Neon Accuracy Test Report

An Evaluation of Lab and In-The-Wild Performance

Chris Baumann and Kai Dierkes
December 22, 2023

Pupil Labs

Abstract. *We present the results of a validation study, assessing the gaze-estimation accuracy of the Neon eye-tracking system by Pupil Labs. We provide results obtained from, 1) a screen-based experiment conducted in a controlled lab environment, and 2) from 'in-the-wild' scenarios. In particular, we show that – without a user-specific calibration – Neon achieves a per-subject median accuracy of 1.8° at 1.3 m depth and across a 60°×35° field of view. By applying a simple offset correction, median per-subject accuracy can be further improved to 1.3°. Furthermore, our results highlight that Neon operates robustly in lighting conditions ranging from total darkness to bright sunlight, for a representative spectrum of eye appearances, as well as for realistic variations in position on the wearer's head. For instance, we show that median accuracy stays within 0.2°– 0.4° between indoor and outdoor settings.*

Citation: *Chris Baumann and Kai Dierkes. Neon Accuracy Test Report. Pupil Labs, 2023. doi: 10.5281/zenodo.10420388.*

1 Introduction

Eye tracking is a well-established methodology used in various research and application fields, including psychology [9], sports research [23, 6, 21], human-computer interaction [7], market research [12, 11], and medical diagnostics [4]. Head-mounted eye trackers, in particular, provide valuable insights into human cognition and physiology in dynamic everyday situations. By allowing researchers and engineers to move beyond the laboratory, these eye trackers enable e.g. the study of social interactions in real-world environments [20, 10], the analysis of human behavior in complex traffic scenarios [18], and even the prevention of falls among elderly individuals in their homes [22, 16, 17].

Traditional head-mounted eye trackers have faced several challenges. These include the need for frequent recalibration after movement of the device relative to the wearer's head, called slippage, and data loss due to sensitivity to direct sunlight.

A new generation of deep-learning-based mobile eye trackers, such as the Neon eye-tracking system offered by Pupil Labs, addresses these challenges. Neon provides accurate gaze estimates without the need for calibration. Leveraging the power of neural networks trained on a diverse dataset, Neon delivers robust eye tracking in all lighting conditions, ranging from total darkness to bright sunlight. Additionally, Neon is resistant to slippage, making it suitable for dynamic use cases.

This report presents the results of a validation study gauging the gaze-estimation accuracy of Neon. We will begin by providing a detailed overview of the Neon eye-tracking system. Then, we will present the accuracy results obtained in two different contexts: a controlled in-lab screen task and an in-the-wild setting. Finally, we will discuss the implications of our findings for different real-world use cases.

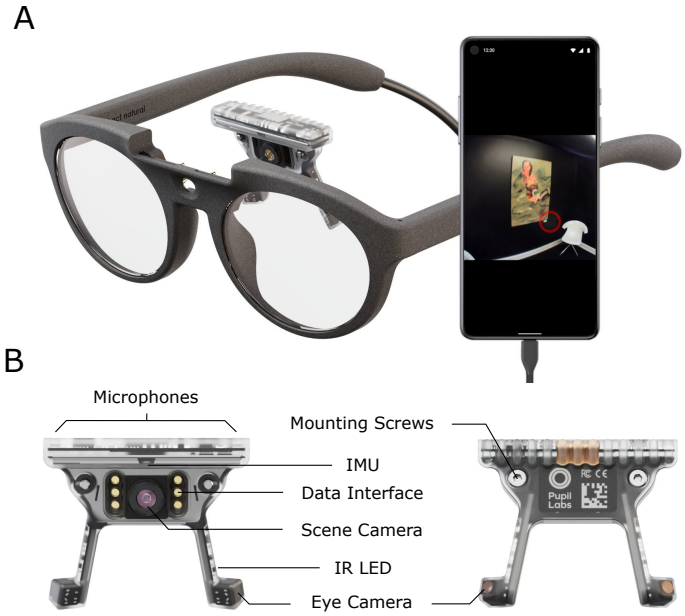

Figure 1: A) The Neon eye-tracking system by Pupil Labs. The Neon module (center, not mounted) contains all eye-tracking and additional sensors and can be combined with different frames depending on the use case. The picture shows the "Just act natural" frame. The headset is connected to a mobile phone running the Neon Companion app, which provides real-time gaze estimation. **B) The Neon module.** A front-facing camera records scene video, while two rear-facing cameras record close-up videos of the eyes. The module is also equipped with a stereo microphone and an inertial measurement unit (IMU).

2 Neon – An Overview

The Neon eye-tracking system consists of the Neon module that can be combined with various frames for different use cases (see Figure 1A).

The Neon module is the central electronic unit, containing all cameras and sensors in a compact form factor (see Figure 1B). It includes a front-facing scene camera that delivers a 30 Hz video stream with a field of view of 132° x 81° and a resolution of 1600 x 1200 pixels. Two infrared cameras capture videos of eye movements at 200 Hz with a resolution of 192 x 192 pixels. The module also features a stereo microphone and an inertial measurement unit (IMU) that reports accelerator, magnetometer, and gyroscope values at 110 Hz. The entire module is encased in bio-compatible silicone for water and sweat resistance.

A range of distinct frames for mounting the Neon module are available. In addition to the general-purpose frame "Just act natural," there are other specialized frames designed for sports and safety ("Better safe than sorry"), frames for children aged 2 to 8 ("All fun and games"), and a frame with swappable diopter lenses ("I can see clearly now"). Additionally, integrations for XR headsets have been released with more in development. Users can also build custom setups using the "Bare metal" frame, which is a mount that can be integrated into their own products.

