# Artistically Structured Images Can Modify Human Eye Gaze Behavior

Nicholas SCHWIER[1], Corrine P. SCHWIER[1,2], Yoshinari KAMEDA[1,3]

*__1.__ University of Tsukuba, Tsukuba, Japan. __2.__ Faculty of Science and Engineering, Waseda University, Tokyo, Japan. __3.__ Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan*

***E-mail:*** *nicholas.schwier@gmail.com*

**Abstract:** This study investigates how compositional structure in artistic images modulates vision behavior. Using VR-integrated eye tracking, we quantified fixation latency, saccadic efficiency, and gaze-salience alignment. Structured images elicited faster fixations and more direct saccades, suggesting that compositional cues guide visual attention beyond low-level salience mechanisms.

**Keywords:** Vision Behavior, Virtual Reality, Eye Tracking, Art

## Introduction

Understanding gaze behavior in complex visual environments is critical to refining visual attention models. Traditional computational frameworks, such as those of Itti and Koch, emphasize low-level feature-driven prominence, but often overlook higher-order influences such as artistic composition [1–3].

Artists have long guided perception through compositional structure, as discussed in Arnheim's Gestalt-based theories and Copley's distinctions between structured and unstructured artworks [4,5]. However, the quantifiable impact of such a structure on eye movements remains underexplored [6,7].

This study investigates how artistic structure modulates gaze dynamics, focusing on fixation acquisition time and saccadic efficiency as representations for visual exploration and object recognition [8]. We hypothesize that structured images facilitate more efficient gaze behavior, reflecting the interaction of composition and visual cognition.

## Methods

Nineteen participants (ages 18 to 35) with normal or corrected vision, no significant ocular conditions, and no formal art training were recruited. This controlled for prior exposure to artistic conventions known to influence gaze behavior. Written informed consent was obtained in accordance with the Ethics Review Board of the Center for Computational Science, University of Tsukuba.

Participants were seated in a 360° rotating chair and viewed stimuli through a VIVE Pro Eye 2 virtual reality (VR) headset with integrated Tobii eye tracking (90 Hz). A UNITY-based virtual environment with a black background displayed stimuli, fixation crosses, and instructions, always centered relative to the participant. Eye tracking data were integrated via SRANIPAL drivers and recorded as CSV files containing timestamps, 3D gaze coordinates, and object labels.

Visual stimuli consisted of photographic images of architectural and natural scenes, carefully selected to exclude humans, animals, and text to minimize cognitive confounders. Two categories of images were used: structured images designed according to established artistic principles to guide visual attention, and unstructured images captured without intentional compositional control.

Participants viewed 19 images (12 structured, 7 unstructured) for 4 seconds each, separated by a central fixation cross. The viewing order and the placement of the lateral image were randomized. Participants were instructed to explore visually without explicit search tasks. Visual salience was computed with DeepGaze II, generating heatmaps of predicted fixation likelihood. Gaze data was overlaid and salience scores were extracted using OpenCV. The metrics included average salience, time to first fixation (interval to first fixation cluster) and path straightness (ratio of cumulative gaze trajectory to direct distance). Fixations and saccades were identified by OPTICS clustering.

A second task involved 15 pairs of structured images varying in compositional magnitude (high-contrast guides vs. subtle negative space), each paired with a corresponding graphic visual path. Filler images minimized recognition effects. Gaze trajectories were extracted and path similarity was quantified by RMSE. Group comparisons used two-sided t tests and ANOVA with post hoc Tukey tests.

## Results

Saliency scores derived from DeepGazeII heat map comparisons differed significantly between image types. Unstructured images produced a higher mean saliency score of 42.138, while structured images produced a lower mean saliency score of 32.637 ($P < 0.001$). The first fixation acquisition times also differed between conditions. Participants exhibited shorter times to the first fixation when viewing structured images compared to unstructured images ($P < 0.032$). Straightness scores, representing the summed vector distance from the initial gaze point to the first fixation, were lower for structured images relative to unstructured images ($P < 0.016$).). Analysis of the similarity of the eye gaze path using RMSE revealed a significant difference based on the magnitude of the image. Small-magnitude image pairs resulted in higher RMSE scores, while large-magnitude image pairs resulted in lower RMSE scores (P <0.0001) No statistically significant differences were observed in time to first fixation between image magnitude conditions (P>0.05).

## Discussion

Our findings show that artistic structure significantly alters gaze behavior, reflected in deviations in fixation prediction accuracy. Specifically, the reduced predictive power of DeepGaze II for structured images suggests that current saliency models overlook key factors that influence visual attention, likely related to intentional compositional design [9].

