SP-EyeGAN: Generating Synthetic Eye Movement Data with Generative Adversarial Networks

Prasse, Paul ; Reich, David Robert ; Makowski, Silvia ; Ahn, Seoyoung ; Scheffer, Tobias ; Jäger, Lena A

Abstract: Neural networks that process the raw eye-tracking signal can outperform traditional methods that operate on scanpaths preprocessed into fixations and saccades. However, the scarcity of such data poses a major challenge. We, therefore, present SP-EyeGAN, a neural network that generates synthetic raw eye-tracking data. SP-EyeGAN consists of Generative Adversarial Networks; it produces a sequence of gaze angles indistinguishable from human micro- and macro-movements. We demonstrate how the generated synthetic data can be used to pre-train a model using contrastive learning. This model is fine-tuned on labeled human data for the task of interest. We show that for the task of predicting reading comprehension from eye movements, this approach outperforms the previous state-of-the-art.

DOI: https://doi.org/10.1145/3588015.3588410

Originally published at:
Prasse, Paul; Reich, David Robert; Makowski, Silvia; Ahn, Seoyoung; Scheffer, Tobias; Jäger, Lena A (2023). SP-EyeGAN: Generating Synthetic Eye Movement Data with Generative Adversarial Networks. In: ETRA '23: 2023 Symposium on Eye Tracking Research and Applications, Tübingen, Germany, 30 May 2023 - 2 June 2023. ACM Digital library, 18.
DOI: https://doi.org/10.1145/3588015.3588410

SP-EyeGAN: Generating Synthetic Eye Movement Data with Generative Adversarial Networks

Paul Prasse*
David R. Reich*
paul.prasse@uni-potsdam.de
david.reich@uni-potsdam.de
University of Potsdam
Potsdam, Germany

Silvia Makowski
silvia.makowski@uni-potsdam.de
University of Potsdam
Potsdam, Germany

Seoyoung Ahn
seoyoung.ahn@stonybrook.edu
Stony Brook University
Stony Book, USA

Tobias Scheffer
tobias.scheffer@uni-potsdam.de
University of Potsdam
Potsdam, Germany

Lena A. Jäger
jaeger@cl.uzh.ch
University of Zurich
Zurich, Switzerland
University of Potsdam
Potsdam, Germany

ABSTRACT

Neural networks that process the raw eye-tracking signal can outperform traditional methods that operate on scanpaths preprocessed into fixations and saccades. However, the scarcity of such data poses a major challenge. We, therefore, present SP-EyeGAN, a neural network that generates synthetic raw eye-tracking data. SP-EyeGAN consists of Generative Adversarial Networks; it produces a sequence of gaze angles indistinguishable from human micro- and macro-movements. We demonstrate how the generated synthetic data can be used to pre-train a model using contrastive learning. This model is fine-tuned on labeled human data for the task of interest. We show that for the task of predicting reading comprehension from eye movements, this approach outperforms the previous state-of-the-art.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**; **Machine learning**; • **Computer systems organization** → **Neural networks**.

KEYWORDS

eye tracking, generative adversarial networks, scanpath, gaze generation, reading comprehension

ACM Reference Format:
Paul Prasse, David R. Reich, Silvia Makowski, Seoyoung Ahn, Tobias Scheffer, and Lena A. Jäger. 2023. SP-EyeGAN: Generating Synthetic Eye Movement Data with Generative Adversarial Networks. In *2023 Symposium on Eye Tracking Research and Applications (ETRA '23), May 30–June 02, 2023, Tubingen, Germany*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3588015.3588410

*Both authors contributed equally to this research.

ETRA '23, May 30–June 02, 2023, Tubingen, Germany

ACM ISBN 979-8-4007-0150-4/23/05.
https://doi.org/10.1145/3588015.3588410

1 INTRODUCTION

Eye tracking data has a wide range of applications, including the assessment of linguistic and cognitive skills [Reich et al. 2022], the detection of conditions such as dyslexia [Haller et al. 2022] or attention deficit hyperactivity disorder [Deng et al. 2022], and even identifying individuals based on their unique patterns of eye movements [Lohr and Komogortsev 2022; Makowski et al. 2021]. In the context of biometric identification, using the raw eye-tracking signal of yaw and pitch angles at the tracker's sampling rate as input to a deep neural network instead of preprocessed and possibly aggregated scanpaths of saccades and fixations has been shown to improve performance by an order of magnitude and enable the use of shorter input sequences [Jäger et al. 2020].

However, data scarcity is a major challenge for developing such neural networks; collecting eye-tracking data is costly in terms of labor and equipment. There is also a risk that personal information such as gender, identity, or ethnicity may be extracted from eye movements, creating a major privacy concern. These problems could potentially be mitigated by using synthetic instead of real-world data to train (or pre-train) machine-learning models. Existing approaches to generating synthetic eye-tracking data are limited in their ability to create realistic data; most known approaches only generate fixation positions and durations [Engbert et al. 2005; Kümmerer and Bethge 2021; Reichle et al. 2003] or use statistical models [Campbell et al. 2014; Duchowski et al. 2016; Fuhl and Kasneci 2018].

In computer vision, biometrics, and other fields, the development of generative adversarial networks (GANs) to generate synthetic data has shown promising results [Bowles et al. 2018]. In this paper, we develop *SP-EyeGAN* (a model to create **S**can **P**aths for **Eye**-tracking data using a **GAN**), a system consisting of two GANs that are capable of generating synthetic eye-tracking data that closely mimics real-world data, and that can be used to overcome

the challenges of data scarcity and privacy. We further provide a proof-of-concept study to demonstrate how the synthetic data generated by SP-EyeGAN can be used to pre-train any neural network architecture that operates on raw eye-tracking data for an arbitrary downstream task. To this end, we train two different neural network architectures that have led to the above mentioned advances in eye-tracking based biometrics. We employ these architectures and investigate whether pre-training them on synthetic data allows us to reach similar advances in other downstream tasks which to date have been proven quite challenging, namely general reading comprehension, text comprehension, text difficulty and nativeness of a reader.

