

CVPR 2020 Workshop on Autonomous Driving 1st Place Solution for Challenge 2: BDD 100K Tracking

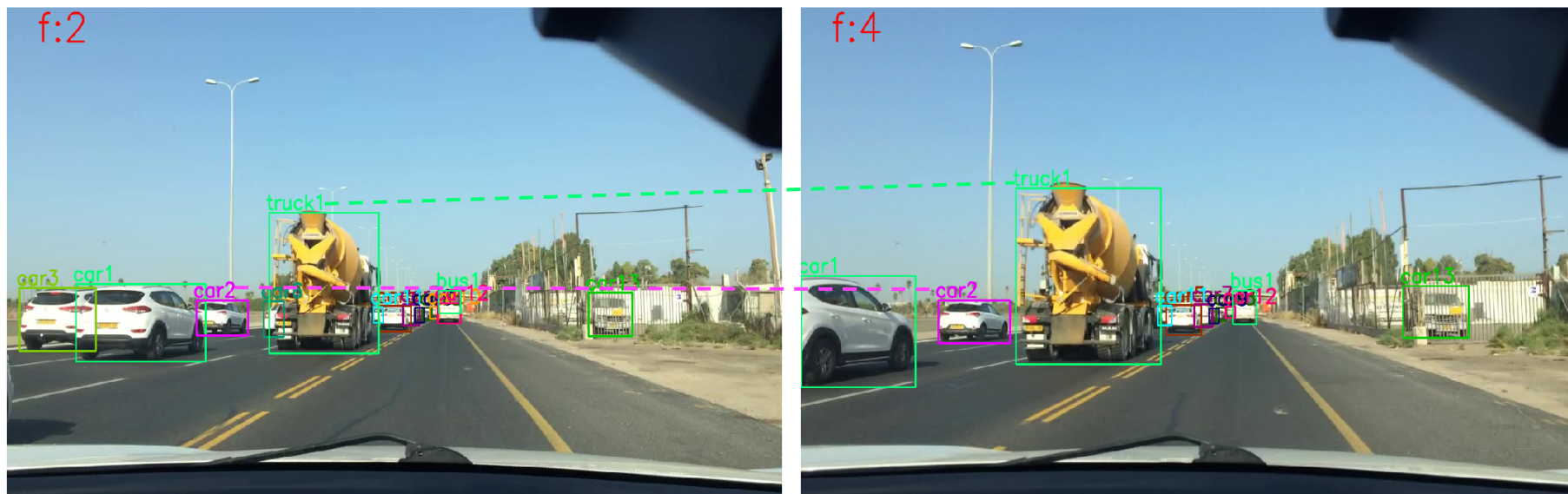
Fan Yang^{1,2}, Xin Chang¹, Sakriani Sakti^{1,2}, Satoshi Nakamura^{1,2}, Yang Wu³

¹Nara Institute of Science and Technology, Japan

²RIKEN Center for Advanced Intelligence Project, Japan

³Kyoto University, Japan

- Problem: detect and track multiple objects in videos, there are 8 categories of objects, as Pedestrian, Rider, Car, Bus, Truck, Train, Motorcycle, Bicycle.
- Input: a video sequence contain that multiple RGB images.
- Output: 2D bounding boxes and corresponding track ID at each frame.

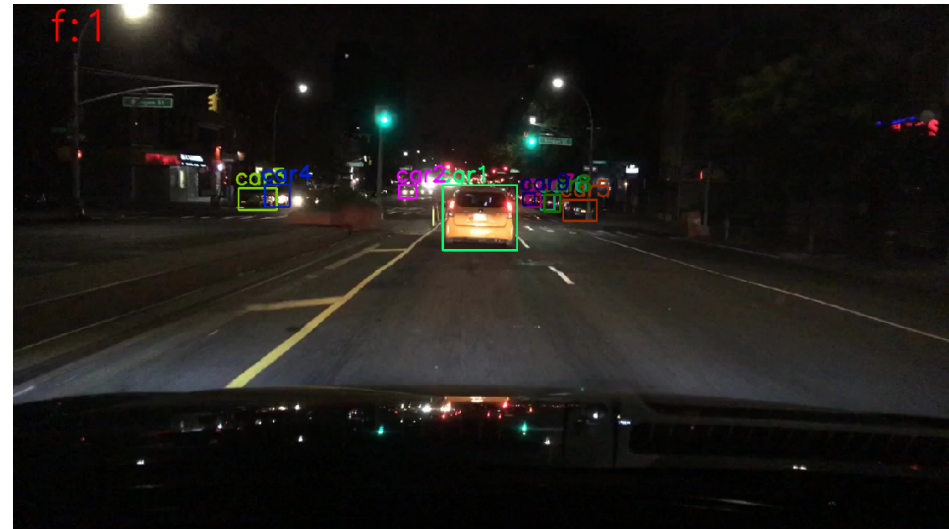


1. Classification ambiguity



pedestrian
or
rider?

2. Weak Illuminations at night scenes



Detector:

Faster_rcnn_X_101_32x8d_FPN of Detectron 2(<https://github.com/facebookresearch/detectron2>)

Training Dataset:

BDD 100K object detection training set, labeled key frame images extracted from the videos at 10th second.

