IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 December 2025, accepted 10 January 2026, date of publication 19 January 2026, date of current version 26 January 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3655327

RESEARCH ARTICLE

Optimizing Within-Domain Gaze Estimation: Insights From a Novel Appearance-Based 2D Model

SARDOR ABDIRAYIMOV ODIL UGLI[1,2], SE-YOUNG BAK[1,3,4], YUBIN KIM[1,3], EUN-HYE CHUNG[1,3], HEEGOO KIM[1,3], EUNYOUNG CHO[3,5], MIRI SUH[3,5], SEYOUNG SHIN[1,3,5], HYEONGMIN JEON[1,3,5], AND MINYOUNG KIM[1,3,5]

[1]Digital Therapeutics Research Team, Department of Research, CHA Bundang Medical Center, Seongnam 13520, Republic of Korea
[2]Department of AI and Big Data, Woosong University, Daejeon-si 34606, Republic of Korea
[3]Department of Rehabilitation Medicine, CHA Bundang Medical Center, CHA University School of Medicine, Seongnam 13496, Republic of Korea
[4]Department of Integrative Medicine, Yonsei University College of Medicine, Seoul 06229, Republic of Korea
[5]Graduate School, CHA University, Pocheon-si, Gyeonggi-do 11160, Republic of Korea

Corresponding authors: HyeongMin Jeon (jhm6302@cha.ac.kr) and MinYoung Kim (kmin@cha.ac.kr)

This research was supported by the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (HR22C 1605) and the Korea Planning and Evaluation Institute of Industrial Technology (KEIT), funded by the Ministry of Trade, Industry and Energy (RS-2025-08762968).

ABSTRACT Appearance-based gaze estimation has emerged as a promising alternative to traditional model-based methods, effectively addressing their limitations in terms of flexibility, cost, and adaptability to unconstrained environments. In this study, the Digital Therapeutics Research Team at Bundang CHA Medical Center developed a novel appearance-based gaze estimation algorithm, CHA-Gaze, by integrating head-pose information into an adaptive feature fusion network (AFF-Net) architecture, which is a widely recognized baseline in the field. To evaluate the effectiveness of CHA-Gaze, we conducted a unified validation using the MPIIFaceGaze dataset, which comprises 37,590 images from 15 participants acquired under semi-natural conditions. The results demonstrated that CHA-Gaze achieved a significantly lower mean Euclidean error of 1.88 cm, compared to 2.59 cm by AFF-Net ($p < 0.001$). These findings indicate that CHA-Gaze offers superior accuracy and improved robustness across various appearances and environmental conditions. This study confirms the effectiveness of architectural refinement within appearance-based gaze estimation frameworks and highlights the potential of CHA-Gaze for real-world deployment in applications, such as digital therapeutics, telehealth, and accessibility technologies. The proposed model provides a scalable, non-intrusive solution using standard webcams, making it suitable for widespread use in both clinical and consumer-grade settings.

INDEX TERMS AFF-Net, appearance-based gaze estimation, deep learning, eye-tracking, gaze estimation, machine learning, multi-task regression module.

I. INTRODUCTION

Gaze estimation is a computational technique designed to predict an individual's point of visual attention, and is typically represented in screen coordinates or directional vectors. This technology has been extensively applied across various domains, including medical diagnosis [1], [2], [3], psychology for the analysis of human behavior and cognition [4], [5], and human-computer interaction [6], [7], [8]. As interdisciplinary interest in gaze estimation has increased, two primary methodological paradigms (model-based and appearance-based) have emerged as the most widely adopted in both research and applied contexts [9].

Model-based gaze estimation, which is utilized in commercial systems such as those developed by Tobii (e.g., Tobii X2-60, Tobii EyeX), integrates the geometric modeling

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

of ocular structures with dedicated hardware components, including infrared cameras and real-time calibration systems. These systems offer high precision and reliability, making them well-suited for laboratory and experimental environments. However, their widespread adoption is limited by high costs, the need for complex calibration procedures, and their dependency on controlled settings, which restrict their accessibility for general users and everyday applications.

To address these limitations, appearance-based gaze estimation has gained traction as a viable and scalable approach. By leveraging deep learning algorithms, these methods directly infer gaze direction from red-green-blue (RGB) images, often captured by standard webcam devices, thus providing a more affordable and user-friendly solution [10], [11], [12], [13], [14], [15]. Notably, the adaptive feature fusion network (AFF-Net), a deep learning model with a feature fusion technique between convolution layers, has demonstrated strong performance in two-dimensional (2D) gaze estimation tasks [10]. However, the existing appearance-based approaches still face challenges in terms of achieving robustness and generalizability, particularly under variable lighting, head poses, and background conditions. Despite ongoing efforts, relatively few studies have systematically addressed these limitations or proposed scalable solutions for deployment in real-world settings [9], [16], [17], [18], [19], [20].

Building on these insights, the Digital Therapeutics team at Bundang CHA Hospital developed a novel appearance-based gaze estimation model, CHA-Gaze, which extends the AFF-Net architecture through additional logic modules and structural refinements. The CHA-Gaze model is specifically designed to enhance estimation accuracy in webcam-based environments, making it suitable for practical applications such as digital therapeutics, telehealth, and assistive technologies. In this study, we introduced and validated the CHA-Gaze model, providing one of the first comparative evaluations of 2D gaze estimation models under a consistent benchmarking protocol. To assess its efficacy, we directly compared the performance of CHA-Gaze with that of AFF-Net [10], LiAGE [21], which is a recent model, and modified versions of CHA-Gaze that varied in network size (i.e., increased or reduced numbers of neurons), using standardized evaluation metrics. This study contributes to the growing body of research aimed at establishing more generalizable gaze estimation technologies for real-world and clinical contexts.

II. RELATED WORKS

Gaze estimation has been a prominent research topic for an extended period and numerous approaches have been proposed to address this challenge. Recent advancements in computer vision technology have facilitated the development of modern eye-tracking devices that typically estimate gaze using eye or face images captured by cameras, including infrared (IR) and RGB cameras. Computer vision-based methods can be categorized into two types: geometry-based and appearance-based. Geometry-based methods typically construct subject-specific geometric eye models to estimate the human gaze direction. These models are fitted with geometric features such as infrared corneal reflections [22], [23], pupil centers [24], and iris contours [25]. In contrast, appearance-based methods directly learn a mapping function from images to the human gaze. A key distinction of appearance-based methods is the absence of dedicated devices for detecting geometric features, instead of using RGB images. Various machine learning algorithms have been employed, including neural networks [26], Gaussian process regression [27], adaptive linear regression [28], convolutional neural networks [11], [15], [29], recurrent neural networks [30], and transformers [31]. Cheng et al. [32] proposed a unidirectional convolutional architecture that improved both the efficiency and accuracy of gaze estimation models. Holla et al. [21] introduced a lightweight adaptive gaze estimator that explicitly models the intrinsic relationships among pupil location, head pose, and geometric priors.

