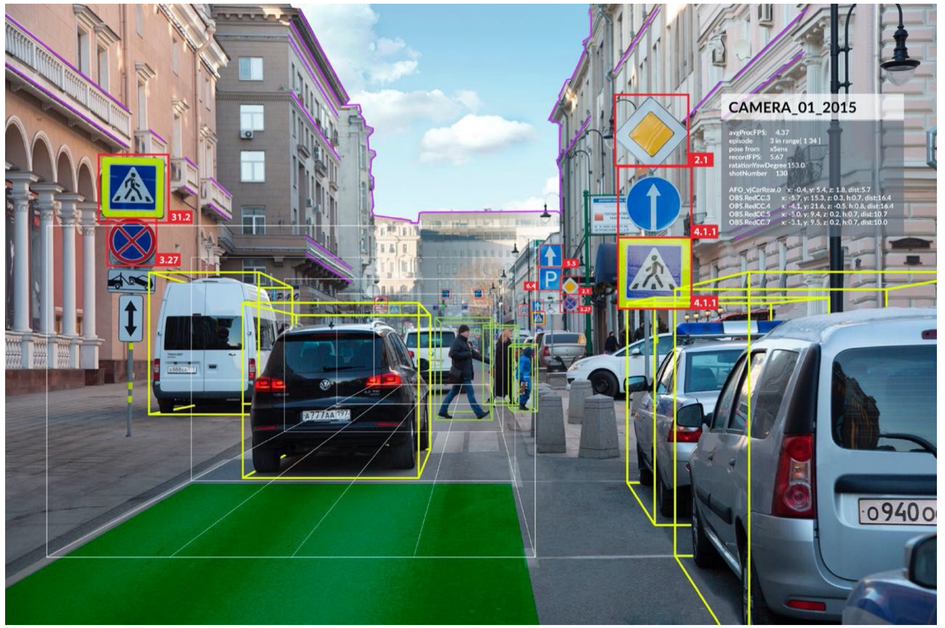
Computer Vision

AI Summer Games

Michiel Bontenbal
HvA

3 July 2024



Hogeschool van Amsterdam

HOCHSCHULE DER MEDIEN

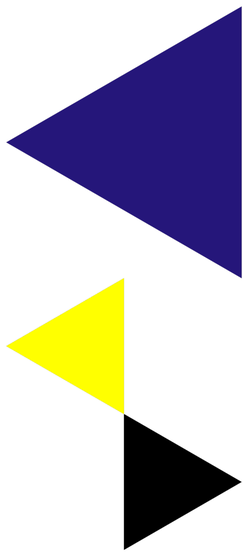
Creating Tomorrow

Agenda

Intro to Computer Vision

Classic CV: **Convolutional Neural Networks**

Modern CV: **CLIP** and **Language Vision Models**



sense maker samst erdam

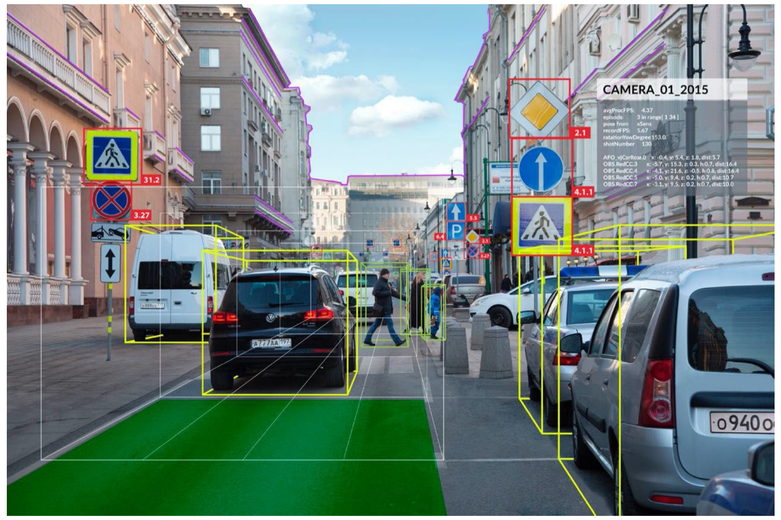
Some questions… raise your hand!

- Who has worked with:
  - Computer Vision
  - CNN
  - Vision Language Models
  - Ollama?

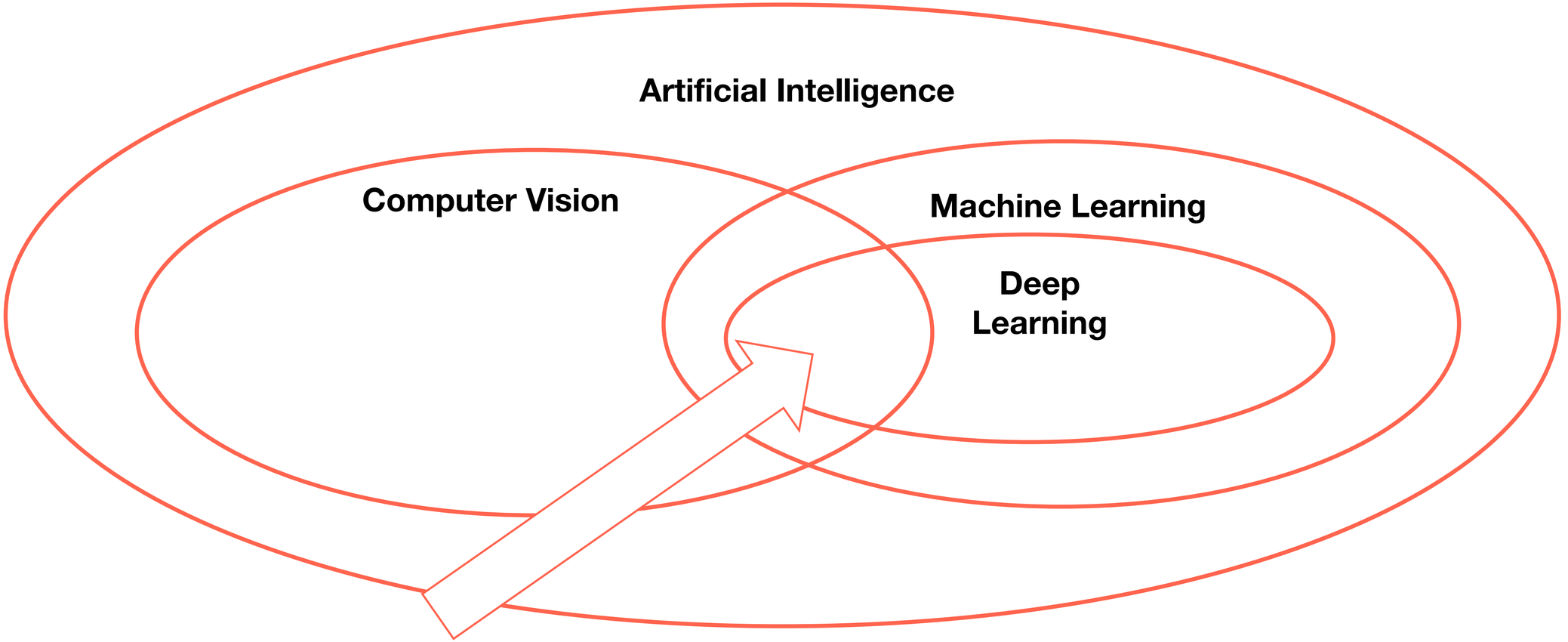
Intro to Computer Vision

What is computer vision?

- Computers + camera’s that can “see”
  - Retrieve information from picture or video
  - Take an action based on this information

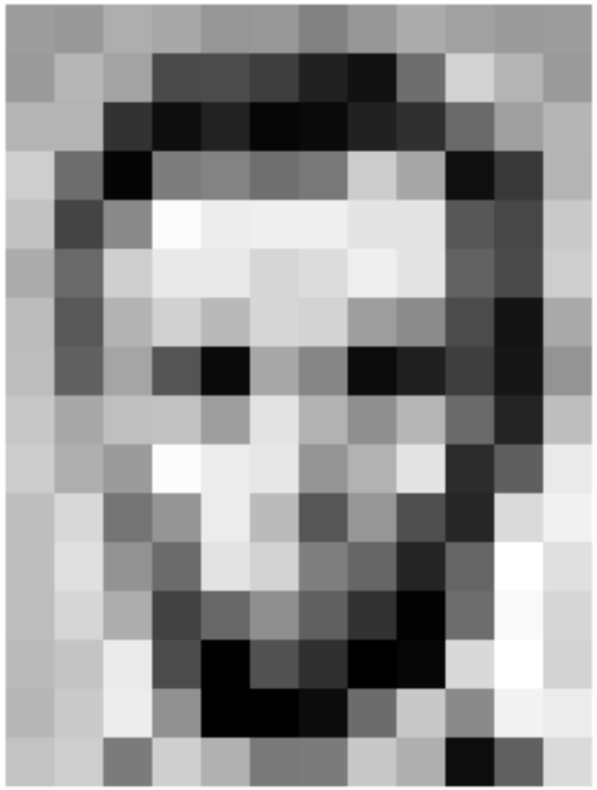


Computer Vision in AI



This lesson is about Computer Vision using Deep Learning.

