AMODAL COMPLETION OF OCCLUDED OBJECTS IN DEEP NEURAL NETWORKS: A PSYCHOPHYSICAL INVESTIGATION

Arnav Shakya & Vaibhav Tripathi *
Department of Cognitive and Brain Sciences
Indian Institute of Technology Gandhinagar
Gandhinagar, 382355, India
{arnav.shakya,vaibhav.tripathi}@iitgn.ac.in

ABSTRACT

Humans have a remarkable ability to complete the shape of an occluded object in a visual scene (Rensink & Enns, 1998). Standard feed-forward architectures of computer vision, however, suffer from diminished performance in classification tasks when objects are partially occluded, whereas humans perform well (Tang et al., 2018). This difference in the ability to classify an occluded object points to different internal mechanisms involved in both visual systems. We do not know the extent of amodal completion possessed by these models, or how human-like this ability is. We tested a neuroscience-inspired visual search model with a VGG16 backbone on four psychophysical tasks involving amodal completion, which are standard in human studies. We found that both a traditional feed-forward model trained on ImageNet, or a fine-tuned model with a recurrent layer, are *not* able to automatically complete occluded objects, suggesting that bridging the gap to natural vision requires more sophisticated mechanisms for extracting and integrating global structural features.

1 INTRODUCTION

When the sun sets behind a hill, we perceive it as full round sphere, not a sliced semi-circle. This phenomenon is known as amodal completion (AC)—the ability of a visual system to fill-in the missing information of an occluded object. It is not merely the perception of occluded parts, but a process of extrapolation or interpolation that can modify the representation of visible (modal) parts of an object (Gerbino, 2020). The human visual system is remarkably adept at AC of occluded objects at a very pre-attentive level (Rensink & Enns, 1998). Furthermore, AC can actively enhance attentional guidance when the search environment provides a consistent context that allows the visual system to form a symmetric, globally completed "target template" in working memory (Chen et al., 2018; Cherian & Arun, 2024).

In the domain of computer vision, Deep Neural Networks (DNNs) have successfully modeled human performance in core object recognition tasks (Serre, 2019; Rajaei et al., 2019). However, it is unknown whether these models also possess an internal mechanism to complete amodal information. Research in this area has been constrained by methodological difficulties, often limiting the scope of investigation. For example, Tang et al. (2018) compared pattern completion in DNNs and humans. However, their study defined completion simply as the ability to classify partial images. Consequently, their findings only reflect robustness of classification in spite of occlusion rather than the presence of a filling-in mechanism.

Several similar studies have tested the recognition performance of DNNs and humans under conditions of occlusion and extreme occlusion (Rajaei et al., 2019; Zhu et al., 2019). It was found that introducing recurrent processing in classical feed-forward models increases performance in recognition tasks (Kang et al., 2020; Rajaei et al., 2019; Tang et al., 2018). Beyond categorization

*Integrative Neuroscience and Technology Lab, IIT Gandhinagar. https://labs.iitgn.ac.in/int/

performance, recent work comparing different architectures found that bio-inspired recurrence, such as predictive feedback and lateral surround suppression, can implement significantly different internal computational strategies than standard task-trained recurrent networks when classifying noisy or degraded images (Lindsay et al., 2022).

A study by Cherian & Arun (2024) compared human amodal completion with deep neural networks. They measured reaction times of humans in visual search tasks with various occluded objects to obtain perceptual distances due to occlusion. These were then compared with the representational distances using feature vectors of those same shapes in the internal layers of pre-trained DNNs. However, the employed method did not allow for a direct comparison of human and DNN performance and mechanisms, as no common psychophysical visual search tasks were incorporated for DNNs.

Our study addresses this limitation by employing psychophysical tasks to rigorously characterize amodal completion in DNNs. We investigate three core questions: first, whether DNNs exhibit the basic effect of amodal completion on psychophysical tasks, and how this compares to human performance; second, what underlying mechanisms drive amodal completion in DNNs, specifically whether it operates through local or global completion; and third, whether recurrent computations are sufficient to give rise to amodal completion ability in DNNs.

2 METHODOLOGY

2.1 MODEL ARCHITECTURE AND IMPLEMENTATION

We utilize the eccNET model following Gupta et al. (2021). The model is designed to take a target image (I_t) and a search image (I_s) as input and generate a sequence of fixations until the target is located. Hence, it allows us to conduct the visual search based psychophysical experiments. Model's architecture is built upon a VGG16 convolutional neural network (CNN) Simonyan & Zisserman (2014), pretrained on the ImageNet classification task Russakovsky et al. (2015); Deng et al. (2009), which serves as a proxy for the primate ventral visual system. Two key biologically-inspired modifications were incorporated into the standard CNN architecture in eccNET:

1. **Eccentricity-Dependent Pooling:** To simulate the difference between foveal (the center of gaze) and peripheral vision, all standard max-pooling layers in the VGG16 architecture were replaced with custom eccentricity-dependent pooling layers. In these layers, the size of the receptive field for each unit increases linearly as a function of its distance from the current point of fixation. This results in high-resolution processing at the fovea and progressively lower-resolution processing in the periphery, mimicking properties of the primate visual system (Freeman & Simoncelli, 2011). The model was not retrained after this modification; original VGG16 weights were preserved to test generalization from object recognition to visual search.
2. **Multi-Layer Top-Down Attention:** To guide search in a goal-directed manner, the model implements a top-down attentional mechanism. At each fixation, feature maps are extracted from both the target and search images at multiple hierarchical levels (layers 9, 13, and 17). The target's feature map at a higher layer is used as a convolutional kernel to modulate the feature map of the search image from the preceding layer, creating a layer-specific attention map. An overall attention map is computed as a weighted combination of these individual maps. A winner-takes-all mechanism selects the location with the highest activation as the locus for the next fixation (Lee et al., 1999; Zhang et al., 2018), while an inhibition-of-return mechanism prevents revisiting locations.

