SIN-CHN
SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2024, 21(4), 602.
https://doi.org/10.62617/mcb602

Article

Ice and snow sports behavior recognition based on multi-scale features and improved CBAM

Chunping Liu

Department of Winter Sports, Jilin Sport University, Changchun 130022, China; 17386821324@163.com

CITATION

Liu C. Ice and snow sports behavior recognition based on multi-scale features and improved CBAM. Molecular & Cellular Biomechanics. 2024; 21(4): 602. https://doi.org/10.62617/mcb602

ARTICLE INFO

Received: 23 October 2024
Accepted: 11 November 2024
Available online: 10 December 2024

Abstract: Accurately identifying and correcting erroneous sports behaviors of athletes or beginners in ice and snow sports can improve the training quality. However, ice and snow sports scenes often have complex motion backgrounds, and the behavioral features during motion are difficult to extract, which affects the recognition accuracy. In order to solve the feature extraction in ice and snow sports behavior recognition, a behavior recognition model based on multi-scale features and improved convolutional block attention module is proposed. The model first utilizes multi-scale features to obtain multi-level features from the collected ice and snow motion images, ensuring that features of different scales in the images can be effectively captured. Then, one-dimensional convolution and spatial random pooling layers are introduced to improve the convolutional attention module, thereby constructing a behavior recognition model. The accuracy of the proposed model in the Ski-Pose dataset was 98.3%, which was 8.2% and 13.7% higher than other recognition models, indicating an obvious gap. The accuracy and $F1$ value were 89.5% and 91.2%, respectively, and the recognition rate for small targets reached 80%, which verified the effectiveness of the model. The research provides new technological support for intelligent monitoring and analysis systems for ice and snow sports.

Keywords: multi-scale features; CBAM; ice and snow sports; behavior recognition; algorithm optimization

1. Introduction

In recent years, ice and snow sports (ICS) have gradually become a highly anticipated sport worldwide. Especially after the Winter Olympics, people's enthusiasm for ICS reached its peak [1,2]. At the same time, the gradual improvement of sports facilities has lowered the threshold for ICS, which has led to a sharp increase in the participants in these sports. With the rapid development of ICS, the recognition and analysis of sports behavior have become increasingly important. Recognizing sports behavior can not only reduce the training difficulty for amateurs when they first encounter ICS, but also enhance the training efficiency [3]. However, due to the dynamic and complex outdoor environment, it is greatly affected by factors such as lighting and weather. It is easy to be interfered by various small target objects during recognition, and a single scale feature is difficult to face these challenges, resulting in low recognition efficiency [4,5]. How to improve the recognition accuracy and robustness of ICS behavior in complex environments has become an urgent problem.

In recent years, the Convolutional Block Attention Module (CBAM), which integrates channel attention and spatial attention, has performed well in various deep learning tasks and has been extensively applied. He et al. [6] proposed a fall

detection approach on the basis of parallel 2D Convolutional Neural Network (CNN) and CBAM. This method used pulse compression to update the image in the radar echo signal. The image was sent to CBAM for recognition and judgment of whether there was a fall or other action. The introduced CBAM achieved multi-domain standards for target behavior recognition, improve image resolution, and effectively reduce the probability of false alarms and missed alarms in fall detection [6]. Feng et al. [7] built a CNN on the ground of CBAM for pneumonia detection in chest X-ray images. This network calculated the channel weights of images by adding CBAM modules, and focused more on the information space part of the feature map, solving the redundancy problem of the feature map. The accuracy and recall on the pneumonia dataset were both higher than 95%, indicating that its discriminative effect on images was good [7]. Deng et al. [8] proposed a learning semi-supervised automatic modulation recognition method on the ground of multi-modal information and domain adversarial networks. This method mined potential knowledge information of unlabeled target domain data by introducing domain adversarial training, and enhanced the network's ability to signify key features by introducing CBAM. Compared with existing schemes, this scheme had higher average classification accuracy and higher adaptability in different network structures [8]. Jiang et al. [9] built a Multi-Scale Feature (MSF) fusion network on the ground of CBAM. This network utilized spatiotemporal convolution blocks to extract temporal and spatial features of EEG signals. The CBAM was applied to process and classify changes in different objects. The performance of the feature fusion network was improved, and the average accuracy of the network in EEG wake-up was as high as 99%, verifying the effectiveness of the scheme [9]. Chang [10] proposed an intelligent bearing fault diagnosis model that combined CBAM module and optimized residual network structure. The model first introduced CBAM to enhance data feature extraction, and then optimized the residual network to reduce the complexity of the model. The feature extraction ability of this model was superior to traditional deep network models, which improved the feature extraction ability [10].

There are also many studies on behavior recognition. Pan et al. [11] proposed an animal behavior recognition method on the basis of CNN. This method used wearable sensors to collect animal behavior and constructed action images in the sensor data stream, distinguishing animal behavior types in a CNN. This method had a high accuracy in behavior monitoring [11]. Pang et al. [12] designed a pedestrian trajectory prediction approach based on adversarial generative networks. This model introduced an adversarial generative network to construct a feature module for processing the collected pedestrian images and making predictions. The model could improve the accuracy and diversity of pedestrian trajectory prediction [12]. Huang et al. [13] designed a real-time driver behavior detection model on the ground of deep learning. This model consisted of an inverted residual network with depth separable convolutions, and introduced attention mechanism into the nonlinear transformation layer. The accuracy of the model on the driving test dataset was 95.17%, and its real-time, accuracy, and reliability were superior to current behavior recognition models [13]. Sun et al. [14] built a car driving behavior management approach on the ground of adaptive soft deep reinforcement learning. This model used high-performance driving behavior recognition to calculate the optimal equivalent

factors for different driving behaviors. An improved multi-learning space adaptive deep reinforcement learning algorithm was introduced to perceive and learn driving behaviors. This model reduced the hydrogen consumption and usage cost of automobiles compared with traditional methods [14]. Yu et al. [15] built a human activity recognition model based on radar and 3D point cloud technology. This model used a multi-input multi-output radar as a static ground sensor to obtain the dense point clouds of human activity behavior. The model could more easily access and obtain human motion information, and had a higher accuracy [15].

In summary, CBAM can be widely applied in various fields, and there have been many studies on behavior recognition. However, there is still relatively little research on behavior recognition in ICS. In order to reduce the misjudgment of ICS behavior recognition, it is necessary to accurately identify it. Based on this, a snow and ice sports behavior recognition model based on MSF and improved CBAM is developed. It is expected to explore more accurate methods for behavior recognition, reduce the misidentification rate, and promote the development of ICS.

The innovation points of this study are as follows. (1) A feature extraction scheme for ICS behavior based on MSF is developed, which improves the recognition ability of sports detail features. (2) The CBAM module is improved to enhance its recognition accuracy and robustness. (3) A recognition model for behavior recognition in ICS is constricted by integrating MSF and improved CBAM. The research is structured from three parts. Part one constructs a new model. The second part is the performance testing. The third part discusses the experimental results.