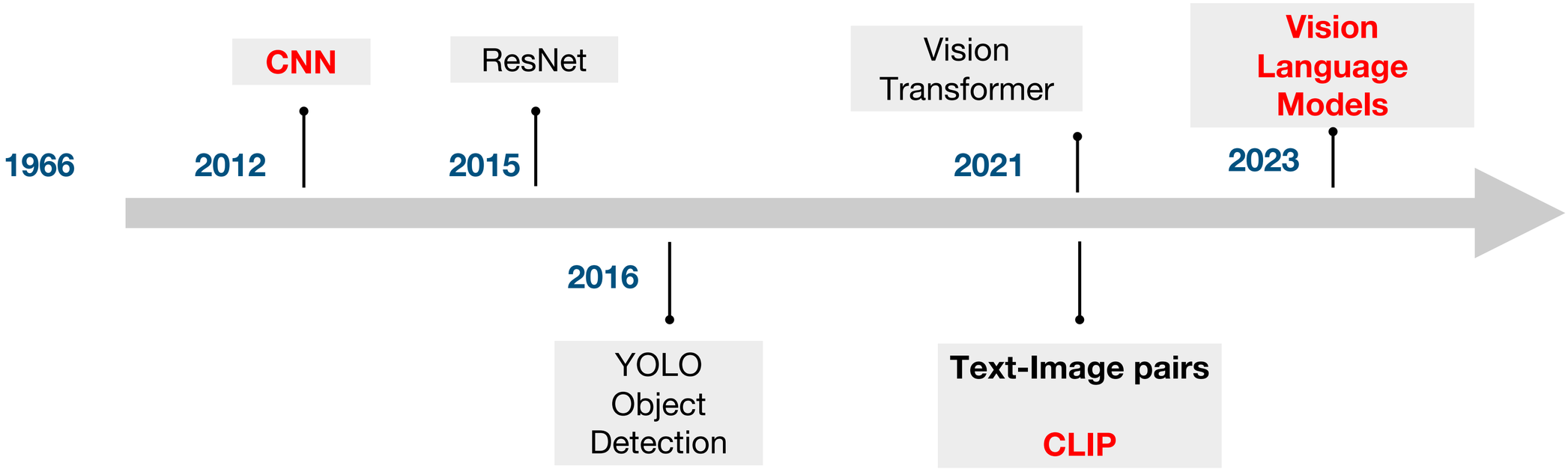
We see images, computers ‘see’ numbers



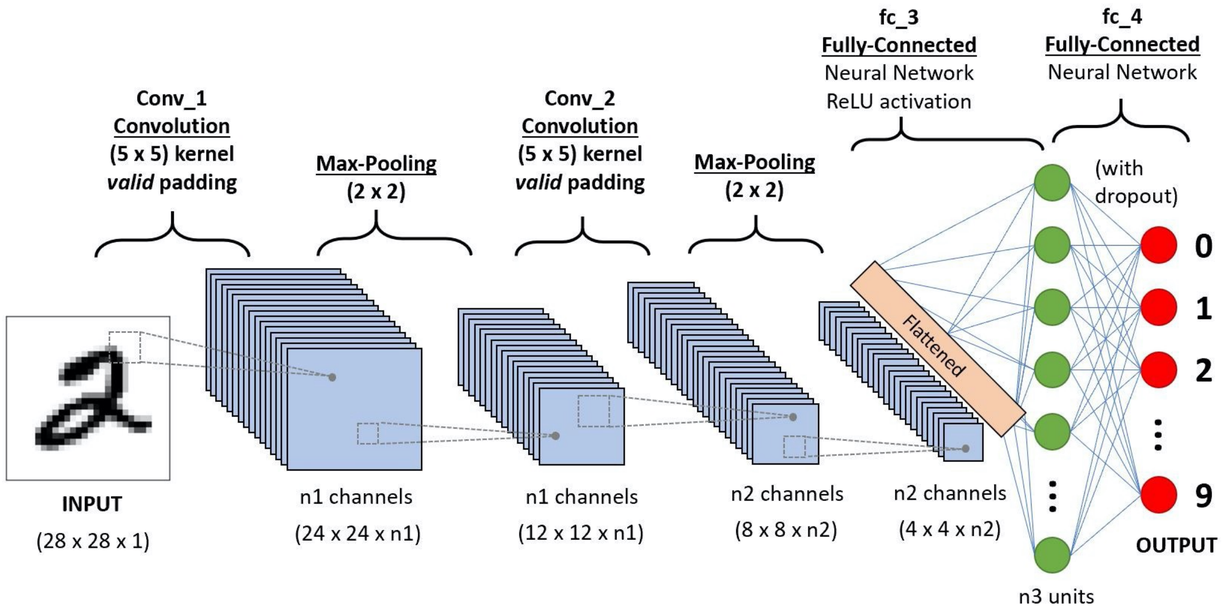
157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	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

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

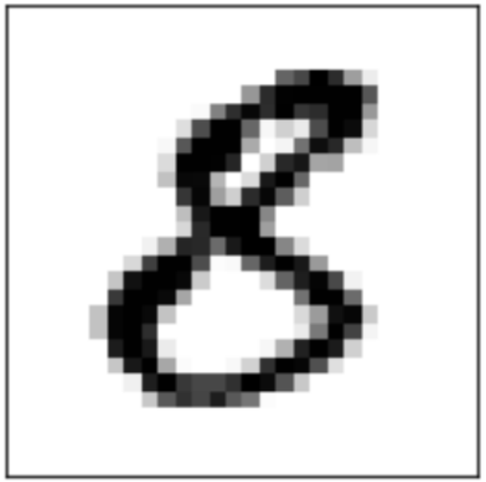
Timeline Computer Vision



Convolutional Neural Netwerks



Pattern recognition

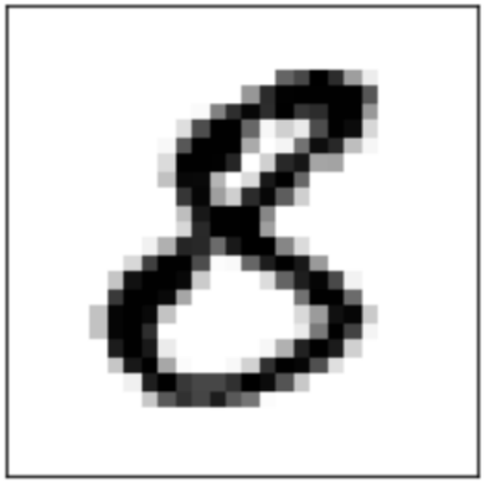


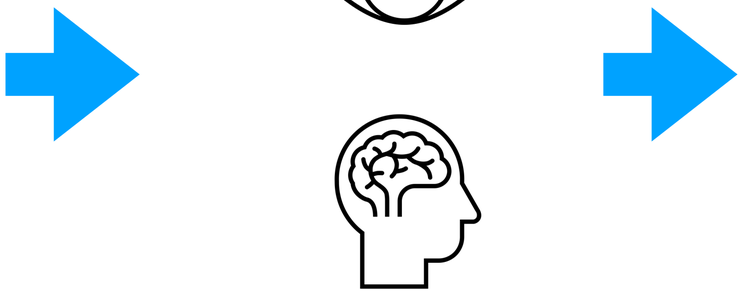
8

Convolutional Neural Network (CNN)

1. What is a Convolution?
2. What's the difference with 'normal' Neural Nets?
3. Typical layers of a CNN?
4. How can a CNN learn?

