SUPREYES: SUPer Resolution for EYES Using Implicit Neural Representation Learning

Chuhan Jiao
University of Stuttgart
Stuttgart, Germany
chuhan.jiao@vis.uni-stuttgart.de

Zhiming Hu*
University of Stuttgart
Stuttgart, Germany
zhiming.hu@vis.uni-stuttgart.de

Mihai Bâce
University of Stuttgart
Stuttgart, Germany
mihai.bace@vis.uni-stuttgart.de

Andreas Bulling
University of Stuttgart
Stuttgart, Germany
andreas.bulling@vis.uni-stuttgart.de

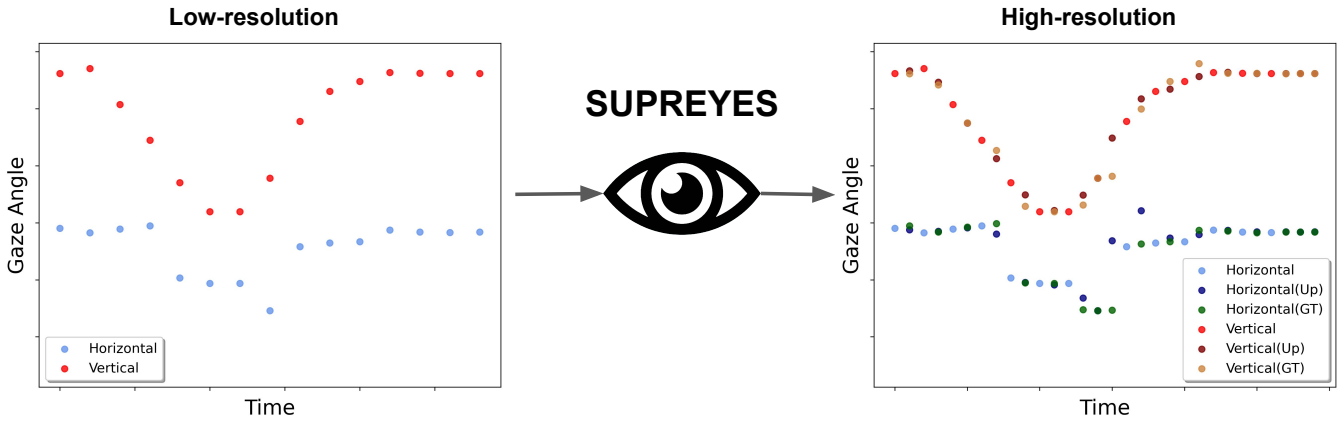

Figure 1: SUPREYES is a novel method for up-sampling gaze data recorded using a low(er)-resolution eye tracker, e.g. at 50 Hz (left), to a multiple times higher spatio-temporal resolution, e.g. 100 Hz (right). Up-sampling is performed in a self-supervised fashion, i.e. without the need for manual annotation, on the raw horizontal (blue) and vertical (red) gaze angles while preserving key characteristics of the signal for downstream tasks. The up-sampled gaze data (Up) and the corresponding ground truth (GT) are illustrated in the right figure.

ABSTRACT

We introduce SUPREYES – a novel self-supervised method to increase the spatio-temporal resolution of gaze data recorded using low(er)-resolution eye trackers. Despite continuing advances in eye tracking technology, the vast majority of current eye trackers – particularly mobile ones and those integrated into mobile devices – suffer from low-resolution gaze data, thus fundamentally limiting their practical usefulness. SUPREYES learns a continuous implicit neural representation from low-resolution gaze data to up-sample the gaze data to arbitrary resolutions. We compare our method with commonly used interpolation methods on arbitrary scale super-resolution and demonstrate that SUPREYES outperforms these baselines by a significant margin. We also test on the sample downstream task of gaze-based user identification and show that our method improves the performance of original low-resolution gaze data and outperforms other baselines. These results are promising as they open up a new direction for increasing eye tracking fidelity as well as enabling new gaze-based applications without the need for new eye tracking equipment.

*Corresponding author

UIST '23, October 29-November 1, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0132-0/23/10…$15.00
https://doi.org/10.1145/3586183.3606780

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; **Machine learning**; • **Human-centered computing**;

KEYWORDS

Gaze Data Super-resolution, Implicit Neural Representation, Up-sampling

ACM Reference Format:
Chuhan Jiao, Zhiming Hu, Mihai Bâce, and Andreas Bulling. 2023. SUPREYES: SUPer Resolution for EYES Using Implicit Neural Representation Learning. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), October 29-November 1, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3586183.3606780

1 INTRODUCTION

Arguably the most important technical property that mobile and stationary eye trackers have been continuously improved on, particularly in recent years, is their spatio-temporal resolution [2, 68]. Spatial resolution refers to the tracker's gaze estimation accuracy, typically measured in pixels of an on-screen 2D gaze location or in degrees of visual angle of a 3D unit gaze vector. Temporal resolution represents the number of gaze samples an eye tracker can record per second (sampling rate). While early stationary eye trackers (except for those for highly specialised applications, e.g. ophthalmologic examinations [7]) only offered sampling rates of a few hundred Hertz, latest systems achieve sampling rates of multiple Kilohertz [2, 68]. While advances have also been achieved for mobile trackers they still lack behind in terms of sampling rate [42]. In parallel, eye tracking accuracy has also continuously improved and is by now well below one degree of visual angle for both stationary and mobile eye trackers [77].

Recent studies in eye tracking research demonstrate that high-resolution gaze data is the key to the success of numerous applications, such as fast eye movement detection [12, 26, 49], gaze-based interaction [2, 18], or gaze-based user identification [33]. These findings suggest the need to upgrade existing low-resolution eye trackers to benefit from such applications. However, hardware upgrades are not only expensive given the high purchase cost of these systems but also time-consuming and cumbersome: First of all, the prices of eye trackers typically scale with hardware capabilities and may range from hundreds dollars for a low-resolution device to hundreds of thousands of dollars for a high-resolution hardware [26, 28]. In addition, research assistants familiar and well-trained on an existing system have to learn to operate the new one and their experience or best practices potentially gathered over years, such as most suitable parameter settings, typically do not transfer to the new system. Furthermore, specifically for mobile eye trackers, the need for wearability and unobtrusive integration pose significant challenges to the use of high-resolution cameras and high-performance onboard processing, and thus the achievable spatio-temporal resolution [67, 70].

In this work, we propose a software solution to increase the spatio-temporal resolution of low(er)-resolution eye trackers (see Figure 1). Our method, SUPREYES, is motivated by advances in representation learning using neural architectures that dominate most computational areas in Computer Science, such as in computer vision or natural language processing [14–16, 31, 32, 71]. Most importantly, recent work on neural radiance fields [56, 59, 60, 75, 79] and image representation learning [3, 9, 20, 41, 64] has demonstrated that implicit neural representation learning is highly effective at identifying and capturing structure in data and that the resulting representations can be used to synthesise new data that matches the statistics and is close to indistinguishable visually from real data. The core idea of SUPREYES is to leverage the power of implicit neural representation learning to up-sample gaze data to arbitrary spatio-temporal resolutions. SUPREYES parameterises human eye gaze as a continuous function with a multilayer perceptron (MLP) that maps a time t and the local features around the time t to the gaze direction expressed as degrees of visual angle. In contrast to previous works in the image domain that rely on pre-trained image feature extractors [9, 34], we design a novel gaze feature extractor for SUPREYES, as no pre-trained feature extractor exists for gaze data. Additionally, by adding an extra loss function for the feature extractor in SUPREYES training, we significantly improve the performance of our method. We compare our method with commonly used interpolation methods on the task of arbitrary scale gaze data super-resolution (e.g. up-sampling 50 Hz gaze data to 100 Hz, 200 Hz, 500 Hz, and 1000 Hz) and demonstrated that SUPREYES outperforms these baselines by a significant margin. We then show how gaze super-resolution can benefit downstream tasks: Using gaze-based user identification as an example, we show that the up-sampled gaze data produced by our method performs significantly better in terms of identification accuracy than the up-sampled data from other methods and the original low-resolution data, validating that our method is not only better than other methods in real applications but can also generate gaze samples that are in line with characteristics of gaze behaviour that are specific to an individual. The specific contributions of our work are three-fold:

(1) We propose SUPREYES – the first method for gaze data super-resolution that learns a continuous implicit neural representation from low-resolution gaze data to up-sample the gaze data to arbitrary resolutions.
(2) Through extensive experiments on arbitrary scale gaze data super-resolution, we demonstrate that our method is effective at capturing the internal structure of the gaze data and significantly outperforms commonly used interpolation baselines.
(3) We demonstrate that SUPREYES can benefit gaze-based downstream tasks. Using gaze-based user identification as an example, we show that our method improves the performance of original low- resolution gaze data and outperforms other baselines. This implies SUPREYES can generate gaze samples that align with user-specific characteristics.

2 RELATED WORK

Our work is related to 1) previous works that demonstrate the importance of high-resolution gaze data and works that propose high-resolution eye tracking solutions, 2) computational methods for super-resolution in time-series signals, and 3) recent advances in implicit neural representation.

