Deep Learning Introduction and Application

Ing. Simone Calderara

AImagelab Computer Vision, Pattern Recognition & Machine Learning

Dief University Of Modena and Reggio Emilia

Contact: simone.calderara@unimore.it

AI Academy

AImage Lab
University of Modena and Reggio Emilia

A Glossary of Artificial-Intelligence Terms

ARTIFICIAL INTELLIGENCE

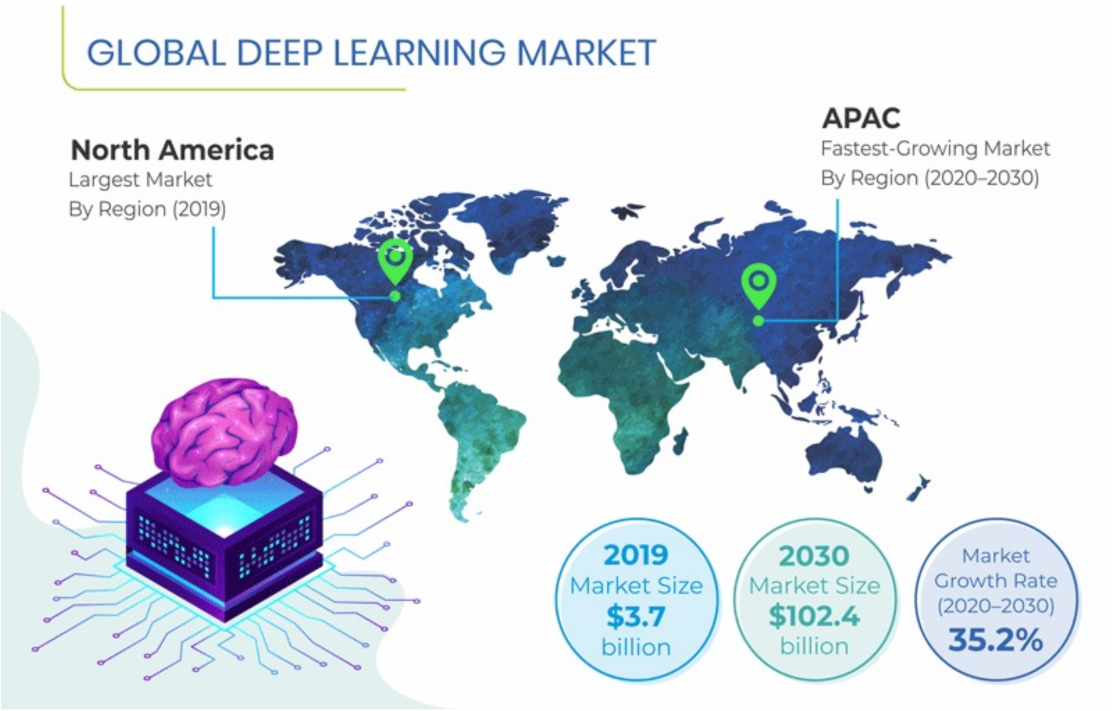
AI is the broadest term, applying to any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision trees, and machine learning (including deep learning).

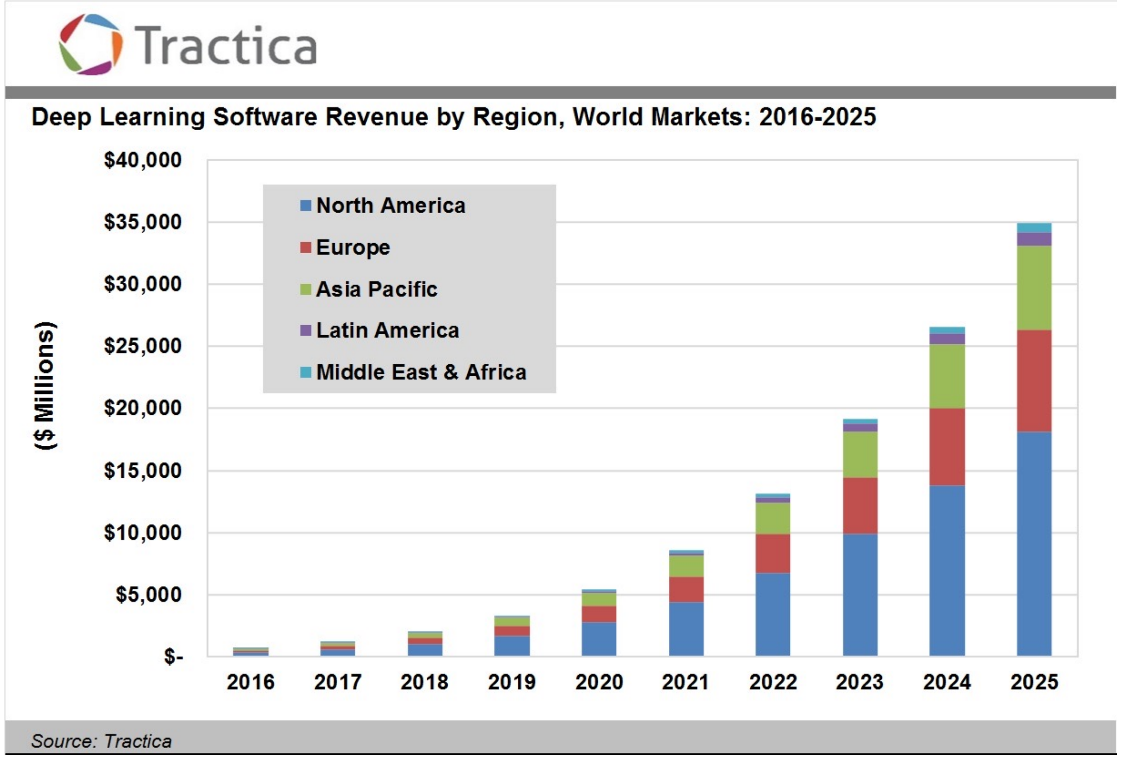
MACHINE LEARNING

The subset of AI that includes abstruse statistical techniques that enable machines to improve at tasks with experience. The category includes deep learning.

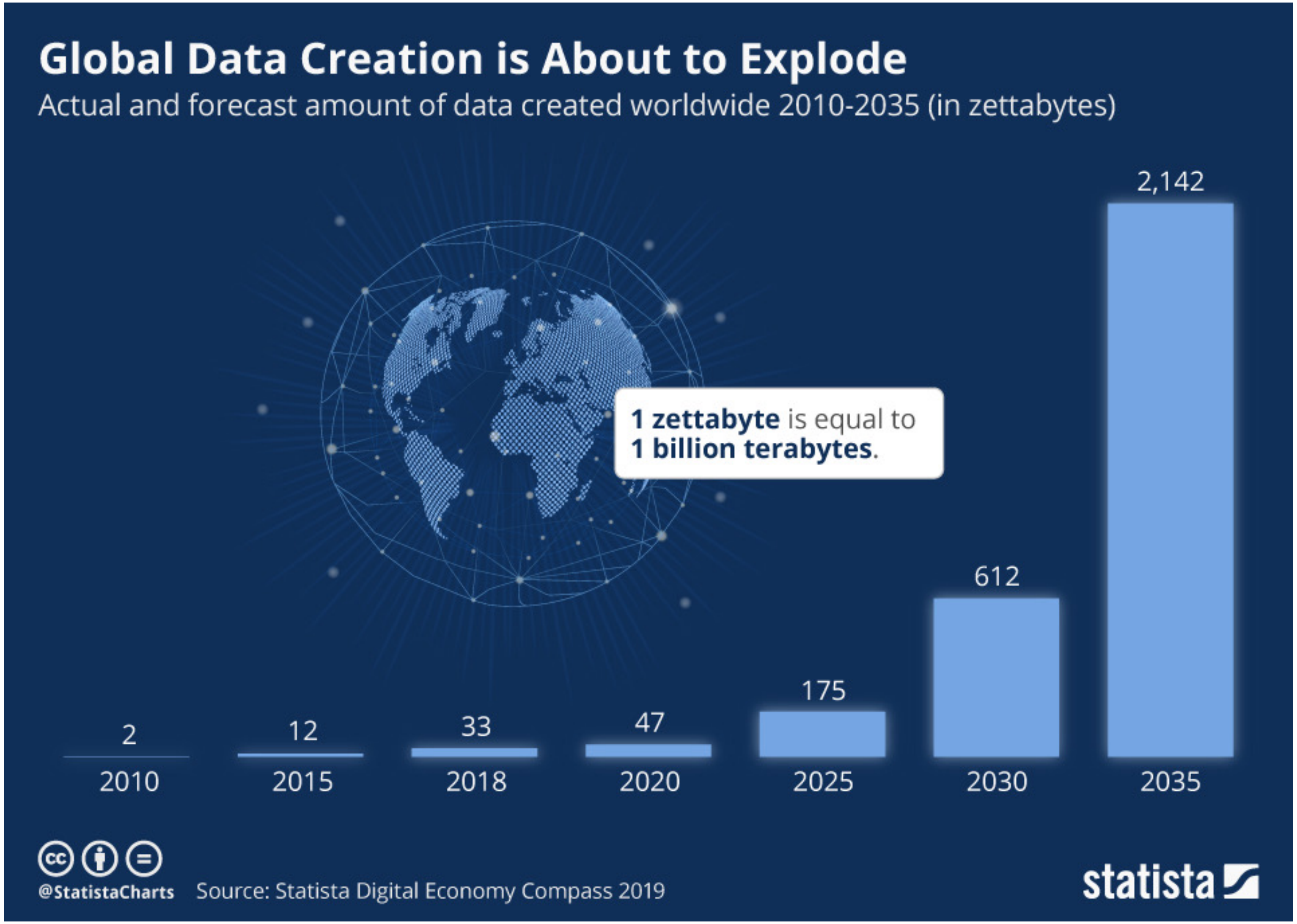
DEEP LEARNING

The subset of machine learning composed of algorithms that permit software to train itself to perform tasks, like speech and image recognition, by exposing **multilayered neural networks to vast** amounts of data.

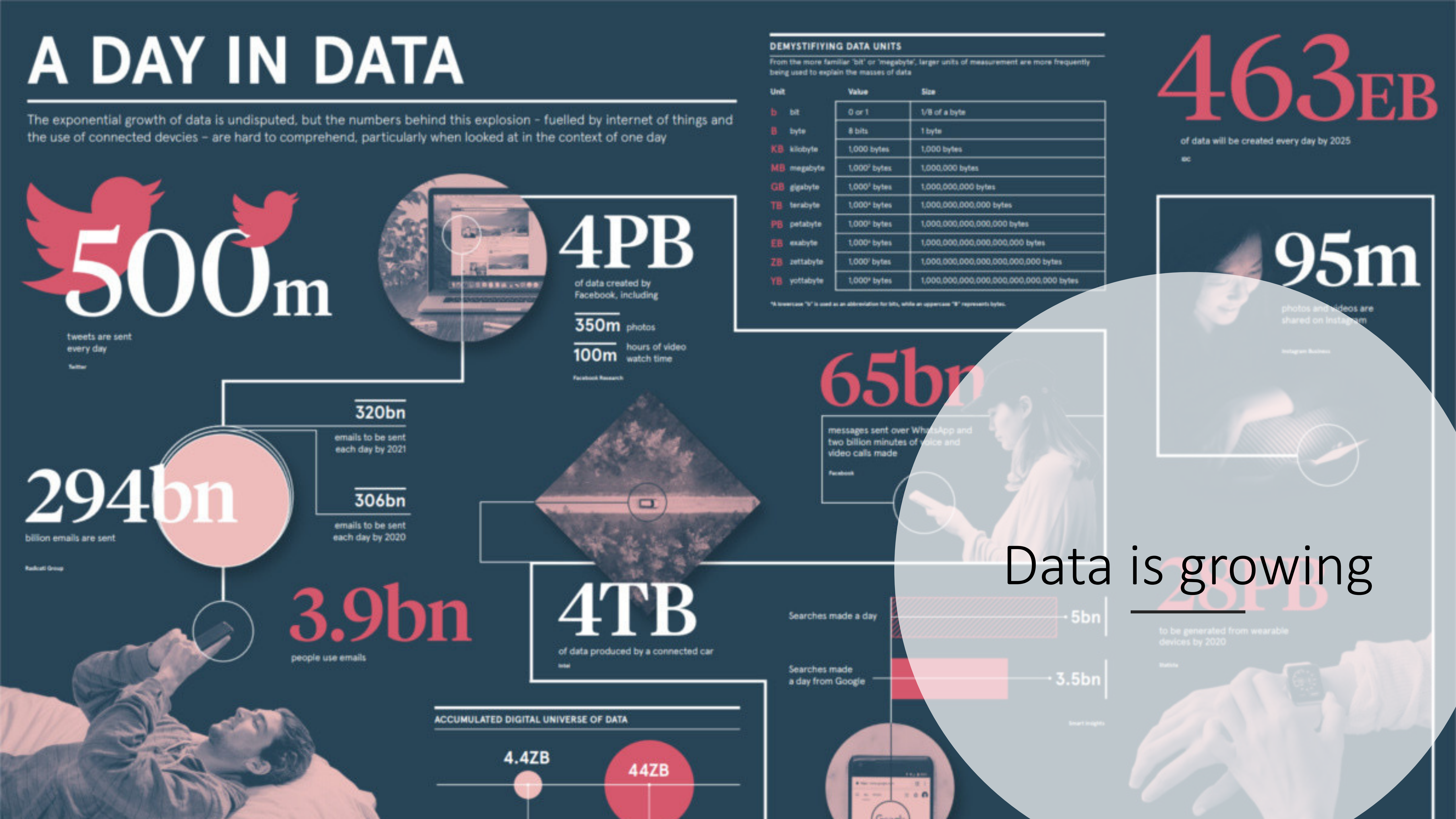

GLOBAL DEEP LEARNING MARKET
North America
Largest Market
By Region (2019)
APAC
Fastest-Growing Market
By Region (2020–2030)
2019
Market Size
$3.7
billion
2030
Market Size
$102.4
billion
Market
Growth Rate
(2020–2030)
35.2%


Tractica
Deep Learning Software Revenue by Region, World Markets: 2016-2025
($ Millions)
$40,000
$35,000
$30,000
$25,000
$20,000
$15,000
$10,000
$5,000
$-
North America
Europe
Asia Pacific
Latin America
Middle East & Africa
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Source: Tractica

Motivation-> Data is changing



Data is growing



Unit		Value	Size
b	bit	0 or 1	1/8 of a byte
B	byte	8 bits	1 byte
KB	kilobyte	1,000 bytes	1,000 bytes
MB	megabyte	$1{,}000^2$ bytes	1,000,000 bytes
GB	gigabyte	$1{,}000^3$ bytes	1,000,000,000 bytes
TB	terabyte	$1{,}000^4$ bytes	1,000,000,000,000 bytes
PB	petabyte	$1{,}000^5$ bytes	1,000,000,000,000,000 bytes
EB	exabyte	$1{,}000^6$ bytes	1,000,000,000,000,000,000 bytes
ZB	zettabyte	$1{,}000^7$ bytes	1,000,000,000,000,000,000,000 bytes
YB	yottabyte	$1{,}000^8$ bytes	1,000,000,000,000,000,000,000,000 bytes

*A lowercase "b" is used as an abbreviation for bits, while an uppercase "B" represents bytes.

