2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 5-8, 2025. Vienna, Austria

Ocular Feature Extraction for Eye Movement Analysis and Neurological Dysfunction Diagnosis

Ziqi Wang[1], Jing Bi[2], Xiaomeng Zhao[†3], Junqi Zhang[2], Jiahui Zhai[2], Hongyao Ma[2], Jinglei Cui[2], Rong Cui[2], Zhipeng Zheng[2], Yuanchen Tang[4], Jiantao Liang[4] and Kai Zhou[3]

[1]School of Software Technology, Zhejiang University, Ningbo 315100, China
[2]College of Computer Science, Beijing University of Technology, Beijing 100124, China
[3]Beijing Neurorient Technology Co., Ltd., Beijing 100000, China
[4]Department of Neurosurgery, Xuanwu Hospital, Capital Medical University, Beijing 100053, China

2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC) | 979-8-3315-3358-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/SMC58881.2025.11342937

***Abstract*—Neurological dysfunction encompasses a variety of diseases resulting from neural damage. Accurate assessment of neurological function is critical for diagnosis and the development of effective treatment plans. A significant number of patients with neurological disorders exhibit ocular abnormalities. Analyzing ocular status through eye movement capture plays a pivotal role in understanding various neurological dysfunctions. However, current methods of analyzing ocular status for neurological function assessments lack precision and objectivity, often relying heavily on physicians' subjective judgment. This work proposes the Ocular-enhanced Face Keypoints Network (OFKNet), a facial keypoint detection model based on deep convolutional neural networks. OFKNet employs ConvNeXt as its backbone network and introduces a multi-scale input enhancement strategy. Additionally, a region enhancement module based on MobileNetV3 is designed to optimize features in the canthus area. Multiscale feature fusion and channel weighting are achieved through an improved Path Aggregation Network and Squeeze-and-Excitation modules. To validate OFKNet's accuracy, we compared it with state-of-the-art models, including MediaPipe FaceLandmarker, InsightFace, Dlib68, and Dlib81, using a patient dataset we collected. Experimental results demonstrate that OFKNet outperforms existing models, particularly in calibration accuracy around the eyes. By monitoring eye movements in real-time, OFKNet ensures high-precision extraction of key points in each frame, accurately reflecting changes in patients' ocular movements.**

***Index Terms*—Computer vision, neurological dysfunction, facial paralysis, and convolutional neural networks.**

I. INTRODUCTION

Neurological dysfunction encompasses a spectrum of diseases resulting from nerve damage, often leading to impairments in movement, sensation, or emotion. Notable examples include Parkinson's disease, Alzheimer's disease, and facial paralysis. A considerable number of patients with neurological disorders experience difficulties in eyelid closure, potentially resulting in severe consequences such as vision impairment or blindness due to exposure keratopathy. Consequently, analyzing eye movements to assess ocular conditions offers vital insights into these neurological dysfunctions. However, the diagnosis of these diseases frequently depends on clinical evaluations and subjective judgments by physicians, highlighting a deficiency in objective and standardized assessment methods [1]. In the context of facial paralysis, there remains an absence of accurate and objective systems for comprehensive assessment of facial nerve function. Current diagnostic practices are primarily based on the physician's expertise and the patient's clinical presentation, with limited availability of quantifiable and standardized objective testing methods.

Clinically, individuals with neurological dysfunction may exhibit incomplete eyelid closure or asymmetry in eyelid movement, leading to complications such as exposure keratopathy or ocular infections, which increase the risk of corneal damage. These ocular manifestations provide valuable diagnostic indicators, enabling physicians to assess the severity of neurological dysfunction and formulate appropriate treatment strategies. By presenting objective visual representations of ocular conditions, patients gain a clearer understanding of their condition, fostering more active engagement in their treatment. Therefore, integrating data visualization techniques into ocular assessments significantly enhances diagnostic precision. However, accurately capturing and analyzing ocular states remains a significant challenge in clinical practice. One of the primary difficulties lies in the eyelid movement's dynamic and subtle nature, which can vary across individuals and fluctuate with factors such as fatigue, lighting conditions, and involuntary reflexes. Traditional assessment methods often rely on manual observation or static imaging, which lack temporal resolution and are prone to subjective interpretation [2].