Recent developments in the field have increasingly shifted from lightweight or geometry-based approaches to generative and foundation-model-based methods to address the challenges of in-the-wild generalization. Ruzzi et al. [33] adapted Neural Radiance Fields (NeRF) for gaze redirection and gaze estimation by modeling the eye region as a continuous volumetric scene conditioned on gaze direction, enabling both novel-view gaze synthesis and accurate estimation. Qin et al. [34] leveraged millions of in-the-wild facial images through masked autoencoding and gaze-related proxy tasks before fine-tuning. Ryan et al. [35] combined frozen large-scale visual encoders (e.g., DINOv2) with a lightweight transformer decoder and person-specific positional prompts to estimate gaze targets in complex-scene images. Vuillecard and Odobez [36] proposed a self-training, weakly supervised framework that integrates precise but limited 3D gaze datasets with abundant 2D gaze-following labels, using pseudo-labeling and consistency regularization. Xiao et al. [37] introduced a deformable and decoupled representation learning framework that uses deformable attention to model individual variations in eye structure, and a dual-branch decoder to explicitly disentangle invariant features (e.g., eyeball geometry) from variable features (e.g., gaze direction and pupil configuration).

AFF-Net is a representative deep learning model designed for appearance-based gaze estimation in 2D space. In a comprehensive review and benchmark study [9], AFF-Net [10] was evaluated against other models using datasets such as the Max Planck Institute for Informatics (MPII) Gaze [11], GazeCapture [15], and EYEDIAP [38]. This study demonstrated that AFF-Net, along with the End-to-end Frame-to-gaze Estimation (EFE) model, achieved superior performance in 2D gaze estimation tasks. Specifically, AFF-Net outperformed MPIIGaze, underscoring its effectiveness in gaze estimation applications. In a study by Bao et al. [10], AFF-Net was compared with models such as iTracker [15], faSt Accurate Gaze trackEr (SAGE) [39], and Tolerant And Talented schema (TAT) [40] on the GazeCapture dataset. The results indicated that AFF-Net outperformed the other methods, achieving lower average errors, particularly for the images captured from the tablets. In another comparative study, AFF-Net was evaluated against models such as RT-Gene, FullFace, and Gaze360 on the MPIIFaceGaze and EYEDIAP datasets. AFF-Net achieved the lowest angular error in the MPIIFaceGaze dataset, indicating its robustness under various conditions. Given the proven performance of AFF-Net in these studies, researchers selected this model as a prototype for CHA-Gaze.

To enhance the accuracy and robustness of gaze estimation, we incorporated head-pose coordinates as auxiliary inputs. This additional framework is particularly advantageous for appearance-based gaze estimation, where subtle variations in eye-region appearance can be more effectively resolved when complementary geometric cues are jointly modeled. By leveraging the inductive bias introduced by related tasks, the model achieves improved generalization and greater resilience to noise and domain shifts, which are the key challenges in real-world gaze estimation. Furthermore, the use of shared feature representations supports more stable and consistent learning, which is particularly beneficial when training data are limited or highly person-specific. In this study, we adopted head-pose features as a more robust alternative to the original geometric variables defined in the pixel coordinate system, resulting in an improved prediction performance under varying head poses and illumination conditions, such as those commonly observed in datasets such as MPIIFaceGaze.

III. METHODS

A. DATASET

MPIIFaceGaze is a subset of the original MPIIGaze dataset [41], a well-established benchmark dataset comprising 213,659 images containing over 37,000 images collected over several months using built-in calibrated laptop cameras. This dataset included images from 15 participants along with additional human-labeled annotations, including pupil center annotations and facial regions. These annotations were created by human annotators and refined using the human facial landmark detection method. For real-world gaze estimation, several authors have suggested that state-of-the-art face models can replace original manual annotations [14]. In this study, we employed the MediaPipe face detection model, which resulted in the exclusion of images that were affected by substantial lighting variations [42]. Unlike synthetic or strictly controlled laboratory datasets, MPIIFaceGaze captures users performing gaze tasks during routine computer use, thereby preserving the natural head and eye movement patterns. Zhang et al. [11] utilized a semi-automatic annotation procedure, in which facial landmarks were first detected algorithmically and then manually corrected when necessary. Accordingly, MPIIFaceGaze serves as a reliable and valuable benchmark for personalized and appearance-based gaze estimation methods, particularly in applications involving human–computer interactions. Each subject contributed 1,500 to 2,930 images, resulting in approximately 37,000 images for training and testing. Although this process may have reduced the naturalism of the images, the dataset remained suitable for detection and processing in the pipeline. Importantly, it captured a wide range of variations in appearance, including head orientation, facial features, and illumination. For this reason, MPIIFaceGaze is widely used as a benchmark dataset for gaze estimation models in real-world human–computer interaction scenarios.

B. MODELS

CHA-Gaze is a competitive deep-learning architecture (Figure 1) designed with four distinct inputs: facial images, images of the left and right eyes, and a unit head pose vector. This architecture incorporates methods from the AFF-Net paper, including 1) Squeeze and Excitation for the fusion of eye features and 2) Adaptive Group Normalization.

1) SQUEEZE-AND-EXCITATION (SE)

Eye feature maps from multiple layers are aggregated channel-wise. Because the left and right eyes share the same shape and structure, their feature maps are integrated by stacking and then passed through convolutional layers to generate the final eye feature vector. To maintain consistency, the right-eye images were flipped horizontally during preprocessing. This ensured alignment of the inner and outer eye corners as well as eyebrow orientation across both eyes. EyeNet [10] was used for eye-feature extraction. The SE layer [43] serves as a powerful attention mechanism, selectively emphasizing important eye features across channels. In this study, SE layers were used to adaptively balance the spatial details and complex representations from both eyes. The mathematical formulation of the SE layer is given by Equation (1).

$$\begin{cases} W\text{weight} = \sigma(\text{FC}(\text{GAP}(\text{Fin}))) \\ F\text{out} = F\text{scale}(W\text{weight, Fin}) \end{cases} \quad (1)$$

The final output F_{out} is a recalibrated tensor with the same shape as that of the input tensor F_{in}. $F_{scale}(\bullet)$ represents the channel-wise multiplication function between input features F_{in} and W_{weight}. The latter variable is derived sequentially from the chain of the Global Average Pooling (GAP), Fully Connected (FC) layers, and sigmoid (σ) layers. The GAP layer is used to reduce the spatial dimensions of feature maps by averaging the values in each feature map. The FC layer is a type of neural network in which every neuron in the layer is connected to every neuron in the previous and subsequent l layers. The sigmoid (σ) is an activation function that squashes the feature map value between 0 and 1.

2) ADAPTIVE GROUP NORMALIZATION (AdaGN)

Building on previous research [44], [45], we implemented recalibration in Group Normalization by incorporating the facial appearance characteristics. AdaGN utilizes concatenated head-pose vector features and facial features as inputs to represent facial appearance characteristics, thereby regulating eye features. This process is illustrated in equation (2). AdaGN adaptively modifies eye feature extraction based on facial appearance characteristics by recalibrating eye features.

$$\begin{cases} [W\text{shift}, W\text{scale}] = \text{LeakyReLU}(\text{FC}(F\text{head}, F\text{face})) \\ F\text{out} = W\text{scale} \cdot \text{GN}(\text{Fin}) + W\text{shift} \end{cases} \quad (2)$$

where F_{in}- stands for the original feature map, F_{out} stands for the final output of the AdaGN function, F_{head} stands for a correct feature extracted by FC layers, F_{face} stands for a face feature, W_{scale} stands for a scale parameter, W_{shift} stands for a shift parameter, FC stands for a neural network type, where each neuron has full connection with previous layers' and next layers' neurons, LeakyReLU is an activation function that returns squash negative values toward 0, and GN (Group Normalization layer) stands for the normalization technique, which normalizes features by dividing input channels into smaller groups and computing the mean and variance for each group independently.