2. Methods and materials

2.1. Feature extraction scheme for ice and snow sports behavior based on MSF

By identifying the behavioral characteristics of athletes in ICS, intelligent posture correction can be provided for athletes to improve exercise efficiency [16]. However, ICS have complex movement characteristics and diverse external environments, such as athlete movement switching, posture changes, and dynamic backgrounds, which pose great challenges to the ICS behavior recognition [17]. Due to the fact that traditional dynamic behavior feature extraction methods mostly use single scale feature extraction methods, these methods cannot capture comprehensive motion details well faced with fast changes in actions and scenes, and may also ignore actions with low discriminability. MSF can effectively capture and parse multiple scale characteristics of data, achieving recognition and fusion of different motion details [18]. Therefore, the study uses MSF to extract behavioral characteristics of ICS. The designed MSF fusion network structure is shown in **Figure 1**.

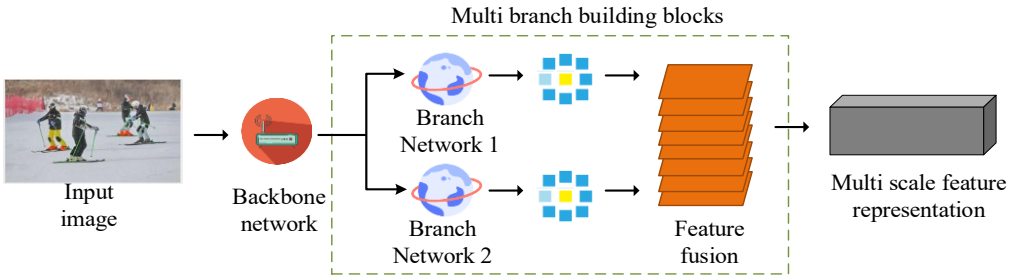

Figure 1. Multi-scale feature fusion network structure.

Figure 1 displays the MSF fusion network for extracting ICS behavior, which is composed of a backbone network and multiple branch modules stacked together. The branch module includes two branch networks, mainly used for selective fusion of multiple feature streams with different receptive field sizes to learn MSF in ice and snow motion images [19]. On this basis, structural units are used to extract MSF from each layer of the network, and MSF from different levels are fused to obtain MSF expressions. The MSF fusion network designed for research is based on the residual network. Given an image input data x, a residual $\hat{x}$ is constructed, and its expression is displayed in Equation (1).

$$\tilde{x} = F(x) \tag{1}$$

In Equation (1), $\hat{x}$ signifies the residual value constructed. $F(x)$ is a numerical value with a mapping function. F is a lightweight convolutional layer used for single scale feature learning of ice and snow motion images. By calculating the value of the mapping function, the feature output y can be obtained, as presented in Equation (2).

$$y = x + \tilde{x} \tag{2}$$

In Equation (2), y is the numerical value of the image feature output. After feature learning, the constructed multi-branch module can consciously extract motion posture features from different ice and snow motion images based on their receptive fields, which can better distinguish the features. The multi-branch module construction of MSF is shown in **Figure 2**.

The construction process diagram of the multi-branch module is presented in **Figure 2**. The process first inputs the input ice and snow motion image into the construction module. Without participating in spatial information fusion, the feature dimension is processed by the contribution network layer. x represents the two independent branches constructed by the multi-branch module, and each independent branch contains a different number of lightweight convolutional layers CONV. At this time, the feature maps are separately input into the branch module to obtain a map of multiple scale features. The scale of the image receptive field can affect the feature fusion efficiency receptive field scales leads to increased computational complexity, smaller receptive fields have higher computational efficiency, and feature fusion needs to consider different feature scales. Therefore, two studies-designed image receptive fields of different sizes. The standard for the change in receptive field scale of collected ICS images refers to the proportional growth model, where there are two types of receptive field sizes for images. One is 3

× 3and the other is 5 × 5. Finally, the image features extracted by the fusion branch are output to a convolutional layer with a size of 1 × 1 for dimensionality reduction. After the specified feature output dimension is obtained, it is fused with the input features. Through this residual neural network-based method, multi-scale representation of ice and snow motion images can be achieved, and the representation contains the complete spatial scale. In summary, a hierarchical iterative algorithm with multiple branch unit structures can achieve feature processing of ice and snow motion images with different sizes. The obtained MSF can contain more complex and rich information [20]. The important step in extracting features of ICS behavior is the feature scale fusion method. This method does not require manual setting, but dynamically calculates the weights of each branch feature on the ground of the input image characteristics. Therefore, the multi-scale residual feature is displayed in Equation (3) [21].

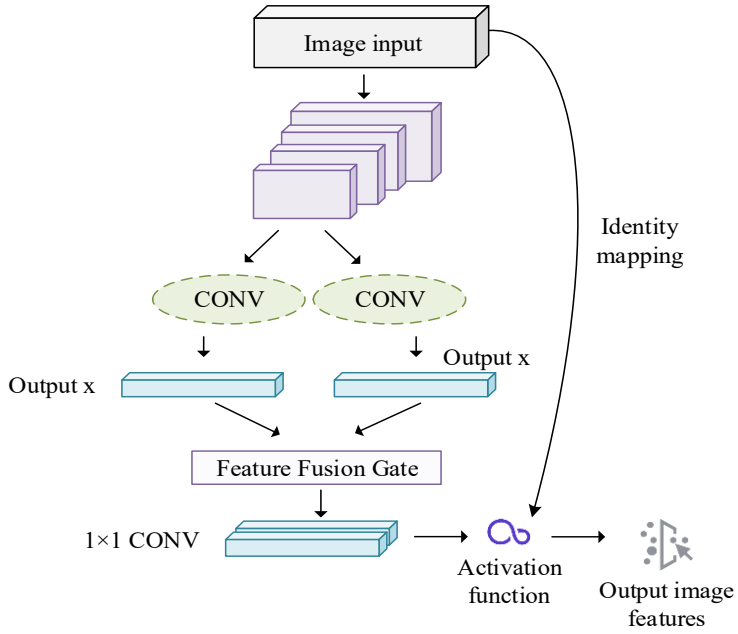

Figure 2. Multi-scale feature fusion network structure.

$$\tilde{x} = \sum_{N}^{n=1} G(X_n) \odot X_n \tag{3}$$

In Equation (3), $\odot$is the Hadamard product. Nsignifies the branches. X_n signifies the size of the feature channel. n is the number of sub-feature channels. $G(\cdot)$ is the vector of the sub-network. The sub-network contains one global average pooling layer and two fully connected layers. $G(X_n)$ is the feature weight of two branches. This method can assign different scale features based on the weight of the input ICS behavior image, reducing the error of manually assigning weights and adaptively expressing and learning MSF of different input images [22]. The expression for extracting MSF of ICS behavior is shown in Equation (4).

$$F_s = \sum_{i=1}^{n} w_i \times F_i \tag{4}$$

In Equation (4), F_s is the final MSF representation of the input image. F_i represents image features extracted at different scales. w_i is the weight of each scale feature. After obtaining the MSF representation of the image, the feature can be convolved, and the expression of the convolution operation is shown in Equation (5).

