2026 International Conference on Automatic Face and Gesture Recognition (FG)

DiffEyeSyn: User-specific Subtle Eye Movement Synthesis Using Diffusion Models

Chuhan Jiao[1], Guanhua Zhang[1], Yeonjoo Cho[1], Zhiming Hu[2], Andreas Bulling[1]
[1] University of Stuttgart
[2] The Hong Kong University of Science and Technology (Guangzhou)

***Abstract*— Simulating realistic human eye movements is important in computer graphics and human-computer interaction. Prior work has mainly focused on low-frequency gaze data or predicting scanpaths from visual stimuli, often overlooking the subtle movements that reflect natural eye behaviour. These subtle dynamics are known containing rich information of an individual. We present DiffEyeSyn, the first computational method for synthesising realistic eye movements that capture these fine-grained details. Our key idea is to model subtle variations in gaze data as a user-specific noise that can be injected into any given sequence. We formulate this injection task as a conditional diffusion process in which the synthesis is conditioned on user-specific embeddings extracted from the gaze data using pre-trained models for user authentication. We propose user identity guidance – a novel loss function that allows the model to produce movements that remain both realistic and consistent with individual characteristics. Experiments on two public datasets show that DiffEyeSyn produces eye movements that are more realistic than baseline methods in terms of velocity distribution and preserving user-specific information. Furthermore, we demonstrate that DiffEyeSyn can synthesise large-scale gaze data and support various downstream tasks, such as gaze-based user identification. As such, our work can serve as a post-processing method for existing scanpath prediction approaches and provides a foundation for applications such as character animation and eye movement biometrics.**

I. Introduction

Synthesising realistic human eye movements presents a longstanding challenge at the intersection of computer graphics, eye tracking, and human-computer interaction (HCI). Realistic eye movement synthesis holds immense potential in several domains, including high-fidelity avatar animation [32], [8], [39], [31], [45], [29], [44], [49], providing an efficient alternative to laborious in-lab eye tracking by simulating human visual attention [24], [2], [55], or augmenting existing eye tracking datasets to advance gaze-based applications [21], [23], [22].

While low-resolution eye gaze recordings (typically under 30 Hz) can fulfil the requirements of some applications like gaze-based activity recognition [5], [20], high-resolution eye gaze signals are key to numerous applications such as fast eye movement detection [33], [15], [7] that detects subtle eye movements like microsaccades from high-resolution gaze recordings as well as gaze-based interaction [12], [1] that reduces the system delay using high-frequency eye gaze data. More importantly, recent studies in eye tracking research have revealed that the high-frequency eye gaze signals contain rich user-specific information than low-frequency ones [19], [21], [37], [36]. For example, the performance of using 60-second 30 Hz eye movements falls far short of the performance using 5-second 1,000 Hz eye movements in gaze-based user authentication [37]. This indicates that high-frequency user-specific information is significant for modelling eye movement behaviours of individual users and has great relevance with various applications such as digital animation and gaze-based biometrics. Despite their importance, existing methods for gaze synthesis have largely focused on low-frequency signals [24] and stimulus-driven scanpath prediction [55], [2], [51], which do not capture the subtle details present in high-resolution data. As a result, current approaches fall short of producing realistic eye movements suitable for high-fidelity applications.

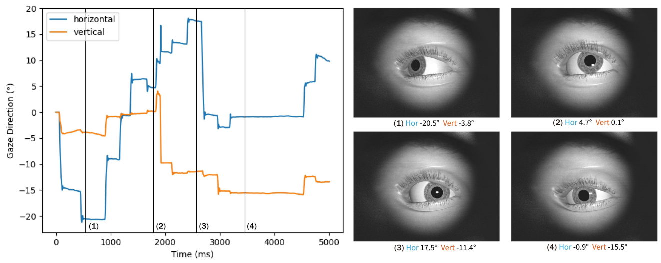

Fig. 1. DiffEyeSyn synthesises eye movements (Left) by injecting subtle eye movements which contain user-specific information into any base sequence, demonstrating the potential for applications such as post-processing of scanpaths, eye movement biometrics, and rendering realistic eye animations (Right, rendered using the tool proposed in [52], with further examples in the supplementary video).

In this work, we present *DiffEyeSyn*—the first computational method to human eye movements that reflect the fine-grained, user-specific dynamics observed in real data. Rather than focusing on where the eye moves in response to visual stimuli, we address the orthogonal challenge of generating how the eye moves, capturing the subtle physiological patterns unique to individuals. Our key idea is to treat these variations as user-specific noise that can be injected into any base eye movement sequence. We formulate this as a conditional diffusion process, where the generation is guided by user-specific embeddings extracted from gaze data using pre-trained user authentication models. In addition, we propose *user identity guidance* – a novel loss function which ensures the generated gaze sequences remain realistic and preserve information of individuals. Experiments on two public eye movement biometric datasets show that the

DiffEyeSyn synthetic eye movements that closely match real human data. Our generated sequences outperform existing baselines in terms of velocity distributions and user identity preservation. In addition, we demonstrate that DiffEyeSyn can increase fidelity for existing gaze prediction methods, synthesise eye tracking data at scale, and support various downstream tasks, including gaze-based user identification.

II. Related Work

A. Eye Movement Synthesis

Eye movement synthesis research focuses on generating realistic eye movement patterns. The synthesised data are valuable for various applications, including video rendering of virtual eyes and gaze data augmentation [13], [30]. Early works primarily originated from general signal processing and computer graphics research and built statistical models [57], [10], [13], [30]. For example, Eyecatch [57] used Kalman filter to simulate mainly saccades and smooth pursuits. Other works attempted to link and generate gaze together with head movements [39], [31]. For instance, Le et al. [31] nonlinearly mapped features of gaze, eyelid motion and head motion to a high-dimensional feature space and then generated these modalities.

Recent works have started to incorporate deep learning methods to generate eye movement data. For example, Fuhl et al. [14] leveraged a variational autoencoder (VAE) to simulate eye-tracking data without specific stimuli. Prasse et al. [46], [47] proposed SP-EyeGAN to synthesise subtle eye movements in-between the known fixation locations. SUPREYES [21] aimed at gaze super-resolution, i.e., synthesised high-frequency eye movement data upon existing low-frequency data. However, this method requires training separate models for different input frequencies. Jiao et al. developed a diffusion-based method, DiffGaze, to synthesise eye movements at 30 Hz given the 360° image [24].

However, existing methods are limited to generating scanpaths or low-frequency eye movements, overlooking the subtle, user-specific dynamics in real human eye movements. In contrast, DiffEyeSyn is the first to synthesise fine-grained eye movements that reflect individual characteristics. A key advantage of DiffEyeSyn is its ability to inject these subtle dynamics into any existing gaze sequence, including scanpaths. This makes it suitable as a post-processing tool for stimulus/task-driven scanpath prediction models. In addition, DiffEyeSyn supports a range of applications, such as gaze-based biometrics and high-fidelity eye movement animation.

