Article

Exploration of Eye Fatigue Detection Features and Algorithm Based on Eye-Tracking Signal [†]

Weifeng Sun [1,2,3], Yuqi Wang [1,3], Bingliang Hu [1,3] and Quan Wang [1,3,*]

[1] Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China
[2] University of Chinese Academy of Sciences, Beijing 101408, China
[3] Key Laboratory of Biomedical Spectroscopy of Xi'an, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China
* Correspondence: wangquan@opt.ac.cn
[†] This article is a revised and expanded version of a paper entitled "Exploring the Connection between Eye Movement Parameters and Eye Fatigue", which was presented in 2023 at the 4th Asia Conference on Computers and Communications (ACCC 2023) on 16 December 2023 by Weifeng Sun et al.

Abstract: Eye fatigue has a fatiguing effect on the eye muscles, and eye movement performance is a macroscopic response to the eye fatigue state. To detect and prevent the risk of eye fatigue in advance, this study designed an eye fatigue detection experiment, collected experimental data samples, and constructed experimental data sets. In this study, eye-tracking feature extraction was completed, and the significance difference of eye-tracking features under different fatigue states was discussed by two-way repeated-measures ANOVA (Analysis of Variance). The experimental results demonstrate the feasibility of eye fatigue detection from eye-tracking signals. In addition, this study considers the effects of different feature extraction methods on eye fatigue detection accuracy. This study examines the performance of machine learning algorithms based on manual feature calculation (SVM, DT, RM, ET) and deep learning algorithms based on automatic feature extraction (CNN, auto-encoder, transformer) in eye fatigue detection. Based on the combination of the methods, this study proposes the feature union auto-encoder algorithm, and the accuracy of the algorithm for eye fatigue detection on the experimental dataset is improved from 82.4% to 87.9%.

Keywords: eye tracking; feature analysis; eye fatigue; deep learning

1. Introduction

Eye fatigue is defined as eye and visual symptoms caused by prolonged use of digital electronic display devices [1], such as sore eyes, swollen eyes, and red eyes, and is one of the major global health problems. Studies have shown that using electronic screens for more than two hours is enough to trigger the risk of eye fatigue [2]. For the different use scenarios, eye fatigue is also known as digital visual fatigue [3]. With the increasing sophistication of current digital electronic display terminal technology, the public's daily life is deeply bound to electronic devices. In addition, online lectures, online offices, and other behaviors are widespread today. This situation further aggravates the public's demand for the use of the internet and electronic equipment. It has led to a large increase in the average time spent on electronic devices.

One of the major direct effects of eye fatigue is damage to the eye's regulatory mechanisms [4,5]. The eye accommodation mechanism is the main mechanism of visual function, which is related to binocular vision, such as the ability of binocular coordination and cooperation. The eye accommodation mechanism contains light perception, color perception, and contrast perception, which is significantly affected by eye fatigue. Another manifestation of eye fatigue is damage to the circulatory mechanisms of the ocular surface, especially direct damage to the tear film homeostasis. This kind of damage appear as symptoms

such as eye dryness and redness, which can lead to ocular disorders such as dry eye and blurred vision [6]. The tear film, as the first refractive medium on the ocular surface and one of the important links in retinal imaging, consists of a lipid, aqueous, and mucin layer 7–40 μm thickness. As the tear film is exposed to air for a long period of time, it is highly susceptible to alteration by air and external factors, and the human body maintains or rearranges the tear film environment on the ocular surface by blinking. However, as the time of electronic screen use and the degree of eye fatigue increases, the frequency of human blinking decreases significantly [7].

In recent years, eye-tracking technology has emerged as a powerful tool in the study of fatigue detection. Eye-tracking technology has multiple advantages over traditional self-report methods and observational techniques, notably including objective and quantitative assessment of fatigue. By precisely and non-invasively measuring eye movements, eye-tracking provides detailed insights into visual and attention behavior during various tasks [8]. This technology enables researchers to gain a deeper understanding of fatigue and develop more effective detection and management strategies. However, within the field of fatigue detection, many researchers have focused their use scenarios on driving fatigue detection, and eye fatigue detection for daily environments is still at a relatively initial stage.

The work content of this study can be divided into two points: 1. In this study, significant differences of eye tracking features in different fatigue states and the performance of eye fatigue detection algorithms based on different feature extraction methods are thoroughly investigated based on eye tracking signal acquisition. 2. This study explores the feasibility of completing eye fatigue detection through eye-tracking signals and combines the performance of multiple algorithms to propose a feature union auto-encoder algorithm used in eye fatigue detection.

2. Related Works

In studies using eye-tracking signals for fatigue detection, the main detection tools available are video signals, electrooculography signals, and multimodal fusion methods. The study of electrooculography (EOG) has been an important aspect of eye fatigue research [9–12]. The EOG signal is a measurement that uses electrodes placed in four directions: above, below, to the left, and to the right of the eye. The purpose of this behavior is to measure the change in the potential difference between different potentials and reference points as a response to the direction and speed of eye movement, as shown in Figure 1a. The degree of eye fatigue is determined by calculating the amplitude changes in the horizontal and vertical electromyography of the eye, measuring the speed of eye movements, and comparing the timing longitudinally [13]. Electrooculography plays a crucial role in studies to reconstruct eye movement trajectories. Most fatigue-monitoring algorithms based on electrical signals use blink frequency and mean eye closure time as indicators to judge the state [14]. There are also studies that perform time-frequency conversion of the acquired EOG signals to extract features from the frequency domain for fatigue analysis [15–17]. However, the data quality of the EOG signals is severely affected by the acquisition process. Any small body movement of the subject can lead to fluctuations in signal quality, thus reducing the confidence and accuracy of the EOG signal. To improve the accuracy and reliability of the eye fatigue detection task, researchers have conducted a lot of studies on EOG-based multimodal fusion detection methods. These studies fuse the EOG signals with electrocardiographic sensors, skin reflectance currents, and skin temperature signals [18–22]. Furthermore, with the rise of brain–machine interface (BCI) technology and the optimization of sensors, researchers have also used steady-state visual evoked potential (SSVEP) methods [23–25] and electrocardiogram (ECG) measurements for eye fatigue detection tasks [26–28]. Although these algorithms can improve the accuracy of the detection task, their signal acquisition is rather difficult. The interpretability of the signal is also poor.
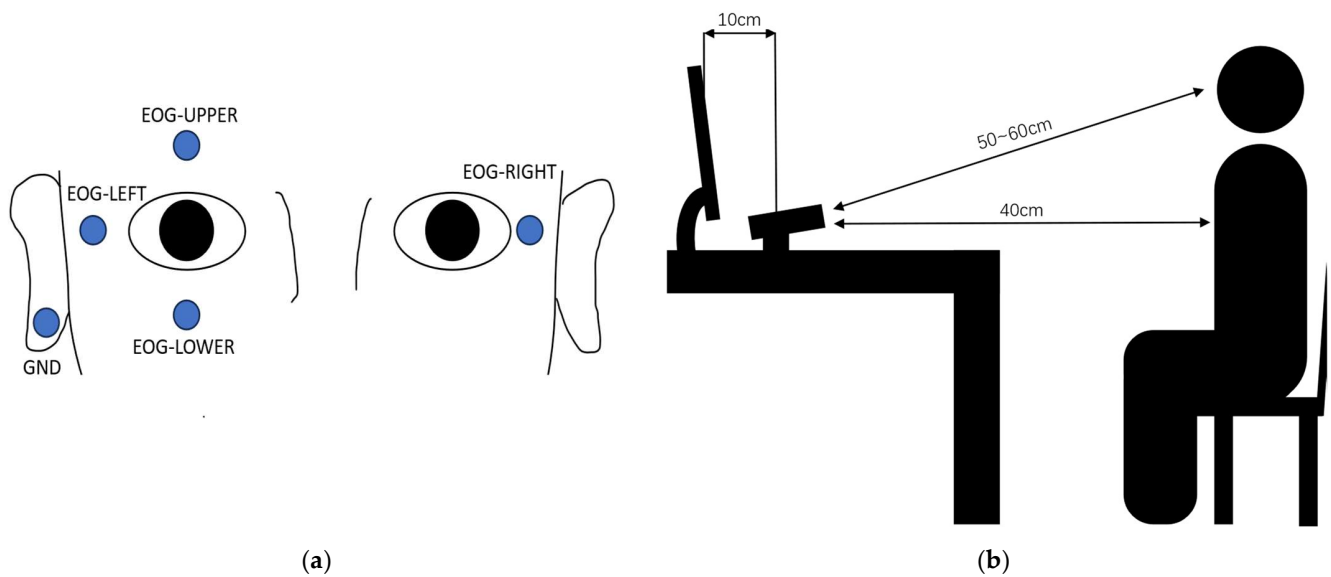
(**a**)

(**b**)

Figure 1. (**a**) Schematic diagram of EOG signal acquisition; (**b**) schematic diagram of the eye fatigue experiment.

