Pattern Recognition and Machine Learning a gentle introduction

Matteo Tomei

Imagelab 4 Computer Vision, Pattern Recognition & Machine Learning

Dief University Of Modena and Reggio Emilia

Contact: **matteo.tomei@unimore.it**

Centro Interdipartimentale di Ricerca
Softech: ICT per le Imprese

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab
University of Modena and Reggio Emilia

A Glossary of Artificial-Intelligence Terms

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

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.

Compare Search terms

machine learning
Search term

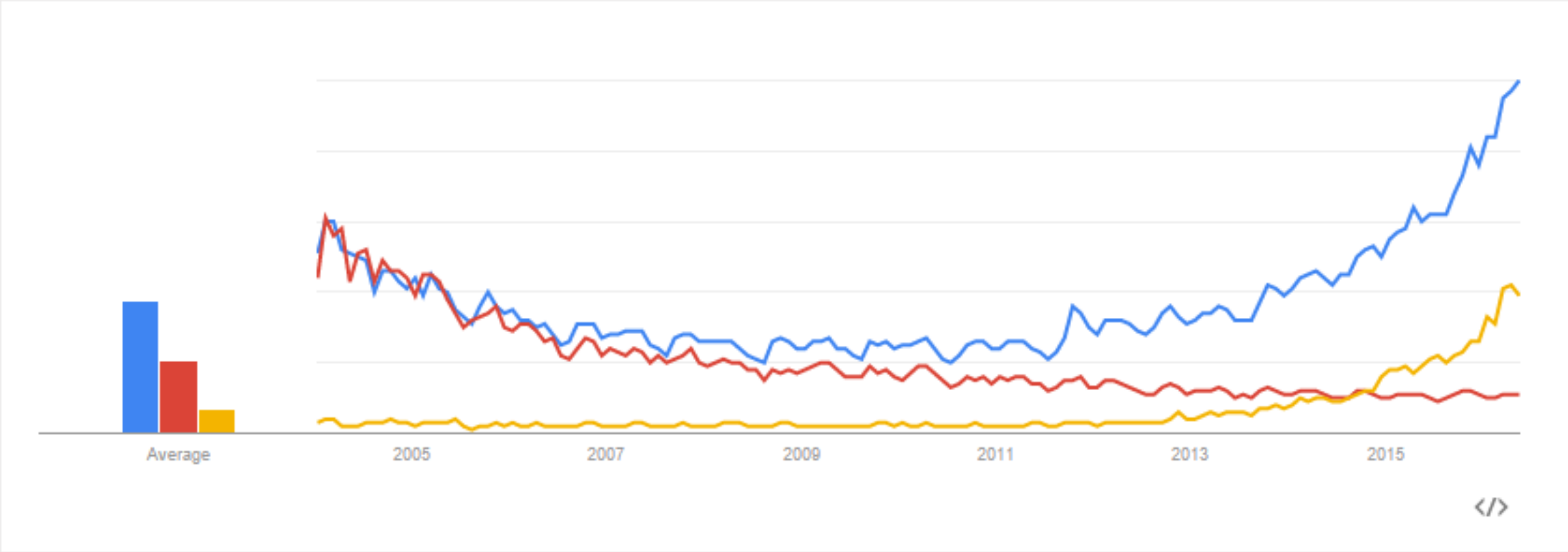
pattern recognitio..
Search term

deep learning
Search term

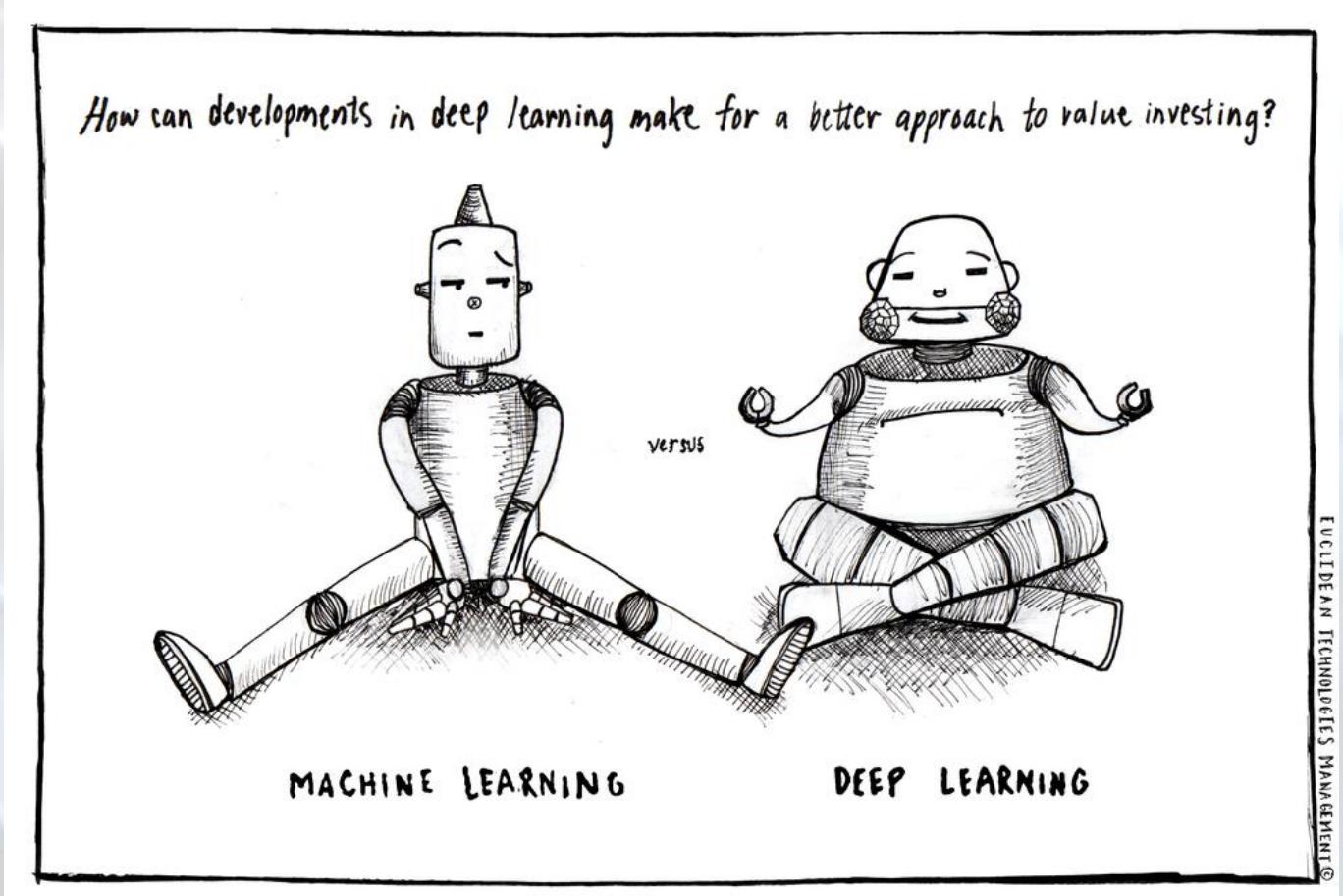
+ Add term

Interest over time

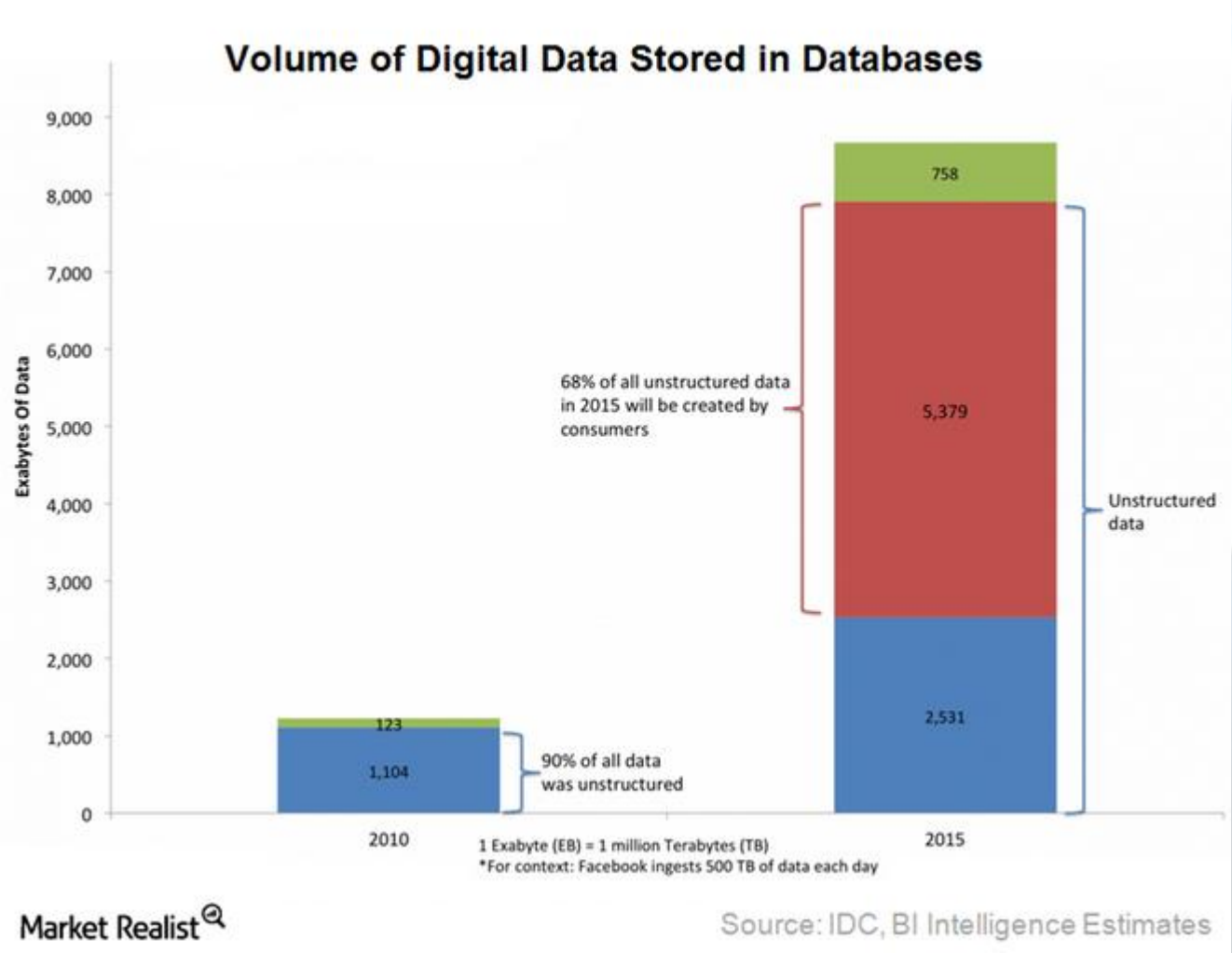
News headlines Forecast



Looking Forward

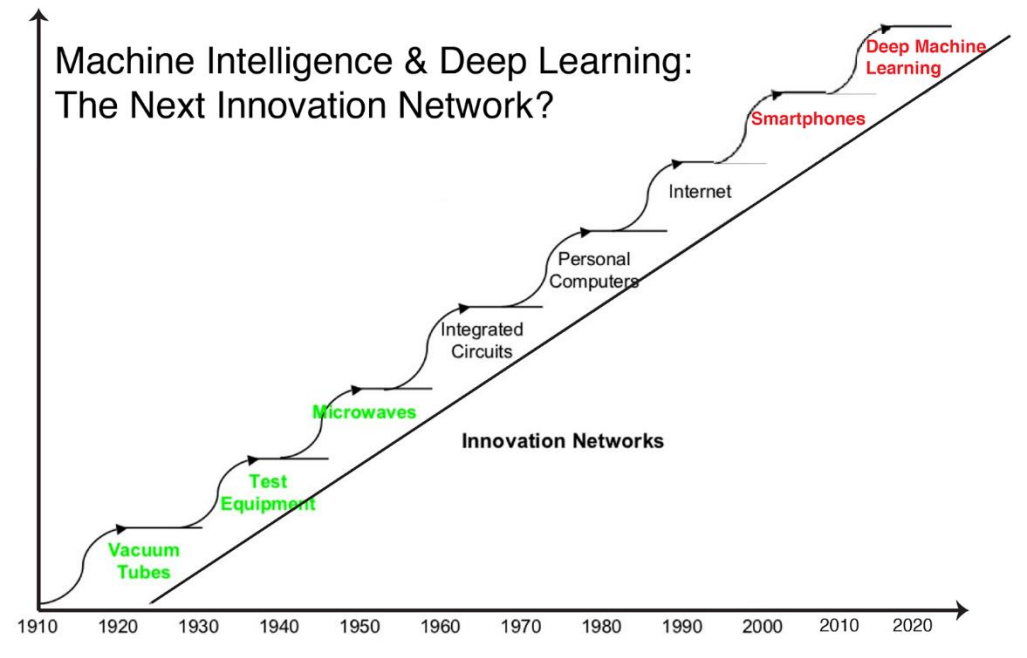


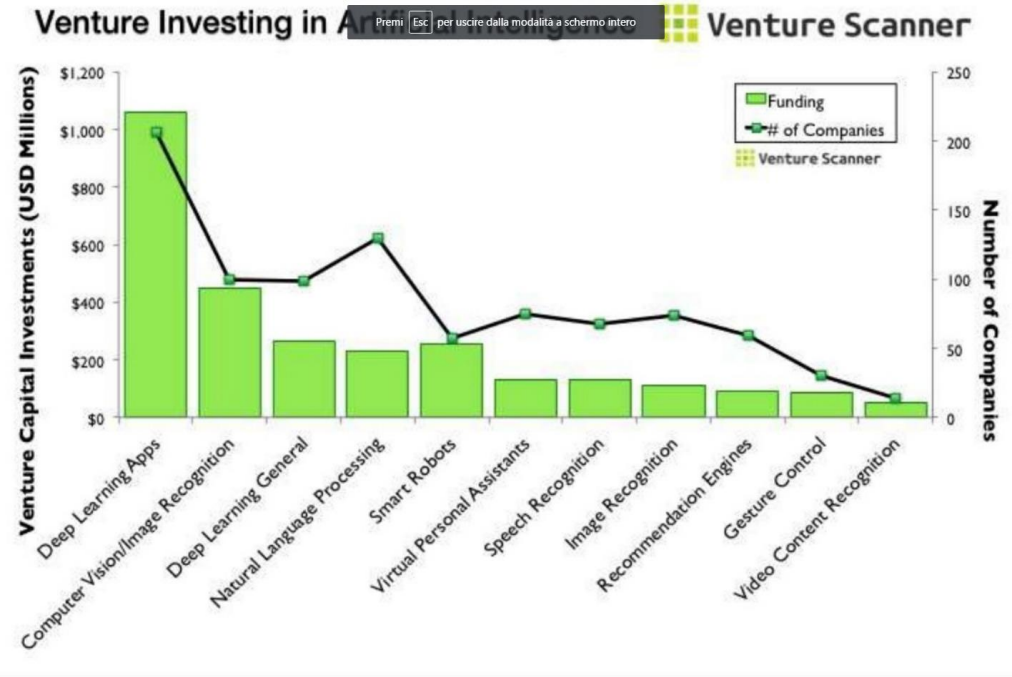
Motivation-> Data is changing



- **Growing in Volume**
- **Becoming unstructured**
- **Sources grow in number**

Investments in the Deep Learning Trend





From The Futures Agency - Silicon Valley Innovations Timeline

To 2016-> and the future

- 2008-2013 Years of theoretical studies and hardware production

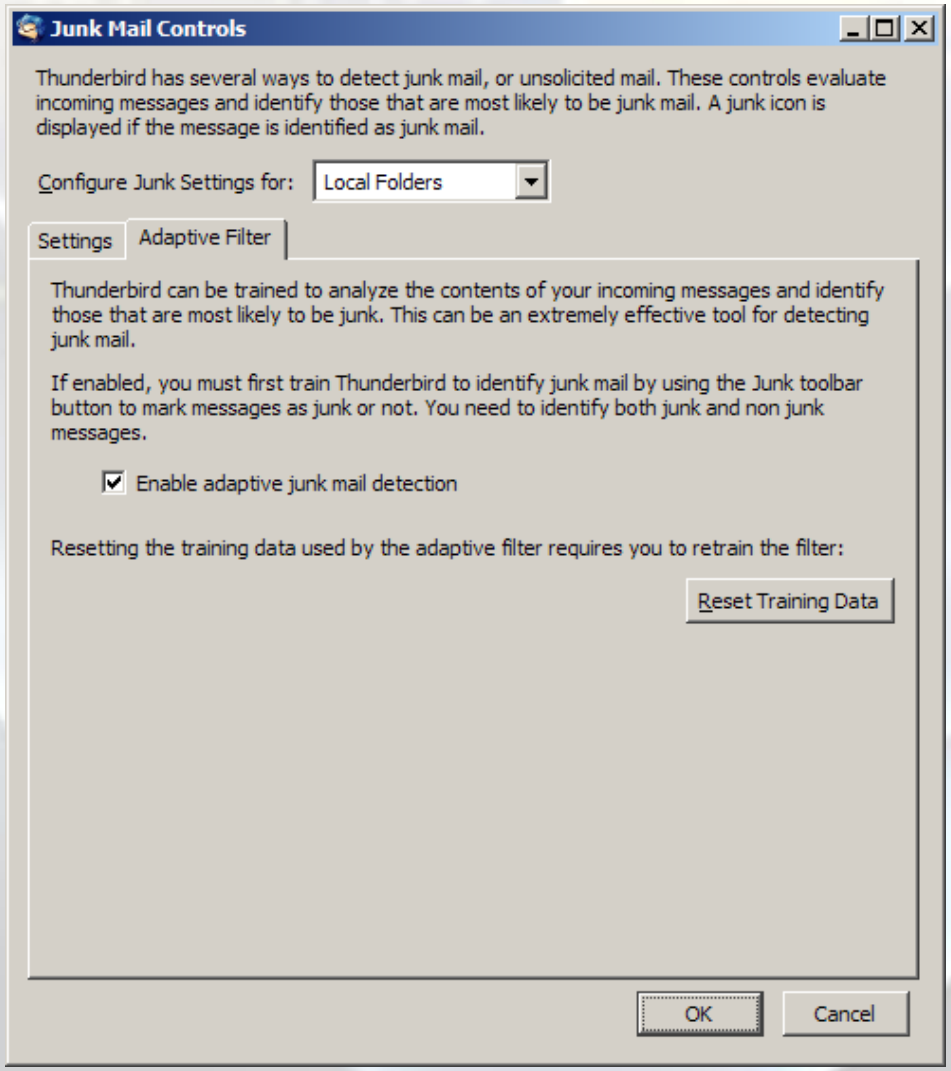
- 2016->…… **Time to bring out the application**

  - “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
  - ….

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

- Using machine learning to detect spam emails.



To: you@gmail.com
GET YOUR DIPLOMA TODAY!
If you are looking for a fast and cheap way to get a diploma, this is the best way out for you. Choose the desired field and degree and call us right now: For US: 1.845.709.8044 Outside US: +1.845.709.8044 "Just leave your NAME & PHONE NO. (with CountryCode)" in the voicemail. Our staff will get back to you in next few days!

ALGORITHM

Naïve Bayes

Rule mining

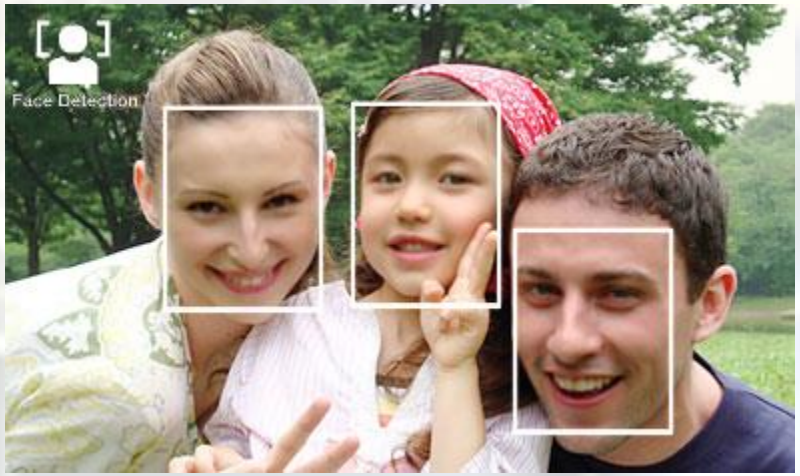
- Using machine learning to recommend books.

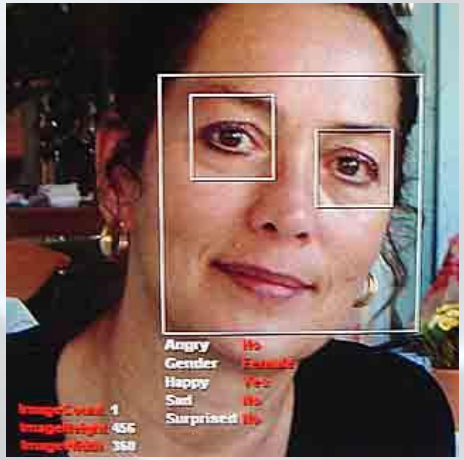
ALGORITHMS

Nearest Neighbour

Clustering

- Using machine learning to identify faces and expressions.



