

Sideview models

Annotations



COCO



After checking base



After bbox extension



Core

Annotations



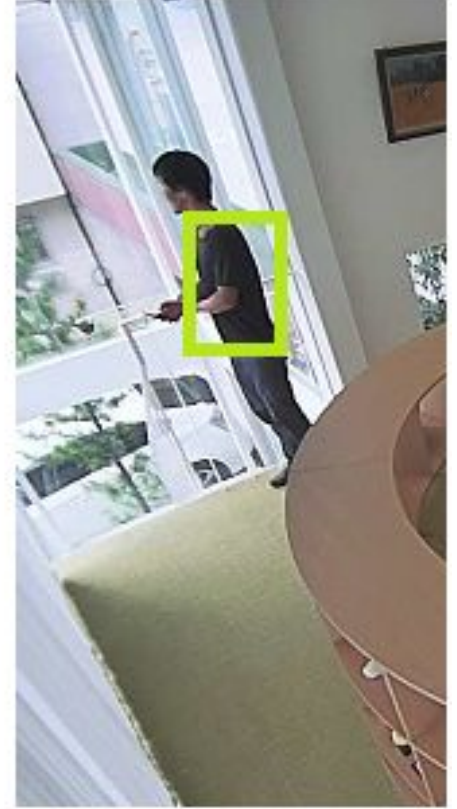
COCO



After checking base



After bbox extension



Core

Fullbody vs Core

FB Training Base	N images	N obs	background
9	~16000		2000
18	4175	11208	0
19	5892	15086	7917
20	5118	13054	7917
21	6619	17052	7917

Core Training Base	N images	N obs	background
18	4186	12456	0
19	5864	16240	7917
20	5092	13758	7917
21	6558	18281	7917

Fullbody vs Core

Validation Base	N images	N obs	background
Fullbody	862	2016	0
Core	858	2034	0

Compare models	F1	Precision	Recall	TP	FP	FN	n_steps
3.9.1	0.68	0.78	0.6	1216	349	800	49345
3.18.1_fb	0.67	0.74	0.61	1221	424	795	18129
3.19.1_fb	0.65	0.73	0.59	1191	430	825	31130
3.20.1_fb	0.64	0.75	0.56	1121	375	895	84329
3.20.2_fb	0.66	0.73	0.6	1200	439	816	124308
2.20.1_fb	0.65	0.79	0.55	1117	292	899	77232
2.20.2_fb	0.64	0.79	0.53	1072	288	944	45741
3.21.1_fb	0.66	0.77	0.58	1172	348	844	44626
3.18.1_core	0.55	0.55	0.55	1115	907	919	24232
3.18.2_core	0.58	0.64	0.52	1065	589	969	66018
3.19.1_core	0.59	0.66	0.53	1085	554	949	59863
3.20.1_core	0.55	0.66	0.47	954	483	1080	20862
3.20.2_core	0.55	0.64	0.48	986	554	1048	56397
2.20.1_core	0.55	0.64	0.48	971	556	1063	16867
2.20.2_core	0.55	0.6	0.5	1027	685	1007	81157
3.21.1_core	0.57	0.56	0.58	1180	932	854	63624
3.21.2_core	0.61	0.69	0.54	1108	489	926	31956 ₆

Evaluation metrics

Multiple Object Tracking Accuracy

$$\text{MOTA} = 1 - \frac{\sum_t (\text{FN}_t + \text{FP}_t + \text{IDSW}_t)}{\sum_t \text{GT}_t}, \quad (1)$$

where t is the frame index and GT is the number of ground truth objects. Note that MOTA can also be negative in cases where the number of errors made by the tracker exceeds the number of all objects in the scene.

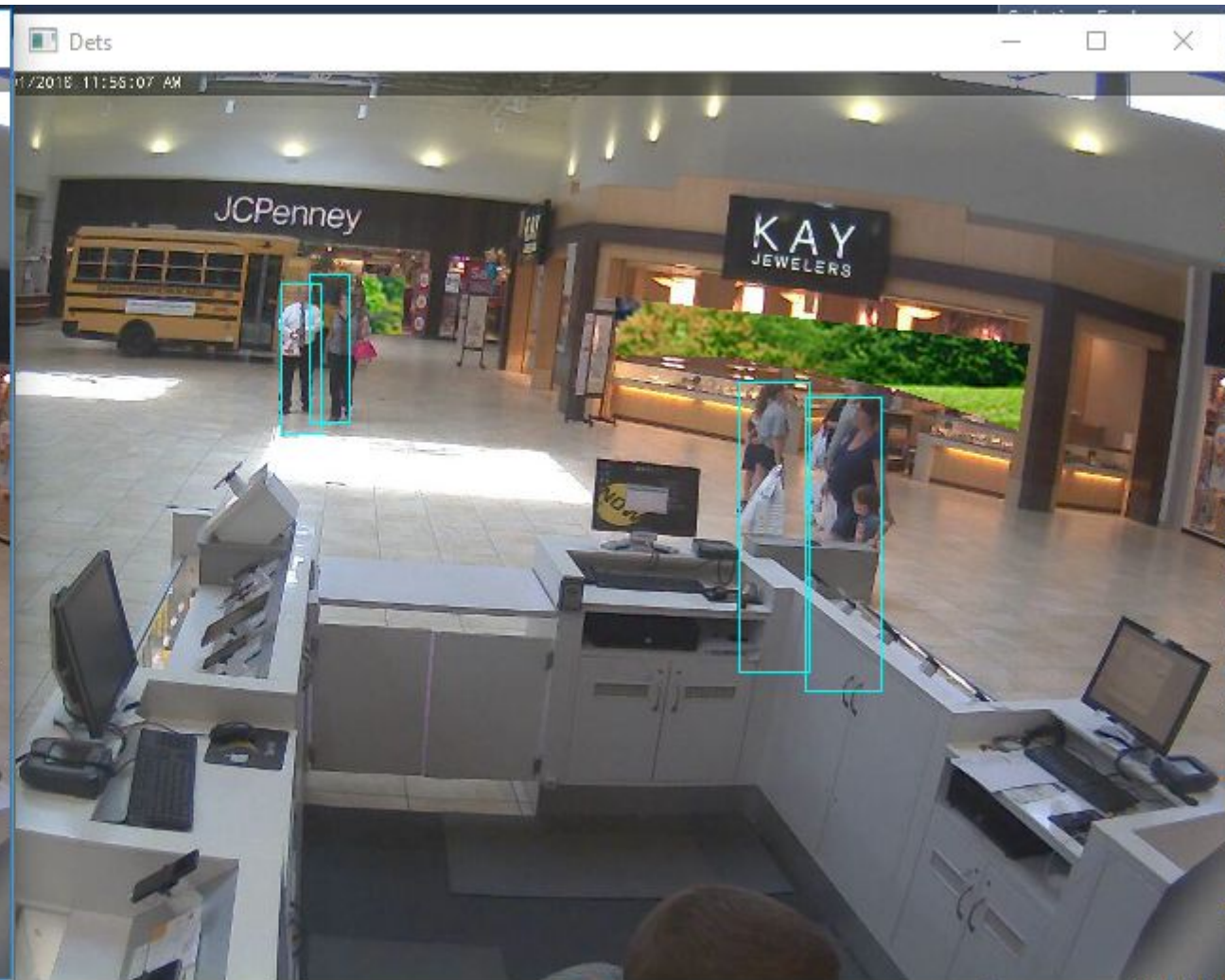
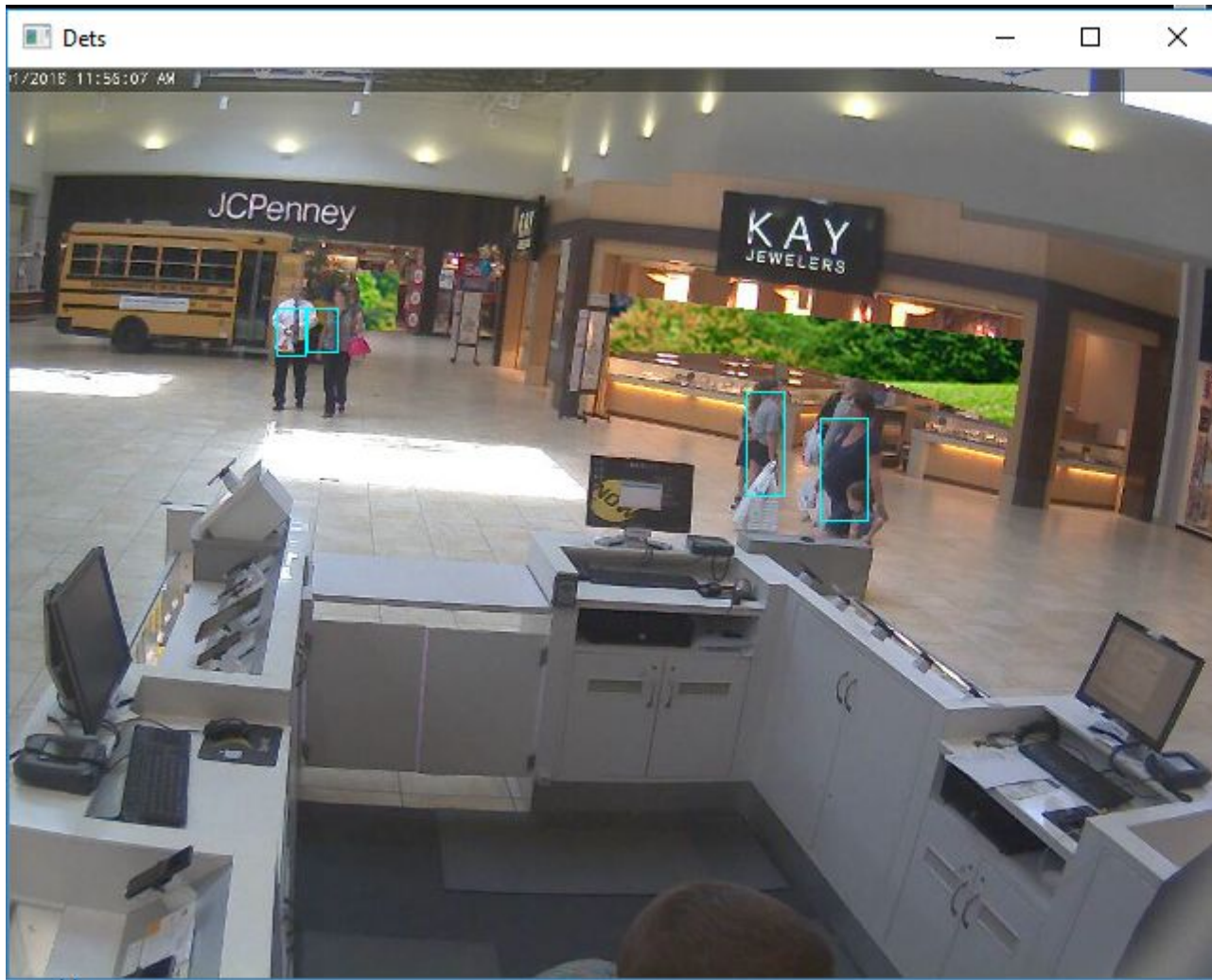
Multiple Object Tracking Precision