Neon is operated by a mobile phone running the Neon Companion app. With a single battery charge, up to four hours of data can be recorded. A web-app streams scene video and gaze estimates directly to any computer, phone, or tablet in the same network. For developers, there is also an API for interacting with the phone (e.g. for starting/stopping recordings) as well as for real-time streaming of eye images, gaze estimates, scene video, and IMU measurements.

Gaze direction is measured in real-time, based on pairs of concurrent left and right eye images, employing a deep-learning-based pipeline, referred to as NeonNet. Resulting gaze estimates are reported in 2D scene camera pixel coordinates. Gaze estimation works "out of the box" and does not require any calibration. However, the Neon Companion app offers the option to personalize gaze estimates using a person-specific additive offset. This offset can be used to account for physiological factors, such as the angle Kappa between the optical and visual axis of the eye [2].

Neon recordings can be uploaded to Pupil Cloud, a web-based eye-tracking platform for storing, managing, viewing, and sharing recordings. Pupil Cloud offers various analysis tools: fixation and blink detection, automatic mapping of gaze to reference images, and more.

3 Neon Accuracy – Screen-based Lab Task

First, we evaluate the Neon eye-tracking system in an indoor lab environment. To do this, we designed a screen task that uses standardized gaze targets. The task was performed by head-stabilized subjects under controlled lighting conditions. Overall, our validation design aimed to replicate a typical eye-tracking experiment conducted in a lab setting.

3.1 Method

The experiment was designed to measure gaze-estimation accuracy for each individual subject over a relevant portion of the field of view. For this, subjects had to fixate gaze targets at various positions. Then, for every gaze target, an error could be calculated by comparing the position of the gaze target to the gaze estimate as output by Neon. Specifically, we developed the following experimental setup.

Over the course of two separate trials, each subject was seated in front of one of two screens (see Figure 2A). We used an 85" Smart TV at a distance of 1.3 m and a 27" computer screen at a distance of 0.4 m. This allowed us to estimate gaze-estimation accuracy at two different depths. Note that validation results at 1.3 m are also indicative of gaze-estimation accuracy at all larger depths. This is because the vergence angle of the eyes and thus eye appearance changes only little beyond this point. In both cases, the screen covered approximately 60° horizontally and 35° vertically of the subject's field of view. To ensure a constant orientation of each subject's head towards the screen, a chinrest was used for all subjects.

To evenly cover the field of view, each trial presented a series of gaze targets. Specifically, 16 consecutive gaze targets were randomly selected from non-repeating bins within a screen-filling rectangular 4x4 grid (see Figure 2B). While the order of the bins was randomized, large jumps between resulting gaze targets were avoided. Once all bins were visited once, a new series of 16 gaze targets was generated.

For the gaze targets, we used a black disk overlaid with a white cross and a small black disk in the center of the cross (see Figure 2C). This design has shown to minimize micro-saccades and promote stable fixations [15]. The entire gaze target covered approximately 1° and the inner circle covered 0.25° of the field of view.

The inner circle of the gaze target periodically flashed red at random intervals uniformly distributed between 1.6 to 2.6 seconds. Subjects were instructed to press a button when the current gaze target flashed red. Only when they pressed the button at the correct time, did the target move to the next position. This gamification method kept subjects engaged in the task and encouraged accurate fixation on each target.

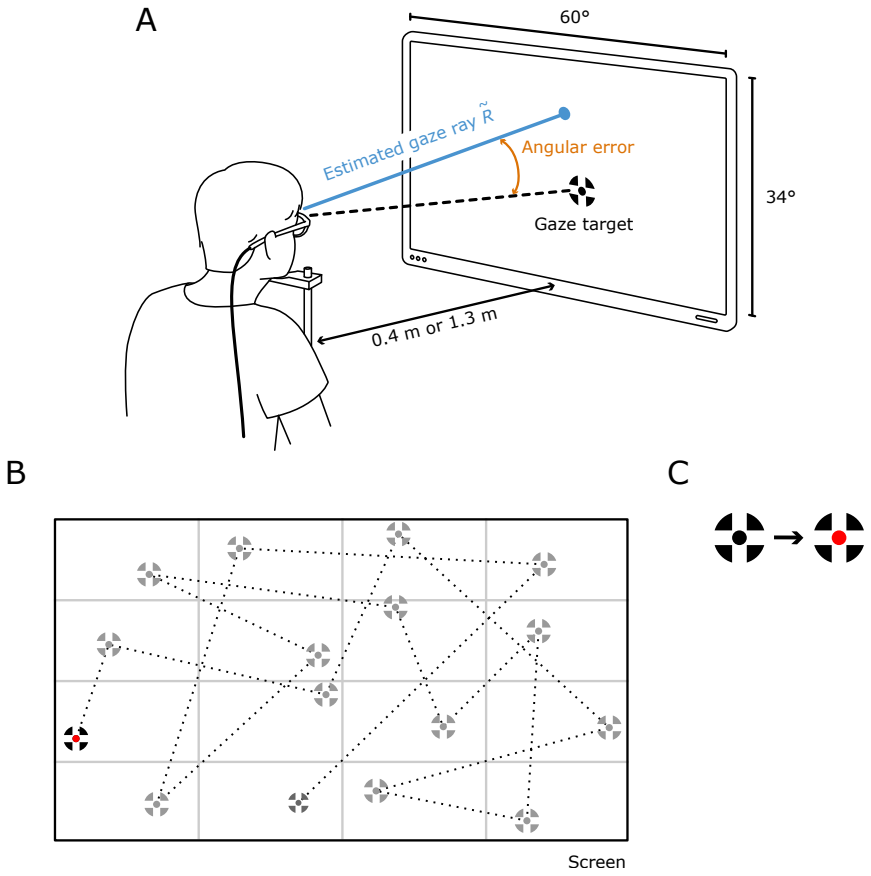

Figure 2: Experimental design of the screen-based accuracy evaluation. **A)** Over the course of two trials, subjects were seated in front of two screens at a distance of 0.4 m and 1.3 m, respectively. **B)** Gaze targets were generated in random order, sampled from a screen-filling grid of 4x4 bins. A single realization of 16 consecutive gaze targets is shown. Lines connect consecutive gaze targets. **C)** Subjects were tasked to press a button every time the center of the gaze target flashed red. Note that for illustration purposes, gaze targets in all panels are drawn enlarged. In reality, the diameter was only 1/60th of the screen width.