We evaluated and compared the performance of two appearance-based gaze estimation models: AFF-Net and CHA-Gaze. AFF-Net served as the baseline model because of its proven performance across multiple benchmark studies on MPIIFaceGaze and other datasets. It employs an adaptive fusion feature extractor to improve the gaze prediction accuracy from RGB images.

C. PREPROCESSING

1) DATA PREPROCESSING

We conducted experiments on the widely used MPIIFaceGaze dataset, which contains approximately 37,000 samples from 15 participants recorded under natural laptop use conditions. To accelerate training, the entire dataset was preprocessed once and stored in a single compressed HDF5 (.h5) file. Preprocessing was performed using MediaPipe Face Mesh, which provides high-accuracy facial landmarks, a 4*4 transformation matrix, and reliable head pose estimation. Images with severe underexposure or landmark detection failure (68 samples, approximately 0.12% of the dataset) were removed.

All face images were resized to 224*224 pixels, and the left- and right-eye patches were cropped around the detected landmarks and resized to 112*112 pixels. Eye blink frames were manually removed from the original MPIIFaceGaze release. From the MediaPipe transformation matrix, we extracted the 3*3 rotation matrix r and converted it into yaw (Equation 3), pitch (Equation 4), and roll angles (Equation 5), which were subsequently transformed into a normalized 3D head-pose vector (Equation 6-9).

$$\varphi = A\tan 2\,(r_{21}, r_{11}) \quad (3)$$

$$\theta = A\tan 2\left(-r_{31}, \sqrt{r_{32}^2 + r_{33}^2}\right) \quad (4)$$

$$\psi = A\tan 2\,(r_{32} + r_{33}) \quad (5)$$

$$\boldsymbol{v} = [x, y, z] \quad (6)$$

$$\boldsymbol{x} = cos\theta \cdot sin\psi \quad (7)$$

$$\boldsymbol{y} = sin\theta \quad (8)$$

$$\boldsymbol{z} = cos\theta \cdot cos\psi \quad (9)$$

2) TRAINING PROCEDURE

We performed a special dataset split to select only 20% of the images from each participant at random and compose the testing dataset. The remaining 80% were used for training. Thus, the total number of testing images was over 6200, and the total number of training images was over 30800. The final training was implemented using the PyTorch framework on a single NVIDIA GeForce RTX 3060 GPU and Intel Core i7 10700 with 64 GB of RAM. The training configuration was as follows: Adam optimizer with a learning rate of 0.003, batch size of 64, and Mean Squared Error (MSE) loss between the predicted and ground-truth 2D gaze coordinates. The number of epochs was set as 30. The MSE Formula is shown in (10).

$$\text{MSE}\,(g, \dot{g}) = \frac{1}{N}\sum\nolimits_{n=1}^{N} ||g - \dot{g}||_2^2 \quad (10)$$

where g is the ground truth in the normalized space and g_hat is the prediction from the CHA-Gaze.

Different loss functions were tested, including the mean Euclidean Loss, mean weighted loss, and mean absolute loss; nevertheless, the MSE was the most influential in the sense that it provides a better magnitude to reach convergence faster.

3) EVALUATION METRICS

Performance was mainly evaluated using Euclidean errors (in centimeters). Smaller values indicate a higher performance. The CHA-gaze outputs normalized (x, y) 2D coordinates between 0 and 1. The model output is expected to be normalized based on the screen size in centimeters. The statistical significance between models was assessed using the Mann-Whitney test [46] to verify whether the improvements from CHA-Gaze were meaningful. In addition, we calculated the angular error to increase the interpretability of the results. Although the CHA-Gaze is a gaze model designed for 2D predictions, it is possible to estimate the angular error. The angular Error (Equation 11) requires two vectors. We calculated the gaze vector as the difference between the two points in the 3D space. We obtained vector $\boldsymbol{a}$ based on the ground-truth gaze position in 3D ($\text{gaze}_{\text{target}}$) and the 3D midpoint between the two eye centers estimated from the person's face ($\text{gaze}_{\text{origin}}$) according to Equation (12). To obtain gaze vector $\boldsymbol{b}$ for CHA-gaze prediction, the 2D screen gaze positions were projected onto the 3D camera space using the camera-to-screen rotation and translation parameters provided by the dataset (Equations 13 and 14).

$$\theta = arccos\frac{a \cdot b}{|a|\,|b|} \quad (11)$$

$$\boldsymbol{a} = \text{gaze}_{\text{target}} - \text{gaze}_{\text{origin}} \quad (12)$$

$$\text{gaze}_{\text{predic}} = R * \text{PoG}_{\text{unnorm}} + \boldsymbol{t}\textbf{vect} \tag{13}$$

$$\boldsymbol{b} = \text{gaze}_{\text{predic}} - \text{gaze}_{\text{origin}} \tag{14}$$

4) STATISTICAL ANALYSIS

Average Euclidean error values were computed from a test set of 15 participants. The normality of the values obtained from the three models, AFF-Net, LiAGE, and CHA-Gaze, was assessed using the Shapiro-Wilk test. As the p-values indicated a deviation from normality ($p < 0.05$), the Mann-Whitney U test [46] was subsequently employed for statistical comparison.

IV. RESULTS

We conducted a quantitative comparison of the proposed CHA-Gaze model with the baseline AFF-Net and LiAGE models using the MPIIFaceGaze dataset. The evaluation metric was the Euclidean error, measured in centimeters (cm), between the predicted and ground-truth gaze points on the screen, with lower values indicating higher accuracy. Tables 1 and 2 summarize the Euclidean and angular errors for all 15 participants, respectively. Overall, CHA-Gaze consistently achieved lower errors than AFF-Net, demonstrating superior performance in predicting gaze positions under natural conditions (Figures 2 and 3). The mean Euclidean error for AFF-Net was 2.59 cm, whereas CHA-Gaze reduced this value to 1.88 cm, corresponding to a 27% improvement in the accuracy.

To assess statistical significance, we used the Mann–Whitney test [46]. The resulting p-value of <0.001 indicates that the performance improvement of CHA-Gaze over AFF-Net is statistically significant and unlikely to be attributable to chance. Comparisons based on the angular error showed trends similar to those observed in the Euclidean error. CHA-Gaze with a reduced network size (fewer neurons) achieved accuracy comparable to the standard CHA-Gaze model, with no statistically significant differences. In contrast, the variant with an increased network size (more neurons) exhibited higher Euclidean and angular errors than LiAGE, CHA-Gaze, and the reduced-neuron variant.

Only the standard CHA-Gaze model and the reduced-neuron version demonstrated accuracy comparable to that of LiAGE (a recent model), with no significant differences. Additionally, both the mean Euclidean and angular errors were lower for CHA-Gaze than for the reduced-neuron model. Table 3 lists the attributes of CHA-Gaze and AFF-Net. In the view of attributes, the two models showed similar values despite of changes in architecture and significant differences in accuracy.

These findings suggest that the architectural enhancements incorporated into CHA-Gaze, particularly its optimized feature-integration modules and multicondition training strategy, substantially improve the reliability of webcam-based gaze estimation. Moreover, the results highlight its potential for practical deployment in applications such as digital therapeutics, telehealth, and accessibility-focused human–computer interactions.

