fMRI-Based Prediction of Eye Gaze During Naturalistic Movie Viewing Reveals Eye-Movement–Related Brain Activity

Le Gao[1], Zhi Wei[1], Bharat B. Biswal[2], Xin Di [2*]

1. Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, 07102, USA

2. Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ, 07102, USA

* Corresponding author:

Xin Di, Ph.D.

604 Fenster Hall, University Height

Newark, NJ, 07102, USA

xin.di@njit.edu

Abstract

Eye gaze and eye movements provide important indices of perceptual and cognitive processes, particularly under naturalistic conditions such as movie viewing. However, concurrent eye tracking is often unavailable in functional MRI (fMRI) studies due to technical and logistical constraints. Recent deep learning approaches have made it possible to estimate eye gaze directly from eyeball signals in fMRI data, offering a potential alternative to camera-based eye tracking. Here, we applied a pre-trained fMRI-based deep neural network model (DeepMReye) to estimate eye gaze during movie watching across three independent fMRI datasets. Model performance was evaluated by comparison with camera-based eye-tracking data when available, as well as by assessing inter-individual correlations, a commonly used benchmark in naturalistic fMRI research. At the individual level, predicted gaze showed modest correspondence with measured data ($r \approx -0.38$ to 0.67). In contrast, group-averaged gaze predictions exhibited substantially higher correlations ($r \approx 0.7$–0.8), indicating improved reliability at the group level. We further derived eye-movement–related time series from the predicted gaze signals and examined their associated brain activity. Consistent with differences in prediction accuracy, individual-level analyses yielded activation patterns largely restricted to visual cortex, whereas group-averaged predictions revealed more widespread activation, including established oculomotor control regions such as frontal and parietal eye fields. Exploratory analyses indicated age-related effects on gaze prediction accuracy and eye-movement–related brain activity, although these effects were not consistent across datasets. Together, these findings demonstrate that group-averaged fMRI-based gaze estimation can support the investigation of eye-movement–related brain activity in naturalistic paradigms, while highlighting current limitations for individual-level inference. The results provide a methodological assessment of fMRI-based gaze prediction and inform its appropriate use in future neuroimaging studies.

Keywords: Convolutional neural network, Frontal eye field, Gaze prediction, Inter-individual correlation, Naturalistic condition.

Introduction

Eye movements are central to human perception and cognition. They guide visual exploration, support attentional selection, and facilitate social communication by revealing what people attend to and how they interpret dynamic visual scenes (Henderson 2003; Tatler et al. 2011). During naturalistic activities such as movie watching, gaze patterns reflect moment-to-moment fluctuations in engagement, comprehension, and social inference (Hasson et al. 2004; Richardson and Saxe 2020). Because of this richness, eye movements are widely used as behavioral markers for studying cognitive processes, including attention, memory, and theory-of-mind reasoning.

Functional MRI (fMRI) provides powerful tools for studying the neural mechanisms underlying these processes, but a major limitation in naturalistic fMRI research is that most datasets do not include simultaneous eye-tracking recordings. This gap is especially evident in large open-access datasets and legacy studies collected before eye-tracking systems became widely available (Sonkusare et al. 2019). Even today, many research sites lack in-scanner eye-tracking hardware, and even when such systems are available, obtaining high-quality recordings is often difficult—particularly in populations such as young children, individuals with neurodevelopmental or psychiatric conditions, or participants who have difficulty maintaining stable head position. As a result, many potentially informative fMRI datasets lack reliable gaze information, complicating the interpretation of neural activity related to visual attention, scene processing, and oculomotor control and leaving valuable datasets under-utilized.

Recent advances in deep learning have introduced a possible solution. Models such as DeepMReye can infer gaze position directly from BOLD fluctuations in the eyeball region (Frey et al. 2021). These approaches build on earlier observations that motion of the eye and surrounding tissues produces systematic fMRI signal changes (Son et al. 2020), which can be decoded using convolutional neural networks. The ability to estimate gaze retrospectively offers a powerful opportunity: large fMRI datasets without eye tracking can be “augmented” with reconstructed gaze trajectories, enabling novel analyses of visual behavior and attention without requiring camera-based systems.

Beyond methodological utility, fMRI-based gaze decoding opens new avenues for developmental and clinical research. Children and adults differ substantially in their visual exploration patterns, attentional control, and consistency of gaze behavior (Kirkorian et al. 2012). These behavioral differences correspond to ongoing maturation of large-scale cortical networks involved in visual attention, cognitive control, and social processing (Casey et al. 2005; Luna et al. 2015). Movie-watching paradigms are particularly well-suited for developmental studies because they elicit reliable neural responses across individuals while also capturing meaningful age-related differences in gaze patterns and comprehension (Cantlon and Li 2013; Petroni et al. 2018).

In this study, we applied DeepMReye to three naturalistic movie-watching fMRI datasets to assess its ability to reconstruct gaze behavior and to characterize the neural correlates of predicted eye movements. We first evaluated prediction accuracy by correlating model-based gaze estimates with simultaneously acquired camera-based eye-tracking data. We also leveraged the well-established phenomenon of inter-individual correlation during shared video viewing (Hasson et al. 2004; Madsen and Parra 2022; Di et al. 2025), and quantified inter-individual correlations of the DeepMReye predictions themselves. If the model accurately captures gaze behavior, we expect both strong correlations with ground-truth eye tracking and high intersubject consistency. In addition, because gaze patterns show robust inter-individual correlations, we computed group-averaged gaze predictions and compared correlations with eye-tracking data at both the group and individual levels. This approach tests the hypothesis that even when individual predictions are noisy, averaging across participants may enhance signal quality and improve correspondence with ground-truth gaze.

In this study, we examined brain activity associated with eye movements by using frame-to-frame gaze displacement (Euclidean distance) as a regressor in a conventional fMRI activation analysis. We hypothesized that we would identify regions involved in oculomotor control, such as the frontal eye fields (Pierrot-Deseilligny et al. 2004). We conducted these analyses using both group-averaged and individual-level gaze predictions to compare their ability to reveal eye-movement–related neural responses.

Together, these analyses provide insight into the performance of DeepMReye and its utility for studying gaze-related brain activity in naturalistic fMRI datasets.

Methods

Participants and Datasets

This study utilized three publicly available fMRI datasets. All datasets included fMRI scans acquired during naturalistic movie-watching tasks, along with demographic information such as age and sex. Across all datasets, participants were included based on data quality criteria, excluding those with excessive imaging artifacts or head motion.

Natural Viewing (NV) Dataset (Telesford et al. 2023): This dataset includes 22 adult participants (ages 23–51 years) and was collected at the Nathan S. Kline Institute. We focused on the naturalistic viewing portion where participants watched "The Present" (TP; 4min 18s) and a clip from "Despicable Me" (DM; 10min). The fMRI data was acquired with a repetition time (TR) of 2.1s. Critically, this dataset also contains simultaneously recorded eye-tracking data (EyeLink 1000 Plus, 250 Hz), which serves as a ground truth for validating the accuracy of the predicted gaze data from DeepMReye.