2.2 INTRODUCTION OF RECURRENT NEURAL NETWORK (RNN) LAYER

To investigate whether recurrent computations are involved in amodal completion, we later extended the base eccNET architecture by replacing the standard fully connected classification head with a Recurrent Neural Network (RNN). As shown by (Tang et al., 2018; Kang et al., 2020; Rajaei et al., 2019; Zhu et al., 2019), the introduction of an RNN in a standard feedforward model can make its classification robust to occlusion, hence we hypothesized that it may also help in AC.

2.2.1 ARCHITECTURE MODIFICATION

The output of the final eccentricity pooling block (Block 5) was flattened and passed to the recurrent layer. We utilized a SimpleRNN layer with the following configuration:

- **Input Structure:** The flattened feature vector from the convolutional backbone is repeated T times using a RepeatVector layer to simulate a sequence of visual inputs, where T represents the number of recurrent steps (set to $T = 5$).
- **Recurrent Layer:** A SimpleRNN layer with 4,096 units and $\tanh$ activation was employed. This layer processes the input sequence, updating its hidden state at each time step to encode historical context from the feature maps.

2.2.2 TRAINING STRATEGY

The training process was divided into two distinct phases to ensure stable convergence:

Phase 1 (Feature Frozen): The VGG16 backbone layers were frozen to retain the pre-trained ImageNet features. Only the newly added RNN layer and the fully connected heads were trained on ImageNet dataset for 10 epochs (Deng et al., 2009; Russakovsky et al., 2015). This phase allowed the RNN to learn how to process the spatial features without disrupting the underlying feature extractors.

Phase 2 (Fine-Tuning): The entire network, including the convolutional backbone and the eccentricity pooling layers, was unfrozen. The model was then fine-tuned end-to-end with a reduced learning rate (1×10^{-5}) to adapt the feature extraction specifically for the temporal dynamics introduced by the RNN.

2.3 PSYCHOPHYSICAL VISUAL SEARCH EXPERIMENTS FOR AMODAL COMPLETION

The psychophysical paradigm of visual search is widely used to study multiple visual perception phenomena in the field of cognitive science. The basic task is for the observer to search for a target amongst distractors, and respond to it via a keypress. The principle is that a target similar to the distractors will be difficult to find, leading to a higher reaction time for the search, whereas, a dissimilar target will be found quickly (Duncan & Humphreys, 1989). Hence, the reaction time (RT) slope of a similar target-distractor pair would be steep, meanwhile, a distinct pair would have a flat slope. The RT slope is also modulated by the number of items in the search display, more items making the search more difficult in the case of similar target-distractor pairs.

We presented the trained eccNET with stimuli from the following psychophysics experiments, typically used in human research on amodal completion:

1. **Basic Effects:** Based on Rensink & Enns (1998), this experiment tests whether early vision performs rapid completion and faster search based on notch distinctiveness.
2. **Depth Changes:** Following Rensink & Enns (1998), we investigate if the perceived depth plane or the existence of occlusion interferes with rapid notch access.

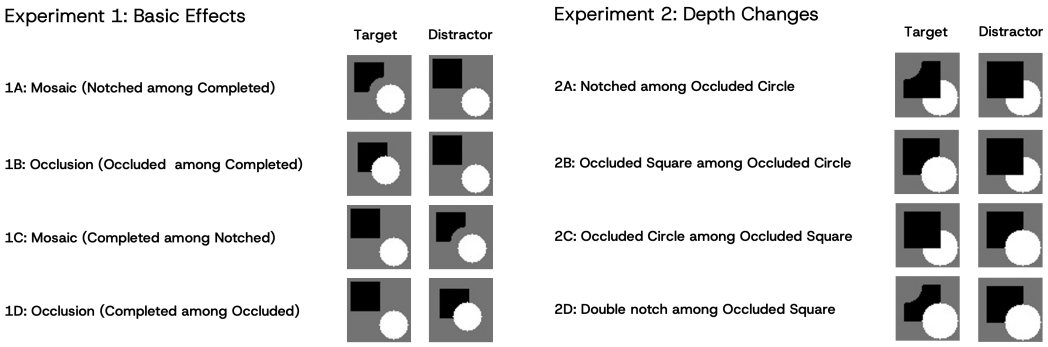

Figure 1: Target and distractor pairs for Experiment 1 (left) and Experiment 2 (right).

3. **Complex Shapes:** Modified from Cherian & Arun (2024), this experiment tests completion of complex shapes based on global statistical properties (spikiness).

4. **Global vs. Local:** Modified from Cherian & Arun (2024), this experiment compares search for global completions (Experiment 4) against local completions (shapes inconsistent with visible global statistics), employing conditions with locally consistent patterned objects.

