Fuzzy Objective Segmentation Optimisation Algorithm for Basketball Video Data

Aixin Yang, Lifu Huang, Haobo Liu

***Abstract*—In daily life, video blurring is often caused by camera shake or damage of physical components, which in turn affects the target segmentation effect in the video. Therefore, in order to solve the problem of video deblurring and target segmentation. In this research, based on the scale recurrent network, Hal 2D wavelet transform and attention mechanism are introduced, and the obtained algorithmic model is tested for performance in a dataset, and at the same time compared with the previous traditional algorithmic model to screen out a more effective deblurring algorithm. Based on the traditional video target segmentation algorithm, the attention mechanism and morphological module are introduced and a new optimisation algorithm for video target segmentation is proposed. The performance of this algorithm is compared with the previous traditional algorithm model, and a more effective target segmentation algorithm is screened. Finally, the screened two optimal algorithms are combined to compare with the previous traditional video deblurring and target segmentation algorithms using basketball videos as experimental objects. The experimental results show that the peak-to-noise ratio of this new deblurring algorithm is as high as 30.55, and the structural similarity is as high as 0.942. The new segmentation algorithm has a region similarity as high as 96.5%, and the contour accuracy is as high as 93%. The processing time of the algorithm is shorter and the effect is better when the two algorithms are combined. In summary, the novel algorithm provides a certain reference value for the further research of subsequent video processing technology.**

***Index Terms*—attentional mechanisms, deblurring, morphology, target segmentation, wavelet transforms**

I. Introduction

In recent years, with the continuous progress of the Internet and information technology, the number of video images and other data as the main medium of information dissemination in the network has shown a leapfrog growth, which has a profound impact on various fields of society [1]. As one of the key technologies of video processing, video deblurring and target segmentation play an important role in video conferencing, video retrieval, video surveillance and so on [2]. Meanwhile, video deblurring is equally important as a preprocessing step for video target segmentation. In China, video deblurring using dynamic filter convolution proposed by Harbin Institute of Technology (HIT) in conjunction with Shangtang Institute, the network timing alignment and feature extraction under this algorithm is faster, but the generalisation is poor [3]. Abroad, Google and Massachusetts Institute of Technology jointly developed a mobile real-time segmentation technique, which can accurately separate the background from the background, while reducing the number of channels and improving the speed, but at this time, the video target segmentation will be affected by the occlusion of the object, violent movement, lighting and so on [4]. Combined with the above background, this research will propose a new deblurring algorithm and target segmentation algorithm based on previous theoretical research, combined with wavelet transform and attention mechanism, morphological module, and test the performance mechanism of these two algorithms under the experimental environment of certain video data samples, aiming at exploring the effect of this algorithm in the application of actual video processing. This research is divided into five parts, the first part is a brief introduction to the overall content of the article, the second part is an analysis and summary of the research of others, the third part describes how the optimisation algorithms for video deblurring and video target segmentation are formed, the fourth part tests the performance of the two algorithms individually and in combination respectively, and the last part is a summary of the article.

Manuscript received January 5, 2024; revised August 21, 2024.

A.X. Yang is lecturer of Institute of Physical Education, Hunan International Economics University, Changsha 410000, China. (e-mail: yaxhlf66@163.com).

L.F. Huang is leader of Physical Education Teaching and Research Group, High School Attached Hunan Normal University, Changsha 410000, China. (corresponding author to e-mail: cshlf55@126.com).

H.B. Liu is associate professor of Institute of Physical Education, Hunan International Economics University, Changsha 410000, China. (e-mail: liuhaobo2023@sina.com).

II. Related Works

Video deblurring has become an indispensable process in daily life as the image of video data can be blurred or degraded in quality due to out-of-focus lenses, shaking, dust masking, camera malfunctions, and so on. De-blurring is able to restore the low quality video images to some extent, which plays a preprocessing role for the subsequent video segmentation [5]. Li et al. proposed an efficient a priori aided blur intensity perception of L-0 structure in order to remove video blur. The scheme estimates content and structural information by generating efficiently clear frames and then aiding separate streams. The experimental results show that this deblurring method is more effective with better video restoration [6]. Zhan et al. proposed a new video deblurring method for the problem of uneven and blurred video images caused by camera shake when shooting with handheld devices. The method introduces a motion estimation and compensation module which estimates the optical flow to remove blurred frames and then restore HD frames. Experimental results show that this novel deblurring method

can promote temporal coherence and reduce the loss of frames for deblurring [7]. Zhang et al. address the two major problems of capturing temporal and spatial topology modules that still exist in video blurring, and propose a new method that combines an attentional inter-frame compensation module and an adaptive residual module, which recovers blurred of the frames. Experimental results show that this combined method has higher adaptive power and better restoration than the traditional method [8].

Video target segmentation is a fundamental task in the field of computer vision and one of the recognised technical difficulties. The purpose of video target segmentation is to segment specific object instances in the whole video sequence, that is, to divide the pixels in a video frame into two sets, foreground target and background region, and at the same time to generate a segmentation mask for the object to eliminate the effect of redundant information in the video, which plays a significant role in the field of computer vision. The general meaning of target segmentation is to simplify the image or video so that the result of the image is more understandable and easy to analyse. Li et al. proposed a target segmentation algorithm using local binary pattern contour coefficients for the field of image detection of target smoke, which is not able to be imaged completely and accurately due to the influence of various natural factors. Experimental results show that the algorithm overcomes the lack of texture information in the binary mode, making the extracted information easier to distinguish [9]. Hu et al. believe that the accurate quantitative analysis of retinal layers is a more difficult part of the process of detecting ocular diseases, so the research team proposes a new target segmentation method that combines Convolutional Neural Networks (CNN) and image search to accurately segment retinal boundaries in images. Experimental results show that the method is able to reduce the probability of misclassification as a target and has strong adaptive power [10]. Lu et al. proposed a background-free retina segmentation method for the time-consuming and laborious problem of medical image segmentation. This method uses a hierarchical clustering algorithm for binary classification, and then inputs the classification results into a neural network for training. Experimental results show that this segmentation method outperforms other state-of-the-art retinal vascular image segmentation methods [11]. Gong et al. proposed a method that uses deep learning and multivariate geometric feature segmentation of target space to accurately solve the problem of targeting and occluding target fruits for the target localisation and mastering of occluded fruits of a fruit picking robot. The experimental results show that the target segmentation accuracy of this method is improved by nearly 10 points compared with the traditional method, and the success rate is high [12].