Based on the aforementioned analysis, this work introduces the Ocular-enhanced Face Keypoints Network (OFKNet) to achieve high-precision detection of 416 facial key points, with a particular emphasis on the accurate localization of ocular key points. OFKNet employs ConvNeXt as its backbone network and integrates a multi-scale input enhancement strategy to improve adaptability across facial features of varying scales. To further refine feature representation, a region enhancement module based on MobileNetV3 is incorporated, leveraging a regional attention mechanism to dynamically adjust the weights of different facial regions, with a particular focus on optimizing the internal canthus area. To enhance multi-scale feature fusion, an improved Path Aggregation Network (PAN) is employed, facilitating both top-down and bottom-up feature propagation. Additionally, a

This work was supported by the Beijing Natural Science Foundation under Grants 4232049, the National Natural Science Foundation of China under Grants 62173013 and 62473014. (Corresponding author: Xiaomeng Zhao)

Squeeze-and-Excitation (SE) module is introduced to apply channel-wise weighting, improving feature discriminability. The keypoint detection process is conducted through a dual-branch approach, comprising a heatmap branch and a regression branch, with joint optimization achieved via an adaptive loss function that balances heatmap loss and regression loss. Comparative experiments demonstrate that OFKNet surpasses existing models in the precise localization of facial key points, particularly in the ocular region. By tracking ocular movements in dynamic temporal sequences, OFKNet ensures high-accuracy feature point extraction for each frame, making it a valuable tool for objective and quantitative neurological assessments.

II. PROPOSED FRAMEWORK

We employ a high-speed camera to capture seven typical facial movements of patients, including neutral face, eyebrow raising, nose scrunching, smiling, teeth showing, lip pucker, and cheek puffing, with a precision capable of recording at 120 frames per second at 4K resolution. OFKNet is proposed to detect 40 specific key points in the ocular region based on the collected images. The overall framework of OFKNet is shown in Fig. 1, which mainly comprises four modules.

1) Image preprocessing module employs ConvNeXt [3] as the backbone network, removing fully connected and pooling layers to preserve the spatial structure. A multiscale input strategy is introduced, adopting three resolutions (256×256, 112×112, and 56×56) of images to enhance the model's adaptability. MobileNetV3 [4] is adopted for facial area segmentation, and a region-enhancement strategy is proposed to adjust the weights of each region.
2) Feature extraction module is based on ConvNeXt. It extracts features at different levels from the pre-processed images. The bottleneck enhances feature representation while reducing the computational cost, and the down-sampling operation generates multi-scale feature maps.
3) Feature enhancement and fusion module receives these multi-scale feature maps and designs an improved PAN to fuse low-level and high-level features through top-down and bottom-up paths. SE further enhances the feature representation ability by weighting the channels and generating a new feature map.
4) Key point detection module conducts key point detection in a parallel manner using the heatmap and regression branches. The heatmap branch generates 2D Gaussian heatmaps to represent the key point probabilities, and the regression branch obtains the key point coordinates. The total loss function combines both to optimize the model and achieve accurate keypoint detection.

The detailed design of each module is then introduced separately in the following sections.

A. Image Preprocessing Module

ConvNeXt is a convolution-based architecture that integrates a hierarchical modular design and efficient computational strategies, enabling superior feature extraction. It has demonstrated outstanding performance in image classification and object detection tasks [3]. However, fully connected layers flatten the feature map into a vector, disrupting the spatial structure information. Since spatial structure is essential for accurately localizing key points in the facial region, fully connected layers are removed to preserve the original spatial layout of the feature map. Additionally, pooling layers are eliminated because they reduce the resolution of the feature map and lead to the loss of fine-grained details from the original image. To enhance the input representation during preprocessing, a multi-scale input enhancement strategy is employed. Three different resolutions, including 256×256, 112×112, and 56×56 are generated for the input image, with one resolution randomly selected during training. Furthermore, MobileNetV3, a lightweight object detection model, is utilized to segment the input image and identify the facial region. To further refine the preprocessing stage, a region enhancement module based on an attention mechanism is designed to dynamically adjust the weights assigned to different regions in the image, thereby improving the model's focus on crucial areas.