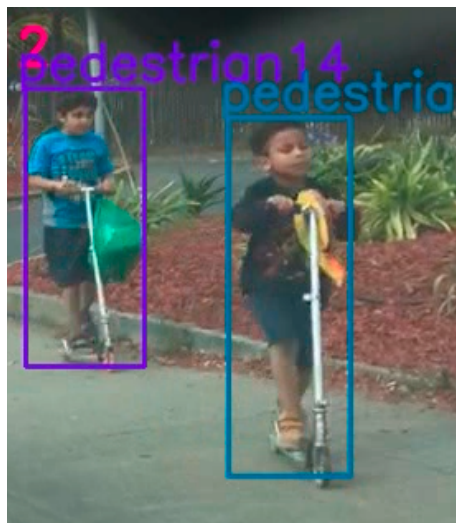
Inference Setting:

Confidence threshold: 0.6

NMS IoU threshold: 0.75

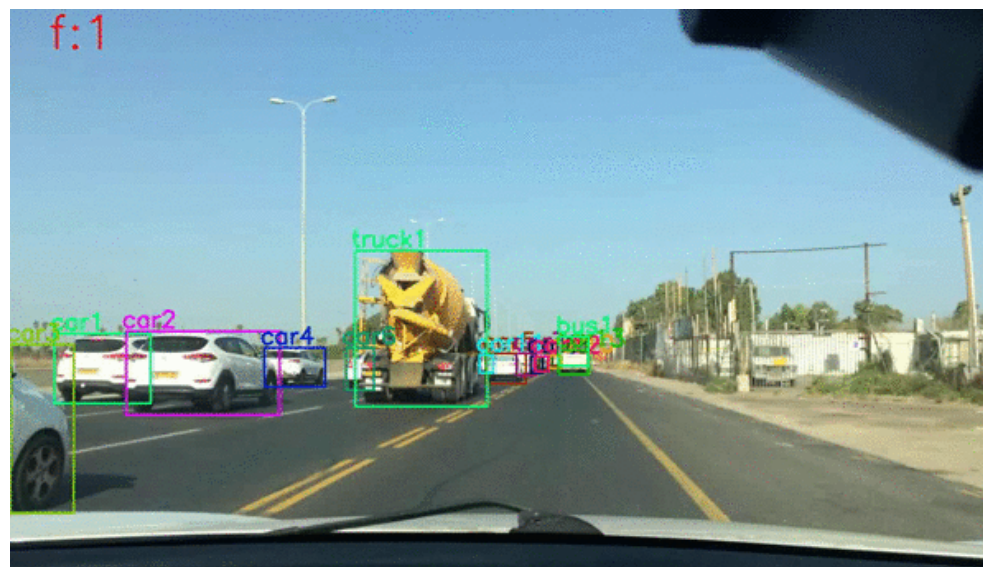
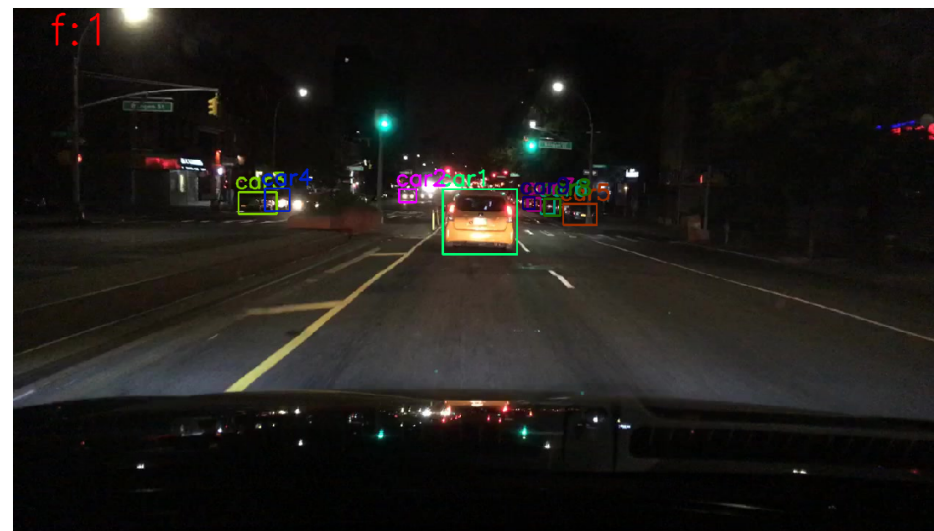
Crop patch size: resize to be 128 x 128 pixels

1. Classification ambiguity



pedestrian
or
rider?

2. Weak Illuminations at night scenes

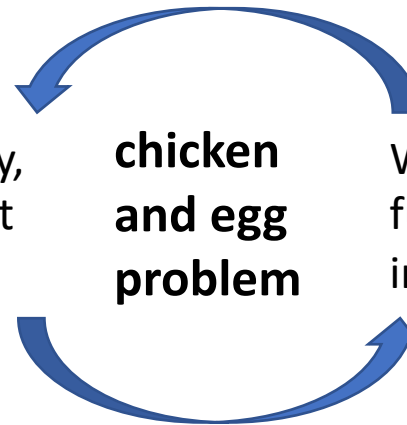


3. Fast motion

Our focus, solving this issue by proposed a robust trajectory initialization method.

To model the trajectory by Kalman Filter or others

Without correct trajectory, cannot predict the correct future position.



Without the correctly predicted future position, cannot correctly initialize the trajectory.

Existing solutions:

Initialization by copy-paste boxes to calculate center distance or IoU (e.g., DeepSORT [1])
Defects: Fast moving objects do not have overlapped positions across frames.

