FONDAZIONE BRUNO KESSLER

UNIVERSITÀ DI PISA

# *Event Detection from Video using Answer Set Programming*

Authors: Abdullah khan, Luciano Serafini, Loris Bozzato, Beatrice Lazzerini

FONDAZIONE BRUNO KESSLER

UNIVERSITÀ DI PISA

# Outline

## Objective

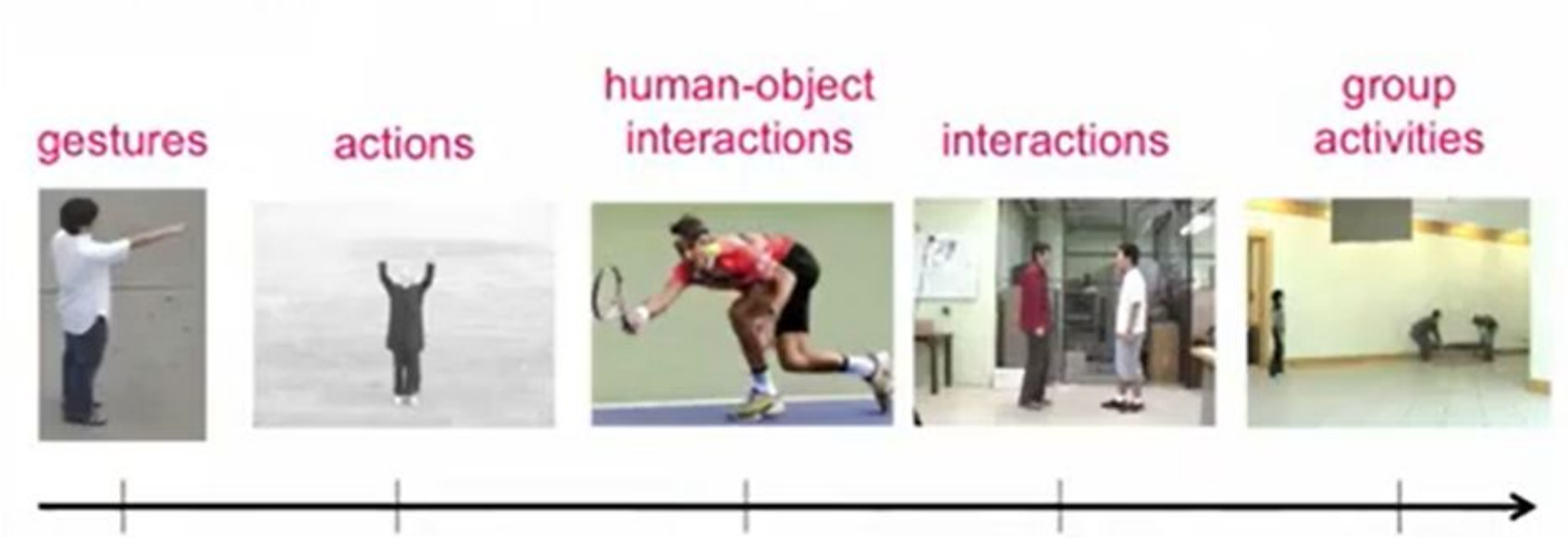
Recognition of complex events from a simple events in videos.