V. DISCUSSION

This study introduced and validated CHA-Gaze, an appearance-based gaze estimation model developed as

TABLE 1. Comparison of Euclidean error among gaze estimation models.

Models		LiAGE	AFF-Net	CHA-Gaze	CHA-Gaze (less neurons)	CHA-Gaze (more neurons)
	Euclidean error (cm)	1.77	2.59	1.88	1.94	2.33
LiAGE	1.77	-	< 0.001***	0.367	0.250	0.002**
AFF-Net	2.59	-	-	< 0.001***	< 0.001***	0.126
CHA-Gaze	1.88	-	-	-	0.567	0.009**
CHA-Gaze (less neurons)	1.94	-	-	-	-	0.023*
CHA-Gaze (more neurons)	2.33	-	-	-	-	-

*$p < 0.05$, **$p < 0.01$, ***, $p < 0.001$; p-value was calculated from the Mann-Whitney test. SD, Standard Deviation. cm, centimeters.

TABLE 2. Comparison of angular error among gaze estimation models.

Models		LiAGE	AFF-Net	CHA-Gaze	CHA-Gaze (less neurons)	CHA-Gaze (more neurons)
	Angular error (degree)	2.05	3.05	2.19	2.26	2.73
LiAGE	2.07	-	< 0.001***	0.305	0.098	< 0.001***
AFF-Net	3.05	-	-	< 0.001***	< 0.001***	0.061
CHA-Gaze	2.19	-	-	-	0.567	0.003**
CHA-Gaze (less neurons)	2.26	-	-	-	-	0.006**
CHA-Gaze (more neurons)	2.73	-	-	-	-	-

**$p < 0.01$, ***$p < 0.001$; p-value was calculated from the Mann-Whitney test. SD, Standard Deviation

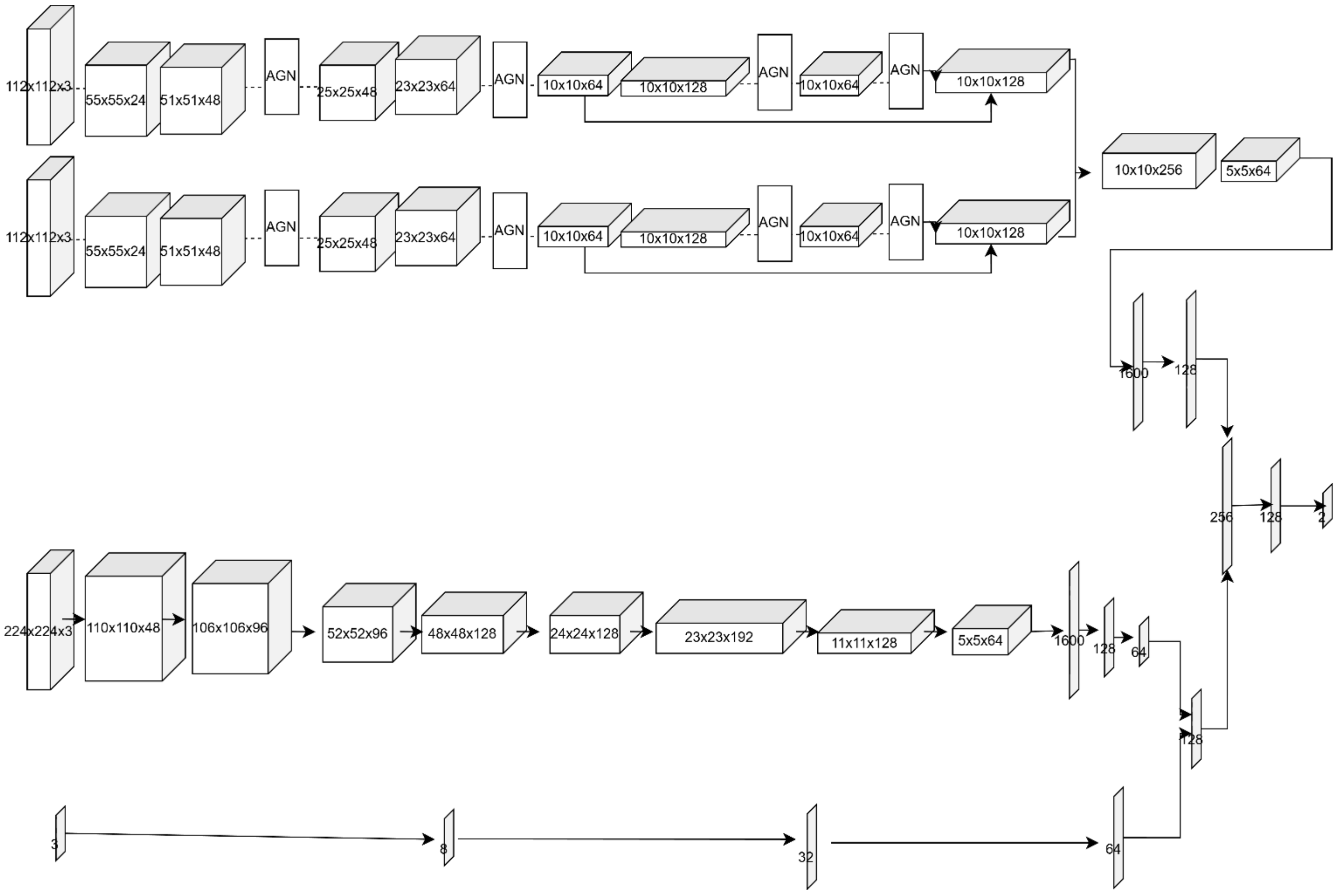


FIGURE 1. Architecture of the proposed gaze estimation network. A shared eye feature extractor processes both left and right eye crops using a lightweight CNN backbone composed of convolutional layers, max-pooling, Squeeze-and-Excitation (SE) blocks, and adaptive normalization layers. In parallel, a face feature extractor encodes the full-face image through convolutional layers with pooling, normalization, and SE blocks. The 3D head pose parameters are projected into a 64-dimensional latent space via fully connected layer sequences. Finally, the gaze estimation head concatenates the eye features, face features, and encoded head pose vector, then progressively reduces the combined representation through a series of linear layers with non-linear activations, yielding a 2D gaze direction normalized to [0,1] that is subsequently denormalized to the original screen resolution (adaptive normalization details are provided in the previous section).

TABLE 3. Comparison of model attributes between CHA-Gaze and Aff-Net.

Model name	GFLOPS	Latency (GPU) in seconds	Latency (CPU) in seconds	Training duration for one epoch over 37,000 images seconds	Size of model (assuming float 32) in MB	Number of Parameters in millions	Number of epochs/ Number of backprop steps
CHA-Gaze	4.84	0.008	0.023	110	7.33	1.91	25/12050
AFF-Net	4.84	0.007	0.025	109	7.43	1.94	25/12050

GFLOPS, GPU floating point operations per second. GPU, graphics processing unit. CPU, central processing unit. MB, megabyte.

an extension of the AFF-Net architecture [10]. Through structural enhancements, such as refined feature-fusion mechanisms and the integration of head-pose vectors, CHA-Gaze was designed to improve gaze estimation accuracy in webcam-based environments. These environments reflect real-world conditions often encountered in digital therapeutics, telehealth services, and assistive human–computer interaction systems.