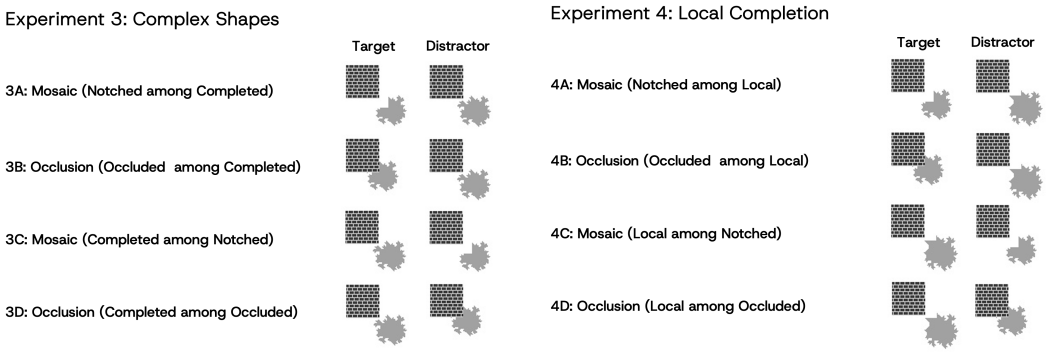

Figure 2: Target and distractor pairs for Experiment 3 (left) and Experiment 4 (right).

Each experiment employed a factorial design across four conditions (A to D). For each condition, stimuli were presented in five set sizes ($1, 3, 6, 9$, and 12 items), with each set size comprising 30 unique trials. This configuration resulted in total 150 trials in each condition and 600 trials in each experiment.

All trials in the experiment had the target present. The model starts each trial with an initial fixation at the center of the search display and generates a scanpath until its fixation lands within the target's bounding box. No task-specific training was performed; performance relied entirely on features learned during ImageNet training.

2.4 Conversion of Fixations to Reaction Time (RT)

The number of fixations (N) required to find the target was converted into a predicted reaction time (RT) in milliseconds using a linear model: $\text{RT} = \alpha \times N + \beta$. The slope, $\alpha = 252.36$ ms/fixation, approximates the combined duration of a single saccade and its subsequent fixation, while the intercept is set at $\beta = 376.27$. These values are adapted from the original eccNET paper(Gupta et al., 2021). A human subject research is underway to collect eye-tracking data to fit task-specific parameters for this conversion.

2.5 Amodal Completion Index

To formally compare the magnitude and direction of completion effects across biological and artificial systems, we introduced the Amodal Completion Index (ACI). This metric normalizes the difference in search efficiency between the Occluded (1B, 3B and 4B) and Notched (1A, 3A and 4A) conditions, providing a scale-invariant measure of completion.

The ACI is calculated as a standard ratio:

$$\text{ACI} = \frac{S_{\text{Occluded}} - S_{\text{Notched}}}{S_{\text{Occluded}} + S_{\text{Notched}}} \tag{1}$$

where S represents the search slope (ms/item). The index ranges from -1.0 to $+1.0$. A positive value indicates amodal completion, while a negative value indicates non-completion.

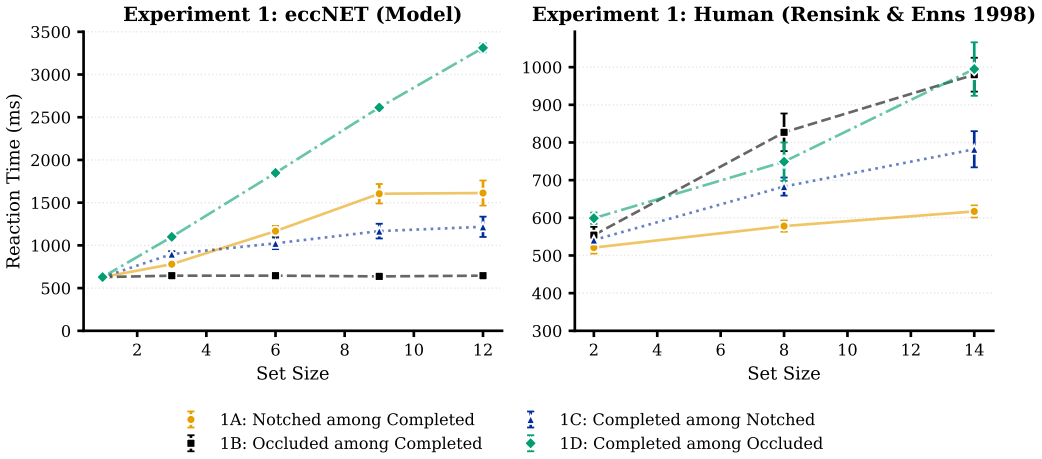

Figure 3: **Search slopes for simple shape completion.** (Left) eccNET performance. (Right) Human performance from Rensink and Enns (1998). The model displays near-zero slopes (easy search) for the occluded condition (1B), contrasting with the significant slowdown observed in humans, who complete the shape.

3 Results

3.1 Non-completion in eccNET

In Experiment 1, we evaluated the search efficiency of the feed-forward eccNET model on simple shape occlusion against human performance established by Rensink & Enns (1998). The aim was to determine if the model exhibits amodal completion effects observed in human early vision. The psychophysical logic employed here is that if there are completion effects, an occluded square will be perceived as similar to a completed square, leading to harder search (high RT). Conversely, the notched square will be perceived as dissimilar to completed square thus it would be easy to find (low RT).

