Utilising Visual Attention Cues for Vehicle Detection and Tracking

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

Insight SFI Research Centre for Data Analytics
Dublin City University

A World Leading SFI Research Centre

Science Foundation Ireland For what's next

Insight
SFI RESEARCH CENTRE FOR DATA ANALYTICS

HOST INSTITUTIONS

PARTNER INSTITUTIONS

FUNDED BY:

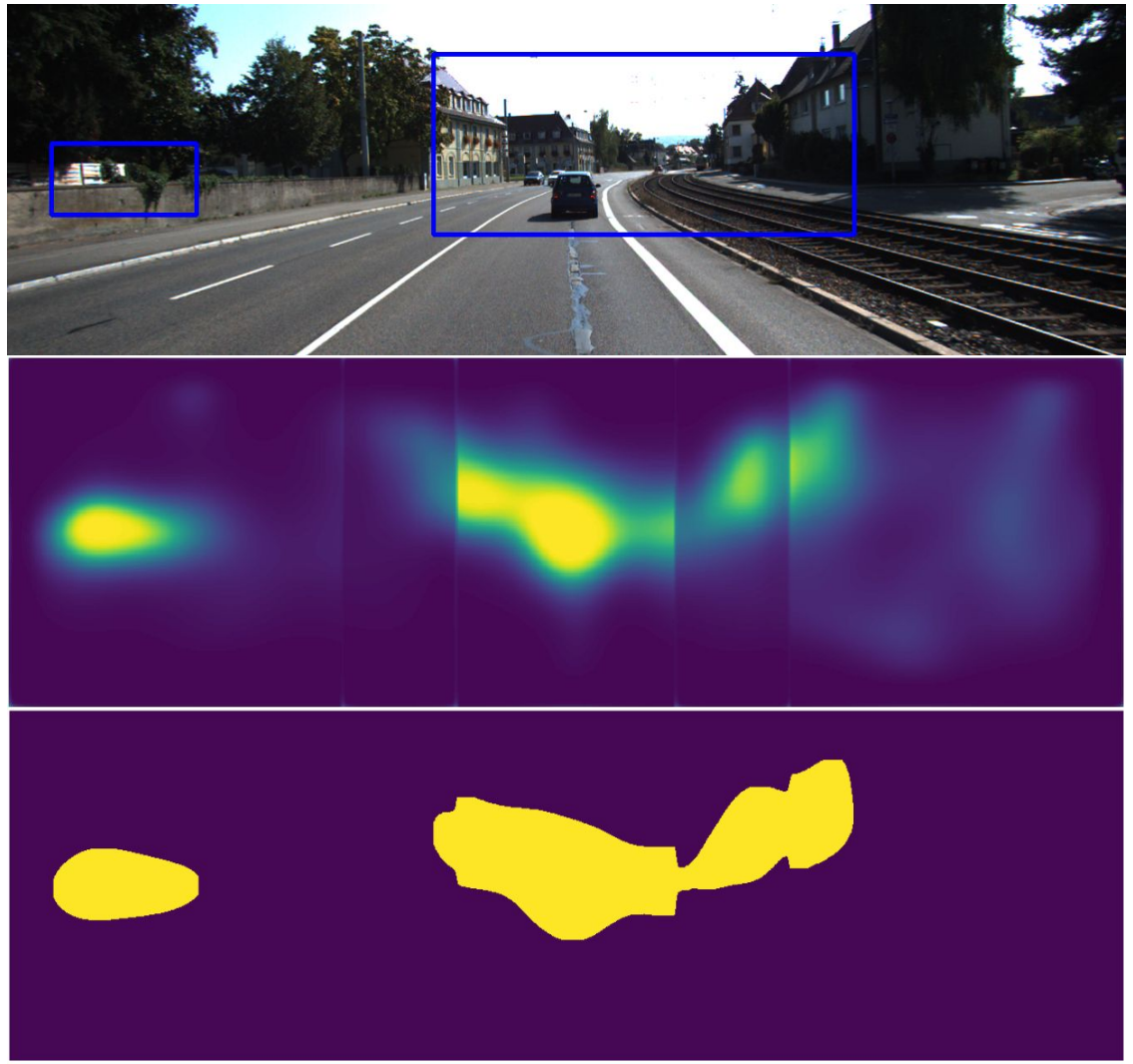
Outlines

- Motivation
- Subjective and Objective Attention
- Visual Attention in Detection
- Visual Attention in Tracking
- Conclusion

Motivation

Saliency detection predicts how humans **perceive and prioritise** visual information.

How can we **use the visual attention cues** efficiently to benefit both detection and tracking in context of Advanced Driver-Assistance Systems (**ADAS**)?