Partly Cloudy (PC) Dataset (Richardson and Saxe 2020): This dataset provided data from two distinct age groups, comprising 53 children (ages 3–12 years) and 29 adults (ages 18–39). All participants viewed a silent version of the animated short film "Partly Cloudy" (5min 36s). The fMRI scanning TR was 2.0s. This dataset does not include concurrent eye-tracking recordings.

Healthy Brain Network (HBN) Dataset (Alexander et al. 2017): This large-scale dataset includes data from children and young adults (ages 5–21 years). To avoid confounding factors from psychopathology, we included only healthy control participants. The stimuli were the same as in the NV dataset (TP and DM). The HBN data were acquired with a higher temporal resolution (TR = 0.8s), providing a more densely sampled BOLD signal. After applying quality control criteria, our final sample consisted of 82 participants for TP and 80 participants for DM. This dataset does not include concurrent eye-tracking recordings.

A summary of the three datasets is provided in Table 1. More details about the study design and MRI acquisitions can be found in the respective papers.

Table 1. Summary of datasets used in this study.

Dataset	N	Age Range (years)	Stimuli	TR (s)	Eye-tracking
NV	22	23–51	TP, DM	2.1	Yes
PC	82	3–39	PC	2	No
HBN	82/80	5–21	TP, DM	0.8	No

Note: TP = "The Present"; DM = "Despicable Me"; PC = "Partly Cloudy". For HBN, sample sizes differ between TP (n = 82) and DM (n = 80) due to data availability after quality control.

fMRI Preprocessing

All fMRI data were preprocessed using SPM12 in MATLAB following a standardized pipeline (Di and Biswal 2023). The procedures for the PC and HBN datasets have been detailed previously (Di and Biswal 2022; Di et al. 2023); a brief summary is provided here. Structural images were segmented and normalized to MNI space. For *The Present* and *Despicable Me* datasets, the first 10 functional volumes were removed (yielding 240 and 740 volumes), whereas all volumes were retained for *Partly Cloudy*. Functional images were realigned, coregistered to the structural scan, normalized using the segmentation-derived deformation fields, and smoothed with an 8-mm FWHM Gaussian kernel.

Gaze Prediction from DeepMReye

Gaze coordinates (horizontal x, vertical y) were predicted from the fMRI time series using DeepMReye, a pre-trained deep neural network The model utilizes eyeball region BOLD signals as input to estimate eye position in degrees of visual angle for each fMRI volume. To isolate the relevant input signals, a standardized eye mask provided with the DeepMReye model was applied to each participant's

preprocessed fMRI data, segmenting the voxels corresponding to the eyeball and surrounding tissues. The output was a time series of (x, y) gaze coordinates for each fMRI run.

Statistical Analysis of Gaze Movement

To quantitatively evaluate model performance, we assessed the correspondence between ground-truth eye-tracking data and predicted gaze movements using Pearson correlation analysis. For each participant and each viewing task, recorded and predicted gaze time series were temporally aligned. Individual correlation coefficients were computed separately for the horizontal (x) and vertical (y) gaze dimensions across all time points within a run. For group-level evaluations, we first computed group-averaged gaze trajectories by averaging the eye-tracking and predicted gaze time series across participants at each time point, yielding a summary time series for tracking and prediction, respectively. Pearson correlation coefficients were then calculated between the group-mean tracking and group-mean prediction trajectories for the x and y dimensions. This group-averaging approach emphasizes shared, stimulus-driven gaze dynamics by reducing idiosyncratic individual variability.

Statistical Analysis of fMRI Data

A two-level, voxel-wise general linear model (GLM) approach was employed to analyze the fMRI data in SPM12. At the first level, subject-specific effects related to gaze were modeled. At the second level, group-level effects and the relations to age and sex were assessed.

First-Level Analysis: Subject-Specific GLM

For each participant, a first-level GLM was specified for each fMRI run to identify brain regions showing activity correlated with moment-to-moment changes in gaze. The design matrix for each participant included a primary regressor of interest and several nuisance regressors. The primary regressor was a parametric modulator representing gaze displacement (d_t) calculated as the Euclidean distance between predicted gaze coordinates at successive time points:

$$d_t = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}$$

where (x_t, y_t) represents the predicted gaze position in degrees of visual angle at time point t. This displacement measure captures the magnitude of eye movement at each time point.

This displacement time series was then convolved with the canonical hemodynamic response function (HRF) to model the expected BOLD signal. To account for motion-related artifacts, 24 nuisance regressors were included in the model. These were derived from the 6 realignment parameters (3 translations and 3 rotations) obtained during preprocessing, their temporal derivatives, and quadratic terms of all 12 parameters. Following the model specification, the GLM was estimated for each subject. The full GLM equation for each voxel can be expressed as:

$$Y = \beta_{gaze} X_{gaze} + \sum_{i=1}^{24} \beta_{motion,i} X_{motion,i} + \beta_{const} X_{const} + \varepsilon$$

where Y is the BOLD signal at a given voxel, β_{gaze} is the parameter estimate for the gaze displacement regressor, $\beta_{motion,i}$ represents the parameter estimate for the i-th motion regressor, β_{const} is the intercept (mean signal level), and ε is the residual error term. This process generated a beta map for each participant, representing the effect size of the relationship between BOLD activity and gaze displacement.

Second-Level Analysis: Group-Level Multiple Regression

To make inferences at the group level, the individual beta maps from the first-level analysis were entered into a second-level, whole-brain multiple regression model. This analysis aimed to identify brain regions where gaze-related activity was significantly associated with age and sex. The model included the following covariates: age (linear term), age² (quadratic term), and sex (coded as 0 = female, 1 = male). For each voxel, the first-level beta value (B_j) for subject j was modeled as:

$$B_j = \beta_0 + \beta_1(Age_j) + \beta_2(Age_j)^2 + \beta_3(Sex_j) + \varepsilon_j$$

where β_0 is the intercept, β_1 and β_2 represent the linear and quadratic effects of age, respectively, β_3 represents the effect of sex, and ε_j is the residual error for subject j. This model allowed us to examine both linear and non-linear relationships between age and brain activity, while controlling for sex.

Region of Interest (ROI) Analysis

To further characterize age-related effects identified in the whole-brain analysis, we conducted a Region of Interest (ROI) analysis. ROIs were defined based on clusters showing significant gaze-related activity (e.g., $p < 0.001$, FWE-corrected) from the group-level analysis, representing core regions of the eye-movement network. For each participant, we extracted the mean beta value from all voxels within each ROI from their first-level results. These values were then plotted against age to visualize the developmental trajectory of gaze-related activity.

Results

DeepMReye predictions vs. Eye-tracking in NV dataset

We first validated DeepMReye predictions using the NV dataset, which included concurrent camera-based eye-tracking data as ground truth. At the group level, the model performed well: group-averaged predicted gaze time series showed high correspondence with the group-averaged ground-truth eye-tracking data. As illustrated in **Figure 1a** and **c**, group-averaged predicted and actual gaze trajectories were closely aligned. For the DM stimulus, correlations were $r = 0.84$ ($p < .001$) horizontally and $r = 0.73$ ($p < .001$) vertically. Similarly, for TP stimulus, Pearson correlations were $r = 0.81$ ($p < .001$) for both horizontal and vertical gaze directions. These results demonstrate that DeepMReye reliably captures the group-level patterns during naturalistic viewing.

