Journal of Computers Vol. 34 No. 5, October 2023, pp. 103-115
doi: 10.53106/199115992023103405008

Two-Stream Spatial–Temporal Transformer Networks for Driver Drowsiness Detection

Qianyi Jiang, Huahu Xu*, Chen Cheng

School of Computer Engineering and Science, Shanghai University,
Shanghai 200444, China
jqy@shu.edu.cn, huahuxu@shu.edu.cn

Received 28 September 2022; Revised 24 January 2023; Accepted 13 March 2023

Abstract. For driver drowsiness detection in the real world, the existing methods have good performance in general. However, when the face is blocked, the light is dim, and the driver's head posture changes, the performance will deteriorate significantly. In this paper, a two-stream Spatial-Temporal transformer network intended to perform diver drowsiness detection task is proposed to solve the above problems. The spatial-temporal graph is extracted from the video and then the results are obtained from 2s-STTN. The model is a two-stream transformer network model. In our model, the Spatial Self-Attention module is used to learn the embedding of different facial landmarks, and the Temporal Self-Attention module is used to learn the correlation between the frames of facial feature points. Different activated facial landmarks are separated and recognized by class activation mapping technology. Each flow recognizes different activated facial features, extracts spatial or temporal features, and integrates the information about facial features, so as to improve the performance of the system. 2s-STTN can not only mine the long-term dependence of driver behavior from video, but also mine the driver drowsiness information provided by the unobstructed facial signs when the face is blocked. By conducting experiments and comparing our model with other models, it is demonstrated that the proposed model has good performance in driver drowsiness detection.

Keywords: driver drowsiness detection, self-attention, graph convolutional neural network, class activation mapping

1 Introduction

With the rapid development of social economy, the number of cars is increasing, and the potential safety hazards are also increasing sharply. According to statistics [1], fatigue driving is the main factor causing traffic accidents, but before fatigue driving leads to traffic accidents, drivers have obvious fatigue characteristics. Giving drivers early warning over fatigue driving in time can greatly reduce the chance of traffic accidents. Many people may experience fatigue driving, and fatigue driving can be predicted. It is of great significance to study how to detect the fatigue characteristics of drivers quickly, accurately, and stably for the improvement of traffic safety driving problems.

Early research usually used head pose estimation [2-3], percentage of eyelid closures (PERCLOS) [4-6] or the number of yawns [7-9] to judge whether fatigue driving occurred. In recent years, various methods based on CNN, LSTM [22-23], and spatial-temporary graph convolution methods [24] have achieved good experimental results. However, it was found out in our tests that when the face is blocked, the light is dim, and the driver's head posture changes, the performance of these methods deteriorates significantly. In these special cases, incorrect results are often obtained. This is also the motivation for us to propose new methods.

In this paper, a two-stream spatial–temporal transformer network (2s-STTN) is proposed for Driver Drowsiness Detection task, which effectively uses spatial-temporal features for drowsiness detection. The framework of 2s-STTN model is shown in Fig. 1. Driver's facial features are extracted from real-time video to construct spatial-temporal data.

Class activation maps (CAM) technology [10] is used to deal with sunglasses occlusion, and CAM technology is applied to graph convolution neural network layer. The different activated facial features are recognized by CAM technology. Each flow recognizes different activated commercial features, extracts spatial or temporal features, and integrates information. To deal with the change of head posture and the change of light, the spa-

* Corresponding Author

tial-temporal self-attention mechanism is adopted. The richer representations of facial landmarks are obtained by introducing spatial-temporal self-attention mechanism. The function of Spatial Self-Attention (SSA) and Temporal Self-Attention (TSA) modules is to build new embeddings through Transformer according to the known graph information. Therefore, graph embeddings cease to depend only on the given graph structure and graph convolution.

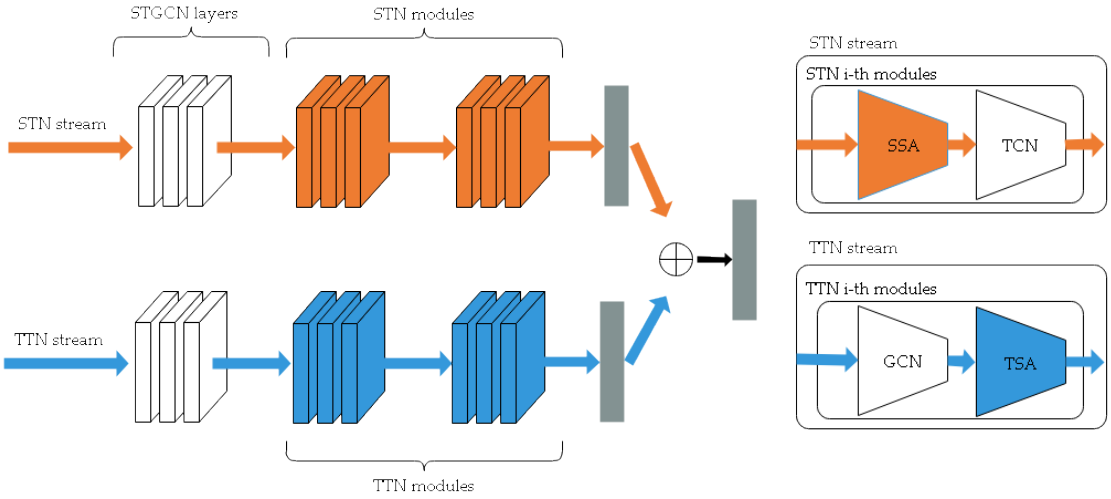

Fig. 1. The framework of 2s-STTN model

There are two streams in the proposed model, namely spatial Transformer Network stream (STN) and temporal Transformer Network (TTN) stream. In the present study, the SSA modules in STN stream and the TSA modules in TTN stream are used.