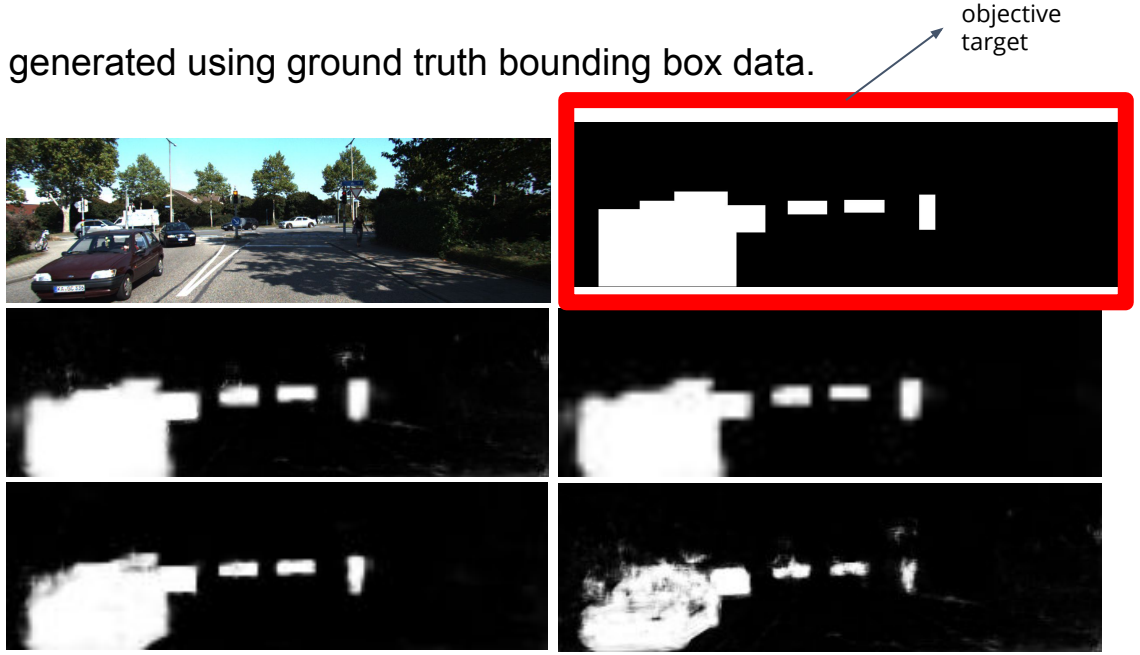


Subjective and Objective Attention

trained with eye fixation data

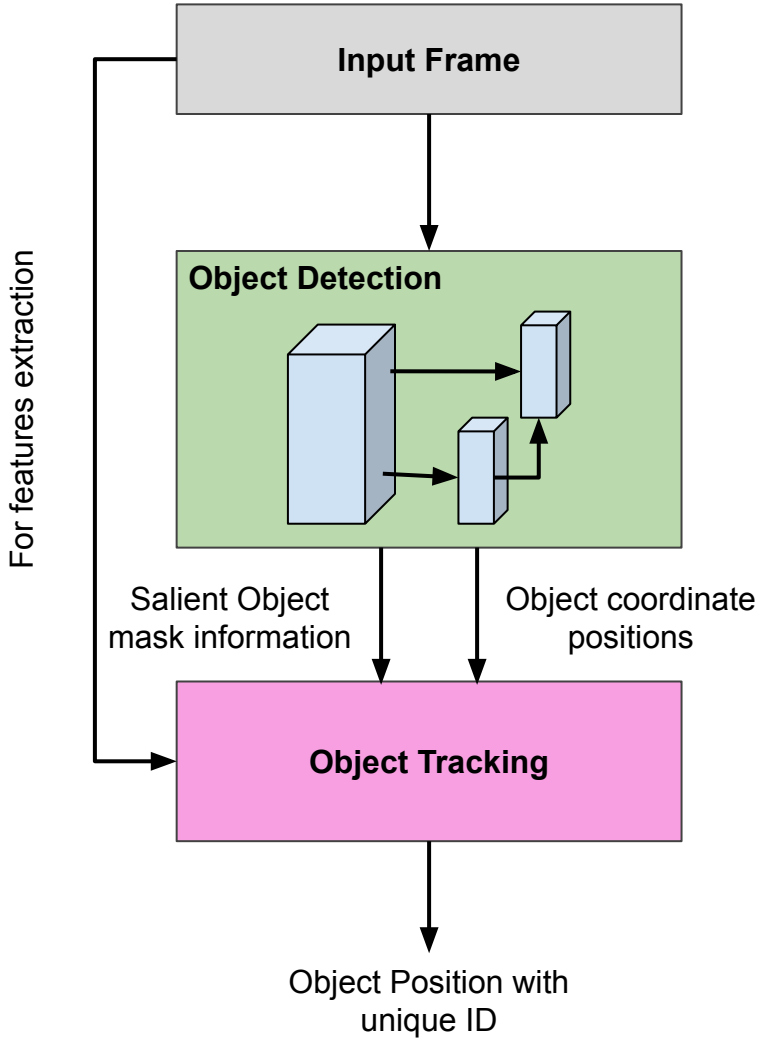
Example of Subjective Attention

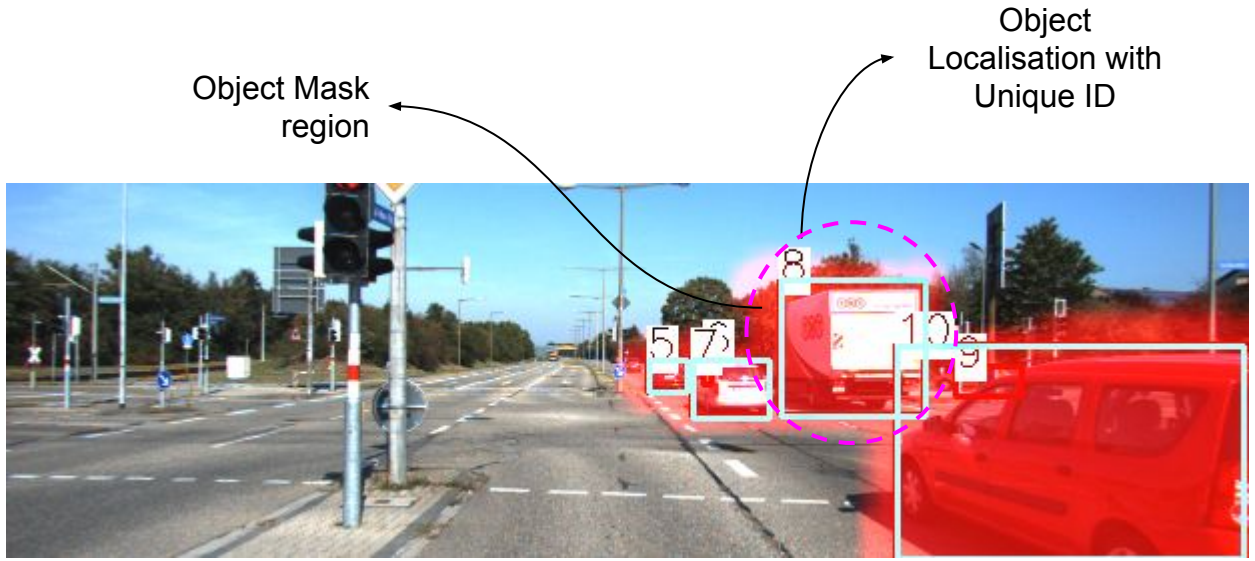
generated using ground truth bounding box data.



