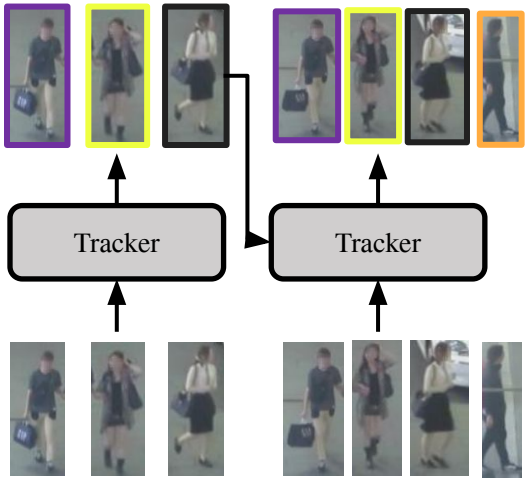
Simple Cues Lead to a Strong Multi-Object Tracker

Jenny Seidenschwarz, Guillem Brasó, Victor Castro Serrano, Ismail Elezi, Laura Leal-Taixé

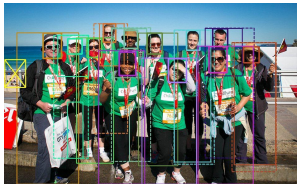
WED-PM-139

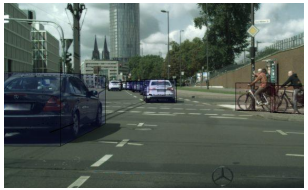
Haunting Completely End-to-End Trackers with GHOST 👻

