A deep learning-based method for nighttime sleep posture recognition

Sheng Wang
Dpartment of Converged communication systems
China Mobile (Hangzhou) Information Technology Co., Ltd
Hangzhou,China
wangshenghz@cmhi.chinamobile.com

Yujia Liu *
Dpartment of Converged communication systems
China Mobile (Hangzhou) Information Technology Co., Ltd
Hangzhou,China
liuyujia63@gmail.com
ORCID: 0000-0002-1561-2582

Xue Xu
Dpartment of Converged communication systems
China Mobile (Hangzhou) Information Technology Co., Ltd
Hangzhou,China
xuxue@cmhi.chinamobile.com

Junjie Zhao
Dpartment of Converged communication systems
China Mobile (Hangzhou) Information Technology Co., Ltd
Hangzhou,China
zhaojunjie@cmhi.chinamobile.com

Bangzhen Xu
Dpartment of Converged communication systems
China Mobile (Hangzhou) Information Technology Co., Ltd
Hangzhou, China
xubangzhen@cmhi.chinamobile.com

2023 IEEE 6th International Conference on Computer and Communication Engineering Technology (CCET) | 979-8-3503-4068-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/CCET59170.2023.10335134

***Abstract*—Sleep is indispensable for human health. Sleeping posture is one of the factors that affect the quality of sleep. Besides, correct sleeping postures could prevent and treat diseases. Conversely, unsuitable sleeping posture would do harm to human health. Therefore, it is of great significance to accurately and quickly distinguish human sleeping posture for improving sleep quality and formulating disease treatment decision. Therefore, in order to further enhance the accuracy of sleeping posture recognition, we used a Infrared night vision camera to collect sleeping data, and proposed a deep learning-based predictive model (DL-model). The DL-model introduced the attention mechanism to and capture the characteristics of significantly related to the target task and reduce the expression of redundant characteristics. Meanwhile, considering that when the blanket is covered, some human features would be blocked, we took whether the blanket covered as auxiliary task. The experimental results showed that the average accuracy of the DL-model reaches 94.62%, which was better than the other state-of-art deep learning methods. Compared with signal task learning, the DL-model using multi-task learning performed well. Also, the DL-model we proposed in this study also has a good performance in the auxiliary task--whether to cover the blanket. To some extend, it confirmed the feasibility of using the monitoring camera for sleeping posture prediction, which is of great significance for realizing the real-time prediction and sleep quality assessment at night.**

Keywords—deep learning; Attention mechanism; multi-task learning; Sleeping posture classification

I. INTRODUCTION

Sleep is important for human health. There are various factors affecting sleep quality [1]-[2]. For example, sleeping time, sleep posture and so on [3]. Correct and good sleeping posture can prevent and treat diseases and improve sleep quality; on the contrary, bad sleeping posture may make people feel tired and uncomfortable, and even damage health [2]. Therefore, accurate recognition of people's sleeping posture is of great significance for the evaluation of sleep quality.

Nowadays, researchers have proposed many different types of sleep monitoring methods to identify a person's sleeping posture and behavior [4]. For example, mattresses with pressure sensors and polysomnogram are useful methods for capture different sleeping postures [5]-[6]. It is undeniable that these studies have achieved satisfactory ability in the sleep posture recognition task, but but these methods required people to wear additional equipment while sleeping, which may damage people's sleep quality.

E-health systems, depending on Internet of things technology (IoT) and 5G Internet infrastructure, provides us with new ideas for sleep position detection. Surveillance cameras have been widely used in our daily live, which could quickly acquire sleep image data, and people do not need to wear additional equipment during sleep. At present, researchers have carried out a series of image-based sleep posture classification studies, and have achieved good results. However, the above researches have some limitations: in real life, the environment of human sleeping is complex and diverse, and when covered with blankets, some body features would be blocked, reducing the available effective image information, which increases the difficulty of recognizing sleeping positions [7].Therefore, in order to improve the accuracy of sleeping posture classification task, the variable whether human cover the blanket or not should be token into account for extracting effective and potential features as much as possible.