The rest of the paper is structured as follows. In Section 2, we discuss related work. Afterwards, in Section 3, we describe SP-EyeGAN and the usability of SP-EyeGAN for contrastive pre-training. Section 4 details our experimental results, which are examined for limitations in Section 5. In Section 6, we provide a more extensive analysis and interpretation of these results, followed by a conclusion in Section 7.

2 RELATED WORK

Existing methods for generating human-like eye-tracking data can be divided into training-free statistical models and trained machine-learning models.

Statistical models. Lee et al. [2002] and Duchowski and Jörg [2015] presented statistical approaches that generate eye movements for rendered, animated faces. Ma and Deng [2009] have developed a method that synthesizes natural eye gaze, given a head-motion sequence as input, by statistically modeling the relationship between gaze and head movements. Le et al. [2012] generate realistic head motion, eye gaze, and eyelid motion simultaneously based on speech input. Wood et al. [2015] present a method that generates eye crops together with gaze vectors. Yeo et al. [2012] proposed a statistical model that generates an eye-tracking sequence of saccades and smooth pursuits for an agent catching a ball. All of these approaches aim at making rendered faces more realistic rather than creating realistic eye-tracking data that include micro- and macro-movements as well as a noise component. An approach of Campbell et al. [2014] creates realistic eye-tracking data based on a statistical model of jointly estimated dynamic properties of eye movements for a known saliency map of the stimulus. Duchowski et al. [2016, 2015] add micro-saccadic jitter, noise, simulated measurement error and pupil unrest to a previously generated eye-tracking sequence. Fuhl and Kasneci [2018] and Fuhl et al. [2018] simulate saccadic movements by gamma distributions and smooth pursuit onsets with the sigmoid function. EyeSyn [Lan et al. 2022] generates fixational movement using Gaussian and pink noise. These two statistical models are used as reference models in our evaluation.

Machine-learning models. Simon et al. [2016] employ a convolutional neural network (CNN) and long short-term memory (LSTM) modules to generate synthetic eye-tracking data; this model is limited to generating eye-tracking data for static images. Assens et al. [2018] proposed a GAN that consumes images as input and generates fixation points but is unable to model saccadic movements. Fuhl and Kasneci [2022] use a hierarchical k-means algorithm, *HPCGen*, that generates eye-tracking data. HPCGen generates random eye-tracking data points not following a specific stimulus with no constant sampling rate, which is not suitable to generate micro-movements and fixations. Fuhl et al. [2021] devised a variational autoencoder (VAE) that generates eye-tracking data, but not for a specific stimulus. We use this model as one of the baselines in our evaluation.

3 METHOD

This section introduces our proposed method *SP-EyeGAN* that generates synthetic eye movement data, and a contrastive pre-training framework to use the generated data to pre-train a neural embedding for eye movement sequences.

3.1 SP-EyeGAN

SP-EyeGAN is composed of two independent, structurally identical generative adversarial networks (GANs) [Goodfellow et al. 2014] for generating fixations (FixGAN) and saccades (SacGAN), respectively, and a module that assembles the generated fixations and saccades into a gaze sequence (see Figure 1, top right). SP-EyeGAN requires a sequence of fixation positions as input. The fixation positions depend on the stimulus and can either be sampled from a saliency map or distribution over word positions [Rayner and McConkie 1976] for text stimuli; or they can be obtained from a cognitive model that generates fixation positions on an image or video frame [Nuthmann et al. 2010] or on a textual stimulus [Engbert et al. 2005; Kümmerer and Bethge 2021; Reichle et al. 2003]. Both GANs use the same architecture shown in the red box in Figure 1 and use the x- and y-velocities of fixations and saccades, respectively, as input. The FixGAN simulates the small (micro-)movements during fixations, while the SacGAN simulates the fast movements within saccades. Both GANs consist of a generative model and a discriminative model. While the generator is used to create synthetic eye movements, the discriminator is trained to distinguish between real and synthetic data. Each GAN is trained by alternating the following steps: In the first step the generator creates some synthetic data. This data is used to train the discriminator using backpropagation with the cross entropy loss. The loss for the generated data points is then used in a backpropagation step to adjust the weights in the generator.

The generator creates a synthetic eye movement sequence by projecting a noise vector into a higher dimension using a fully connected layer followed by batch normalization and a LeakyReLU activation. This output is then reshaped to match the required sequence length (100 ms for fixations and 30 ms for saccades). The reshaping layer is followed by 3 deconvolutional blocks. Each block consists of a deconvolution (filter size f, kernel size k) followed by batch normalization and an optional LeakyReLU activation.