Initialization trajectory by appearance features (e.g., MOTbyReid [2])
Defects: Object may have the similar appearance features, especially for cars at night scenes.

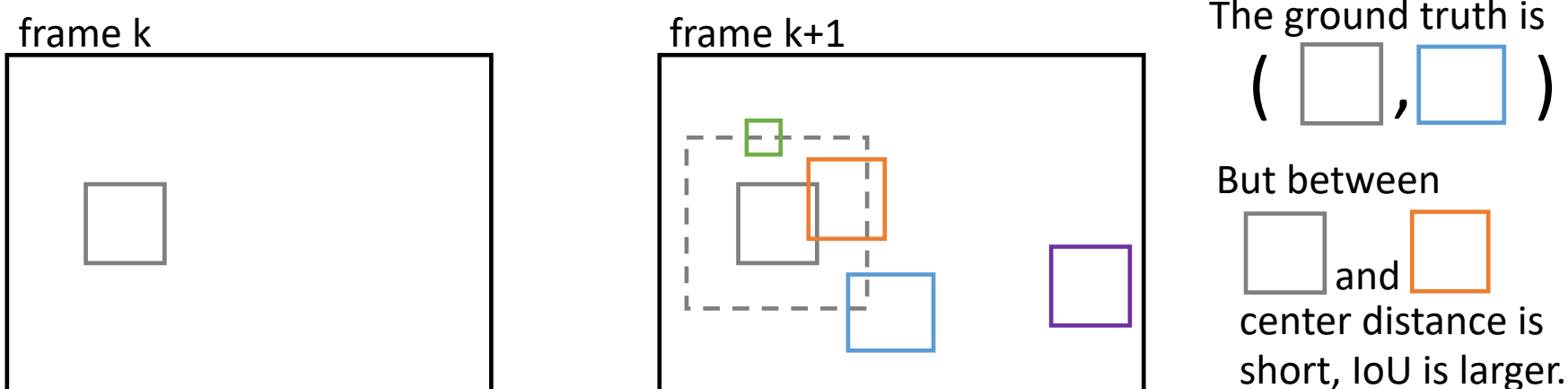
Our solution:


Initialization trajectory by hard fusing appearance features and IoU of expanded boxes.

[1] Wojke et al. "Simple Online and Realtime Tracking with a Deep Association Metric" (ICIP17)

[2] Tang et al. "Multiple People Tracking by Lifted Multicut and Person Re-identification" (CVPR17)


Initializing trajectory by hard fusing appearance features and IoU of expanded boxes.






It is the first time to observe  at frame k, we cannot predict its position at frame k+1.

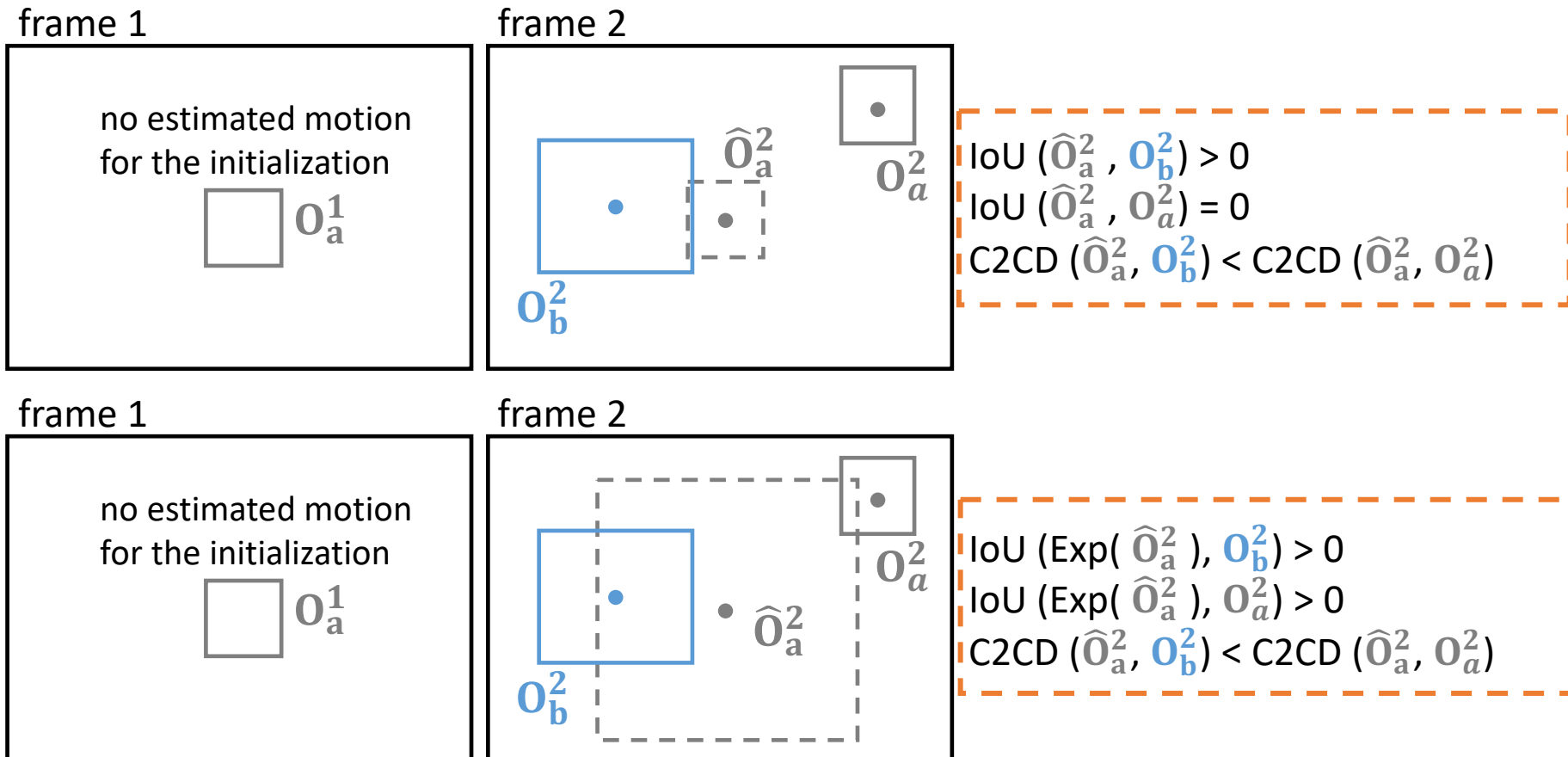
We expand  to be  and check its IoU with observations at frame k+1.