In contrast, video signals are much more convenient for information acquisition compared to other physiological sensors, such as EOG. The interpretability of video signals is also significantly better than that of signal sources such as EOG due to the close relationship between the visualization features of video signals and the physiological structure of the eye. However, in the current research on fatigue detection based on video eye-tracking signals, most studies are oriented towards driving fatigue and fatigue from the use of virtual reality technologies [29–32]. These studies usually use only the number of blinks and mean eye closure time as discriminators and do not delve into the effects of eye fatigue status on eye movement. While most fatigue detection algorithms can demonstrate excellent fatigue detection accuracy and reliability in driving fatigue detection, there is still less research on eye fatigue detection in daily life. Previous studies have shown that eye movement metrics are valuable for eye fatigue detection tasks [33]. Therefore, this study designs an experiment to explore the small changes in eye movement features reflected by video signals in different eye fatigue status. This study also completes eye fatigue detection on this basis by deeply researching algorithms based on different feature extraction methods.

In conclusion, this paper will focus on the use of electronic screens in daily environments as the target scenario. The research objectives of this paper are to study the detection methods of eye fatigue from the perspective of accommodation mechanism and circulatory mechanism damage. Specifically, the research objective of this paper is to experimentally verify the specific manifestations of the impairment of regulatory and circulatory mechanisms triggered by eye fatigue in the target scene. This research aims to verify and test the response of eye-tracking features in the detection of eye fatigue and to propose a series of complete, reliable, and interpretable eye fatigue detection features. This study designs an automatic detection algorithm of eye fatigue to capture small changes in fatigue status that are difficult for humans to detect and provide guidance and reference value for fatigue research in other domains.

3. Methodology

3.1. Experimental Setup

In the process of eye movement signal acquisition, this research adopted a 1920×1080 resolution electronic display and used a self-developed desktop near-infrared eye tracker. The eye tracker consists of an infrared camera and two sets of infrared fill-in lights placed at ten centimeters on both sides of the camera. The eye-tracking system also adopts a

nine-point calibration method. This system provides near-infrared images with a resolution of 3840 × 2160 pixels and an acquisition frequency of 25 frames/second. The eye tracker was placed horizontally below the electronic display during the experimental interval, and subjects were required to face the camera and to minimize head rotation during the experiment. In addition, the distance between the subject's eyes and the eye-tracker was strictly controlled to be between 50 and 60 cm. The calibration and eye-tracking signal acquisition started from the end of the questionnaire and visual acuity test. A schematic diagram of the experiment is shown in Figure 1b.

The Chinese Ethics Review Committee requires ethical approval for research activities related to the prevention, diagnosis, treatment, and rehabilitation of diseases, as well as interventional research involving the collection or storage of biological samples related to the life sciences through epidemiological and sociological methods. Since this research only use infrared cameras to collect subjects' eye movements in our study, there was no contact and intervention behavior, there was no harm to subjects, and it was not a disease-related study, so this research did not need to apply for ethics committee approval. There are many similar examples in previous experimental studies of eye tracking [34–36]. All experiments in this study were conducted in strict accordance with the relevant laws and the principles embodied in the Helsinki Declaration, and each participant personally signed an informed consent form.

3.2. Feature Extraction

Based on the experimental design and task requirements, various features were extracted from different experiments to capture the overall characteristics. These features can be categorized into three main aspects:

- Pupil-related features: number of blinks (BN), average duration of eye closures (BD), maximum pupil sizes (PMAX), minimum pupil sizes (PMIN), mean velocity of pupil change (PV), response time for the pupil to adjust to screen changes (PL);
- Eye movement features: number of fixations (FN), average duration of fixations (FD), fixation as a percentage of total time (FR), number of saccades (SN), average distance covered during saccades (SD);
- Task performance features: percentage of the duration of subjects' fixation on the search target during the search phase in relation to the duration of the entire search phase in the visual search task (Target Fixation Ratio, TFR), response time for the fixation point to move from the center of the screen to the area where the search target is located after the start of the search phase of the visual search task (Target Search Latency, TSL), percentage of experiments in which subjects' fixation points were within the target area at the end of the search phase of the visual search task (Task Success Ratio, TSR), proportion of time subjects looked at the center of the screen in the FF, SF, LG, and CF tasks as a proportion of the total time spent on each task (Center Fixation Ratio, CFR).

3.3. Methods of Statistical Analyses

3.3.1. Two-Way Analysis of Variance (Two-Way ANOVA)

Two-way ANOVA generally defines the influencing factors in an experiment as fac-tor *A* and factor *B*. In data processing, two-way ANOVA requires that the interrelationships between factor *A* and factor *B* be presupposed before analyzing. The different analytical ideas should be used for the data in which *A* and *B* do or do not have an interaction. The former explores the influence of factor *A* and factor *B* on the experimental results and the significance of the difference in the data. However, the latter is more concerned with whether *A* and *B* produce a new effect that affects the experimental data by combining them. The correspondence between the formulae for the source of variation and the interaction for the two-factor ANOVA method is shown in the following equation.

$$SS_A = \frac{1}{2n}\left(A_1{}^2 + A_2{}^2\right) - C \tag{1}$$

$$SS_B = \frac{1}{2n}\left(B_1{}^2 + B_2{}^2\right) - C \tag{2}$$

$$SS_{AB} = \frac{1}{n}\left(T1^2 + T2^2 + T3^2 + T4^2\right) - C - SS_A - SS_B \tag{3}$$

In the above equation, *A*, *B* denote the influences; *T* denotes the sample group under the joint influence of *A*, *B*; and *C* is a constant calculated by summing the squares of *A* and *B* to correct for the effect of the interaction.

The hypothesis of no interaction between factors *A* and *B* was used in this study. Fac-tor *A* indicates the difference between the subjects themselves currently participating in the experiment, and factor *B* indicates the difference between the subjects' eye fatigue status. The focus of this study was on exploring the analysis of the between-group differences of the different fatigue levels on eye movement features for the various tasks.

3.3.2. Repeated-Measures Analysis of Variance (RM ANOVA)

RM ANOVA methods are used by researchers to assess the effect of one or more factors on a subject's multiple measurements. This method compares the differences in the variables within each group as well as the differences in the variables between each group and is a method of comparing the differences in the means of variables within multiple groups. Unlike traditional ANOVA, repeated-measures ANOVA considers multiple measurements of the same subjects and allows for a more accurate assessment of the effect of factors on the measurements. In the present study, repeated-measures ANOVA was used in the analysis of this part of the data due to the presence of multiple flicker trials in the fast flicker, slow flicker, luminance gradient, color flicker, and visual search tasks.

3.4. Eye Fatigue Detection Algorithm

During the exploration of eye fatigue detection algorithms, this study explored both machine learning algorithms for eye-tracking features and deep learning algorithms for automatic feature extraction. For the machine learning algorithms of eye-tracking features, this study successively used support vector machine (SVM), decision tree (DT), random forest (RF), and extremely randomized trees (ET), and tested the performance for different key parameters.

In the exploration of deep learning algorithms, this study tried three typical networks: convolutional neural network (CNN), auto-encoder (AE), and transformer.

3.5. Auto-Encoder Feature Union Algorithm (AEFU)

Considering the interpretability of eye movement features in eye fatigue classification tasks, this research deeply fused automatically extracted features and manually extracted features. This research updated the network architecture to explore the fusion method of automatically extracted features and manually extracted features. This research also proposed a new eye fatigue detection algorithm based on feature fusion strategy based on the auto-encoder algorithm, called the auto-encoder feature union algorithm.