$$F_{conv} = W \times X + b \tag{5}$$

In Equation (5), F_{conv} is the convolved feature map. W signifies the weight matrix of the convolution kernel. X signifies the input image or feature map. b is the bias term. Performing depth separable convolution on the obtained feature map can separate the channels and regions of the image, reducing the number of image transformations and computational power. As a result, the network takes less time to process massive image data [23,24]. The equation expression for depth separable convolution is shown in Equation (6).

$$F_{sep} = W_{depth} \times \left(W_{point} \times X\right) \tag{6}$$

In Equation (6), F_{sep} is the feature map after depth separable convolution. W_{depth} is the weight of the deep convolution kernel. W_{point} signifies the weight of the point convolution kernel. Normalizing the feature map can improve the recognition ability of complex ICS behaviors, as displayed in Equation (7).

$$A_{norm} = \frac{F - \mu}{\sigma} \tag{7}$$

In Equation (7), A_{norm} signifies the normalized feature map. F signifies the original feature map. μ signifies the mean of the feature map. σ signifies the standard deviation of the feature map. In summary, the flowchart of the ICS behavior feature extraction scheme based on MSF is shown in **Figure 3**.

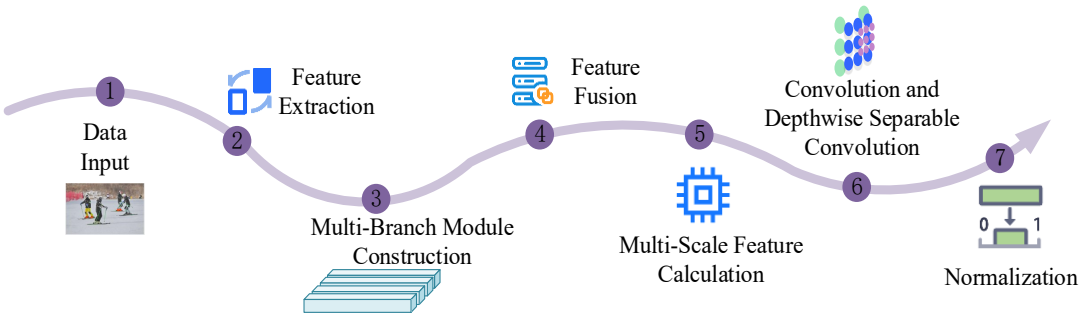

Figure 3. Flow chart of feature extraction scheme for ice and snow sports behavior based on multi-scale features.

Figure 3 shows the flowchart of the ice and snow motion feature extraction scheme based on MSF. Firstly, the collected ice and snow motion behavior images are input into the residual neural network for feature extraction and the construction of a multi-branch module architecture. After extracting motion features of different scales, these features are fused and calculated. Finally, the obtained feature map is normalized to output the final MSF map.

2.2. Construction of ice and snow sports behavior recognition model based on MSF-ICBAM

Due to the diversity of target objects in ICS scenes and the presence of a large number of small target objects, traditional behavior recognition methods have low recognition accuracy and robustness when dealing with this scene. Common ICS scenes are shown in **Figure 4**, which is from the Common Objects in Context (COCO) dataset (https://cocodataset.org/#download). This dataset is a publicly available dataset widely used for research and development in the field of computer vision. The currently widely used small object recognition model is CBAM, which can simultaneously process channel and spatial information and optimize the model's expressive and generalization abilities. In addition, it can adaptively learn features of different regions in the input image, increasing the model's attention to important features [25]. However, CBAM still has limitations such as unclear feature extraction for small targets and high computational complexity [26]. To improve the recognition ability of small targets in ICS, and optimize the robustness and generalization ability of the recognition model, the CBAM is optimized. A model on the basis of Multi-scale Features and improved Convolutional Block Attention Module (MSF-ICBAM) for ICS behavior recognition is constricted.

Figure 4. Ice and snow sports scene.

The study first optimizes the channel attention mechanism in CBAM. The optimized CBAM channel attention mechanism module calculation is displayed in Equation (8).

$$Mc(F) = \sigma(f_{1\times1}(\text{MaxPool}(F)) + f_{1\times1}(\text{LocalAvgPool}(F)) + f_{1\times1}(\text{StoPool}(F))) \quad (8)$$

In Equation (8), σ is the sigmoid activation function. $f_{1\times1}$ is a one-dimensional convolution operation. MaxPool is the global max pooling. LocalAvgPool is local average pooling. StoPool is global random pooling. Local averaging and global random pooling are used to calculate the proportion of attention channels, so that the MSF-based ICS behavior features can be further divided and weighted, thereby improving the attention mechanism's focus on small targets [27]. In addition, introducing one-dimensional convolution to achieve information exchange between multiple channels reduces the number of calculations and improves the efficiency of operations. Local averaging and global random pooling in CBAM provide a more comprehensive capture of inter-channel relationships and a more robust feature representation, enhance the model's perception of local details and global context information, thus improving the accuracy of attention allocation, optimizing small target recognition, and enhancing the model's generalization

ability. Finally, the behavior recognition performance and computational efficiency of the model in complex scenarios are improved. After local average pooling and global random pooling, the calculation performance of the model has been improved. When dealing with outdoor sports such as ice and snow sports, the target identification can be more timely and effectively completed. The structural diagram of improving the channel attention mechanism of CBAM through the above operation is shown in **Figure 5**.

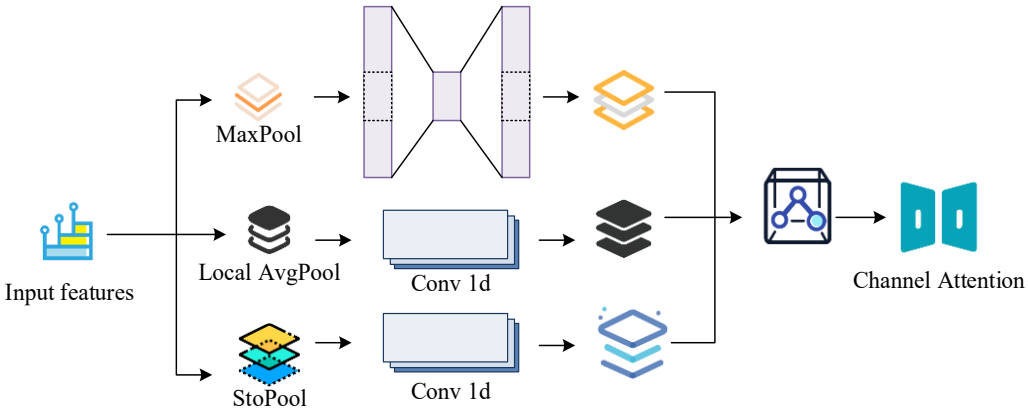

Figure 5. Structural diagram of improved CBAM channel attention mechanism.