Combined with the above many practical cases and experimental conclusions, it is not difficult to find that most of the previous experimental objects are based on the de-blurring of images, proposed and experimented with many solutions. And there is still some room for development in the research of video deblurring methods. This research firstly solves the problem of video deblurring, and secondly re-proposes a video target segmentation method based on the existing algorithm to extract and reorganise the video data in an attempt to provide certain new value for the research of high-definition video restoration methods.

III. Research on Deblurring Algorithm and Target Segmentation Optimisation Algorithm for Video Data

Video as a special transmission medium, in which video deblurring and video segmentation have become a hot research topic nowadays. In Section I, an algorithm using scale recurrent network is proposed for video deblurring, based on which Hal's two-dimensional wavelet transform is introduced for feature transformation extraction, and finally the attention mechanism is introduced to achieve better video deblurring effect. Section II video oriented segmentation proposes an optimisation algorithm for video object segmentation using morphological and attention mechanism (Attention and Morphology Video Object Segmentation, AMVOS), which is used in One-Shot Video Object Segmentation, OSVOS), which also introduces the attention mechanism on the basis of One-Shot Video Object Segmentation (OSVOS), in particular, the algorithm makes use of the morphological base module to filter the non-target information so as to get a better target segmentation.

A. Research on Video Data Oriented Deblurring Algorithm

S Scale-recurrent Network (SRN) can reduce the input video image into feature maps with more channels, and later reduce it to the original size of the video. This feature not only reduces the complexity of the network framework, but also allows for the development of innovations by integrating with multiple mechanisms. This research will take a deeper step in the development of the framework of this network. In daily life, video images are blurred due to camera shake or other unavoidable factors, so it is increasingly important to study how to remove video blur. In which the process of blur generation is shown in Equation (1).

$$y = x * k + n \tag{1}$$

In Equation (1), y denotes fuzzy video image; x denotes de-sharpened video image; k denotes fuzzy kernel; n denotes noise; and $*$ denotes convolution operator. The Haar 2D wavelet transform is a fast discrete wavelet transform that is fast and efficient [13]. It is suitable for image compression and enhancement. The function mapping relationship of the Haar 2D wavelet transform as preprocessing is shown in Fig. 1.

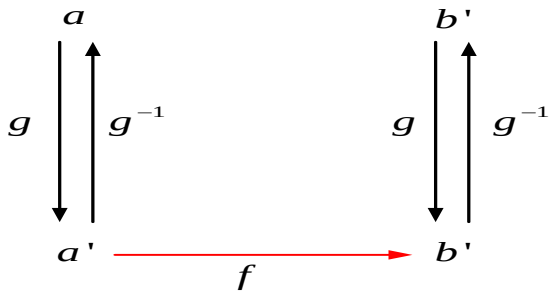


Fig. 1. Wavelet transform mapping relationship.

In Fig. 1, a denotes the input fuzzy video image to be processed; b denotes the completed processed video image; g, f are two different mapping functions, and g is a 2D wavelet function. In this way, the video processing becomes more diversified while reducing the complexity of the network framework. After the introduction of 2D wavelets, the cyberspace results at this point need to be more expressive, so this research adds the Spatial-Channel Attention Block (SCAB) [14-15]. The network framework diagram of SCAB is shown in Fig. 2.

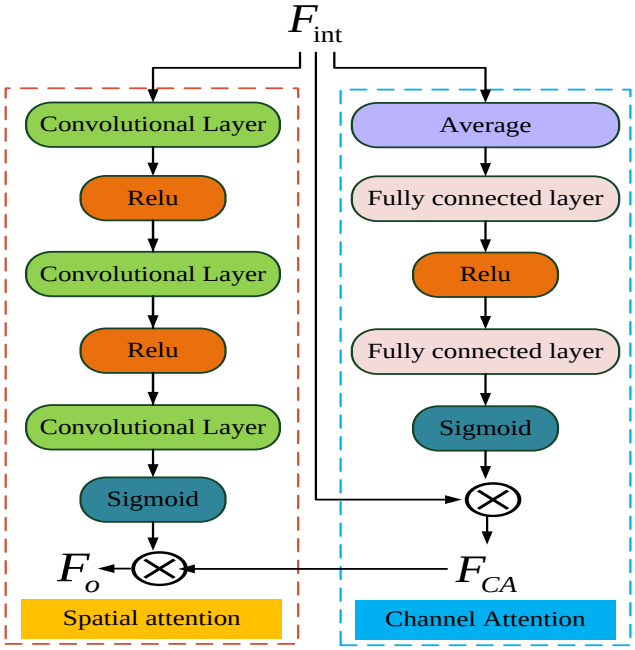


Fig. 2. Network framework diagram of SCAB.

In Fig. 2, F_o denotes the output of SCAB; F_{int} denotes the input of SCAB; F_{CA} denotes the output of attention; F_i denotes the input of channel attention; and F_{SA} denotes the output of spatial attention. As can be seen in Fig. 2, the output value of spatial attention is obtained after the input of F_{int}, which is calculated by the spatial attention module, convolutional layer and Relu function, and then by the convolutional layer and sigmoid function. The branch F_i is input into the average pooling, and the output is obtained from F_{CA} after all-connected layer, Relu function, all-connected layer, and sigmoid function. Finally, multiplying the outputs of spatial attention and channel attention gives the final output value of SCAB F_o. The specific formula is shown in Equation (2) [16-17].

$$F_o = F_{SA} \otimes \left(F_{CA} \otimes F_i \right) \tag{2}$$

In Equation (2), F_o represents the output of SCAB; F_{SA} represents the output of spatial attention branch; F_{CA} the output of attention branch; and $\otimes$ represents the multiplication. Inception module can aggregate different filter features and further collect different features under the action of the residual grid to make up for the lack of spatio-temporal features extracted by the network framework. Therefore, on this basis, it is proposed to embed the SCAB module into the structure of jump connection block to form Skip Spatial-Channel Attention (SSCA), and into the residual block to form Residual Spatial-Channel Attention, RSCA), embedded in the Inception module to form the corresponding residual Inception Spatial-Channel Double Attention block (Residual Inception Spatial-Channel Attention, RISCA). Where the process of RISCA module can be represented by Equation (3).

$$F_o = SCAB(f^{2\times2}(f^{2\times2}(F_i) \oplus (f^{4\times4}(f^{2\times2}(F_i))) \oplus (f^{4\times4}(f^{4\times4}(F_i))))) + F_i \tag{3}$$

In Equation (3), F_o denotes the RISCA output value; F_i denotes the input value; $f^{2\times2}$ is the convolutional layer with 2x2 convolutional kernel; $f^{4\times4}$ denotes the convolutional layer with 4 × 4 convolutional kernel; $\oplus$ is the cascade level; and $SCAB(\bullet)$ is the application function of SCAB module.

