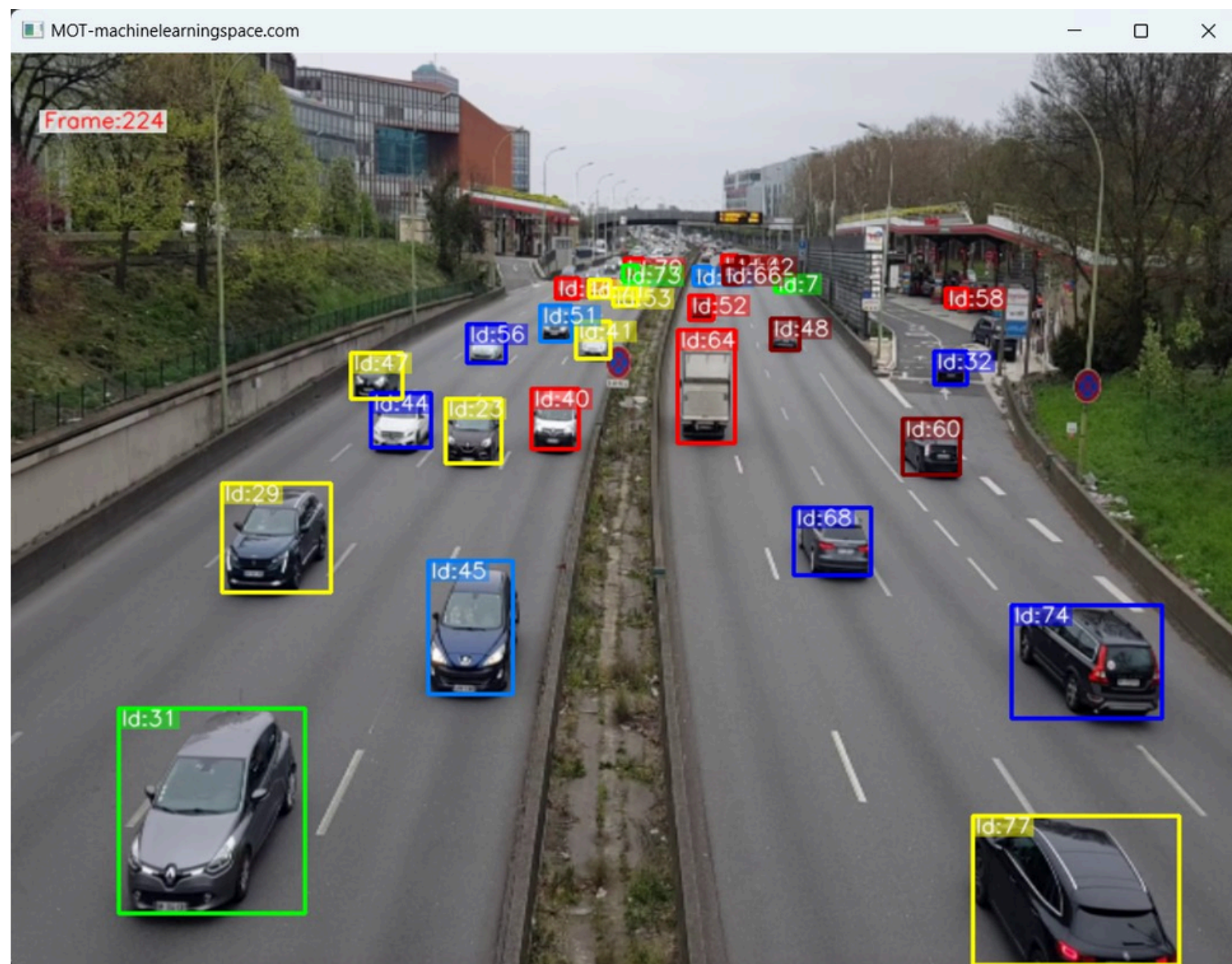


موضوع

MOT

Multi Object Tracking



MOT چندین شیء را به طور همزمان شناسایی و ردیابی می کند و برای هر کدام یک شناسه (ID) پایدار نگه می دارد.