The optimization process of CBAM is as follows: First, the channel attention mechanism was optimized by combining global max pooling and global random pooling, as well as introducing one-dimensional convolution, which allowed the model to more accurately assign attention weights between channels and improve computational efficiency. Second, the spatial attention mechanism enhanced the model's attention to key targets by replacing the original convolution kernel with an expansion convolution kernel and adding a random pooling layer, while reducing the focus on unimportant information. Additionally, to address the problem of imbalanced sample categories, the focus loss was introduced, which made the model pay more attention to difficult-to-classify samples. Finally, the model's stability in dynamic changing environments was improved by introducing affine transformation matrices and Kalman filters. In addition to optimizing the channel attention mechanism in CBAM, the study also improves the spatial attention mechanism. In order to reduce the attention mechanism's focus on unimportant information and increase the attention weight on key targets to achieve accurate spatial relationship mapping, the convolution kernel with a horizontal and vertical size of 7 in the original structure is replaced with a dilated convolution with a size of 3, and the dilation rate of this convolution is set to 2. In addition, a random pooling layer is added to the module. The optimized structure diagram of the spatial attention module in CBAM is displayed in **Figure 6**.

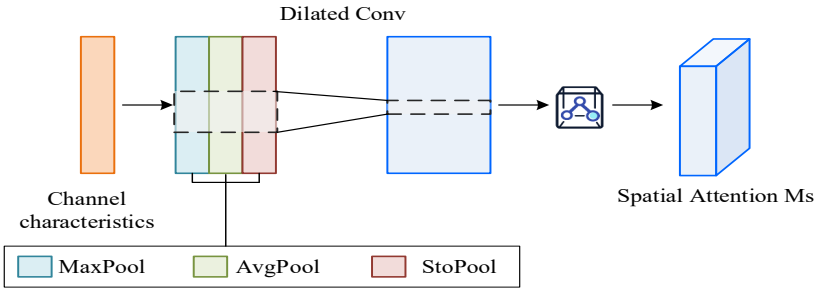

Figure 6. Structural diagram of the spatial attention mechanism improvement for CBATM.

Figure 6 shows the spatial attention structure diagram with the added dilated convolution kernels and random pooling layers, and its calculation is shown in Equation (9).

$$Ms(F) = \sigma(f_{3\times3}^{\mathrm{dw}}([\mathrm{MaxPool}(F); \mathrm{AvgPool}(F); \mathrm{StoPool}(F)])) \tag{9}$$

In Equation (9), $Ms(F)$ is the output value of the spatial attention mechanism. $f_{3\times3}^{\mathrm{dw}}$ represents using a convolution kernel with a horizontal and vertical size of 3 and a dilated ratio of 2 for dilated convolution operation. In order to solve the imbalanced sample categories during the training process of the improved CBAM, the study introduces focus loss to improve it. The calculation is shown in Equation (10) [28,29].

$$\mathrm{AFL}(p_t) = -\alpha_t(1-p_t)^{\gamma} log(p_t) \tag{10}$$

In Equation (10), p_t is the predicted probability. α_t is a factor used to balance positive and negative samples. γ is the parameter for adjusting the weights of difficult and easy samples. α_t is used to balance the attention of a small number of categories in the dataset, and is usually set between [0, 1]. Too high of this parameter will cause excessive attention to a small number of categories, so the study sets it at 0.25. γ is used to adjust the model's attention to the difficulty degree of the sample, which is usually set between [1,5]. The larger the parameter, the higher the model's attention to the difficulty sample. Focusing entirely on difficult samples will cause the model to be difficult to converge and the amount of computation will increase. Therefore, this parameter is set to 2. $log(p_t)$ is the logarithmic part of the probability. The adaptive factor calculation for focus loss is shown in Equation (11) [30,31].

$$\beta_t = \frac{\sum_{i=1}^{N}\alpha_i p_i}{\sum_{i=1}^{N} p_i} \tag{11}$$

In Equation (10), j is the value of the adaptive factor in the focus loss. p_i signifies the predicted probability of the i-th sample image. N signifies the number of image samples of ICS behavior. By calculating the channel statistics of convolution, the feature expression ability of the model at different scales is improved, as displayed in Equation (12).

$$S_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U(i,j) \tag{12}$$

In Equation (12), S_c represents the statistical information of the channel. H signifies the height of the feature image. W signifies the width of the feature image. $U(i,j)$ signifies the pixel value of the i-th row and j-th column in feature map U. The fully connected layer of the improved CBAM dimension after reduction is displayed in Equation (13).

$$z = \delta(W_s B(S)) \tag{13}$$

In Equation (13), z signifies the output channel attention weight vector. δ is the ReLU activation function. W_s is the weight matrix. $B(S)$ is the image feature value after batch normalization of the improved CBAM. In the recognition model of ICS behavior, since the target is in motion, to optimize the stability and accuracy of recognition, the affine transformation matrix is calculated [32,33]. The calculation is shown in Equation (14).

$$A = [M \mid T] \tag{14}$$

In Equation (14), A is the affine transformation matrix. M is the rotated part of the affine matrix. T is a translation vector. To compensate for the stability during ICS, a Kalman filter update formula is introduced to optimize it. The expression of this update formula is shown in Equation (15).

$$P_t^{'} = M_k P_{t-1} M_k^T \tag{15}$$

In Equation (15), $P_t^{'}$ is the updated covariance matrix. M_k is the transformation matrix. P_{t-1} is the prior covariance matrix. In summary, a snow and ice sports behavior recognition model based on MSF-ICBAM can be constructed, and the flowchart is displayed in **Figure 7**.

Figure 7. Flow chart of ice and snow sports behavior recognition model based on MSF-ICBAM.

The model structure diagram shown in **Figure 7** first inputs images of ICS behavior, and uses an MSF extraction module to extract motion features at different scales. Then, the improved CBAM module enhances the attention to small targets and integrates MSF to output recognition results.

3. Results

3.1. The effectiveness evaluation of the feature extraction scheme for ice and snow sports behavior based on MSF

The experiment first evaluates the effectiveness of the ICS behavior feature extraction scheme based on MSF. The study uses publicly available datasets as image data input, namely COCO and Ski-Pose. The initial parameter settings for the hardware environment, software, and model used are shown in **Table 1**.

Table 1. Experimental environment setup.

Experimental hardware setup		**Experimental software settings**	
Project	Set up	Project	Set up
CPU	Intel(R)Core(TM)i7-10700K	Python Version	3.8.10
GPU	GTX3060	Pytorch Version	1.9
Operating system	Ubuntu 18.04	Number of training iterations	500
CUDA Version	10.2	Optimizer selection	Adam
CUNN Version	81	Number of warm-up training batches	15
Memory	512 GB SSD	Initial learning rate	0.01

To verify the effectiveness of the ICS behavior feature extraction based on MSF, it is compared with other commonly used feature extraction methods, including Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT). **Figure 8** displays the average loss curves of the three feature extraction methods for ICS behavior recognition at different training times.