For each trial, subjects performed the described task for a total of 2.5 minutes, yielding on average 54 different gaze targets at each of the two depths, respectively.

For high-quality accuracy estimates, it is essential to disregard gaze targets that were not actually looked at. To this end, we preprocessed the data as follows. For each gaze target that elicited a correct button press, we considered all gaze estimates collected in the one-second window preceding the button press as relevant. By applying a dispersion-based criterion, we then algorithmically verified that during this time the participant held their head stable and fixated steadily on the gaze target. More specifically, we discarded the gaze target if the standard deviation of the estimated gaze signal was above 1°. Fluctuations due to noise intrinsic to the gaze-estimation pipeline were found to have an amplitude of about 0.3° during fixations (estimated via the root-mean-square amplitude of sample-to-sample deviations, RMS-S2S [8]); therefore a standard deviation above 1° implied that the subject was not fixating the center of the gaze target correctly, which only spanned 0.25° of their field of view. Similarly, we disregarded a gaze target due to head motion if the standard deviation of the gaze target location in the scene camera was above 0.25°. This threshold is approximately thrice the observed average standard deviation. In addition, visual inspection showed that discarded fixations indeed exhibited pronounced head motion.

In total, we removed only 4.8% of gaze targets due to incorrect fixations and 5.6% due to head motion.

To minimize the impact of blinks, we also disregarded the target if blinks were detected within the window. For this, we used the Pupil Labs blink detector [1].

This procedure resulted in a set of valid gaze targets for each subject and trial, with 200 gaze estimates per gaze target.

Given a gaze target at position (x, y) in 2D pixel coordinates, we wish to calculate the error associated with an individual gaze estimate $(\tilde{x}, \tilde{y})$ in degrees of visual angle. To this end, note that each point $P = (x, y)$ in 2D pixel space corresponds to a 3D ray, R, in the spatial coordinate system of the camera, such that R is projected onto $P = (x, y)$ (see Figure 2A). Therefore, the angular gaze error can be determined by the angle enclosed by the two rays, R and $\tilde{R}$, corresponding to the gaze target (x, y) and the gaze estimate $(\tilde{x}, \tilde{y})$, respectively. From this, the accuracy with respect to a given gaze target and subject was calculated as the average angular error over all associated gaze estimates. Finally, the overall gaze-estimation accuracy of Pupil Neon for a specific subject, also referred to as *per-subject accuracy*, was calculated as the average of all gaze-target accuracies in a given trial.

Neon provides gaze estimates without the need for a subject-specific calibration. However, in the Neon Companion app, the user has the option to apply an offset correction to all gaze estimates. This offset helps further reduce systematic errors that are consistent in both magnitude and direction across the field of view. To test the potential impact of this personalization mechanism, we determined an offset for each subject and trial based on the first 16 valid gaze targets. Specifically, we calculated the average error vector $\Delta = \frac{1}{n} \sum_i \Delta_i$ for these samples, where $\Delta_i = (\tilde{x}_i, \tilde{y}_i) - (x_i, y_i)$. The remaining gaze estimates were adjusted by Δ, and per-subject accuracy was calculated based on these adjusted gaze estimates.

We recorded data from a total of N=206 subjects. To assess the potential impact of various general appearance characteristics on gaze-estimation accuracy, a diverse subject pool was recruited. This pool included both male and female subjects, individuals with and without contact lenses, and represented a wide age range from 18 to 85 years. The subjects also had typical interpupillary distance (IPD) values ranging from 55 to 74 mm, and covered a variety of racial appearances and choices of facial makeup. Importantly, none of the data from these subjects was used to train Neon-Net. Therefore, the accuracies reported in the following section can be considered as true validation results, representative of the wider population.

3.2 Results

We will start by presenting population statistics for the entire pool of N=206 subjects. In Figure 3, we display the results for the two depths we tested at. In the respective upper panels (see Figures 3A and B), the data are represented as point clouds, indicating the distribution of per-subject accuracies without calibration (blue points) and with an additive offset (orange points). The lower panels (see Figures 3C and D) show the corresponding cumulative distribution functions, which illustrate the percentage of subjects with an accuracy lower than a given value. Tick values for the shared horizontal axis are displayed in lower panels.

Without calibration, the median accuracy is similar at both distances, measuring 2.1° at 0.4 m and 1.8° at 1.3 m (see Figure 3A and B). For about 80% of subjects at both depths, an accuracy better than 3.0° was recorded (see Figures 3C and D).

By applying an additive calibration offset (as discussed in the previous section), the median accuracy is further improved to 1.4° at 0.4 m and 1.3° at 1.3 m, respectively.