 is excluded due to the IoU ( , )=0.

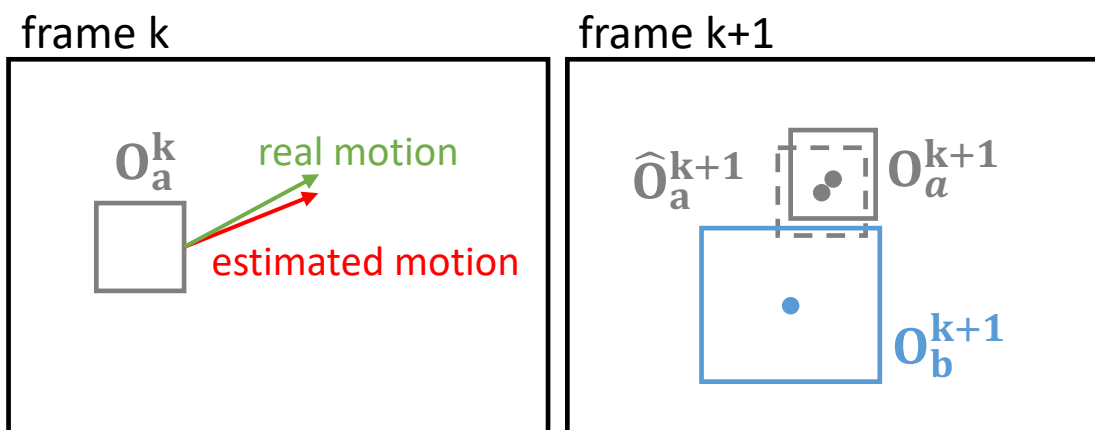
 is excluded due to both of the height and width has reduced 2 times (approx. 100 m).

We compare the appearance similarity from  to  and  only to decide the association by linear assignment.

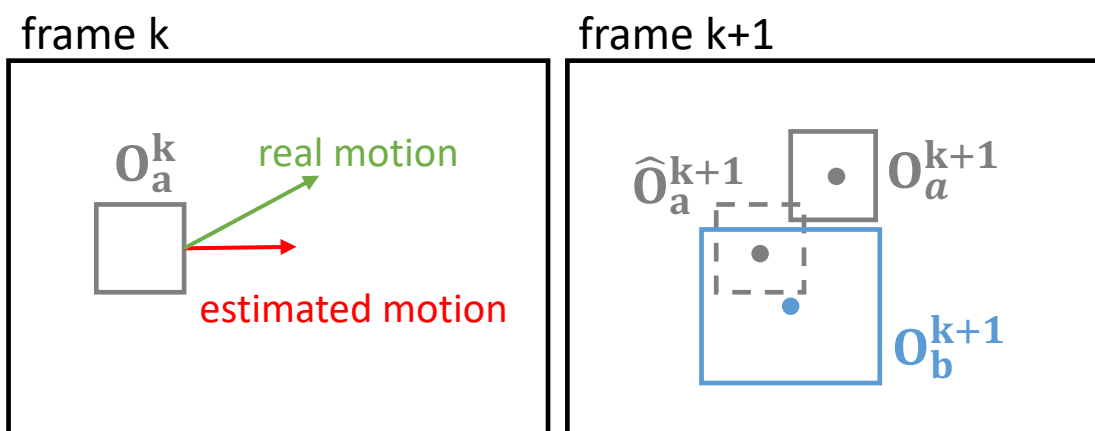
It is challenging to estimate the accurate motion.
High IoU does not equal the high correlation