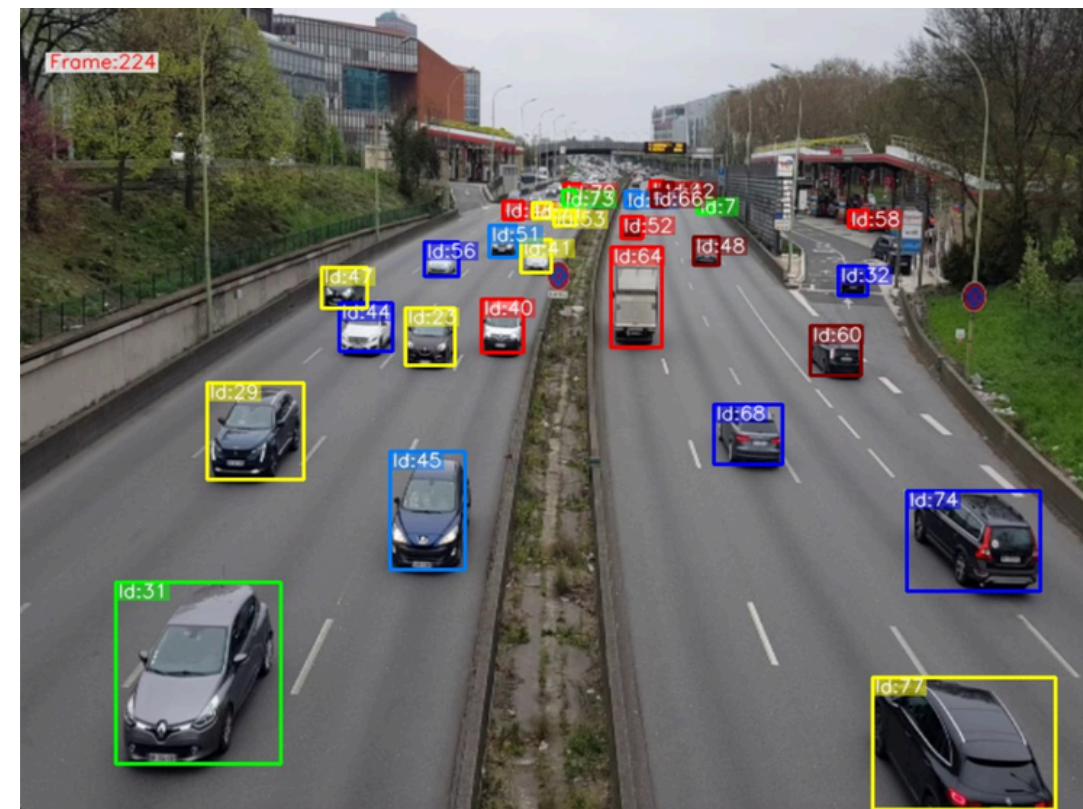


SOT

یک شی

پیدا کردن همان شیء اولیه در هر فریم

معمولا: فقط در فریم اول Object Detection انجام می شود.



MOT

چندین شی همزمان

شناسایی، ردیابی و حفظ ID همه اشیاء در فریم‌ها

معمولا: در همه ی فریم ها (یا هر چند فریم) باید OD انجام شود.

Video



سیستم یک ویدیو یا فریم‌های زنده از دوربین دریافت می‌کند.

Object Detection

در هر فریم، اشیاء (افراد، ماشین‌ها، ...) با استفاده از یک مدل تشخیص (مثلاً YOLO) شناسایی می‌شوند. $[x1, y1, x2, y2, confidence, class_id]$

Filtering

- حذف detection‌های کم‌اعتماد (confidence پایین)
- انتخاب کلاس‌های خاص (مثلاً فقط انسان)

Difference Detector(Appearance Feature Extraction)

- برای هر جعبه (Bounding Box)، یک شبکه‌ی Re-ID ویژگی بصری استخراج می‌کند.
 - هدف: اگر ظاهر شیء ثابت باشد، بهتر بتوان آن را ردیابی کرد.
- 128-dimensional vector (embedding)

Motion Prediction

- با استفاده از Kalman Filter، موقعیت احتمالی اشیاء در فریم جدید پیش‌بینی می‌شود.