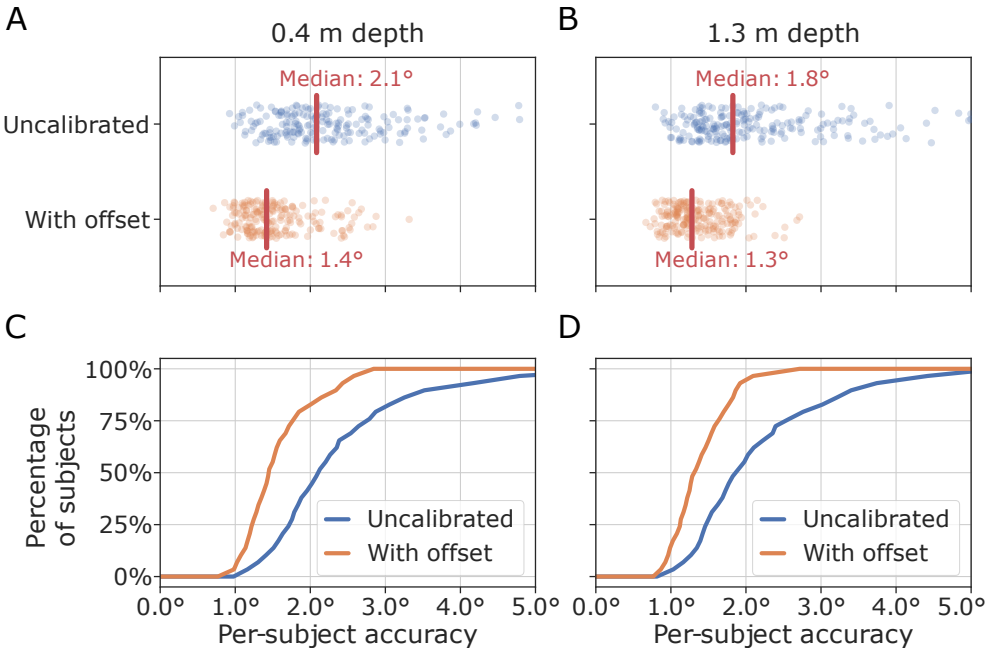

Figure 3: Population statistics of Neon gaze-estimation accuracy. In the left column (panels A and C), results at 0.4 m depth are shown. In the right column (panels B and D), at 1.3 m. **Upper panels (A and B):** distributions of per-subject accuracies as point clouds (blue: uncalibrated, orange: with offset). **Lower panels (C and D):** corresponding cumulative distribution functions (same colors as in upper panels). Without calibration, median accuracy is similar between depths (2.1° at 0.4 m; 1.8° at 1.3 m). With offset calibration, accuracy in both settings is further improved (1.4° at 0.4 m and 1.3° at 1.3 m). For 94% of subjects, accuracy with offset at 1.3 m depth is better than 2°.

Notably, there are almost no subjects with an accuracy worse than 2°. Specifically, at a distance of 1.3 m, 94% of subjects have an accuracy better than 2° (see Figure 3D). The results at a depth of 1.3 m also provide an indication of the expected accuracy at larger distances, as vergence and eye appearance change only little beyond this point.

Next, we will show that the gaze-estimation accuracy of Neon is uncorrelated with extraneous factors influencing eye appearance (e.g. makeup, age, IPD, racial appearance, etc.). To this end, we split the pool of subjects along a number of dimensions and for each split calculated the median accuracy for each resulting subgroup with offset correction at 1.3 m. In Figure 4, we show results for gender appearance, makeup, wearing or not wearing contact lenses, age, IPD, and racial appearance. Subgroup sizes are as indicated, with subgroup sizes within splits adding up to N=206. We observe that differences in median accuracy due to the considered attributes are within tenths of degrees from the median accuracy of 1.3° reported above (see Figure 3B) and are thus indeed small.

In summary, the results from our controlled in-lab screen task demonstrate that Neon, without calibration, offers gaze estimation with a median accuracy that is better than 2° at all depths. Through a simple calibration procedure, accuracy can be further optimized to a median of 1.3°, with almost all subjects exhibiting accuracies better than 2°. Furthermore, accuracy is shown to be robust with respect to various external variables influencing eye appearance.