B. High-Frequency Eye Movement

The peak speed of human eye movement can reach up to 700°$/s$ in real life [11]. High-frequency eye movement is hence pivotal in capturing the rapid and subtle movements of human eyes, e.g., microsaccades that typically last for a very short time like 25 ms [28]. It is also essential for reducing the latency of real-time gaze-based interaction [1].

Moreover, high-frequency eye movements are also key to user biometrics [40], [42], [41], [3], [6], [34]. Previous works have demonstrated that high-frequency eye movements (e.g. 1,000 Hz) perform significantly better than low(er)-frequency eye movements (e.g. 30 Hz) with the same duration in gaze-based user authentication and identification [36], [37], [21], [19].

Since the primary difference between high- and low-frequency eye movement recordings lies in subtle eye dynamics, our approach focuses on synthesising these fine-grained movements and injecting them into any base gaze sequence. Building on earlier findings that these subtle movements encode rich, user-specific information, we are the first to leverage pre-trained user authentication models for subtle eye movement synthesis and its corresponding evaluation.

III. Background: Denoising Diffusion Probabilistic Models

Denoising diffusion probabilistic models (DDPMs) [18] is a type of generative model that has demonstrated state-of-the-art performance on various data synthesis tasks, such as image [9], [48], [4], [27], [24]. DDPMs consist of two computational processes: A forward process and a reverse process.

In the forward process, the goal is to shift the input data distribution to a normal distribution by adding random noise according to

$$q\left(x_t \mid x_0\right)=\mathcal{N}\left(x_t ; \sqrt{\bar{\alpha}_t} x_0,\left(1-\bar{\alpha}_t\right) I\right), \tag{1}$$

where x_0 is the original input data, x_t is the noisy data at timestep t, and $\bar{\alpha}_t=\prod_{s=1}^t \alpha_s$ is a hyper-parameter that can be calculated given a predefined noise schedule.

During the reverse process, given noisy data x_t, a deep learning model f_θ is trained to predict the added noise at timestep t: The training objective of the model is to minimise the difference between the predicted and the actual noise at each timestep:

$$\mathcal{L}_{simple}=\|\epsilon_t-\hat{\epsilon}_t\|_2, where\ \hat{\epsilon}_t=f_\theta(x_t,t). \tag{2}$$

The generated data is initialised from the normal distribution x_t, and realistic data is obtained after T denoising steps. The reverse process can be formalised as:

$$p_\theta\left(x_{t-1} \mid x_t\right)=\mathcal{N}\left(x_{t-1} ; \mu_\theta(x_t, t), \sigma_\theta(x_t, t) I\right), \tag{3}$$

where $\mu_\theta(x_t,t)$ can be determined given the predicted noise $\hat{\epsilon}_t$ while $\sigma_\theta(x_t,t)$ is typically a constant value to simplify the optimisation.

IV. DiffEyeSyn

A. Problem Definition

Realistic simulation of human eye movements requires capturing the subtle variations that can be captured by commercial eye trackers, but are often overlooked in prior eye movement synthesis works. Prior work in eye movement biometrics (EMB) has shown that these fine-grained dynamics contain rich user-specific information [21], [37], [42]. Models trained on high-frequency eye movements (e.g., 1,000 Hz) significantly outperform those trained on lower-frequency data (e.g., 30 Hz) with the same duration for tasks

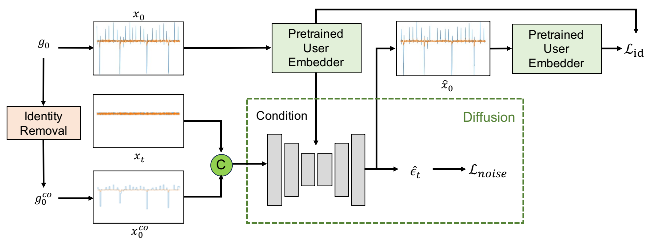

Fig. 2. Pipeline of training DiffEyeSyn. DiffEyeSyn is trained in a self-supervised way. The original eye movement data g_0 and its identity-removed variant g_0^{co} are converted into velocities x_0 and x_0^{co}. The objective is to train a diffusion model that learns to reintroduce the subtle, user-specific variations removed from x_0^{co}, thereby reconstructing x_0. We use a pretrained user embedder to extract the user-specific embedding from the x_0. At each diffusion timestep t, given the x_0^{co} and the user embedding as the condition, DiffEyeSyn predicts the noise $\hat{\epsilon}_t$ that converts the x_t to x_0. Instead of only optimising DiffEyeSyn with the normal diffusion loss which minimises the difference between the predicted and ground truth noise, we propose user identity guidance $\mathcal{L}_{id}$ - a novel loss function to constraint the synthesised data contains the given user-specific information. More specifically, we estimate the cleaned eye movement velocity $\hat{x_0}$ by denoising x_t with the predicted noise $\hat{\epsilon}_t$. The proposed user identity guidance maximises the cosine similarity between the embedding of the $\hat{x_0}$ and x_0.

such as user authentication and identification [37]. Building on these insights, we frame the task of simulating realistic fine-grained eye movements as a process of injecting user-specific noise into a given eye movement sequence.

Diffusion models are well-suited to this task, as they are designed to gradually add and remove noise in a learnable way, and have shown success across a range of generative problems [18], [24], [48], [54], [27]. We formulate the injection task as a conditional diffusion process that synthesises eye movement sequences guided by both a base input and a user-specific embedding. The embedding is extracted from a short sequence of gaze data using a pretrained user authentication model (see Section IV-B for details).

An overview of the training pipeline is shown in Figure 2. First, we use a pretrained EMB model to extract a user embedding from the original eye movements. Next, we remove the user-specific characteristics from the original signal to create a base for injection (see Section V-A). The model is trained to reconstruct the personalised eye movements from the base input, conditioned on the user embedding. At inference time, given a user embedding and a base sequence of eye movements, the trained model injects the fine-grained, user-specific eye movements into the sequence. (see Section IV-E).

To support compatibility with existing EMB models, which often operate on velocity domain [36], [37], DiffEyeSyn is designed to synthesise gaze velocities. Gaze positions can then be recovered by integrating these velocities from a known starting location.

B. Conditional Diffusion Model for Eye Movement Synthesis

We formulate the task described in Section IV-A as a conditional diffusion process trained in a self-supervised manner. The training pipeline is illustrated in Figure 2. Let x_0 denote the velocity signal of the ground-truth gaze data, which includes the subtle, user-specific eye movement patterns. To serve as a base for the injection, we construct x_0^{co}, the corresponding velocity sequence from which identity-related information (i.e., subtle eye movements) has been removed. We employ a pretrained user authentication model E to extract the user embedding from the original gaze data, denoted as $E(x_0)$. DiffEyeSyn takes both the identity-removed gaze sequence x_0^{co} and the user embedding $E(x_0)$ as conditional inputs. The objective is to learn a modelled distribution $p_\theta(x_0 \mid x_0^{co}, E(x_0))$ that approximates the true conditional distribution $q(x_0 \mid x_0^{co}, E(x_0))$. In other words, the model learns to reconstruct realistic gaze data that includes subtle, user-specific dynamics, given a reference sequence and a user embedding.

