Examining how aided target recognition (AiTR) with varying reliabilities affects visual search performance

Alex Ridolfi

Mentored by Dr. Chloe Callahan-Flintoft

Introduction

Humans are constantly scanning the world around them, using their eyes and the subconscious process of visual search to gather information or identify a target. But what if the items they were searching for were highlighted for them? Would it be detrimental to them if the wrong item was highlighted? Being able to accurately identify friendlies and enemies is paramount in battle and so any advantage computers can provide should be seized. However, if the computer makes a mistake, the life of a friendly soldier could be at stake. Findings show radiologists to be less likely to discover lesions if they were not marked by a computer-aided detection (CAD) system (Drew et al., 2012). In contrast, when cued to a certain location, subjects were able to detect a subsequent target at that location faster than without any cues (Wolfe et al., 2021). Identifying the level of accuracy and reliability needed for effective computer-aided detection is one of the most vital aspects of implementing these systems. Further understanding the process of visual search overall will help make the development of these computer systems more targeted and effective.

The purpose of this project is to find out how aided target recognition systems (AiTR) with different accuracy levels affect the mental and optical behavioral performance results of a visual search task.

Materials and Methods

Designing and running this experiment was broken down into three steps: deciding what strength AiTR systems to simulate, writing and designing the experiment in PsychoPy 2022.2.4, and running the experiment on subjects at Aberdeen High School (AHS) ($n = 24$) and Army Research Lab (ARL) ($n = 6$). Ultimately, the AiTR accuracies chosen were 85% and 55%, with a control condition with no AiTR, seen in Figure 1. This means that in their respective blocks, the simulation will highlight the correct soldier for that percentage of the trials. Parameters that were chosen for the experiment include a two-second time limit per trial, 56 soldiers per trial image, and 42 trials per block. Subjects were also given an example trial before the test to become familiar with the task, in Figure 2.

Figure 1 (right): The flow of how the experiment chose each image it used. The percentages indicate how often that event occurred.
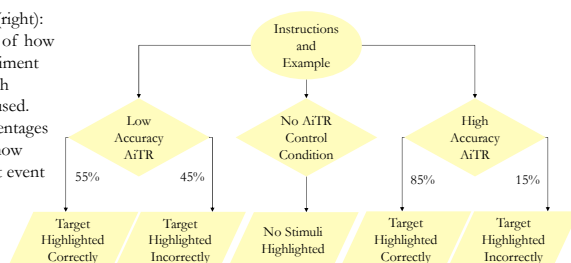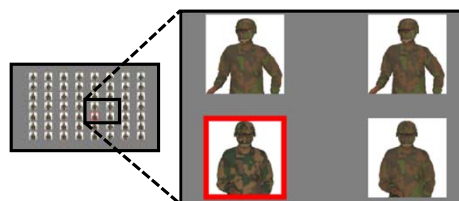

Materials and Methods (continued)

Figure 2 (left): A scene of soldiers that a participant in the study may see. In this case, the AiTR system is correct, highlighting the enemy amongst friendlies.

For each subject, the order in which the AiTR reliabilities were tested was randomized. The control block was always set as the second condition tested to act as a mental reset in system confidence.

Thirty subjects were tested in total, 24 being at AHS between the ages 17–18, and six were tested at ARL with the addition of a Tobii Eye-Tracking system. For the subjects tested at AHS, two main variables were measured by the PsychoPy program, reaction time and click location. At ARL, these two variables were once again measured, but eye-tracking data was recorded as well. All potentially confounding variables were held constant throughout the tests such as the use of a mouse and mousepad, screen brightness, and the instructions of the task.

Results

A one-way ANOVA test was done on all the dependent variables that were measured by the experiment. Each subject experienced all three conditions.
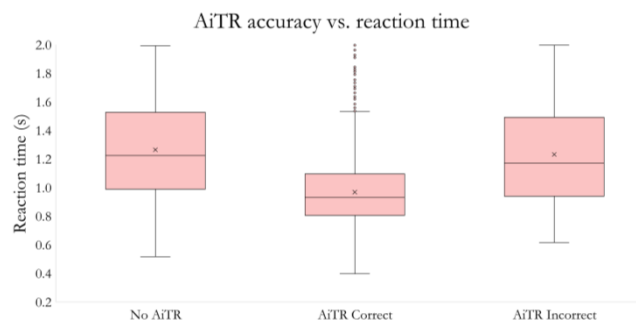
Graph 1 (above): A box and whisker plot representing the mean reaction times for each subject for each of the three AiTR accuracies. There was an extremely significant difference in reaction time (seconds) when the AiTR was correct ($M = 0.969$, $SD = 0.239$) compared to when the AiTR was incorrect ($M = 1.233$, $SD = 0.352$) and when there was no AiTR ($M = 1.266$, $SD = 0.353$), $F(2, 2882) = 332.57$, $p < .001$.

Results (continued)

There was also a significant difference in click accuracy when the AiTR was correct ($M = 0.885$, $SD = 0.319$) versus when the AiTR was incorrect ($M = 0.548$, $SD = 0.498$) and no AiTR ($M = 0.598$, $SD = 0.490$), $F(2, 3567) = 228.61$, $p < .001$. Finally, there was a significant difference in the number of fixations within the 85% accurate AiTR condition ($M = 3.53$, $SD = 2.90$) versus the 55% accurate AiTR condition ($M = 5.37$, $SD = 4.33$) and the no AiTR condition ($M = 6.68$, $SD = 4.83$), $F(2, 663) = 30.02$, $p < .001$.

Figure 3 (right): A heat map depicting the most common fixation points for this image over all trials where the AiTR was incorrect. Higher concentrations of fixations are depicted closer to a pinker color.

Conclusion

The purpose of this project was to learn about how AiTR accuracy can affect visual search performance. The subject's performance was extremely sensitive to the AiTR's accuracy. When the AiTR is correct, there is a jump in performance, indicated by a shorter mean reaction time, seen in Graph 1, higher mean accuracy, and a lower mean number of fixations compared to the no AiTR condition. In contrast, when the AiTR was incorrect, there was a decrease in accuracy and an increase in the number of fixations when compared to the no AiTR condition. Figure 3 depicts how an incorrect AiTR can cause an increase in fixations around a distractor item.

Ultimately, there is a direct relationship between AiTR accuracy and search performance, which must be considered when developing these technologies. This project suggests that there can be major consequences if these systems are incorrect for soldiers, as well as radiology imaging and security scans, but incredibly beneficial when it is correct. The development of these technologies for such purposes must be highly accurate and eliminate the chance of system failure to prevent missed targets. Further testing can be done to validate these results and expand the theme into the other fields mentioned above.

References

Drew, T., Cunningham, C., & Wolfe, J. M. (2012). When and why might a computer-aided detection (CAD) system interfere with visual search? An eye-tracking study. *Academic Radiology, 19*(10), 1260–1267. https://doi.org/10.1016/j.acra.2012.05.013

Wolfe, B., Kosovicheva, A., Stent, S., & Rosenholtz, R. (2021). Effects of temporal and spatiotemporal cues on detection of dynamic road hazards. *Cognitive Research: Principles and Implications, 6*(80), 1–15. https://doi.org/10.1186/s41235-021-00348-4