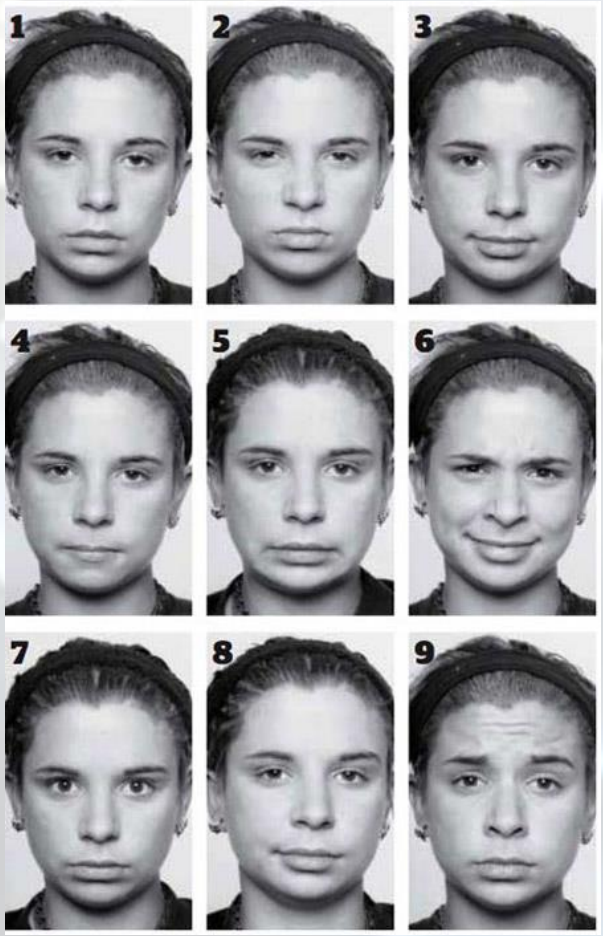


ALGORITHMS

Decision Trees

Adaboost

- Using machine learning to identify vocal patterns



ALGORITHMS

Feature Extraction
Probabilistic Classifiers
Support Vector Machines
HMM

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- ML for working with social network data: detecting fraud, predicting click-thru patterns, targeted advertising, etc etc etc .

facebook

amazon.com

ebay

Lloyds TSB

ALGORITHMS

Support Vector Machines
Rule mining algorithms
Semi-supervised Learning

New massive applications

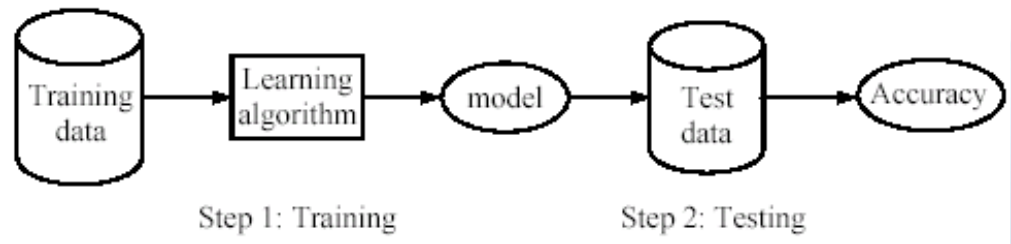
- **Data Security**
  Malware is a huge — and growing — problem. In 2014, Kaspersky Lab said it had detected 325,000 new malware files *every day*. But, institutional intelligence company Deep Instinct says that each piece of new malware tends to have almost the same code as previous versions — only between 2 and 10% of the files change from iteration to iteration.
- **Personal Security**
  Machine learning is proving that it can be an asset to help eliminate false alarms and spot things human screeners might miss in security screenings at airports, stadiums, concerts, and other venues.
- **Financial Trading**
  machine learning algorithms are getting closer all the time. Many prestigious trading firms use proprietary systems to predict and execute trades at high speeds and high volume. Many of these rely on probabilities.
- **Healthcare**
  Machine learning algorithms can process more information and spot more patterns than their human counterparts.
  The more you can understand about your customers, the better you can serve them, and the more you will sell.
- **Fraud Detection**
  Machine learning is getting better and better at spotting potential cases of fraud across many different fields. PayPal, for example, is using machine learning to fight money laundering. The company has tools that compare millions of transactions and can precisely distinguish between legitimate and fraudulent transactions between buyers and sellers.
- **Recommendations**
  Services like Amazon or Netflix. Intelligent machine learning algorithms analyze your activity and compare it to the millions of other users to determine what you might like to buy or binge watch next.
- **Online Search**
  Perhaps the most famous use of machine learning, Google and its competitors are constantly improving what the search engine understands.
- **Natural Language Processing (NLP)**
  Machine learning algorithms with natural language can stand in for customer service agents and more quickly route customers to the information they need.
- **Smart Cars**
  IBM recently surveyed top auto executives, and 74% expected that we would see smart cars on the road by 2025.

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



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

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

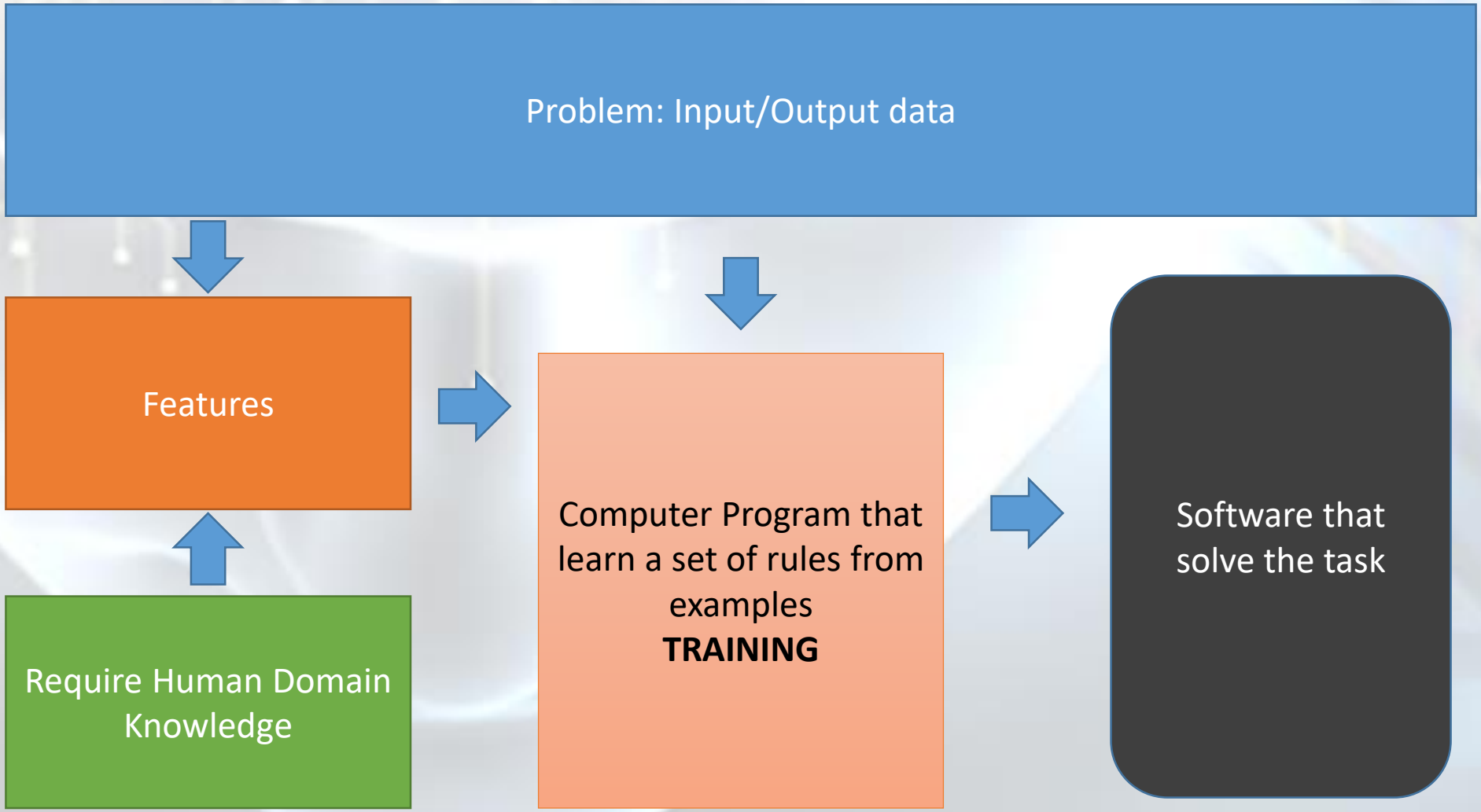
Assumption: The distribution of training examples is identical to the distribution of test examples (including future unseen examples).

- 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.

Supervised Learning well known techniques

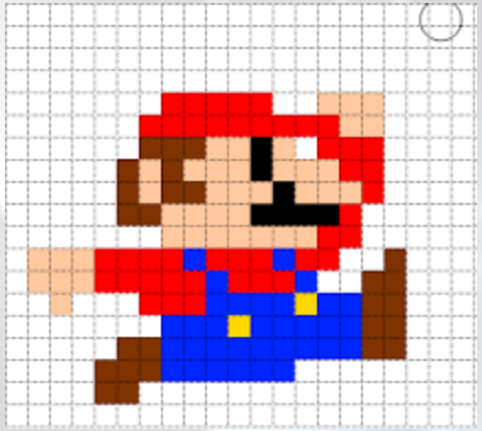
- **Bayes Calssifier** -> Learn a posteriori probability or pdf
- **Linear classifiers LDA, Logistic regression-**> Learn a set of coefficient of a separating plane in a hyperspace
- **SVM** -> linear/non-linear learn coefficients
- **Ensemble methods** -> learn to combine linear classifiers

Learning Pipeline



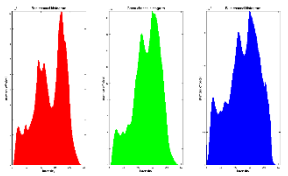
Supervised learning example

- From Data to Features (Design Choice)

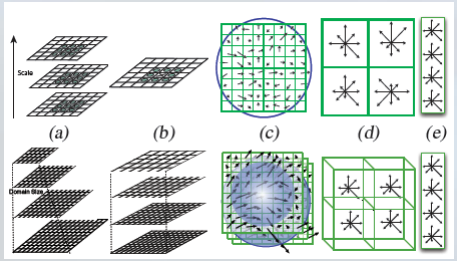


Possible representations:

- **Pixels {[R,G,B]1,[RGB],2 ….}**
- **Pixels+coordinates {[R,G,B,x,y]1….,}**
- **Histograms:**



- **Complex Features:**



Domain dependant and effectiveness varies from problem to problem

Supervised learning example cont

- Features are numbers
- Features are vectors of numbers

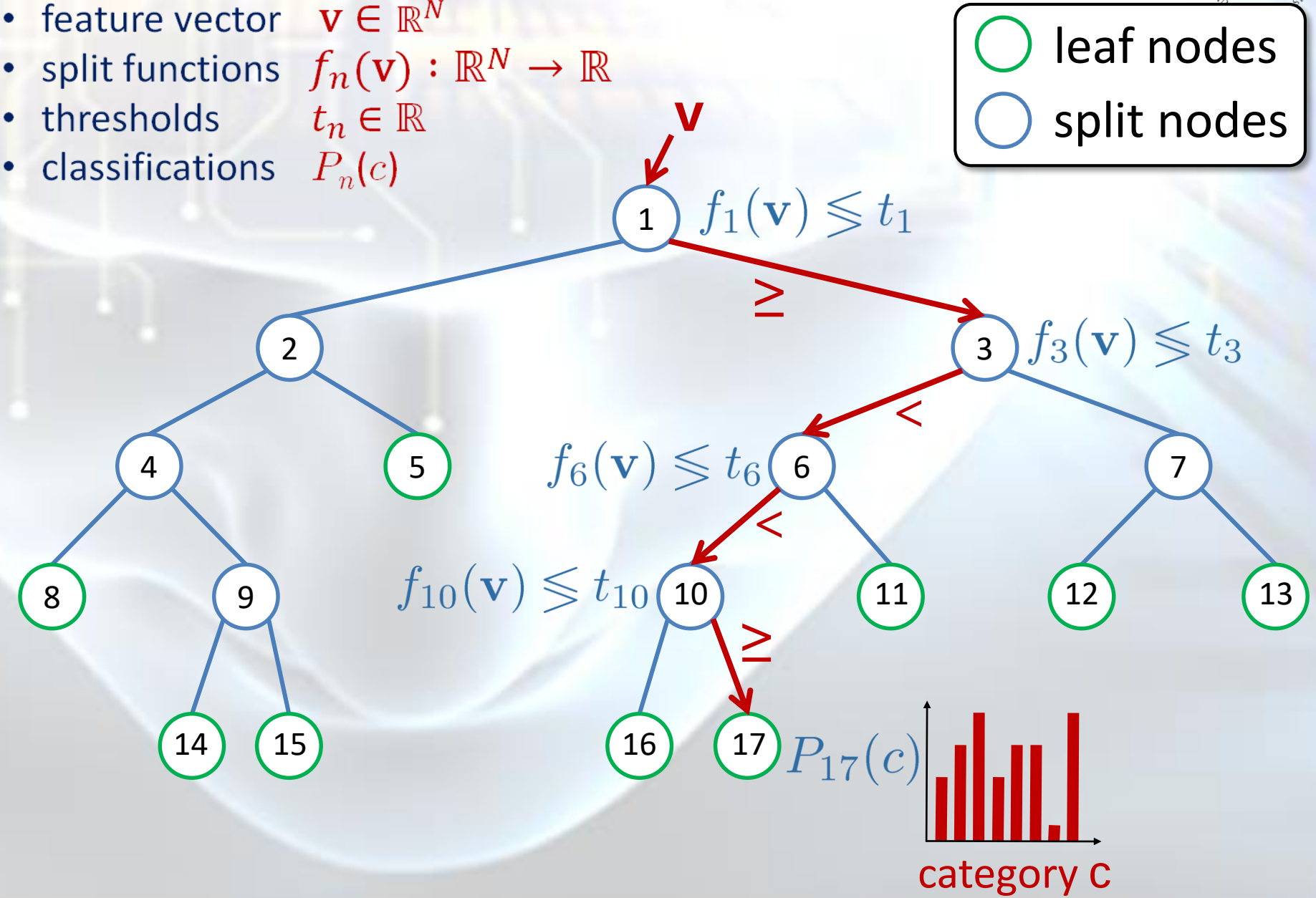
- Every classifier combine features and parameters in order to take a decision (often binary)-> e.g. «Is Mario in the Image?» Yes or No

Optimization problem over a dataset

The Basics: **Binary Decision Trees**

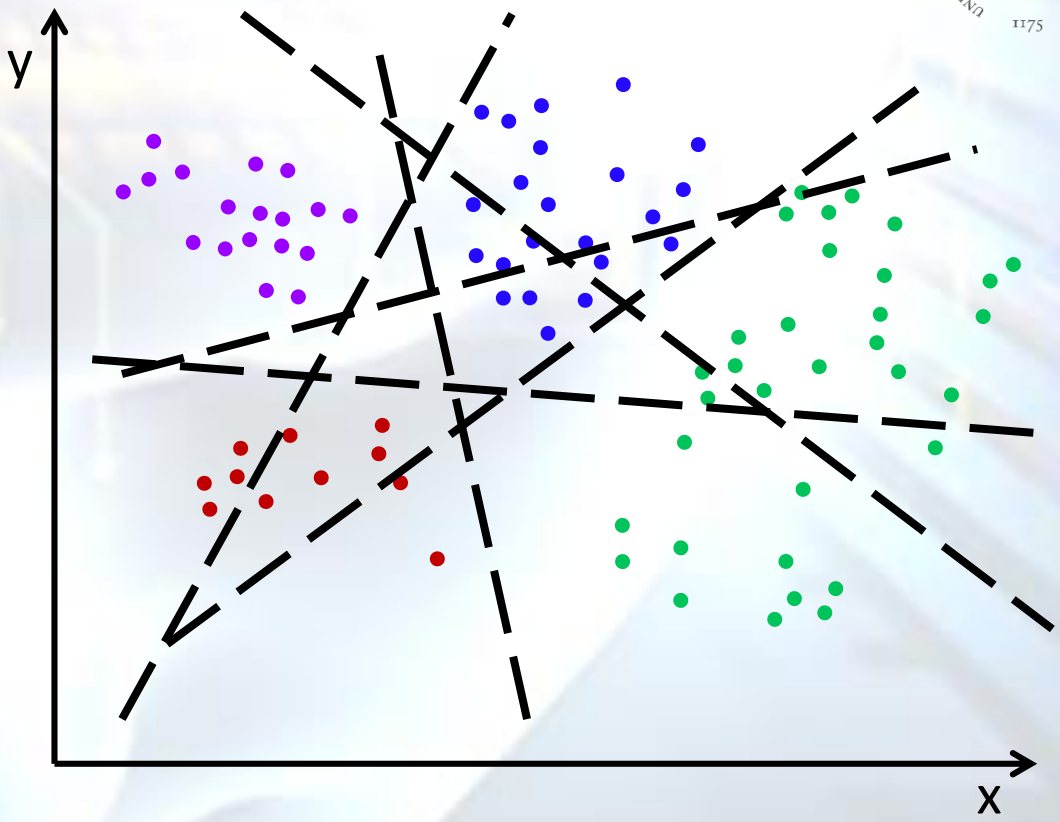
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- feature vector $\mathbf{v} \in \mathbb{R}^N$
- split functions $f_n(\mathbf{v}) : \mathbb{R}^N \rightarrow \mathbb{R}$
- thresholds $t_n \in \mathbb{R}$
- classifications $P_n(c)$



Toy Learning Example

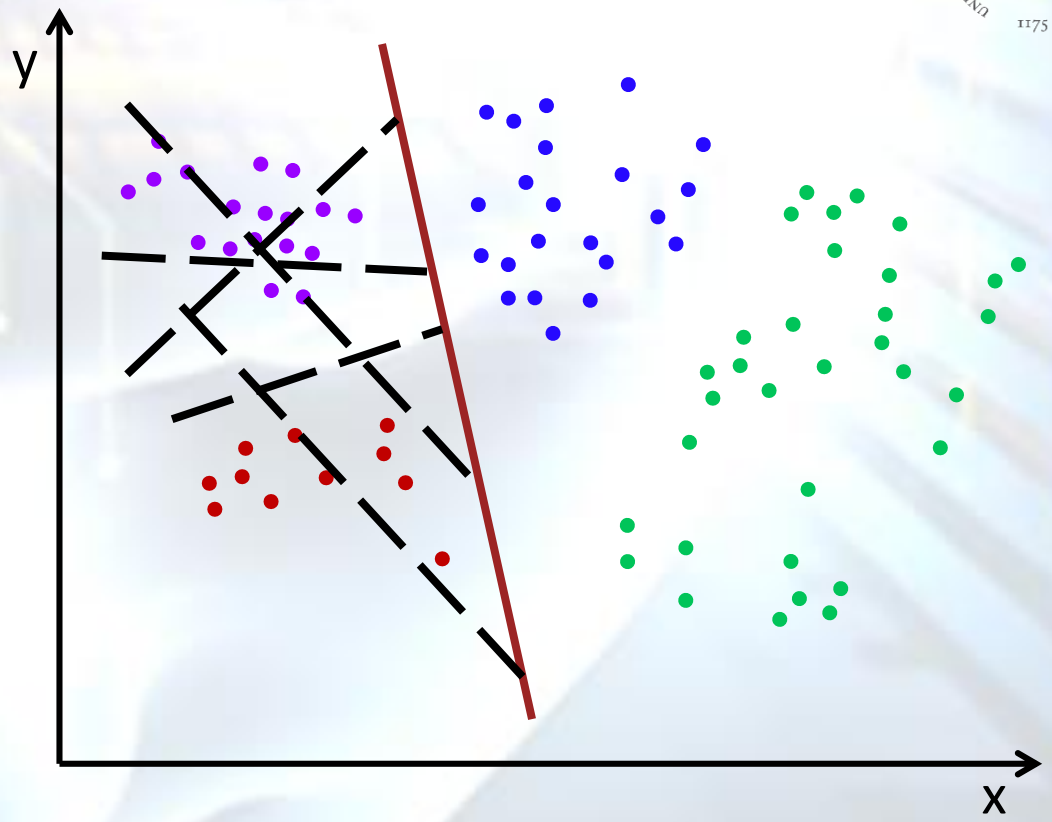
- **Try several lines, chosen at random**
- **Keep line that best separates data**
  - information gain
- **Recurse**



- feature vectors are x, y coordinates: $\mathbf{v} = [x, y]^T$
- split functions are lines with parameters a, b: $f_n(\mathbf{v}) = ax + by$
- threshold determines intercepts: t_n
- four classes: purple, blue, red, green

Toy Learning Example

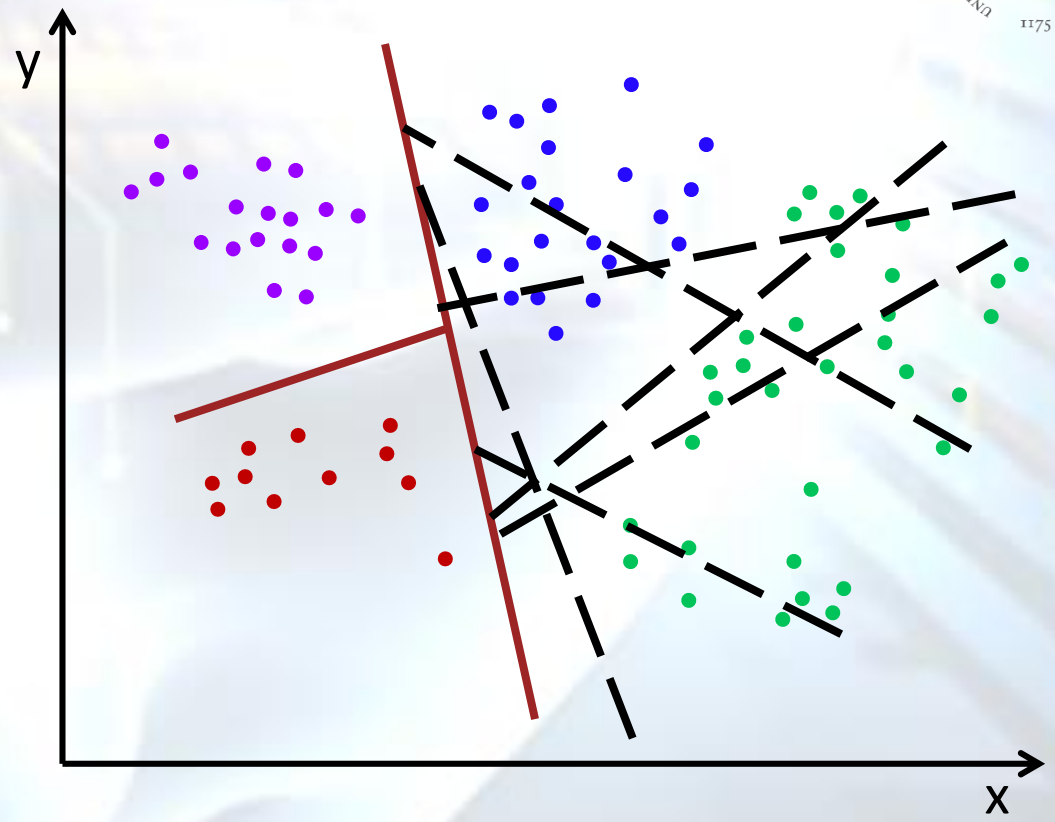
- **Try several lines, chosen at random**
- **Keep line that best separates data**
  - information gain
- **Recurse**



- feature vectors are x, y coordinates: $\mathbf{v} = [x, y]^T$
- split functions are lines with parameters a, b: $f_n(\mathbf{v}) = ax + by$
- threshold determines intercepts: t_n
- four classes: purple, blue, red, green