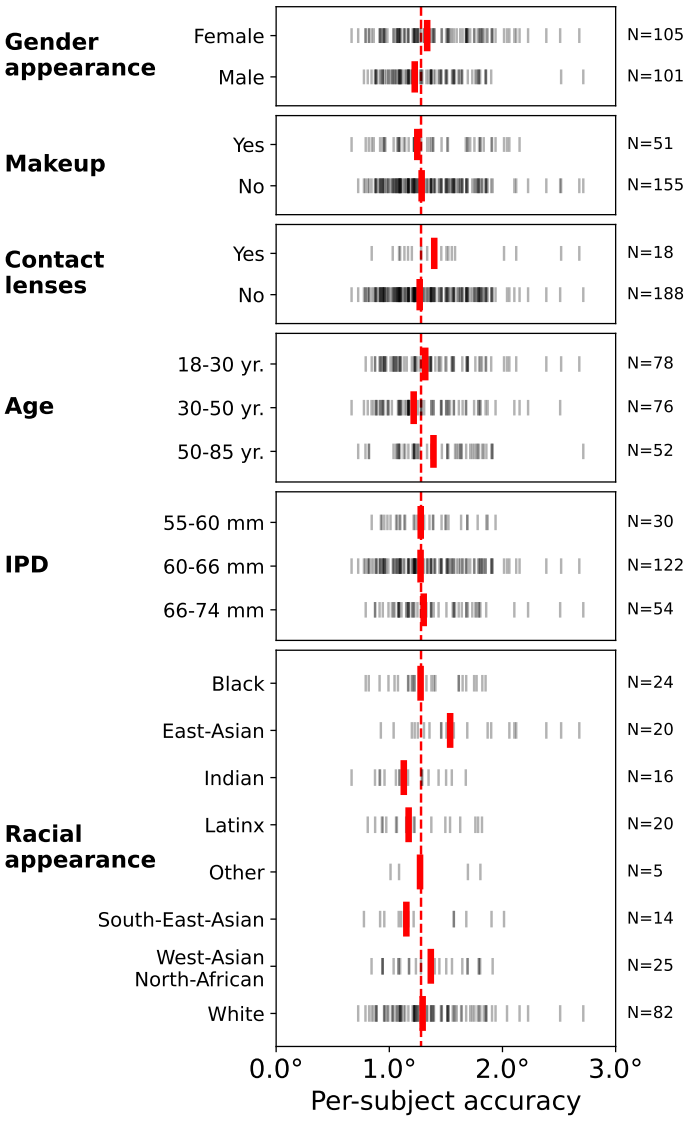

Figure 4: Comparison of median per-subject accuracy between various demographics. The distributions of per-subject accuracies (at 1.3 m, with offset correction; black dashes) are shown for various splits of the subject pool. Subgroup sizes are given on the right. For each subgroup, the red dash indicates the respective median accuracy. For comparison, vertical red dashed lines mark the median per-subject accuracy over the full subject pool (cf. Figure 3B).

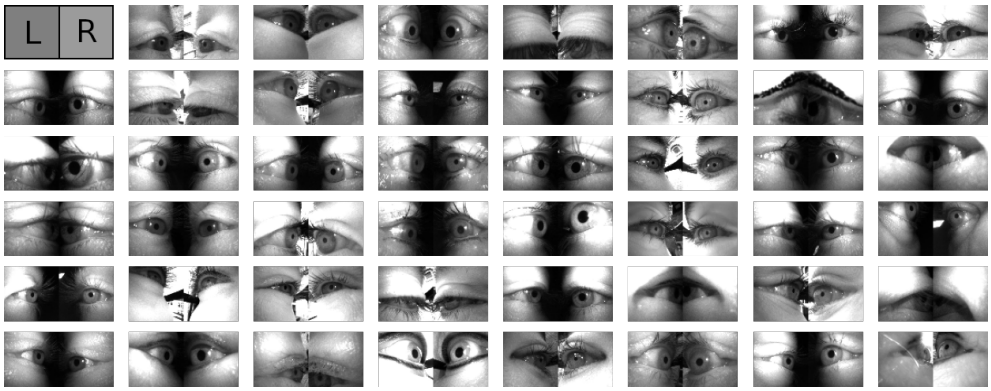

Figure 5: Example eye images. Pairs of concurrent left (L) and right (R) eye images from our in-the-wild dataset. All images correspond to the same gaze direction. Note the high variability in appearance, lighting, and slippage state.

4 Neon Accuracy – In-The-Wild

Head-mounted eye-tracking systems extend gaze estimation beyond controlled laboratory settings. In this section, we therefore present validation results that assess the accuracy of the Neon eye-tracking system in more diverse and challenging situations. To this end, we have collected a comprehensive in-house dataset in realistic in-the-wild scenarios. Specifically, we have focused on maximizing the variability of the dataset in terms of subject appearance, lighting conditions, and headset slippage.

4.1 Method

Altogether, N=402 randomly selected subjects were recruited from 53 different countries all over the world. Participants were 18 to 85 years old at the time of the study. No restrictions were imposed on eye appearance, makeup, or headwear. In particular, the dataset includes recordings with subjects wearing pronounced eye makeup, artificial eyelashes, facial masks, as well as headscarves (see Figure 5). As for the screen task, none of the data from these subjects was used to train NeonNet.

In order to obtain a diverse set of fixations on known gaze targets, we developed a recording choreography reflective of the real-world usage of Neon. More specifically, each subject recorded several short clips fixating predefined gaze targets in their field of view. By changing their distance (0.3 m - 4.0 m) and overall orientation with respect to the gaze targets, a wide range of samples throughout the accessible 3D viewing volume was obtained. Clips were recorded both indoors and outdoors, the latter often including direct sunlight from diverse directions. From each clip, corresponding pairs of gaze-target positions in the subject's 3D field of view and Neon gaze estimates were extracted.

To account for the variability of slippage configurations, subjects were instructed to randomly move the Neon eye-tracking system on their face between clips. This resulted in a rich selection of different slippage states.

The Neon module can be combined with a variety of frames. To reflect this variability, three different frames were included in the collection of our in-the-wild dataset. More specifically, the "Just act natural", "Better safe than sorry", and "Ready set go!" frames were used in a randomized manner. The frames chosen cover a wide range of use cases, from everyday activities to sports and workplace settings.

