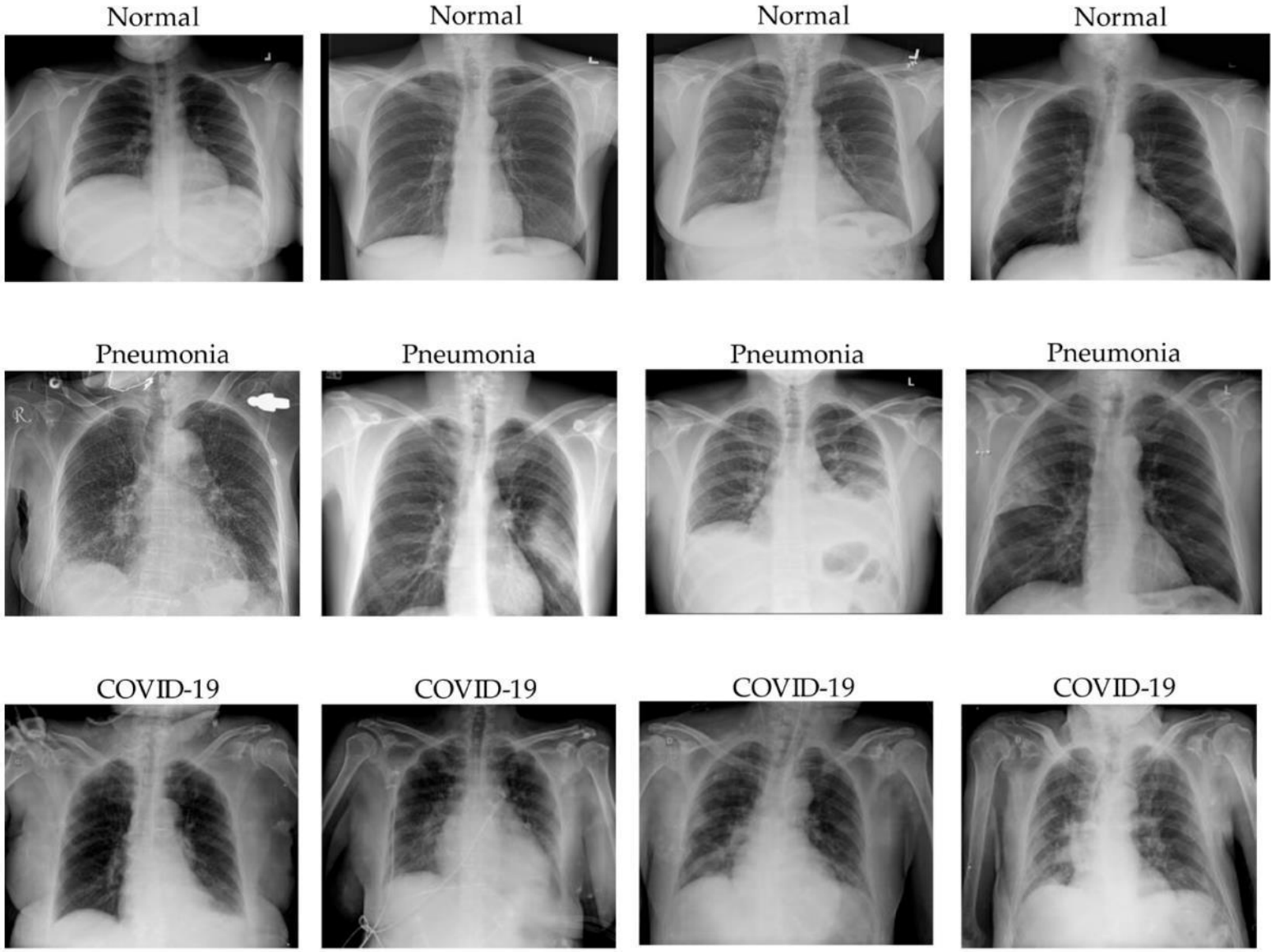
Take actions or make recommendations based on that information.



Mascagni, P., Alapatt, D., Sestini, L. et al. Computer vision in surgery: from potential to clinical value. npj Digit. Med. 5, 163 (2022)

Computer vision

Classification: predict the class(es) an image or video belongs to.

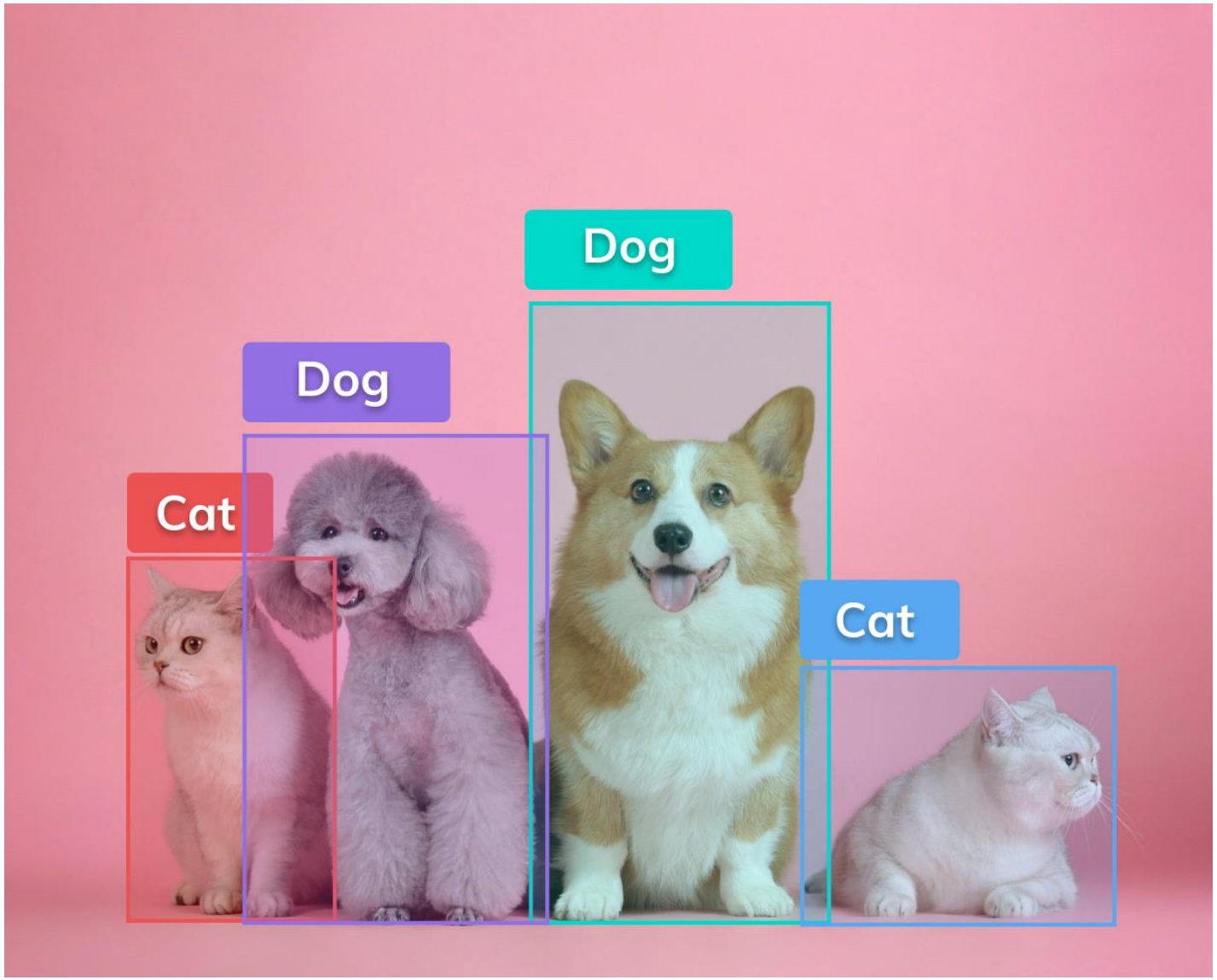


Le Dinh, T, et al. COVID-19 Chest X-ray Classification and Severity Assessment Using Convolutional and Transformer Neural Networks. Appl. Sci. 2022, 12, 4861

Computer vision

Classification: predict the class(es) an image or video belongs to.

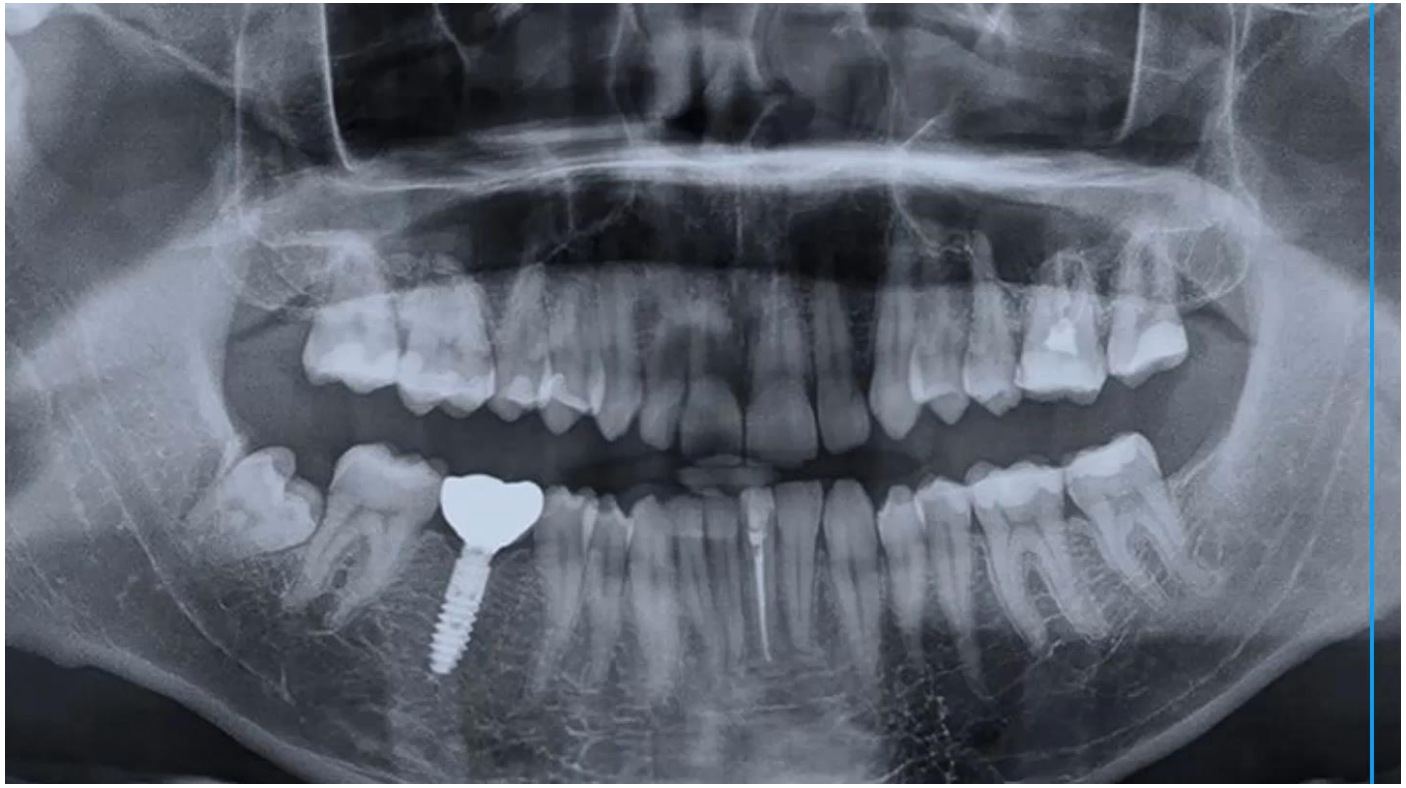
Object detection: identify a certain class of image and then detect and tabulate their appearance in an image or video.