Toy Learning Example

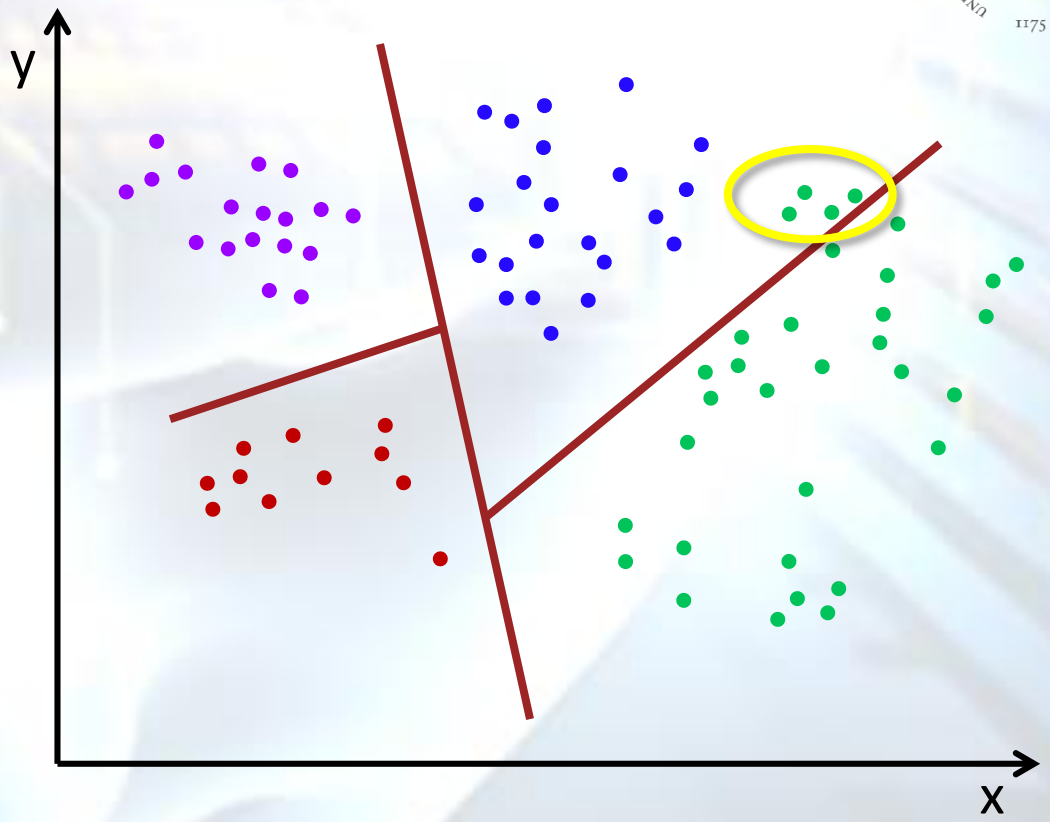
- **Try several lines, chosen at random**
- **Keep line that best separates data**
  - information gain
- **Recurse**



- feature vectors are x, y coordinates: $\mathbf{v} = [x, y]^T$
- split functions are lines with parameters a, b: $f_n(\mathbf{v}) = ax + by$
- threshold determines intercepts: t_n
- four classes: purple, blue, red, green

Toy Learning Example

- Try several lines, chosen at random
- Keep line that best separates data
  - information gain
- Recurse



- feature vectors are x, y coordinates: $\mathbf{v} = [x, y]^T$
- split functions are lines with parameters a, b: $f_n(\mathbf{v}) = ax + by$
- threshold determines intercepts: t_n
- four classes: purple, blue, red, green

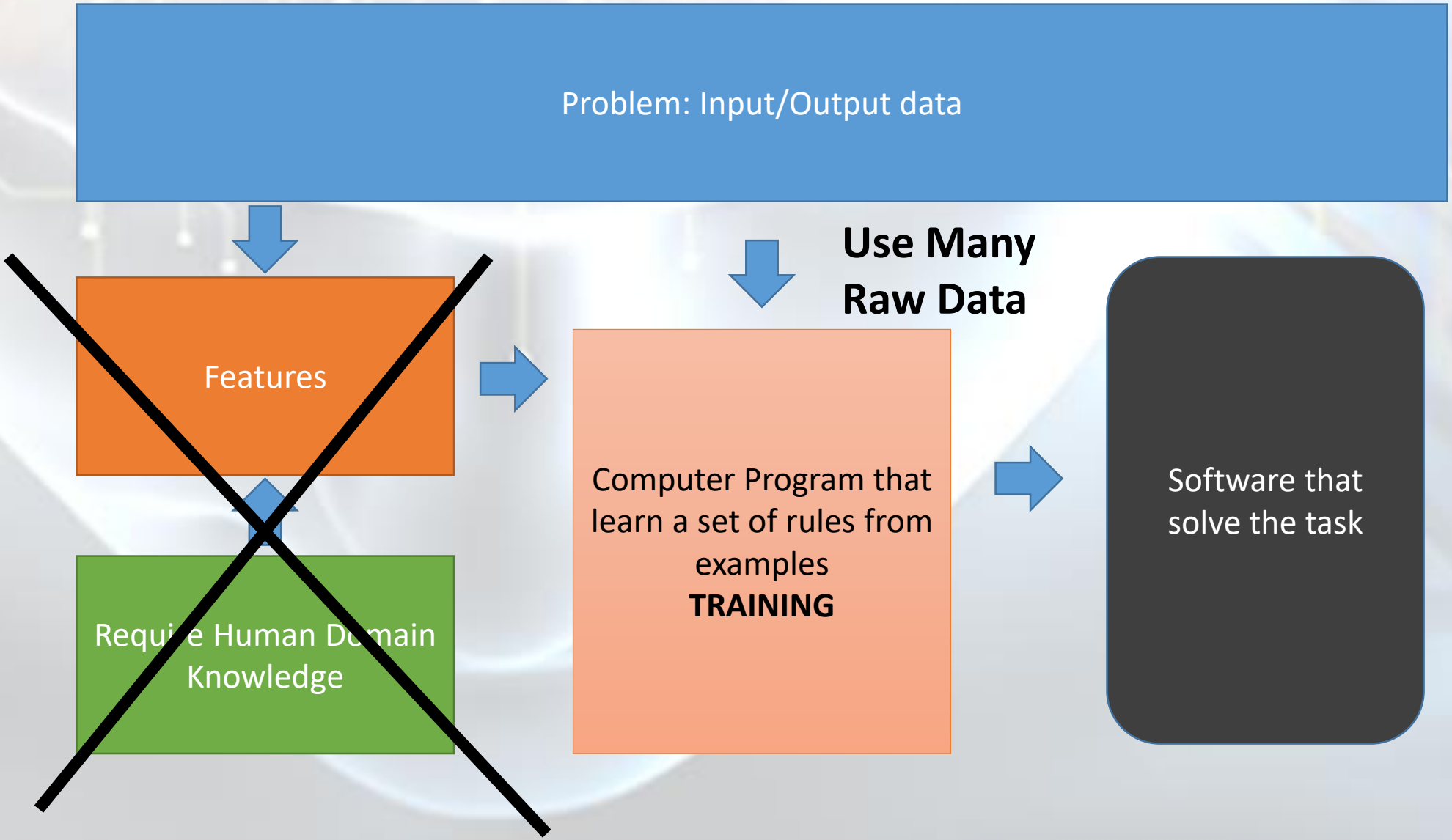
Flaw of the new millennium

- There is a failure in the process:

«Features are a decent representation of the data but contains human knowledge that bias the results»

DEEP Learning innovation

Innovation of Deep Learning



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

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

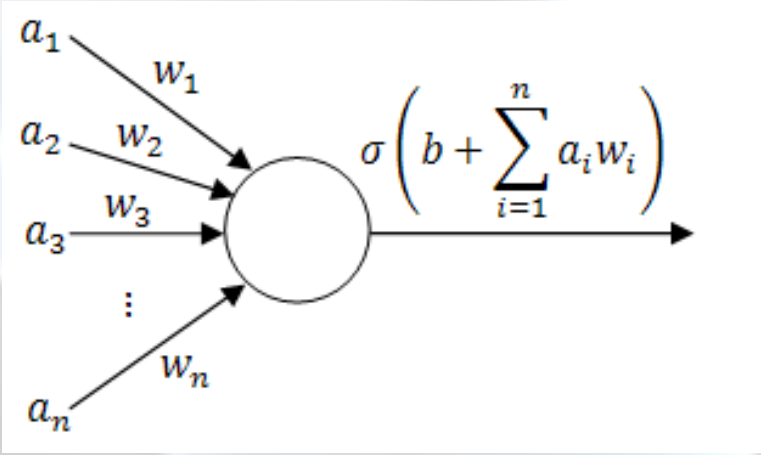
Technology	Accuracy
Tesseract eng language	40%
Tesseract trained language	60%
DEEP neural network(NN)	**98%**

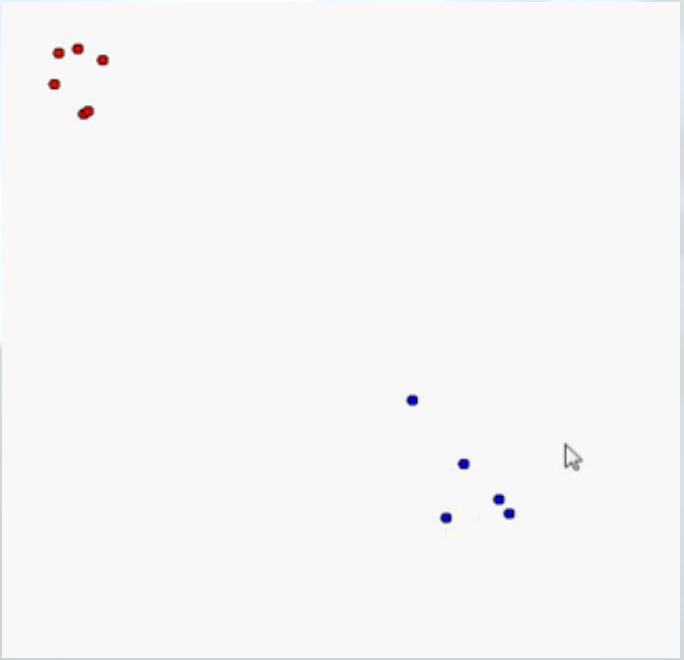
DEMO code @
http://christopher5106.github.io/computer/vision/2015/09/14/comparing-tesseract-and-deep-learning-for-ocr-optical-character-recognition.html

From the Perceptron to …….

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

Perceptron is a linear Classifier





How to Train it?

Gradient Descent and the Delta Rule

- The key idea of delta rule: to use *gradient descent* to search the space of possible weight vector to find the weights that best fit the training examples. This rule is important because it provides the *basis* for the backpropagration algorithm, which can learn networks with many interconnected units.

- The delta training rule: considering the task of training an unthresholded perceptron, that is a *linear unit*, for which the output o is given by:

$$o = w_0 + w_1 x_1 + \cdots + w_n x_n \quad (1)$$

- Thus, a linear unit corresponds to the first stage of a perceptron, without the threshold.

- In order to derive a weight learning rule for linear units, let specify a measure for the ***training error*** of a weight vector, relative to the training examples. The Training Error can be computed as the following squared error

$$E[\vec{w}] \equiv \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2 \qquad (2)$$

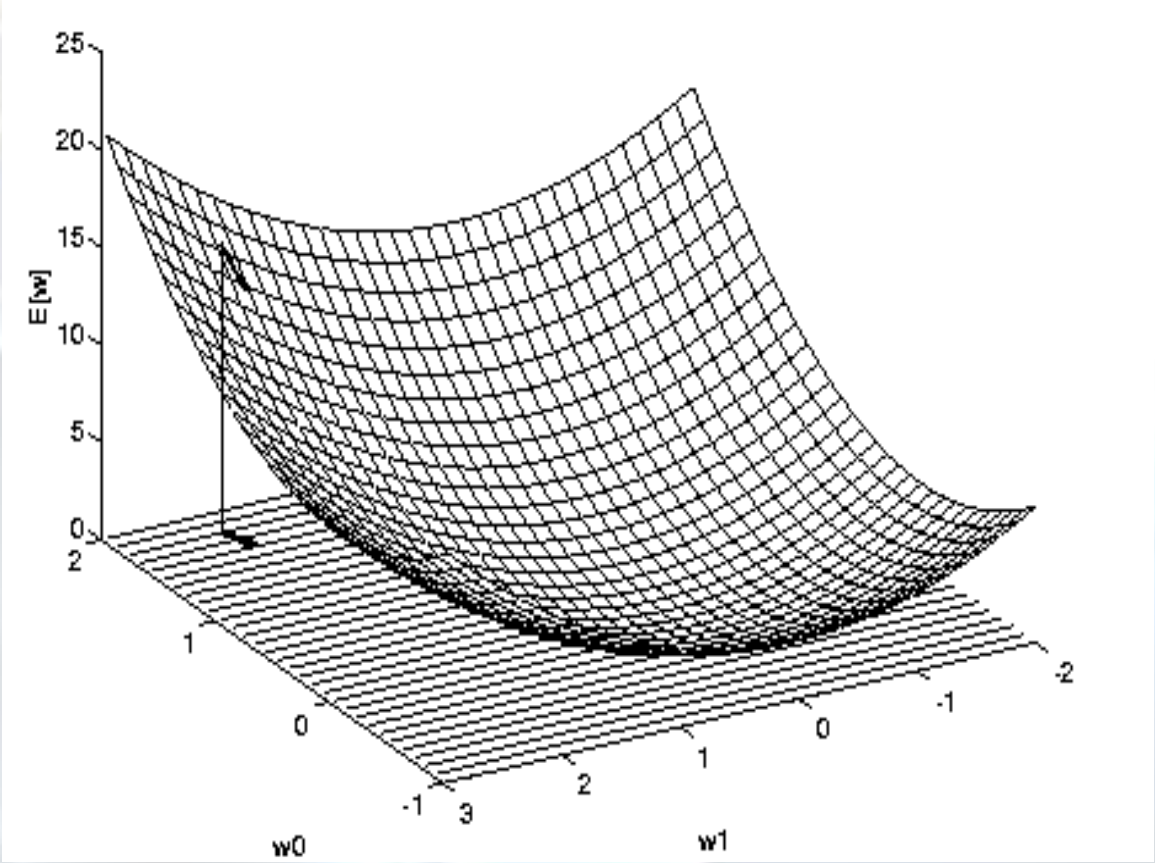
where D is set of training examples, t_d is the target output for the training example d and o_d is the output of the linear unit for the training example d.

Here we characterize E as ***a function of weight vector*** because the linear unit output O depends on this weight vector.

Hypothesis Space

- To understand the gradient descent algorithm, it is helpful to visualize the entire space of possible weight vectors and their associated *E* values, as illustrated in Figure 5.
  - Here the axes w_0,w_1 represents possible values for the two weights of a simple linear unit. The w_0,w_1 plane represents the entire hypothesis space.
  - The vertical axis indicates the error *E* relative to some fixed set of training examples. The error surface shown in the figure summarizes the desirability of every weight vector in the hypothesis space.
- For linear units, this **error surface** must be **parabolic** with a single global minimum. And we desire a weight vector with this minimum.

Figure 5. The error surface



How can we calculate the direction of steepest descent along the error surface? This direction can be found by computing the derivative of E w.r.t. each component of the vector w.

Derivation of the Gradient Descent Rule

- This vector derivative is called the *gradient* of E with respect to the vector $\langle w_0,\ldots,w_n\rangle$, written ∇E .

$$\nabla E[\vec{w}] \equiv \left[\frac{\partial E}{\partial w_0}, \frac{\partial E}{\partial w_1}, \cdots \frac{\partial E}{\partial w_n}\right] \quad (3)$$

Notice ∇E is itself a vector, whose components are the partial derivatives of E with respect to each of the w_i. When interpreted as a vector in weight space, the gradient specifies the **direction** that produces the steepest increase in E. The negative of this vector therefore gives the direction of steepest decrease.

Since the gradient specifies the direction of steepest increase of E, the training rule for gradient descent is

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

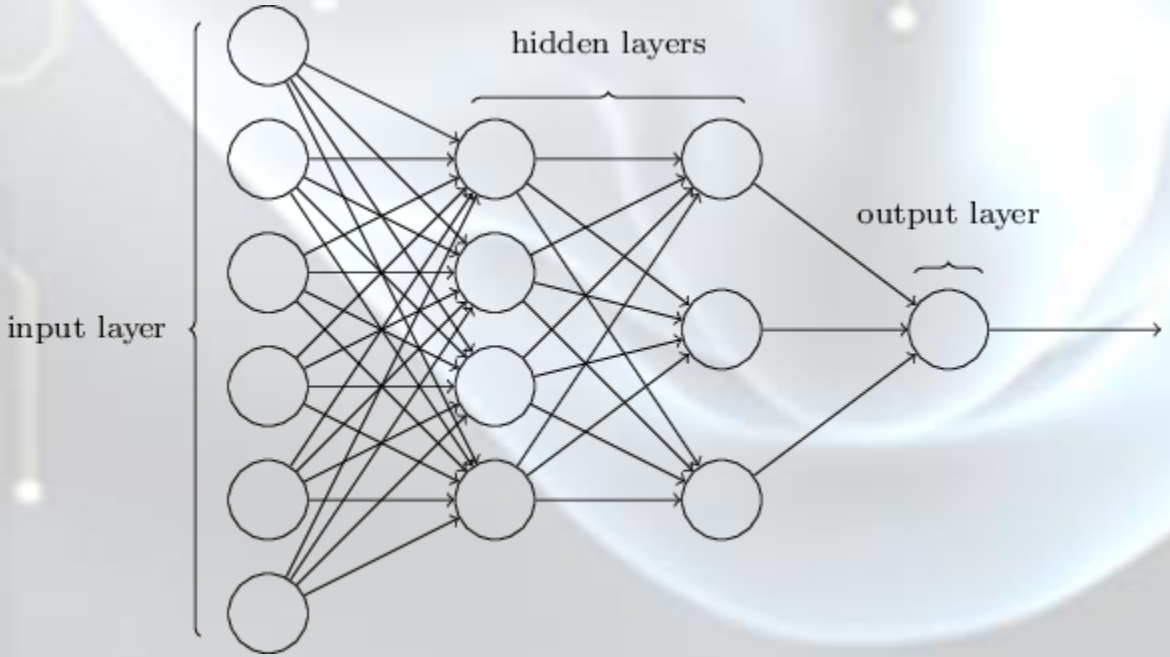
where

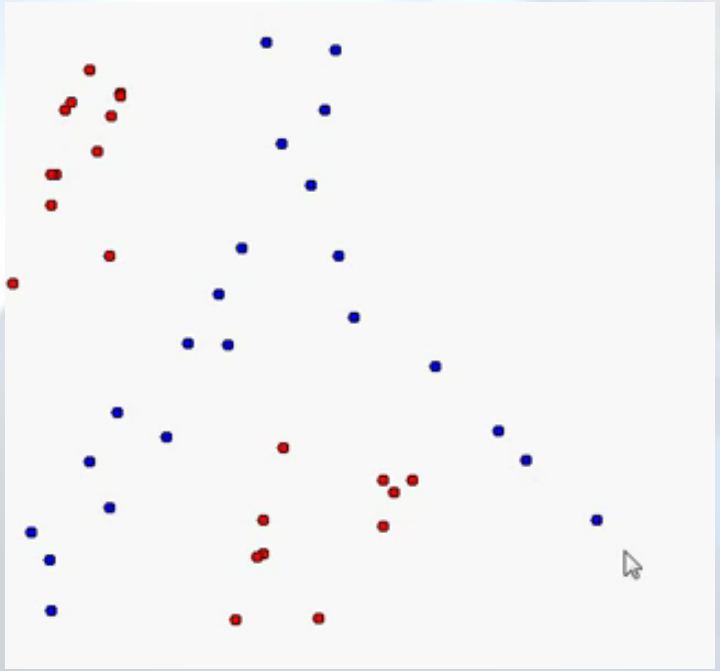
$$\Delta\vec{w} = -\eta\nabla E[\vec{w}] \quad (4)$$

Multilayered Neural Network

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

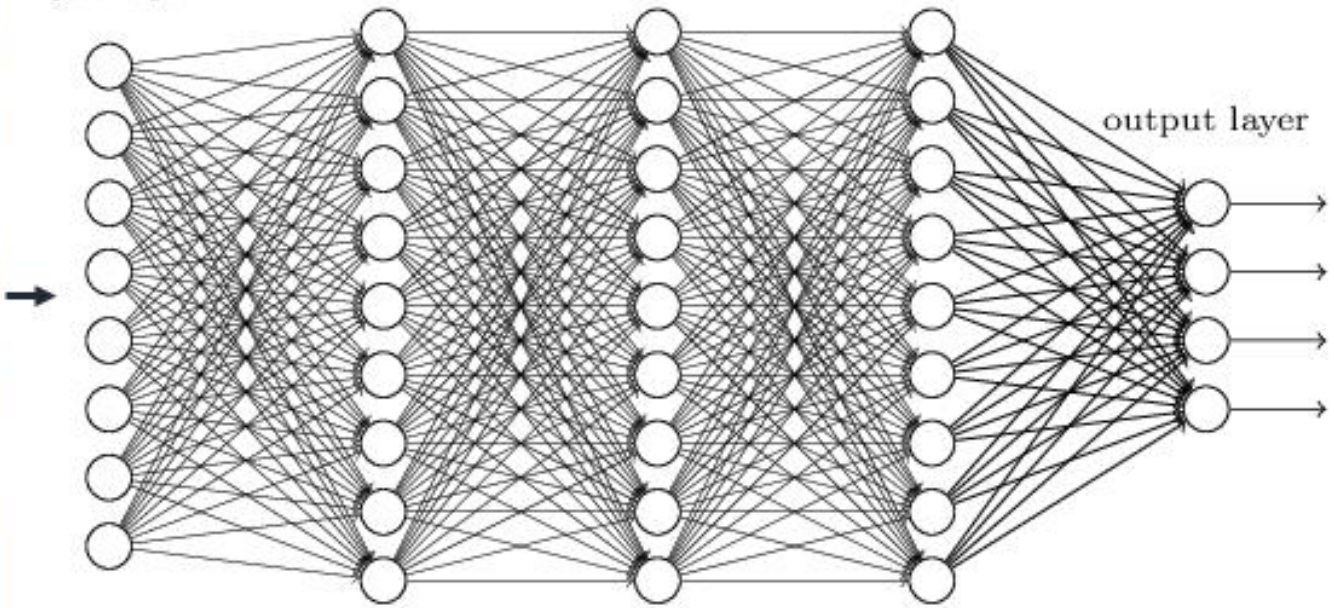
Network With 3 layers is a non-linear classifer