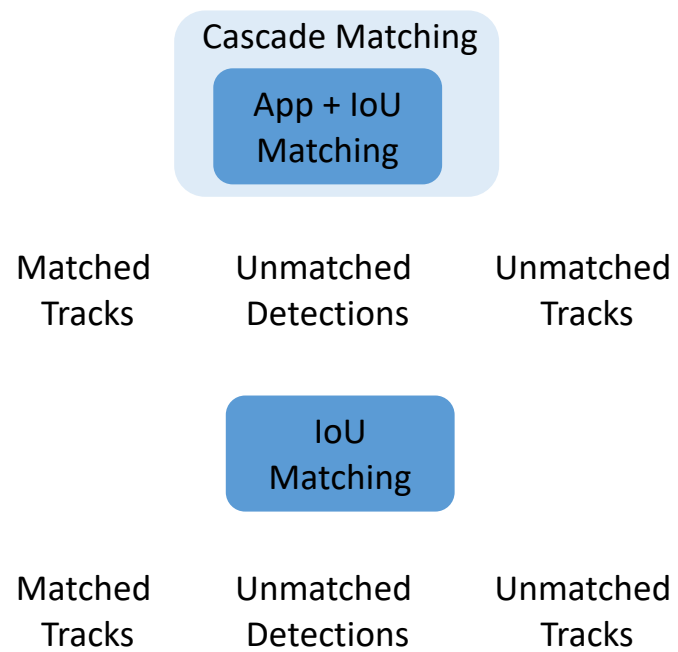
It is challenging to estimate the accurate motion.
High IoU does not equal the high correlation



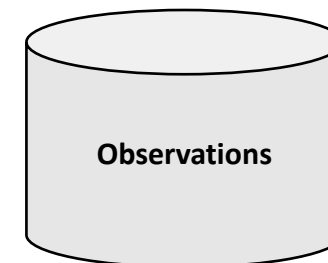
$$\begin{aligned} \text{IoU}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_b^{k+1}) &< \text{IoU}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_a^{k+1}) \\ \text{C2CD}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_b^{k+1}) &< \text{C2CD}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_a^{k+1}) \\ \text{SoR}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_b^{k+1}) &< \text{SoR}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_a^{k+1}) \end{aligned}$$

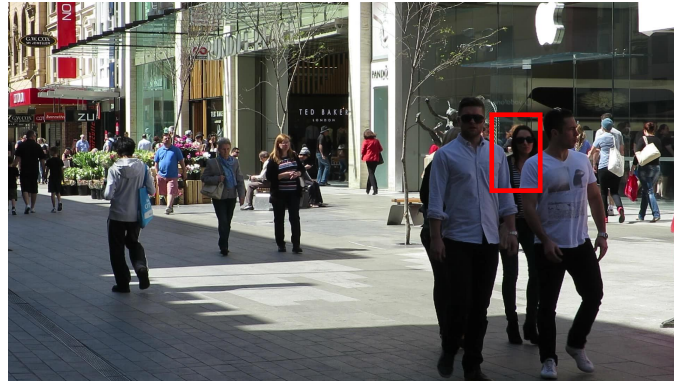


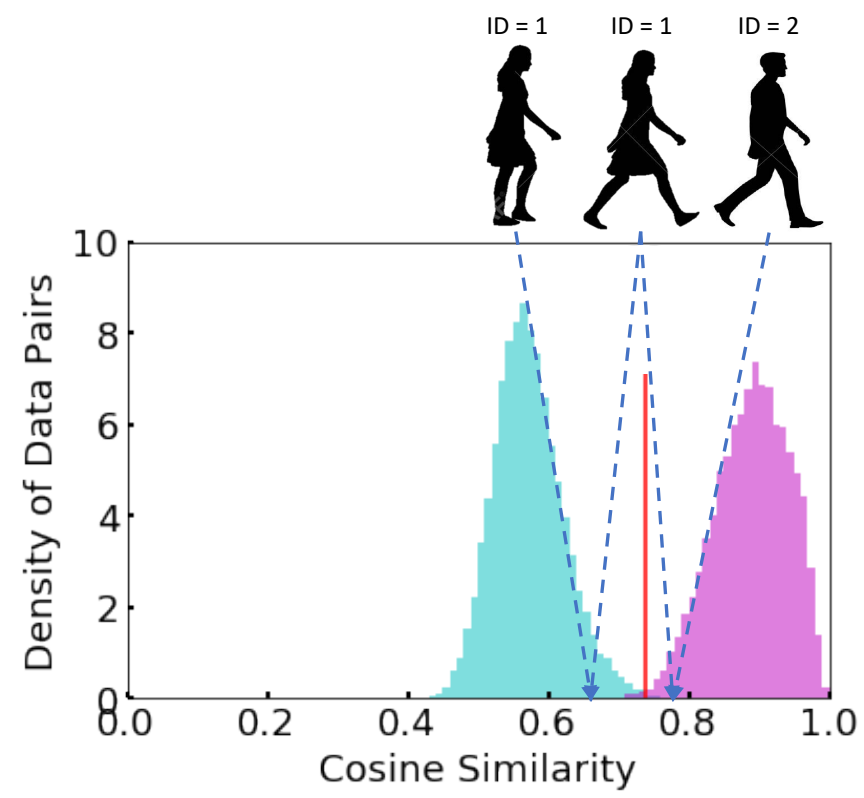
$$\begin{aligned} \text{IoU}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_b^{k+1}) &> \text{IoU}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_a^{k+1}) \\ \text{C2CD}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_b^{k+1}) &> \text{C2CD}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_a^{k+1}) \\ \text{SoR}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_b^{k+1}) &< \text{SoR}(\hat{\mathbf{O}}_a^{k+1}, \mathbf{O}_a^{k+1}) \end{aligned}$$