## Methodology

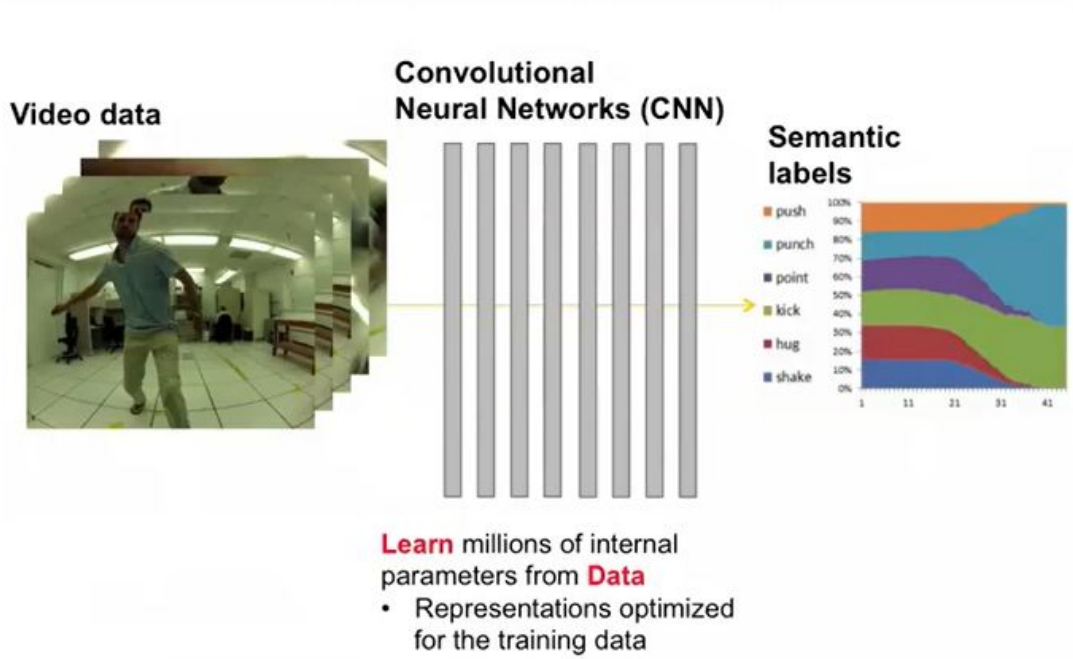
1. Object detection and tracking in videos
2. Logical Framework (Event Calculus) for event recognition
3. Answer set programming (reason about the logical rules).

# What is event recognition?

Given an input video/image, perform some appropriate processing, and output the “action label”.



# State of the art in video event detection



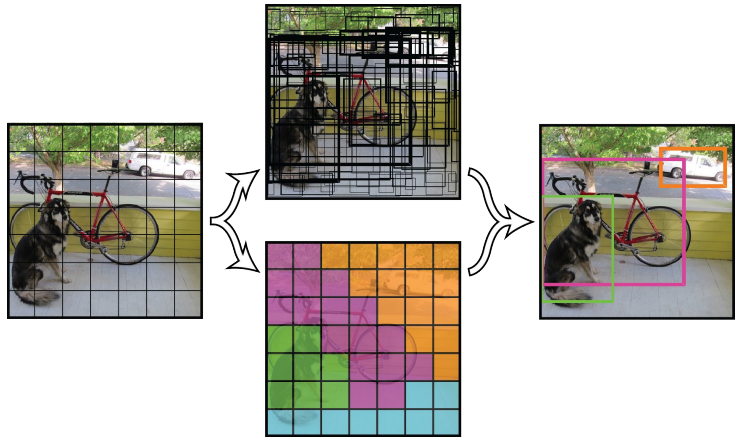
# YOLO Object detection and tracking?

Divide image into SxS grid

Within each grid cell predict:

