

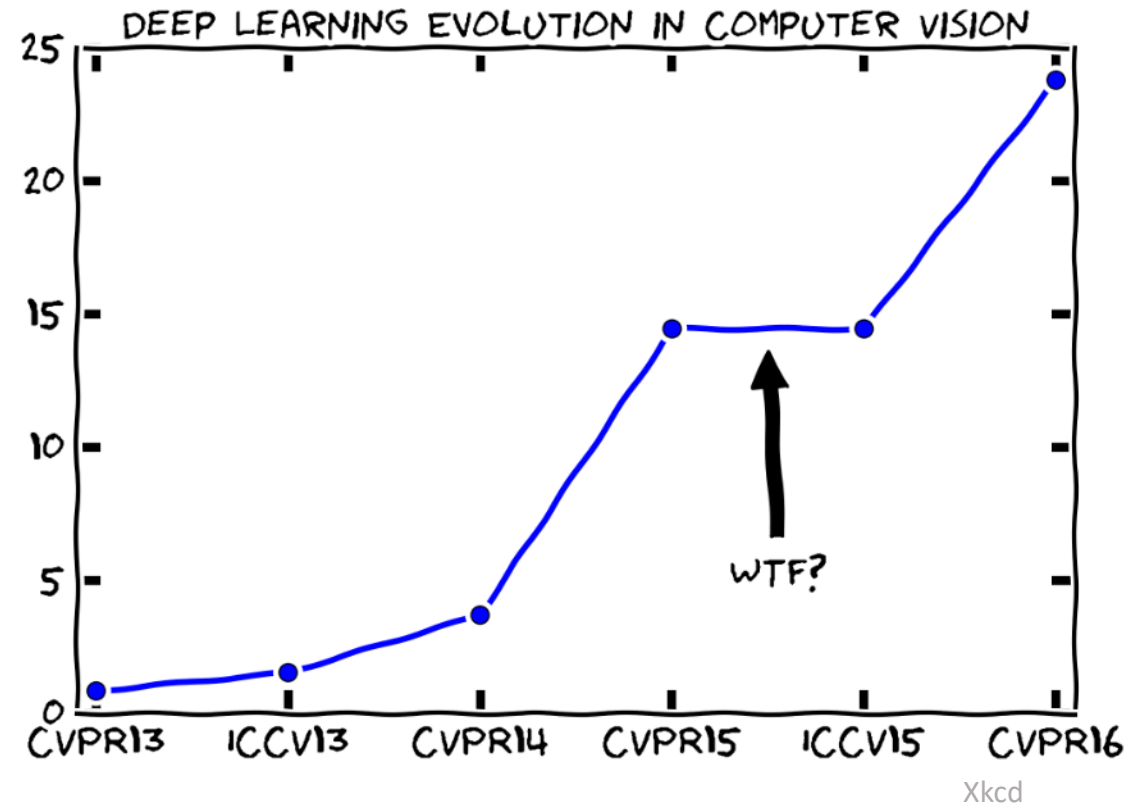
Image data



Xkcd, 2015

IN CS, IT CAN BE HARD TO EXPLAIN
THE DIFFERENCE BETWEEN THE EASY
AND THE VIRTUALLY IMPOSSIBLE.

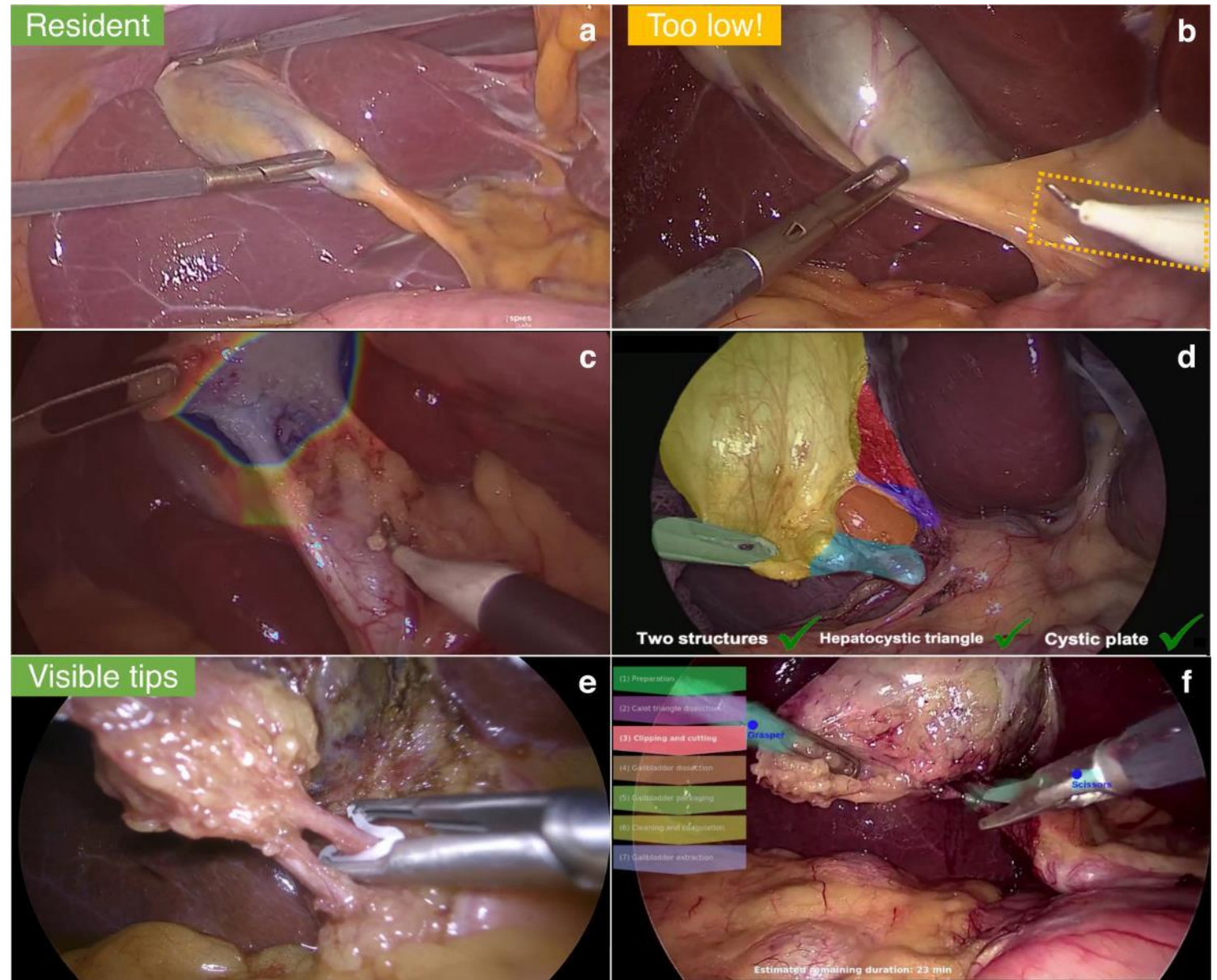
Image data



Computer vision

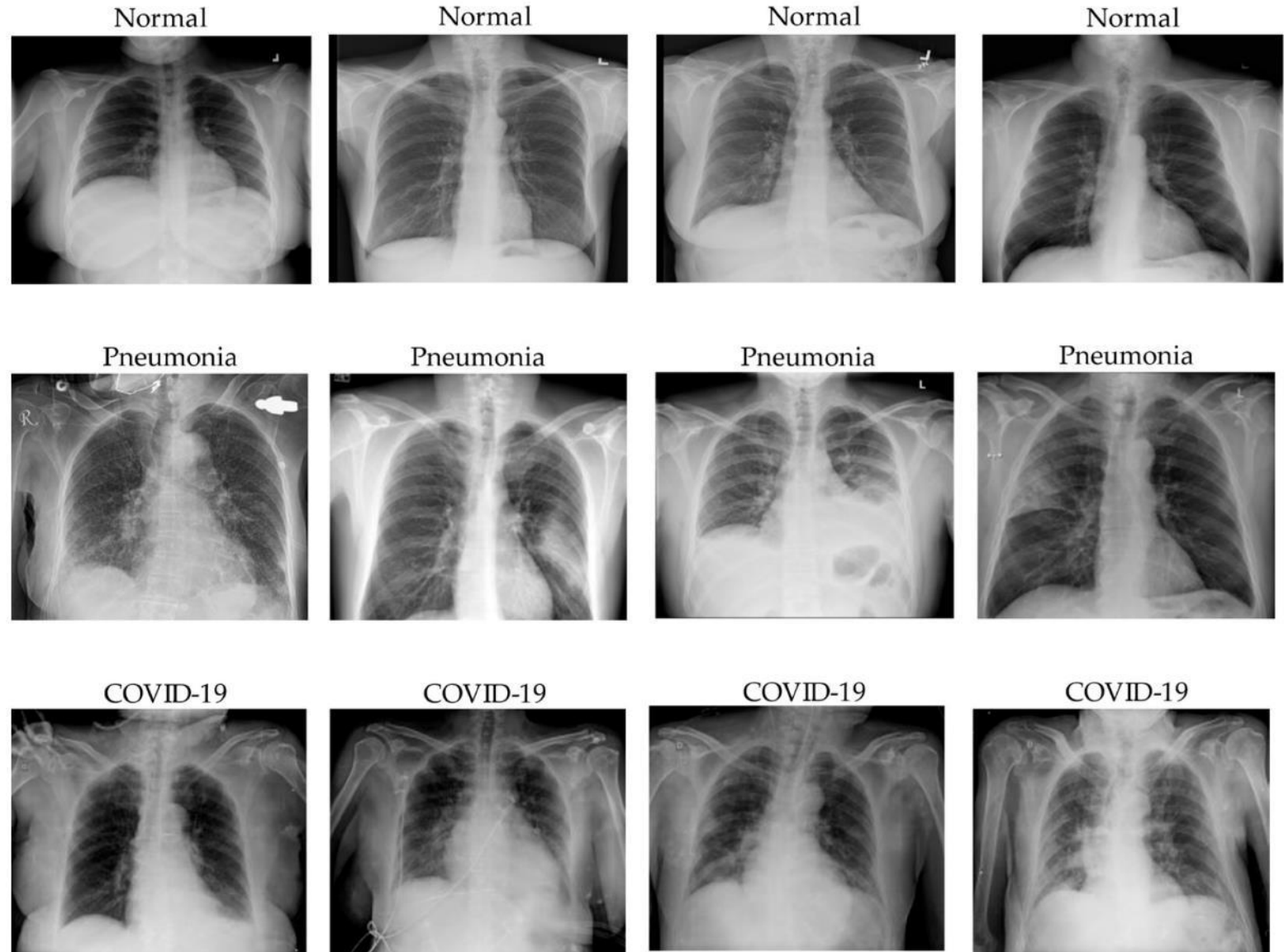
Derive meaningful information from digital images, videos and other visual inputs.

Take actions or make recommendations based on that information.