The accurate localization of the canthus is pivotal for the diagnosis, treatment, and prognostic evaluation of facial nerve movement disorders, as this region often exhibits subtle yet clinically significant asymmetries or deformations. Therefore, an optimization procedure is designed for ocular regions to enhance ocular feature representation in OFKNet [5]. Specifically, the internal canthus region is designated as a critical area, and targeted cropping is applied to ensure that the extracted image preserves essential features surrounding the canthus. Subsequent adjustments are made based on the precise canthus location, focusing on capturing deformed features in this region. During color adjustment, contrast and color saturation in the canthus region are selectively enhanced to improve the distinguishability of its color distribution. Additionally, an image blur operation is applied while maintaining relative clarity in the canthus region. For other key regions beyond the canthus, conventional enhancement techniques, including cropping, flipping, and color adjustment are employed, albeit with slightly lower intensity compared to the canthus region. In contrast, only mild enhancement operations are applied to background areas to ensure that feature optimization remains concentrated on the internal canthus region. This region-aware enhancement strategy not only improves the robustness and sensitivity of keypoint detection in high-resolution videos but also aligns with clinical priorities by emphasizing subregions with the highest diagnostic value.

B. Feature Extraction Module

After preprocessing, the input image is denoted as x_p, which undergoes multiple convolutional operations to extract features at different levels. In OFKNet, the convolution operation of each layer can be expressed by a general convolution operation. For the convolution layer l, its input and output are denoted as F_{l-1} and F_l, respectively. The kernel size is $k\times k$, the stride of the convolution kernel is s, the padding at the

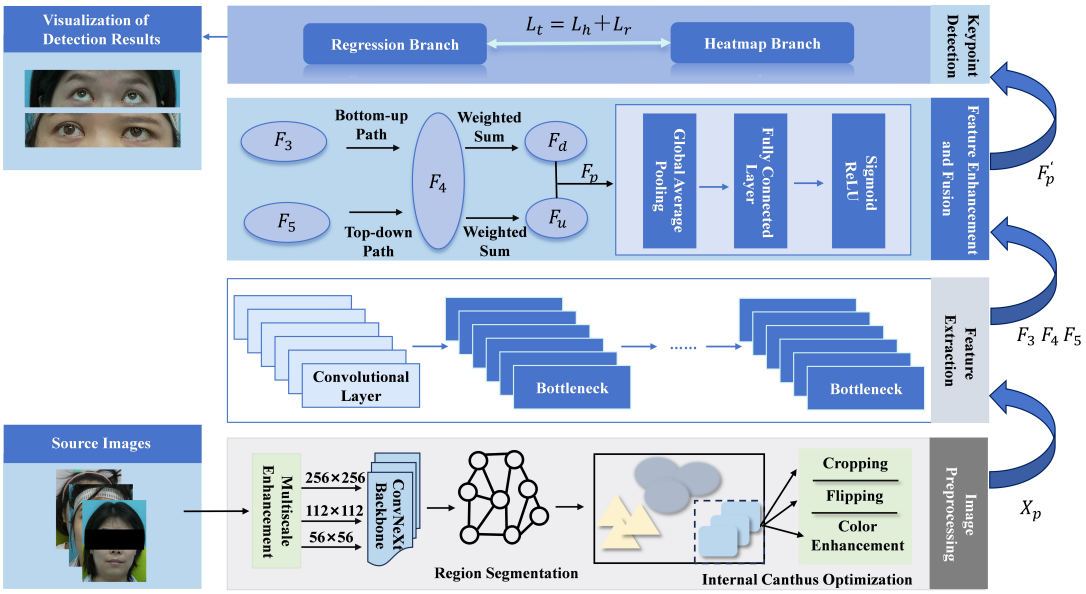


Fig. 1. Framework of OFKNet.