The main contributions of this paper are as follows. Firstly, a new two-stream Spatial-Temporal transformer network is proposed for Driver Drowsiness Detection task, with both temporal and spatial information collected from drivers' continuous facial features aggregated. Secondly, in the spatial dimension, the SSA module is introduced to establish connections between facial landmarks, while in the temporal dimension, the TSA module is introduced to learn dynamic facial features. Lastly, each spatial-temporal flow is recognized through class activation mapping to improve the robustness of the model in the event of face occlusion.

The rest of this paper is organized as follows. In the second part, the related work of driver drowsiness detection is introduced. In the third part, the method and its implementation are described in detail. The fourth part presents the experimental results. Finally, the fifth section summarizes this paper.

2 Related Work

At present, there are three commonly used methods of driver drowsiness detection, which are based on physiological signal detection, driving behavior, and computer vision. To implement the driving drowsiness detection method based on physiological signal, it is necessary to collect the driver's physiological signal through special electronic facilities, and to judge the driver's driving state through the physiological signal. Despite the high detection accuracy of this method, the relevant detection equipment can interfere with the driver and shows poor practicability. The driving drowsiness detection methods based on driving behavior is not invasive, but it is greatly affected by different uncertain factors such as driver's characteristics and external environment. Besides, its robustness is weak.

2.1 Drowsiness Detection Method Based on Physiological Signal

The physiological signals of drivers mainly include EEG (which measures brain activity), ECG (which measure impulses in the heart), EMG (which records the electrical activity of muscles), EOG (which records the electrical activity of eyes), and other electrical signal information. Researchers usually extract the characteristic indexes that can reflect fatigue according to the collected electrical signal information, and treat these characteristic indexes as the input characteristics of drowsiness detection model.

Fu [13] et al. first preprocessed the EEG signal data by conducting independent component analysis, extracted the features of the preprocessed data through fast Fourier transform (FFT), and finally trained the self-organizing fuzzy neural network to establish the model. On the basis of EEG signals, MB Kurt [14] and others combined three different physiological parameters of Ophthalmology, EEG and ECG to carry out study, and applied wavelet transform algorithm and neural network technology to analyze the changes of physiological parameters in the process of driving from wakefulness to fatigue. RN khushaba [15] and others proposed a feature extraction method of wavelet transformation by using the fuzzy mutual information combined with the three physiological parameters of driver's ECG, EEG, and EOG signals. Also, they compared the classification and prediction results of drivers' driving state by means of linear discriminant analysis, KNN, support vector machine and other algorithms. Dr. Bueno [16] et al. studied the EMG signal indexes related to fatigue, including the average rating rate (MNF), median frequency, Dimitrov spectrum index, root mean square, and zero crossing detection. The model is established on the basis of Gaussian mixture model technology.

The invasive problems with physiological signal detection still need to be solved. Although many researchers such as Kobayashi [17] and klingeberg [18] adopted wireless technologies such as ZigBee and Bluetooth, the accuracy of the noninvasive system is relatively low due to motion artifacts and improper electrical level contact. Therefore, the drowsiness detection system based on physiological signals can not solve the problems of low invasiveness and accuracy at the same time.

2.2 Drowsiness Detection Method Based on Driving Behavior

The drowsiness detection based on driving behavior is reliant mainly on the relevant indicators such as vehicle speed, acceleration, lane offset and steering wheel parameters. The degree of driving fatigue is determined by studying the law of these indicators.

Sandberg [19] et al. measured the vehicle speed, lateral position, steering wheel angle, and other vehicle state data during driving, so as to establish the drowsiness driving detection model based on a feedforward neural network. Sayed [20] et al. extracted the steering wheel angle frequency as the discrimination index and established the detection model based on Neural Network. Siegmund [21] et al. collected the operation behavior data of steering wheel and throttle through experiments, selected 46 discrimination indexes, and finally established the drowsiness detection model based on multiple regression. The detection algorithm based on driving behavior usually treats vehicle speed, acceleration, lane offset distance, and so on as the evaluation indexes. However, its reliability is poor due to the different road properties and drivers' personal driving levels.

2.3 Drowsiness Detection Method Based on Computer Vision

Since the above two research methods are not widely used due to their respective limitations, the current research focus on drowsiness detection is based on vision, or the driver's facial features to be exact. Computer vision-based drowsiness detection methods usually use cameras to collect the features of face image to judge the driver's fatigue state. These face features include glasses state, head movement, blinking frequency, and yawning. Researchers extract facial features from camera images, and then use machine learning technologies to establish a fatigue prediction model, such as support vector machine, hidden Markov model or convolutional neural network.

