Camera based Eye State Estimation for ICU Patients: A Pilot Clinical Study

Haowen Wang[1], Weijun Huang[1], Jia Huang[2], Guowei Wang[2], Hongzhou Lu[2,*], Wenjin Wang[1,*]

[1]Department of Biomedical Engineering, Southern University of Science and Technology, China.
[2]Department of Intensive Care Unit, The Third People's Hospital of Shenzhen, China.
[*]Corresponding authors: Hongzhou Lu, Wenjin Wang

Abstract

In the Intensive Care Unit (ICU), the awakening of patients from comas is indicative of recovery. This article investigates the feasibility of using conventional and deep learning-based methods for eye state estimation based on videos recorded by a CCTV-camera installed in the ICU. For handcrafted feature-based methods, geometric, HOG and RGB features are combined as the input of the SVM classifier to classify the eye state as open and closed. For deep learning-based methods, the eye and face images were used as joint input for classification. The clinical trial involved 48 critically-ill ICU patients. The results show that the HOG-RGB based method achieved an accuracy of 91.39%, while the deep learning-based method achieved an accuracy of 89.35%. These findings highlight the chances of using CCTV cameras to estimate the eye state of ICU patients, which can be a useful mean to provide information regarding the patient consciousness for clinicians to assess their recovery.

1. Introduction

The Glasgow Coma Scale (GCS) is a commonly used metric to assess the level of consciousness of a patient in medical settings [1]. One criterion of GCS is the state of eye-opening, which reflects a patient's level of arousal and information processing by the cerebral cortex. In the context of the Intensive Care Unit (ICU), monitoring the status of a patient's eyes can provide valuable information about their consciousness, which can indicate improvement or deterioration in their neurological status. This information is critical for caregivers to make decisions on patient's recovery or treatment, i.e. it can help doctors to determine appropriate dosages of sedation and analgesia, reducing the risk of over-treatment or under-treatment.

Recently, CCTV cameras have been used for physiological monitoring of patients in an ICU setting [2]. In addition to physiological measurement, it is also valuable to explore the functions of an imaging device for semantic measurements, such as the eye state. Vision-based eye state estimation has been explored in a wide range of applications, including driver fatigue monitoring [3], human-computer interaction [4] and human biometrics [5]. Vision-based techniques are commonly classified into two categories: handcrafted feature-based and deep learning-based, depending on the employed learning scheme [6]. In the classification of eye states based on handcrafted features, the open and closed eye states provide different appearance features around the eyes, which can be extracted as descriptors for binary classification (closed or open). Geometrical and appearance-based features are two typically used features in this category. The geometrical methods extract features such as the shape of the pupil and eyelid, and calculate the degree of eye opening [6], for example, calculating the proportion of pupil diameter and distance between the eyelids and thresholding it as binary classification [7]. Soukupova and Cech [8] proposed to use the eye aspect ratio (EAR), which is the ratio of height and width of the eye, for classification. In comparison, appearance-based techniques exploit more sophisticated textural features in the eye region. Features like Gabor, local binary patterns and histograms of oriented gradients (HOG) depict the texture and shape features of the eye image. Leveraging the advanced machine learning algorithms, satisfactory results have been achieved in the application of eye state estimation [9–12].

In addition to handcrafted feature-based techniques, deep learning-based methods, relying on convolutional neural networks (CNNs) for feature extraction [6], are used in a trainable end-to-end fashion. Kim et al. [13] proposed a deep residual CNN structure to estimate the eye open and closed states to cope with the environmental challenges (e.g. image blurring, illumination changes and image resolution differences) that may degrade the quality of eye images. Compared to other CNN structures, this method achieved a smaller classification error on the database combined of ZJU-database, DDBD-DB1-database and DDBD-DB2-database. In [14], Gou et al. developed a cascade regression framework for detecting eye and estimating eye states, which shows significant improvement on top of other mainstream methods for both the eye localization and eye state estimation based on BioID and GI4E databases. Next to the successful applications in non-clinical scenarios, it is interesting to know if the eye state estimation can create a new value stream for clinical scenarios such as in ICU, as a tool for estimating the level of patient consciousness.

In this study, the aim is to develop and evaluate the methods for eye state estimation in ICU patients, based on handcrafted features and deep learning features. Videos were collected by CCTV cameras installed in a real ICU setting, and manual annotation under expert-guidance was done to label the eye states as "open or closed" in the video sequences. The methods are developed to exploit geometric features, texture features and data-driven deep learning features for classifications. For handcrafted feature-based methods, the basic shape and texture information such as EAR, HOG and color were analyzed by resorting to a Support Vector Machine (SVM) classifier. For deep learning-based methods, we use eye images and face images as the input of a CNN model pre-trained on the CEW dataset [11]. The eye image and face image are fed to the CNN model to extract deep features respectively, and the feature weights are adaptively assigned to output the prediction results. The evaluations on ICU image datasets and individual videos indicate that both the handcrafted feature based SVM and deep learning-based classifier have reliable performance for eye state estimation.

