



Humans and Neural Networks Learn to Use Background Structures for Efficient Visual Search

VIU
vision • image
understanding

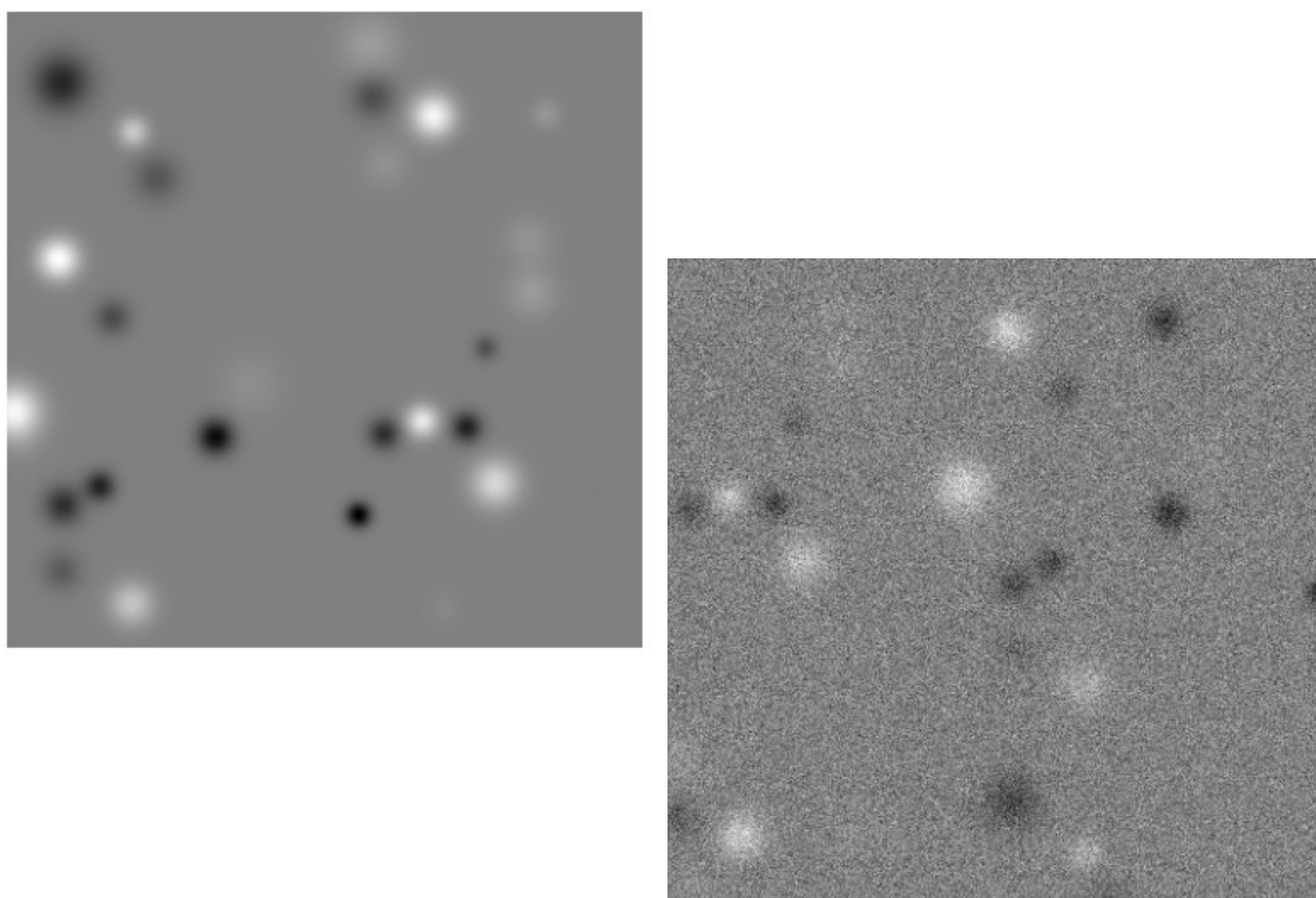
Tanish Mendki (tmendki@ucsb.edu) and Miguel Eckstein
Dept. of Psychological and Brain Sciences, UCSB

