

# Accuracy and Response Speed of Eye Center Annotation Using Eye Movement Models: Validating the Effectiveness of Eyesight Detection

Xinzhe An\*, Xiaofan Xu, Zhenwei Ye

Jinan University, Guangzhou, China  
Email: \*woyagou60@163.com

**How to cite this paper:** An, X.Z., Xu, X.F. and Ye, Z.W. (2026) Accuracy and Response Speed of Eye Center Annotation Using Eye Movement Models: Validating the Effectiveness of Eyesight Detection. *Open Journal of Applied Sciences*, 16, 584-592.

<https://doi.org/10.4236/ojapps.2026.162036>

**Received:** January 19, 2026

**Accepted:** February 11, 2026

**Published:** February 14, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Eye center annotation is vital for ophthalmic diagnostics and surgery. However, existing algorithms often require specialized equipment and face challenges in real-time performance, particularly under varying lighting. This study evaluates four widely used facial landmarking algorithms—Mediapipe, Dlib, Haar Cascade, and RetinaFace—in the task of eye iris center annotation. The optimal algorithm is employed to validate the effectiveness in optokinetic nystagmus (OKN) detection and eyesight assessment. The results demonstrate that Mediapipe outperforms the other algorithms, offering superior real-time performance, high accuracy, and robust adaptability to different lighting conditions. Additionally, this study validates its potential in eyesight detection.

## Keywords

Eye Center Annotation, Mediapipe, Dlib, Haar Cascade, RetinaFace, Accuracy, Eyesight

## 1. Introduction

Eye iris center annotation holds significant value in ophthalmic diagnostics and surgery [1]. Accurate real-time eye annotation not only supports the early diagnosis, long-term monitoring, and auxiliary treatment of ophthalmic diseases but also provides essential support for certain oculomotor research. For instance, in the early screening of amblyopia in children, precise annotation of the eye center can detect subtle eye tremors, enabling early detection and intervention.

Optokinetic Nystagmus (OKN) is a natural reflexive eye movement in oculo-

motor studies, reflecting the health status of the visual system. Through accurate eye center annotation, physicians can observe the minute variations in eye tremors, allowing for early detection of abnormalities and providing a basis for subsequent treatment.

Eyesight refers to the ability of the human eye to distinguish the minimum distance between two points, which reflects the ability of the fovea centralis to resolve the minimum spacing between two points. It is usually measured by the minimum angle of resolution (MAR), which can be converted into logarithmic visual acuity (LogMAR) for quantitative comparison [2]. Eyesight examination is an important aspect of ophthalmic examinations, helping doctors evaluate the eye health of patients. Traditional eyesight detection methods are mainly subjective, using visual acuity charts such as the Snellen chart and E-chart [3] [4]. Although widely used [5], they rely on the patient's language ability and active cooperation [6], leading to an error rate of up to 30% in infants and young children, individuals with intellectual disabilities, and uncooperative adults (such as malingerers). In addition, affected by factors such as letter spacing and chart lighting, the accuracy and repeatability of traditional subjective visual acuity test results are often questioned.

Existing studies have shown a correlation between OKN and eyesight [7].

Mediapipe, developed by Google, is a cross-platform framework that provides efficient facial landmarking and other computer vision tasks [8]. It uses deep learning models to achieve high-precision facial feature point detection [9] [10]. Dlib is a widely used open-source library that provides facial landmarking, face recognition, and face detection capabilities. It uses machine learning-based methods for facial feature point detection [11]. Haar Cascade is a traditional computer vision method provided by OpenCV, widely used for face detection. It uses Haar features and a cascade classifier for object detection [11]. RetinaFace is a deep learning-based facial detection method that uses an efficient Convolutional Neural Network (CNN) for face detection [12].

## 2. Objective

This study aims to comprehensively compare and analyze the accuracy and response time of four widely used facial landmarking algorithms—Mediapipe, Dlib, Haar Cascade, and RetinaFace—in eye iris center annotation tasks. Additionally, it intends to establish an objective eyesight detection method based on collecting Optokinetic Nystagmus (OKN) responses and explore its application value in the adult population.

## 3. Methodology

### 3.1. Self-Collected Dataset

This study uses a dataset of eye images that includes a variety of ages, genders, and lighting conditions. Each image in the dataset is annotated with the true eye center position, which serves as the ground truth for comparing the algorithm's annota-

tion results.