To our best knowledge, this is the first work that studies the feasibility of using cameras to estimate the eye state of ICU patients and its potential for consciousness indication. Despite the challenges posed in an ICU environment (e.g. illumination, shadows, occlusions and unpredictable patient movements), our system demonstrates the feasibility of this emerging application. This system could improve the quality of care for ICU patients, also may improve the efficiency of clinical decision-making, by enabling timely detection of the patient's awakening state.

2. Clinical setup and data collection

The clinical trial was conducted in the ICU of Shenzhen Third People's Hospital, China, which involved 48 critically-ill patients. The study was approved by the hospital's Institutional Review Board (IRB) and informed consents were obtained from the patients or their guardians. The patient was monitored by a CCTV camera (Panasonic WV-SC384H Fast Ball Network Camera) installed approximately 3 meters and 45 degrees in front of the bed, recording the scene in the RGB format at 25 frames per second with a resolution of 1080×1920 pixels. Fig. 1 shows the scenario of ICU recordings.

Among the 48 ICU patients, 29 were annotated as asleep or unconscious. In this study, each ICU patient was continuously monitored for a short period of 10 minutes. To ensure that the experiments could reflect the real conditions in an ICU environment, its setting was not adjusted nor tuned for video recording, resulting in a challenging dataset that included various practical challenges such as poor illumination, blurring, shadows, occlusions of clinicians during operation, etc. Typically the challenges like the uneven distribution of illumination on face, head pose variations, and the occlusion of the medical tape on the face, are difficult for analyzing facial attributes. Moreover, due to the needs of patient monitoring, the focal length of CCTV cameras were not standardized, which varied in different units.

Figure 1. The ICU setup for the clinical study, where a surveillance CCTV camera is used to record the patient face and upper body.

The created dataset has 1200 open-eye faces and 1450 closed-eye faces, which are further separated into the images of left-eye and right-eye. Each face image corresponds to a left-eye image and a right-eye image. Prior to the analysis, the eye images were resampled to a uniform size of 24×24 pixels, while the face images were resampled to 224×224 pixels with aspect ratio unchanged. Fig. 2 exemplifies snapshots of the facial and eye images in our dataset (the facial images are blurred to protect the privacy).

3. Methodology

The purpose of this study is to evaluate the accuracy of using different facial features for classifying the eye state of ICU patients. Specifically, classical handcrafted feature-based methods and a lightweight deep learning method pre-trained on a public dataset were used. The model generated after using five-fold cross-validation during training is applied to assess the patient's eye state (open or closed) in video sequences. In the video, a label is created each 5 seconds, with 1 indicating eye-opening and 0 indicating eye-closed. The video sequences were annotated under the

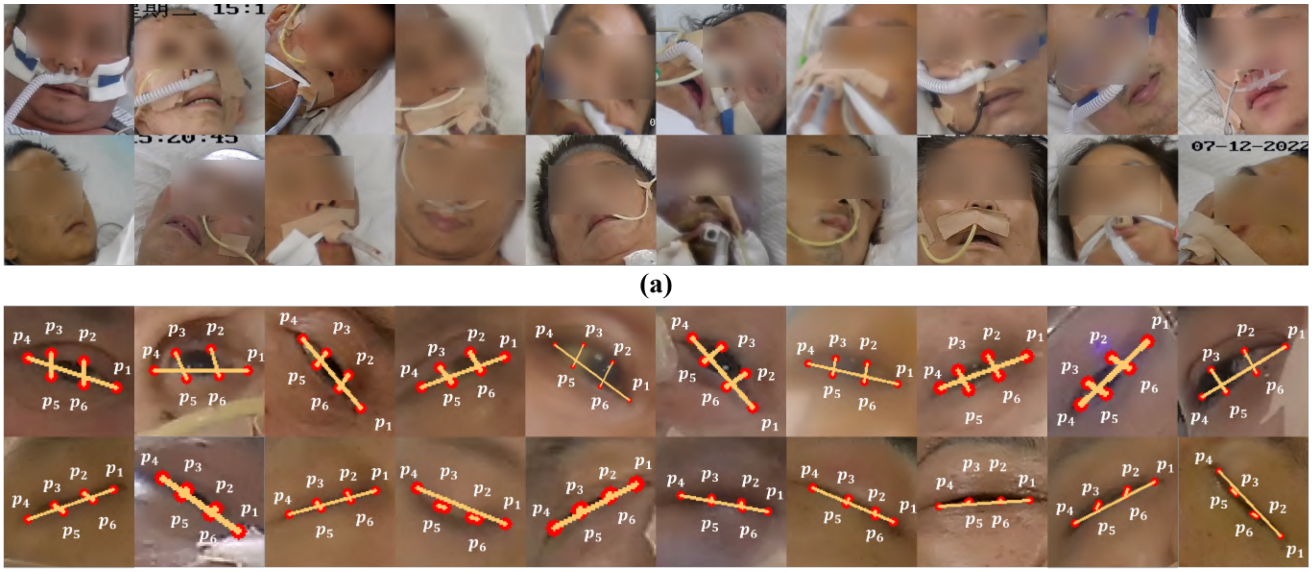

Figure 2. Exemplified images in our ICU dataset: (a) images of patient face and (b) images of patient eyes (with landmark detection and distance calculation).