Investments in the Deep Learning Trend

The Most Common Use Cases of AI

Estimated % of AI spending by use case in 2025

Use Case	% of Spending
Algorithmic financial trading	17%
Image recogntion and tagging	16%
Patient data processing	15%
Predictive maintenance	10%
Content distribution on social media	8%
Text query of images	8%
Automated geophysical feature detection	7%
Object identification and tracking	7%
Object detection / classification	6%
Contract analysis	6%

Source: Tractica Research's spending estimates for Top 10 AI use cases in 2025

From the Past to the future

- 2008-2013 Years of theoretical studies and hardware production
- 2016->……… **Time to bring out the application**
  - “Google and Movidius who have teamed up to increase adoption (of deep learning technology) within mobile devices.”
  - Google changed the «Page Rank» algorithm with «Rank Brain» Deep learning based
  - Facebook «face recognition» is deep learning based
  - Google and Apple cars use DL to drive autonomous vehicles
  - Toyota is spending $1 billion on AI in Silicon Valley for autonomous cars
  - GPT3 is the current standard de-facto for text analysis

What is learning and a learning machine

Types of Learning

Supervised (inductive) learning	Training data includes desired outputs
Unsupervised learning	Training data does not include desired outputs
Semi-supervised learning	Training data includes a few desired outputs
Reinforcement learning	Rewards from sequence of actions

Supervised vs. unsupervised Learning

Supervised learning: classification is seen as supervised learning from examples.

Supervision: The data (observations, measurements, etc.) are labeled with pre-defined classes. It is like that a "teacher" gives the classes (supervision).

Test data are classified into these classes too.

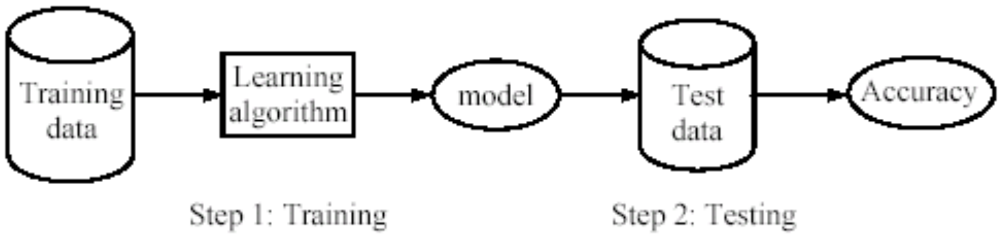
Unsupervised learning (clustering)

Class labels of the data are unknown

Given a set of data, the task is to establish the existence of classes or clusters in the data

Supervised learning process: two steps

- Learning (training): Learn a model using the training data
- Testing: Test the model using unseen test data to assess the model accuracy



$$Accuracy = \frac{\text{Number of correct classifications}}{\text{Total number of test cases}},$$

Classification measures

- Accuracy is **only one measure** (error = 1-accuracy).
- Accuracy is **not suitable** in some applications.
- In text mining, we may only be interested in the documents of a particular topic, which are only a small portion of a big document collection.
- In classification involving skewed or highly imbalanced data, e.g., network intrusion and financial fraud detections, we are interested only in the minority class.
  - High accuracy does not mean any intrusion is detected.
  - E.g., 1% intrusion. Achieve 99% accuracy by doing nothing.
- The class of interest is commonly called the positive class, and the rest negative classes.

What do we mean by learning?

- Given
  - a data set *D*,
  - a task *T*, and
  - a performance measure *M*,

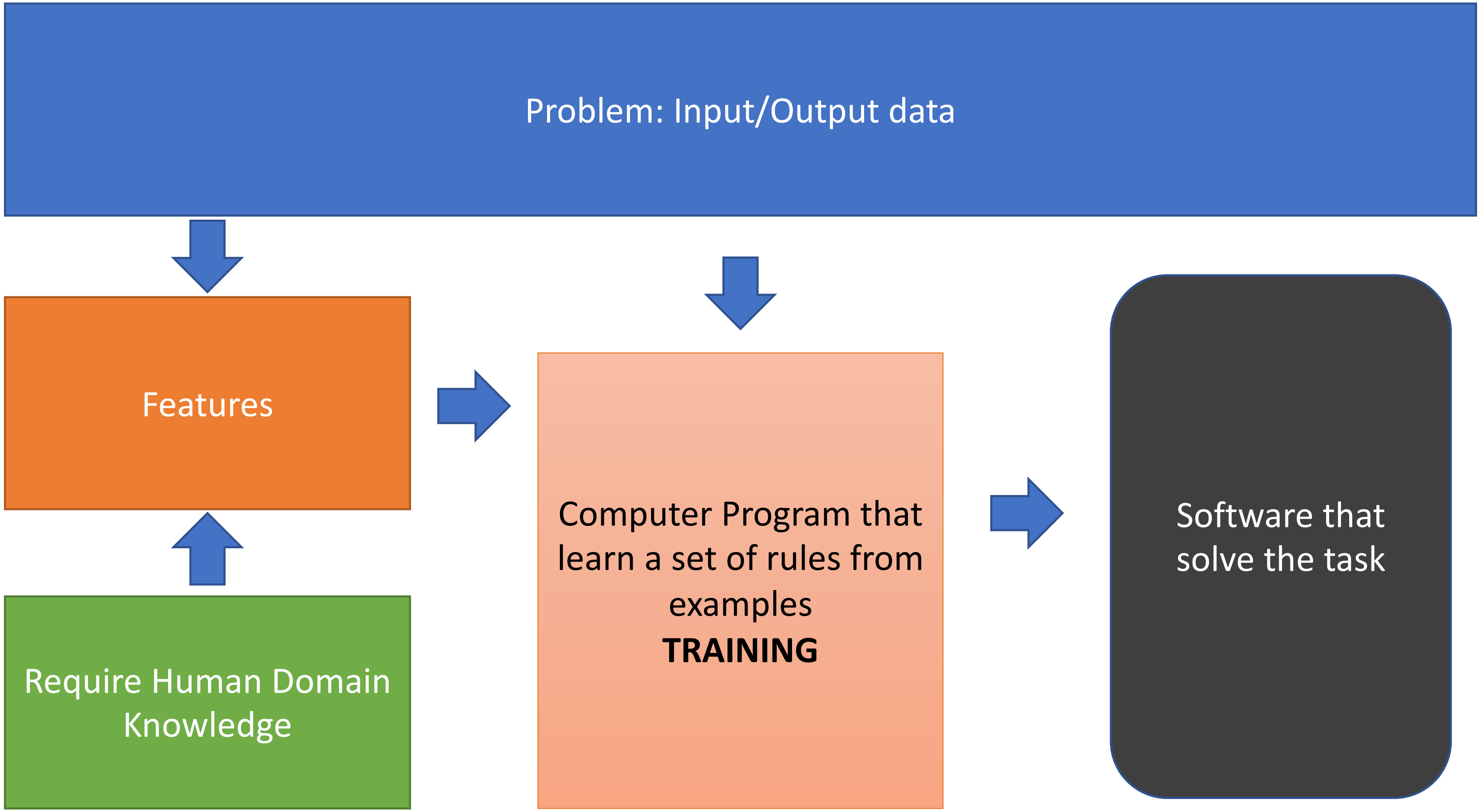
  a computer system is said to **learn** from *D* to perform the task *T* if after learning the system's performance on *T* improves as measured by *M*.
- In other words, the learned model helps the system to perform *T* better as compared to no learning.

Fundamental assumption of learning

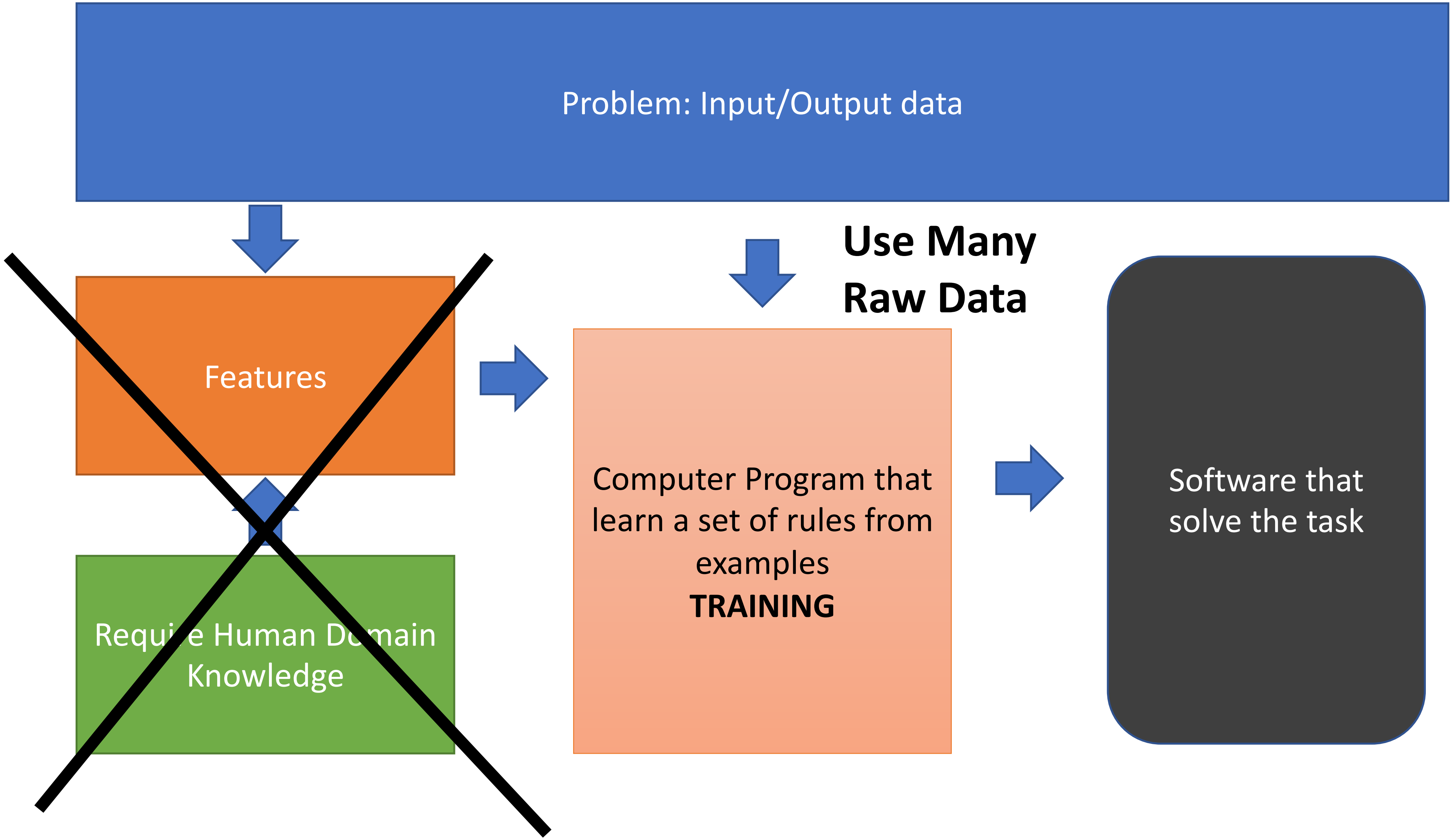
- Assumption: The distribution of training examples is identical to the distribution of test examples (including future unseen examples).
- In practice, this assumption is often violated to certain degree.
- Strong violations will clearly result in poor classification accuracy.
- To achieve good accuracy on the test data, training examples must be sufficiently representative of the test data.