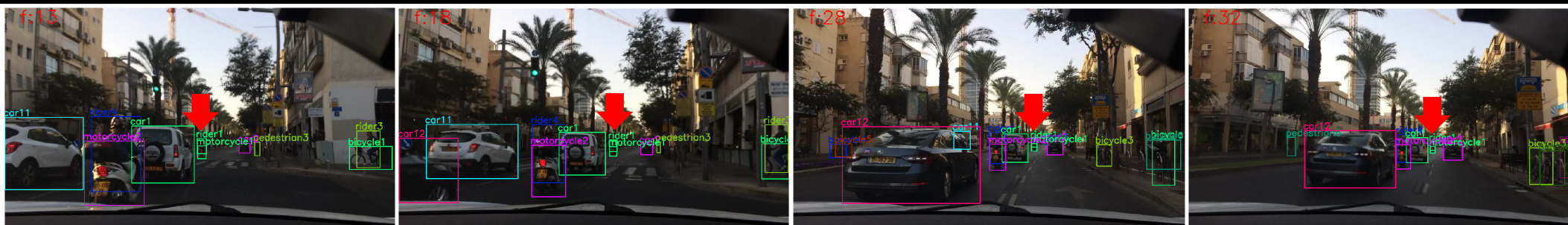


Cascade
Matching









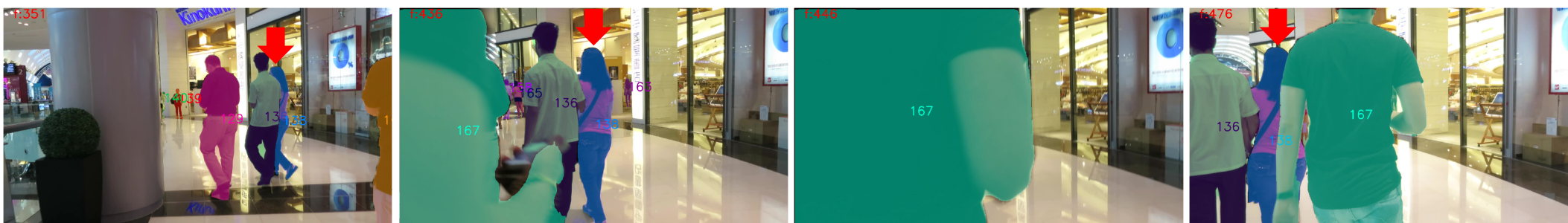
MOTS20-01 (static camera)



MOTS20-06 (stroller-mounted camera)

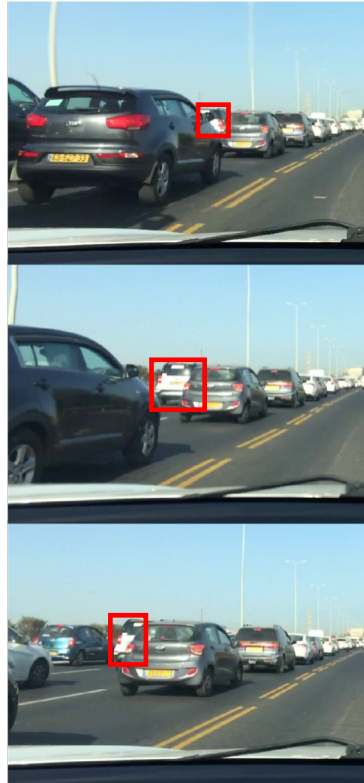


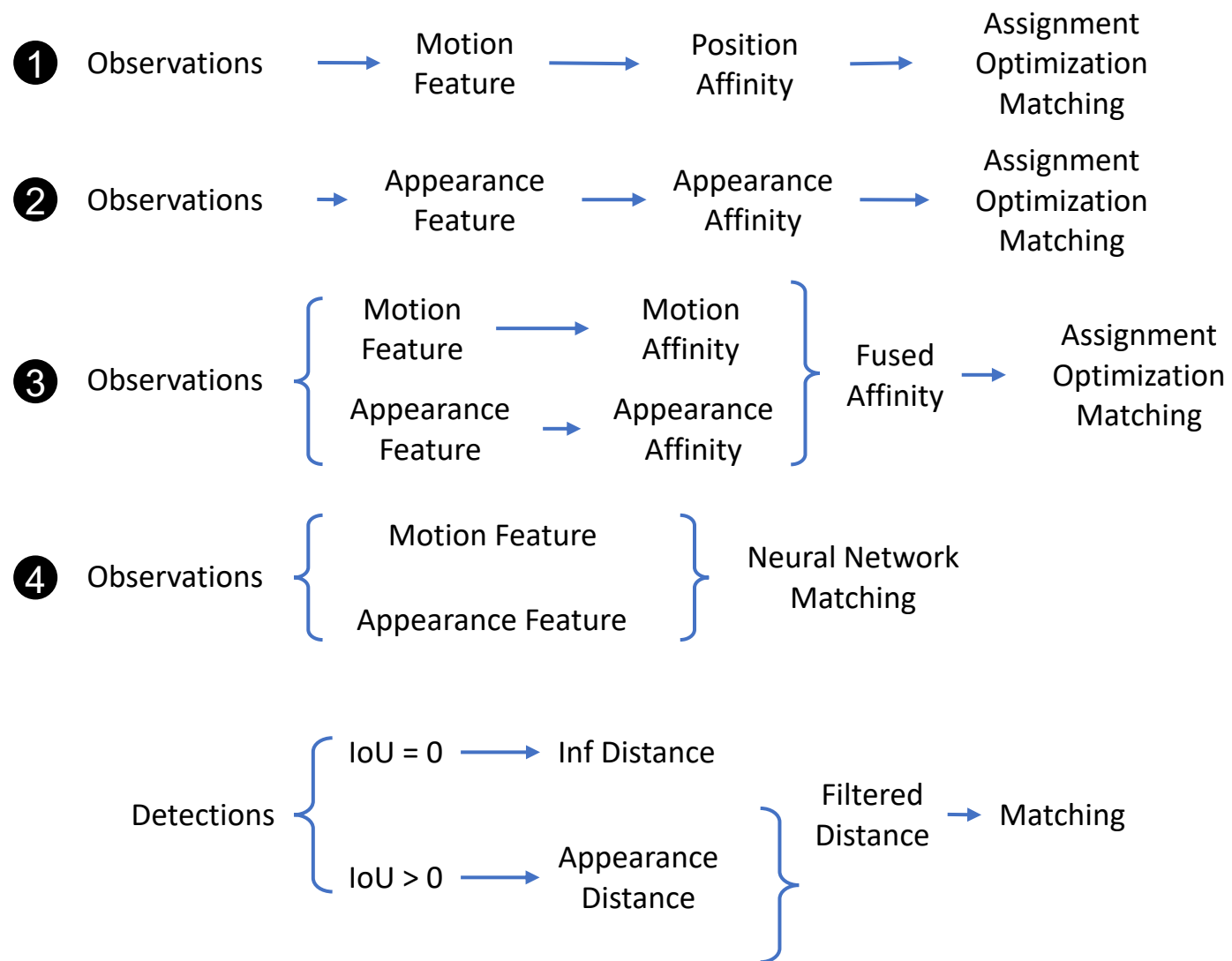
MOTS20-12 (stroller-mounted camera)