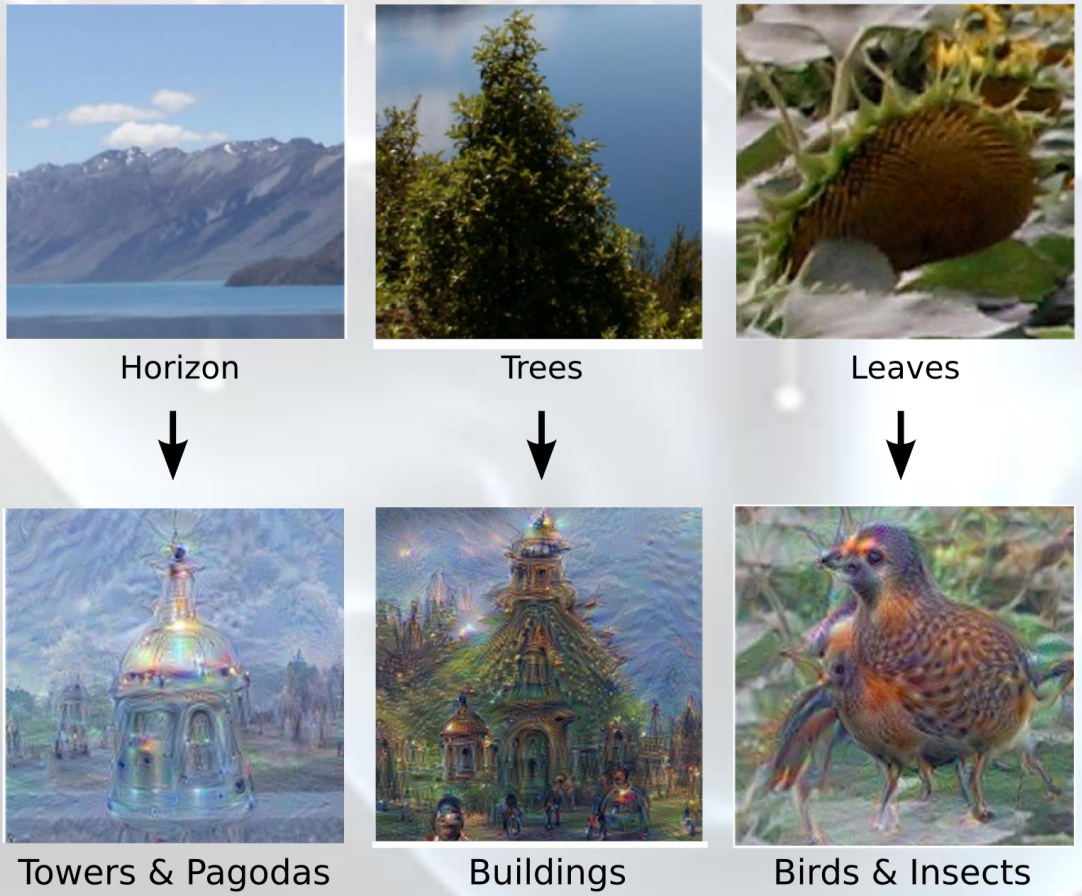
Going Deep

Deep neural networks learn hierarchical feature representations



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

How the network learn the world



Google inception program
https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Deep Networks For:

Numerical Data -> Deep Neural Network

Applications: Production management, Prediction, Controls and Robotics

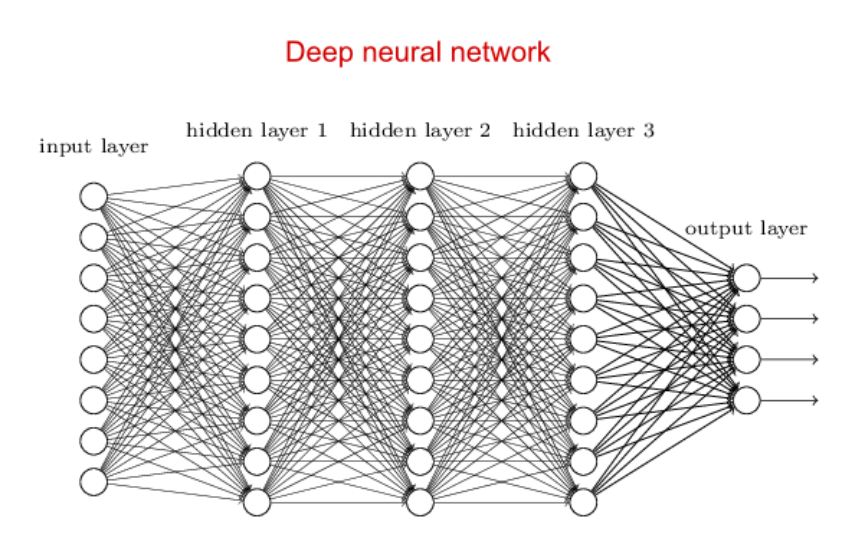
Multimedia Data-> Convolutional Network

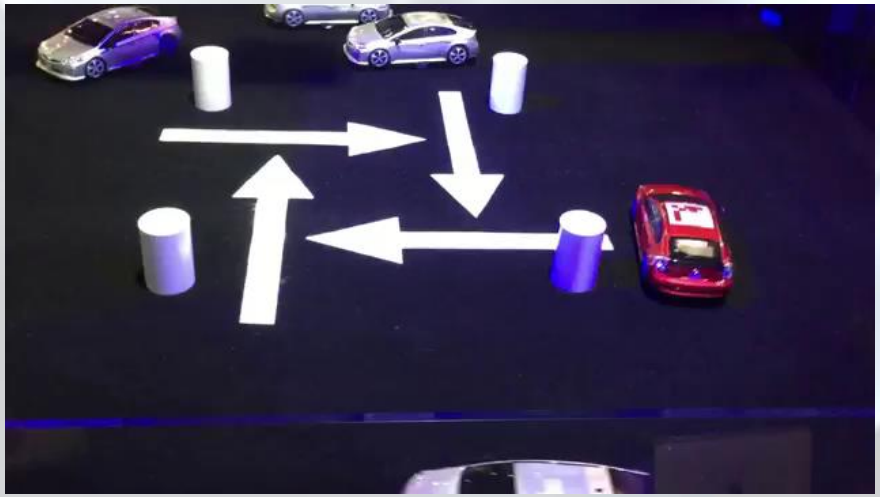
Applications: Image and Video classification, Face recognition, Licence Plate Detection, OCRs..

Time series -> Recurrent Neural Network

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

Numerical Data -> Deep Neural Network





Pros:

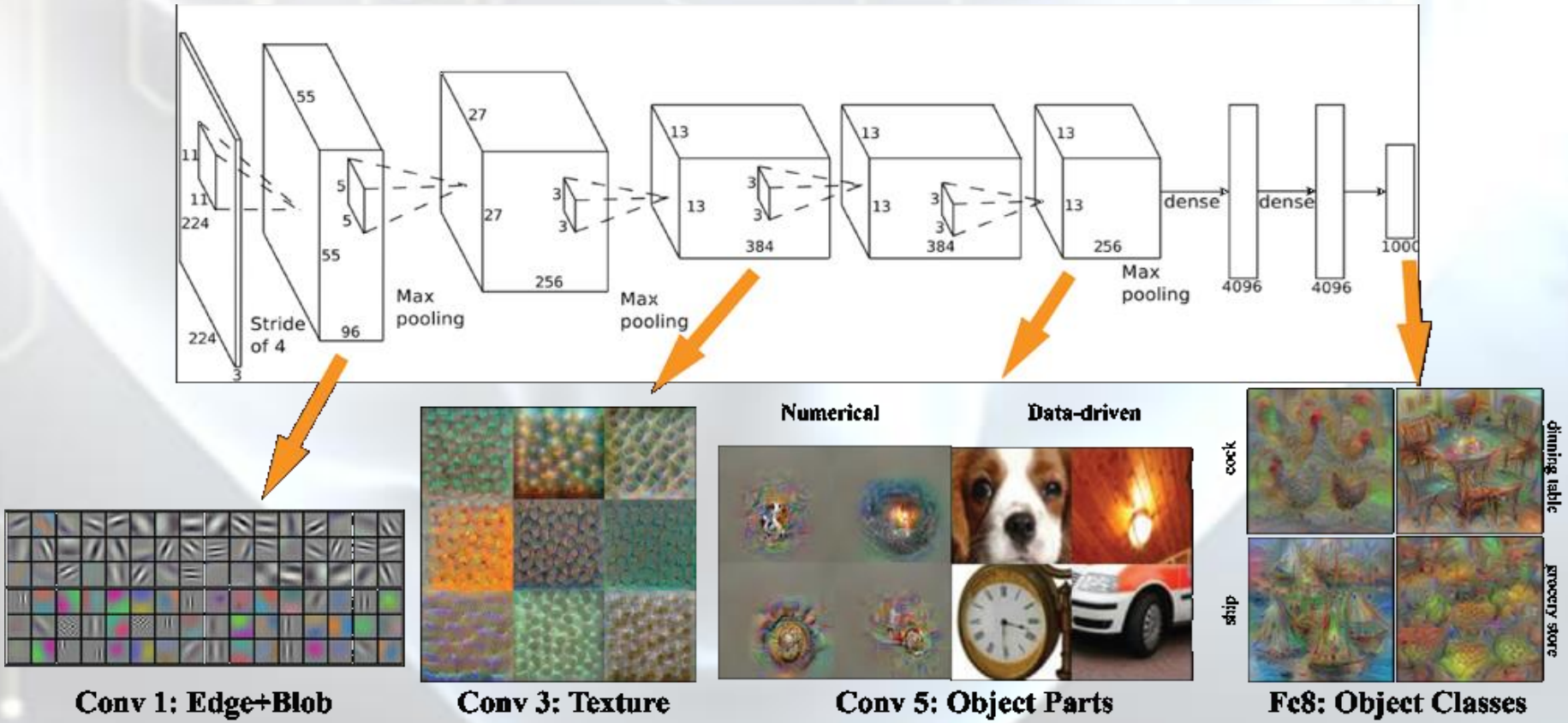
- Use Digital Sensors data as input
- Theoretically can learn every classification function
- Can predict a flexible number of outcomes

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

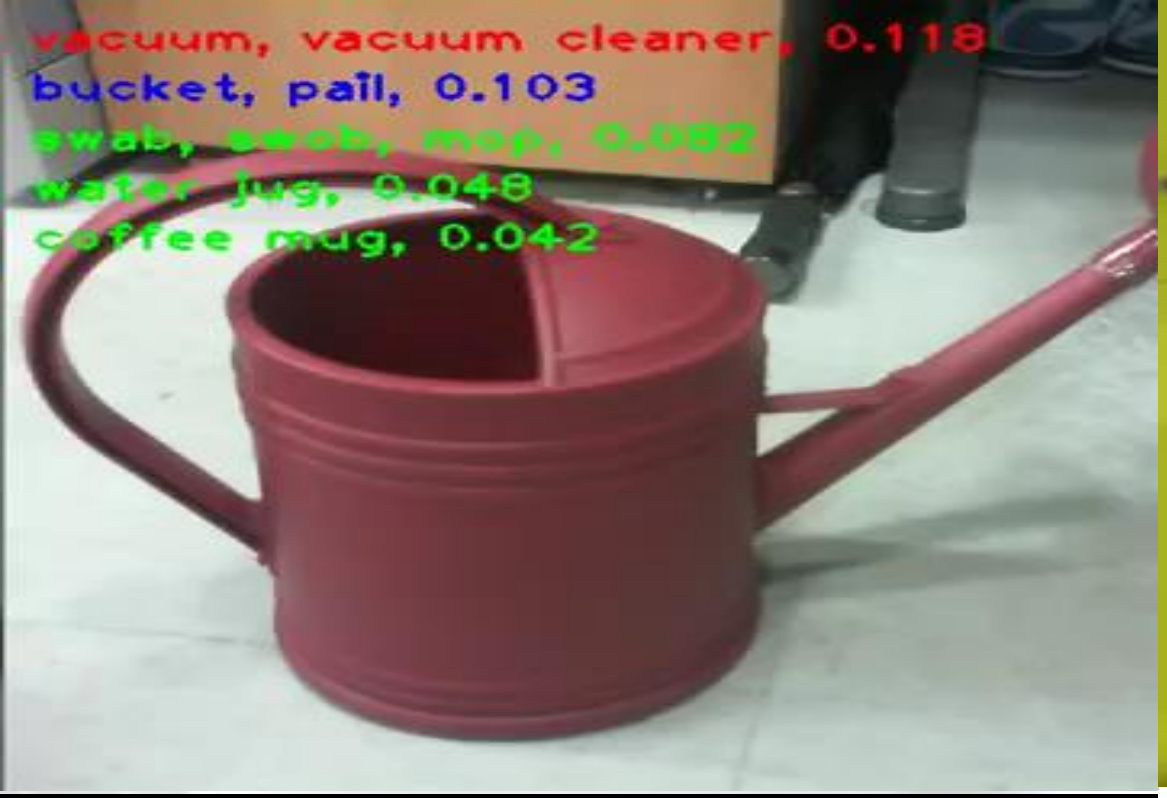


Pros:

- 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

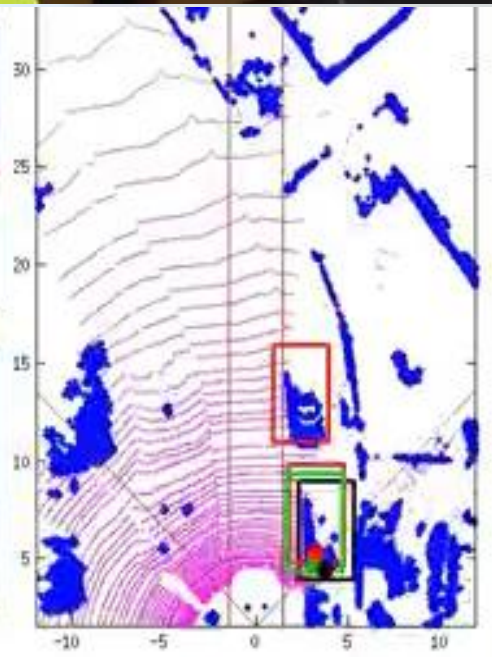


Google DeepMind's Deep Q-learning

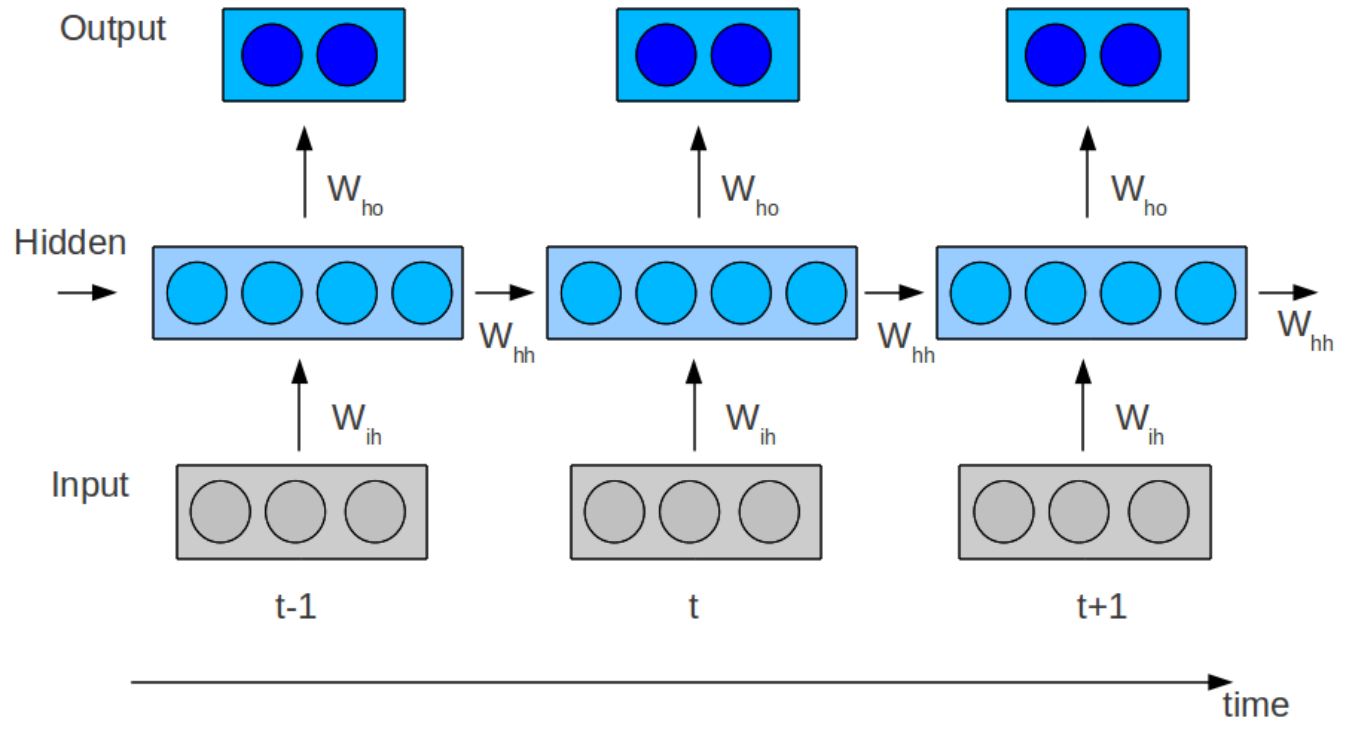
The algorithm will play Atari breakout.

The most important thing to know is that all the agent is given is sensory input (what you see on the screen) and it was ordered to maximize the score on the screen.

No domain knowledge is involved! This means that the algorithm doesn't know the concept of a ball or what the controls exactly do.



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

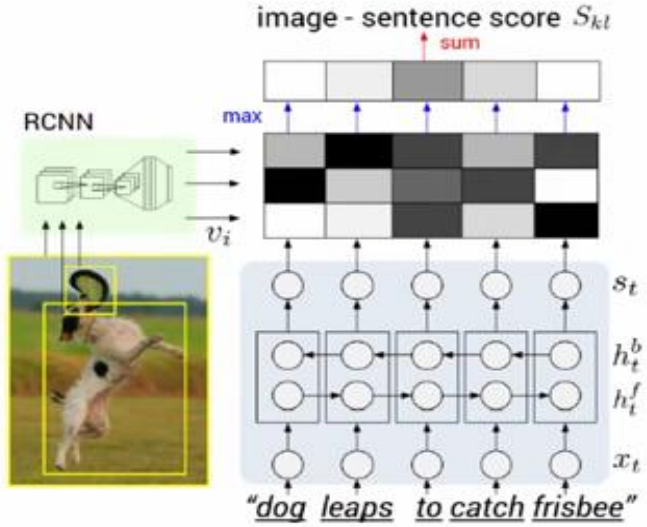
Source: Andrej Karpathy

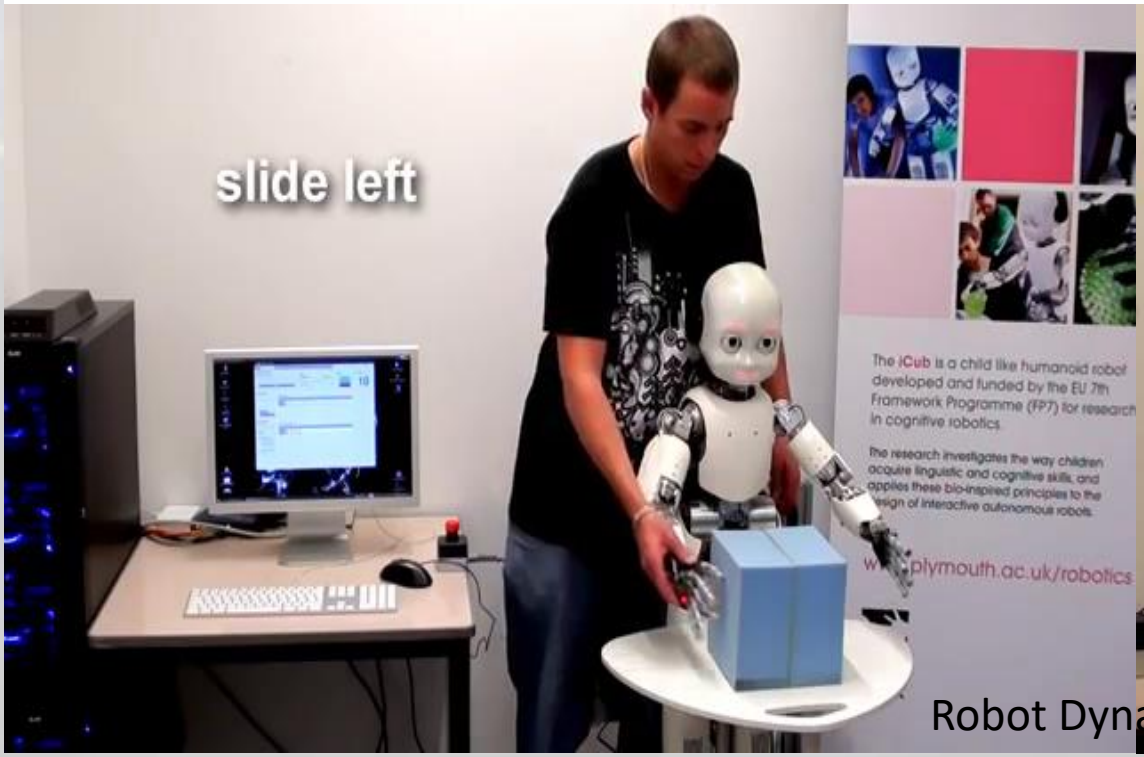
Text and Music Writing

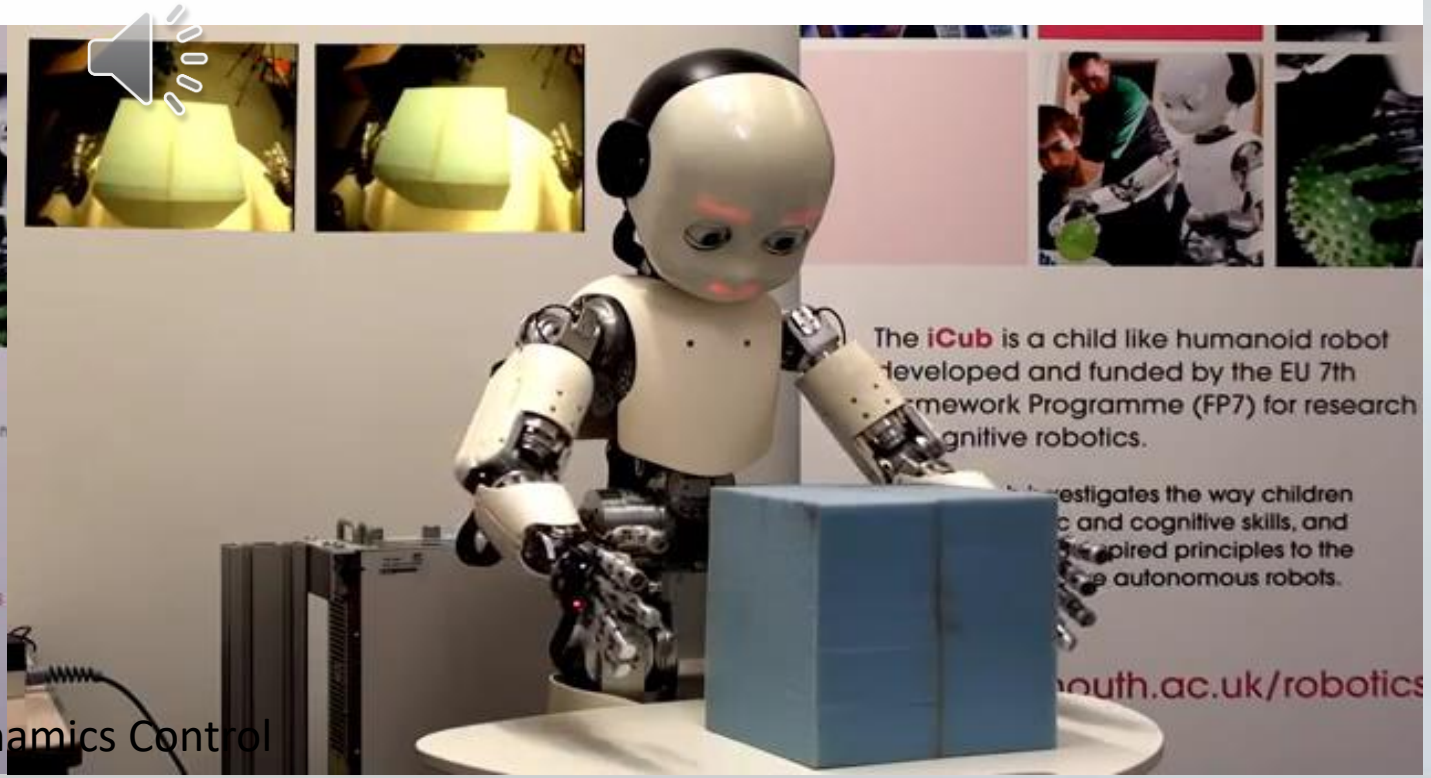
```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