However, prediction accuracy declined substantially at the individual level. **Figure 1b** and **d** present raincloud plots of individual-participant Pearson correlation coefficients between predicted and ground-truth gaze for horizontal (x) and vertical (y) directions. Individual-level correlations showed greater variability and were generally lower than those observed at the group level. For the DM stimulus, median correlations were r = 0.32 (IQR = 0.26–0.47) for horizontal and r = 0.28 (IQR = 0.17–0.34) for vertical gaze directions. For the TP stimulus, median correlations were r = 0.37 (interquartile range [IQR] = 0.16–0.47) for horizontal and r = 0.24 (IQR = 0.15–0.33) for vertical gaze directions. While a subset of participants exhibited relatively high correlations between predicted and ground-truth gaze, many others

showed substantially lower correspondence. Together, these results indicate that individual-level prediction performance is more variable and generally weaker than group-level performance.

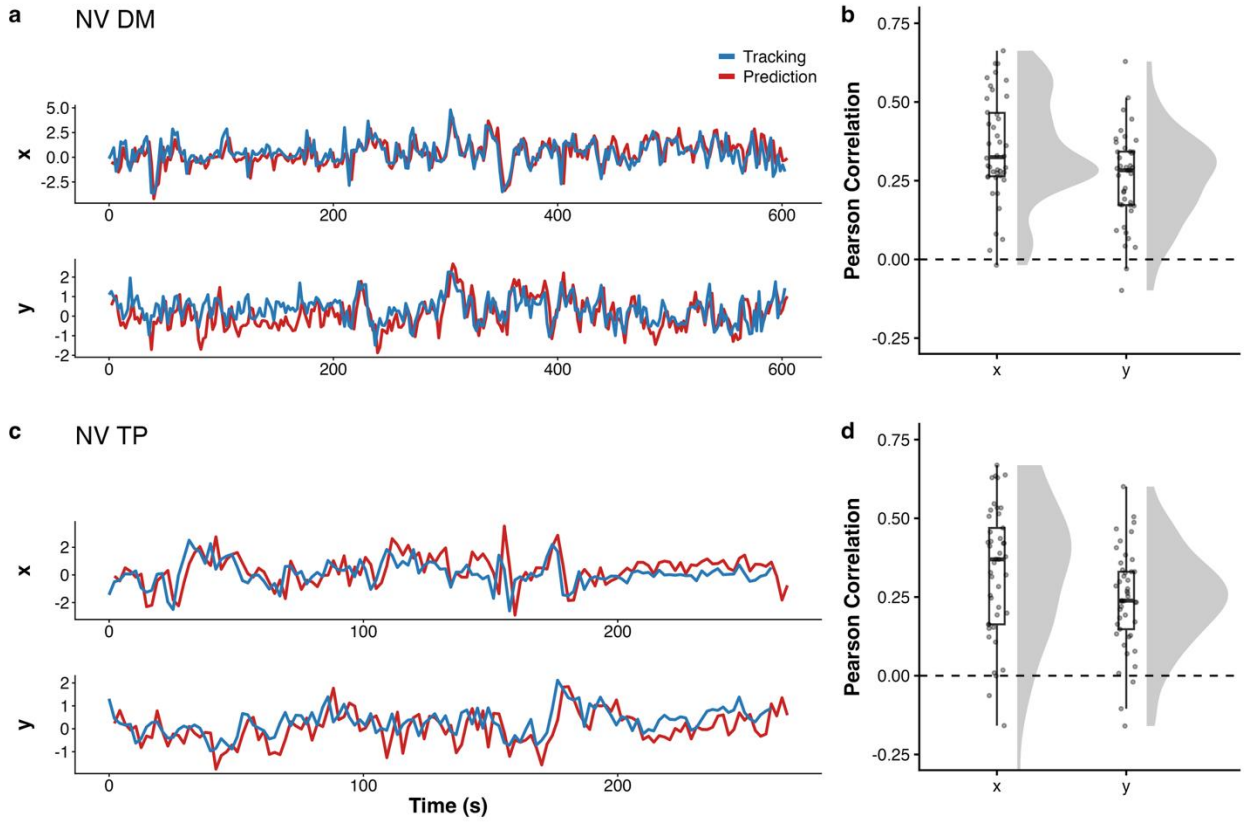

Figure 1. Validation of DeepMReye gaze predictions against ground-truth eye-tracking data. (a) Group-averaged eye-tracking (blue) and predicted gaze positions (red) over time for the x and y coordinates in the NV DM task, with Pearson correlations of 0.84 and 0.73, respectively. (b) Distribution of individual-level Pearson correlations between tracking and predicted positions for x and y coordinates in NV DM. (c) Group-averaged eye-tracking (blue) and predicted gaze positions (red) over time for x and y coordinates in the NV TP task, with Pearson correlations of 0.81 for both dimensions. (d) Distribution of individual-level Pearson correlations for x and y coordinates in NV TP.

Cross-Dataset Validation between NV and HBN datasets

To further assess the generalizability of DeepMReye predictions, we performed a cross-dataset validation by comparing group-averaged predictions from the HBN dataset with ground-truth eye-tracking data from

the NV dataset. Both datasets used identical stimuli (TP and DM), allowing direct comparison despite differences in participant samples and scanning parameters. As shown in **Figure 2a** and **d**, the predicted gaze trajectories from HBN participants closely matched the ground-truth data from NV participants. For the DM stimulus, Pearson correlations were $r = 0.87$ for horizontal and $r = 0.80$ for vertical gaze directions. For the TP stimulus, correlations were $r = 0.90$ for horizontal and $r = 0.86$ for vertical gaze directions. These results indicate strong correspondence between group-averaged DeepMReye predictions and ground-truth gaze trajectories across independent datasets.

We further quantified this relationship by computing the correlation between each participant's predicted gaze trajectory and the NV group-mean tracking gaze trajectory. During the DM task, positive associations were observed for both horizontal ($r = 0.068$, $p = 0.524$) and vertical ($r = 0.087$, $p = 0.412$) gaze components, though these effects did not reach statistical significance (**Figure 2b** and **c**). In contrast, during the TP task, stronger age-related increases in correlation were observed, particularly for the vertical gaze dimension ($r = 0.298$, $p = 0.004$), with a weaker trend for the horizontal component ($r = 0.181$, $p = 0.085$) (**Figure 2e** and **f**).

These results indicate that individual gaze predictions increasingly align with normative group-level viewing patterns as age increases, especially under structured viewing conditions. This finding suggests improved model generalization and developmental stabilization of gaze behavior in later childhood and adolescence.