2.1 Importance and Solutions of High-resolution Eye Tracking

Human eyes can move very quickly in real life, sometimes reaching up to a peak speed of $700^\circ/s$ [19]. In this context, high-resolution eye tracking is fundamental to the recording of realistic eye movements and is the key to the success of numerous eye gaze-based research and applications. More specifically, high-resolution eye trackers have demonstrated superior performance on detecting fast eye movements than low-resolution ones [12, 26, 49], and especially some subtle eye movements like microsaccades can only

be detected by high-resolution eye trackers because they usually last for an extreme short period of time, e.g. 25 ms [45]. In addition, in real-time applications like gaze-based interaction [18, 63], high-resolution eye tracking is the key to reduce the latency of the systems [2]. Furthermore, for biometric purposes such as user identification, high-resolution eye gaze data (e.g. 250 Hz) has also demonstrated significant better performance than low-resolution data (e.g. 30 Hz) [33].

To obtain high-resolution eye gaze data, commercial eye trackers like EyeLink 1000 Plus[1] used high-speed cameras to track human eyes, making the system much more expensive and power hungry than low-resolution ones [2, 26, 28]. Some researchers used dynamic vision sensors to achieve high-resolution eye tracking by adaptively sampling when the eye moves [2, 68]. However, this solution also puts requirement on special-purpose hardware, making it difficult to generalise to commonly used eye tracking systems which generally only use ordinary cameras. In contrast, our method produces high-resolution eye gaze data directly from low-resolution ones without using any additional equipment, which means our method can be easily applied to existing commonly used low-resolution eye tracking systems to improve their sampling frequency. Existing low-resolution eye tracking datasets [36–39] can also benefit from SUPREYES by applying it to improve their resolutions.

2.2 Super-resolution in Time-series Signals

Computational methods for super-resolution in time-series signals have been investigated by many researchers in the past few decades. Chen et al. focused on wind-induced pressure time series signals and proposed to use artificial neural networks to capture the complex variations of the pressure time series for data interpolation [8]. Liu et al. concentrated on turbulent flow data and proposed two convolutional neural networks (CNNs) to reconstruct high-resolution signals from low-resolution coarse flow data [51]. Su et al. studied the ocean's subsurface dynamic parameters and used a convolutional neural network and a light gradient boosting machine to reconstruct high-resolution dynamic parameters from low-resolution ones [69]. In addition to above time-series signals, audio signal has also been extensively studied and many deep learning-based methods have been proposed to reconstruct high-resolution audio signals, including convolutional neural networks [47, 50, 76], generative adversarial networks [21, 35], or diffusion-based methods [30, 48]. More recently, a more promising avenue for representation learning of any continuous signals is implicit neural representation learning.

2.3 Implicit Neural Representation

Implicit neural representation (INR) is a novel approach that, in contrast to prior works that discretise the input into, e.g. pixels or voxels, learns a continuous representation of the input signal. By leveraging the expressive power of neural networks to model such functions, INR can handle arbitrary topologies and resolutions and, as such, is well-suited for a number of applications including super-resolution. For example, Local Implicit representation for Super resolution of Arbitrary scale (LISA) [43] is an approach that can reconstruct the high-level components of an audio signal from low-level components by using information from neighbouring chunks. A similar approach are Local Deep Implicit Functions (LDIF) [27], which take one or more depth images as input to learn a 3D shape representation from multiple DIF functions with overlapping images. Local Implicit Image Function (LIIF) [9] is another approach that uses local information that, given a 2D coordinate in the image and 2D local deep features, it can predict the RGB value for any coordinate at arbitrary resolutions. INR methods have been proposed for photorealistic view synthesis from images [56] or videos [10, 17, 78]. Furthermore, INRs have also been used in the time-series domain. Jeong and Shin [40] used INR for detecting anomalies in multivariate time-series data. Franceschi et al. [23] proposed a causal dilated convolutions based encoder to obtain general-purpose representations for variable length and multivariate time series. Fons et al. [22] analysed implicit neural representations of time-series data and proposed HyperTime, an INR based hypernetwork for time-series generation.

While implicit neural representations have shown promising results on several domains, to the best of our knowledge, we propose the first approach for eye gaze data super-resolution. Our method draws inspiration from LIIF [9] and incorporates local gaze information to learn a continuous representation of eye tracking data. Unlike LIIF, which utilises a pretrained global feature extractor, we develop our own global feature extractor as there is currently no pretrained model for gaze data. In addition, we add a self-supervised reconstruction objective to the MLP output that enables the global gaze feature extractor to represent low-resolution gaze data better.

3 IMPLICIT NEURAL REPRESENTATION LEARNING OF GAZE DATA

Mobile and stationary eye trackers commonly output continuous (time-stamped) gaze data in the form of 2D x and y (on-screen gaze position) or 3D x, y, and z (gaze vector) coordinates. While on-screen gaze data can directly be associated with the visual stimulus, e.g. buttons of a graphical user interface, they depend on the physical properties of the recording setup, such as the screen size and resolution as well as the distance between user and display. 2D gaze data is, as such, not directly comparable across recording setups. 3D gaze vectors, in contrast, are independent of these factors and, if needed, they can be mapped to 2D on-screen gaze pixel coordinates. Given these advantages in terms of generalisability, SUPREYES uses 3D gaze data in terms of degrees of visual angle (dva) as input. The degrees of visual angle contain the angle of x axis and the angle of y axis.

In our method, each input sequence of eye tracking data is represented as global features $\mathbf{F}$. The key idea of our method is to use an MLP network to approximate a continuous gaze representation function for each input gaze data using its global features. The continuous representation can not only reconstruct the input gaze data but has the ability to up-sample the input gaze data to arbitrary resolutions. The approximation takes the form:

$$h_\theta : (\mathbf{F}, t) \rightarrow (x, y) \tag{1}$$

where θ are the parameters of MLP h, t is a time coordinate in the continuous time domain of the input gaze data, and (x, y) is the

[1] https://www.sr-research.com/eyelink-1000-plus/

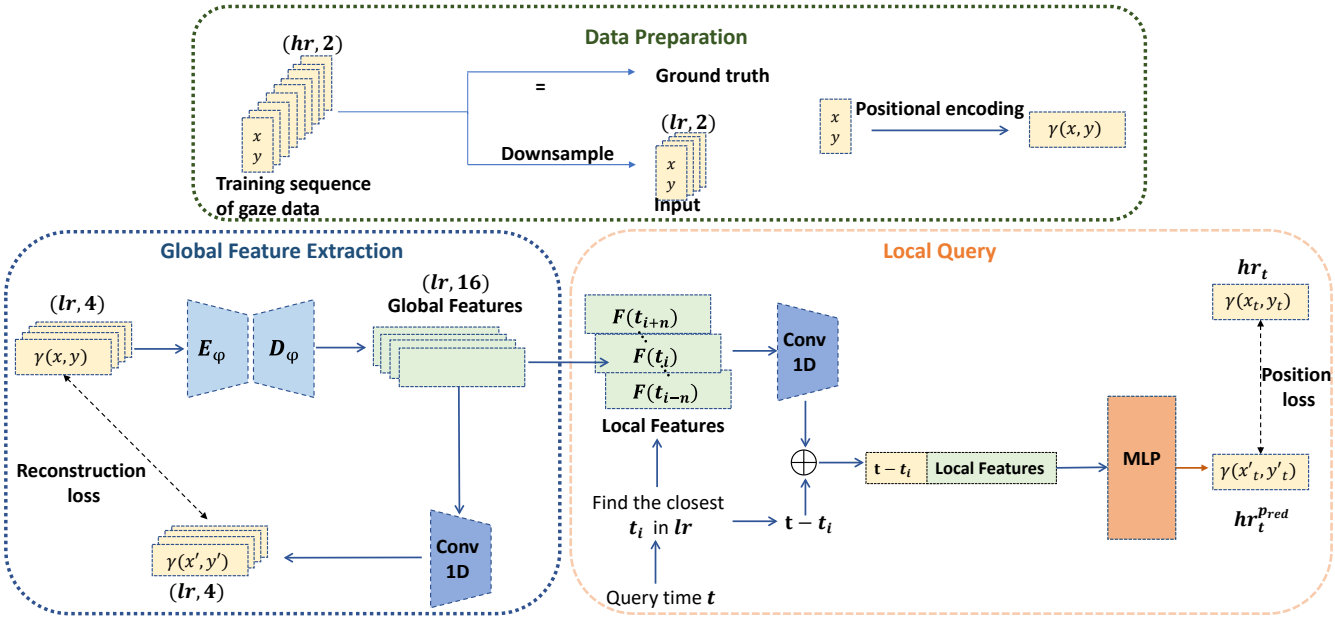

Figure 2: SUPREYES learns continuous representations from low-resolution eye tracking data. In data preparation, the high-resolution eye tracking data is first down-sampled into low-resolution inputs. We then apply positional encoding to convert degrees of visual angle into sine and cosine values for both high-resolution and low-resolution eye movements. Given a sequence of positional encoded low-resolution eye movements, we employ an encoder-decoder architecture, denoted as E_φ and D_φ, to extract its global features. Given a query time t, we find the closest time t_i in the input sequence and use the features around t_i to form the local features. The local features and the relative time coordinate $t - t_i$ are concatenated and input to a MLP to query the eye position at the time t. In training, global feature extraction and local query are jointly trained with reconstruction loss plus position loss.

gaze position at the time t represented as degrees of visual angle of x-axis and y-axis.