To evaluate its effectiveness, we conducted a comparative analysis using the MPIIFaceGaze dataset, which is a widely adopted benchmark for appearance-based gaze estimation. CHA-Gaze demonstrated a substantial improvement over AFF-Net, achieving a mean Euclidean error of 1.88 cm, corresponding to a 27% reduction relative to AFF-Net's 2.59 cm. Statistical analysis using the Mann–Whitney test [46] confirmed that this improvement was significant ($p < 0.001$), indicating that the performance gain can be attributed to meaningful architectural refinements rather than random variation. Furthermore, the mean Euclidean and angular error results suggest that the architecture requires an optimized number of neurons. The Euclidean error obtained for AFF-Net in our study was lower than the 3.9 cm

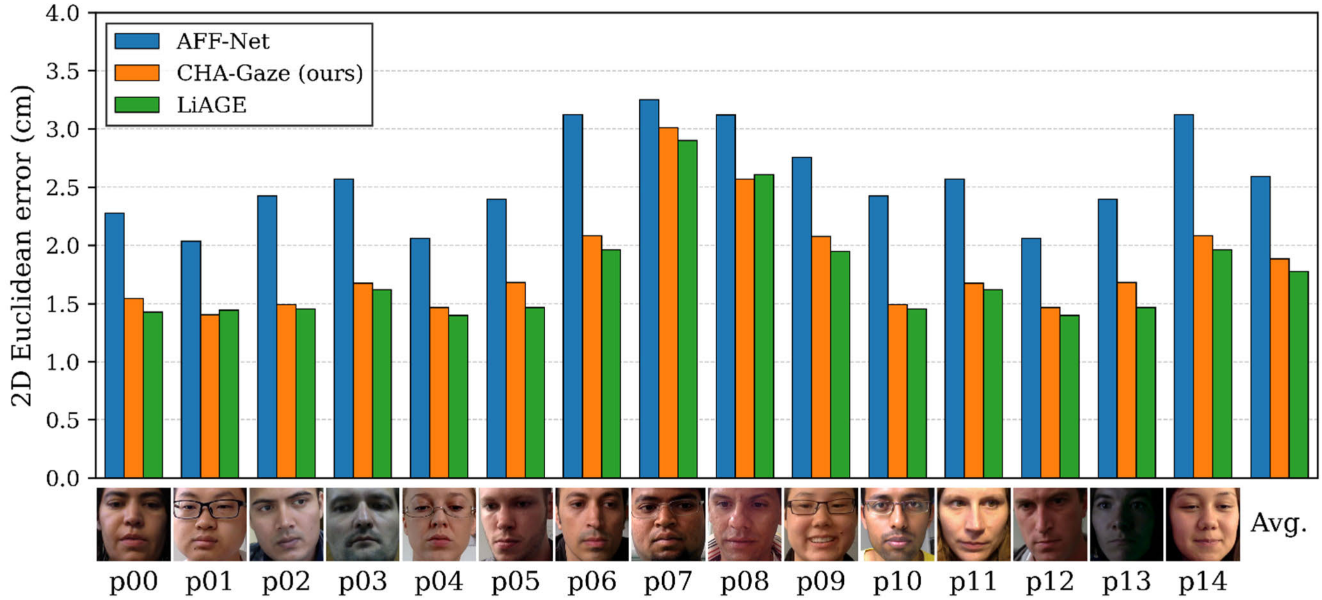


FIGURE 2. Bar graphs of Euclidean errors from AFF-Net,CHA-Gaze and LiAGE for whole participants in MPIIFaceGaze dataset. 2D, two dimensions. cm, centimeters.p, participant. Avg, Average.

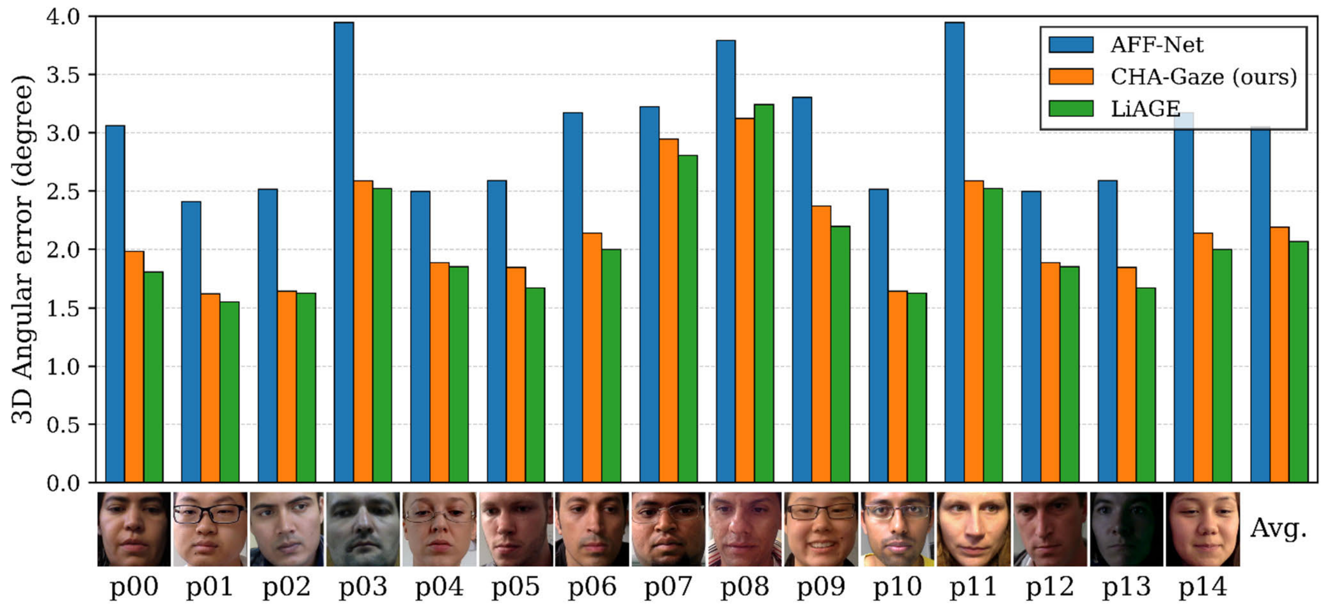


FIGURE 3. Bar graphs of angular errors from AFF-Net,CHA-Gaze and LiAGE for whole participants in MPIIFaceGaze dataset. 3D, three dimensions. p, participant.Avg, Average.

previously reported [10], likely due to our use of 37,000 images from the MPIIFaceGaze dataset combined with Mediapipe-based preprocessing, which provides robust face detection and alignment [42]. These findings highlight the importance of preprocessing pipelines when assessing gaze estimation performance in practical environments.

The decision to focus on within-domain evaluation was motivated by the study's primary objective: to isolate and quantify the performance gains attributable to architectural modifications under controlled conditions. Existing gaze benchmarks differ substantially in head-pose distributions, camera intrinsic, image resolution, and gaze-range coverage, which introduces pronounced dataset bias and can obscure model-specific improvements. By maintaining a consistent evaluation domain, this study provides a clearer attribution of performance differences to model design choices. Nevertheless, we acknowledge that cross-dataset generalization remains a critical requirement for real-world deployment and should be addressed in subsequent work.

