Multiple Object Tracking Systems

Javier Berneche

Senior Machine Learning Engineer

Tryolabs

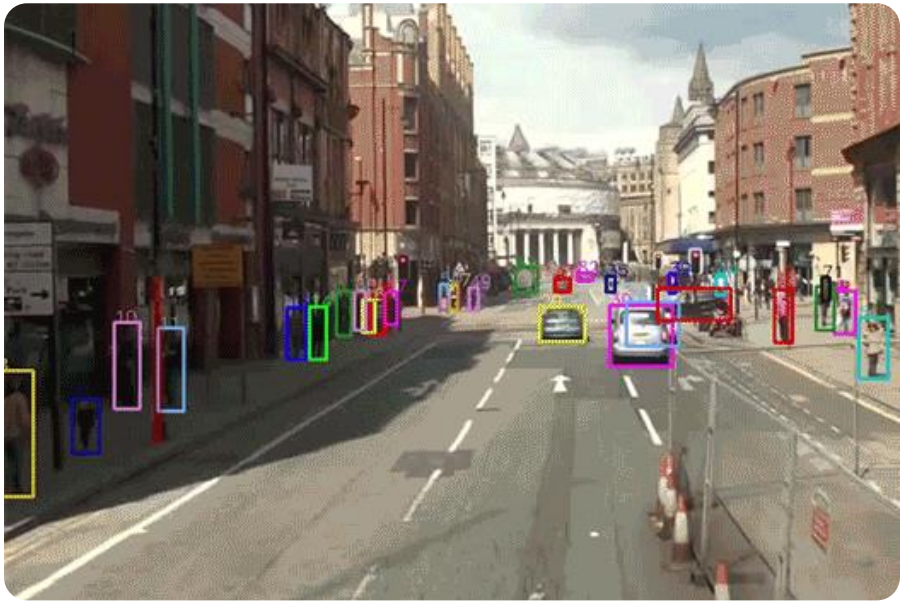
tryo·labs

Agenda

1. Introduction to Multiple Object Tracking (MOT)
2. Building blocks
3. Challenges
4. Evaluation and promising research

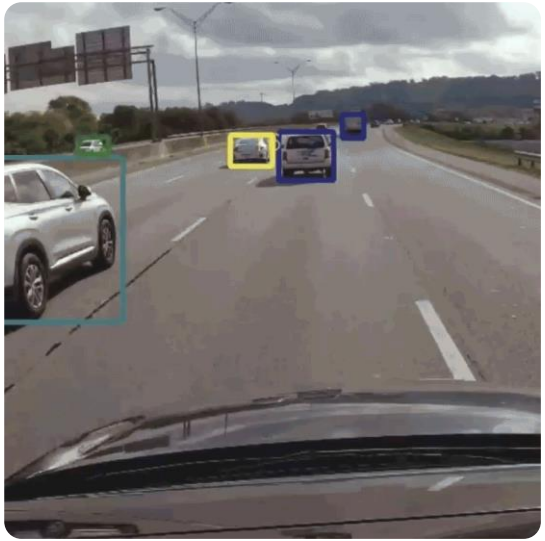
Definition

Multiple Object Tracking (MOT) is the problem of identifying **multiple objects** in a video or live feed and representing them as a set of **trajectories**