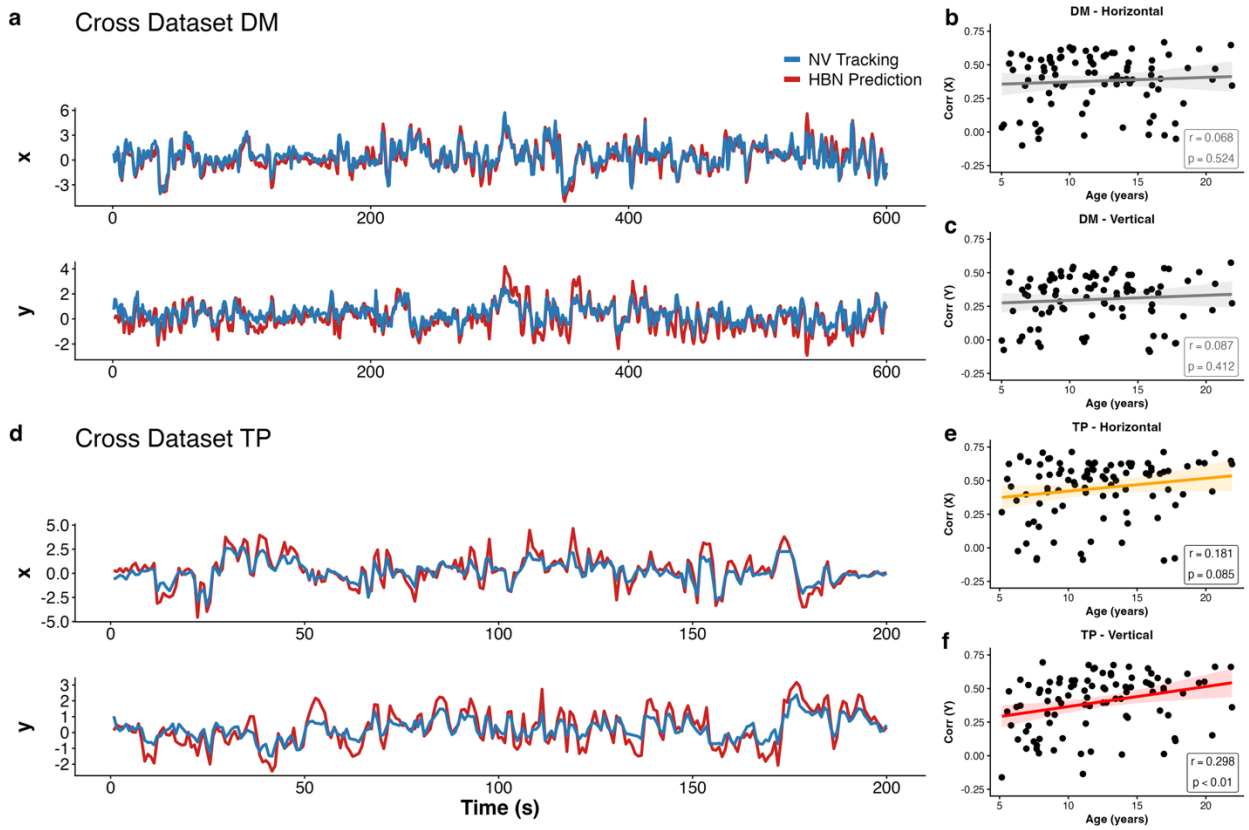

Figure 2. Cross-dataset comparison of group-mean gaze trajectories and age-related correlation patterns between NV ground-truth and HBN-predicted gaze.

(a–c) Group and individual level comparisons for horizontal (x) and vertical (y) gaze coordinates between NV group-averaged ground-truth tracking (blue) and HBN gaze predictions (red) during the DM task. (d–f) Corresponding results for the TP task. Panels (b, c, e, f) show the correlation between the NV group-mean tracking gaze and the individual HBN predictions as a function of age. Each dot represents one participant. Solid lines indicate linear regression fits, and shaded regions represent 95% confidence intervals. Red lines indicate statistically significant age effects ($p < 0.05$), whereas orange lines indicate marginally significant effects ($p < 0.1$). Pearson correlation coefficients (r) and associated p-values are shown in each panel.

Intersubject correlations of predicted gaze in PC dataset

To further examine developmental effects on gaze behavior, we analyzed intersubject similarity and variability in the PC dataset (**Figure 3**). Correlation matrices revealed increasingly structured alignment between the same group of individual predictions with increasing age (**Figure 3a** and **b**). The distinct high-correlation cluster among adults in the lower-right quadrant of the matrices, together with the weaker correlations involving younger participants, highlights age-related differences in gaze synchrony during movie viewing.

Quantitative analyses demonstrated a significant positive relationship between age and gaze similarity for both horizontal ($r = 0.352$, $p = 0.001$) and vertical ($r = 0.391$, $p < 0.001$) components (**Figure 3c** and **d**), indicating that older participants exhibited stronger correspondence with normative gaze patterns. It suggests that there are significant age-related increases in gaze, potentially reflecting maturation of visual exploration strategies and improved coordination between attentional control and eye movement dynamics.

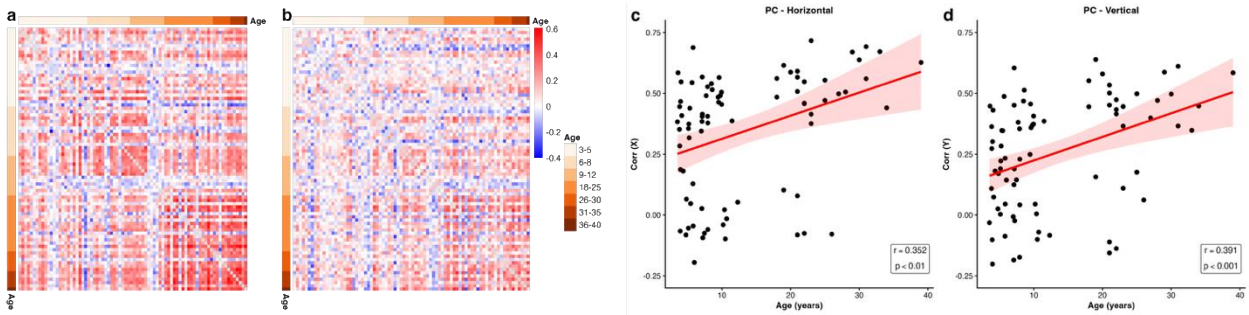

Figure 3. Intersubject Similarity and Variability of Gaze Patterns in the Partly Cloudy Dataset. (a–b) Intersubject correlation matrices within the PC dataset revealed increasingly structured alignment of predicted gaze trajectories with increasing age for horizontal (a) and vertical (b) gaze components. (c–d) Scatter plots showing the relationship between age and correlation strength for horizontal (c) and vertical (d) gaze components. Solid lines indicate linear regression fits, and shaded regions represent 95% confidence intervals. Red lines indicate statistically significant age effects ($p < 0.05$).

Brain activations associated with Eye Movement

To identify the core neural network associated with eye movements during naturalistic movie viewing, we construct eye-movement time series by calculating the Euclidean distance between gaze positions between consecutive time points, using either individual-level predicted gaze or group-averaged gaze predictions. Group-level analyses were performed separately for each movie.