Computer vision

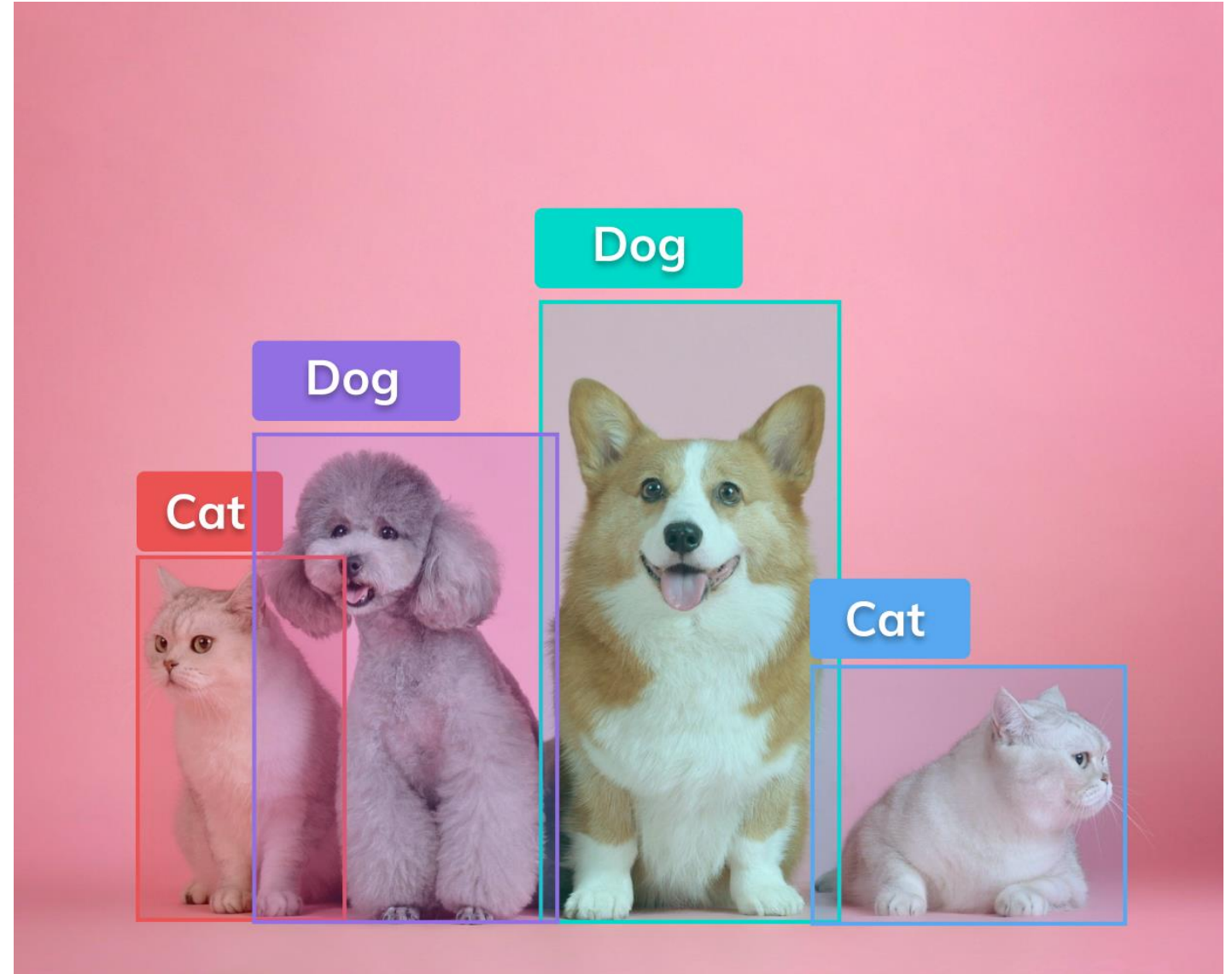
Classification: predict the class(es) an image or video belongs to.



Computer vision

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

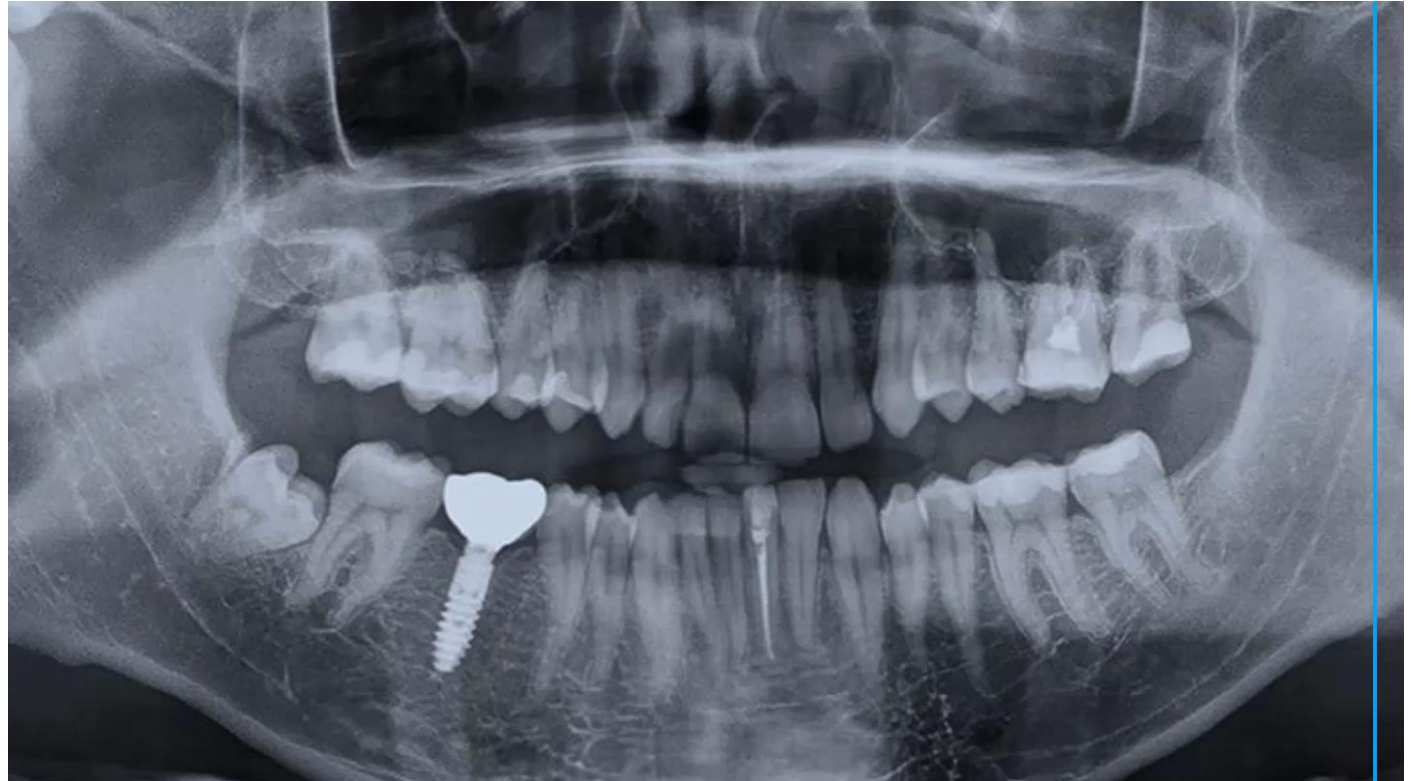
Object detection: identify a certain class of image and then detect and tabulate their appearance in an image or video.



Computer vision

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

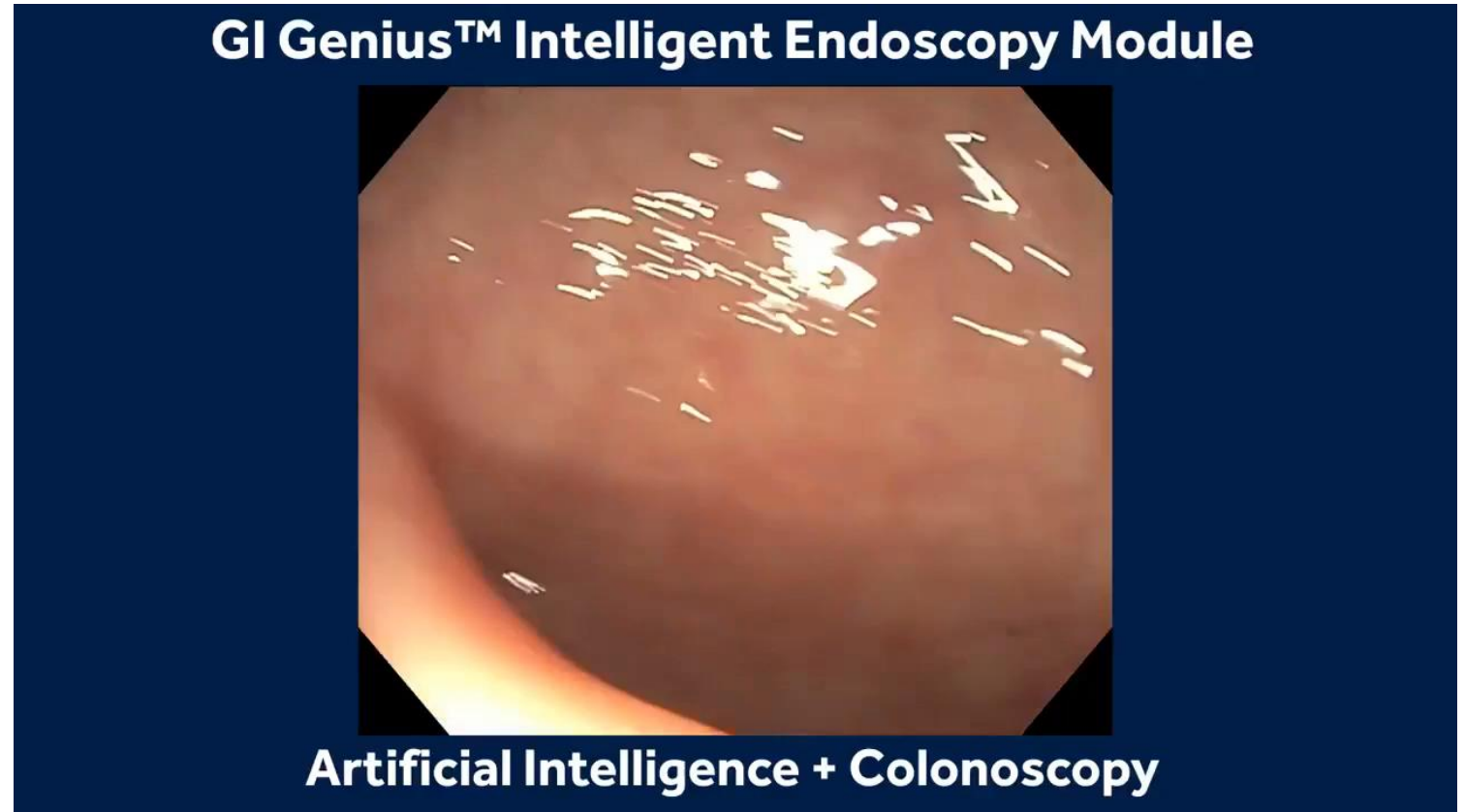
Object detection: identify a certain class of image and then detect and tabulate their appearance in an image or video.



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.