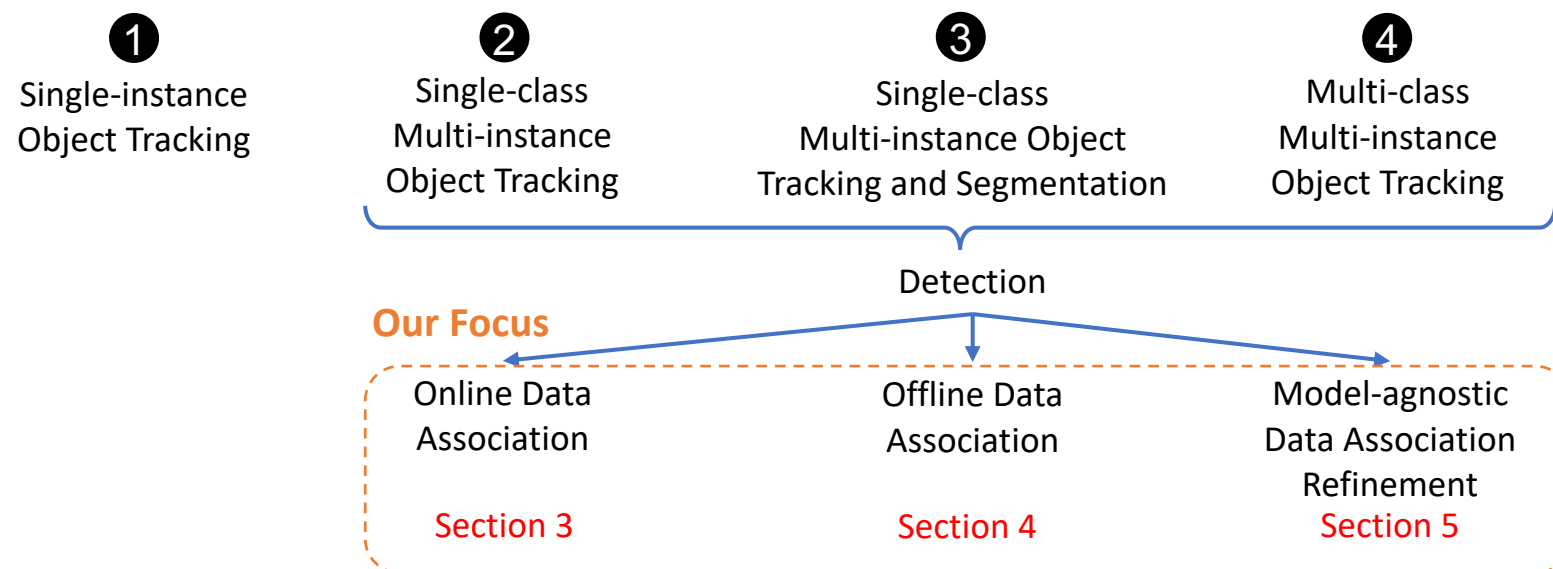
**KITTI-MOTS-0003 (car-mounted camera, turning scene)**

KITTI-MOTS-0018 (car-mounted camera, pedestrian-car scene)