To achieve this, we extend the reverse processes of the DDPM to conditional ones. The objective is the same as the original DDPM that minimises the difference between the predicted noise $\hat{\epsilon}_t$ and the ground truth noise ϵ_t:

$$\mathcal{L}_{noise} = \|\epsilon_t - \hat{\epsilon}_t\|_2, where\ \hat{\epsilon}_t = f_\theta(x_t, t \mid (x_0^{co}, E(x_0))). \tag{4}$$

Furthermore, the reverse process in Equation 3 is extended to:

$$p_\theta\left(x_{t-1} \mid x_t\right) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t \mid (x_0^{co}, E(x_0))), \sigma_\theta(x_t, t \mid (x_0^{co}, E(x_0)))I\right). \tag{5}$$

C. User Identity Guidance

Training the model with $\mathcal{L}_{noise}$ enables it to learn the general distribution of human eye movements. However, this does not explicitly encourage the model to preserve user identity, which is strongly associated with subtle eye movement patterns. To address this, we introduce *user identity guidance*, inspired by classifier-guided diffusion models in image generation [9], that leverages a pretrained user authenticator E to help DiffEyeSyn capture user-specific, subtle variations.

At timestep t, we obtain x_t by adding a sampled noise ϵ_t on the target x_0. We then predict the added noise $\hat{\epsilon}_t$. With the predicted noise $\hat{\epsilon}_t$, we can estimate the target $\hat{x_0}$ by denoising x_t using:

$$\hat{x_0} = \frac{x_t - \sqrt{1-\alpha_t}\hat{\epsilon}_t}{\sqrt{\alpha_t}} \tag{6}$$

To preserve user-specific features, we constrain the user embedding of the reconstructed signal $\hat{x_0}$ to remain close to that of the original x_0. We use cosine similarity in the embedding space of the pretrained user authenticator E, following [37]. The identity guidance loss $\mathcal{L}_{id}$ is defined as:

$$\mathcal{L}_{id} = 1 - \text{CosineSimilarity}(E(x_0), E(\hat{x_0})). \tag{7}$$

Therefore, the training objective is

$$\mathcal{L} = \mathcal{L}_{noise} + \lambda\mathcal{L}_{id} \tag{8}$$

D. Data Preprocessing

The output of eye trackers is commonly in the form of continuous 2D on-screen gaze locations (x, y) or 3D gaze vectors (x, y, z). However, the 2D on-screen gaze locations vary across different experimental settings. To increase the generalisability, DiffEyeSyn uses 3D gaze data in the form of degrees of visual angle, which is represented by the angle of the x-axis and the angle of the y-axis.

Given a sequence of ground truth eye movements, we first remove the subtle eye movements containing user identity information (Details are described in Section V-A). To obtain the x_0 and x_0^{co} for our model, we estimate the velocities (in the unit of degrees per second) of the original eye movements and noisy eye movements in both x-axis and y-axis following [37]. The summarisation of data preprocessing and DiffEyeSyn training can be found in our supplementary materials.

E. Inference

At inference time, DiffEyeSyn injects subtle movements that encode user-specific information into any given base eye movement sequence, including scanpaths. This enables the generation of realistic, identity-preserving fine-grained gaze signals based on a target user's characteristics. A summary of DiffEyeSyn inference can be found in our supplementary.

F. Architecture

The main challenge of the task is handling very long gaze trajectories within a feasible time frame. For example, a 5-second 1,000 Hz gaze trajectory contains 5,000 samples, which makes it extremely slow for training transformer-based architectures such as DiffGaze [24]. To solve this issue, we built DiffEyeSyn upon DiffWave [27], leveraging bidirectional dilated convolutions (Bi-DilConv) for memory- and time-efficient training and inference for very long input sequences.

The main difference to DiffWave architecture is the condition processing. DiffWave employs a mel-spectrogram as its only condition. In contrast, DiffEyeSyn has two conditions: an observation x_0^{co} and a user embedding $E(x_0)$. Given x_0's length L, we merge the noisy input x_t with x_0^{co} along the feature axis. The concatenated tensor with shape $(4, L)$ is then processed by a 1D convolutional layer with the output channel of C. In addition, a fully connected layer broadcasts the user embedding $E(x_0)$ to length L before feeding into the residual layer. Details of DiffEyeSyn architecture can be found in our supplementary materials.

V. Experiments

A. Experimental Setup

Pretrained User Authenticator. Because subtle eye movements are known to reflect user-specific behavioural signatures, a key goal of our method is synthesise user-specific eye movements. As such, it is essential to evaluate whether the generated sequences retain user-specific characteristics.

However, there is no direct method to assess whether a generated sequence belongs to a particular user. Inspired by the Fréchet Inception Distance (FID) [17], which evaluates image realism using a pretrained Inception model [53], we adopt a similar strategy: using a pretrained gaze-based user authentication model to extract user embeddings, which implicitly measure identity consistency.

Specifically, we employ EKYT [37], the current state-of-the-art in eye movement-based user authentication. Given a 5-second sequence sampled at 1,000 Hz, EKYT produces a 512-dimensional embedding that captures user-specific patterns, including those reflected in subtle movements. Importantly, EKYT generalises to unseen users, enabling identity-aware evaluation even for novel subjects. We use the official pretrained weights and compute cosine similarity between EKYT embeddings of the generated and ground truth sequences to quantify identity preservation.

Datasets. We used two public eye movement biometric datasets, GazeBase dataset [16] and JuDo1000 dataset [40].

GazeBase contains 1,000 Hz eye movements from 322 participants while performing seven different tasks: fixations (FXS), horizontal (HSS) and random (RAN) saccades, reading (TEX), watching two videos (VD1 and VD2), and playing gaze-driven games (BLG). To avoid the information leak, we adopted the dataset split described in the EKYT paper, as the EKYT model was trained on this dataset. There is *no overlap of users* between the training and test sets. These ensure zero information leakage from the pretrained model and we are able to test DiffEyeSyn performance on synthesising user-specific eye movements for unseen users during training.

JuDo1000 dataset contains recordings of 150 participants at 1,000 Hz while performing a single task. The task is similar to the RAN task in GazeBase dataset. We used all the recordings from JuDo1000 for cross-dataset evaluation.

Implementation Details. As we chose the pretrained EKYT model to offer user identity guidance, DiffEyeSyn produces 5-second 1,000 Hz eye movements to match the pretrained EKYT input format. We set the total diffusion steps $T = 50$ and we used a linear noise schedule $[1 \times 10^{-4}, 0.05]$. We trained DiffEyeSyn on GazeBase training set for $1M$ steps using Adam optimizer with a batch size of 32 and learning rate 2×10^{-4}.