guidance of a medical practitioner. The classifier output was compared with the annotations for quantifying the performance. The ultimate goal is to provide a real-time assessment of ICU patient wakefulness based on the eye state estimation.

3.1. ROI extraction of eyes

To extract the eye features, the MediaPipe [15], an advanced deep-learning based facial landmark detector, was employed to locate the eye keypoints. The first step is to detect the facial region and locate 468 facial landmarks. Once the landmarks were identified, the positions of the left and right eyes were determined using six key-points for both eyes, i.e. the key-points were (387, 373, 384, 381, 362, 263) for the left eye, for right eye were (160, 144, 157, 154, 33, 133). Finally, the leftmost and rightmost landmarks of both eyes are used to extract the eye patches. These patches were resampled to 24×24 pixels and used for eye state estimation, as shown in Fig. 3.

3.2. Handcrafted features with SVM

Fig. 3 shows the proposed methodology used for eye state recognition. The proposed scheme consists of several steps such as facial landmark detection, handcrafted feature extraction and eye state estimation.

Geometric features (EAR). EAR is a commonly used metric used for detecting the state of a patient's eye (open or closed). It is calculated as the ratio between the vertical distance (of four landmarks on the upper and lower eyelids) and the horizontal distance (of two other landmarks on the inner corner and outer corner of the eye), expressed as:

$$\mathrm{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\,\|p_1 - p_4\|}, \tag{1}$$

where p_1 and p_4 are landmarks of eye corners, p_2, p_3, p_5, p_6 are landmarks of upper and lower eyelids [8], as illustrated in Fig. 2 (b).

In the EAR+SVM approach, a model was learned to recognize different EAR values and landmark distances that can differentiate open eyes and closed eyes. The SVM classifier is trained on a set of labeled data, with the EAR value and six distances (vertical and horizontal) between eye landmarks as the input feature.

Texture features (HOG). The HOG feature extraction is utilized to construct a histogram that characterizes the texture pattern of the eye region, by calculating the gradient orientation of pixel intensities [16]. This histogram is used as a feature representation for eye state classification. The texture features of the grayscale images were extracted using HOG, with a cell size of 10×10 pixels and a block size of 2×2 cells. Instead of applying HOG to the eye patch, we extract the HOG features from six landmarks of an eye and crop a square region with half of the eye's length as its side length. Subsequently, we extract a 144-dimensional histogram from each region. It solves the problem when applying HOG on the whole patch of the eye. In the ideal case, the pupil (the anchor) should be positioned at the center of the eye images, and the corresponding block can extract the eye features, as open and closed eyes have distinct differences. However, as shown in Fig. 2 (a), the orientation of a

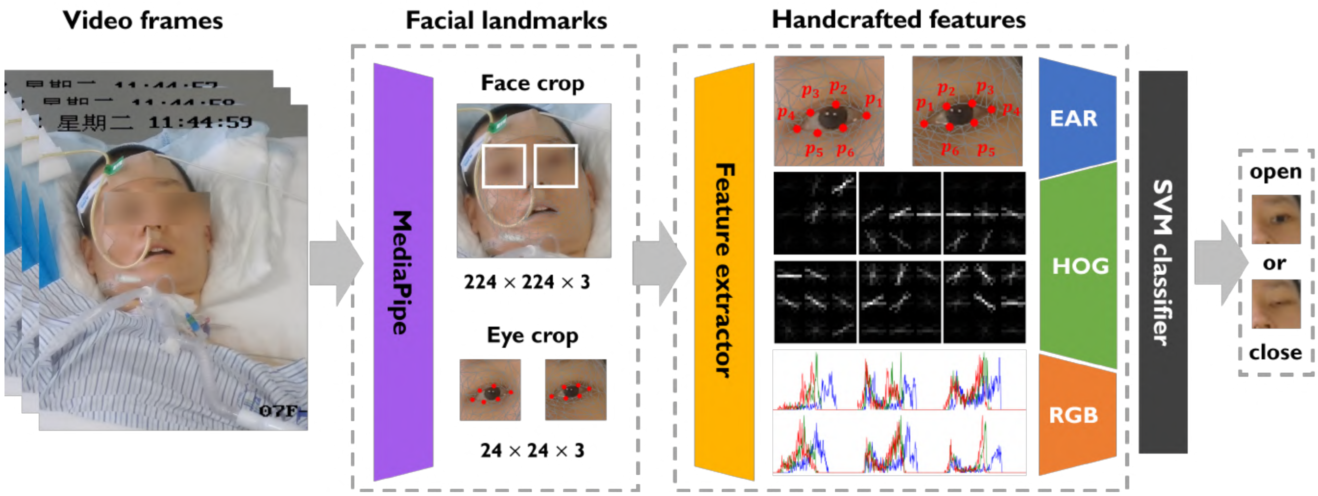

Figure 3. The classification model based on handcrafted features, i.e. a hybrid of geometric features (EAR), texture features (HOG) and color histogram features, for eye state estimation.