This study continuously adapted the structure of the auto-encoder model. Under the optimal performance, the structure of the auto-encoder model is a five-layer encoder structure. The length of the output implicit feature variables is 128 dimensions. This study used the implicit features as the output part of the auto-encoder algorithm. To combine the manually extracted eye movement feature data and the feature data learned by the model autonomously, we calculated the eye movement feature data in the deep learning dataset using the experimental eye movement feature calculation method. The eye movement feature computation used the standard definitions of fixation and saccade, using time data and pupil size data to compute pupil-related features, and using time data and fixation point data to compute eye movement features and task performance features. In the

calculation of pupil-related features, we recorded the start of a blink as the time when the pupil size plummeted below 200 pixel points. Based on this, we concatenated the two types of feature data and used the auto-encoder model for further feature extraction and the eye fatigue detection task, which had previously performed the best. We adjusted the structure of the whole model, especially the auto-encoder structure at the back and the length of the further extracted implicit feature vectors. Finally, we chose a two-layer auto-encoder with 64-dimensional implicit features, and its overall structure is shown in Figure 2.
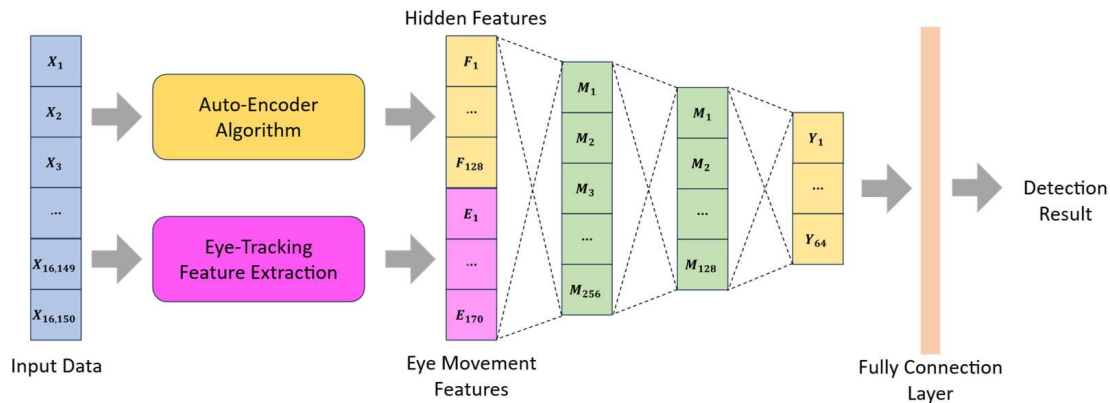

Figure 2. The optimal model architecture of AEFU.

4. Experiments

4.1. Eye Detection Experimental Task

4.1.1. Fatigue Discrimination Questionnaire (T/CVIA-73-2019)

Considering the ease of use and relative objectivity, this study adopted the "Visual Fatigue Testing and Assessment Methods for Display Terminals—Part 2 Scale Assessment Methods" published by the China Video Industry Association (CVIA) on 19 July 2019 as an eye fatigue assessment scale, No. T/CVIA 73-2019 [37]. This scale has been widely used in fatigue studies at home and abroad and has performed well in terms of accuracy and reliability. The scale contains a total of 20 questions covering the three major ocular mechanisms of eye fatigue. The questionnaire focuses on the subjective perception and visual condition of the eyes, including questions on eye fatigue, eye tightness, eye tearing, dry eyes, eye pain, and blurred vision, supplemented by questions on vertigo, headache, and neck pain, which are affected by extra-ocular mechanisms. In contrast, questions about the attentional mechanisms caused by eye fatigue were answered by the subjects based on their current sensations, such as difficulty concentrating, slow thinking, distraction, unresponsiveness, and lethargy. To quantify and differentiate the level of fatigue of the subjects in this paper, the scale gives five levels for each question for the subjects to choose from, corresponding to a score of 0, 25, 50, 75, and 100, respectively.

A part of the scale content and the scatter plot of the scale scores is shown in Figure 3. In this study, the average score of all questions was used as the scale outcome. To eliminate the possible influence of subjectivity in the eye fatigue assessment scale and fit as closely as possible the non-fatigued and fatigued states of the subjects, this research restricted the experimental time by asking the subjects to conduct the experiment before they started and finished working. From the results, the scale score of 15 is a clear cut-off between fatigued and non-fatigued states.
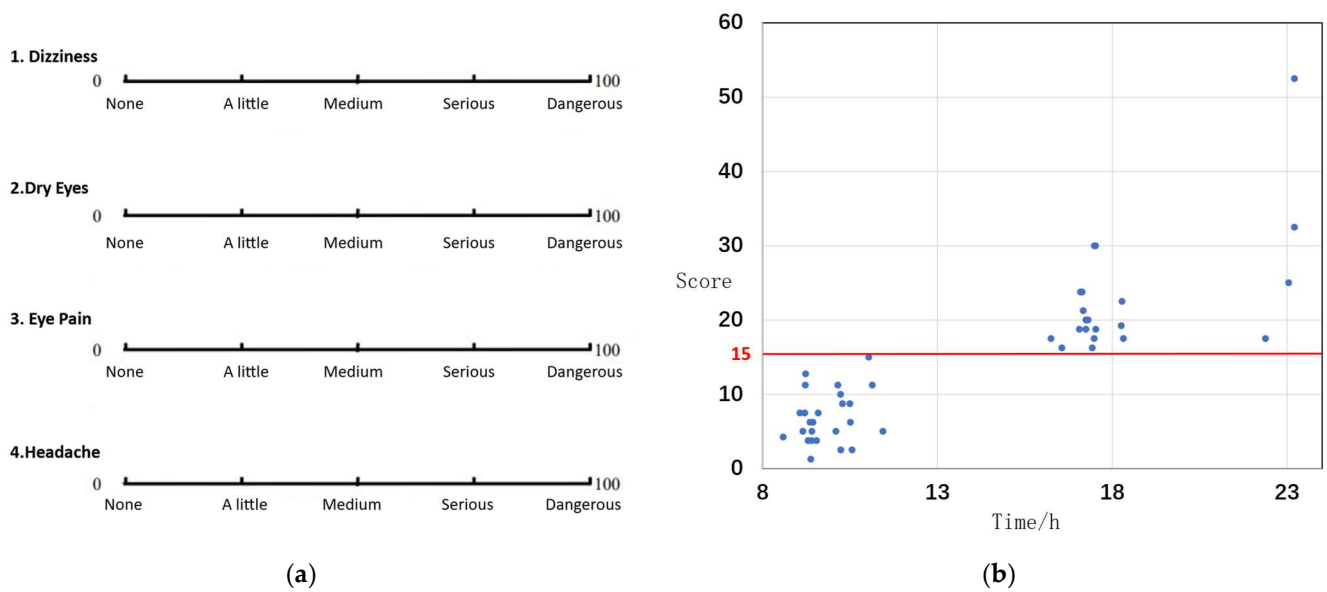
(**a**)   (**b**)

Figure 3. (**a**) Excerpts from the T/CVIA 73-2019; (**b**) scatter plot of fatigue scale.