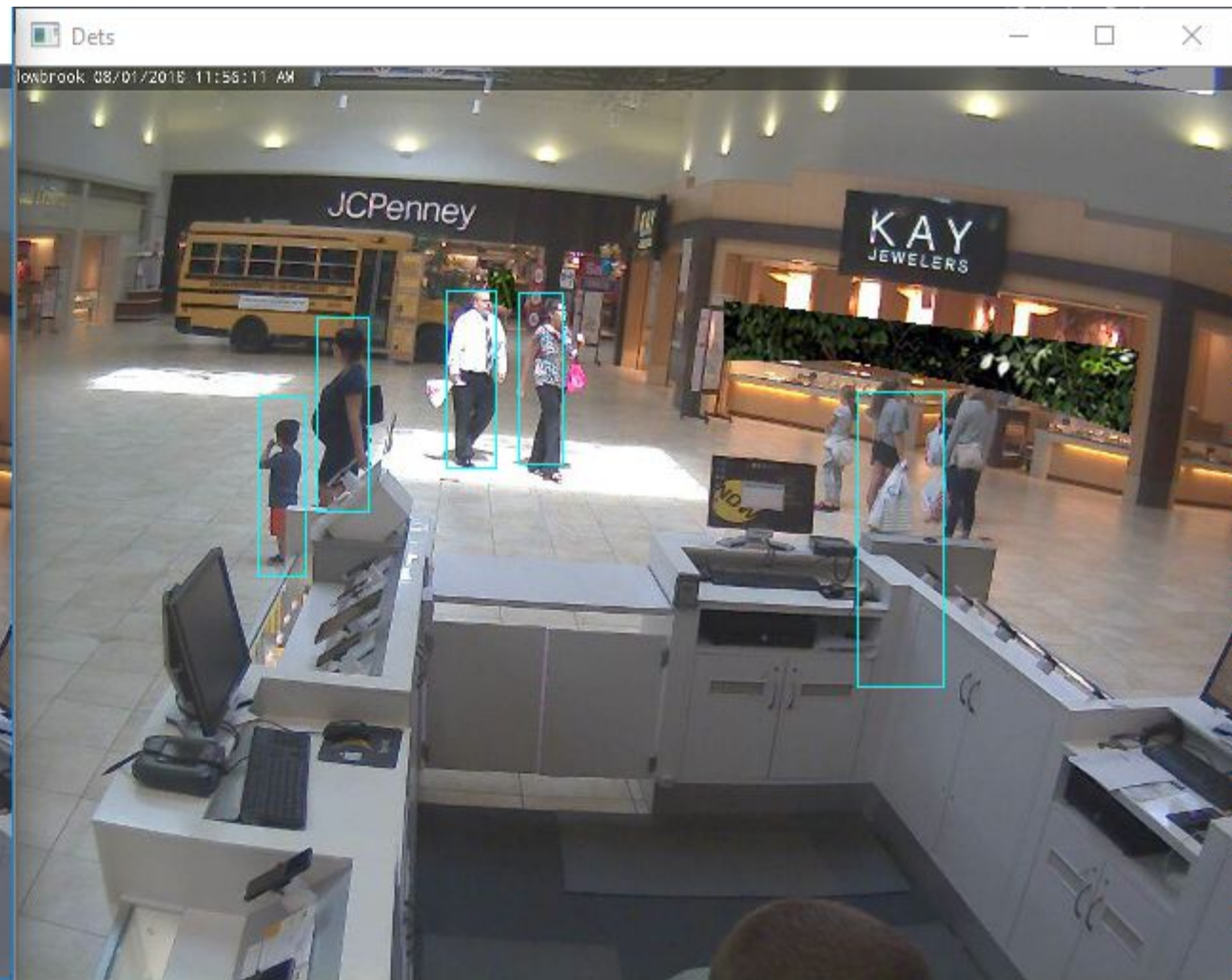
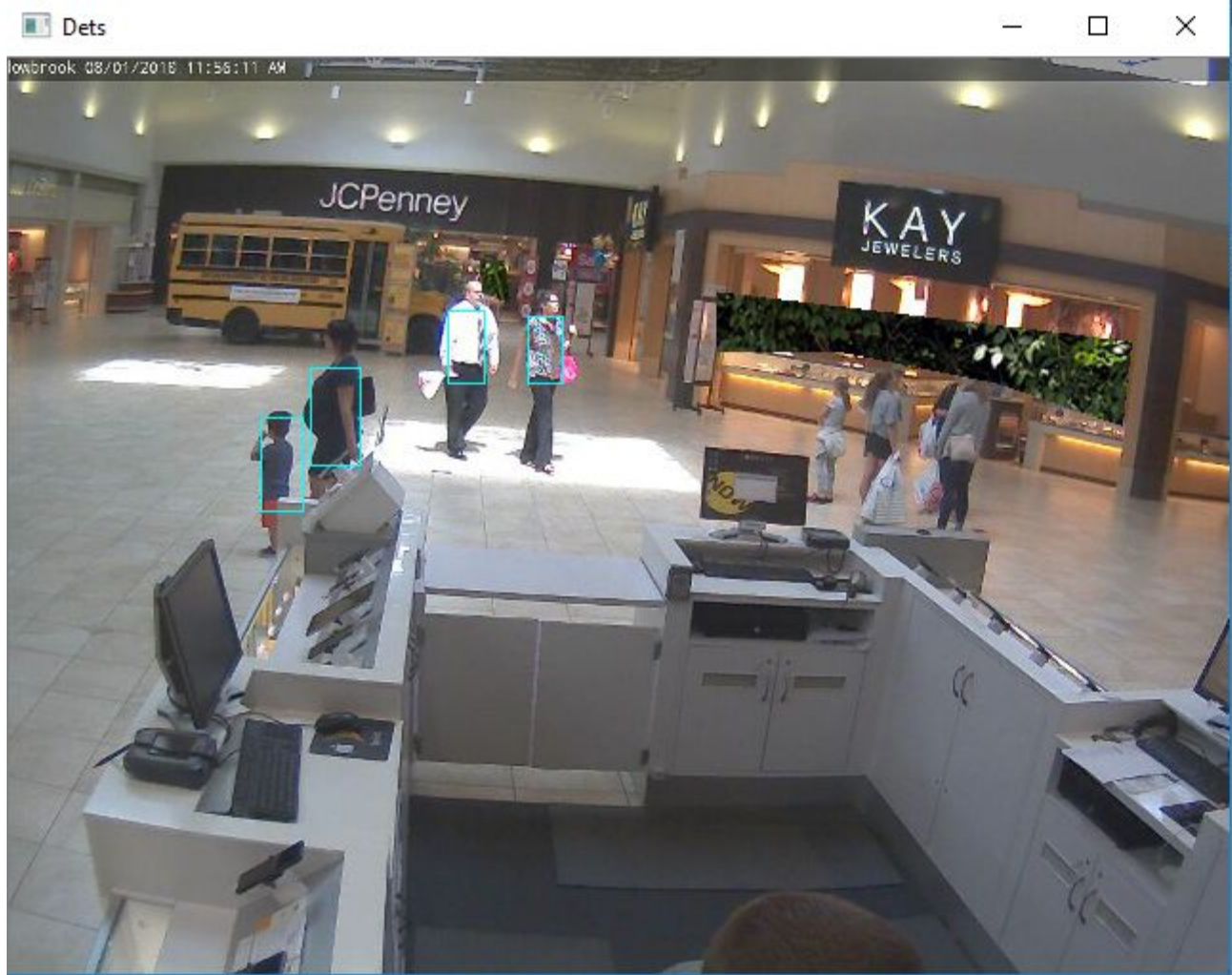
$$\text{MOTP} = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t}, \quad (2)$$