SSCA is a spatial-channel attention module embedded in a hopping connection that connects both the encoding and decoding network frameworks for nonlinear changes for better noise suppression to extract feature images. The process of SSCA module can be represented by Equation (4).

$$s = F_i + SCAB(f^{4\times4,2}(f^{4\times4,2}(f^{4\times4,1}(F_i)))) \tag{4}$$

In Equation (4), s denotes the output value; F_i denotes the input value; $f^{4\times4}$ denotes the convolutional layer with 4 × 4 convolutional kernel; and $SCAB(\bullet)$ is the application function of the SCAB module. RSCA is the addition of the SCAB module after the residual layer in the original multiscale network, which further enriches the framework of the network. The process of the RSCA module can be represented by Equation (5).

$$F_o = F_i + SCAB(f^{4\times4,2}(f^{4\times4,1}(F_i))) \tag{5}$$

In Equation (5), F_i denotes the input value; F_o denotes the output value; $f^{4\times4}$ denotes the convolutional kernel as a 4 × 4 convolutional layer; and $SCAB(\bullet)$ is the application function of SCAB module. The loss function is used as another reference index to monitor the accuracy of video restoration, and the traditional loss function is shown in Equation (6).

$$L_2 = \sum_{n=1}^{N} \frac{1}{N} \left\| V^n - V_*^n \right\|_2^2 \tag{6}$$

In Equation (6), L_2 denotes the traditional loss function;

N denotes the number of samples; V denotes the clear video image; and $V_{\bullet}$ denotes the output video image. Since L_2 is not robust, it is necessary to introduce the perceptual loss function, whose expression is shown in Equation (7).

$$L_p = \frac{1}{W_{s,q} H_{s,q}} \sum_{A=1}^{W_{s,q}} \sum_{B=1}^{H_{s,q}} (\phi_{s,q}(V)_{A,B} - \phi_{s,q}(V_*)_{A,B})^2 \quad (7)$$

In Equation (7), $W_{s,q} \times H_{s,q}$ is the feature map size and $\phi_{A,B}$ is the feature map obtained from the Ath maximisation layer trained on the dataset by the Bth convolution. The edge loss function is introduced to further extract the edge features and its expression is shown in Equation (8).

$$L_e = \frac{1}{N} \sum_{n=1}^{N} (\left\| \nabla_x V^n - \nabla_x V_* \right\|_1 + \left\| \nabla_y V^n - \nabla_y V_*^n \right\|_1) \quad (8)$$

In Equation (8), ∇y denotes the vertical difference; ∇x denotes the horizontal difference; and the sum of Equation (6), Equation (7), and Equation (8) is the loss function L_t used in this study, the expression of which is shown in Equation (9).

$$L_t = \alpha L_2 + \beta L_p + \delta L_e \quad (9)$$

In Equation (9), α , β , and δ denote the weight coefficients of the loss functions L_2 , L_p , and L_e respectively. In summary, this study proposes the HAVD algorithm (Haar and Attention Video Deblurring, HAVD), which makes use of the algorithm of scaled recurrent network, and introduces the Haar 2D wavelet transform as a preprocessing based on this method, while combining the attention mechanism and the channel, then we get a new type of video deblurring process as shown in Fig. 3.

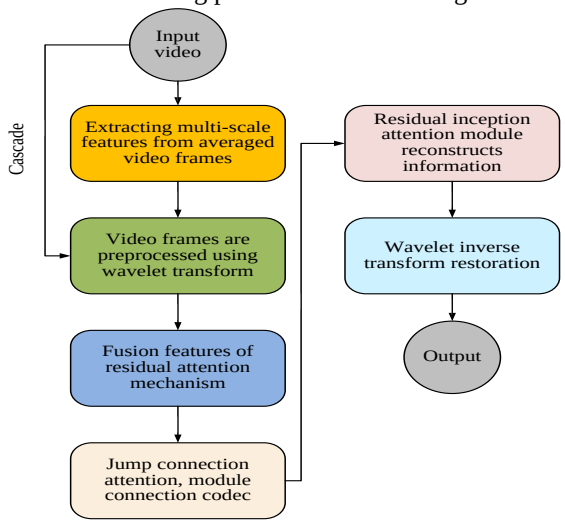


Fig. 3. Video deblurring process.

In Fig. 3, the de-blurring process before improvement can be decomposed into the following operations. Firstly, the original video is input, after preprocessing and noise suppression effect through wavelet transform. Then after the extraction of features in the residual attention module, the extracted features are passed through the module encoder, after which the encoded information is recombined using the dual attention module, and finally the recombined information is integrated and output using the wavelet transform, and the final video result is obtained. The improved deblurring process introduces the Haar wavelet transform after inputting the original video, and inputs the video into the wavelet transform module for processing, and the whole process is simpler than before the improvement due to the sparsity of the 2D wavelet.

B. Research on Optimisation Algorithm for Objective Segmentation of Video Data

Nowadays, target segmentation of video data covers a wide range of areas, such as target motion capture under video conferencing, separation of different codes in video compression, etc. The OSVOS algorithm is a single-frame video segmentation algorithm, which is only applicable under good lighting environment. The OSVOS algorithm is a single-frame video segmentation algorithm, which is only applicable to the background of good lighting environment [18]. However, the segmentation performance of OSVOS algorithm decreases dramatically under more complex regional backgrounds. Therefore, it can be said that the target segmentation efficiency of OSVOS algorithm is changed by the influence of interference factors. It so happens that the Non-Local mechanism is able to capture the critical region of the video frame and also belongs to a kind of self-attention mechanism. In this context, this study combines the OSVOS algorithm with the Non-Local mechanism to explore a new video segmentation method. The Non-local mechanism can be expressed as Equation (10).

$$y_i = \frac{1}{C(X)} \sum_{\forall j} f(X_i, X_j) g(X_j) \quad (10)$$

In Equation (10), X denotes the input signal; i denotes the index of the feature element; and j denotes the index of all the points obtained using the enumeration method. The function $C(X)$ operates similarly to softmax with the aim of normalisation and in the proposed algorithm $C(X)$ is set to $\sum_{\forall j} f(X_i, X_j)$. On this basis it can be derived that the attention mechanism module of this new algorithm can be expressed as Equation (11) [19-20].

$$y_i = \frac{1}{\sum_{\forall j} e^{\theta(X_i)T\phi(X_j)}} e^{\theta(X_i)T\phi(X_j)} W_g X_j \quad (11)$$

In Equation (11), the function and ϕ function denote 1 × 1 convolution, and $g(X_i)$ denotes the unitary input function. The function $e^{\theta(X_i)^T \phi(X_j)}$ computes the similarity between

the i th position and the j th position $f\left(X_i, X_j\right)$.