The discriminator consumes a sequence of eye movement data and decides whether a sequence originates from real recorded eye movements or was generated by one of the generators. It consists of three convolutional blocks. Each convolutional block consists of a convolution (filter size f, kernel size k) followed by batch normalization and a LeakyReLU activation. The output of the last convolutional block is flattened and is fed into a fully connected

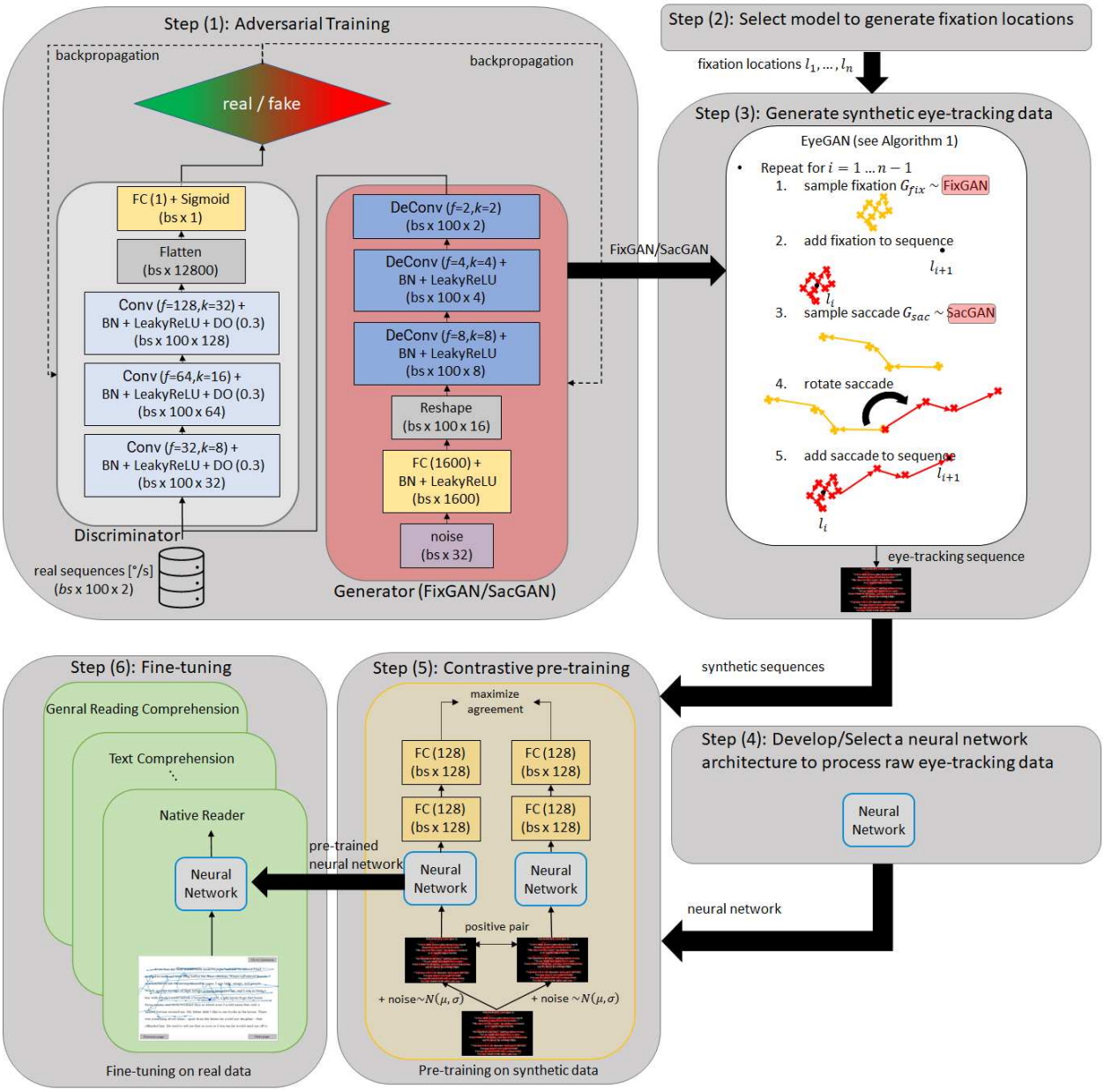


Figure 1: SP-EyeGAN overview. This figure shows the complete pipeline to train models to generate fixations and saccades (step 1) that are used to create synthetic data (step 3). Steps 5 and 6 depict the contrastive pre-training. The GANs from step 1 are trained with batch size bs and consist of fully connected layers (denoted as FC), batch normalization (denoted as BN), convolutional/deconvolutional layers (denoted as Conv/DeConv with filter size f, kernel size k and dilation d), and the leaky rectified linear unit (LeakyReLU) activation function. Numbers in brackets show the dimensions of the data after each layer.

layer, followed by a sigmoid activation to create the output (the distinction between real and generated sequences).

SP-EyeGAN, shown in Algorithm 1 generates a complete eye movement sequence S of fixations and saccades by sampling fixations and saccades using the trained FixGAN/SacGAN. The algorithm creates a synthetic eye movement sequence given the mean μ_{fix} and standard deviation σ_{fix} for fixation durations, the mean μ_{sac} and standard deviation σ_{sac} for saccade durations and n fixation locations $F = l_1 \ldots l_n$ as shown in Figure 1 (top right). Each sequence starts with a fixation on the first fixation location created using the FixGAN clipped to the sampled fixation duration d_{fix} (lines 4, 6). Each fixation is added to the generated sequence S after its generation (line 7). The amplitude of the preceding saccade at iteration i is determined by the distance between the two fixation

locations l_i and l_{i+1} (line 3). In the next step, the algorithm samples a saccade matching the amplitude and the saccade duration d_{sac} (lines 5, 8). This saccade can point to another direction and therefore has to be rotated accordingly (line 9) before being added to the sequence.

Algorithm 1 The SP-EyeGAN algorithm generates a synthetic eye-movement sequence for given fixation locations.