Bhargava Reddy [22] et al. proposed a convolutional neural network model integrating the left and right eyes, face, and mouth states. In order to improve the running speed, they established a fatigue driving prediction model based on the left eye and mouth states. On this basis, they performed network distillation to further compress the model. A. Jamshidi [23] et al. proposed the hierarchical deep drowsiness detection network. This method uses RESNET to classify the driver's condition and LSTM to extract the time feature, which improves the accuracy of fatigue detection. In spite of this, it can not fully exploit the driver's long-term facial features. J. Bai [24] et al.

proposed the two-stream spatial-temporary graph convolutional network. This method uses facial feature points to construct spatial graphs and temporal graphs. However, the graphical topology representing the face is fixed for all graph convolutional layers and all poses. Also, it does not have a solution to drivers' facial occlusion.

3 Richly Activated Graph Convolutional Network

As shown in Fig. 2, the proposed model involves two steps. The first step is to construct graphs according to the two-dimensional coordinate sequence of facial landmarks. The second step is to train the 2s-STTN model according to the driver's continuous sequence of facial landmarks.

In Section 3.1, the facial landmark detection in the data preprocessing stage will be introduced. In Section 3.2, it will be explained how extracted facial landmarks can be used to construct temporal graphs and spatial graphs. In Section 3.3, the concept of graph convolution will be introduced. In Section 3.4, Transformer Self-Attention will be introduced. In Section 3.5, the spatial self-attention module and the temporal self-attention module will be described. In Section 3.6, it will be described how to CAM technology can be use to perform driver fatigue detection under the context of face occlusion.

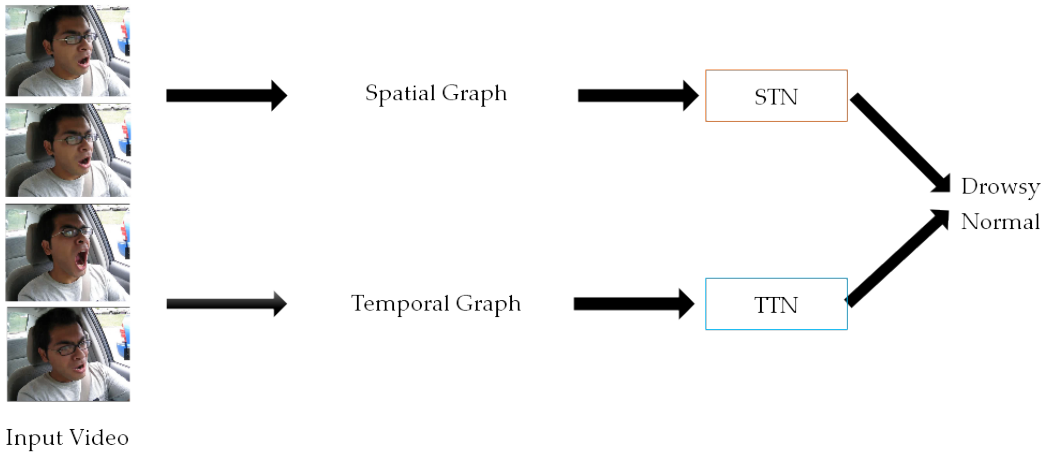

Fig. 2. Our model includes two steps

3.1 Facial Landmark Detection

In the past 10 years, facial landmark point detection algorithms [25-27] have experienced some significant development, their robustness is getting better and better, and there are even good results achieved in some complex scenes.

The facial feature point detection algorithm provides two-dimensional coordinates (x, y) for 70 facial feature points. An array of 70 facial feature points is used to represent the frame sequence of a group of facial landmarks, as shown in Fig. 3.

Fig. 3. An array of 70 facial feature points is used to represent the frame sequence of a group of facial landmarks

3.2 Graph Construction

The edge between adjacent vertices is represented as a vector. Spatial graphs are composed of vertices and temporal graphs are composed of vectors.

As for the construction of spatial graphs, spatial graphs are composed of n facial feature points. Graph G, which shows the connection between vertices, is an adjacency matrix. The vertex set is defined as $V=\{V_i \mid i = 1, \ldots, n\}$. The vertex set consists of all the facial landmarks. The generated facial feature points are usually expressed in the form of two-dimensional coordinates which can be expressed as (x, y).

The constructed spatial map of facial feature points is shown in Fig. 4. In the process of yawning, the driver's mouth and nose in the spatial map provide valuable information for drowsiness detection.

Fig. 4. Spatial graphs are composed of vertices

As for the construction of temporal graphs, temporal graphs represent the connection between video frames, which are composed of vectors. The vertex in frame t is denoted as V_{it}. The vertex in frame t+r is denoted as $V_{i(t+r)}$. Let $V_{it} = (X_1, Y_1)$, $V_{i(t+r)} = (X_2, Y_2)$, then vector $V_{it}V_{i(t+r)} = (X_2 - X_1, Y_2 - Y_1)$. In Fig. 5, the vector connects the same facial feature points in two adjacent frames.