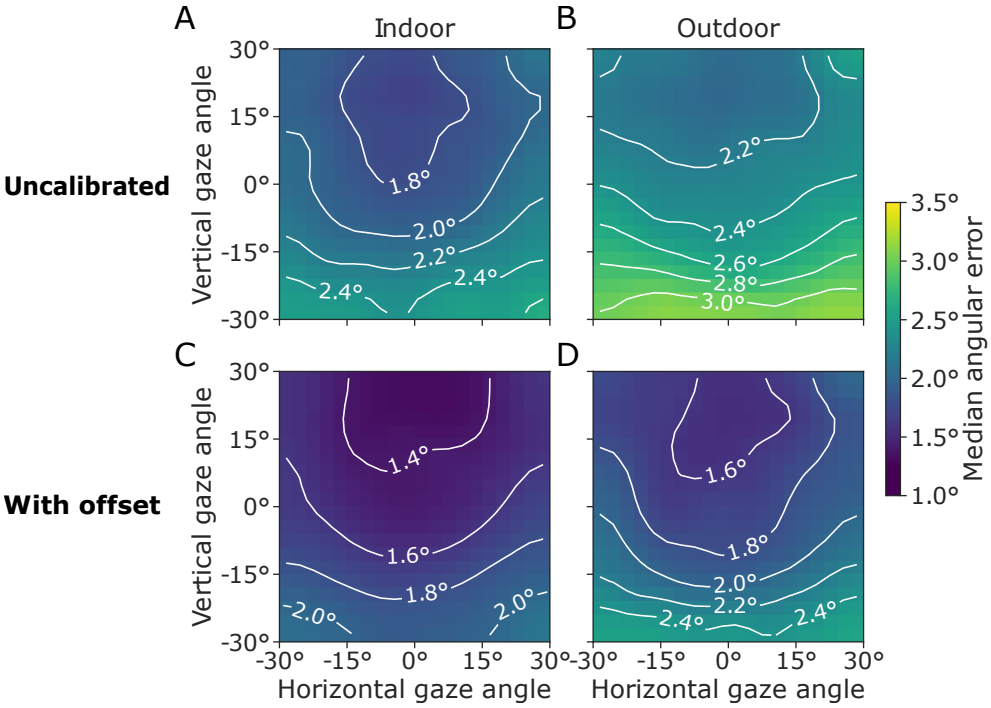

Figure 6: Distribution of the median gaze estimation accuracy over the field of view. Upper panels (A and B): "Out of the box" results without calibration for indoor and outdoor settings. **Lower panels (C and D):** Results with offset correction. Plots show results calculated over all depths from 30 cm to 4 m using our in-the-wild dataset.

In summary, our validation dataset approximates the expected variability of in-the-wild recordings made with Neon. Figure 5 displays a random selection of example eye images from our dataset, demonstrating its rich diversity regarding slippage state, lighting conditions, and eye appearance.

4.2 Results

In order to aggregate the collected data, first the angular error associated with each pair of corresponding gaze-target position and gaze estimate was calculated as described earlier (see Section 3.1). Conditioned on gaze direction, we then calculated the median accuracy over all samples and subjects. For better interpretability, in this section and the next, we express gaze direction in terms of its horizontal and vertical gaze angle.

The latter angles are defined such that a gaze direction with 0° both vertically and horizontally, corresponds to gazing straight ahead. By construction, the scene camera is tilted approximately 5° down relative to the average vertical gaze angle (across different activities). Thus, the gaze direction $(0°, 0°)$ corresponds to a point slightly above the center of the scene camera video. To account for this, the gaze angles are defined in the 3D coordinate system defined by the scene camera, rotated by 5° around the x-axis. In this coordinate system, a given 2D gaze position (x, y) corresponds to a ray R. The horizontal and vertical components of a gaze angle are then the angles between R and the y-z-plane or x-z-plane, respectively.

We show the median gaze-estimation accuracy as a heat map for a regular grid of gaze directions in Figure 6. The grid was chosen to extend 60° horizontally and vertically, effectively covering most of a typical subject's field of view. White contour

lines indicate levels of identical accuracy. To gauge the effect of indoor vs. outdoor lighting, we show results separately for indoor (see Figure 6, left panels) and outdoor clips (see Figure 6, right panels). Both settings are evaluated uncalibrated (upper panels) and with a person-specific offset correction (lower panels). The latter results were derived by first fitting a single offset for every subject using one quarter of the subject's gaze targets and then evaluating on the other three quarters. Note, the same offsets were used indoors and outdoors. Using the same method as in in Section 3.1, the offsets were fitted by calculating the average error vector across all calibration samples. The clips used for calibration were disjoint from the ones used for validation, which ensured slippage configurations were different between settings.

First, consider the uncalibrated setting. In both lighting conditions, accuracy varies by only about 1° over the whole field of view. Accuracy is best close to the center of the upper half of the field of view, achieving optimal values of 1.6° indoors and 1.9° outdoors. In particular, indoor accuracy close to the center of the field of view is consistent with our screen-task results presented earlier (c.f. Figure 3B). Accuracy towards the lower end of the field of view is slightly reduced, most likely as a result of systematic changes in eye appearance due to occlusions by eyelids and eyelashes. The effect of lighting conditions is small, amounting to a spatially uniform deterioration of only 0.4° in outdoor vs. indoor environments.

Second, consider the setting with offset correction. In both lighting conditions, performance improves significantly and consistently across the whole field of view. Indoor accuracy improves by 0.4° and outdoor performance by 0.6°. As a result, the performance difference between the two settings gets even smaller as compared to the uncalibrated case and is now only 0.2°. Again, the results in the center are consistent with the accuracies obtained during the screen-based lab task (c.f. Figure 3B). While the beneficial impact of offset correction was even higher during the screen task, we hypothesize that this is due to the more complex data in the in-the-wild dataset. Not only does the latter contain many different slippage configurations for every subject, but slippage was minimized during the screen task. Also, the in-the-wild dataset contains a high variability in outdoor lighting conditions, whereas the screen task was recorded only indoors.

Overall, the results presented in this section show that Neon also provides reliable gaze estimates outside of controlled lab environments, even without a user-specific calibration. In particular, our data attests to the robustness of Neon to both in-the-wild slippage and lighting conditions.

5 Implications for Real-World Tasks

In the last section, we have discussed the gaze-estimation accuracy of the Neon eye-tracking system as a function of gaze direction. The field of view covered by our in-the-wild dataset was deliberately chosen to be large enough to cover a major portion of the physiologically accessible range. Typically, however, gaze is not uniformly distributed and tends to exhibit a pronounced center bias, i.e. humans tend to look straight ahead more often than to the side [3, 13, 19]. Also, the distribution of gaze over the field of view changes for different activities [14, 5]. In this section, we briefly discuss the accuracy results from the previous section in light of these task-specific gaze distributions.