User Identity Removal. Previous studies [37], [36], [21] suggest that low-frequency gaze data at 20-30 Hz performs poorly in gaze-based user identification and authentication. In addition, Jiao et al. [21] demonstrate that upsampling the low-frequency gaze data to a high frequency using traditional interpolation methods leads to a performance drop in user identification compared with the original low-frequency data. Based on these findings, we removed user-specific subtle eye movements in the ground truth by downsampling and interpolation. Specifically, we downsample the 1,000 Hz gaze sequences to 20 Hz and then upsample them back to 1,000 Hz using nearest interpolation. This process effectively removes fine-grained user traits while preserving coarse eye movement patterns. To verify its effectiveness, we compare EKYT embeddings before and after subtle eye movement removal. The average cosine similarity remains below 0.1 across all datasets: 0.081 (GazeBase train), 0.090 (GazeBase test), and 0.091 (JuDo1000), confirming that user-specific characteristics have been successfully attenuated (see Table 1 in the supplementary materials).

TABLE I

THE COSINE SIMILARITY BETWEEN EKYT EMBEDDINGS OF THE GROUND TRUTH EYE MOVEMENT DATA AND SYNTHETIC DATA OF DIFFERENT METHODS IN USER IDENTITY RECOVERY (R) AND USER IDENTITY MANIPULATION (M). THE RESULTS THAT ARE BETTER THAN THE HUMAN WITHIN-USER BASELINE THE MODELS ARE BOLDED. COSINE SIMILARITY BEING 1 MEANS THAT TWO EMBEDDINGS ARE EXACTLY THE SAME, WHILE 0 MEANS TWO EMBEDDINGS ARE ORTHOGONAL, I.E., NOT RELATED.

Dataset	Method	Task							
		HSS	RAN	TEX	FXS	VD1	VD2	BLG	ALL
GazeBase	Human (within-user)	0.514 ± 0.159	0.512 ± 0.154	0.506 ± 0.161	0.403 ± 0.153	0.481 ± 0.158	0.491 ± 0.160	0.468 ± 0.160	0.498 ± 0.160
	Human (cross-user)	0.111 ± 0.168	0.110 ± 0.171	0.113 ± 0.168	0.135 ± 0.159	0.117 ± 0.168	0.114 ± 0.170	0.099 ± 0.168	0.113 ± 0.169
	SP-EyeGAN [46], [47]	0.114 ± 0.136	0.134 ± 0.137	0.157 ± 0.142	0.106 ± 0.150	0.128 ± 0.145	0.121 ± 0.144	0.135 ± 0.133	0.129 ± 0.141
	High-pass Filter (R)	0.390 ± 0.170	0.407 ± 0.178	0.320 ± 0.155	0.477 ± 0.241	0.362 ± 0.178	0.362 ± 0.174	0.324 ± 0.178	0.375 ± 0.180
	DiffEyeSyn w.o.$\mathcal{L}_{id}$ (R)	0.404 ± 0.161	0.348 ± 0.173	0.352 ± 0.169	0.176 ± 0.211	0.269 ± 0.196	0.298 ± 0.193	0.407 ± 0.171	0.322 ± 0.182
	DiffEyeSyn (R)	**0.820** ± 0.063	**0.800** ± 0.064	**0.823** ± 0.079	**0.705** ± 0.104	**0.779** ± 0.081	**0.791** ± 0.076	**0.791** ± 0.081	**0.787** ± 0.078
	High-pass Filter (M)	0.181 ± 0.148	0.178 ± 0.151	0.178 ± 0.157	0.223 ± 0.143	0.190 ± 0.152	0.194 ± 0.150	0.153 ± 0.145	0.182 ± 0.151
	DiffEyeSyn w.o.$\mathcal{L}_{id}$ (M)	0.235 ± 0.178	0.219 ± 0.179	0.209 ± 0.186	0.100 ± 0.169	0.164 ± 0.181	0.171 ± 0.183	0.214 ± 0.183	0.187 ± 0.179
	DiffEyeSyn (M)	**0.668** ± 0.101	**0.664** ± 0.098	**0.696** ± 0.106	**0.454** ± 0.144	**0.629** ± 0.129	**0.644** ± 0.121	**0.643** ± 0.110	**0.628** ± 0.116
JuDo1000	Human (within-user)	-	-	-	-	-	-	-	0.685 ± 0.104
	Human (cross-user)	-	-	-	-	-	-	-	0.367 ± 0.157
	SP-EyeGAN [46], [47]	-	-	-	-	-	-	-	0.141 ± 0.136
	High-pass Filter (R)	-	-	-	-	-	-	-	0.534 ± 0.162
	DiffEyeSyn w.o.$\mathcal{L}_{id}$ (R)	-	-	-	-	-	-	-	0.226 ± 0.126
	DiffEyeSyn (R)	-	-	-	-	-	-	-	**0.695** ± 0.072
	High-pass Filter (M)	-	-	-	-	-	-	-	0.424 ± 0.139
	DiffEyeSyn w.o.$\mathcal{L}_{id}$ (M)	-	-	-	-	-	-	-	0.119 ± 0.126
	DiffEyeSyn (M)	-	-	-	-	-	-	-	0.605 ± 0.091

− not applicable

B. *User Identity Recovery*

We evaluate whether DiffEyeSyn can inject subtle, user-specific eye movement variations, crucial cues for identity, into sequences where such information has been removed.

Task Definition. This task simulates an idealised scenario: given an identity-removed input and the user embedding extracted from the original gaze data, can we recover the user's unique subtle eye movements? Although impractical in real-world use (as it requires ground truth), this setting provides an upper bound for evaluating identity fidelity.

Experimental Details. Our experiments were conducted on two datasets: the GazeBase test set, comprising 73,011 non-overlapping 5-second 1,000 Hz gaze sequences, and the JuDo1000 dataset, containing 7,200 non-overlapping 5-second 1,000 Hz gaze sequences. None of the users in the two sets are shown during training. Given that DiffEyeSyn is a generative method, we synthesised five samples per gaze sequence to mitigate sampling bias.

Evaluation Metrics and Baselines. We report the mean cosine similarity between EKYT embeddings of generated and ground truth sequences and the standard deviation. We compare DiffEyeSyn with:

- High-pass filter: a Butterworth high-pass filter to extract frequency components above 20 Hz from the ground truth, then added these back to the identity-removed eye movements.
- SP-EyeGAN [46], [47]: An existing eye movement synthesis method. It requires predefined fixation locations, mean, and standard deviation for fixation and saccade durations as input for eye movement synthesis. We consider it as a random generation baseline and trained one SP-EyeGAN model per GazeBase task. For Judo1000, we employed the SP-EyeGAN model trained on the RAN task.
- DiffEyeSyn without identity guidance loss.

Moreover, we established human baselines *within-user cosine similarity* and *cross-user cosine similarity* to better explain the results. Specifically, within-subject cosine similarity measures how similar a sequence of eye movements is to other sequences from the same user. For each user's sequence of eye movements, we calculated cosine similarities between its EKYT embedding and those of other sequences from the same user, then averaged these values across all users. Achieving an average cosine similarity close to the within-user cosine similarity suggests a method can simulate a user's average subtle gaze behaviour. An average cosine similarity higher than the within-user cosine similarity indicates that the method can simulate fine-grained movements similar to the sequence used for extracting the user embeddings.