Tracking-by-Detection



Haunting Completely End-to-End Trackers with GHOST 👻



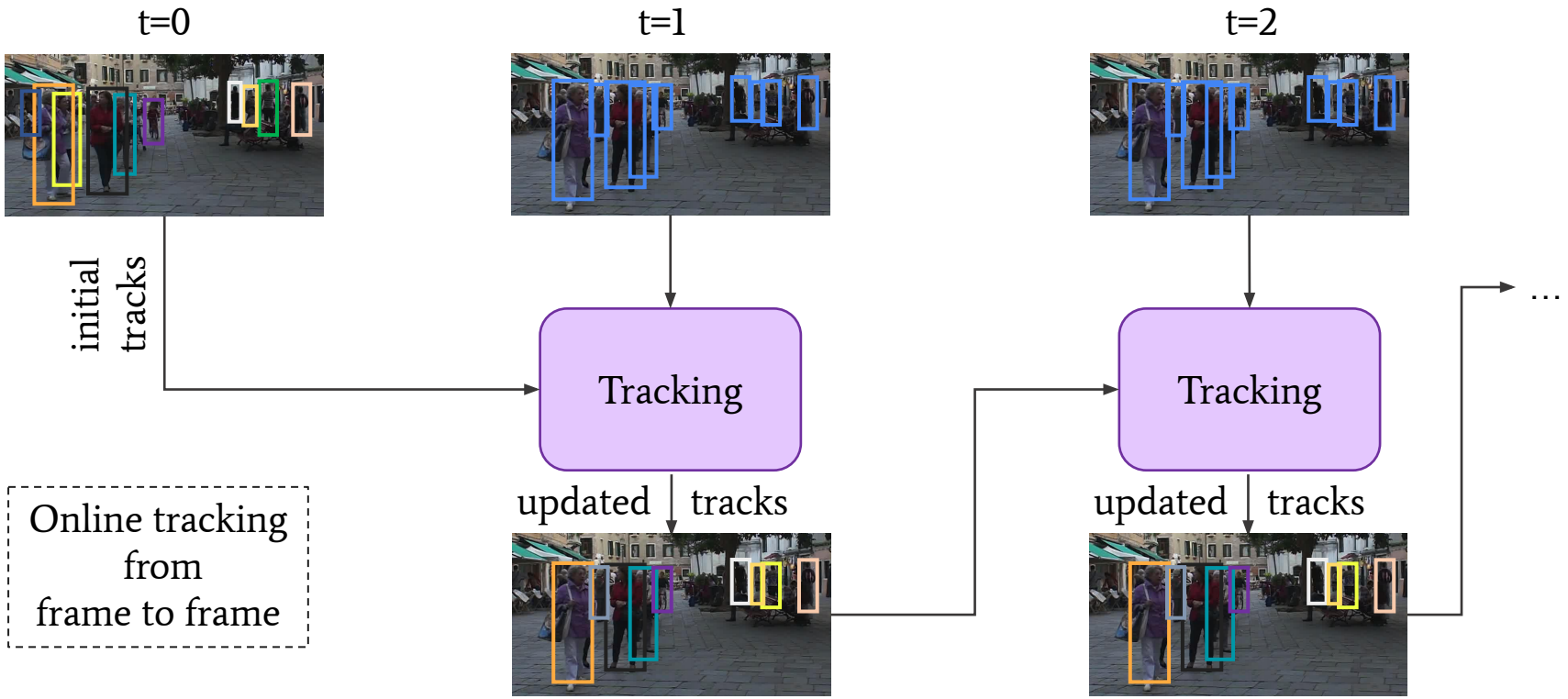


End-to-End Trainable

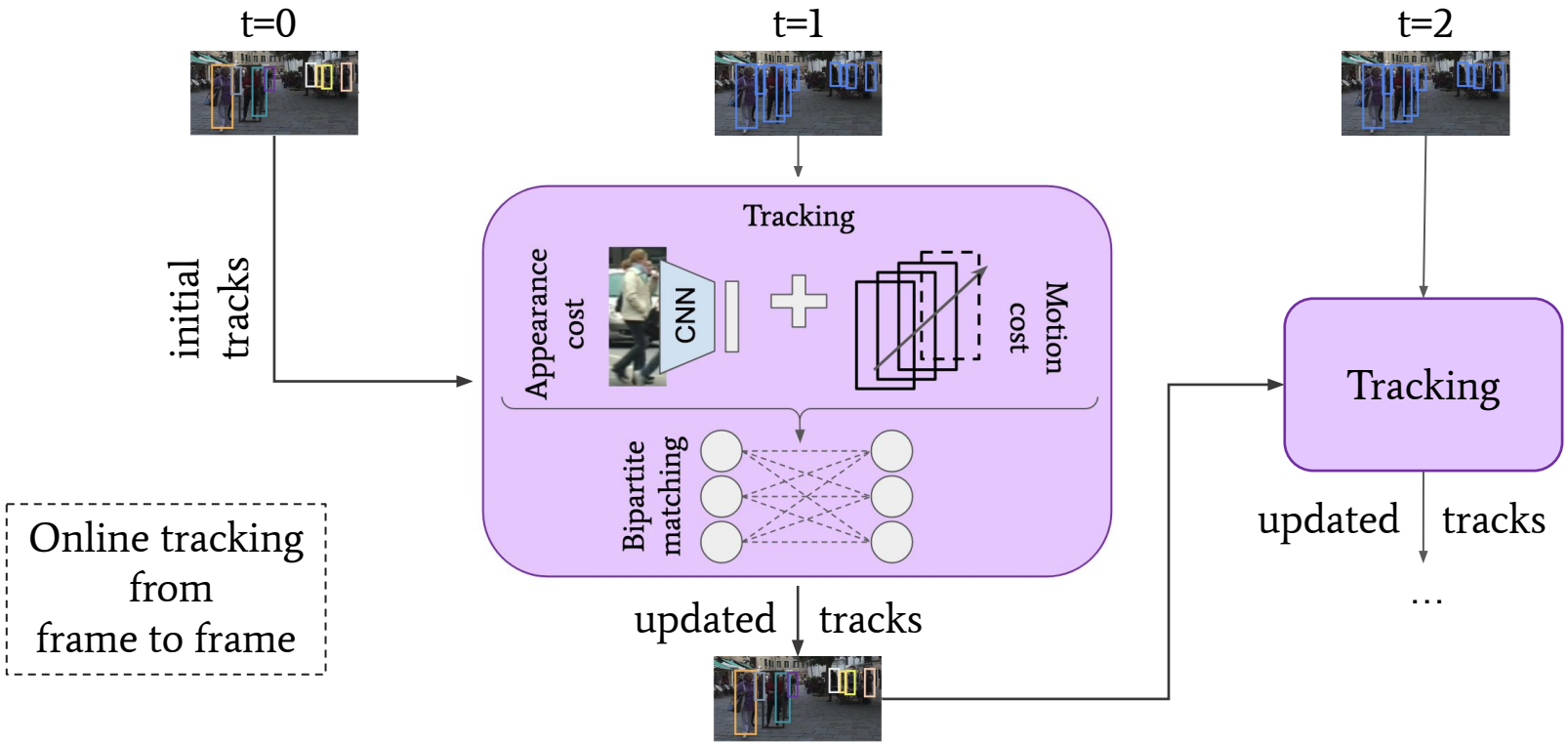
Haunting Completely End-to-End Trackers with GHOST 👻