Fig. 5. Temporal graphs are composed of vectors

3.3 Graph Convolution

In Reference [28], the convolution of spatial graph for matrix V_i is defined as

$$f_{out}(v_i) = \sum_{v_j \in S_i} \frac{1}{Z_{ij}} f_{in}(v_j) \cdot w(l_i(v_j)). \tag{1}$$

Where f represents the feature map, S represents the convolution region of V_i, w represents the weight, l_i represents the mapping function, and Z_{ij} represents the cardinality of the corresponding subset. The temporal graph convolution on vector $\overrightarrow{V_{it}V_{i(t+\tau)}}$ can be expressed as

$$f_{\text{out}}\left(\vec{e}_i\right) = \sum_{\vec{e}_j \in \mathcal{T}_i} f_{\text{in}}\left(\vec{e}_j\right) \cdot w\left(l\left(\vec{e}_{\text{j}}\right)\right). \tag{2}$$

$$\vec{e}_i = \overrightarrow{v_{it}v_{i(t+\tau)}}. \tag{3}$$

$$\vec{e}_j = \overrightarrow{v_{jt}v_{j(t+\tau)}}. \tag{4}$$

3.4 Transformer Self-Attention

Vaswani [37] et al. adopted the self-attention mechanism in Transformer model. Initially, the self-attention mechanism was used to enrich word embeddings in the NLP task. The word embeddings in Transformer are compared with neighboring words and their embeddings are mixed based on their relevance. Self-attention can be used to

extract better semantic information and dynamically establish the relationship of each word. Self-attention mechanism can be written as:

$$\text{Attention}(Q,K,V) = \text{soft max}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \ . \tag{5}$$

Where Q, K, and V represent the matrices containing query, key and value vectors, respectively; and d_k refers to the channel dimension of the key vectors.

3.5 Spatial Self-Attention and Temporal Self-Attention Model

As shown in Fig. 6, the spatial self-attention module extracts the relationship between the embedded facial landmarks by calculating the correlation between the facial landmarks in each frame while the temporal self-attention module studies each facial landmark along all frames. The implementation of spatial self-attention module is similar to that of temporal self-attention module.

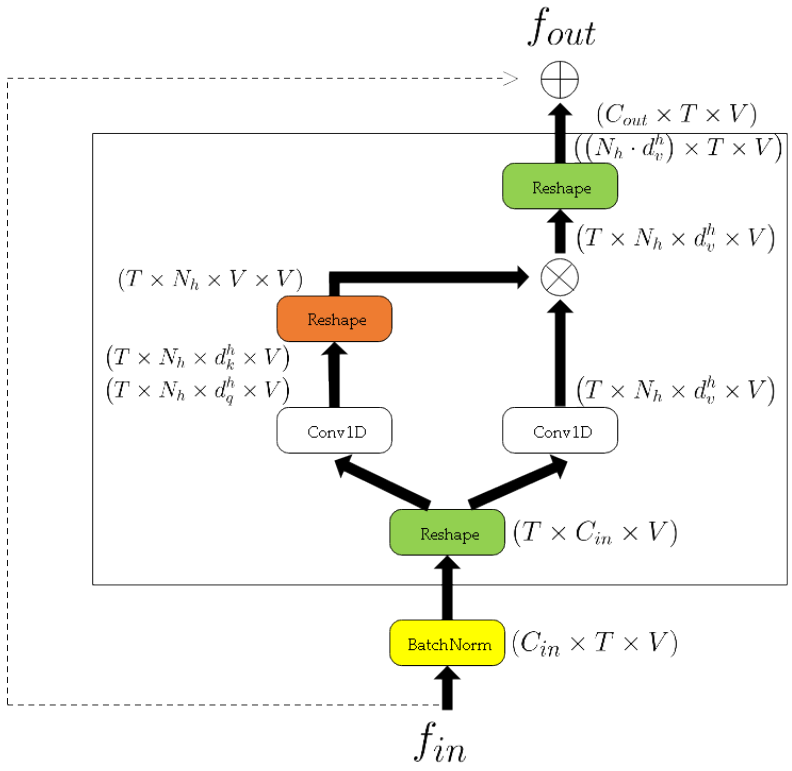

Fig. 6. The implementation of SSA and TSA (⊗ denotes the matrix multiplication.)

For each facial landmark V_i, a query vector q_i, a key vector k_i and a value vector v_i are initialized by applying affine transformation to the embedding of the facial landmark. For facial landmarks (V_i, V_j), a score α_{ij} represents the correlation between two facial landmarks.

$$\alpha_{ij} = \mathbf{q}_i \cdot \mathbf{k}_j \ . \tag{6}$$

$$\mathbf{z}_i = \sum_j softmax_j\left(\frac{\alpha_{ij}}{\sqrt{d_k}}\right)\mathbf{v}_j \quad . \tag{7}$$

Where z_i represents a new embedding for facial landmark v_i. Multi-head attention is applied by repeating this embedding extraction process N_h times.