Matching

- ترکیب شباهت مکانی (IOU) و شباهت ظاهری (Cosine Distance)
- اختصاص detection جدید به مسیر قبلی با الگوریتم Hungarian

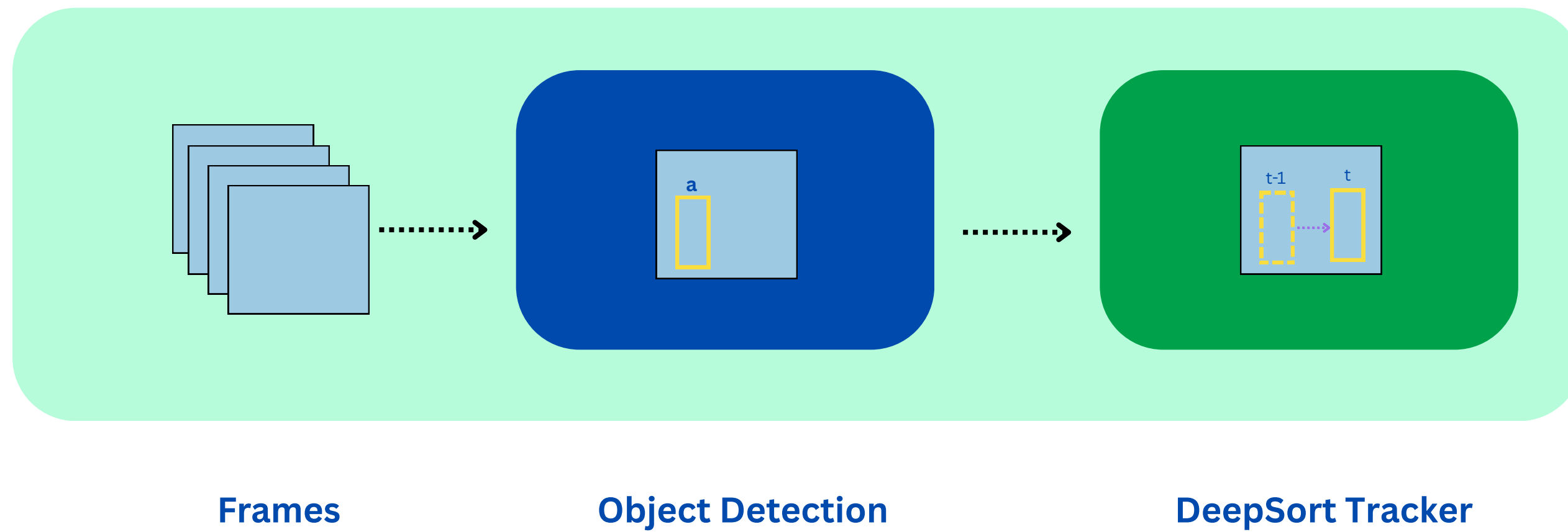
Track Management

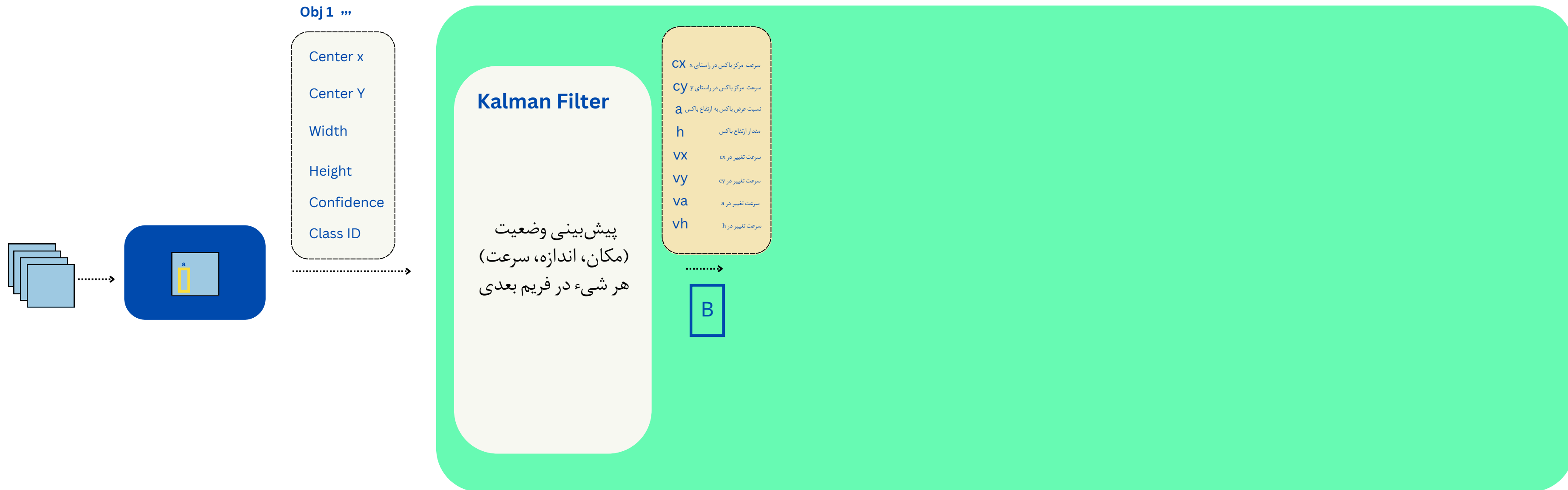