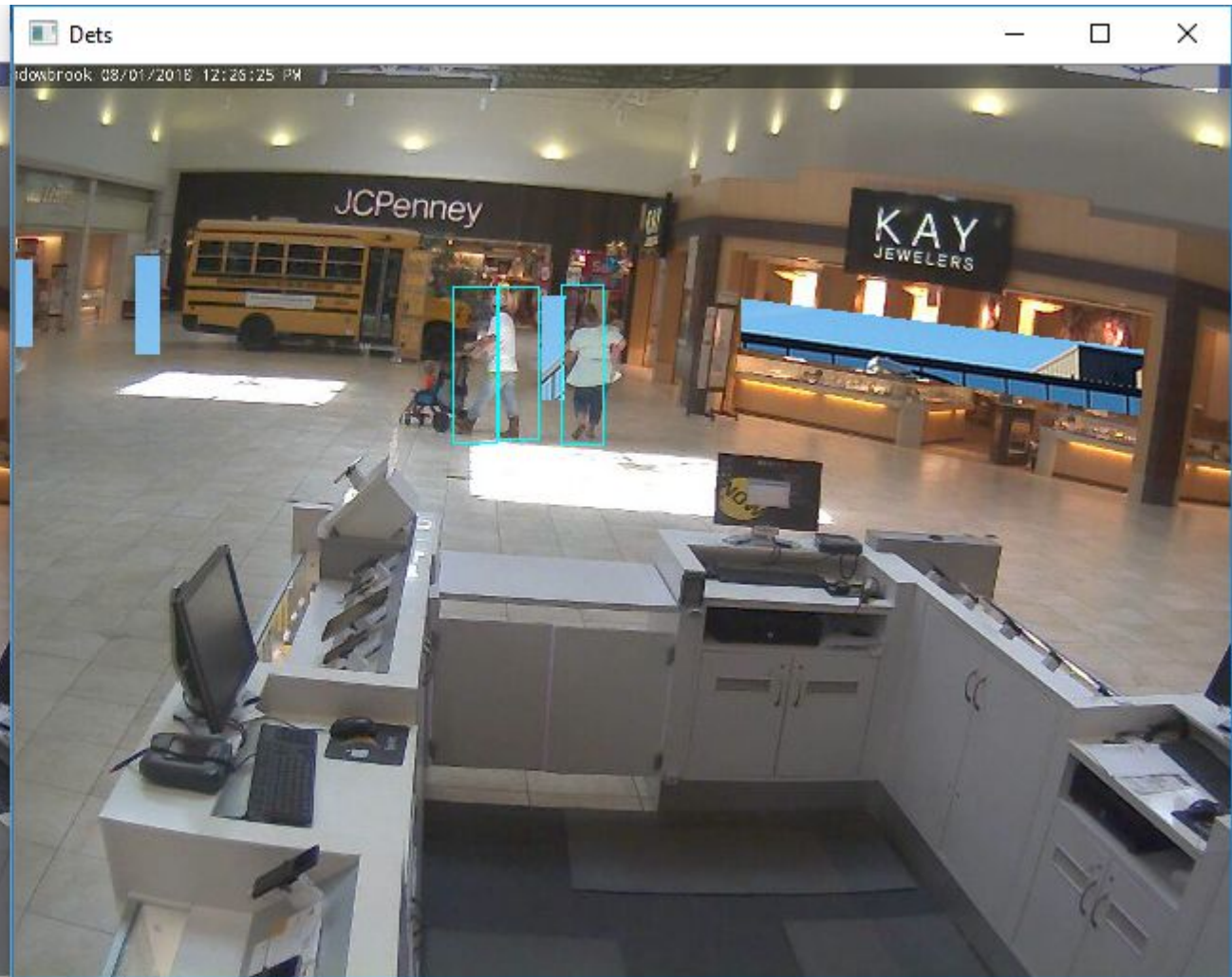
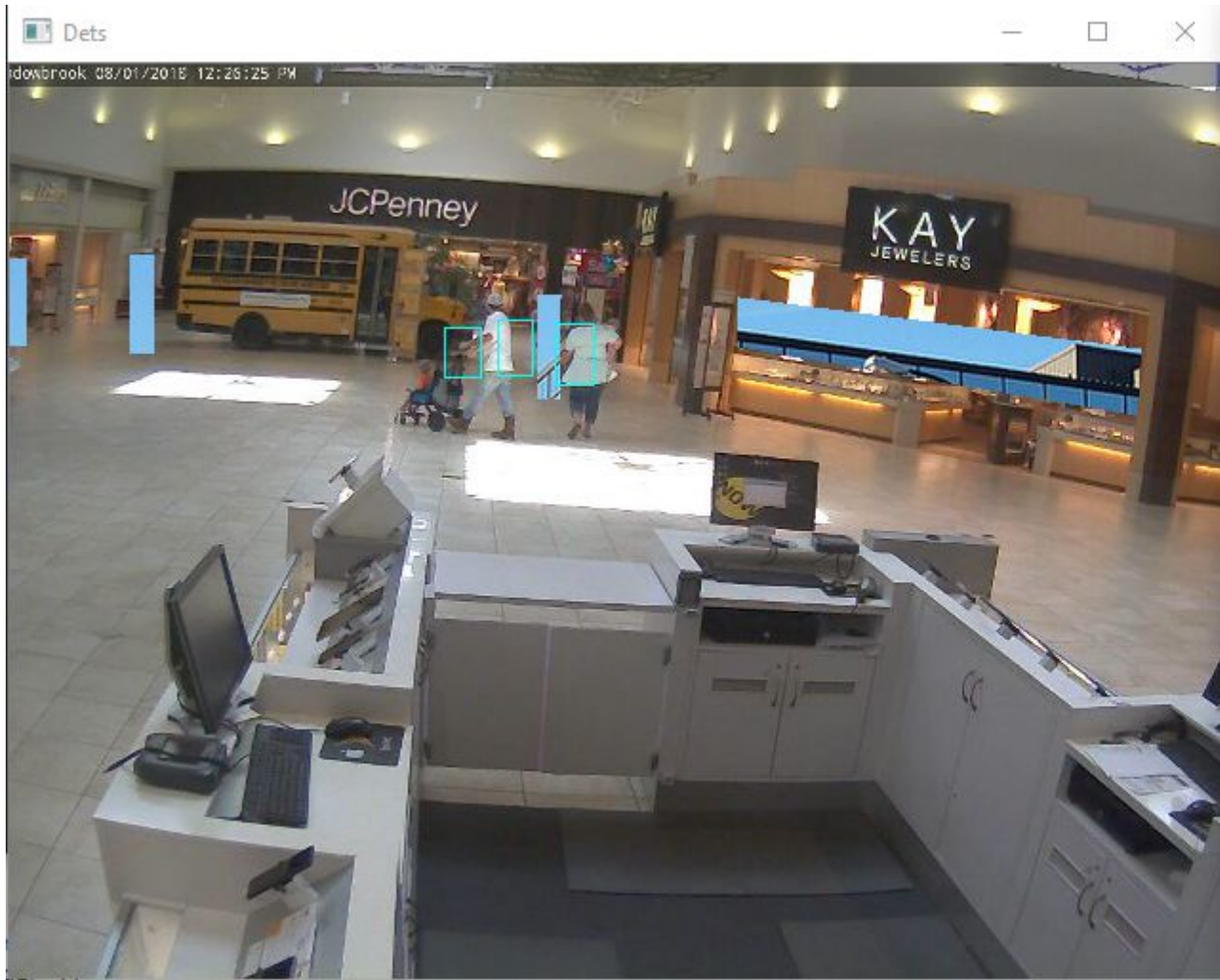
where c_t denotes the number of matches in frame t and $d_{t,i}$ is the bounding box overlap of target i with its assigned ground truth object.

