object tracking

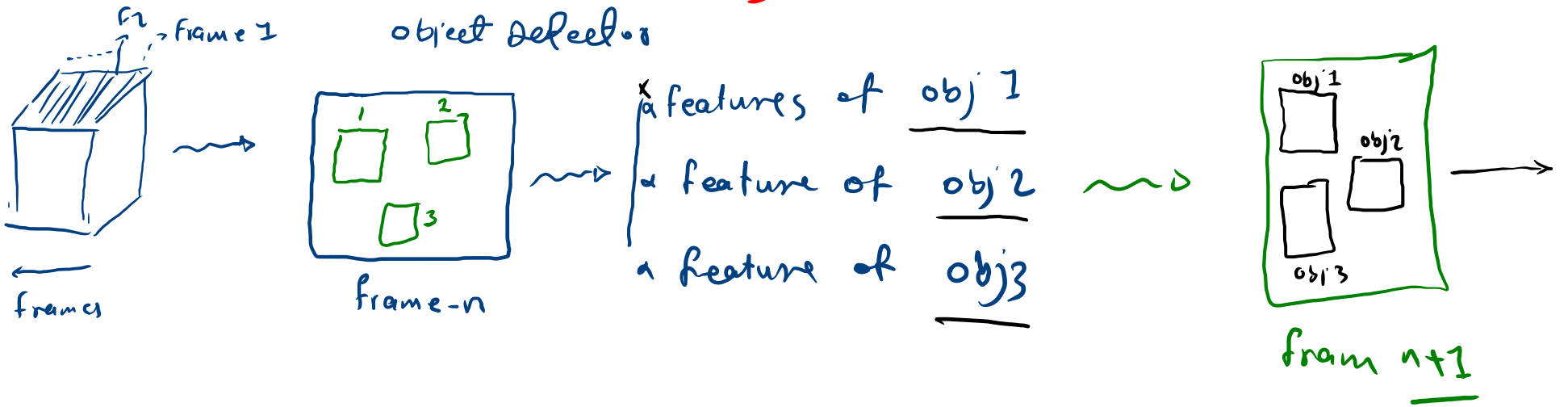
[دنبال کردن یک یا چند پیش‌نمایش در طی فرایند ویدئو]