With the rapid development of computer vision, deep learning as an effective feature extraction and analysis method were applied into image recognition, segmentation and other fields [8]-[10]. Deep learning method may solve the above problems. Multi-task learning (MTL) could simultaneously implement two or more prediction tasks. It acquires potential

relationship between different tasks and filter noise features utilizing feature sharing and finally achieves the effect of improving the performance of the model and enhancing the robustness of the model. This method has achieved remarkable results in many applications such as computer vision, medical image analysis and speech recognition [11]-[12]. In addition, inspired by human perception, researchers have explored the attention mechanism to strengthen the representative features and weaken the redundant features [13]-[14]. Common attention mechanisms include channel attention, spatial attention mechanism, and convolution block attention that fuses channel attention and spatial attention networks and so on [15]-[16]. These attention mechanisms have been proven to be effective in improving model performance.

Therefore, in order to improve the performance of sleeping posture at night, we used the technologies for IoT and 5G to acquire the human sleeping postures dataset at night, and proposes a new prediction model based on deep learning method. The deep learning-based model (DL-model) used the attention mechanism to learn more effective features from the target area; at the same time, put "whether the blanket is covered or not" as an auxiliary task to further improve the sleep position prediction performance of the model. The experimental results showed that the DL-model proposed could accurately recognize sleeping postures the under poor nighttime lighting conditions. In addition, when taking the auxiliary task of "whether to cover the blanket", the performance of the model in predicting sleeping postures has been further improved.

II. METHODS

A. Datasets building and Pre-procession

In this study, a surveillance camera with infrared night vision function was used to collect a human sleep image dataset at night in a real family situation. Considering that sleep images taken at different locations are different, in this study we selected two positions to collect sleep images: at the end of the bed and at the side of the bed.

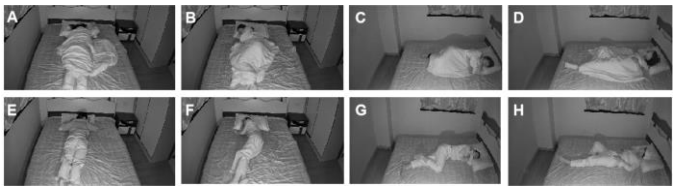

Fig. 1. Examples of sleeping images token by surveillance camera surveillance camera in different perspectives

The processes of data collection were as follows: firstly, fixing the camera at the end of the bed or on the side of the bed; then, using the surveillance camera to record the human sleeping images, and saved them in the form of video. It should be noted that the collected images of sleeping positions include four types: supine, prone, left side and right side, and each sleeping position has two situations: covering the quilt and not covering the quilt. In order to obtain clear sleep data, during data collection, subjects can randomly change sleeping positions, and each subject was required to maintain the duration of each sleeping position for about 2 seconds. Fig. 1A-H were the example data of sleeping images collected by the surveillance camera. The size of the image collected by the surveillance camera is 2304×1296.

Since the position of the camera was fixed during data collection, we intercept video images second by second, and then use a method based on image difference method to automatically obtain the target image area, and delete redundant images in the video at the same time. Specifically, by comparing and analyzing the image difference between the 1.5 seconds before and after in the video, when the difference in the variation range of the pixel values in the same area of the two images exceeds a certain threshold, it was considered that the point has changed.

$$D_{(x,y)} = \begin{cases} 1, \left| I_{(x,y)} - I'_{(x,y)} \right| > T_1 \\ 0, \left| I_{(x,y)} - I'_{(x,y)} \right| \le T_1 \end{cases} \quad (1)$$

$I'_{(x,y)}$ represents the image of a previous second, $I_{(x,y)}$ represents the current image, and $D_{(x,y)}$ represents the binary image obtained by subtracting the image matrix of the two seconds before and after. The pixel point in the image is 1, which means that after the image in the area changes.

$$\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} D_{(x,y)} \ge T_2 \quad (2)$$

When the cumulative number of changed pixels exceeds T_2, the image was considered to be different from the previous image, and then included into the data set.

Finally, there were 3906 sleeping posture images selected from 16 people, including left side (819 pictures), right side (868 pictures), supine (1109 pictures) and prone (1110 pictures). Noted that each image has one person.