Cross-user cosine similarity assesses the similarity between a user's eye movements and those of other users in the dataset. We calculated cosine similarities between a user's sequence of eye movements and those of sequences from other users, then averaged these values across all users. Comparing cross-user cosine similarity with within-user cosine similarity helps verify the efficacy of the pretrained user authenticator (EKYT). A gap between these two values indicates the pretrained user authenticator's ability to differentiate eye movements from different users.

Quantitative Results. Table I presents the results for both datasets. The within-user baseline strongly exceeds cross-user similarity (GazeBase: 0.498 vs. 0.113; JuDo1000: 0.685 vs. 0.367), showing EKYT reliably captures individual-specific subtle eye movements.

SP-EyeGAN performs near or below the cross-user baseline, indicating weak identity retention. DiffEyeSyn without $\mathcal{L}_{id}$ improves over SP-EyeGAN but still underperforms: on GazeBase, it trails the within-user baseline by 0.176, and on JuDo1000 it drops below even the cross-user baseline. The high-pass filter performs slightly better, but still fails to reach within-user similarity, confirming that user identity is embedded not only in high-frequency components (above 20 Hz) but also in low-frequency components.

Full DiffEyeSyn surpasses the within-user baseline across all tasks, outperforming it by at least 0.3 on GazeBase. On JuDo1000, it generalises well, achieving 0.695 vs. 0.685 for human within-user similarity. These results highlight DiffEyeSyn 's ability to recover high-fidelity, user-specific subtle eye movements.

Qualitative Results. Figure 3 shows two examples; more are in the supplementary. Compared to the identity-removed inputs and high-pass filter outputs, DiffEyeSyn produces more realistic trajectories. The high-pass filter baseline often yields unstable or invalid values, because the original input might not contain frequency components above the threshold, while DiffEyeSyn reliably injects user-specific subtle eye movement patterns, demonstrating robustness across diverse sequences.

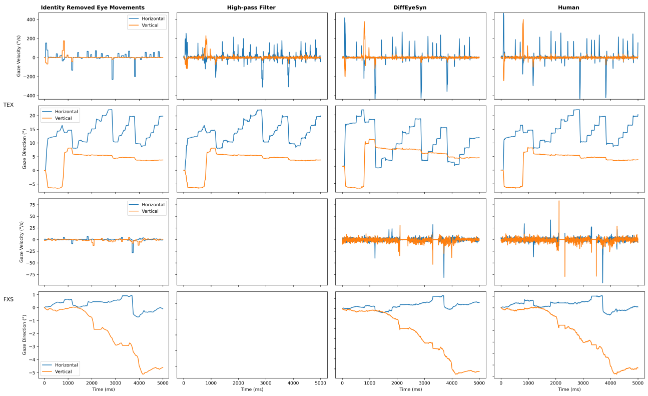

Fig. 3. Qualitative comparison between the identity removed eye movements, High-pass filter synthetic eye movements, DiffEyeSyn synthesised eye movements, and ground truth human eye movements in two example tasks within a 5-second time window. For each sequence of eye movements, we visualise its velocities (above) and gaze direction (below). A figure is empty means the method fails to produce valid results.

TABLE II

JENSEN-SHANNON DIVERGENCE BETWEEN VELOCITY DISTRIBUTION OF GROUND TRUTH HUMAN EYE MOVEMENTS AND VELOCITY DISTRIBUTION OF SYNTHETIC EYE MOVEMENTS OF DIFFERENT METHODS FROM EXPERIMENTS MENTIONED IN SECTION V-B AND V-C. THE BEST RESULTS ARE BOLDED AND THE SECOND BEST ARE UNDERLINED. R: USER IDENTITY RECOVERY, M: USER IDENTITY MANIPULATION.

Dataset	Synthetic Eye Movements	JS divergence ($\times 10^{-2}$) $\downarrow$ Fixation	Saccade	ALL
GazeBase	SP-EyeGAN [46], [47]	1.28	7.98	1.59
	High-pass Filter (R)	15.68	4.68	15.43
	DiffEyeSyn (R)	1.10	0.35	1.05
	High-pass Filter (M)	18.44	3.94	18.17
	DiffEyeSyn (M)	**1.01**	**0.27**	**1.00**
JuDo1000	SP-EyeGAN [46], [47]	5.79	2.87	4.63
	High-pass Filter (R)	6.66	4.35	6.69
	DiffEyeSyn (R)	**2.26**	**0.10**	**2.17**
	High-pass Filter (M)	7.25	8.81	7.43
	DiffEyeSyn (M)	2.69	0.36	2.59

C. User Identity Manipulation

Task Definition. This task evaluates whether DiffEyeSyn can inject target-specific identity signals—particularly subtle eye movement dynamics—into sequences from other users. Given a target user and one of their 5-second eye movement sequences, we aim to synthesise new sequences that reflect this user's unique patterns by conditioning on their identity embedding and using eye movements from a different user as the base. This simulates the real-world application scenario.

Experimental Setup. For each sequence in the GazeBase test set, we treated its user as the identity target. We then selected seven sequences from different users and tasks with near-zero cosine similarity ($\in [0, 0.05]$) to the target, ensuring the base inputs were highly dissimilar in both content and identity. This created 511,077 manipulation instances. A similar process was applied to JuDo1000, where we selected one dissimilar sequence per target. To remove subtle user-specific movements, we downsampled each base sequence to 20 Hz.

Metrics and Baselines. We used the same metrics and baselines as in Section V-B. For the high-pass filter baseline, we extracted frequency components above 20 Hz from the target user's eye movements and injected them into the downsampled base sequences.

Quantitative Results. Table I shows the performance of DiffEyeSyn and baselines. As with the recovery task, DiffEyeSyn without identity guidance fails to inject meaningful user information. Its performance on GazeBase slightly exceeds the cross-user baseline but remains far below the within-user baseline. On JuDo1000, it even underperforms the cross-user baseline.

The high-pass filter performs poorly here—only marginally above the cross-user baseline—highlighting its instability in cross-identity synthesis. In contrast, full DiffEyeSyn consistently outperforms human within-user baselines across all GazeBase tasks (by at least 0.05, avg. 0.13), confirming its ability to inject user-specific characteristics, particularly subtle gaze dynamics, into unrelated sequences. On JuDo1000, DiffEyeSyn achieves 0.605 cosine similarity—below the within-user baseline (0.685), but well above the cross-user baseline (0.367) and high-pass filter (0.424). This demonstrates DiffEyeSyn's generalisability ability to inject meaningfully user-specific, subtle eye movement patterns.

Qualitative Results. Figure 4 visualises two examples; more are in the supplementary. The high-pass filter often fails to produce valid outputs, generating unnatural jitter. In contrast, DiffEyeSyn yields more realistic sequences with plausible velocity dynamics and consistent spatial structure.