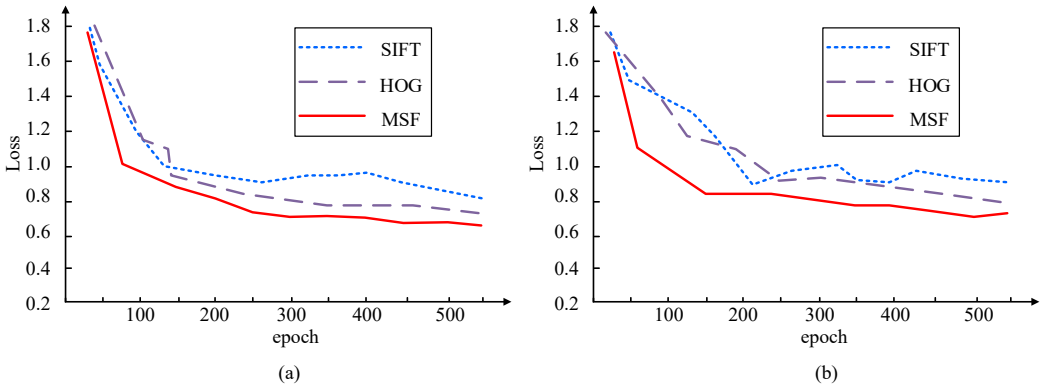

Figure 8. The average loss curve of three feature extraction methods for recognizing ice and snow sports behavior under different training times. (**a**) The average loss curves of MSF, HOG, and SIFT in the COCO datasetSF, HOG, and SIFT in the COCO dataset; (**b**) The average loss curves of MSF, HOG, and SIFT in the Ski Pose training set.

Figure 8a shows the average loss curves of three feature extraction methods for ICS behavior feature extraction on the COCO dataset. The loss curve of MSF decreased rapidly, rapidly decreasing after 100 training iterations and then slowly decreasing to 0.71. The loss curves of HOG and SIFT decreased slowly, and the average loss function gradually approached the lowest value when the training iterations were 200–300. When trained 500 times, the average loss function of MSF

decreased by 0.12 and 0.24 compared with HOG and SIFT. **Figure 8b** shows the average loss curves of three feature extraction methods for ICS behavior feature extraction on the Ski-Pose dataset. In the figure, as the training iterations approached 150, the average loss of MSF decreased from 1.6 to 1.1, then rapidly decreased to 0.84, and then slowly decreased to 0.81 before leveling off. The average loss curves of HOG and SIFT decreased slowly, and the loss was higher than MSF at 500 training iterations. The experimental results show that MSF converges faster compared with HOG and SIFT under training on different datasets. It has been demonstrated that the proposed approach can effectively capture MSF in ICS behavior images, providing more accurate feature extraction results. It is more suitable for application in ICS behavior feature extraction. The Mean Average Precision (mAP) of MSF, HOG, and SIFT trained on COCO and Ski-Pose datasets is shown in **Figure 9**.

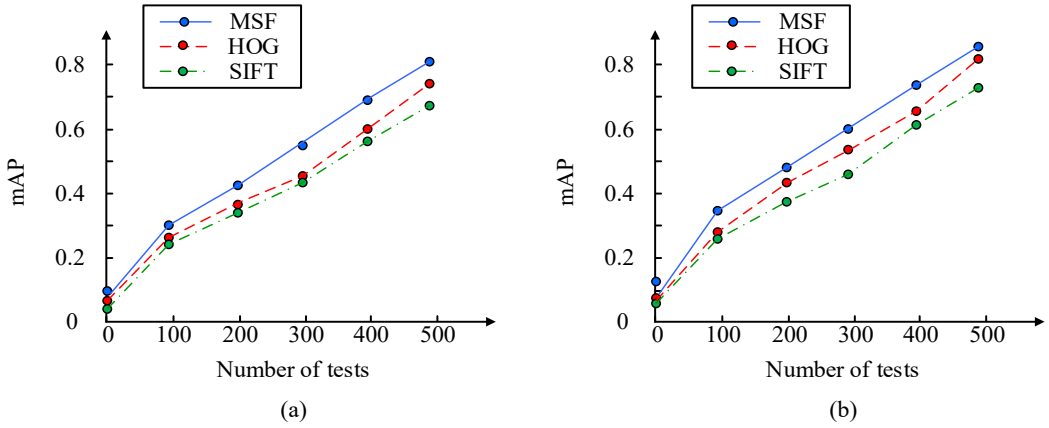

Figure 9. The average loss curve of three feature extraction methods for recognizing ice and snow sports behavior under different training times. (**a**) MAP results of three feature extraction models on COCO dataset; (**b**) MAP results of three feature extraction models on Ski-Pose dataset.

Figure 9a shows the mAP curve trends of three feature extraction methods on the COCO. As the tests increased, the mAP values also increased. At 500 tests, the mAP values of MSF, HOG, and SIFT were 0.80, 0.72, and 0.65, respectively, with MSF having a higher value than the other two algorithms. **Figure 9b** shows the mAP curve trends of three feature extraction methods on the Ski-Pose dataset. The mAP values of all three feature extraction methods gradually increased with the increase of testing times, with HOG and SIFT showing slower growth rates than MSF. When tested 500 times, the mAP values of MSF, HOG, and SIFT were 0.85, 0.75, and 0.68, respectively, with MSF having a higher value than the other two feature extraction methods. The experimental results indicate that the motion behavior feature extraction scheme based on MSF is more suitable for ICS, and MSF can accurately capture multi-dimensional information. The precision of three feature extraction methods on COCO and Ski-Pose datasets with different training times is shown in **Figure 10**.

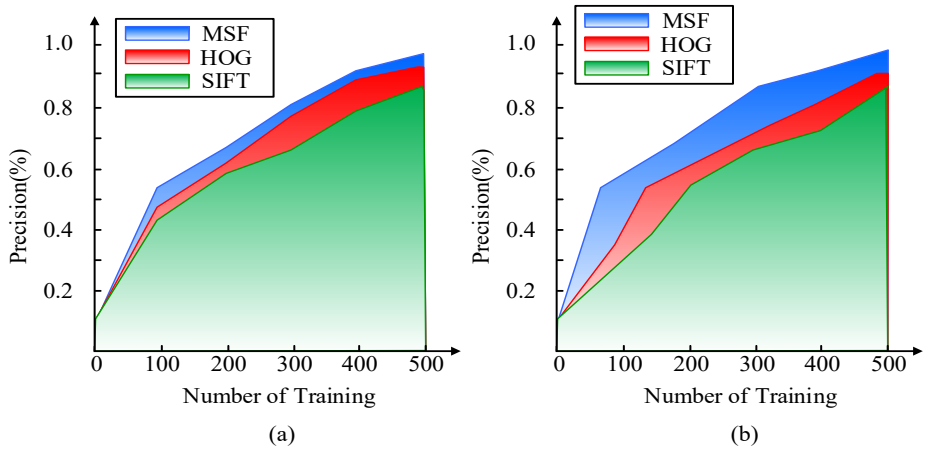

Figure 10. Experimental data on the precision of MSF, HOG, and SIFT under different training times on COCO and Ski-Pose datasets. (**a**) Comparison of precision of three feature extraction models in COCO dataset; (**b**) Comparison of precision of three feature extraction models in Ski-Pose dataset.