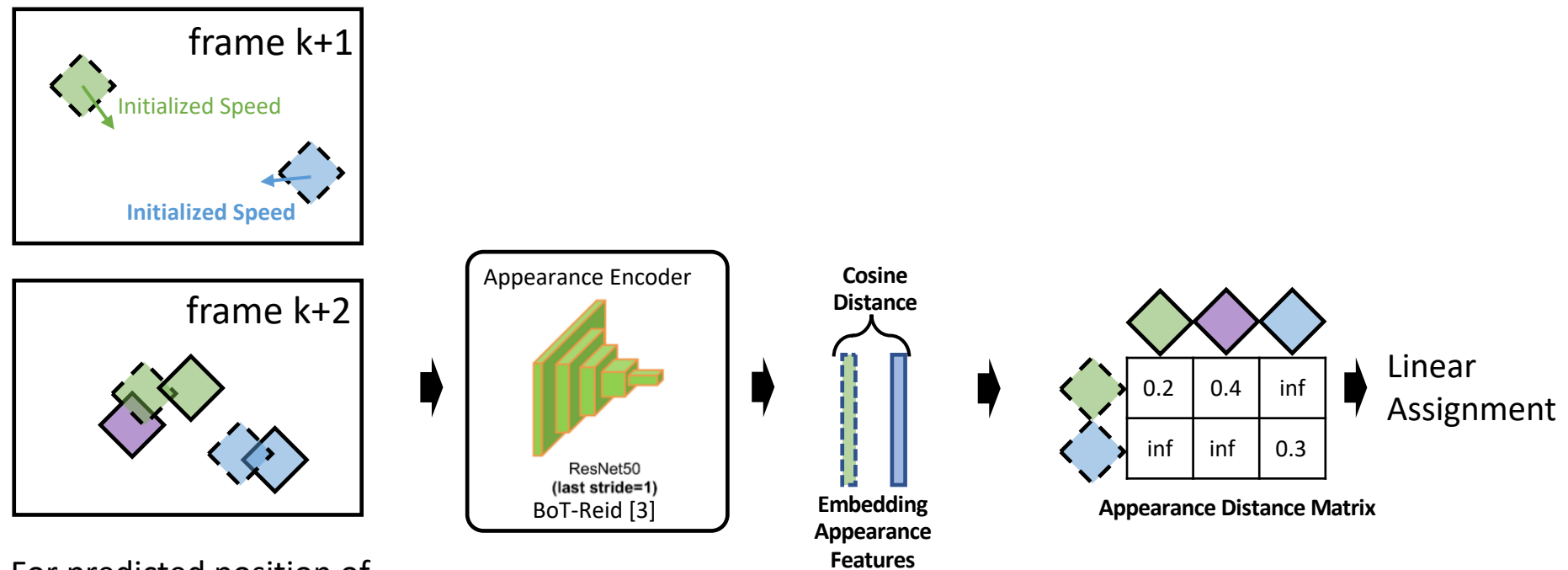








Our strategy to fuse appearance feature and trajectory:
Always decide the association by appearance, but constrained by box trajectory.



For predicted position of frame k, consider all of IoU > 0 observations of frame k+1 for matching

[3] Luo et al. "Bag of Tricks and a Strong Baseline for Deep Person Re-Identification" (TMM20)

Overall													
#	User	Entries	Date of Last Entry	mMOTA ▲	mMOTP ▲	MOTA ▲	MOTP ▲	Misses ▲	FP ▲	Switch ▲	Mostly Tracked ▲	Mostly Lost ▲	Partially Tracked ▲
1	madamada	11	06/13/20	33.63 (1)	81.06 (2)	59.76 (1)	82.78 (3)	209339.00 (1)	76612.00 (2)	42901.00 (5)	16774.00 (1)	5004.00 (1)	10353.00 (1)
2	DeepBlueAI	7	06/12/20	31.64 (2)	82.38 (1)	56.85 (3)	84.76 (1)	292063.00 (3)	35401.00 (1)	25186.00 (4)	10296.00 (3)	12266.00 (3)	9569.00 (5)
3	bdd100k	1	05/21/20	26.40 (3)	78.92 (3)	58.27 (2)	82.90 (2)	224083.00 (2)	100868.00 (4)	16047.00 (1)	15739.00 (2)	6506.00 (2)	9886.00 (3)
Super-category: Person Challenges in correctly detect objects													
#	User	Entries	Date of Last Entry	MOTA ▲	MOTP ▲	Misses ▲	FP ▲	Switch ▲	Mostly Tracked ▲	Mostly Lost ▲	Partially Tracked ▲		
1	madamada	11	06/13/20	44.59 (1)	77.67 (2)	44811.00 (1)	9180.00 (2)	7912.00 (5)	1922.00 (1)	1344.00 (1)	2608.00 (1)		
Super-category: Vehicle													
#	User	Entries	Date of Last Entry	MOTA ▲	MOTP ▲	Misses ▲	FP ▲	Switch ▲	Mostly Tracked ▲	Mostly Lost ▲	Partially Tracked ▲		
1	madamada	11	06/13/20	67.18 (1)	83.37 (3)	140039.00 (1)	49082.00 (2)	38558.00 (5)	15191.00 (1)	2785.00 (1)	7556.00 (4)		
Super-category: Bike													
#	User	Entries	Date of Last Entry	MOTA ▲	MOTP ▲	Misses ▲	FP ▲	Switch ▲	Mostly Tracked ▲	Mostly Lost ▲	Partially Tracked ▲		
1	madamada	11	06/13/20	29.77 (1)	76.87 (2)	7108.00 (1)	969.00 (3)	277.00 (5)	123.00 (1)	278.00 (1)	288.00 (1)		

- Good trajectory initialization are important for MOT when objects are moving fast.
- We proposed an robust trajectory initialization approach by hard fusing appearance features and IoU of expanded boxes.

Thanks for your listening