```
"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

Welcome to ImageLab

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Who

4 Staff people, 8 Phd Students, 6 Research assistants

What

research in some AI-based fields: computer vision, pattern recognition, machine learning, deep learning; sensors data and image processing , HCI

When

Since 1998 @ UNIMORE in Modena DIEF, since 2011 within Modena Technopole

www.imagelab.unimore.it

Facebook AI Research Launches Partnership Program

Facebook AI Research (FAIR)

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Imagelab

- Who
  - 4 Staff people, 8 Phd Students, 6 Research assistants
- What
  - research in some AI-based fields: computer vision, pattern recognition, machine learning, deep learning; sensors data and image processing , HCI
- When
  - Since 1998 @ UNIMORE in Modena DIEF, since 2011 within Modena Technopole




Projects and Collaboration (2017)

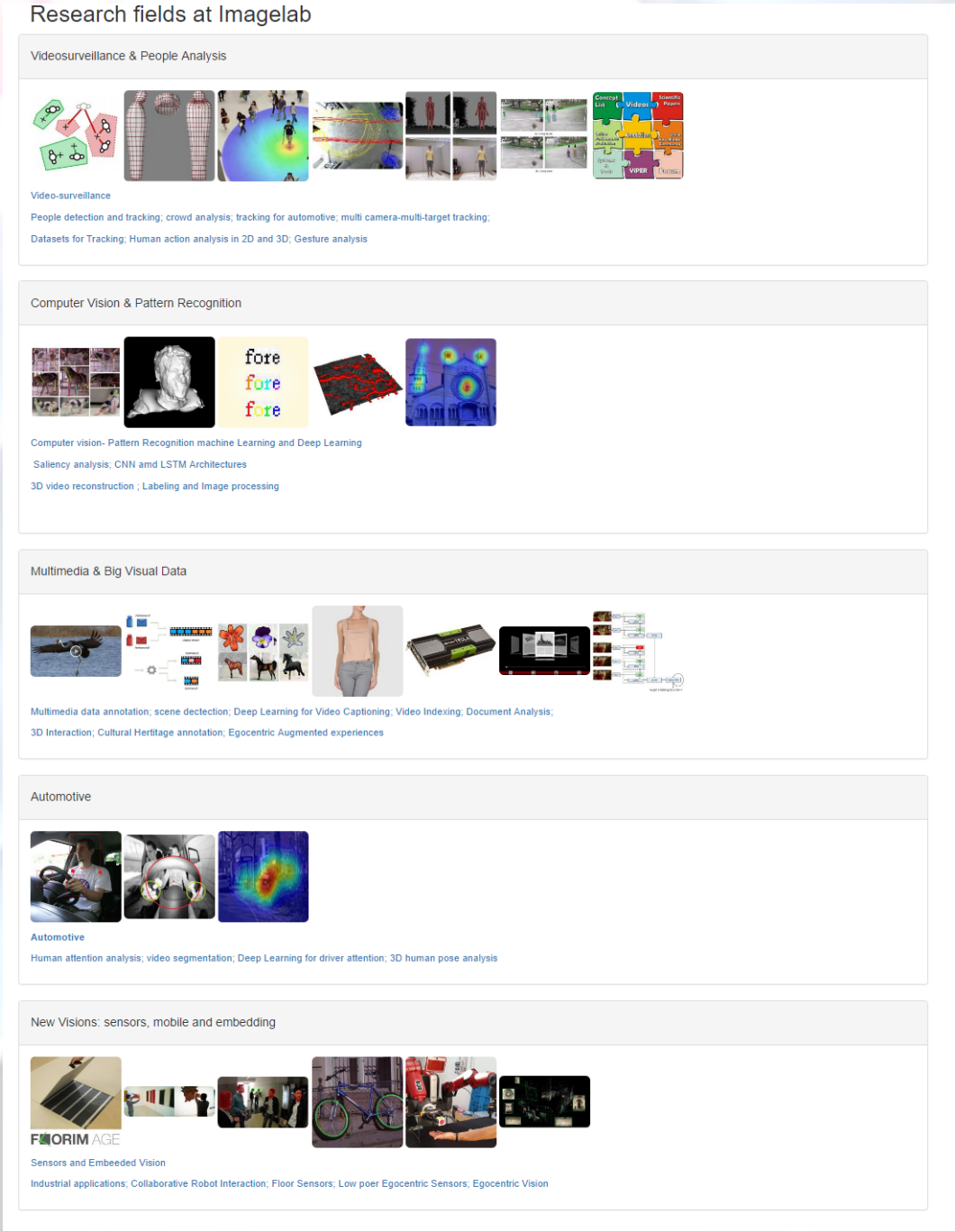
http://imagelab.ing.unimore.it/imagelab/projects.asp

- Industrial Projects:
  - Maserati, Ferrari, Voilap, Sata, …
- Regional projects:
  - Jump, Sacher, (EU FESR 2015-2017)
  - Drivid Driver attention (UNIMORE- Stanford project)
  - HixIA
  - Vaex Augmented experience in cultural heritage
- National Projects:
  - Cluster Smart City and Community « Città Educante» (Italian MIUR- Project 2016-2018)
  - MIUR Progetti Rilevante Interesse Nazionali COSMOS (2017-2019)
- International collaborations:
  - **Facebook Artificial Intelligence Research** : selected as 15 FAIR European Labs
  - ISCRA **Italian SuperComputing Resource Allocation – CINECA 2016.2017**,
  - **EU Horizon 2020** Advance
  - Panasonic CA, USA Phd Student exchange

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Research activities

- Videosurveillance & Human Behavior Understanding
- Computer Vision, Pattern recognition Machine Learning and Deep Learning
- Multimedia and Visual Big Data
- Automotive
- Sensors mobile and embedded vision



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Case Studies:
ML models for reproducing **Human Behavior**

- Gaze and Attention -> bottom up and task driven
- Attitude-> Spotting prejudice with camera sensors
- Scene understanding -> from Video to Caption
- Security and Surveillance -> People Guessing

Gaze and Attention

What is Saliency?

- The saliency of an item (an object, a person, a pixel, etc.) is the state or quality by which it stands out relative to its neighbors.
- Classical algorithms for saliency prediction focused on identifying the **fixation points** that human viewer would focus on at first glance.

Original Image

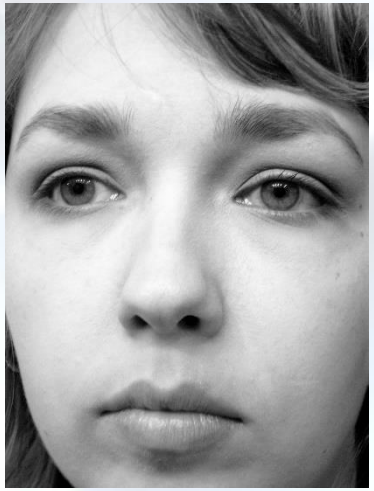
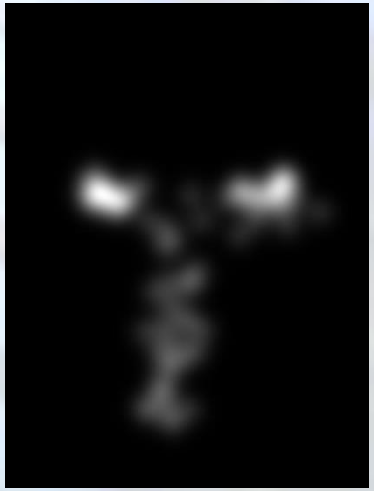


Image with fixation points

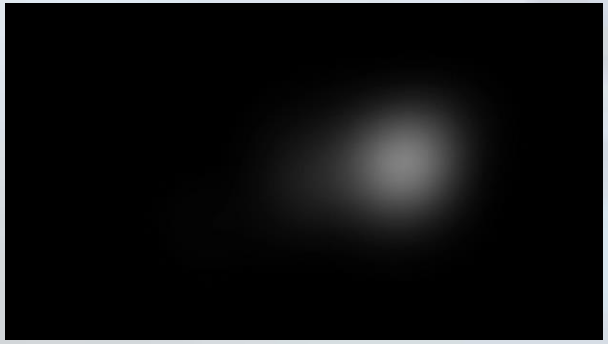
Saliency Map



Original Video

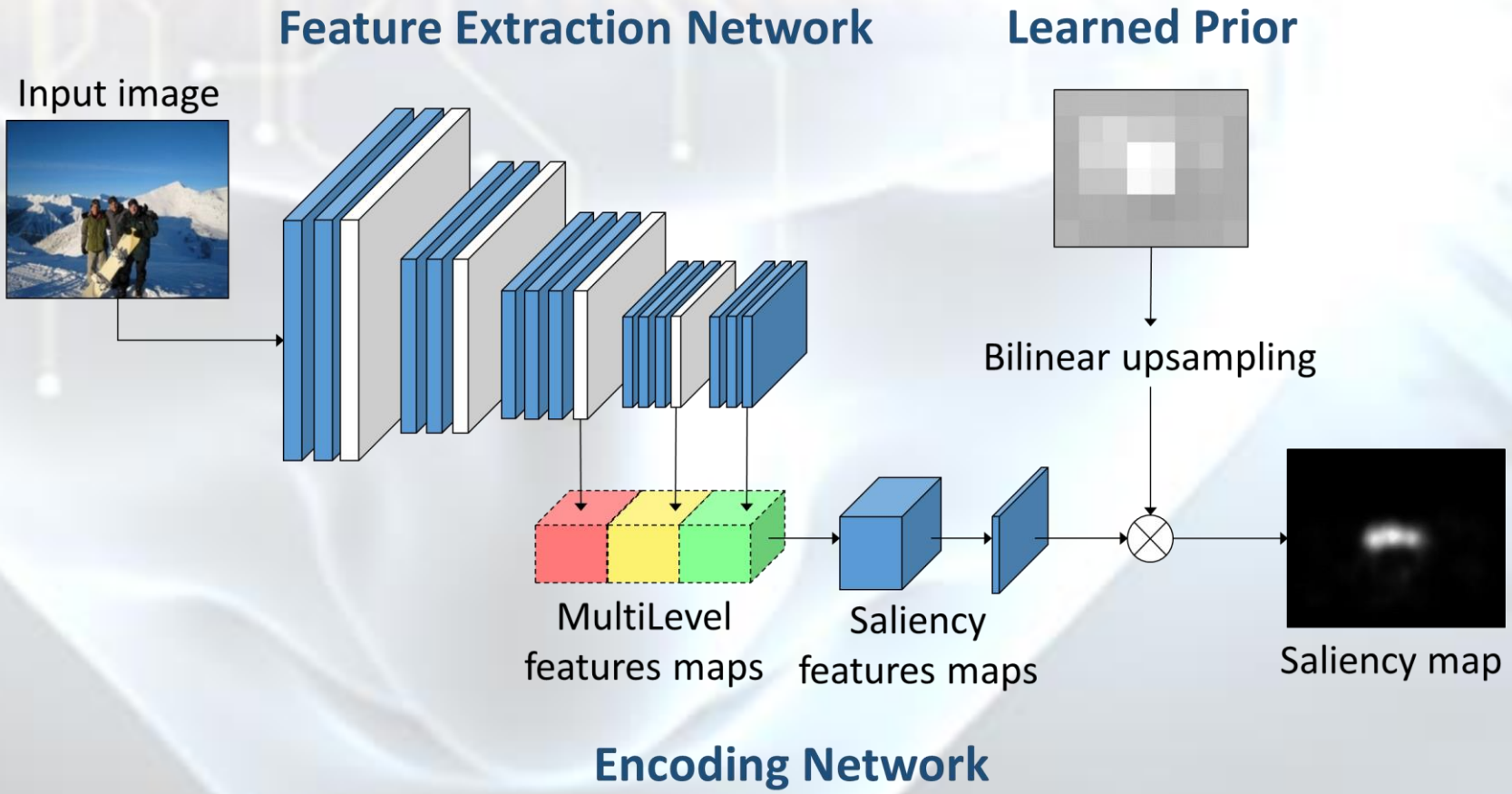
Video with fixation points

Saliency Map



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

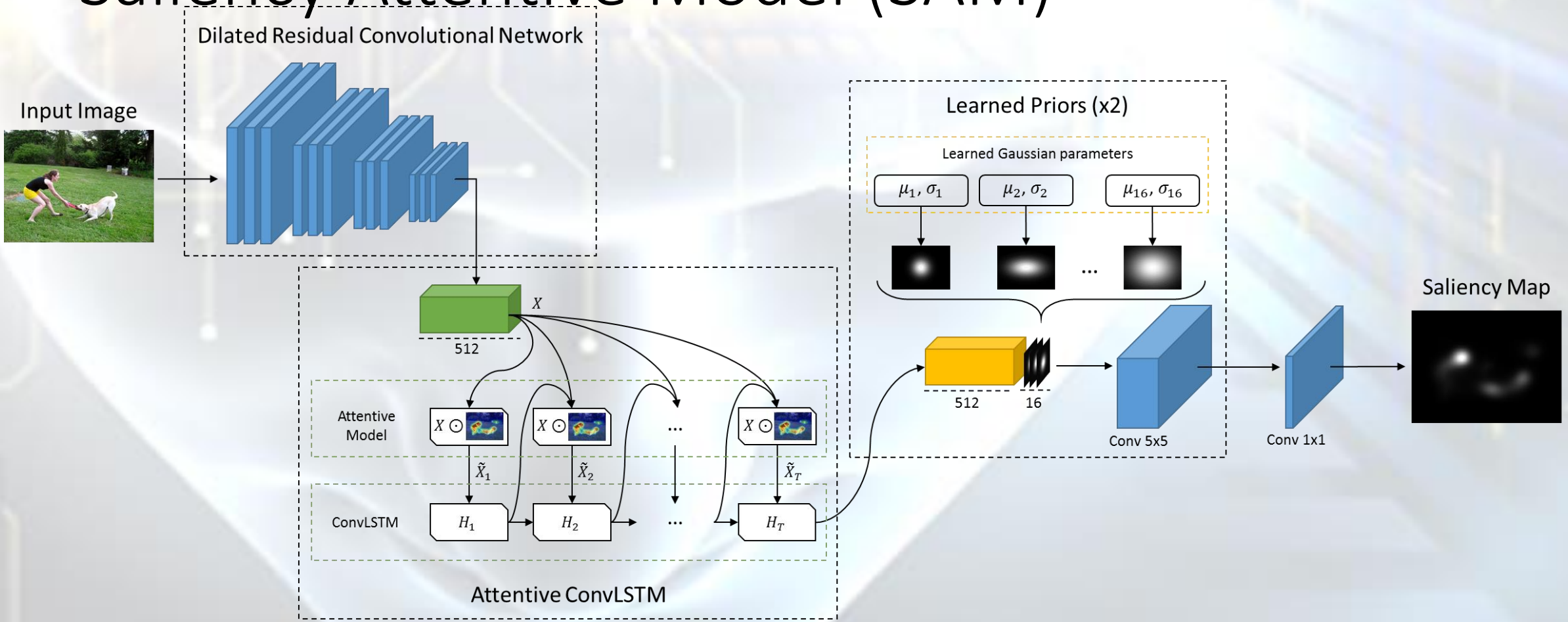
Multi-Level Network (ML-Net) –



Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, Rita Cucchiara. "A Deep Multi-Level Network for Saliency Prediction." In Proceedings of the 23rd International Conference on Pattern Recognition, 2016.

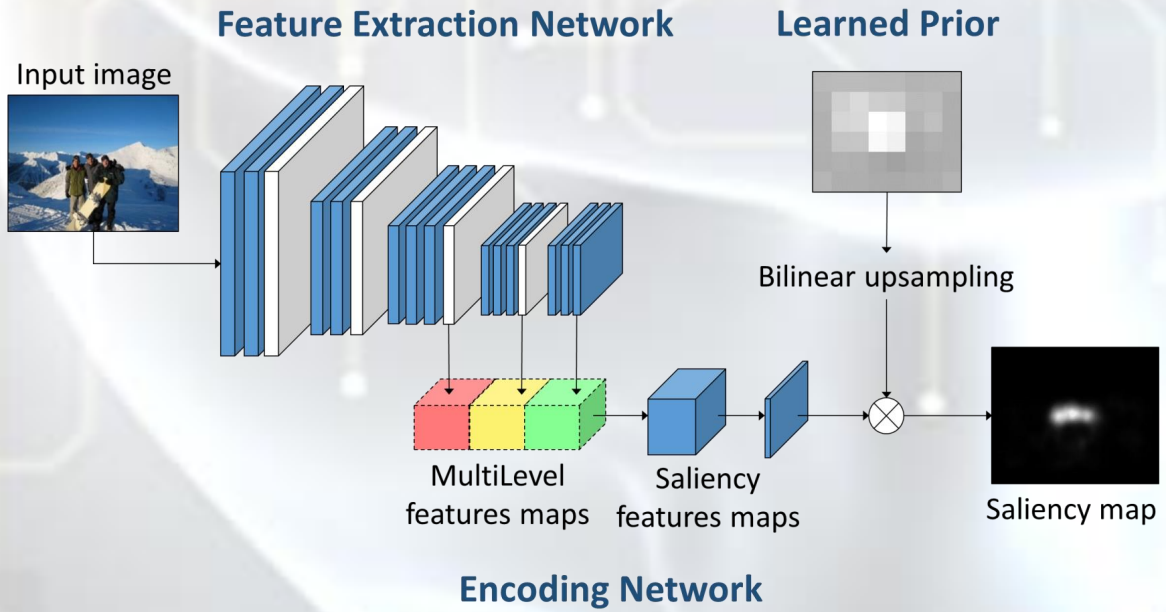
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

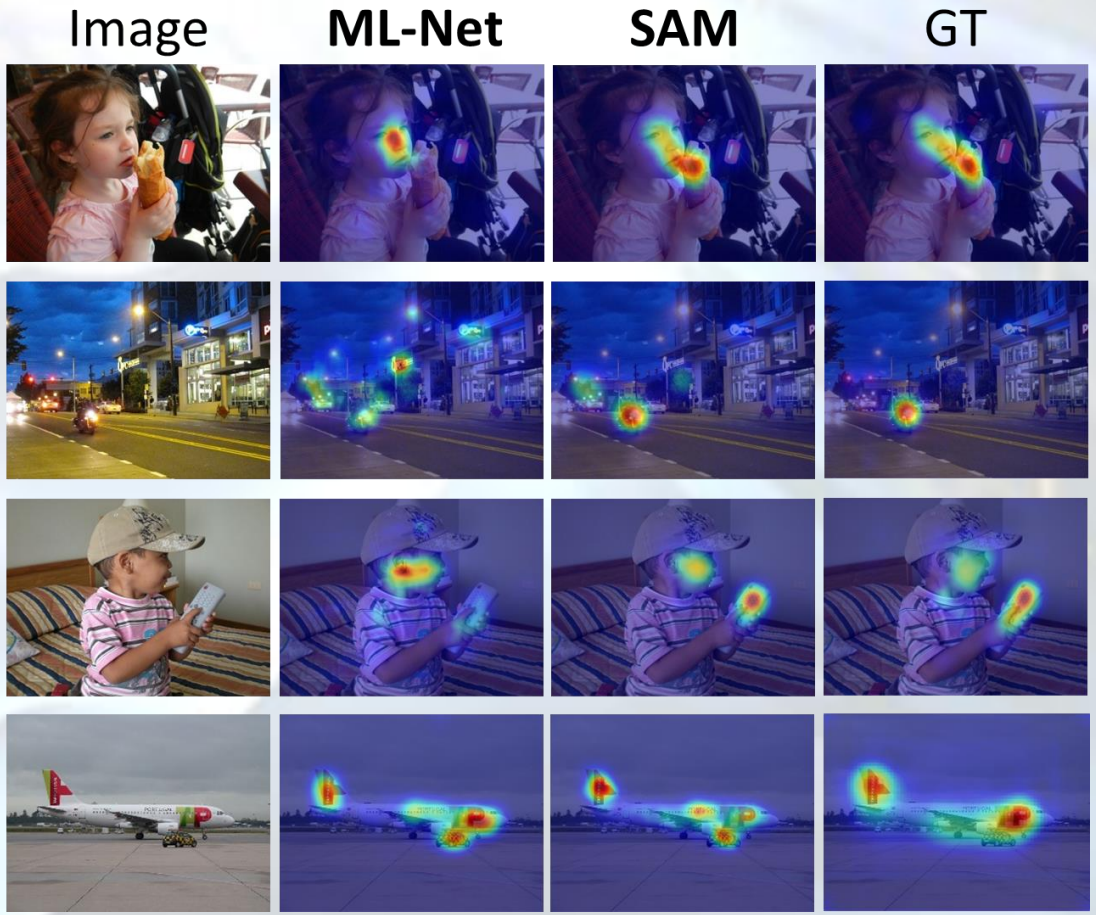
Saliency Attentive Model (SAM)



Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, Rita Cucchiara. "Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model." arXiv preprint arXiv:1611.09571, 2016.

