Utilising Visual Attention Cues for Vehicle Detection and Tracking

Insight

SFI RESEARCH CENTRE FOR DATA ANALYTICS

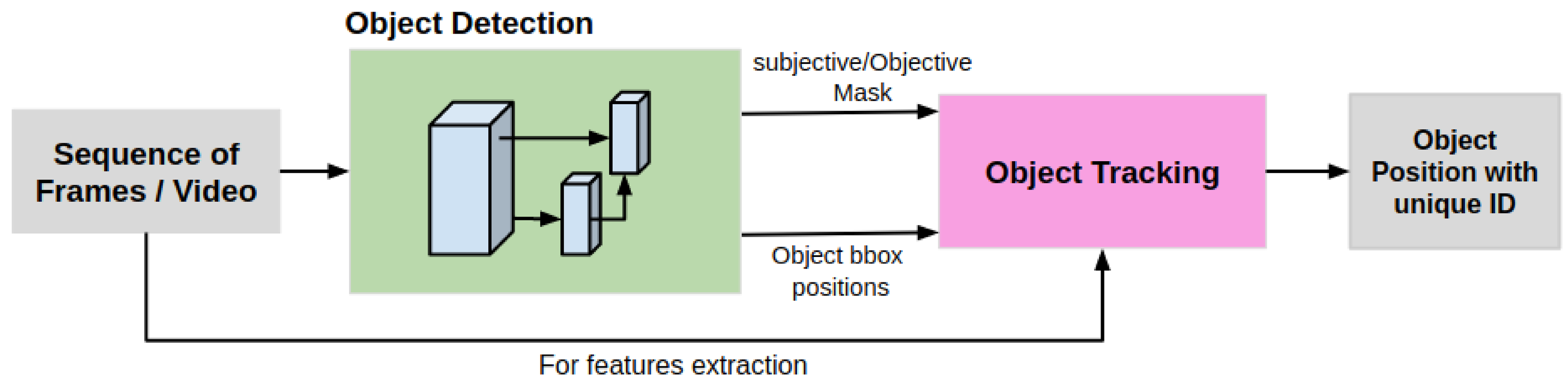
Feiyan Hu, Venkatesh G M, Noel E. O'Connor, Alan F. Smeaton and Suzanne Little

