C14200: FROM ZERO TO ONE - DEEP LEARNING WITH PYTORCH

JOE SPISAK
PRODUCT MANAGER

FRANCISCO MASSA
RESEARCH ENGINEER

WHO AM I?

Francisco Massa
fmassa

CURRENT: RESEARCH ENGINEER - PYTORCH

PREVIOUS:
- COMPUTER VISION RESEARCH ENGINEER @TWITTER
- PHD STUDENT AT ECOLE DES PONTS - FRANCE

AGENDA

01
PRIMER ON COMPUTER VISION

02
GOING BEYOND IMAGE CLASSIFICATION

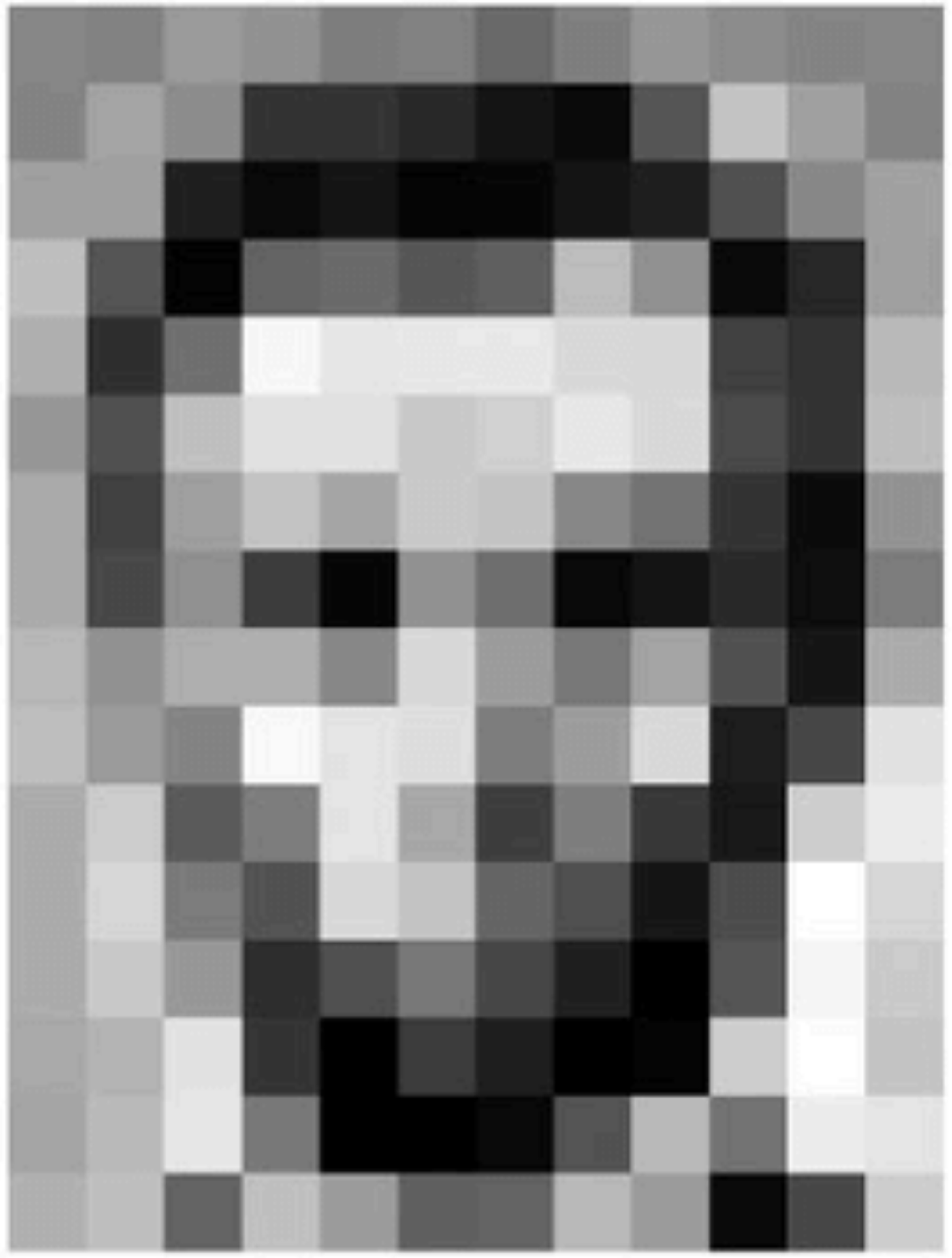
03
COMPUTER VISION WITH TORCHVISION

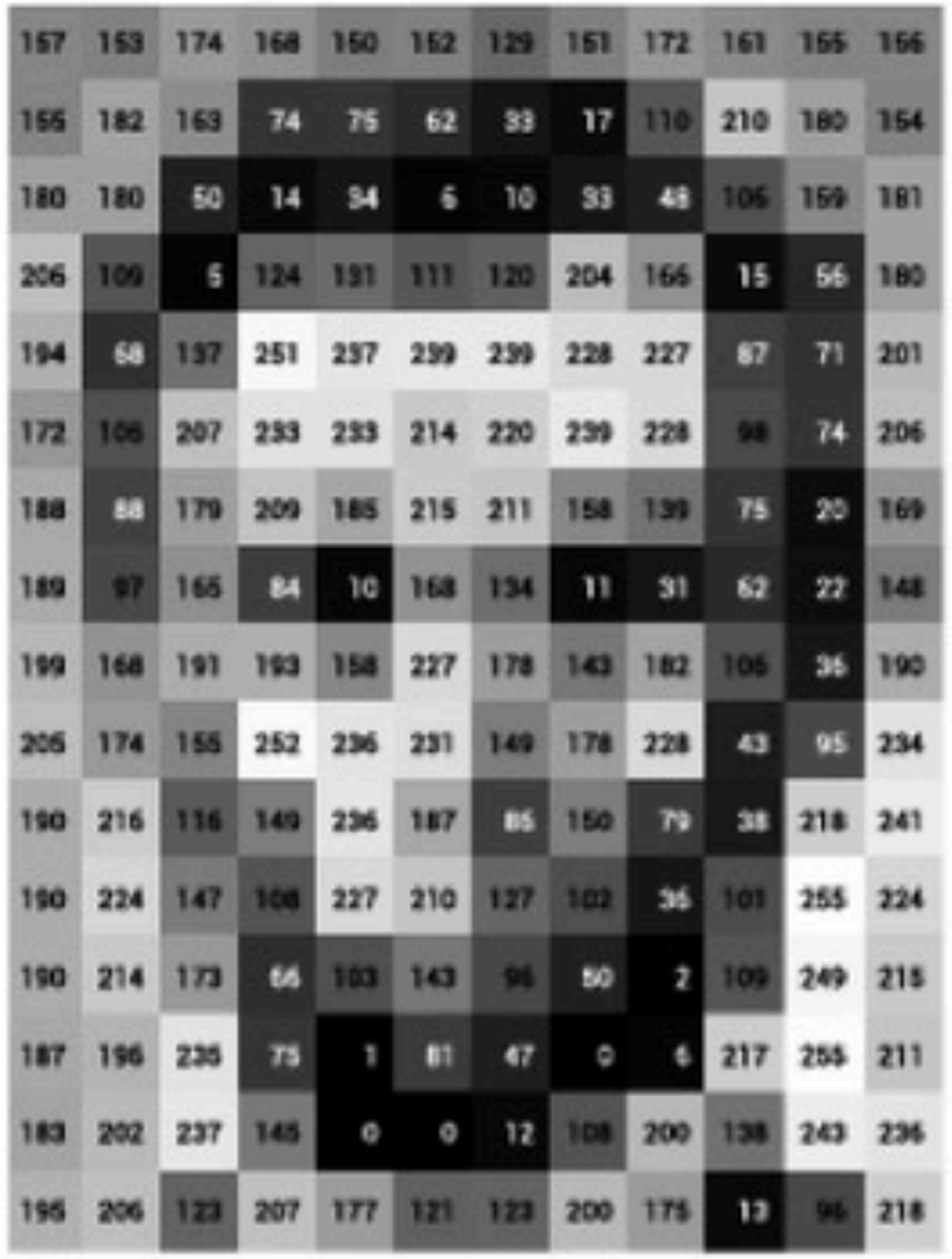
04
HANDS ON WITH DEEPLAB V3

05
HANDS ON WITH KEYPOINT R-CNN

PRIMER ON COMPUTER VISION

HOW ARE IMAGES REPRESENTED IN A COMPUTER



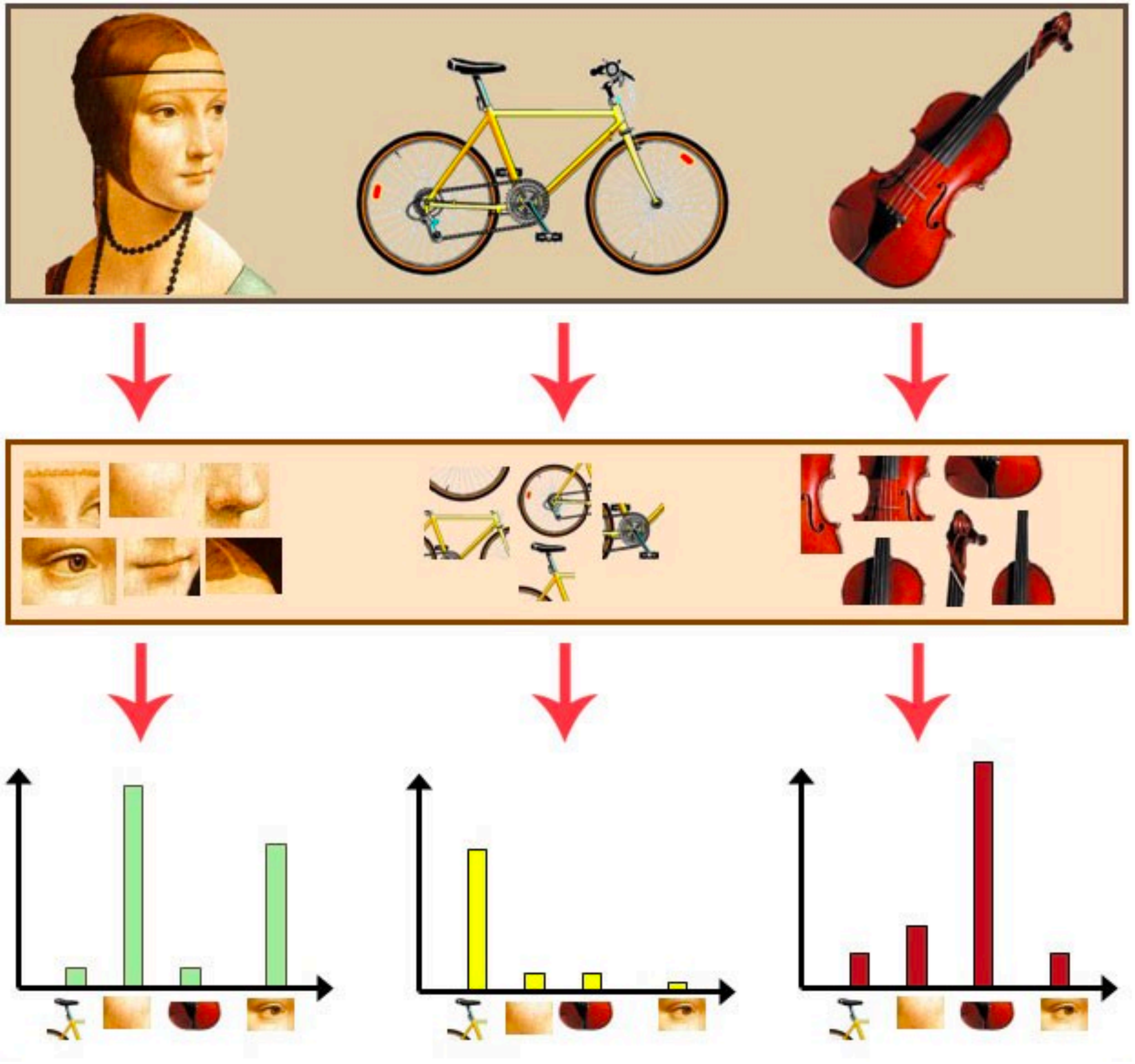


157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