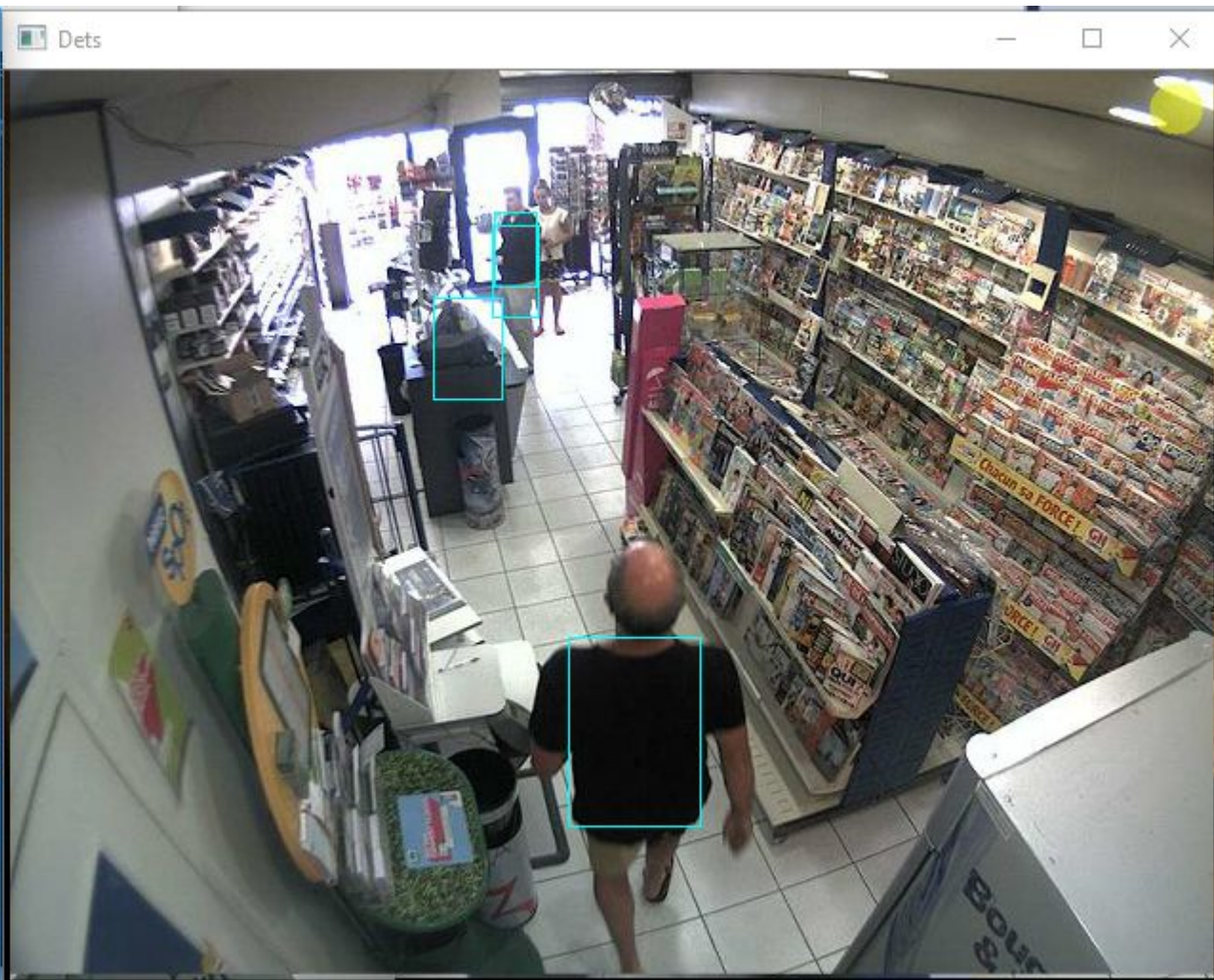
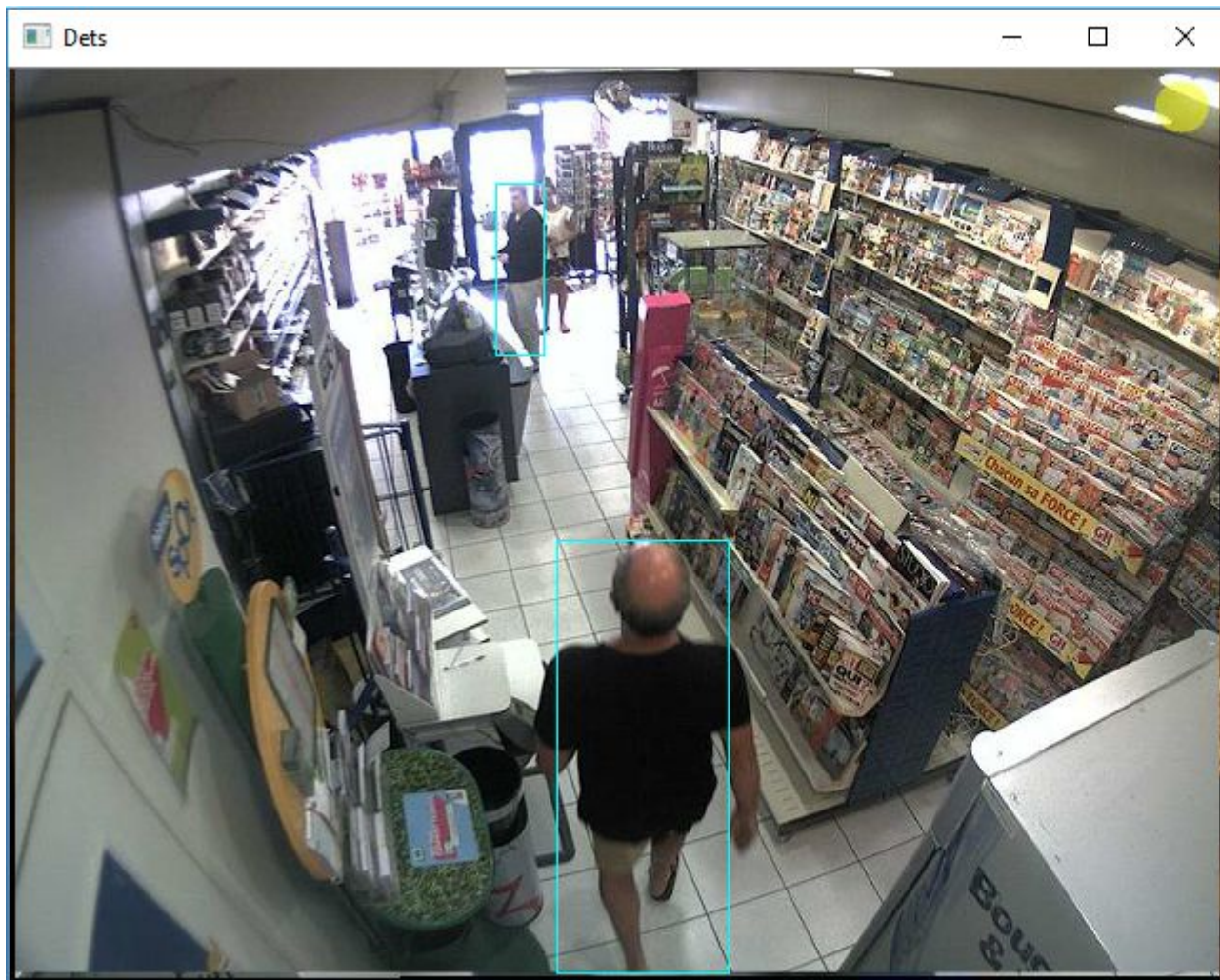
Fullbody vs Core