As illustrated in **Figure 4a** and **c**, group-averaged eye-movement regressors produced highly similar whole-brain activation patterns for the HBN dataset during the DM task and for the PC dataset during movie viewing. Robust activation was consistently observed in a distributed oculomotor control network, with prominent involvement of frontal and parietal regions implicated in the planning and control of eye movements, including the frontal eye fields (FEF) and the intraparietal sulcus (IPS). These regions constitute core components of the dorsal attention and oculomotor networks that support saccade generation, visuospatial attention, and top-down control of visual exploration.

In addition to this control network, widespread activation was observed in occipital visual cortices, encompassing the cuneus, lingual gyrus, and primary visual cortex, reflecting the expected engagement of early visual processing during continuous gaze shifts. Notably, despite substantial differences between the HBN and PC datasets—including movie stimuli, participant age distributions, and fMRI acquisition parameters—the oculomotor control network associated with gaze displacement was largely conserved across datasets. This robust main effect provides a stable foundation for subsequent analyses examining how age modulates gaze-related brain activity. Compared with the DM task, the TP task in the HBN dataset exhibited fewer and more spatially restricted activation clusters at the same statistical threshold (**Figure 4b**), potentially reflecting reduced statistical power due to shorter task duration and more constrained viewing dynamics.

In contrast, regressors derived from individual-level eye-movement predictions resulted in significantly smaller and more localized activation clusters across all three conditions. As illustrated in **Figures 4d** and **4e**, the HBN-DM and HBN-TF movie-watching tasks still showed positive activation in visual cortical regions (e.g., lingual gyrus and cuneus), but these effects were substantially less extensive.

The PC task showed minimal and sparse activation at the same statistical threshold (**Figure 4f**). Overall, individual-level regressors yielded less extensive and less consistent group-level activation patterns compared to group-averaged regressors.

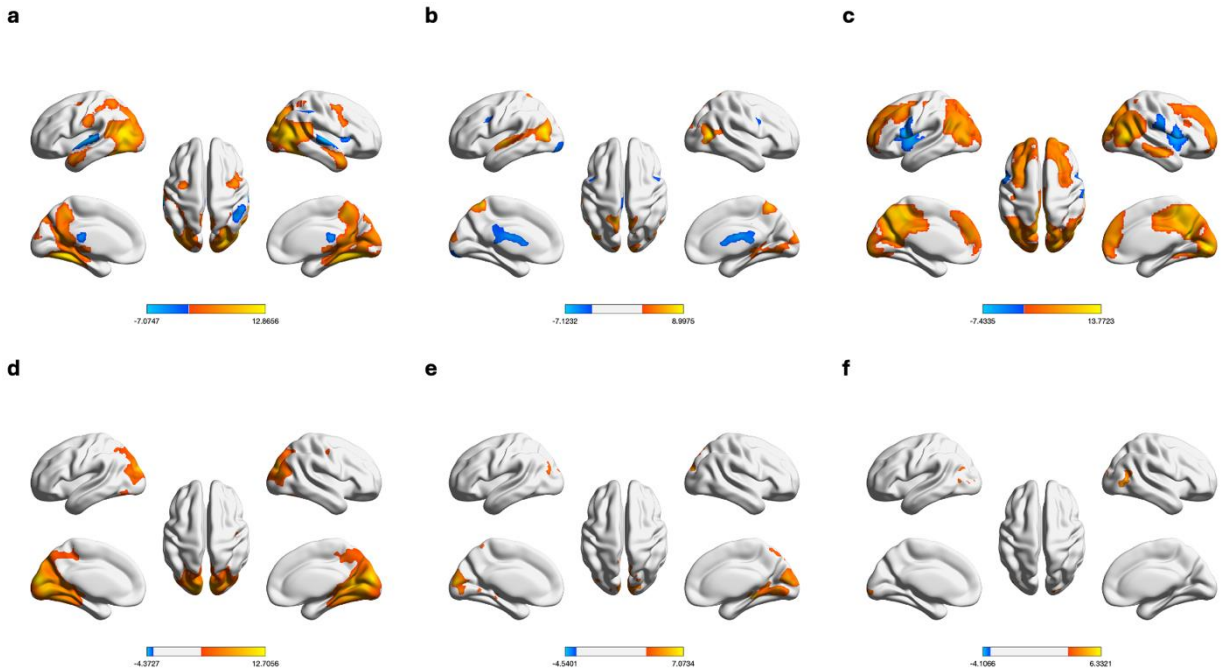

Figure 4. Main Effect of Gaze Displacement on Whole-Brain Activation. The figure shows regions where brain activity was significantly and positively correlated with individual and group mean of eye movement difference. Activation using group-averaged gaze prediction of (a) subjects in the HBN dataset DM task, (b) subjects in the HBN dataset TP task, and (c) subjects in the PC dataset. Activation using individual gaze prediction of (d) subjects in the HBN dataset DM task, (e) subjects in the HBN dataset TP task, and (f) subjects in the PC dataset. Warm colors indicate a stronger positive correlation, and cold colors indicate negative correlations.

Age Effects on Eye-Movement related Brain activations

We further examined whether eye-movement–related brain activity was modulated by age by including age as a covariate in the second-level analyses. Age effects were evaluated separately for each dataset and for regressors derived from group-averaged versus individual-level gaze predictions. We found that the

relationship between age and brain activity demonstrated different patterns across the two datasets, which featured different movie stimuli and participant cohorts. Similarly, we also compared the activation results derived from the beta values generated based on individual and group mean predictions.

Age Effects in the Partly Cloudy (PC) Dataset

In the PC dataset, group-level analyses using group-averaged gaze predictions revealed significant positive age-related effects in the Postcentral Gyrus, the Precuneus, and the Middle Frontal Gyrus (**Figure 5a**). This pattern findings systematic age-related modulation of gaze-related neural activity in regions implicated in sensory integration, visuospatial processing, and higher-order cognitive control during viewing of this stimulus. In contrast, no significant age-related activation was detected in analyses based on individual-level gaze predictions. When individual-level gaze predictions were used, age-related effects were weaker and spatially more restricted, highlighting the greater robustness of group-averaged gaze estimates for detecting developmental effects.

Age Effects in the Healthy Brain Network (HBN) Dataset

The DM task in the HBN dataset, which involved viewing a clip from movie DM with complex dialogue and social scenes, revealed distributed pattern of age-related effects based on group-averaged gaze predictions (Figure **5b**). Significant positive age-related associations were observed primarily in posterior visual regions, including the lingual gyrus and cuneus, as well as in the middle frontal gyrus. These regions are implicated in visual processing, visuospatial integration, and higher-order cognitive control, suggesting age-related modulation of gaze-related neural activity during naturalistic viewing. In contrast, no significant age-related effects were detected for the TP task.

ROI analysis

To assess age-related effects within these regions, we performed ROI-based analyses using quadratic generalized linear models, with mean beta values extracted from each ROI as the dependent variable and both linear (Age) and quadratic (Age^2) terms included as predictors (Mean Beta ~ Age + Age^2). Models were fitted separately for each ROI to characterize potential non-linear developmental trajectories.