Video: MOT Challenge

Applications

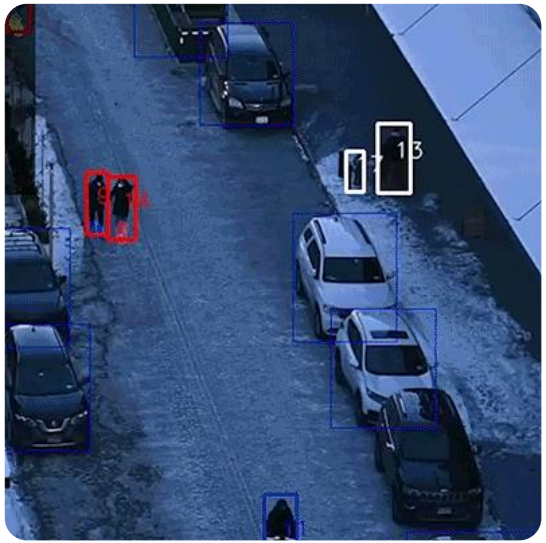


Autonomous navigation systems

Analyze and monitor congestion

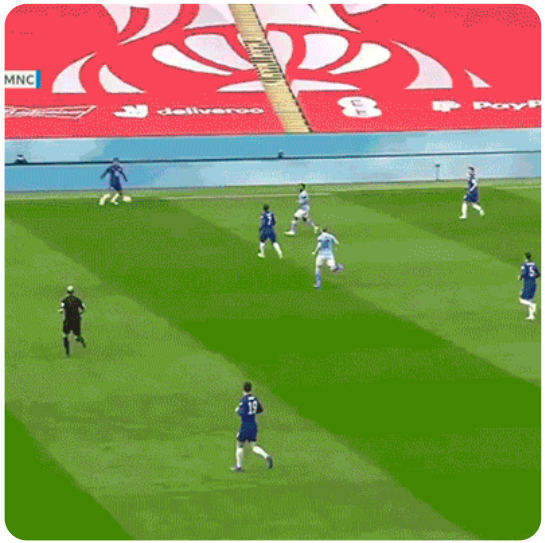
Augmented reality

Applications



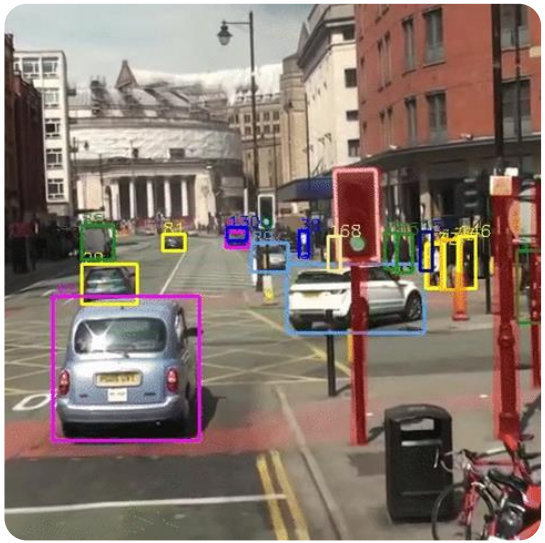
Surveillance

Crowd analysis

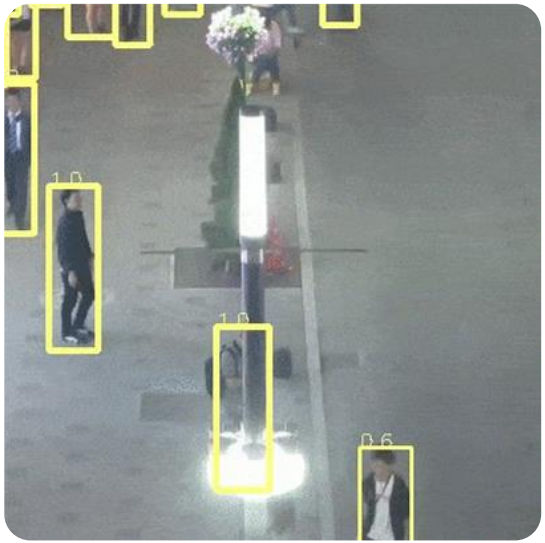


Sports analytics

Challenges



Changes in appearance