Computer vision

Classification: predict the class(es) an image or video belongs to.

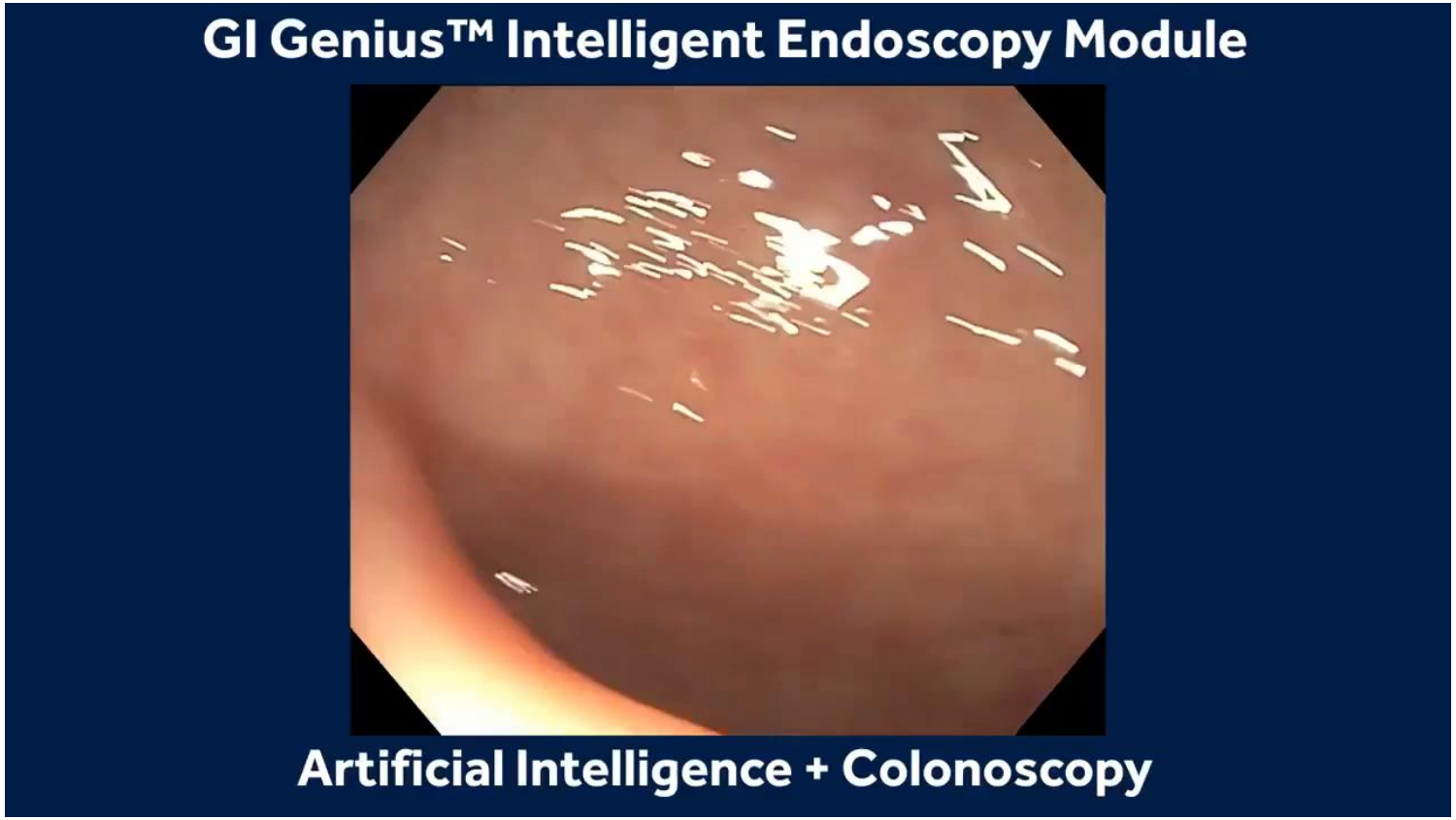
Object detection: identify a certain class of image and then detect and tabulate their appearance in an image or video.



Computer vision

Classification: predict the class(es) an image or video belongs to.

Object detection: identify a certain class of image and then detect and tabulate their appearance in an image or video.

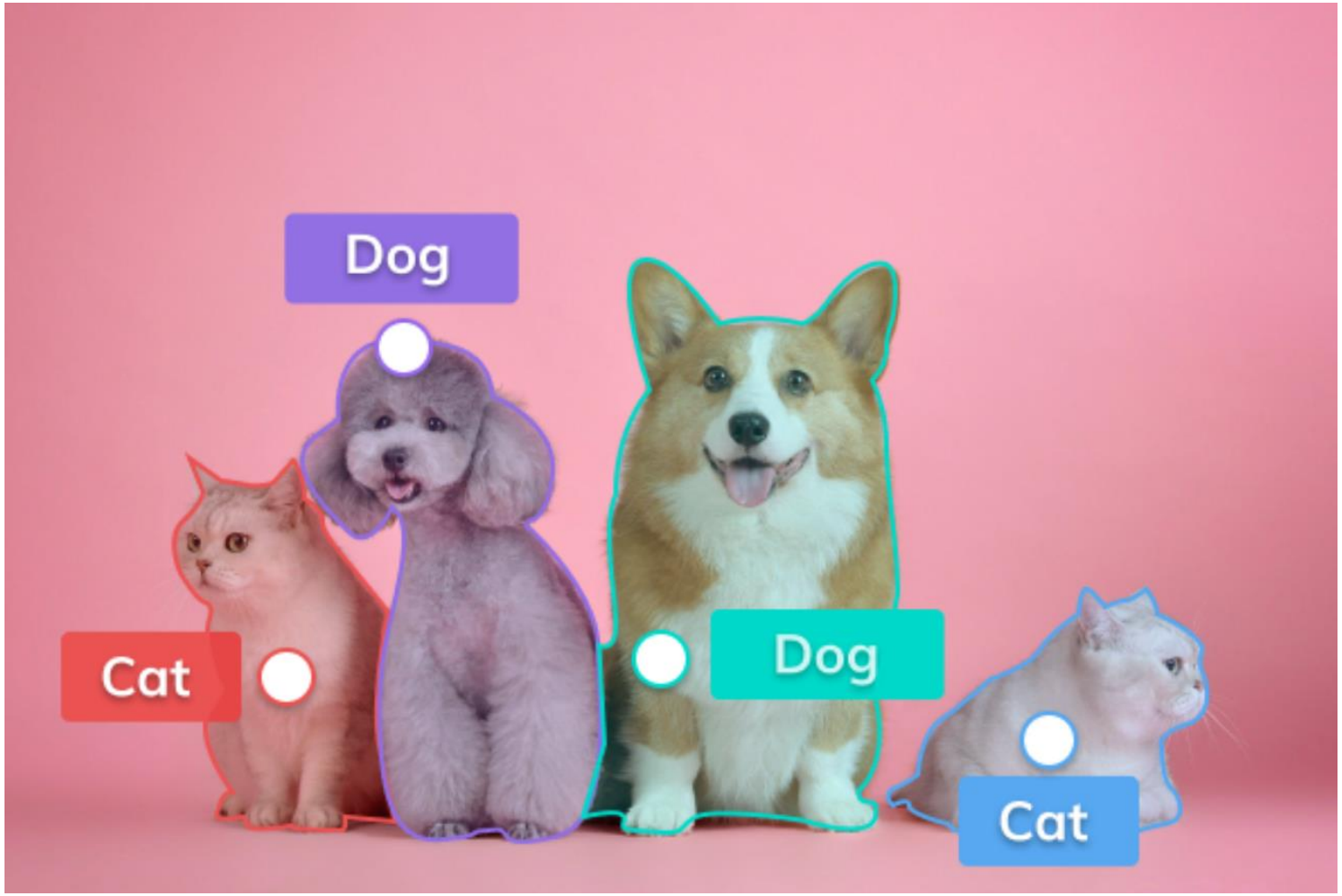


Computer vision

Classification: predict the class(es) an image or video belongs to.

Object detection: identify a certain class of image and then detect and tabulate their appearance in an image or video.

Segmentation: the process of partitioning a digital image into multiple image segments (regions).

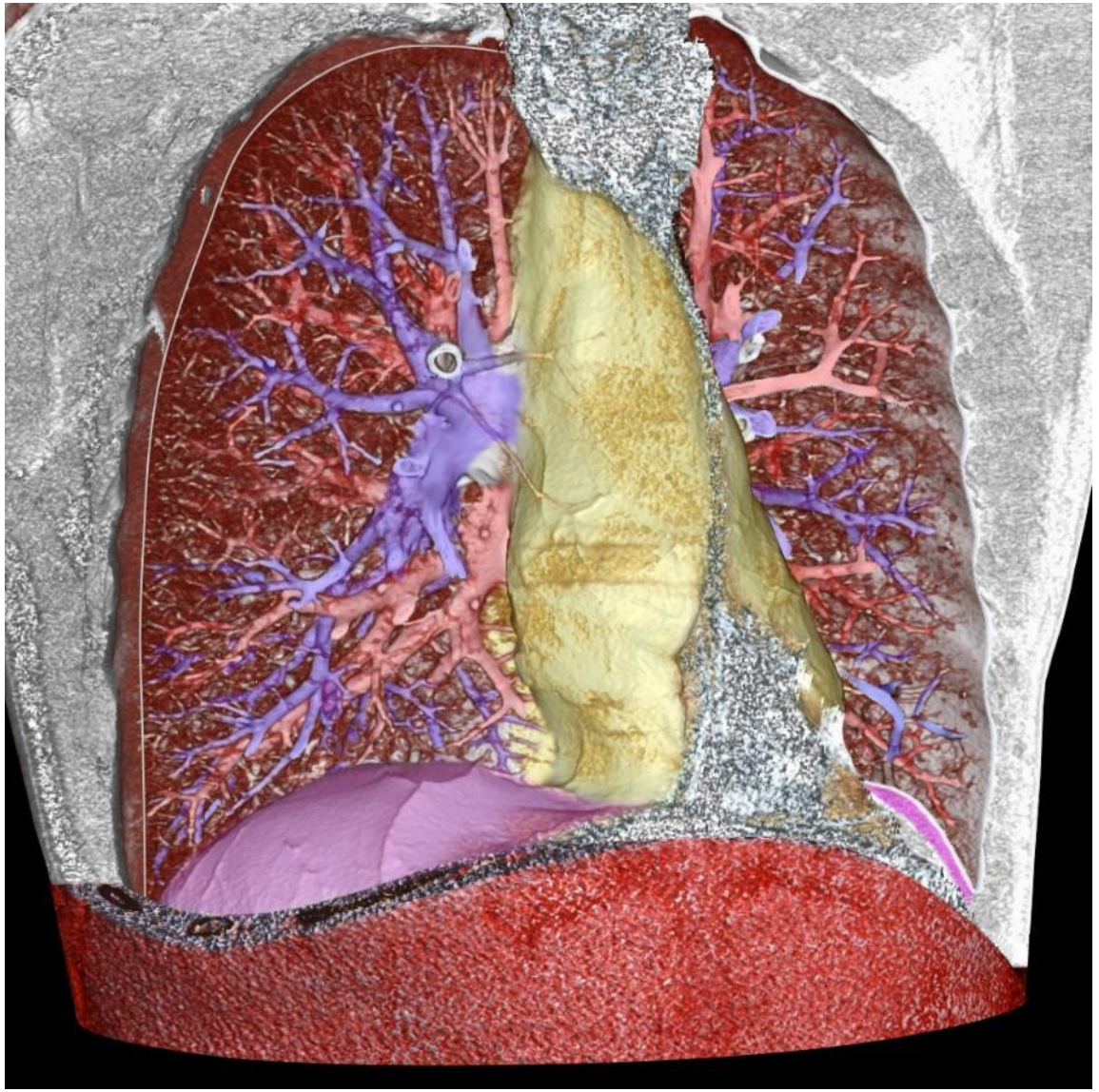


Computer vision

Classification: predict the class(es) an image or video belongs to.