### 3.2. Algorithm Selection and Experimental Setup

We selected four facial landmarking algorithms for the experimental tests: Mediapipe, Dlib, Haar Cascade, and RetinaFace. Custom programs were written to conduct the tests and collect results.

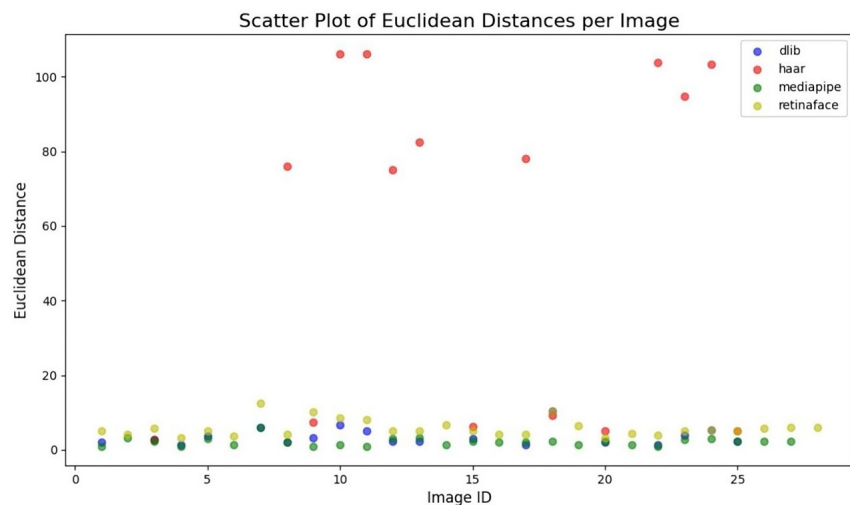
### 3.3. Evaluation Metrics

The accuracy is quantified by calculating the Euclidean distance, Mean Squared Error (MSE), and Mean Absolute Error (MAE) to measure the difference between the algorithm's annotation and the true eye center position. The real-time processing capability of each algorithm is measured by the frames per second (FPS), assessing its performance in real-time video streams. The detection rate is calculated by dividing the number of successfully detected images by the total number of images. Collection and correlation verification of OKN signals and eyesight: First, use the optimal algorithm obtained from the comparison to annotate the eye center and extract OKN signals; then, pair the OKN signals with the subjective eyesight test results (Snellen visual acuity chart) of the corresponding subjects; finally, input the paired data into different machine learning models for training and verification to explore the correlation between OKN signals and eyesight.

## 4. Experimental Results and Analysis

This section presents the experimental results of the four facial landmarking algorithms—Mediapipe, Dlib, Haar Cascade, and RetinaFace—on the eye center annotation task, and provides a detailed analysis of their accuracy and response time.

### 4.1. Annotation Accuracy



**Figure 1.** Scatter plot of Euclidean distances of each algorithm per image.

By analyzing the experimental data, we evaluated the accuracy of the four algorithms in eye center annotation. The following are the specific results based on various charts and evaluation metrics:

The scatter plot (Figure 1) intuitively displays the annotation error for each algorithm on different images. The distance values for Haar Cascade show significant fluctuations, especially on certain images where the annotation error is notably higher than that of the other models. This indicates that Haar Cascade performs inconsistently when handling changes in facial angles. In contrast, Dlib, RetinaFace, and Mediapipe show more stable annotation errors, with their distributions being relatively close to each other.

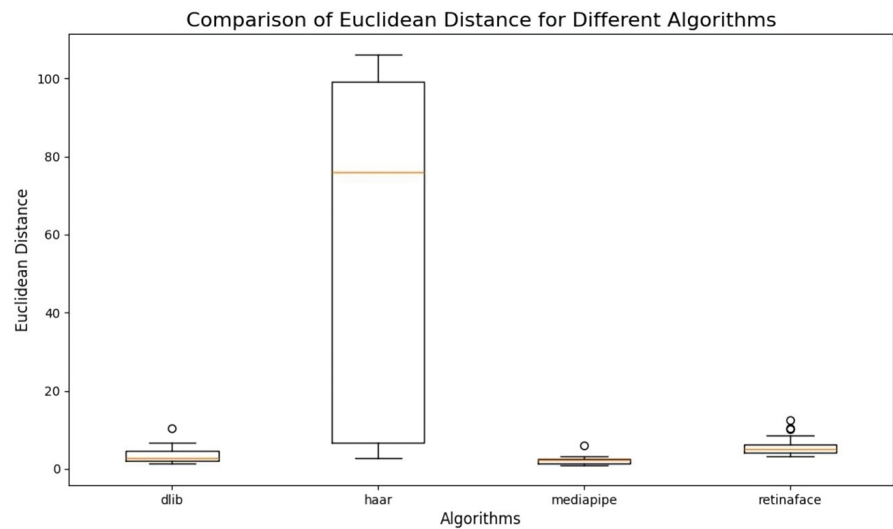


Figure 2. Box plot of Euclidean distances of each algorithm.