3.6 Activation Module

As shown in Fig. 7, the STGCN layers use the method of class activation map-ping to distinguish the activation nodes of each flow, and then accumulate the activation mapping of the previous flow to guide the learning process of the new flow. At first, CAM technology is applied to locate the area where the neural network pays attention to the mesh in a forward propagation. In this context, Mc is defined as the class activation graph of class C. For each spatial point, the following formula applies:

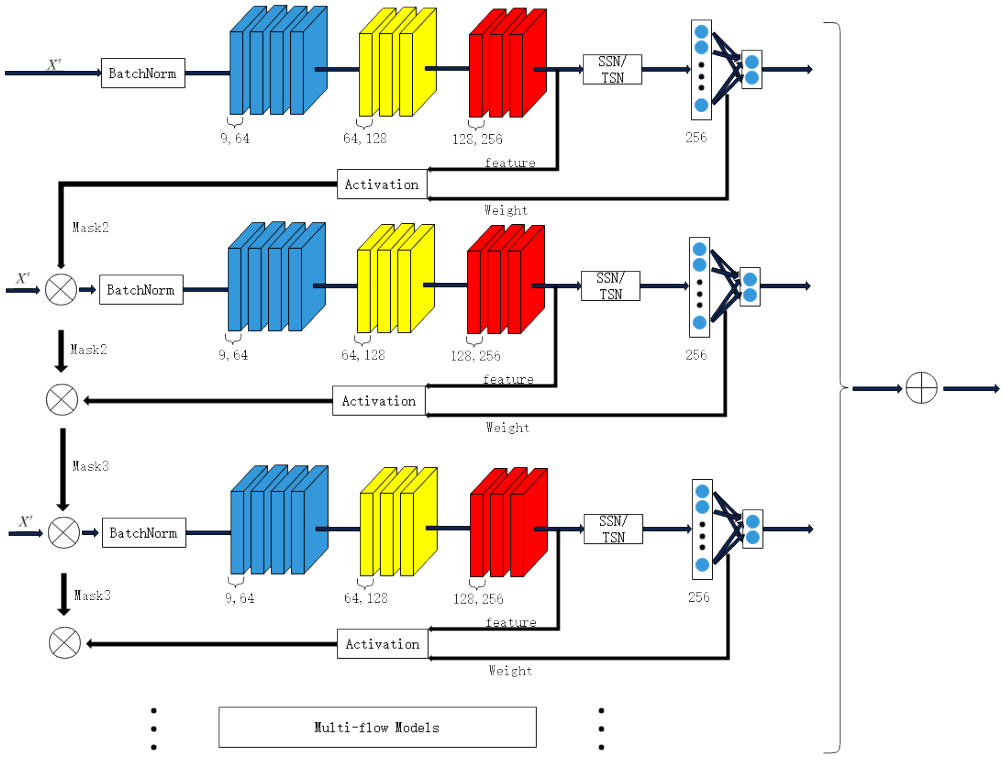

Fig. 7. Class activation mapping technique applied in STGCN layers

(**X**′ represents the temporal graph or the spatial graph after data processing. Each flow is inputted into a single 10 layer graph convolution network. ⊗ and ⊕ denote element-wise multiplication and concatenation, respectively.)

$$M_c(x,y) = \sum_k w_k^c f_k(x,y) \,. \tag{8}$$

Where $f_k(x, y)$ represents the feature map before the global average pooling operation and the weight of class C in the k-th channel. In this paper, the coordinates (x, y) in the image is replaced with the number of frames T and facial feature points I in the face sequence, so as to locate the activated facial feature points. These facial feature

points can also be regarded as the nodes of flow attention. In this context, since class C is a correct classification result, the following formula applies to the mask matrix of flow:

$$\text{mask}_s = \left(\prod_{i=1}^{s-1} \text{mask}_i \right) \otimes \left(1 - \text{Soft}\max\left(M_c^{s-1} \right) \right). \tag{9}$$

$\prod$ represents the element product of all mask matrices before the s-th flow. The initialization of the mask matrix is a matrix of all 1. The input of flow s can be expressed as:

$$\mathbf{x}_s = \mathbf{x}' \otimes \text{mask}_s . \tag{10}$$

Where $\mathbf{x}'$ is the facial landmark data after preprocessing. The input of flow s contains only the facial landmarks that have not been activated by the previous flow. Therefore, 2s-STTN fully explores the identification features of all facial feature points. In the case of face occlusion, 2s-STTN fully mines the facial feature point information of the unobstructed part, so as to improve the robustness of the model.

4 Experiment and Discussion

The experiment was carried out on a Windows 10 system. The CPU is Intel i7 10700k which has 8 cores, 16 threads, a frequency of 3.8GHz, and a memory of 32GB. There are two NVIDIA GeForce GTX 1080 graphics cards with 8GB of memory each. The operating system is Ubuntu 18.04, with the PyTorch neural network framework used.