General structure of a learning model

Learning Pipeline



Innovation of Deep Learning



Much more Data are needed: OCR Example

Tesseract Google OCR

- 800 Chars needed for Training
- Avg Trainig Time 10 minutes
- Core i7 PC NO GPU

Deep Neural Network

- 5000 chars needed for Training
- Avg Training time 30 minutes
- Core i7 PC + NVIDIA GPU CARD

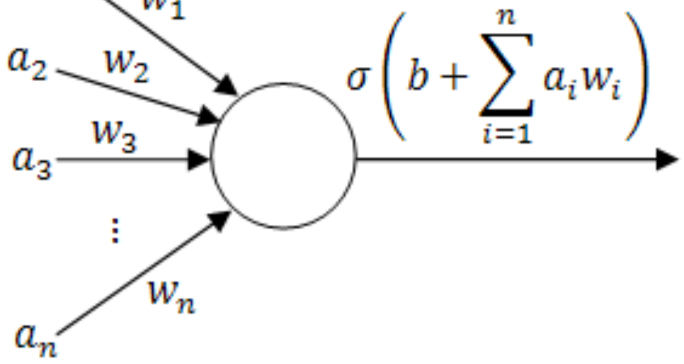
Technology	Accuracy
Tesseract eng language	40%
Tesseract trained language	60%
DEEP neural network(NN)	**98%**

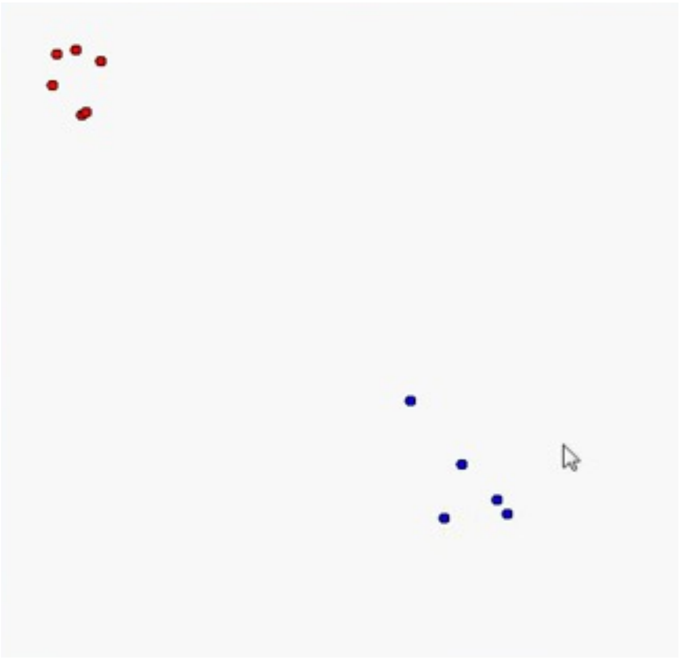
DEMO code @
http://christopher5106.github.io/computer/vision/2015/09/14/comparing-tesseract-and-deep-learning-for-ocr-optical-character-recognition.html

From the Perceptron to …….

- Perceptron is the analogous of a neuron
- Computational model -> perform linear classification

Perceptron is a linear Classifier

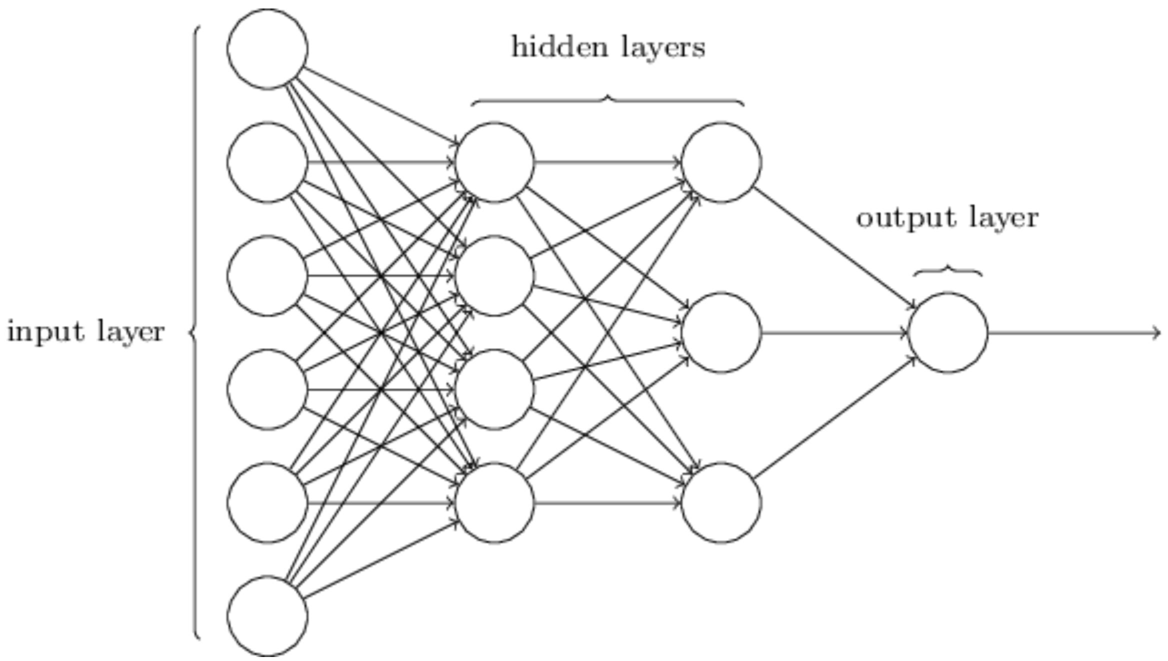


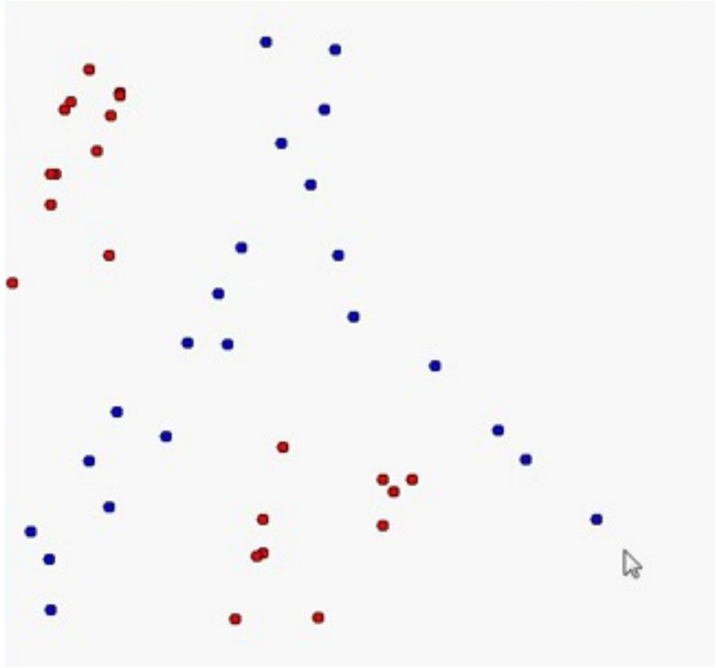


Multilayered Neural Network

- Stacking perceptrons **vertically** we obtain a layer
- Stacking layers **horizontally** we obtain a network

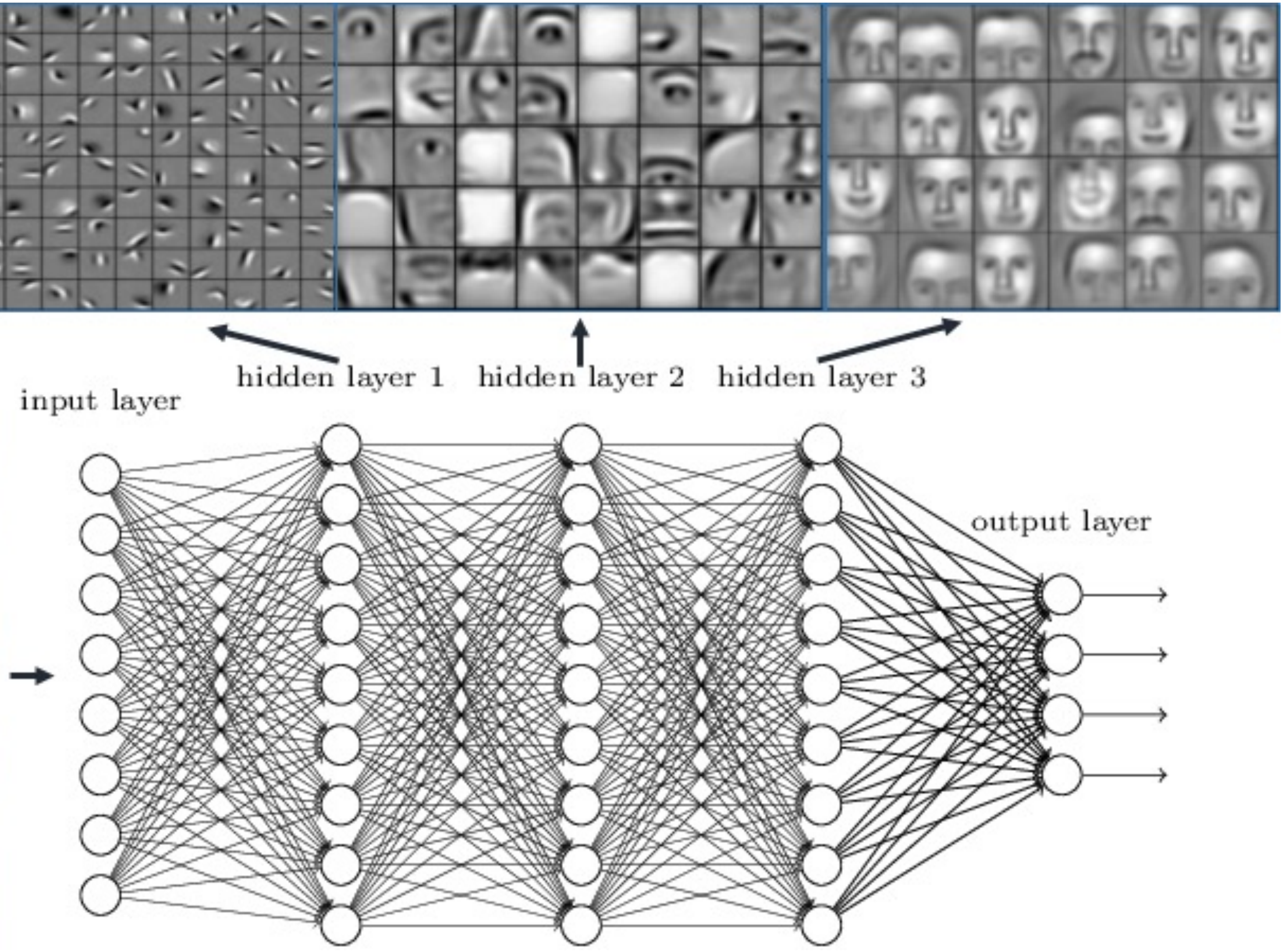
Network With 3 layers is a non-linear classifer



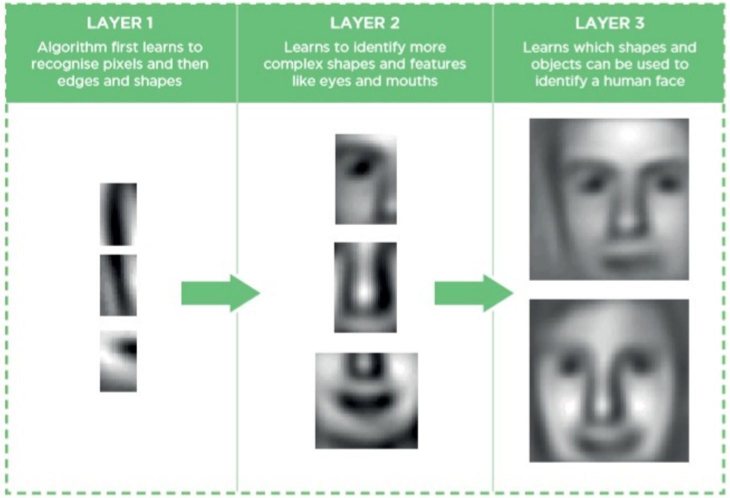


Going Deep

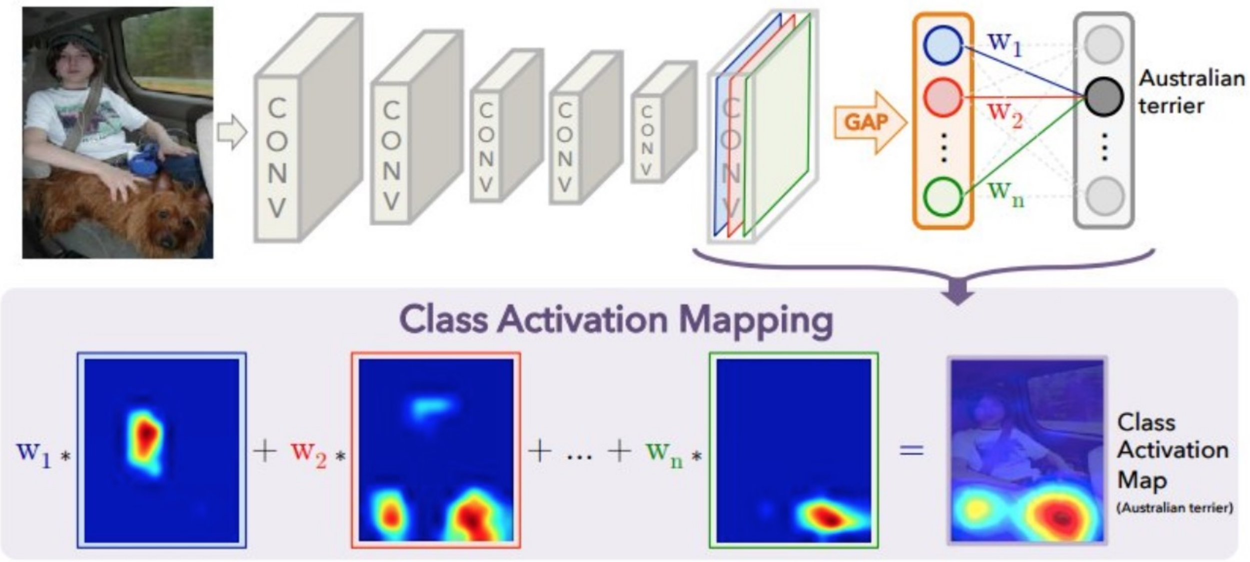
Deep neural networks learn hierarchical feature representations