Assume a sequence of eye gaze data that corresponds to a long duration, to query the gaze position at time t using the global features may introduce redundant information because only the information from a certain period of time around the time t have an impact on the gaze position at the time. Therefore, we follow prior works [9, 43] to extract the local features $\mathbf{F}_{local}$ from the global features $\mathbf{F}$. We first find the closest time coordinate to t in $\mathbf{F}$, denoted as t_i. The local features $\mathbf{F}_{local}$ are formed by the features around t_i. We use $\mathbf{F}_{local}$ and the relative time coordinate $t - t_i$ for the query. Therefore, the new objective of the MLP is defined as:

$$h_\theta : (\mathbf{F}_{local}, t - t_i) \rightarrow (x, y) \tag{2}$$

Figure 2 provides an overview of our method. It consists of three components: data preparation, global feature extraction, and local query.

Data preparation. For each training high-resolution sequence of gaze data with shape $(hr, 2)$, the low-resolution input to our method with shape $(lr, 2)$ is generated by down-sampling the high-resolution sequence. hr and lr refer to the number of samples in high-resolution gaze data and low-resolution gaze data respectively. The high-resolution sequence is the ground truth that we use to supervise the MLP.

Since there are no clear bounds for degrees of visual angle, the model have problems converging with this kind of training data. To alleviate this issue, we apply a standard positional encoding function γ on each gaze position (x, y) in the training data:

$$\gamma(x, y) = [sin(x), cos(x), sin(y), cos(y)]^\intercal \tag{3}$$

After the encoding, all the training data are limited to the range of $[-1, 1]$ and the input data to the model and the ground truth are converted to the shapes of $(lr, 4)$ and $(hr, 4)$, respectively.

Global feature extraction. We use a fully 1D convolutional encoder-decoder architecture to obtain the global features $\mathbf{F}$ of each input low-resolution gaze data. The encoder consists of three 1D convolutional layers with the channel size of 16, 32, 64 and the kernel size of 3, 3, 1. The decoder consists of three transposed 1D convolutional layers to make the global features $\mathbf{F}$ have the same length as the low-resolution input. By having the same length, we are able to easily control how much local information we want to use in the local query. The final layer of the decoder has 16 channels. Therefore, the shape of the global features is $(lr, 16)$.

To make sure the low-resolution input is well represented by the global features $\mathbf{F}$, we add an additional 1D convolutional layer to reconstruct the low-resolution input from $\mathbf{F}$. The reconstruction loss $\mathcal{L}_{rec}$ is defined as the L2 distance between the reconstructed

sequence $\gamma(\mathbf{x}'_{lr}, \mathbf{y}'_{lr})$ and the ground truth sequence $\gamma(\mathbf{x}_{lr}, \mathbf{y}_{lr})$:

$$\mathcal{L}_{rec} = \left\|\gamma(\mathbf{x}'_{lr}, \mathbf{y}'_{lr}) - \gamma(\mathbf{x}_{lr}, \mathbf{y}_{lr})\right\|_2 \quad (4)$$

Local query. To query the gaze position at a given time coordinate t, we first find the closest time coordinate t_i in the input low-resolution sequence. Then, given a time window n, the local features $\mathbf{F}_{local}$ are formed by $2n+1$ closest features of time t_i:

$$\mathbf{F}_{local} = Concat([\mathbf{F}(t_{i-n}) \cdots \mathbf{F}(t_i) \cdots \mathbf{F}(t_{i+n})]) \quad (5)$$

The $\mathbf{F}_{local}$ undergoes a subsequent 1D convolutional layer and is flattened prior to concatenating with the local time coordinate $t - t_i$. As per Equation 2, we utilise a 5-layer MLP with hidden layers of dimensions 256, 256, 256, 256, 4. The first 4 layers are followed by ReLU activations and the last layer is followed by a tanh activation function. This MLP takes the aforementioned concatenation as input and predicts the gaze position $\gamma(x'_t, y'_t)$ at time t. We use the L2 loss to penalise the difference between the predicted gaze position and the ground truth gaze position:

$$\mathcal{L}_{pos} = \left\|\gamma(x'_t, y'_t) - \gamma(x_t, y_t)\right\|_2 \quad (6)$$

where $\gamma(x_t, y_t)$ is the ground truth from the high-resolution eye gaze data.

Objective function and optimisation. The time length of the training samples is consistent and is set to one second in our experiments. For every training sequence, we assume it represents the gaze data in time range $(-1, 1)$. Assume the high-resolution eye tracking data is captured with sampling frequency hr, a set of time coordinates for queries is generated, with the starting coordinate $t_0 = -1 + \frac{1}{2*hr}$ and other coordinates $t_i = t_0 + \frac{i}{hr}$. The time coordinates for the points inside the low-resolution input is down-sampled from the high-resolution time coordinates by using the same method in data preparation. Our proposed model is trained on the task of reconstructing every single point in high-resolution eye tracking data from the down-sampled low-resolution input and the corresponding time coordinates. Two parts of our method, the global feature extraction and the local query are jointly trained in an end-to-end fashion with a summed loss term

$$\mathcal{L} = \mathcal{L}_{rec} + \mathcal{L}_{pos} \quad (7)$$

We pick L2 loss instead of L1 loss for both loss terms because L2 loss is more sensitive to the outliers in the training data. In eye tracking data, the majority of points belong to fixations, only few belong to saccades. Using L2 loss helps our model better represent fast eye movements.

4 GAZE DATA SUPER-RESOLUTION

4.1 Experimental Setup

Dataset. While numerous publicly available eye tracking datasets exist, most were not collected using high-frequency eye trackers [58, 66] or do not provide a substantial amount of gaze data from various tasks although with high-frequency gaze data [1, 65, 80]. In order to provide high-frequency supervision to SUPREYES and allow it to learn a broad range of gaze data representations, we utilize the GazeBase dataset [29] for all of our evaluations. GazeBase is a large-scale, longitudinal eye movement dataset collected using an Eyelink 1000 eye tracker with a sampling frequency of 1,000 Hz. The dataset contains 12,334 eye tracking recordings from 322 individuals, covering a variety of eye movements performed over nine rounds of recordings spanning 37 months. Each round consists of two consecutive sessions, with each session comprising seven visual tasks, including fixations, horizontal and random saccades, reading, two video viewings, and gaze-driven gaming. Further details on these tasks can be found in the original paper on the GazeBase dataset [29]. Participants were exclusively recruited from those who had previously completed the previous round(s), providing the opportunity for participants to redo a round of study. We split the GazeBase dataset into training, validation, and test sets based on the number of participants in each round. Specifically, we randomly selected 80% of the participants in each round and used their recordings for training our model. The remaining 20% of participants were split into the validation and test sets, each containing 10% of the total participants. Our training, validation, and test set contained 9,814, 1,218, and 1,302 recordings, respectively.

When participants are blinking or their pupils cannot be detected, the eye tracker returns NaN as the current eye position. Instead of filling NaNs with a fixed value or interpolation methods [25, 52, 53], to prevent the effect of filling values on learning human eye movement representations, we cut each recording into segments by removing the NaNs. Then we keep the segments that are longer than 3 seconds. In total, there are 49,772 training segments, 6,139 validation segments, and 6,797 test segments.

Implementation details. We fixed the total time length of the input eye tracking data to our model to 1 second. Given an input low frequency and a target frequency, we first obtain the input low-resolution eye movements and ground truth high-resolution eye movements by downsampling the 1000 Hz segments in our dataset. Meanwhile, the query coordinates of the target frequency are downsampled from the time coordinates of 1000 Hz as well.

We train our model with Adam optimizer [44] with initial learning rate $1e^{-4}$ for 500 epochs with batch size 64. We apply learning rate decay every 200 epochs with decay factor 0.5. At each epoch, we repeat the segments in our training set 20 times, and each time we randomly crop a 1-second input sequence from each segment.

For validation and testing, we also crop the segments from the validation set and test set into 1-second short segments. For each segment, we start cropping from the beginning, the time interval between every two cropping starting points is 0.25 seconds. In total, there are 214,833 1-second segments in validation and 236,236 segments in testing.