Object detection: identify a certain class of image and then detect and tabulate their appearance in an image or video.

Segmentation: the process of partitioning a digital image into multiple image segments (regions).



pulmonary arteries: blue; pulmonary veins (and also the abdominal wall): red; the mediastinum: yellow; the diaphragm: violet

Computer vision

Classification: predict the class(es) an image or video belongs to.

Object detection: identify a certain class of image and then detect and tabulate their appearance in an image or video.

Segmentation: the process of partitioning a digital image into multiple image segments (regions).

Object tracking: follow the movement of an object in a video.

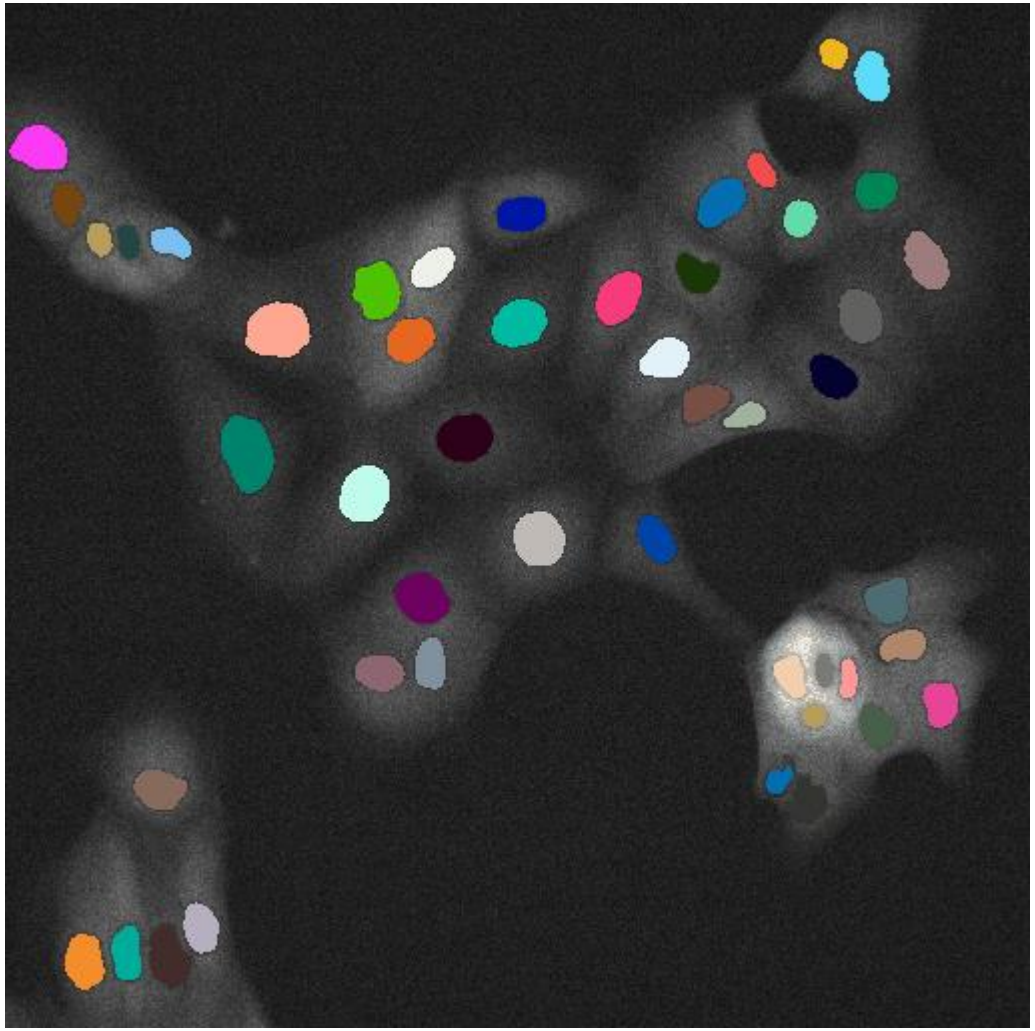
Computer vision

Classification: predict the class(es) an image or video belongs to.

Object detection: identify a certain class of image and then detect and tabulate their appearance in an image or video.

Segmentation: the process of partitioning a digital image into multiple image segments (regions).

Object tracking: follow the movement of an object in a video.



https://imaging.cs.msu.ru/en/research/cell-tracking

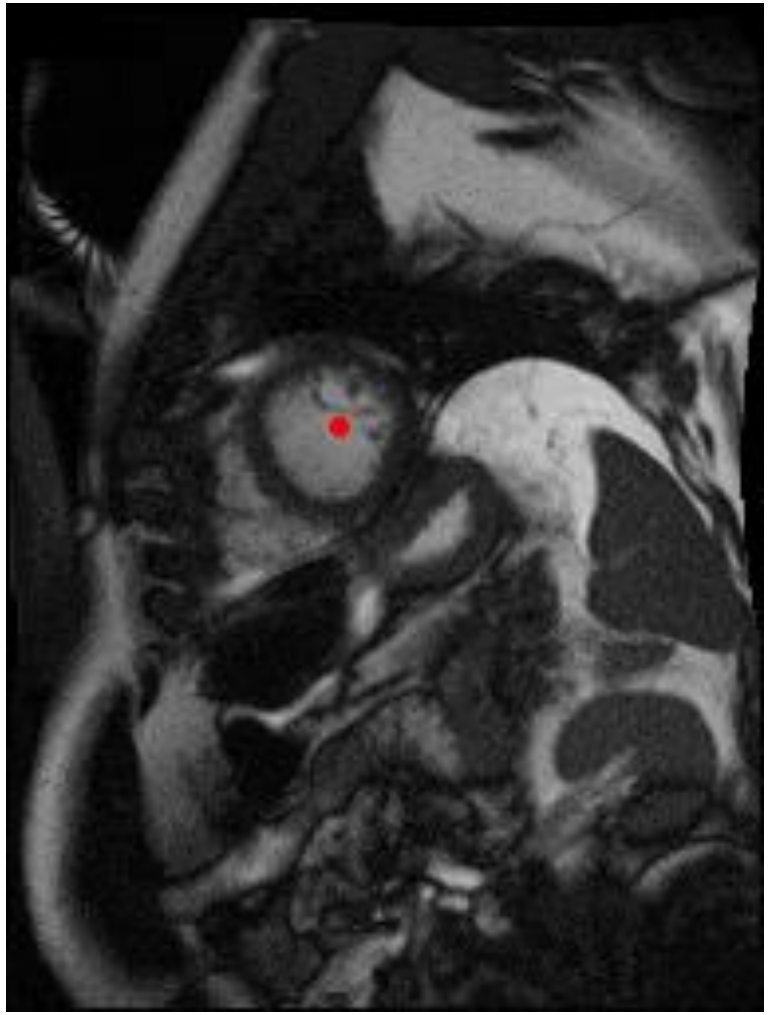
Computer vision

Classification: predict the class(es) an image or video belongs to.

Object detection: identify a certain class of image and then detect and tabulate their appearance in an image or video.

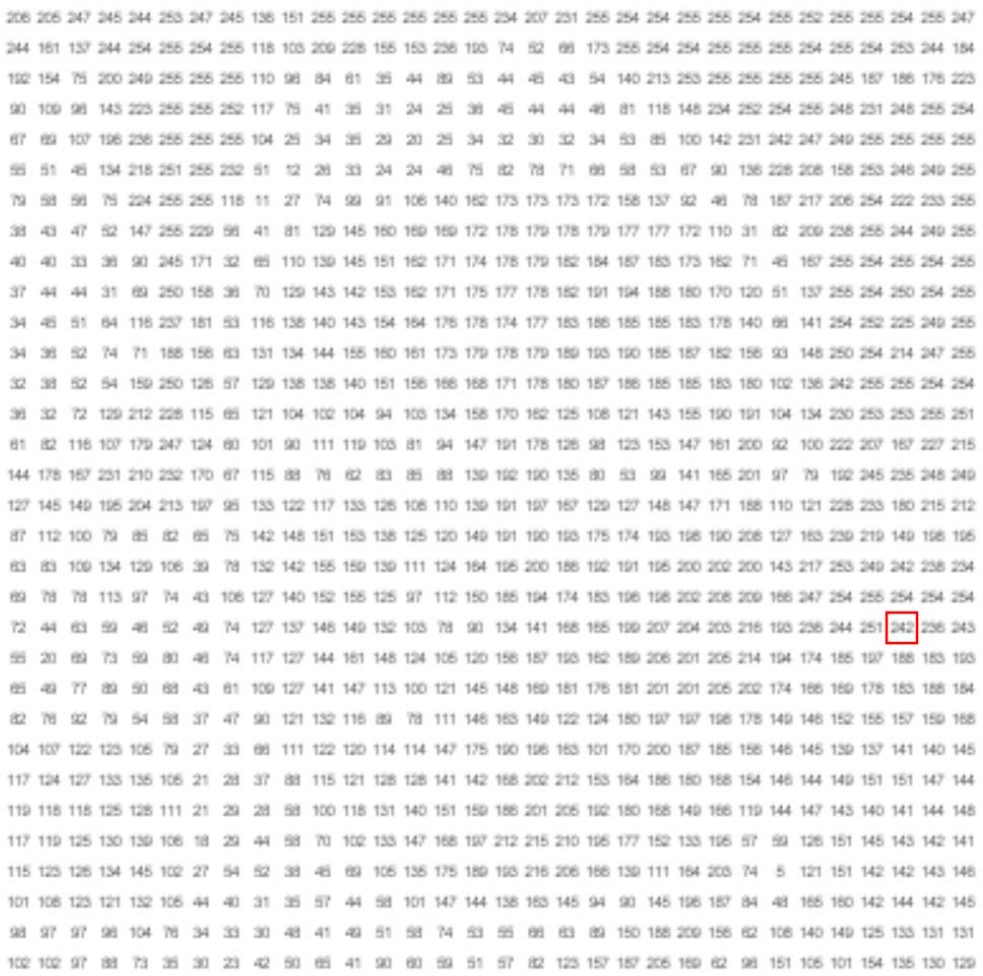
Segmentation: the process of partitioning a digital image into multiple image segments (regions).