patient's head can cause the pupils to deviate from the eye centroid, resulting in different HOG features even though the parameters are the same, i.e. a rotation of HOG features that may render the misclassification. Our approach tends to avoid this kind of error by placing HOG features on eye key-points, enabling more reliable extraction of HOG features regardless of their positions in an image.

To capture the eye states that are indicative of consciousness, separate classifiers were trained for the left and right eyes. We determined the eye state based on the presence of an open eye, assuming that if either the left eye or right eye is open, the patient is considered to have eyes open.

Color histogram features. As shown in the feature extraction module of Fig. 3, our proposed method attempts to fuse different features to determine the eye state. Due to the difference between eye-tone and skin-tone, color-based features have also been extensively used for eye state classification [17, 18]. To extract RGB features, we first calculate a histogram of color distribution of the eye image in each of RGB channels. To make the histogram more informative, we divide each channel into a certain number of bins. Each channel is divided into 5 bins, leading to a total of 125 vectors for histogram representation. We integrated RGB features and HOG features to provide complementary information to the classifier.

3.3. Deep learning features with transfer learning

CNN model architecture. Fig. 4 shows the overview of the designed CNN architecture fusionModel, consisting of eyeModel, faceModel, adaptive weights layer, and linear classifier. The inputs to the whole network are images of the face and both eyes obtained from the Mediapipe, and the outputs are the predictions of the states of the eyes.

For the eye images of 24×24 pixels from MediaPipe, it is not very suitable to extract image features with an overly complex network model. Therefore, in the eyeModel, we designed two parallel lightweight 5-layer CNN layers with the same architecture to extract the image features of the left eye and right eye, respectively. The obtained features are then concatenated and fed into the linear classification layer for eye state estimation. We added a dropout layer to the linear classification layer to reduce the risk of overfitting. In addition, for facial images with 224×224 pixels, if we directly build a CNN for facial feature extraction, the network model is prone to overfitting and it is difficult to obtain robust facial features. Therefore, we choose ResNet101 as the backbone for deep feature extraction and initialize the parameters of the network model using the parameters pre-trained on ImageNet. The facial feature extraction network trained on a large public dataset was further fine-tuned using the images of the ICU dataset.

Finally, we concatenate the features of the left and right eyes obtained from the eyeModel and the facial features obtained from the faceModel. The neural network was used to assign weights for the concatenated features through the adaptive weight assignment layer, and the weighted features were used as the input of the fully-connected layer for final decision making.

Transfer learning. In deep learning, the size and diversity of the dataset often affect the recognition accuracy, if our fusionModel is trained directly on the ICU dataset, it is prone to overfitting and lack of generalizability. Therefore, we employ a migration learning scheme to address such limitations.

We also used an external eye condition dataset, called CEW [11], to pre-train our eyeModel and faceModel (see