Participants demonstrated faster fixation acquisition and more direct gaze paths when viewing structured images, supporting the hypothesis that compositional structure facilitates scene parsing and object recognition. This is consistent with previous work showing that specific visual characteristics can improve fixation efficiency and reduce exploratory saccades [10]. Together, these results indicate that artistic structure provides visual cues that guide perception, reducing cognitive load during scene interpretation.

Our approach combined eye-tracking metrics with computational saliency analysis to quantify how artistic intent shapes visual behavior. Graded effects emerged based on the visibility of the underlying structure: participants' gaze paths aligned more closely with intended guides in large-magnitude images, while small-magnitude images elicited more search behavior, as reflected by increased RMSE deviation.

These results extend previous research on guided attention in cinema, suggesting that similar mechanisms apply to static images [11]. However, the use of images from a single artist, while controlling for stylistic variability, limits generalizability. Future studies should incorporate diverse artistic styles to validate and expand these findings. Moreover, while DeepGaze II represents an advanced saliency model, its inability to account for higher-order semantic or intentional cues highlights the need for models that better integrate cognitive influences on gaze behavior.

Future work should incorporate neural measures (eg, fMRI, EEG, NIRS) to identify the cortical correlations of altered gaze behavior in response to artistic structure. Such research will clarify whether the observed behavioral changes stem from reduced processing demands, improved perceptual grouping, or other mechanisms, as suggested by Brunnström et al. [12]. Exploring parallels with cinematic continuity editing may further illuminate how the brain integrates complex visual information.

This study provides initial empirical support for the idea that artistic structure systematically shapes visual exploration, highlighting gaps in current models, and offering a pathway for interdisciplinary research bridging vision science and the arts.

## Conclusions

This study demonstrates that artistic structure in images systematically alters gaze behavior, guiding visual attention along intended paths, and facilitating early object recognition. These effects challenge purely stimulus-driven visual salience models by highlighting the role of compositional intent in shaping perception.

Consistent gaze patterns among participants suggest that deliberate structuring of visual elements reduces perceptual ambiguity and cognitive load, promoting efficient scene understanding. These results have broad implications for vision science, art perception, and the development of more comprehensive models of attention.

Future research should integrate neural imaging, expand artistic sampling, and refine computational models to fully characterize the mechanisms underlying this "artistic gaze effect." Such efforts promise to deepen our understanding of how art engages and shapes the human perceptual system.

## Acknowledgments

This research was made possible and funded in part by University of Tsukuba, Center for Computational Science, and School of Integrated Global Majors.

## References

[1] L. Itti, "Models of bottom-up attention and saliency," *Neurobiol. Atten.*, 2005. doi: 10.1016/b978-012375731-9/50098-7.
[2] F. Baluch and L. Itti, "Mechanisms of top-down attention," *Trends Neurosci.*, vol. 34, 2011. doi: 10.1016/j.tins.2011.02.003.
[3] B. J. White et al., "Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video," *Nat. Commun.*, vol. 8, 2017. doi: 10.1038/ncomms14263.
[4] G. E. Myers and R. Arnheim, Art and Visual Perception: *The Psychology of the Creative Eye, Philos. Phenomenol. Res.*, vol. 16, 1956. doi: 10.2307/2104731.
[5] L. Alcopley, "Drawings as structures and nonstructures," Leonardo, vol. 1, 1968. doi: 10.2307/1571900.
[6] M. Hatt et al., *Art History: A Critical Introduction to Its Methods* Manchester, UK: Manchester Univ. Press, 2006.
[7] M. Pelowski et al., "Visualizing the impact of art: An update and comparison of current psychological models of art experience," *Front. Hum. Neurosci.*, vol. 10, 2016. doi: 10.3389/fnhum.2016.00160.
[8] A. T. Duchowski, "Eye movement analysis," in *Eye Tracking Methodology*. Cham, Switzerland: Springer, 2017. doi: 10.1007/978-3-319-57883-5_13.
[9] M. Kümmerer et al., "DeepGaze II: Predicting fixations from deep features over time and tasks," *J. Vis.*, vol. 17, 2017. doi: 10.1167/17.10.1147.
[10] J. L. Liu, "A comparative study of visual salient features affect object recognition," *Appl. Mech. Mater.*, vol. 713–715, 2015. doi: 10.4028/www.scientific.net/AMM.713-715.2165.
[11] T. J. Smith, "The attentional theory of cinematic continuity," *Projections*, vol. 6, 2012. doi: 10.3167/proj.2012.060102.
[12] K. Brunnström et al., "Active fixation for scene exploration," *Int. J. Comput. Vis.*, vol. 17, 1996. doi: 10.1007/BF00058749.