We train three different models with input low frequencies of 25 Hz, 50 Hz, 100 Hz and target frequencies of 250 Hz, 250 Hz, 500 Hz. We choose these three input frequencies because the eye movements at these frequencies are easy to obtain from our 1000 Hz ground truth compared with 30 Hz or 60 Hz. Moreover, many previous eye tracking studies have used eye trackers with these sampling frequencies, such as the Dikablis mobile eye tracker (25 Hz) [6, 46, 72, 73, 81] and Tobii Pro Glasses (50 Hz or 100 Hz) [4, 57, 61]. We choose the target frequencies based on [33], where the authors suggest that 250 Hz or higher gaze data is better for biometric purposes. Besides, there are commercial eye trackers with these target frequencies, e.g. SMI RED eye tracker has 250 Hz and 500 Hz sampling rates. The time window n in the local query are

	Input	Metrics	Methods							
			linear	quadratic	cubic	nearest	previous	PCHIP	cubic spline	SUPREYES
Val	25 Hz	MAE↓	0.102	0.111	0.132	0.141	0.230	0.088	0.132	**0.079**
		MSE↓	0.185	0.222	0.451	0.463	1.196	0.158	0.451	**0.077**
		sDTW↓	39.186	73.955	188.472	138.570	138.568	44.780	188.513	**24.344**
	50 Hz	MAE↓	0.042	0.040	0.044	0.074	0.112	0.036	0.440	**0.034**
		MSE↓	0.039	0.037	0.051	0.128	0.317	0.033	0.052	**0.021**
		sDTW↓	10.628	11.971	19,358	45.452	45.452	10.326	19.466	**6.879**
	100 Hz	MAE↓	0.019	**0.017**	0.018	0.040	0.062	**0.017**	0.018	0.022
		MSE↓	0.014	0.014	0.016	0.042	0.098	0.013	0.016	**0.010**
		sDTW↓	7.975	7.968	9.911	28.629	28.629	7.474	9.938	**5.733**
Test	25 Hz	MAE↓	0.104	0.113	0.135	0.143	0.234	0.090	0.135	**0.079**
		MSE↓	0.187	0.223	0.459	0.481	1.254	0.156	0.459	**0.071**
		sDTW↓	38.214	74.075	191.989	143.446	143.446	43.378	192.129	**21.885**
	50 Hz	MAE↓	0.042	0.040	0.044	0.075	0.114	0.036	0.044	**0.034**
		MSE↓	0.035	0.031	0.044	0.127	0.327	0.027	0.044	**0.015**
		sDTW↓	9.153	10.166	16.446	46.279	46.279	8.756	16.557	**5.198**
	100 Hz	MAE↓	0.019	**0.017**	0.018	0.041	0.063	**0.017**	0.018	0.022
		MSE↓	0.011	0.011	0.012	0.039	0.097	0.010	0.012	**0.006**
		sDTW↓	5.830	5.584	6.602	27.924	27.924	5.439	6.697	**3.462**

Table 1: Quantitative comparison on the validation and test sets. The target resolution is 250 Hz for 25 Hz and 50 Hz inputs, and 500 Hz for 100 Hz inputs. The best results are given in bold, second-best results are underlined.

Input	Method	Target														
		50 Hz			100 Hz			200 Hz			500 Hz			1000 Hz		
		MAE↓	MSE↓	sDTW↓	MAE↓	MSE↓	sDTW↓	MAE↓	MSE↓	sDTW↓	MAE↓	MSE↓	sDTW↓	MAE↓	MSE↓	sDTW↓
25 Hz	linear	0.073	0.136	12.889	0.096	0.170	18.340	0.103	0.184	31.250	0.105	0.192	73.839	0.106	0.196	143.972
	quadratic	0.080	0.128	12.196	0.104	0.185	27.572	0.112	0.217	58.431	0.115	0.238	154.442	0.116	0.244	313.120
	PCHIP	0.063	0.099	9.544	0.082	0.136	18.733	0.089	0.153	34.983	**0.091**	0.162	86.112	**0.092**	0.165	170.340
	cubic spline	0.092	0.194	18.622	0.122	0.343	59.364	0.133	0.438	147.004	0.139	0.503	419.630	0.140	0.525	874.774
	SUPREYES	**0.057**	**0.050**	**4.312**	**0.079**	**0.065**	**9.376**	**0.087**	**0.072**	**17.526**	0.093	**0.077**	**42.045**	0.095	**0.079**	**83.080**
50 Hz	linear				0.031	0.027	4.960	0.041	0.034	7.844	0.044	0.037	15.694	0.044	0.038	27.314
	quadratic				0.030	0.022	3.783	0.039	0.030	8.231	0.042	0.034	19.551	0.043	0.035	36.248
	PCHIP				0.027	0.019	3.427	0.035	0.026	7.169	**0.038**	0.029	15.625	**0.038**	0.030	27.830
	cubic spline				0.032	0.026	4.511	0.042	0.041	12.648	0.046	0.051	35.945	0.047	0.054	73.293
	SUPREYES w.o. $\mathcal{L}_{rec}$				0.216	0.234	36	1.041	6.142	2151	0.990	6.077	5737	1.197	7.272	13843
	SUPREYES				**0.024**	**0.011**	**1.777**	**0.033**	**0.014**	**4.156**	0.041	**0.017**	**9.954**	0.045	**0.018**	**19.011**
100 Hz	linear							0.015	0.009	3.042	0.019	0.011	5.830	0.020	0.011	9.268
	quadratic							**0.013**	0.008	2.342	**0.017**	0.011	5.685	**0.018**	0.011	9.500
	PCHIP							**0.013**	0.008	2.269	**0.017**	0.010	5.349	**0.018**	0.010	8.897
	cubic spline							**0.013**	0.009	2.359	0.018	0.012	6.697	0.019	0.013	12.233
	SUPREYES							0.021	**0.005**	**1.400**	0.022	**0.006**	**3.462**	0.034	**0.008**	**8.335**

Table 2: Quantitative comparison on arbitrary scale super-resolution and ablation study on the test sets. The best results are given in bold, second-best results are underlined.

2, 4, 8 for 25 Hz, 50 Hz, 100 Hz inputs respectively to make every model have local features from the same time range.

Evaluation. To assess the similarity of our generated high-resolution eye tracking data to the human ground truth, we first evaluate SUPREYES from a time-series perspective. We apply three popular time-series metrics, Mean Absolute Errors (MAE), Mean Square Errors (MSE), and soft Dynamic Time Warping (sDTW) [11], to measure the performance of our model. These three metrics complement each other, with MAE measuring the overall numerical similarity, MSE focusing on penalizing the differences in fast eye movements, and sDTW measuring the similarity of shape between two time series. In this paper, we use sDTW divergence proposed in [5] with a smoothing factor 0.001:

$$sDTW_{div}(s', s) = sDTW(s', s) - \frac{1}{2}(sDTW(s, s) + sDTW(s', s')) \quad (8)$$

where s' is a predicted time series and s is the ground truth. The advantage of using sDTW divergence instead of sDTW or DTW is that the minimal value 0 is obtained when $s' = s$. We refer to sDTW divergence as sDTW throughout the paper for simplicity.

4.2 Comparison with Interpolation Methods

We follow the default of super-resolution tasks in other application domains that the super-resolution includes both interpolation and extrapolation [9, 43] (e.g. in our task 50hz to 500hz, the method

has to extrapolate the last 9 high-res data points.). We compare SUPREYES with commonly used interpolation methods that also have extrapolation ability, including linear, quadratic, cubic, nearest neighbour, Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) [24], cubic spline [13], and naive approaches that simply return the previous value of the point. All baselines are implemented using SciPy [74] built-in functions.

Because each SUPREYES model is trained on a certain target frequency, we first present our model evaluation results on its target frequencies in training in Table 1. Since interpolation methods do not need to be trained and validated, we also show the results on the validation set. Both in the validation set and the test set, our SUPREYES outperforms all interpolation baselines in all three time-series metrics for 25 Hz and 50 Hz inputs with a significant margin. For 100 Hz inputs, SUPREYES surpasses all interpolation methods in two out of three time-series metrics, except for MAE (The explanation for this is discussed in the second last paragraph of Section 4.2). Moreover, we observe that the nearest neighbour and previous interpolations perform much worse than other methods, and cubic and cubic spline have similar performance. Therefore, for further evaluations, we only keep linear, quadratic, PCHIP, and cubic spline interpolations as baselines.

SUPREYES aims to learn continuous gaze data representations from input gaze data which can be used to up-sample the input gaze data to arbitrary resolutions. To evaluate the quality of the learned continuous gaze data representations, we evaluate the ability of SUPREYES to perform arbitrary scale super-resolution and compare it with the two selected baseline methods. Specifically, the task of arbitrary scale super-resolution refers to the resolution of input data is fixed we up-sample the input data to multiple different resolutions and run evaluations. We choose target frequencies of 50 Hz, 100 Hz, 200 Hz, 500 Hz, and 1000 Hz, as these frequencies are easy to obtain with original 1000 Hz gaze data, and there are commercial eye trackers with these sampling frequencies, such as Tobbi Pro Glasses (50 Hz, 100 Hz), Pupil Core and VT3 mini remote eye tracker (200 Hz), SMI RED (500 Hz), and Eyelink 1000 (1000 Hz). We up-sampled the 25 Hz, 50 Hz, and 100 Hz gaze data to the higher target frequencies using our trained models and two baseline methods. Figure 1 presents a qualitative example of using our method to up-sample 50 Hz gaze data to 100 Hz, we can observe our up-sampled data is close to human ground truth. Table 2 shows the results of the arbitrary scale evaluation. We can see that SUPREYES achieves better performances on MSE and sDTW in all experiments. For MAE, SUPREYES performs better than baseline interpolation methods in most of the tasks with 25 Hz inputs and 50 Hz inputs, but interpolation methods perform better on other tasks.