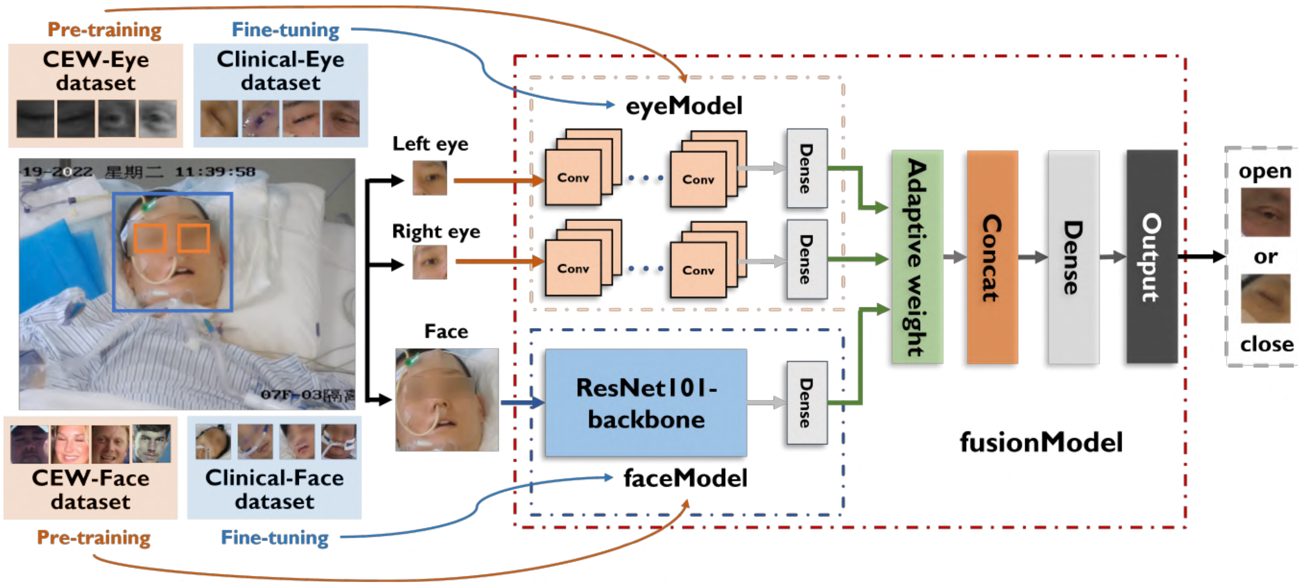

Figure 4. The deep learning model designed for eye state estimation, which fuses eye features and facial features for decision making.

Fig. 4). Although the face model was initialized using ImageNet's pre-training parameters, we fine-tuned the feature extraction layer with a smaller learning rate (e.g. 1e-6) to specialize this generic feature extractor into a face feature extractor. After the pre-training was completed, we migrated the learning weights of the eyeModel and faceModel feature extraction parts to fusionModel, respectively, and then fine-tuned the fusionModel model end-to-end on the ICU dataset.

4. Results and discussion

We compared the performance of three deep learning models with and without the pre-training strategy in Table 1. It is clear that the models that have been pre-trained on the CEW dataset have an overall better performance than the models trained on the ICU dataset only. We also observed that the models without pre-training were prone to the overfitting when trained directly on the ICU dataset, which is in line with our expectation. Therefore, we consider that the performance of the target model can be improved to some extent by pre-training it on a preliminary model through an external dataset and then migrating the pre-trained network weights to the target scenario (e.g. ICU).

To validate the effectiveness of our fusionModel architecture, we compared the performance of eyeModel, faceModel and fusionModel under the same training scheme with the five-fold cross-validation. From Table 1, we can see that for eyeModel, the overall performance is poor due to the limited size of the model, i.e. only the eye region is exploited for analysis. The faceModel can achieve an accuracy of 93.30% when pre-trained on the CEW dataset

Method	ACC.(%)	CR(%)	OR(%)
EAR+SVM	82.25	77.28	87.83
HOG+SVM	87.44	94.66	79.32
HOG+RGB+SVM	**91.75**	90.11	93.58
eyeModel	64.28	78.72	49.93
eyeModel$_{pre}$	72.20	78.80	64.85
faceModel	74.80	99.80	38.18
faceModel$_{pre}$	93.30	94.83	87.62
fusionModel	72.95	99.93	34.37
fusionModel$_{pre}$	**94.09**	95.27	90.18

Table 1. Evaluation of different architectures. pre denotes the method pre-trained on the CEW dataset. (Acc. - Accuracy, CR - Closed-eye correctly rate, OR - Open-eye correctly rate)

and fine-tuned on the ICU dataset. fusionModel achieves the best performance using the same training strategy. The main advantage of fusionModel is the fusion of eye features and facial features using adaptive weight assignment.

Different feature sets are thoroughly compared in Table 1 using similar metrics of [12]. The accuracy of approaches using features of EAR, HOG, HOG-RGB and deep model are 82.25%, 87.44%, 91.75% and 94.09% respectively. However, it should be noted that the method using EAR feature may not be effective in the ICU application, particularly for the patients with brain injuries or other issues whose eyes cannot be fully closed. It shows that the geometric features are not sufficient in such challenging cases without the learning of eye appearance. Therefore, features that are capable of measuring local textual information (e.g. deep data-driven features) are more suitable for the ICU condition.

PATIENT	EAR+SVM			HOG+SVM			HOG+RGB+SVM			Deep Learning: fusionModel$_{pre}$		
	ACC.(%)	CR(%)	OR(%)	ACC.(%)	CR(%)	OR(%)	ACC.(%)	CR(%)	OR(%)	ACC.(%)	CR(%)	OR(%)
1	93.33	100.00	90.00	99.17	97.5	100.00	97.50	97.50	97.50	97.50	100.00	96.25
2	79.17	69.64	87.50	93.33	94.64	92.19	93.33	100.00	87.50	87.50	100.00	76.56
3	100.00	100.00	100.00	85.83	90.91	76.74	100.00	100.00	100.00	99.17	98.70	100.00
4	78.33	100.00	76.36	95.00	70.00	97.27	85.00	90.00	84.55	80.00	30.00	84.55
5	81.67	100.00	73.17	99.17	100.00	98.78	97.50	92.11	100.00	97.50	94.74	98.78
6	91.67	100.00	56.52	90.00	100.00	47.83	92.50	100.00	60.87	95.00	100.00	73.91
7	76.67	100.00	41.67	80.00	70.83	93.75	75.00	95.83	43.75	77.50	100.00	43.75
8	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
9	65.00	100.00	43.24	50.00	4.35	78.38	81.67	100.00	70.27	71.67	100.00	54.05
Overall	85.09	96.51	75.72	88.06	83.78	91.57	**91.39**	98.36	85.67	**89.54**	97.95	82.63