Multi-Level Network (ML-Net)

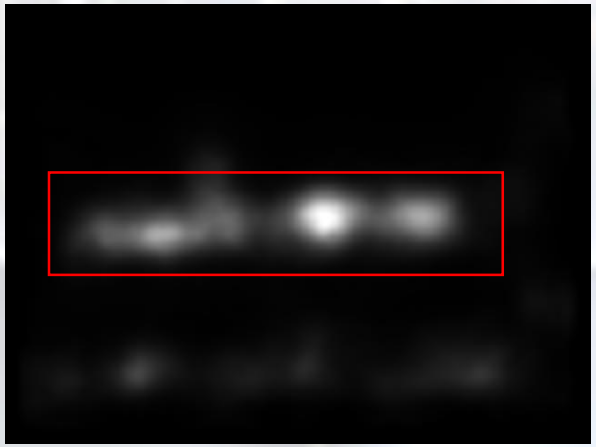


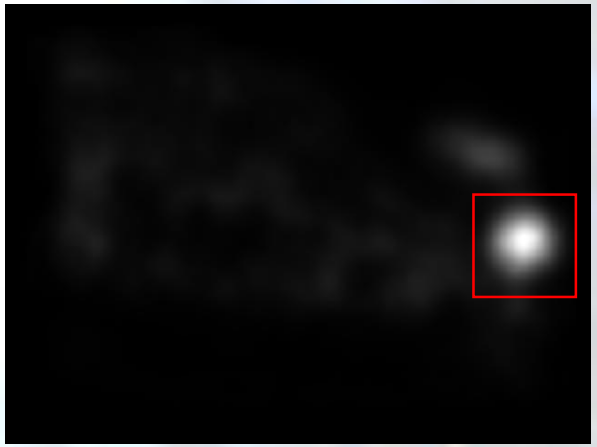


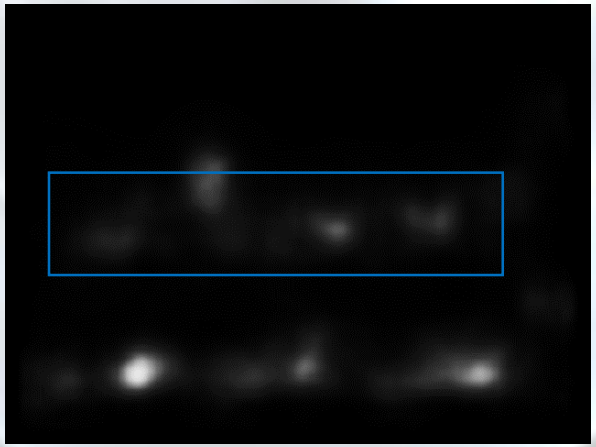
Experiment

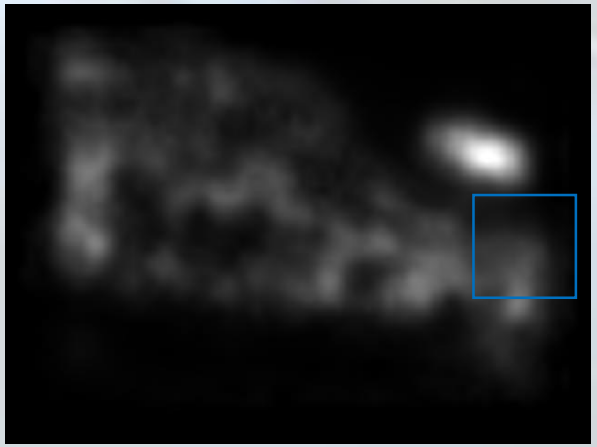
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Image | Predicted map | Image | Predicted map







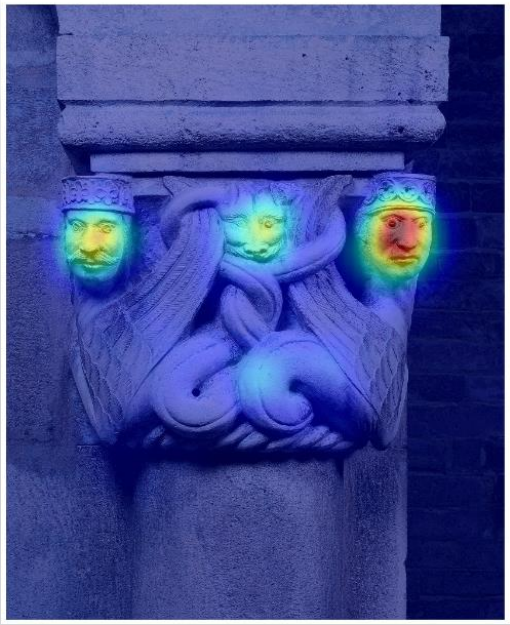


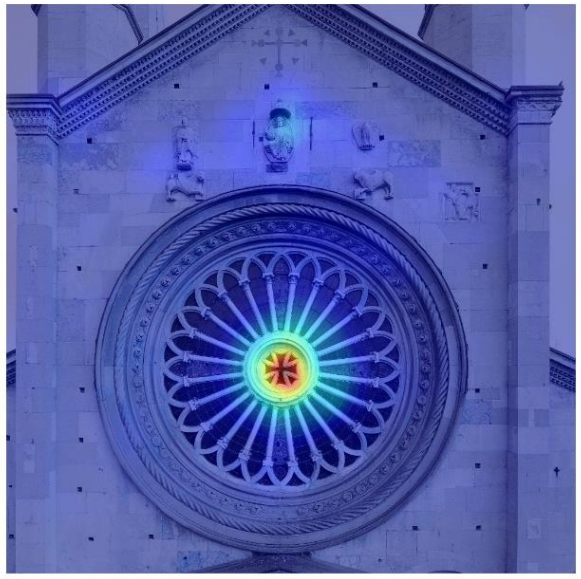
Example: Cultural Heritage Saliency

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA







Gaze and Attention by Adding task: Reproducing driver gaze

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

DRIVER DISTRACTION is an important safety problem. Between 13% and 50% of crashes are attributed to driver distraction, resulting in as many as 5000 fatalities and $40 billion in damages each year [1]–[3]. Increasing use of

and growing problem with global dimensions. A recent study by World Health Organization mentions that annually, over 1.2 million fatalities and over 20 million serious injuries occur worldwide [1]. Enhancement of traffic safety is pursued

Human error is the main cause of more than 90 percent

prompt safe decisions about driving maneuvers. Every year, traffic accidents result in approximately 1.2 million fatalities worldwide; without novel prevention measures, this number could increase by 65% over the next two decades [2]. In the U.S. alone, more than 43 000 fatalities are projected this year due to traffic accidents, with up to 80% of them due to driver inattention [3], [4]. To counter the effect of inattention,

is due to drivers with a diminished vigilance level. In the trucking industry, 57% of fatal truck accidents are due to driver fatigue. It is the number one cause of heavy truck crashes. Seventy percent of American drivers report driving fatigued.

- Existing dataset?

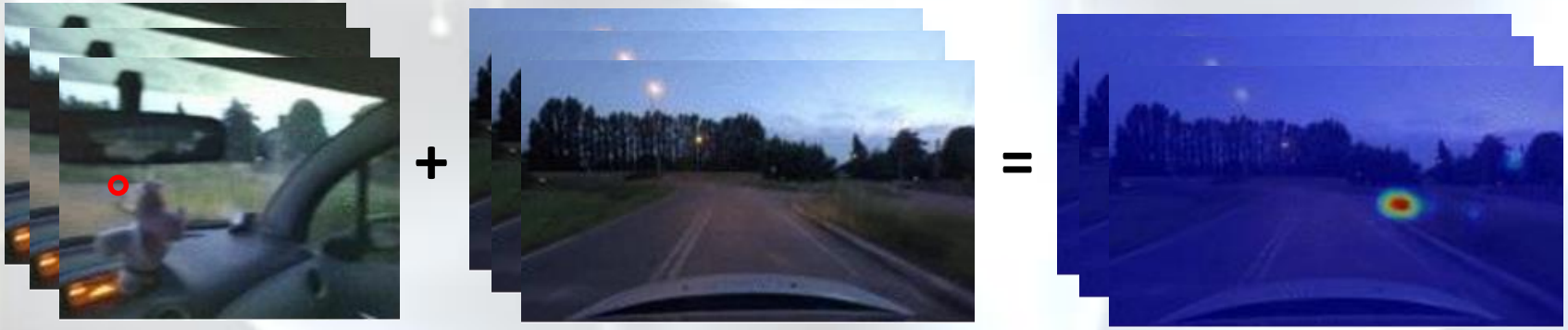
Dataset	Frames	Drivers	Scenarios	Annotations	Real-world	Public
Pugeault *et al.* [17]	158,668	n.d.	Countryside, Highway Downtown	9 classes in Environment Road, Junction, Attributes	Yes	No
Simon *et al.* [19]	40	30	Downtown	Gaze Maps	No	No
Underwood *et al.* [23]	120	77	Urban Motorway	n.d.	No	No
Fridman *et al.* [6]	1,860,761	50	Highway	6 Gaze Location Classes	Yes	No
Dr(eye)ve	555,000	8	Countryside, Highway Downtown	Gaze Maps	Yes	Yes

DR(eye)VE dataset: overview

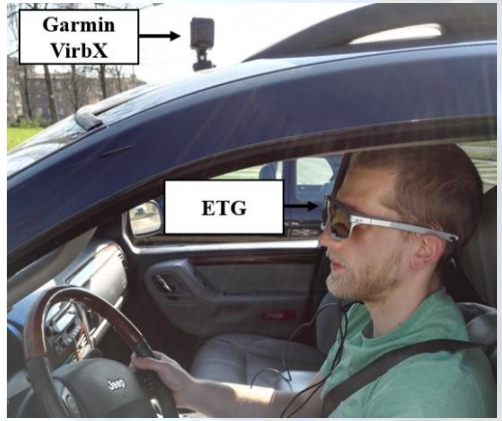
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

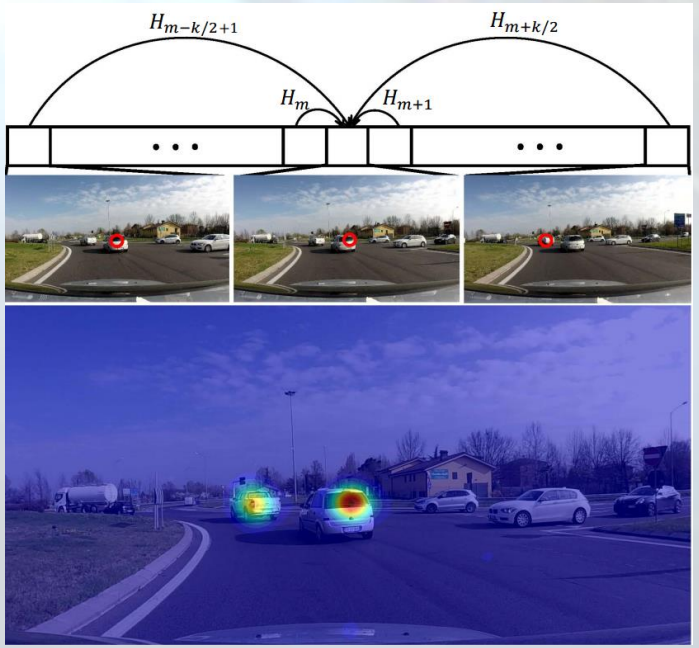
Acquisition rig:

- Car mounted camera: Garmin VirbX 1080p/25fps, embedded GPS
- Eye tracker POV: SMI ETG HD camera 720p/30fps



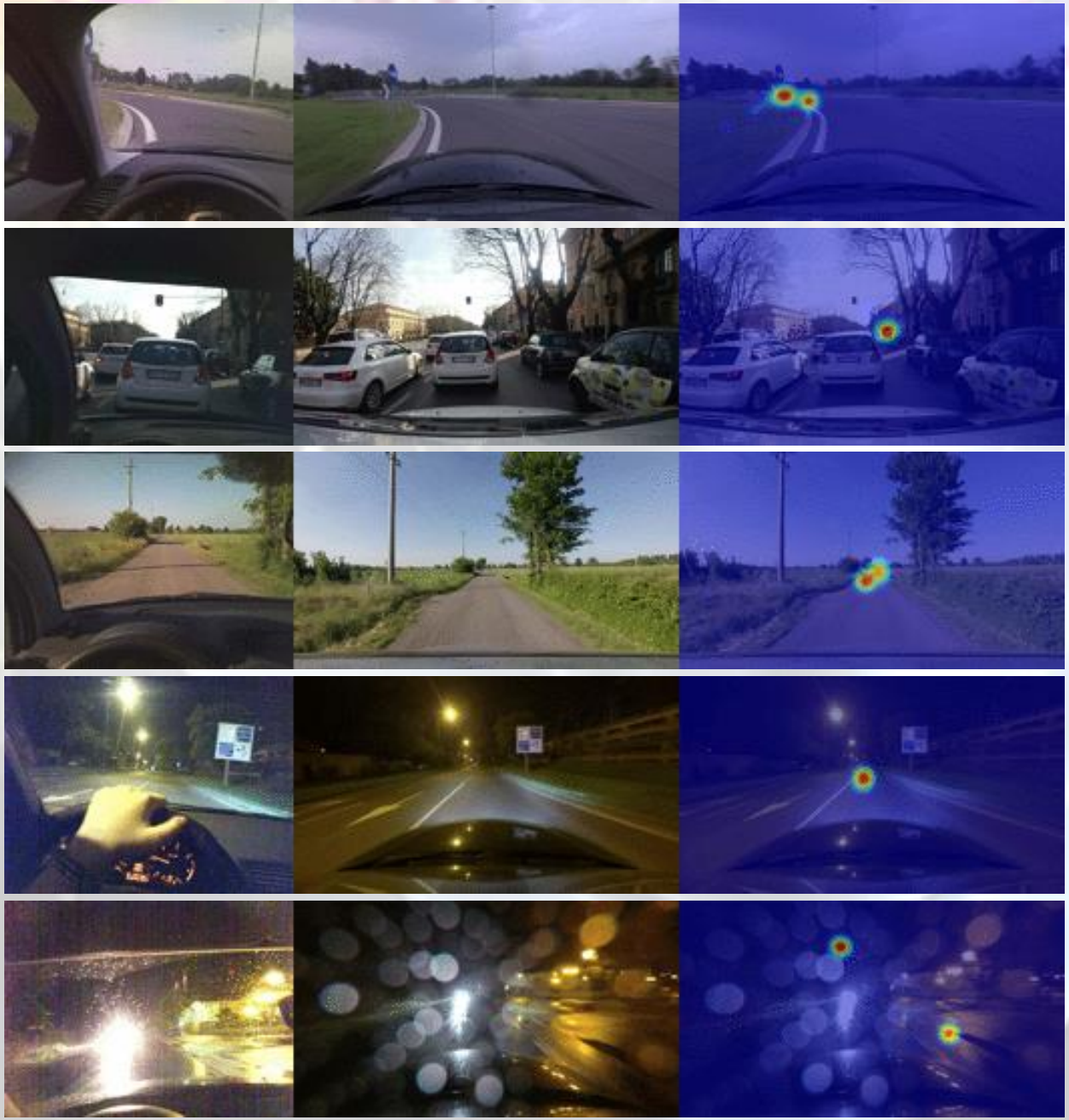
Gaze position projected on the video of the roof-mounted camera





Ground Truth:

- Attentional map integrated over 25 frames (1 sec)
- Speed/GPS and driving course information



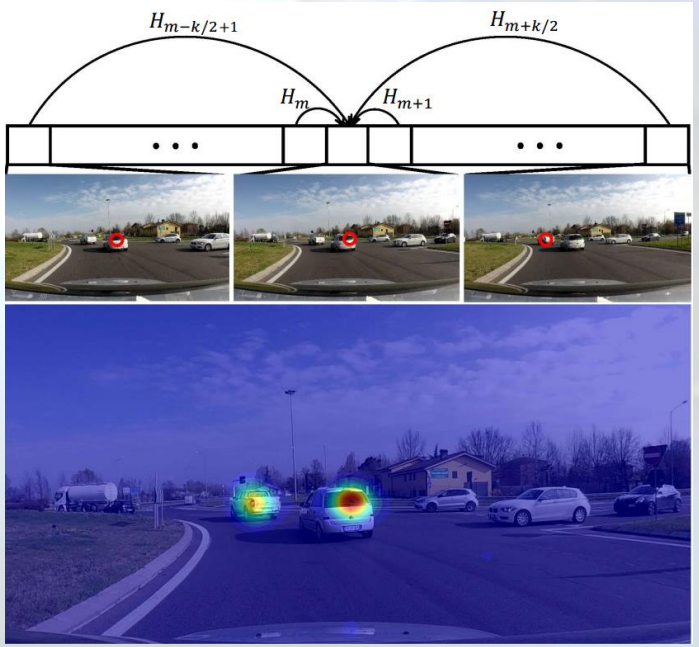
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- 8 different drivers
- 3 different landscapes
  {Highway, Countryside, Downtown}
- 3 different weather's conditions:
  {Sunny, Cloudy, Rainy}
- 3 different light's conditions:
  {Morning, Evening, Night}

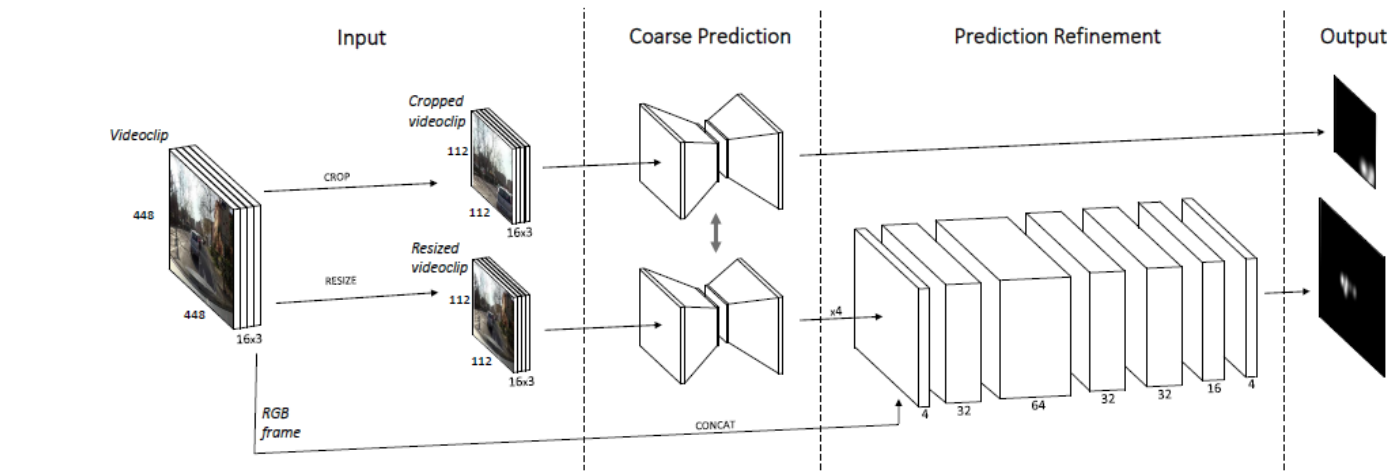
74 videos of 5 minutes each!

Data Acquisition and Fusion

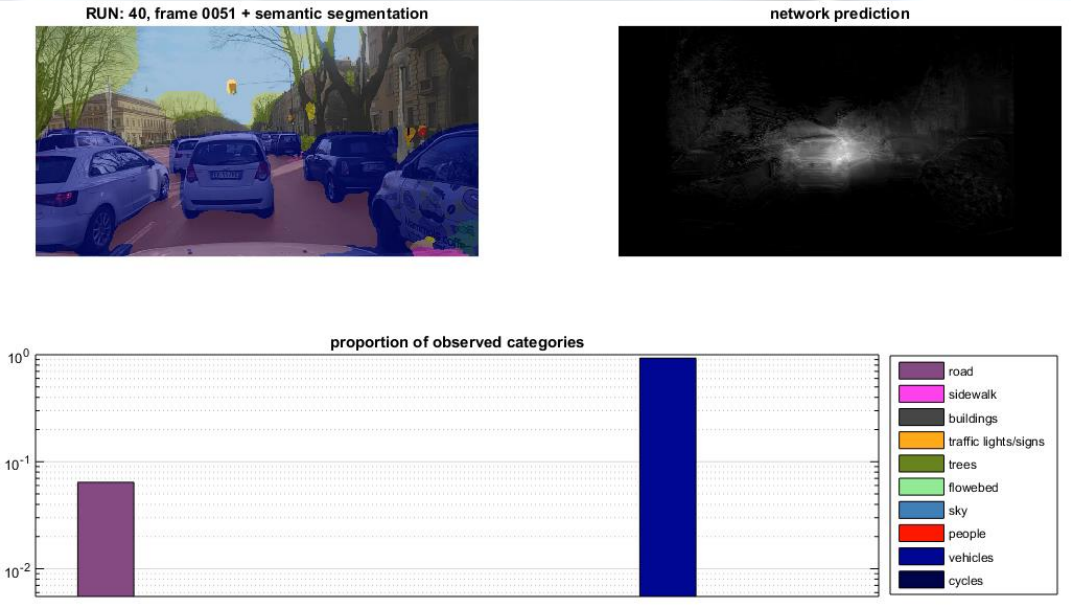
Warping Views through Image Homography



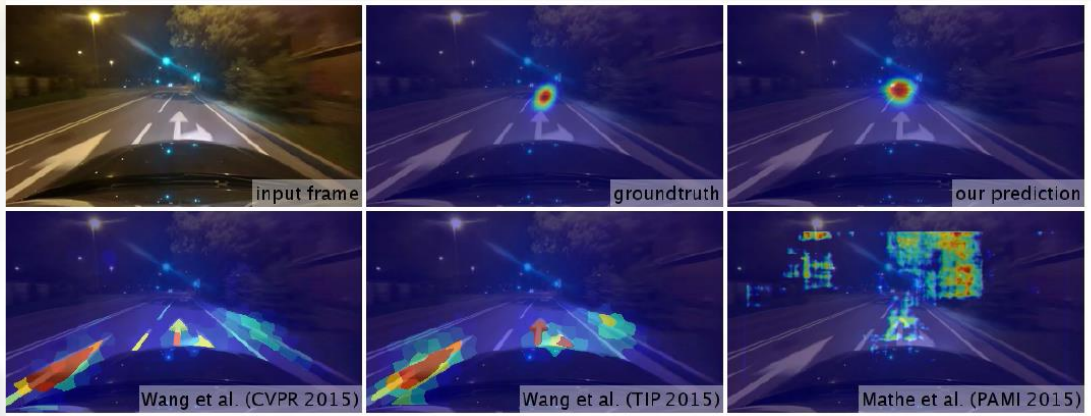
Prediction and Semantics Deep Model



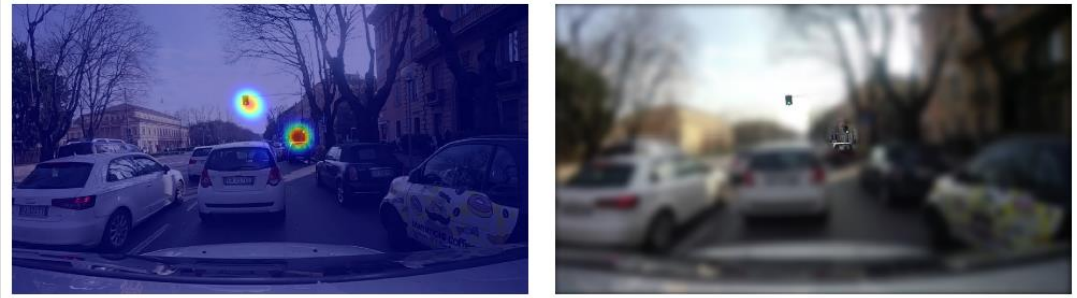
- **Coarse Prediction**: using crops to escape central bias
- **Fine Prediction**: Refinement at full scale



Is the Network Capturing the task?