Tracking-by-Detection is highly **general** and shows **SOTA** performance if we follow **key observations** obtained by in-depth **analysis of simple cues**

Simple Online Tracking



Simple Online Tracking



Observations Appearance Model

1. different challenges **active** and **inactive** tracks

person gets
...
occludes

track inactive

person
...
reappears

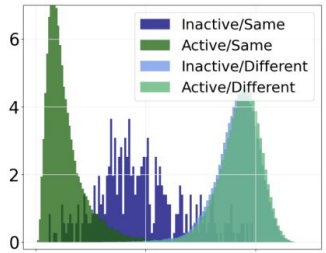
Observations Appearance Model

1. different challenges **active** and **inactive** tracks

person gets
...
occludes

track inactive

person
...
reappears



(a) Appearance

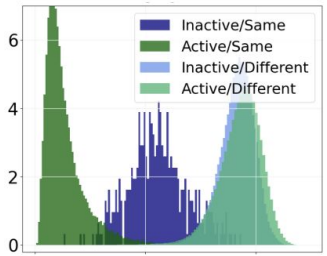
Observations Appearance Model

1. different challenges **active** and **inactive** tracks

person gets
...
occludes

track inactive

person
...
reappears



(b) Proxy Appearance

$$d_{i,k} = \frac{1}{N_k} \sum_{n=1}^{N_k} d(f_i, f_k^n)$$

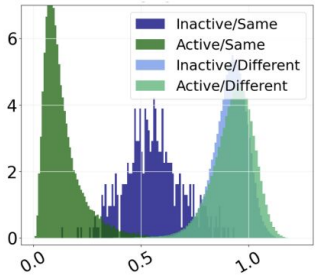
Observations Appearance Model

1. different challenges **active** and **inactive** tracks

person gets ... occludes

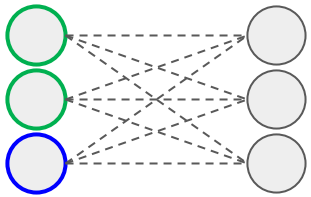
track inactive

person ... reappears

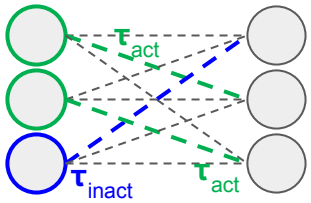


(b) Proxy Appearance

$$d_{i,k} = \frac{1}{N_k} \sum_{n=1}^{N_k} d(f_i, f_k^n)$$



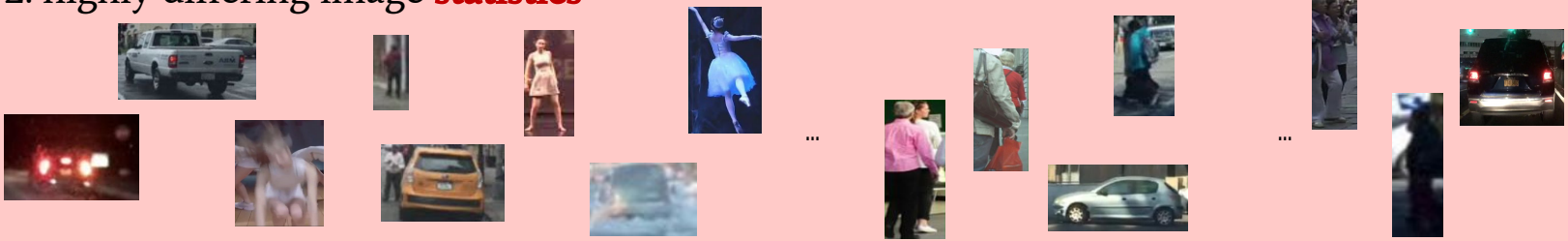