4.1.2. Visual Acuity Test

The visual acuity test is used to check whether the current visual level and response level of the subject is normal. As there were subsequent experimental tasks in which color changes were the main stimulus, the visual acuity test also included a check of the subject's color vision level. The test was divided into two phases, as shown in Figure 4.
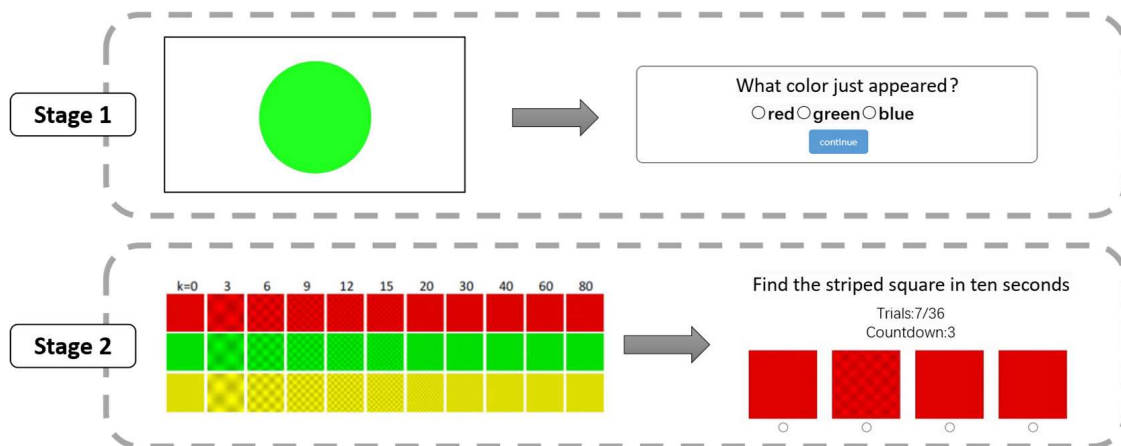

Figure 4. Schematic diagram of the visual acuity test.

In the first phase, a circular pattern with colors was flashed on a white screen, and the colors of the circular pattern were red, green, and blue randomly. The time that the circle stayed on the screen decreased with the number of trials. The subjects were asked to choose the color of the circle that had just appeared within 5 s after the circle disappeared. This test was repeated 10 times. The second phase consisted of 36 trials. Each trial resulted in a pattern of four squares, with a base color appearing on the screen. One of the squares was marked with a grid of stripes of the corresponding number of items. Subjects were required to correctly select the square with the grid within 10 s of the appearance of the square. The number of stripes increased in this phase as the number of experiments increased. Due to the limited resolution and refresh rate of the electronic monitors we used in our experiments, and to ensure that the experiments were not too simple to avoid decreasing the subjects' attentional focus, we allowed for errors in the final stage of the visual acuity

experiment. The subjects were required to achieve 90% and 85% correctness in the two stages of the visual acuity test before their experimental data could be included in this study.

4.1.3. Flicker Test

In this study, a flash task was utilized to examine potential differences in eye-tracking metrics across varying levels of fatigue. Previous research has indicated that the features of blink and pupil changes are more pronounced in scenes with significant variations [38]. If the ability of the eye muscles to adjust during fatigue is diminished, the range of pupil variation should narrow, and the response rate may decrease in an environment with rapidly changing stimuli.

The study employed a four-part scintillation test, which consisted of the following components: the fast flicker test (FF), slow flicker test (SF), luminance gradient test (LG), and color flicker test (CF). Initially, participants underwent five fast flash tests. In each trial, the screen rapidly blinked at a frequency of 5 Hz for ten seconds, with five seconds of rest between groups. This procedure was repeated for a total of five sessions. Subsequently, the slow flash test was conducted using the same timing and intervals, but with the screen blinking at a frequency of 1 Hz.

The third part of the task involved stimuli that were significantly reduced compared to the previous two parts. In this phase, the screen brightness changed from 0 to 255 at a frequency of 20 Hz. Each experiment was repeated five times, with five seconds of rest between groups.

The final part of the task examined the response of the eyes to different colors. The screen cycled through the colors blue, green, red, yellow, magenta, and cyan at a frequency of 1 Hz. Throughout the experimental phase, participants were instructed to focus their gaze on a cross positioned at the center of the screen.

4.1.4. Visual Search Test

The test was referenced from a study on attention deficits by Rommelse [39]. Considering that the state of eye fatigue can have an impact on subjects' attention, and to complement the acquisition of pupil scaling in response to small changes in the screen, this research modified and incorporated this task into the eye fatigue experimental paradigm. This test was divided into three phases:

The first phase was the preparation phase of the test, in which the screen was black. This phase consisted only of a yellow dot with a radius of 3 pixels in the center of the screen, which lasted for 2 s and required the subject to look at the center of the screen.

The second phase was the pre-search stage, in which six red circles with a radius of 30 pixels and a thickness of 2 pixels were generated around the center dot, based on the screen from the first stage. In this phase, subjects were still asked to look at the center of the screen, and the focus of attention and pupil movements were examined for a total of 2 s when the screen changed.

The third phase was a search phase, in which one of the six circles that appeared in the second phase was randomly made to be filled with grey. In this phase, subjects were asked to search for the location of the target circle and shift their fixation from the center of the screen to the target circle within 1 s. The speed of the shift and the accuracy of the search were analyzed as target features.

All three phases were used as one complete trial of the visual search test, and the whole test flow is shown in Figure 5. Each group of experiments contained eight trials, and other circles appeared as interference items in the third phase of the last four trials, while they did not appear the first four times. The whole experiment went through five sets of experiments, with a rest of 5 s between sets.
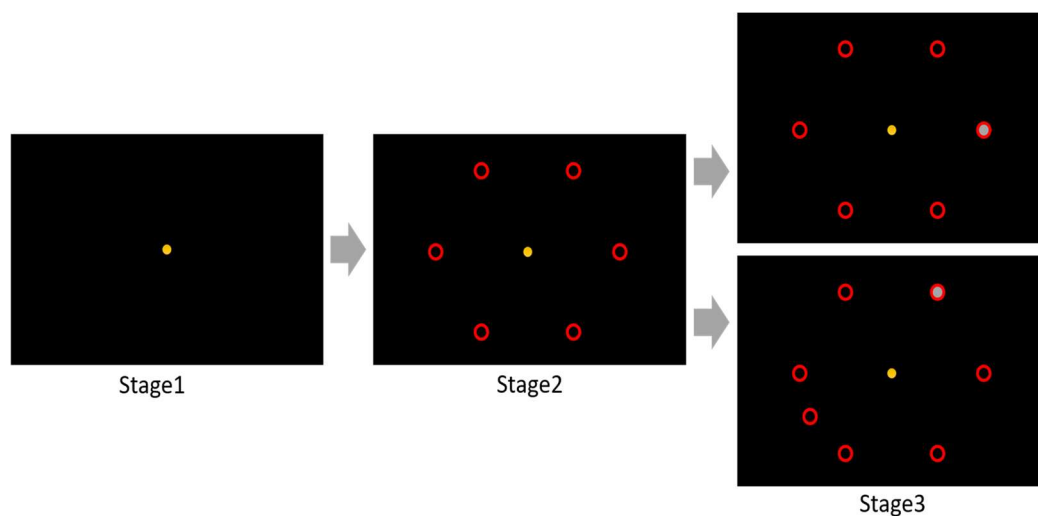
Figure 5. Schematic diagram of the visual search test.

4.1.5. Reading Test

The main role of the reading task is to simulate the daily use of an electronic screen and to explore whether there are differences in subjects' eye movement behaviors between eye fatigue and non-fatigue states in a free eye movement environment. During this phase, the screen presented a black-on-white text passage of approximately 500 words, which was a partial excerpt from a well-known novel without emotional guidance. The subjects were questioned about their reading before the start of the experiment to make sure that they had no prior knowledge of the passage. The test lasted 60 s each time, with a 5 s break between sets, and was administered three times.

4.2. Participants

A total of 19 participants, including 13 males and 6 females, were recruited for this study, all of whom were enrolled in high school or master's degree programs. All subjects were fully informed of the purpose of this study before the experiment and signed a written consent form before starting the experiment. The inclusion criteria for participant selection were as follows: (a) refractive errors in the primary and nonprimary eyes that met specific requirements for each group, with a cylindrical diopter of less than 0.50 D and an equivalent spherical diopter difference between the eyes of less than 1.00 D; (b) corrected binocular visual acuity of 1.0 or higher; (c) students between the ages of 14 and 26 years enrolled in either a high school program or a master's program; (d) normal performance on visual acuity tests; and (e) use of electronic screens for a long time before the experiment with eye fatigue status.

Participants with the following conditions were excluded from the study: (a) use of medications that may affect changes in pupil size or mechanisms of ocular surface regulation; (b) non-myopic eye diseases, such as retinal detachment and strabismus, that can be corrected with lenses; (c) presence of hereditary developmental disorders or cognitive impairments; or (d) history of alcohol consumption, drug abuse, or use of medications that affect cognition, such as antiepileptics, antipsychotics, or anticholinergics.