A World Leading SFI Research Centre

Science Foundation Ireland For what's next

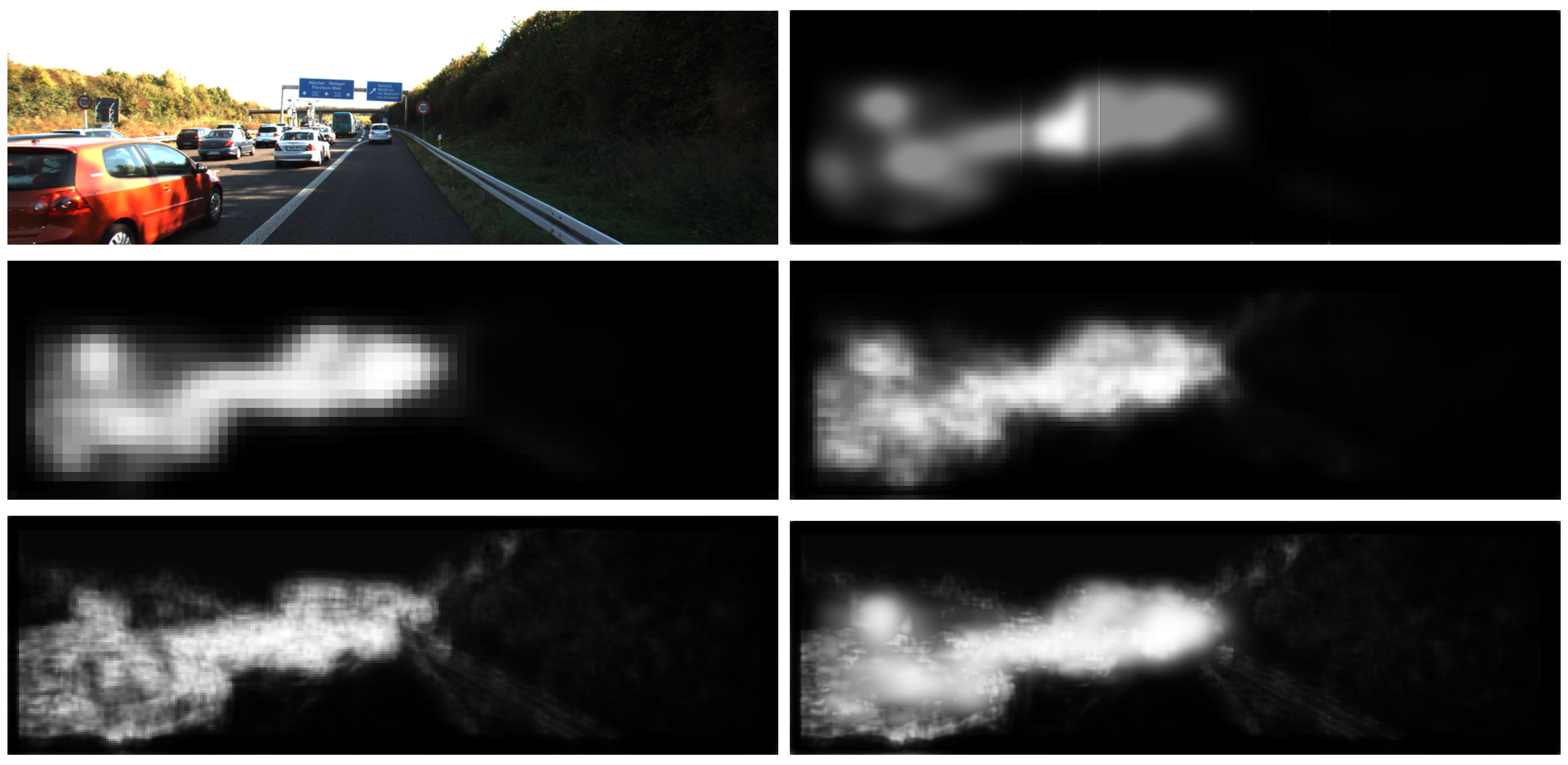
Introduction

An accurate, robust and safe Advanced Driver Assistance Systems (ADAS) detects and tracks other road users (objects) using sensors incorporated into the vehicles. This includes visual sensors such as video that results in very high volumes of input to be processed and interpreted in near real-time. Human drivers do not focus on all objects at all times but rather focus on the salient or critical regions in their field of view. We can focus and divert attention based on task priority. Similarly, in computer vision, visual saliency can predict how our visual perception ranks the importance of visual information, whether low level features or high level semantics. We propose a detection and tracking system that could prioritizing image regions based on subjective and/or objective visual attention cues.