After the target segmentation of a video is performed, the accuracy usually decreases or some non-target regions appear due to various factors [21, 22]. So based on this an optimisation algorithm using morphology is proposed to remove these non-target regions. The algorithm removes the non-target regions using ecology algorithm is shown in Fig. 4.

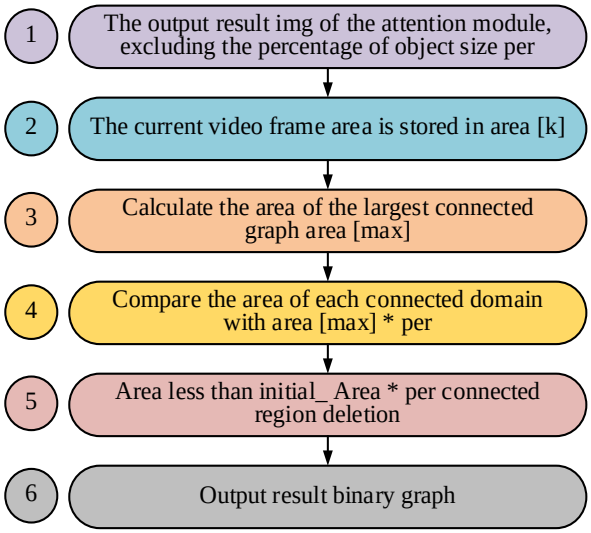


Fig. 4. Morphology based algorithm for locating non target regions.

In Fig. 4, the output of the Attention Mechanism module is first taken as an input, and then the area of each video region is calculated using the formula, and the result is accumulated and placed in the area of the plurality of region concatenation. The largest area of the concatenated map is filtered and each concatenated region is compared with this largest area. Finally, thresholds smaller than this area are removed to get the final output. Taken together, this research proposes a new AMVOS algorithm using OSVOS, which combines the CSVOS algorithm with the introduction of a self-attention mechanism to extract the key information of the video frames. Finally, the results of the algorithm are optimised by combining the ecological aids. The flow of video target segmentation under AMVOS algorithm is shown in Fig. 5.

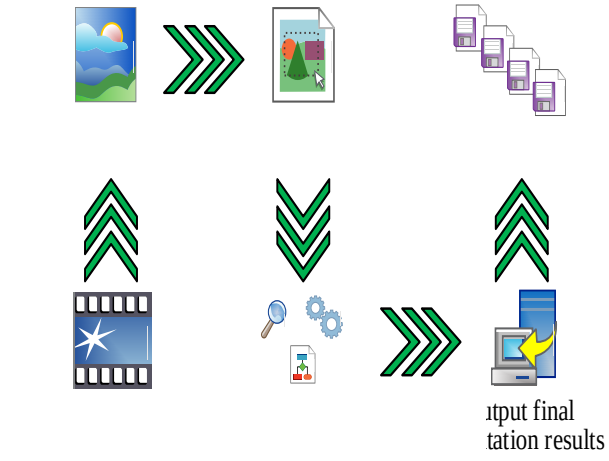


Fig. 5. Video object segmentation process.

In Fig. 5, the video frames that have been deblurred are firstly input, and then the first frame given beforehand in the mechanism is used for online fine-tuning training, and then the preliminary segmentation results are inputted into the morphological module for the target and non-target region refinement, and finally the final segmentation results are obtained.

IV. Performance Test of Fuzzy Goal Segmentation Optimisation Algorithm for Basketball Video Data

The video deblurring optimisation algorithm and the target segmentation optimisation algorithm proposed in the methods section above are combined to work together on basketball videos under a video dataset, and the peak signal to noise ratio (PSNR), structural similarity (SSIM), region similarity J and contour accuracy are used as the measurement metrics to try to investigate the performance of the combined performance of these two optimisation algorithms, accuracy F as measurements to investigate the combined performance of these two optimisation algorithms.

A. Performance Tests of De-Fuzzy Optimisation Algorithm and Goal Segmentation Optimisation Algorithm

A suitable experimental environment was set up for this experiment, and the number of samples in the training set accounted for 90% of the COCO dataset, while the number of samples in the test set accounted for 10% of it. The GPU was NVIDIA GeForce RTX 2060 SUPER, and the deep learning framework was Caffe. The three weighting coefficients, e, were set to 100,000, d, 35, and g, and the number of iterations was 5,000. number is 3,000.

Different modules in the HAVD algorithm have some influence on the overall performance, the experiment defines the algorithm model of multi-scale network as MSN, the algorithm model of single-scale network as SCN, the algorithm model with wavelet transform added on top of MSN as MSNH, and the continued addition of dual-attention mechanism on top of MSNH as HAVD. The peak signal-to-noise ratio and the structural similarity are also used as reference indexes for the evaluation of the algorithm model. As shown in Fig. 6.

Fig. 6(a) shows the Peak to Noise Ratio (PSNR) plots for various algorithms with different number of samples, and Fig. 6(b) shows the similarity of results with different number of samples for various algorithms with similar motion. In Fig. 6, the higher PSNR score indicates the better deblurring effect of the video, and similarly, the higher SSIM score indicates the better recovery of the video. The HAVD algorithm improves over the above three algorithms in terms of both PSNR and SSIM metrics, where the PSNR value can be exceeded up to 30.55 and the SSIM value can be improved up to 0.942. The HAVD algorithm performs best in both metrics because it combines several advanced techniques, including multi-scale network structure, wavelet transform and spatial channel attention mechanism. The multi-scale network structure can capture the different scale features of the video image more comprehensively, and the wavelet transform effectively separates and processes the low-frequency and high-frequency components of the image, resulting in excellent performance in denoising and detail enhancement. In particular, the spatial channel attention mechanism further

enhances the clarity and detail recovery of the image by adaptively adjusting the weights in the feature map, highlighting the key features, and suppressing the interference of non-target regions. Thus, it can be shown that the HAVD algorithm can maintain high image recovery quality when processing large-scale data, and the denoising of basketball video with this algorithm can obtain clearer images, so as to restore the movement changes of basketball players.