Example of Objective Attention

Pipeline

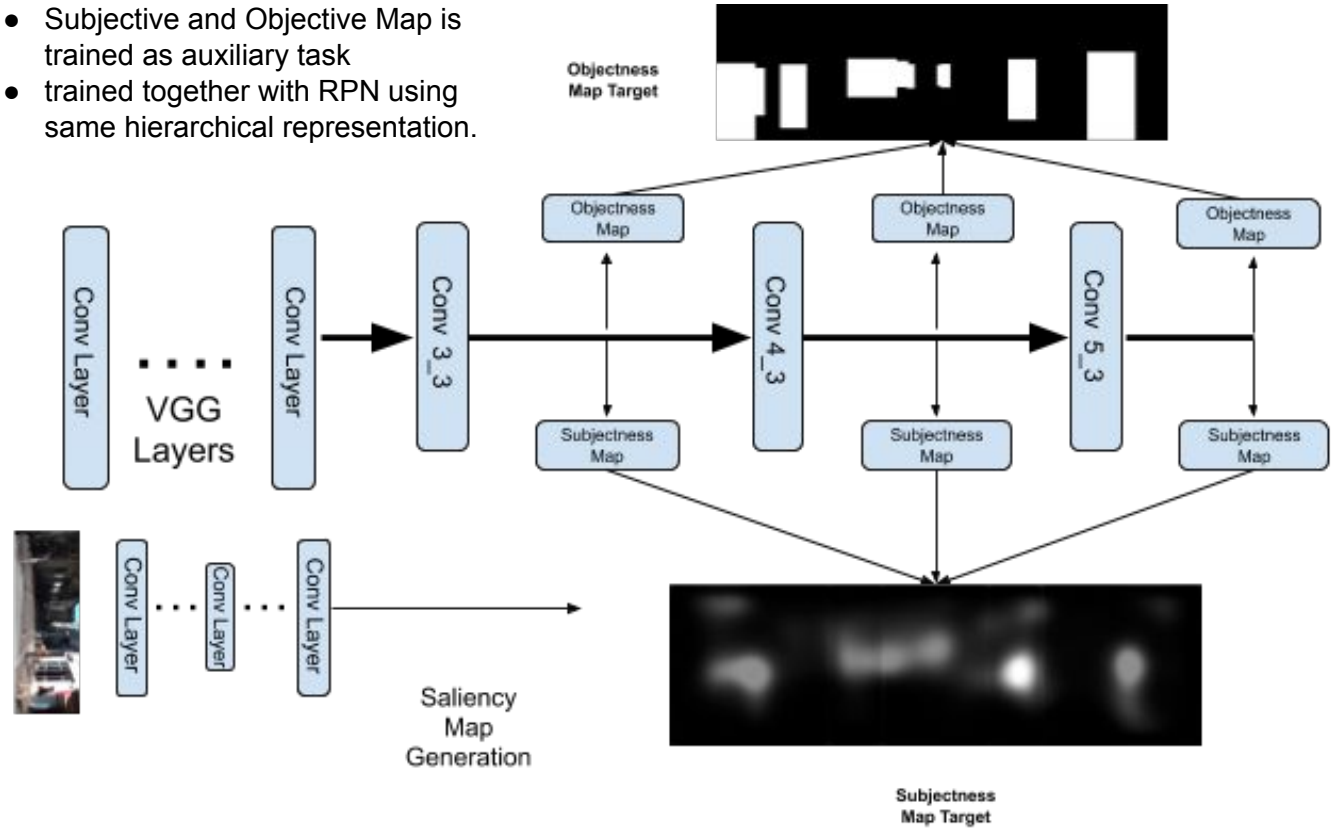

Input Frame
Object Detection
For features extraction
Salient Object mask information
Object coordinate positions
Object Tracking
Object Position with unique ID