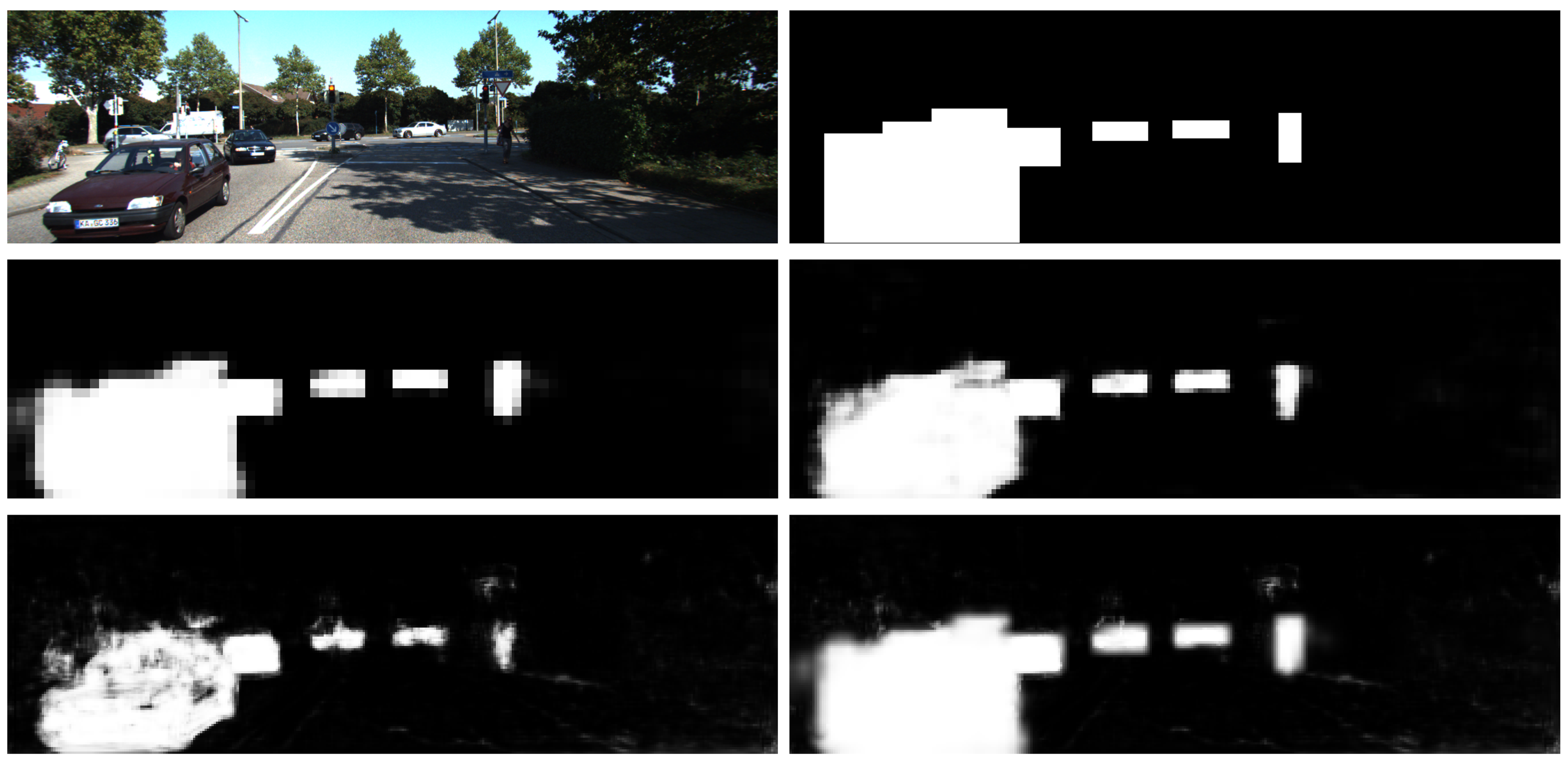


Subjective and Objective Attention

Subjective attention or saliency models are normally trained with eye fixation data collected when experiment participants view images. The images displayed to the participants normally contain broad concepts and generic object classes. We use SalGAN to generate subjective attention map as auxiliary target trained along with object detection task.