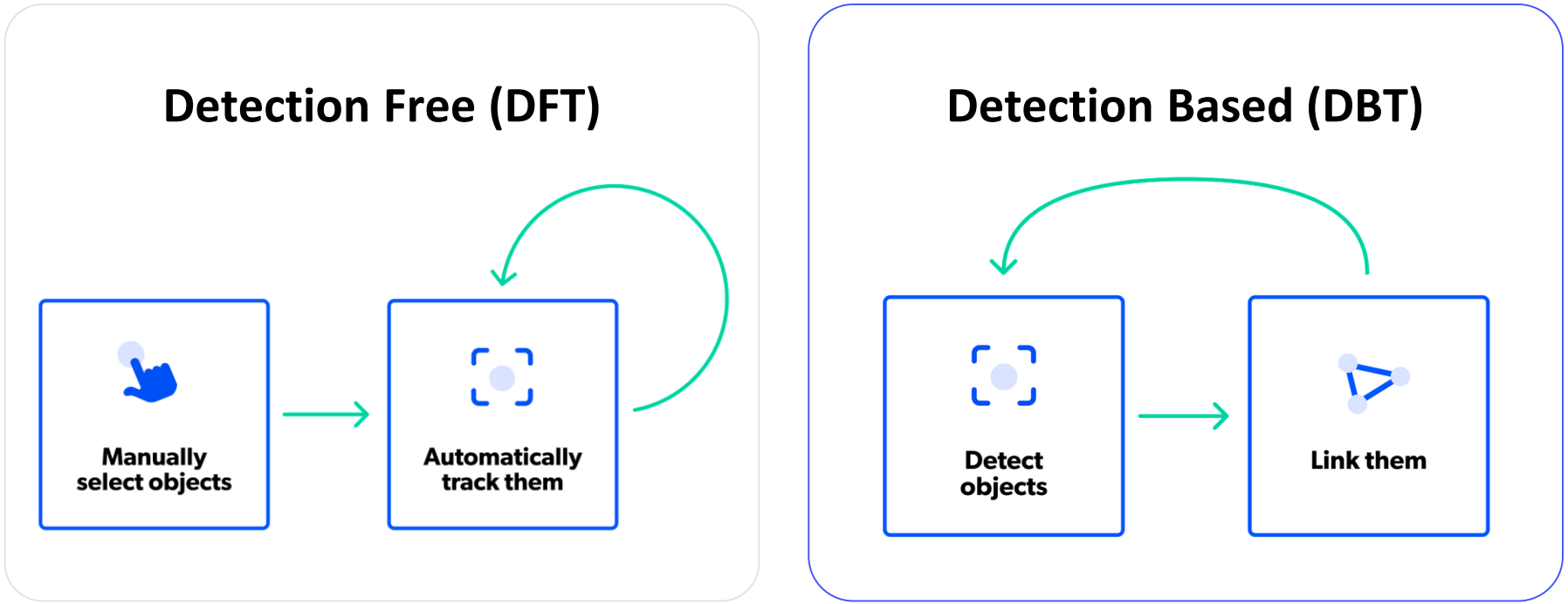
Occlusions

Crowded scenes

Building Blocks

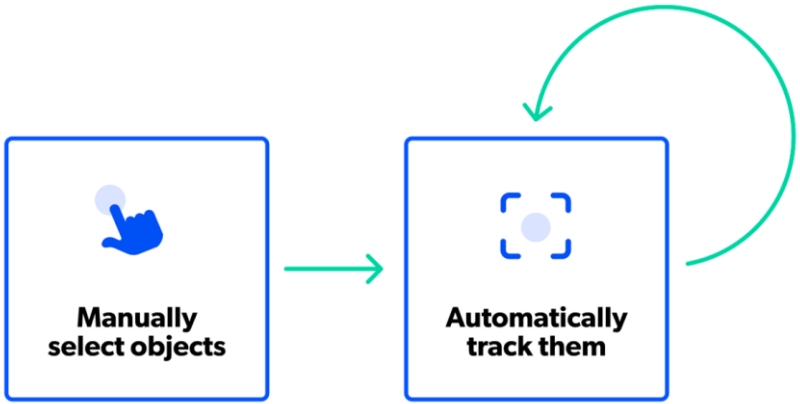
tryo·labs

Building Blocks: Initialization



Building Blocks: Initialization

Detection Free (DFT)

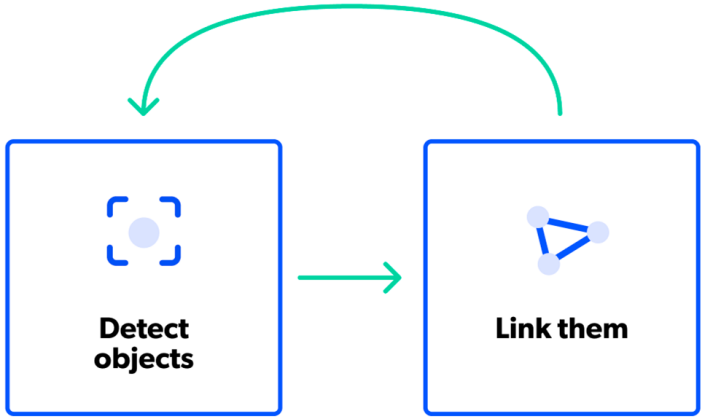


- Automatically track **manually** selected objects
- Can track **any** type of object
- **Can’t** handle new objects coming into the scene