How can we make the computer understand what is in the image?

Image from https://www.researchgate.net/publication/330902210/figure/fig1/AS:878026619375622@1586349267376/mage-of-Abraham-Lincoln-as-a-matrix-of-pixel-values.ppm

OBJECTS AS A BAG OF VISUAL WORDS



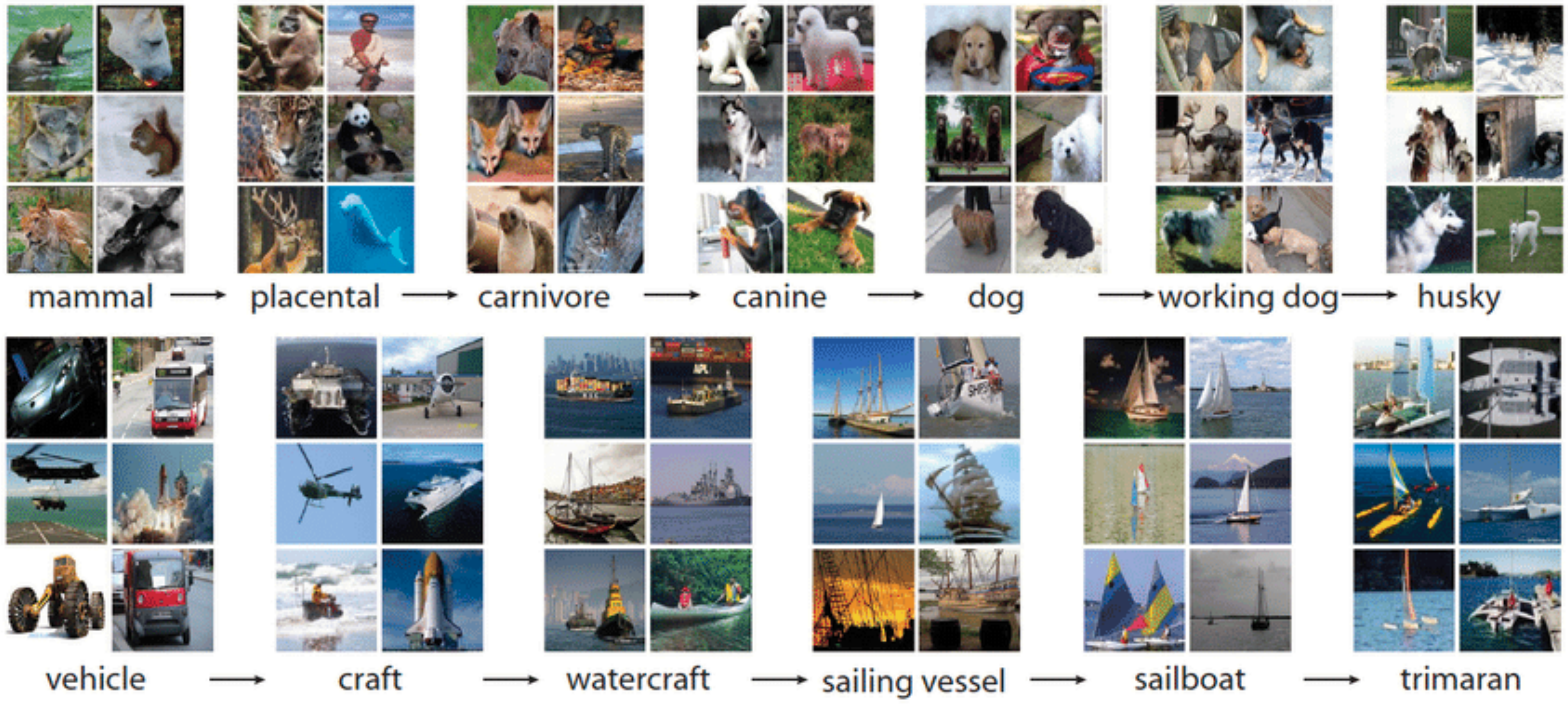
Prototype patches (visual words)