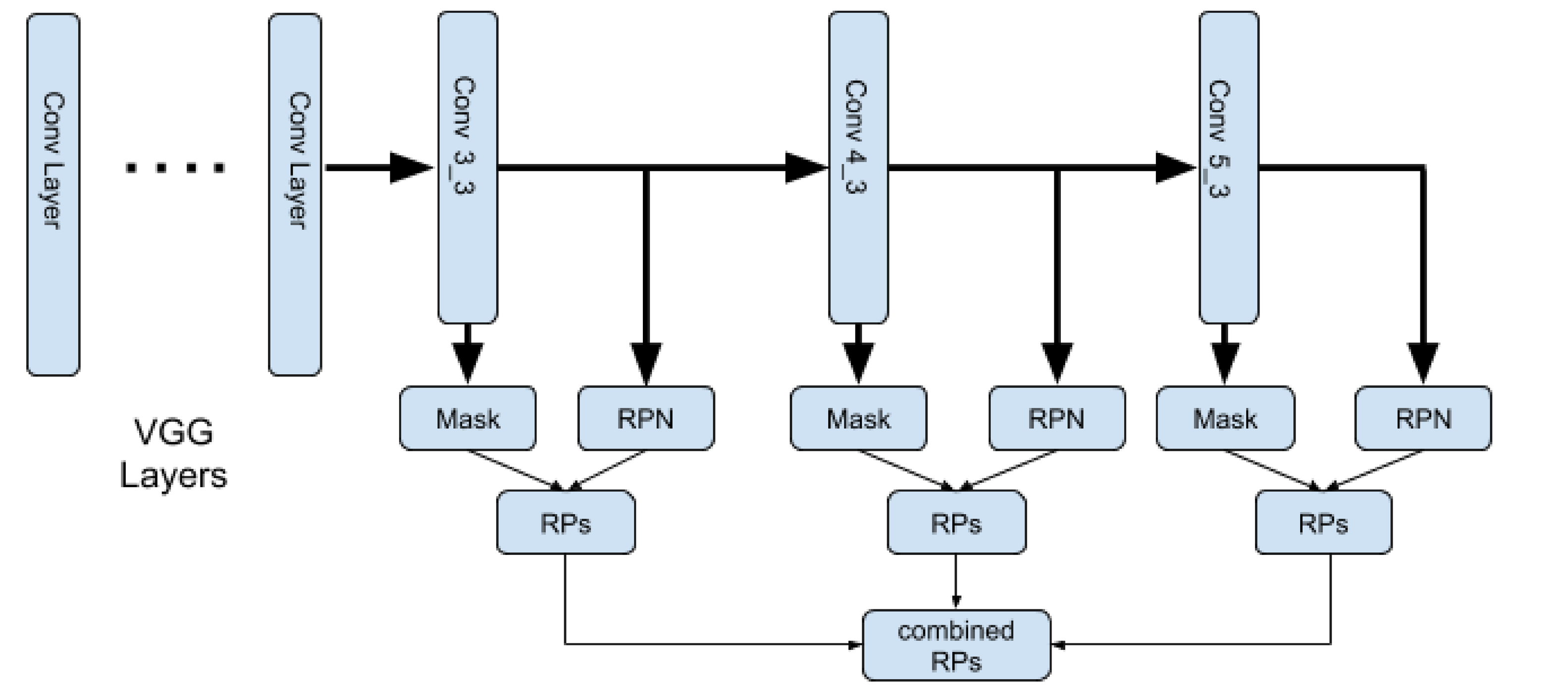


An objectness map is more object oriented and is generated using ground truth bounding box data. It separates foreground and background and thus identifies possible coarse locations for objects.



Utilising Attention in Detection

- Faster RCNN based network (RPN and ROIAlign).
- Shared visual representation for 3 tasks (Objection detection, subjective map and objective map generation)
- Hierarchical features for 3 tasks.
- Attention based Region Proposals Filtering.



Experiments using Attention RP Filtering

To train the object detection model, we have used the KITTI object detection benchmark dataset and DETRAC vehicle dataset. The following resutls are reported on split KITTI training set. The experiments show that about **10%** of the total area of features maps are contributing to the detection of objects.

HOST INSTITUTIONS

PARTNER INSTITUTIONS

FUNDED BY:

This work has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/16/SP/3804 and SFI/12/RC/2289 P2.

Utilising Visual Attention Cues for Vehicle Detection and Tracking

Insight

SFI RESEARCH CENTRE FOR DATA ANALYTICS

Feiyan Hu, Venkatesh G M, Noel E. O'Connor, Alan F. Smeaton and Suzanne Little

A World Leading SFI Research Centre

Saliency map or subjective filtering achieved better performance than All RPs and objectness map for both "Car" and "Pedestrian" classes, when we using RPs with top 4 saliency scores. "OM1" and "OM4" are top 1 and 4 proposals with highest objectness scores for objectness map. "SM" is subjective map or saliency map.

Training		All RPs					Random RPs(OM, SM, All)				
Testing		OM1	SM1	OM4	SM4	All	OM1	SM1	OM4	SM4	All
Car	E	90.8	90.8	90.8	90.8	90.8	90.8	90.9	90.8	90.8	90.8
	M	90.6	90.6	90.6	90.6	90.6	90.6	90.7	90.6	90.7	90.7
	H	88.4	88.9	89.2	89.4	89.3	87.9	88.8	89.2	89.4	89.2
	mAP	81.2	81.3	81.3	81.3	81.3	81.2	81.3	81.3	**87.0**	81.3
Ped.	E	67.9	68.1	72.8	74.1	73.1	68.3	68.6	73.9	75.0	74.4
	M	57.5	58.3	58.8	61.2	58.8	57.7	58.9	59.4	62.5	59.4
	H	50.1	50.4	50.7	55.4	50.6	50.4	50.8	50.8	55.7	51.2
	mAP	49.5	50.0	50.4	52.9	50.3	49.7	50.5	50.5	**54.0**	50.8
% of RPs		0.86	1.10	3.43	4.41	100	0.87	1.11	3.48	4.44	100