From an application perspective, CHA-Gaze demonstrates competitive performance for 2D gaze estimation using only RGB webcam inputs, underscoring its potential utility for scalable, non-intrusive gaze-based interfaces. This is particularly relevant in digital therapeutics, telehealth services, and assistive human–computer interaction systems, where conventional hardware-based eye trackers (e.g., Tobii) may be impractical due to cost, calibration burden, or limited operating range. Importantly, the model's architecture remains lightweight relative to transformer-heavy alternatives, suggesting favorable computational scalability for real-time deployment on consumer-grade hardware. Hence, the authors believe CHA-Gaze aligns well with the increasing demand for accessible digital health technologies that require minimal calibration and no specialized hardware, including remote rehabilitation, cognitive monitoring, and accessibility services for individuals with mobility or communication impairments [47].

Despite these promising results, several limitations of this study remain to be acknowledged. The model was evaluated exclusively in healthy adults with predominantly frontal head poses without ocular abnormalities. Its generalizability to underrepresented populations, including children, older adults, and individuals with neurological or ophthalmological conditions (e.g., strabismus, ptosis, nystagmus, or partial pupil occlusion), remains unknown, as such characteristics are either absent or severely underrepresented in current public datasets.

Cross-dataset evaluation was not performed in this study because of substantial domain shifts across existing benchmarks, including differences in head-pose distributions, recording devices (e.g., laptops, tablets, and high-resolution cameras), and gaze-range coverage. These discrepancies introduce a dataset bias that limits the validity of the direct comparisons. Nonetheless, a rigorous assessment of generalization, robustness, and fairness across demographics and hardware platforms remains essential. Therefore, future work will include comprehensive cross-dataset experiments, such as training on large-scale datasets and evaluation of MPIIFaceGaze, ColumbiaGaze, and TabletGaze under standardized protocols.

Additional research directions include incorporating lightweight attention mechanisms or hybrid CNN–transformer architectures, exploring multitask learning strategies, and integrating personalized few-shot calibration to improve performance under extreme head poses or occlusion. Ultimately, this study contributes to ongoing efforts to democratize gaze estimation technologies by demonstrating that careful architectural refinement of existing models can yield meaningful performance gains without reliance on specialized hardware. CHA-Gaze shows strong potential as a practical, accurate, and accessible solution for gaze-based interactions for both clinical and consumer applications.

Future work should focus on (1) collecting ethically approved, diverse clinical datasets, (2) developing few-shot or zero-shot adaptation techniques for 3D gaze estimation under challenging ocular conditions, (3) generating high-fidelity synthetic data for children and older adults, and (4) conducting extensive real-world validations across patient populations. We are currently pursuing IRB-approved data collection with the goal of releasing a more inclusive and comprehensive gaze estimation benchmark.

VI. CONCLUSION

In this study, we present CHA-Gaze, an innovative appearance-based gaze estimation model developed to enhance the AFF-Net. To assess its performance, we conducted a comparative analysis using the MPIIFaceGaze dataset, with a focus on Euclidean error (cm) as the primary metric. The results indicate that CHA-Gaze consistently surpassed AFF-Net, exhibiting a significantly lower average Euclidean error (1.88 cm vs. 2.59 cm, $p < 0.001$). The evaluation confirmed the efficacy of the architectural modifications introduced in CHA-Gaze under unconstrained webcam-based conditions. These findings provide a validated foundation within the domain for further development and deployment of CHA-Gaze in gaze-driven applications, where accessibility, scalability, and performance are critical. Future research should include broader evaluations across various datasets, user populations, and task conditions to assess the generalizability and practical utility of the model.

ACKNOWLEDGMENT

The authors express their sincere appreciation to them of AFF-Net for releasing a clean and well-documented code base, which substantially facilitated the implementation of their method for comparative analysis. They also thank them of LiAGE for providing a clear and comprehensive description of their model, which greatly supported their implementation. This research was supported by the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (HR22C 1605) and the Korea Planning and Evaluation Institute of Industrial Technology (KEIT), funded by the Ministry of Trade, Industry and Energy (RS-2025-08762968).

(Sardor Abdirayimov Odil Ugli and Se-Young Bak contributed equally to this work.)

REFERENCES

[1] S. A. Chung, J. Choi, S. Jeong, and J. Ko, "Block-building performance test using a virtual reality head-mounted display in children with intermittent exotropia," *Eye*, vol. 35, no. 6, pp. 1758–1765, Jun. 2021, doi: 10.1038/s41433-020-01160-y.

[2] M. S. Corin, T. S. Elizan, and M. B. Bender, "Oculomotor function in patients with Parkinson's disease," *J. Neurological Sci.*, vol. 15, no. 3, pp. 251–265, Mar. 1972, doi: 10.1016/0022-510x(72)90068-8.

[3] M. Elsabbagh, E. Mercure, K. Hudry, S. Chandler, G. Pasco, T. Charman, A. Pickles, S. Baron-Cohen, P. Bolton, and M. H. Johnson, ''Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism,'' *Current Biol.*, vol. 22, no. 4, pp. 338–342, Feb. 2012, doi: 10.1016/j.cub.2011.12.056.

[4] S. L. Brown and M. Richardson, ''The effect of distressing imagery on attention to and persuasiveness of an antialcohol message: A gaze-tracking approach,'' *Health Educ. Behav.*, vol. 39, no. 1, pp. 8–17, Feb. 2012, doi: 10.1177/1090198111404411.

[5] K. R. Newman and C. R. Sears, ''Eye gaze tracking reveals different effects of a sad mood induction on the attention of previously depressed and never depressed women,'' *Cognit. Therapy Res.*, vol. 39, no. 3, pp. 292–306, Jun. 2015, doi: 10.1007/s10608-014-9669-x.

[6] A. Laddi and N. R. Prakash, ''Eye gaze tracking based directional control interface for interactive applications,'' *Multimedia Tools Appl.*, vol. 78, no. 22, pp. 31215–31230, 2019.

[7] C. H. Morimoto and M. R. M. Mimica, ''Eye gaze tracking techniques for interactive applications,'' *Comput. Vis. Image Understand.*, vol. 98, no. 1, pp. 4–24, Apr. 2005, doi: 10.1016/j.cviu.2004.07.010.

[8] J. Song, S. Cho, S.-Y. Baek, K. Lee, and H. Bang, ''GaFinC: Gaze and finger control interface for 3D model manipulation in CAD application,'' *Computer-Aided Design*, vol. 46, pp. 239–245, Jan. 2014, doi: 10.1016/j.cad.2013.08.039.

[9] Y. Cheng, H. Wang, Y. Bao, and F. Lu, ''Appearance-based gaze estimation with deep learning: A review and benchmark,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 7509–7528, Dec. 2024, doi: 10.1109/TPAMI.2024.3393571.

[10] Y. Bao, Y. Cheng, Y. Liu, and F. Lu, ''Adaptive feature fusion network for gaze tracking in mobile tablets,'' in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 9936–9943, doi: 10.1109/ICPR48806.2021.9412205.

[11] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, ''Appearance-based gaze estimation in the wild,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4511–4520.

[12] H. Balim, S. Park, X. Wang, X. Zhang, and O. Hilliges, ''EFE: End-to-end frame-to-gaze estimation,'' 2023, *arXiv:2305.05526*.

[13] Y. Sugano, Y. Matsushita, and Y. Sato, ''Learning-by-synthesis for appearance-based 3D gaze estimation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1821–1828.

[14] S. Park, S. D. Mello, P. Molchanov, U. Iqbal, O. Hilliges, and J. Kautz, ''Few-shot adaptive gaze estimation,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9367–9376.