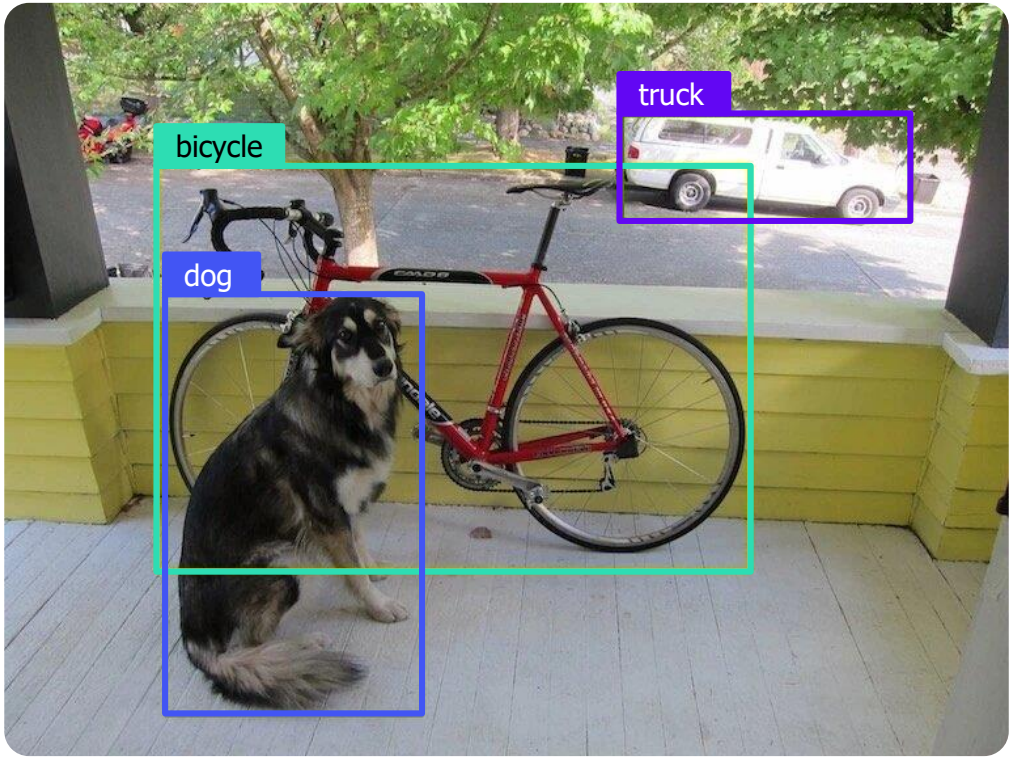
Building Blocks: Initialization

- Detect objects in **each** frame
- New objects are **automatically** discovered
- Object classes need to be **predefined**

Detection Based (DBT)