Bboxes:4 coordinates + confidence

Direct prediction using a CNN



# Datasets: UCF / J-HMDB

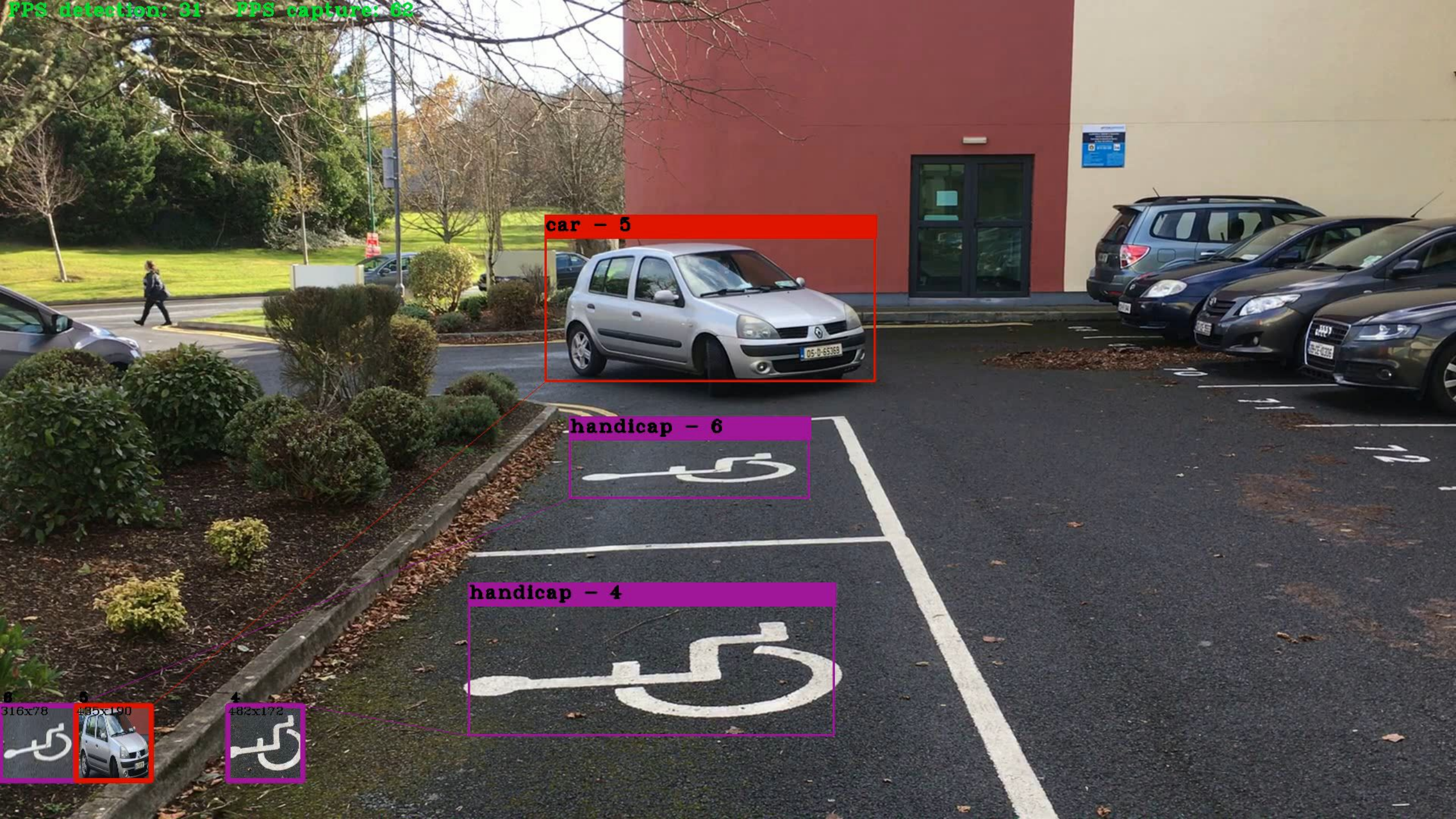
- UCF-Sports
  - 10 action categories
  - 150 videos
  - Trimmed
- J-HMDB-21 (subset)
  - 21 action categories
  - 928 videos
  - Trimmed
- UCF-101-24 (subset)
  - 24 action categories
  - Multiple instances, not whole video duration