The use of MAE alone as a metric for evaluating gaze data super-resolution techniques is limited as it only measures the numerical difference between two time-series. Interpolation methods, which interpolate points between given points in the inputs, can achieve low MAE values as the number of given points in the input increases. However, this does not necessarily indicate whether the up-sampled gaze data is similar to the human gaze data from a time-series perspective. Therefore, the use of additional metrics such as MSE and sDTW is necessary to evaluate the similarity of fast eye movements and overall shape between the up-sampled gaze data and the human gaze data.

In general, SUPREYES achieves higher similarity to human gaze data than all interpolation baselines from a time-series perspective in the super-resolution task when considering the combination of these metrics. However, it should be noted that time-series metrics alone may not be sufficient for evaluating gaze data enhancement techniques. In particular, whether the up-sampled gaze data is in line with characteristics of gaze behaviour of the subject is important in gaze data super-resolution. To further evaluate SUPREYES and other baseline methods, we also perform evaluations on a gaze-based user identification task (details in Section 5).

4.3 Ablation Study

Most works in implicit neural representation learning utilise pre-trained neural networks to extract global features or latent codes or any other useful information from inputs, which are then used as inputs for MLPs [9, 34, 59]. However, for areas such as audio processing where pretrained models may not be available, researchers design their own extractors and jointly trained them with the final output loss of MLPs [43]. We argue that extractors trained only with the loss of the final outputs may not be sufficient to extract useful information from inputs. To confirm our hypothesis, we train an 50 Hz SUPREYES model without the reconstruction loss $\mathcal{L}_{rec}$ with 250 Hz target frequency. We evaluate this model on the arbitrary scale super-resolution task described in Section 4.2. The results, shown in Table 2, indicate that the SUPREYES model without $\mathcal{L}_{rec}$ performs extremely poorly on all three time-series metrics, suggesting that without $\mathcal{L}_{rec}$, the global feature extractor fails to extract robust global features from the inputs.

5 GAZE-BASED USER IDENTIFICATION

Maintaining users' identity in gaze data super-resolution is crucial for enhancing mobile eye trackers, as high-resolution gaze data generated by a method should retain the same user identity as their low frequency input during practical eye-tracking data collection. However, there is no explicit way to evaluate this. Alternatively, we evaluate our method on the task of gaze-based user identification. Our assumption is that if the up-sampled gaze data achieves better performance than the original low-resolution gaze data in user identification, it implies that the up-sampling method generated gaze samples that align with the specific characteristics of an individual's gaze behaviour.

5.1 Experimental Setup

Problem Definition and Data preparation. We merge the validation and test sets described in Section 4.1 to form the dataset for closed-set user identification. The merged dataset comprises eye-tracking data from 130 subjects. Our task involves predicting which user a given gaze data input belongs to, among the entire pool of 130 users. Each subject completed two consecutive sessions, comprising seven tasks in total, as part of the GazeBase dataset [29]. However, we exclude data from the fixation task since in this task participants were asked to fixate on a static target at the screen centre and the results of the previous work show that the model trained with the gaze data from the fixation task only has the worst user authentication performance [52]. Specifically, we utilise data from session one (S1) for training and session two (S2)

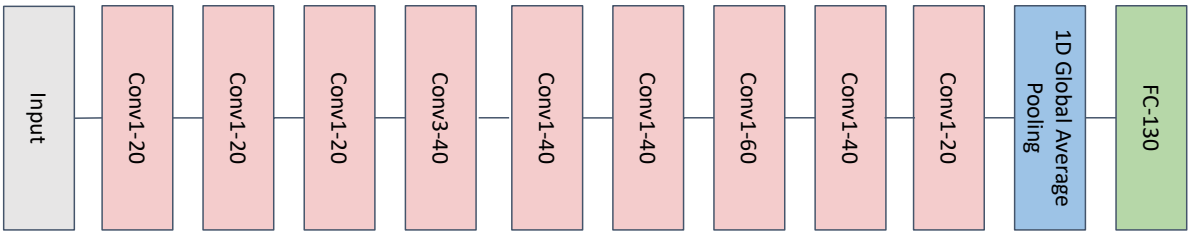

Figure 3: CNN-based model for user identification. The numbers inside CNN blocks reflect the output channels of CNN layers, the number inside the fully-connected layer is the size of the hidden layer. Each CNN layer is followed with ReLU activation. The first 6 CNN layers are with kernel size 3 and padding 1, and the last 3 CNN layers are with kernel size 3,1,1 and no padding. The output of the final fully-connected layer is the probabilities of the input eye movement data belonging to the 130 users.

for validation purposes. And we downsample the eye tracking data to different frequencies, including 25 Hz, 50 Hz, 100 Hz, and 200 Hz, and then use SUPREYES and different interpolation methods to up-sample 50 Hz and 100 Hz data to higher frequencies (the reason of choosing these frequencies is presented in Section 5.3 and the details of up-sampling are presented in Section 5.2). We keep the 25 Hz data here because we believe the user identification performance on the original 25 Hz data can reveal the lower bound of the performance of 50 Hz up-sampled data. Following prior work on user authentication [53], we set the duration of the input gaze data to 5 seconds. To this end, each eye tracking data is cut into non-overlapping 5-second segments starting from the beginning of the data. NaNs in these segments are filled with the previous valid value, and if there is no previous valid value, they are filled with the next valid value. Overall, we obtain 15,544 training segments and 14,947 validation segments for the dataset at different frequencies.

User Identification Method. CNN-based methods have shown great achievements in user authentication/identification in recent years [52, 53, 55]. In our experiment, we employed the 9-layer CNN architecture proposed in [52]. The original model proposed in [52] has too many parameters, making the network easy to get overfitting. Therefore, we reduced the number of channels for each CNN layer in our implementation. We also changed the flatten layer in [52] into a 1D global average pooling layer to ensure fair comparisons between the eye tracking data at different resolutions. Since the models for inputs at different resolutions have the same number of learnable parameters, the results change only because of the different input frequencies. The model architecture of our implementation is shown in Figure 3. The model takes the positional encoded (see Equation 3) 5-second eye tracking segments as input and outputs the probabilities of the input eye movement data belonging to the 130 users.

Implementation details. To train the CNN model on eye tracking data at different frequencies, we employ cross-entropy loss and Adam optimizer [44] with an initial learning rate of $1e^{-3}$. We trained the models for 500 epochs using a batch size of 64. Since the models possess the same number of learnable parameters, the low-resolution input model converges faster than the high-resolution input model. To avoid overfitting in the low-resolution input model, we apply early stopping during training. Specifically, if the performance on the validation set, does not improve for ten consecutive epochs we stop the training.

Evaluation. We evaluate all models trained with the up-sampled eye movements on corresponding ground-truth high-resolution validation sets with two evaluation metrics - equal error rate (EER) and the classification accuracy of user identification. EER has been widely used in evaluating biometric systems [52–55, 62], it refers to a point on a receiver operating characteristic (ROC) curve where the false acceptance rate (FAR) is equal to the false rejection rate (FRR). The lower EER indicates the better performance of a biometric system.

5.2 Up-sampling Strategy

Given the superior performance in MAE of the interpolation methods over SUPREYES in the super-resolution task of 100 Hz inputs (as discussed in Section 4.2), we utilise these four interpolation methods as baselines for user identification. We up-sample the 50 Hz and 100 Hz eye tracking data in our training set to (100 Hz, 200 Hz) and 200 Hz, respectively, using both the baseline methods and our proposed method.

The baseline methods are implemented as previously described in Section 4.2. We directly apply the baseline methods to the degrees of visual angle in the low-resolution training sets to generate the up-sampled high-resolution training sets. For SUPREYES, we first apply positional encoding to the degrees of visual angle and then perform the up-sampling. We explore three methods for up-sampling eye tracking data using our proposed method: single-stage up-sampling, fine-tuning for frequent users, and multi-stage up-sampling.

Single-stage up-sampling. Single-stage up-sampling refers to directly applying SUPREYES to up-sample the low-resolution eye movements to target resolution. This method is denoted as SUPREYES (ss) for simplicity.

Fine-tuning for frequent users. Fine-tuning for frequent users is an upgrade to SUPREYES (ss). Since SUPREYES is trained on the eye tracking data from many subjects, in up-sampling, it performs eye movements like an average or aggregated person of these subjects. However, in practical use cases, it is often necessary to fine-tune the model to better approximate the eye movements of a specific user. Assume we have a frequent user of the mobile eye tracker that is going to be enhanced. To make our model act more

like this user, we can fine-tune SUPREYES on previous gaze recordings of this user. In our experiment, we fine-tune SUPREYES for each user using their own low-resolution data in our training set for 10 epochs, and up-sample the low-resolution data to target resolution using the fine-tuned model. We refer to this approach as SUPREYES (ss+ft).

Multi-stage up-sampling. The up-sampling of gaze data can also be done in a multi-stage manner, which we refer to as SUPREYES (ms). This is motivated by the observation that as the target resolution increases, the similarities of up-sampled eye movements towards the ground truth human eye movements from a time series perspective drop significantly, as shown in Table 2. However, we found that in the case of up-sampling scale ×2, SUPREYES is able to perform similarly to a real human. In our experiment, we use SUPREYES (ms) to up-sample the 50 Hz eye tracking data to 200 Hz. Specifically, we first apply the 50 Hz SUPREYES model to up-sample the data to 100 Hz, and then use the 100 Hz SUPREYES model to up-sample the data to the final resolution of 200 Hz. Note that the SUPREYES (ms) can be a combination of SUPREYES (ss) and SUPREYES (ss+ft). We use SUPREYES (ss+ft) plus SUPREYES (ss) for up-sampling 50 Hz to 200 Hz in our experiment.