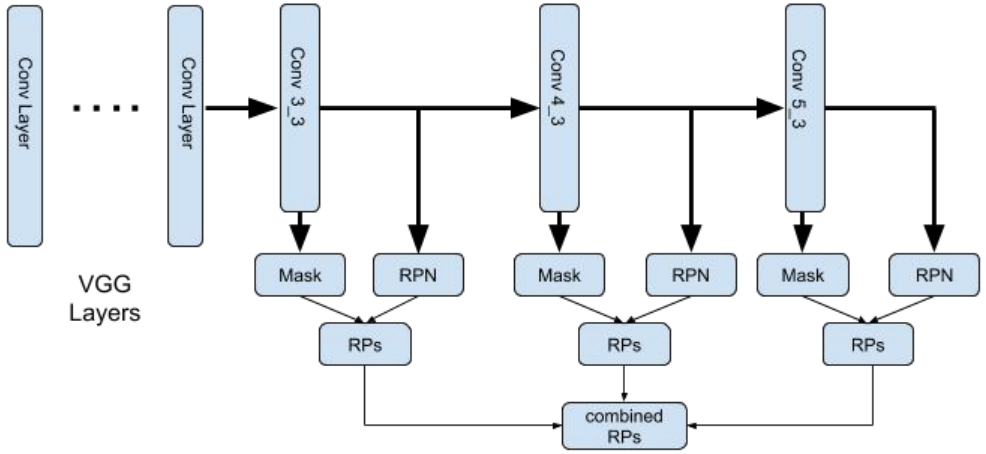

Object Mask region
Object Localisation with Unique ID

Attention masks used to filter RPs

Visual Attention in Detection

- Subjective and Objective Map is trained as auxiliary task
- trained together with RPN using same hierarchical representation.





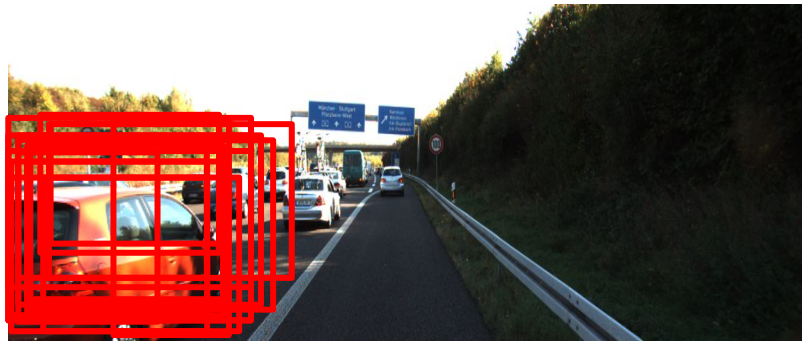
- Hierarchical features
- Shared representation for subjective, objective attention prediction and object detection.
- SalGAN[1] as saliency teacher.
- Attention cues are used to filter/reduce Region Proposals (RPs).

1. Pan, Junting, et al. "Salgan: Visual saliency prediction with adversarial networks." CVPR Scene Understanding Workshop (SUNw). 2017.