In the PC dataset, quadratic modeling revealed significant age-related curvature in several ROIs, as indicated by a significant quadratic (Age^2) term (Figure 5c). Specifically, gaze-related activation within these regions exhibited a non-linear developmental trajectory, characterized by higher activation in younger participants followed by a gradual decrease with increasing age. This pattern suggests that age-related changes in gaze-related brain activity in the PC dataset are not well described by a simple linear trend but instead reflect systematic non-linear modulation across development. In the HBN DM dataset, ROI-wise quadratic models likewise revealed significant quadratic age effects in multiple ROIs (Figure 5d). In contrast to the PC dataset, these regions showed inverted U-shaped relationships, with gaze-related activation increasing from childhood into adolescence and subsequently decreasing into young adulthood. The presence of significant quadratic terms indicates that age-related modulation of gaze-related neural activity during the DM task varies across developmental stages rather than progressing monotonically with age. Taken together, these ROI-based analyses demonstrate that age-related changes in gaze-related brain activity are non-linear and context-dependent, differing across datasets and stimulus conditions. Whereas the PC dataset is characterized primarily by a decreasing trajectory with age, the HBN DM dataset exhibits developmentally specific peak activation during adolescence, underscoring the importance of modeling non-linear age effects when examining gaze-related neural development.

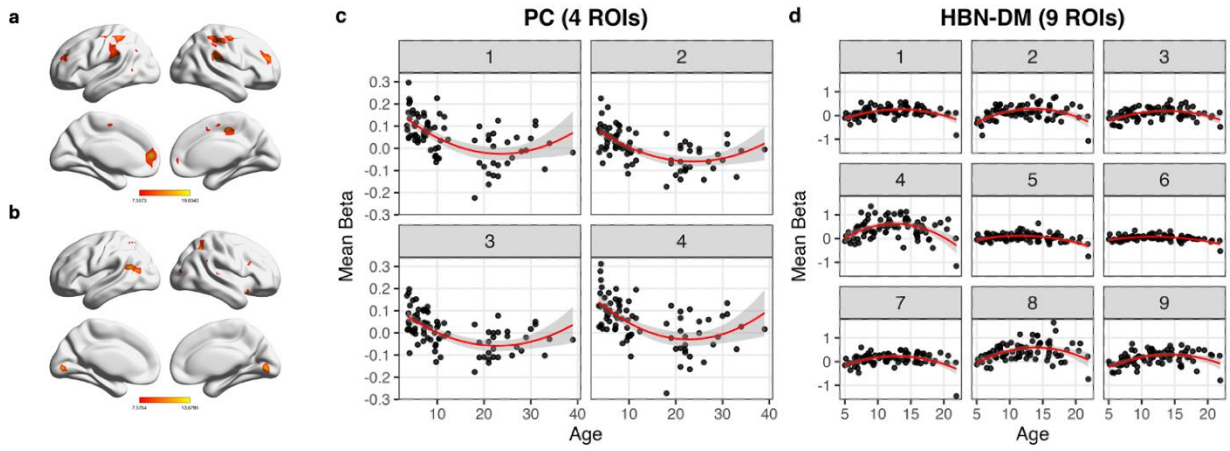

Figure 5. Modulation of Gaze-Related Brain Activity by Age. (a) and (b) shows regions where brain activity exhibited a significant positive correlation with age using group-averaged gaze prediction of subjects in the PC dataset and the HBN dataset DM task. Scatter plots of mean beta values are versus age in significant ROIs based on results from the PC dataset (c) and the HBN dataset DM task (d). Solid lines depict ROI-wise quadratic fits estimated using generalized linear models (Mean Beta ~ Age + Age^2), with shaded areas indicating 95% confidence intervals.

Discussion

In this study, we evaluated the ability of DeepMReye to recover gaze-related neural and behavioral patterns during naturalistic movie viewing across independent datasets spanning childhood to adulthood. At the behavioral level, DeepMReye predictions closely matched ground-truth eye-tracking data at the group level, both within and across datasets, while individual-level prediction accuracy was more variable. At the neural level, gaze-related activity consistently engaged a canonical oculomotor control network encompassing frontal and parietal regions, including the frontal eye fields and intraparietal sulcus, alongside expected activation in visual cortices. Importantly, age-related modulation of gaze-related brain activity differed across datasets, with non-linear developmental trajectories observed in ROI-based analyses.

As established in our validation, the group-averaged predictions demonstrated substantially higher fidelity to the ground-truth data. By averaging across participants, idiosyncratic variations and model prediction noises are attenuated, thereby enhancing the shared, stimulus-driven component of the gaze signal. This improvement in signal-to-noise ratio was directly mirrored in our neuroimaging results: brain activation maps derived from the group-mean regressor revealed a comprehensive and neuroanatomically plausible network, whereas maps from individual regressors yielded more restricted and variable activations.

The strong performance observed at the group level likely reflects the dominant influence of shared, stimulus-driven gaze patterns in naturalistic viewing. Prior work has demonstrated that movies

strongly synchronize attention and gaze across viewers, producing robust inter-subject correlations in both behavior and neural responses (Hasson et al. 2004; Dorr et al. 2010; Mital et al. 2011). By averaging predicted gaze across participants, idiosyncratic noise arising from anatomical variability, voxel-level signal heterogeneity in the orbit, and motion-related artifacts is attenuated, yielding a clearer estimate of the shared component of gaze behavior. These findings highlight a key strength of fMRI-based gaze prediction: despite the relatively coarse spatiotemporal resolution of fMRI, the shared component of visual attention elicited by naturalistic stimuli can be reliably recovered using DeepMReye.

In contrast, predictive correspondence at the individual level was markedly lower and more variable. Multiple factors may contribute to this limitation, including inter-individual differences in eye anatomy, orbit geometry, EPI distortions, and head motion, all of which can affect the extraction of reliable features from signals surrounding the eyeballs (Petit and Haxby 1999; Beauchamp 2003). In addition, because the pre-trained model was not fine-tuned on our datasets, differences in scanner characteristics, acquisition parameters, and population demographics may reduce generalizability (Tran et al. 2025; Botvinik-Nezer et al. 2020). Further, the temporal resolution of fMRI constrains the detection of rapid eye movements, making it difficult to capture saccades or subject-specific patterns of visual exploration (Engbert and Kliegl 2003). Together, these factors suggest that while the model captures shared, stimulus-driven gaze components, it does not yet adequately capture individualized gaze dynamics. This limitation is further supported by the weak or absent eye-movement–related activations observed when individual-level predictions were used.

These findings have important methodological implications for fMRI studies that lack camera-based eye tracking. Group-averaged gaze predictions can serve as a practical substitute for hardware-based eye tracking in studies focused on population-level gaze dynamics or stimulus-driven attentional processes, particularly in large naturalistic datasets where eye tracking was not acquired (e.g., HCP, NSD). However, the current approach is not suitable for applications requiring precise individual-level gaze measurement, such as quality control of fixation compliance, fine-grained behavioral analyses, or clinical assessments at the single-subject level, where camera-based eye tracking remains essential.