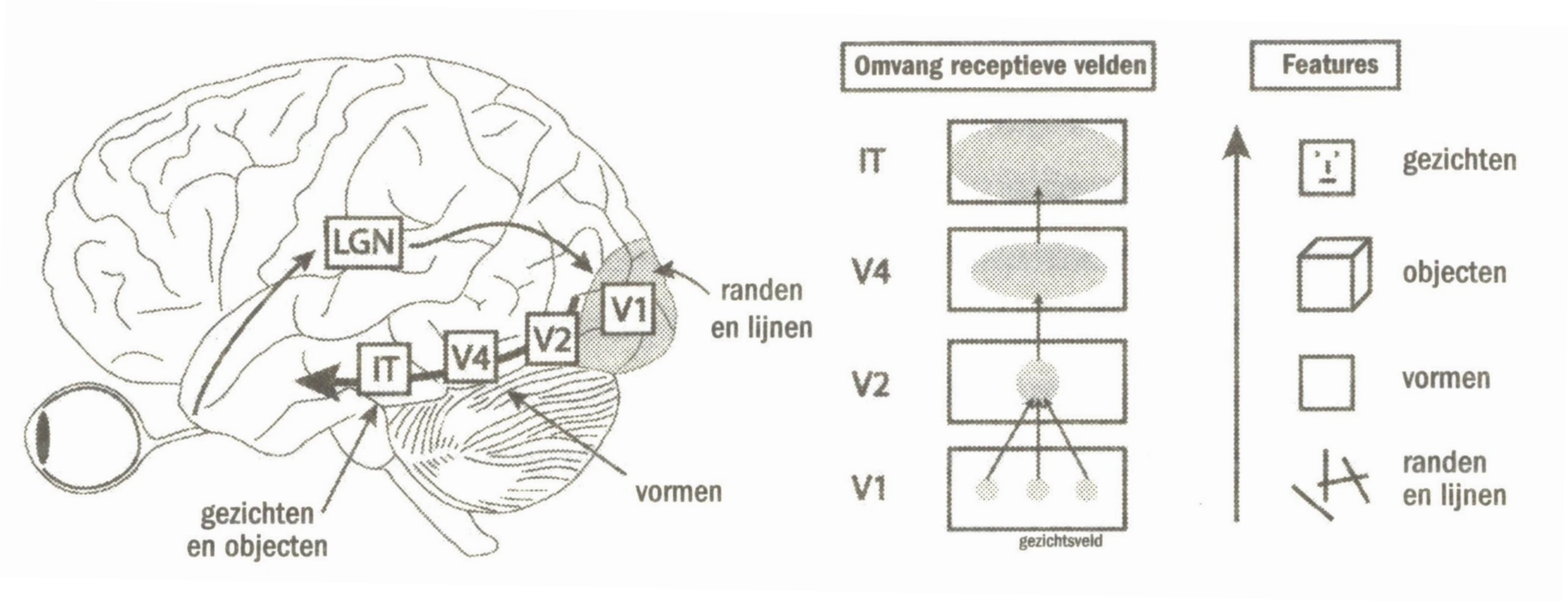
How do we as humans recognize this?





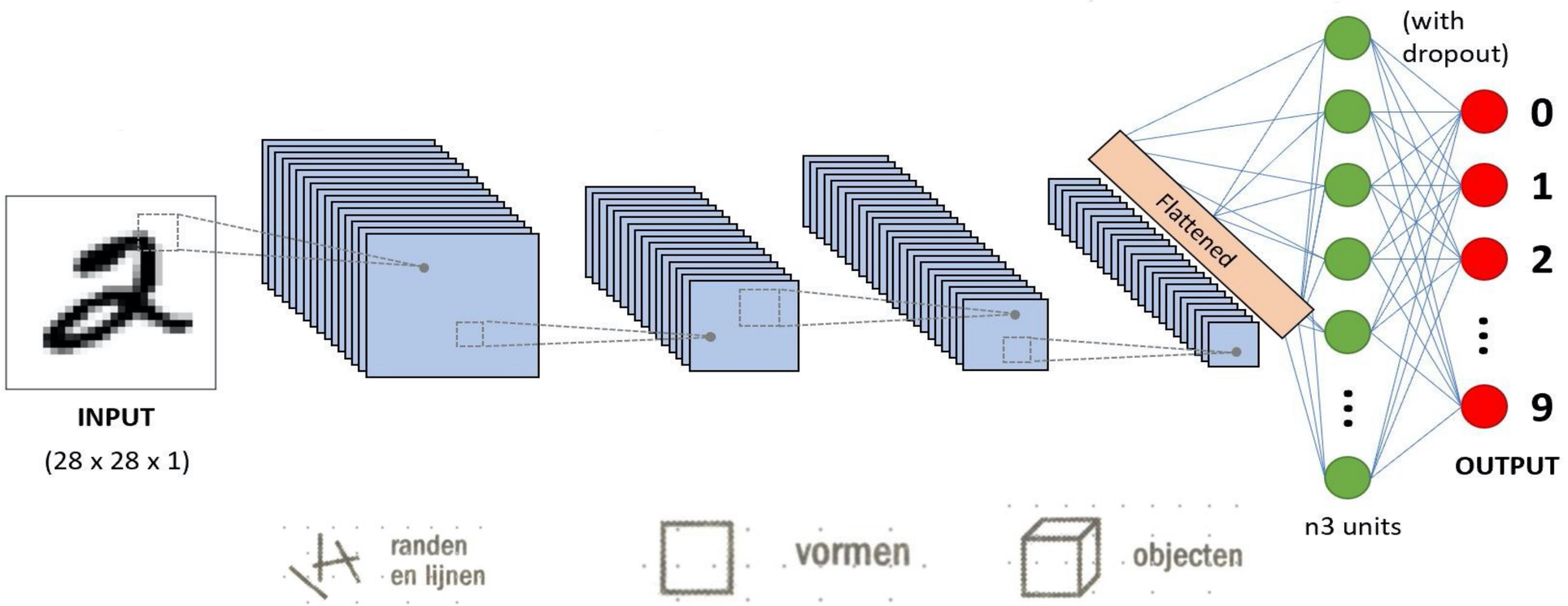
8

Hubel & Wiesel – hierarchical vision (Nobel prize ‘81)