edges of the input feature map is p, and the number of output channels is denoted as C_l. Then, the output of convolution layer l is

$$F_l(i,j,c)=\text{ReLU}\left(\sum_{m=0}^{k-1}\sum_{n=0}^{k-1}\sum_{d=0}^{C_{l-1}} w_l(m,n,d,c) \cdot F_{l-1}(i+m\cdot s-p, j+n\cdot s-p, d)\right) \quad (1)$$

where $w_l(m,n,d,c)$ is the weight of the convolution kernel l, represented as a four-dimensional tensor. m and n are position indices within the $k\times k$ convolution kernel, d is the channel index of the input feature map, c is the channel index of the output feature map, and (i,j) denotes the spatial position of the feature map. ReLU$(\cdot)$ denotes the activation function.

When considering the convolution layer's computation, the spatial size of the feature map is changed for each convolution operation. Let $H_l\times W_l$ denote the output feature map size of the layer l, and its relation to the previous layer's output feature map size is calculated as

$$H_l=\left\lfloor \frac{H_{l-1}-k+2p}{s} \right\rfloor +1 \quad (2)$$

$$W_l=\left\lfloor \frac{W_{l-1}-k+2p}{s} \right\rfloor +1 \quad (3)$$

OFKNet incorporates a bottleneck module, which improves the expressive capacity of the network by expanding and compressing the number of channels, enhancing the representation of characteristics while reducing the computational cost. The input feature of the bottleneck module is denoted as x_t. After a 1×1 convolutional layer, the number of channels is expanded to get the intermediate feature x_m. Then, the feature x_n is obtained after a 3×3 convolutional layer. Finally, the number of channels is restored through a 1×1 convolutional layer to obtain the output feature x_o:

$$x_m=Conv_{1\times1}(x_t, C_i\times t) \quad (4)$$

$$x_n=Conv_{3\times3}(x_m, C_i\times t, s=1) \quad (5)$$

$$x_o=Conv_{1\times1}(x_n, C_i) \quad (6)$$

where C_i is the number of input channels of x_t, and t is the expansion factor. OFKNet performs downsampling operations to convert high-resolution feature maps into low-resolution ones. Let F_h denote the input high-resolution feature map and F_l denote the output low-resolution ones. It is obtained as:

$$F_l(i,j,c)=Conv\left(F_h(i\times r, j\times r, c), k, s=r\right) \quad (7)$$

where r denotes the down-sampling factor. Based on this operation, the network extracts features at multiple scales, generating feature maps F_3, F_4, and F_5 at different resolutions, better capturing facial key points of varying sizes.

C. Feature Enhancement and Fusion Module

Subsequently, for the extracted multi-scale feature maps F_3, F_4, F_5, an improved PAN is employed to integrate features of different scales. PAN effectively merges low-level and high-level features through top-down and bottom-up paths. The high-level feature map F_5 is upsampled to the size of F_4 and weighted summed with F_4 to obtain an intermediate feature map F_u:

$$F_u=\psi(F_5)+F_4 \quad (8)$$

where $\psi(\cdot)$ denotes the upsample process. Similiarly, the low-level feature map F_3 is upsampled to the size of F_4 and weighted summed with F_4 to get the feature map F_d:

$$F_d=\psi(F_3)+F_4 \quad (9)$$

Then, the top-down and bottom-up feature maps F_u and F_d are fused to obtain the final multi-scale feature map F_P:

$$F_P{=}F_u{+}F_d \tag{10}$$

In this case, PAN facilitates information flow between feature maps. It enables features at different scales to complement each other, ensuring that the network considers both local details and global structural information. SE is then adopted by using PAN's output to further enhance the feature representation ability. SE applies global average pooling (ϕ) to obtain global features for each channel. It then computes channel importance z_c and channel weights s_c using fully connected layers, *i.e.*,

$$z_c{=}\phi(F_{P,c}) \tag{11}$$

$$s_c = \text{ReLU}\left(W_2{\cdot}\text{ReLU}\left(W_1{\cdot}z_c\right)\right) \tag{12}$$

where W_1 and W_2 are the weight matrices of the fully connected layers. Each channel's weight s_c is multiplied with the input feature to obtain the final feature map F'_P:

$$F'_P{=}F_P{\cdot}s_c \tag{13}$$

The combination of SE and PAN enhances the representation ability of the new feature map F'_P, allowing it to better adapt to facial key points at different scales and complexities.