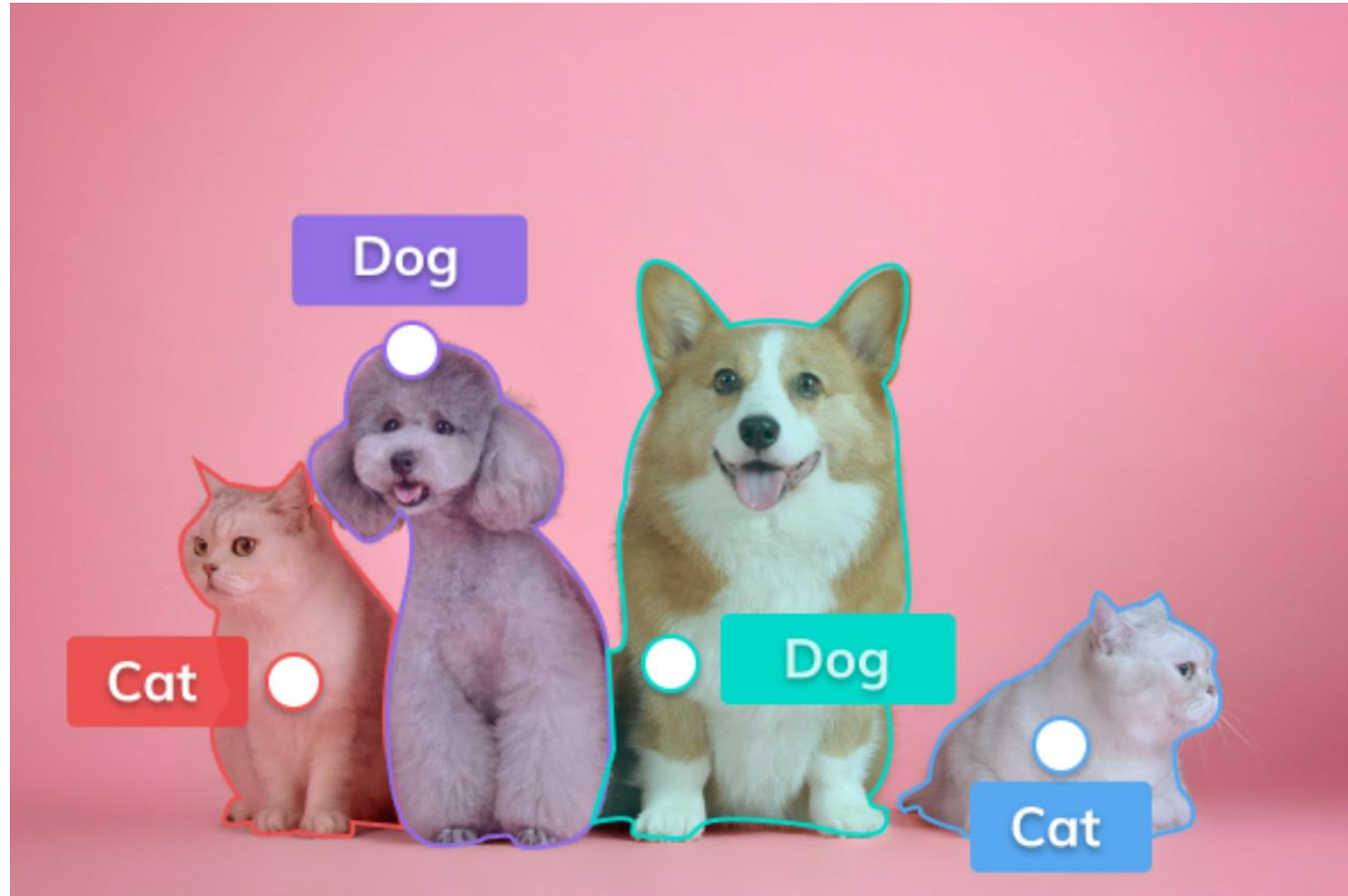


Computer vision

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).

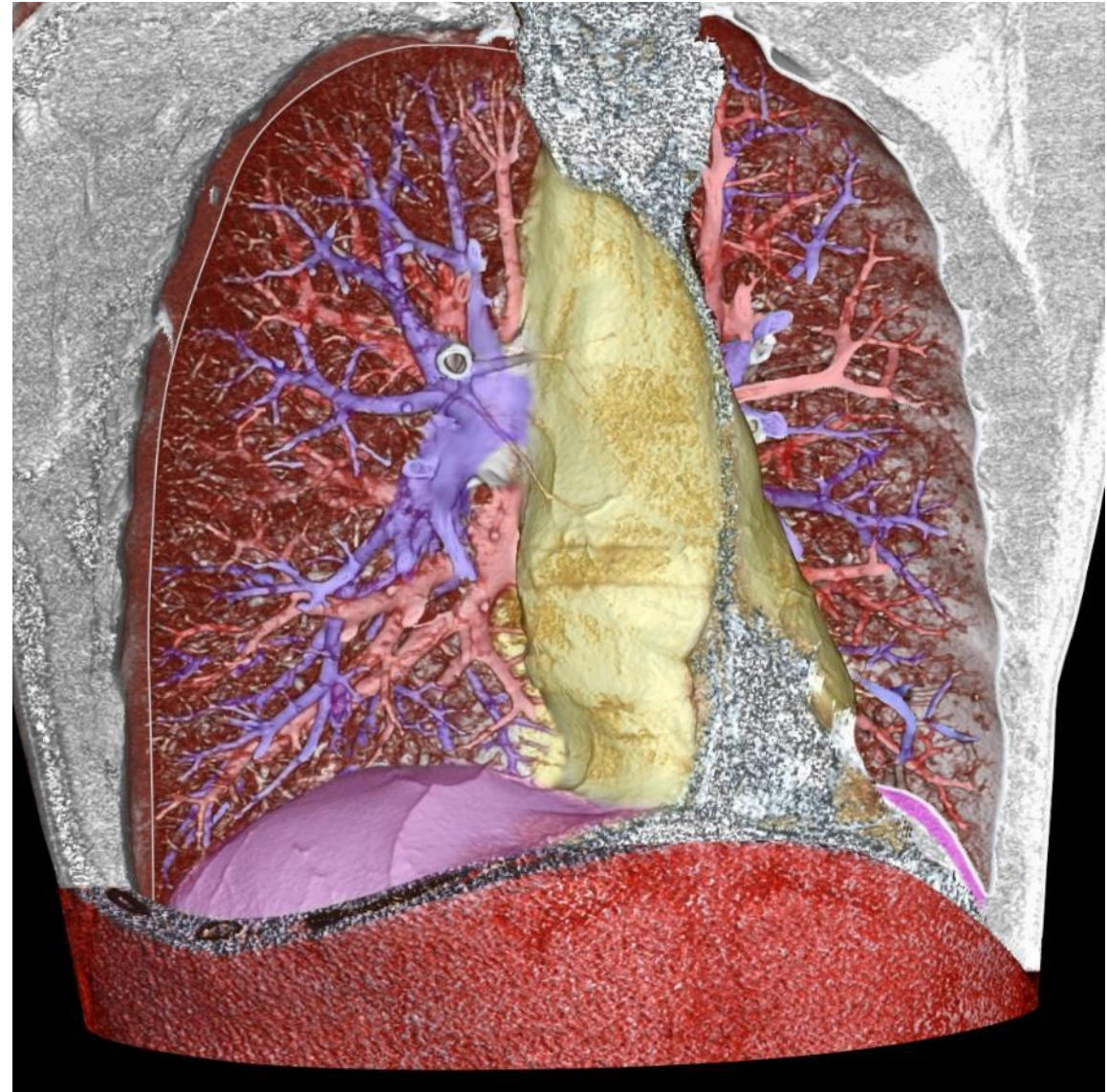


Computer vision

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.



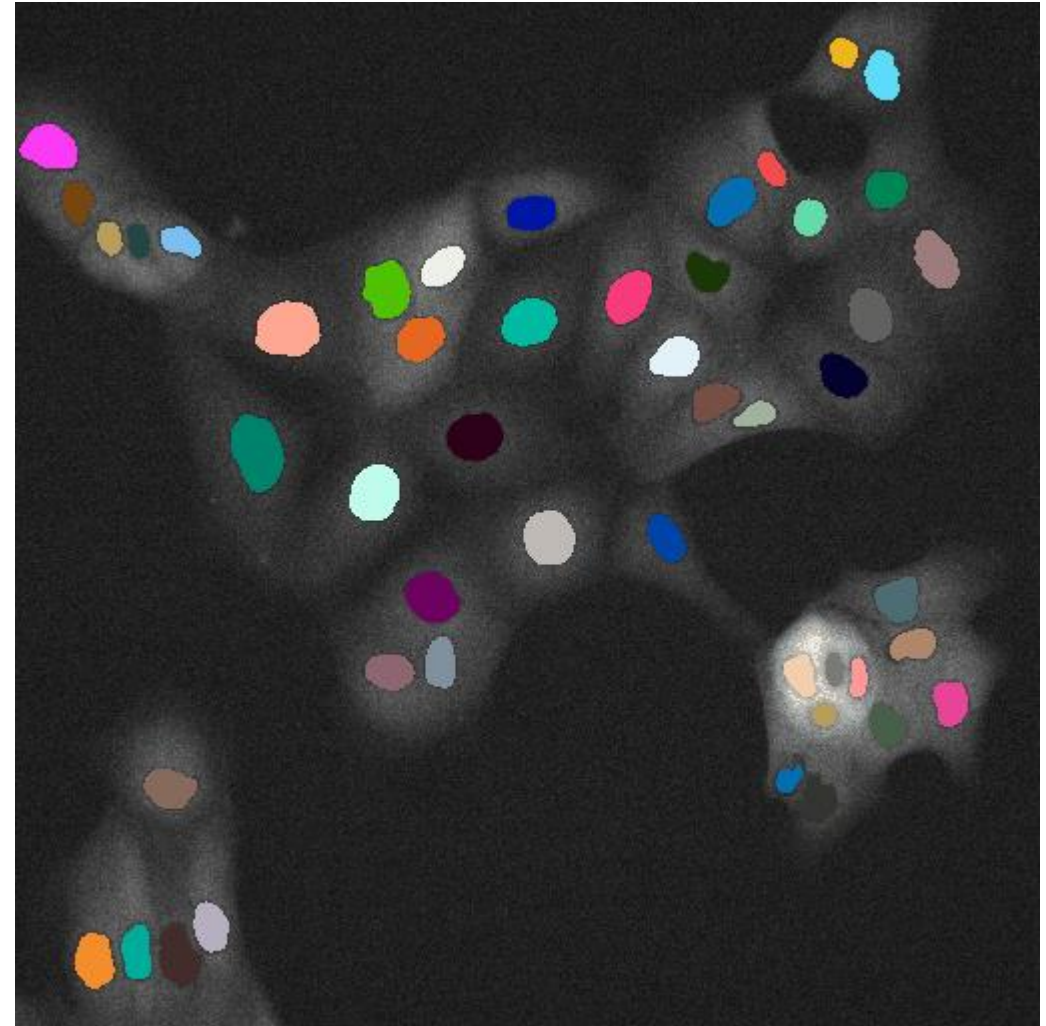
Computer vision

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>

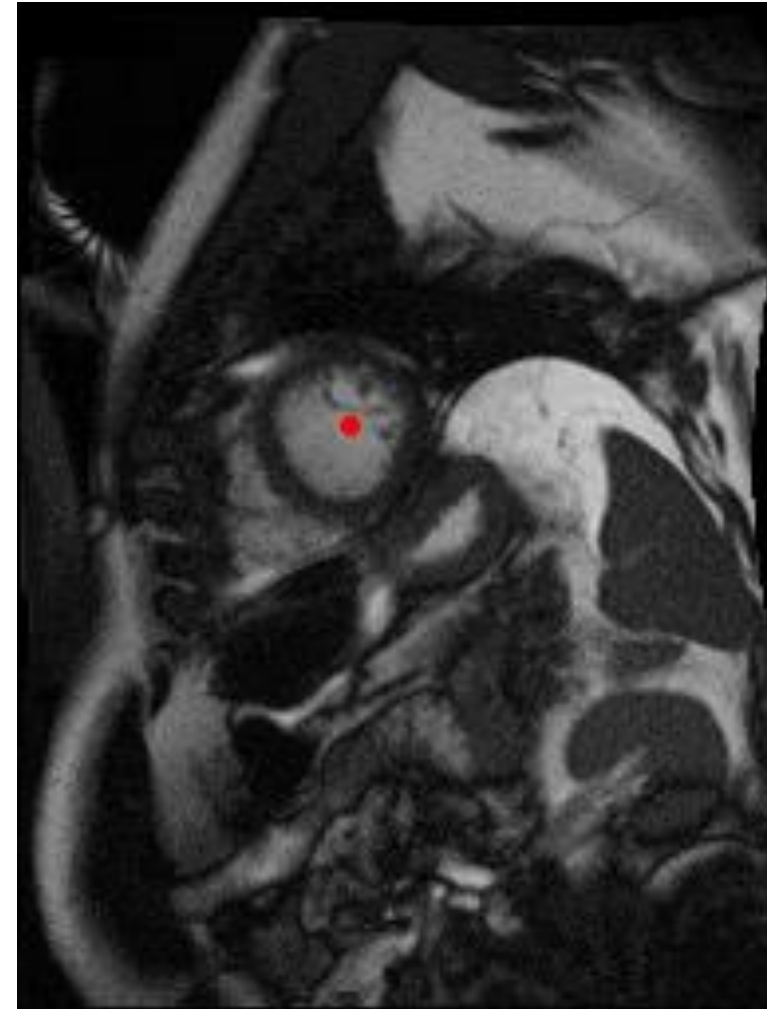
Computer vision

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).