4.3. Eye Fatigue Dataset

Since deep learning models usually require a large number of data, this research performed data augmentation and constructed the dataset from the original collected eye movement signal dataset, and the main process was as follows:

1.   Normalization of each subject's data. This research collected pupil data from each subject in the resting state during the experiment preparation phase, taking the mean

value as the pupil size benchmark. This research also calculated the pupil size change throughout the trial phase sequentially as the ratio to the pupil size benchmark. The differences in the features of the eyes themselves between different subjects were reduced by normalization, which served as the basis for the subsequent construction of the eye movement dataset for eye fatigue.

2. Sample slicing. We defined samples judged to be fatigued and non-fatigued as positive and negative samples, respectively. All the positive and negative samples were numbered separately, and the data were sliced according to the starting and ending points of the test. The data of each sample were sliced into six segments, corresponding to the six tasks of the fast flicker test, slow flicker test, luminance gradient test, visual search test, color flicker test, and reading test in the eye fatigue experiment;

3. Random combination and generation of datasets. After slicing the sample, the random sampling method was used for dataset construction. Positive and negative sample data were obtained. Considering the practicality of the dataset, this research finally generated 10,000 non-repeated data points for each of the positive and negative samples as model training. The number of training, validation, and test sets used in training was 8:1:1.

In the process of randomly generating the dataset, this research also recorded the original sample number used for each piece of data so as to ensure that there were no duplicate samples in the dataset.

5. Results

5.1. Feature Analysis

In the process of data analysis, this research carried out correspondence matching for different stages of different tasks and corresponding features, and the results of the analysis are shown in Tables 1–4. Table 1 presents the overall amount of eye fatigue and non-fatigue sample data collected in this study. Table 2 shows the results of the analysis of BN and BD in the whole test and other pupil-related features in the flicker test using the two-way RM ANOVA. Table 3 shows the statistical analysis performance of eye movement features in all tests under eye fatigue and non-fatigue. Table 4 presents the statistical analysis performance of the full task performance features, including TFR, TSL, and TSL in the visual search test and CFR in the flicker test.

Table 1. Number of tested samples for data analysis.

Tested Sample	Positive (eye-fatigue)	18
	Negative (non-fatigue)	20

Table 2. Data analysis of pupil-related features.

Statistical Magnitude	Feature	Tasks					
		FF	SF	LG	VS	CF	RT
Mean Positive		7.100	25.150	42.450	67.300	84.450	60.050
Mean Negative	BN	10.056	24.278	50.278	79.333	96.556	58.388
p		0.908	0.973	0.758	0.636	0.634	0.948
F		0.014	0.001	0.095	0.224	0.227	0.004
Mean Positive		5.600	12.700	25.900	47.700	30.556	38.950
Mean Negative	BD	8.500	13.889	31.389	51.167	27.350	37.722
p		0.803	0.953	0.636	0.765	0.782	0.916
F		0.063	0.004	0.224	0.089	0.076	0.011

Table 2. *Cont.*

Statistical Magnitude	Feature	Tasks		
		FF	SF	LG
Mean Positive	PMAX Bright	482.818	526.209	550.694
Mean Negative		509.175	565.160	594.286
p		0.059	**0.039 ***	0.060
F		3.814	4.571	3.763
Mean Positive	PMAX Dark	488.442	540.845	543.968
Mean Negative		519.402	580.475	584.265
p		**0.023 ***	**0.046 ***	0.080
F		5.655	4.255	3.248
Mean Positive	PMIN Bright	461.943	423.206	422.970
Mean Negative		491.634	451.658	455.277
p		**0.029 ***	0.056	**0.046 ***
F		5.155	3.896	4.256
Mean Positive	PMIN Dark	466.053	414.568	408.962
Mean Negative		496.686	439.468	441.652
p		**0.024 ***	0.071	0.054
F		5.591	3.468	3.960
Mean Positive	PL Bright	1.217	10.484	13.666
Mean Negative		1.505	16.500	14.033
p		0.515	0.109	0.930
F		0.433	2.706	0.008
Mean Positive	PL Dark	1.332	16.969	42.384
Mean Negative		1.606	18.559	43.350
p		0.553	0.618	0.856
F		0.359	0.253	0.034
Mean Positive	PV Bright	0.028	3.674	−0.799
Mean Negative		0.231	4.061	−1.084
p		0.631	0.487	0.094
F		0.235	0.493	2.961
Mean Positive	PV Dark	−0.484	−4.342	0.914
Mean Negative		−1.222	−5.038	1.164
p		0.235	0.301	0.195
F		1.457	1.100	1.747

Table 3. Data analysis of eye movement features.

Statistical Magnitude	Feature	Tasks					
		FF	SF	LG	VS	CF	RT
Mean Positive	FN	36.950	33.150	85.800	177.800	36.950	205.100
Mean Negative		29.500	31.500	81.667	186.278	45.667	207.444
p		0.870	0.916	0.791	0.586	0.576	0.880
F		0.027	0.011	0.071	0.297	0.314	0.023
Mean Positive	FD	3005.930	2147.282	1919.841	861.563	1837.863	469.763
Mean Negative		2053.71	1950.482	1789.268	881.868	1533.476	817.670
p		0.655	0.715	0.809	0.970	0.573	0.519
F		0.200	0.133	0.059	0.001	0.319	0.417
Mean Positive	FR	0.848	0.759	0.761	0.650	0.733	0.557
Mean Negative		0.730	0.679	0.728	0.586	0.593	0.555
p		0.079	0.234	0.625	0.343	**0.038 ***	0.972
F		3.105	1.425	0.239	0.901	4.370	0.001
Mean Positive	SN	25.950	32.150	84.800	176.800	35.950	204.100
Mean Negative		28.500	30.500	80.667	185.278	44.667	206.444
p		0.870	0.916	0.791	0.586	0.576	0.880
F		0.027	0.011	0.071	0.297	0.314	0.023
Mean Positive	SD	78.667	94.899	84.509	237.651	86.738	400.718
Mean Negative		99.100	94.898	99.204	229.992	110.643	314.108
p		0.452	1.000	0.589	0.778	0.379	**0.002 ****
F		0.567	0.000	0.293	0.080	0.776	10.187

Table 4. Data analysis of task performance features.

Feature	Task	Mean Positive	Mean Negative	p	F
TFR	VS	0.466	0.385	0.069	3.508
TSL		264.483	289.306	0.519	0.423
TSR		0.868	0.714	**0.043***	4.406
CFR	FF	0.610	0.514	0.337	0.927
	SF	0.540	0.532	0.934	0.007
	LG	0.524	0.484	0.688	0.161

From the results of data analysis, among the pupil-related features, PMAX and PMIN showed significant differences in all tasks, PL in the bright stage of SF and PV in all stages of FF. In contrast, eye movement features performed poorly, with only FR in CF and SD in RT (reading test) showing significant differences. Among the task performance features, TSR showed significant differences. TFR, TSL, and CFR did not show significant differences.