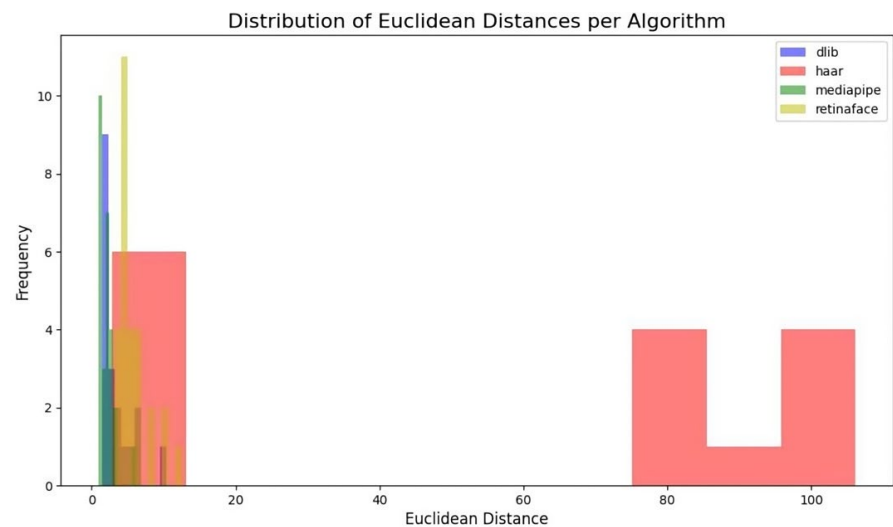


Figure 3. Histogram of Euclidean distance distribution of each algorithm.

The box plot (Figure 2) further confirms the instability of Haar Cascade in terms of annotation accuracy. Haar Cascade exhibits the largest fluctuation range in an-

notation errors, with the median of its distance values being higher than the other models, indicating poor robustness in eye annotation. In contrast, Dlib and Mediapipe show lower median errors with narrower error distribution ranges, validating their superior accuracy. RetinaFace ranks just behind these two algorithms.

The histogram (Figure 3) shows the distribution of Euclidean distances for different algorithms. Haar Cascade’s distance values exhibit a bimodal distribution, indicating significant bias and instability in its annotation results. In contrast, Dlib, RetinaFace, and Mediapipe have most of their distance values concentrated within a smaller range, validating their consistency in annotation accuracy.

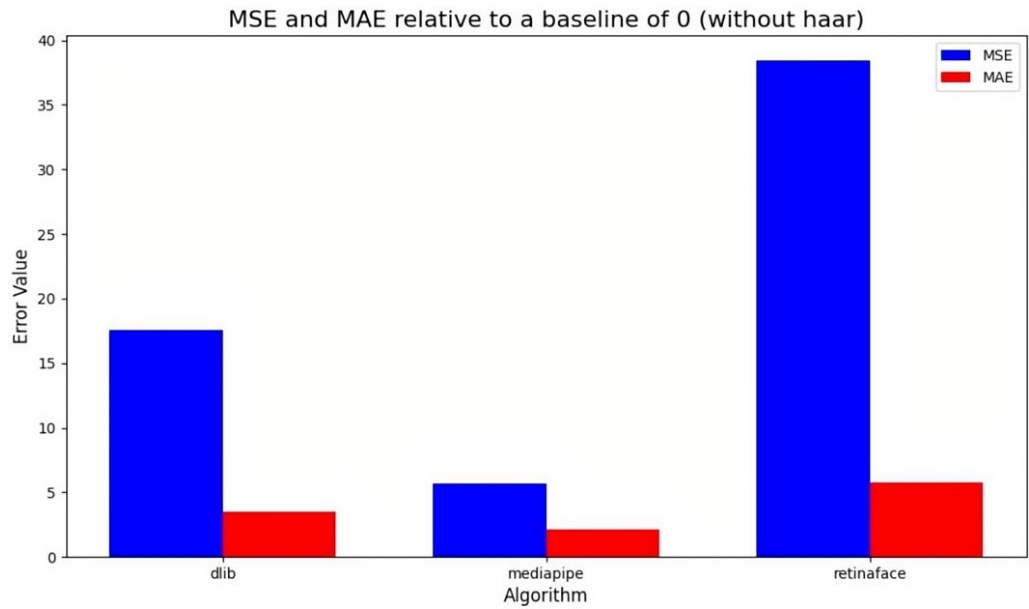


Figure 4. Bar Chart of MSE and MAE of Dlib, RetinaFace, and Mediapipe.