Visual Attention in Detection

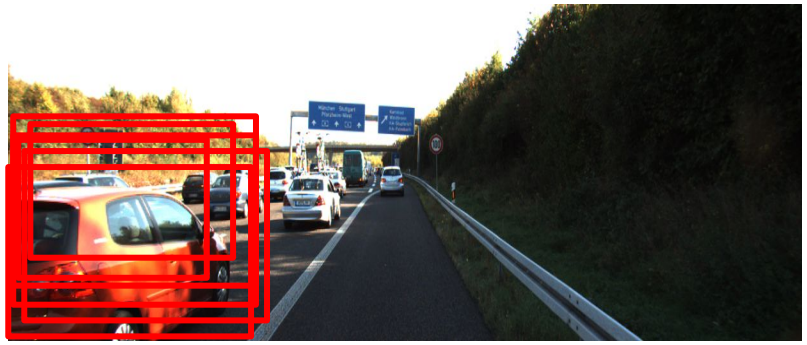
RP: Region Proposals

All RPs



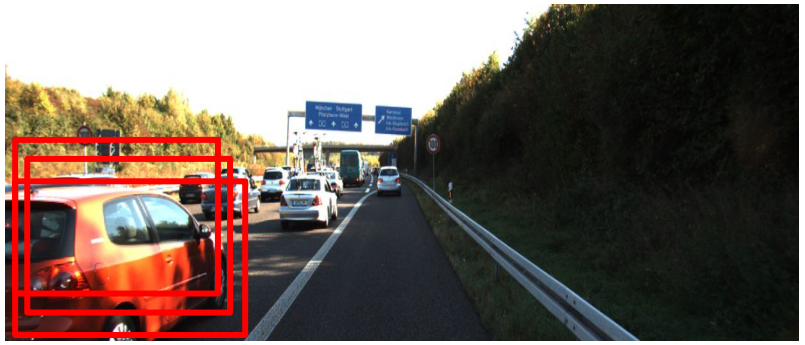
% of RPs **100%**

Top 4



~4%

Top 1



~1%

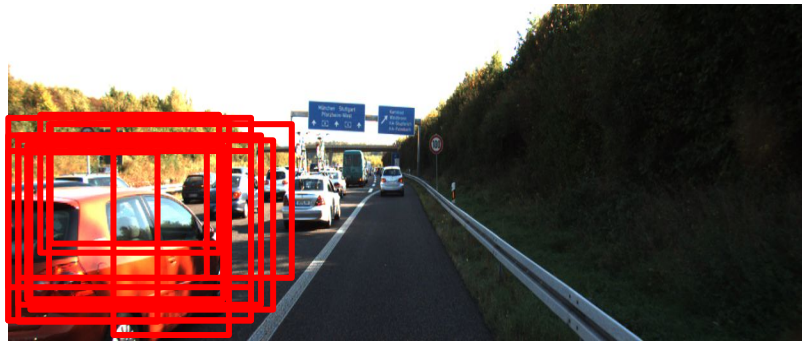
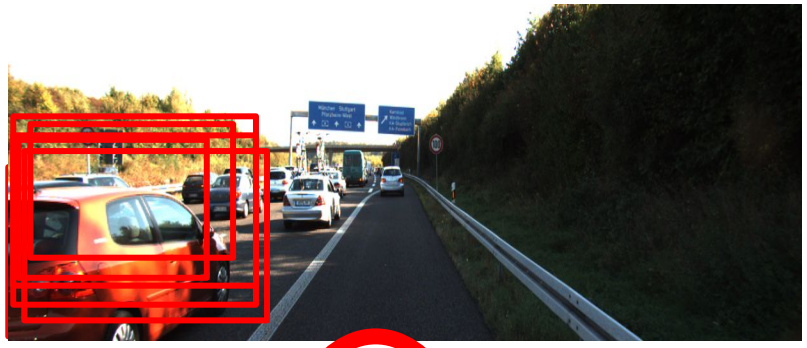
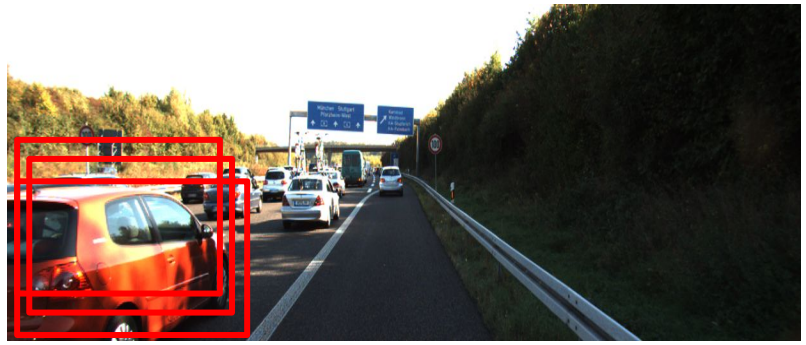
Training		All RPs					Random RPs(OM, SM, All)				
Testing		OM1	SM1	OM4	SM4	All	OM1	SM1	OM4	SM4	All
Car	E	90.84	90.84	90.84	90.84	90.84	90.83	90.85	90.83	90.84	90.83
	M	90.63	90.64	90.64	90.64	90.64	90.62	90.65	90.64	90.67	90.65
	H	88.40	88.91	89.20	89.44	89.25	87.92	88.76	89.23	89.42	89.23
	mAP	81.19	81.26	81.28	81.31	81.27	81.21	81.27	81.27	87.03	81.27
Pedestrian	E	67.93	68.09	72.76	74.08	73.06	68.32	68.64	73.88	74.95	74.36
	M	57.50	58.28	58.78	61.18	58.78	57.69	58.85	59.39	62.51	59.37
	H	50.08	50.44	50.73	55.41	50.62	50.37	50.80	50.80	55.67	51.22
	mAP	49.48	49.95	50.43	52.93	50.33	49.67	50.50	50.50	54.02	50.83
% of RPs		0.86%	1.10%	3.43%	4.41%	100%	0.87%	1.11%	3.48%	4.44%	100%

Results reported on KITTI 2D object detection dataset.
OM is filtered with Objective Map.
SM is filtered with Subjective Map.
1 or **4** represents selected top 1 or 4 RPs with highest attention scores.

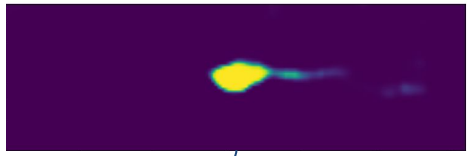
Insight
SFI RESEARCH CENTRE FOR DATA ANALYTICS

Visual Attention in Detection

RP: Region Proposals

	All RPs	Top 4	Top 1
			
% of RPs	100%	~4%	~1%
Car (mAP)	81.27%	87.03%	81.27%
Pedestrian (mAP)	50.83%	54.02%	50.50%

Visual Attention for Tracking



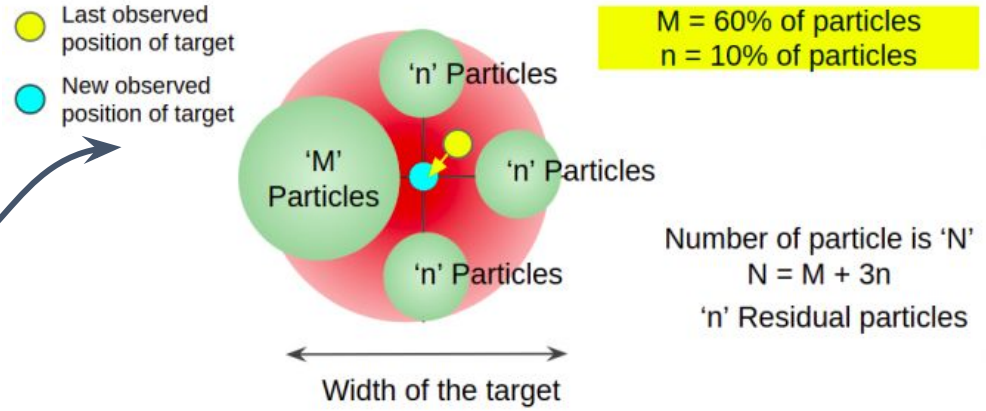
Use saliency for distribution of particles around the object for estimating the new position of the object

Object Tracking:

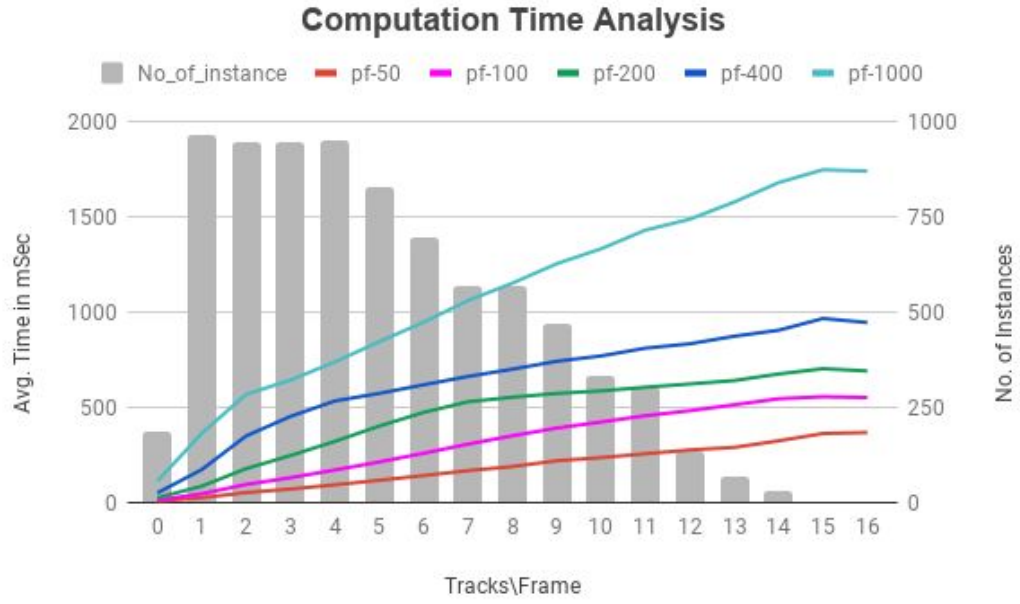
- **Input:**
  - Object Position [x,y,w,h]
  - Input Frame
  - Saliency map
- **Output:**
  - Object Position with ID

Particle-PHD filter Highlights

- **Initialisation**: Use of multi-peak Gaussian distribution by a number of particles with randomised weights
- **State Prediction and weight computation:** based on the weighted IoU, distance metric computed on the temporal histogram. particle with maximum weight is considered as the final state.
- **Particle Resampling:** based on the motion cues along with Kalman Gain
- **Refining and update using visual attention cues:** Correction mechanism will guide validated particles in the particle-PHD filter to converge.



Choosing number of particles for tracking

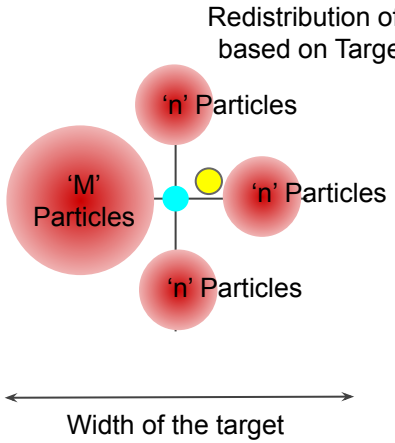


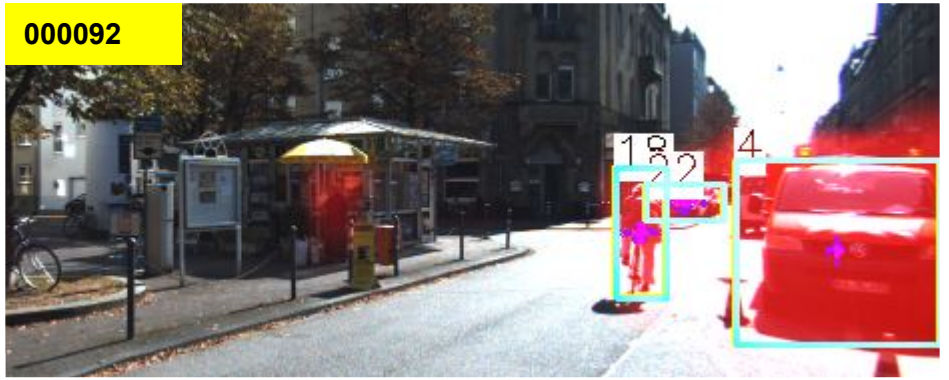
Insight
SFI RESEARCH CENTRE FOR DATA ANALYTICS