Object tracking: follow the movement of an object in a video.



kaggle.com/second-annual-data-science-bowl (2015)

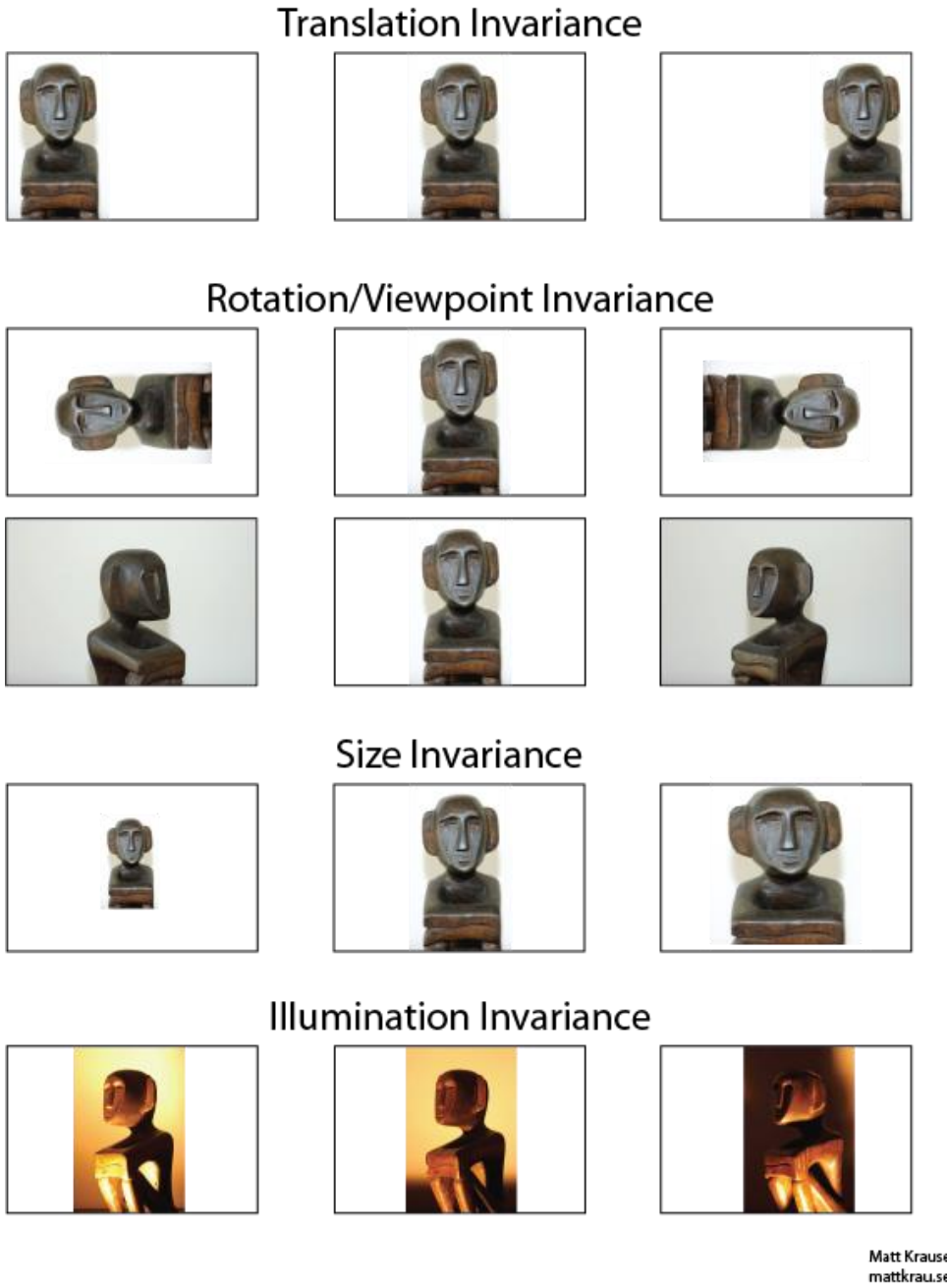
What computers see



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

Invariances

Invariance means that we can recognize an object as an object, even when its appearance varies in some way.



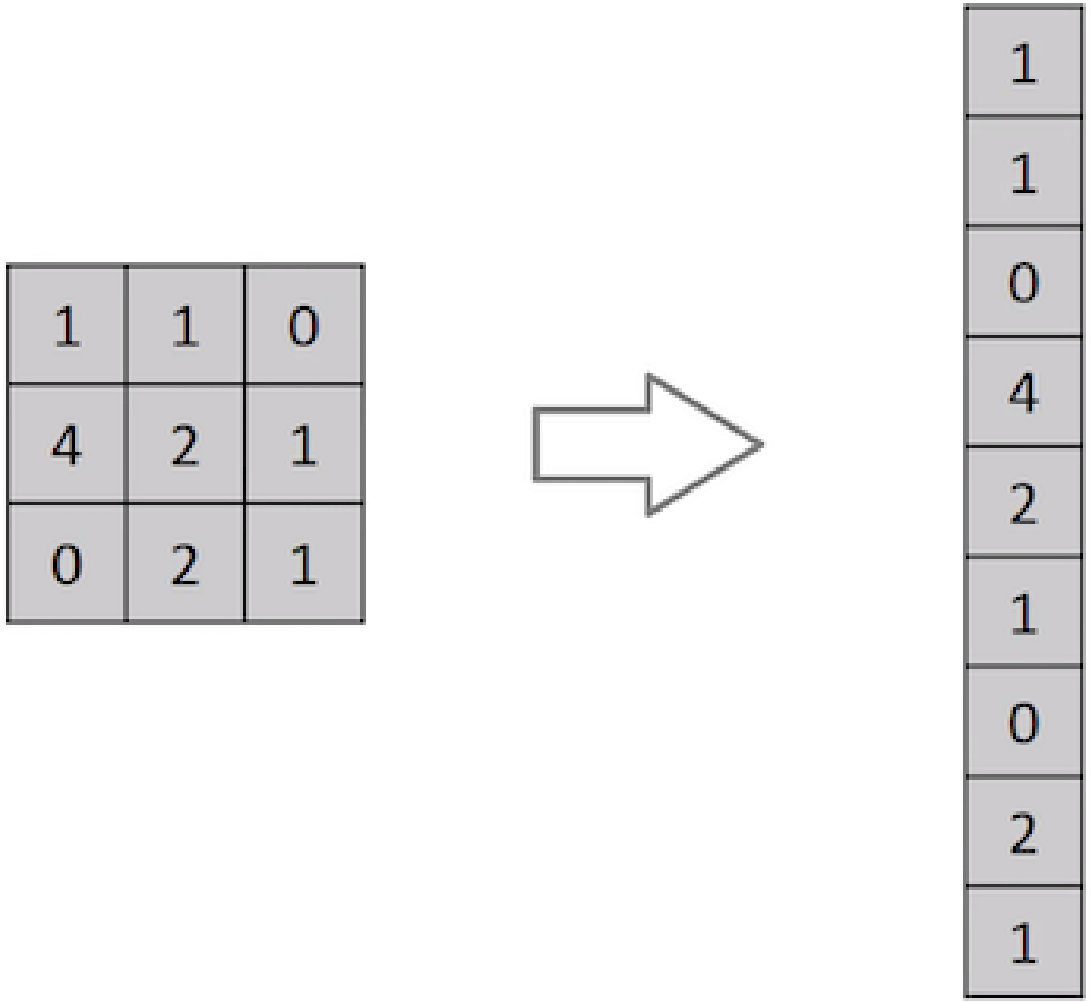
Flatten

We could just represent each pixel as a feature.

This would mean that we **flatten** the 2D matrix to a 1D feature vector.

The learning algorithm then needs to learn the 2D **spatial correlations** from the 1D representation.

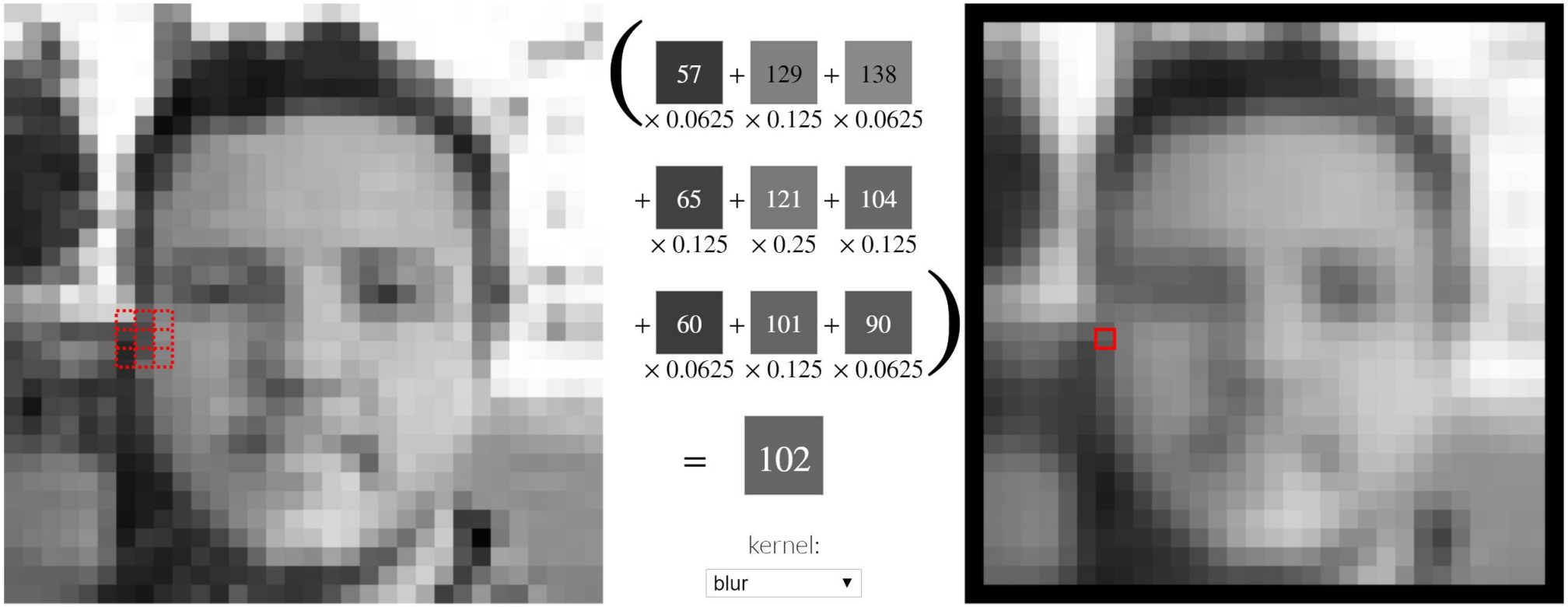
And what with the invariances?



Convolutional filter

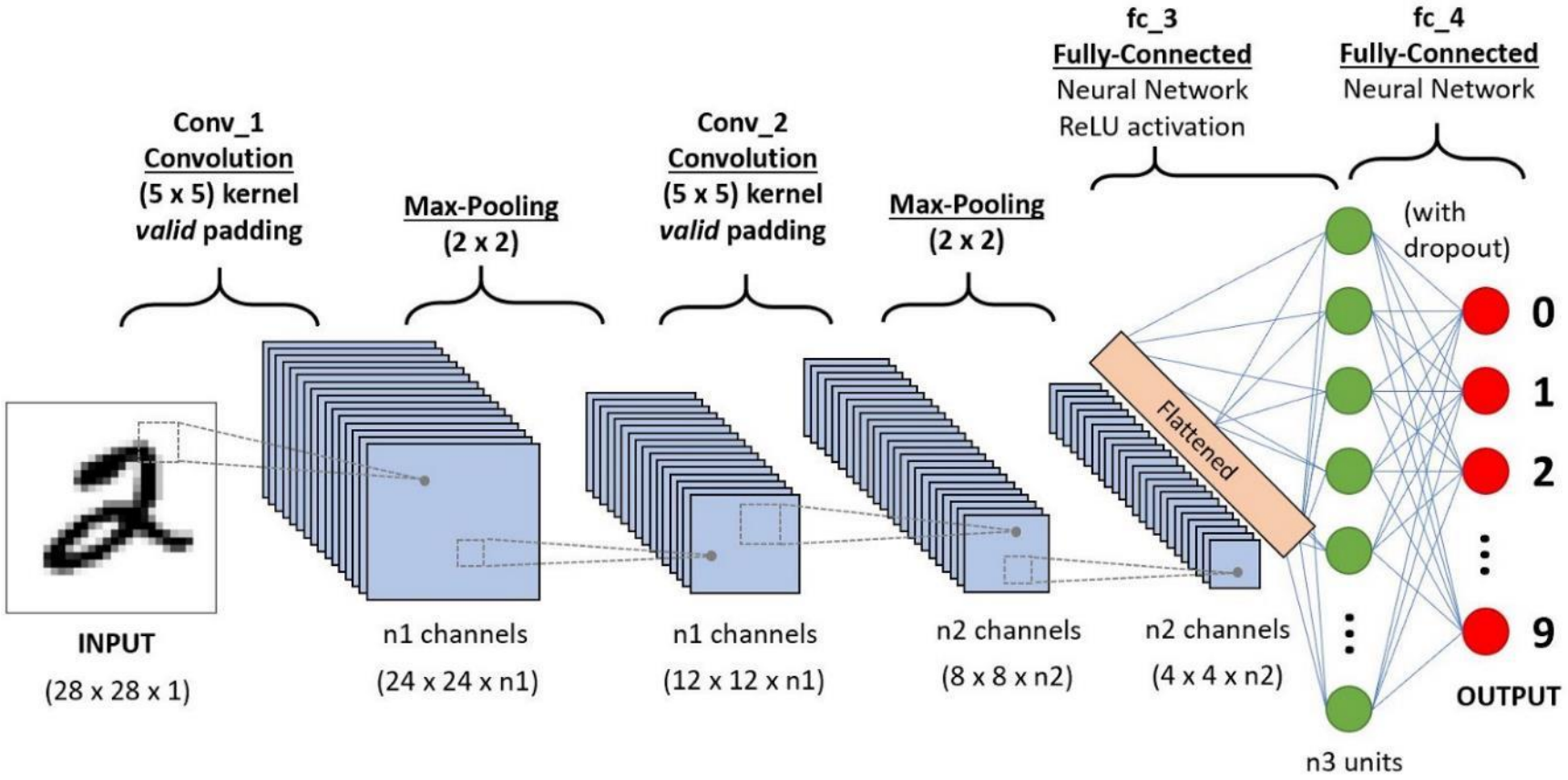
Convolutional filters are small matrices that “slide” over an image.