- به‌روزرسانی یا حذف مسیر

Visualization / Saving

YOLOv8 + DeepSORT

MOT

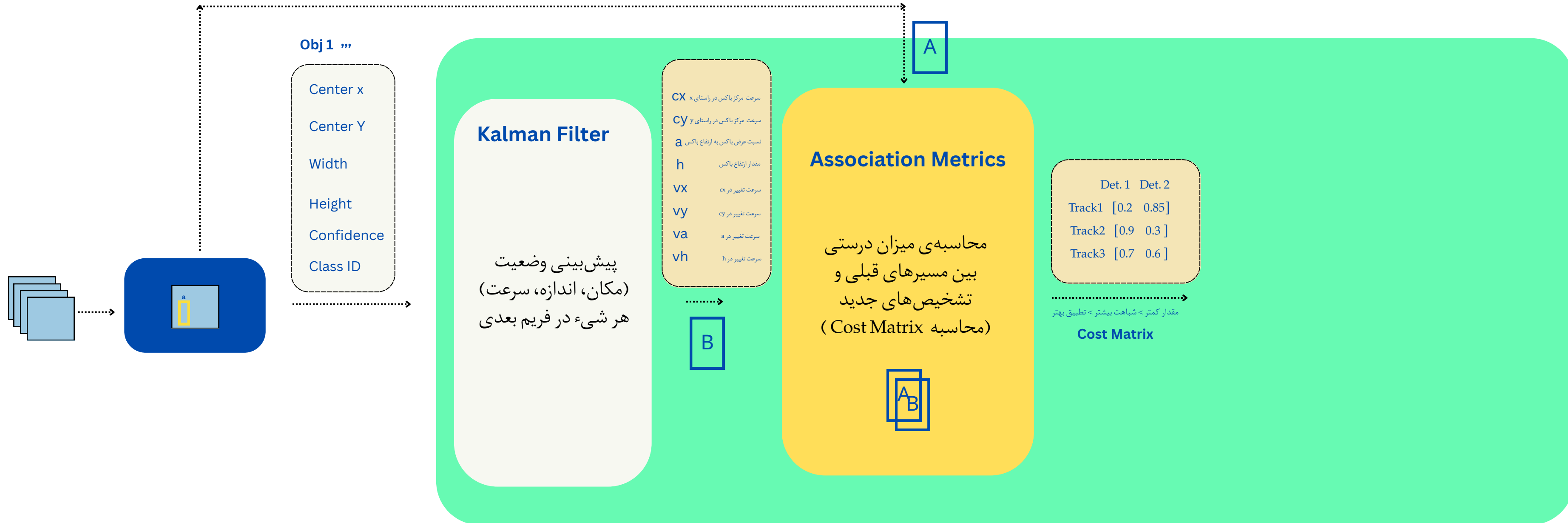




Frames

Object Detection
YOLOV8

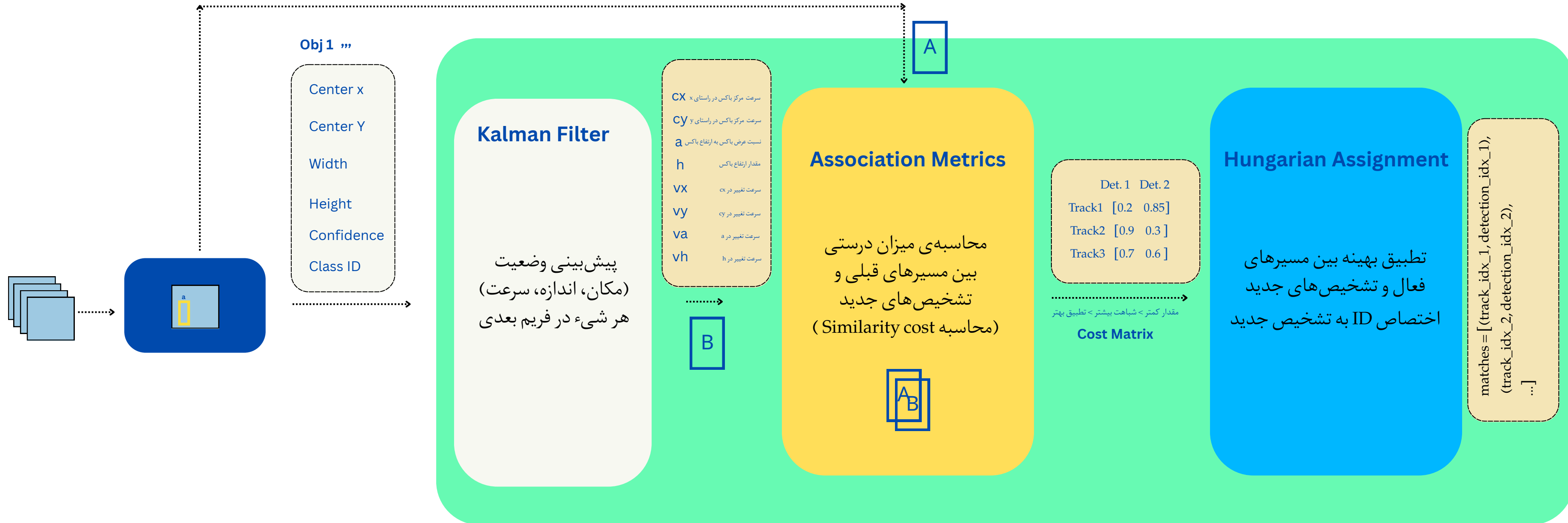
DeepSort Tracker



Frames

Object Detection

DeepSort Tracker



Frames

Object Detection

DeepSort Tracker

خروجی Yolo

D

Appearance-based Metric

Motion-based Metric

T

خروجی KF

Association Metrics

محاسبه‌ی هزینه‌ی تطبیق
(Matching Cost) بین هر
مسیر فعال (track) و هر
detection جدید