Require: $\mu_{fix}, \sigma_{fix}, \mu_{sac}, \sigma_{sac}$, FixGAN, SacGAN, fixation locations $F = l_1, \ldots, l_n$
Ensure: Synthetic eye movement sequence $S = s_1, \ldots, s_m$
1: $S = l_1$ ▷ start location is first fixation location
2: **for** $i \in [1 \ldots n-1]$ **do**
3: $a_{sac} = computeSaccadeAmplitude(l_i, l_{i+1})$ ▷ compute saccade amplitude for jump from l_i to l_{i+1}
4: $d_{fix} = \mathcal{N}(\mu_{fix}, \sigma_{fix})$ ▷ sample duration for next fixation
5: $d_{sac} = \mathcal{N}(\mu_{sac}, \sigma_{sac})$ ▷ sample duration for next saccade
6: $G_{fix} = generateFixation(\text{FixGAN}, d_{fix})$ ▷ generate fixation $[^\circ/s]$ with duration d_{fix}
7: $S = S + dva(G_{fix})$ ▷ add fixation converted to degrees of visual angle to sequence
8: $G_{sac} = generateSaccade(\text{SacGAN}, d_{sac}, a_{sac})$ ▷ generate saccade $[^\circ/s]$ with duration d_{sac} and amplitude a_{sac}
9: $G^{rot}_{sac} = rotateSaccade(G_{sac}, S[-1], l_{i+1})$ ▷ rotate generated saccade to end at new fixation location l_{i+1}
10: $S = S + dva(G^{rot}_{sac})$ ▷ convert rotated saccade to degrees of visual angle and add to eye movement sequence
11: **end for**

3.2 Contrastive Learning

We further evaluate the usefulness of synthetic data for pre-training. We employ SP-EyeGAN to generate raw eye movement sequences which serve as input for a neural network that is pre-trained using the self-supervised technique *contrastive learning* [Bautista and Naval 2020; Chen et al. 2020]. A major benefit of contrastive learning is its ability to be implemented without the need for labeled data, making it a suitable approach for learning representations from synthetic data. The goal of contrastive learning, as shown in Figure 1 (step 5), is to train the neural network to differentiate between positive and negative pairs of sequences. In our study, we define positive (i.e., similar) pairs as the same synthetic sequence augmented with Gaussian noise, while dissimilar pairs are composed of different sequences, also augmented with Gaussian noise. The two sequences that constitute a (positive or negative) pair are fed into the neural network, which computes a hidden representation whose dimension is then reduced using two bottleneck layers. The objective for the neural network during the contrastive-learning process is to maximize the agreement between the hidden representations in positive pairs and minimize the agreement in negative pairs. Following contrastive pre-training, the neural network can then be fine-tuned for a specific downstream task using a potentially small amount of real data containing the labels of interest, see Figure 1 (green boxes in step 6).

To summarize, our method is comprised of the following steps:

(1) Adversarial training of FixGAN and SacGAN using unlabeled human eye movement data;
(2) Selection of a model (e.g., a cognitive model) that generates fixation locations for a given stimulus;
(3) Generation of synthetic raw eye-tracking data using SP-EyeGAN together with the fixation location model;
(4) Development or selection of a neural network architecture suitable to process raw eye-tracking data for the downstream task at hand;
(5) Pre-training of the neural network on the synthetic data using contrastive learning;
(6) Fine-tuning of the neural network with (potentially small amounts of) labeled human data for the task.

Note that for training a given neural network on a new task, only the last step needs to be re-done; for training a novel network architecture for any task, only the last two steps need to be performed.

4 EXPERIMENTS

This section shows the experiments we conducted to evaluate our approach and reports on the results. All code to reproduce the results and create synthetic eye movement data can be found online[1].

4.1 Metrics

In our evaluation, we use the *Jensen-Shannon divergence* to measure the similarity between two probability distributions. It is used to evaluate the quality of generated eye movement data by comparing properties of generated fixations and saccades with the same properties of human eye movement data [Manning and Schütze 1999]. For discrete probability distributions P and Q defined on the same sample space $\mathcal{X}$ the Jensen-Shannon divergence is defined as $JSD(P||Q) = \frac{1}{2}KL(P||M) + \frac{1}{2}KL(Q||M)$, where $M = \frac{1}{2}(P+Q)$ and $KL(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log_2 \left(\frac{Q(x)}{P(x)}\right)$.

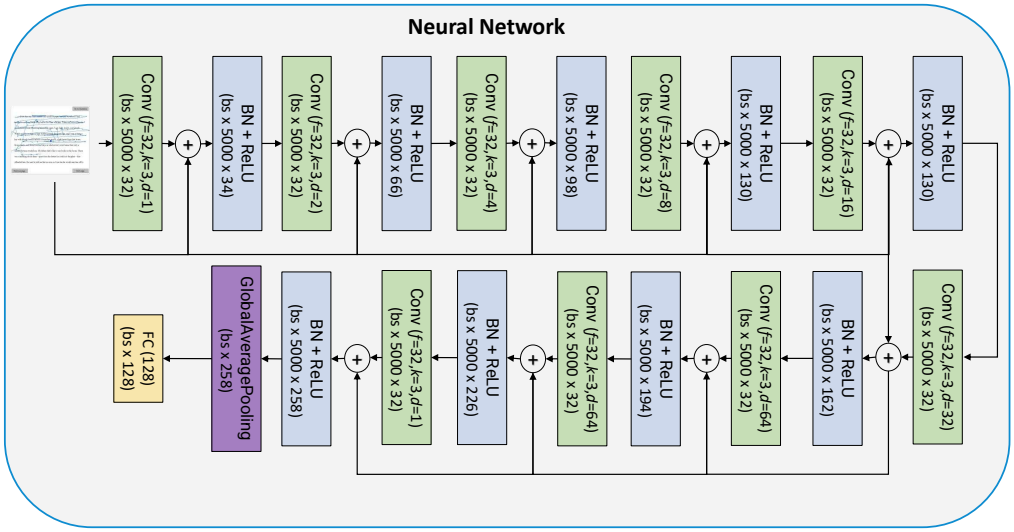
We evaluate the performance of models trained on a downstream task in terms of the area under the receiver operating characteristic curve (AUC), which is a quantitative indicator of classification performance. Independently of the class ratios, the AUC ranges from 0.5 for random guessing to 1 for perfect separation. The ROC curve plots the true positive rates versus false-positive rates by varying the decision threshold for a learned model.