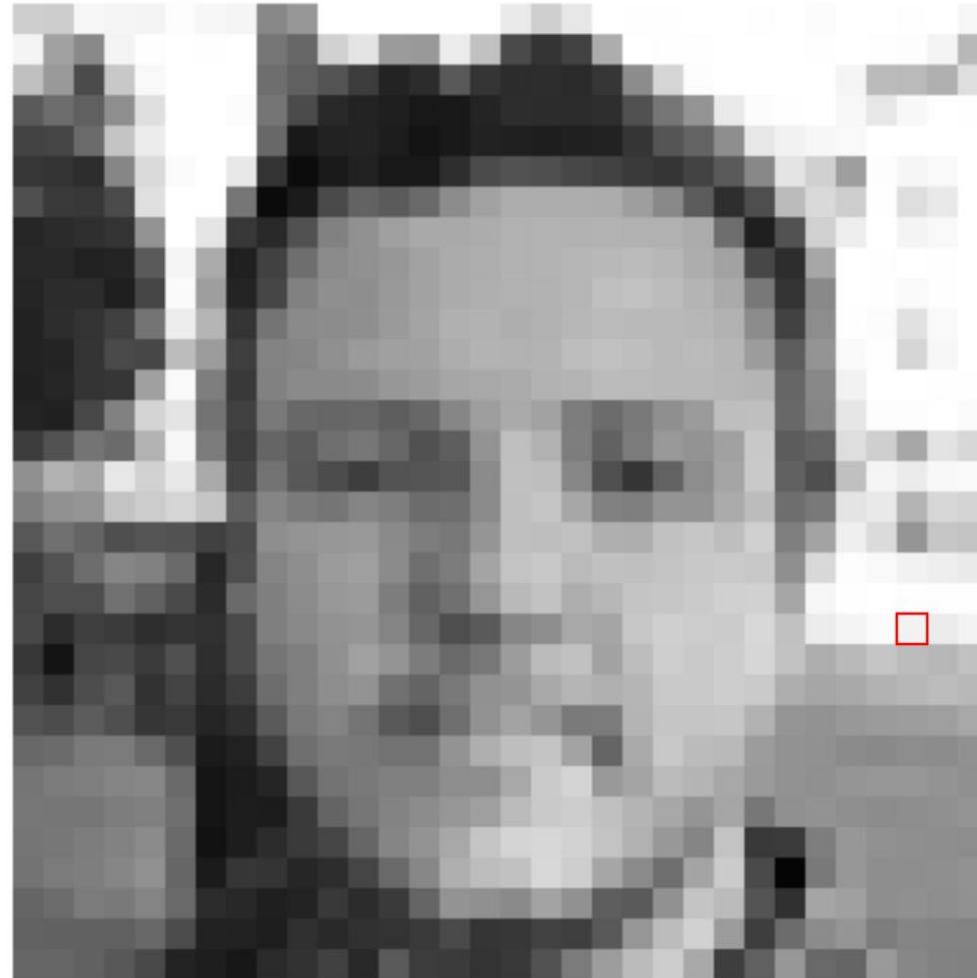
Object tracking: follow the movement of an object in a video.



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

What computers see

206 206 247 245 244 253 247 245 136 151 255 255 255 255 255 255 254 207 231 255 254 254 255 255 254 255 252 255 255 254 255 247
244 161 137 244 254 255 254 255 118 103 209 228 155 153 236 105 74 52 69 173 255 254 254 255 255 255 254 255 254 253 244 184
192 154 75 200 249 255 255 255 110 98 84 81 35 44 89 93 44 45 43 54 140 213 253 255 255 255 255 245 187 186 178 223
90 109 98 143 223 255 255 252 117 75 41 35 31 24 25 38 45 44 44 48 81 118 148 234 252 254 255 248 231 248 255 254
67 69 107 196 236 255 255 255 104 25 34 35 29 20 25 34 32 30 32 34 53 85 100 142 231 242 247 249 255 255 255 255
55 51 45 134 215 251 255 232 51 52 28 33 24 24 48 75 82 78 71 68 58 53 67 90 138 228 208 158 253 248 249 255
79 58 58 75 224 255 255 118 11 27 74 99 91 106 140 162 173 173 173 172 158 137 92 48 78 187 217 206 254 222 233 255
38 43 47 52 147 255 229 58 41 81 129 145 160 169 169 172 178 179 178 179 177 177 172 110 31 82 209 238 255 244 249 255
40 40 33 38 90 245 171 32 65 110 139 145 151 162 171 174 178 179 182 184 187 183 173 162 71 45 167 255 254 255 254 255
37 44 44 31 69 250 158 38 70 129 143 142 153 162 171 175 177 178 182 191 194 188 180 170 120 51 137 255 254 250 254 255
34 45 51 64 116 237 181 53 116 138 140 143 154 164 176 178 174 177 183 186 185 185 183 178 140 69 141 254 252 225 249 255
34 36 52 74 71 188 156 63 131 134 144 155 160 161 173 179 178 179 189 193 190 185 187 182 156 93 148 250 254 214 247 255
32 38 52 54 159 250 126 57 129 138 138 140 151 156 166 166 171 178 180 187 186 185 185 183 180 102 136 242 255 255 254 254
36 32 72 129 212 228 115 65 121 104 102 104 94 103 134 158 170 162 125 108 121 143 155 190 191 104 134 230 253 253 255 251
61 82 116 107 179 247 124 60 101 90 111 119 103 81 94 147 191 178 126 98 125 153 147 161 200 92 100 222 207 167 227 215
144 178 167 231 210 232 170 67 115 88 76 62 83 85 88 139 192 190 135 80 53 99 141 166 201 97 79 192 245 235 248 249
127 145 149 155 204 213 197 95 133 122 117 133 126 108 110 139 191 197 167 129 127 148 147 171 188 110 121 228 233 180 215 212
87 112 100 79 85 82 65 75 142 148 151 153 138 125 120 149 191 190 193 175 174 193 198 190 208 127 163 239 219 149 158 155
63 83 109 134 129 106 39 78 132 142 155 159 139 111 124 164 155 200 186 192 191 195 200 202 200 143 217 253 249 242 238 234
69 78 78 113 97 74 43 106 127 140 152 155 125 97 112 150 185 194 174 183 196 198 202 206 209 166 247 254 255 254 254 254
72 44 63 59 48 52 49 74 127 137 146 149 132 103 78 90 134 141 168 165 199 207 204 203 216 193 236 244 251 242 236 243
55 20 69 73 59 80 48 74 117 127 144 161 148 124 105 120 156 187 193 162 189 206 201 205 214 194 174 185 197 188 183 193
65 49 77 89 90 88 43 81 109 127 141 147 113 100 121 145 148 169 181 178 181 201 201 205 202 174 166 169 178 183 188 184
82 76 92 79 54 58 37 47 90 121 132 116 89 78 111 146 163 149 122 124 180 197 197 198 178 149 146 152 155 157 159 168
104 107 122 123 105 79 27 33 68 111 122 120 114 114 147 175 190 196 163 101 170 200 187 185 156 146 145 139 137 141 140 145
117 124 127 133 135 105 21 28 37 88 115 121 128 128 141 142 168 202 212 153 164 186 180 188 154 146 144 149 151 151 147 144
119 118 118 125 128 111 21 29 28 58 100 118 131 140 151 159 186 201 205 192 180 168 149 166 119 144 147 143 140 141 144 145
117 119 125 130 139 106 18 29 44 58 70 102 133 147 168 197 212 215 210 165 177 152 133 195 57 89 126 151 145 143 142 141
115 123 126 134 145 102 27 54 52 38 45 69 105 135 175 189 193 216 206 166 139 111 164 203 74 5 121 151 142 142 143 146
101 108 123 121 132 105 44 40 31 35 57 44 59 101 147 144 138 163 145 94 90 145 196 187 84 48 165 180 142 144 142 145
98 97 97 98 104 76 34 33 30 48 41 49 51 58 74 53 55 68 63 89 150 188 209 156 62 106 140 149 125 133 131 131
102 102 97 88 73 35 30 23 42 50 65 41 90 60 59 51 57 82 123 157 187 205 169 62 98 151 105 101 154 135 130 129



<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.

Translation Invariance



Rotation/Viewpoint Invariance



Size Invariance



Illumination Invariance



Flatten

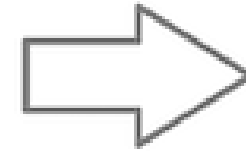
We could just represent each pixel as a feature.

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?