Table 2. Comparison of benchmarked methods on 9 ICU patients recorded in video sequences. The boldface denotes the best method in each category of conventional machine learning and deep learning.

Figure 5. The temporal changes of left and right eye states of patient 4 with brain injury.

In addition to the validation on images of the ICU dataset, we collected 10-minute videos of nine ICU patients to perform long-term eye state assessment using benchmarked methods. Table 2 presents a comparison of classification results using four different features. The SVM classifier with EAR, HOG, and HOG-RGB features achieved an overall accuracy of 85.09%, 88.06%, and 91.39%, respectively, while the deep learning-based method achieved an overall accuracy of 89.54%. However, since the closed-eye and open-eye periods of each patient in videos may vary, a single accuracy metric is insufficient to evaluate the performance. The CR and OR metrics can more intuitively reflect the detection rate of open eyes and closed eyes. Table 2 shows that the SVM trained on multivariate features (e.g. HOG and RGB histogram) outperforms the ones trained on a single feature (e.g. HOG only), which is reasonable as HOG and RGB histogram are complementary, and their fusion can further rich the description of eye regions.

The deep learning method shows good performance as compared with the SVM-based methods in video-based analysis. The single feature based classification is not effective in handling the challenges of the ICU due to the occluded faces or changes in patient's head orientation, which could lead to persistent incorrect predictions. Overall, detecting open eyes for ICU patients is more challenging. For example, patient 4 was diagnosed with acute cerebral infarction, who is unable to fully close the eyes. Additionally, the patient with nystagmus symptoms has uncontrolled eye movements, leading to the changes of eye state even in the period of unconsciousness (see Fig. 5). Patient 7 experienced eyelid ptosis due to sedative use, making it difficult to open the eyes. During the certain periods of wakefulness, the close proximity of the upper and lower eyelids may contribute to misclassifications (see Fig. 7).

To understand the effectiveness of deep networks for eye-related feature learning, the activation heatmaps obtained by networks trained on different datasets are shown in Fig. 6, to illustrate which image parts are highlighted for task-related feature extraction. The heatmap is presented as a color-coded image, where redder indicates the regions with higher weights or attentions and bluer means less important regions. It is clear that the network trained on ImageNet only does not have correct focus of task-related feature extraction for nine patients. After pre-training on the CEW dataset, the heatmaps can highlight the eye region that is closely related to the training label (wake and sleep), though some patients have slight drift between the eyes and highlighted areas, i.e. three patients have only a single eye highlighted. This suggests that the training on the CEW dataset was functional but not optimal. The fine-tuning on the ICU dataset leads to further improvement where the eye areas are more pronounced in the heatmaps. It proves that the designed network is indeed valid for ICU patients.

Fig. 7 shows the prediction results of each classifier on an individual basis. The red area denotes the patient's wake period, while the blue area denotes the sleep period. As can be

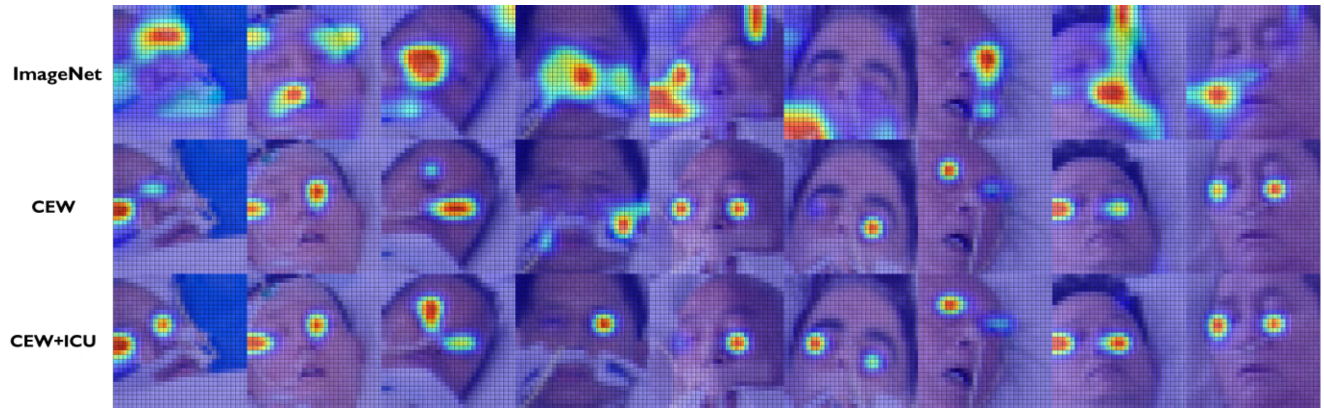

Figure 6. Activation heatmaps obtained by networks trained with three different datasets, e.g. ImageNet, ImageNet + CEW, and CEW+ICU.