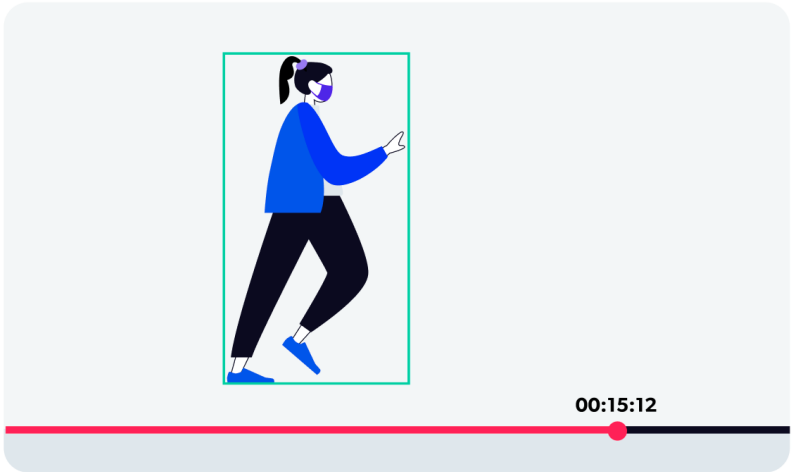


Building Blocks: Detector

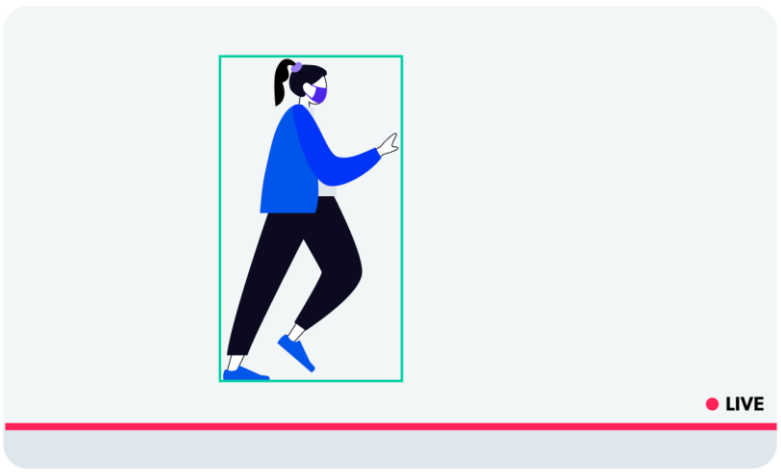


Building Blocks: Processing

Offline

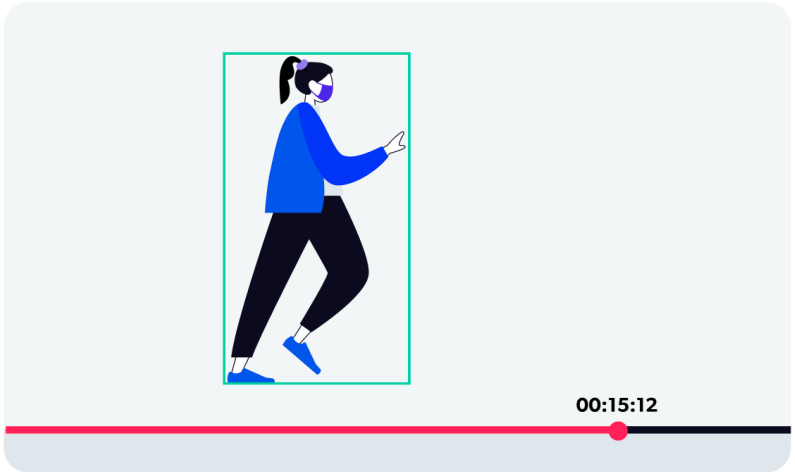


Online



Building Blocks: Processing

Offline

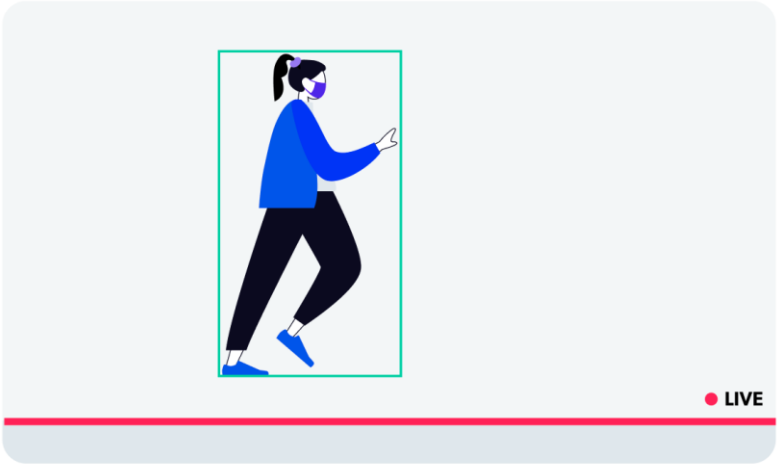


- Information from **future** frames can be used
- **Interpolate** trajectories
- Keep different **hypotheses**
- Can be more **accurate**

Building Blocks: Processing

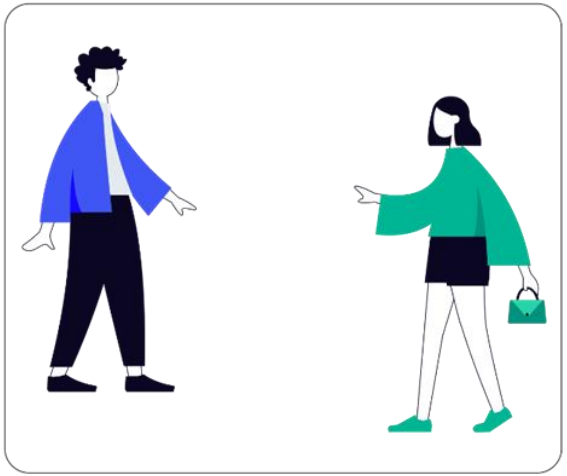
- Only **past** information is used
- **Extrapolate** trajectories
- **Decide** on every frame
- Can be used in **real-time** applications