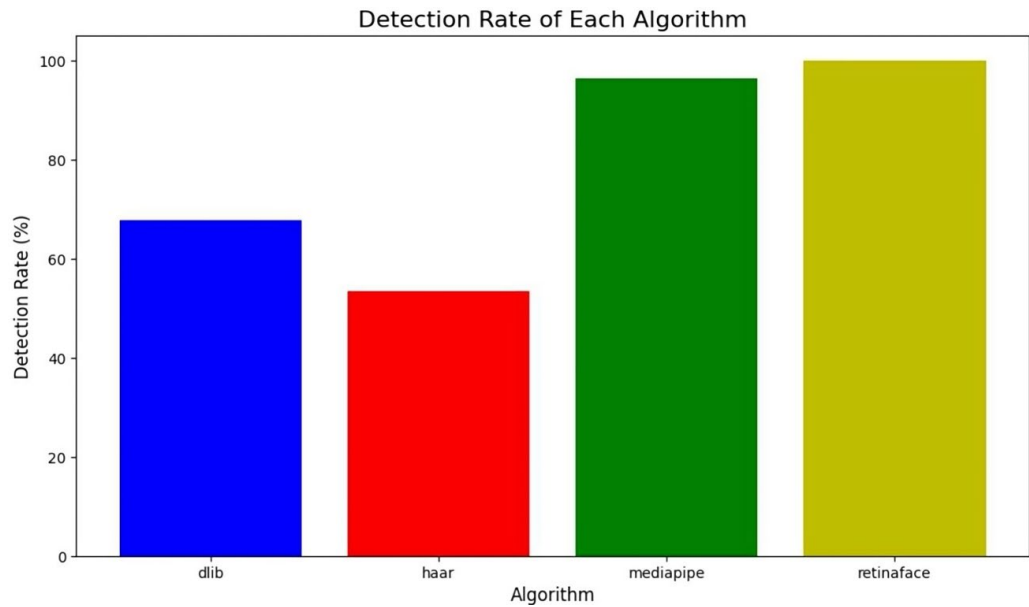


Figure 5. Bar chart of detection rate of each Algorithm.

Due to Haar Cascade's larger errors in eye annotation, we compare the MSE (Mean Squared Error) and MAE (Mean Absolute Error) bar charts only for the three models: Dlib, RetinaFace, and Mediapipe. The bar charts (Figure 4) reveal that RetinaFace's error values are significantly higher than those of the other models, further highlighting its disadvantage in accuracy. In contrast, Mediapipe shows the lowest error values, demonstrating its superior performance in eye center annotation.

The detection rate bar chart (Figure 5) shows that both Mediapipe and RetinaFace achieved a detection rate of 100%, demonstrating excellent performance. In contrast, Haar Cascade had the lowest detection rate, only 50%. This result further confirms Haar Cascade's poor performance in complex scenarios. Dlib's success rate was also below 80%, while Mediapipe and RetinaFace were able to consistently complete the eye annotation task.

**Table 1.** Statistical table of Euclidean distance under different lighting conditions.

	Mean Load Time (s)	Mean Detect Time (s)	Mean Total Time (s)	Std Load Time (s)	Std Detect Time (s)	Std Total Time (s)	Frame Rate Detect (fps)
Dlib	0.0031	0.037	0.040	0.0038	0.021	0.022	27.06
haar	0.0036	0.038	0.041	0.0017	0.020	0.022	26.54
mediapipe	0.0027	0.0056	0.0083	0.0033	0.0024	0.0054	177.08
retinaface	0.0038	3.01	3.0	0.0019	0.89	0.89	0.33

For each algorithm, we calculated its loading time, detection time, and total processing time, and analyzed the frames per second (FPS) as a measure of response speed. The following are the detailed statistics: From the table and chart (Table 1), it can be seen that all four models perform well in terms of loading time. Mediapipe shows the best FPS performance, reaching 177.08 FPS, significantly higher than the other algorithms, making it suitable for real-time annotation. Haar Cascade and Dlib achieve frame rates of 26.54 and 27.07 FPS, respectively. While they can handle typical real-time tasks, their performance seems insufficient for rapid eye movement tracking. RetinaFace, with an FPS of only 0.33, has an extremely slow response time and is unsuitable for annotation tasks in real-time video streams.