4.2 Data

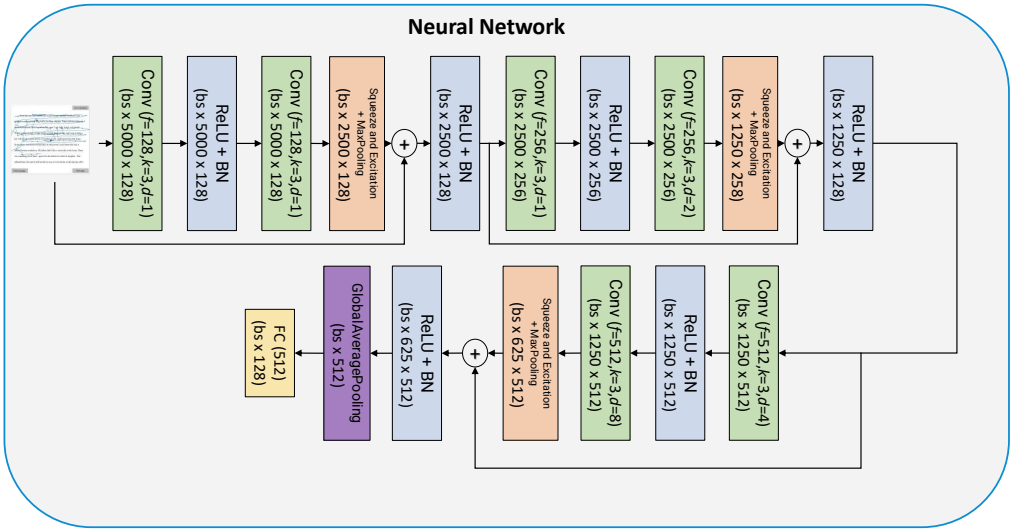
To train SP-EyeGAN, we use eye movement data from a reading experiment taken from the *GazeBase* database [Griffith et al. 2021]. *GazeBase* consists of gaze recordings from 322 college-aged participants recorded monocularly with an EyeLink 1000 eye tracker at a sampling frequency of 1,000 Hz. The participants were recorded in up to nine sessions over several months, and the recordings were taken while reading a poem. We use fixational data as training data for FixGAN and saccadic data as training data for SacGAN, respectively. We extract fixations and saccades using the Dispersion-Threshold Identification algorithm [Salvucci and Goldberg 2000].

For the downstream tasks, we use the raw eye movement recordings of the Stony Brook Scholastic Assessment Test (*SB-SAT*) dataset [Ahn et al. 2020]. *SB-SAT* consists of eye movement data

[1] https://github.com/aeye-lab/sp-eyegan



(a) EKYT model used for downstream tasks. See Lohr and Komogortsev [2022] for more details.



(b) CLRGaze model used for downstream tasks. See Bautista and Naval [2020] for more details.

Figure 2: Neuronal Networks used to pre-train and fine-tune a model on the downstream task. Figure 2a and2b depict the model architectures trained with batch size bs consisting of fully connected layers (denoted as FC), batch normalization (denoted as BN), convolutionanl layers (denoted as Conv with filter size f, kernel size k and dilation d), and the rectified linear unit (ReLU) activation function. The numbers in brackets show the dimensions of the data after each layer.

from 95 undergraduate students reading SAT texts, followed by comprehension questions recorded at a sampling rate of 1,000 Hz. Table 1 shows descriptive statistics for the datasets used in this study.

4.3 Synthetic Data Quality

We evaluate the quality of the generated synthetic data by comparing generated and real eye-movement events in terms of descriptive features. In order to measure the quality of generated fixations, we

Table 1: Dataset statistics. Descriptive statistics for datasets used to train and evaluate EyeGAN.

Dataset	Number of participants	Eye tracking device	Sampling frequency
GazeBase [Griffith et al. 2021]	322 (151 female, 171 male)	EyeLink 1000	1,000 Hz
SB-SAT [Ahn et al. 2020]	95	EyeLink 1000	1,000 Hz

Table 2: Quality of generated fixations in terms of the Jensen-Shannon divergence between human eye movement data and data generated by the model. For reference, the table also shows the divergence between two different parts of human eye movement data, denoted as *real*. Bold values indicate the best model.

	Jensen-Shannon divergence ↓		
Method	Velocity	Mean velocity	Dispersion
Statistical model [Fuhl et al. 2018]	0.283	0.679	–
VAE [Fuhl et al. 2021]	0.201	0.946	0.722
EyeSyn [Lan et al. 2022]	0.064	0.785	0.989
SP-EyeGAN	**0.029**	**0.295**	**0.271**
Real	0.0	0.026	0.045

Table 3: Quality of generated saccades in terms of the Jensen-Shannon divergence between human data and data generated by the model. For reference, the table also shows the divergence between two different sequences of human eye movement data, denoted as *real*. Bold values indicate the best model.