D. Key Points Detection Module

Based on the processed feature map F'_P, key point detection is performed using both the heatmap and the regression branches. The heatmap branch generates more discriminative features that produce accurate 2D Gaussian heatmaps, making the probability distribution of key points locations more precise. The regression branch learns the mapping between key points' coordinates and features, leading to more accurate key point coordinate detection. The heatmap branch detects the 2D Gaussian heatmap for each key point through convolution operations. It is assumed that the input feature F'_P is denoted as x , and the heatmap is obtained through two convolution operations:

$$y_1{=}\text{ReLU}\left(Conv\left(x, 3{\times}3, 512\right)\right) \tag{14}$$

$$y_2{=}\text{ReLU}\left(Conv\left(y_1, 3{\times}3, 256\right)\right) \tag{15}$$

$$h{=}\text{Sigmoid}\left(Conv\left(y_2, 1{\times}1\right)\right) \tag{16}$$

where y_1 is obtained by subjecting x to a single convolutional transformation and a ReLU activation, and y_2 is obtained by subjecting y_1 to another convolution and a ReLU activation. h is the heatmap for each key points, ranging from $[0, 1]$, indicating the probability distribution of the point in space. The regression branch directly detects the (x, y) coordinates of key points. When the input features are y_1 and y_2, the coordinates of the key points are detected through the convolutional layer as $\hat{x}_i$ and $\hat{y}_i$:

$$\hat{x}_i, \hat{y}_i{=}Conv\left(y_2, 1{\times}1\right) \tag{17}$$

The total loss function L_t consists of heatmap loss and regression loss. The heatmap loss L_h is based on the mean squared error, ensuring that the detected heatmap is close to the true heatmap. The regression loss function L_r ensures accurate coordinate detection, with weights w_i distinguishing the importance of different key points. They are obtained as:

$$L_h{=}\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i{-}y_i)^2 \tag{18}$$

$$L_r{=}\sum_{i=1}^{M} w_i(\hat{x}_i{-}x_i)^2 \tag{19}$$

$$L_t{=}L_h{+}L_r \tag{20}$$

where x_i and y_i are the true coordinates of the samples, N represents the number of samples used for calculating the heatmap loss, and M represents the number of samples used for calculating the regression loss.

III. Experimental Results and Discussion

A. Experimental Setting and Evaluation Metrics

The experiment focuses on real data from seven clinical tests, each with seven actions: neutral face, eyebrow-raising, nose scrunch, smiling, teeth showing, lip pucker, and cheek puffing[1]. These actions are selected for their ability to engage distinct facial muscle groups directly controlled by the facial nerve, each of which also influences the eye region's muscle. This work aims to evaluate the robustness and accuracy of different key point detection models for eye-related feature points under pathological conditions. Two manually annotated datasets are selected to achieve this: one with 14 eye feature points and another with 40 eye feature points [6]. Five different facial landmark detection models are selected for comparative analysis, including OFKNet, MediaPipe FaceLandmarker [13], InsightFace [15], Dlib68, and Dlib81 [7]. By assessing the performance of these models in detecting eye feature points, our objective is to understand their applicability and precision in diagnosing facial nerve dysfunction. Specifically, InsightFace, Dlib68, and Dlib81 cannot detect all 40 eye feature points. Therefore, to ensure fairness and comparability in the experiment, the dataset with 14 eye feature points is also adopted for comparison with models that cannot handle all 40 eye feature points. Table I provides detailed descriptions of the 14 and 40 eye feature points.