1. Background and Introduction

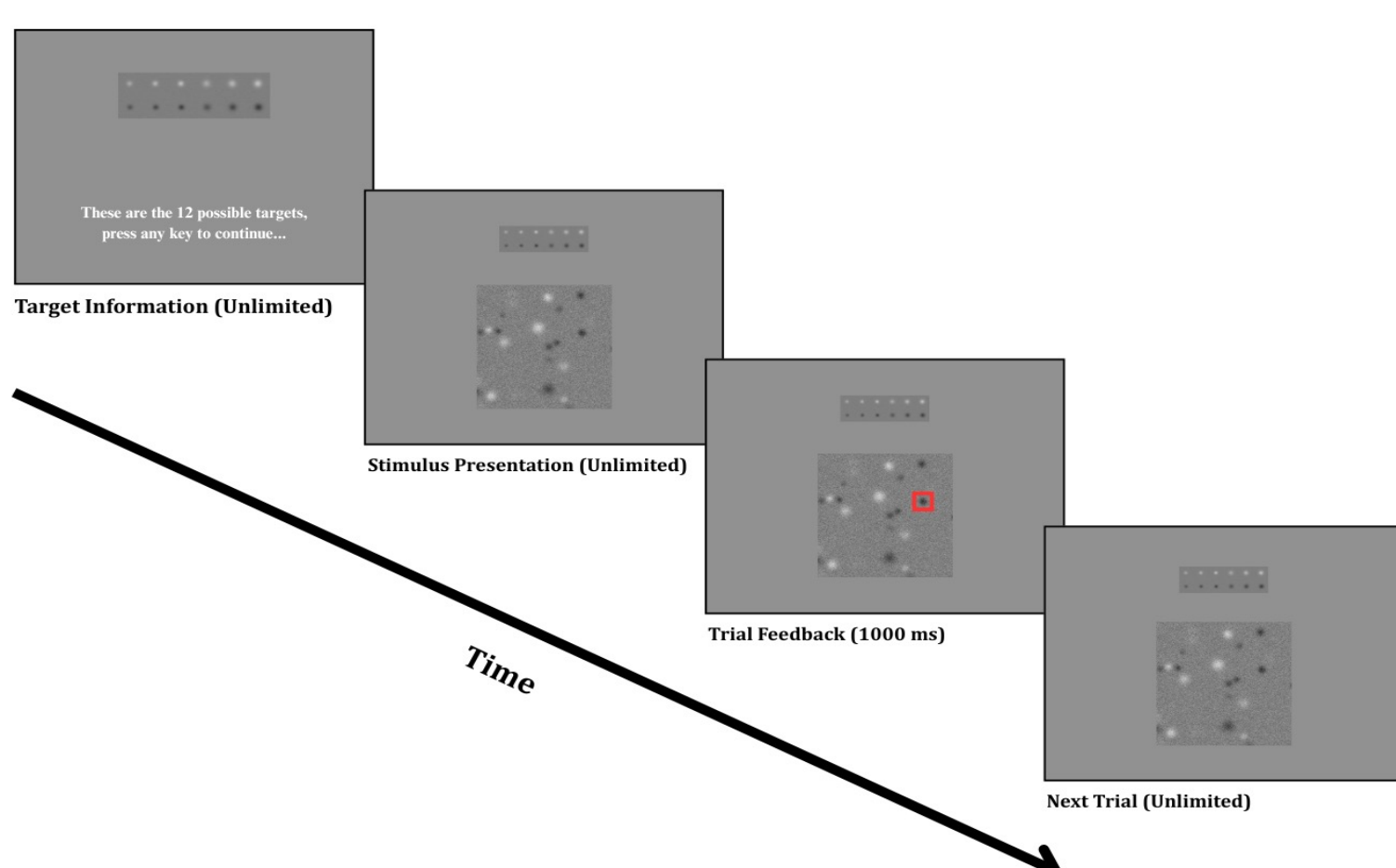
- The ability to visually find targets in any given scene is limited by the random variability in the environment/background that can make the target confusable with any distractor. However, environments can also contain background structures that are highly predictable and detectable despite containing some random variability, such as in the case of medical images.
- In cases such as medical imaging, these random background structures must be detected yet efficiently discounted in search of targets. Unlike classically considered noise processes, these distractors need to be learned in order to efficiently complete the task.
- Here, we use both human data along with behavioral signatures of a Convolutional Neural Network (CNN) to show how such learning might take place and inform future studies on medical imaging and search involving dependable distractors.