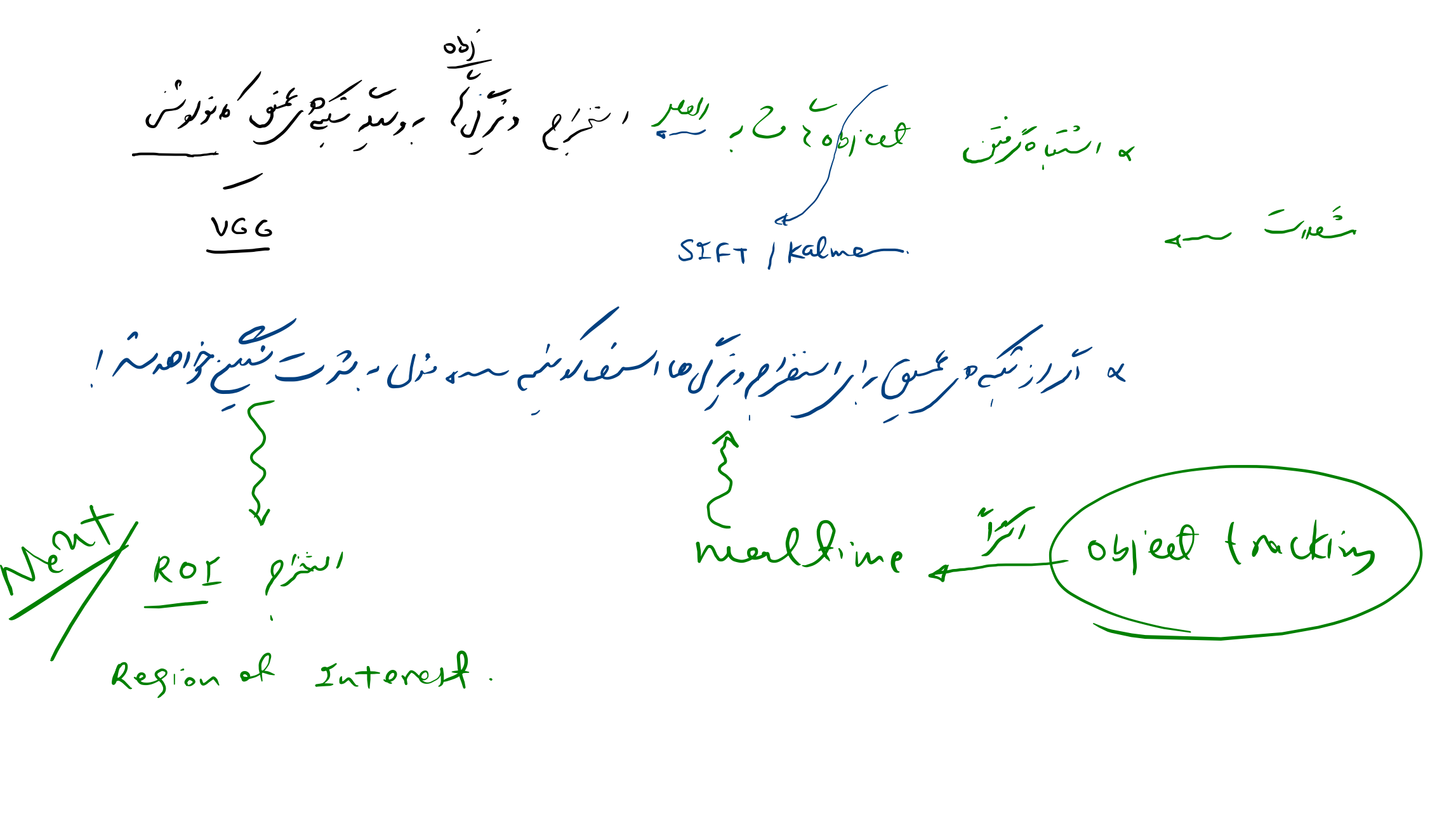
[object tracking]

(به صورت کلی)

اشعه ریتم

object detector





* اشتباه گرفتن object ها به العلم استخراج ویژگی‌ها (obj) به وسیله شبکه‌های عمیق کانولوشن

VGG

SIFT / Kalman

مشکلات

* اگر از شبکه‌های عمیق برای استخراج ویژگی‌ها استفاده کنیم مدل به شدت سنگین خواهد شد!