Histogram of visual words

Image from https://towardsdatascience.com/bag-of-visual-words-in-a-nutshell-9ceea97ce0fb

ALEXNET - A REVOLUTION IN COMPUTER VISION

ImageNet Image Classification Challenge



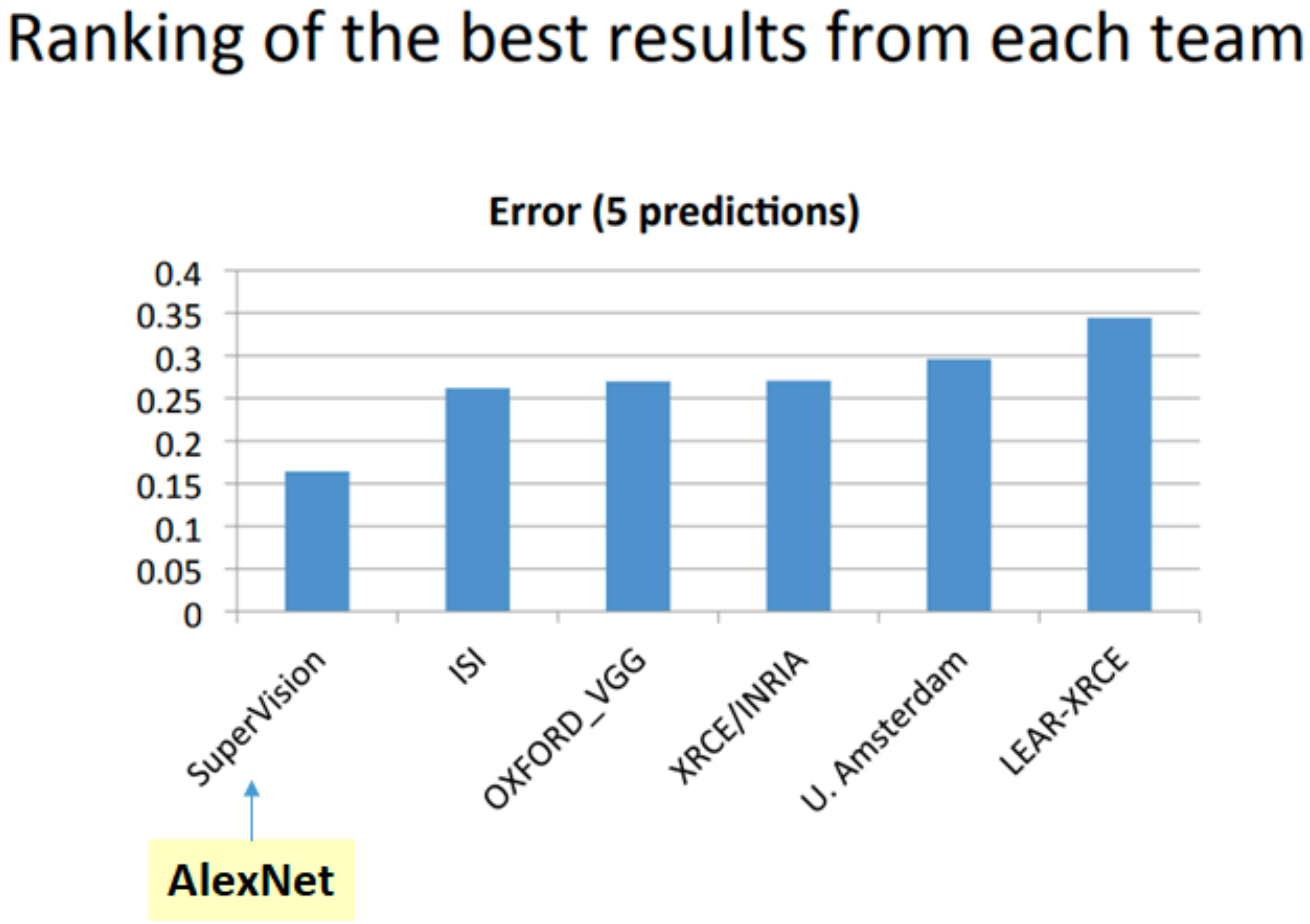
1M Images

1000 categories

Image from https://medium.com/@atharvabarve24/breakthrough-in-computer-vision-c318fedba4d2

ALEXNET - A REVOLUTION IN COMPUTER VISION

ImageNet Image Classification Challenge



Key ingredients:

- Deep Convolutional Neural Networks
- Lot's of training data
- ReLU and dropout
- GPUs

Image from https://medium.com/@atharvabarve24/breakthrough-in-computer-vision-c318fedba4d2

CONVOLUTIONAL NEURAL NETWORKS

What is a convolution?