2. Materials and Methods

Stimuli: 400×400 pixel images containing 30 distractors (2-D gaussians, varying in contrast and standard deviation) to form a structured background alongside a randomly placed target (2-D gaussians). Distractors maintain their relative spatial positions to each other but are all rigidly shifted with wrap-around, such that the overall structure is preserved.

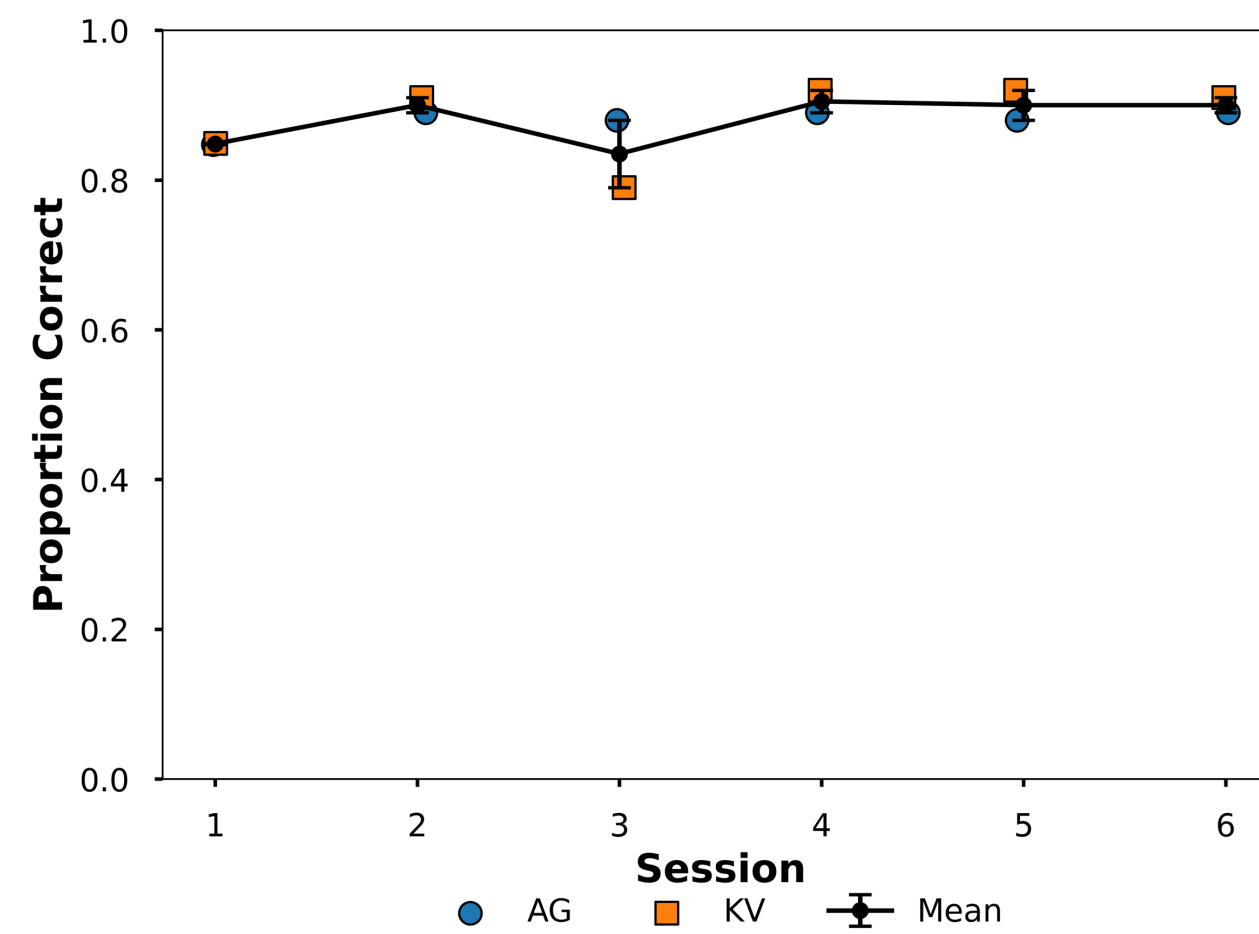


Experiment Paradigm: Two observers completed three sessions of 100 trials per day across two consecutive days. They first performed a target-in-white-noise search, followed by a background-plus-noise search. On each trial, a test image with a randomly sampled target was shown until the observer selected a location by clicking. A higher contrast reference image of possible targets was displayed above. Responses within 0.376° of the target center were considered correct. Feedback was provided with a red square marking the correct location.