On the other hand, combining the HAVD algorithm and the various types of deblurring algorithms previously proposed by scholars are compared horizontally under a Gopro dataset, and the evaluation results are shown in Fig. 7.

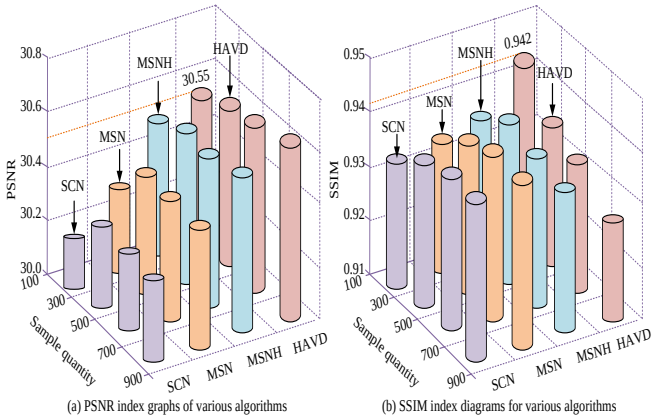


Fig. 6. The impact of different modules on the network.

Fig. 7(a) shows the peak-to-noise ratio curves for each type of algorithm and Fig. 7(b) shows the structural similarity curves for each type of algorithm. Algorithm 1 in Fig. 7 is the Wiener filtering algorithm; Algorithm 2 is the regular filtering method; and Algorithm 3 is the blind convolution method. As the number of samples increases, the newly proposed HAVD algorithm consistently scores higher than the other three compared algorithms in terms of PSNR and SSIM. When the sample size is 50, the PSNR and SSIM of HAVD are as high as 30.72 and 0.95, respectively. The HAVD algorithm is significantly higher than traditional algorithms in both metrics, which is mainly attributed to the multi-level feature extraction and advanced attention mechanism of the HAVD algorithm. Traditional Wiener filtering and conventional filtering methods have limited effect in dealing with complex motion and high-frequency details, which is easy to cause image blurring and loss of details, while blind convolution methods can deal with unknown blurring kernels, but the computational complexity is high and the effect is unstable in dynamic videos. In contrast, the HAVD algorithm effectively separates the low-frequency and high-frequency components of the image through the Haar 2D wavelet transform, and combines the attention mechanism to adaptively adjust the weights of the feature maps, which highlights the key features and reduces the interference of the non-target information, thus achieving better noise removal and detail recovery. This enables the HAVD algorithm to balance noise removal and detail recovery when processing video deblurring, and significantly improves the PSNR and SSIM values, demonstrating its excellent performance in video processing.

TABLE I
BENCHMARK PERFORMANCE OF VARIOUS DEBLURRING ALGORITHMS

Algorithm	Precision	Recall	F1 value
SCN	0.85	0.81	0.82
MSN	0.86	0.85	0.85
MSNH	0.88	0.84	0.87
Algorithm 1	0.86	0.83	0.85
Algorithm 2	0.90	0.93	0.92
Algorithm 3	0.91	0.89	0.90
HAVD	0.98	0.97	0.97

The benchmark performance of the seven deblurring algorithms in Fig. 6 and Fig. 7 is given in Table I. From Table I, it can be seen that the proposed deblurring optimisation algorithm HAVD has higher values of deblurring precision, recall and F1 compared to SCN, MSN, MSNH, Algorithm 1, Algorithm 2 and Algorithm 3. The precision, recall, and F1 values of HAVD are as high as 0.98, 0.97, and 0.97, respectively, while SCN has the lowest precision, recall, and F1 values, which are as low as 0.85, 0.81, and 0.82, respectively.

The AMVOS algorithm also consists of three mechanism modules, and this experiment will explore each type of mechanism module separately. The first category is the traditional OSVOS algorithm; the second category is the OSVOS + attention mechanism module; the third category is the OSVOS + morphology module; and the fourth category is the newly proposed AMVOS algorithm. The Davis dataset is divided into two categories: test set and training set, and the region similarity J, contour accuracy F and temporal stability T are introduced as reference indexes in it, and the specific test results are shown in Table II.

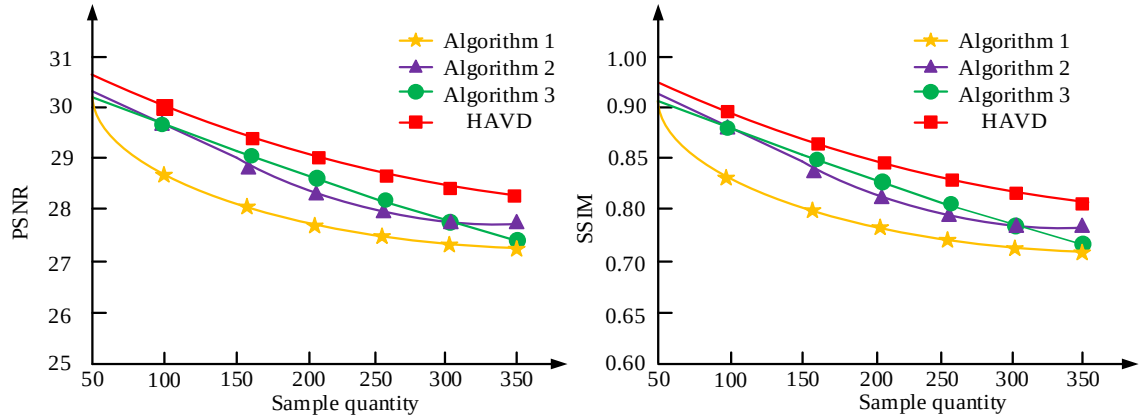


Fig. 7. Peak to noise ratio and structural similarity curves of various algorithms.

TABLE II
PERFORMANCE INDICATOR RESULTS OF VARIOUS ALGORITHMS' TEST AND TRAINING SETS

/	OSVOS	OSVOS + Attention	AMOVS
J Test set	93.6	92.2	94.9
J Training set	91.3	90.8	94.5
F Test set	92.6	91.9	93.2
F Training set	93.5	92.8	93.1
T Test set	37.6	31.9	31.2
T Training set	35.5	33.4	30.8

From Table II, it can be seen that the larger values of region similarity J and contour accuracy F indicate better feature extraction and target segmentation of the video, while the smaller value of temporal stability T indicates that the transition from the previous frame to the next frame tends to be more stable. The traditional OSVOS algorithm has low region similarity and contour accuracy and poor temporal stability. And after adding the attention mechanism and morphological module respectively, the newly proposed AMVOS algorithm has the highest result score and better temporal stability, while the scores of the three types of indexes in the training set and the test set are similar, which indicates that this video target segmentation method not only enhances the selection of features, but also operates very stably and improves the region similarity and contour accuracy. According to the test results and training results in Table II, AMVOS algorithm shows a high degree of consistency and stability on the two data sets, which indicates that the algorithm has a better benchmark performance under different conditions, and has higher reliability when processing subsequent basketball video data. This is because the AMVOS algorithm optimises the segmentation of target and non-target regions by adaptively adjusting the feature weights in the video frames by combining the attention mechanism and the morphological module, which significantly improves the region similarity and contour accuracy. Meanwhile, the algorithm is able to maintain consistent high performance on different datasets, which proves its reliability and robustness in handling video target segmentation tasks.