During the training process, the image data of the training set is amplified to avoid over-fitting of the model , to enhance the generalization of the model. Amplify the data set by random cropping, random scaling and vertical flipping of the image, and then randomly crop it to a size of 225×225 and input it into the model for training. In the test set, only the image was resized Processing. In addition, use the z-zero method to normalize the collected grayscale image data, and use the processed image matrix as output, and send it to the network model to speed up the convergence of the network and improve the performance of the model.

$$I'_{(x,y)} = \frac{I_{(x,y)} - u}{\delta} \quad (3)$$

u represents the mean value of the pixels of the image matrix; δ represents the variance of all pixels in the image,

$I'_{(x,y)}$ represents the normalized image, $I_{(x,y)}$ represents the original image.

By the way, the images of the target area segmented from the image was resized into the same size (224×224), and then sent to the model for feature extraction.

B. Deep learning-based predicting model construction

The basic structure of the model is shown in Fig. 2A. Our DL-model could comprise public feature extraction module and multi-task learning module. The public feature extraction model was used to extract the image features significantly related to the target task from images through weight sharing, and the multi-task learning module were applied for extracting the specific features for sleeping posture and "whether the blanket covered or not" classification task. Inspired by Resnet [17], each of module were made up of two Resnet blocks. In addition, each residual block contained several convolution units, and a convolution units include a convolution layers, a batch normalization layer and a ReLU activation layer. There was a skip connection between every two convolution units.

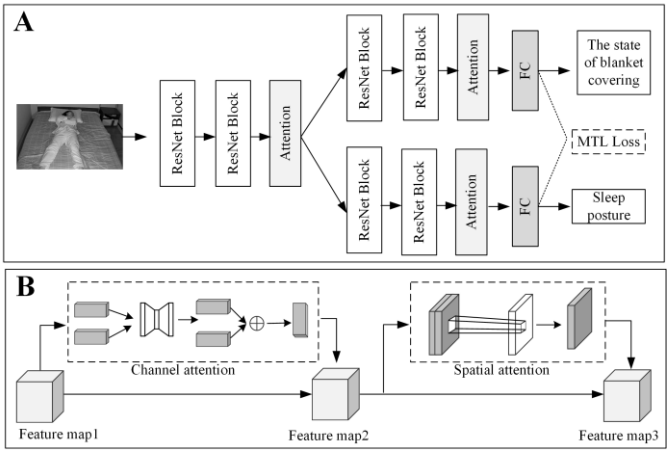

Fig. 2. Overview of the DL-model we constructed.

Besides, we inserted an attention module to enhance the relevance of the extracted features to the target task. The attention module was divided into channel attention module and spatial attention module. The channel attention module could find the correlations between feature maps, and the spatial attention module has advantage in enhancing specific target features by regions of interest and weakening irrelevant background regions. Previous studies have shown that a convolution block attention module (CBAM) that fused channel attention and spatial attention performed well in classification and detection tasks [18]. Therefore, we applied the into our DL-model, which was embedded CBAM module behind the last residual block of the public feature extraction module and multi-task learning module, respectively. The CBAM model could enhance the extracted features and make the model focus more on the features that was significantly relevant to the target task and suppress the representation of redundant features. Fig.2B shows the detailed structure of the attention module.

The prediction network proposed in this study mainly consists of two tasks. For the sleeping posture recognition task, a multivariate cross-entropy loss was used, while for the binary prediction task of "whether the blanket covered or not", it was used a binary cross-entropy loss function. Considering the difference in the range of loss changes between different tasks, different weights were used for different tasks, and its multi-task loss function ($Loss_{MTL}$) could be expressed as:

$$Loss_{MTL} = \sum_{i=1}^{n} w_i \cdot L_i = w_1 \cdot L_1 + w_2 \cdot L_2 \tag{4}$$

In this study, the loss weight sums of the two tasks wer set to 0.5 and 1, respectively. L_1 was the loss function for the prediction task of "Whether to cover the blanket", and L_2 was the loss function for the prediction of the sleeping posture task.

$$L_1 = -\frac{1}{N} \cdot \sum \left[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i) \right] \tag{5}$$

$$L_2 = -\frac{1}{N} \sum_{i} \sum_{c}^{M=4} y_{ic} \log(p_{ic}) \tag{6}$$

where, N indicated the number of samples, y_{ic} and y_i represented the category predicted by the model, p_i and p_{ic} indicated the probability predicted by the model.