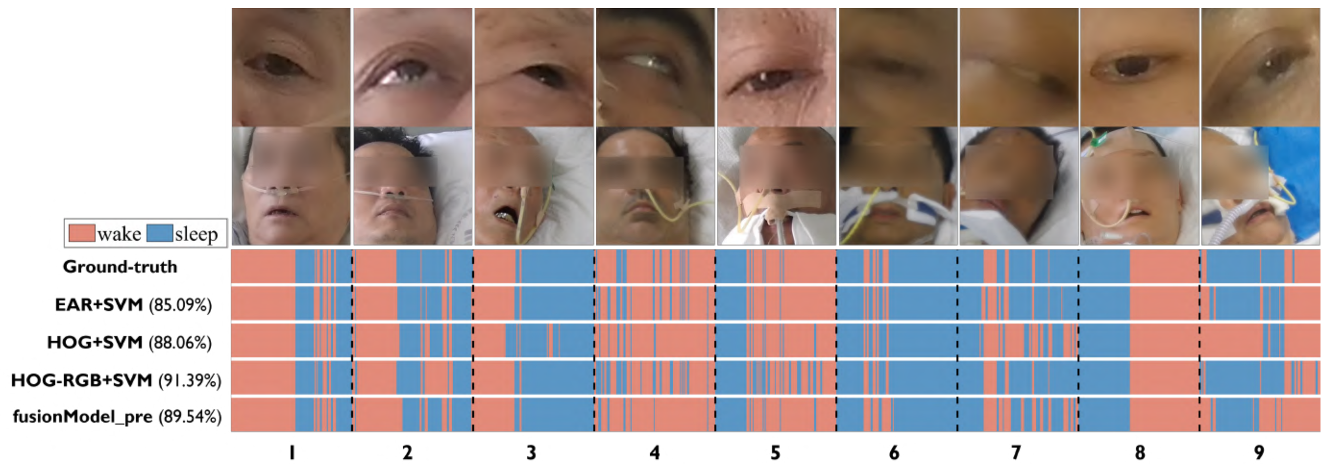

Figure 7. Continuous classification of wake and sleep states of ICU patients over a period of time in a video.

seen in Fig. 7, some patients have intermittent spontaneous eye opening (e.g. patients 4, 6, 7 and 9). Intermittent spontaneous eye opening refers to the occurrence of short periods of eye opening during a longer period of continuous eye closure. It may be due to the patient's brain function injury that leads to an inability to control the eyeball movements. Sedation and analgesia may also influence the patient's consciousness and thus the controlling of eyeball movements. In addition, ICU patients are generally faced with sleep difficulties, which are characterized by sleep fragmentation, decreased sleep efficiency, and frequent arousal [19]. Based on eye state estimation, future work can be focused on using deep learning techniques to further distinguish the status between sleep, awake and unconsciousness, following the medical guidelines that can be interpreted by clinicians and involving their visions to solve the real needs.

In summary, this study aims to explore the feasibility of eye state estimation for ICU patients using CCTV cameras. To this end, we developed different classification models based on handcrafted features and deep learning features. The results demonstrated that HOG-RGB features and deep learning-based features are effective for this assignment, outperforming other features in terms of classification accuracy, also CR and OR. However, detecting open eyes for ICU patients is more challenging due to a variety of challenges such as facial occlusions and different face orientations (e.g. non-frontal face). Additionally, intermittent spontaneous eye opening was observed for some patients, making it difficult to distinguish between sleep and wakefulness.

As the next step, we may further investigate the opportunity of integrating the functions of eye state estimation and physiological measurement (e.g. heart-rate, respiration rate, SpO2) on a single camera sensor, to provide a more comprehensive measurement for ICU patients. It is especially interesting to explore if the hybrid approach including both the physiological and semantic information can enable more accurate calculation of early warning scores. This would be particularly beneficial for critical patients that need intensive monitoring in acute settings like ICU.

5. Conclusions

We conducted a clinical research in ICU to evaluate various vision-based methods for eye state estimation, focused on assessing their performance in the classification of closed eye and open eye. The experimental results showed that both the HOG-RGB model and the fusionModel (deep learning-based) have good detection accuracy in eye state estimation for ICU patients. Additional efforts are needed to improve the localization of facial landmarks for non-frontal faces in specific sleep positions. Real-time monitoring of a patient's eye state is crucial in the ICU for detecting changes in alertness and enabling prompt medical interventions. As the videos were acquired by existing CCTV cameras in ICU, integrating the proposed system into the Internet of Medical Things system would not require additional space or infrastructure in the ward.