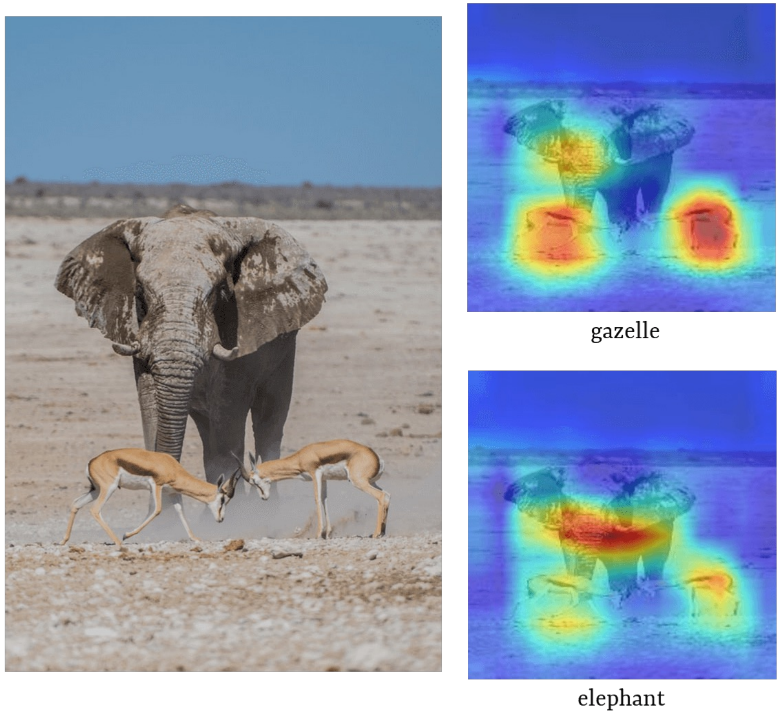


How the network learn the world



- Inner Representation is compositional
- Semantic representation must be inspected with additional rechniques e.g. Class Activation Maps





Deep Networks For:

Numerical Data -> Deep Neural Network

Applications: Production management, Prediction, Controls and Robotics

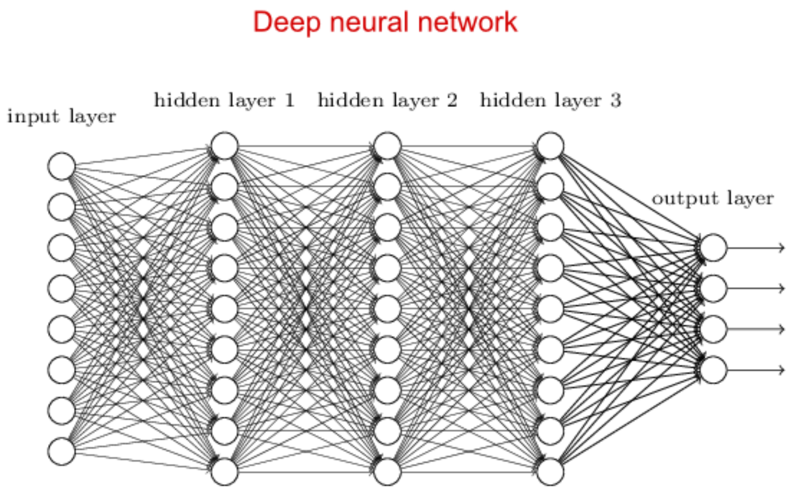
Multimedia Data-> Convolutional Network

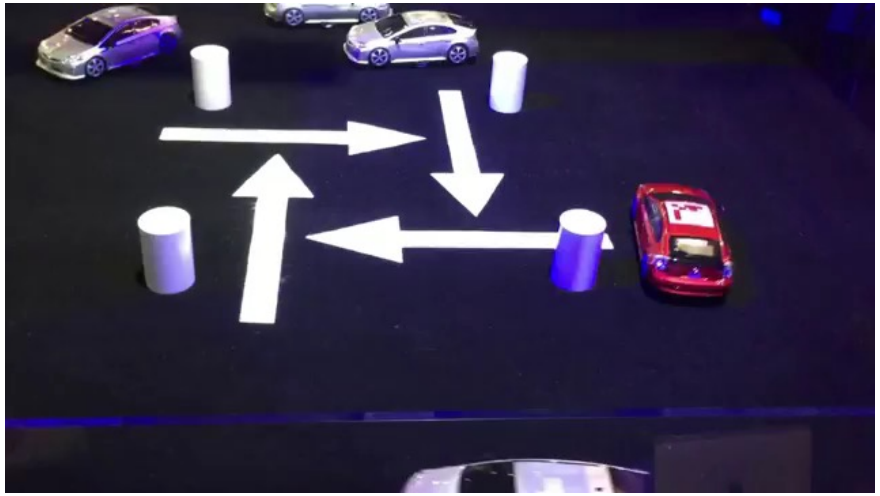
Applications: Image and Video classification, Face recognition, Licence Plate Detection, OCRs..

Time series -> Recurrent Neural Network

Applications: Financial Analysis, Audio and Speech analysis, Text analysis and traslation, Forecasting

Numerical Data -> Deep Neural Network





Pros:

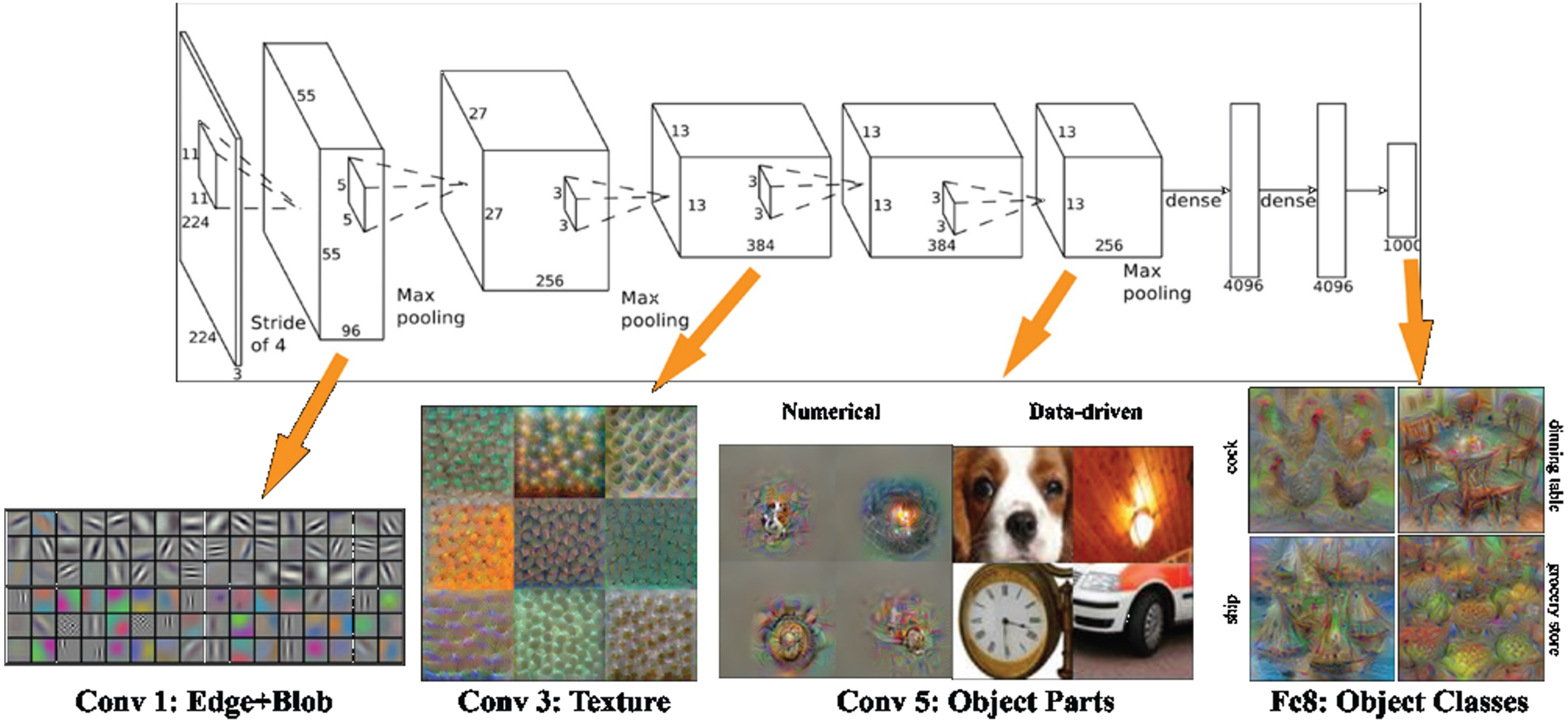
- Use Digital Sensors data as input
- Theoretically can learn every classification function
- Can predict a flexible number of outcomes

Cons:

- Many parameters to be learned
- Many training data needed
- Input dimension must be kept small

From CES2016 Red car is human guided

Multimedia Data-> Convolutional Network

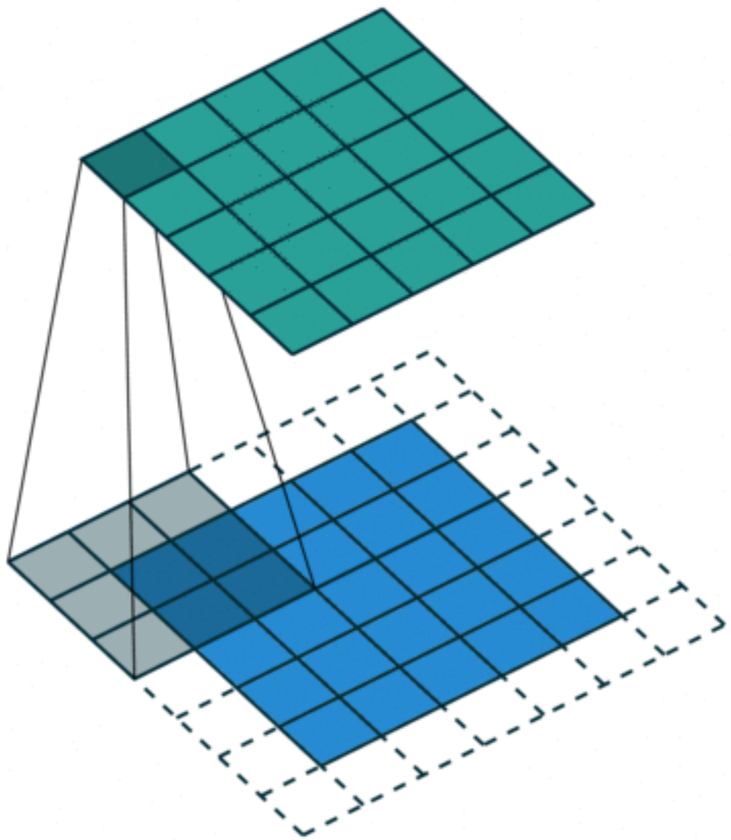


Pros:
- Use Image as Raw Data
- Can predict a flexible number of outcomes
- Use convolutions to reduce the number of parameters