FONDAZIONE BRUNO KESSLER

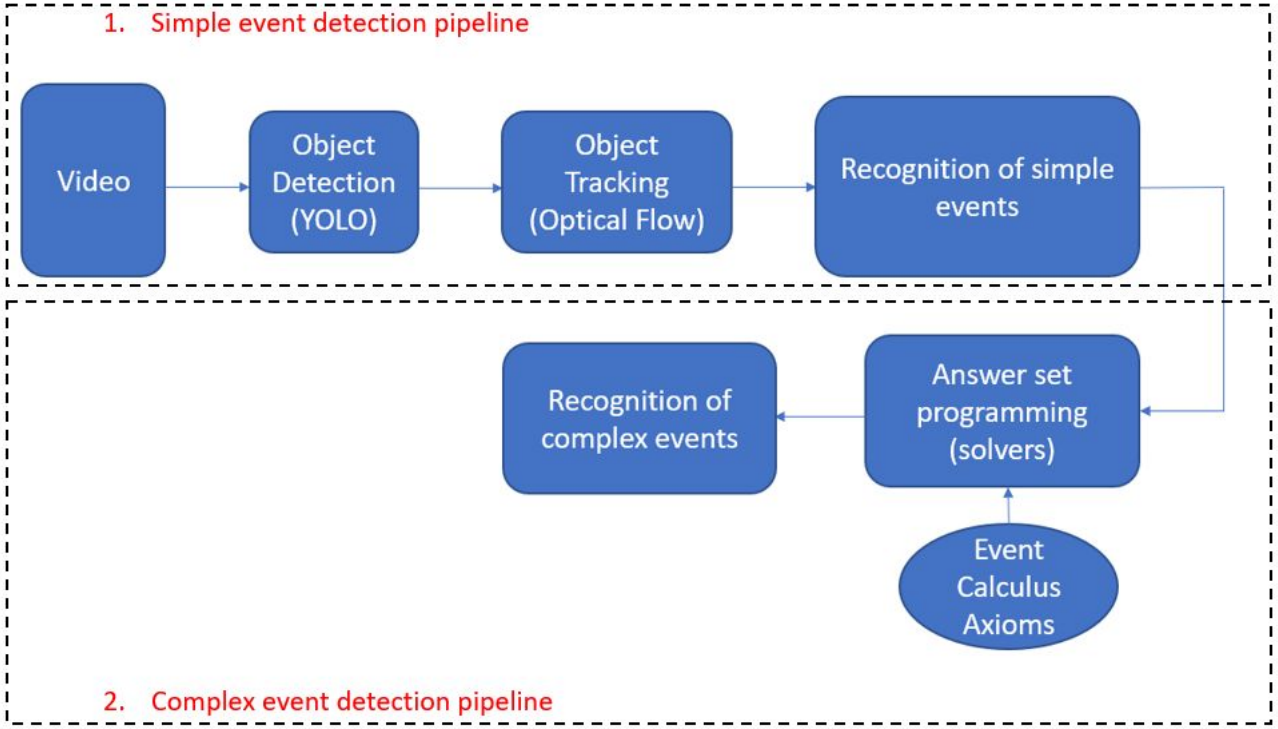
UNIVERSITÀ DI PISA

# Use-case (Handicap Parking Detection)

- 4 min long video, consisting of approximately 6.5k manually annotated frames.
- Objects are detected and tracked from every single frame using the state-of-the-art object detector (YOLO).

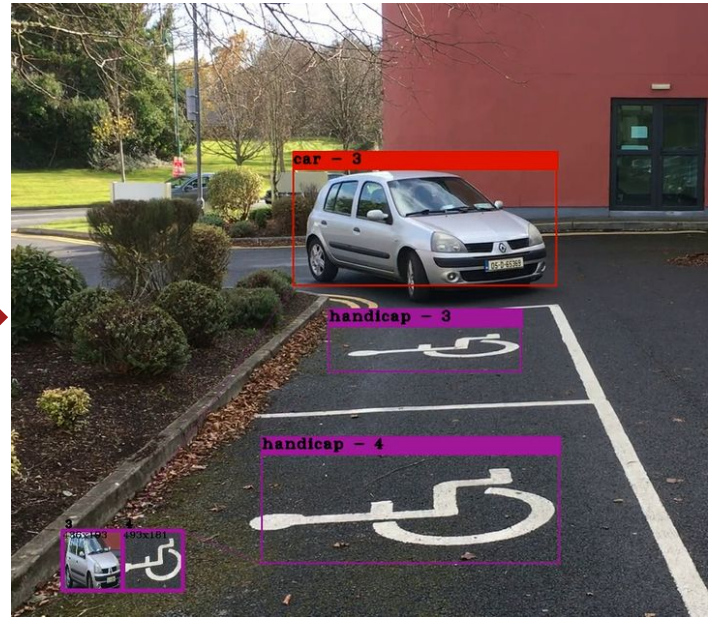

FPS detection: 31 FPS capture: 62
car – 5
handicap – 6
handicap – 4
6
316x78
5
485x190
4
482x172

# Proposed Architecture



# YOLO (You Only Look Once)

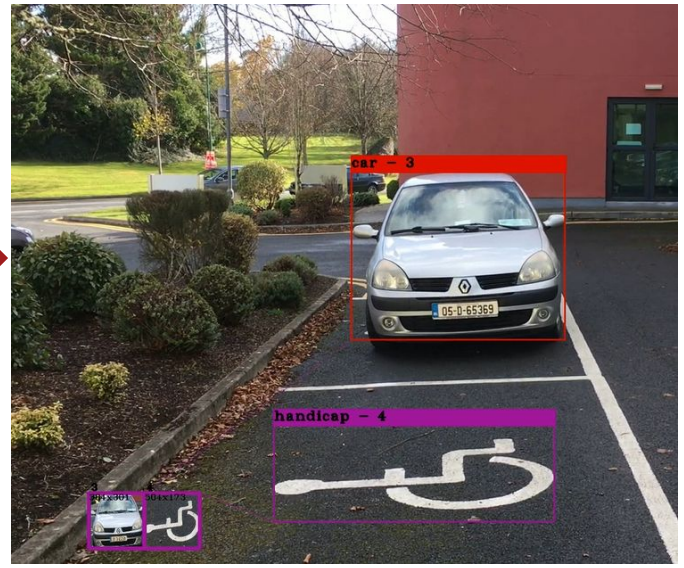
Input video → YOLO (Object Detection/Tracking) (YOLO) →