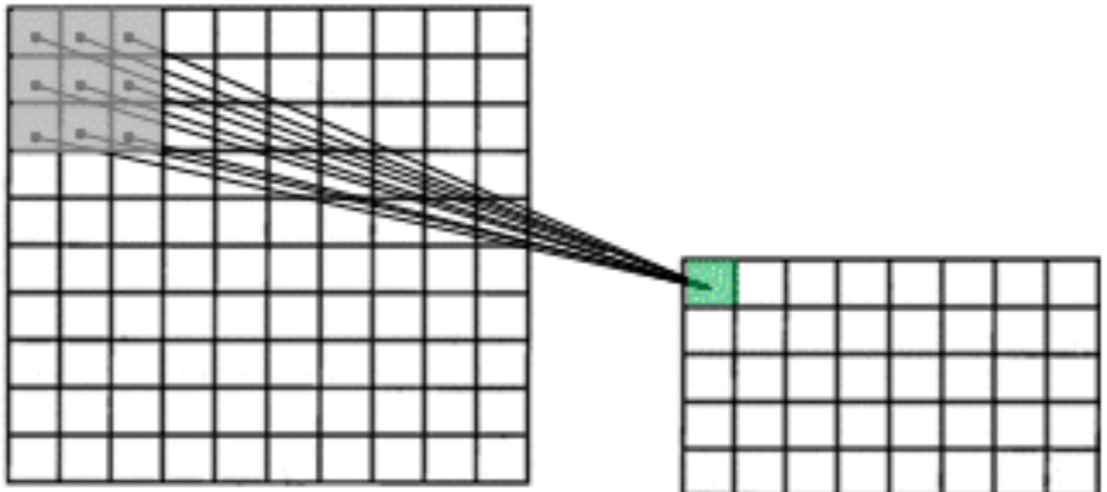


Source: A Brief history of Intelligence - Max Bennett

CNN : Feature maps and NN



Convolution2D Layer



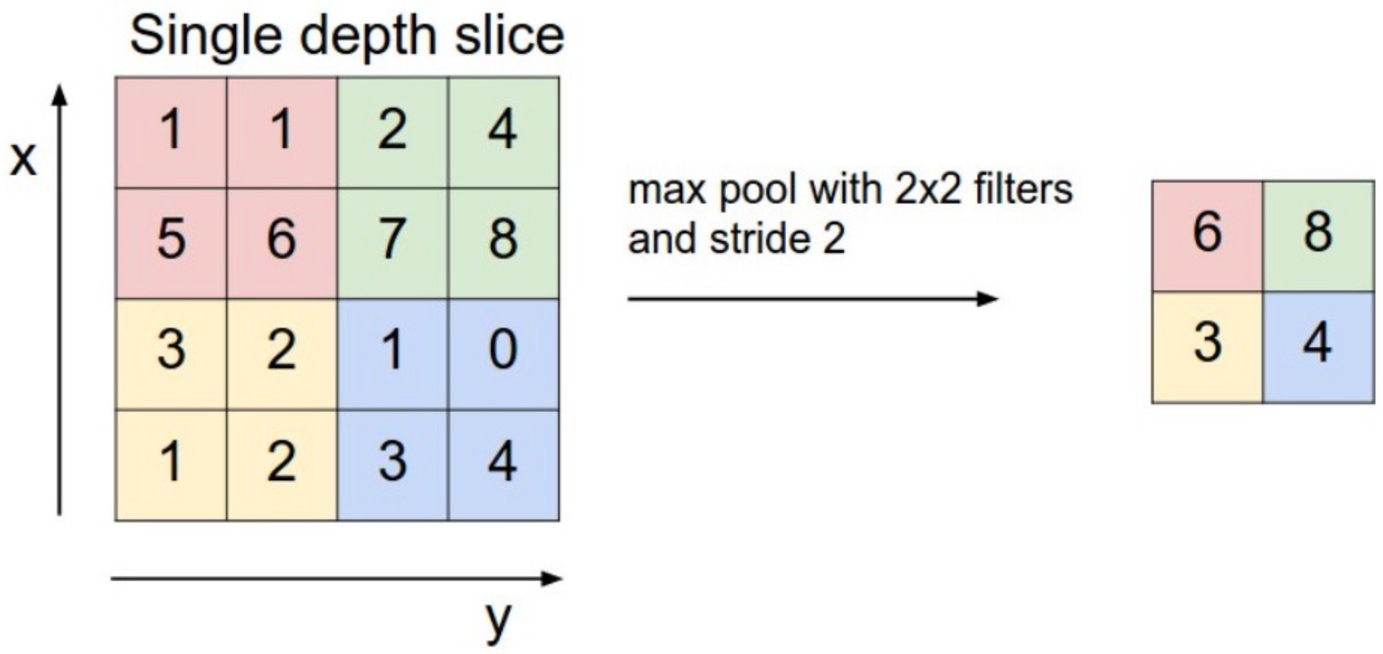
Kernel

Hidden neuron

(Ctrl)

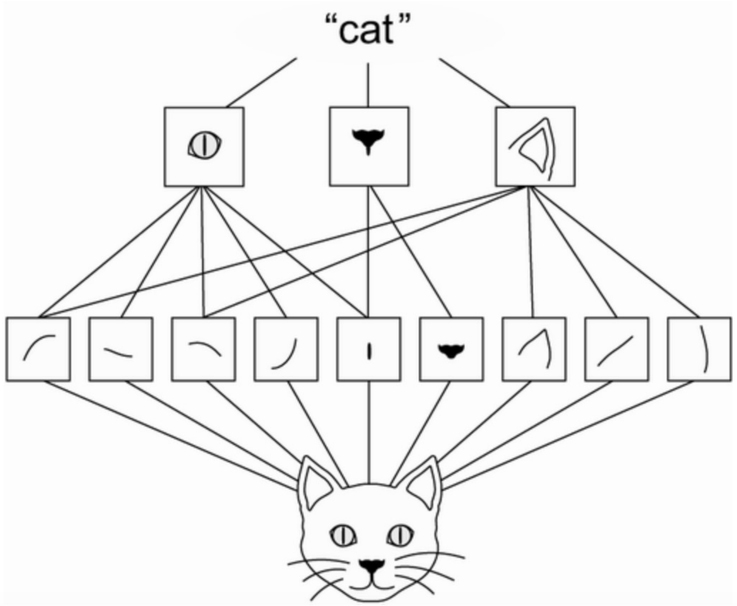
Pooling layer

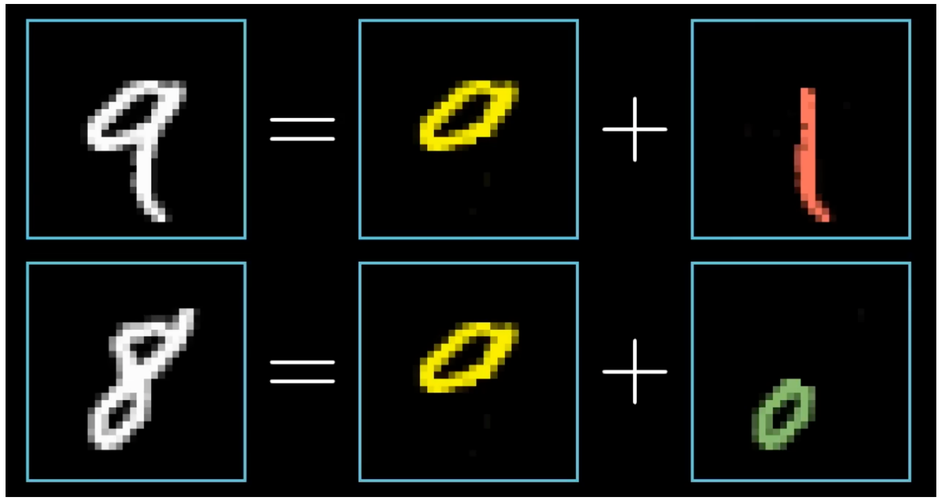
max pooling, average pooling



Source: https://cs231n.github.io/ (Stanford CNN course)

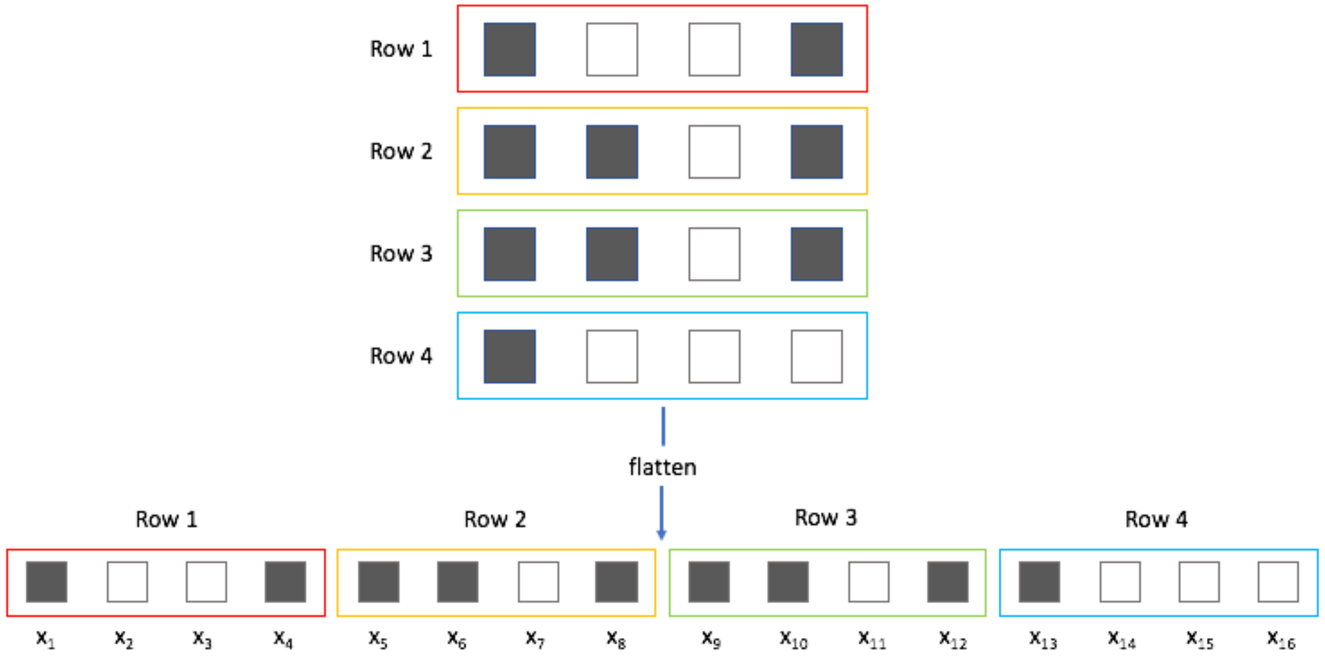
By repeating conv2d + pooling, the model can recognize larger features