Human observers, based on data from Rensink & Enns (1998), exhibit a strong positive amodal completion index (ACI $= +0.67$) for simple shapes, reflecting the substantial cost amodal completion imposes on the target search. They find isolated notched object rapidly (1A: 7 ms/item), it popped-out. When these fragments are occluded (1B), search slows significantly (36 ms/item). This slowdown is attributed to amodal completion as it indicates the filling-in of the occluded part and reduction in the ability to tease apart the target from the distractor. In contrast, eccNET exhibited a strongly negative index (ACI $= -1.0$), indicating the inversion of human visual processing. The model detected occluded targets instantaneously (1B: Slope $= 0.00$ ms/item), while the search for the isolated notched square was slower (1A: Slope $= 99$ ms/item). This means, unlike humans, for whom occlusion hinders search, occlusion acts as a *distinctive feature* for eccNET.

Furthermore, a two-way ANOVA was conducted to examine the effects of condition (4 levels) and set size (5 levels) on reaction time (ms) in Experiment 1. Results revealed a significant main effect of condition, $F(3, 580) = 398.45$, $p < .001$; a significant main effect of set size, $F(4, 580) = 220.56$, $p < .001$; and a significant condition $\times$ set size interaction, $F(12, 580) = 63.16$, $p < .001$.

Targets and distractors of condition 1A and 1B were swapped to get condition 1C and 1D respectively. As seen in figure 3, the slope of condition 1B (occluded among completed) and condition 1D (completed among occluded) revealed a stark asymmetry: detection of occluded objects among completed was quick while finding completed objects among occluded took longer. On the other hand, target detection times for condition 1A (notched among completed) and 1C (completed among notched) were symmetric, suggesting that notched objects were not a distinctive feature for eccNET.

There was a possibility that depth ordering (e.g., finding a target ”in front” vs. ”behind”) of two shapes might have influenced these results. We tested this possibility through Experiment 2. Human

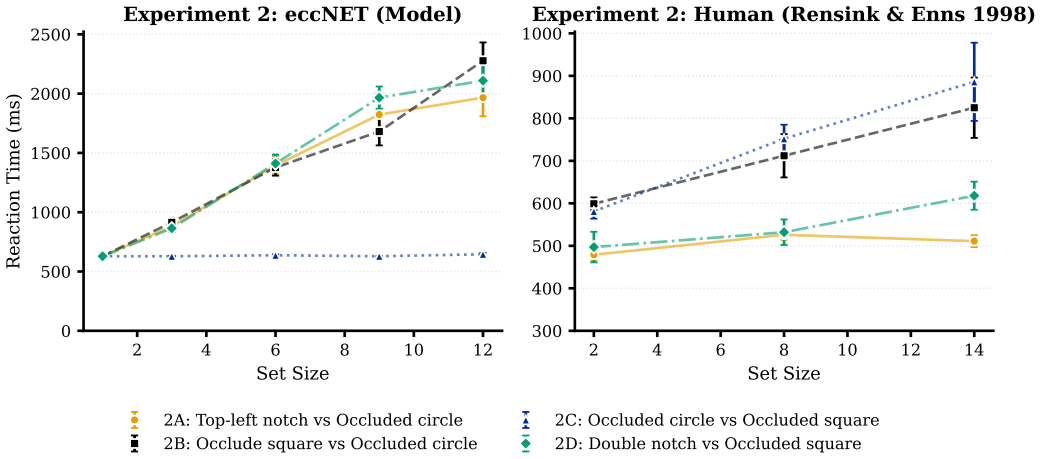

Figure 4: **Search slopes for depth ordering effects (Experiment 2).** (Left) eccNET performance. (Right) Human performance from Rensink and Enns (1998). The model fails to replicate the human pop-out effect for notched features (2A & 2B), exhibiting steep search slopes regardless of depth ordering, whereas humans find the notch rapidly in all depth configurations.

observers, as reported in Rensink & Enns (1998), show robustness of target search to depth ordering. The search for a notched square remains rapid whether it appears in front of a disk (condition 2A: 6 ms/item) or behind it (condition 2D: 5 ms/item). This indicates that a distinctive notch remains accessible across all depths. Furthermore, there was no significant search asymmetry based on depth ordering. In contrast, eccNET failed to replicate the human pop-out effect for the notched feature. In condition 2A (notched square in front), the model yielded a slope of 128 ms/item, and in condition 2D (notched square behind), performance worsened to 145 ms/item. This indicates that the model could not extract the notch as a salient feature in either depth configuration, treating the target and distractors as indistinguishable.

Furthermore, a two-way ANOVA was conducted to examine the effects of condition and set size on reaction time (ms) in Experiment 2 for eccNET. Results revealed a significant main effect of condition, $F(3, 580) = 102.81$, $p < .001$; a significant main effect of set size, $F(4, 580) = 132.28$, $p < .001$; and a significant condition $\times$ set size interaction, $F(12, 580) = 15.42$, $p < .001$.

Another divergence from human performance was observed in the asymmetry between conditions 2B and 2C. While humans show relatively slow search in both directions (19 vs. 25 ms/item), eccNET displayed a massive asymmetry. Finding an occluded circle among squares (condition 2C) was instantaneous (slope = 1.21 ms/item), whereas finding an occluded square among circles (condition 2B) was laborious (slope = 145.20 ms/item). This suggests that eccNET finds objects with simple curvature features like circles particularly distinct from squares.

Next, we investigate if there can be completion of complex shapes in eccNET through Experiment 3. The occluded object contained high-frequency edge information. Based on Cherian & Arun (2024), we can infer that human observers have positive ACI for complex shapes, similar to simple shapes. They can detect the notched spiky shape in condition (3A) rapidly, but search would slow down when the notch was occluded (3B), because the completion process hides the target feature. Interestingly, eccNET yielded a positive index (ACI $= +0.59$).