	Jensen-Shannon divergence ↓				
Method	Peak velocity	Mean velocity	Peak acceleration	Mean acceleration	Amplitude
Statistical model [Fuhl et al. 2018]	0.346	**0.226**	0.921	0.856	–
VAE [Fuhl et al. 2021]	0.924	0.929	0.912	0.907	0.915
SP-EyeGAN	**0.33**	0.23	**0.263**	**0.22**	**0.214**
Real	0.026	0.021	0.235	0.031	0.03

Figure 3: Generated eye movement sequences using EyeGAN. The fixation locations are sampled using a statistical model [Rayner and McConkie 1976].

measure the Jensen-Shannon divergence between real and generated fixations in terms of velocities, mean velocities, and dispersion, respectively. The quality of generated saccades is determined by comparing the peak velocity, mean velocity, peak acceleration, mean acceleration, and the amplitude of a saccade. We compare SP-EyeGAN to the statistical models proposed by Fuhl et al. [2018] and Lan et al. [2022], and the neural network approach of Fuhl et al. [2021].

Table 2 compares generated fixations. Note that the statistical model only creates velocities without directions so we can not compute the dispersions. We can conclude that the fixation profiles

generated by SP-EyeGAN are more similar to real human eye-movement sequences than fixations generated by baseline methods. The results for comparing model generated saccades can be seen in Table 3. We are not able to compute saccade amplitudes for the statistical model proposed by [Fuhl et al. 2018], because it only creates velocities without directions. From the results we can conclude that SP-EyeGAN creates saccades that are more similar to human saccades than the baseline methods. The statistical model beats SP-EyeGAN when investigating the mean velocities for saccades, but performs worse comparing the other attributes of a saccade. Figure 3 shows four different synthetic eye movement sequences generated using SP-EyeGAN that look like eye movement data generated by humans while reading a text.

4.4 Exemplary Downstream Tasks

In order to quantify the benefit of pre-training a model on synthetic data generated by SP-EyeGAN, we investigate four downstream tasks that have proven quite challenging [Ahn et al. 2020; Berzak et al. 2018; Reich et al. 2022]: The prediction of i) general reading comprehension skills, ii) text comprehension, iii) experienced text difficulty and iv) whether the reader is a native speaker. We compare the performance of deep neural models [Bautista and Naval 2020; Lohr and Komogortsev 2022] that are able to process raw eye-tracking data, and work exceptionally well for biometric identification, on these four exemplary downstream tasks in two settings: when being trained from scratch on human data only and when being first pre-trained on synthetic eye movement sequences generated by SP-EyeGAN and then fine-tuned on the human data. The labels extracted for each task are: overall comprehension score across all passages (General Reading Comprehension), text-based comprehension accuracy (Text Comprehension), a subjective difficulty rating (Text Difficulty), and whether the presented text was the first language of the reader (Native Reader).

4.4.1 Results for Downstream Tasks. We evaluate the effect of contrastive pre-training with SP-EyeGAN-generated synthetic data on two neural network models that are designed to process raw eye movement data: CLRGaze [Bautista and Naval 2020] and EKYT [Lohr and Komogortsev 2022] (see Figure 2). All models are compared to the current state-of-the-art BEyeLSTM [Reich et al. 2022] (which is not able to process raw data, but only preprocessed fixations). We apply 5-fold cross-validation splitting the training and test data by reader, that is, only readers not seen during training are used for testing. Note that splitting by readers rather than by texts has been found to be the more challenging evaluation setting since it assesses the models' ability to generalize to novel readers [Makowski et al. 2019; Reich et al. 2022].

An overview of the results can be found in Table 4. We find that contrastive pre-training significantly improves performance compared to models trained from scratch in three cases and appears to improve the performance in the remaining five cases. For three out of four downstream tasks, our approach establishes a new state-of-the-art. For the fourth downstream task, BEyeLSTM—that processes engineered features of the fixated text which our models have no access to—remains the state of the art.

5 LIMITATIONS

Although SP-EyeGAN shows promising results in generating synthetic scanpaths, there are still limitations to consider.

First, while our model is able to generate eye movements (raw samples) during fixations and saccades, it does not generate fixation and saccade durations directly, but rather complements any model that generates fixation locations and durations. In our approach, we sample the durations using a statistical model. Future research is needed to explore the integration of duration information into our model.

Second, our model was only tested on a single data set and a relatively homogenous sample of participants - predicting reading comprehension of US college students. Therefore, its generalizability to other eye-tracking data sets or populations is unknown. Our model's performance on other tasks and different data sets and populations remains to be evaluated in future work.

Another limitation of our model is that it is currently not able to process the viewed stimulus as input and hence does not take it into account for the generation of the scan path. While the advantage of this approach is the model's ability to generalize across stimulus types, the pay-off is that the model is not optimized for specific tasks where stimulus content is a crucial factor affecting the characteristics of the fixational and saccadic dynamics. In order to alleviate this short-coming, in future work, we plan to explore ways to incorporate stimulus information into our model.

Finally, our model does not account for smooth pursuits, which are an important oculomotor event that occurs while viewing a moving stimulus. Given training data containing smooth pursuits, it is straight forward to extend our model to include smooth pursuit movements.

Despite these limitations, our model represents a significant step forward in using machine learning to generate synthetic raw eye-tracking data. Future studies can build on our work to address these limitations and further improve the accuracy and generalizability of models generating eye movement data.

