Humans vs. AI in Detecting Vehicles and Humans in Driving Scenarios

Alice Yumeng Yang, Guoyang Liu, Yunke Chen, Ruoxi Qi, Jindi Zhang, Janet H. Hsiao

Purpose

Examining the similarities and differences between humans and AI (Yolo-v5s) in performance and attended features in detecting vehicles and humans in driving scenarios

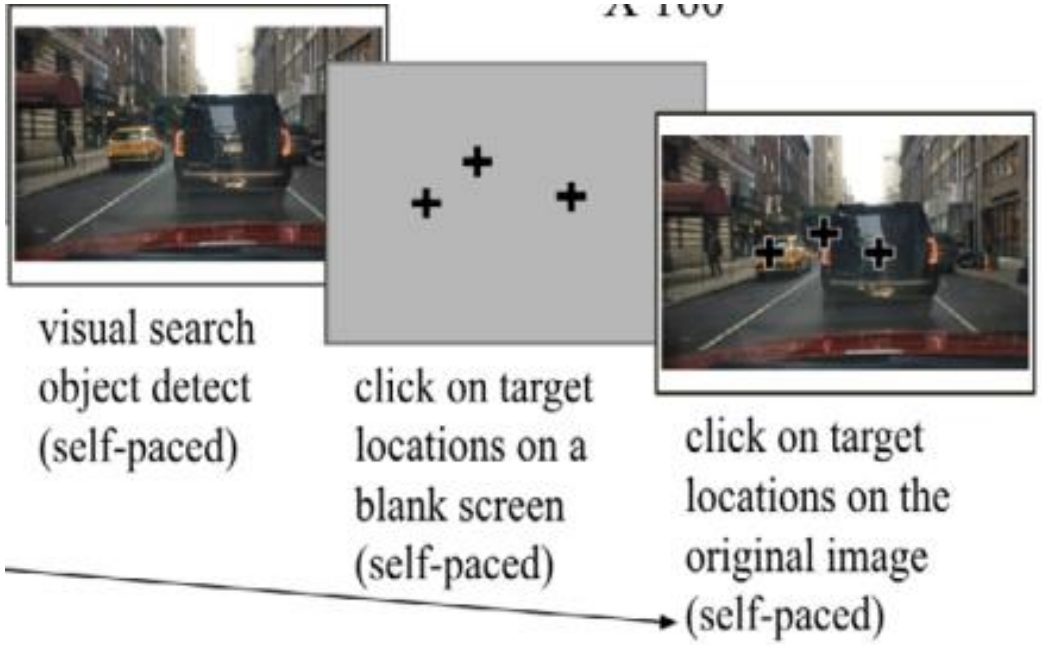
Yolo-v5s: a real-time object detector

Study 1: Humans vs. AI in Detecting Vehicles

Identifying representative attention strategies from humans, using a data-driven method, EMHMM (Eye Movement analysis with Hidden Markov Models) with co-clustering

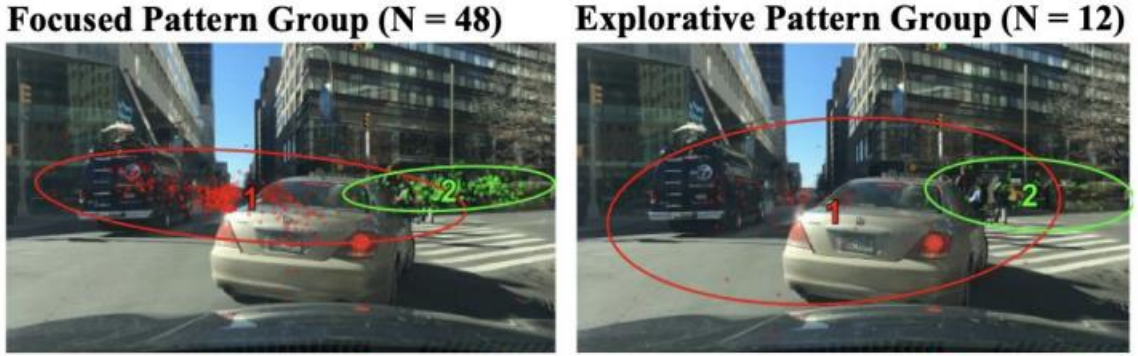
Comparing the strategies with AI

1. detect all vehicle targets in the image and press a key as soon as they thought they had detected all. Their eye movements during the visual search before the key press were used for data analysis
2. immediately after the key press, participants were asked to use a mouse click to place a marker at each detected target location on a blank screen
3. click again on the same objects they had clicked previously on the original image to confirm their selection