To this end, we had four subjects wear the Neon eye-tracking system while each performing a different real-world activity. The activities chosen were cooking, going for a walk, playing table tennis, and biking. The resulting recordings had durations

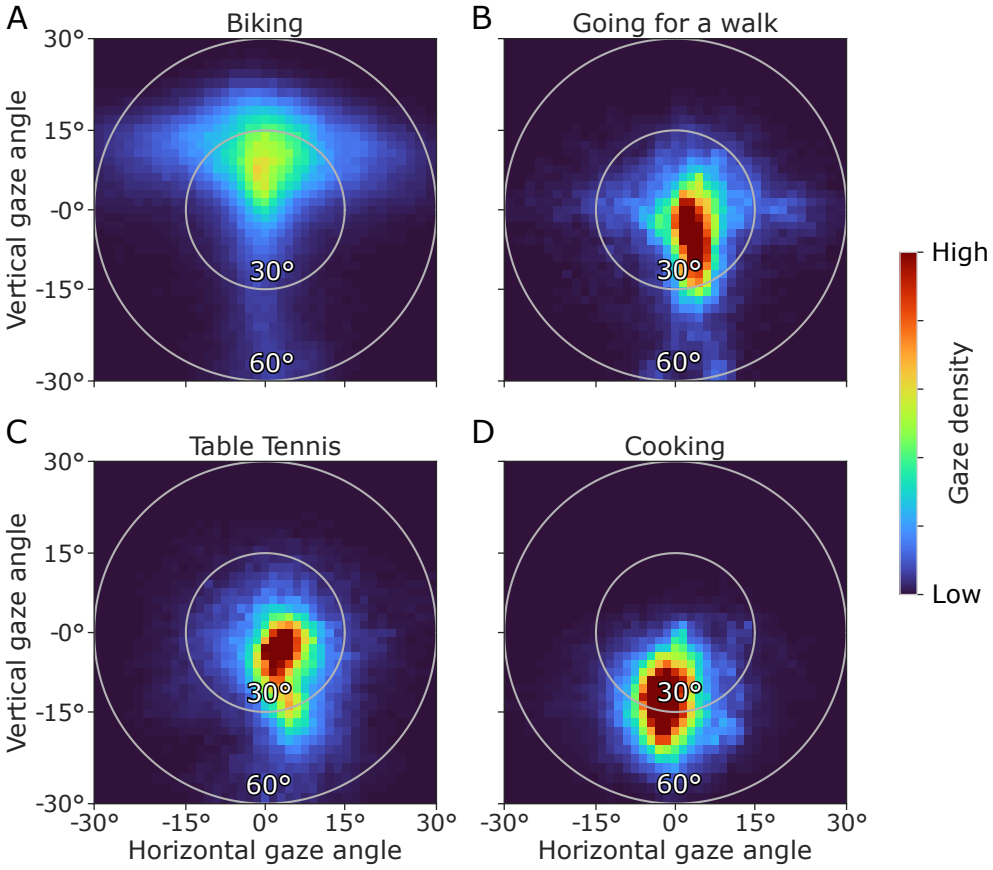

Figure 7: Gaze-direction densities for four different activities. Warm colors indicate directions in which people tended to look more often. White circles delineate a 30° and 60° field of view, respectively, relative to the central axis of the Neon scene camera.

from 39 minutes (table tennis) to 3.5 hours (biking). Based on the gaze estimates reported by Neon, for each recording we calculated the associated distribution of gaze directions, shown as heat maps in Figure 7.

The resulting distributions clearly reflect the demands associated with the respective task. For example, when biking (see Figure 7A), subjects tend to lean forward and thus need to look slightly upwards to see the street ahead ($\rightarrow$ high density region shifted upwards from the center). At the same time, they need to scan their surroundings predominantly along the horizontal plane ($\rightarrow$ distribution spread out along the horizontal axis). On the other hand, while cooking (see Figure 7D), subjects frequently gaze downwards, e.g. when cutting vegetables or frying food in a pan ($\rightarrow$ high-density region focused downwards from the center).

Since gaze-estimation accuracy varies with gaze direction (cf. Figure 6), the effective gaze-estimation accuracy relative to a given task depends on the associated distribution of gaze over the field of view. Combining the results from Figure 6 and Figure 7, we derive an effective accuracy, which we deem a good proxy for measuring Neon gaze-estimation accuracy relative to a specific real-world task. More specifically, we average the spatial accuracy distribution from Figure 6 weighted by the task-specific gaze distribution from Figure 7. The resulting effective accuracies are given in Table 1. For all real-world tasks considered here, we find an uncalibrated accuracy between 2.1° and 2.4°, which improves to 1.7°– 1.8° with an offset correction. These results are only slightly higher than for the screen task in controlled lab conditions presented in Section 3.

Activity	Effective accuracy		Lighting condition
	Uncalibrated	With offset	
Table tennis	2.2°	1.7°	Indoor + Outdoor
Biking	2.3°	1.7°	Outdoor
Going for a walk	2.4°	1.8°	Outdoor
Cooking	2.1°	1.7°	Indoor

Table 1: Effective gaze-estimation accuracy for different activities. Values are calculated as averages of the accuracy distributions from Figure 6, weighted by the task-specific gaze distributions from Figure 7. The table also indicates the lighting condition assumed for the respective task (indoor, outdoor, or both).