To investigate the effect of user embeddings, Figure 5 presents synthesis results using the same base input with four different target embeddings. The base gaze trajectory remains largely intact, but fine-grained differences, such as saccade magnitude, clearly appear across users. This supports that DiffEyeSyn preserves high-level gaze trajectory

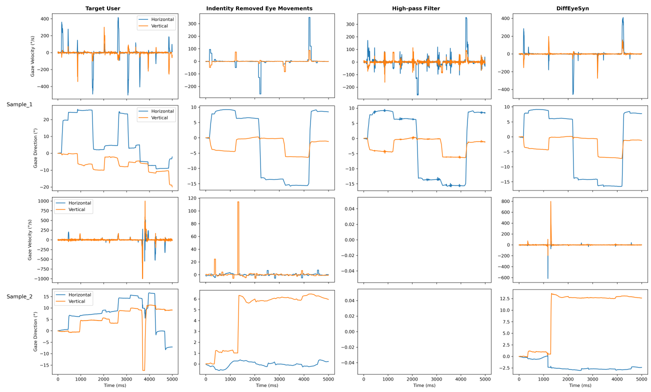

Fig. 4. Two examples of the user identity manipulation task. Left: the eye movements used to extract the target user embedding. Middle 1: the eye movements from different users that DiffEyeSyn injects the target subtle eye movements. Middle 2: High-pass filter synthesised eye movements. Right: DiffEyeSyn synthesised eye movements. For each sequence of eye movements, we visualise its velocities (above) and gaze direction (below). A figure is empty means the method fails to produce valid results.

while injecting user-specific information through subtle eye movement variations.

D. Comparison of Synthetic and Human Velocity Distributions

We further evaluated the realism of synthesised gaze data by comparing its velocity distribution against that of real eye movements, following [46], [47].

Evaluation Metrics and Baselines. Same as prior works [46], [47], [43], we used the Jensen-Shannon divergence to measure the similarity between two distributions. The JS is in the range of $[0, 1]$ where lower values indicate greater similarity. We evaluated DiffEyeSyn against SP-EyeGAN and the high-pass filter baseline using the synthesised data described in Section V.

Exprimental Details. Given the large scale of the two datasets we used, we adopted a sampling approach by selecting 1,000,000 velocity samples from each data distribution for distribution comparison. This process was repeated ten times to mitigate sampling bias, and we recorded the average JS value.

Additionally, we employed the I-VT algorithm [50] to identify fixations ($< 100°/\text{s}$) and saccades ($> 300°/\text{s}$). In line with our approach for computing the average JS divergence between the velocity distributions of all synthetic samples and the human ground truth, we also calculated and reported the JS divergence between the velocity distributions of synthetic fixations and real fixations, as well as between synthetic and human saccades.

This additional evaluation is motivated by the fact that the majority of human eye movements fall within the fixation category, which generally involves slower velocities. Consequently, a method might achieve favourable JS scores across all samples by simply modelling low-velocity fixations well, while failing to capture the more dynamic saccadic behaviour. By examining fixations and saccades separately, we gain a clearer picture of the method's ability to capture the full range of subtle and rapid behaviours.

To compute these distributions, we used the `histogram` function from `NumPy`. As velocities were clipped at a maximum of $1000°/\text{s}$, we selected `bins = 500` for all samples, `bins = 350` for saccades, and `bins = 50` for fixations. This ensured a consistent bin width of 2 across all histograms, allowing for direct comparison between JS divergence scores.

Quantitative Results. The results are summarised in Table II. Across both tasks and datasets, DiffEyeSyn consistently achieved the lowest JS divergence. This indicates that DiffEyeSyn generates synthetic data that most closely approximates the velocity distribution of real human eye movements. Notably, DiffEyeSyn retains this performance advantage not only when considering the full set of samples but also when fixations and saccades are evaluated separately. This suggests that DiffEyeSyn is capable of faithfully modelling both slow and rapid eye movement behaviours. In contrast, the high-pass filter exhibits markedly higher divergence across all categories, reflecting its limited capacity to replicate human-like eye movement subtle dynamics, which is consistent with the qualitative observations in Figure 4. While SP-EyeGAN achieves comparable performance to

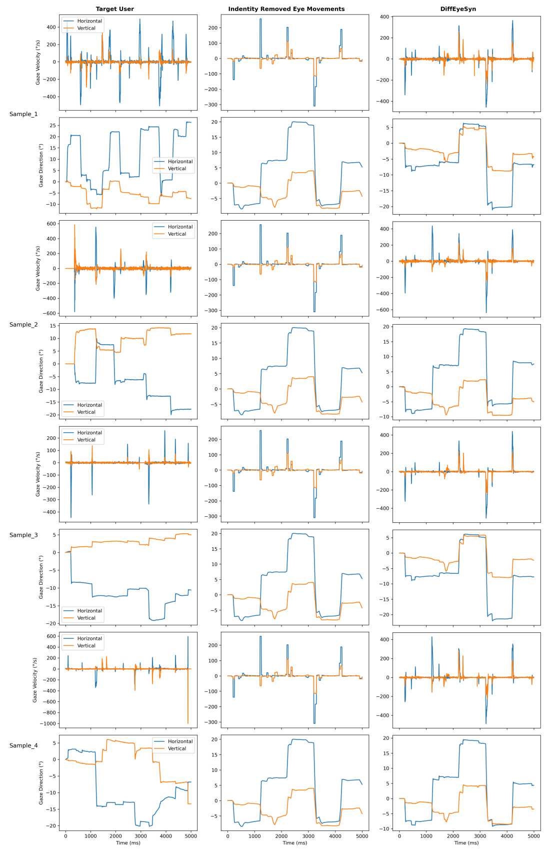

Fig. 5. Four examples of the user identity manipulation task with the same base eye movement sequence. Left: the eye movements used to extract the target user embedding. Middle: the eye movements that DiffEyeSyn injects the target user's subtle eye movements. Right: DiffEyeSyn synthesised eye movements. For each sequence of eye movements, we visualise its velocities (above) and gaze direction (below).

DiffEyeSyn on fixation data, it performs noticeably worse in modelling saccadic movements, further highlighting the strengths of our approach in capturing fast eye movement behaviour.

E. Augmenting Gaze-based User Identification

We evaluated whether DiffEyeSyn-generated trajectories improve user identification accuracy by augmenting real gaze datasets.

Experimental Details. We conducted experiments with 59 users from the GazeBase test set. For each user, we split the data 50/50 into training and testing. We then augmented the training set by adding seven synthetic samples per real trajectory, generated via the user identity manipulation. We extended the EKYT model with a final linear classification layer, and trained two models from scratch: one using only real data, and another using the augmented set. Training was run for 100 epochs with Adam (learning rate 0.001, batch size 128), repeated across five random seeds to minimise training variability. We report the average best classification accuracy and equal error rate (EER) following [21].