Next

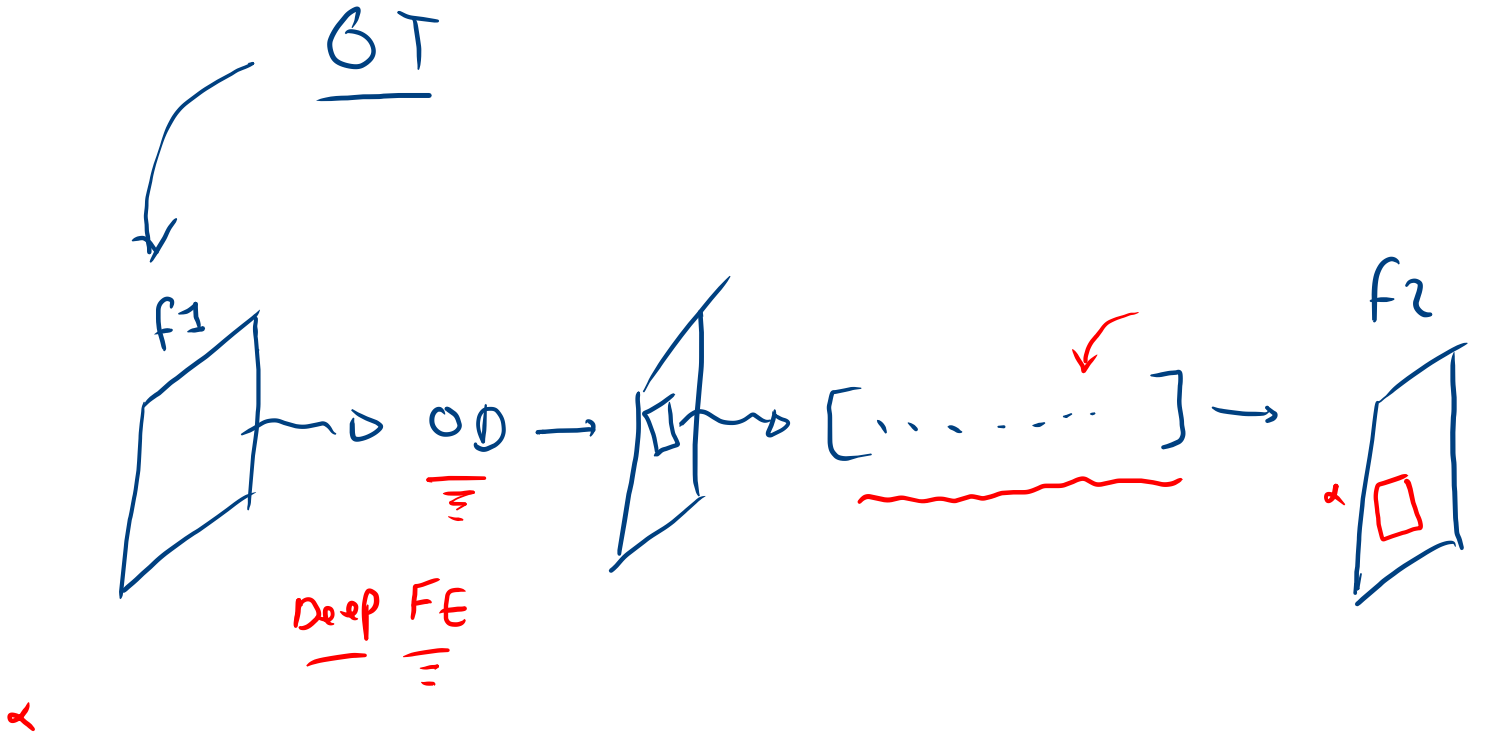
ROI استخراج

Region of Interest.