1	1	0
4	2	1
0	2	1

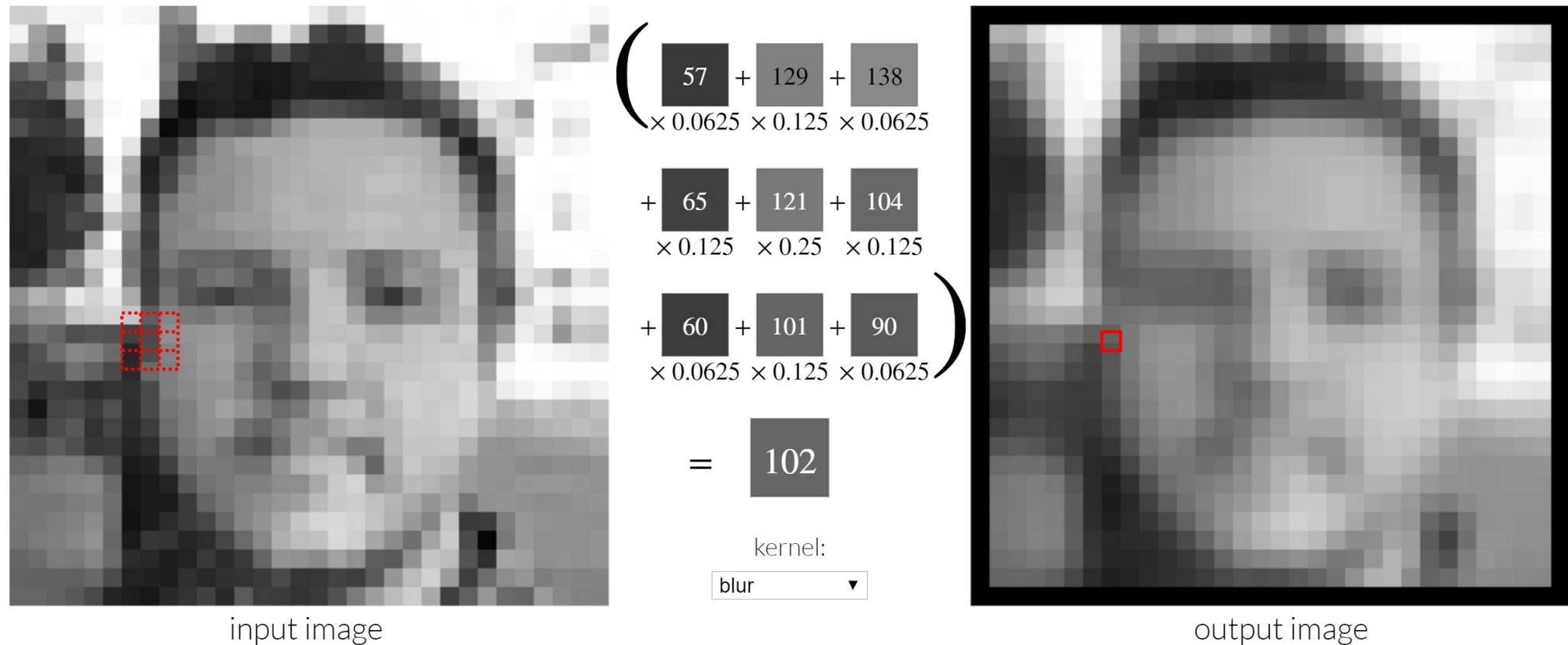


1
1
0
4
2
1
0
2
1

Convolutional filter

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

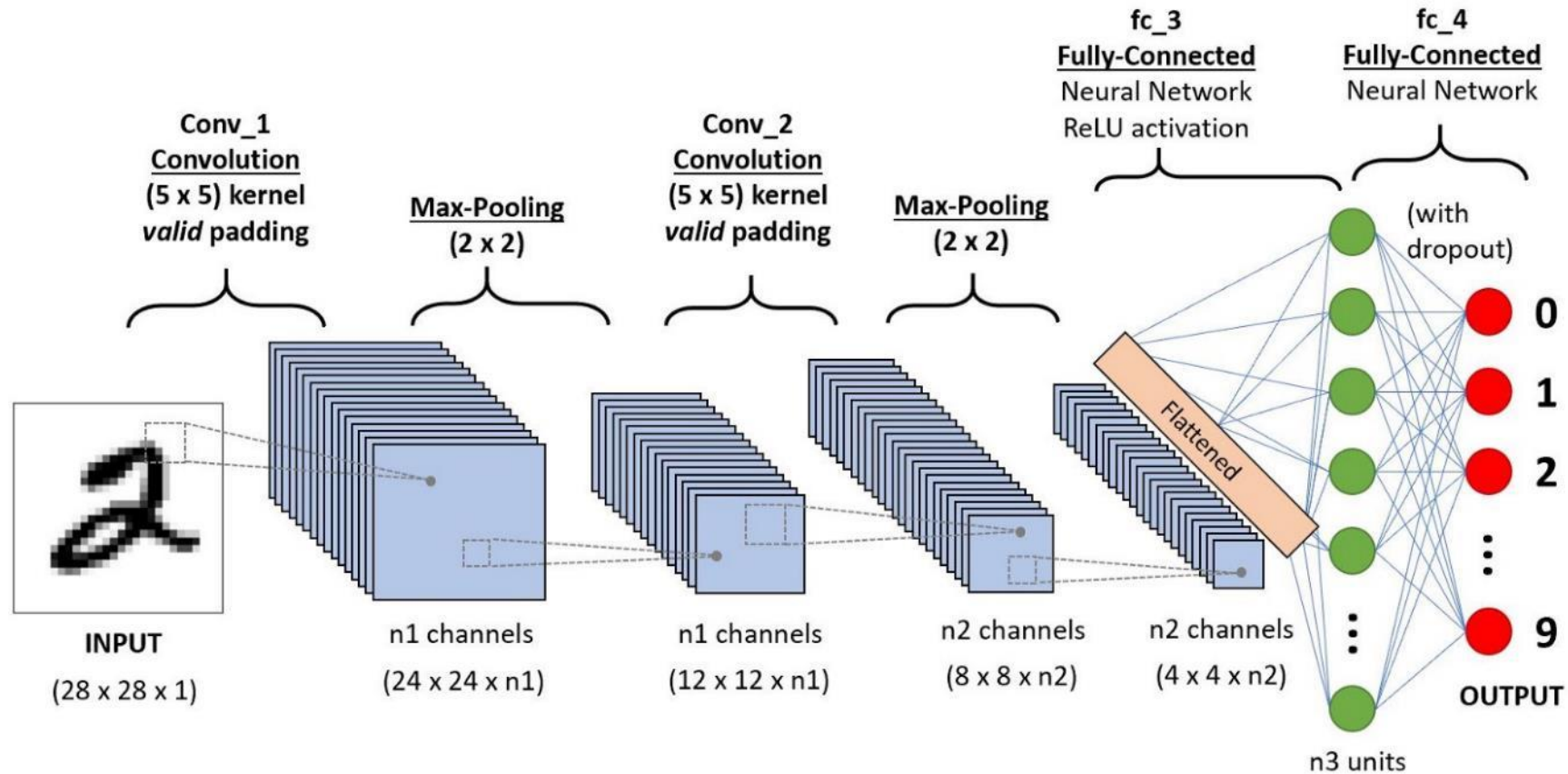
They provides a measure for how close a patch or a region in the image resembles a **feature**.



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.

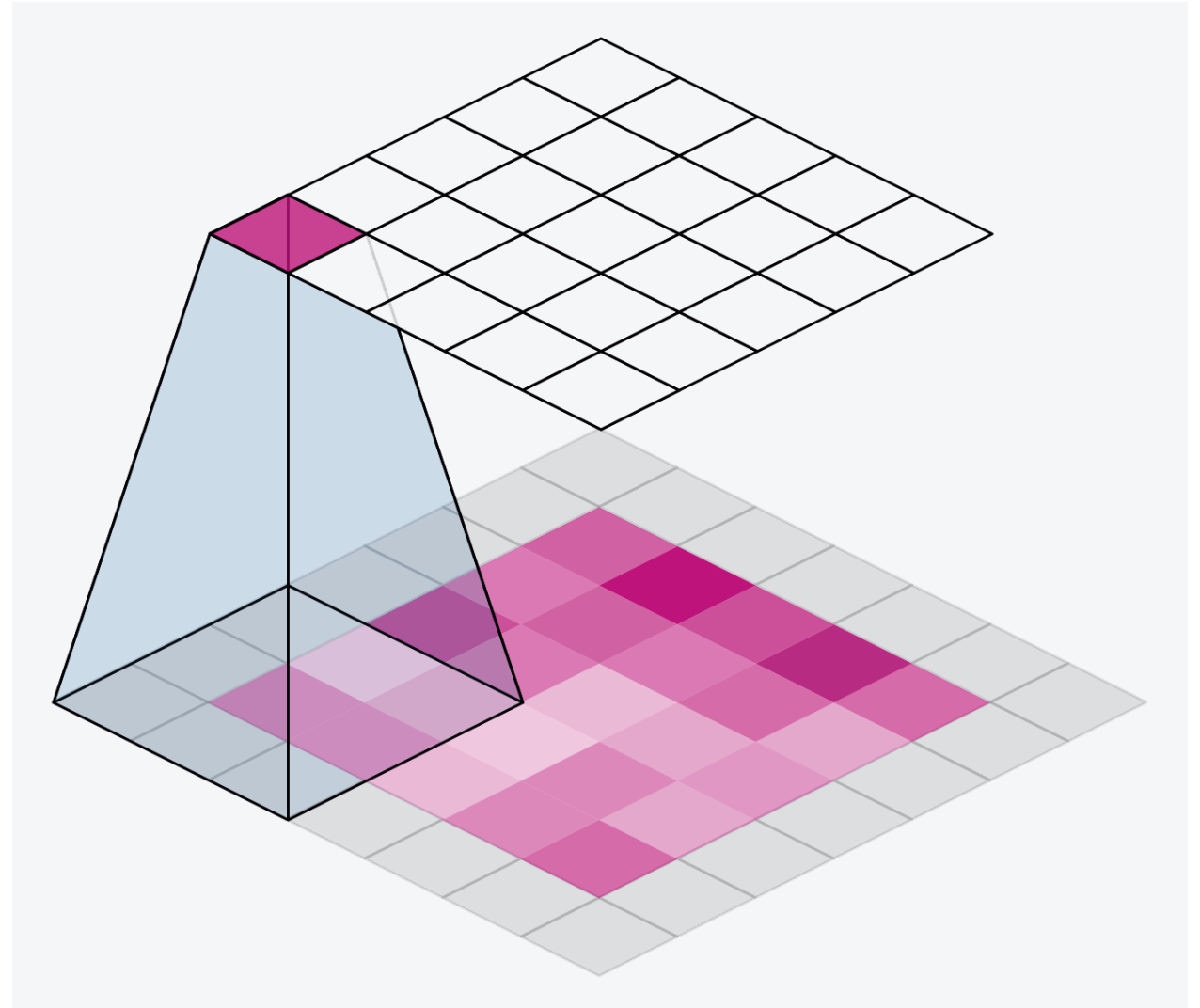


Feature map

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

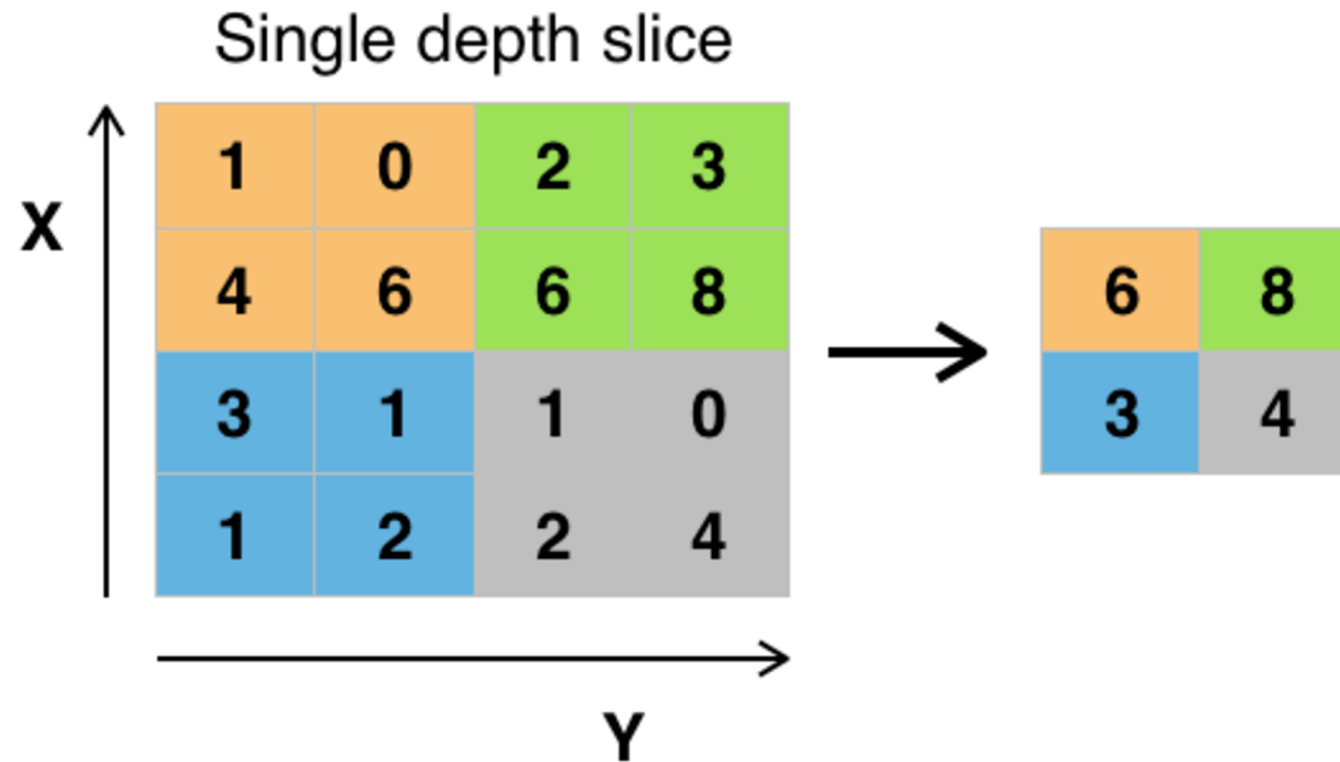
The convolution is defined by the model parameters that are learned from the data.

So, in each feature maps, the model parameters 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.