A central neuroscientific finding of this study is that age-related modulation of gaze-associated brain activity is non-linear and highly context dependent. In the HBN dataset, which spanned childhood through young adulthood and involved viewing a socially complex movie clip (DM), ROI analyses inverted U-shaped developmental trajectories in multiple regions within a distributed posterior–frontal network. These trajectories were characterized by increasing gaze-related activation from childhood into adolescence, followed by a subsequent decrease into young adulthood. Such patterns are consistent with prior work highlighting non-linear maturation of attentional and visual systems across development (Madden 2007). In contrast, the PC dataset exhibited a different pattern of age-related modulation, with gaze-related activity showing a predominantly decreasing trajectory across the sampled age range. Given the limited adolescent coverage in this dataset, these effects likely reflect differences between age groups rather than a fully resolved continuous developmental trajectory. Together, these findings emphasize that apparent age effects in naturalistic viewing are strongly shaped by both stimulus properties and sampling structure and are not adequately captured by simple linear models.

The presence of age-related differences in group-level analyses primarily suggests that the model may be sensitive to broad developmental trends in gaze behavior (Fernandes et al. 2024), but insufficiently precise to detect subject-specific effects amid individual noise. Whether these age effects reflect true changes in visual attention or methodological artifacts remains unclear. Future work should examine how demographic variables influence model performance and whether targeted model adaptation can improve sensitivity to individual differences.

Our findings revealed a notable dissociation between horizontal and vertical eye movement synchronization across development. While horizontal eye movement synchrony remained relatively stable across age groups, vertical eye movement synchrony showed a significant positive correlation with age ($r = 0.298$, $p = 0.004$). This differential developmental trajectory may reflect the distinct neural substrates underlying horizontal and vertical saccade generation. The prolonged maturation of vertical eye movement control suggested by these findings aligns with previous literature suggesting that certain oculomotor functions continue to develop throughout childhood and adolescence (Grönqvist et al. 2006).

These results indicate that age-related differences in gaze-related brain activity are strongly shaped jointly by the developmental stage and the stimulus context. Group-averaged model-derived gaze predictions are sufficiently robust to reveal population-level developmental patterns, whereas individual-level predictions remain too noisy to support reliable subject-specific inference. This highlights an important practical boundary of current fMRI-based gaze decoding approaches: deep-learning–based MRI gaze decoding may be valuable for identifying population-level developmental trajectories, but its current accuracy limitations constrain its utility for fine-grained individual differences work.

Several limitations should be noted. Our analyses were based on a limited number of movie stimuli and datasets, with camera-based eye tracking available only for one dataset. The use of a pre-trained model without dataset-specific fine-tuning may have constrained performance particularly in populations that differ substantially from the training sample. In addition, the temporal resolution of fMRI inherently limits the model's sensitivity to rapid eye-movement events. These limitations motivate further methodological innovation.

Future work should explore strategies for improving individual-level prediction accuracy, including incorporating subject-specific calibration (Kassner et al. 2014), retraining or fine-tuning models on larger or more diverse datasets (He et al. 2016), integrating high-resolution structural data of the orbit, and developing hybrid models that jointly use anatomical and functional information. Future work should also explore extending fMRI-based gaze decoding to predict discrete gaze events (e.g., fixations and saccades) rather than continuous position estimates, as well as integrating multimodal signals such as pupillometry or eyelid dynamics. Such improvements will be essential for enabling individualized inference and expanding the applicability of fMRI-based gaze prediction to clinical and developmental research, particularly for studying individual differences in clinical populations such as autism spectrum disorder (ASD) or attention-deficit/hyperactivity disorder (ADHD), where gaze behavior represents a key behavioral marker.

Conclusion

In summary, this study demonstrates that fMRI-based gaze prediction provides a viable and powerful framework for investigating gaze-related brain activity at the group level during naturalistic viewing, while also highlighting current limitations for reliable individual-level inference. By validating the DeepMReye model across independent developmental datasets, we show that model-derived gaze estimates recover biologically meaningful oculomotor and attentional networks and reveal systematic, age-related modulation of gaze-related brain activity. Importantly, these developmental effects are not uniform across contexts but vary as a function of stimulus properties, task demands, and population characteristics, underscoring the context-dependent nature of neurodevelopment during naturalistic viewing. Together, these findings highlight the value of integrating data-driven computational approaches with large-scale developmental neuroimaging datasets to move beyond static, task-isolated paradigms, offering a promising path toward understanding how dynamic visual behavior and its neural substrates evolve across development, particularly in settings where traditional eye tracking is unavailable.

Data and Code Availability

MRI and eye-tracking data were obtained from publicly available sources: the Natural Viewing (NV) dataset (https://fcon_1000.projects.nitrc.org/indi/retro/nat_view.html), the Healthy Brain Network (HBN) dataset (http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/), and the OpenNeuro repository (https://openneuro.org/datasets/ds000228). Analysis scripts will be shared on OSF.

Author Contributions

L.G. designed the study, implemented the gaze prediction, performed the analyses, and drafted the manuscript. X.D. provided overall conceptual guidance and contributed to the study design. B.B.B and Z.W. offered critical feedback and revisions throughout the project. B.B.B. and X.D. provided resources and secured funding for the project. All authors contributed to the interpretation of results and approved the final version of the manuscript.

Funding

This study was supported by grants from (US) National Institutes of Health for Xin Di (R15MH125332) and Bharat B. Biswal (5R01MH131335, 5R01AG085665 and 4R01NS124778) and by (NJ) Governor's Council for Medical Research and Treatment of Autism for Xin Di (CAUT25BRP005).

Declaration of Competing Interests

The authors declare that there is no competing interest.

Reference

Alexander, Lindsay M., Jasmine Escalera, Lei Ai, et al. 2017. "An Open Resource for Transdiagnostic Research in Pediatric Mental Health and Learning Disorders." *Scientific Data* 4 (1): 1–26. https://doi.org/10.1038/sdata.2017.181.

Beauchamp, Michael S. 2003. "Detection of Eye Movements from fMRI Data." *Magnetic Resonance in Medicine* 49 (2): 376–80. https://doi.org/10.1002/mrm.10345.

Botvinik-Nezer, Rotem, Felix Holzmeister, Colin F. Camerer, et al. 2020. "Variability in the Analysis of a Single Neuroimaging Dataset by Many Teams." *Nature*, May 20, 1–7. https://doi.org/10.1038/s41586-020-2314-9.

Cantlon, Jessica F., and Rosa Li. 2013. "Neural Activity during Natural Viewing of Sesame Street Statistically Predicts Test Scores in Early Childhood." *PLOS Biology* 11 (1): e1001462. https://doi.org/10.1371/journal.pbio.1001462.

Casey, B. J., Nim Tottenham, Conor Liston, and Sarah Durston. 2005. "Imaging the Developing Brain: What Have We Learned about Cognitive Development?" *Trends in Cognitive Sciences*, Special issue: Developmental cognitive neuroscience, vol. 9 (3): 104–10. https://doi.org/10.1016/j.tics.2005.01.011.