Source: Deep Learning with Python, Francois Chollet, 8.1 Intro to Convnets

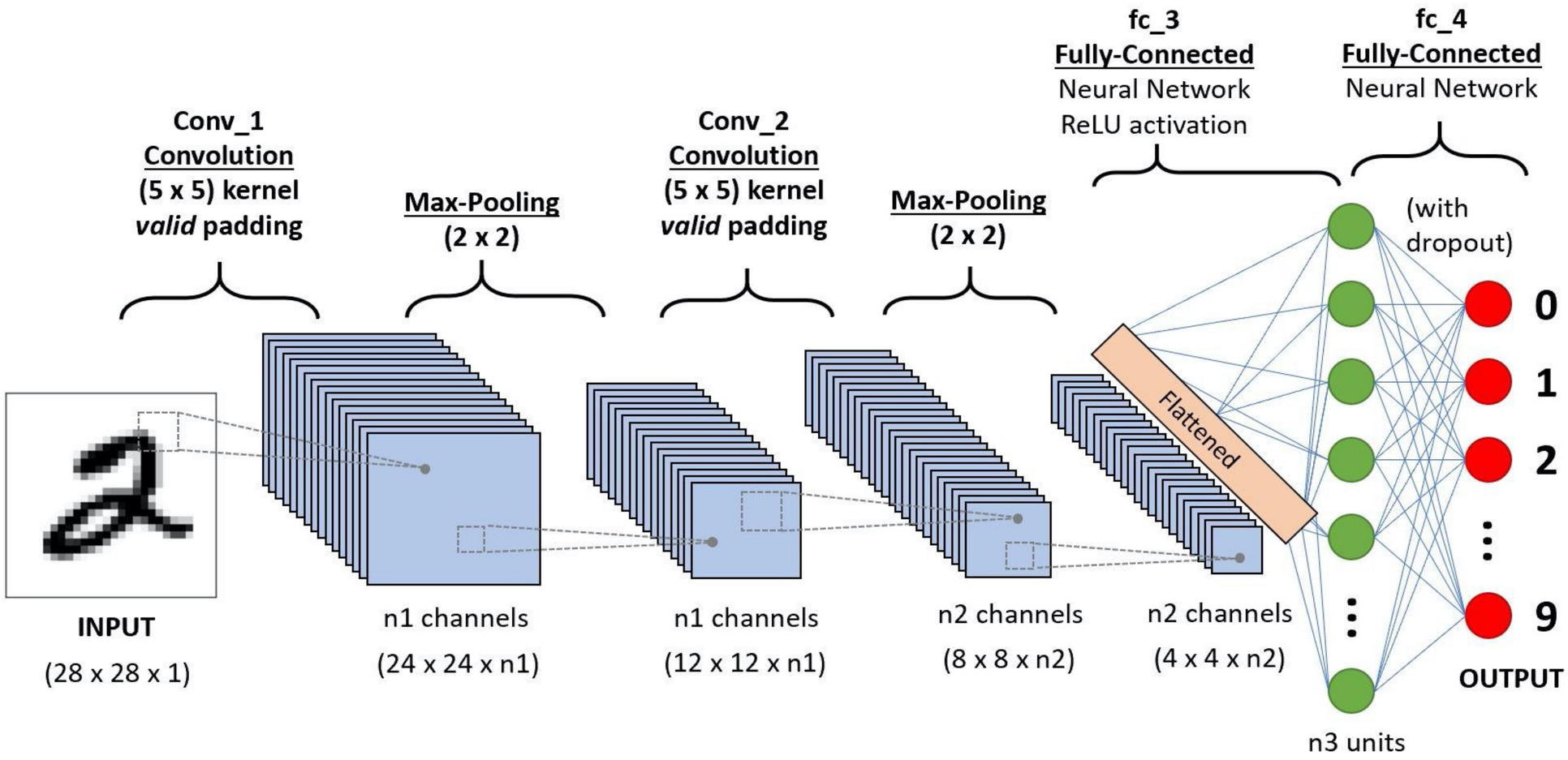
Flatten(): 2D image => 1D input layer



"flatten()"

Convert 1 pixel => 1 node of input layer.
For MNIST 28 x 28 = 784 node

Convolutional Neural Network



CNN in Keras

```
model = Sequential()

model.add(Conv2D(20, kernel_size=(4, 4),
                 activation='relu',
                 input_shape=(28,28,1)))
model.add(MaxPooling2D(pool_size=(2, 2), strides=(2,2)))

model.add(Conv2D(50, kernel_size=(5, 5), activation='sigmoid'))
model.add(MaxPooling2D(pool_size=(2, 2), strides=(2,2)))

model.add(Flatten())
model.add(Dense(500, activation='relu'))

model.add(Dense(10, activation='softmax'))
```

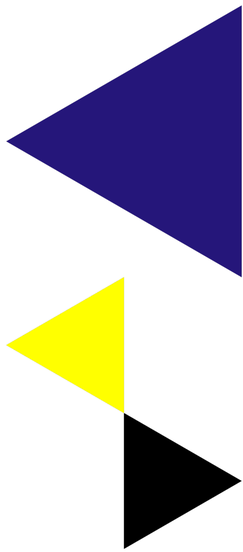
CNN in Keras

model.summary()