Figure 10a shows the feature extraction precision of MSF, HOG, and SIFT methods on the COCO dataset. The MSF-based ice and snow motion behavior feature extraction method had a higher precision on the COCO dataset than HOG and SIFT, and the graphic area was 20%–30% larger than the other two feature extraction methods. When the training frequency was 500 times, the precision values of MSF, HOG, and SIFT were 0.94, 0.87, and 0.82, respectively, with MFS having a higher precision value. **Figure 10b** shows the feature extraction precision of MSF, HOG, and SIFT methods on the Ski-Pose dataset. From the graph area in the figure, the precision of the three feature extraction methods increased with the increase of training times. In addition, MSF had a higher precision value than the other two methods. When the training frequency was 500 times, the precision values of MSF were 0.12 and 0.17 higher than those of HOG and SIFT. The feature extraction method based on MSF has higher robustness and precision in extracting complex ice and snow motion scene features due to the HOG and SIFT methods.

3.2. Performance verification of ice and snow sports behavior recognition model based on MSF-ICBAM

To verify the recognition effect of the proposed model in ICS behavior, common behavior recognition models are compared with it, including the Inflated 3D ConvNet (I3D), Two-stream CNN, ResNet + Long Short-Term Memory (ResNet + LSTM), and Temporal Segment Networks (TSN). Firstly, the accuracy and recall of the three different models are experimentally verified, as displayed in **Figure 11**.

Figure 11a shows the accuracy of three models in recognizing ICS behavior on the Ski-Pose dataset. From the figure, all three algorithms showed a rapid upward trend after reaching 100 iterations, and then tended to flatten out. The accuracy of MSF-ICBAM exceeded the other two methods. When the iteration reached 500, the accuracy of MSF-ICBAM, TSN, and I3D was 98.3%, 90.1%, and 84.6%, respectively. The accuracy of MSF-ICBAM was 8.2% and 13.7% higher than that of MSF-ICBAM, indicating a significant difference. **Figure 11b** shows the recall results of three models for recognizing ICS behavior. The recall rates of all three

algorithms showed a rapid upward trend at 200 iterations, while the recall rate values slowly increased and then flattened out at 200–500 iterations. The recall rate of MSF-ICBAM was 20%–30% higher than that of TSN and I3D at different iterations. The designed model has high effectiveness in identifying small targets in ICS behavior. The model exhibits high recognition error and recognition ability. The training speed and inference speed of different models are presented in **Figure 12**.

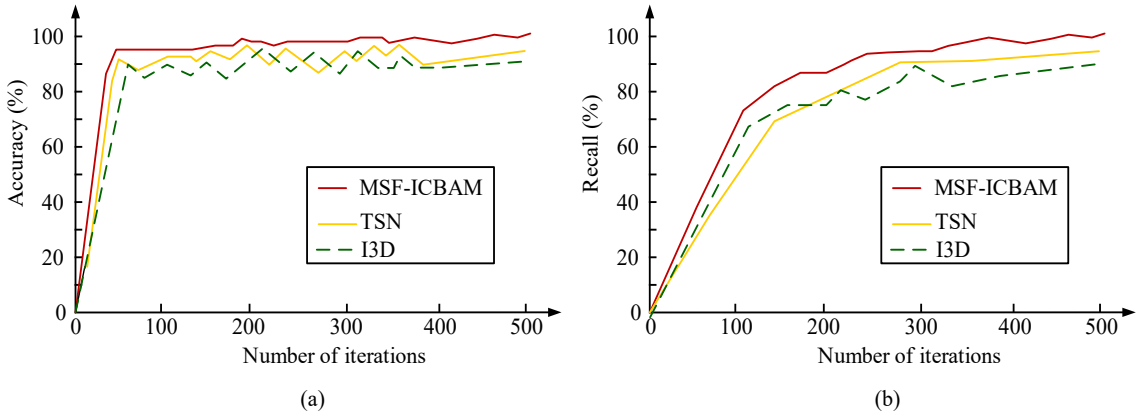

Figure 11. Comparison of accuracy and recall rates of different models for recognizing ice and snow sports behavior. (**a**) Accuracy of Different Models in Recognizing Ice and Snow Sports Behavior; (**b**) Recall of Different Models in Recognizing Ice and Snow Sports Behavior.

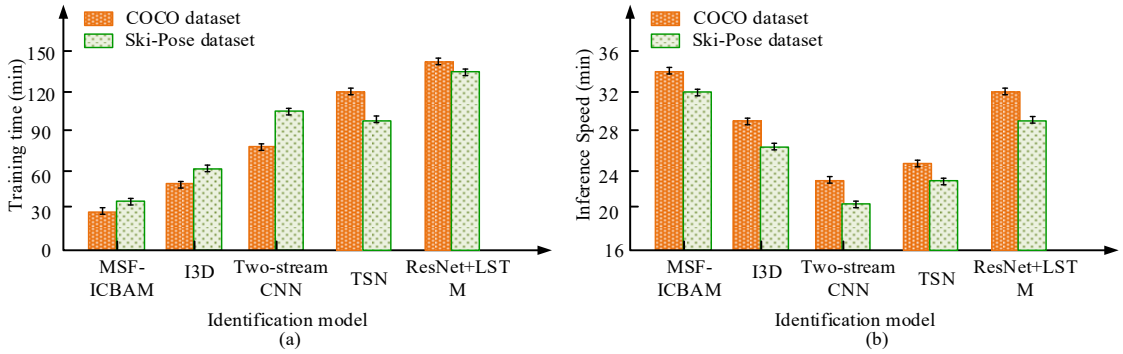

Figure 12. Experimental results on training speed and inference speed of different models. (**a**) Comparison results of training time for different models; (**b**) Comparison results of inference speed among different models.

Figure 12a shows the training time of different models for ICS behavior recognition in two training sets. The training time of MSF-ICBAM, I3D, Two-stream CNN, ResNet + LSTM, and TSN recognition models on the COCO dataset was 29, 54, 73, 117, and 136 minutes, respectively. The training time on the Ski-Pose dataset was 31, 62, 103, 96, and 128 minutes, respectively. The training time of the proposed model was relatively short, with a difference of 30%–60% compared with other models. **Figure 12b** shows the inference time results of different models for ICS behavior recognition in two training sets. From the height of the bar chart shape, MSF-ICBAM had the fastest inference time, exceeding other models. The inference speeds of MSF-ICBAM, I3D, Two-stream CNN, ResNet + LSTM, and TSN recognition models on the COCO dataset were 34, 28, 22, 25, and 30 FPS, respectively. The inference speeds on the Ski-Pose dataset were 32, 26, 20,

23, and 28 FPS, respectively. MSF-ICBAM had the highest inference speed on both datasets. The experimental results demonstrate that MSF-ICBAM has shorter training time and higher training efficiency on different datasets. In addition, the inference speed of MSF-ICBAM is superior to other models, indicating that it has high efficiency and good real-time performance, which can still quickly and accurately identify motion behavior in complex ICS environments. The study conducts experiments on the model size, parameter quantity, floating-point operation count, accuracy, and *F*1 value on two datasets, as displayed in **Table 2**.

Table 2. Comparison results of different models.