matching



thresholding

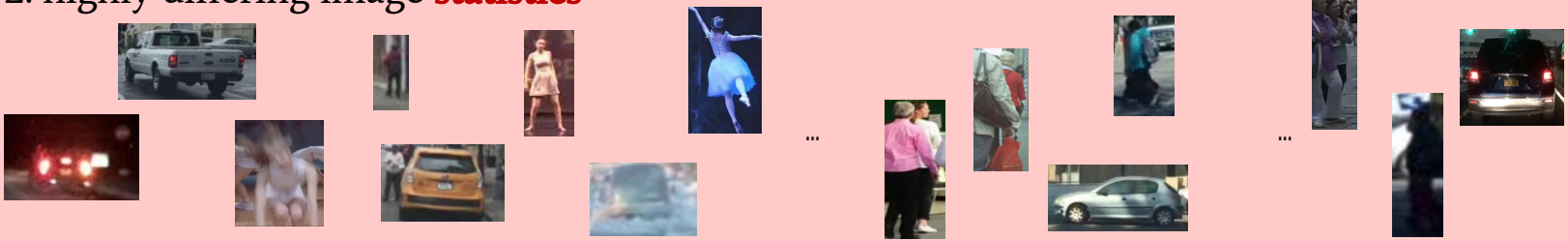
Observations Appearance Model

2. highly differing image **statistics**



Observations Appearance Model

2. highly differing image **statistics**



Frame-wise statistics

D detections

$$\mu_b = \frac{1}{D}\sum_{i=0}^{D} x_i$$

$$\sigma_b = \frac{1}{D}\sum_{i=0}^{D} (x_i - \mu_b)^2$$

Features adapted to specific frame

$$\hat{x_i} = \gamma \frac{x_i - \mu_b}{\sqrt{\sigma_b + \epsilon}} + \beta$$

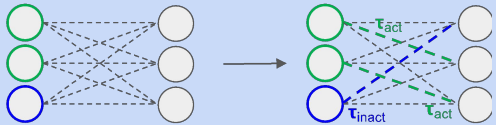
Observations Appearance Model

1. different challenges **active** and **inactive** tracks

proxy distance

$$d_{i,k} = \frac{1}{N_k}\sum_{n=0}^{N_k} d(f_i, f_k^n)$$

thresholding



Spiced up appearance model better suited for MOT

			MOT 17			BDD		
diff τ	IP	DA	HOTA ↑	IDF1 ↑	MOTA ↑	HOTA ↑	IDF1 ↑	MOTA ↑
			61.6	72.2	69.5	41.3	47.7	42.0
✓			61.8	72.7	69.6	41.9	48.6	41.9
✓	✓		62.4	73.6	69.6	42.9	50.4	43.7
✓	✓	✓	63.3	75.3	69.6	43.7	51.5	43.9

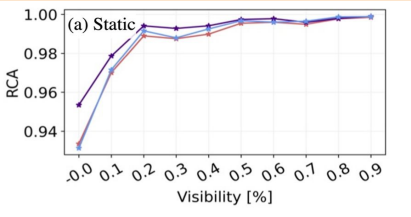
2. highly differing image **statistics**

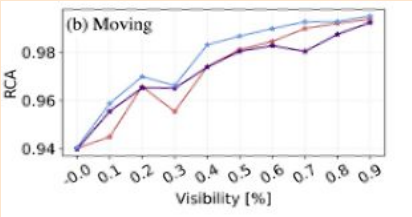
Features adapted to specific frame

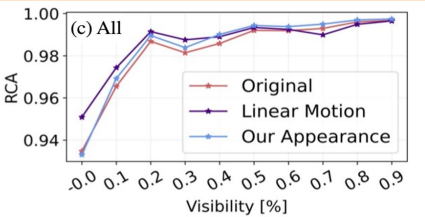
$$\hat{x}_i = \gamma\frac{x_i - \mu_b}{\sqrt{\sigma_b + \epsilon}} + \beta$$