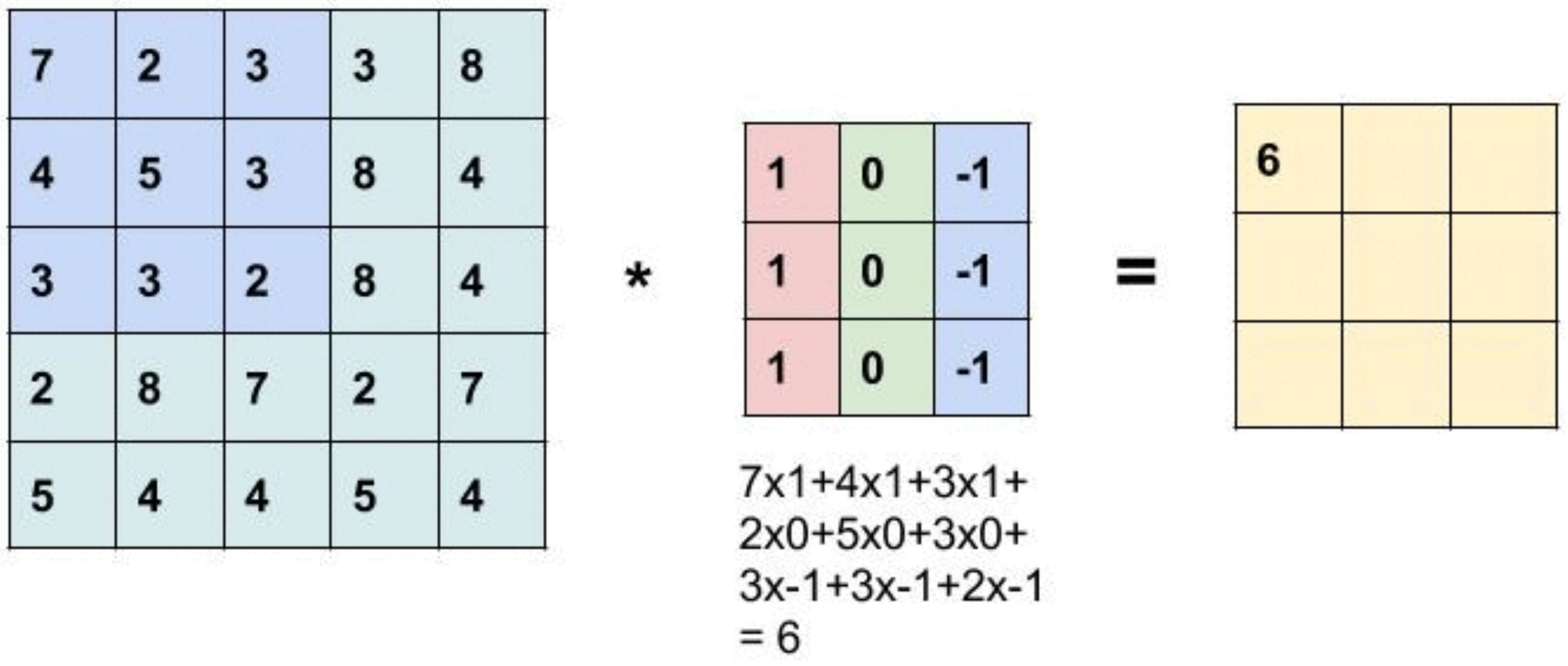


Image from https://medium.com/datadriveninvestor/convolutional-neural-networks-3b241a5da51e

Neural networks that uses the convolution operator

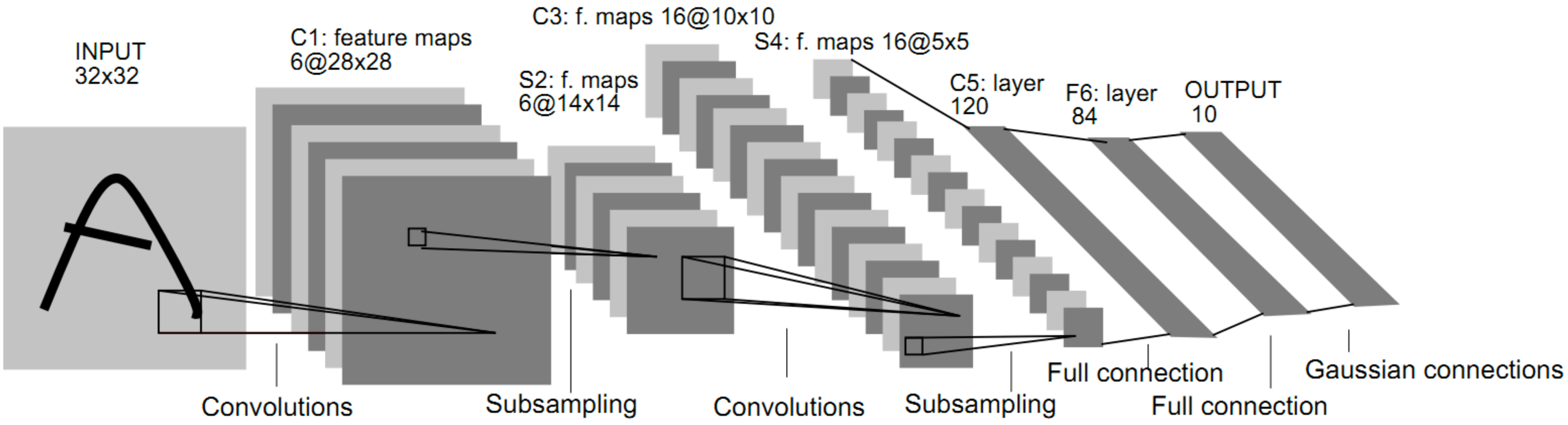


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Image from https://pythonmachinelearning.pro/introduction-to-convolutional-neural-networks-for-vision-tasks/