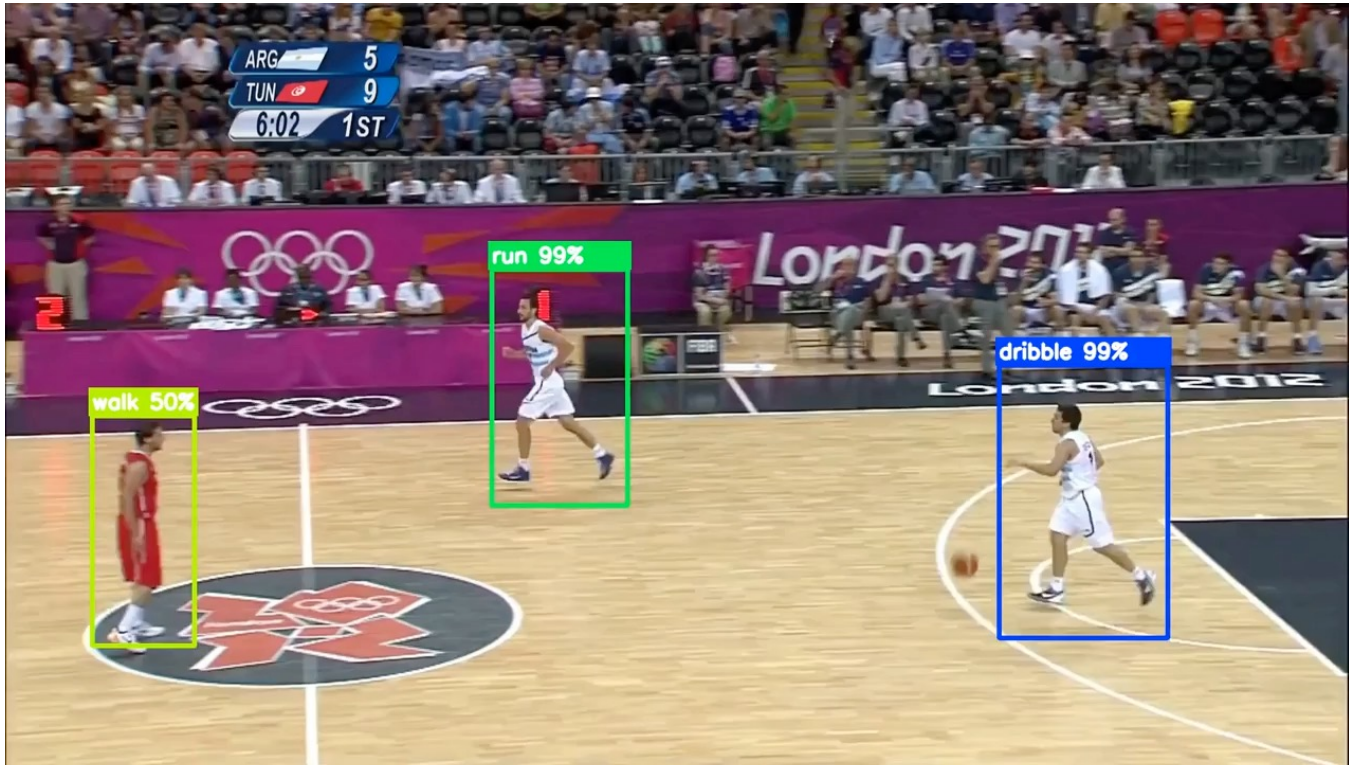
Cons:
- Image Scaling must be handled
- Input has «mostly» fixed shape
- Annotating images costs

Convolution

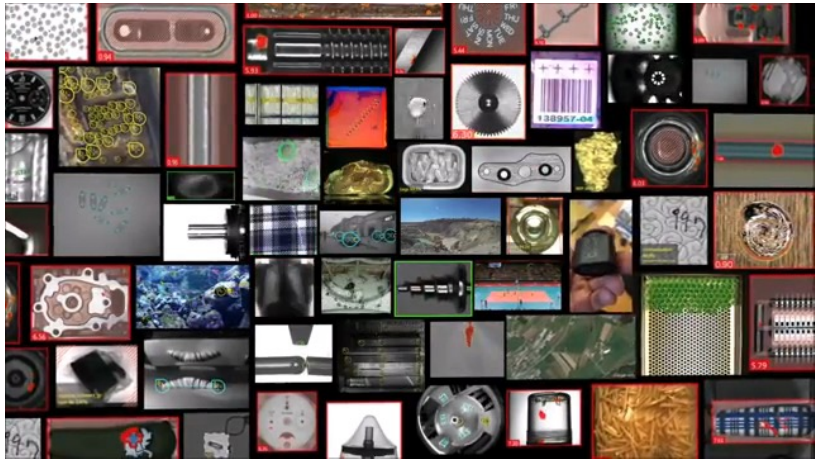
- Is a **spatial operation** with a learnable kernel
- Kernel shift spatially over data support
- Implies **translation invariance**



https://medium.com/syncedreview/a-guide-to-receptive-field-arithmetic-for-convolutional-neural-networks-42f33d4378e0


ARG 5
TUN 9
6:02 1ST
run 99%
walk 50%
dribble 99%

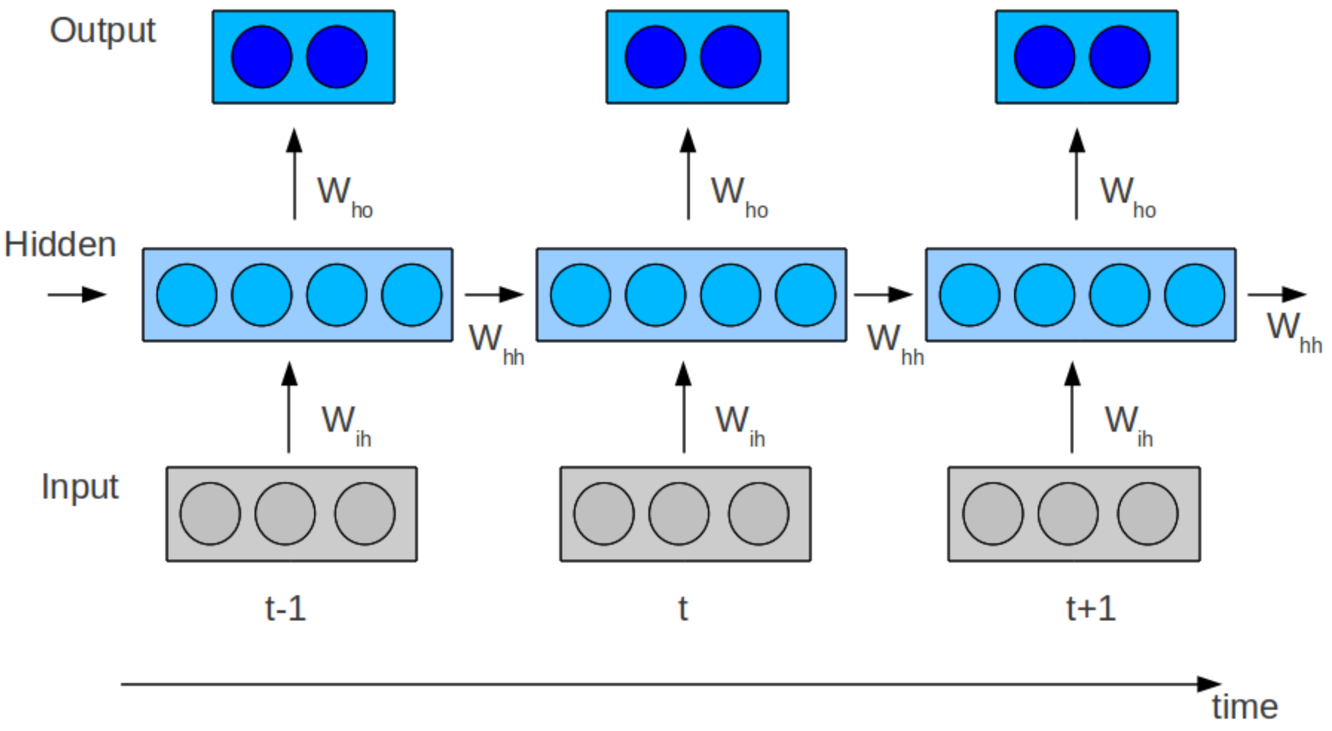

30
25
20
15
10
5
-10
-5
0
5
10



Inspection time

+/- 200 msec.

Time series -> Recurrent Neural Network



Pros:
- Use Temporal Data
- Has memory of the past
- Can predict future outcomes

Cons:
- Hard to train
- It forgets!
- Parameters grows as time grows

cybernaut@Cyberion: ~/neuralengine_3stream

```
cybernaut@Cyberion:~/neuralengine_3stream$ th demo2.lua

Loading data. Please wait...
constructing clones inside the LanguageModel
```

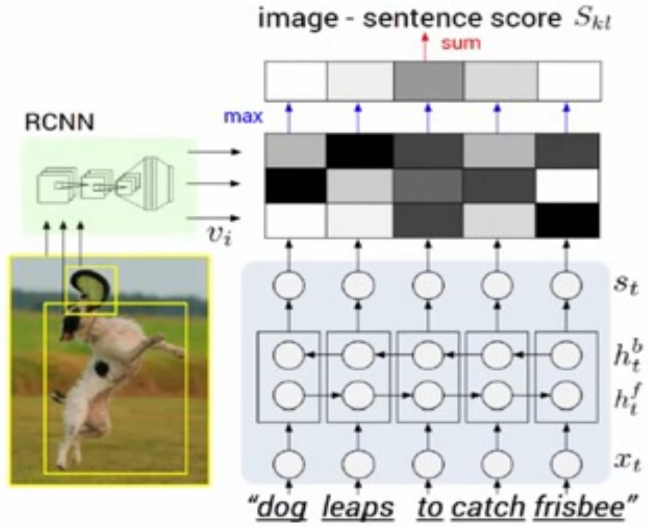
Source: Andrej Karpathy

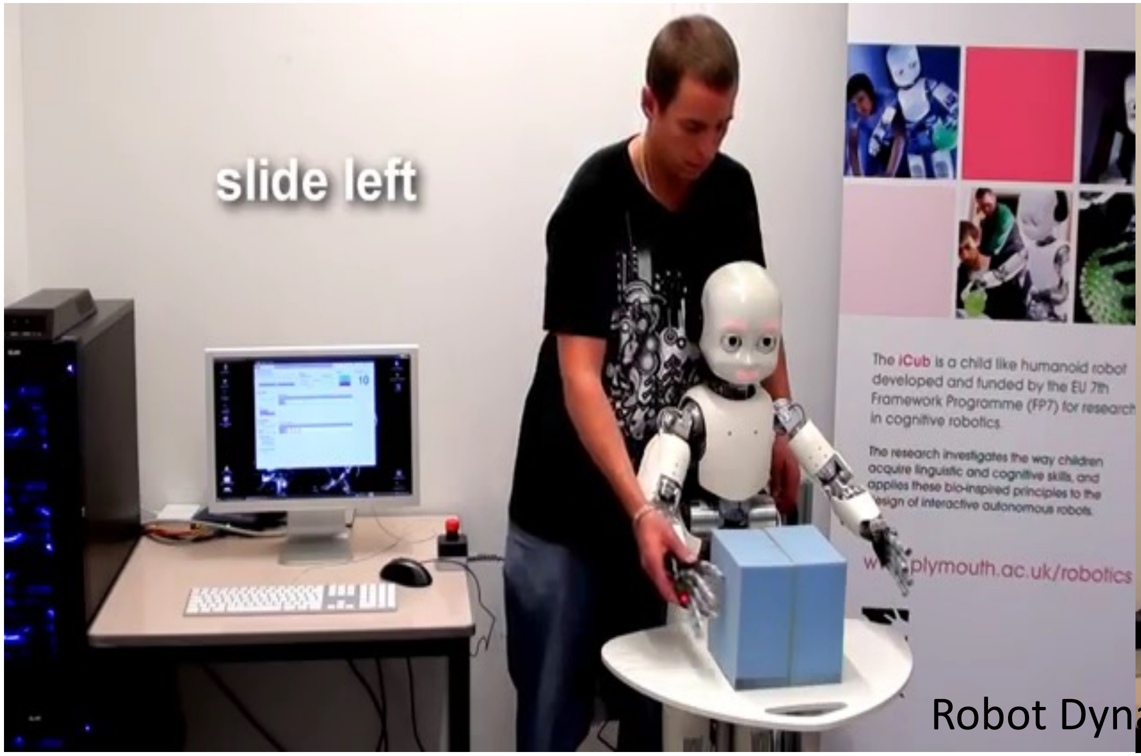
Text and Music Writing

```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