They provides a measure for how close a patch or a region in the image resembles a **feature**.



Convolutional neural networks (CNN)

A **convolutional neural network (CNN)** module stacks modules that consist of
- **feature maps** (or **channels**) that each learn a relevant convolutional filter with specific dimensions, and
- a **pooling layer** that allows for the location invariance of the learned features.

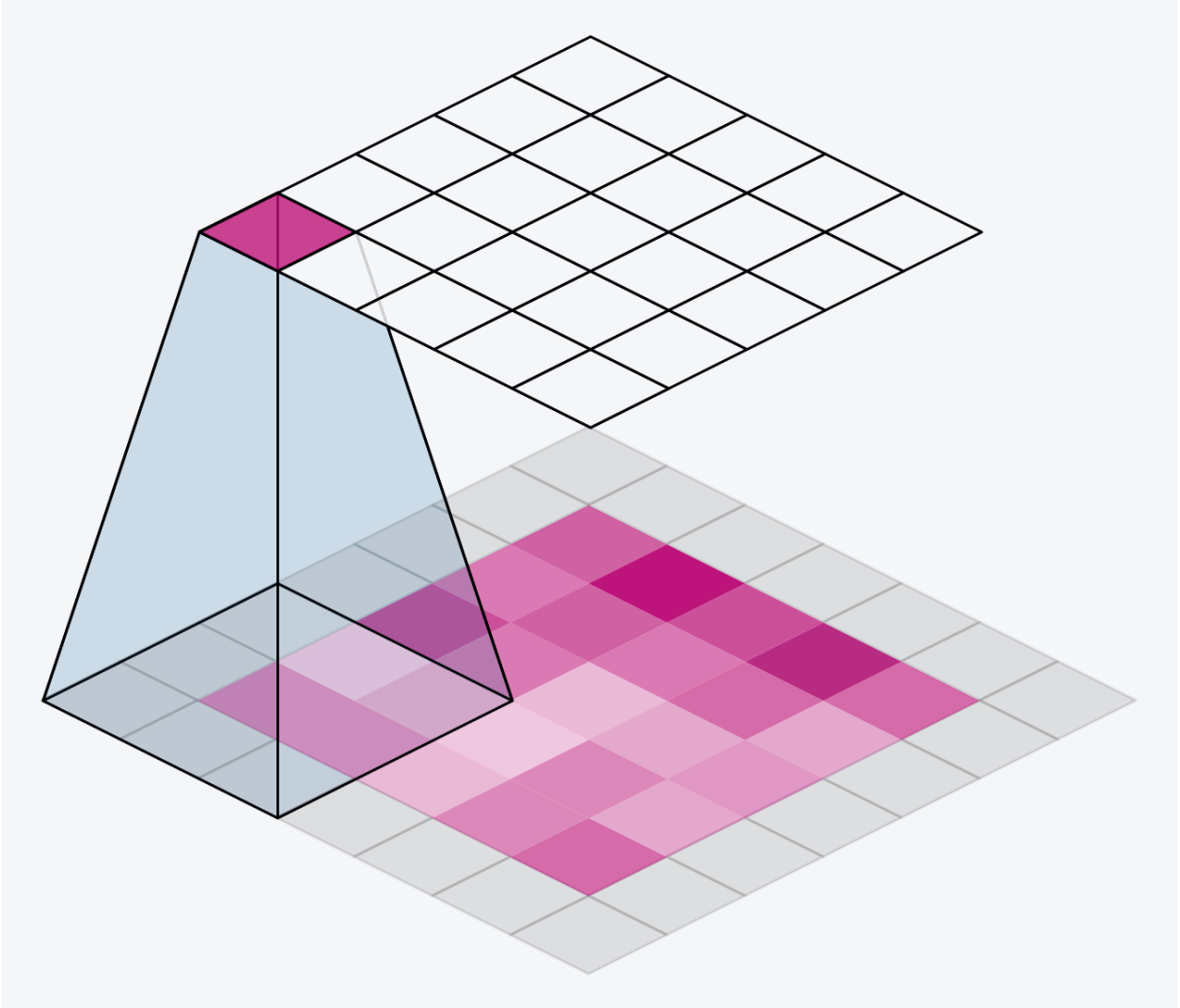


Feature map

Each neuron looks at a different (overlapping) part in the image and performs a convolution.

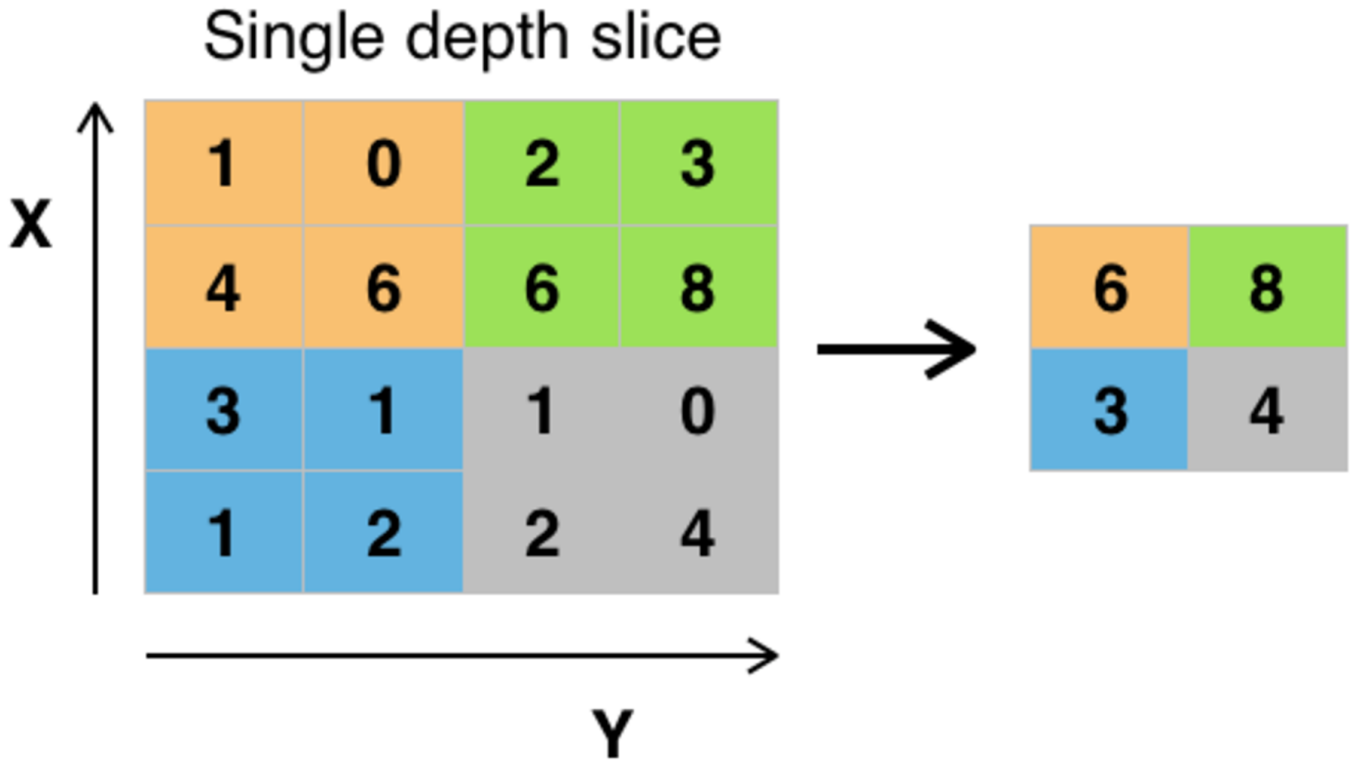
The convolution is defined by the modelparameters that are learned from the data.

So, in each feature maps, the modelparameters for each neuron have the same value, i.e. they are shared. The only difference is in what part of the image the neuron looks at.



Pooling

Maximum pooling, or max pooling, is a pooling operation that calculates the maximum, or largest, value in each patch of each feature map.



Example of Maxpool with a 2x2 filter and a stride of 2

Invariances: data augmentation

CNNs are

translation invariant,

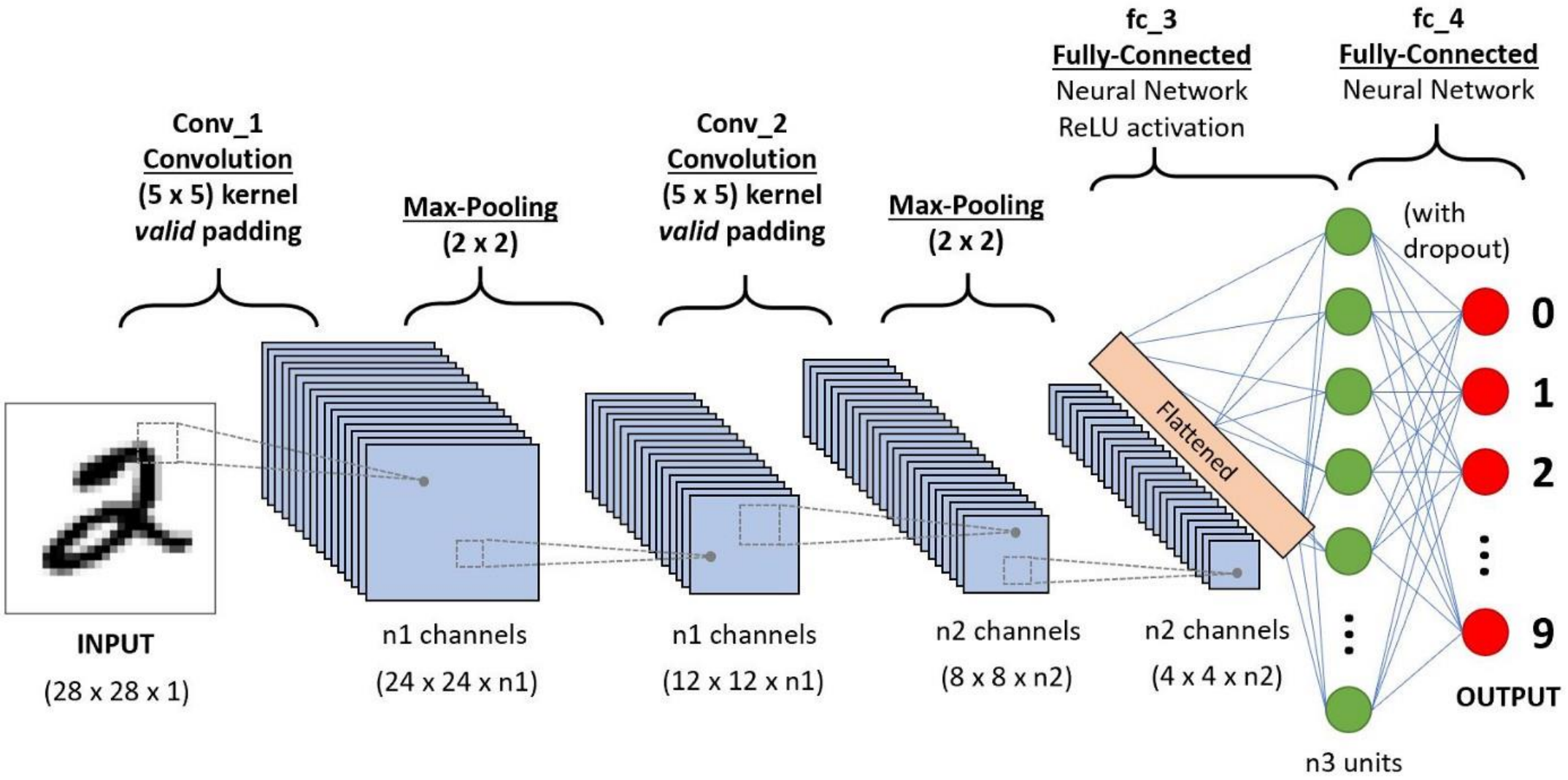
scale invariant to some degree (learned from data, not in filters),