4.1 Dataset

There are two kinds of data sets in the YAW-DD data set, in which the drivers have different facial features. Videos were collected under different lighting conditions. All videos are in AVI format with a resolution of 640x480 and 30 frames. There is no audio material involved. Each video is provided with a corresponding label, which includes normal driving, speaking and yawning. The first data set contains 322 videos and the second data set contains 29 videos. The sample frames of the Yaw-DD dataset are shown in Fig. 8.

Fig. 8. Exemplary frames of the YAW-DD dataset

The NTHU-DDD data set is composed of five scenes: drivers under normal conditions, drivers wearing glasses, drivers under night conditions, drivers wearing glasses at night and drivers wearing sunglasses. The NTHU-DDD dataset was captured by an infrared camera, and the videos are all in AVI format with a resolution of 640X480. The NTHU-DDD data set provides a label of driver drowsiness detection for each frame.

Since 2s-STTN is a video-based drowsiness detection model, both YAW-DD and HTHU-DDD datasets were processed into video clips.

For the YAW-DD dataset, multiple T-frame videos were clipped from the dataset. Since the labels of the YAW-DD dataset need to meet the requirements of 2s-STTN, all video clips were relabeled. In each video clip, video frames have the same label, either awake or tired. Then, the facial feature points of the video clip were generated by using the facial feature point detection algorithm.

For the HTHU-DDD dataset, multiple T-frame segments were cut from the video according to the given label of a single frame. In each clips, video frames have the same label, either awake or tired. Then, the facial feature points of the video clip were generated by using the facial landmark detection algorithm.

4.2 Details

2s-STTN was implemented with Pytorch. Training was performed using random gradient descent, with momentum set to 0.9. The learning rate decay mechanism was adopted to train the weights finely. Attenuation mechanism means that the learning rate was set to a high value and then attenuated dynamically during training. During the experiment, the learning rate was set to 0.5 and the number of iterations was set to 100. The learning rate attenuated to 0.05, 0.005, 0.0005 and 0.00005 respectively at the end of 20, 40, 60 and 80 iterations of training. Due to limited video memory, batch of training was set to 28. The loss function is the cross entropy loss function.

4.3 Experimental Analysis

The YAW-DD data set was purposed mainly to test the performance of algorithm and model in yawn detection. Yawn is an important feature of driver drowsiness detection. By default, driver fatigue is a positive sample and driver wakefulness is a negative sample. As shown in Table 1, the call rate is greater than the accuracy rate, which indicates that the proposed model is more prone to false alarm, which is beneficial to the safety of drivers.

Table 1. Experimental results on the YAW-DD dataset

True positive rate	Precision	Accuracy	F1 Measure
0.994	0.906	0.943	0.948

Table 2 presents the experimental results of other advanced algorithms on the YAW-DD dataset and the proposed model. Compared with those previous methods, the proposed model can distinguish drivers' states not only in the context of long-term dependency but also in the context of occlusion, which makes the proposed algorithm perform better.

Table 2. Comparison between different methods on the YAW-DD dataset

Method	Platform	Accuracy
Tayyaba et al. [29]	-	83.7%
Mona et al. [30]	CPU-i5	75.0%
Weiwei et al. [31]	-	88.6%
Weiwei et al. [32]	CPU-i7 4700MQ	92.0%
2s-STGCN [24]	GPU-GTX1080	93.4%
2s-STTN (ours)	GPU-GTX1080	94.3%

The NTHU-DDD dataset consists of five scenarios: normal drivers, drivers wearing glasses, drivers wearing glasses at night and drivers wearing sunglasses at night. Table 3 lists the experimental results of the algorithms. Table 4 presents experimental results of other advanced algorithms on the NTHU-DDD dataset and the proposed algorithm. Compared with those previous methods, 2s-STTN is advantageous due to its ability to aggregate the temporal and spatial information collected from different graph convolution streams.

Table 3. Experimental results on NTHU-DDD dataset

Scenario	Drivers under normal conditions	Drivers wearing glasses	Drivers under night conditions	Drivers wearing glasses at night	Drivers wearing sunglasses	Overall
True Positive Rate	0.96	0.95	0.95	0.94	0.94	0.95
Precision	0.88	0.88	0.87	0.89	0.87	0.88
Accuracy	0.94	0.93	0.92	0.92	0.92	0.93
F1 Measure	0.93	0.92	0.91	0.92	0.92	0.92

Table 4. Comparison of different methods on NTHU-DDD dataset

Method	Platform	Spatial features	Temporal features	Accuracy
Tayyaba et al. [29]	-	-	-	78.4%
Park et al. [33]	-	DDD Network	SVM	73.1%
Huynh et al. [34]	GPU-Titan X	3D-CNN	3D-CNN	87.5%
Jie et al. [35]	GPU-M40	MCNN	LSTM	90.1%
2s-STGCN [24]	GPU-GTX1080	Spatial GCN	Temporal GCN	92.7%
2s-STTN (ours)	GPU-GTX1080	Spatial Transformer	Temporal Transformer	93.0%

According to the experimental results, the proposed method is obviously superior to the traditional method of fatigue detection based on facial features. This method can capture the correlation of long-term fatigue movements and make full use of the temporal and spatial information of facial feature points, which is unavailable to most fatigue detection methods. Different solutions are adopted for shading by sunglasses, light changes and angle changes respectively. 2s-STTN model uses temporal self-attention mechanism and spatial self-attention mechanism to improve the accuracy of model prediction under the context of light changes and angle changes. 2s-STTN model uses CAM technology to improve the accuracy of model prediction under the shading by sunglasses. Presumably, that's why our model gets higher scores.