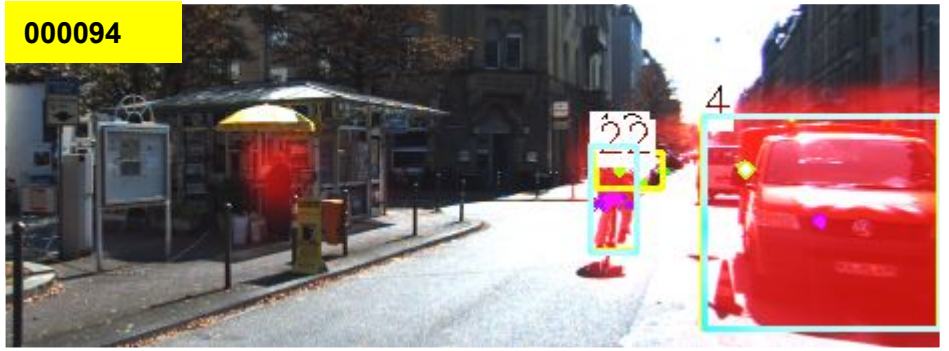
KITTI Sequence : 0000

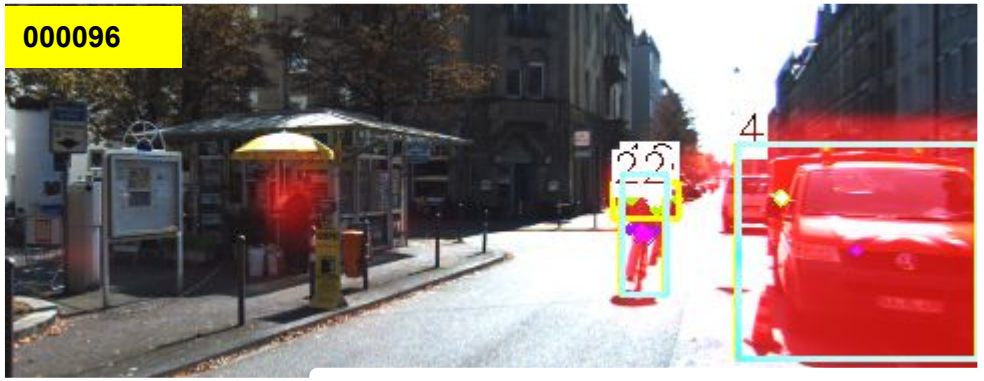
Why ?

Why Visual attention is helpful ?







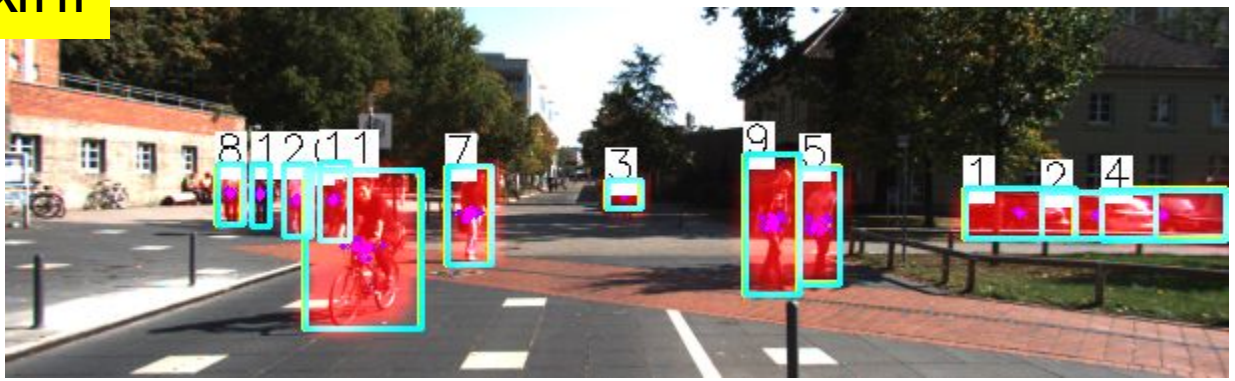


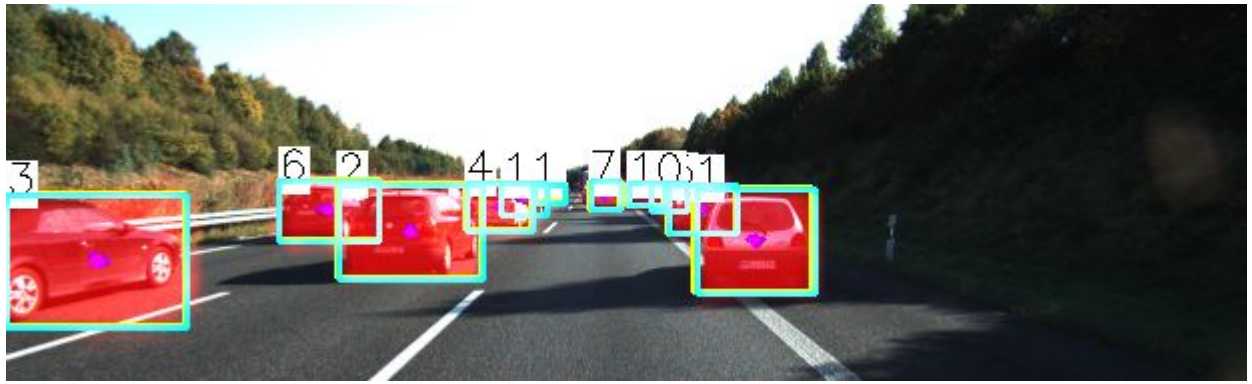
Insight
SFI RESEARCH CENTRE FOR DATA ANALYTICS

Multiple Target Tracking on KITTI & DeTRAC Dataset

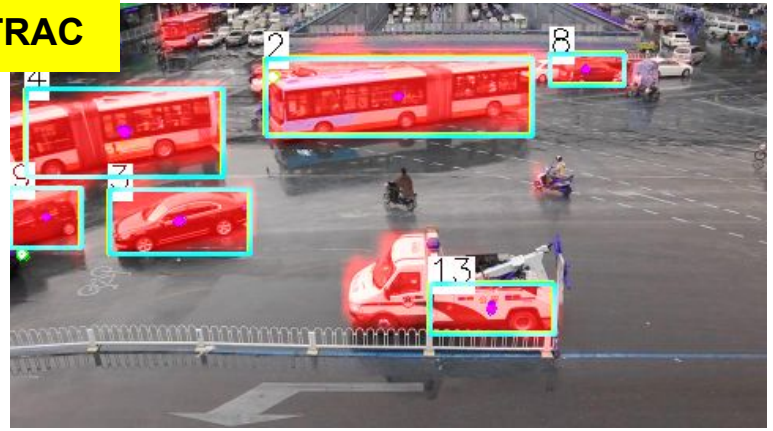
using subjective hierarchy model with objectness mask