Acknowledgements

This work is supported by the National Key R&D Program of China (2022YFC2407800), General Program of National Natural Science Foundation of China (62271241), Guangdong Basic and Applied Basic Research Foundation (2023A1515012983), and Shenzhen Fundamental Research Program (JCYJ20220530112601003).

References

[1] Bryan Jennett and Graham Teasdale. Aspects of coma after severe head injury. *The lancet*, 309(8017):878–881, 1977. 1

[2] Haowen Wang, Jia Huang, Guowei Wang, Hongzhou Lu, and Wenjin Wang. Surveillance camera-based cardio-respiratory monitoring for critical patients in icu. In *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 1–4. IEEE, 2022. 1

[3] Luis Miguel Bergasa, Jesús Nuevo, Miguel A Sotelo, Rafael Barea, and María Elena Lopez. Real-time system for monitoring driver vigilance. *IEEE Transactions on intelligent transportation systems*, 7(1):63–77, 2006. 1

[4] Kristen Grauman, Margrit Betke, Jonathan Lombardi, James Gips, and Gary R Bradski. Communication via eye blinks and eyebrow raises: Video-based human-computer interfaces. *Universal Access in the Information Society*, 2:359–373, 2003. 1

[5] Mariusz Szwoch and Paweł Pieniażek. Eye blink based detection of liveness in biometric authentication systems using conditional random fields. In *Computer Vision and Graphics: International Conference, ICCVG 2012, Warsaw, Poland, September 24-26, 2012. Proceedings*, pages 669–676. Springer, 2012. 1

[6] Sumeet Saurav, Prashant Gidde, Ravi Saini, and Sanjay Singh. Real-time eye state recognition using dual convolutional neural network ensemble. *Journal of Real-Time Image Processing*, 19(3):607–622, 2022. 1

[7] Ioana Bacivarov, Mircea Ionita, and Peter Corcoran. Statistical models of appearance for eye tracking and eye-blink detection and measurement. *IEEE transactions on consumer electronics*, 54(3):1312–1320, 2008. 1

[8] Jan Cech and Tereza Soukupova. Real-time eye blink detection using facial landmarks. *Cent. Mach. Perception, Dep. Cybern. Fac. Electr. Eng. Czech Tech. Univ. Prague*, pages 1–8, 2016. 1, 3

[9] Hai Yan Yang, Xin Hua Jiang, Lei Wang, and Yong Hui Zhang. Eye statement recognition for driver fatigue detection based on gabor wavelet and hmm. In *Applied Mechanics and Materials*, volume 128, pages 123–129. Trans Tech Publ, 2012. 1

[10] Lubing Zhou and Han Wang. Open/closed eye recognition by local binary increasing intensity patterns. In *2011 IEEE 5th International Conference on Robotics, Automation and Mechatronics (RAM)*, pages 7–11. IEEE, 2011. 1

[11] Fengyi Song, Xiaoyang Tan, Xue Liu, and Songcan Chen. Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. *Pattern Recognition*, 47(9):2825–2838, 2014. 1, 2, 4

[12] Yanchao Dong, Yan Zhang, Jiguang Yue, and Zhencheng Hu. Comparison of random forest, random ferns and support vector machine for eye state classification. *Multimedia Tools and Applications*, 75:11763–11783, 2016. 1, 5

[13] Ki Wan Kim, Hyung Gil Hong, Gi Pyo Nam, and Kang Ryoung Park. A study of deep cnn-based classification of open and closed eyes using a visible light camera sensor. *Sensors*, 17(7):1534, 2017. 1

[14] Chao Gou, Yue Wu, Kang Wang, Kunfeng Wang, Fei-Yue Wang, and Qiang Ji. A joint cascaded framework for simultaneous eye detection and eye state estimation. *Pattern Recognition*, 67:23–31, 2017. 1

[15] Valentin Bazarevsky, Yury Kartynnik, Andrey Vakunov, Karthik Raveendran, and Matthias Grundmann. Blazeface: Sub-millisecond neural face detection on mobile gpus. *arXiv preprint arXiv:1907.05047*, 2019. 3

[16] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005. 3

[17] Zhonglong Zheng, Jie Yang, and Limin Yang. A robust method for eye features extraction on color image. *Pattern Recognition Letters*, 26(14):2252–2261, 2005. 4

[18] Chun-Wei Tan and Ajay Kumar. Automated segmentation of iris images using visible wavelength face images. In *Cvpr 2011 Workshops*, pages 9–14. IEEE, 2011. 4

[19] Randall S Friese. Sleep and recovery from critical illness and injury: a review of theory, current practice, and future directions. *Critical care medicine*, 36(3):697–705, 2008. 7