Translation Invariance



Rotation/Viewpoint Invariance



Size Invariance

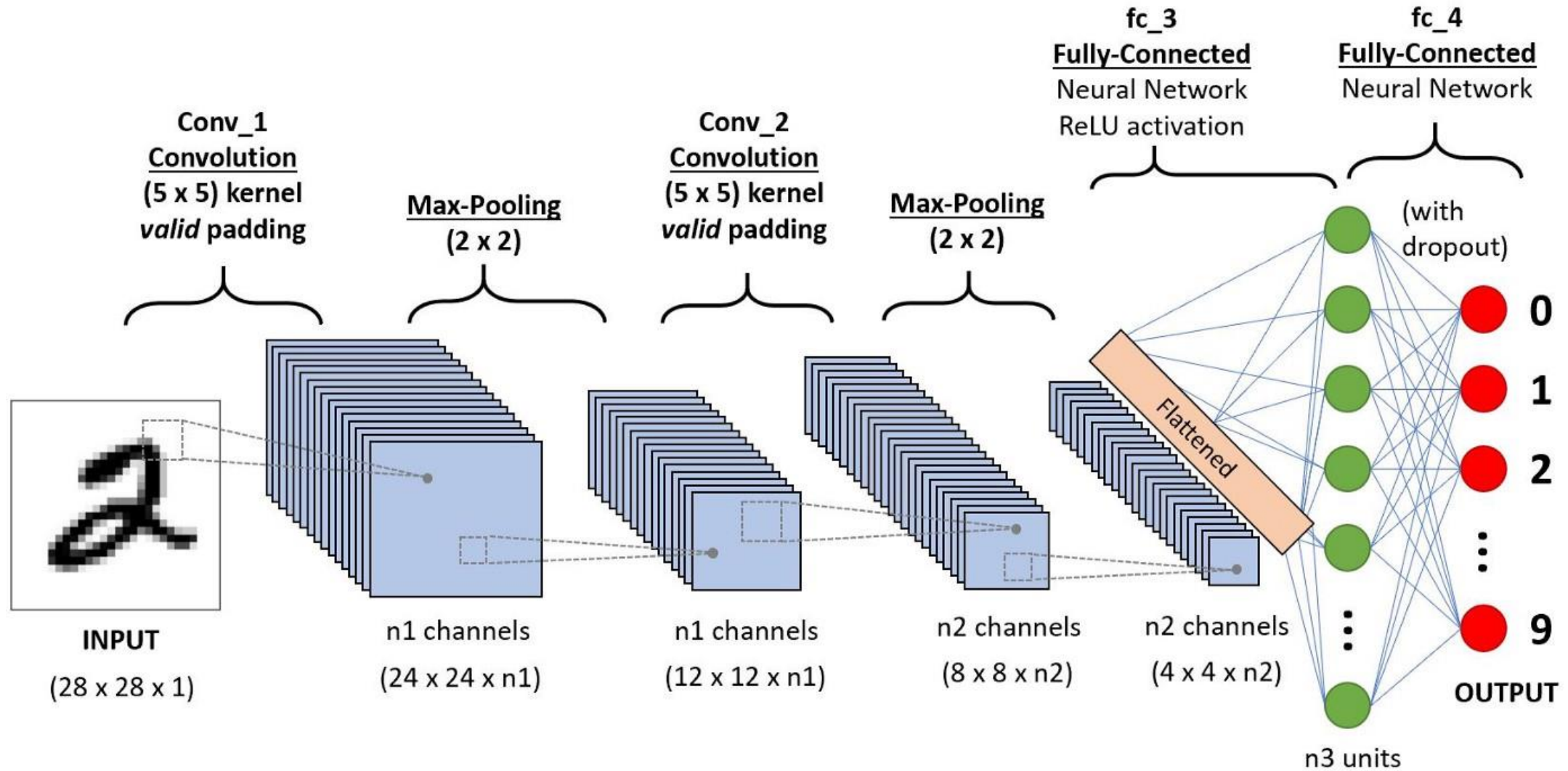


Illumination Invariance

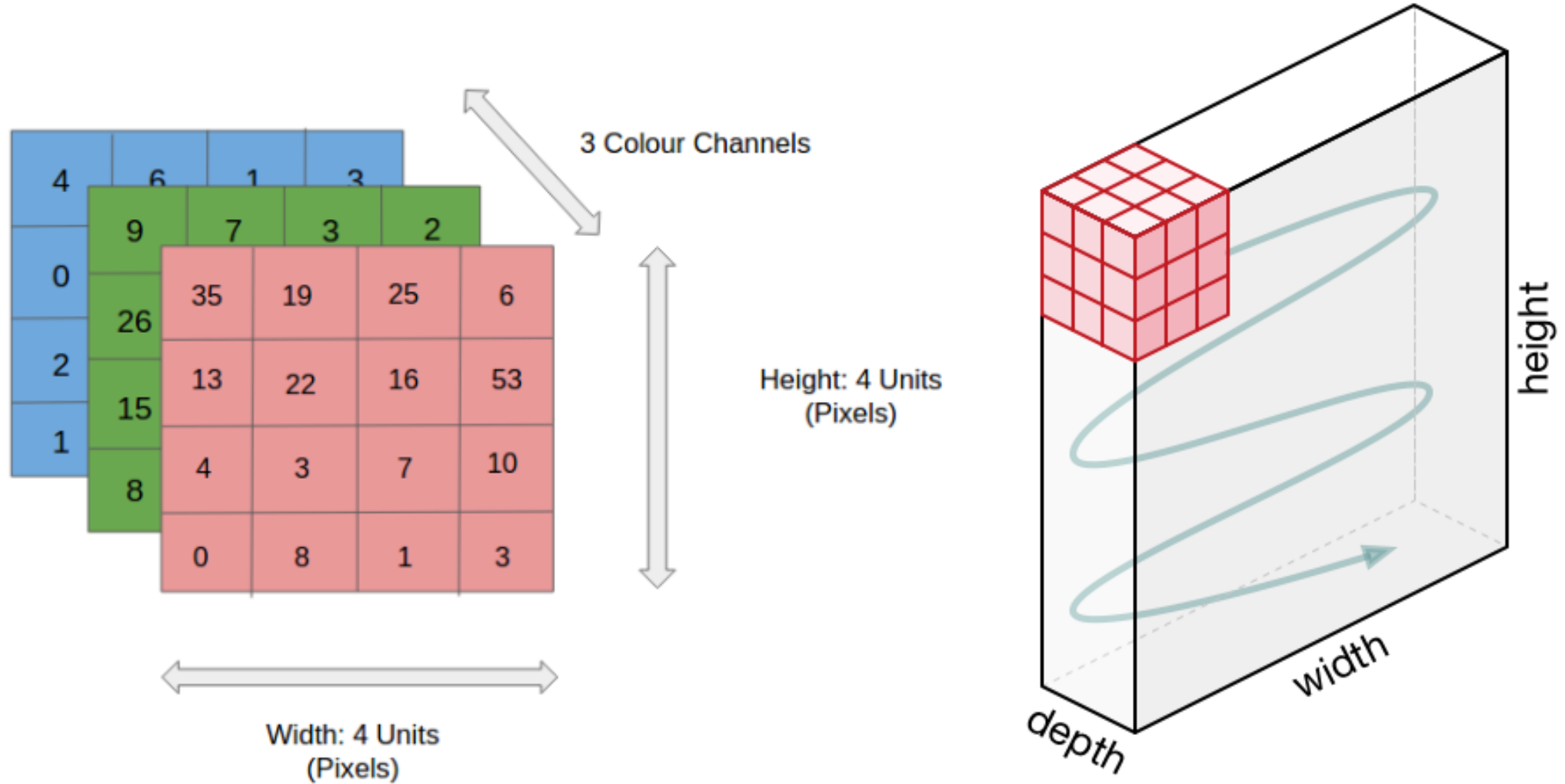


Matt Krause
mattkrause

CNN



CNN: channels



Neural network



`mnist_pytorch.ipynb`

ImageNet

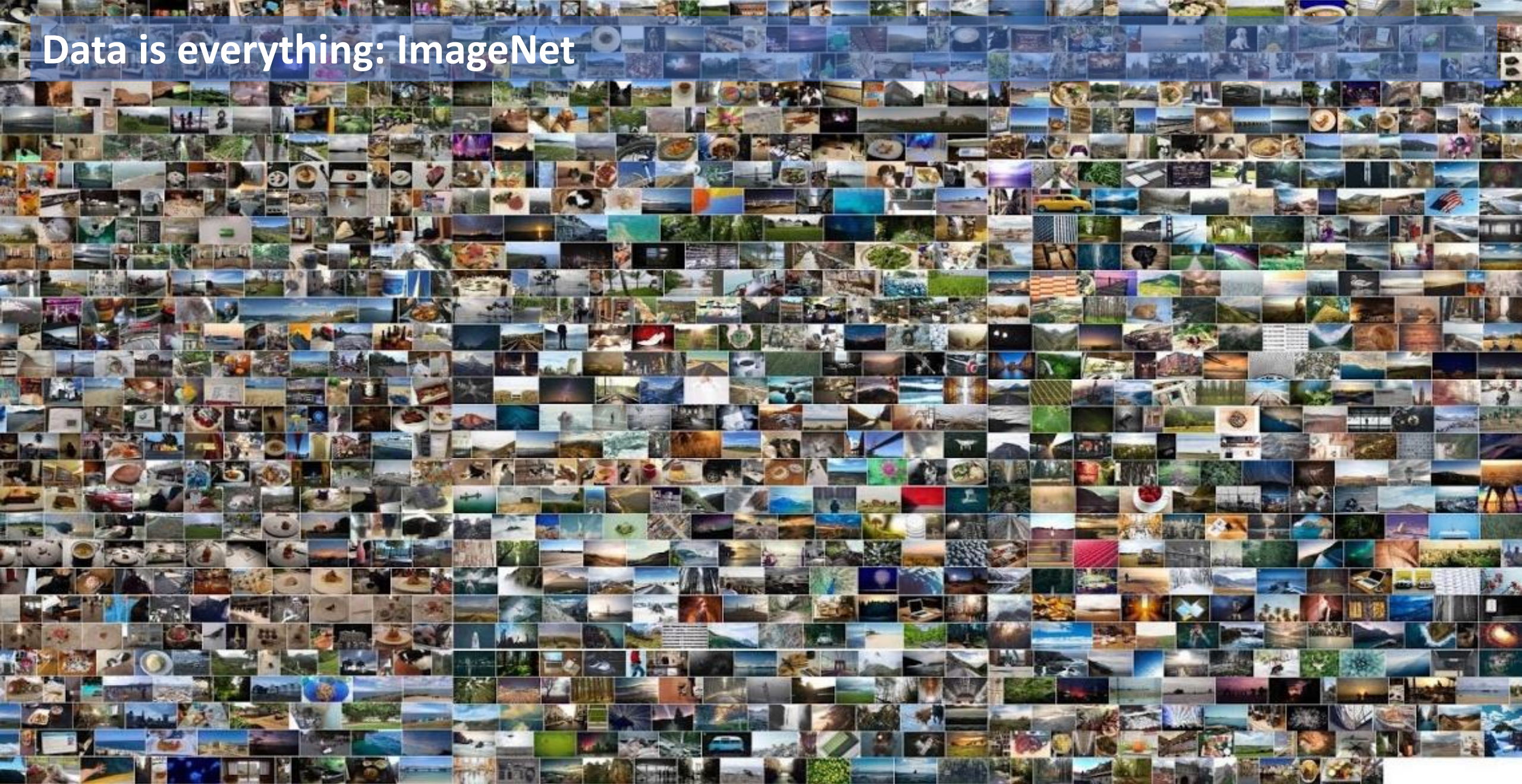
ImageNet is a large-scale visual database designed for use in visual object recognition research, containing millions of labeled images across thousands of categories.

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.