6 DISCUSSION

We have introduced SP-EyeGAN, a method that generates realistic raw eye-tracking data. Fixational micro-movements can be generated around fixation locations taken from any model of eye movement control—be it a statistical model, a machine-learning based model, or a cognitive model. SP-EyeGAN connects these fixations with realistic saccadic movements. The synthetic raw eye gaze sequences can be used to pre-train neural networks that are designed to process raw eye movement data for any downstream task. In this pre-training step, the downstream neural network learns to compute informative neural representations of eye movement sequences—at first, independently of the downstream task. In a final step, the neural network is fine-tuned with human eye-tracking data for any downstream task of the researcher's choice.

As a proof of concept, we have investigated four downstream prediction tasks that have recently attracted attention in eye-tracking-while-reading research. Although we used neural network architectures that were originally developed for other tasks, we found that pre-training on SP-EyeGAN-generated synthetic data improved their performance significantly in some and appeared to improve

Table 4: AUC ± standard error reported for 5-fold CV. Contrastively pre-trained models are indicated by (CP), all others are trained from scratch. A star denotes models with a performance significantly better than random guessing and a pre-trained model marked with a † indicates a model that is significantly better than its variant trained from scratch.

Method	Task			
	General Reading Comprehension	Text Comprehension	Text Difficulty	Native Reader
BEyeLSTM	0.608±0.037*	0.542±0.015*	**0.710±0.017***	0.670±0.025*
EKYT	0.585±0.015*	0.566±0.020*	0.494±0.021	0.550±0.014*
EKYT (CP)	**0.622±0.029***	0.574±0.024*	0.545±0.006*†	**0.721±0.061***†
CLRGaze	0.569±0.065	0.560±0.055	0.516±0.034	0.528±0.046
CLRGaze (CP)	0.577±0.033	**0.592±0.032***	0.566±0.018*	0.704±0.050*†

their performance in other cases. For three of the four downstream tasks, our approach establishes a new reference performance.

To date, most researchers focus on methods that operate on preprocessed scanpaths of fixations and saccades, often using engineered fixational and saccadic features. Recent research in eye-tracking-based biometrics [Jäger et al. 2020], however, has shown that the raw eye-tracking signal contains valuable information that is lost by preprocessing. Since neural networks that are designed to process raw eye-tracking data typically have even more parameters, data scarcity is a major obstacle. Our proposed approach opens the possibility to develop deep neural networks with large numbers of parameters since potentially infinite amounts of synthetic data are available for (pre)-training.

Besides our approach's advantages for training neural networks, it has also important advantages for *privacy*. In recent years, it has been shown that in many cases, it is possible to reconstruct the training data from a neural network's final parameters [Carlini et al. 2021], which can violate the privacy of donors of training data: it may be possible to infer the training users' identity, gender or other sensitive attributes [Lahey and Oxley 2021; Lohr and Komogortsev 2022; Makowski et al. 2021]. The inclusion of synthetic training data dilutes any potentially identifiable traits.

7 CONCLUSION

We have developed and evaluated an approach for generating synthetic raw eye movement data that outperforms previous statistical and machine-learning based approaches in terms of the statistical likeness of the generated data with human eye-tracking data. We have further found that using these synthetic data for contrastive pre-training of neural networks that process raw eye-tracking data for downstream tasks in many cases improves the performance on these downstream tasks, often establishing new performance benchmarks. Thereby, our approach paves the way to training better-performing, higher-capacity models for a wealth of eye-tracking-related problems.

ACKNOWLEDGMENTS

This work was partially funded by the German Federal Ministry of Education and Research under grant 01| S20043.

REFERENCES

Seoyoung Ahn, Conor Kelton, Aruna Balasubramanian, and Greg Zelinsky. 2020. Towards predicting reading comprehension from gaze behavior. In *ACM Symposium on Eye Tracking Research and Applications*. Association for Computing Machinery, Stuttgart, Germany, 1–5.

Marc Assens, Xavier Giro-i Nieto, Kevin McGuinness, and Noel E O'Connor. 2018. PathGAN: Visual scanpath prediction with generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 0–0.

Louise Gillian C. Bautista and Prospero C. Naval. 2020. CLRGaze: Contrastive Learning of Representations for Eye Movement Signals. *2021 29th European Signal Processing Conference (EUSIPCO)* (2020), 1241–1245.

Yevgeni Berzak, Boris Katz, and Roger Levy. 2018. Assessing language proficiency from eye movements in reading. In *Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, New Orleans, Louisiana, 1986–1996.

Christopher Bowles, Liang Chen, Ricardo Guerrero, Paul Bentley, Roger Gunn, Alexander Hammers, David Alexander Dickie, Maria Valdés Hernández, Joanna Wardlaw, and Daniel Rueckert. 2018. GAN augmentation: Augmenting training data using generative adversarial networks. *arXiv preprint arXiv:1810.10863* (2018).

Daniel J. Campbell, Joseph Chang, Katarzyna Chawarska, and Frederick Shic. 2014. Saliency-based bayesian modeling of dynamic viewing of static scenes. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. 51–58.

Nicholas Carlini, Florian Tramer, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, Ulfar Erlingsson, et al. 2021. Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)*. 2633–2650.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the 37th International Conference on Machine Learning (ICML'20)*. JMLR.org, Article 149, 11 pages.

Shuwen Deng, Paul Prasse, David R. Reich, Sabine Dziemian, Maja Stegenwallner-Schütz, Daniel Krakowczyk, Silvia Makowski, Nicolas Langer, Tobias Scheffer, and Lena A. Jäger. 2022. Detection of ADHD based on eye movements during natural viewing. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*. Springer, Grenoble, France.

Andrew T. Duchowski and Sophie Jörg. 2015. Modeling physiologically plausible eye rotations. In *Proceedings of Computer Graphics International*.