On the other hand, the AMVOS algorithm is also compared with various types of video target segmentation algorithms previously proposed by scholars under the same davis2016 dataset, which is tested with region similarity J and contour accuracy F as reference metrics. The evaluation results are shown in Fig. 8.

Fig. 8(a) shows the region similarity curve of each type of segmentation algorithm, and Fig. 8(b) shows the contour accuracy curve of each type of segmentation algorithm. In Fig. 8, Algorithm 1 is an automatic segmentation algorithm; Algorithm 2 is a semi-automatic segmentation algorithm; and Algorithm 3 is an interactive segmentation method. From Fig. 8, it can be seen that the AMVOS algorithm acts on the target segmentation of video sequences after the region similarity is up to 96.5% and the contour accuracy is up to 93%. The region similarity and contour accuracy of the AMVOS algorithm in Fig. 8 show a slight decline on some video sequences, which may be due to the difficulty in feature extraction caused by object occlusion or illumination changes in the video. However, compared with other traditional video object segmentation algorithms, the score of AMVOS algorithm is still higher. Therefore, the superiority of AMVOS algorithm in the field of video object segmentation can be proved.

B. Performance Effectiveness Test of De-Fuzzy Optimisation Algorithm and Objective Segmentation Optimisation Algorithm Applied to Basketball Video

The HAVD algorithm and AMVOS algorithm are combined together and used in the field of deblurring and target segmentation of basketball videos. Due to the fact that general videos contain many elements, this time, in order to achieve the uniqueness and accuracy of the experiment, only the action categories in basketball videos are used as the test objects of this experiment. There are many basketball action categories, which are listed here as shooting, layup, dribbling in place, caps, running without the ball, and so on. The Gygo dataset is used as the experimental dataset, which contains 30 videos of the training set and 20 videos of the test set, and the frame sizes of the selected videos are all 854 × 480. The region similarity J and contour accuracy F are used as the reference rating metrics. The performance test of the deblurring and target segmentation optimisation algorithm for basketball videos is shown in Fig. 9.

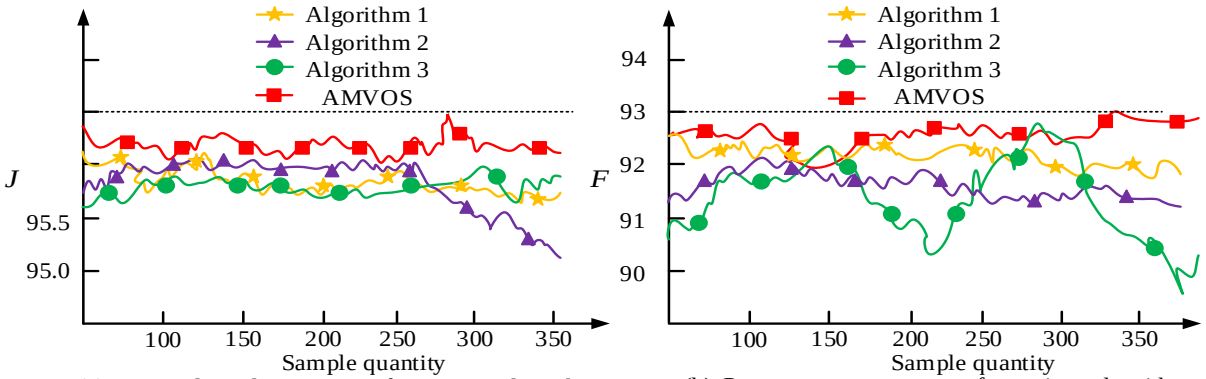
As can be seen from Fig. 9, the official truth value of video segmentation is about 1.5 s, and the truth value of deblurring and target segmentation of video is about 3.1 s. For the target segmentation algorithms of video, the mean value of the target segmentation time of AMOVS algorithm is closest to 1.5 s for the different types of basketball action categories, while the video processing time under HAVD + AMOVS algorithm is closer to 3.1 s. The actual performance test of deblurring and target segmentation optimisation algorithms can be further illustrated in terms of four reference metrics: effectiveness, robustness, recall and temporal efficiency, as shown in Table III.

TABLE III
PERFORMANCE INDICATORS TESTING OF VARIOUS ALGORITHMS

/	HAVD	OSVOS	AMOVS	HAVD + OSVOS	HAVD + AMOVS
Effectiveness	8.7	7.4	9.1	16.1	17.8
Robustness	8.6	8.5	8.9	17.1	17.5
Recall	8.3	6.8	8.8	15.1	17.1
Spatiotemporal efficiency	9	7.2	8.7	16.2	16.7

As can be seen from Table III, the combined effectiveness and recall scores of HAVD + AMOVS algorithm are 17.8 and 17.1, respectively, which are far more than other algorithmic models. Meanwhile, combining the above experimental data information, the video deblurring effectiveness and robustness of HAVD + AMOVS algorithm exceed the traditional video processing algorithms. It shows that the newly proposed HAVD deblurring algorithm and AMOVS target segmentation algorithm are more suitable for the

deblurring preprocessing of action and segmentation of action feature targets in basketball video. This is mainly due to the excellent performance of HAVD algorithm in deblurring and the efficiency and accuracy of AMVOS algorithm in target segmentation. By combining the multi-scale feature extraction, attention mechanism and morphological module, the HAVD + AMVOS algorithm is able to achieve more efficient deblurring and target segmentation when processing complex dynamic video data, which significantly improves multiple metrics such as effectiveness, robustness, recall, and spatio-temporal efficiency, and proves its superior performance and practical application value in the field of basketball video processing.



(a) Regional similarity curves for various algorithms (b) Contour accuracy curves for various algorithms

Fig. 8. Regional similarity and contour accuracy curves of various algorithms.