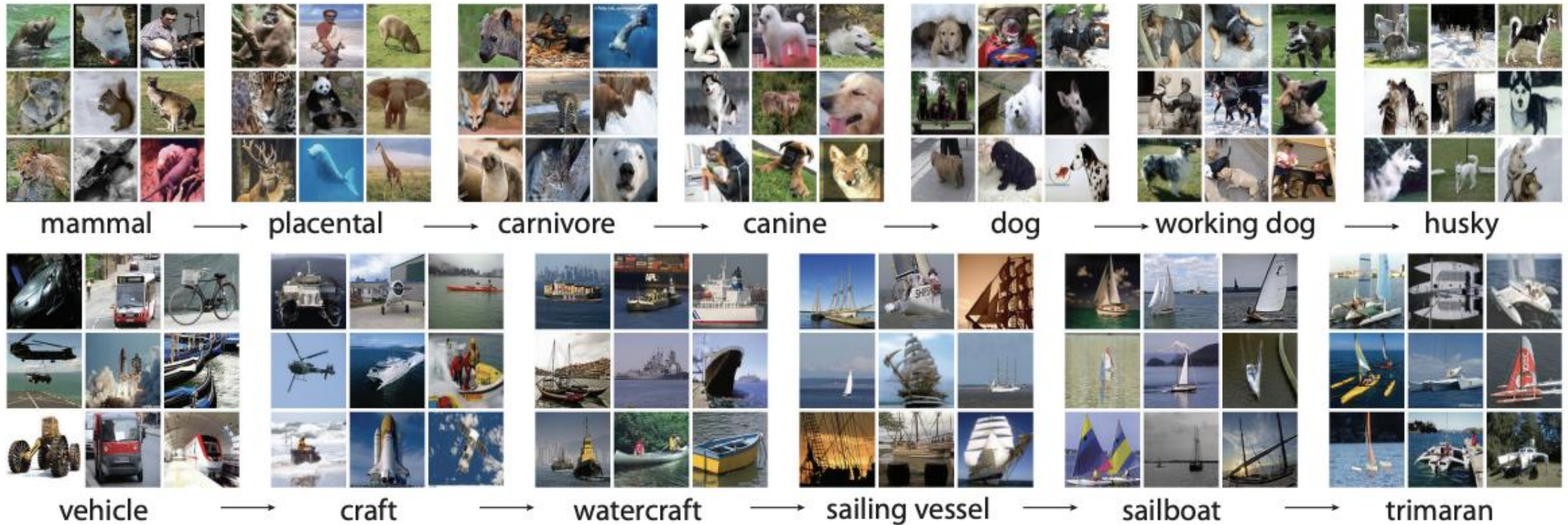
Data is everything: ImageNet



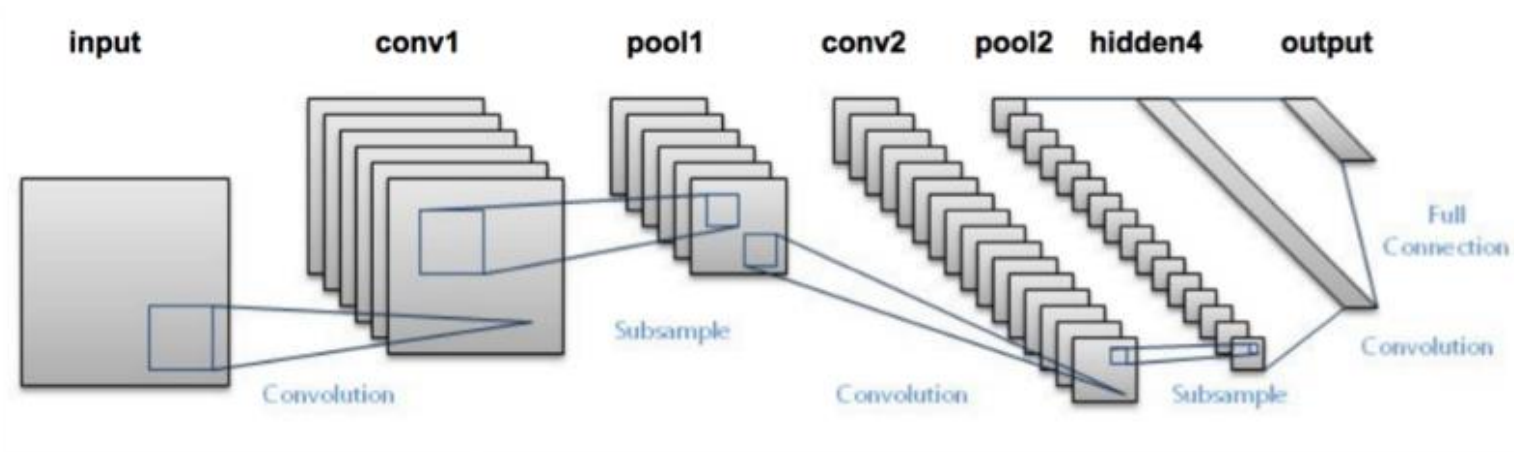
J. Deng et al., "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition

Machine Learning Methods for Biomedical Data (D012554)

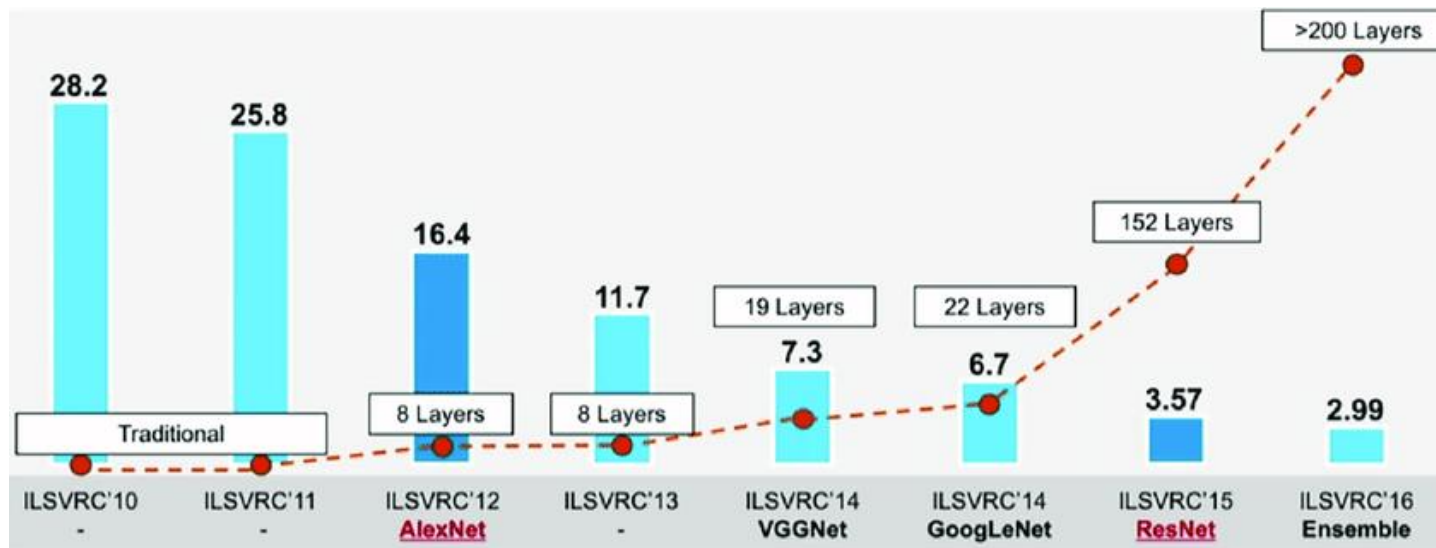
Data is everything: ImageNet



ImageNet: computer vision is solved

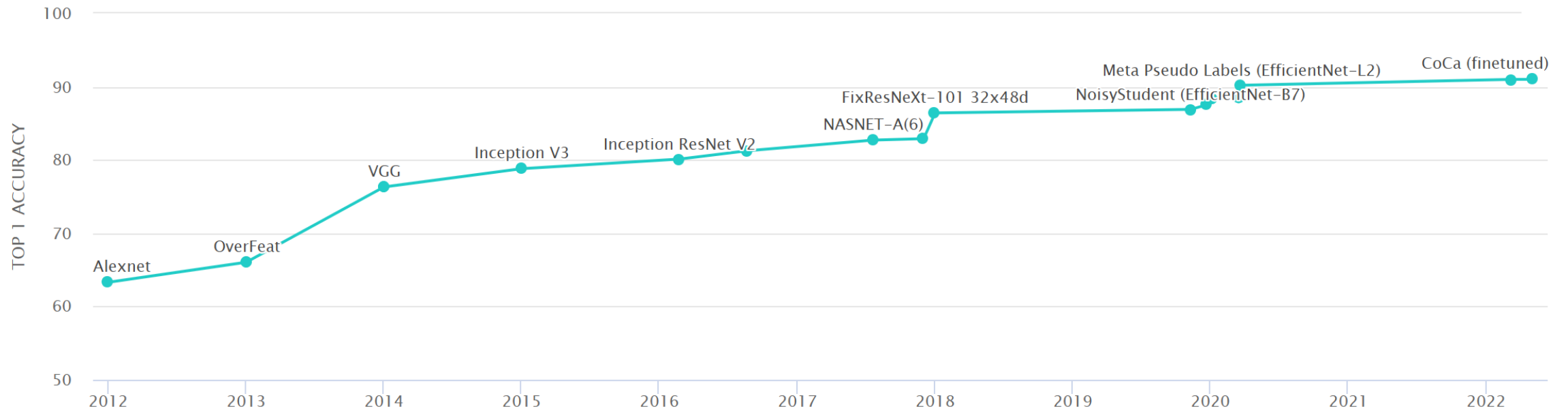


LeNet-5 (1998)



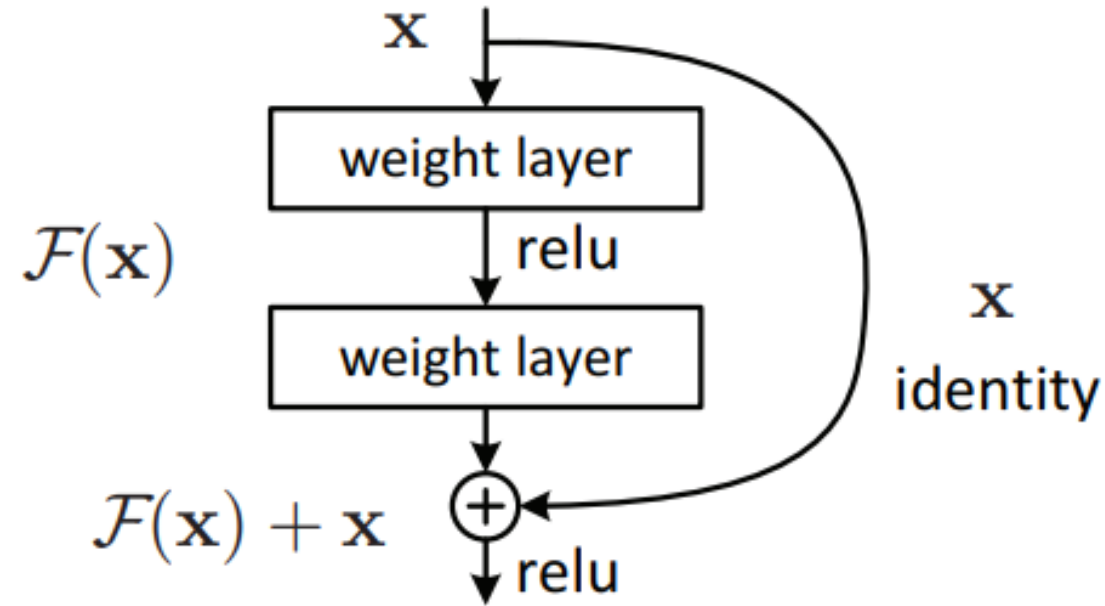
ImageNet Large Scale Visual Recognition Challenge: <https://www.image-net.org/challenges/LSVRC/>

ImageNet: computer vision is solved



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.