Blurred Perception Field



- The task shifts **attention towards different objects**
- Behavior is influenced by the **cognitive effort** of the human in performing the task -> Focusing Gaze is **not effort free**

Example Chasing Car in GTA V

Attitude: Spotting prejudice with camera sensors

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

A new synergy:

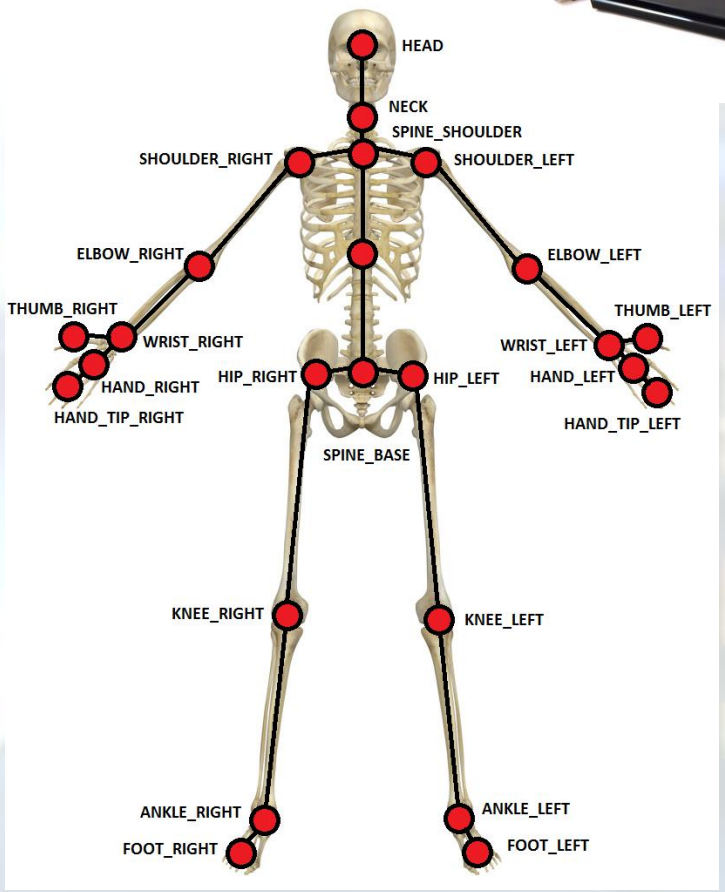
- Experimental studies in *Social Psychologies*
- Technologies and models of *Computer Vision*
- Integration of multi-sensorial data through *Machine Learning*

More disciplines involved:

Department of Science and Education
Prof. L. Vezzali,
Dipartimento di Ingegneria «Enzo Ferrari»
***Ing. S.Calderara**, Ing. N.Bicocchi, Ing. A.Palazzi, Prof. R. Cucchiara*

- **Research questions:**
  - Can social interaction be detected and recognized through Computer Vision?
  - Can we quantitatively measure the presence of prejudice in an interaction though multi-sensorial data?
  - Could these techniques lead to a better society?

Experimental Setup



Pool of participants

- 60 students of University of Modena and Reggio Emilia

Questionnaires:

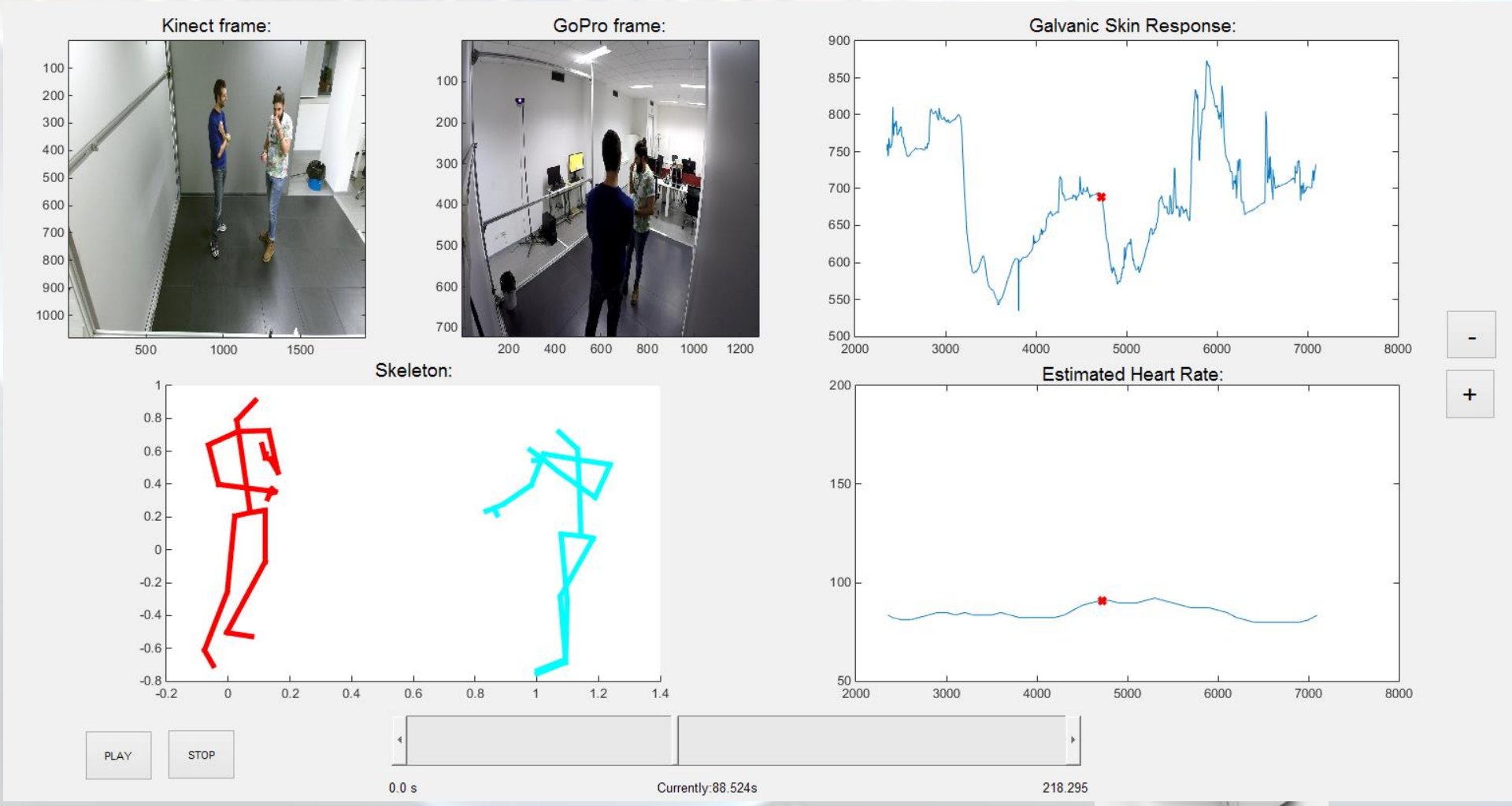
- Both *Explicit* and *Implicit* prejudice towards black people are tested through appropriate tests and questionnaires

Interactions:

- Participants interact both with white and black people. Interactions are recorded by multiple sensors.

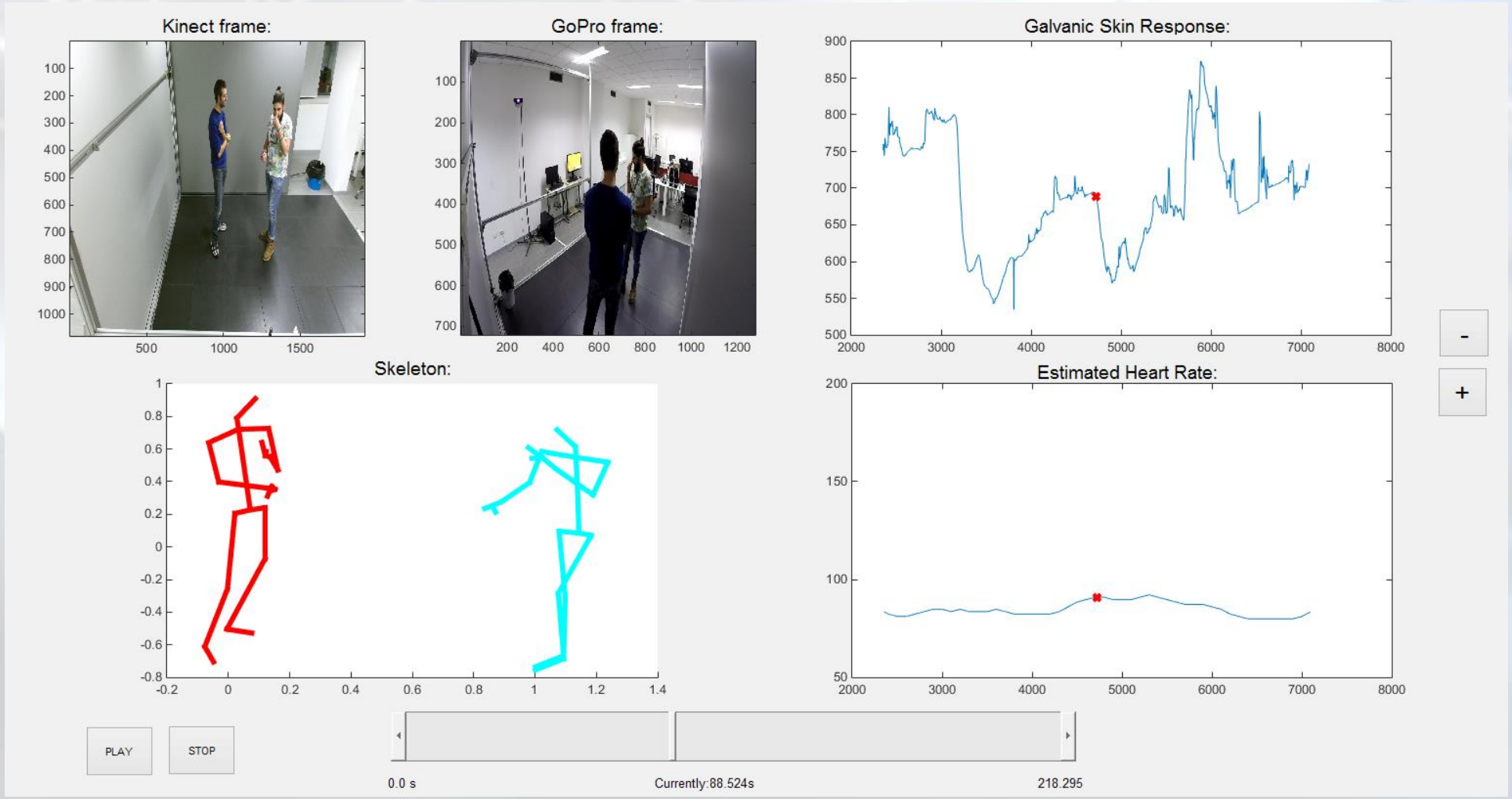
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Data Acquisition



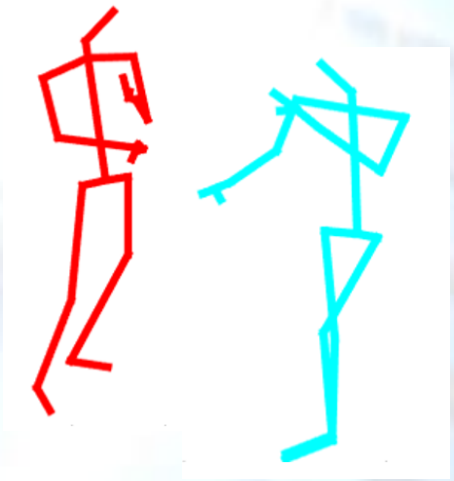
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

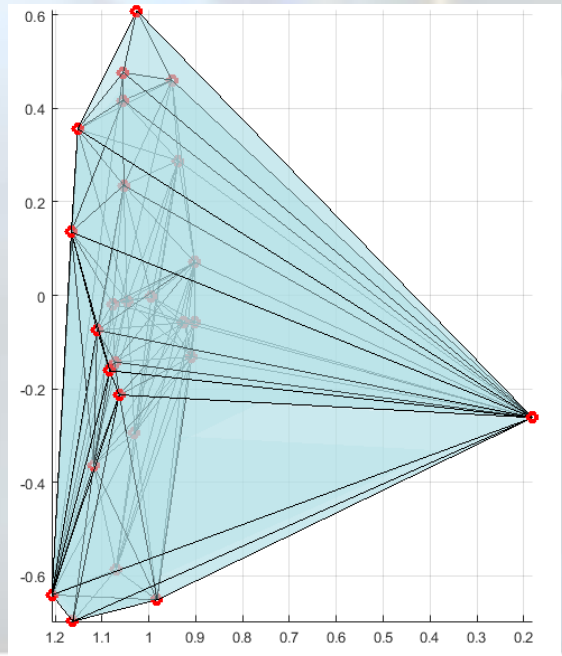
A View of the Experimental Setting



Features Extracted

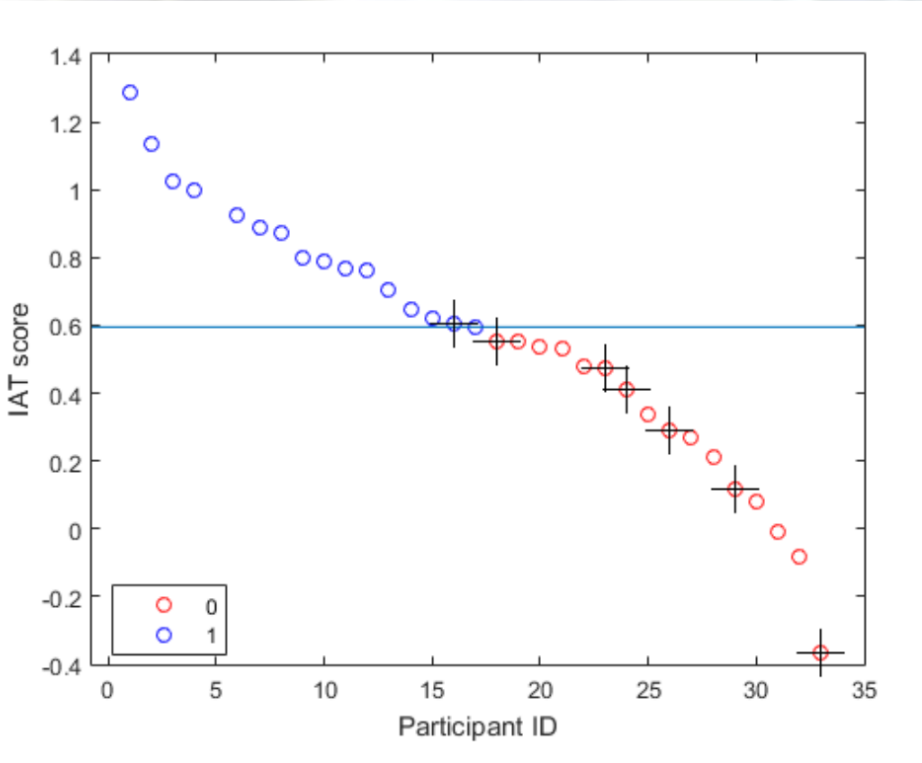
- *Motion, Speed, Acceleration* for each joint
- *Mutual distance* between interlocutors
- *Volume* between interlocutors
- *Pauses* in the dialogue
- *Biometric features*



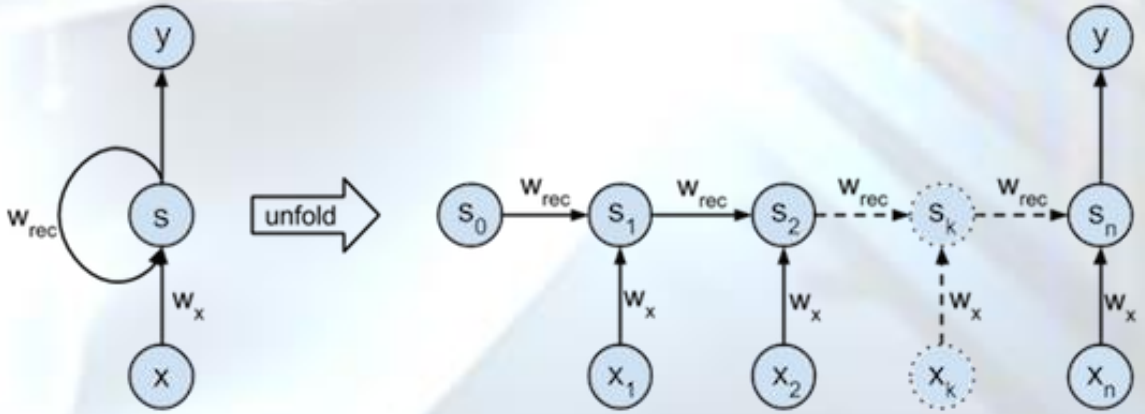


Results

Using a Vanilla **Recurrent NN Classifier** on extracted features



IAT trigger Pos vs Neg

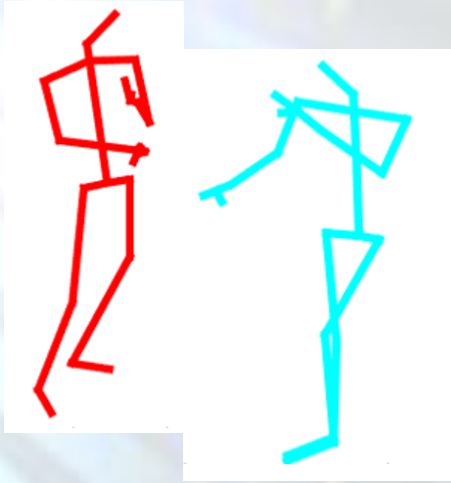


Extracted Features sequence

Scheme of classification results

Light blue line is the median IAT score, while black crosses represent an error for the classifier.

Publications and Press



- **Publications**

  The following work has been accepted (oral) to the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp) 2016 (Heidelberg)

  Palazzi, A., Calderara, S., Bicocchi, N., Vezzali, L., di Bernardo, G. A., Zambonelli, F., & Cucchiara, R. (2016, September). Spotting prejudice with nonverbal behaviours. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 853-862). ACM.

- **Press Release**

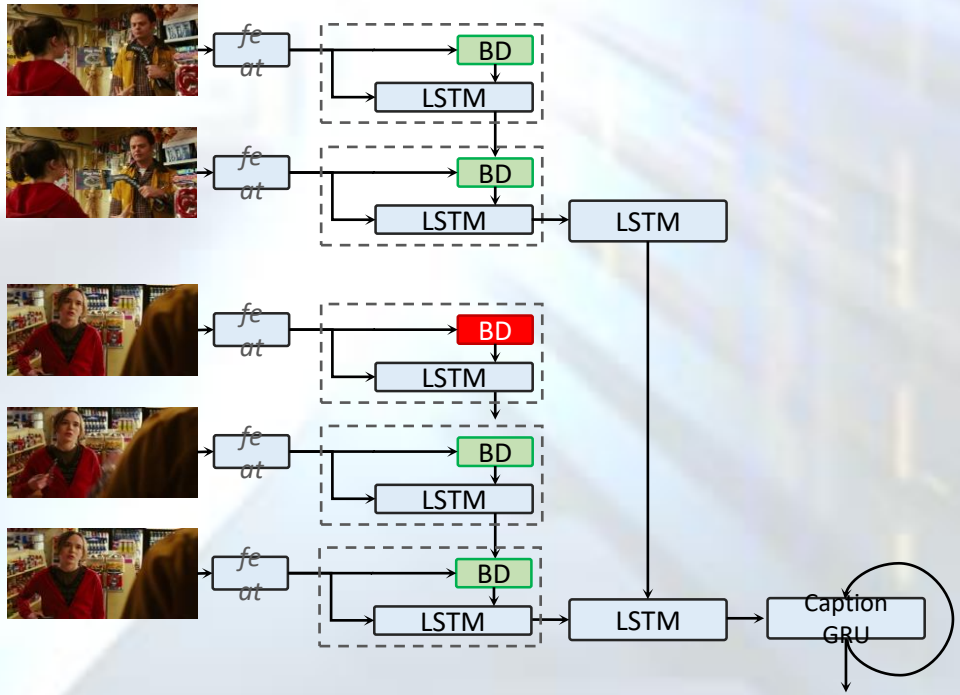
  - The international science magazine New Scientist, based in UK, has talked about this work in one of its cover articles.