TABLE I
Eye Feature Points Index Description for Two Datasets

14 Eye Feature Points		40 Eye Feature Points	
Index	**Description**	**Index**	**Description**
0-5	Right eye contour	0	Right internal canthus
6-11	Left eye contour	1-18	Right eye contour
12	Right pupil center	19	Left internal canthus
13	Left pupil center	20-37	Left eye contour
-	-	38	Right pupil center
-	-	39	Left pupil center

For the 40 feature point dataset, OFKNet and MediaPipe FaceLandmarker are evaluated, as they can detect a more

[1]Neurorient supports motion capture videos.

detailed set of eye feature points, including the complete eye contour, pupil position, and other key points. This dataset is chosen to assess their performance in facial nerve dysfunction diagnosis, providing a comprehensive description of eye features in patients with facial nerve dysfunction. In contrast, the dataset with 14 eye feature points is used for a broader comparison across all models. In particular, Dlib68 and Dlib81 do not detect the pupil center. To quantitatively compare detection accuracy, three evaluation metrics are employed: mean Euclidean distance (MED), normalized mean error (NME), and eye opening degree (EO). MED calculates the deviation between the detected and manually annotated key points using Euclidean distance, quantifying the error for each key point. NME extends MED by considering the influence of inter-individual eye size differences. It normalizes the error by using the intercanthal distance, thus eliminating individual differences [8]. EO is an application-oriented metric that quantifies eye closure by analyzing changes in the eye fissure area, aiding in model sensitivity assessment for eye state changes in facial nerve dysfunction diagnosis [9]–[12]. Specifically, MED is defined as:

$$\text{MED} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i^{\text{p}} - x_i^{\text{g}})^2 + (y_i^{\text{p}} - y_i^{\text{g}})^2} \tag{21}$$

where N is the total number of key points. $(x_i^{\text{p}}, y_i^{\text{p}})$ denote the predicted coordinates of the key point i at a given frame. x_i^{g}, and y_i^{g} represent the corresponding ground truth coordinates. NME is defined as:

$$\text{NME} = \frac{d_{\text{a}}}{\text{dist}(P_{\text{l}}, P_{\text{r}})} \tag{22}$$

where d_{a} is the average predicted deviation for the given image, P_{l} and P_{r} are the coordinates of the left and right internal canthus points, and $\text{dist}(P_{\text{l}}, P_{\text{r}})$ is the Euclidean distance between the internal canthus points of both eyes. Finally, EO is computed as:

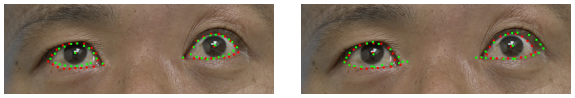
$$\text{EO} = \frac{S(P_1, P_2, \ldots, P_{17})}{\|P_1 - P_{10}\|^2} \tag{23}$$

where P_1 is the external canthus point, P_{10} is the internal canthus point, and $S(P_1, P_2, \ldots, P_{17})$ is the area of the eye fissure formed by these points.

B. Experiment Results

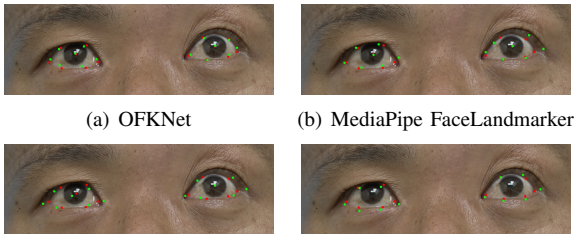
One patient is selected as an example to show eyebrow raise and voluntary blink actions as experimental data. Figs. 2(a) and 2(b) compare the errors between the detected 40 eye landmarks and the manually annotated points for OFKNet and MediaPipe FaceLandmarker [14]. Red points represent the manually annotated eye landmarks, while the green points represent the detected points. It is shown that the results of the two models in detecting the eye contour and pupil location are different, with OFKNet exhibiting higher detection accuracy.

Fig. 3 shows the detection results using the 14 eye feature points dataset for different models. We present the results for all models. It shows that OFKNet and MediaPipe FaceLandmarker demonstrate higher accuracy, especially in detecting the pupil center and eye internal/external canthus points. Table II shows the NME results, including the median (Median), mean (Mean), and standard deviation (Std) for each model. It is demonstrated that OFKNet outperforms compared models in both 40 and 14 eye feature point detection tasks, achieving the lowest NME.