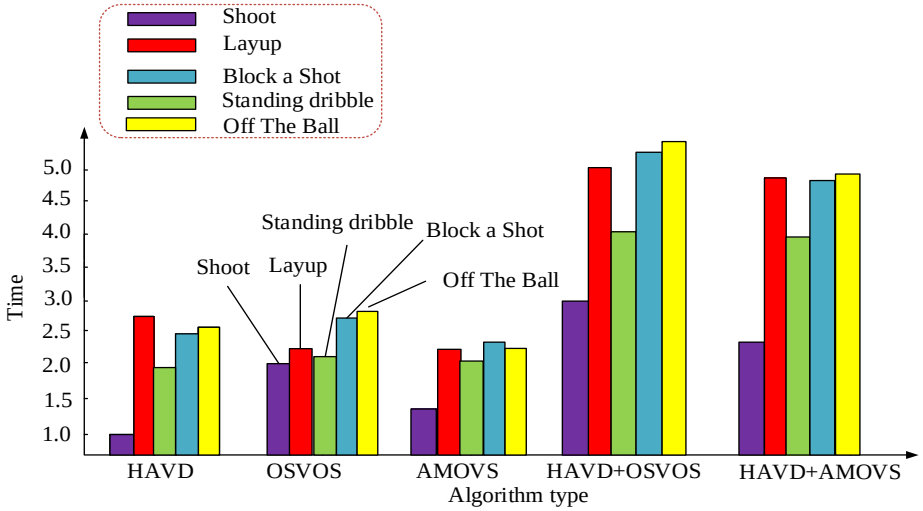
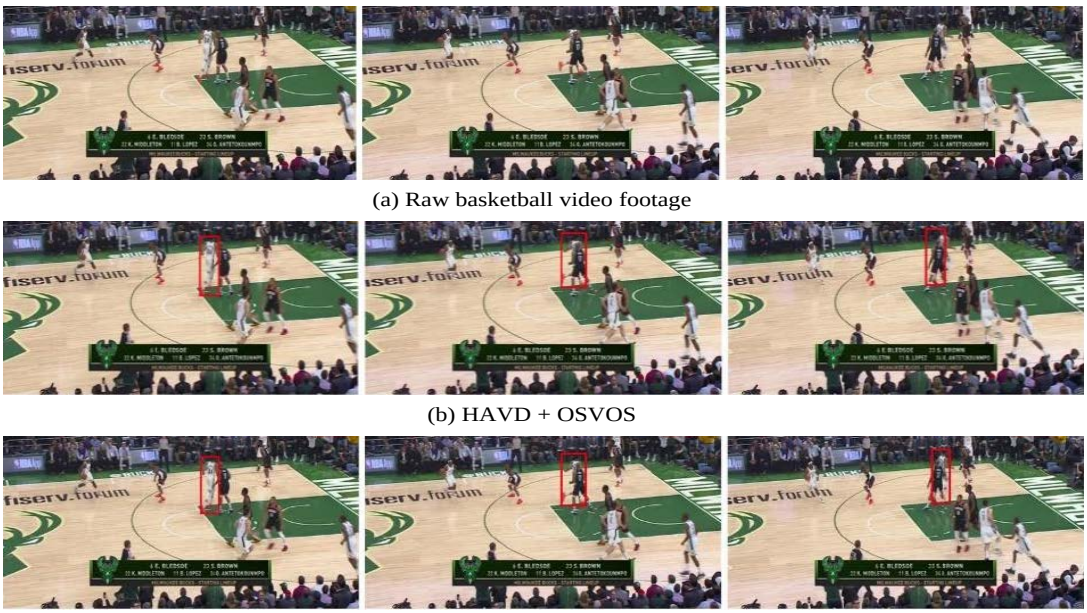


Fig. 9. Algorithm performance test graph for basketball video.



(a) Raw basketball video footage

(b) HAVD + OSVOS

(c) HAVD + AMOVS

Fig. 10. Algorithm performance test graph for basketball video.

Fig. 10 compares the actual analysis effects of two different algorithms in basketball data sets. Fig. 10 (a) shows a continuous basketball video, while Fig. 10 (b) and Fig. 10 (c) show the segmentation results of basketball video under the two algorithms HAVD+OSVOS and HAVD+AMOVS respectively. By comparing Fig. 10 (b) and Fig. 10 (c), it can be seen that the HAVD+AMOVS algorithm can accurately track the target object in a continuous period of time, which indicates that the algorithm has strong defuzzification and segmentation capabilities.

V. Conclusion

In order to find an optimisation algorithm for deblurring and target segmentation for basketball videos, which can be used to solve the current deficiencies in the field of video deblurring and target segmentation, such as poor processing clarity, slow processing time, and low accuracy of target segmentation. In this research, HAVD deblurring optimisation algorithm is proposed by introducing Hal 2D wavelet and attention mechanism on the basis of existing scaled recurrent network. The AMVOS target segmentation optimisation algorithm is proposed by combining the attention mechanism and morphological algorithm on the basis of the existing OSVOS algorithm. The experimental results show that the HAVD algorithm has the highest peak-to-noise ratio up to 30.55, and the highest structural similarity up to 0.942. The two test metrics are higher than each individual algorithm module. The performance of the HAVD algorithm was tested against previous conventional algorithms, and it was found that both PSNR and SSIM metrics of the HAVD algorithm were higher than any other algorithms, indicating the best de-blurring reduction. Based on the existing OSVOS algorithm, the AMOVS algorithm is proposed by introducing the Attention Mechanism Module and Morphology Module, which is used for the optimal segmentation of video targets. The experimental results show that after AMOVS algorithm acts on the target segmentation of video data, it obtains region similarity up to 96.5% and contour accuracy up to 93%. The final experiment applies the HAVD + AMVOS algorithm to the processing of basketball video, and the experimental results show that the processing time using the HAVD + AMVOS algorithm is closer to the true value, which demonstrates that the combination of these two algorithms results in faster and better video deblurring and target segmentation. However, due to the existence of certain errors in the index measurement data of this experiment, the experimental results may have certain limitations and incompleteness, which is also a place where the subsequent experiments can be improved.

References

[1] Z. Wang, X. S. Li, X. J. Zhang, Y. R. Bai, C. C. Zheng, “Blind image deblurring for a close scene under a 6-DOF motion path,” Sensor Review, vol. 41, no. 2, pp. 216-226, 2021.

[2] X. Peng, Y. Sui, J. D. Trzasko, K. J. Glaser, J. Huston, R. L. Ehman, J. G. Pipe, “Fast 3D MR elastography of the whole brain using spiral staircase: data acquisition, image reconstruction, and joint deblurring,” Magnetic Resonance in Medicine, vol. 86, no. 4, pp. 2011-2024, 2021.