https://github.com/AlexeyAB/darknet

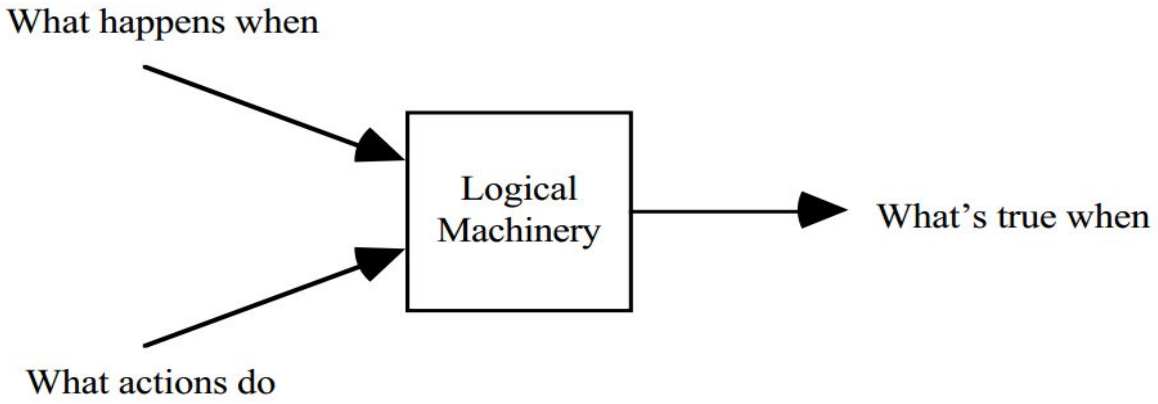
# YOLO (Continued)

Input video → YOLO (Object Detection/Tracking) →




https://github.com/AlexeyAB/darknet

# Logical reasoning on Complex events(Event Calculus)




- EC distinguishes three kind of objects. *Events*, *fluents*, *time-points*.
- ***Fluents*** are relations whose truth values varies with time.

| Basic Predicates | Description |
|---|---|
| $holdsAt(f,t)$ | fluent $f$ is true at time-point $t$ |
| $happens(e,t)$ | event $e$ occurs at time-point $t$ |
| $initiates(e,f,t)$ | if event $e$ occurs at time-point $t$, then fluent $f$ will be true after $t$. |
| $terminates(e,f,t)$ | if event $e$ occurs at time-point $t$, then fluent $f$ will be false after $t$ |

FONDAZIONE BRUNO KESSLER

UNIVERSITÀ DI PISA

# Simple and complex events

| **Simple event** | **Description** |
|---|---|
| $appearsCar(A, T)$ | The object corresponding to car $A$ enters the scene at time $T$ |
| $disappearsCar(A, T)$ | The object corresponding to car $A$ leaves the scene at time $T$ |
| $appearsSlot(L, T)$ | The object corresponding to parking slot $L$ appears in the scene at time $T$ |
| $disappearsSlot(L, T)$ | The object corresponding to parking slot $L$ disappears from the scene at time $T$ |
| **Complex event** | **Description** |
| $covers(A, L, T)$ | The object car $A$ covers the slot $L$ at time $T$ |
| $uncovers(A, L, T)$ | The object car $A$ uncovers the slot $L$ at time $T$ |

# Encoding of simple and complex events using EC

## Simple events using EC formalism

$$initiates(appearsCar(A), visibleCar(A), T) \leftarrow agent(A), time(T).$$
$$terminates(disappearsCar(A), visibleCar(A), T) \leftarrow agent(A), time(T).$$
$$initiates(appearsSlot(L), visibleSlot(L), T) \leftarrow location(L), time(T).$$
$$terminates(disappearsSlot(L), visibleSlot(L), T) \leftarrow location(L), time(T).$$

*We are currently assuming a simple scenario with one car and one slot in the scene*

FONDAZIONE BRUNO KESSLER

UNIVERSITÀ DI PISA

# Encoding of simple and complex events using EC

**Complex events derived from simple events using EC formalism**

$$
\begin{aligned}
happens(covers(A, L), T) \leftarrow\ & agent(A), location(L), time(T), \\
& happens(disappearsSlot(L), T), \\
& holdsAt(visibleCar(A), T).
\end{aligned}
$$