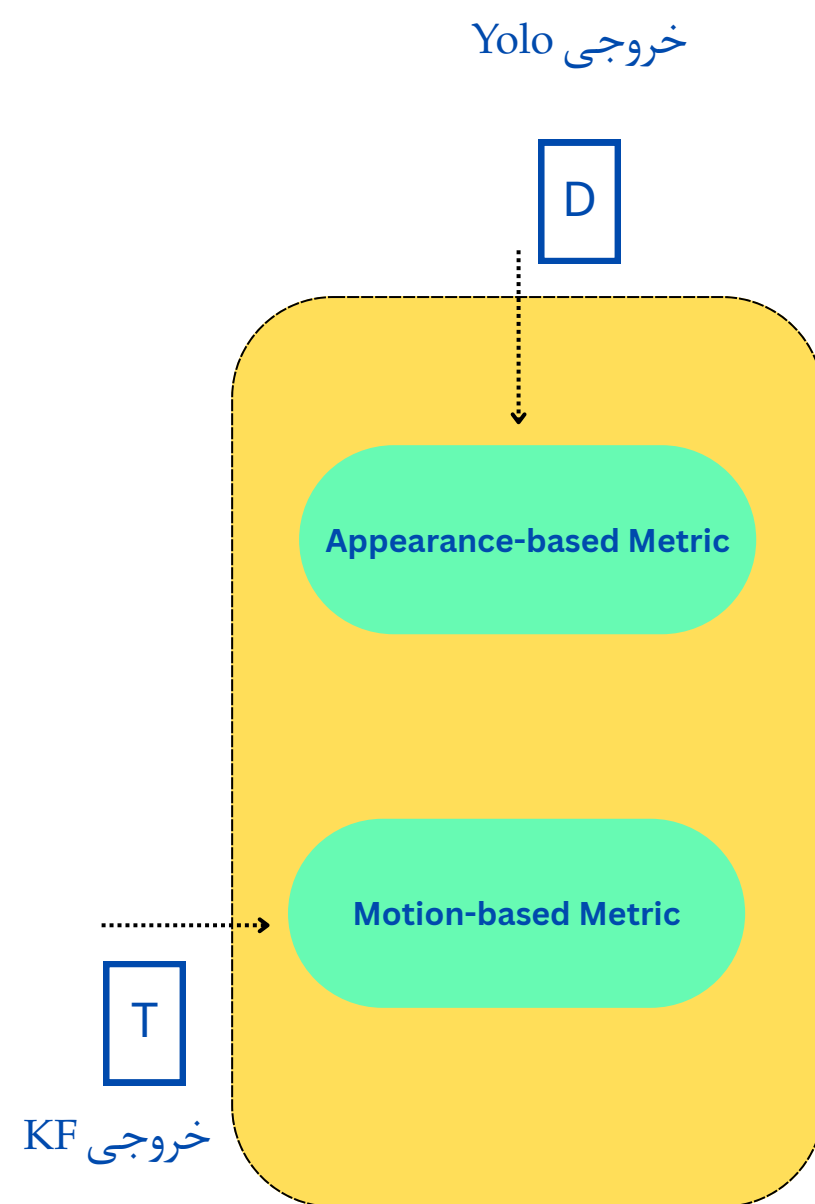
.....→ Cosine distance $= 1 - \text{cosine_similarity}(f_{\text{track}}, f_{\text{det}})$

.....→ Mahalanobis distance $d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y})}$

x : مکان تشخیص داده شده

y : مکان پیش‌بینی شده توسط Kalman

S : ماتریس کوواریانس پیش‌بینی Kalman



Association Metrics

محاسبه‌ی هزینه‌ی تطبیق
(Matching Cost) بین هر
مسیر فعال (track) و هر
detection جدید

$$\text{cost}[i][j] = \lambda * \text{motion_cost} + (1 - \lambda) * \text{appearance_cost}$$

$\lambda = 0.98$ در محیط‌های با حرکت کم
 $\lambda = 0.5$ در محیط‌های شلوغ و با occlusion زیاد

Hungarian Assignment

تطبیق بهینه بین مسیرهای
فعال و تشخیص‌های جدید
اختصاص ID به تشخیص جدید

	$D1$	$D2$	$D3$
$T1$	4	1	3
$T2$	2	0	5
$T3$	3	2	2

matches = [(0, 0), (1, 1), (2, 2)]

◀ **بریم سراغ کدش!**

MOT