Implementation details of up-sampling with SUPREYES. SUPREYES takes 1-second gaze data as input, we need to preprocess the 5-second low-frequency gaze data before passing it to SUPREYES. One naive approach is to cut the 5-second data into 5 1-second segments. However, in this way, the contextual information in the 5-second data is not sufficiently used, which may lead to lower fidelity. To better use the contextual information, we cut the data into 1-second segments that overlap with each other. Specifically, we start cutting the input from the very beginning, and the time interval between each cutting starting point is 0.8 seconds. If the last 1-second data is not in the current obtained segments, we also include the last 1-second data in the segments. These segments are then upsampled to higher frequencies with SUPREYES. After obtaining the up-sampled 1-second data, we merge each adjacent up-sampled 1-second data by replacing the data of the last 0.1 second in the first segment with the data from 0.1 second to 0.2 second in the second segment, except for the last 1-second segment. For the last segment, we not only just replace the last 0.1 seconds data in its previous segments with the corresponding part in the augmented last segment, but also concatenate the rest part in the augmented last segment after that. This merging process effectively reduces the discontinuity caused by segmenting the data.

5.3 Results

While previous work suggests that biometric purposes require sampling frequencies of 250 Hz or higher [33], our experimentation reveals that using 250 Hz inputs do not significantly improve performance compared to 200 Hz data in our experimental settings. As such, we opt to utilise 200 Hz as the maximum eye tracker sampling frequency in our study. Conversely, 25 Hz inputs lack sufficient information to accurately identify users. Our results, as shown in Table 3, indicate that 25 Hz data perform poorly. Therefore, we limit our up-sampling to 50 Hz and 100 Hz eye movements in our experiment

The user identification results on original frequencies are used as a criterion to evaluate whether an up-sampling method maintains user identity in the up-sampling process. After up-sampling, if the user identification results on the high testing frequency are superior to those on the original input frequency, we assume that the up-sampling method generates meaningful user-specific gaze data to some extent.

Table 3 presents the quantitative results of user identification on our validation set for three different up-sampling tasks: 50 Hz to 100 Hz, 50 Hz to 200 Hz, and 100 Hz to 200 Hz. We compare the performance of our proposed method SUPREYES with four baseline interpolation methods, linear and quadratic, PCHIP, and cubic spline interpolation.

We observe that for all three up-sampling tasks, all baseline methods lead to an increase in equal error rate (EER) compared to the original input frequency. Additionally, only one baseline (cubic spline) gets a tiny improvement in classification accuracy in the task 100 Hz to 200 Hz. This suggests that these interpolation methods fail to approximate user-specific gaze behaviour during super-resolution. However, for the up-sampling scale of ×2 (50 Hz to 100 Hz and 100 Hz to 200 Hz), SUPREYES (ss) leads to improvements in both evaluation metrics compared to the original 50 Hz and 100 Hz results. Furthermore, by incorporating fine-tuning, SUPREYES (ss+ft) achieves even better performance, with improvements of 19% in EER and 15% in classification accuracy when up-sampling 50 Hz gaze data to 100 Hz. Similarly, when up-sampling 100 Hz data to 200 Hz, SUPREYES (ss+ft) improves EER by 17% and classification accuracy by 7%. Notably, the results for the task of up-sampling 100 Hz to 200 Hz rival those of the original 200 Hz eye tracking data. These findings implicitly demonstrate the effectiveness of our proposed method in approximating the unique gaze behaviour traits of each individual during super-resolution of gaze data with up-sampling scale ×2.

When the up-sampling scale goes higher to ×4 (50 Hz to 200 Hz), the task becomes more challenging. SUPREYES (ss) fails to improve the performance in up-sampling 50 Hz data to 200 Hz, and the performance of SUPREYES (ss+ft) deteriorates even further. Our proposed SUPREYES (ms) is used to address this challenge. With the help of multi-stage up-sampling, SUPREYES (ms) slightly improves the results, but still falls far short of the performance achieved on the original high-frequency data. Considering these results, our proposed SUPREYES (ms) approach represents an important step forward in addressing the difficulties of super-resolution tasks at high up-sampling scales.

6 DISCUSSION

6.1 Factors that Affect Model Performance

Although SUPREYES outperforms existing interpolation baselines and helps increase user identification results by preserving users' identity throughout the up-sampling process. However, as shown in both Table 2 and Table 3, the performance of our method is highly correlated to the input frequency and the target frequency. Specifically, when the input frequency is fixed, the performance of SUPREYES decreases as the target frequency increases. This is due to the fact that SUPREYES is trained on eye-tracking data from many different subjects, it learns the average gaze data representations of

Original input frequency	Testing frequency	Upsampling method	EER(%)↓	Classification accurarcy(%)↓
25 Hz	25 Hz	-	16.17	12.55
50 Hz	50 Hz	-	3.13	35.56
100 Hz	100 Hz	-	1.46	55.78
200 Hz	200 Hz	-	1.20	61.93
50 Hz	100 Hz	linear	7.78	16.93
50 Hz	100 Hz	quadratic	3.97	31.14
50 Hz	100 Hz	PCHIP	4.11	29.60
50 Hz	100 Hz	cubic spline	4.62	27.72
50 Hz	100 Hz	SUPREYES (ss)	2.77	38.30
50 Hz	100 Hz	SUPREYES (ss+ft)	**2.52**	**40.89**
50 Hz	200 Hz	linear	9.87	16.56
50 Hz	200 Hz	quadratic	4.65	26.96
50 Hz	200 Hz	PCHIP	4.75	28.78
50 Hz	200 Hz	cubic spline	5.74	25.14
50 Hz	200 Hz	SUPREYES (ss)	4.68	31.14
50 Hz	200 Hz	SUPREYES (ss+ft)	5.72	26.98
50 Hz	200 Hz	SUPREYES (ms)	**3.12**	**36.03**
100 Hz	200 Hz	linear	2.64	44.10
100 Hz	200 Hz	quadratic	1.58	50.36
100 Hz	200 Hz	PCHIP	1.59	52.01
100 Hz	200 Hz	cubic spline	1.52	55.88
100 Hz	200 Hz	SUPREYES (ss)	1.27	57.91
100 Hz	200 Hz	SUPREYES (ss+ft)	**1.21**	**59.67**

Table 3: The quantitative results of user identification on the validation set. The user identification results on the high testing frequency which outperform the results on the original input frequency are underlined. The best results of up-sampling methods of each (original input frequency, test frequency) pairs are bolded.

these subjects. When the up-sampling scale is high, the number of points that need to be generated by the model exceeds the number of points in the given input. Consequently, our method interpolates these points by utilising the average knowledge learned from the subjects in the training set, resulting in a drop in performance in generating realistic gaze data of the input user. Additionally, we find that the higher the input frequency, the better the performance in eye-tracking fidelity. With up-sampling scale ×2, up-sampling 100 Hz gaze data with SUPREYES to 200 Hz rivals the original 200 Hz data in user identification. However, the results of the 50 Hz to 100 Hz gaze data are still far from reaching the performance of the original 100 Hz gaze data. This is because 100 Hz gaze data contains more user-specific information than 50 Hz data, which helps SUPREYES maintain user consistency during the up-sampling process. Overall, SUPREYES works well with the up-sampling scale ×2 for low-resolution inputs.

6.2 Applications of SUPREYES

Enhancing low-resolution eye trackers. In Sections 4 and 5, SUPREYES has shown promising results in generating human-like gaze data and preserving identity consistency during super-resolution, particularly in the upsampling scale of ×2. As a result, it can be utilized as a post-processing technique to enhance low-resolution eye trackers when higher frequency information is required.

Augment existing eye tracking datasets. SUPREYES can not only enhances low-frequency eye-trackers but also enables the augmentation of existing eye-tracking datasets. With the development of technologies, we believe there will be huge improvements in increasing the sampling frequency of mobile eye trackers. However, the datasets collected by these low-resolution mobile eye trackers cannot upgrade simultaneously. Re-collecting these datasets requires considerable time and effort. Our method offers a convenient solution by enabling the easy augmentation of these datasets to higher sampling frequencies without additional effort. Consider the scenario of a gaze-based user authentication task, where a user authentication model for a VR helmet is trained on gaze data collected by its built-in 100 Hz eye tracker. Suppose the user wants to upgrade to a new VR helmet equipped with a 200 Hz eye tracker while still using a gaze-based authentication model. One naive solution is to use the new device for a long time and collect enough new 200 Hz data to train a model. Restricting the capacity of the new eye tracker to 100 Hz is a suboptimal solution, the previous model can directly adapted, but this defeats the purpose of upgrading to a new VR helmet. Our proposed method, SUPREYES, is highly beneficial in such cases, as it allows for the direct up-sampling of the 100 Hz gaze data captured by the previous device to 200 Hz with almost no loss in identity information. This enables users to seamlessly apply a new 200 Hz user authentication model without having to collect new data. Additionally, during the usage, the 200 Hz data collected by the new device can be merged into the up-sampled dataset, further improving the authentication performance.