Observations Motion Model

3. **Linear motion** model complements appearance well

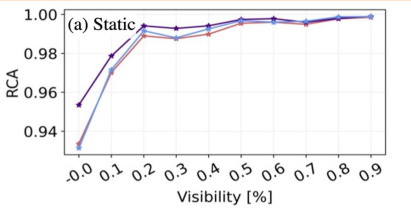


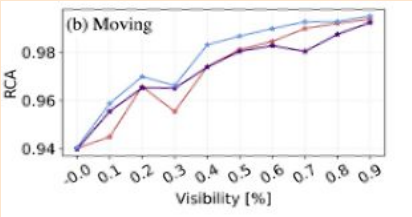


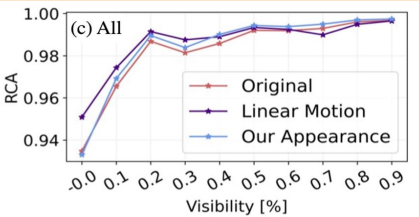


Observations Motion Model

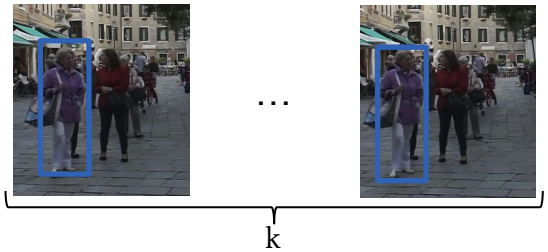
3. **Linear motion** model complements appearance well



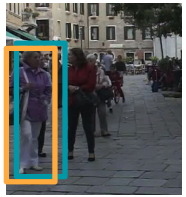




Motion still struggles with moving camera → adaptive frames k for velocity computation



$$v_j = \frac{1}{k}\sum_{i=0}^{k}\frac{\Delta p}{\Delta}$$



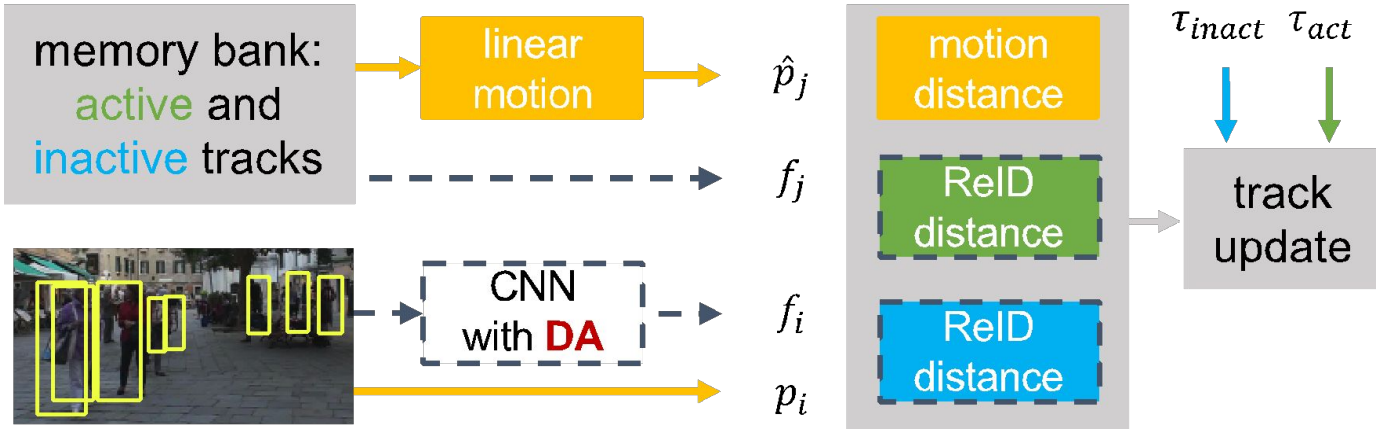
p predicted

d detection

$$d_m = IoU(p, d)$$

Good Old Hungarian Simple Tracker or GHOST* 👻

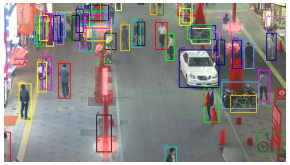
*the order of the letters of the acronym does not change the product



SOTA on highly differing datasets w/o training on tracking data!