[15] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, ''Eye tracking for everyone,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2176–2184.

[16] S. Huynh, R. K. Balan, and J. Ko, ''IMon: Appearance-based gaze tracking system on mobile devices,'' *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 4, pp. 1–26, Dec. 2021, doi: 10.1145/3494999.

[17] X. Zhang, S. Park, T. Beeler, D. Bradley, S. Tang, and O. Hilliges, ''ETH-XGaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation,'' in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 365–381, doi: 10.48550/arxiv.2007.15837.

[18] M. Heck, C. Becker, and V. Deutscher, ''Webcam eye tracking for desktop and mobile devices: A systematic review,'' in *Proc. 56th Hawaii Int. Conf. Syst. Sci.* Germany: University of Mannheim, 2023, pp. 6820–6829.

[19] H.-R. Yang, X. Chen, and C.-X. Ren, ''A generalized label shift perspective for cross-domain gaze estimation,'' 2025, *arXiv:2505.13043*.

[20] L. Falch and K. S. Lohan, ''Webcam-based gaze estimation for computer screen interaction,'' *Frontiers Robot. AI*, vol. 11, Apr. 2024, Art. no. 1369566, doi: 10.3389/frobt.2024.1369566.

[21] N. S. Holla, A. Kushwaha, C. Sanchi, and S. Balaji, ''LiAGE : Light-weight adaptive gaze estimation,'' in *Proc. 15th Indian Conf. Comput. Vis. Graph. Image Process.*, India, Dec. 2024, pp. 1–8, doi: 10.1145/3702250.3702261.

[22] Z. Zhu and Q. Ji, ''Novel eye gaze tracking techniques under natural head movement,'' *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2246–2260, Dec. 2007, doi: 10.1109/TBME.2007.895750.

[23] E. D. Guestrin and M. Eizenman, ''General theory of remote gaze estimation using the pupil center and corneal reflections,'' *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1124–1133, Jun. 2006, doi: 10.1109/TBME.2005.863952.

[24] R. Valenti, N. Sebe, and T. Gevers, ''Combining head pose and eye location information for gaze estimation,'' *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 802–815, Feb. 2012, doi: 10.1109/TIP.2011.2162740.

[25] K. A. Funes Mora and J.-M. Odobez, ''Geometric generative gaze estimation (G3E) for remote RGB-D cameras,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1773–1780.

[26] S. Baluja and D. Pomerleau, ''Non-intrusive gaze tracking using artificial neural networks,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 6, 1993, pp. 753–760.

[27] O. Williams, A. Blake, and R. Cipolla, ''Sparse and semi-supervised visual mapping with the S^3GP,'' in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jul. 2006, pp. 230–237, doi: 10.1109/CVPR.2006.285.

[28] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, ''Adaptive linear regression for appearance-based gaze estimation,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2033–2046, Oct. 2014, doi: 10.1109/TPAMI.2014.2313123.

[29] A. A. Akinyelu and P. Blignaut, ''Convolutional neural network-based methods for eye gaze estimation: A survey,'' *IEEE Access*, vol. 8, pp. 142581–142605, 2020, doi: 10.1109/ACCESS.2020.3013540.

[30] G. Ren, Y. Zhang, and Q. Feng, ''Gaze estimation based on attention mechanism combined with temporal network,'' *IEEE Access*, vol. 11, pp. 107150–107159, 2023, doi: 10.1109/ACCESS.2023.3317013.

[31] Y. Cheng and F. Lu, ''Gaze estimation using transformer,'' in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 3341–3347, doi: 10.1109/ICPR56361.2022.9956687.

[32] Z. Cheng and Y. Wang, ''Multi-task gaze estimation via unidirectional convolution,'' 2024, *arXiv:2411.18061*.

[33] A. Ruzzi, X. Shi, X. Wang, G. Li, S. De Mello, H. J. Chang, X. Zhang, and O. Hilliges, ''GazeNeRF: 3D-aware gaze redirection with neural radiance fields,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 9676–9685.

[34] J. Qin, X. Zhang, and Y. Sugano, ''UniGaze: Towards universal gaze estimation via large-scale pre-training,'' 2025, *arXiv:2502.02307*.

[35] F. Ryan, A. Bati, S. Lee, D. Bolya, J. Hoffman, and J. M. Rehg, ''Gaze-LLE: Gaze target estimation via large-scale learned encoders,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2025, pp. 28874–28884.

[36] P. Vuillecard and J. Odobez, ''Enhancing 3D gaze estimation in the wild using weak supervision with gaze following labels,'' in *Proc. Comput. Vis. Pattern Recognit. Conf.*, 2025, pp. 13508–13518.

[37] Y. Xiao, X. Bai, B. Chen, H. Su, H. He, L. Xie, and E. Yin, ''De2Gaze: Deformable and decoupled representation learning for 3D gaze estimation,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2025, pp. 3091–3100.

[38] K. Funes Mora, F. Monay, and J.-M. Odobez, ''Eyediap database: Data description and gaze tracking evaluation benchmarks,'' Idiap Res. Inst., Tech. Rep. Idiap-RR Idiap-RR-08-2014, 2014.

[39] J. He, K. Pham, N. Valliappan, P. Xu, C. Roberts, D. Lagun, and V. Navalpakkam, ''On-device few-shot personalization for real-time gaze estimation,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1149–1158.

[40] T. Guo, Y. Liu, H. Zhang, X. Liu, Y. Kwak, B. I. Yoo, J.-J. Han, and C. Choi, ''A generalized and robust method towards practical gaze estimation on smart phone,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, p. 0.

[41] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, ''It's written all over your face: Full-face appearance-based gaze estimation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 2299–2308.

[42] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Guang Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, ''MediaPipe: A framework for building perception pipelines,'' 2019, *arXiv:1906.08172*.

[43] J. Hu, L. Shen, and G. Sun, ''Squeeze-and-excitation networks,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[44] S. Ioffe and C. Szegedy, ''Batch normalization: Accelerating deep network training by reducing internal covariate shift,'' in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[45] Y. Wu and K. He, ''Group normalization,'' in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 128, 2018, pp. 3–19.

[46] T. W. MacFarland and J. M. Yates, ''Mann–Whitney U test,'' *Corsini encyclopedia Psychol.*, pp. 103–132, Jul. 2016.

[47] P. S. Holzman, ''Eye-tracking dysfunctions in schizophrenic patients and their relatives,'' *Arch. Gen. Psychiatry*, vol. 31, no. 2, pp. 143–151, Aug. 1974, doi: 10.1001/archpsyc.1974.01760140005001.

SARDOR ABDIRAYIMOV ODIL UGLI was born in Qashqadaryo, Uzbekistan, in 2003. He received the B.S. degree in AI and big data from Woosong University, Daejeon, Republic of Korea, in 2025.

Since 2024, he has been a Researcher with the Digital Therapeutics Research Team, CHA Bundang Medical Center, Republic of Korea. His research interests include the application of artificial intelligence, deep learning, and embedded software with machine learning integration.

SE-YOUNG BAK was born in Seoul, Republic of Korea, in 1996. He received the B.S. degree in physical therapy from Yonsei University, Wonju, in 2021, and the M.S. degree in rehabilitation science from Korea University, Seoul, in 2023. He is currently pursuing the Ph.D. degree in integrative medicine with Yonsei University, Seoul.