6.3 Limitations and Future work

Simulating real low-resolution eye tracker through down-sampling. To facilitate the training of our model using low-resolution gaze data and corresponding high-resolution ground truth, we down-sampled the high-resolution gaze data to simulate the output of a real low-resolution eye tracker. By mimicking the behaviour of a low-resolution camera, our simulated data closely resembles the characteristics of real low-resolution gaze data within an ideal data collection setting. However, it is important to note that practical usage of low-resolution eye trackers introduces additional factors that can impact data quality, such as motion blur, varying lighting conditions, hardware noise, and other sources of noise inherent in the eye tracker itself. Our simulation only accounts for the ideal scenario. How these factors affect our model's performance remains to be explored.

Running additional evaluations. In addition to the time-series evaluation that we showed in Section 4, we plan to conduct further evaluations from an eye movement perspective in furture work. This includes analysing time-series errors for different eye movement types (saccades, smooth pursuits) to identify areas for improvement. We also aim to assess whether the generated eye movements exhibit similar attributes, such as the distribution of eye movement types and velocity profiles, as observed in real human data. Furthermore, we will conduct user identification evaluation on real low-resolution gaze datasets. These evaluations will provide valuable insights to improve the performance of our approach.

Improving the performance of SUPREYES. In future work, we plan to improve SUPREYES in up-sampling low-resolution gaze data with higher up-sampling scales. The multi-stage up-sampling method SUPREYES (ms) proposed in Section 5.2 successfully increases the performance of single-stage up-sampling method SUPREYES (ss) in up-sampling scale ×4. This indicates that the multi-scale information of eye tracking data helps with large-scale up-sampling. However, currently, we train separate models for the inputs at different scales, e.g. we up-sample 50 Hz data to 200 Hz by up-sampling 50 Hz to 100 Hz first and using 100 Hz model the continue up-sampling to 200 Hz. This is not ideal when the up-sampling scale goes higher, since we have to train many models with different input frequencies. However, our fully convolutional encoder-decoder used for extracting global features of low-frequency inputs theoretically enables us to develop a unified model for inputs of different frequencies merging the multi-scale gaze representations. Furthermore, we plan to investigate how SUPREYES can help with other downstream tasks that can benefit from high-resolution gaze data, such as learning-based eye movement detection. Additionally, we aim to make SUPREYES run in real-time, as this is crucial for enhancing low-resolution eye trackers during data collection. Currently, our method can be seen as a post-processing method and cannot be integrated with eye trackers to enhance the sampling-frequencies in real-time.

7 CONCLUSION

This paper presented SUPREYES, the first learning-based approach for up-sampling low-resolution gaze data to arbitrary target resolutions. We conducted extensive experiments to evaluate the performance of SUPREYES in arbitrary scale super-resolution and gaze-based user identification, demonstrating its superiority over commonly used interpolation baselines. In particular, SUPREYES can generate gaze samples that are in line with user-specific characteristics during the up-sampling process and can provide high-frequency information that benefits gaze-based applications. We believe that SUPREYES has enormous potential to enhance low-resolution eye trackers and existing eye tracking datasets without requiring any additional hardware.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon Europe research and innovation funding programme under grant agreement No. 101072410. Mihai Bâce was funded by a Swiss National Science Foundation (SNSF) Postdoc.Mobility Fellowship (grant number 214434). Zhiming Hu was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 – 390740016.

Funded by
the European Union

REFERENCES

[1] Richard Andersson, Linnea Larsson, Kenneth Holmqvist, Martin Stridh, and Marcus Nyström. 2017. One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior research methods* 49 (2017), 616–637.

[2] Anastasios N Angelopoulos, Julien NP Martel, Amit P Kohli, Jorg Conradt, and Gordon Wetzstein. 2021. Event-Based Near-Eye Gaze Tracking Beyond 10,000 Hz. *IEEE Transactions on Visualization & Computer Graphics* 27, 05 (2021), 2577–2586.

[3] Mojtaba Bemana, Karol Myszkowski, Hans-Peter Seidel, and Tobias Ritschel. 2020. X-Fields: Implicit Neural View-, Light- and Time-Image Interpolation. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2020)* 39, 6 (2020). https://doi.org/10.1145/3414685.3417827

[4] Jeroen S Benjamins, Roy S Hessels, and Ignace TC Hooge. 2018. GazeCode: Open-source software for manual mapping of mobile eye-tracking data. In *Proceedings of the 2018 ACM symposium on eye tracking research & applications*. 1–4.

[5] Mathieu Blondel, Arthur Mensch, and Jean-Philippe Vert. 2021. Differentiable divergences between time series. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 3853–3861.

[6] Christian Braunagel, Enkelejda Kasneci, Wolfgang Stolzmann, and Wolfgang Rosenstiel. 2015. Driver-activity recognition in the context of conditionally autonomous driving. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, 1652–1657.

[7] Melinda Y Chang, Nandini Gandhi, and Mary O'Hara. 2019. Ophthalmologic disorders and risk factors in children with autism spectrum disorder. *Journal of American Association for Pediatric Ophthalmology and Strabismus* 23, 6 (2019), 337–e1.

[8] Y Chen, GA Kopp, and D Surry. 2002. Interpolation of wind-induced pressure time series with an artificial neural network. *Journal of Wind Engineering and Industrial Aerodynamics* 90, 6 (2002), 589–615.

[9] Yinbo Chen, Sifei Liu, and Xiaolong Wang. 2021. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8628–8638.

[10] Zeyuan Chen, Yinbo Chen, Jingwen Liu, Xingqian Xu, Vidit Goel, Zhangyang Wang, Humphrey Shi, and Xiaolong Wang. 2022. VideoINR: Learning Video Implicit Neural Representation for Continuous Space-Time Super-Resolution. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), 2037–2047.

[11] Marco Cuturi and Mathieu Blondel. 2017. Soft-dtw: a differentiable loss function for time-series. In *International conference on machine learning*. PMLR, 894–903.

[12] Edwin Dalmaijer. 2014. *Is the low-cost EyeTribe eye tracker any good for research?* Technical Report. PeerJ PrePrints.

[13] Carl De Boor and Carl De Boor. 1978. *A practical guide to splines*. Vol. 27. springer-verlag New York.

[14] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

[17] Michail Christos Doukas, Stylianos Ploumpis, and Stefanos Zafeiriou. 2023. Dynamic Neural Portraits. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 4073–4083.

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

[19] Andrew T Duchowski and Andrew T Duchowski. 2017. *Eye tracking methodology: Theory and practice*. Springer.

[20] Emilien Dupont, Adam Goliński, Milad Alizadeh, Yee Whye Teh, and Arnaud Doucet. 2021. Coin: Compression with implicit neural representations. *arXiv preprint arXiv:2103.03123* (2021).

[21] Sefik Emre Eskimez and Kazuhito Koishida. 2019. Speech super resolution generative adversarial network. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3717–3721.

[22] Elizabeth Fons, Alejandro Sztrajman, Yousef El-Laham, Alexandros Iosifidis, and Svitlana Vyetrenko. 2022. HyperTime: Implicit Neural Representations for Time Series. In *NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research*. https://openreview.net/forum?id=DZ2FaoMhWRb

[23] Jean-Yves Franceschi, Aymeric Dieuleveut, and Martin Jaggi. 2019. Unsupervised scalable representation learning for multivariate time series. *Advances in neural information processing systems* 32 (2019).

[24] Frederick N Fritsch and Judy Butland. 1984. A method for constructing local monotone piecewise cubic interpolants. *SIAM journal on scientific and statistical computing* 5, 2 (1984), 300–304.

[25] Wolfgang Fuhl, Yao Rong, and Enkelejda Kasneci. 2021. Fully Convolutional Neural Networks for Raw Eye Tracking Data Segmentation, Generation, and Reconstruction. In *2020 25th International Conference on Pattern Recognition (ICPR)*. 142–149. https://doi.org/10.1109/ICPR48806.2021.9413268

[26] Gregory Funke, Eric Greenlee, Martha Carter, Allen Dukes, Rebecca Brown, and Lauren Menke. 2016. 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*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 1240–1244.

[27] Kyle Genova, Forrester Cole, Avneesh Sud, Aaron Sarna, and Thomas A. Funkhouser. 2019. Local Deep Implicit Functions for 3D Shape. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 4856–4865.

[28] Lisa Graham, Julia Das, Jason Moore, Alan Godfrey, and Samuel Stuart. 2022. The Eyes as a Window to the Brain and Mind. In *Eye Tracking: Background, Methods, and Applications*. Springer, 1–14.

[29] Henry Griffith, Dillon Lohr, Evgeny Abdulin, and Oleg Komogortsev. 2021. GazeBase, a large-scale, multi-stimulus, longitudinal eye movement dataset. *Scientific Data* 8, 1 (2021), 184.

[30] Seungu Han and Junhyeok Lee. 2022. NU-Wave 2: A general neural audio upsampling model for various sampling rates. *arXiv preprint arXiv:2206.08545* (2022).