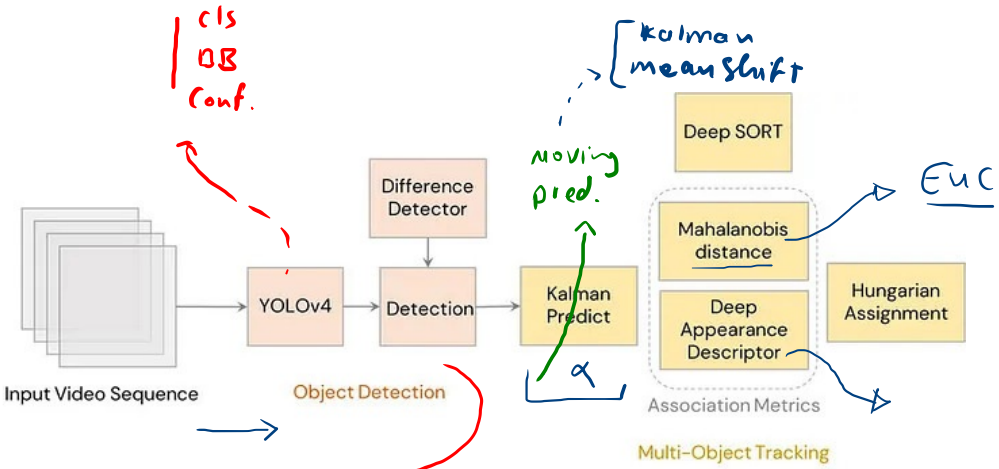
real time

اکثراً

object tracking


GT
f1
f2
Deep FE

deepsort



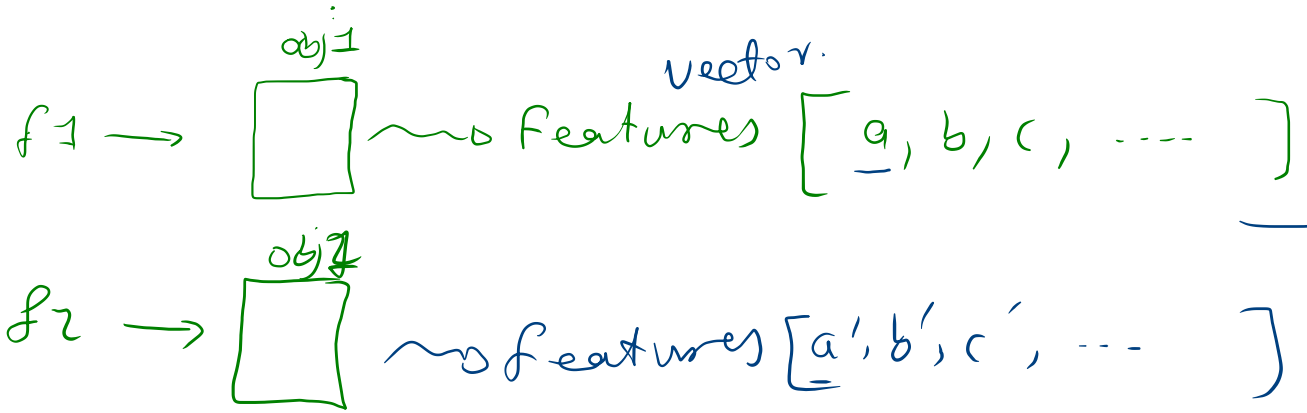
predicted position.

meanshift

frame 1 → OD → objs / BB → predict new position of objs. (in next frame)

frame 2 → OD → objects / BB → A vs B → motion cost, app cost

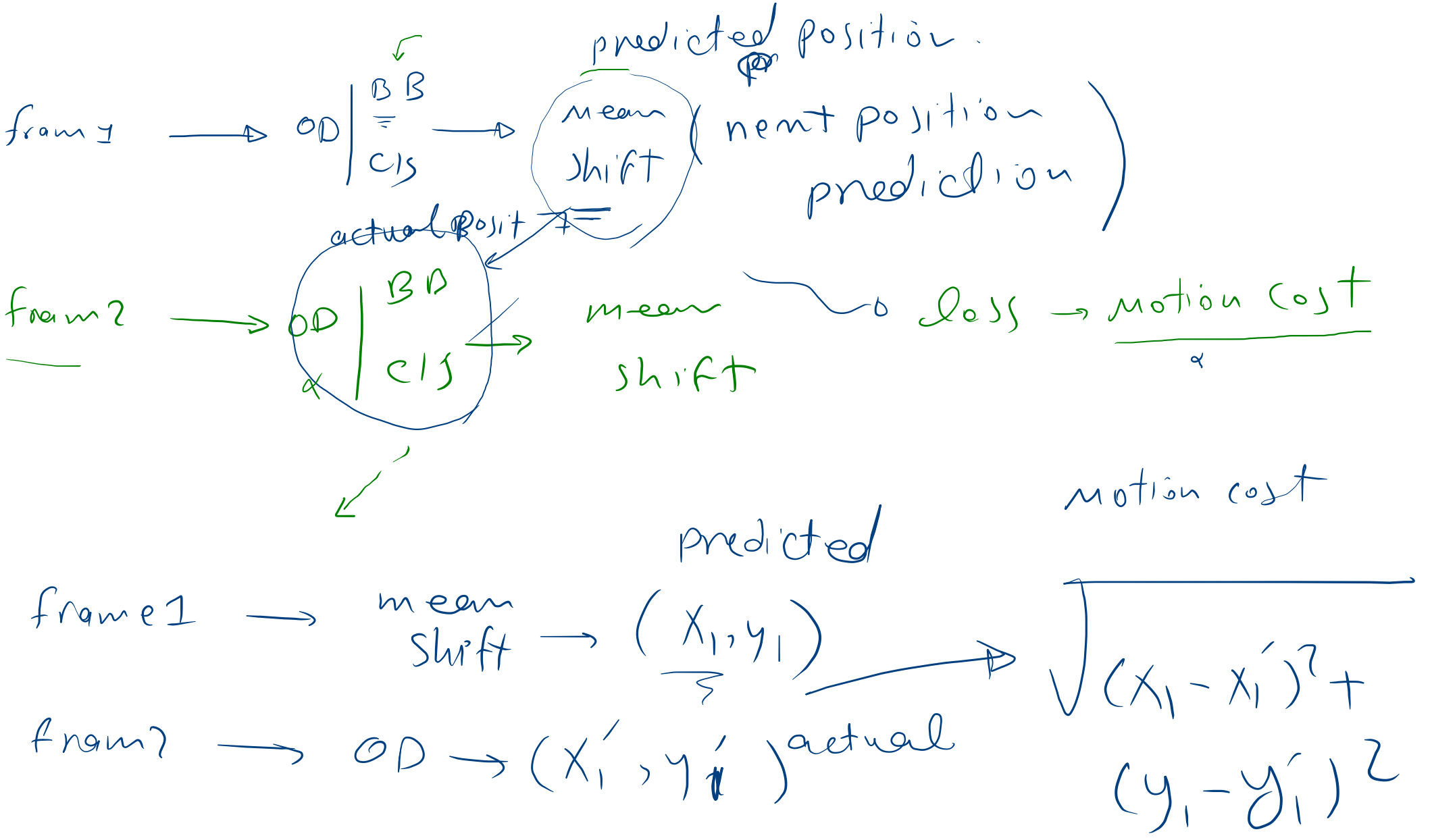
Deep Appearance Descriptor.