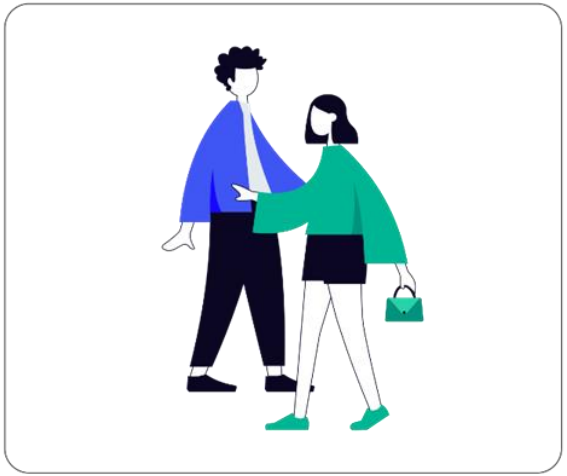
Online

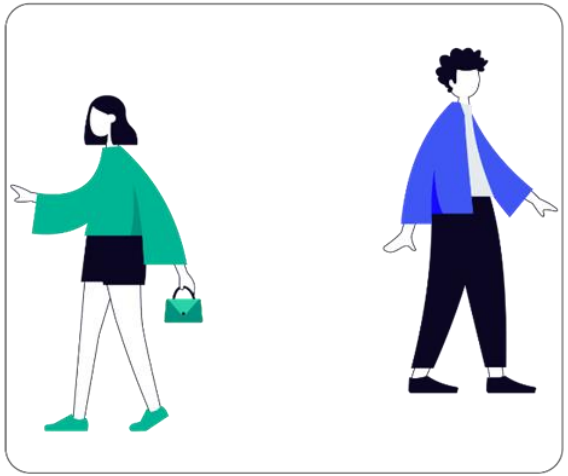


Building Blocks: Positional Cues

Frames





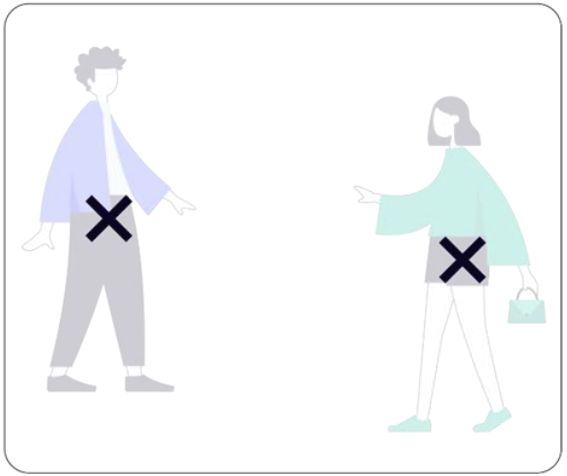


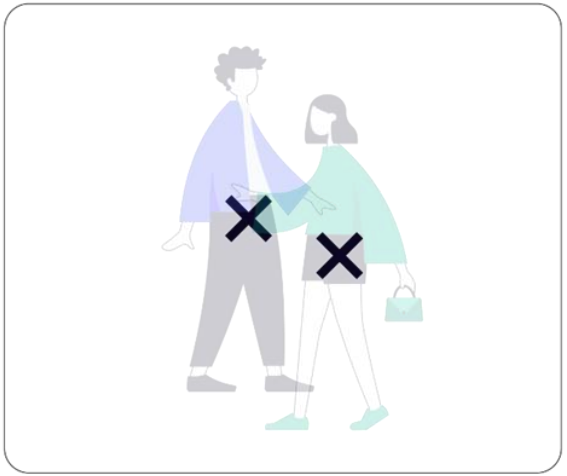
Building Blocks: Positional Cues

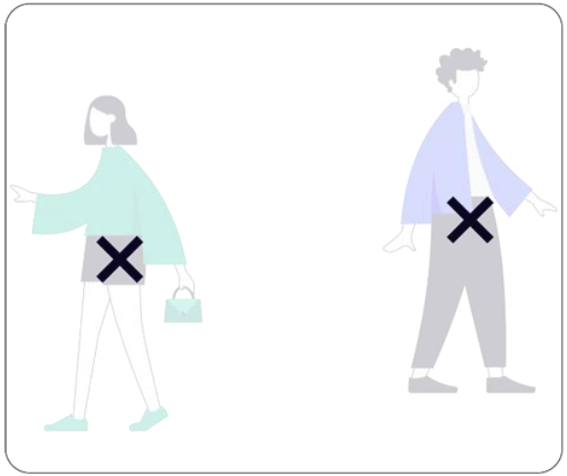
Detection

Building Blocks: Positional Cues

Centroids

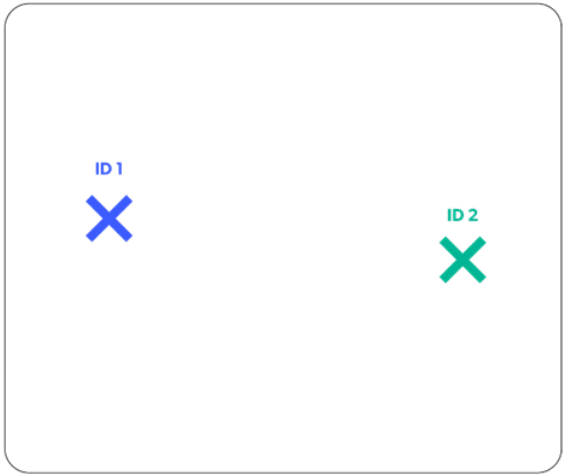


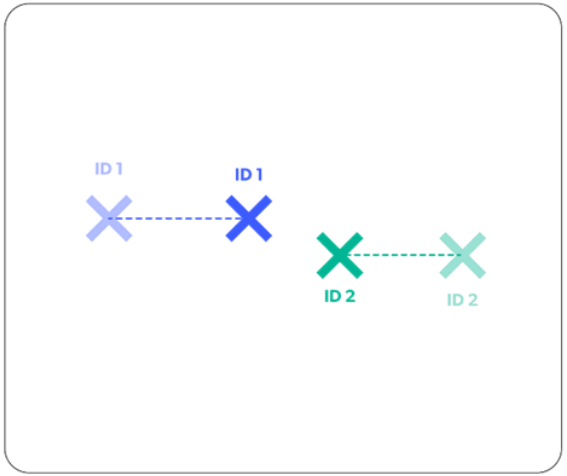


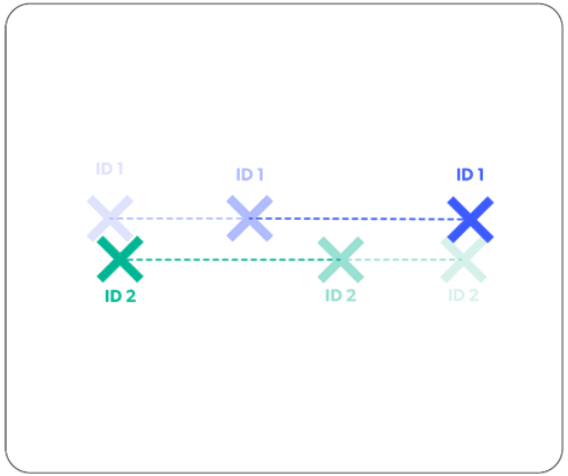


Building Blocks: Positional Cues

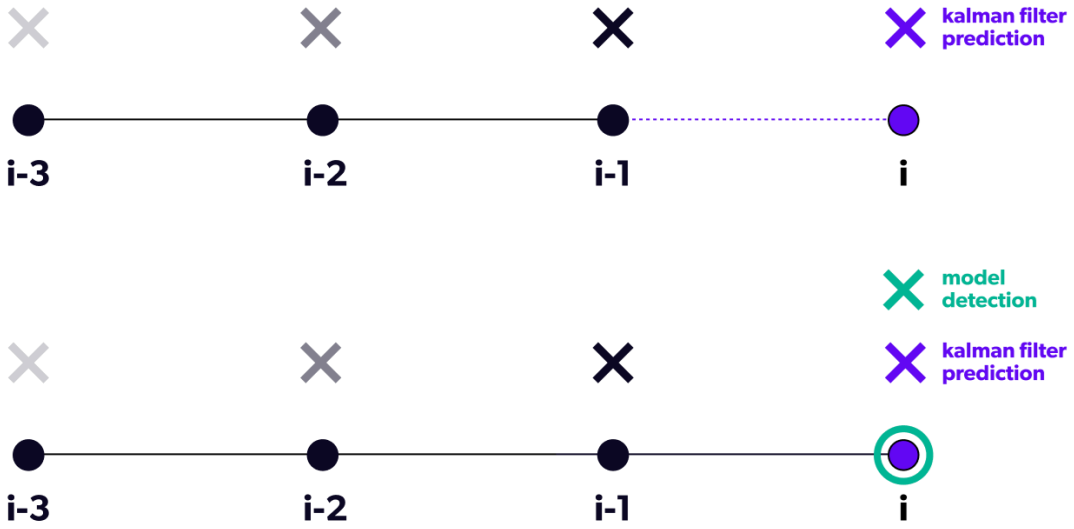
Assignment







Positional Cues: Kalman Filter



Positional Cues: Assignment

Greedy

- Simpler
- Faster

Hungarian

- Minimizes global distance
- Slower

Visual Cues

Important for **recovering** from occlusions and collisions

Helps with **long-term** tracking

Generally used to **complement** the positional cues

Visual Cues: Vectors

Classical approaches can be used

Usually deep-learning **embeddings** are used

Cosine distance is usually used to compare the embeddings

Visual Cues: History

- Decide how to represent an object's embedding considering all **past embeddings**
- Rolling **averages**
- **Clustering** to maintain different versions of the object
- Memory usage and computational **cost** of comparison