The distribution of features with significant differences is shown in Figure 6. The asterisk is used in the tables and figures to indicate whether the data are statistically significant or not. One asterisk is used for $p < 0.05$ and two asterisks for $p < 0.01$.
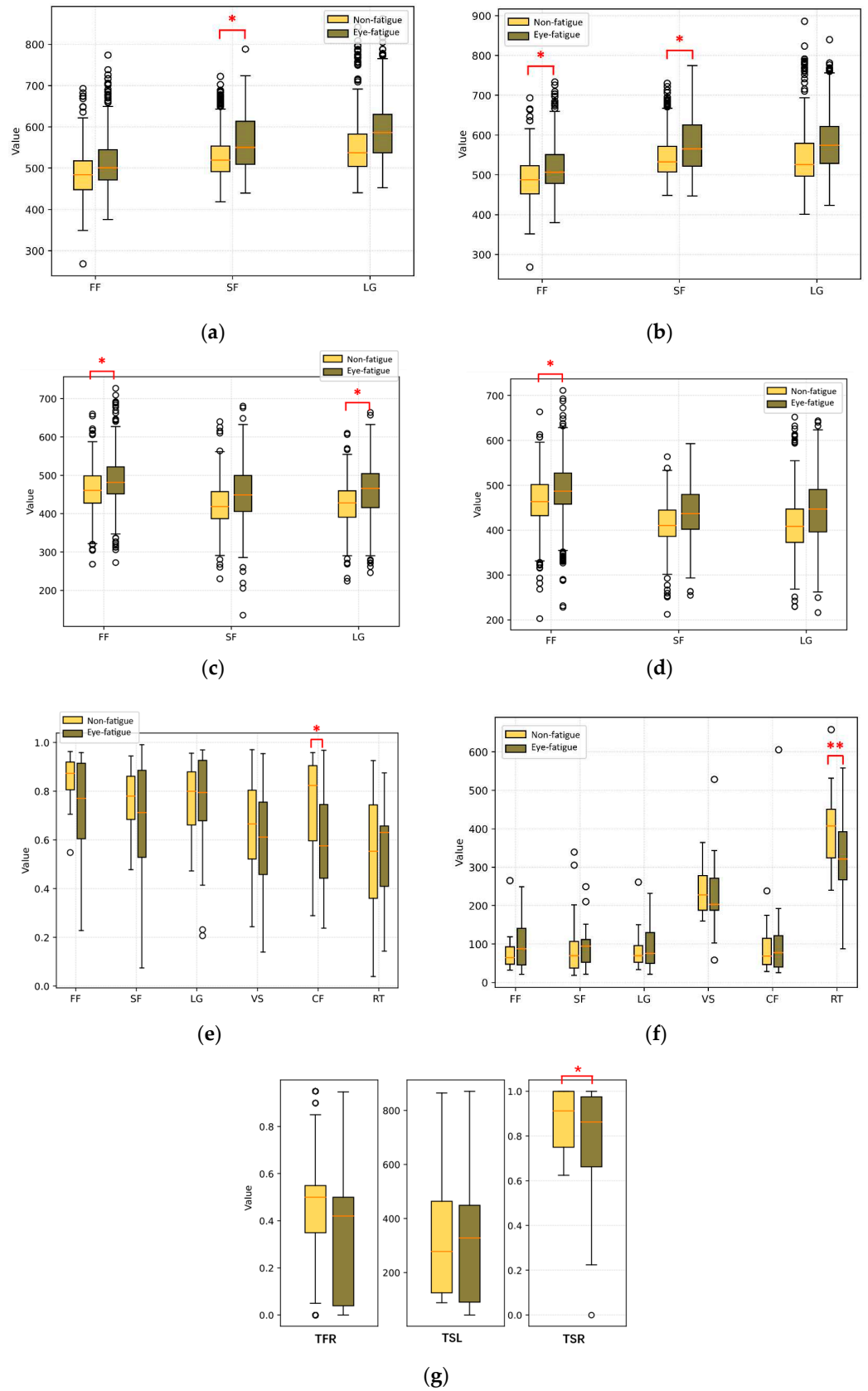
Figure 6. Distribution of features with significant differences: (**a**) PMAX-Bright; (**b**) PMAX-Dark; (**c**) PMIN-Bright; (**d**) PMIN-Dark; (**e**) FR; (**f**) SD; (**g**) TFR, TSL, TSR.

5.2. Detection Algorithm

Based on the statistical analyses, this study used eye movement features manually extracted from the raw data and an eye fatigue dataset constructed from the raw data to complete the eye fatigue detection task, respectively. In this study, we used scores from the eye fatigue questionnaire to assess the current level of eye fatigue in the subjects. In our overall research program, the part of the study described in this paper belongs to a more preliminary stage, and therefore we only performed a binary classification of whether fatigue was detected in the fatigue detection task and did not further discriminate the level of fatigue. As manually extracted eye movement features of raw signals are relatively readily comprehensible, this study adopted traditional machine learning classification models such as SVM, DT, RF, and ET for eye fatigue discrimination and adjusted the key parameters therein, such as the choice of kernel function in SVM and the classification strategy in DT. In the detection task in the eye fatigue dataset, this study optimized the structure of the algorithm several times, especially the number of network layers.

Since the ratio of positive and negative samples of the original data was approximate to 1:1 and the ratio of positive and negative samples of the eye fatigue dataset strictly conformed to 1:1, this study adopted the accuracy rate for the algorithm's performance evaluation after comprehensively considering the performance evaluation criteria of the classification algorithms., whose mathematical expression is shown in the following equation:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

TP (true positive): values that are positive and predicted as positive,
FP (false positive): values that are negative but predicted as positive,
FN (false negative): values that are positive but predicted as negative,
TN (true negative): values that are negative and predicted as negative.

The detection accuracies of the different models used in this study as well as the different parameters are shown in Table 5. Parameter "-" indicates that the optimal structure has been adjusted. Machine learning algorithms have significantly lower detection accuracy than deep learning algorithms. The best performance among the classical deep learning algorithms was the auto-encoder algorithm, with 82.4%, and its optimal network structure consisted of a five-layer encoder. Transformer was slightly lower than the auto-encoder, at 81.7%. The machine learning algorithm based on manually extracted features was significantly less effective in detection than the deep learning algorithm with automatically extracted features. The auto-encoder feature union algorithm, which incorporates eye-tracking features based on the optimal model, achieved 87.9% detection accuracy.

Table 5. Detection accuracy of different algorithms.

Algorithm	Parameter	Accuracy
SVM	Linear	57.6
	Poly	55.2
	Rbf	68.4
DT	ID3	51.8
	C4.5	54.1
	CART	57.8
RF	-	65.8
ET	-	63.5
CNN	3-convolutional layer	72.9
	5-convolutional layer	74.1
	10-convolutional layer	71.3

Table 5. *Cont.*

Algorithm	Parameter	Accuracy
AE	3-encoder layer	78.5
	5-encoder layer	82.4
	10-encoder layer	80.3
Transformer	3-encoder layer	79.3
	5-encoder layer	81.7
	10-encoder layer	80.4
AEFU	5-2-encoder layer	**87.9**

The variation of the training loss function for the auto-encoder algorithm and the auto-encoder feature union algorithm is shown in Figure 7. From the change in the loss function, both the convergence speed and the degree of overfitting of the model were improved after combining manually calculated eye features. The number of epochs needed to train the model from the beginning to near convergence dropped from 40 to 25.

(**a**)　　　　　　　　　　　　　　(**b**)
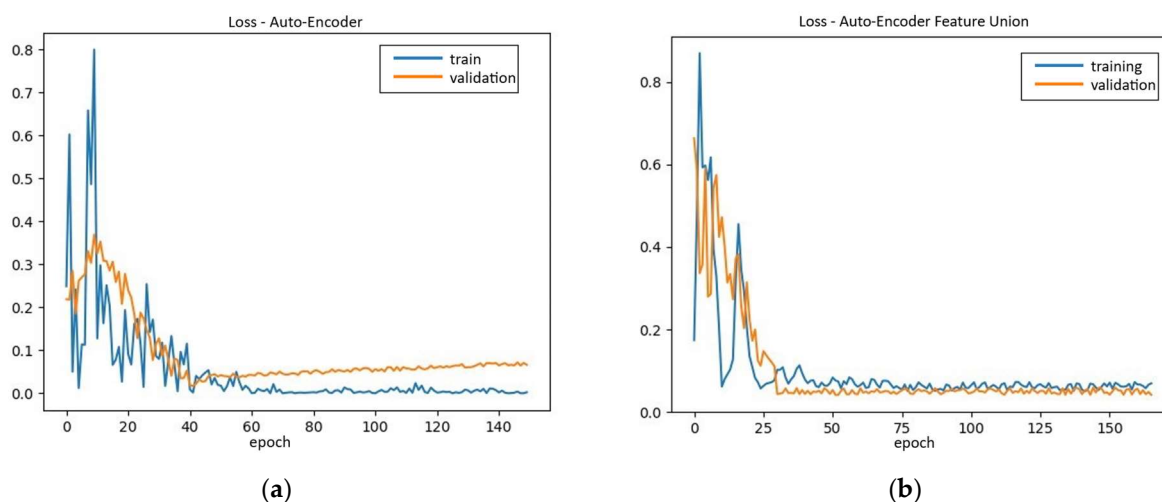
Figure 7. (**a**) Algorithm loss of auto-encoder; (**b**) algorithm loss of auto-encoder feature union algorithm.