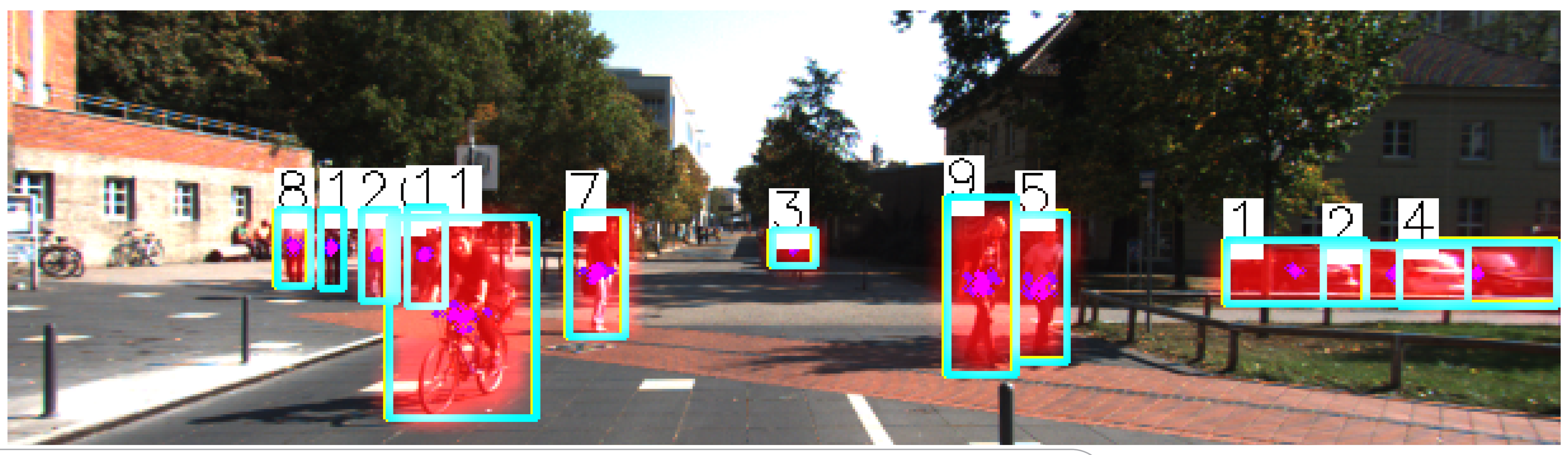
Utilising Attention in Tracking

In this paper, a tracking-by-detection framework is employed to track detected object in the scene using a modified sequential Monte Carlo probability hypothesis density(PHD) filter utilising the attention maps generated by the detection module. The subjectness or objectness maps assist the tracker to correct the predicted position of the targets during detection failure.

Modified Particle-PHD filter with visual attention cues

Specifically, in the proposed filter, Kalman gain along with the visual cues is used to compute the inter-frame displacement of the objects to facilitate the particle distribution and re-sampling process.

- Weighted IoU and distance metric to compute state estimation. Track history and visual attention cues to compute temporal histogram.
- Motion cues to resample particles.
- Additional correction mechanism when occupancy below 30% of the area of the predicted box. In this case, attention cues are used to correct prediction.