$$\text{loss} = \sqrt{(a-a')^2 + \cdots}$$

α

تضمین کند که object در frame 1 و object در frame 2 است!

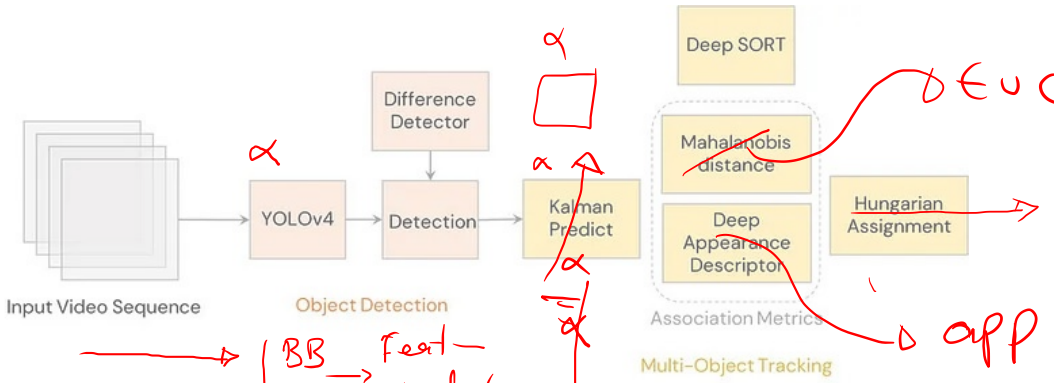

predicted position
frame 1
OD
BB
cls
mean shift
next position prediction
actual posit
frame 2
OD
BB
cls
mean shift
loss
motion cost
frame 1
mean shift
predicted
(x_1, y_1)
frame 2
OD
(x_1', y_1') actual
motion cost
$\sqrt{(x_1 - x_1')^2 + (y_1 - y_1')^2}$

Motion cost ← تضمین می‌کند که BB به جزئی object مد نظر را دنبال می‌کند!

Appearance cost ← تضمین می‌کند که object نمت صفر داده شده و دنبال شده

در frame ۲ همان object، مزیر ①

است!