6. Discussion

In general, this study verified the support of eye-tracking features for eye fatigue detection. For all the eye-tracking signal features, this study analyzed their response to the eye fatigue status and whether the difference between the non-fatigued and eye fatigue states was significant. In addition to considering the influence of the ocular damage mechanism of the eye fatigue on the features, this research also considered the influence of the experimental design and the subjects themselves on the experimental results and the significance of the differences in the features.

Pupil-related features performed better in the analysis of variance of repeated measures of variance, especially the maximum and minimum pupil sizes. In the data analysis, maximum pupil sizes showed a significant difference in the slow flicker task where the screen became brighter, as well as in the fast flicker and slow flicker tasks where the screen became darker. Meanwhile, minimum pupil sizes also showed significant differences in the screen-brightening in the fast flicker and luminance gradient tasks, as well as in the screen darkening in the fast flicker task. In addition, both features showed near-significant differences in other tasks and other situations, which perhaps related to the number of data. This phenomenon suggests that as eye fatigue increases, pupillary control by the pupillary sphincter and pupillary dilator muscles decreases significantly and the pupil dilates accordingly. Therefore, changes in pupil size is one of the reliable ocular motility

features of eye fatigue. In contrast, mean velocity of pupil changes and response time for the pupil to adjust to screen changes performed poorly and did not show significant differences. Overall, although there were no significant differences in some of the pupil-related features, they were still relatively well-performing eye movement features that can be used as discriminators of eye fatigue.

In contrast to the performance of pupil-related features, the reliability of fixation-related features was poorer. Only the percentage of fixation duration in the color flicker task and the mean saccade distance in the reading task showed significant differences. Our explanation for this phenomenon is that because eye movement behavior is a direct reflection of subjects' attention, eye movement behavior is seriously influenced by subjects' subjective consciousness. The problem of eye fatigue is inherently a problem that may be masked by the subjective fatigue feelings of the human brain, and fixation-related features that were seriously affected by subjective influences did not show a good response effect. Although most of the features related to fixation showed no significance difference, the mean saccade distance showed significant differences in the reading task. The average saccade distance, which is an oculomotor feature that represents the sweeping distance of the subjects in reading, can objectively reflect the current speed of information reception of the subjects to a certain extent. As the level of eye fatigue increases, subjects' ability to receive information becomes weaker, so it is difficult for them to maintain long-distance saccade at the same level of information reception. The subjects' number of fixations was significantly lower in the color flicker task. This result demonstrates that there was no significant difference in the performance of this feature in relatively mild stimuli. However, fixation behavior was difficult to maintain, and the difference was instead significant for more intense stimuli. We conjecture that the performance of task-related features is largely influenced by subjects' subjective awareness. In other words, subjects' performance of task features is largely a reflection of their subjective awareness rather than an objective manifestation of eye strain. For this issue, further analysis of the performance of eye movement behavioral features in eye fatigue detection experiments is needed. It is important to verify whether the influence of subjective factors is real by modifying the experimental paradigm and other methods. Among the task features, only the TSR feature showed significant variability, and the rest of the features did not show significant differences. This phenomenon seems to prove from another perspective that eye movement behavior is largely influenced by subjective factors of the subjects.

The proportion of subjects' attention to the target, reaction time, and accuracy decreased to different degrees as fatigue increased, which was in line with the experimental expectation. It still did not show significant changes under the influence of subjective awareness. Consideration should be given to reducing the influence of subjects' subjective factors from the perspective of experimental design in subsequent studies.

In the exploration of eye fatigue detection algorithms, the excellent performance of the auto-encoder among many algorithms is surprising, which suggests that eye-tracking signals are suitable for feature extraction using the auto-encoder. In addition, the transformer is an upgraded version of the auto-encoder, but the detection performance was slightly lower than that of the auto-encoder algorithm in this study, which may be related to the fact that eye-tracking data possess a strict temporal order. Since the time of acquisition and the time of occurrence of the experimental stimuli strictly correspond to each other, the position encoding in the transformer instead resulted in information redundancy, which affected the classification results. This point needs to be explored in depth in subsequent studies.

It can be seen from the changes in the loss function curves that the addition of eye-tracking features not only accelerated the model learning efficiency and convergence speed but also solved the overfitting problem to a certain extent. The enhancement effect of the AEF model proves that the features extracted from eye movement signals were still of some guiding significance and practical value. However, due to the relatively small number of subjects collected in this study, the setting of the experimental task was quite specific and the results obtained in this study did not have high generalization performance.

Although this paper achieved better results during the research process, its generalizability for eye fatigue detection in daily use environments needs to be further considered. In this paper, the eye movement features in the experimental scenario were used for eye fatigue detection, and the real-time performance of the detection model also needs to be optimized. In addition, in the process of feature fusion, this study only directly connected the eye movement feature data with the implicit features of the self-encoder, and the method of feature fusion also needs to be studied.

7. Conclusions

This study focuses on the eye fatigue detection system and discrimination algorithm under the environment of daily electronic screen use. Aiming at the physiological factors that can directly produce and potentially affect the problem of eye fatigue, the feasibility of eye fatigue detection through eye-tracking signals captured by video is explored through analyses of the features extracted from the eye movement signals by using statistical methods. The supporting effect of different eye movement features on eye fatigue assessment was also discussed. In addition, this study explores the establishment of eye fatigue detection algorithms based on auto feature extraction and eye movement features and based on different feature extraction methods. Various traditional machine learning models and deep learning algorithms are investigated, and an auto-encoder feature union algorithm based on the feature fusion coding model is proposed, which realizes a fast and high-precision eye fatigue detection method. In the future research program, a total of two aspects are envisioned for this study. On the one hand, based on the research of this paper, the performance of eye-tracking features in the eye fatigue detection task will be further explored by changing the feature calculation method and optimizing the experimental setting. On the other hand, this study only uses a simple link method for feature fusion, so it is necessary to conduct in-depth research on feature fusion methods and feature selection proportion.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Blehm, C.; Vishnu, S.; Khattak, A.; Mitra, S.; Richard, W.Y. Computer vision syndrome: A review. *Surv. Ophthalmol.* **2005**, *50*, 253–262. [CrossRef] [PubMed]
2. American Optometric Association. Computer Vision Syndrome. 2017. Available online: http://www.aoa.org/patients-and-public/caring-for-your-vision/protecting-your-vision/computer-vision-syndrome?ss0=y (accessed on 17 March 2024).
3. Smith, W.J. A review of literature relating to visual fatigue. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **1979**, *23*, 362–366.
4. Portello, J.K.; Rosenfield, M.; Bababekova, Y.; Estrada, J.M.; Leon, A. Computer-related visual symptoms in office workers. *Ophthalmic Physiol. Opt.* **2012**, *32*, 375–382. [CrossRef] [PubMed]
5. Sheppard, A.L.; Wolffsohn, J.S. Digital eye strain: Prevalence, measurement and amelioration. *BMJ Open Ophthalmol.* **2018**, *3*, e000146. [CrossRef] [PubMed]
6. Hirota, M.; Uozato, H.; Kawamorita, T.; Shibata, Y.; Yamamoto, S. Effect of incomplete blinking on tear film stability. *Optom. Vis. Sci.* **2013**, *90*, 650–657. [CrossRef] [PubMed]
7. Himebaugh, N.L.; Begley, C.G.; Bradley, A.; Wilkinson, J.A. Blinking and tear break-up during four visual tasks. *Optom. Vis. Sci.* **2009**, *86*, E106–E114. [CrossRef] [PubMed]
8. Kim, D.; Sunghwan, C.; Sangil, P.; Kwanghoon, S. Stereoscopic visual fatigue measurement based on fusional response curve and eye-blinks. In Proceedings of the 2011 17th International Conference on Digital Signal Processing (DSP), Corfu, Greece, 6–8 July 2011.