Furthermore, a two-way ANOVA was conducted to examine the effects of condition (4 levels) and set size (5 levels) on reaction time (ms) in Experiment 3. Results revealed a significant main effect of condition, $F(3, 580) = 300.88$, $p < .001$; a significant main effect of set size, $F(4, 580) = 289.74$, $p < .001$; and a significant condition $\times$ set size interaction, $F(12, 580) = 47.78$, $p < .001$.

A stark asymmetry was observed between condition 3B (occluded vs completed) and condition 3D (completed vs occluded). While finding the completed target among occluded distractors (3D)

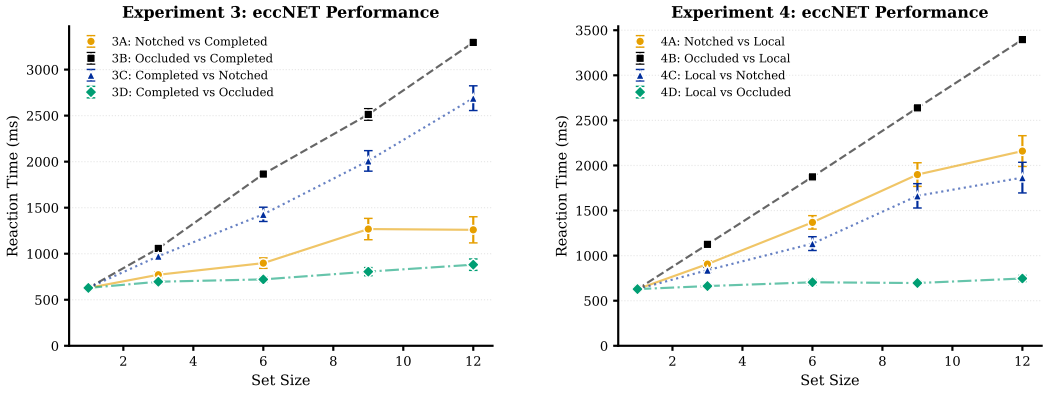

Figure 5: **eccNET search slopes for complex shapes, Experiment 3 (Global) & 4 (Local).** The model shows no difference in search efficiency between global completion and local completion.

was relatively easy (slope = 21.88 ms/item), the inverse task of finding an occluded target among completed distractors (3B) was hard (slope = 242.67 ms/item).

To further investigate whether the completion of complex shapes is merely based on global patterns (likely/plausible) or if local pattern (unlikely) completion also occurs, we designed Experiment 4. We evaluated the search efficiency of the feed-forward eccNET model on local completion—shapes that are locally completed but globally inconsistent. This served as a control to Experiment 3: while Experiment 3 tested detection of global (plausible) shapes, Experiment 4 tested if the model could distinguish purely statistical patches (local) from notches and occlusions, hence assessing whether the model differentiates between *meaningful* and *nonsensical* completion.

Human observers (based on Cherian & Arun (2024)) typically treat local completions as distinct oddballs because they violate the global structural prediction generated by the visual system. However, eccNET seems Plausible Blind, failing to differentiate the local patches from the notched shapes based on structural integrity. The search for a local target among notched distractors (condition 4C) yielded a slow slope of 117.96 ms/item, and the inverse condition (notched vs. local, condition 4A) was similarly inefficient (144.67 ms/item). Unlike Experiment 3, where the smooth global shape was relatively easier to distinguish from a notch (62 ms/item), here the 'local' patch was statistically confusable with the notch, leading to severe feature congestion.

Furthermore, a two-way ANOVA was conducted to examine the effects of condition (4 levels: notched vs. local, occluded vs. local, local vs. notched, local vs. occluded) and set size (5 levels: 1, 3, 6, 9, 12) on reaction time (ms) in Experiment 4. Results revealed a significant main effect of condition, $F(3, 580) = 237.75$, $p < .001$; a significant main effect of set size, $F(4, 580) = 245.09$, $p < .001$; and a significant condition $\times$ set size interaction, $F(12, 580) = 35.62$, $p < .001$.

A strong asymmetry was present between conditions 4D and 4B. When the task involved finding a local target among occluded distractors (condition 4D), search was near-instantaneous (slope = 9.50 ms/item). Conversely, finding an occluded target among local distractors (condition 4B) was very hard (slope = 251.83 ms/item). Post-hoc Tukey HSD tests confirmed that condition 4B was significantly slower than all other conditions ($p < .001$), while the difference between the conditions 4A and 4C was not significant ($p = 0.21$). This conclusively demonstrates that the model does not validate the shape of the completion; it merely exploits the absence of occlusion artifacts to pop out the local target, just as it did for the global target in Experiment 3.

3.2 No Aid through Recurrent Computations

To test if recurrent computations help in gaining features that can amodaly complete the object under occlusion, we introduced a recurrent layer in eccNET. While the recurrent layer changed the absolute reaction times (fixations), the general pattern of results mirrored the feed-forward model, confirming a persistent non-completion.