- Focused Pattern Group: scan the horizon where relevant targets typically occur
- Explorative Pattern Group: scan a broader area
- Performance was assessed by hit rate: the number of correctly detected targets divided by the total number of targets
- the focused strategy was associated with better performance
- Yolo-v5s' attended features had higher similarity to the focused group than the explorative group

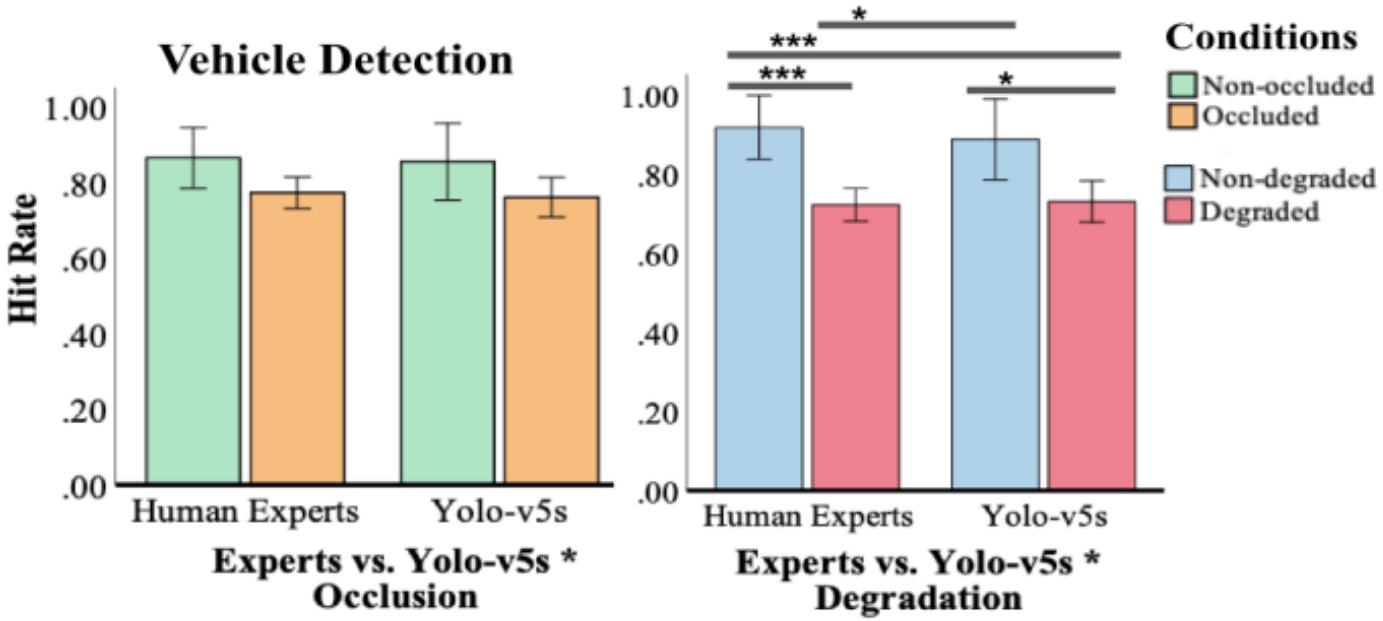
Priors: probabilities of the first fixation landing on each Region of Interest (ROI)
Transition matrices: transition probabilities among the ROIs



Group A	To R	To G
Priors	.96	.04
From Red	.90	.10
From Green	.30	.70

Group B	To R	To G
Priors	.99	.01
From Red	.96	.04
From Green	.40	.60

- Images with any target hided by any other object or the image boundary were rated as ‘occluded.’
- Images containing any target whose feature identification was influenced by night vision, motion blur, uneven illumination, shadow, or light reflection were rated as ‘degraded
- (compared with the focused group)