not rotation invariant (learned from data),

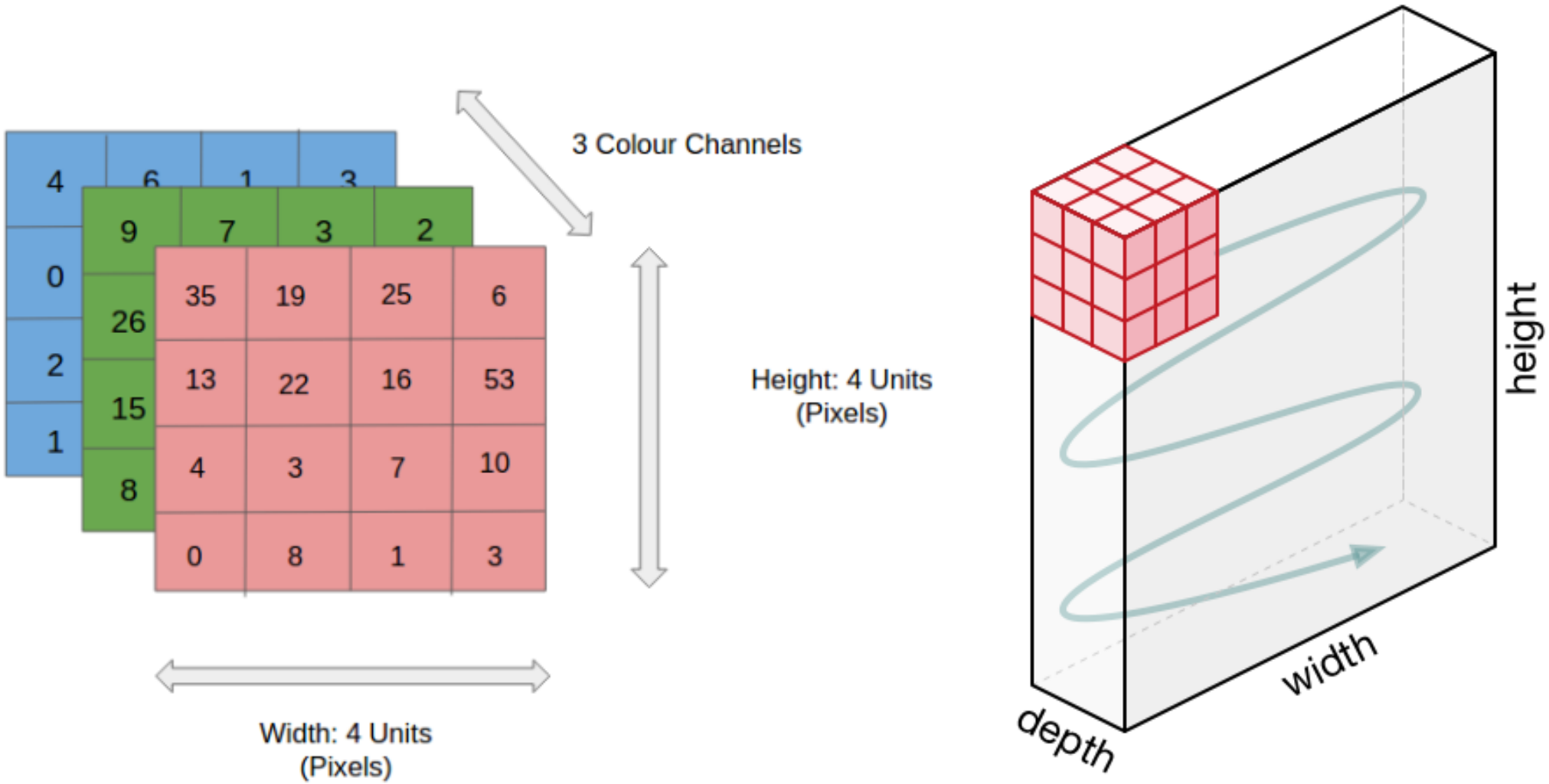
not illumination invariant (learned from data).

We can generate artificial data, by rotating, scaling, illuminating training images. A process called **data augmentation**. This is a form of model **regularization**.

CNN


Conv_1
Convolution
(5 x 5) kernel
valid padding
Max-Pooling
(2 x 2)
Conv_2
Convolution
(5 x 5) kernel
valid padding
Max-Pooling
(2 x 2)
fc_3
Fully-Connected
Neural Network
ReLU activation
fc_4
Fully-Connected
Neural Network
(with dropout)
Flattened
INPUT
(28 x 28 x 1)
n1 channels
(24 x 24 x n1)
n1 channels
(12 x 12 x n1)
n2 channels
(8 x 8 x n2)
n2 channels
(4 x 4 x n2)
n3 units
0
1
2
9
OUTPUT

CNN: channels



Neural network

mnist_pytorch.ipynb

ImageNet

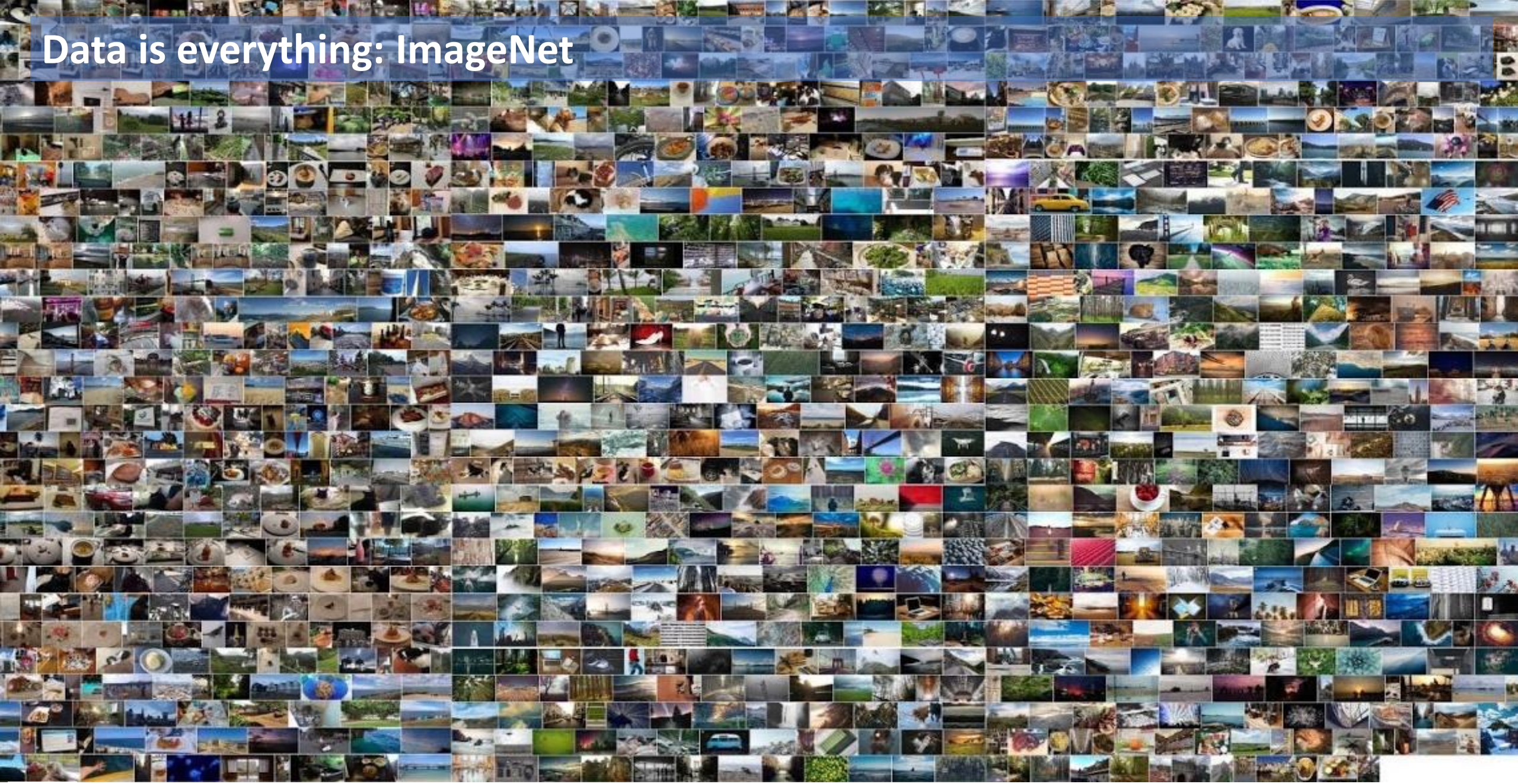
ImageNet is a large-scale visual database designed for use in visual object recognition research, containing millions of labeled images across thousands of categories.

It was introduced by Fei-Fei Li and her team in 2009 and played a crucial role in advancing deep learning, particularly in convolutional neural networks (CNNs).

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC), held annually from 2010 to 2017, drove major breakthroughs in AI and computer vision, with models like AlexNet, VGG, ResNet, and EfficientNet achieving state-of-the-art performance.