Model	Data set	Model size (MB)	Parameter quantity (in millions)	Floating-point operations (GFLOPs)	Accuracy (%)	*F*1 value (%)
MSF-ICBAM	COCO	120	11.7	145	89.5	91.2
I3D		215	12.4	120	86.1	87.9
Two-stream CNN		190	35.4	110	84.8	86.4
ResNet + LSTM		150	25.6	135	87.4	89.1
TSN		160	16.3	130	85.2	87.0
MSF-ICBAM	Ski-Pose	118	11.7	145	88.9	90.7
I3D		210	12.4	120	85.6	87.2
Two-stream CNN		188	35.4	110	83.9	85.5
ResNet + LSTM		148	25.6	135	86.9	88.4
TSN		159	16.3	130	84.7	86.5

Table 2 shows the model size, parameter quantity, floating-point operation count, accuracy, and *F*1 value of different models on two datasets. From the data in the table, MSF-ICBAM outperformed the other four models in various metrics, especially in accuracy and *F*1 value. In the COCO and Ski-Pose datasets, the model sizes of MSF-ICBAM were 120 MB and 118 MB, which were 44.18% and 43.8% smaller than I3D, respectively. The parameter quantity and floating-point operation count of MSF-ICBAM were 117 million and 145, respectively, indicating that it can maintain efficient computing capability in different ICS scenarios. In the COCO dataset, the accuracy and *F*1 value of MSF-ICBAM were 89.5% and 91.2%, respectively, which were 3.4% and 4.7% higher than I3D and Two-stream CNN, respectively. In the Ski-Pose dataset, the accuracy and *F*1 score of MSF-ICBAM were also higher than other models. The results verify the effectiveness of the model in processing ICS behavior recognition tasks. In order to further analyze the parameters and computational complexity of MSF-ICBAM in practical applications, the number of parameters of the model in different scenarios and the usage of computing resources of the device are studied and counted, and the results are shown in **Table 3**.

Table 3. Number of parameters and computing resource usage in different scenarios.

Map	Parameter quantity (in millions)	Computing resources (%)
1	143	32.5
2	126	28.7
3	124	28.7
4	136	29.6
5	128	29.1

As can be seen in **Table 3**, the MSF-ICBAM model can control the number of parameters between 120 million and 150 million in different scenarios. The MSF-ICBAM model maintains a low level of device computing resource consumption in different scenarios. When the number of parameters in the scenario reaches 143 million, the MSF-ICBAM model consumes only 32.5% of device computing resources. The study uses the I3D model and the proposed model to identify ICS behavior, as displayed in **Figure 13**.

(a) (b)

Figure 13. The recognition effect of two recognition models on ice and snow sports behavior. (**a**) The recognition effect of I3D on ice and snow sports behavior; (**b**) The effectiveness of the method proposed by the research institute in recognizing ice and snow sports behavior.

From **Figure 13**, the ICS behavior recognition effect of the proposed model was better than that of the I3D model, and the recognition rate for small targets reached 80%. After obtaining the skiing athlete's movement posture area and body contour, the research model uses ICBAM to extract the athlete's movement posture area and body contour features. Based on this, MSF are used to fuse the complementary features of the two, ensuring the recognition effect of ICS behavior.

4. Discussion and conclusion

Through experimental analysis, the effectiveness of the ICS behavior recognition model on the basis of MSF-ICBAM was verified. The study compared and evaluated the recognition performance of MSF-ICBAM models on the Ski-Pose and COCO datasets, demonstrating significant advantages in recognition accuracy, recall, training speed, and inference time compared with other traditional recognition models. By integrating MSF and ICBAM, the model demonstrated good adaptability in complex environments and small target recognition in ICS, especially with high robustness in accurately identifying small target athlete postures and dynamic

background changes. The results can provide effective support for the automated analysis of ICS behavior, and provide real-time basis for sports posture management and correction.

ICS is usually conducted in outdoor environments, and the scenes of their movements are often dynamic and complex, easily affected by factors such as a large number of small targets and lighting weather. In order to improve the recognition accuracy and robustness of ICS behavior, a model for ICS behavior recognition based on MSF-ICBAM was proposed. Relevant experiments were conducted to verify the effectiveness. The loss curve of MSF decreased rapidly in the COCO and Ski-Pose datasets, and the average loss function dropped to the lowest of 0.71 and 0.83 respectively when the number of tests reached 500, significantly lower than other comparison feature extraction methods. In the COCO dataset, when the number of tests was 500, the mAP values of MSF, HOG, and SIFT were 0.80, 0.72, and 0.65, respectively, with MSF having a higher value than the other two algorithms. In the Ski-Pose dataset, the accuracy of MSF-ICBAM, TSN, and I3D was 98.3%, 90.1%, and 84.6%, respectively. The accuracy of MSF-ICBAM was 8.2% and 13.7% higher than that of TSN and I3D, indicating a significant difference. The accuracy and $F1$ value of MSF-ICBAM were 89.5% and 91.2%, respectively, which were 3.4% and 4.7% higher than I3D and Two-stream CNN, respectively. The recognition rate for small targets reached 80%. Ice and snow sports are usually carried out outdoors, portable equipment resources are limited, and high-performance models often need more computing resources. Therefore, the model designed by research needs to face the problem of model performance degradation caused by resource constraints in practical applications. The model can be integrated into the intelligent monitoring system to provide technical support for the management and analysis of ice and snow sports, such as athlete performance tracking and training effect evaluation. In the competition environment, the model can help referees more accurately judge whether the athletes' actions are compliant, and also provide data support for competition analysis. The results have verified the effectiveness of the model and can be well applied to ICS behavior recognition, with good adaptability to complex and varied ICS scenes. Although the model can improve its performance in ICS behavior recognition, its computational efficiency will decrease in real-time behavior recognition applications. In the future, the parallel computing technology will be explored to enhance the computational efficiency of the model and achieve real-time motion behavior recognition.

Ethical approval: Not applicable.

Conflict of interest: The author declares no conflict of interest.

Abbreviations

ICS	Ice and Snow Sports
CBAM	Convolutional Block Attention Module
CNN	Convolutional Neural Network
MSF	Multi-Scale Feature
COCO	Common Objects in Context

MSF-ICBAM	Convolutional Block Attention Module
HOG	Histogram of Oriented Gradients
SIFT	Scale-Invariant Feature Transform
I3D	Inflated 3D ConvNet
ResNet+LSTM	ResNet+Long Short-Term Memory
TSN	Temporal Segment Networks