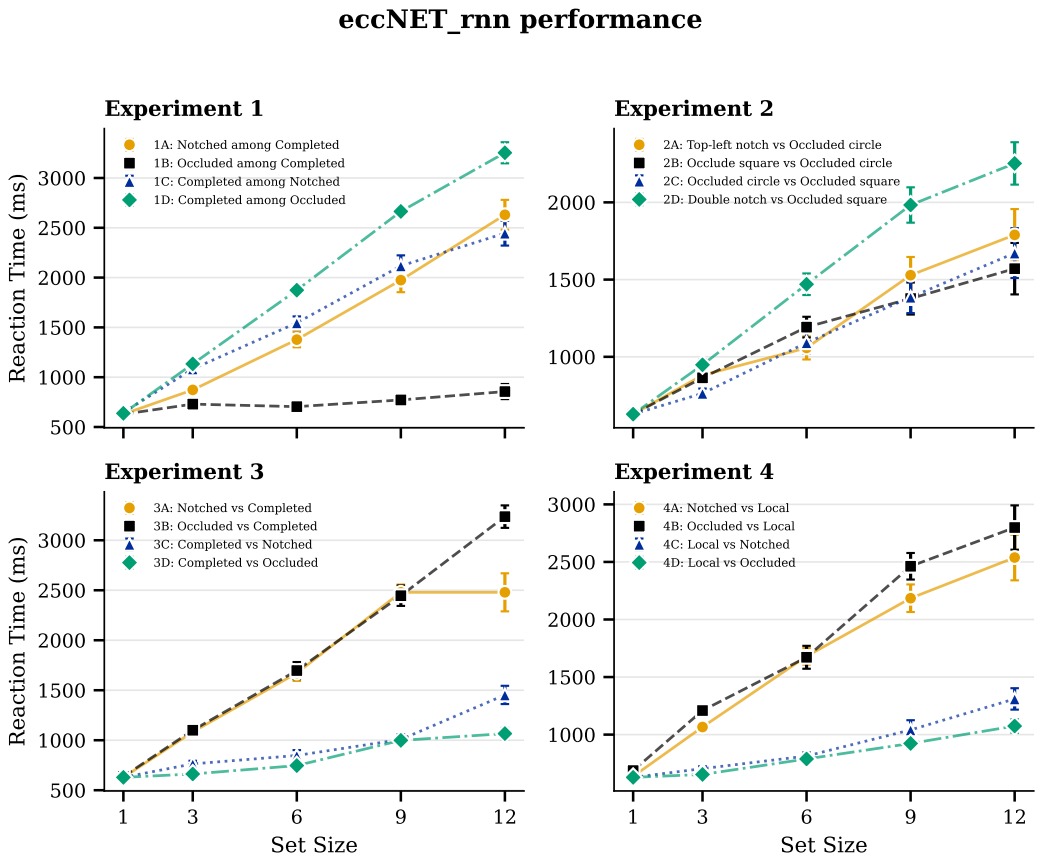

Figure 6: **Search slopes for the recurrent model (eccNET_rnn) across all four experiments.** The results are shown for (Top-Left) Experiment 1, (Top-Right) Experiment 2, (Bottom-Left) Experiment 3, and (Bottom-Right) Experiment 4. The introduction of a recurrent layer fails to facilitate amodal completion, instead mirroring the Non-Completion effects and feature asymmetries observed in the feed-forward architecture.

In Experiment 1, similar to the feed-forward model, eccNet_rnn maintained a negative Amodal Completion Index (ACI $= -0.83$), confirming that occlusion artifacts facilitate rather than hinder search. However, a critical divergence occurred in the notched condition (1A). While the feed-forward model processed isolated notches efficiently (34.7 ms/item), the RNN suffered a severe slowdown (183.38 ms/item).

In Experiment 2, results observed were similar to eccNET, with the RNN failing to isolate features in depth (Conditions 2A & 2D slopes $>$ 100 ms/item). Notably, the recurrence suppressed the 'curvature pop-out' effect seen in the feed-forward model. In Condition 2C (occluded circle vs. square), the slope degraded from 1.2 ms/item (feed-forward) to 97.04 ms/item (RNN), indicating that recurrent iterations likely *smoothed over* the unique curvature signal that drove the feed-forward efficiency.

In complex shapes, the ACI dropped to $+0.13$ (compared to $+0.59$ in eccNET) for global patterns. This reduction was not driven by better occlusion handling, but by a performance drop in the notched baseline (3A: 180.88 ms/item vs. eccNET's 62.7 ms/item). For local patterns (Experiment 4), a similar trend was observed. Finding a locally completed target among notched distractors was faster (61.25 ms/item), the inverse remained inefficient (174.86 ms/item).

For all experiments on eccNET_rnn, two-way ANOVAs confirmed significant main effects of condition ($p < .001$) and set size ($p < .001$), as well as significant interactions ($p < .05$).

4 DISCUSSION

Our results reveal a divergence between human vision and deep neural networks. While humans have a robust ability to amodaly complete occluded objects, models tested here — both feed-forward and recurrent — do not show such an ability.

In Experiment 1, the stark contrast in the Amodal Completion Index (humans: $+0.67$; models: -0.83 to -1.0) confirms that the models do not 'fill in' missing information of simple shapes. This divergence is critical as it is inconsistent with what Cherian & Arun (2024) found using the representation distance method. One possibility that explains these divergent results is that the internal representation of completion, and the ability to use the completed representation in deploying attention for visual search, might be different. Evidence from Wolfe et al. (2011); Wolfe & Horowitz (2017) indicates a dissociation between perception and guidance in humans, where a perceptually compelling orientation created through amodal completion often fails to guide the deployment of attention efficiently in visual search tasks.