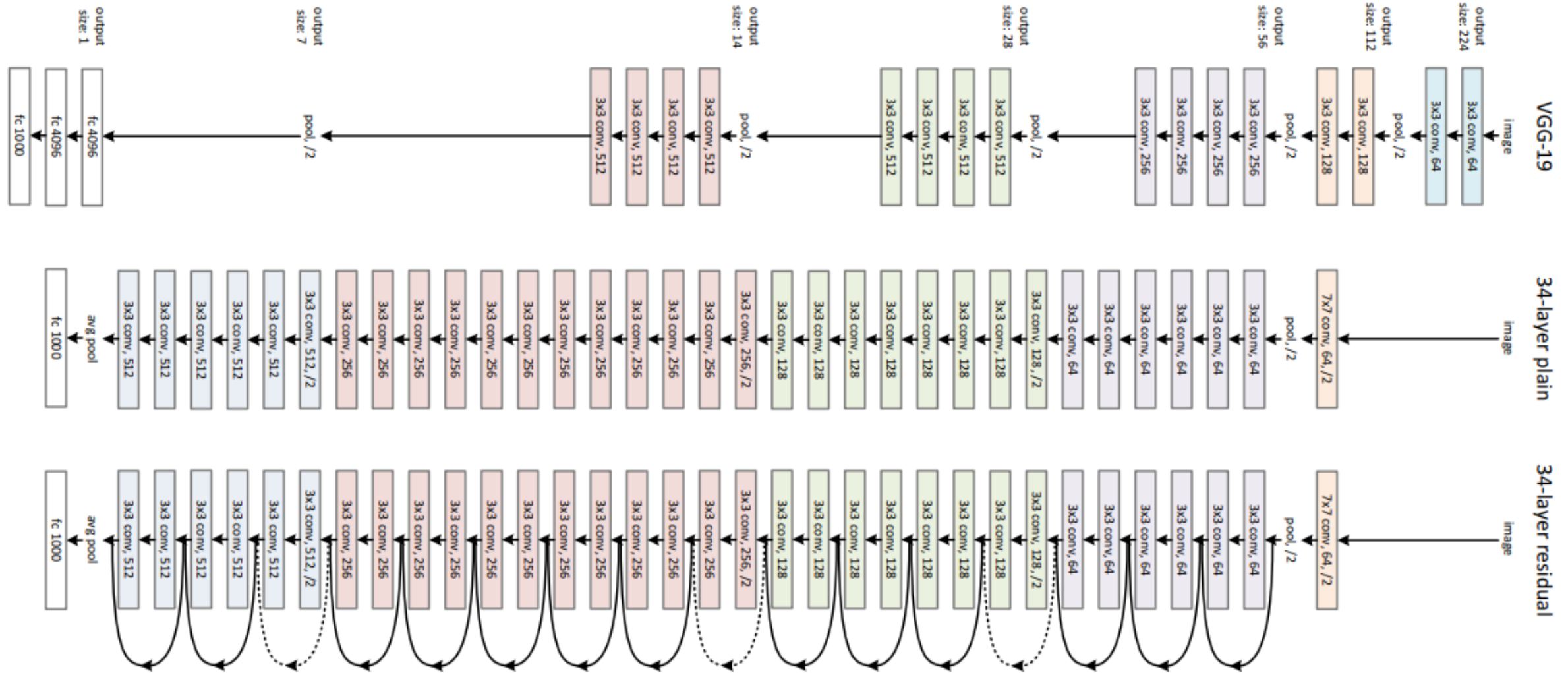
ResNet (Residual Network)

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

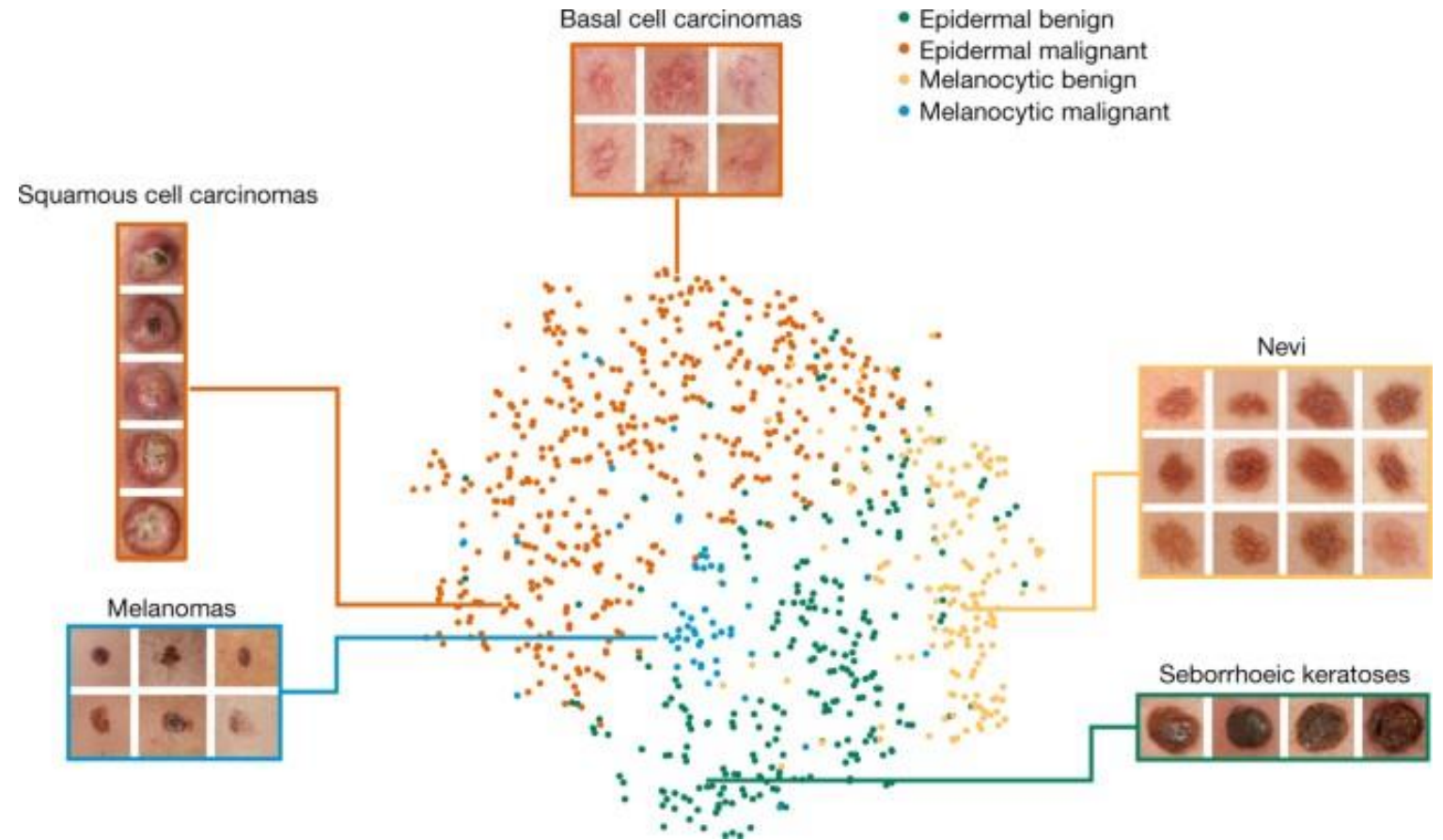
Instead of learning the full transformation, ResNet learns the **residual (difference) between input and output**, making optimization easier and improving convergence.

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

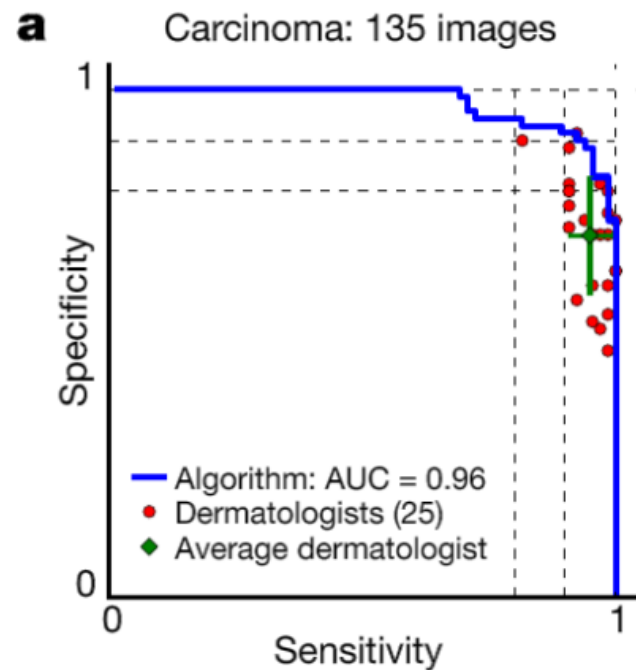
ResNet (Residual Network)



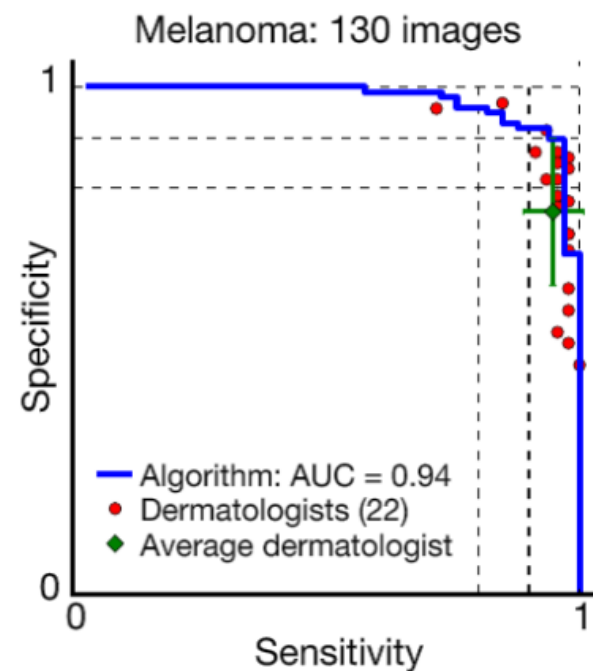
Remember this?



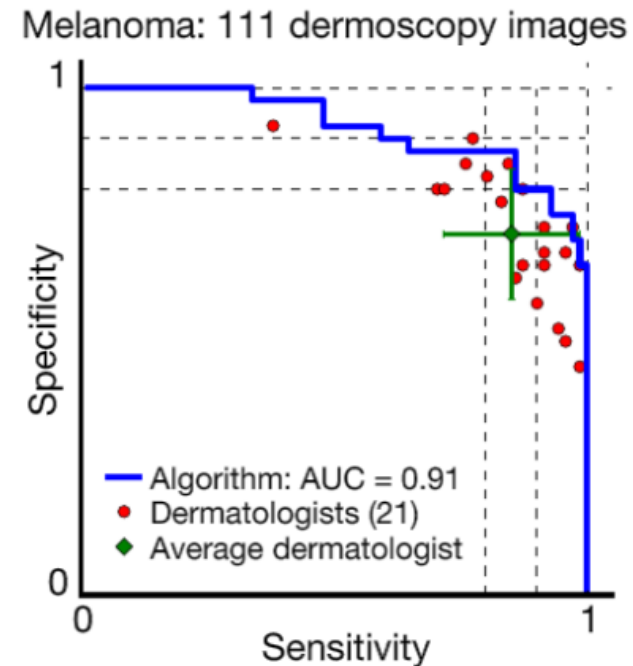
Thrun et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature (2017)



keratinocyte carcinomas: 65
benign seborrheic keratoses: 75



malignant melanomas: 33
benign nevi: 97



malignant melanomas: 71
benign nevi: 40