6 Summary

The Neon eye-tracking system by Pupil Labs is engineered to provide accurate gaze estimates, without necessitating a calibration prior to its use. In this report, we have corroborated this claim by means of a multi-pronged validation study. Bringing to bear data recorded from hundreds of subjects, we have gauged Neon gaze-estimation accuracy both in and outside of the lab. Accuracy was found to be similar in both contexts, attesting in particular to Neon's robustness to lighting conditions and movement-induced device slippage. More specifically, a median per-subject accuracy of 1.8° at 1.3 m depth and across a wide field of view was measured in the framework of an in-lab screen task. By means of a simple subject-specific offset correction, median per-subject accuracy could be further improved to 1.3°.

References

[1] Pupil Labs blink detector whitepaper. `https://assets.pupil-labs.com/pdf/Pupil_Labs_Blink_Detector.pdf`.

[2] David A. Atchison. *Optics of the Human Eye*. Multidisciplinary and Applied Optics. CRC Press, 2023.

[3] Guy T. Buswell. How people look at pictures: A study of the psychology and perception in art. *Univ. Chicago Press*, 1935.

[4] Katarzyna Harezlak and Pawel Kasprowski. Application of eye tracking in medicine: A survey, research issues and challenges. *Computerized Medical Imaging and Graphics*, 65:176–190, 2018. Advances in Biomedical Image Processing.

[5] Christopher Kanan, Dina N. F. Bseiso, Nicholas A. Ray, Janet H. Hsiao, and Garrison W. Cottrell. Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision Research*, 108:67–76, 2015.

[6] Yuki Kishita, Hiroshi Ueda, and Makio Kashino. Eye and head movements of elite baseball players in real batting. *Frontiers in Sports and Active Living*, 2, 2020.

[7] Yaxiong Lei, Shijing He, Mohamed Khamis, and Juan Ye. An end-to-end review of gaze estimation and its interactive applications on handheld mobile devices. *ACM Computing Surveys*, 56(2):1–38, 09 2023.

[8] Diederick C. Niehorster, Raimondas Zemblys, Tanya Beelders, and Kenneth Holmqvist. Characterizing gaze position signals and synthesizing noise during fixations in eye-tracking data. *Behavior Research Methods*, 52:2515–2534, 2020.

[9] Rima-Maria Rahal and Susann Fiedler. Understanding cognitive and affective mechanisms in social psychology through eye-tracking. *Journal of Experimental Social Psychology*, 85:103842, 2019.

[10] Shane L. Rogers, Craig P. Speelman, Oliver Guidetti, and Melissa Longmuir. Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific reports*, 8(1):4271, 2018.

[11] Matthias Rothensee and Philipp Reiter. Neuromarketing. *Eye Movement Research: An Introduction to its Scientific Foundations and Applications*, pages 819–855, 10 2019.

[12] Rene Santos, Jorge Oliveira, Jessica Rocha, and Janaina Giraldi. Eye tracking in neuromarketing: A research agenda for marketing studies. *International Journal of Psychological Studies*, 7, 02 2015.

[13] Benjamin W. Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):4–4, 11 2007.

[14] Benjamin W. Tatler, Nicholas J. Wade, Hoi Kwan, John M. Findlay, and Boris M. Velichkovsky. Yarbus, eye movements, and vision. *i-Perception*, 1(1):7–27, 2010.

[15] Lore Thaler, Alexander C. Schütz, Melvyn A. Goodale, and Karl R. Gegenfurtner. What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research*, 76:31–42, 2013.

[16] Neil M. Thomas, Timmion K. Skervin, Richard J. Foster, Johnny V. Parr, Mark G. Carpenter, Thomas D. O'Brien, Constantinos N. Maganaris, Vasilios Baltzopoulos, Carolyn Lees, and Mark A. Hollands. Influence of step-surface visual properties on confidence, anxiety, dynamic stability, and gaze behaviour in young and older adults. *Human movement science*, 77:102774, 2021.

[17] Isabelle de Thysebaert, Thijs Ackermans, Natasha Francksen, Neil M. Thomas, Mark Hollands, Thomas O'Brien, Christopher McCrum, Kenneth Meijer, and Constantinos Maganaris. Gaze behaviour of older adults during stair negotiation: The effect of step dimensions. *OSF Preprints*, 2023.

[18] Darja Topolšek, Igor Areh, and Tina Cvahte. Examination of driver detection of roadside traffic signs and advertisements using eye tracking. *Transportation Research Part F: Traffic Psychology and Behaviour*, 43:212–224, 2016.

[19] Po-He Tseng, Ran Carmi, Ian G. M. Cameron, Douglas P. Munoz, and Laurent Itti. Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision*, 9(7):4–4, 07 2009.

[20] Niilo V. Valtakari, Ignace T. C. Hooge, Charlotte Viktorsson, Pär Nyström, Terje Falck-Ytter, and Roy S. Hessels. Eye tracking in human interaction: Possibilities and limitations. *Behavior Research Methods*, pages 1–17, 2021.

[21] Pieter Vansteenkiste, David Van Hamme, Peter Veelaert, Renaat Philippaerts, Greet Cardon, and Matthieu Lenoir. Cycling around a curve: The effect of cycling speed on steering and gaze behavior. *PLOS ONE*, 9(7):1–11, 07 2014.

[22] William R. Young and Mark A. Hollands. Can telling older adults where to look reduce falls? Evidence for a causal link between inappropriate visual sampling and suboptimal stepping performance. *Experimental brain research*, 204:103–113, 2010.

[23] Gal Ziv, Ronnie Lidor, Sima Zach, Stephanie Brams, and Werner F. Helsen. Gaze behavior of referees in sport – a review. *Frontiers in Sports and Active Living*, 2, 2020.