From March 2023 to September 2023, he was an Intern with Naver Healthcare Laboratory, Naver Inc., Seongnam, Republic of Korea. Since November 2023, he has been a Researcher with the Department of Rehabilitation Medicine, CHA Bundang Medical Center, Seongnam. His research interests include rehabilitation, physical therapy, digital health care, and biomechanics. He is a Regular Member of Korean Society of Biomechanics and Korean Society for Precision Engineering.

YUBIN KIM received the B.S. degree in occupational therapy from Dongnam Health University, Suwon, South Korea, in 2022. She is currently pursuing the M.S. degree in translational medicine with CHA University, Seongnam, Republic of Korea.

She was previously a Research Assistant with the Department of Rehabilitation Medicine, Soonchunhyang University Hospital, Seoul, Republic of Korea. She is a Research Assistant with the Department of Rehabilitation Medicine, CHA Bundang Medical Center, Seongnam. Her research interests include stroke, functional recovery, and brain imaging.

EUN-HYE CHUNG received the B.S. degree in biomedical mechatronics engineering from Soonchunhyang University, Asan, South Korea, in 2021, and the M.S. degree in medical device industry from the Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Seoul, South Korea, in 2023.

From 2021 to 2023, she was a Graduate Researcher with the NUERI Laboratory, Samsung Medical Center, focusing on human motion analysis using sEMG, 3D motion capture, and metabolic data in robotic rehabilitation studies. Since 2023, she has been a Researcher with the CHA Bundang Medical Center, where biomechanical motion analysis and clinical research were conducted. Her research interests include biomechanics, digital therapeutics, and rehabilitation technology.

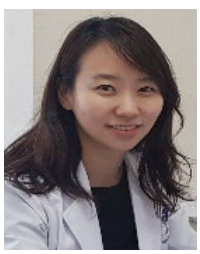
Ms. Chung was awarded the Best Paper Award from the Korean Society of Biomechanics and won the Grand Prize at the SAIHSTer Aqua competition for her independent research.

HEEGOO KIM received the B.S. degree in physical therapy from Yonsei University, Wonju, South Korea, in 2015, and the integrated M.S. and Ph.D. degree in health sciences and technology from the Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, South Korea, in 2023. Her major field of study is neurorehabilitation.

She is currently a Senior Researcher with the Digital Therapeutics Research Team, CHA Bundang Medical Center, Seongnam, South Korea. Her research focuses on neurorehabilitation technologies, including non-invasive brain stimulation and functional near-infrared spectroscopy (fNIRS). She has contributed to multiple interdisciplinary research initiatives aimed at enhancing the early prediction of stroke recovery. She has actively participated in national research collaborations in the fields of digital therapeutics and neurotechnology and has co-authored several peer-reviewed articles.

EUNYOUNG CHO received the Ph.D. degree in clinical psychology from Duksung Women's University, Seoul, South Korea, in 2021.

Since 2022, she has been the Chief of Psychology of the Department of Rehabilitation Medicine, CHA Bundang Medical Center, Gyeonggi-do, South Korea. She is also a Cogstate Clinician, conducting centralized assessment reviews for global clinical trials, and an Adjunct Professor with CHA University and Duksung Women's University. She is a licensed Clinical Psychologist and a Clinical Psychology Specialist in South Korea, in 2012 and 2013, respectively. She completed the ADOS-2 Introductory and Toddler Module Training and the Advanced Training at the University of California, San Francisco, in 2023 and 2024, respectively. Her recent publications include studies on Korean Version of the Oxford Cognitive Screen (K-OCS), including a normative study Annals of Rehabilitation Medicine, in 2024, Validation Research, in 2025, and a review of post-stroke cognitive assessment tools *Journal of Rehabilitation Psychology*, in 2020. Her research interests focus on neuropsychological assessment, digital cognitive rehabilitation, and early autism screening using eye-tracking technology. She is a member of Korean Clinical Psychology Association and Korean Psychological Association. She received the Award for Outstanding Project from Korea Health Industry Development Institute, in 2019, and a Special Research Grant from Korean Academy of Rehabilitation Medicine, in 2023.



MIRI SUH received the M.D. and M.S. degrees in medicine from Yonsei University, Seoul, Republic of Korea, in 2010 and 2016, respectively.

She was a Clinical Research Assistant Professor with Yonsei Gangnam Severance Hospital, Seoul, from 2016 to 2018, and a Clinical Assistant Professor with the Department of Rehabilitation Medicine, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea, from 2018 to 2019. Since 2019, she has been an Assistant Professor with the Department of Rehabilitation Medicine, CHA Bundang Medical Center, CHA University. She has published her experiences in more than 15 journals related to her research fields. Her research interests include tissue regeneration and pediatric neurorehabilitation.

SEYOUNG SHIN received the M.D. and M.S. degrees in medicine from Korea University, Seoul, Republic of Korea, in 2016 and 2020, respectively, and the Ph.D. degree in medicine from Sungkyunkwan University, Suwon, Republic of Korea, in February 2025.

She has been a Board-Certified Specialist in Rehabilitation Medicine, since 2021. She is currently an Assistant Professor with the Department of Rehabilitation Medicine, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea. Her research interests include the prediction of stroke prognosis and rehabilitation. She is a member of Korean Academy of Rehabilitation Medicine and Korean Society for Neurorehabilitation. She has not yet received any awards or served on the IEEE committees or publications.

HYEONGMIN JEON received the B.S. and integrated M.S. and Ph.D. degrees in biomedical engineering from Konkuk University, Republic of Korea, in 2014 and 2020, respectively.

From 2020 to 2021, he was a Postdoctoral Fellow in biomedical engineering with Konkuk University. Since 2022, he has been a Team Leader and a Professor with the Digital Therapeutics Research Team, CHA Bundang Medical Center, and the Medical School of CHA University, Republic of Korea. His research interests include the development of rehabilitation and medical technologies utilizing digital health. He is the Academic Director of Korean Society for Precision Engineering and Korean Society of Biomechanics and holds five patents. His awards and honors include the best paper and presentation, and the Young Outstanding Engineer Award.

MINYOUNG KIM received the M.D. degree from Yonsei University College of Medicine, Republic of Korea, in 1991, and the M.S. and Ph.D. degrees in medicine from Yonsei University, Seoul, Republic of Korea, in 2002. She also undertook a Postdoctoral Training with the Department of Neurosciences, University of California, San Diego, in 2007 and 2008. She completed the Internship and Resident Ship of Rehabilitation Medicine at Severance Hospital, Seoul, and became a Board-Certified Specialist, in 1996. She conducted clinical and research activities as an Instructor with Yonsei University College of Medicine, from 1996 to 1999, and was a Senior Invited Research Scientist with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea, from 2000 to 2002. Since 2003, she has been a Professor with the Department of Rehabilitation Medicine, CHA Bundang Medical Center, CHA University School of Medicine, Republic of Korea. Her clinical and research interests include brain rehabilitation and pediatric developmental abnormalities. Her clinical trials and translational research on children with developmental disorders. She has served as the Chairperson of the Board of the Academy of Pediatric Rehabilitation and as a Board Member of Korean Academy of Rehabilitation Medicine, Korean Society for Neurorehabilitation, and Korean Society for Digital Therapy. She is also a Regular Member of the National Academy of Medicine of Korea and a Scientific Committee Member of the International Alliance of Academies of Childhood Disability.