(a) OFKNet (b) MediaPipe FaceLandmarker

Fig. 2. Error Comparison for 40 Eye Landmarks Detection.



(a) OFKNet (b) MediaPipe FaceLandmarker

(c) InsightFace (d) Dlib68

(e) Dlib81

Fig. 3. Error comparison for 14 eye landmarks detection.

TABLE II
NME FOR EYE LANDMARK DETECTION ACROSS DIFFERENT MODELS

Model	40-Point Eye Landmark NME			14-Point Eye Landmark NME		
	Median	Mean	Std	Median	Mean	Std
OFKNet	0.026285	0.028996	0.009228	0.033222	0.034307	0.00943
MediaPipe FaceLandmarker	0.080603	0.083615	0.01862	0.045109	0.049395	0.015408
InsightFace	-	-	-	0.335803	0.354827	0.081584
Dlib68	-	-	-	0.208735	0.203527	0.022459
Dlib81	-	-	-	0.218109	0.216568	0.01873

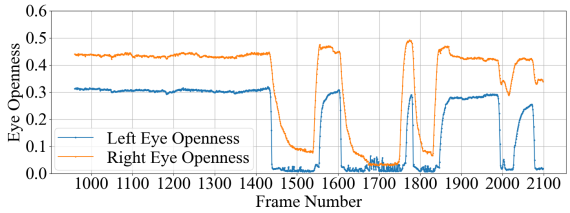


Fig. 4. EO Curves for OFKNet during voluntary blinking.

Figs. 4–8 illustrate the eye-opening degree over time for representative voluntary blinking sequences, comparing OFKNet with all four baseline models. It is shown that OFKNet produces a more precise and nuanced eye-opening degree curve than other methods. In all cases, the OFKNet curve closely tracks the subtle changes in eyelid position throughout the blink cycle, including minor partial closures

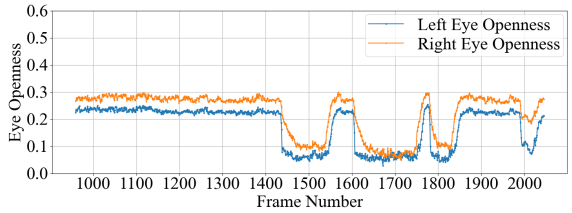


Fig. 5. EO Curves for FaceLandmarker during voluntary blinking.

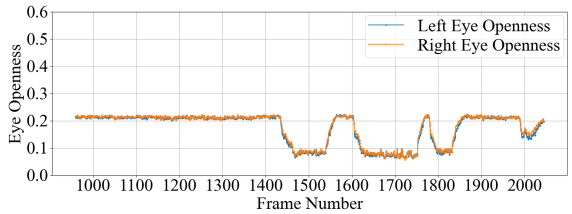


Fig. 6. EO Curves for InsightFace during voluntary blinking.

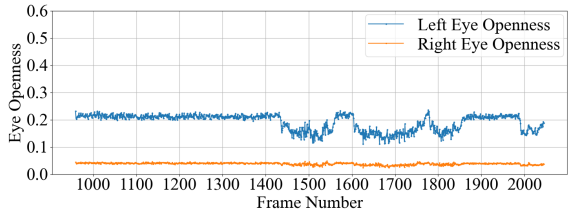


Fig. 7. EO Curves for for Dlib68 during voluntary blinking.

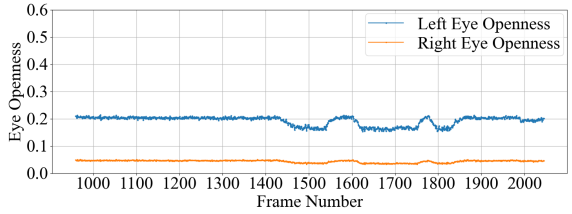


Fig. 8. EO Curves for Dlib81 during voluntary blinking.

and re-openings, resulting in a smooth, detailed waveform for the eye-opening degree. In contrast, the compared methods yield coarser or less responsive curves. Their outputs tend to miss or flatten minor variations in eye-opening, showing only significant dips corresponding to complete blinks but not finer oscillations during slight or incomplete blinks. These visual comparisons highlight the superiority of OFKNet in capturing minor variations in eye-opening degrees during voluntary blinking.