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 24, 24, 20)        520
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 12, 12, 20)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 8, 8, 50)          25050
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 4, 4, 50)          0
_________________________________________________________________
flatten_1 (Flatten)          (None, 800)               0
_________________________________________________________________
dense_1 (Dense)              (None, 500)               400500
_________________________________________________________________
dense_2 (Dense)              (None, 10)                5010
=================================================================
Total params: 431,080
Trainable params: 431,080
Non-trainable params: 0
_________________________________________________________________
```

Keras – www.keras.io

- Python library for deep learning.
- Developed by Francois Chollet at Google.
- Current version 3.0
- High level library
- Uses Tensorflow / JAX / Pytorch as backend



Summary CNN

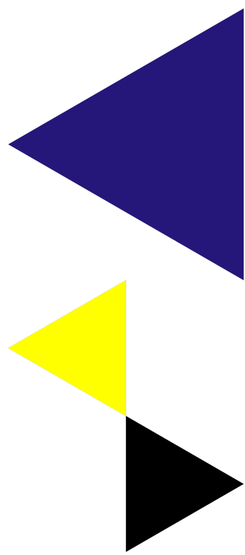
1. **What is a “convolution”?**
   - A ‘convolution kernel’ is a matrix that goes step-by-step over the image and calculates pixel values.

2. **What is difference with ‘normal’ NN’s?**
   - Conv2D + Pooling Layers

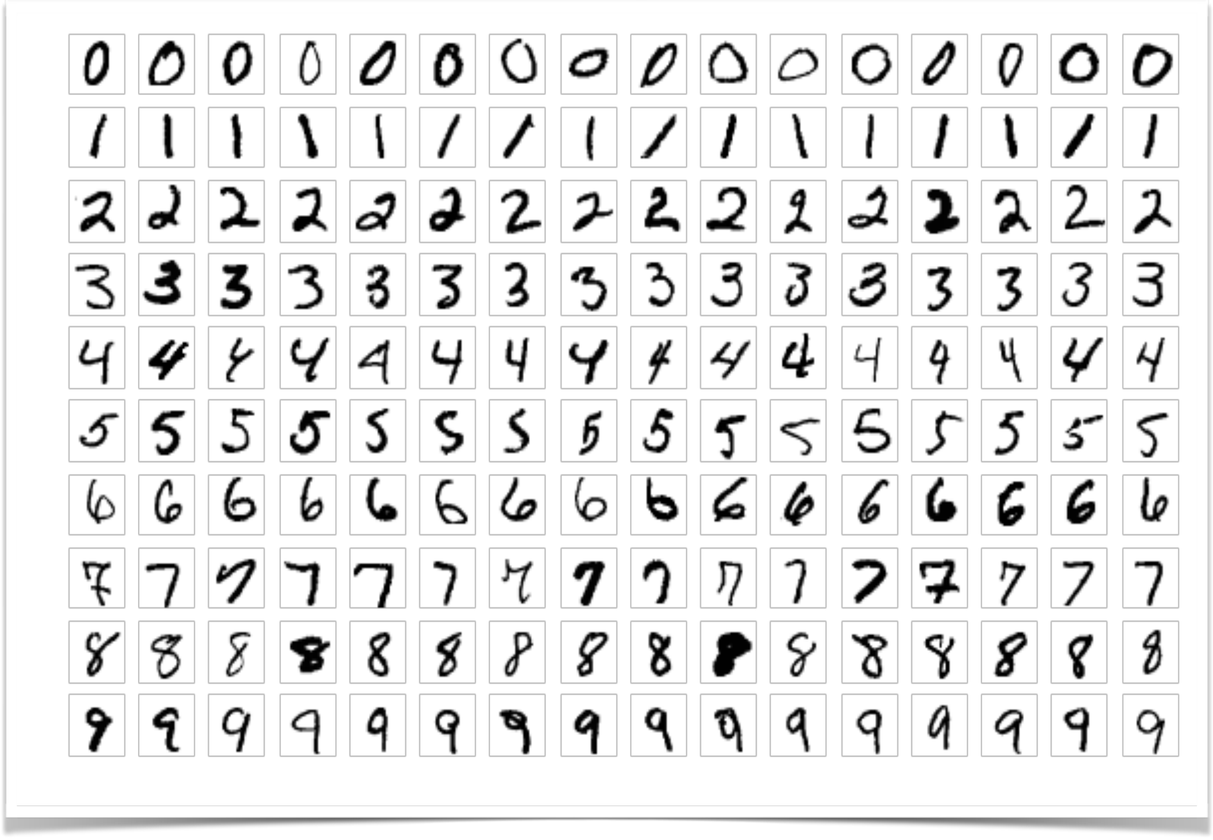
Summary CNN

3. **What are the layers of a CNN?**
   - Conv2D, Pooling, Flatten, Dense

4. **Summary: how can a CNN ‘learn’**