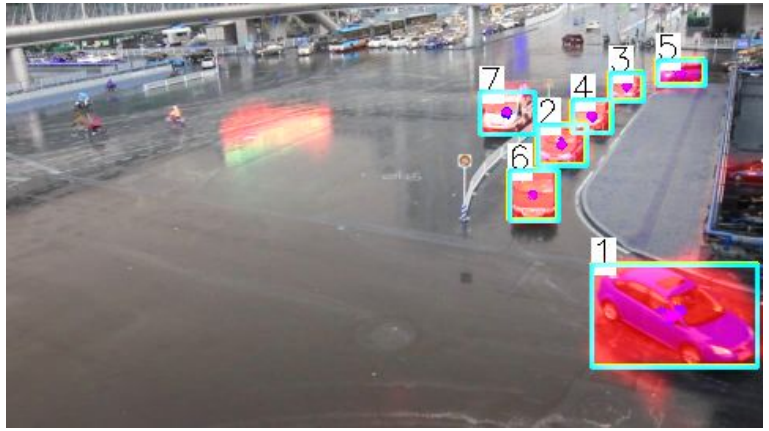
KITTI

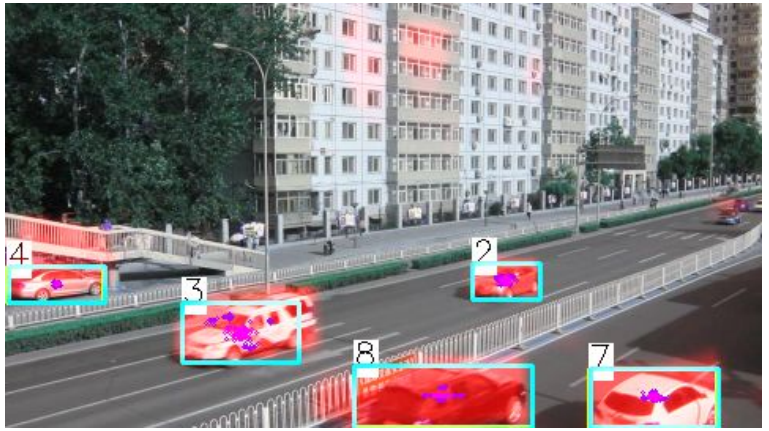




DeTRAC







Multiple Target Tracking Accuracy on KITTI Dataset

For Car & Pedestrian with Number of Particles=100,
OM: Objectness Map

Class	Method	MOTA	MOTP	Rcll	Prcn	F1	FAR	MT	PT	ML	IDs	FM
Car	Baseline	79.13	80.69	85.31	85.31	90.91	**8.25**	65.43	29.79	4.79	**244**	655
	mh-sub-OM	**84.82**	81.88	90.94	**96.79**	**93.77**	10.52	77.13	19.86	**3.01**	268	553
Pedestrian	Baseline	58.56	75.05	65.97	92.17	76.90	7.85	35.93	54.49	9.58	**147**	547
	mh-sub-OM	**64.49**	76.93	**71.81**	92.88	**80.99**	7.70	**48.50**	**44.91**	6.59	161	516

Detection and Tracking Accuracy on DETRAC Dataset

For Car & Pedestrian with Number of Particles=100,
SM: Subjective Map, OM: Objectness Map

PERFORMANCE ON DETRAC USING MODEL WITH HIERARCHICAL FEATURES TRAINED WITH ALL REGION PROPOSALS. TESTED WITH ALL AND ATTENTION FILTERED RPS. ALL* IS RESULTS ACHIEVED USING OFFICIAL TESTING DATA AND CODE WITH ALL RPS.

Detection

	All	E	M	H	Cloudy	Night	Rainy	Sunny	RPs
All*	71.50	90.23	77.62	57.81	81.39	69.19	58.19	85.55	100%
All	70.98	90.02	79.30	60.00	80.03	70.99	60.69	80.74	100%
OM1	70.98	90.02	79.25	53.12	80.04	70.94	60.64	80.75	1.59%
SM1	70.89	89.84	79.13	59.81	80.04	71.06	60.21	80.73	1.10%
OM4	70.97	90.01	79.22	59.41	80.05	70.79	60.65	80.75	6.36%
SM4	70.86	89.79	79.02	59.56	80.03	70.88	60.21	80.71	4.39%

Tracking

using subjective hierarchy model with objectness mask

Method	PR-MOTA	PR-MOTP	PR-MT	PR-ML	PR-IDs	PR-FRAG	PR-FP	PR-FN
frcnn+6thAI*	**30.7**	37.4	28.7	23.2	**143.3**	1183.1	13387.9	195193.9
mh-sub-OM	24.85	36.67	28.99	**6.59**	1583	3054	52286	**105165**

Conclusion

- Utilisation of visual attention cues can improve the processing efficiency of object detection and performance of tracking module
- Use of **knowledge distillation** to train a detector that can simultaneously generate **objectness and saliency maps** using joint image representation
- **10% of the total area of features maps** are contributing to the detection of objects. The obtained results shows similar performance is achieved using only with **~1% of RPs** by choosing the RPs having highest objectness score from each feature map.
- Modified particle-probability hypothesis density (PHD) filter is explored utilising the visual attention map during **particle resampling, distribution and update** process while tracking.


THANK
YOU