Di, Xin, and Bharat B. Biswal. 2022. "Principal Component Analysis Reveals Multiple Consistent Responses to Naturalistic Stimuli in Children and Adults." *Human Brain Mapping* 43 (11): 3332–45. https://doi.org/10.1002/hbm.25568.

Di, Xin, and Bharat B. Biswal. 2023. "A Functional MRI Pre-Processing and Quality Control Protocol Based on Statistical Parametric Mapping (SPM) and MATLAB." *Frontiers in Neuroimaging* 1: 1070151.

Di, Xin, George B. Hanna, and Bharat B. Biswal. 2025. "Timing Is Everything: Temporal Dynamics of BOLD Responses to Naturalistic Features in Movie Watching." Preprint, bioRxiv, November 10. https://doi.org/10.1101/2025.11.07.687226.

Di, Xin, Ting Xu, Lucina Q. Uddin, and Bharat B. Biswal. 2023. "Individual Differences in Time-Varying and Stationary Brain Connectivity during Movie Watching from Childhood to Early Adulthood: Effects of Age, Sex, and Behavioral Associations." Preprint, bioRxiv, February 2. https://doi.org/10.1101/2023.01.30.526311.

Dorr, Michael, Thomas Martinetz, Karl R. Gegenfurtner, and Erhardt Barth. 2010. "Variability of Eye Movements When Viewing Dynamic Natural Scenes." *Journal of Vision* 10 (10): 28. https://doi.org/10.1167/10.10.28.

Engbert, Ralf, and Reinhold Kliegl. 2003. "Microsaccades Uncover the Orientation of Covert Attention." *Vision Research* 43 (9): 1035–45. https://doi.org/10.1016/S0042-6989(03)00084-1.

Fernandes, Eunice G., Benjamin W. Tatler, Gillian Slessor, and Louise H. Phillips. 2024. "Age Differences in Gaze Following: Older Adults Follow Gaze More than Younger Adults When Free-Viewing Scenes." *Experimental Aging Research* 50 (1): 84–101. https://doi.org/10.1080/0361073X.2022.2156760.

Frey, Markus, Matthias Nau, and Christian F. Doeller. 2021. "Magnetic Resonance-Based Eye Tracking Using Deep Neural Networks." *Nature Neuroscience* 24 (12): 12. https://doi.org/10.1038/s41593-021-00947-w.

Grönqvist, Helena, Gustaf Gredebäck, and Claes von Hofsten. 2006. "Developmental Asymmetries between Horizontal and Vertical Tracking." *Vision Research* 46 (11): 1754–61. https://doi.org/10.1016/j.visres.2005.11.007.

Hasson, Uri, Yuval Nir, Ifat Levy, Galit Fuhrmann, and Rafael Malach. 2004. "Intersubject Synchronization of Cortical Activity during Natural Vision." *Science (New York, N.Y.)* 303 (5664): 1634–40. https://doi.org/10.1126/science.1089506.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. "Deep Residual Learning for Image Recognition." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June, 770–78. https://doi.org/10.1109/CVPR.2016.90.

Henderson, John M. 2003. "Human Gaze Control during Real-World Scene Perception." *Trends in Cognitive Sciences* 7 (11): 498–504. https://doi.org/10.1016/j.tics.2003.09.006.

Kassner, Moritz, William Patera, and Andreas Bulling. 2014. "Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-Based Interaction." *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, September 13, 1151–60. https://doi.org/10.1145/2638728.2641695.

Kirkorian, Heather L., Daniel R. Anderson, and Rachel Keen. 2012. "Age Differences in Online Processing of Video: An Eye Movement Study." *Child Development* 83 (2): 497–507. https://doi.org/10.1111/j.1467-8624.2011.01719.x.

Luna, Beatriz, Scott Marek, Bart Larsen, Brenden Tervo-Clemmens, and Rajpreet Chahal. 2015. "An Integrative Model of the Maturation of Cognitive Control." *Annual Review of Neuroscience* 38 (July): 151–70. https://doi.org/10.1146/annurev-neuro-071714-034054.

Madden, David J. 2007. "Aging and Visual Attention." *Current Directions in Psychological Science* 16 (2): 70–74. https://doi.org/10.1111/j.1467-8721.2007.00478.x.

Madsen, Jens, and Lucas C Parra. 2022. "Cognitive Processing of a Common Stimulus Synchronizes Brains, Hearts, and Eyes." *PNAS Nexus* 1 (1): pgac020. https://doi.org/10.1093/pnasnexus/pgac020.

Mital, Parag K., Tim J. Smith, Robin L. Hill, and John M. Henderson. 2011. "Clustering of Gaze During Dynamic Scene Viewing Is Predicted by Motion." *Cognitive Computation* 3 (1): 5–24. https://doi.org/10.1007/s12559-010-9074-z.

Petit, Laurent, and James V. Haxby. 1999. "Functional Anatomy of Pursuit Eye Movements in Humans as Revealed by fMRI." *Journal of Neurophysiology* 82 (1): 463–71. https://doi.org/10.1152/jn.1999.82.1.463.

Petroni, Agustin, Samantha S. Cohen, Lei Ai, et al. 2018. "The Variability of Neural Responses to Naturalistic Videos Change with Age and Sex." *eNeuro* 5 (1). https://doi.org/10.1523/ENEURO.0244-17.2017.

Pierrot-Deseilligny, Charles, Dan Milea, and René M. Müri. 2004. “Eye Movement Control by the Cerebral Cortex.” *Current Opinion in Neurology* 17 (1): 17.

Richardson, Hilary, and Rebecca Saxe. 2020. “Development of Predictive Responses in Theory of Mind Brain Regions.” *Developmental Science* 23 (1): e12863. https://doi.org/10.1111/desc.12863.

Son, Jake, Lei Ai, Ryan Lim, et al. 2020. “Evaluating fMRI-Based Estimation of Eye Gaze During Naturalistic Viewing.” *Cerebral Cortex* 30 (3): 1171–84. https://doi.org/10.1093/cercor/bhz157.

Sonkusare, Saurabh, Michael Breakspear, and Christine Guo. 2019. “Naturalistic Stimuli in Neuroscience: Critically Acclaimed.” *Trends in Cognitive Sciences* 23 (8): 699–714. https://doi.org/10.1016/j.tics.2019.05.004.

Tatler, Benjamin W., Mary M. Hayhoe, Michael F. Land, and Dana H. Ballard. 2011. “Eye Guidance in Natural Vision: Reinterpreting Salience.” *Journal of Vision* 11 (5): 5. https://doi.org/10.1167/11.5.5.

Telesford, Qawi K., Eduardo Gonzalez-Moreira, Ting Xu, et al. 2023. “An Open-Access Dataset of Naturalistic Viewing Using Simultaneous EEG-fMRI.” *Scientific Data* 10 (1): 1. https://doi.org/10.1038/s41597-023-02458-8.

Tran, Anh T., Tal Zeevi, and Seyedmehdi Payabvash. 2025. “Strategies to Improve the Robustness and Generalizability of Deep Learning Segmentation and Classification in Neuroimaging.” *BioMedInformatics* 5 (2): 20. https://doi.org/10.3390/biomedinformatics5020020.