As shown in Table III, models trained on DiffEyeSyn-augmented data achieve higher accuracy and lower EER. This highlights DiffEyeSyn's ability to inject identity-preserving signals, particularly subtle eye movements, into any base sequences, making them valuable for downstream tasks.

TABLE III

USER IDENTIFICATION RESULTS OF THE SAME MODEL TRAINED ON THE GAZEBASE (HUMAN) DATASET AND THE DIFFEYESYN AUGMENTED DATASET. THE BEST RESULTS ARE BOLDED. ACC: CLASSIFICATION ACCURACY.

Training data	Acc. (%) ↑	EER (%) ↓
Human	57.77	1.72
Augmented	**59.63**	**1.70**

VI. DISCUSSION

Applications. A key strength of DiffEyeSyn is its ability to inject subtle eye movement dynamics into any base sequence. This enables a novel post-processing strategy to enhance outputs from existing scanpath or gaze prediction models, particularly in applications requiring naturalistic gaze behaviour, such as virtual humans (see Figure 1 and supplementary video). Although the 1,000 Hz sampling rate used in this work exceeds the refresh rates of current commercial displays, we selected this setting for practical reasons: it allows us to leverage the pretrained user authenticator EKYT to reconstruct fine-grained, user-specific subtle eye movements. Achieving comparable EKYT performance with lower-frequency gaze data (e.g., 72 Hz) requires binocular recordings from approximately 9,200 unique users [38], data currently only available in private datasets [38]. Given access to such data, DiffEyeSyn could be retrained to synthesise binocular eye animations at 60–120 Hz by adjusting the input dimensions, thereby making it directly compatible with modern display hardware.

As shown in Section V-E, the subtle eye motions are not only important for making eye movements look natural, but they can also help in tasks like user authentication/identification. This means DiffEyeSyn could be used to augment training data for eye movement biometric systems.

Performance upper bound. DiffEyeSyn uses a pretrained model to guide how it adds subtle eye movement features. In our case, we use EKYT [37], but the framework can work with other models as well. The quality of the synthetic eye movements, especially the fine-grained eye movements containing characteristics of individuals, depends on how good this guiding model is. As better user embedding models are developed in the future, DiffEyeSyn can benefit directly by using them for better control and output quality of synthetic eye movements.

Quantitative evaluation of synthetic eye movements. Evaluating the realism of synthesised eye movements remains an open challenge, particularly for subtle behaviours that are hard to capture with existing metrics. We follow prior work [46], [47] in using the Jensen-Shannon divergence over velocity distributions as a proxy for realism, and further distinguish between fixations and saccades to separate fast movements from slow ones. To assess the preservation of fine-grained behavioural cues, we additionally evaluate the synthesised data using a pretrained user recognition model, since the subtle dynamics are known to contain rich user-specific information. While EKYT is used during both training and evaluation, this is common in generative modelling [25], [26], [56], [35], where a shared model (e.g., CLIP) is used for both control and evaluation. As the field advances, stronger and more generalisable embedding models could serve as both improved loss functions and evaluation tools for measuring eye movement synthesis quality.

Limitation and future work. We evaluated DiffEyeSyn on two high-quality datasets recorded with high-speed eye trackers and fixed head positions, which helped the model learn and reproduce subtle eye movement patterns. However, many modern eye trackers in head-mounted devices work at lower sampling rates and capture both eye and head motion, which is different from our current evaluation setting. We plan to explore adapting DiffEyeSyn to VR/AR setting in future work.

VII. CONCLUSION

We presented DiffEyeSyn, the first method for synthesising eye movements that closely mirror the fine-grained patterns seen in real human behaviour. Through extensive evaluation, we showed that DiffEyeSyn captures subtle variations and can inject these variations into any given base sequence. Our results demonstrate our synthetic eye movements are more realistic than baselines, Moreover, the synthetic sequences can enhance downstream tasks like gaze-based user identification, suggesting broader potential for applications in virtual reality, eye movement animation, and human-computer interaction. As such, DiffEyeSyn sets a foundation in high-fidelity eye movement synthesis.

VIII. ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon Europe research and innovation funding program under grant agreement No. 101072410.

REFERENCES

[1] A. N. Angelopoulos, J. N. Martel, A. P. Kohli, J. Conradt, and G. Wetzstein. Event-based near-eye gaze tracking beyond 10,000 hz. *IEEE Transactions on Visualization & Computer Graphics*, 27(05):2577–2586, 2021.

[2] M. Assens, X. Giro-i Nieto, K. McGuinness, and N. E. O'Connor. Pathgan: Visual scanpath prediction with generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.

[3] A. Bayat and M. Pomplun. Biometric identification through eye-movement patterns. In *Advances in Human Factors in Simulation and Modeling: Proceedings of the AHFE 2017 International Conference on Human Factors in Simulation and Modeling, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8*, pages 583–594. Springer, 2018.

[4] T. Brooks, B. Peebles, C. Holmes, W. DePue, Y. Guo, L. Jing, D. Schnurr, J. Taylor, T. Luhman, E. Luhman, C. Ng, R. Wang, and A. Ramesh. Video generation models as world simulators. 2024.

[5] A. Bulling, C. Weichel, and H. Gellersen. Eyecontext: Recognition of high-level contextual cues from human visual behaviour. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 305–308, 2013.

[6] N. V. Cuong, V. Dinh, and L. S. T. Ho. Mel-frequency cepstral coefficients for eye movement identification. In *2012 ieee 24th international conference on tools with artificial intelligence*, volume 1, pages 253–260. IEEE, 2012.

[7] E. Dalmaijer. Is the low-cost eyetribe eye tracker any good for research? Technical report, PeerJ PrePrints, 2014.

[8] Z. Deng. Automated eye motion using texture synthesis. *IEEE Computer Graphics and Applications*, 25(2):24–30, 2005.

[9] P. Dhariwal and A. Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[10] A. Duchowski, S. Jörg, A. Lawson, T. Bolte, L. Świrski, and K. Krejtz. Eye movement synthesis with 1/f pink noise. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, pages 47–56, 2015.

[11] A. T. Duchowski. *Eye tracking methodology: Theory and practice*. Springer, 2017.

[12] A. T. Duchowski. Gaze-based interaction: A 30 year retrospective. *Computers & Graphics*, 73:59–69, 2018.

[13] A. T. Duchowski, S. Jörg, T. N. Allen, I. Giannopoulos, and K. Krejtz. Eye movement synthesis. In *Proceedings of the ninth biennial ACM symposium on eye tracking research & applications*, pages 147–154, 2016.

[14] W. Fuhl, Y. Rong, and E. Kasneci. Fully convolutional neural networks for raw eye tracking data segmentation, generation, and reconstruction. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 142–149. IEEE, 2021.

[15] G. Funke, E. Greenlee, M. Carter, A. Dukes, R. Brown, and L. Menke. Which eye tracker is right for your research? performance evaluation of several cost variant eye trackers. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, volume 60, pages 1240–1244. SAGE Publications Sage CA: Los Angeles, CA, 2016.