CONVOLUTIONAL NEURAL NETWORKS TODAY

More layers work better

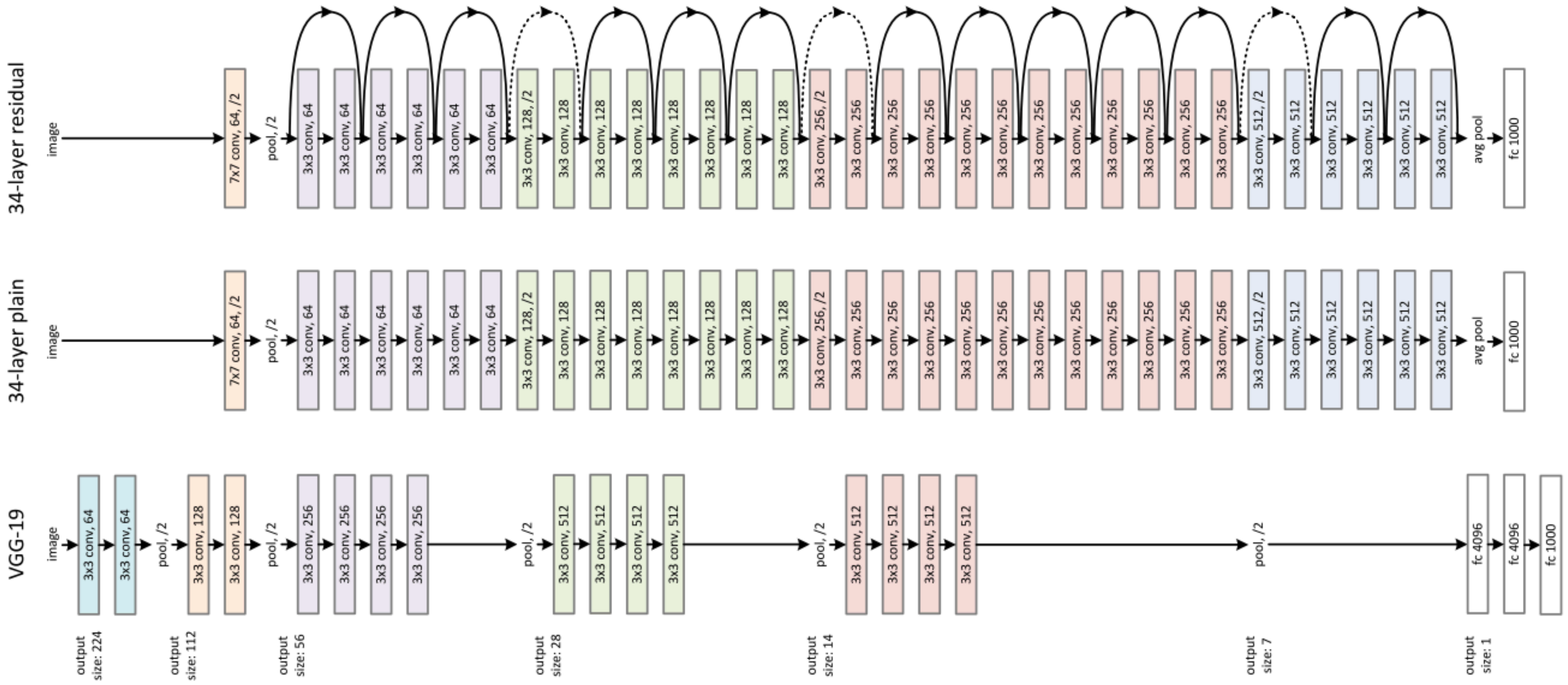
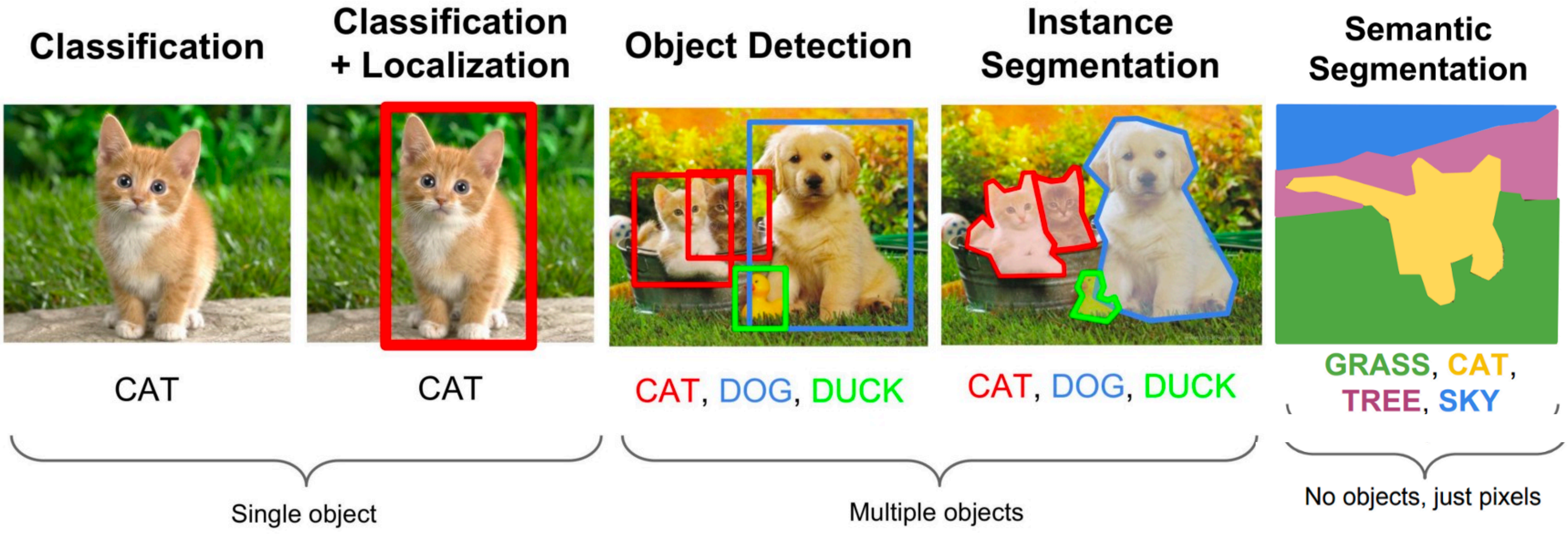


Image from https://medium.com/@pierre_guillou/understand-how-works-resnet-without-talking-about-residual-64698f157e0c

GOING BEYOND IMAGE CLASSIFICATION

WHAT IS IN THE IMAGE, AND WHERE?