```
"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."
```

Source: [Karpathy and Fei-Fei 2015]





Robot Dynamics Control

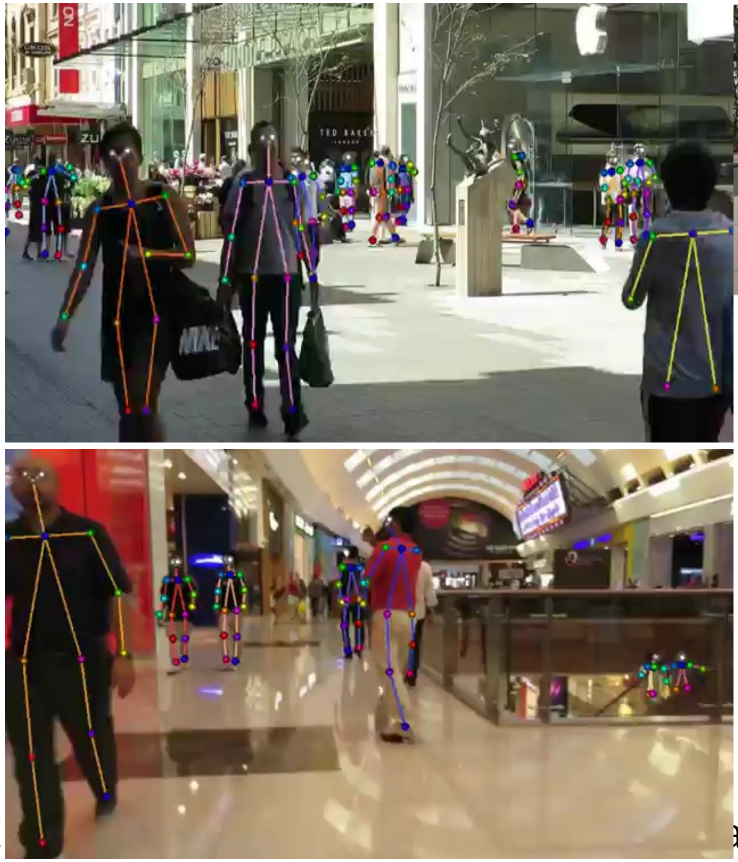
Applications made in ER

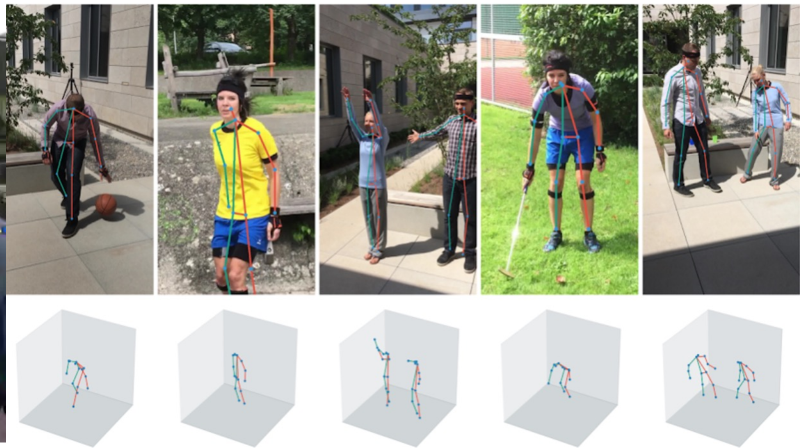
AImage Lab

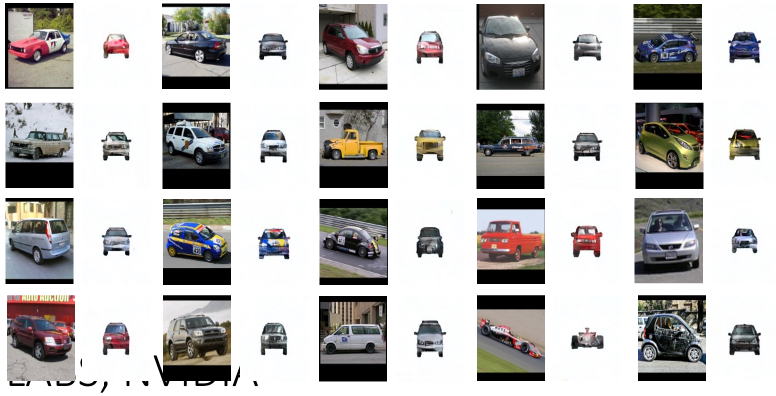
https://aimagelab.ing.unimore.it/

ML and CV for People Analysis in Smart Urban Environment

- **Study Deep Learning techniques** for:
- People tracking
- People detection 2D and 3D
- Human Behavior understanding
- Anomaly Detection
- Vehicles-human interaction
- Geometric view synthesis
- Conferences and Journals:
- CVPR, ICCV, TPAMI, TIP, TMM





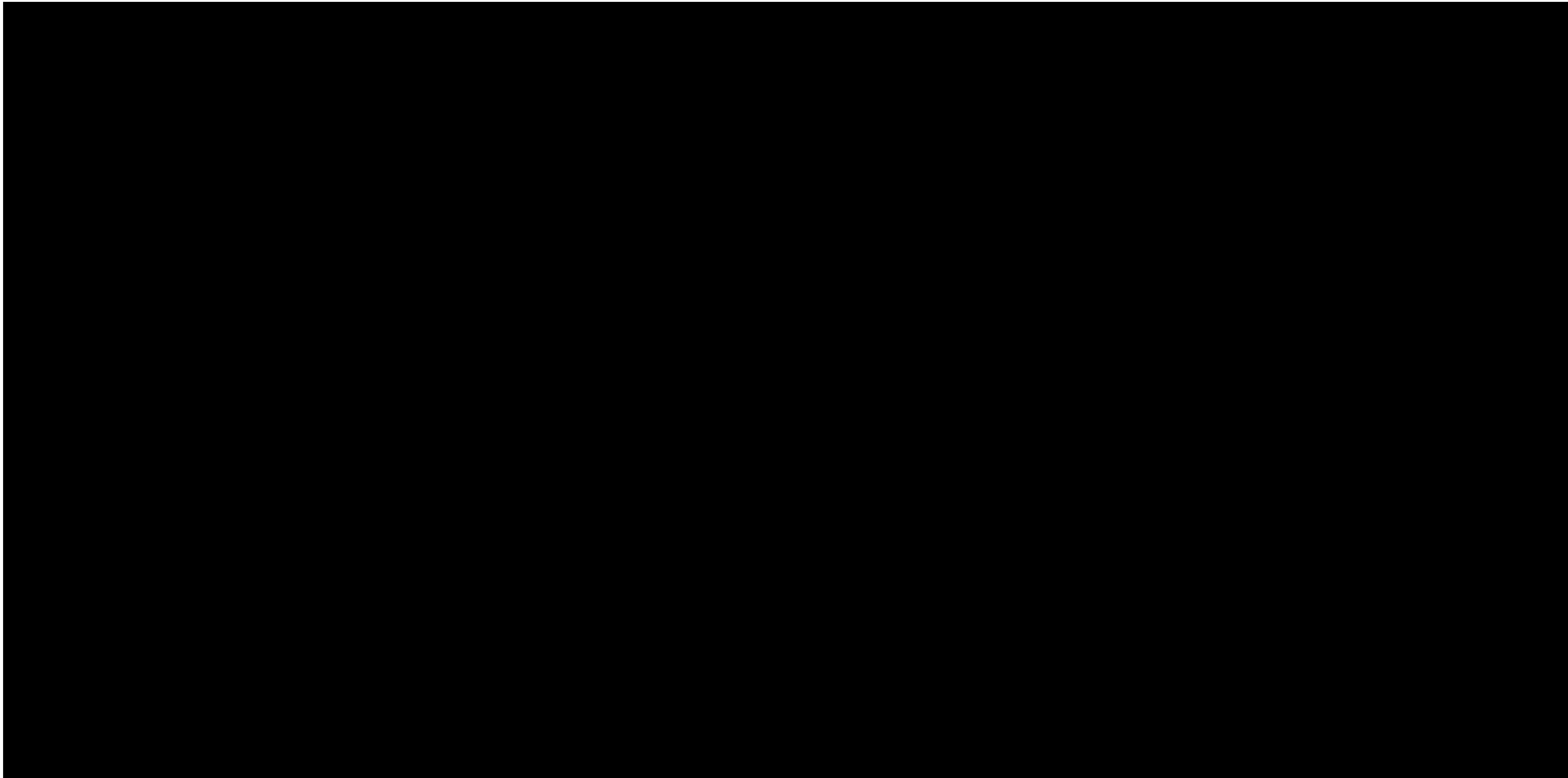


- Projects and Collaborations:
- PRIN COSMOS and PREVUE, EU PRYSTINE, E

- AImageLab Group: Rita Cucchiara, Roberto Vezzani, Simone Calderara
- http://aimagelab.ing.unimore.it

AImage Lab
University of Modena and Reggio Emilia

Car perception and analysis



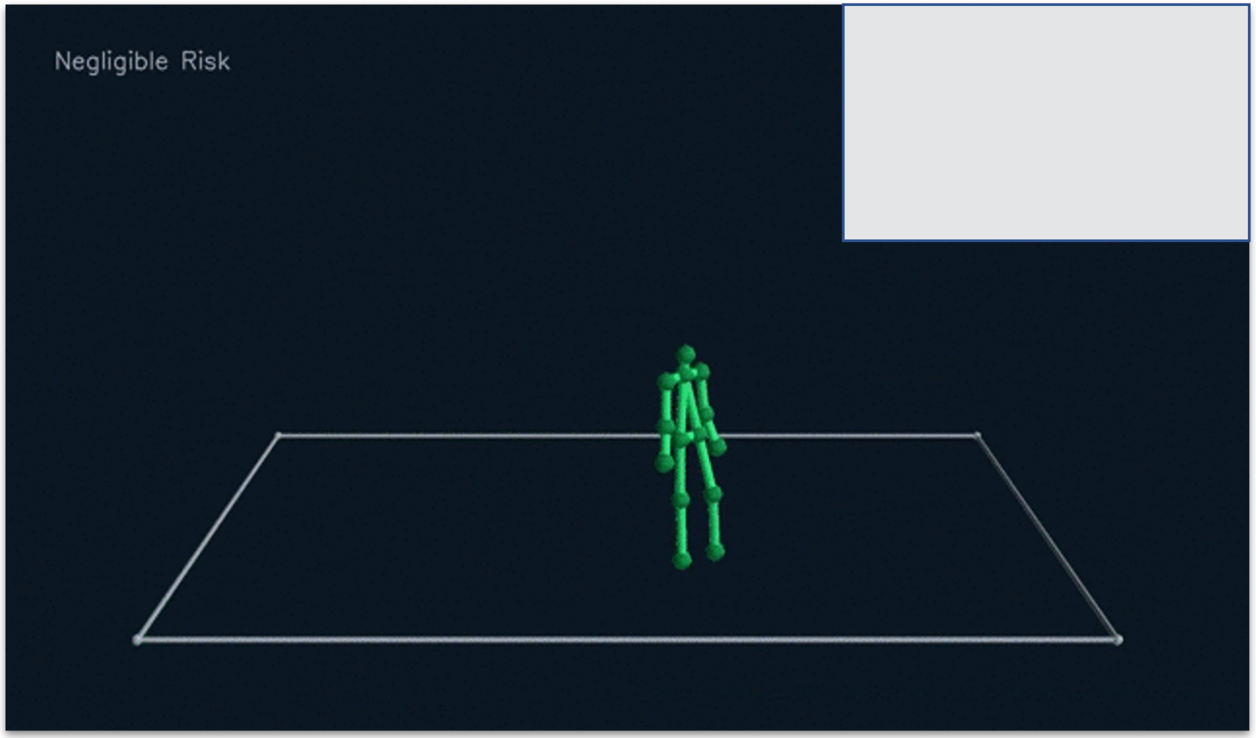
AI academy

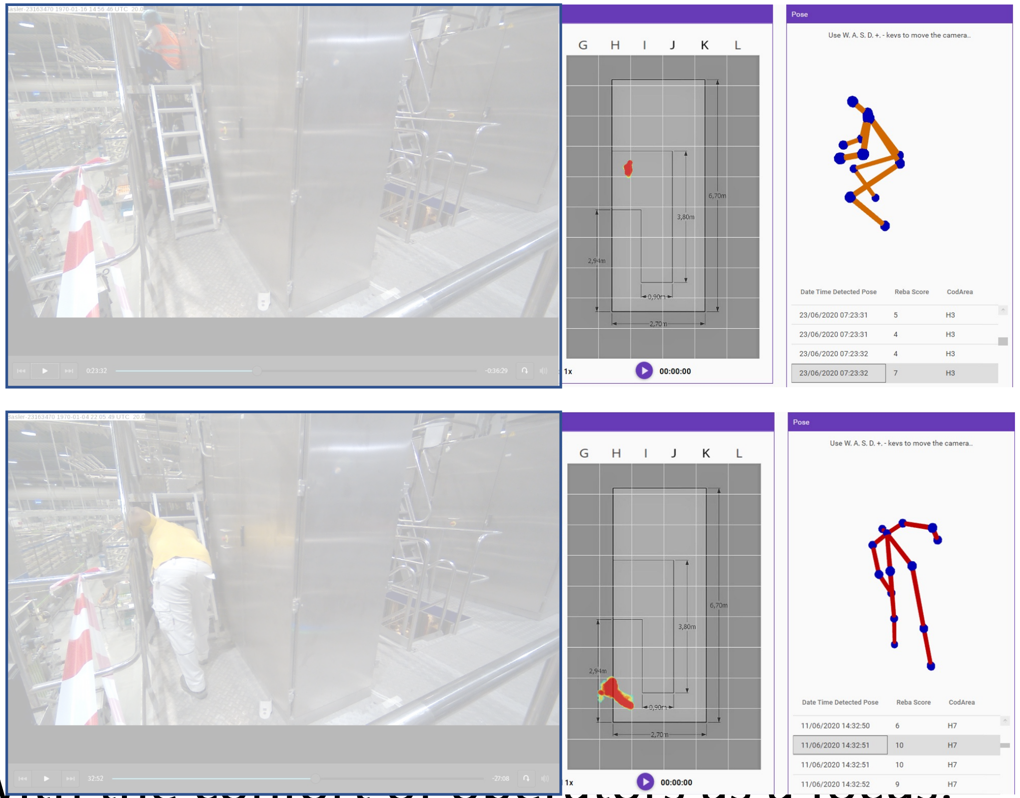
EyeCandy Results

Learning human in industrial environment

Can virtual data generalize to different environments?

Can we understand how people work and interact with machines?



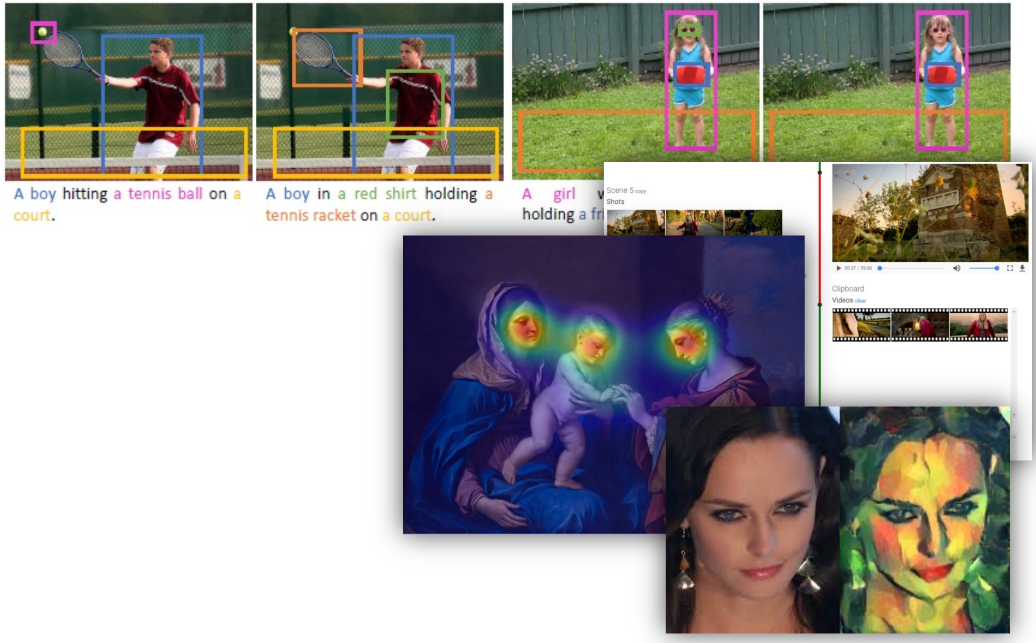


Machines of the future can reconfigure themselves with the comfort of operators as a focus.

Activity in conjunction with Tetra Pak (Automation & Digital) (Modena Lab) Emilia Romagna

Embodied AI and Digital Humanities

- **Embodied AI**: Integration between Vision, Language and Action