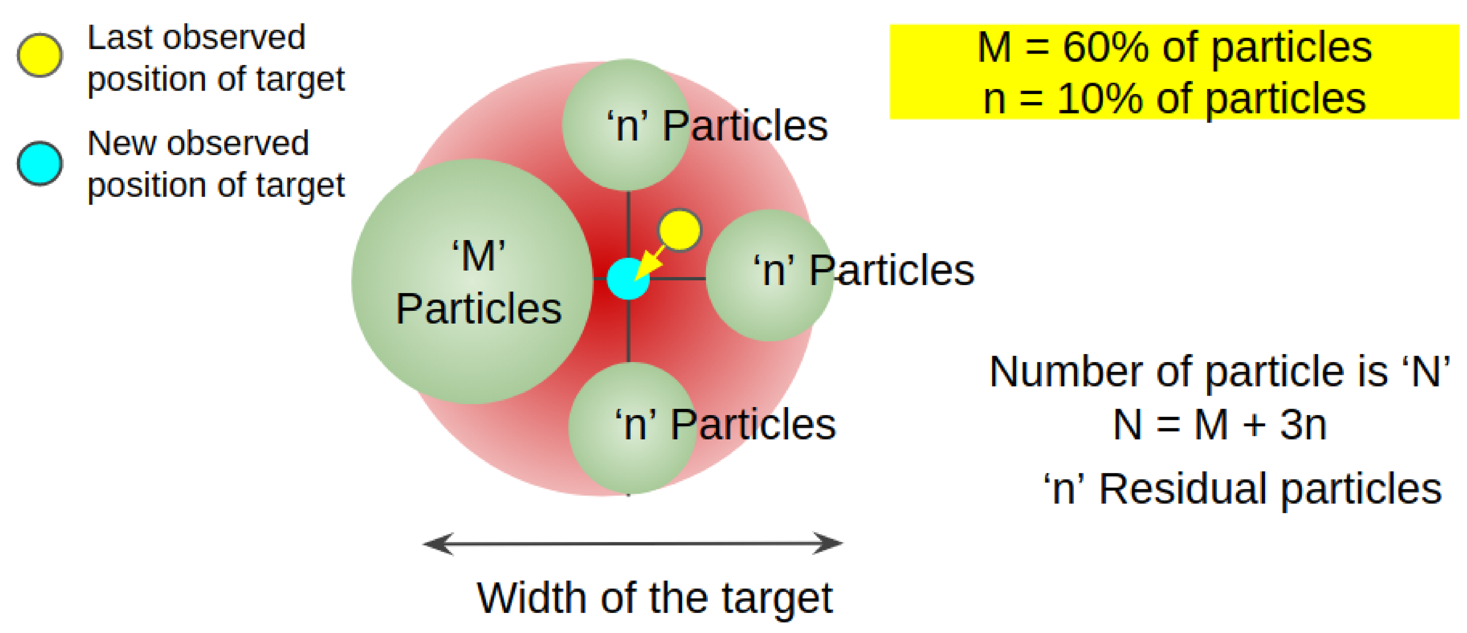


Figure: multi-peak Gaussian particle distribution based on object motion

Experiments using Attention Cues

Class	Method	M.A	M.P	Rcll	Prcn	F1	FAR	MT	PT	ML	IDs	FM
Car	Baseline	79.1	80.7	85.3	85.3	90.9	**8.3**	65.4	29.8	4.8	**244**	655
	all-SM	83.6	82.0	90.2	96.3	93.2	11.9	75.2	21.6	3.2	246	534
	all-OM	83.7	81.9	90.1	96.6	93.2	11.1	74.5	22.2	3.4	251	526
	sub-SM	84.6	81.9	**91.0**	96.5	93.7	11.5	**77.3**	**19.5**	3.2	265	538
	sub-OM	**84.8**	81.9	90.9	**96.8**	**93.8**	10.5	77.1	19.9	**3.0**	268	553
	obj-SM	83.0	82.1	89.7	96.4	92.9	11.8	74.8	21.5	3.7	246	**521**
	obj-OM	83.2	8**2.1**	89.6	96.6	93.0	11.1	74.5	22.0	3.6	252	531
Ped.	Baseline	58.6	75.1	66.0	92.2	76.9	7.9	35.9	54.5	9.6	**147**	547
	all-SM	62.0	77.2	69.8	92.2	79.5	8.3	42.5	50.3	7.2	160	488
	all-OM	62.7	**77.4**	69.6	**93.2**	79.7	**7.1**	44.9	47.9	7.2	154	**479**
	sub-SM	63.2	77.0	71.7	91.4	80.4	9.4	**48.5**	**44.9**	**6.6**	159	520
	sub-OM	**64.5**	76.9	**71.8**	92.9	**81.0**	7.7	**48.5**	**44.9**	6.6	161	516
	obj-SM	60.9	77.3	69.1	91.5	78.7	8.9	42.5	50.9	**6.6**	149	481
	obj-OM	61.8	**77.4**	69.0	92.9	79.2	7.4	42.5	50.9	**6.6**	163	496

Conclusion

In the paper, we describe an object detector and a tracker that take full advantage of visual attention cues for improved processing efficiency. We train a detector that can simultaneously generate objectness and saliency maps using joint image representation. The visual attention cues is used as guidance to filter out the region proposals that are not in important/salient regions. Multiple object tracking using a modified sequential Monte Carlo probability hypothesis density (PHD) filter is explored utilising the visual attention map during particle resampling and distribution process while tracking. The experiments show that attention maps could be a very good heuristic to select region of interest and generate region proposals for effective object detection.

HOST INSTITUTIONS

PARTNER INSTITUTIONS

FUNDED BY:

This work has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/16/SP/3804 and SFI/12/RC/2289 P2.