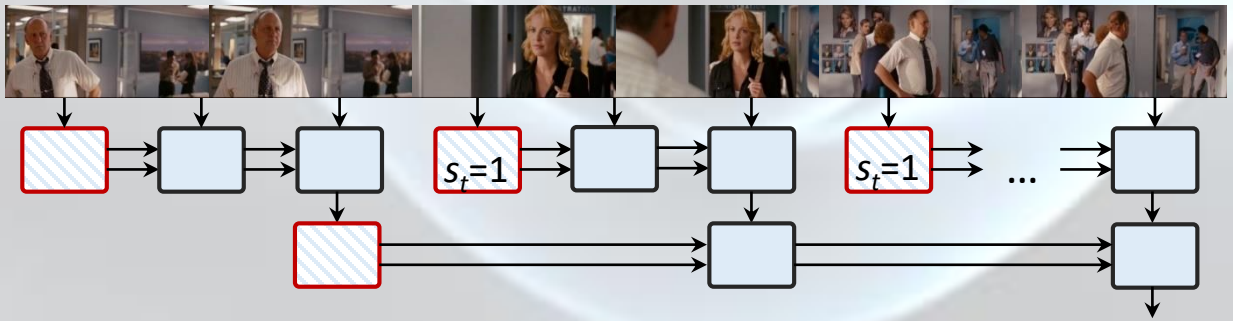
    https://www.newscientist.com/article/mg23130933-200-camera-spots-your-hidden-prejudices-from-your-body-language/

Scene understanding: from Video to Caption

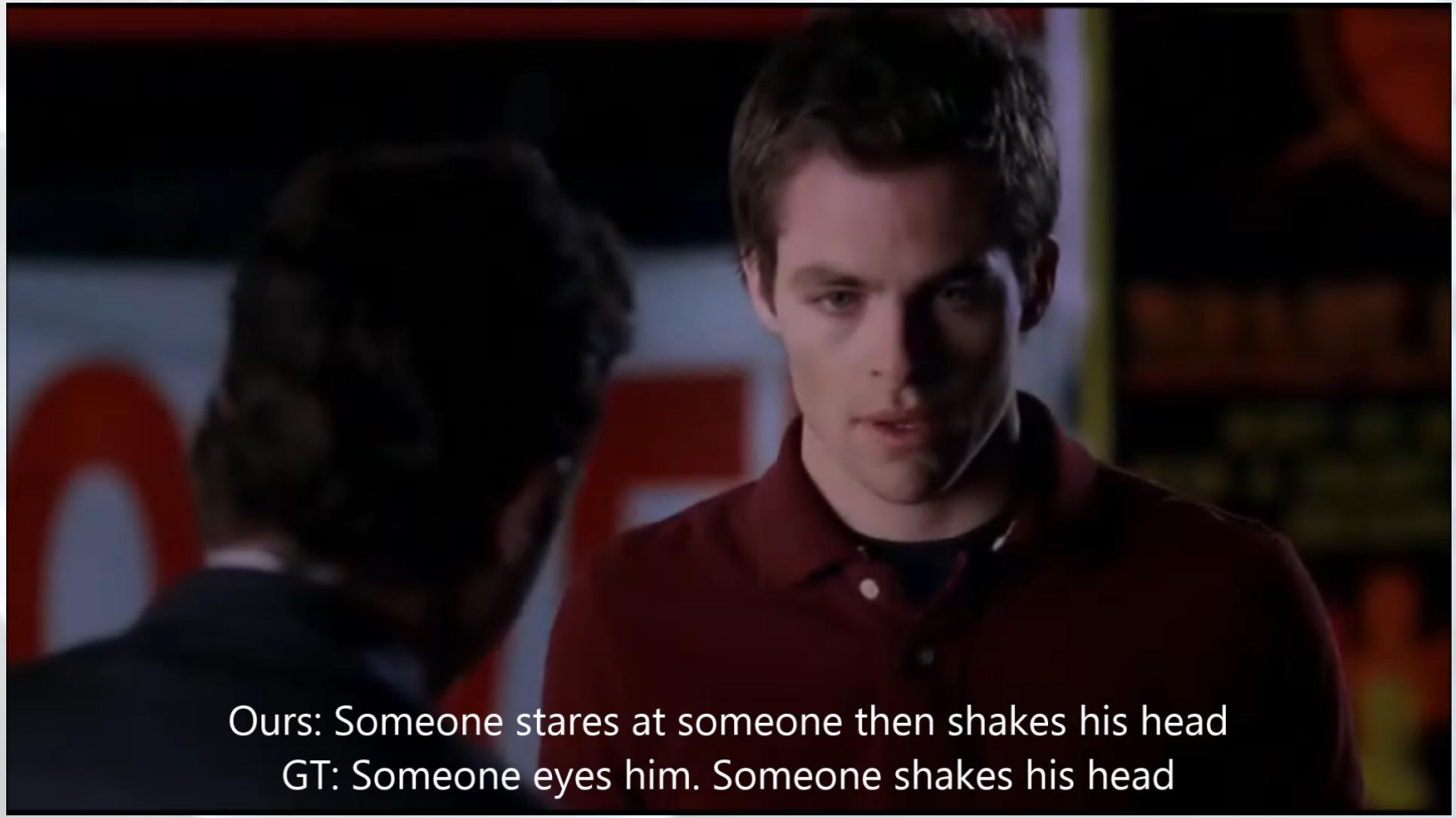
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- We propose a video encoding network which can adaptively modify its structure to improve video captioning.
- The result is a variable length and adaptive encoding of the video, whose length and granularity depends on the input video itself.





Scene understanding:
from Video to Caption Results

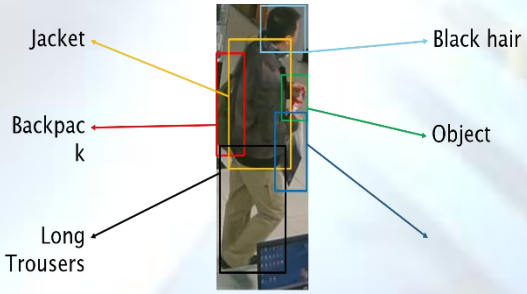


Security and Surveillance: Attribute Classification

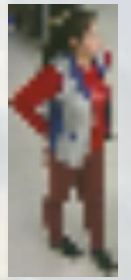
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Attribute classification from surveillace images

- Detect and extract attribute from small images
- Important security task subject to problems of:
  - Occlusions
  - Resolution



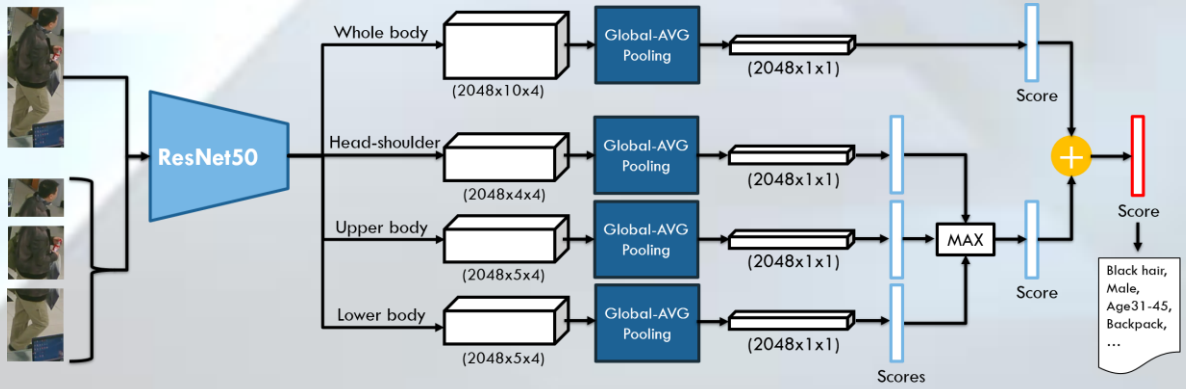
Occlusion



Low Resolution

Using Generative Models for guess the true appearance

1. Train a generative model on the concept of human in surveillance
2. Capture the most probable subject appearance
3. Try to 'guess' images missing part and resolution enhancement
4. Classify attribute on guessing

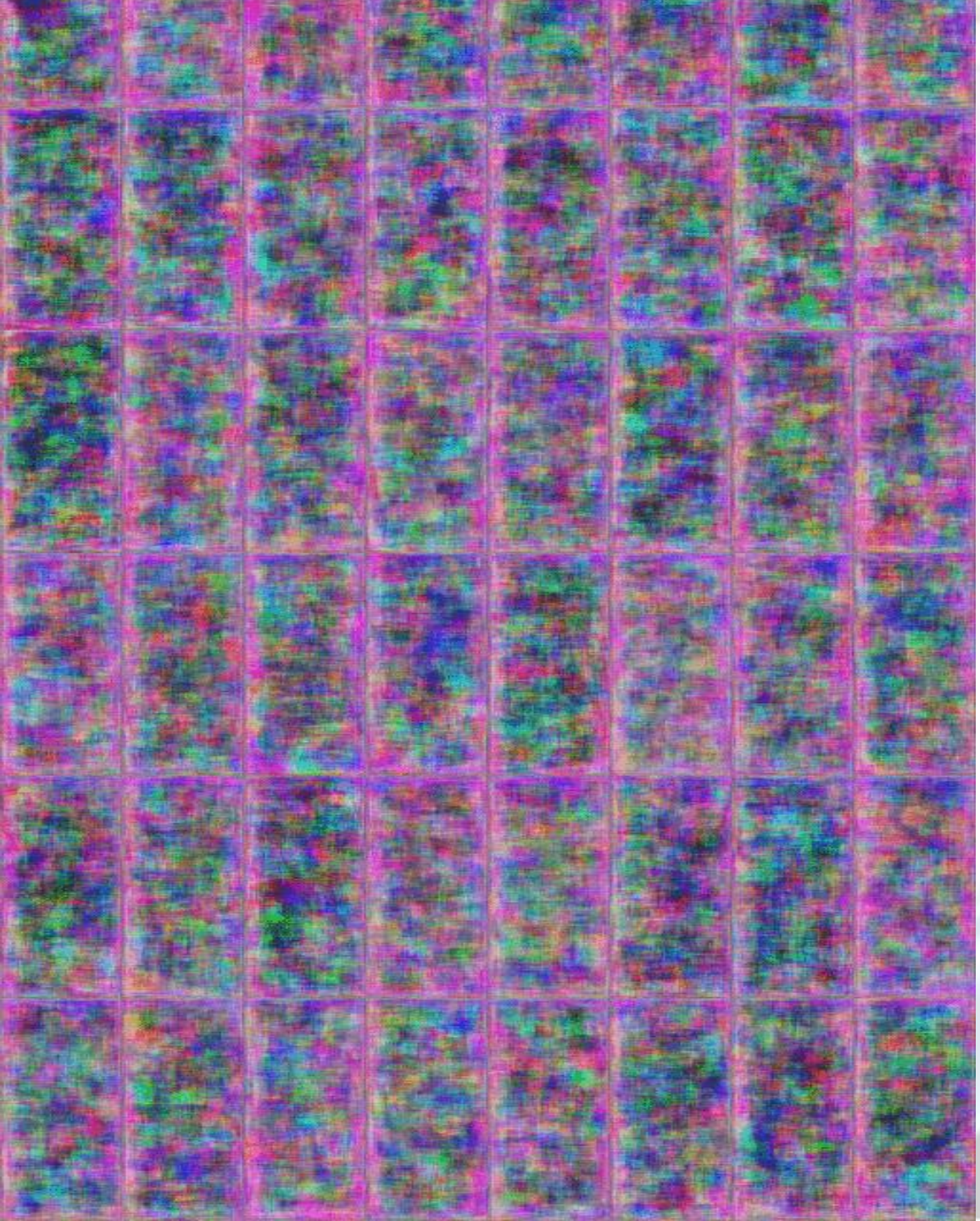


UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Securit esolution

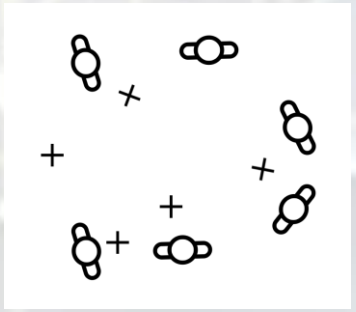
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Security ing Part

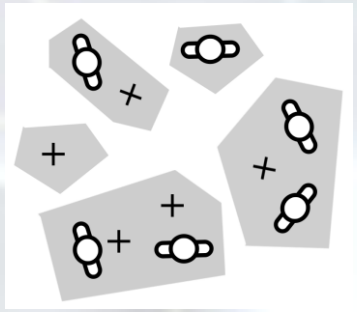


Security and Surveillance: Track People

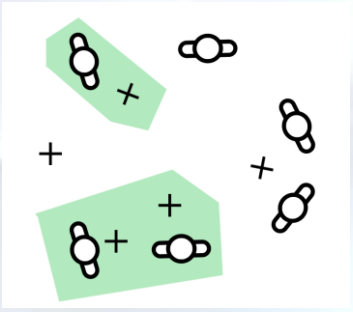
TARGETS and DETECTIONS (+)



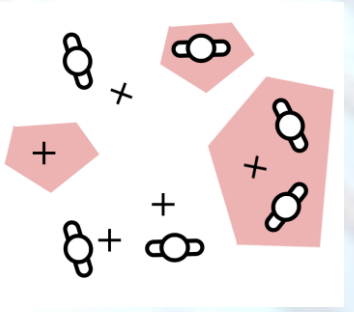
CLUSTERING STEP



ASSOCIATE ON DISTANCE FEATURES ONLY



ASSOCIATE ON DISTANCE AND MORE COMPLEX FEATURES



Why is it important?

- To complete higher level analysis, identities are often mandatory.

Our approach

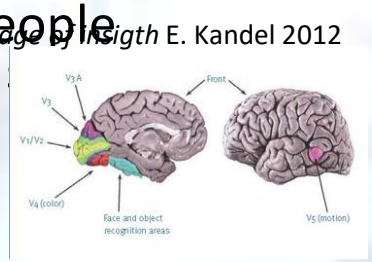
- Evolution has taught us to prefer spatial information over surface features (patterns, colors, …) or motion
- Position is always meaningful, while other features benefit changes from scene to scene
- At each frame, split detections and tracks in locally compact clusters. The key idea is that some of these will be really easy to solve and position will be enough.
- More complex features can still be used in more difficult cases.

Security and Surveillance: Track People

Cognitive Visual Tracking

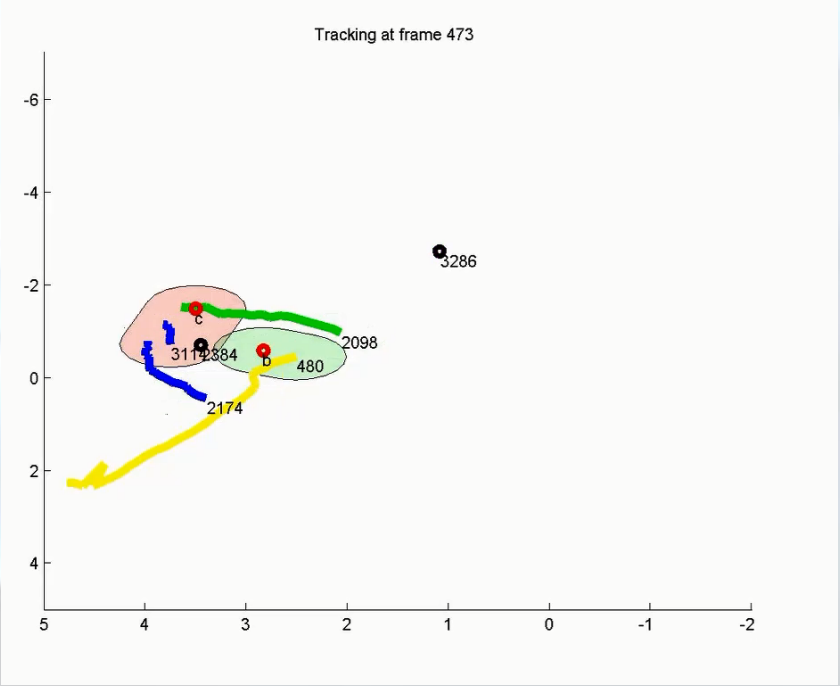
- From **neuroscience** : there are detection areas (expecially for people and faces) and areas for localizing target independently by their
- From **perceptual psychology** : the object file theory
- (Kahnemann, Treisman, Gibbs 1995)
- We use **distance** only when is possible
- Motion prediction and appearance is a plus when usef

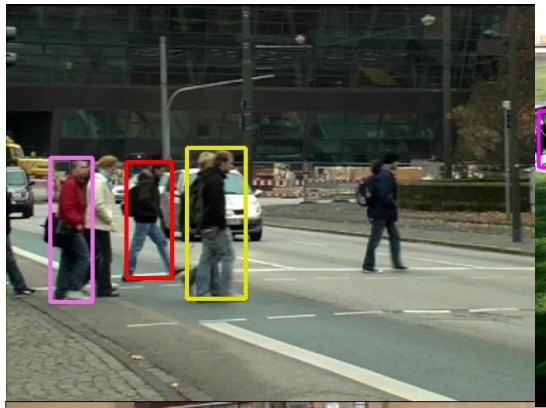
The Age of Insigth E. Kandel 2012

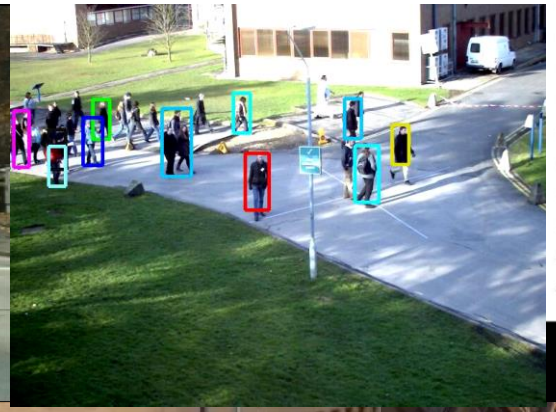


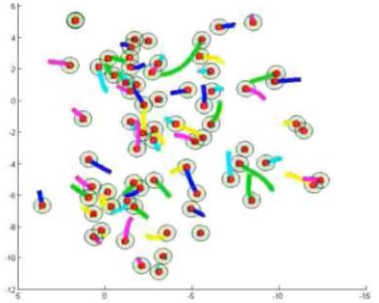
Thus? Cognitive tracking with latent structual svms

1. Split the crowd in **influence zones** (latent knowledge)
2. Decide whether those **zones are ambiguous** (also latent)
3. Solve unambiguous associations with distance only
4. Employ different level features in ambiguous cases (ask for shapes, color.. edges.. motion)



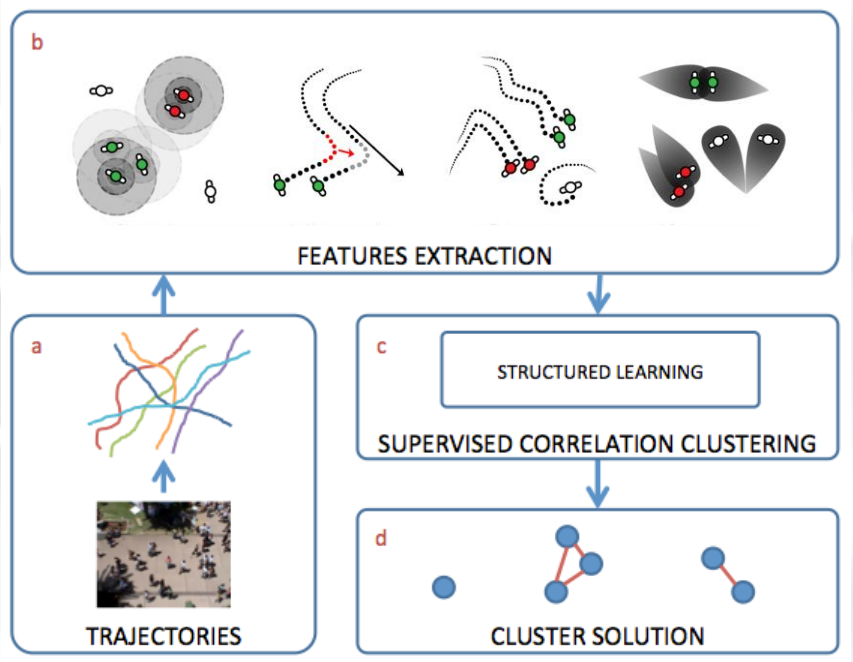






0% detection error rate

Security and Surveillance :Group People



Why is it important?

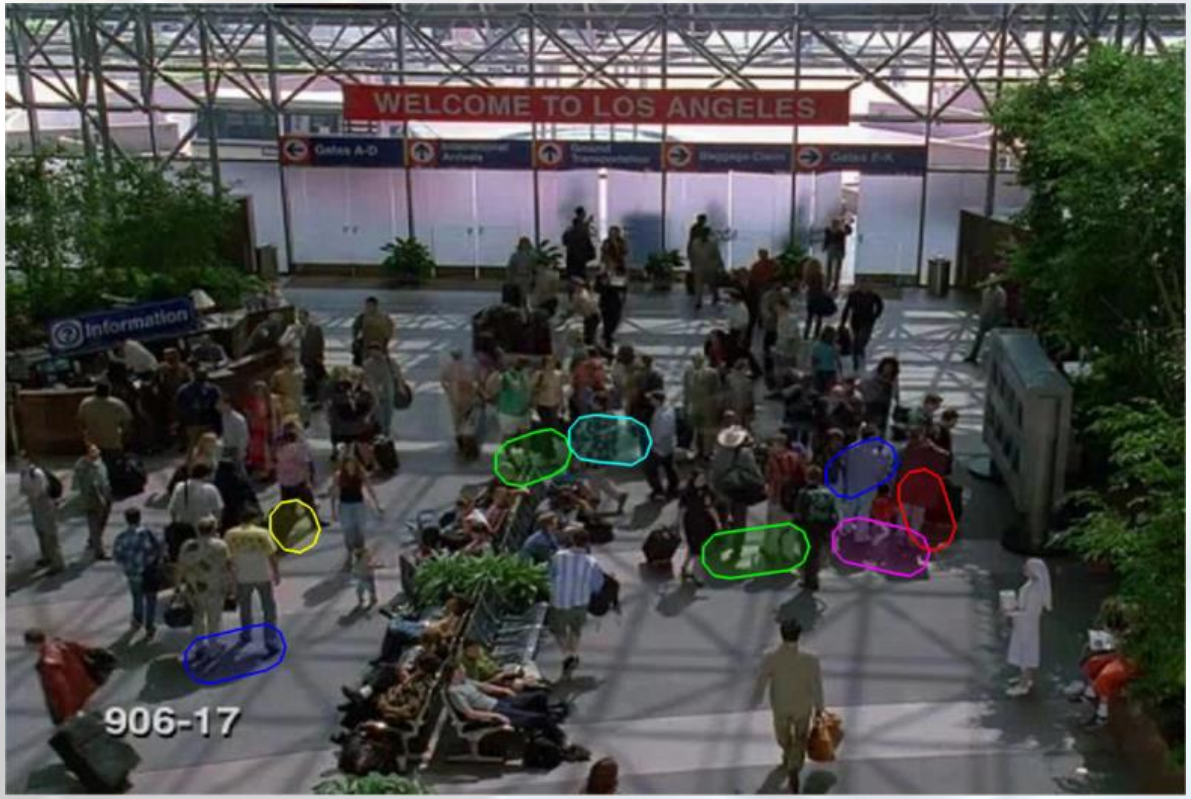
- Benefit simulation through more realistic crowd structure
- Increase surveillance systems awareness of relationships between observed targets
- Discover strange behaviors, e.g. thieves getting close to an already established group

Our approach

- Define socially grounded features
- We cast it as a correlation clustering task where affinity measure is learned depending on the scenario
- We introduce a loss function specifically designed to obtain plausible social groups and a way to optimize it

Security and Surveillance :Group People

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Drive a Car @imagelab