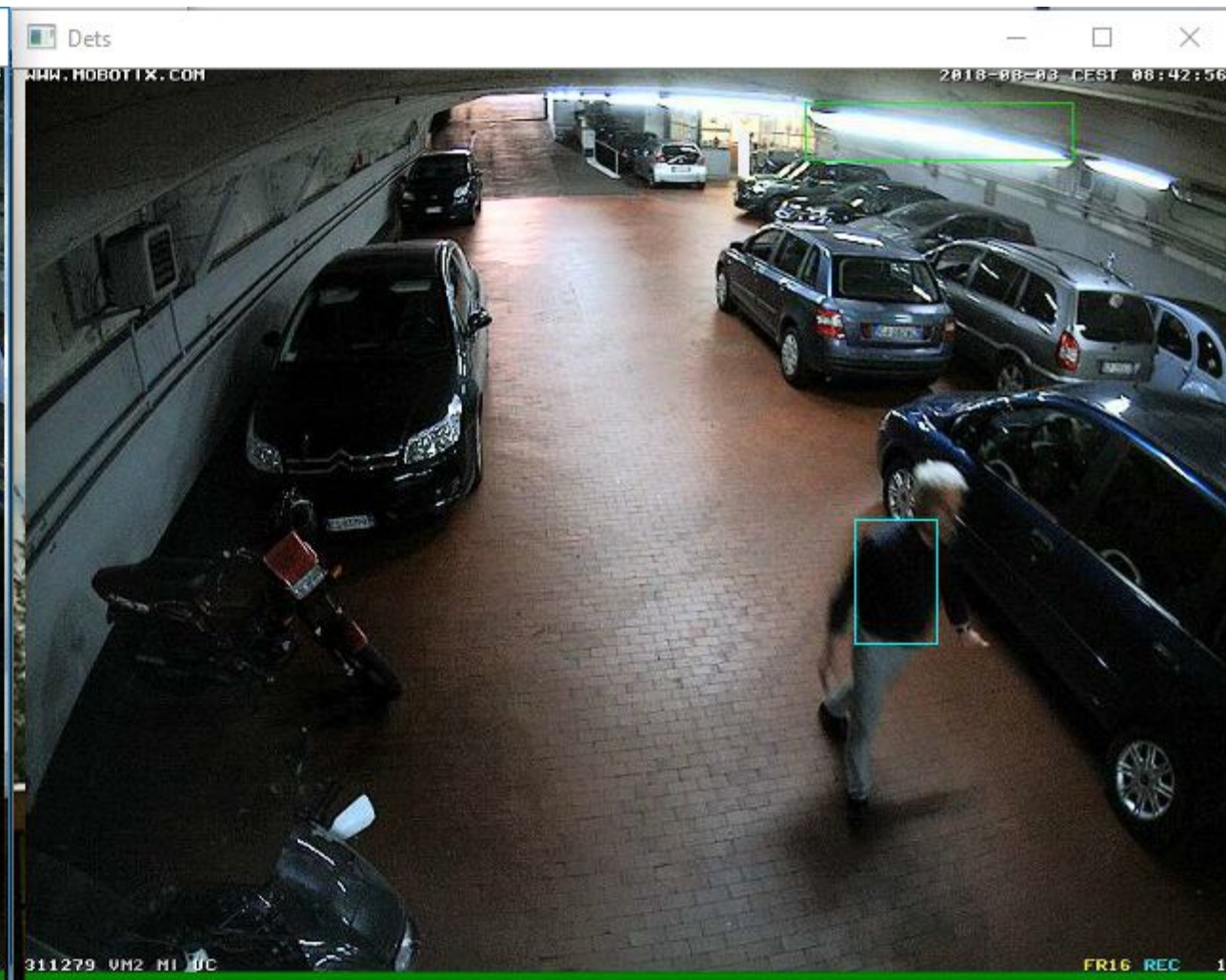
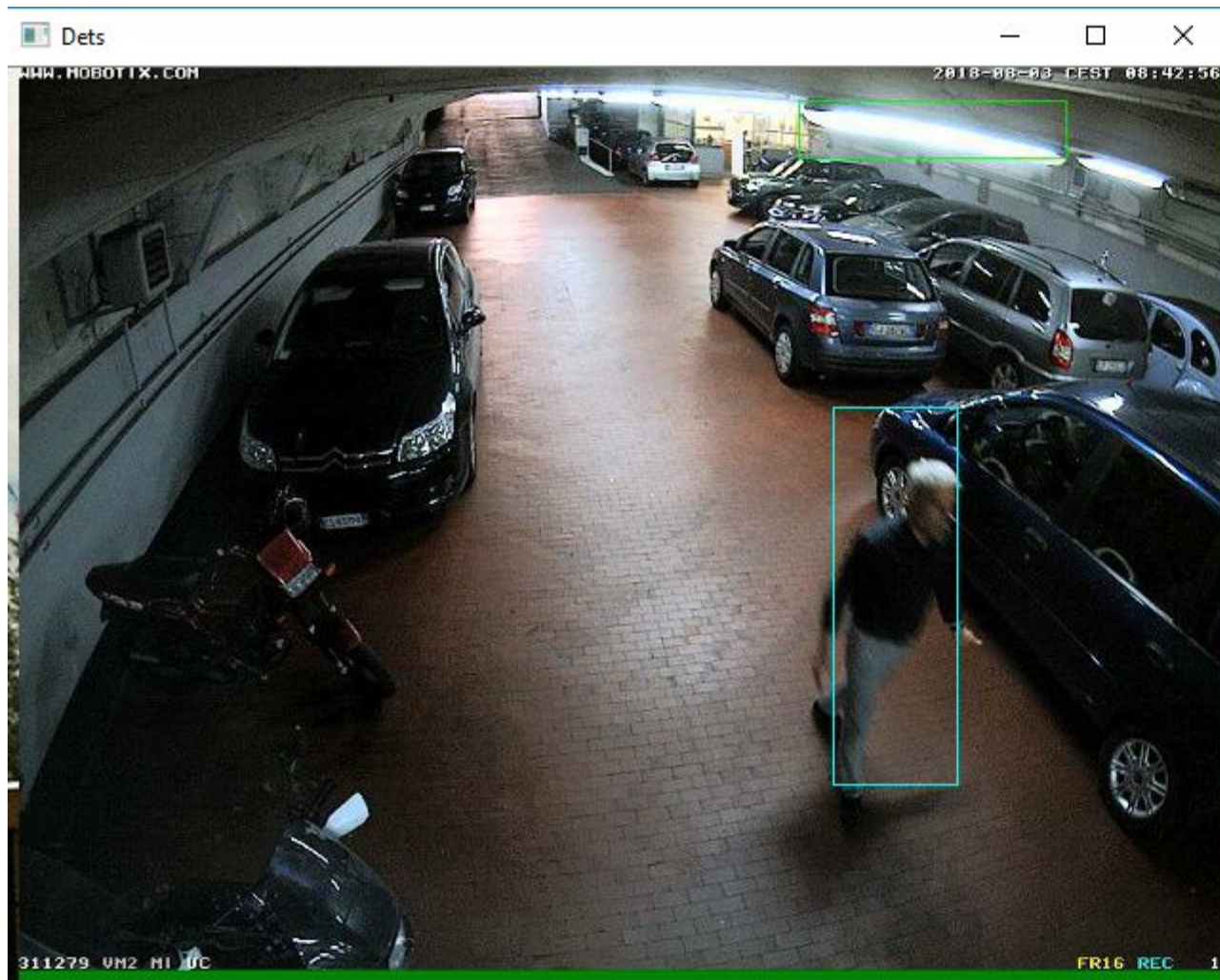
3.20.1_fb	MOTA	MOTP	3.20.1_core	MOTA	MOTP
Barker_Ewing_Whitewater_01	0,236	0,25	Barker_Ewing_Whitewater_01	0,389	0,742
Barker_Ewing_Whitewater_02	0,934	0,192	Barker_Ewing_Whitewater_02	0,484	0,746
Barker_Ewing_Whitewater_04	0,878	0,184	Barker_Ewing_Whitewater_03	0,671	0,662
Barker_Ewing_Whitewater_05	0,372	0,249	Barker_Ewing_Whitewater_04	0,907	0,699
Jackson_Hole_Whitewater_Rafting_01	0,715	0,254	Barker_Ewing_Whitewater_05	0,546	0,759
Maple_Supermarket_00	0,717	0,278	Jackson_Hole_Whitewater_Rafting_01	0,704	0,739
Maple_Supermarket_03	0,651	0,194	Maple_Supermarket_00	-0,457	0,664
Metro_PCS_Store_Robbery	0,745	0,257	Maple_Supermarket_03	0,771	0,636
OVERALL	0,586	0,244	Metro_PCS_Store_Robbery	0,432	0,666
			OVERALL	0,547	0,724