III. Experimental Results and Analysis

The sleeping dataset we collected was randomly divided according to the ratio of 7:3, and there were 2734 images (795 supine images, 762 prone images, 562 left-lying images and 615 right-lying images) in the training cohort, and there were 1151 (347 supine, 315 prone, 257 left and 253 right) images in validation cohort.

A prediction model based on the attention mechanism and multi-task learning was established on the collected nighttime sleep dataset. The average accuracy of the sleeping position recognition model was 94.62%, among which the recognition accuracy of lying on the right side reached 99.21%. It surpassed the other studies [5]-[6]. The prediction effect of task cover was 98.81%. The results showed that the model proposed in this study could accurately identify the sleeping position of people at night and whether the blanket were covered or not. In other words, the model could well distinguish the sleeping position of the human body under the condition of covering and not covering the blanket. In order to further verify the performance of the model, common deep learning networks such as Resnet [17] and DenseNet [19] would be used to establish a sleeping posture prediction model, As shown in Table 1, the comparison showed that the prediction accuracy of the method proposed in this study was the highest.

TABLE I. THE ACCURACY OF THE SLEEPING POSITION MODLE BUILDING BY COMMON CLASSIFICATION NETWORK MODELS

newtwork	*Average accuray*	*prone*	*supine*	*Left side*	*Right side*
ResNet	*91.21%*	*98.84%*	*88.23%*	*90.66%*	*84.98%*
DenseNet	*89.33%*	*78.96%*	*93.02%*	*96.89%*	*91.30%*
DL-model	*94.62%*	*91.07%*	*95.87%*	*99.39%*	*99.21%*

In addition, we also conducted ablation experiments to verify the effectiveness of the proposed model. As shown in TABLE II, the model proposed in this study has the best effect in the recognition of sleeping postures, and also has a good performance in the recognition of the auxiliary task of being covered. By comparison, it was found that the predictive performance of the model that does not introduce attention mechanism or only uses multi-task learning was slightly lower than the model proposed in this study, but higher than the basic network without attention mechanism and multi-task learning. Compared with the single-task sleeping posture prediction model, the prediction accuracy of the model proposed in this study could be improved by 3%. Therefore, the model proposed in this study that using attention mechanism and multi-task learning could improve the accuracy of sleeping posture recognition to a certain extent.

TABLE II. ABLATION EXPERIMENTS: PERFORMANCE OF PREDICTIVE MODELS BUILT USING DIFFERENT MODULES

Model	*Accuracy*
Baseline	*91.21%*
Baseline + MTL	*93.38%*
*Resnet+attention+MTL**	*94.62%*

Resnet+attention+MTL* means the proposed DL-model.

In addition, we also explored the loss weights of each task in the multi-task learning module. We fixed the weight ω in the sleeping posture recognition task and set it to 1, and only change the weight of the cover quilt recognition task. Fig. 3 shows the sleep position recognition prediction effect of the model when different values were selected. When it was set to 0.5, the accuracy of the average sleep position detection was the highest, so the model under this weight was selected as the final model of this study.

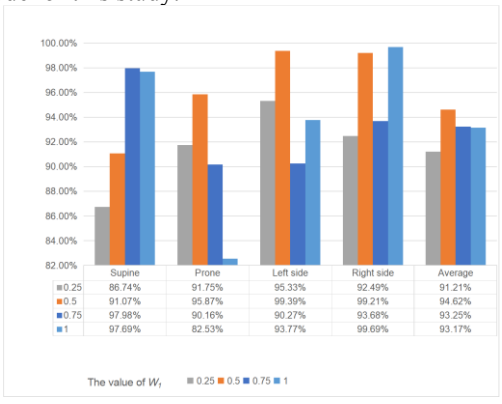

Fig. 3. The performance of Multi-task learning model under different W weights

IV. EXPERIMENTAL SETUP

A large number of previous experiments have shown that this method can improve the classification effect and speed up the model training. Therefore, when training the model in this study, ImageNet was firstly used for pre-training,which could speed up convergence and avoid over-fitting of the model. Epoch was set to 80. Batch size was set as 64. We used the Adam optimizer to optimize the model, the initial learning rate was set to 1e-4, and the learning rate was adjusted by exponential decay, besides, the learning rate was adjusted every 20 steps. The magnitude of the decrease was 0.8.