Experiment 2 explored the influence of depth ordering and found that while humans can perform target search efficiently regardless of whether an object is perceived to be in front of or behind an occluder, the model's search efficiency was heavily modulated by local curvature and depth-specific artifacts. The failure to replicate human pop-out effects for notched features in various depth configurations confirms that the model lacks the 'functional filling-in' mechanism that humans have.

We also observed critical asymmetries in search across all experiments under some conditions. These findings are consistent with other human studies in various visual search experiments (Rosenholtz, 2001; Carrasco et al., 1998; Yamani & McCarley, 2010; Shen & Reingold, 2001), and the study on the model (Gupta et al., 2021). Another critical observation seen across experiments 1 and 2 is that occluded simple shapes created a pop-out effect for eccNET; they were perceived as having a distinct feature. This might be due to the fact that training data (ImageNet) does not contain occluded images, leading to occlusion itself becoming a distinctive feature that stands out. Evidence suggests that in humans, familiarity with the target is decremental to search efficiency while unfamiliar objects pop-out, ultimately resulting in search asymmetries (Shen & Reingold, 2001). However, we need to train the model on an dataset which contains occluded objects to test this interpretation in DNNs.

In the experiments with complex shapes (Experiment 3 & 4), the model initially appeared to show a positive ACI ($+0.59$), which might superficially suggest completion. However, the results provided by Experiment 4 (local completion) hint towards a more complex interpretation. The model failed to differentiate between *meaningful* global completions and *nonsensical* local patches, treating any absence of high-contrast occlusion artifacts as a reason for pop-out. This indicates that the perceived completion in complex shapes was actually an artifact of feature congestion and the model's inability to extract spiky features, rather than a true reconstruction of the occluded statistical properties.

These results suggest that the mechanisms used by DNNs in visual search might be entirely based on local statistical contrast rather than global structural integrity of the shapes. In Experiment 4, the model's failure to distinguish between globally consistent shapes and statistically confusable local patches demonstrates that it lacks a top-down or structural prior for what constitutes a *plausible* or *likely* object. While humans use global statistical properties—such as the spikiness of an object to predict hidden parts, the model might be 'plausibility blind', relying on the presence or absence of occlusion-induced edges to drive attention. We are currently conducting a more granular investigation into the results of Experiments 3 and 4 by dissecting other factors that might be playing a role.

In biological vision, the completion of occluded structure is not just about extrapolating missing pixels, rather, it is about creating a proto-object or representation that becomes the primary unit of attentional access. This process preempts the fragments, making it difficult for the observer to see the notched shape once it is perceived as behind an occluder (Sekuler & Palmer, 1992; Rauschenberger & Yantis, 2001). The models examined here do not seem to exhibit this preemption. Their ability to detect occluded targets faster than notched ones shows that they maintain access to the notched boundaries and artifacts throughout the hierarchy.

The research is under progress to train eccNET model with a recurrent layer from scratch instead of using fine-tuned weights, and also, on a dataset with occluded objects. We also plan to collect human eye-tracking and psychophysical data for all the experiments.

ACKNOWLEDGMENTS

We acknowledge National Supercomputing Mission (NSM) for providing computing resources of 'PARAM Ananta' at IIT Gandhinagar, which is implemented by C-DAC and supported by the Ministry of Electronics and Information Technology (MeitY) and Department of Science and Technology (DST), Government of India.

REFERENCES

Marisa Carrasco, Tracy L. McLean, Svetlana M. Katz, and Karen S. Frieder. Feature asymmetries in visual search: Effects of display duration, target eccentricity, orientation and spatial frequency. *Vision Research*, 38(3):347–374, 1998. doi: 10.1016/S0042-6989(97)00138-0.

S. Chen, L. Schnabl, H. J. Müller, and M. Conci. Amodal completion of a target template enhances attentional guidance in visual search. *i-Perception*, 9(4):2041669518796240, 2018. doi: 10.1177/2041669518796240.

T. Cherian and S. P. Arun. What do we see behind an occluder? Amodal completion of statistical properties in complex objects. *Attention, Perception, & Psychophysics*, 86(8):2721–2739, 2024. doi: 10.3758/s13414-024-02948-w.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, 2009.

J. Duncan and G. W. Humphreys. Visual search and stimulus similarity. *Psychological Review*, 96 (3):433–458, 1989. doi: 10.1037/0033-295x.96.3.433.

J. Freeman and E. P. Simoncelli. Metamers of the ventral stream. *Nature Neuroscience*, 14(9): 1195–1201, 2011. doi: 10.1038/nn.2889.

W. Gerbino. Amodal completion revisited. *i-Perception*, 11:2041669520937323, 2020. doi: 10.1177/2041669520937323.

S. K. Gupta, M. Zhang, C.-C. Wu, J. M. Wolfe, and G. Kreiman. Visual search asymmetry: Deep nets and humans share similar inherent biases. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6946–6958, 2021.

Byungwoo Kang, Benjamin Midler, Feng Chen, and Shaul Druckmann. Recurrent connections facilitate occluded object recognition by explaining-away. *bioRxiv*, 2020. doi: 10.1101/2020.12.16.422991. URL `https://doi.org/10.1101/2020.12.16.422991`.

Dale K. Lee, Laurent Itti, Christof Koch, and Jochen Braun. Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, 2(4):375–381, 1999. doi: 10.1038/7267.

Grace W. Lindsay, Thomas D. Mrsic-Flogel, and Maneesh Sahani. Bio-inspired neural networks implement different recurrent visual processing strategies than task-trained ones do. *bioRxiv*, 2022. doi: 10.1101/2022.03.07.483196.