  - Automatic description of Images and Video
  - Natural Language and multi-modal retrieval
  - Vision and Language Navigation
  - Navigation of embodied agents in unseen environments
- Applications in Cultural Heritage and Digital Humanities



- **Conferences and Journals:**
- CVPR, ICCV, TPAMI, TIP, TMM

- **Projects and Collaborations:**
  - IDEHA, CULTMEDIA, AI4CH, AI4DH
  - Facebook AI Research, NVIDIA, University of Haifa (Israel)

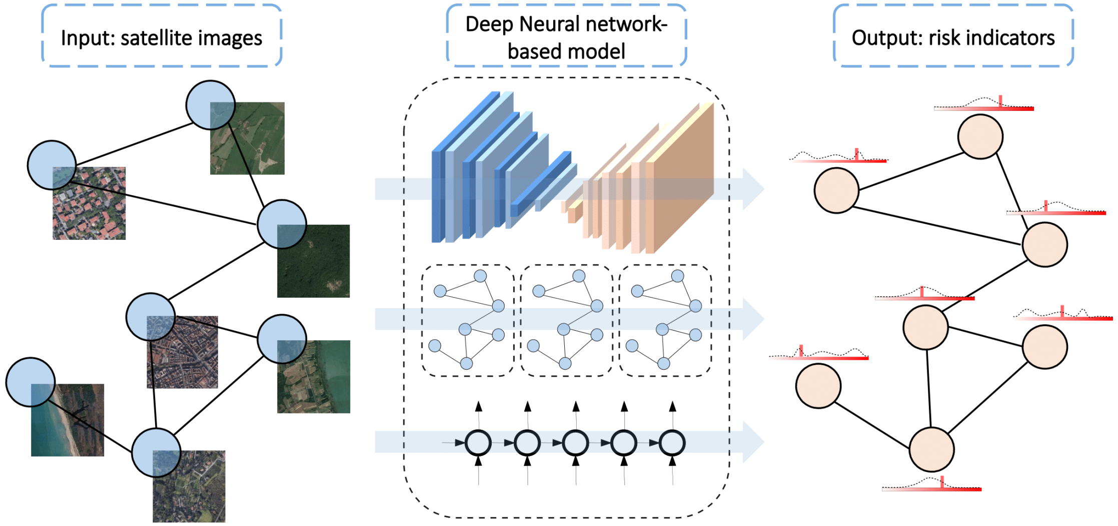
- AImageLab Group: Rita Cucchiara, Lorenzo Baraldi, Marcella Cornia
- http://aimagelab.ing.unimore.it

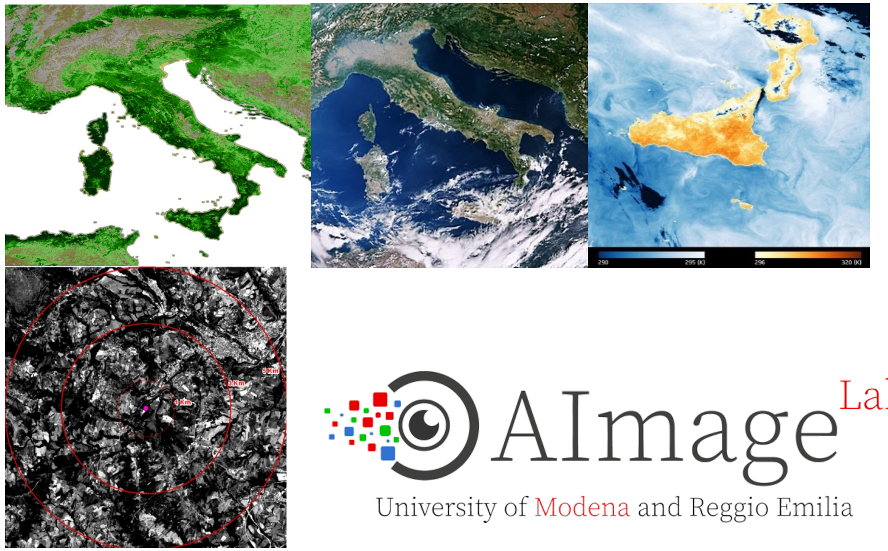
AImage Lab
University of Modena and Reggio Emilia

Machine learning for Earth Observation

- Deep Learning and Graph based analysis for:
- Satellite Images self-supervised feature extraction
- Inference of physical phenomena from EO
- Epidemic and vectors analysis using temporal EO

- **Projects and Collaborations:**
- AI4VECT Italian Ministry of Health,
  AIDEO European Spatial Agency

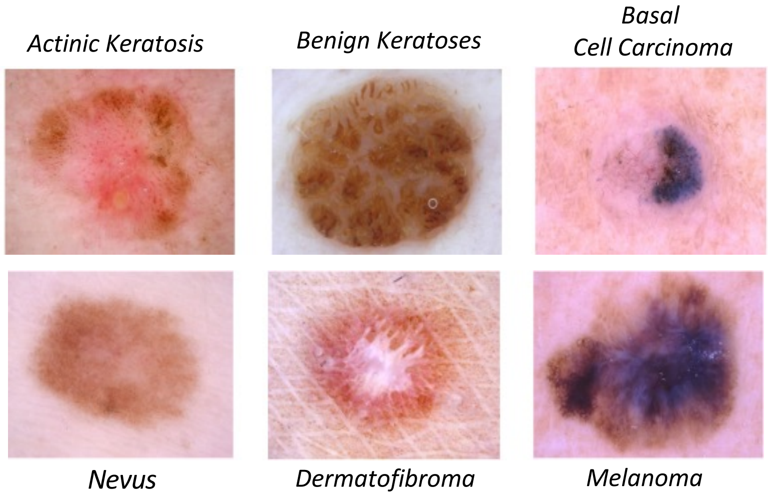




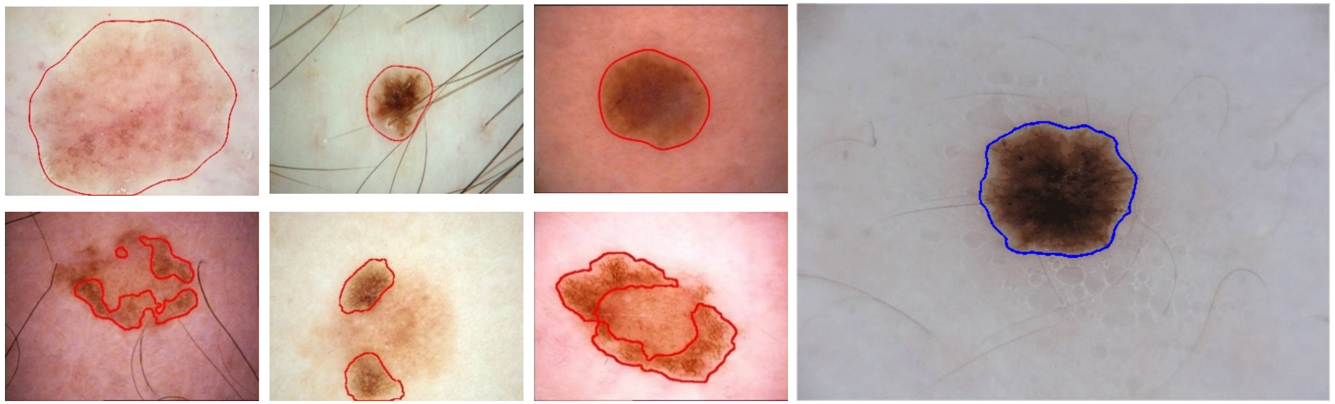
AImageLab Group: Simone Calderara, Angelo Porrello
http://aimagelab.ing.unimore.it

AImage Lab
University of Modena and Reggio Emilia

Medical Imaging



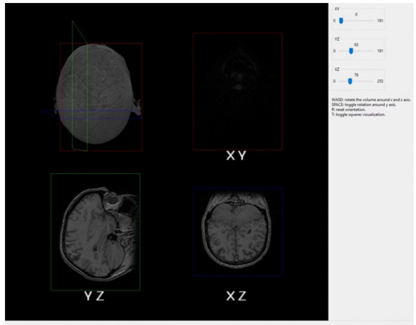
Lesion diagnosis

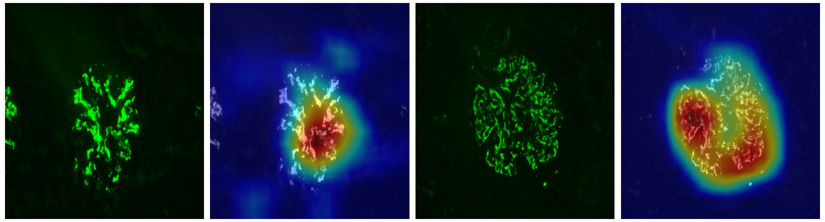


Lesion boundary segmentation and attribute detection

- **Third-place** (out of 64 research groups) at the 2019 **international ISIC challenge** (lesion diagnosis)