   - from small features to larger, more complex features than compare them to labeled ground truth.



You will use the MNIST dataset



- Handwritten digits 0 - 9
- 1994
- 70.000 images
- 28 x 28 pixels
- Yann LeCunn

Notebook CNN

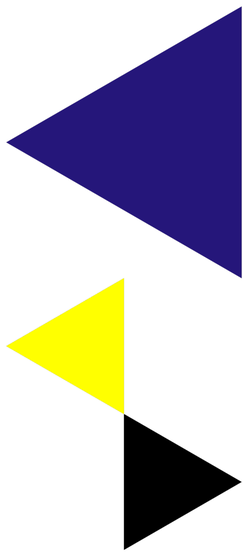
- Notebook:
  `2024_07_03_MNIST_CNN.ipynb`

- Challenge: Haribo for the highest score!

kNN	**ca. 0.91**
NN	**ca. 0.94**
CNN	**> 0.99**

World record: **0.9983**

Name	Score CNN



Read more

Book

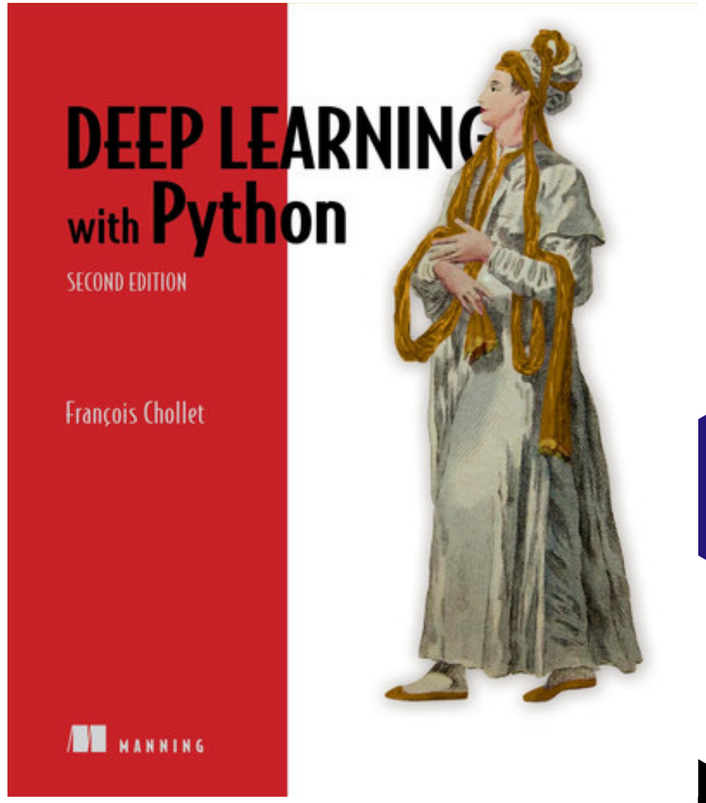
‘Deep learning with python’ – Francois Chollet
https://learning-oreilly-com.rps.hva.nl/library/view/deep-learning-with/9781617296864/

YouTube

But what is a neural network? https://www.youtube.com/watch?v=aircAruvnKk&

www.keras.io

Tutorials & code examples op www.keras.io



Bonus: Can cats see lines from birth or do they develop their vision by ‘training’ their brain?

- https://youtu.be/QzkMo45pcUo?t=198

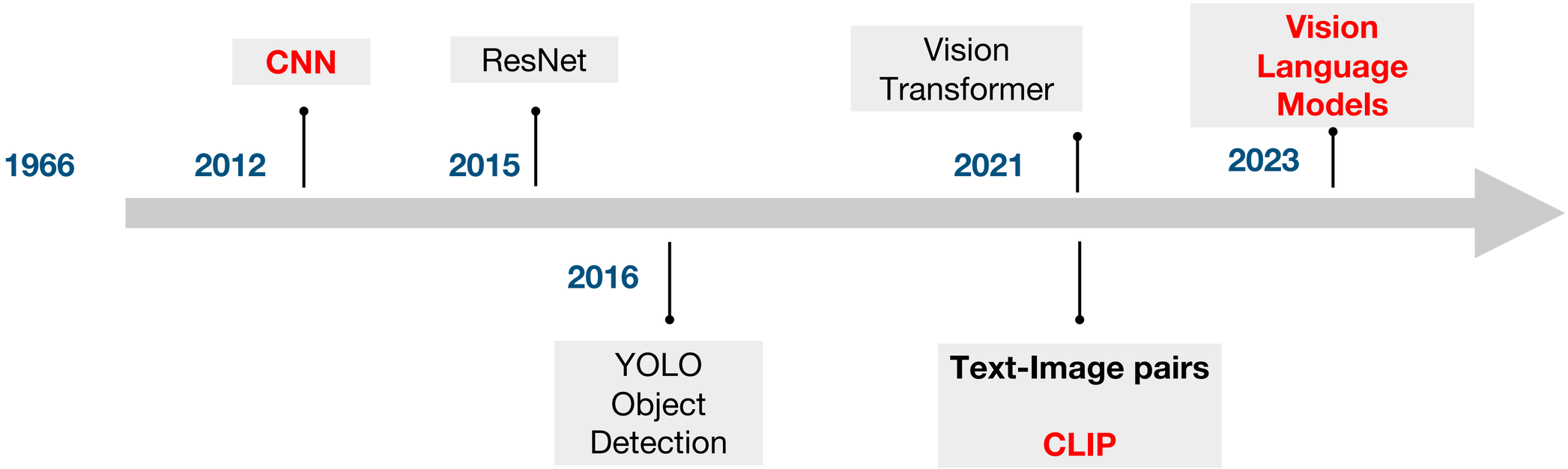
Modern CV

Text and images

- CLIP

- Vision Language Models

Timeline Computer Vision



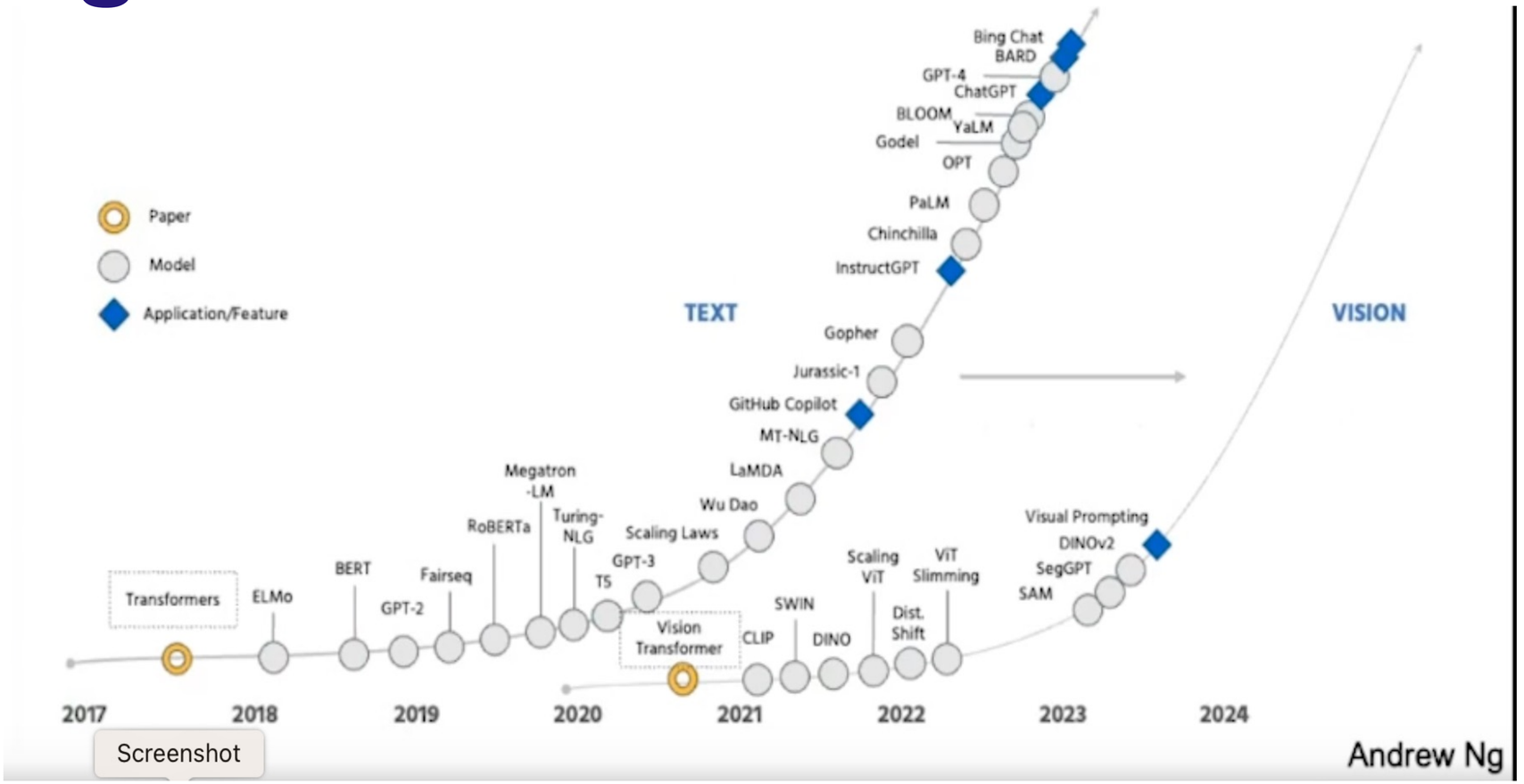
What’s happening here?

???

“What is unusual about this image?”

“The image shows a person ironing clothes on the back of a moving vehicle,…”

The ChatGPT revolution is coming to vision!



Screenshot

Large Vision Models =. Large Multimodal Models

How can computers find this relationship?

‘Paris’

‘vector embeddings’

Read more: https://www.geeksforgeeks.org/what-are-embeddings-in-machine-learning/

We use an ‘embedding model’ and compare the vectors. Vectors capture the meaning.