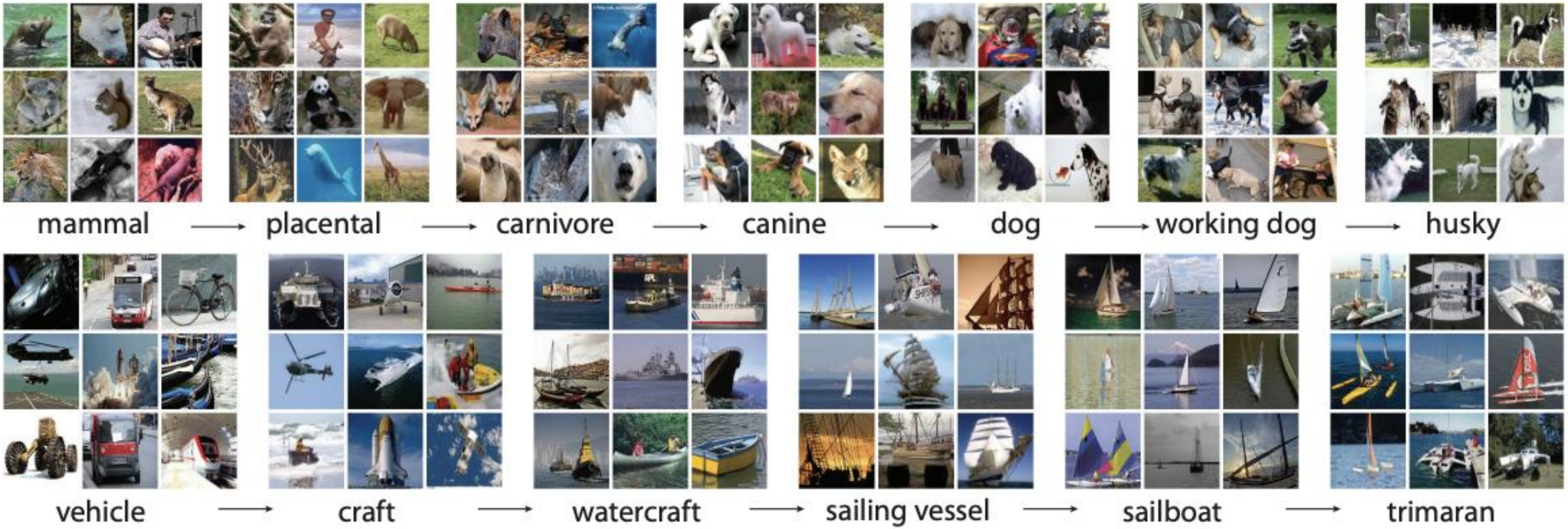
Despite its influence, ImageNet has limitations, including dataset bias, ethical concerns related to labeling, and challenges in real-world generalization beyond the controlled dataset.

Data is everything: ImageNet

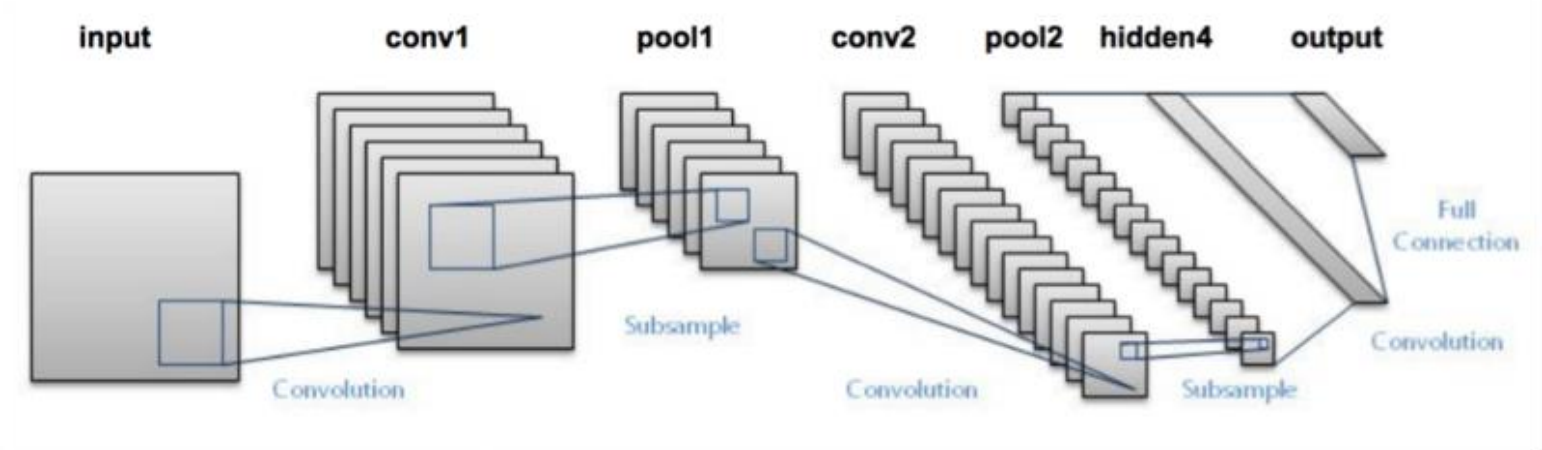


J. Deng et al., "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition

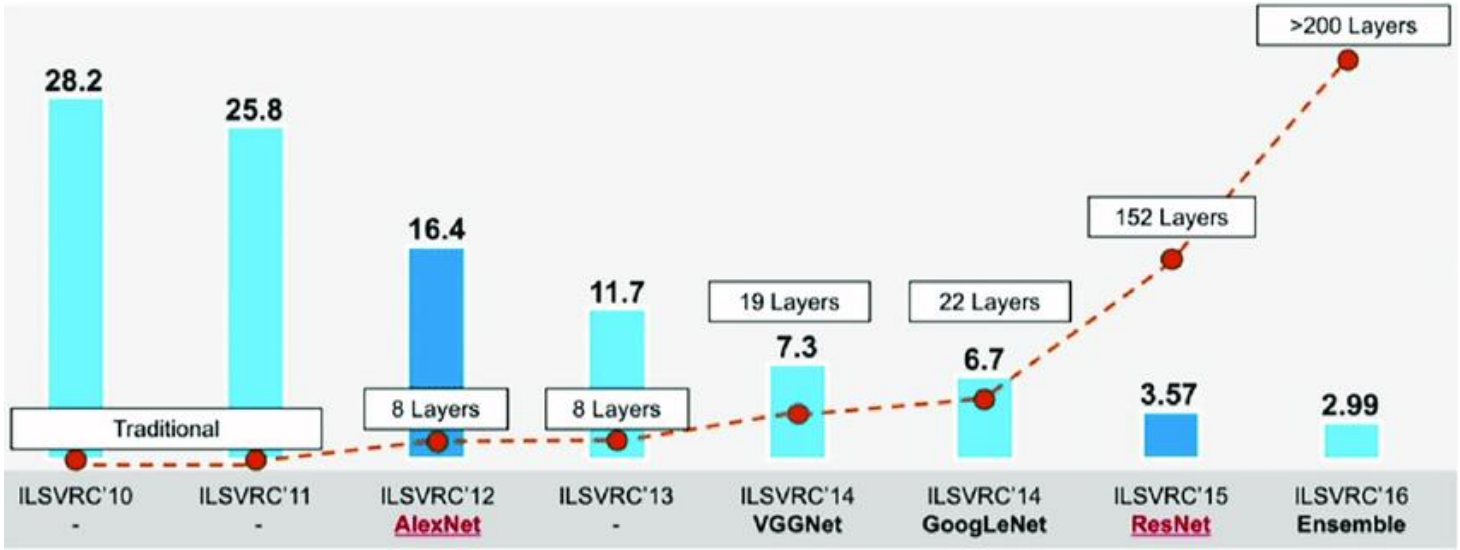
Data is everything: ImageNet



ImageNet: computer vision is solved

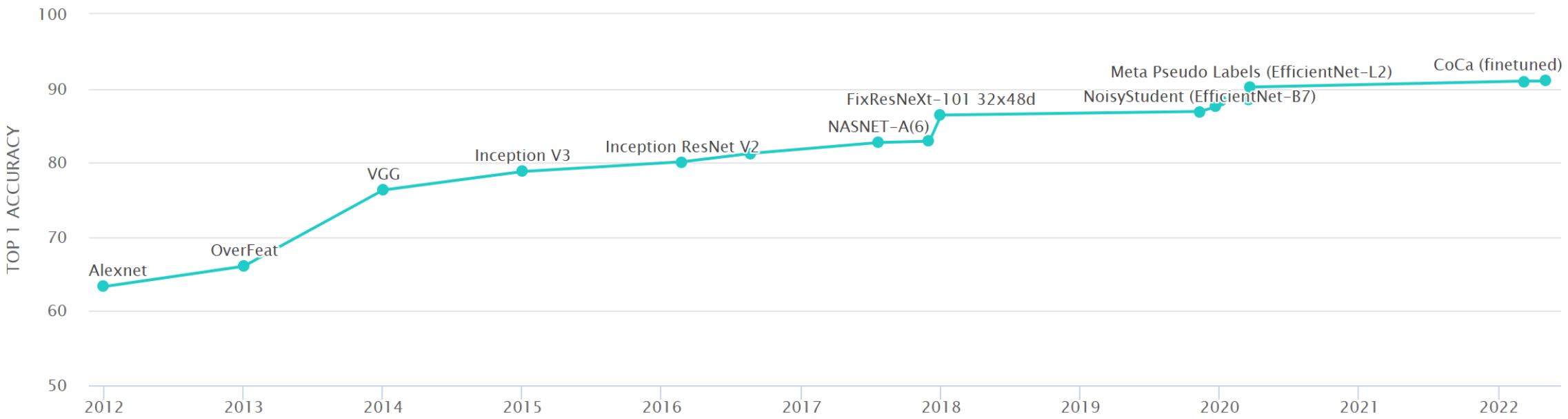


LeNet-5 (1998)



ImageNet Large Scale Visual Recognition Challenge: https://www.image-net.org/challenges/LSVRC/

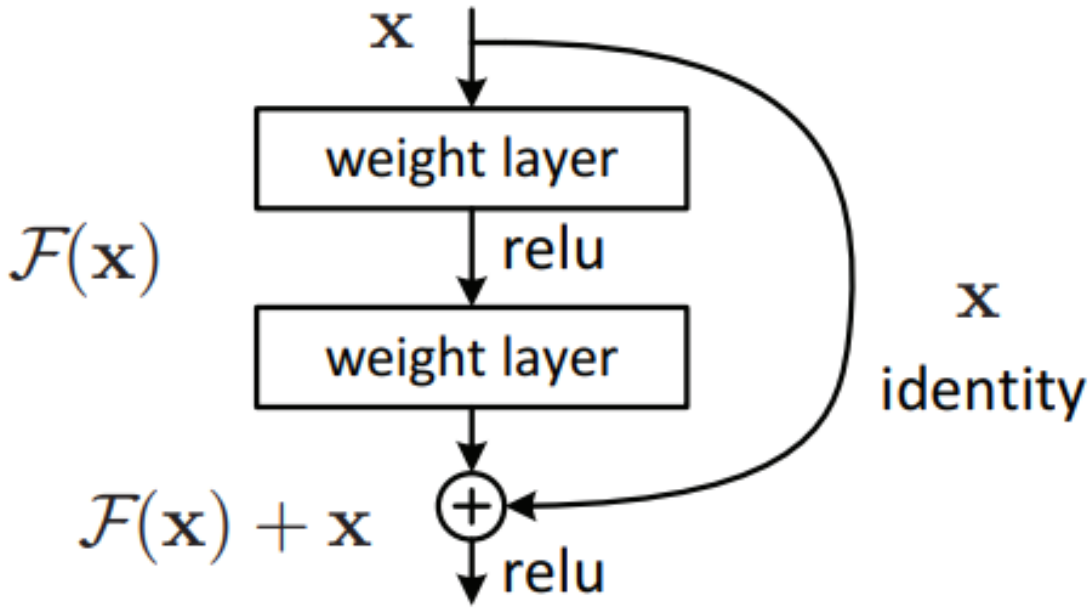
ImageNet: computer vision is solved



https://paperswithcode.com/sota/image-classification-on-imagenet

ResNet (Residual Network)

ResNet introduces **skip connections**, allowing gradients to bypass certain layers, solving the vanishing gradient problem and enabling the training of very deep networks.



He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. CVPR, 770-778.