Study 2: Humans vs. AI in Detecting Humans

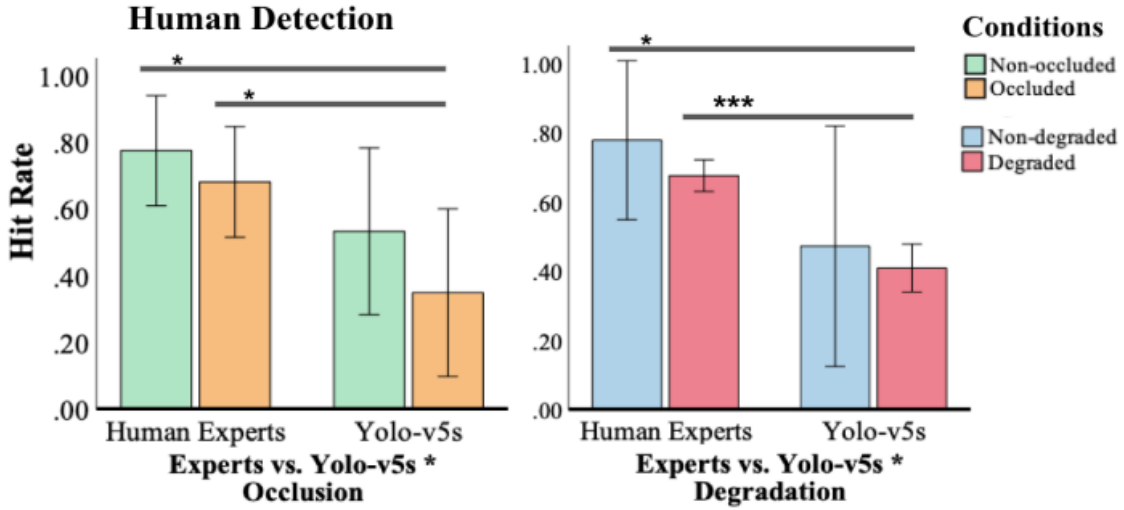
Human Detection

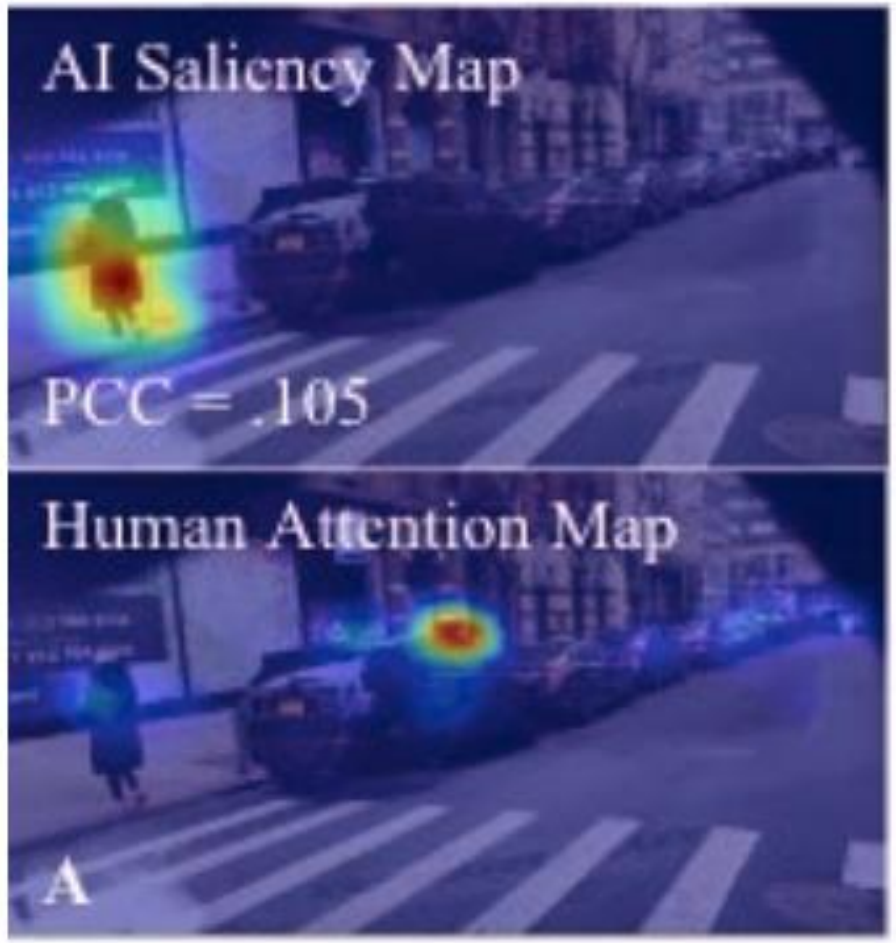
Focused Pattern Group (N = 48)

Group A	To R	To G
Priors	1.0	.00
From Red	.97	.03
From Green	.48	.52

Explorative Pattern Group (N = 12)

Group B	To R	To G
Priors	1.0	.00
From Red	.98	.02
From Green	.65	.35





Conclusion

higher similarity in attended features to the focused than the explorative strategies, and achieved a similar performance level to human experts in vehicle detection

➔human attention could be used to guide AI design