IV. CONCLUSIONS

Accurate evaluation of facial nerve function is essential for diagnosing and treating neurological disorders. Among facial expressions, ocular states serve as key indicators of facial nerve dysfunction. This work proposes an ocular-enhanced facial key point detection algorithm based on deep convolutional neural networks, named Ocular-enhanced Face Keypoints Net (OFKNet) to facilitate data-driven facial feature analysis. OFKNet significantly enhances the accuracy of eye-related key point calibration, particularly in tracking dynamic eye movements. It effectively captures and visualizes ocular data, aiding physicians in assessing critical indicators such as eye-opening degree, eyelid closure ability, and interocular synchronization. By precisely calibrating facial key points and providing visual evidence, OFKNet offers a more objective and accurate tool for evaluating neurological dysfunctions, with substantial potential for analyzing ocular impairments and supporting prognostic assessments.

REFERENCES

[1] V. Skaramagkas, A. Pentari, Z. Kefalopoulou and M. Tsiknakis, "Multi-Modal Deep Learning Diagnosis of Parkinson's Disease—A Systematic Review," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 2399–2423, May 2023.

[2] N. A. Sachs, E. L. Chang, N. Vyas, B. N. Sorensen and J. D. Weiland, "Electrical Stimulation of the Paralyzed Orbicularis Oculi in Rabbit," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 1, pp. 67–75, Mar. 2007.

[3] Z. Liu, H. Mao, C. -Y. Wu, C. Feichtenhofer, T. Darrell and S. Xie, "A ConvNet for the 2020s," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, pp. 11966–11976, Jun. 2022.

[4] A. Howard, M. Sandler, B. Chen, W. Wang, L. -C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam and Q. Le, "Searching for MobileNetV3," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), pp. 1314–1324, Jun. 2019.

[5] M. C. Moncaliano, P. Ding, J. M. Goshe, D. J. Genther, P. J. Ciolek and P. J. Byrne. "Clinical Features, Evaluation, and Management of Ophthalmic Complications of Facial Paralysis: A Review," *Journal of Plastic, Reconstructive & Aesthetic Surgery*, vol. 87, pp. 361–368, Dec. 2023.

[6] L. Ding and A. M. Martinez, "Features versus Context: An Approach for Precise and Detailed Detection and Delineation of Faces and Facial Features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2022–2038, Nov. 2010.

[7] D. E. King, "Dlib-ml: A Machine Learning Toolkit," *The Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, Dec. 2009.

[8] K. Wagner and M. Doroslovački, "Proportionate-Type Normalized Least Mean Square Algorithms With Gain Allocation Motivated by Mean-Square-Error Minimization for White Input," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2410–2415, May 2011.

[9] B. Mandal, L. Li, G. S. Wang and J. Lin, "Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 545–557, March 2017.

[10] S. Chen and J. Epps, "Efficient and Robust Pupil Size and Blink Estimation From Near-Field Video Sequences for Human–Machine Interaction," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2356–2367, Dec. 2014.

[11] M. Aloudat, M. Faezipour and A. El-Sayed, "Automated Vision-Based High Intraocular Pressure Detection Using Frontal Eye Images," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–13, Jun. 2019.

[12] S. Liu, "Alterations in Patients With First-Episode Depression in the Eyes-Open and Eyes-Closed Conditions: A Resting-State EEG Study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1019–1029, 2022.

[13] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg and M. Grundmann, "MediaPipe: A Framework for Perceiving and Processing Reality," *Proc. of the Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019.

[14] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-Time Facial Surface Geometry from Monocular Video on Mobile GPUs," 2019, arXiv:1907.06724.

[15] J. Deng, J. Guo, X. An, Z. Zhu and S. Zafeiriou, "Masked Face Recognition Challenge: The InsightFace Track Report," *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada, 2021.