ResNet (Residual Network)

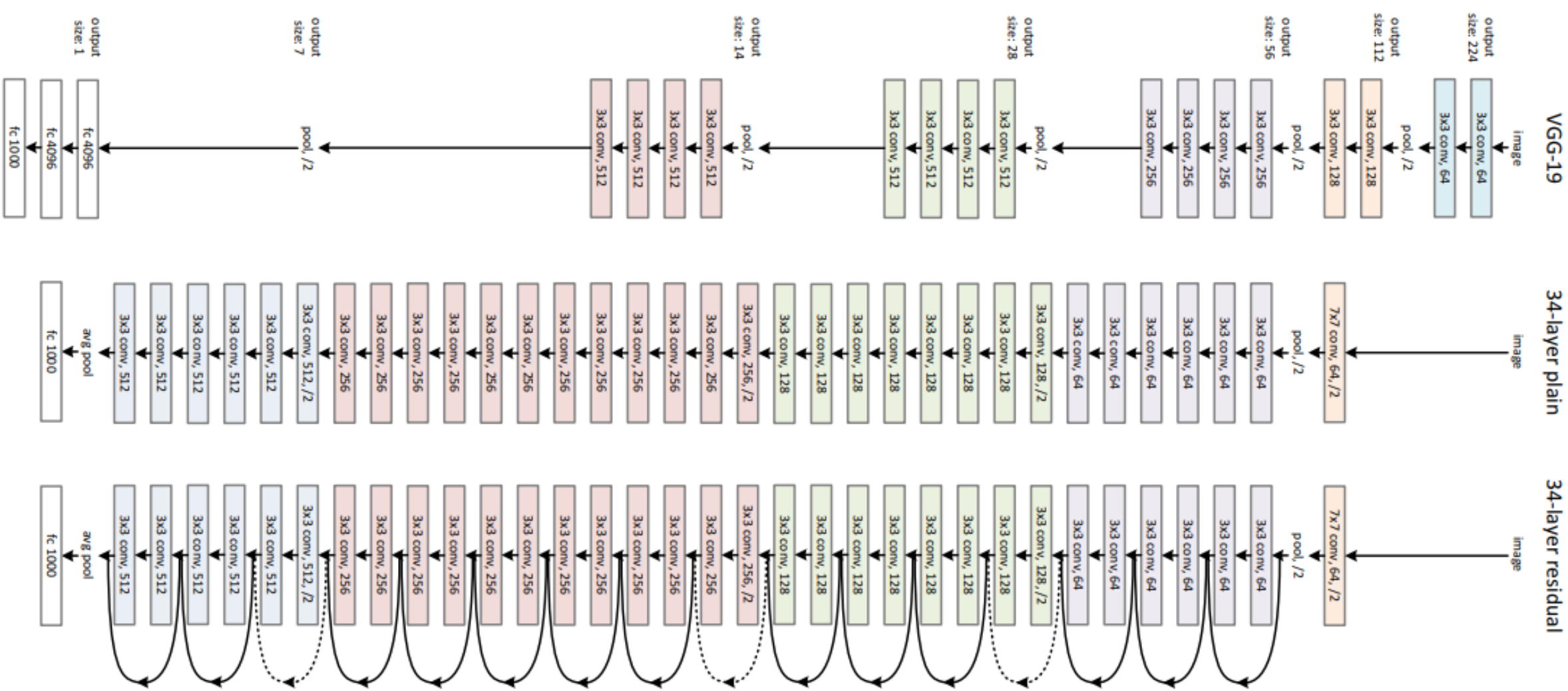
ResNet introduces **skip connections**, allowing gradients to bypass certain layers, solving the vanishing gradient problem and enabling the training of very deep networks.

Instead of learning the full transformation, ResNet learns the **residual (difference) between input and output**, making optimization easier and improving convergence.

ResNet enables training of networks with **hundreds or even thousands of layers**, significantly improving performance in image recognition tasks without degradation.

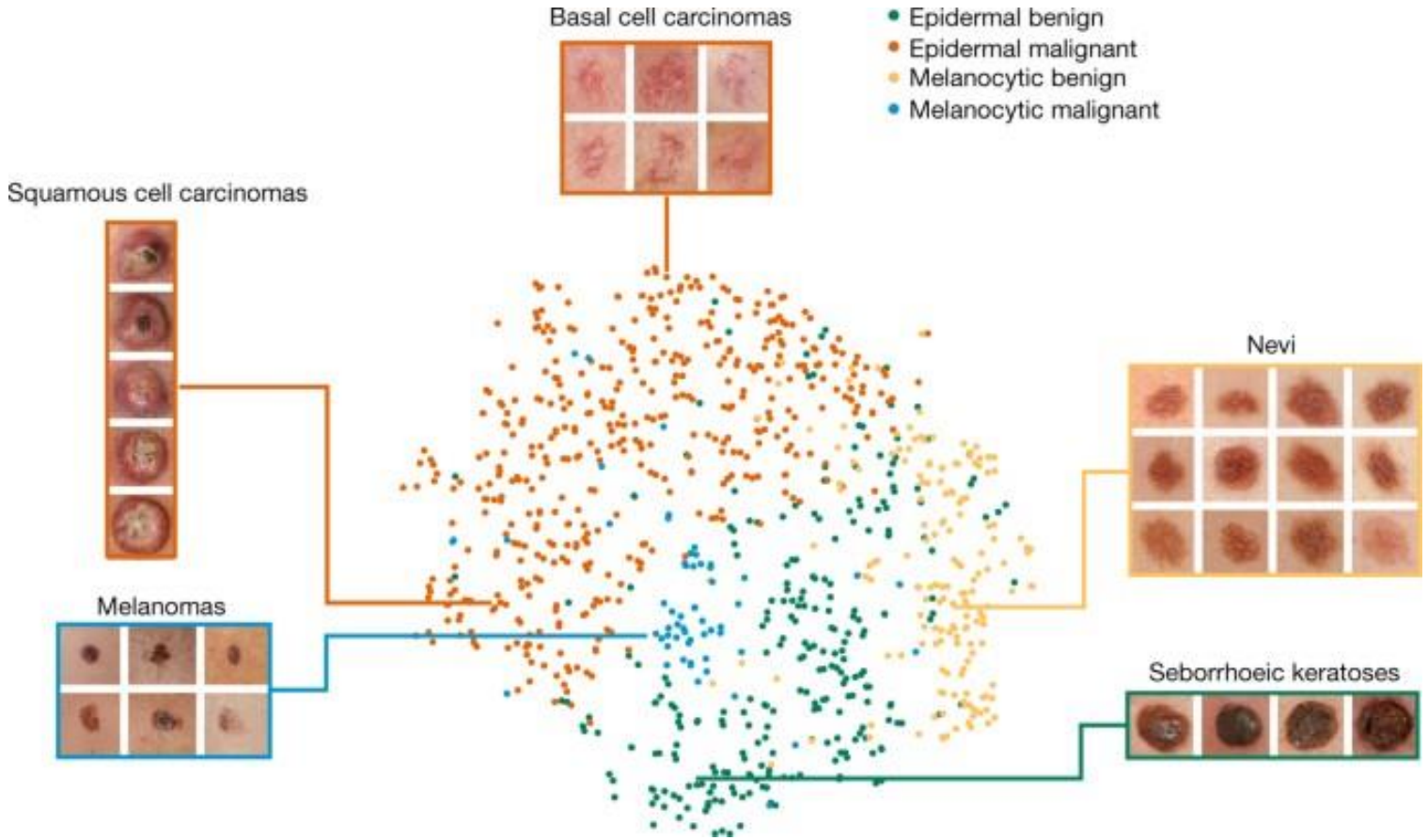
He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. CVPR, 770-778.

ResNet (Residual Network)

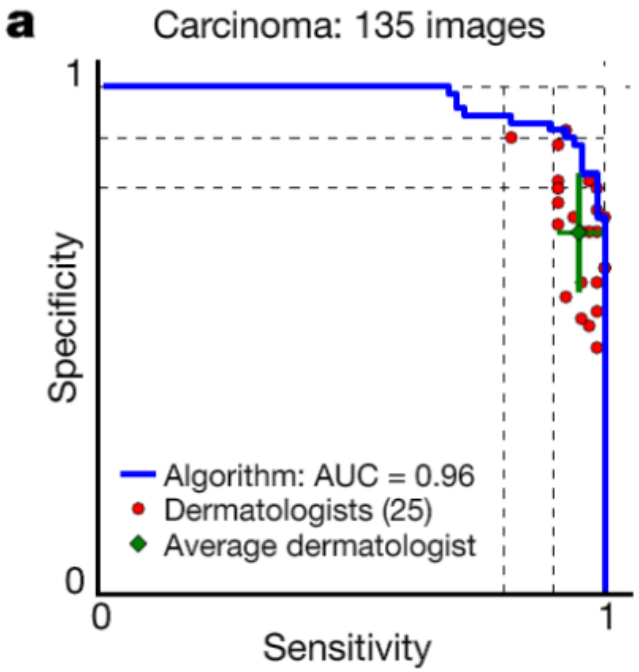


He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. CVPR, 770-778.

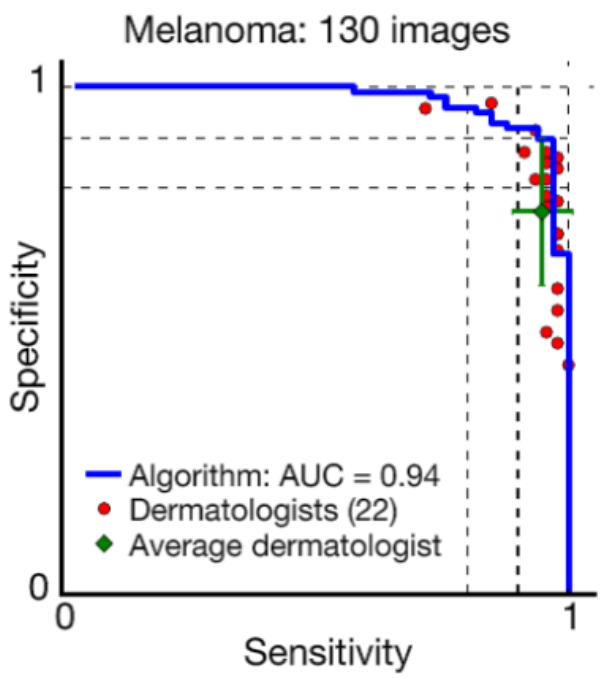
Remember this?



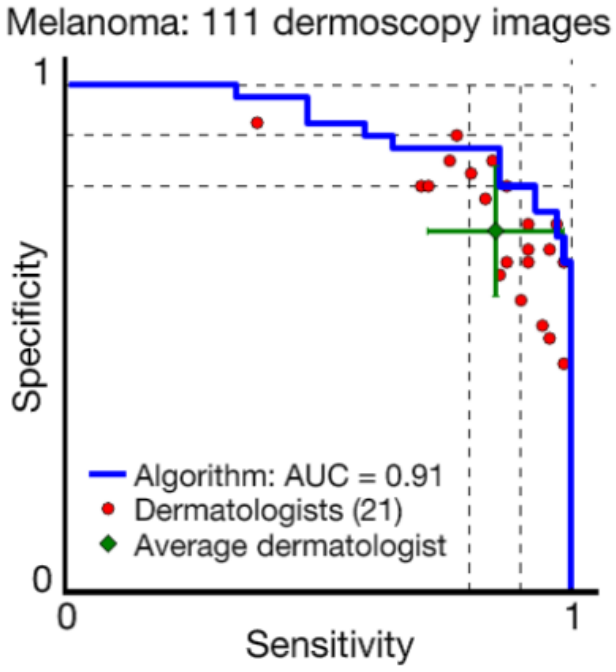
Thrun et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature (2017)



keratinocyte carcinomas: 65
benign seborrheic keratoses: 75



malignant melanomas: 33
benign nevi: 97



malignant melanomas: 71
benign nevi: 40

Thrun et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature (2017)