5 Summary and Outlook

In this paper, a driver drowsiness detection model is proposed on the basis of STGCN and class activation map technology. Our model has various class activation maps. Different activated facial features are recognized by the class activated map technology. Each flow recognizes different activated facial features, extracts spatial or temporal features, and integrates facial feature information to improve the performance of the system.

During the fatigue driving system test, it was found out there were some reasons that led to the decline of prediction results and consideration was given to how the model can be improved according to the problem itself. Different solutions were adopted for sunglasses occlusion, light change and angle change. From the experimental results, it can be seen that this method has good performance in driver drowsiness detection, and the proposed model can be applied in the fatigue detection system. Since this paper uses face detection and key point location algorithm based on the driver's face image, the algorithm may fail when there is a large range of occlusion on the face. In the future, it is necessary to continue optimizing the driver's face detection and key point location algorithm, so as to ensure that the face can be detected and key points can be located in case of large range occlusion. In this paper, the driver is only tested for fatigue driving. In the actual driving environment, there are still various dangerous behaviors such as making phone calls and smoking. In the next step, the fatigue detection can be combined with the driver's posture detection to detect whether there are dangerous behaviors such as making phone calls and smoking, and the driver's risk warning can be issued together with the fatigue driving behavior. In the future, the network model will be further optimized and data sets will be collected from real drivers to support our further research.

References

[1] J.M. Owens, T.A. Dingus, F. Guo, Y. Fang, M. Perez, J. McClafferty, B.C. Tefft, Prevalence of Drowsy Driving Crashes: Estimates from a Large-Scale Naturalistic Driving Study. <https://aaafoundation.org/prevalence-drowsy-driving-crashes-estimates-large-scale-naturalistic-driving-study/>, 2018.