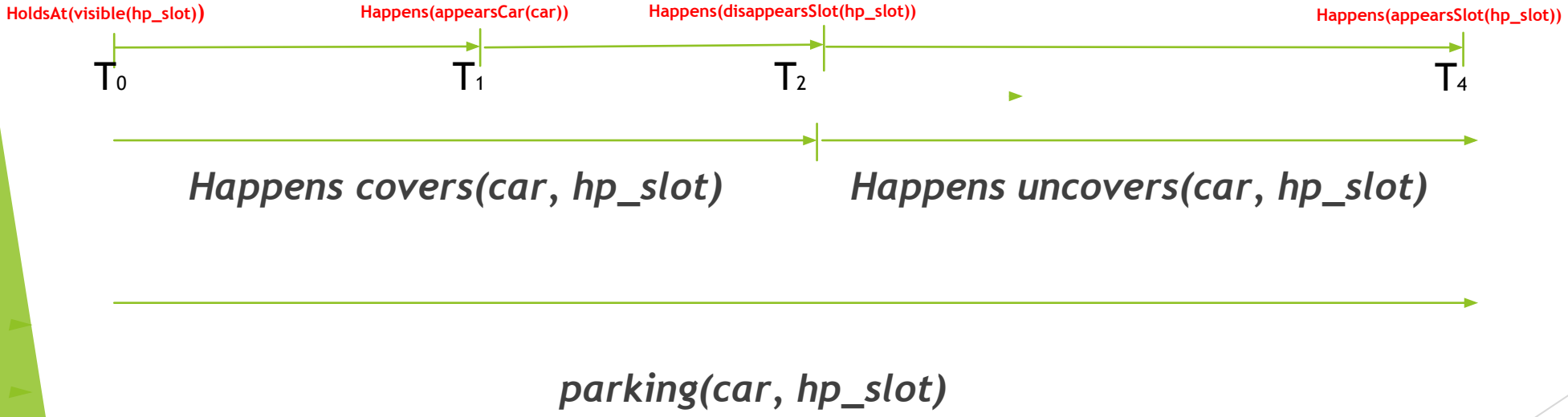
$$
\begin{aligned}
happens(uncovers(A, L), T) \leftarrow\ & agent(A), location(L), time(T), \\
& happens(appearsSlot(L), T), \\
& holdsAt(visibleCar(A), T).
\end{aligned}
$$

# Encoding of simple and complex events using EC

By these rules, we recognize that a car ***covers*** a slot if the car is visible at the time that the slot disappears. Similarly, the ***uncovers*** event occurs when a slot appears, and the car is still visible. By combining the information on complex events, we can define that a ***parking*** from time $T_1$ to time $T_2$ is detected whenever a car covers a slot at time $T_1$, uncovers the slot at time $T_2$ and it stands on the slot for at least a number of frames defined by ***parkingframes***.

$$parking(A, L, T1, T2) \leftarrow happens(covers(A, L), T1), happens(uncovers(A, L), T2), \\ parkingframes(N), T3 = T1 + N, T2 >= T3.$$

# Simple and complex events via Timeline



# Query on basic facts from tracker Output

$$holdsAt(visibleSlot(hp\_slot), 0).$$
$$happens(appearsCar(car), 1).$$
$$happens(disappearsSlot(hp\_slot), 2).$$
$$happens(appearsSlot(hp\_slot), 4).$$
$$happens(disappearsCar(car), 5).$$

Query: if there is a parking in the video? which objects and at what time?

*parking(A,L,T1,T2) ?*

*car, hp_slot, 2, 4.*

# Evaluation

we run the program on DLV using the output of the tracker from previous step. We were able to

detect complex events for some of the video sequences (e.g. *car 3* covers the *handicap slot 3* at

time-point 87 and uncovers the slot at time-point 107). Unfortunately, we could not apply the

method to the whole video: the reason stands in the ambiguities of tracker output (e.g. multiple

labelling of the same object, incorrect disappearance of objects) which produce unclean data.

FONDAZIONE BRUNO KESSLER

UNIVERSITÀ DI PISA

# Conclusion

The overall goal of this work is the integration of knowledge representation and computer vision:

1) Visual processing pipeline for detection-based object tracking, leading to the extraction of simple events.

(2) Answer set programming-based reasoning to derive complex events

# Future work

For the future work we aim to manage inaccuracies of the tracker output by a (possibly logical based) data cleaning step. We also want to apply and evaluate the presented method in different scenarios e.g (sports videos)

FONDAZIONE BRUNO KESSLER

UNIVERSITÀ DI PISA

# THANK YOU