Images from https://medium.com/zylapp/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852 and https://medium.com/ml-research-lab/what-is-object-detection-51f9d872ece7

CNN OVER REGIONS (R-CNN)

Image from https://www.halifaxhumanesociety.org/20-Little-Known-Facts-About-Cats-and-Dogs-1-38.html

COMPUTER VISION WITH TORCHVISION

WHAT IS TORCHVISION?

A library built to facilitate research and experimentation in the field of Computer Vision

DATASETS

COMMON DATASETS

MODELS

PRE-TRAINED MODELS

OPS

EFFICIENT OPERATORS

TRANSFORMS

DATA TRANSFORMATION

IO

EFFICIENT VIDEO READER

REFERENCES

TRAINING SCRIPTS

DATASETS

TRANSFORMS

MODELS

OPS

IO

IMAGENET 64X64

DATASETS

TRANSFORMS

MODELS

OPS

IO

```
impor                        for

T.Com
     T                       224
     T                       ip(
     T
     T                       486

])
```

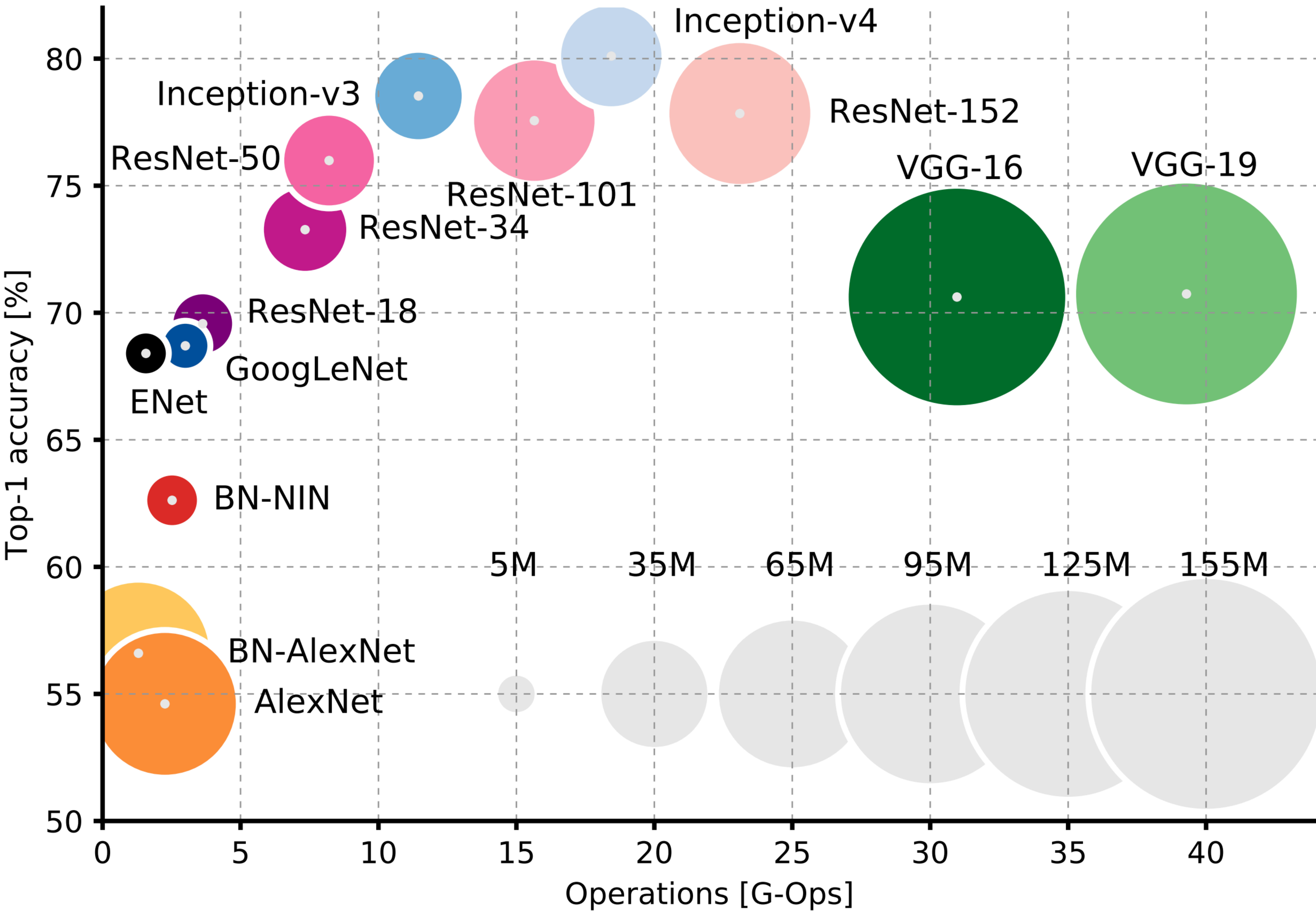
DATASETS

TRANSFORMS

MODELS

OPS

IO



An Analysis of Deep Neural Network Models for Practical Applications

Alfredo Canziani, Adam Paszke, Eugenio Culurciello

DATASETS

TRANSFORMS

MODELS

OPS

IO

```
import torchvision.ops

torchvision.ops.box_iou(...)
torchvision.ops.roi_align(...)
torchvision.ops.nms(...)
torchvision.ops.roi_pool(...)
```