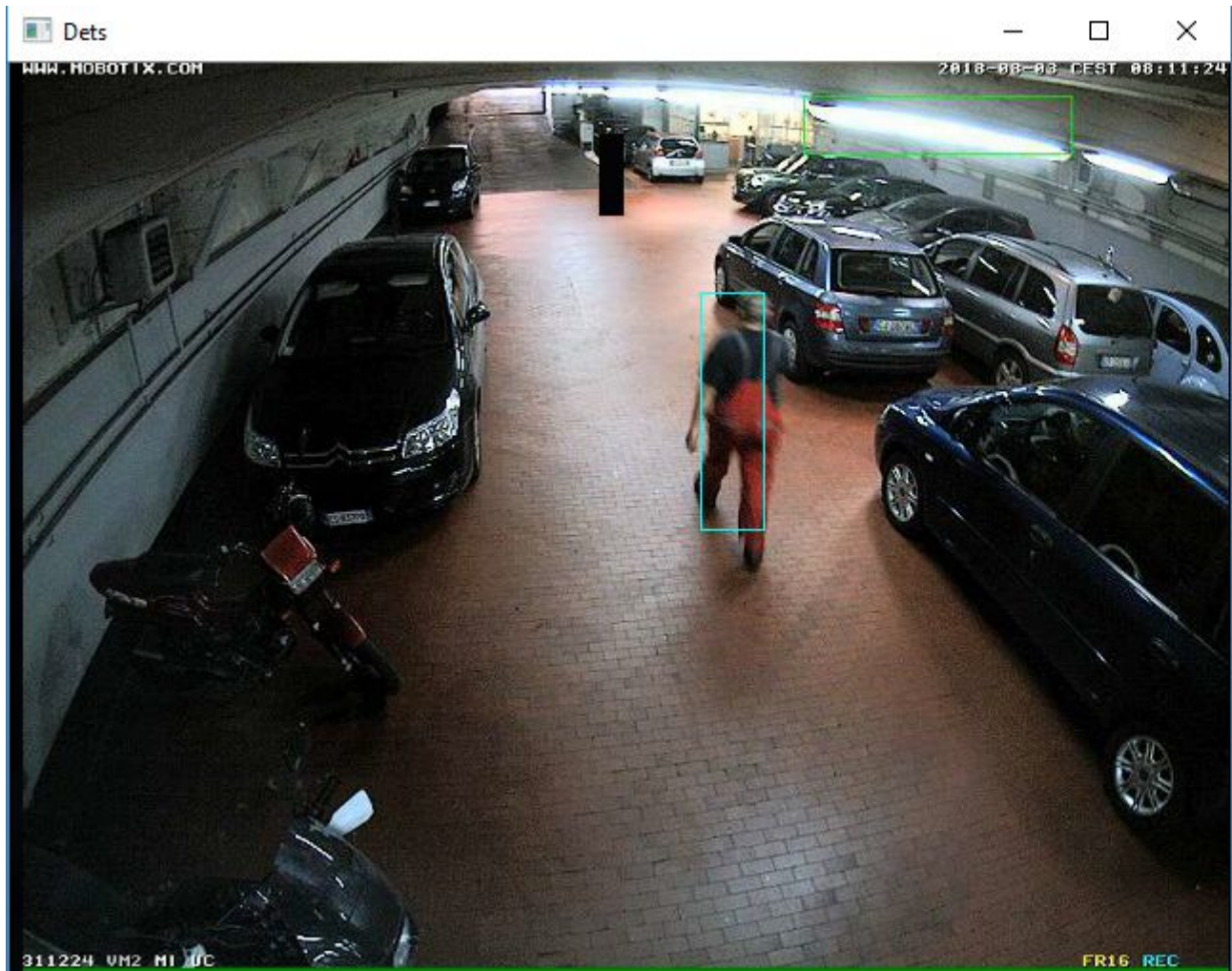


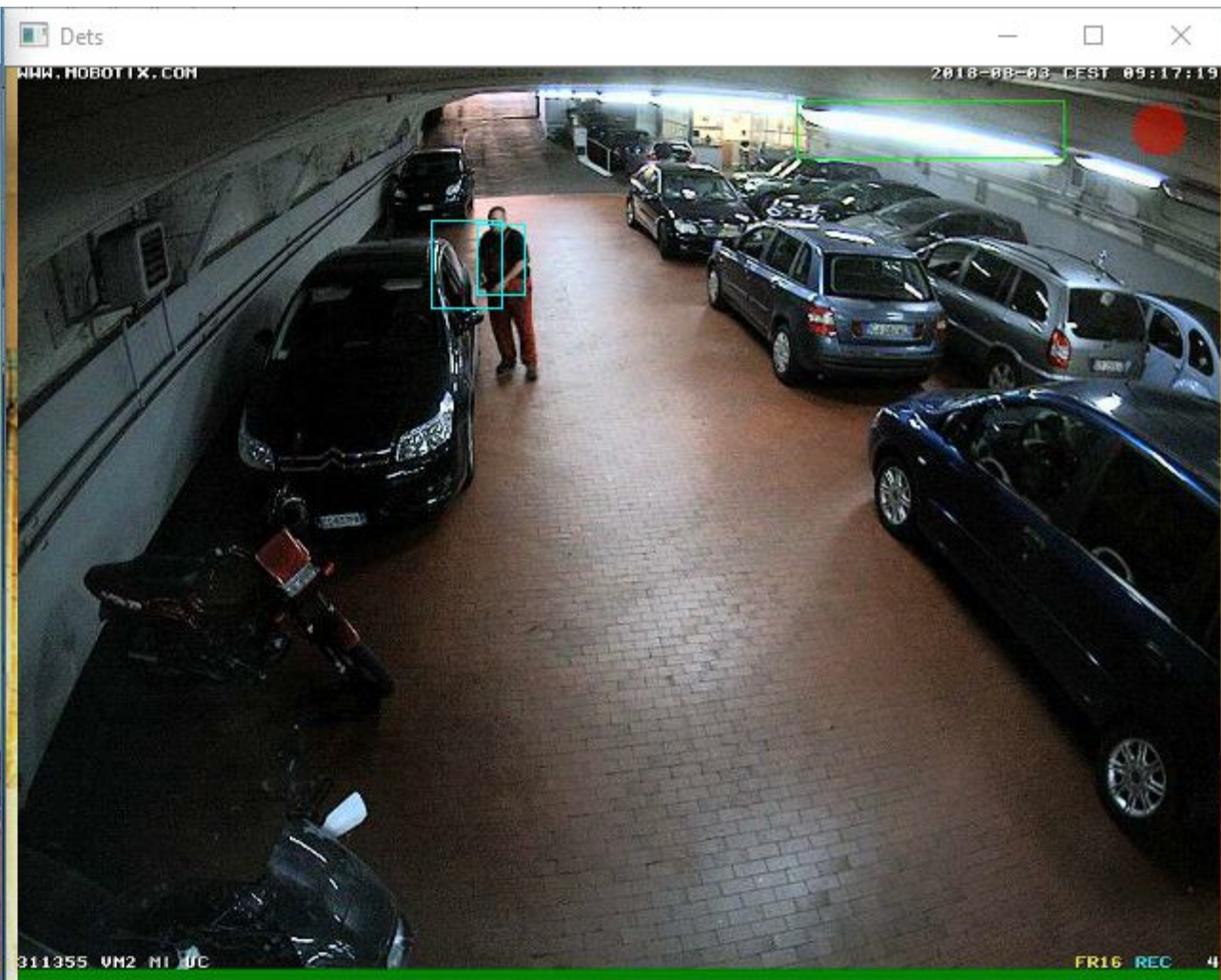