tryo·labs

Recap

A complete tracking system

Object Detector

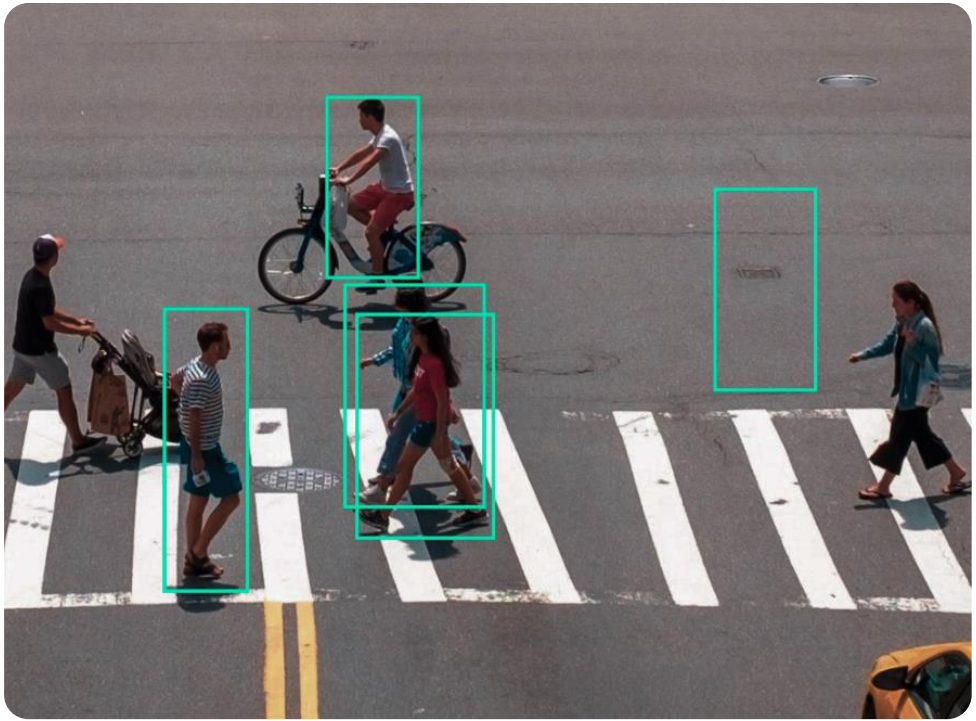
Positional cues for short-term tracking

Visual cues for long-term tracking

Challenges

tryo·labs

Challenges: Detection Quality



False positives

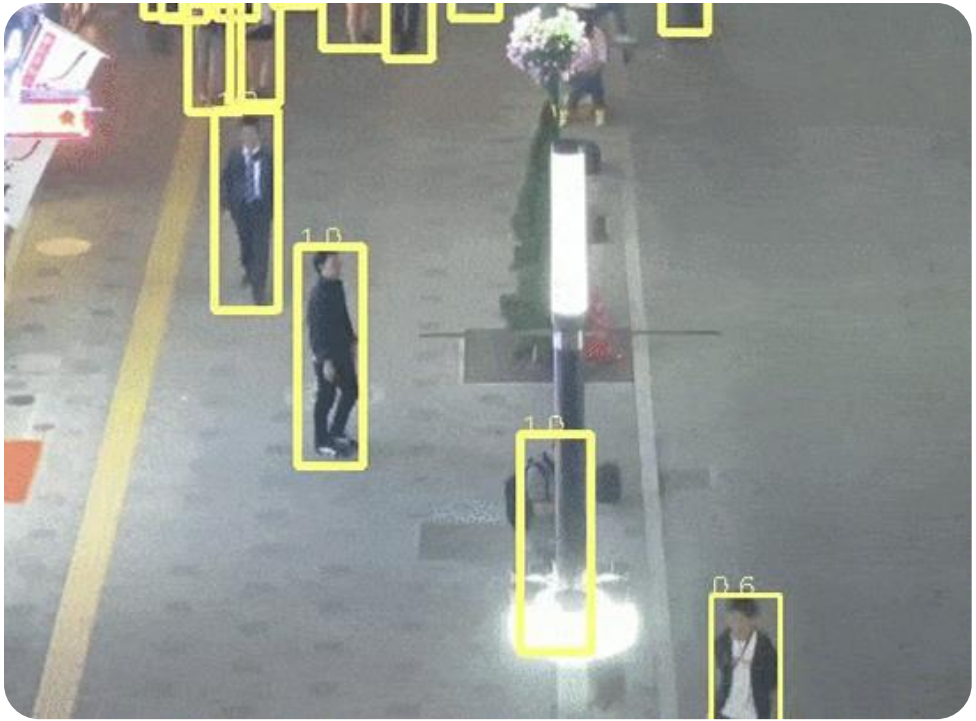
False negatives

Challenges: Movement

Erratic movement of the objects

Camera movement

Challenges: Occlusions



- Causes more false negatives
- Positional tracking can fall apart
- Embeddings of partially occluded objects can be bad

Challenges: Embeddings

Object detectors usually do not yield good **embeddings**

Need to add a **second model** for embeddings

Partial occlusions

No obvious model to start with

Evaluation & Research

tryo·labs

Promising Research

- Improvements to embeddings
- A single model for detections and embeddings
- One-stage models
- More datasets

MOT-Challenge Benchmark

Go-to benchmark for MOT

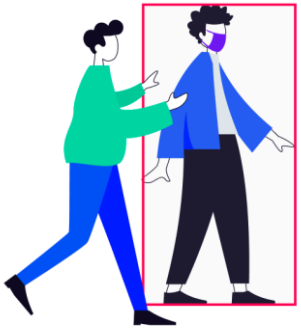
More datasets are added periodically

Few classes

Videos are short

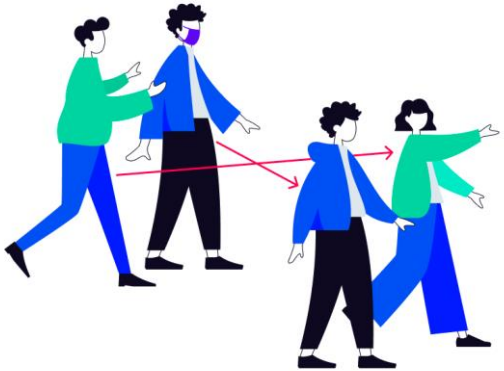
Evaluation Metrics

Detection error



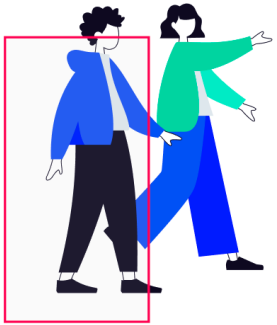
Multi-Object Tracking Accuracy (MOTA)

Association error



Identity F1 (IDF1)

Localization error



High Order Tracking Accuracy (HOTA)

Conclusions

MOT has a huge variety of **applications**

The problem is **challenging**

Solutions involve a number of **components**

Lots of promising **research**

Resources

- MOTChallenge
- Luo et al. 2021 Literature review
- Laura Leal-Taixe
- Object detection
- Hungarian method
- Greedy matching
- Embeddings
- Kalman filter

Open-Source Tools

Trackers

- ByteTrack
- Norfair
- SORT
- DeepSORT

Tools

- MOTMetrics
- YOLO
- OpenMMLab

Trackers

- SORT
- DeepSORT
- BoT-SORT
- ByteTrack
- SMILEtrack
- SUSHI

Thank You!

tryo·labs