DATASETS

TRANSFORMS

MODELS

OPS

IO

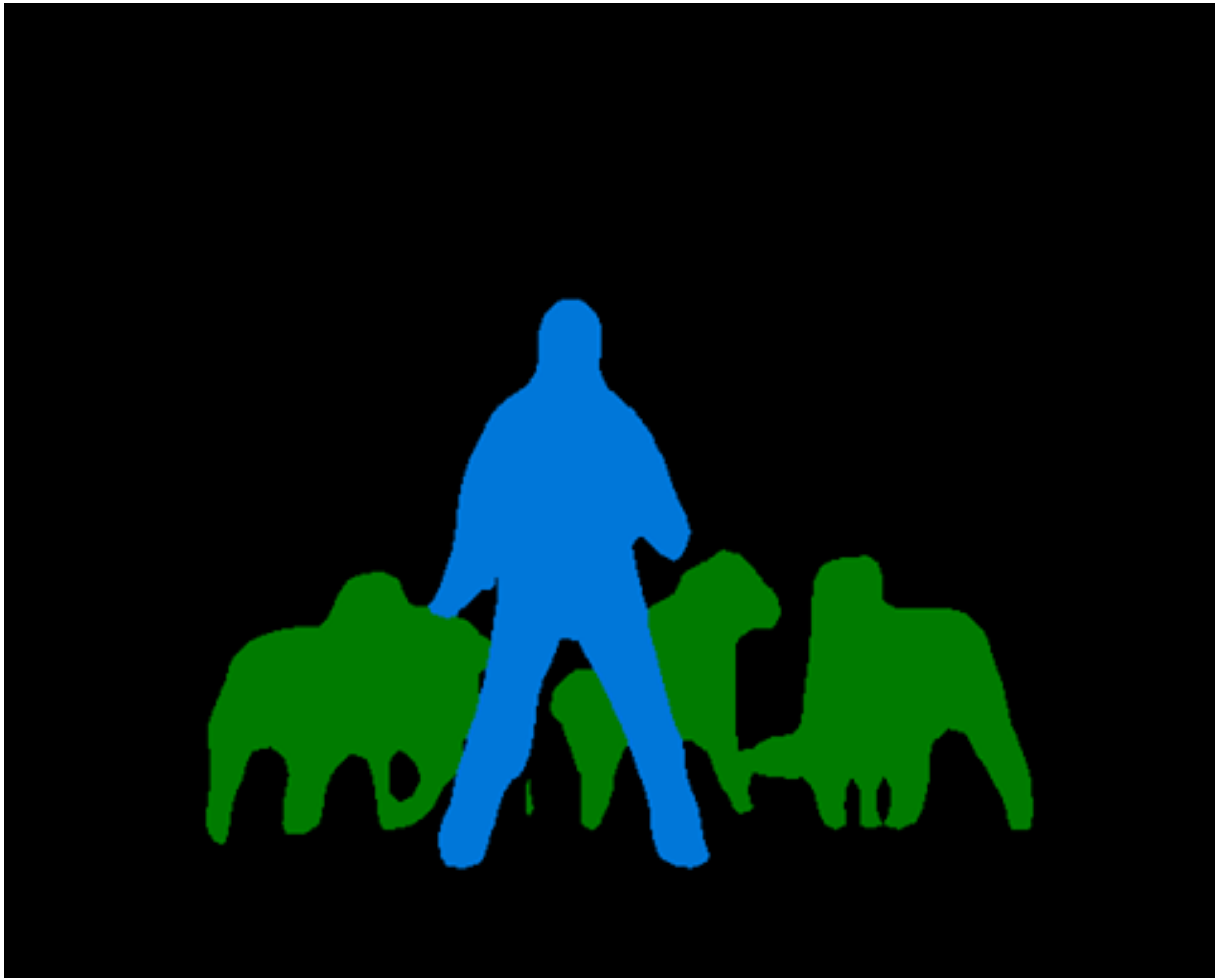
```
import torchvision.io

torchvision.io.read_video(filename,
                          start_pts=0,
                          end_pts=None)

torchvision.io.read_video_timestamps(filename)

torchvision.io.write_video(filename,
                           video_array,
                           fps,
                           video_codec='libx264',
                           options=None)
```

PRE-TRAINED MODELS




KeypointRCNN
MaskRCNN
DeepLabV3
person: 1.00
person: 1.00
sheep: 0.98
dog: 0.89

USING TORCHVISION MODELS

```
import torchvision

model = torchvision.models.detection.maskrcnn_resnet50_fpn(pretrained=True)
# set it to evaluation mode, as the model behaves differently
# during training and evaluation
model.eval()

image = PIL.Image.open('/path/to/an/image.jpg')
image_tensor = torchvision.transforms.functional.to_tensor(image)

# pass a list of (potentially different sized) tensors
# to the model, in 0-1 range. The model will take care of
# batching them together and normalizing
output = model([image_tensor])
# output is a list of dict, containing the post processed predictions
```

HANDS ON WITH TORCHVISION