Andrew T. Duchowski, Sophie Jörg, Tyler N Allen, Ioannis Giannopoulos, and Krzysztof Krejtz. 2016. Eye movement synthesis. In *Proceedings of the ninth biennial ACM symposium on eye tracking research & applications*. 147–154.

Andrew T. Duchowski, Sophie Jörg, Aubrey Lawson, Takumi Bolte, Lech Świrski, and Krzysztof Krejtz. 2015. Eye movement synthesis with 1/f pink noise. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*. 47–56.

Ralf Engbert, Antje Nuthmann, Eike M Richter, and Reinhold Kliegl. 2005. SWIFT: a dynamical model of saccade generation during reading. *Psychological review* 112, 4 (2005), 777.

Wolfgang Fuhl and Enkelejda Kasneci. 2018. Eye movement velocity and gaze data generator for evaluation, robustness testing and assess of eye tracking software and visualization tools. *arXiv preprint arXiv:1808.09296* (2018).

Wolfgang Fuhl and Enkelejda Kasneci. 2022. HPCGen: Hierarchical K-Means Clustering and Level Based Principal Components for Scan Path Genaration. In *2022 Symposium on Eye Tracking Research and Applications*. 1–7.

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)*. IEEE Computer Society, Los Alamitos, CA, USA, 142–149. https://doi.org/10.1109/ICPR48806.2021.

9413268

Wolfgang Fuhl, Thiago Santini, Thomas Kuebler, Nora Castner, Wolfgang Rosenstiel, and Enkelejda Kasneci. 2018. Eye movement simulation and detector creation to reduce laborious parameter adjustments. *arXiv preprint arXiv:1804.00970* (2018).

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger (Eds.), Vol. 27. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf

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), 1–9.

Patrick Haller, Andreas Säuberli, Sarah E. Kiener, Jinger Pan, Ming Yan, and Lena A. Jäger. 2022. Eye-tracking based classification of Mandarin Chinese readers with and without dyslexia using neural sequence models. In *Proceedings of the Workshop on Text Simplification, Accessibility, and Readability*. Association for Computational Linguistics, Abu Dhabi, UAE.

Lena A Jäger, Silvia Makowski, Paul Prasse, Sascha Liehr, Maximilian Seidler, and Tobias Scheffer. 2020. Deep Eyedentification: Biometric Identification using Micro-Movements of the Eye. In *ECML/PKDD 2019*. 299–314.

Matthias Kümmerer and Matthias Bethge. 2021. State-of-the-art in human scanpath prediction. *arXiv preprint arXiv:2102.12239* (2021).

Joanna N Lahey and Douglas R Oxley. 2021. Discrimination at the Intersection of Age, Race, and Gender: Evidence from an Eye-Tracking Experiment. *Journal of Policy Analysis and Management* 40, 4 (2021), 1083–1119.

Guohao Lan, Tim Scargill, and Maria Gorlatova. 2022. 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)*. 233–246. https://doi.org/10.1109/IPSN54338.2022.00026

Binh Le, Xiaohan Ma, and Zhigang Deng. 2012. Live speech driven head-and-eye motion generators. *IEEE transactions on visualization and computer graphics* 18, 11 (2012), 1902–1914.

Sooha Park Lee, Jeremy B Badler, and Norman I Badler. 2002. Eyes alive. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. 637–644.

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

Xiaohan Ma and Zhigang Deng. 2009. Natural Eye Motion Synthesis by Modeling Gaze-Head Coupling. In *2009 IEEE Virtual Reality Conference*. IEEE Computer Society, Los Alamitos, CA, USA, 143–150. https://doi.org/10.1109/VR.2009.4811014

Silvia Makowski, Lena A Jäger, Ahmed Abdelwahab, Niels Landwehr, and Tobias Scheffer. 2019. A discriminative model for identifying readers and assessing text comprehension from eye movements. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 209–225.

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* (2021).

Christopher Manning and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. MIT press.

Antje Nuthmann, Tim J Smith, Ralf Engbert, and John M Henderson. 2010. CRISP: a computational model of fixation durations in scene viewing. *Psychological review* 117, 2 (2010), 382.

Keith Rayner and George W McConkie. 1976. What guides a reader's eye movements? *Vision research* 16, 8 (1976), 829–837.

David R. Reich, Paul Prasse, Chiara Tschirner, Patrick Haller, Frank Goldhammer, and Lena A. Jäger. 2022. Inferring Native and Non-Native Human Reading Comprehension and Subjective Text Difficulty from Scanpaths in Reading. In *2022 Symposium on Eye Tracking Research and Applications* (Seattle, WA, USA) *(ETRA '22)*. Association for Computing Machinery, New York, USA, Article 23, 8 pages.

ED Reichle, K Rayner, and Pollatsek A. 2003. The E-Z reader model of eye-movement control in reading: comparisons to other models. *The Behavioral and Brain Sciences* 26 (2003), 445–526. Issue 4.

Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *ETRA 2020*. 71–78.

Daniel Simon, Srinivas Sridharan, Shagan Sah, Raymond Ptucha, Chris Kanan, and Reynold Bailey. 2016. Automatic scanpath generation with deep recurrent neural networks. In *Proceedings of the ACM Symposium on Applied Perception*. 130–130.

Erroll Wood, Tadas Baltrusaitis, Xucong Zhang, Yusuke Sugano, Peter Robinson, and Andreas Bulling. 2015. Rendering of eyes for eye-shape registration and gaze estimation. In *Proceedings of the IEEE international conference on computer vision*. 3756–3764.

Sang Hoon Yeo, Martin Lesmana, Debanga R Neog, and Dinesh K Pai. 2012. Eyecatch: Simulating visuomotor coordination for object interception. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 1–10.