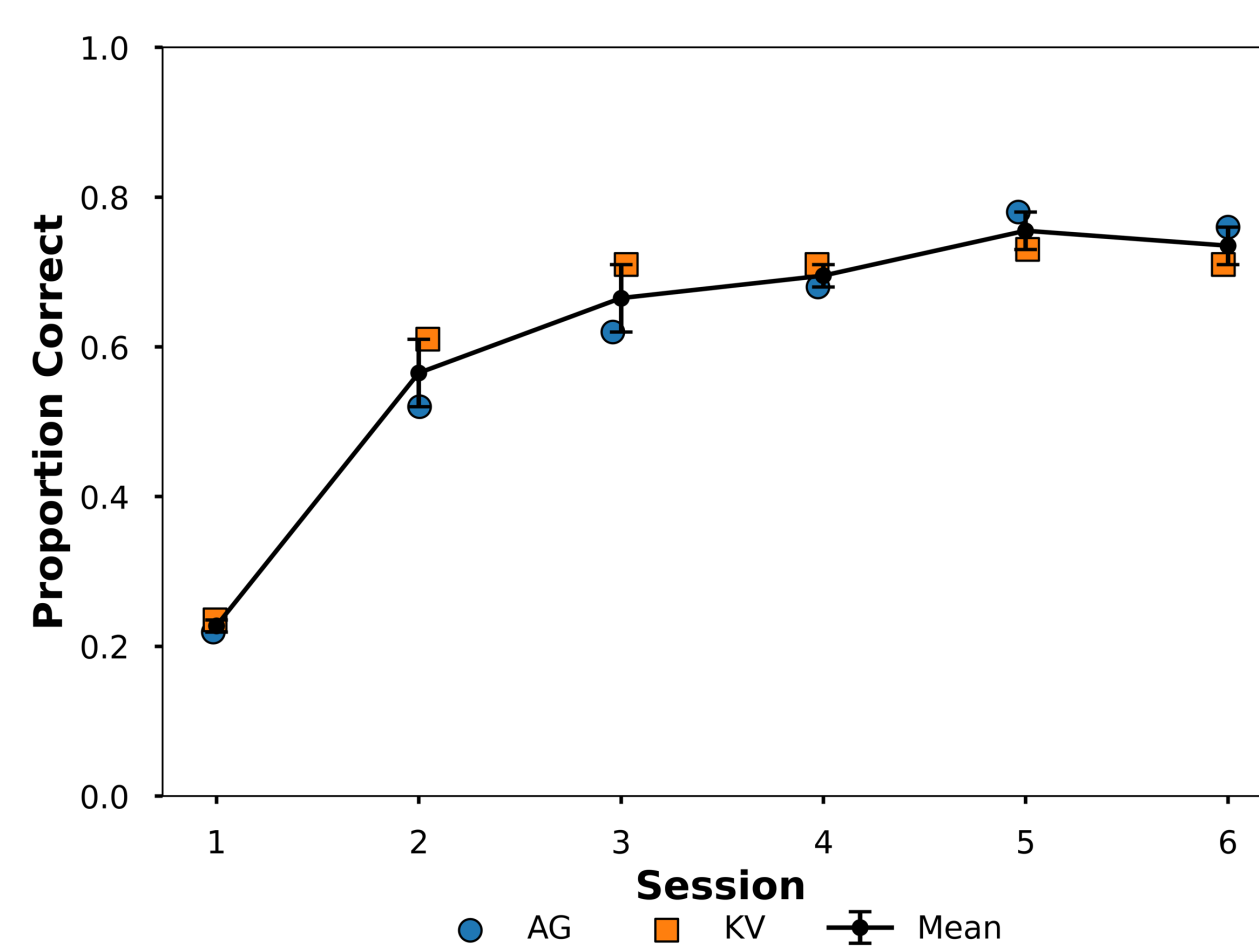
3. Results

Target Only Trials



Participants' accuracy during the *target-in-white-noise-only* search task remained high across sessions. Both observers completed three sessions per day over two consecutive days. While a modest drop in performance was observed during the third session on Day 1, accuracy recovered by Day 2 and remained stable thereafter. Overall, performance was consistent, indicating rapid learning or stable task engagement after initial exposure.