References

1. Zhou F, Li J, Dai Y, Liu L, Qin H, Jiang Y, et al. Time-Series Fusion-Based Multicamera Self-Calibration for Free-View Video Generation in Low-Texture Sports Scene. IEEE Sens J. 2023;23(3):2956-2969.
2. Wang G, Pang Z, Wang F, Chen Y, Dai H, Wang B. Urban Fiber Based Laser Interferometry for Traffic Monitoring and Analysis. J Lightwave Technol. 2023;41(1):347-354.
3. Li Y, Wu L. A Wireless Self-Powered Sensor Network Based on Dual-Model Convolutional Neural Network Algorithm for Tennis Sports. IEEE SENS J, 2023;23(18):20745-20755.
4. Wang J, Li W, Gao Y, Zhang M, Tao R, Du Q. Hyperspectral and SAR Image Classification via Multiscale Interactive Fusion Network. IEEE T Neur Net Lear. 2023;34(12):10823-10837.
5. Axi N, Yu Z, Chaoning Z, Jinqiu S, Pei W, Inso K, et al, MS2Net: Multi-Scale and Multi-Stage Feature Fusion for Blurred Image Super-Resolution. IEEE T Circ Syst Vid, 2022;32(8):5137-5150.
6. He J, Ren Z, Zhang W, Jia Y, Guo S, Cui G. Fall Detection Based on Parallel 2DCNN-CBAM With Radar Multidomain Representations. IEEE Sens J, 2023;23(6):6085-6098.
7. Feng Y, Yang X, Qiu D, Zhang H, Wei D, Liu J. PCXRNet: Pneumonia Diagnosis From Chest X-Ray Images Using Condense Attention Block and Multiconvolution Attention Block. IEEE J Biomed Health. 2022;26(4):1484-1495.
8. Deng W, Wang X, Huang Z, Xu Q. Modulation Classifier: A Few-Shot Learning Semi-Supervised Method Based on Multimodal Information and Domain Adversarial Network. IEEE Commun Lett. 2023;27(27):576-580.
9. Jiang Y, Xie S, Xie X, Cui Y, Tang H. Emotion Recognition via Multiscale Feature Fusion Network and Attention Mechanism. IEEE Sens J. 2023;23(10):10790-10800.
10. Chang M, Yao D, Yang J. Intelligent Fault Diagnosis of Rolling Bearings Using Efficient and Lightweight ResNet Networks Based on an Attention Mechanism. IEEE Sens J. 2023;23(9):9136-9145.
11. Pan Z, Chen H, Zhong W, Wang A, Zheng C. A CNN-Based Animal Behavior Recognition Algorithm for Wearable Devices. IEEE Sens J. 2023;23(5):5156-5164.
12. Pang SM, Cao JX, Jian MY, Lai J, Yan ZY. BR-GAN: A Pedestrian Trajectory Prediction Model Combined With Behavior Recognition. IEEE T Intell Transp. 2022;23(12):24609-24620.
13. Huang T, Fu R, Chen Y, Sun Q. Real-Time Driver Behavior Detection Based on Deep Deformable Inverted Residual Network With an Attention Mechanism for Human-Vehicle Co-Driving System. IEEE T Veh Technol. 2022;71(12):12475-12488.
14. Sun H, Tao F, Fu Z, Gao A, Jiao L. Driving-Behavior-Aware Optimal Energy Management Strategy for Multi-Source Fuel Cell Hybrid Electric Vehicles Based on Adaptive Soft Deep-Reinforcement Learning. IEEE T Intell Transp. 2023;24(4):4127-4146.
15. Zheqi Y, Ahmad T, William T, Adnan Z, Khalid R, Hadi H, et al. A Radar-Based Human Activity Recognition Using a Novel 3-D Point Cloud Classifier. IEEE Sens J. 2022;22(19):18218-18227.
16. Meng C, Hui FS, Hsin YL, Yifan L, Ho-Yin C, Yufan W, et al. Phase-Based Quantification of Sports Performance Metrics Using a Smart IoT Sensor. IEEE Internet Things. 2023;10(18):15900-15911.
17. Ergeneci M, Carter D, Kosmas P. sEMG Onset Detection via Bidirectional Recurrent Neural Networks With Applications to Sports Science. IEEE Sens J. 2022;22(19):18751-18761.
18. Tang Y, Zhang L, Min F, He J. Multiscale Deep Feature Learning for Human Activity Recognition Using Wearable Sensors. IEEE T Ind Electron. 2023;70(2):2106-2116.
19. Ye Y, Pan C, Wu Y, Wang S, Xia Y. MFI-Net: Multiscale Feature Interaction Network for Retinal Vessel Segmentation. IEEE J Biomed Health. 2022;26(9):4551-4562.

20. Preethi P, Mamatha HR. Region-Based Convolutional Neural Network for Segmenting Text in Epigraphical Images. Artif. Intell. Appl. 2023;1(2):119-127.
21. Wang G, Peng C, Gu Y, Zhang J, Wang H. Interactive Multi-Scale Fusion of 2D and 3D Features for Multi-Object Vehicle Tracking. IEEE T Intell Transp. 2023;24(10):10618-10627.
22. Bhosle K, Musande V. Evaluation of Deep Learning CNN Model for Recognition of Devanagari Digit. Artif. Intell. Appl. 2023;1(2):114-118.
23. Kunchang L, Yali W, Junhao Z, Peng G, Guanglu S, Yu L. UniFormer: Unifying Convolution and Self-Attention for Visual Recognition. IEEE T Pattern Anal. 2023;45(10):12581-12600.
24. Min S, Jialin S, Qingming Y, Jian W, Zunkai H, Aiwen L. LMFFNet: A Well-Balanced Lightweight Network for Fast and Accurate Semantic Segmentation. IEEE T Neur Net Lear. 2023;34(6): 3205-3219.
25. Lu Z, Bian Y, Yang T, Ge Q, Wang Y. A New Siamese Heterogeneous Convolutional Neural Networks Based on Attention Mechanism and Feature Pyramid. IEEE T Cybernetics. 2024;54(1):13-24.
26. Li G, Fan W, Xie H, Qu X. Detection of Road Objects Based on Camera Sensors for Autonomous Driving in Various Traffic Situations. IEEE Sens J. 2022;22(24):24253-24263.
27. Xu S, Zhang L, Tang Y, Han C, Wu H, Song A. Channel Attention for Sensor-Based Activity Recognition: Embedding Features into all Frequencies in DCT Domain. IEEE T Knowl Data En. 2023;35(12):12497-12512.
28. Li X, Lv C, Wang W, Li G, Yang L, Yang J. Generalized Focal Loss: Towards Efficient Representation Learning for Dense Object Detection. IEEE T Pattern Anal. 2023;45(3):3139-3153.
29. Jian S, Yanan Z, Huajian L, Zeguang Z, Kexin Z, Kun Q. Depression Recognition From EEG Signals Using an Adaptive Channel Fusion Method via Improved Focal Loss. IEEE J BIOMED HEALTH, vol. 27, no. 7, pp. 3234-3245, July 2023, DOI: 10.1109/JBHI.2023.3265805.
30. Tang Y, Xie Y, Zhang W. Affine Subspace Robust Low-Rank Self-Representation: From Matrix to Tensor. IEEE T Pattern Anal. 2023;45(8):9357-9373.
31. Pareek G, Nigam S, Singh R. Modeling transformer architecture with attention layer for human activity recognition. Neural Computing and Applications, 2024, 36(10): 5515-5528.
32. Tang J, Gong S, Wang Y, Liu B, Du C, Gu B. Beyond coordinate attention: spatial-temporal recalibration and channel scaling for skeleton-based action recognition. Signal, Image and Video Processing, 2024, 18(1): 199-206.
33. Ha M H. Top-heavy capsnets based on spatiotemporal non-local for action recognition. Journal of Computing Theories and Applications, 2024, 2(1): 39-50.