The experiments were mainly implemented using the Pytorch-based development framework; all programs were developed in Python (version 3.8.5, https://www.python.org/) 3.8.5.

V. CONCLUSION

In order to realize accurate recognition of sleeping posture at night, this study collects and constructs a dataset of nighttime sleeping posture recognition based on surveillance cameras in real family scenes, and proposes a sleeping posture recognition model based on attention mechanism and multi-task. Experiments showed that the model proposed in this study had a good effect on the recognition of sleeping posture at night, and its prediction performance was better than that of common classification deep learning methods. The model could not only realize the accurate recognition of sleep posture, but also has the auxiliary function of cover detection, which improves the utilization rate of the model.

Therefore, this study confirms the feasibility of using 5G and IoT technologies to identify human sleeping positions. In the future, we will study direct analysis of surveillance video to predict sleep position changes at night in real time. Also, we would apply the IoT and 5G technology to the field of human sleep quality to explore the correlation between sleeping posture and sleep quality, and establish a non-invasive human sleep quality assessment model.

REFERENCES

[1] Cudney L E, Frey B N, McCabe R E, et al. Investigating the relationship between objective measures of sleep and self-report sleep quality in healthy adults: a review[J]. Journal of Clinical Sleep Medicine, 2022, 18(3): 927-936.

[2] Xing Z, Gao W, Chuai G. Research on sleeping position recognition algorithm based on human body vibration signal[C]//2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA). IEEE, 2022: 403-406.Tang K, Kumar A, Nadeem M, et al. CNN-based smart sleep posture recognition system[J]. IoT, 2021, 2(1): 119-139.

[3] Ye, S.; Eum, S. Implement the system of the Position Change for Obstructive sleep apnea patient. J. Korea Inst. Info. Comm. Eng. 2017, 21, 1231–1236.

[4] Islam S M M, Lubecke V M. Sleep posture recognition with a dual-frequency microwave Doppler radar and machine learning classifiers[J]. IEEE Sensors Letters, 2022, 6(3): 1-4.

[5] Enayati M, Skubic M, Keller J M, et al. Sleep posture classification using bed sensor data and neural networks[C]//2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018: 461-465.

[6] Han P, Li L, Zhang H, et al. Low-cost plastic optical fiber sensor embedded in mattress for sleep performance monitoring[J]. Optical fiber technology, 2021, 64: 102541.

[7] Tam A Y C, So B P H, Chan T T C, et al. A blanket accommodative sleep posture classification system using an infrared depth camera: A deep learning approach with synthetic augmentation of blanket conditions[J]. Sensors, 2021, 21(16): 5553.

[8] Li Y. Research and application of deep learning in image recognition[C]//2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA). IEEE, 2022: 994-999.

[9] Zaidi S S A, Ansari M S, Aslam A, et al. A survey of modern deep learning based object detection models[J]. Digital Signal Processing, 2022: 103514.

[10] Gul S, Khan M S, Bibi A, et al. Deep learning techniques for liver and liver tumor segmentation: A review[J]. Computers in Biology and Medicine, 2022: 105620.

[11] Nazir M, Ali M J, Tufail H Z, et al. Multi-task learning architecture for brain tumor detection and segmentation in MRI images[J]. Journal of Electronic Imaging, 2022, 31(5): 051606.

[12] Amyar A, Modzelewski R, Vera P, et al. Multi-task multi-scale learning for outcome prediction in 3D PET images[J]. Computers in Biology and Medicine, 2022, 151: 106208.

[13] Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: A survey[J]. Computational Visual Media, 2022, 8(3): 331-368.

[14] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[15] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[16] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[17] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[18] Jiang M, Song L, Wang Y, et al. Fusion of the YOLOv4 network model and visual attention mechanism to detect low-quality young apples in a complex environment[J]. Precision Agriculture, 2022: 1-19.

[19] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.