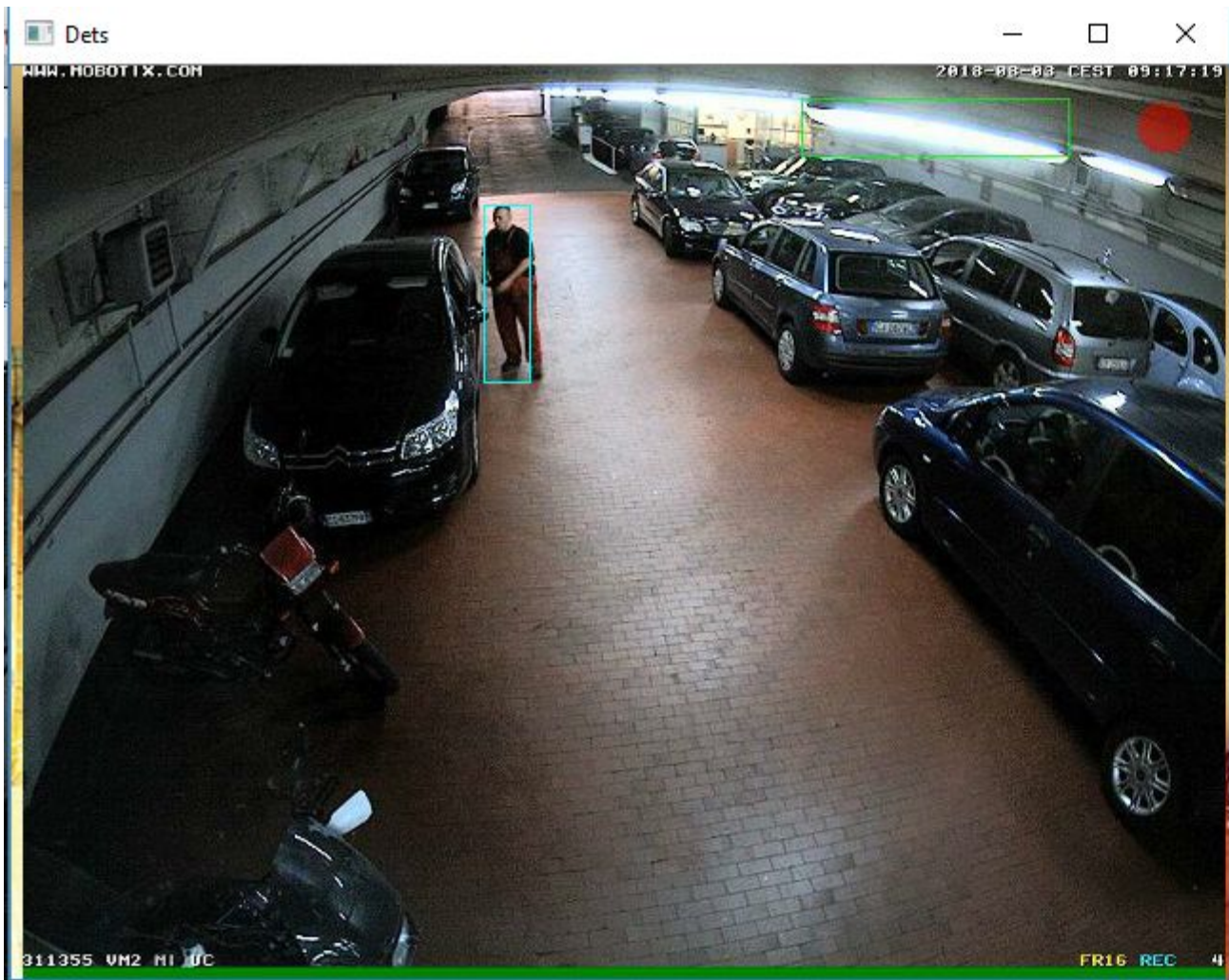