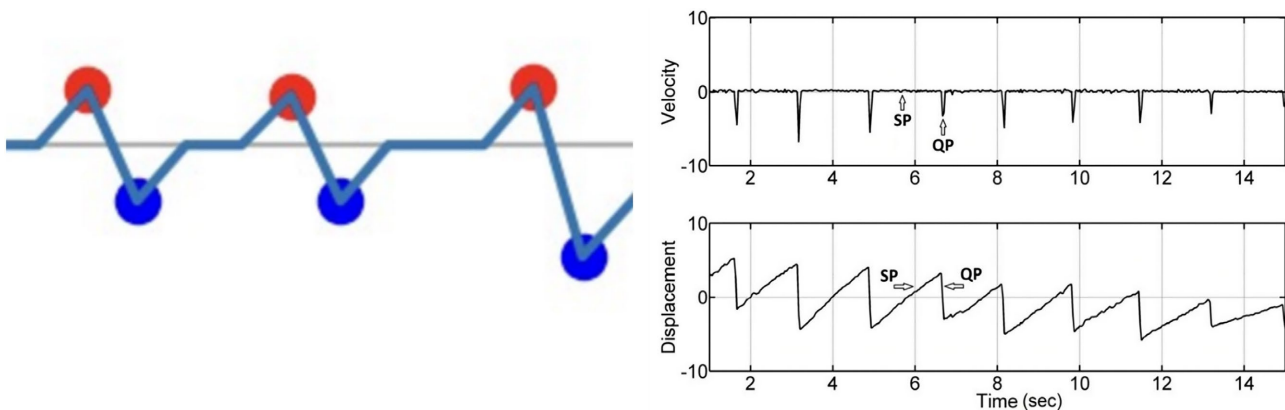
## 4.2. Further Experiments

To further assess Mediapipe's performance, we manually adjusted the brightness and darkness of images to verify the algorithm's adaptability under different lighting conditions. The test results showed that Mediapipe performed with higher accuracy under brightened images, while the accuracy was slightly lower under darkened images. By using multi-threaded processing and Haar Cascade ROI calibration, Mediapipe achieved a detection rate of 96.43% on the adjusted dataset.

**Table 2.** Statistical table of Euclidean distance under different lighting conditions.

	mean	std	min	25%	50%	75%	max
Normal vs Dark	1.15	1.13	0	0	1	1.41	4.12
Normal vs Bright	0.85	0.678	0	0	1	1.19	2.03

Additionally, we conducted repeatability tests on normal, brightened, and darkened images. The results (Table 2) showed a 100% repeatability rate across all three tests. This indicates that Mediapipe performs consistently under different processing conditions, with all three test results achieving a 100% repeatability rate, further proving the robustness and consistency of Mediapipe.



**Figure 6.** OCN waveform.

In real-time detection of dynamic video streams, we also conducted real-time annotation tests on Mediapipe. The results showed that it provided stable annotation results and produced a standard OCN waveform (Figure 6), demonstrating its feasibility for ophthalmic applications (Table 3).

**Table 3.** Evaluation results of machine learning models for eyesight detection.

Model	Mean Squared Error (MSE)	Mean Absolute Error (MAE)
Regression Tree	0.043	0.139
Random Forest Regression	0.042	0.141
Support Vector Machine Regression	0.055	0.162
KNN Regression	0.056	0.171

### 5. Discussion

This study compared four widely used facial landmarking algorithms—Mediapipe, Dlib, Haar Cascade, and RetinaFace—assessing their accuracy and response time in eye iris center annotation tasks. Mediapipe’s core strengths lie in outstanding real-time processing, efficient facial feature annotation, and strong robustness

under varying lighting conditions; integrating deep learning with hardware acceleration, it delivers high-precision, low-latency eye annotation while maintaining high FPS in dynamic video streams, which is crucial for long-term ophthalmic home monitoring, and it balances accuracy, speed and low hardware resource demands, though its detection rate is not 100%, calling for future optimization to cut computational overhead and boost performance on resource-constrained devices. This study also has certain limitations. First, the dataset does not include samples of patients with ophthalmic diseases, and the applicability of the algorithm in patients with eye diseases needs to be further verified. Second, the algorithm's performance in occlusion scenarios (such as wearing glasses, squinting, and eye closure) is not tested, and future research should supplement relevant experiments. Third, OKN signal collection may be interfered by eye movement artifacts, and more effective signal preprocessing methods need to be explored to improve signal quality.