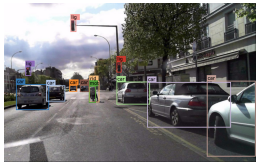
MOT17

	HOTA ↑	IDF1 ↑	MOTA ↑	IDSW ↓
Private MOT17				
CenterTrack [16]	52.2	64.7	67.8	3039
TraDeS [12]	52.7	63.9	69.1	3555
QDTrack [7]	53.9	66.3	68.7	3378
FairMOT [15]	59.3	72.3	73.7	3303
MeMOT [3]	56.9	72.5	69.0	2724
GTR [17]	59.1	71.5	75.3	2859
MOTR [13]	57.8	68.6	73.4	2439
ByteTrack* [14]	**62.8**	**77.1**	**78.9**	2363
ByteTrack* [14]	63.1	77.3	80.3	2196
GHOST	**62.8**	**77.1**	78.7	**2325**



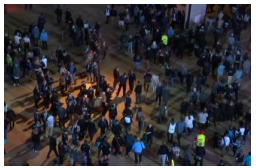
BDD100k

	mHOTA ↑	mIDF1 ↑	mMOTA ↑	HOTA ↑	IDF1 ↑	MOTA ↑
validation						
Yu et. al. [69]	-	44.5	25.9	-	66.8	56.9
ByteTrack [73]	45.4	54.6	**45.2**	61.6	70.2	**68.7**
QDTrack [41]	41.7	51.5	36.3	60.9	**71.4**	63.7
MOTR [71]	-	43.5	32.0	-	-	-
TETer [29]	-	53.3	39.1	-	-	-
GHOST	**45.7**	**55.6**	44.9	**61.7**	70.9	68.1
test						
Yu et. al. [69]	-	44.7	26.3	-	68.2	58.3
ByteTrack [73]	-	55.8	**40.1**	-	71.3	**69.6**
QDTrack [41]	-	52.3	35.5	-	**72.3**	64.3
GHOST	**46.8**	**57.0**	39.5	**62.2**	72.0	68.9



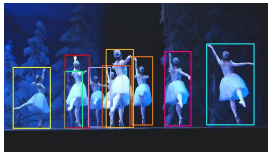
MOT20

	HOTA ↑	IDF1 ↑	MOTA ↑	IDSW ↓
Private MOT20				
GSDT [10]	53.6	67.5	67.1	3230
FairMOT [15]	54.6	67.3	61.8	5243
MeMOT [3]	54.1	66.1	63.7	1938
MTrack [13]	-	69.2	63.5	6031
ByteTrack* [14]	60.4	74.5	**74.2**	**925**
ByteTrack* [14]	61.3	75.2	77.8	1223
GHOST	**61.2**	**75.2**	73.7	1264



DanceTrack

	HOTA ↑	IDF1 ↑	MOTA ↑	DetA ↑	AssA ↑
CenterTrack [79]	41.8	35.7	86.8	78.1	22.6
FairMOT [74]	39.7	40.8	82.2	66.7	23.8
QDTrack [41]	54.2	50.4	87.7	80.1	36.8
TraDeS [64]	43.3	41.2	86.2	74.5	25.4
MOTR [71]	54.2	51.5	79.7	73.5	**40.2**
GTR [80]	48.0	50.3	84.7	72.5	31.9
ByteTrack [73]	47.7	53.9	89.6	71.0	32.1
GHOST	**56.7**	**57.7**	**91.3**	**81.1**	39.8



SOTA on highly differing datasets w/o training on tracking data! 👻