[16] H. Griffith, D. Lohr, E. Abdulin, and O. Komogortsev. Gazebase, a large-scale, multi-stimulus, longitudinal eye movement dataset. *Scientific Data*, 8(1):184, 2021.

[17] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[18] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[19] C. D. Holland and O. V. Komogortsev. Biometric verification via complex eye movements: The effects of environment and stimulus. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 39–46. IEEE, 2012.

[20] Z. Hu, A. Bulling, S. Li, and G. Wang. Ehtask: Recognizing user tasks from eye and head movements in immersive virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(4):1992–2004, 2023.

[21] C. Jiao, Z. Hu, M. Bâce, and A. Bulling. Supreyes: Super resolutin for eyes using implicit neural representation learning. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–13, 2023.

[22] C. Jiao, Z. Hu, and A. Bulling. Hagi++: Head-assisted gaze imputation and generation. *arXiv preprint arXiv:2511.02468*, 2025.

[23] C. Jiao, Z. Hu, and A. Bulling. Hagi: Head-assisted gaze imputation for mobile eye trackers. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, pages 1–14, 2025.

[24] C. Jiao, Y. Wang, G. Zhang, M. Bâce, Z. Hu, and A. Bulling. Diffgaze: A diffusion model for modelling fine-grained human gaze behaviour on 360 images. *ACM Transactions on Interactive Intelligent Systems*, 16(1):1–23, 2026.

[25] G. Kim, T. Kwon, and J. C. Ye. Diffusionclip: Text-guided diffusion models for robust image manipulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2426–2435, 2022.

[26] Y. Kim, J. Lee, J.-H. Kim, J.-W. Ha, and J.-Y. Zhu. Dense text-to-image generation with attention modulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7701–7711, 2023.

[27] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. *arXiv preprint arXiv:2009.09761*, 2020.

[28] R. Kothari, Z. Yang, C. Kanan, R. Bailey, J. B. Pelz, and G. J. Diaz. Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. *Scientific reports*, 10(1):2539, 2020.

[29] K. Krejtz, A. Duchowski, H. Zhou, S. Jörg, and A. Niedzielska. Perceptual evaluation of synthetic gaze jitter. *Computer Animation and Virtual Worlds*, 29(6):e1745, 2018.

[30] G. Lan, T. Scargill, and M. Gorlatova. Eyesyn: Psychology-inspired eye movement synthesis for gaze-based activity recognition. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 233–246. IEEE, 2022.

[31] B. H. Le, X. Ma, and Z. Deng. Live speech driven head-and-eye motion generators. *IEEE transactions on visualization and computer graphics*, 18(11):1902–1914, 2012.

[32] S. P. Lee, J. B. Badler, and N. I. Badler. Eyes alive. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 637–644, 2002.

[33] A. Leube and K. Rifai. Sampling rate influences saccade detection in mobile eye tracking of a reading task. *Journal of Eye Movement Research*, 10(3), 2017.

[34] C. Li, J. Xue, C. Quan, J. Yue, and C. Zhang. Biometric recognition via texture features of eye movement trajectories in a visual searching task. *PloS one*, 13(4):e0194475, 2018.

[35] M. Li, Y. Duan, J. Zhou, and J. Lu. Diffusion-sdf: Text-to-shape via voxelized diffusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12642–12651, 2023.

[36] D. Lohr, H. Griffith, and O. V. Komogortsev. Eye know you: Metric learning for end-to-end biometric authentication using eye movements from a longitudinal dataset. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(2):276–288, 2022.

[37] D. Lohr and O. V. Komogortsev. Eye know you too: Toward viable end-to-end eye movement biometrics for user authentication. *IEEE Transactions on Information Forensics and Security*, 17:3151–3164, 2022.

[38] D. Lohr, M. J. Proulx, and O. Komogortsev. Establishing a baseline for gaze-driven authentication performance in vr: A breadth-first investigation on a very large dataset. In *2024 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2024.

[39] X. Ma and Z. Deng. Natural eye motion synthesis by modeling gaze-head coupling. In *2009 IEEE Virtual Reality Conference*, pages 143–150. IEEE, 2009.

[40] S. Makowski, L. A. Jäger, P. Prasse, and T. Scheffer. Biometric identification and presentation-attack detection using micro-and macro-movements of the eyes. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–10. IEEE, 2020.

[41] S. Makowski, P. Prasse, L. A. Jäger, and T. Scheffer. Oculomotoric biometric identification under the influence of alcohol and fatigue. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9. IEEE, 2022.

[42] S. Makowski, P. Prasse, D. R. Reich, D. Krakowczyk, L. A. Jäger,

and T. Scheffer. Deepeyedentificationlive: Oculomotoric biometric identification and presentation-attack detection using deep neural networks. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(4):506–518, 2021.

[43] C. Manning and H. Schutze. *Foundations of statistical natural language processing*. MIT press, 1999.

[44] S. Masuko and J. Hoshino. Head-eye animation corresponding to a conversation for cg characters. In *Computer Graphics Forum*, volume 26, pages 303–312. Wiley Online Library, 2007.

[45] T. Pejsa, B. Mutlu, and M. Gleicher. Stylized and performative gaze for character animation. In *Computer Graphics Forum*, volume 32, pages 143–152. Wiley Online Library, 2013.

[46] P. Prasse, D. R. Reich, S. Makowski, S. Ahn, T. Scheffer, and L. A. Jäger. Sp-eyegan: Generating synthetic eye movement data with generative adversarial networks. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, pages 1–9, 2023.

[47] P. Prasse, D. R. Reich, S. Makowski, T. Scheffer, and L. A. Jäger. Improving cognitive-state analysis from eye gaze with synthetic eye-movement data. *Computers & Graphics*, 119:103901, 2024.

[48] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models, 2021.

[49] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. In *Computer graphics forum*, volume 34, pages 299–326. Wiley Online Library, 2015.

[50] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78, 2000.

[51] D. Shi, Y. Wang, Y. Bai, A. Bulling, and A. Oulasvirta. Chartist: Task-driven eye movement control for chart reading. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1–14, 2025.

[52] L. Świrski and N. Dodgson. Rendering synthetic ground truth images for eye tracker evaluation. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 219–222, 2014.

[53] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[54] Y. Tashiro, J. Song, Y. Song, and S. Ermon. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. In *Advances in Neural Information Processing Systems*, 2021.

[55] Y. Wang, M. Bâce, and A. Bulling. Scanpath prediction on information visualisations. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, pages 1–15, 2023.

[56] B. Yang, S. Gu, B. Zhang, T. Zhang, X. Chen, X. Sun, D. Chen, and F. Wen. Paint by example: Exemplar-based image editing with diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18381–18391, 2023.

[57] S. H. Yeo, M. Lesmana, D. R. Neog, and D. K. Pai. Eyecatch: Simulating visuomotor coordination for object interception. *ACM Transactions on Graphics (TOG)*, 31(4):1–10, 2012.