euc-dist

motion cost

motion

loss = cost

+ app-cost

BB
cls → feat-vector

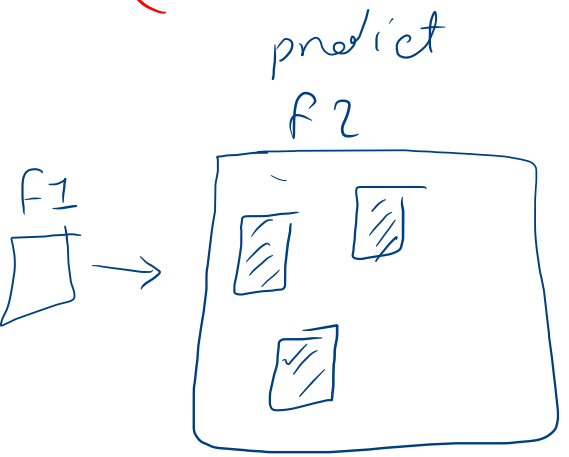
F2

BB
cls

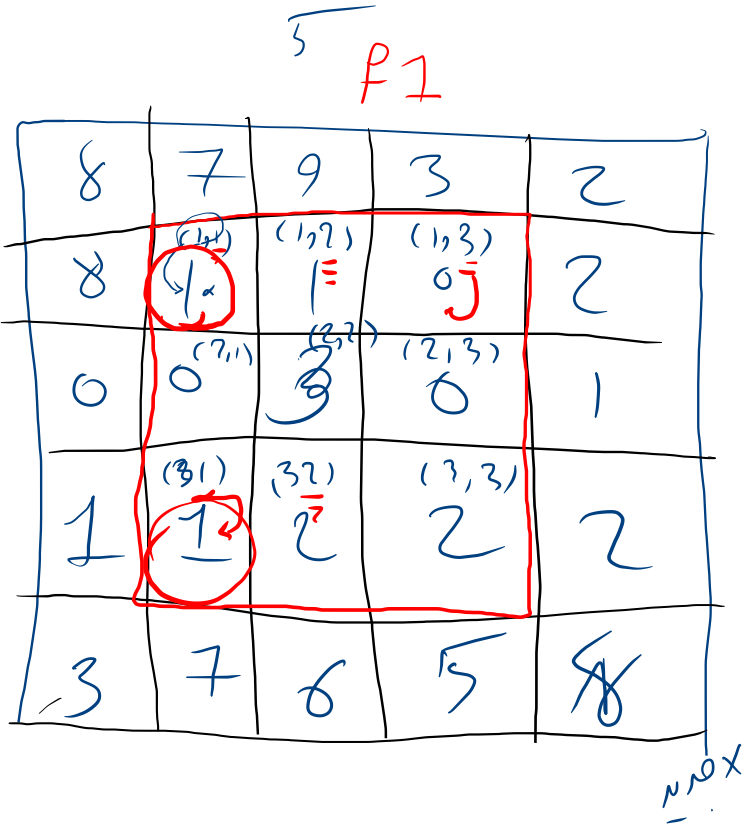
app-cost

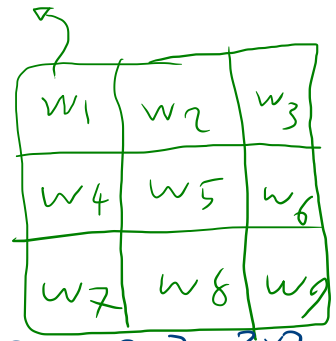
position predict

[position prediction]



[mean shift]





$w_1(1,1)$

$w_2(1,2)$

$w_3(1,3)$

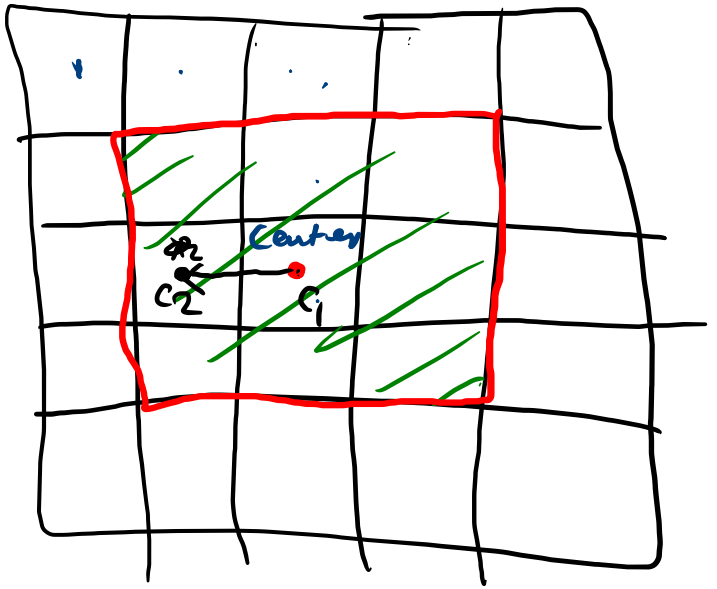
$$\text{mean } X = \frac{1\times1 + 1\times1 + 1\times0 + 2\times0 + 2\times3 + 2\times0 + 3\times1 + 3\times2 + 3\times2}{1+1+3+1+2+2}$$