9.   Wang, Y.; Zhang, L.; Fang, Z. Eye fatigue detection through machine learning based on single channel electrooculography. *Algorithms* **2022**, *15*, 84. [CrossRef]

10.   Gao, X.-Y.; Zhang, Y.-F.; Zheng, W.-L.; Lu, B.-L. Evaluating driving fatigue detection algorithms using eye tracking glasses. In Proceedings of the 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), Montpellier, France, 22–24 April 2015.

11.   Zhang, Y.-F.; Gao, X.-Y.; Zhu, J.-Y.; Zheng, W.-L.; Lu, B.-L. A novel approach to driving fatigue detection using forehead EOG. In Proceedings of the 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), Montpellier, France, 22–24 April 2015.

12.   Tian, C.J. Fatigue driving detection based on electrooculography: A review. *EURASIP J. Image Video Process.* **2021**, *2021*, 33. [CrossRef]

13.   Bang, J.W.; Choi, J.-S.; Heo, H.; Park, K.R. A fuzzy-based fusion method of multimodal sensor-based measurements for the quantitative evaluation of eye fatigue on 3D displays. *Sensors* **2015**, *15*, 10825–10851. [CrossRef] [PubMed]

14.   Ileri, R.; Latifoglu, F.; Demirci, E. A novel approach for detection of dyslexia using convolutional neural network with EOG signals. *Med. Biol. Eng. Comput.* **2022**, *60*, 3041–3055. [CrossRef]

15.   Kołodziej, M.; Tarnowski, P.; Sawicki, D.J.; Majkowski, A.; Rak, R.J.; Bala, A.; Pluta, A. Fatigue detection caused by office work with the use of EOG signal. *IEEE Sens. J.* **2020**, *20*, 15213–15223. [CrossRef]

16.   So, G.J.; Kim, S.H.; Kim, J.Y. A Study of Testing Tools for the Fatigue Caused by the Stereoscopic Video Considering Temporal and Spatial Characteristics. *Int. J. Comput. Theory Eng.* **2013**, *5*, 712. [CrossRef]

17.   Manabe, H.; Fukumoto, M.; Yagi, T. Direct gaze estimation based on nonlinearity of EOG. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1553–1562. [CrossRef]

18.   Murata, K.; Araki, S.; Yokoyama, K.; Yamashita, K.; Okumatsu, T.; Sakou, S. Accumulation of VDT work-related visual fatigue assessed by visual evoked potential, near point distance and critical flicker fusion. *Ind. Health* **1996**, *34*, 61–69. [CrossRef] [PubMed]

19.   Chen, C.; Wang, J.; Li, K.; Wu, Q.; Wang, H.; Qian, Z.; Gu, N. Assessment visual fatigue of watching 3DTV using EEG power spectral parameters. *Displays* **2014**, *35*, 266–272. [CrossRef]

20.   Zheng, W.-L.; Lu, B.-L. A multimodal approach to estimating vigilance using EEG and forehead EOG. *J. Neural Eng.* **2017**, *14*, 026017. [CrossRef]

21.   Wang, H.; Wu, C.; Li, T.; He, Y.; Chen, P.; Bezerianos, A. Driving fatigue classification based on fusion entropy analysis combining EOG and EEG. *IEEE Access* **2019**, *7*, 61975–61986. [CrossRef]

22.   Huo, X.-Q.; Zheng, W.-L.; Lu, B.-L. Driving fatigue detection with fusion of EEG and forehead EOG. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016.

23.   Kalyani, P.S.; Holland, G.N.; Fawzi, A.A.; Arantes, T.E.F.; Yu, F.; Sadun, A.A. Association between retinal nerve fiber layer thickness and abnormalities of vision in people with human immunodeficiency virus infection. *Am. J. Ophthalmol.* **2012**, *153*, 734–742.e731. [CrossRef] [PubMed]

24.   Peng, Y.; Wong, C.M.; Wang, Z.; Rosa, A.C.; Wang, H.T.; Wan, F. Fatigue detection in SSVEP-BCIs based on wavelet entropy of EEG. *IEEE Access* **2012**, *9*, 114905–114913. [CrossRef]

25.   Mun, S.; Park, M.-C.; Park, S.; Wang, M. SSVEP and ERP measurement of cognitive fatigue caused by stereoscopic 3D. *Neurosci. Lett.* **2012**, *525*, 89–94. [CrossRef]

26.   Mahmood, H.R.; Irshayyid, M.S.; Hussien, M.K.; Al-Shammari, B.K.J. Measuring the Fatigue of Persons Based on EOG Signal Processing Mmethods. In Proceedings of the 2022 5th International Conference on Engineering Technology and Its Applications (IICETA), AI-Najaf, Iraq, 31 May–1 June 2022.

27.   Němcová, A.; Janoušek, O.; Vítek, M.; Provazník, I. Testing of features for fatigue detection in EOG. *Bio-Med. Mater. Eng.* **2017**, *28*, 379–392. [CrossRef]

28.   Kan, J.; Li, A.; Zou, H.; Chen, L.; Du, J. A machine learning based dose prediction of lutein supplements for individuals with eye fatigue. *Front. Nutr.* **2020**, *7*, 577923. [CrossRef] [PubMed]

29.   Souchet, A.D.; Philippe, S.; Lourdeaux, D.; Leroy, L. Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. *Int. J. Hum.–Comput. Interact.* **2022**, *38*, 801–824. [CrossRef]

30.   Zargari Marandi, R.; Madeleine, P.; Omland, Ø.; Vuillerme, N.; Samani, A. Eye movement characteristics reflected fatigue development in both young and elderly individuals. *Sci. Rep.* **2018**, *8*, 13148. [CrossRef] [PubMed]

31.   Li, J.; Li, H.; Umer, W.; Wang, H.; Xing, X.; Zhao, S.; Hou, J. Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. *Autom. Constr.* **2020**, *109*, 103000. [CrossRef]

32.   Benedetto, S.; Drai-Zerbib, V.; Pedrotti, M.; Tissier, G.; Baccino, T. E-readers and visual fatigue. *PLoS ONE* **2013**, *8*, e83676. [CrossRef] [PubMed]

33.   Weifeng, S.; Yuqi, W.; Bingliang, H.; Quan, W. Exploring the Connection between Eye Movement Parameters and Eye Fatigue. In Proceedings of the 2023 The 4th Asia Conference on Computers and Communications (ACCC 2023), Shanghai, China, 15–17 December 2023.

34.   Wang, Q.; Bolhuis, J.; Rothkopf, C.A.; Kolling, T.; Knopf, M.; Triesch, J. Infants in control: Rapid anticipation of action outcomes in a gaze-contingent paradigm. *PLoS ONE* **2012**, *7*, e30884. [CrossRef] [PubMed]

35. Valliappan, N.; Dai, N.; Steinberg, E.; He, J.; Rogers, K.; Ramachandran, V.; Xu, P.; Shojaeizadeh, M.; Guo, L.; Kohlhoff, K.; et al. Accelerating eye movement research via accurate and affordable smartphone eye tracking. *Nat. Commun.* **2020**, *11*, 4553. [CrossRef] [PubMed]

36. Hickson, S.; Dufour, N.; Sud, A.; Kwatra, V.; Essa, I. Eyemotion: Classifying facial expressions in VR using eye-tracking cameras. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, 7–11 January 2019.

37. China Video Industry Network. Testing and Evaluation Method of Vision Fatigue Part 2, Scale Evaluation Method. 2019. Available online: http://www.cvianet.org.cn/download/?page=2 (accessed on 17 March 2024).

38. Alamia, A.; VanRullen, R.; Pasqualotto, E.; Mouraux, A.; Zenon, A. Pupil-linked arousal responds to unconscious surprisal. *J. Neurosci.* **2019**, *39*, 5369–5376. [CrossRef]

39. Van der Stigchel, S.; Rommelse, N.N.J.; Deijen, J.B.; Geldof, C.J.A.; Witlox, J.; Oosterlaan, J.; Sergeant, J.A.; Theeuwes, J. Oculomotor capture in ADHD. *Cogn. Neuropsychol.* **2007**, *24*, 535–549. [CrossRef]