[2] K.H. Lee, W. Kim, H.K. Choi, B.T. Jang, A Study on Feature Extraction Methods Used to Estimate a Driver's Level of Drowsiness, in: Proc. 2019 21st International Conference on Advanced Communication Technology (ICACT), 2019.
[3] R.O. Mbouna, S.G. Kong, M.G. Chun, Visual Analysis of Eye State and Head Pose for Driver Alertness Monitoring, IEEE Transactions on Intelligent Transportation Systems 14(3)(2013) 1462-1469.
[4] A. Dasgupta, A. George, S.L. Happy, A. Routray, A Vision-Based System for Monitoring the Loss of Attention in Automotive Drivers, IEEE Transactions on Intelligent Transportation Systems 14(4)(2013) 1825-1838.
[5] B. Mandal, L. Li, G.S. Wang, J. Lin, Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State, IEEE Transactions on Intelligent Transportation Systems 18(3)(2017) 545-557.
[6] L.M. Bergasa, J. Nuevo, M.A. Sotelo, R. Barea, M.E. Lopez, Real-time system for monitoring driver vigilance, IEEE Transactions on Intelligent Transportation Systems 7(1)(2006) 63-77.
[7] Y. Xie, K. Chen, Y.L. Murphey, Real-time and Robust Driver Yawning Detection with Deep Neural Networks, in: Proc. 2018 IEEE Symposium Series on Computational Intelligence (SSCI), 2018.
[8] W.H. Gu, Y. Zhu, X.D. Chen, L.F. He, B.B. Zheng, Hierarchical CNN-based real-time fatigue detection system by visual-based technologies using MSP model, IET Image Processing 12(12)(2018) 2319-2329.
[9] N. Alioua, A. Amine, M. Rziza, Driver's Fatigue Detection Based on Yawning Extraction, International Journal of Vehicular Technology (2014) 678786.
[10] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning Deep Features for Discriminative Localization, in: Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (Cvpr), 2016.
[11] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, B. Hariri, YawDD: A yawning detection dataset, in: Proc. 2014 ACM Multimedia Systems (MMSys), 2014.
[12] C.H. Weng, Y.H. Lai, S.H. Lai, Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network, in: Proc. Computer Vision- ACCV 2016 International Workshops, 2017.
[13] F.C. Lin, L.W. Ko, C.H. Chuang, T.P. Su, C.T. Lin, Generalized EEG-Based Drowsiness Prediction System by Using a Self-Organizing Neural Fuzzy System, IEEE Transactions on Circuits and Systems I-Regular Papers 59(9)(2012) 2044-2055.
[14] M.B. Kurt, N. Sezgin, M. Akin, G. Kirbas, M. Bayram, The ANN-based computing of drowsy level, Expert Systems with Applications 36(2)(2009) 2534-2542.
[15] R.N. Khushaba, S. Kodagoda, S. Lal, G. Dissanayake, Driver Drowsiness Classification Using Fuzzy Wavelet-Packet-Based Feature-Extraction Algorithm, IEEE Transactions on Biomedical Engineering 58(1)(2011) 121-131.
[16] D.R. Bueno, J.M. Lizano, L. Montano, Muscular fatigue detection using sEMG in dynamic contractions, in: Proc. 2015 Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015.
[17] H. Kobayashi, EMG/ECG acquisition system with online adjustable parameters using ZigBee wireless technology, Electronics and Communications in Japan 96(5)(2013) 1-10.
[18] T. Klingeberg, M. Schilling, Mobile wearable device for long term monitoring of vital signs, Computer Methods and Programs in Biomedicine 106(2)(2012) 89-96.
[19] D. Sandberg, M. Wahde, Particle swarm optimization of feedforward neural networks for the detection of drowsy driving, in: Proc. 2008 IEEE International Joint Conference on Neural Networks, 2008.
[20] R. Sayed, A. Eskandarian, Unobtrusive drowsiness detection by neural network learning of driver steering, Proceedings of the Institution of Mechanical Engineers Part D-Journal of Automobile Engineering 215(D9)(2001) 969-975.
[21] G.P. Siegmund, D.J. King, D.K. Mumford, Correlation of Heavy-Truck Driver Fatigue with Vehicle-Based Control Measures, International Truck & Bus Meeting & Exposition (1995) 441-468.
[22] B. Reddy, Y.H. Kim, S. Yun, C. Seo, J. Jang, Real-time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks, in: Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (Cvprw), 2017.
[23] S. Jamshidi, R. Azmi, M. Sharghi, M. Soryani, Hierarchical deep neural networks to detect driver drowsiness, Multimedia Tools and Applications 80(10)(2021) 16045-16058.
[24] J. Bai, W.T. Yu, Z. Xiao, V. Havyarimana, A.C. Regan, H.B. Jiang, L.C. Jiao, Two-Stream Spatial-Temporal Graph Convolutional Networks for Driver Drowsiness Detection, IEEE Transactions on Cybernetics 52(12)(2022) 13821-13833.
[25] M.L. Zhu, D.M. Shi, M.J. Zheng, M. Sadiq, Robust Facial Landmark Detection via Occlusion-adaptive Deep Networks, in: Proc. 2019 IEEE/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr 2019), 2019.
[26] H.W. Zhang, Q. Li, Z.N. Sun, Y.F. Liu, Combining Data-Driven and Model-Driven Methods for Robust Facial Landmark Detection, IEEE Transactions on Information Forensics and Security 13(10)(2018) 2409-2422.
[27] F.M. Sukno, J.L. Waddington, P.F. Whelan, 3-D Facial Landmark Localization With Asymmetry Patterns and Shape Regression from Incomplete Local Features, IEEE Transactions on Cybernetics 45(9)(2015) 1717-1730.
[28] S.J. Yan, Y.J. Xiong, D.H. Lin, Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition, in: Proc. 2018 Thirty-Second AAAI Conference on Artificial Intelligence/ Thirtieth Innovative Applications of Artificial Intelligence Conference/ Eighth Aaai Symposium on Educational Advances in Artificial Intelligence, 2018.
[29] T. Azim, M.A. Jaffar, A.M. Mirza, Fully automated real time fatigue detection of drivers through Fuzzy Expert Systems, Applied Soft Computing 18(2014) 25-38.

[30] M. Omidyeganeh, S. Shirmohammadi, S. Abtahi, A. Khurshid, M. Farhan, J. Scharcanski, B. Hariri, D. Laroche, L. Martel, Yawning Detection Using Embedded Smart Cameras, IEEE Transactions on Instrumentation and Measurement 65(3)(2016) 570-582.
[31] W. Zhang, Y.L. Murphey, T.Y. Wang, Q.J. Xu, Driver Yawning Detection based on Deep Convolutional Neural Learning and Robust Nose Tracking, in: Proc. 2015 International Joint Conference on Neural Networks (Ijcnn), 2015.
[32] W.W. Hang, J.Y. Su, Driver Yawning Detection based on Long Short Term Memory Networks, in: Proc. 2017 IEEE Symposium Series on Computational Intelligence (Ssci), 2017.
[33] S. Park, F. Pan, S. Kang, C.D. Yoo, Driver Drowsiness Detection System Based on Feature Representation Learning Using Various Deep Networks, in: Proc. 2016 Computer Vision - Accv 2016 Workshops, 2016.
[34] X.P. Huynh, S.M. Park, Y.G. Kim, Detection of Driver Drowsiness Using 3D Deep Neural Network and Semi-Supervised Gradient Boosting Machine. in: Proc. 2016 Computer Vision - Accv 2016 Workshops, 2016.
[35] J. Lyu, Z. Yuan, D. Chen, Long-term Multi-granularity Deep Framework for Driver Drowsiness Detection. <https://arxiv.org/abs/1801.02325>, 2018.
[36] K.M. He, X.Y. Zhang, S.Q. Ren, J. Sun, Deep Residual Learning for Image Recognition. in: Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (Cvpr), 2016.