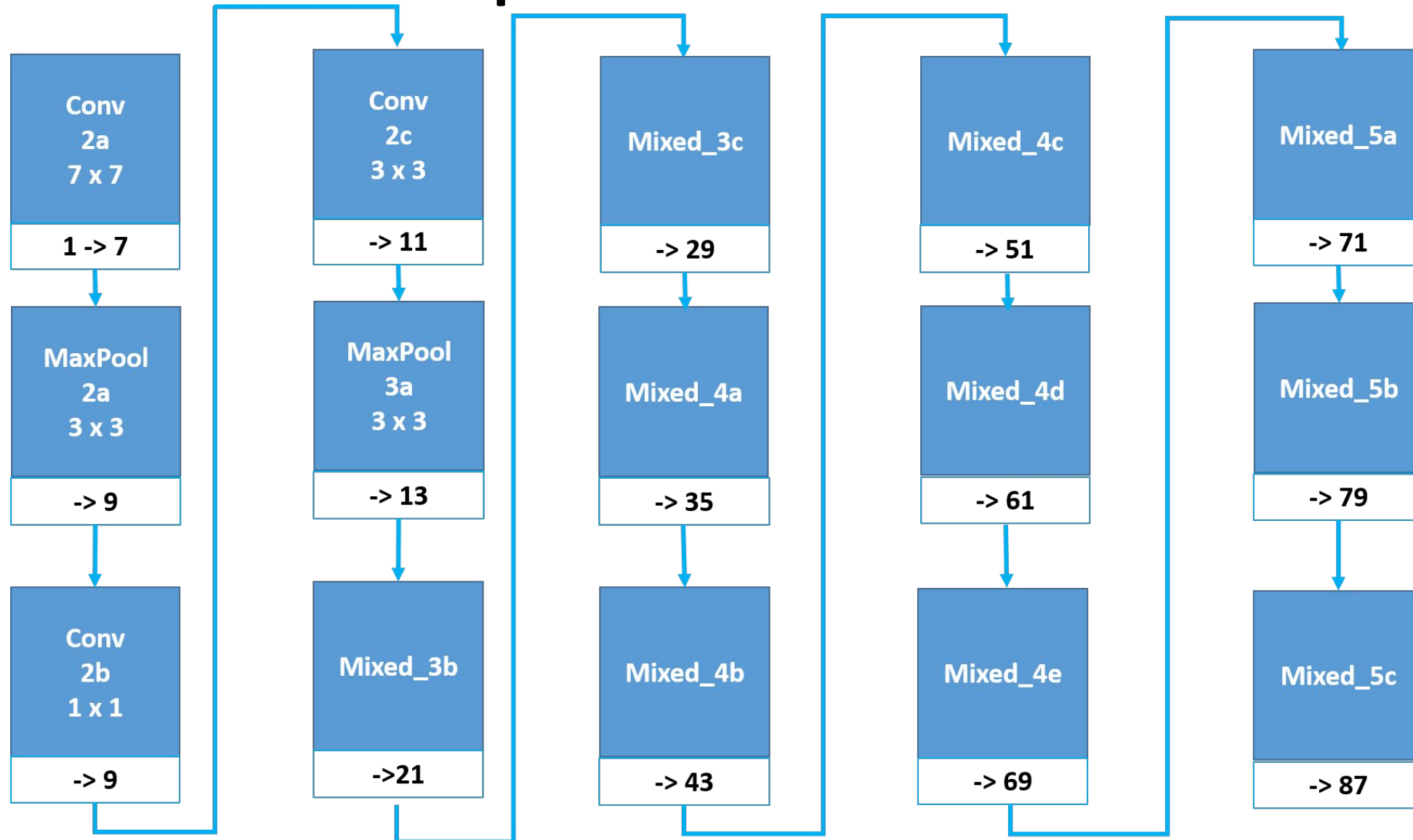
Test with different size of training base

Models	F1	Precision	Recall	TP	FP	FN	n_steps	val_loss	val_loss_class	val_loss_loc
0,25a	0,76	0,82	0,71	2857	616	1185	39345	4,19	2,98	0,68
0,25b	0,78	0,84	0,73	2939	552	1103	34308	4,23	3,01	0,68
0,50a	0,8	0,86	0,74	2989	478	1053	60226	4,09	2,9	0,65
0,50b	0,78	0,85	0,73	2937	535	1105	42082	4,15	2,95	0,66
0,75a	0,77	0,86	0,7	2810	474	1232	74969	4,29	3,08	0,7
0,75b	0,76	0,82	0,71	2871	649	1171	80323	4,3	3,1	0,69
0,95a	0,75	0,84	0,68	2729	518	1313	56859	4,53	3,23	0,77
0,95b	0,74	0,82	0,67	2706	596	1336	100802	4,54	3,24	0,76

Future work

- Look for a more suitable architecture
 - Inception modifications
 - MobileNet with Feature Pyramid Networks

Inception model v2



Test with different image size

Models	IDF1	IDP	IDR	Rcll	Prcn	MOTA	MOTP
160x120	0,406	0,475	0,354	0,588	0,788	0,424	0,286
240x180	0,542	0,551	0,533	0,736	0,76	0,5	0,253
320x240	0,622	0,686	0,568	0,761	0,919	0,69	0,23
400x300	0,556	0,595	0,522	0,756	0,862	0,629	0,234
480x360	0,621	0,707	0,554	0,732	0,933	0,676	0,228
560x420	0,607	0,658	0,563	0,768	0,898	0,677	0,222
640x480	0,578	0,623	0,539	0,696	0,805	0,524	0,222