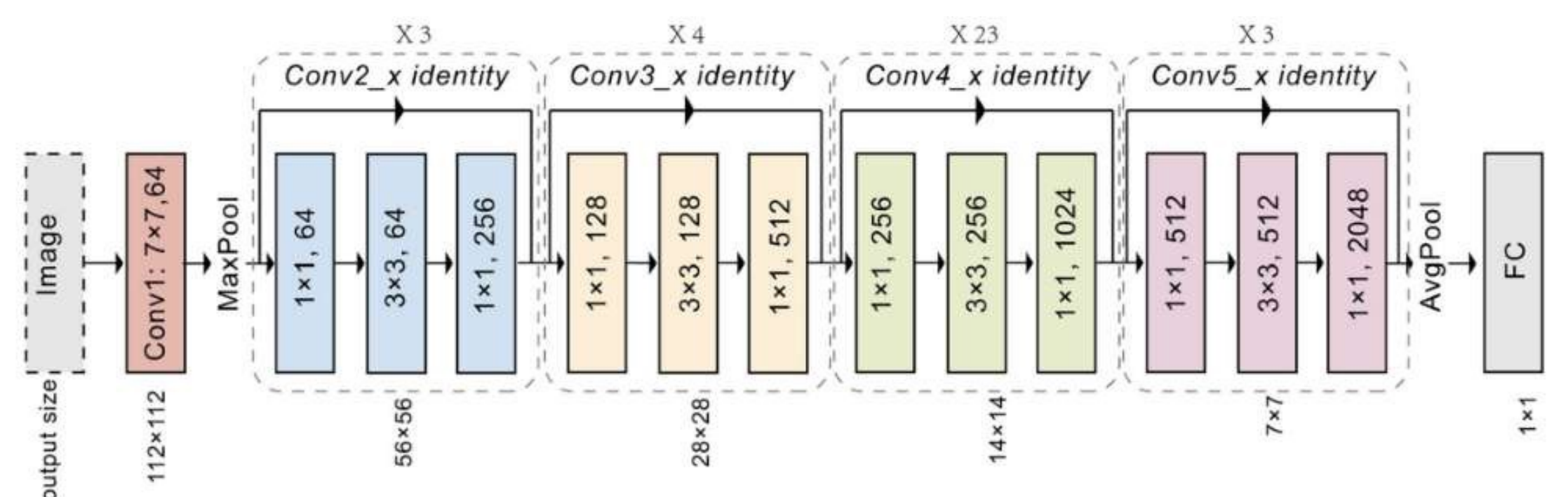
Background + Target Trials



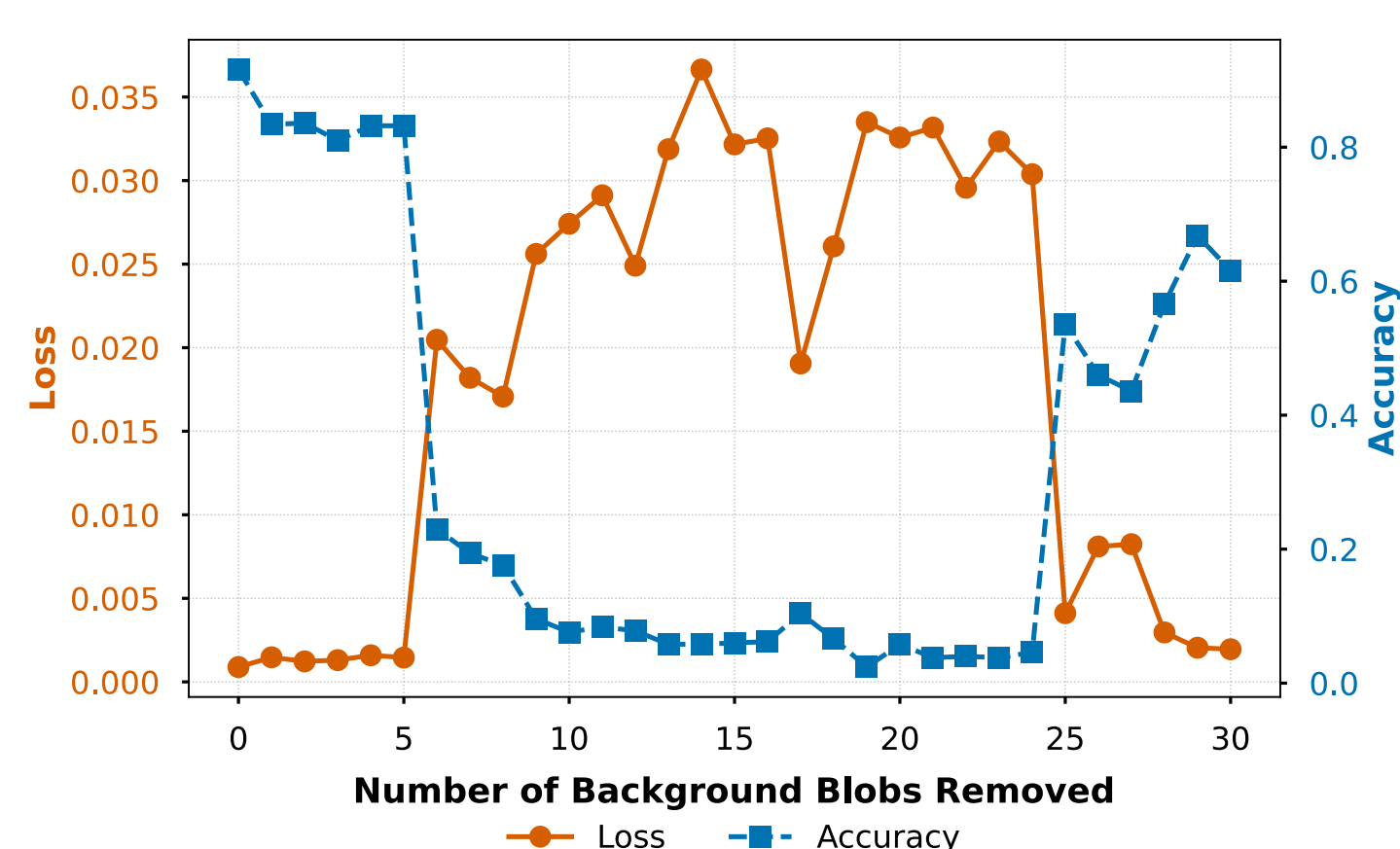
In the *background plus target in white noise* search task, participants showed clear learning across sessions. Accuracy started low on Day 1 but steadily improved within and across days, with performance approaching an asymptote of 75% by the final sessions on Day 2. This pattern suggests observers adapted to the background, showing substantial gains in accuracy with practice.

4. Convolutional Neural Network

Model Architecture



Effects of Background Structure Removal



Removing background structure from images affected target detection accuracy in the CNN trained with background-present stimuli. Accuracy dropped sharply when background blobs were removed, indicating that the network relied on structured background information to locate the target. As more background was removed, accuracy eventually recovered, suggesting that once background structure was minimal, the task became easier, and the network relied more on target features alone.

5. Conclusions

- Background structures, even when initially treated as distractors, can become reliable cues that facilitate target localization. Human observers and CNNs both demonstrated evidence of learning to incorporate structured background information to improve search performance.
- In CNNs, removing these structured backgrounds disrupted performance, confirming their role as learned, informative features rather than mere noise. These findings challenge classical views of background variability as purely detrimental, highlighting instead how dependable distractors can support efficient search.
- Analyzing CNN behavior in this context offers a promising path toward understanding neural mechanisms underlying expert visual search, such as in medical image reading. Future work training networks on varied and complex environments may help reveal how the brain learns to balance the detection and discounting of structured background information during challenging perceptual tasks.