[31] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. 2022. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16000–16009.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[33] Corey D Holland and Oleg V Komogortsev. 2012. 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)*. IEEE, 39–46.

[34] Yang Hong, Bo Peng, Haiyao Xiao, Ligang Liu, and Juyong Zhang. 2022. HeadNeRF: A Real-time NeRF-based Parametric Head Model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[35] Shichao Hu, Bin Zhang, Beici Liang, Ethan Zhao, and Simon Lui. 2020. Phase-aware music super-resolution using generative adversarial networks. *arXiv preprint arXiv:2010.04506* (2020).

[36] Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. 2021. FixationNet: Forecasting Eye Fixations in Task-Oriented Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 27, 5 (2021), 2681–2690. https://doi.org/10.1109/TVCG.2021.3067779

[37] Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. 2022. EHTask: recognizing user tasks from eye and head movements in immersive virtual reality. *IEEE Transactions on Visualization and Computer Graphics* (2022).

[38] Zhiming Hu, Sheng Li, Congyi Zhang, Kangrui Yi, Guoping Wang, and Dinesh Manocha. 2020. DGaze: CNN-based gaze prediction in dynamic scenes. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 1902–1911.

[39] Zhiming Hu, Congyi Zhang, Sheng Li, Guoping Wang, and Dinesh Manocha. 2019. SGaze: a data-driven eye-head coordination model for realtime gaze prediction. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (2019), 2002–2010.

[40] Kyeong-Joong Jeong and Yong-Min Shin. 2022. Time-series anomaly detection with implicit neural representation. *arXiv preprint arXiv:2201.11950* (2022).

[41] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2021. Alias-Free Generative Adversarial Networks. In *Proc. NeurIPS*.

[42] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Adj. Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. 1151–1160. https://doi.org/10.1145/2638728.2641695

[43] Jaechang Kim, Yunjoo Lee, Seunghoon Hong, and Jungseul Ok. 2022. Learning Continuous Representation of Audio for Arbitrary Scale Super Resolution. In *ICASSP*.

[44] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[45] Rakshit Kothari, Zhizhuo Yang, Christopher Kanan, Reynold Bailey, Jeff B Pelz, and Gabriel J Diaz. 2020. Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. *Scientific Reports* 10, 1 (2020), 2539.

[46] Thomas C Kübler, Colleen Rothe, Ulrich Schiefer, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2017. SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior research methods* 49 (2017), 1048–1064.

[47] Volodymyr Kuleshov, S Zayd Enam, and Stefano Ermon. 2017. Audio super resolution using neural networks. *arXiv preprint arXiv:1708.00853* (2017).

[48] Junhyeok Lee and Seungu Han. 2021. Nu-wave: A diffusion probabilistic model for neural audio upsampling. *arXiv preprint arXiv:2104.02321* (2021).

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

[50] Teck Yian Lim, Raymond A Yeh, Yijia Xu, Minh N Do, and Mark Hasegawa-Johnson. 2018. Time-frequency networks for audio super-resolution. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 646–650.

[51] Bo Liu, Jiupeng Tang, Haibo Huang, and Xi-Yun Lu. 2020. Deep learning methods for super-resolution reconstruction of turbulent flows. *Physics of Fluids* 32, 2 (2020), 025105.

[52] Dillon Lohr, Henry Griffith, and Oleg V. Komogortsev. 2022. 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 (2022), 276–288. https://doi.org/10.1109/TBIOM.2022.3167633

[53] Dillon Lohr and Oleg V. Komogortsev. 2022. Eye Know You Too: Toward Viable End-to-End Eye Movement Biometrics for User Authentication. *IEEE Transactions on Information Forensics and Security* 17 (2022), 3151–3164. https://doi.org/10.1109/TIFS.2022.3201369

[54] Dillon J Lohr, Samantha Aziz, and Oleg Komogortsev. 2020. Eye movement biometrics using a new dataset collected in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications*. 1–3.

[55] Silvia Makowski, Paul Prasse, David R. Reich, Daniel Krakowczyk, Lena A. Jäger, and Tobias Scheffer. 2021. DeepEyedentificationLive: Oculomotoric Biometric Identification and Presentation-Attack Detection Using Deep Neural Networks. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 3, 4 (2021), 506–518. https://doi.org/10.1109/TBIOM.2021.3116875

[56] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.

[57] Diederick C Niehorster, Thiago Santini, Roy S Hessels, Ignace TC Hooge, Enkelejda Kasneci, and Marcus Nyström. 2020. The impact of slippage on the data quality of head-worn eye trackers. *Behavior research methods* 52 (2020), 1140–1160.

[58] Seonwook Park, Emre Aksan, Xucong Zhang, and Otmar Hilliges. 2020. Towards End-to-end Video-based Eye-Tracking. In *European Conference on Computer Vision (ECCV)*.

[59] Sida Peng, Yuanqing Zhang, Yinghao Xu, Qianqian Wang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. 2021. Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9054–9063.

[60] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. 2020. D-NeRF: Neural Radiance Fields for Dynamic Scenes. *https://arxiv.org/abs/2011.13961* (2020).

[61] Shane L Rogers, Craig P Speelman, Oliver Guidetti, and Melissa Longmuir. 2018. Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific reports* 8, 1 (2018), 1–9.

[62] Ronald Salloum and C-C Jay Kuo. 2017. ECG-based biometrics using recurrent neural networks. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2062–2066.

[63] Lei Shi, Cosmin Copot, and Steve Vanlanduit. 2021. Gazeemd: Detecting visual intention in gaze-based human-robot interaction. *Robotics* 10, 2 (2021), 68.

[64] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. 2020. Implicit Neural Representations with Periodic Activation Functions. In *Proc. NeurIPS*.

[65] Ekta Sood, Fabian Kögel, Florian Strohm, Prajit Dhar, and Andreas Bulling. 2021. VQA-MHUG: A gaze dataset to study multimodal neural attention in visual question answering. *arXiv preprint arXiv:2109.13116* (2021).

[66] Julian Steil and Andreas Bulling. 2015. Discovery of Everyday Human Activities From Long-term Visual Behaviour Using Topic Models. In *Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. 75–85. https://doi.org/10.1145/2750858.2807520

[67] Julian Steil, Marc Tonsen, Yusuke Sugano, and Andreas Bulling. 2019. InvisibleEye: Fully Embedded Mobile Eye Tracking Using Appearance-Based Gaze Estimation. *GetMobile: Mobile Computing and Communications* 23, 2 (2019), 30–34.

[68] Timo Stoffregen, Hossein Daraei, Clare Robinson, and Alexander Fix. 2022. Event-based kilohertz eye tracking using coded differential lighting. In *Proceedings of the 2022 IEEE Winter Conference on Applications of Computer Vision*. 2515–2523.

[69] Hua Su, An Wang, Tianyi Zhang, Tian Qin, Xiaoping Du, and Xiao-Hai Yan. 2021. Super-resolution of subsurface temperature field from remote sensing observations based on machine learning. *International Journal of Applied Earth Observation and Geoinformation* 102 (2021), 102440.

[70] Marc Tonsen, Julian Steil, Yusuke Sugano, and Andreas Bulling. 2017. Invisibleeye: Mobile eye tracking using multiple low-resolution cameras and learning-based gaze estimation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21.

[71] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[72] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2011. Analysing EOG signal features for the discrimination of eye movements with wearable devices. In *Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction*. 15–20.

[73] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2012. Detection of smooth pursuits using eye movement shape features. In *Proceedings of the symposium on eye tracking research and applications*. 177–180.

[74] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17 (2020), 261–272. https://doi.org/10.1038/s41592-019-0686-2

[75] Huan Wang, Jian Ren, Zeng Huang, Kyle Olszewski, Menglei Chai, Yun Fu, and Sergey Tulyakov. 2022. R2L: Distilling Neural Radiance Field to Neural Light Field for Efficient Novel View Synthesis. In *ECCV*.

[76] Heming Wang and DeLiang Wang. 2021. Towards robust speech super-resolution. *IEEE/ACM transactions on audio, speech, and language processing* 29 (2021), 2058–2066.

[77] Katarzyna Wisiecka, Krzysztof Krejtz, Izabela Krejtz, Damian Sromek, Adam Cellary, Beata Lewandowska, and Andrew Duchowski. 2022. Comparison of Webcam and Remote Eye Tracking. In *Proceedings of the 2022 Symposium on Eye Tracking Research and Applications*. 1–7.

[78] Wenqi Xian, Jia-Bin Huang, Johannes Kopf, and Changil Kim. 2020. Space-time Neural Irradiance Fields for Free-Viewpoint Video. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 9416–9426.

[79] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. 2020. pixelNeRF: Neural Radiance Fields from One or Few Images. *https://arxiv.org/abs/2012.02190* (2020).

[80] Ruohan Zhang, Calen Walshe, Zhuode Liu, Lin Guan, Karl Muller, Jake Whritner, Luxin Zhang, Mary Hayhoe, and Dana Ballard. 2020. Atari-head: Atari human eye-tracking and demonstration dataset. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 6811–6820.

[81] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2011. Discrimination of gaze directions using low-level eye image features. In *Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction*. 9–14.