K. Rajaei, Y. Mohsenzadeh, R. Ebrahimpour, and S. M. Khaligh-Razavi. Beyond core object recognition: Recurrent processes account for object recognition under occlusion. *PLOS Computational Biology*, 15:1–30, 2019.

R. Rauschenberger and S. Yantis. Masking unveils pre-amodal completion representation in visual search. *Nature*, 410(6826):369–372, 2001. doi: 10.1038/35066577.

R. A. Rensink and J. T. Enns. Early completion of occluded objects. *Vision Research*, 38:2489–2505, 1998. doi: 10.1016/S0042-6989(98)00052-1.

Ruth Rosenholtz. Search asymmetries? What search asymmetries? *Perception & Psychophysics*, 63(3):476–489, 2001. doi: 10.3758/BF03194413.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

Allison B. Sekuler and Stephen E. Palmer. Perception of partly occluded objects: A microgenetic analysis. *Journal of Experimental Psychology: General*, 121(1):95–111, 1992. doi: 10.1037/0096-3445.121.1.95. URL https://doi.org/10.1037/0096-3445.121.1.95.

T. Serre. Deep learning: The good, the bad, and the ugly. *Annual Review of Vision Science*, 5: 399–426, 2019. doi: 10.1146/annurev-vision-091718-014951. URL https://doi.org/10.1146/annurev-vision-091718-014951.

Jiye Shen and Eyal M. Reingold. Visual search asymmetry: The influence of stimulus familiarity and low-level features. *Perception & Psychophysics*, 63(3):464–475, 2001. doi: 10.3758/BF03194412.

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

H. Tang, M. Schrimpf, W. Lotter, C. Moerman, A. Paredes, J. Ortega Caro, W. Hardesty, D. Cox, and G. Kreiman. Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences*, 115(35):8835–8840, 2018. doi: 10.1073/pnas.1719397115.

J. M. Wolfe and T. S. Horowitz. Five factors that guide attention in visual search. *Nature Human Behaviour*, 1(3):0058, 2017. doi: 10.1038/s41562-017-0058.

J. M. Wolfe, E. Reijnen, T. S. Horowitz, R. Pedersini, Y. Pinto, and J. Hulleman. How does our search engine "see" the world? The case of amodal completion. *Attention, Perception, & Psychophysics*, 73(4):1054–1064, 2011. doi: 10.3758/s13414-011-0103-0.

Yusuke Yamani and Jason S. McCarley. Visual search asymmetries within color-coded and intensity-coded displays. *Journal of Experimental Psychology: Applied*, 16(2):124–134, 2010. doi: 10.1037/a0019623.

M. Zhang, J. Feng, K. T. Ma, J. H. Lim, Q. Zhao, and G. Kreiman. Finding any Waldo with zero-shot invariant and efficient visual search. *Nature Communications*, 9(1):3730, 2018. doi: 10.1038/s41467-018-06217-x. URL https://doi.org/10.1038/s41467-018-06217-x.

H. Zhu, P. Tang, J. Park, S. Park, and A. Yuille. Robustness of object recognition under extreme occlusion in humans and computational models. *arXiv preprint arXiv:1905.04598*, 2019.

A Appendix: Post-hoc Analysis Results

This appendix provides the detailed results of the Tukey HSD post-hoc tests ($\alpha = 0.05$) for all experiments.

Table 1: Tukey HSD comparisons for Condition in Experiment 1

Group 1	Group 2	Meandiff	p-adj	Lower	Upper
1A_Notched	1B_occluded	-518.18	0.000	-705.45	-330.91
1A_Notched	1C_sep_Notched	-171.60	0.086	-358.87	15.66
1A_Notched	1D_sep_occluded	741.94	0.000	554.67	929.20
1B_occluded	1C_sep_Notched	346.57	0.000	159.31	533.84
1B_occluded	1D_sep_occluded	1260.12	0.000	1072.85	1447.38
1C_sep_Notched	1D_sep_occluded	913.54	0.000	726.28	1100.81

Table 2: Tukey HSD comparisons for Condition in Experiment 2

Group 1	Group 2	Meandiff	p-adj	Lower	Upper
2A_notch	2B_occ_sq	37.01	0.962	-158.33	232.36
2A_notch	2C_occ_circ	-704.93	0.000	-900.27	-509.58
2B_occ_sq	2C_occ_circ	-741.94	0.000	-937.28	-546.60
2C_occ_circ	2D_db_notch	762.13	0.000	566.78	957.47

Table 3: Tukey HSD comparisons for Condition in Experiment 3

Group 1	Group 2	Meandiff	p-adj	Lower	Upper
3A_Notched_n	3B_occ_occ	906.81	0.000	692.53	1121.10
3A_Notched_n	3C_Notched_c	580.43	0.000	366.15	794.71
3B_occ_occ	3C_Notched_c	-326.39	0.001	-540.67	-112.10
3C_Notched_c	3D_occ_c	-799.14	0.000	-1013.42	-584.86

Table 4: Tukey HSD comparisons for Condition in Experiment 4

Group 1	Group 2	Meandiff	p-adj	Lower	Upper
4A_notched	4B_occluded	540.05	0.000	318.76	761.34
4A_notched	4C_local	-166.56	0.213	-387.85	54.73
4B_occluded	4C_local	-706.61	0.000	-927.90	-485.32
4C_local	4D_occluded	-538.37	0.000	-759.66	-317.08