## 6. Outlook and Future Work

This study experimentally compared the performance of four facial landmarking algorithms—Mediapipe, Dlib, Haar Cascade, and RetinaFace—in eye center annotation tasks, evaluating their accuracy, response time, and robustness. Nevertheless, Mediapipe still has room for improvement, particularly in terms of robustness in complex environments and computational resource consumption. Therefore, future research could focus on improving and expanding the algorithm in the following areas:

Future work could incorporate the Multi-Task Learning (MTL) framework to jointly optimize facial feature annotation and eye center annotation tasks. By sharing parts of the network layers and feature representations, the algorithm can simultaneously improve the performance of multiple related tasks. Enhancing Detection of Other Key Information While Processing Eye Annotation While processing eye annotation, enhancing the ability to detect other key information will further improve the comprehensiveness and accuracy of ophthalmic diagnostic systems.

Also, expand the dataset to include samples of patients with various ophthalmic diseases, and conduct more in-depth research on the correlation between OKN signals and eyesight, so as to further improve the effectiveness of Mediapipe in eyesight detection and promote its clinical application [13].

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Zhang, Y. and Li, X. (2020) Face Detection Using Deep Learning: A Survey. *Computer Vision and Image Understanding*, **191**, Article 102871.
- [2] Falkenstein, I.A., Cochran, D.E., Azen, S.P., Dustin, L., Tammewar, A.M., Kozak, I.,

- et al. (2008) Comparison of Visual Acuity in Macular Degeneration Patients Measured with Snellen and Early Treatment Diabetic Retinopathy Study Charts. *Ophthalmology*, **115**, 319-323. <https://doi.org/10.1016/j.ophtha.2007.05.028>
- [3] Suh, D.W. and Shahraki, K. (2023) Vision Screening Claims for Young Children in the United States. *Pediatrics*, **152**, e2023062804. <https://doi.org/10.1542/peds.2023-062804>
- [4] Ambrosino, C., Dai, X., Antonio Aguirre, B. and Collins, M.E. (2023) Pediatric and School-Age Vision Screening in the United States: Rationale, Components, and Future Directions. *Children*, **10**, Article 490. <https://doi.org/10.3390/children10030490>
- [5] Bailey, I.L. and Lovie-Kitchin, J.E. (2013) Visual Acuity Testing. from the Laboratory to the Clinic. *Vision Research*, **90**, 2-9. <https://doi.org/10.1016/j.visres.2013.05.004>
- [6] US Preventive Services Task Force (2017) Vision Screening in Children Aged 6 Months to 5 Years: US Preventive Services Task Force Recommendation Statement. *Journal of the American Medical Association*, **318**, 836-844.
- [7] Garcia, F. and Soto, R. (2021) Enhancements of Mediapipe for Real-Time Eye Tracking and Gaze Estimation. *Journal of Computer Vision*, **59**, 129-142.
- [8] Liao, M. and Wang, H. (2019) Efficient Real-Time Eye Tracking Using Haar Cascades and Deep Learning. *Vision Technology*, **52**, 1124-1135.
- [9] Gupta, S. and Roy, D. (2020) Real-Time Multi-Face and Eye Detection with Dlib and Open CV. In: *Proceedings of the International Conference on Computer Vision*, Springer, 45-50.
- [10] Wu, P. and Zhang, H. (2021) Retina Face: A Practical Single-Stage Dense Face Localization in the Wild.
- [11] Aigbe, S. and Zhang, Z. (2022) Improving Eye Center Annotation Accuracy in Real-Time Systems Using Mediapipe. *Journal of Machine Learning Research*, **23**, 111-123.
- [12] King, D.E. (2009) Dlib-ML: A Machine Learning Toolkit. *Journal of Artificial Intelligence Research*, **2**, 1-6.
- [13] Sahoo, B. and Li, L. (2020) Challenges and Improvements in Facial Landmark Detection for Robust Eye Center Annotation. *IEEE Transactions on Image Processing*, **29**, 7845-7857.