[3] Y. Cao, L. J. Sun, C. Han, J. Guo, “Attention-based video object segmentation algorithm,” IET Image Processing, vol. 15, no. 8, pp. 1668-1678, 2021.

[4] Q. M. Peng, Y. M. Cheung, “Automatic video object segmentation based on visual and motion saliency,” IEEE Transactions on Multimedia, vol. 21, no. 12, pp. 3083- 3094, 2019.

[5] W. Huang, H. M. Yu, “Cooperative object segmentation and recognition via restricted Boltzmann machine,” Electronics Letters, vol. 56, no. 8, pp. 378-380, 2020.

[6] C. Li, L. Song, R. Xie, W. J. Zhang, “L0 structure-prior assisted blur-intensity aware efficient video deblurring,” Neurocomputing, vol. 483, no. 28, pp. 195-209, 2022.

[7] Z. Q. Zhan, X. Yang, Y. H. Li, C. Pang, “Video deblurring via motion compensation and adaptive information fusion,” Neurocomputing, vol. 341, no. 14, pp. 88-98, 2019.

[8] X. Q. Zhang, R. H. Jiang, T. Wang, P. C. Huang, L. Zhao “Attention-based interpolation network for video deblurring,” Neurocomputing, vol. 453, no. 17, pp. 865-875, 2020.

[9] Q. R. Li, H. Liu, J. P. Zhang, P. F. Zeng, “Target segmentation of industrial smoke image based on LBP Silhouettes coefficient variant (LBPSCV) algorithm,” IET Image Processing, vol. 14, no. 12, pp. 2879-2889, 2020.

[10] K. Hu, B. W. Shen, Y. Zhang, C. H. Cao, F. Xiao, X. P. Gao, “Automatic segmentation of retinal layer boundaries in OCT images using multiscale convolutional neural network and graph search,” Neurocomputing, vol. 365, no. 6, pp. 302-313, 2019.

[11] Z. Lu, D. L. Chen, D. Y. Xue, “Weakly supervised retinal vessel segmentation algorithm without groundtruth,” Electronics Letters, vol. 56, no. 23, pp. 1235-1237, 2020.

[12] L. Gong, W. J. Wang, T. Wang, C. L. Liu, “Robotic harvesting of the occluded fruits with a precise shape and position reconstruction approach,” Journal of Field Robotics, 39: 69-84, 2022.

[13] H. Zadhoush, A. Giannopoulos, “Optimising GPR time-zero adjustment and two-way travel time wavelet measurements using a realistic three-dimensional numerical model,” Near Surface Geophysics, vol. 20, no. 2, pp. 208-226, 2021.

[14] X. M. Zhao, M. Y. Qi, Z. W. Liu, S. H. Fan, C. Li, M. Dong, “End-to-end autonomous driving decision model joined by attention mechanism and spatiotemporal features,” IET Intelligent Transport Systems, vol. 15, no. 9, pp. 1119-1130, 2021.

[15] X. Y. Wang, M. H. Cheng, J. Eaton, C. J. Hsieh, S. F. Wu, “Fake node attacks on graph convolutional networks,” Journal of Computational and Cognitive Engineering, vol. 1, no. 4, pp. 165-173, 2022.

[16] R. S. Al-Gounmeein, M. T. Ismail, B. N. Al-Hasanat, A. M. Awajan “Improving models accuracy using kalman filter and holt-winters approaches based on ARFIMA Models,” IAENG International Journal of Applied Mathematics, vol. 53, no. 3, pp. 869-878, 2023.

[17] Y.B. Wang, Y. Duan, Y.Y. Li, H.Y. Wu, “Smoke recognition based on dictionary and bp neural network,” Engineering Letters, vol. 31, no. 2, pp. 554-561, 2023.

[18] L. Z. Wu, L. C. He, W. Chen, X. H. Hao, “Recognition method of voltage sag sources based on rmt-cnn model,” IAENG International Journal of Applied Mathematics, vol. 53, no. 3, pp. 934-944, 2023.

[19] W. Liu, M.Y. Cui, L.Y. Ma, “A novel deep learning-based disocclusion hole-filling approach for stereoscopic view synthesis,” IAENG International Journal of Applied Mathematics, vol. 53, no. 2, pp. 524-533, 2023.

[20] X. Zhang, Y. Zhong, B. Zhang, S. Nie, “Optimization of the NSGA-III algorithm using adaptive scheduling,” Engineering Letters, vol. 31, no. 2, pp. 459-466, 2023.

[21] Z. Hu, L. Wang, Y. Luo, Y. Xia, H. Xiao, “Speech emotion recognition model based on attention CNN Bi-GRU fusing visual information,” Engineering Letters, vol. 30, no. 2, pp. 427-434, 2022.

[22] M. Zaini, R. T. Yuliarto, G. P. Gusti, Y. Agustira, “The Influence of video tutorial learning media on improving financial literacy knowledge: a study for e-commerce user students,” Malaysian E Commerce Journal, vol. 6, no. 2, pp. 72-75, 2022.

Aixin Yang received Master's degree in Physical Education and Training from Shanghai Sport University in 2007 and 2010. She has been working as a full-time physical education teacher at Hunan University of International Studies since June 2010. Teaching professional courses in the School of Physical Education, including badminton, health education, and sports research methods; Public physical education courses: basketball, Tai Chi,

aerobics, sports dance, etc.

Lifu Huang graduated from Hunan Agricultural University in June 2015 with a Master's degree in Physical Education.
He is currently the leader of the Physical Education Teaching and Research Group at Hunan Normal University Affiliated Middle School, a first-class coach of the International Association of Athletics Federations, and a Chinese physical fitness trainer. He has won the first prize in Hunan Province's teaching curriculum and the special prize in Changsha's micro course competition. Currently serving as the coach of the athletics team at Hunan Normal University Affiliated Middle School

Haobo Liu is an associate professor and a basketball professional teacher at the School of Physical Education of Hunan University of International Economics; Graduated from Hunan University of Technology in June 2015 with a Master's degree in Sports. Received the first prize in the Hunan Provincial Teaching Competition, the top ten teaching plans of the School of Physical Education at Hunan University of Foreign Economics, a first level basketball referee, a second level volleyball referee, and an outstanding basketball coach in Hunan Province. Led Hunan University of Foreign Economics to achieve excellent results in the Hunan University Basketball Competition.