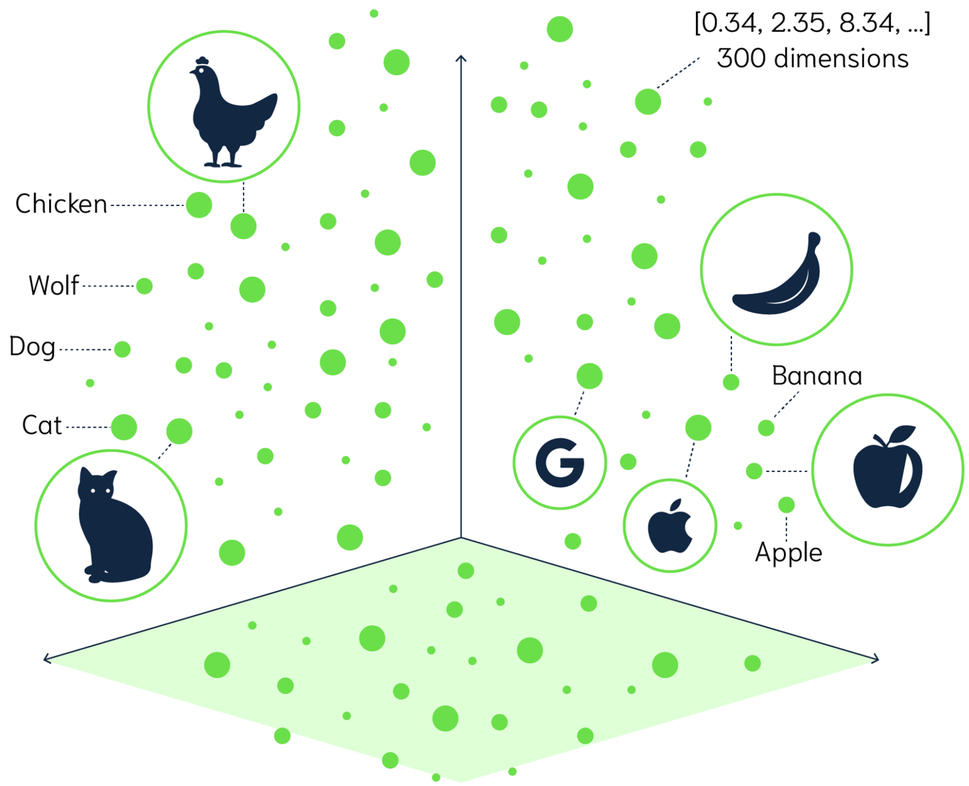
‘Paris’

[0.5, 0.3, …, 0.7]

[0.4, 0.3, …, 0.6]

Compare vectors

Texts and images in vector space



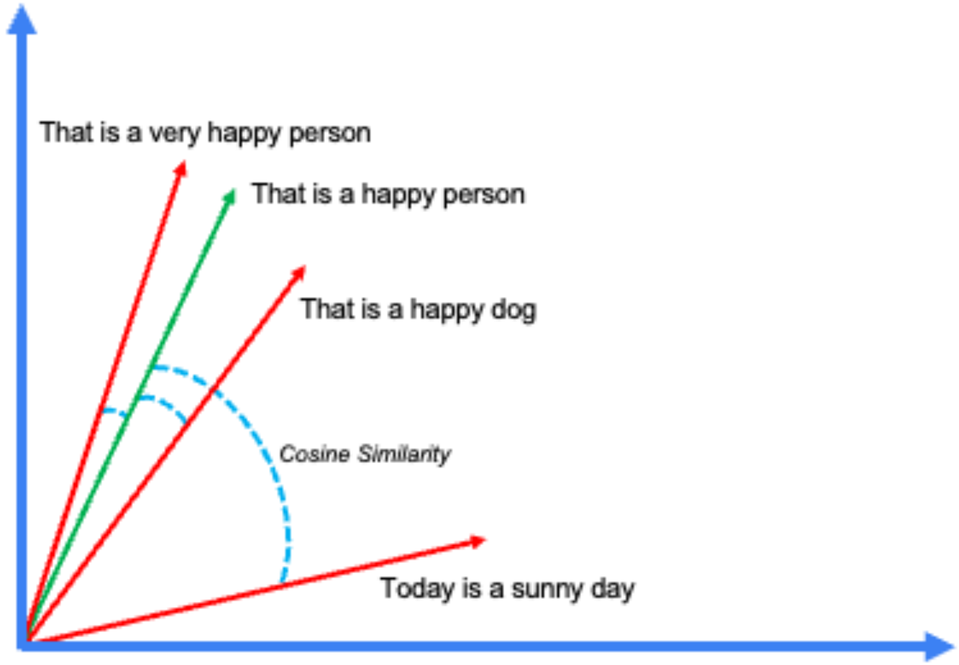
Words and images with same meaning are close in vector space.

Compare vector embeddings to find relationships

- Comparing vector embeddings is known as ‘similarity search’

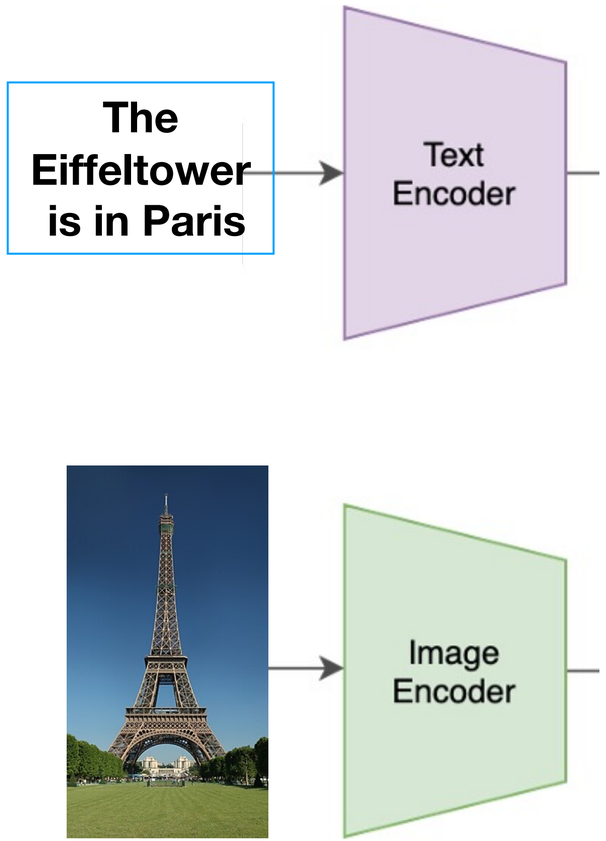
- Most often we use ‘cosine similarity’.
  - See notebook.

- It gives meaning to search – not just ‘strings’.



We will use OpenAI's CLIP with text-image pairs

Contrastive Language-Image Pre-training



- Pairs of captions with images
- Trained on 400 million pairs. Published in 2021.

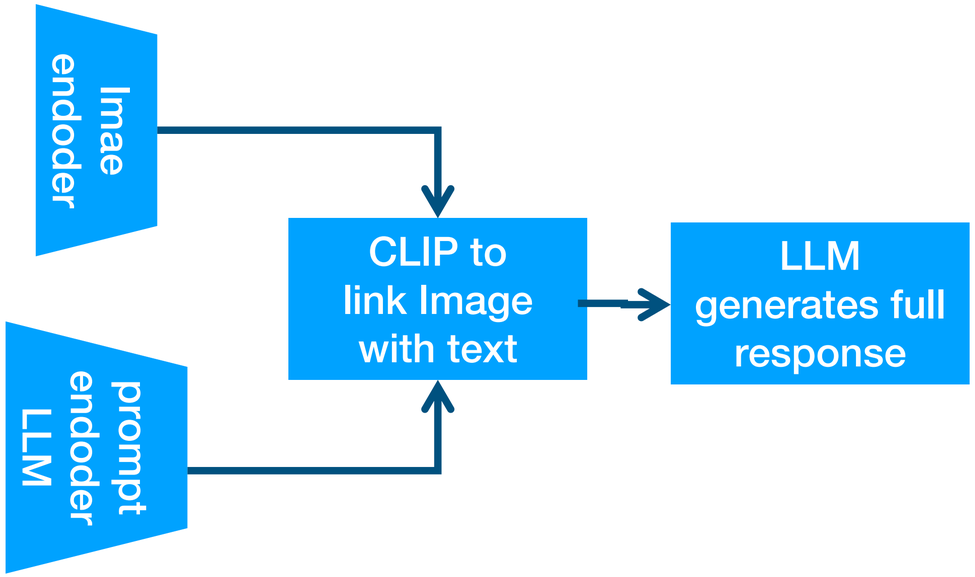
- Use CLIP to describe images => '**image2text**'
- Dall-e is the reverse => **text2image**

Sources:

- https://github.com/OpenAI/CLIP
- https://openai.com/research/clip

LLaVA: Vision Language Model

Combines CLIP with a Large Language Model

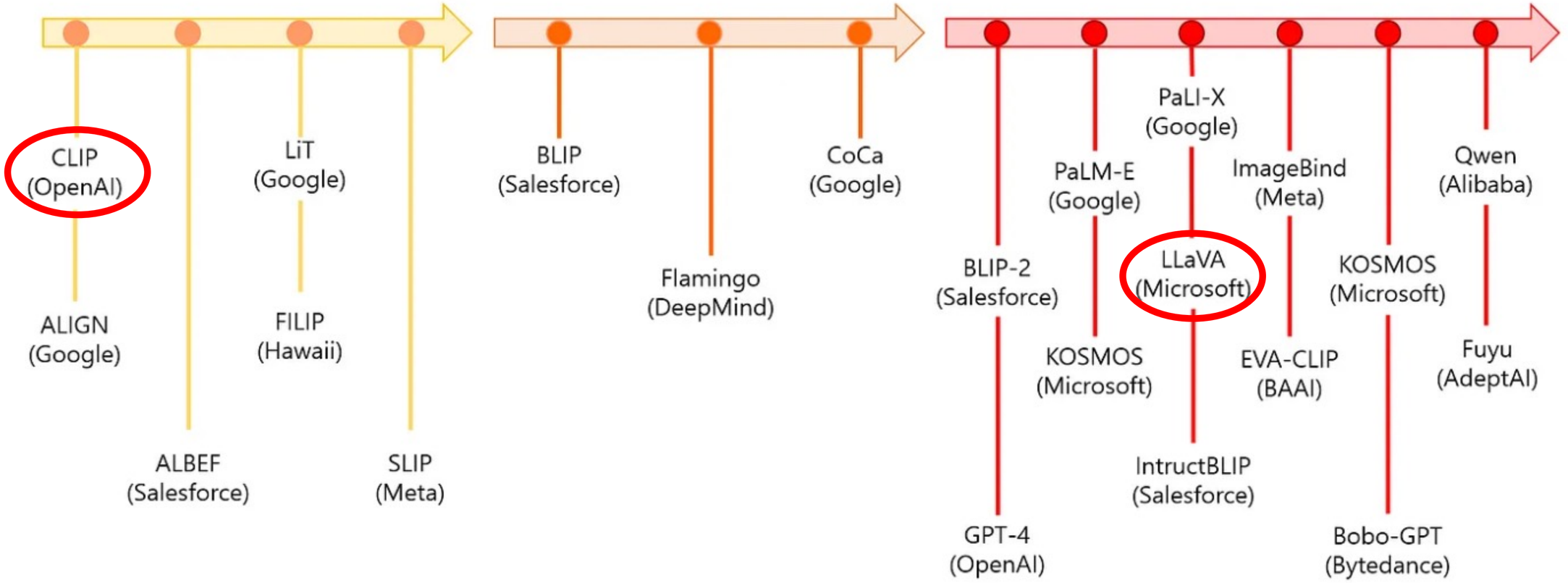


“What is unusual about this image?”

“The image shows a person ironing clothes on the back of a moving vehicle,…”

This is also called a ‘neck & head architecture’.

Vision Language Models



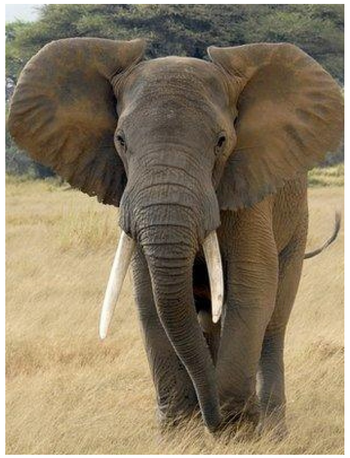
Computer Vision challenges

Describe screenshots

Security webcam

`ollama_llava_challenges.ipynb`

Find similar images (CLIP)



`Find_similar_images_CLIP.ipynb`

sense
maker
samst
erdam

ollama

- Download and install via www.ollama.com (Mac / Windows / Linux)

- Ollama is a tool that you can use to run models locally.