$$\text{mean } X = \frac{23}{9}$$

$$\text{mean } y = \frac{1\times1 + 2\times1 + 3\times0 + 1\times0 + 2\times3 + 3\times0 + 1\times1 + 2\times2 + 3\times2}{1+1+3+1+2+2}$$

$$\text{mean } y = 2{,}6$$

f 1



BB-center $\overset{x\ y}{(3,3)}$ $\xrightarrow[f2]{\text{mean shift}}$ $(2,3)$

f

predicted

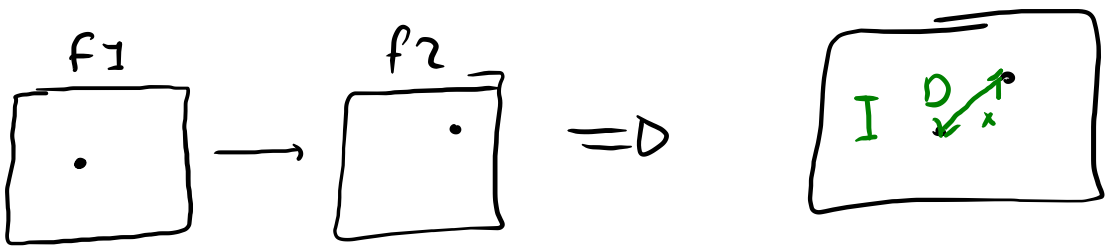
$f2 \rightarrow$ BB-center $= (2,4)$

$\rightarrow$ motion cost $= \sqrt{(2-2)^2+(4-3)^2}$

$= \sqrt{0+1} = 1$

نحوه استخراج سرعت در ویدئو

video → FPS → $\frac{\text{frames}}{\text{second}}$ ⇒ 30 $\frac{\text{frames}}{\text{sec}}$



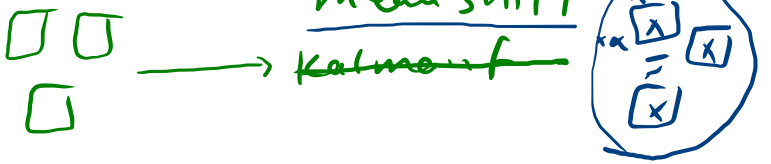
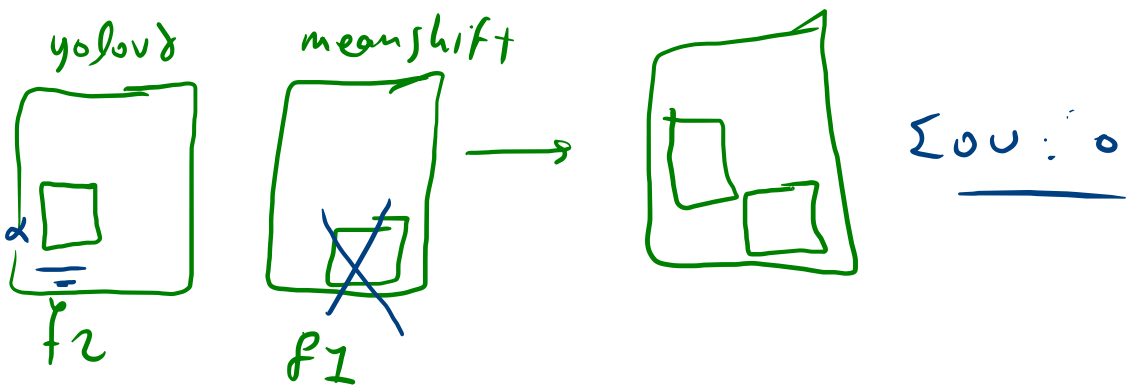
$D = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$

$\frac{D}{\text{time}} \rightarrow V = \frac{D \times \text{coef.}}{t}$

فاصله بین فریم‌ها : $\frac{1}{30}$ s

inference

yolov8



yolov8