- DeepHealth (H2020): Deep-Learning and HPC to Boost Biomedical Applications for Health. **Period:** 2019 - 2021 **Budget:** 14.64 M€



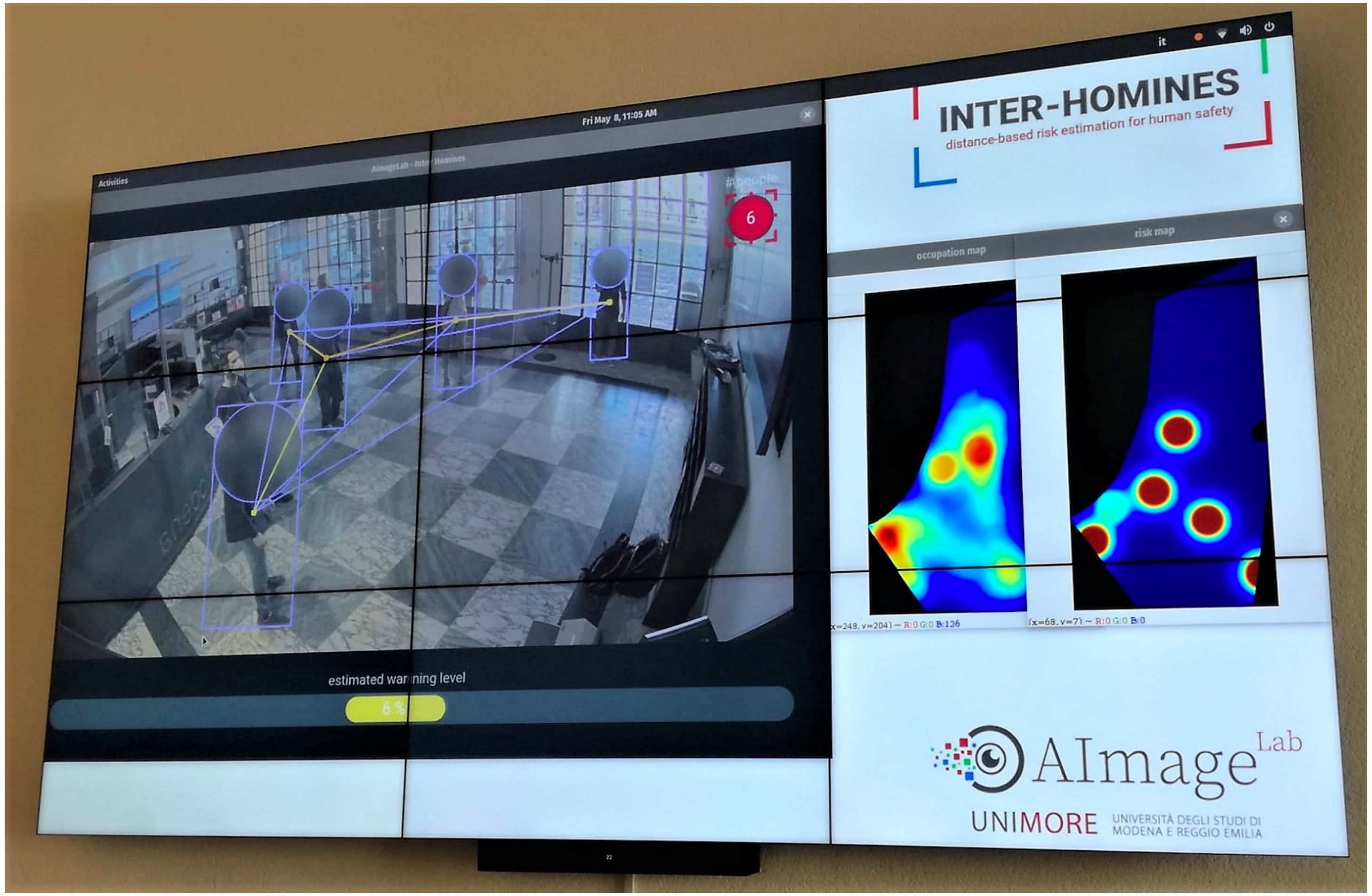


- Intensity, patterns and locations of antibody deposits in immunofluorescence images from renal biopsies

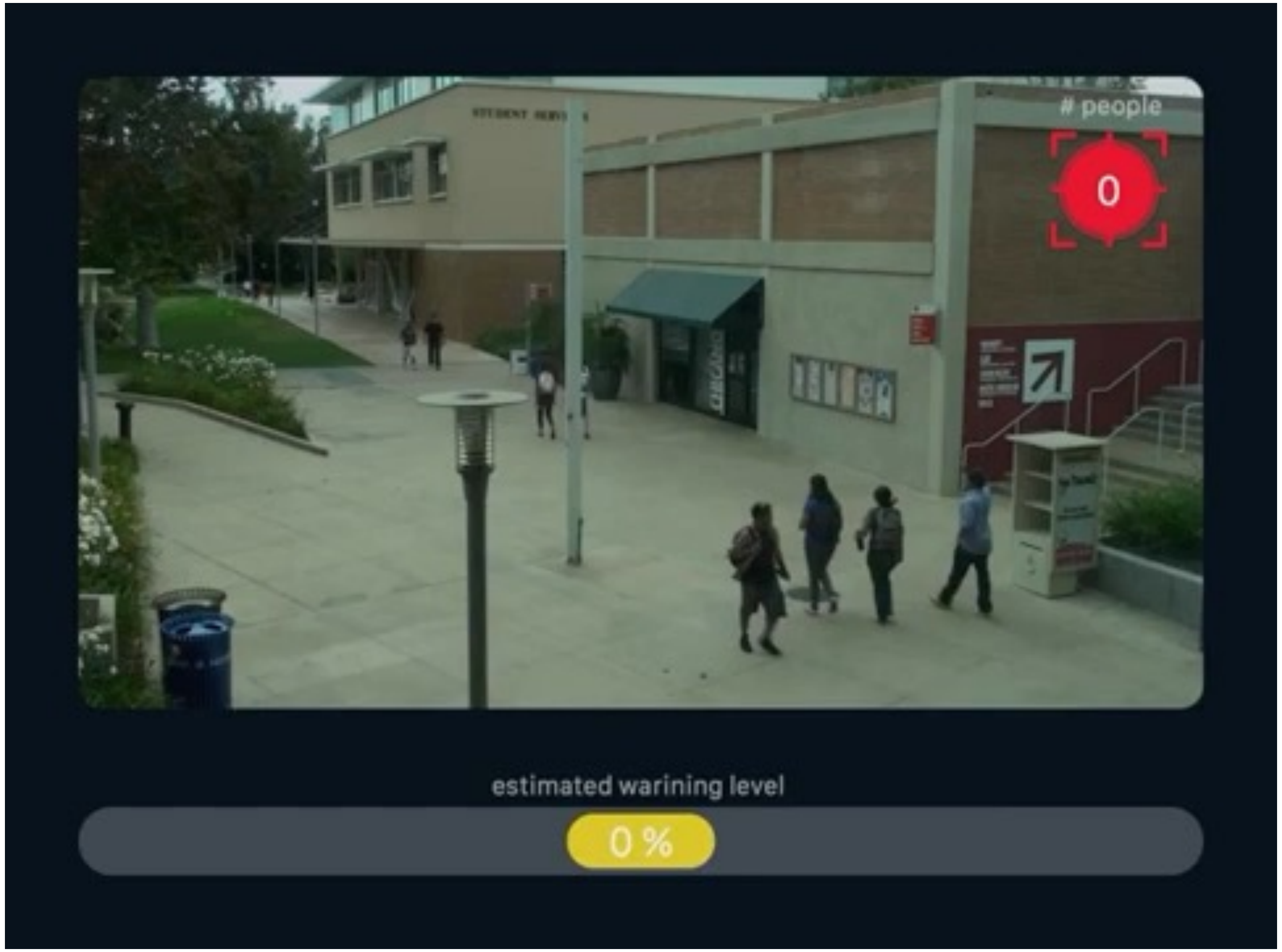
AImageLab Group: Costantino Grana, Federico Bolelli
http://aimagelab.ing.unimore.it

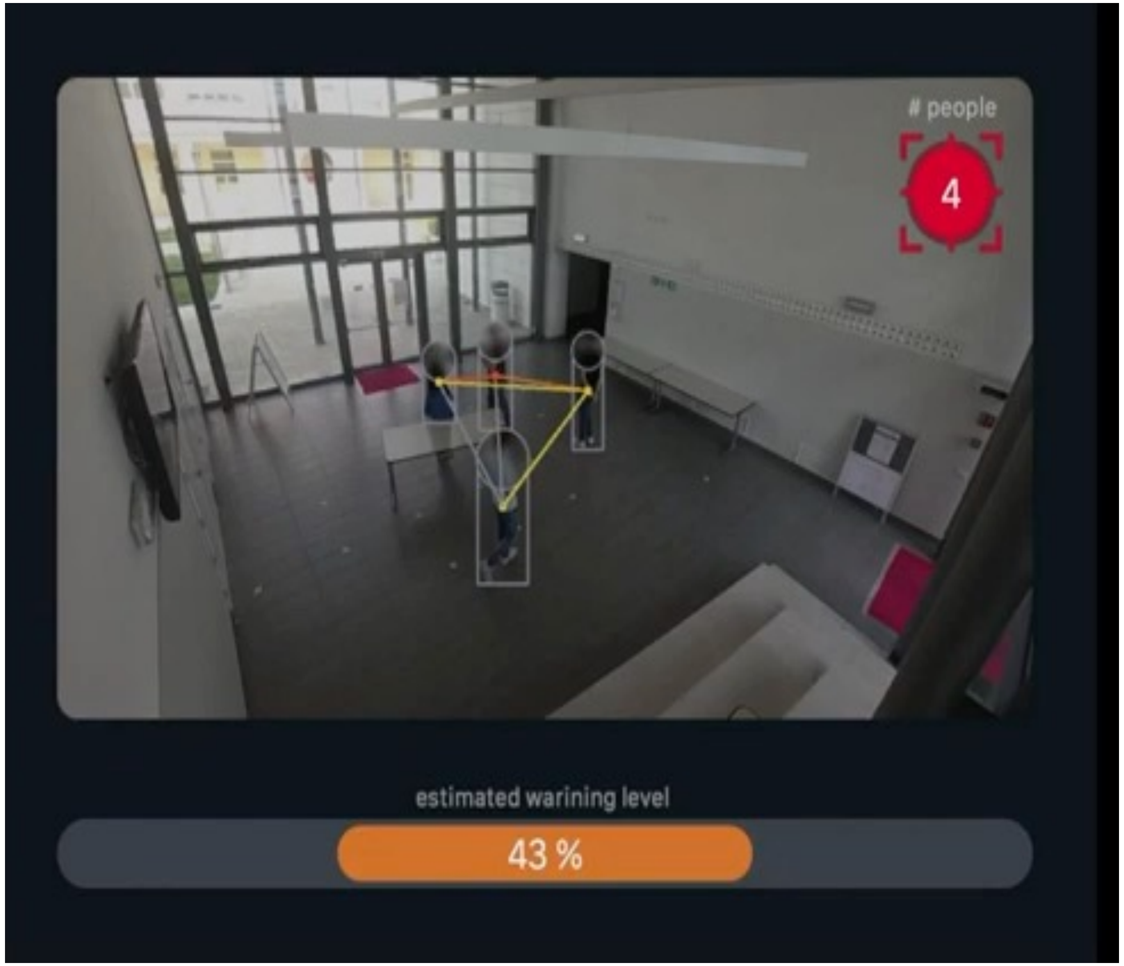
Inter-Homines - Distance-based risk estimation for human safety

- Using cameras and AI to detect interpersonal distances
- Assess the risk of an area
- Sophisticated behavior analysis models for system robustness and risk evaluation



For more info https://aimagelab.ing.unimore.it/imagelab/project.asp?idprogetto=82
Prof. Rita Cucchiara rita.cucchiara@unimore.it Director of the Project
Dr. Matteo Fabbri matteo.fabbri@unimore.it


people
0
estimated warining level
0 %


people
4
estimated warining level
43 %

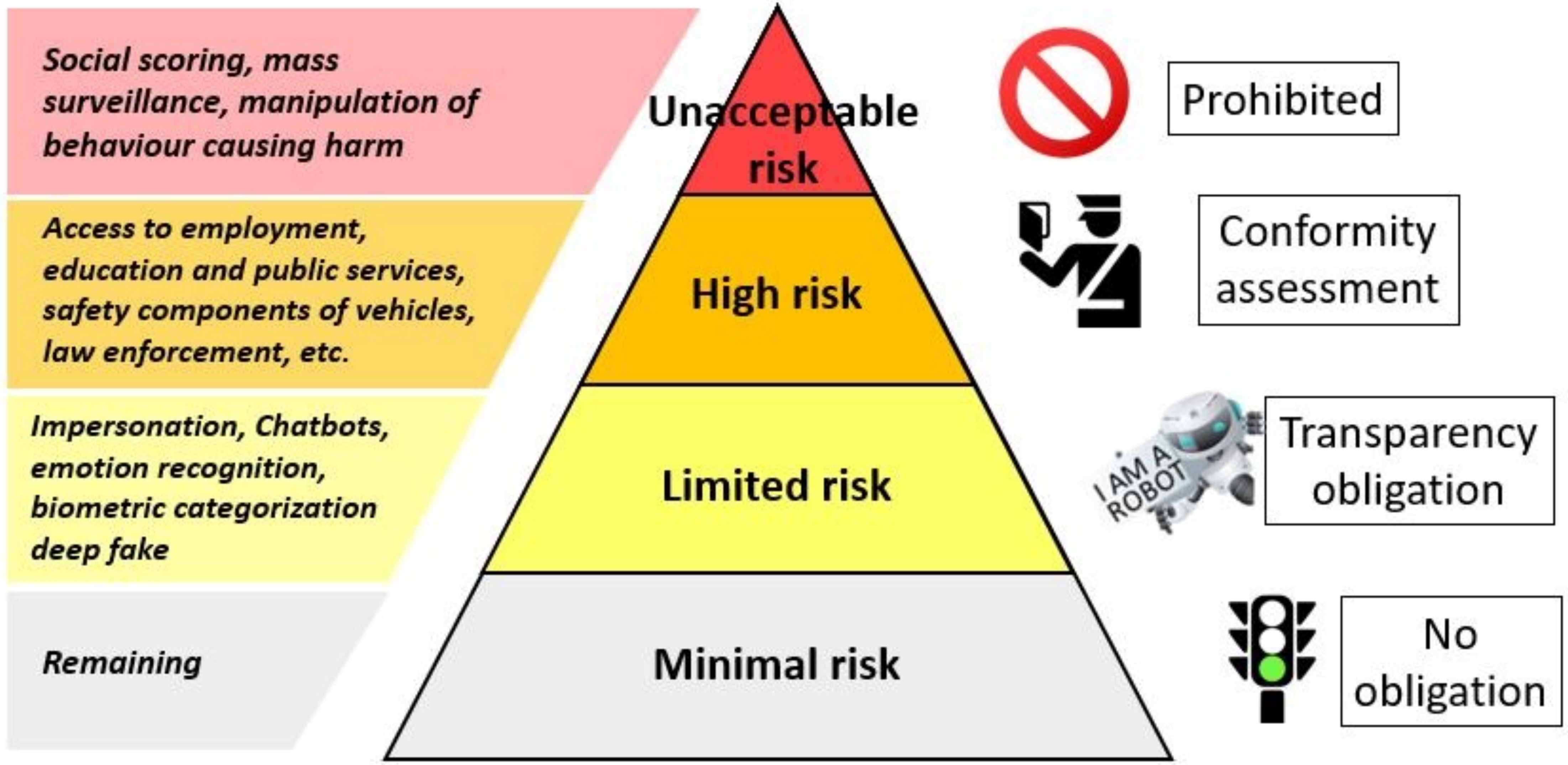
Future

WE ARE STILL MISSING ABSTRACTION

WE ARE MISSING INTERPRETABILITY

WE ARE MISSING AUTONOMY OF REASONING

EU Artificial Intelligence Act: Risk levels



Thank you for your attention

- **More on us**

- **Research:**
- AImageLab Research Group: http://aimagelab.unimore.it
- **Ellis Unit UNIMORE: https://ellis.eu/units/modena-unimore**

- **Tech Transfer and Life Long Learning:**
- AIAcademy UNIMORE: http://aiacademy.unimore.it/

UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AI Academy

AImage Lab