

Martial Arts Moves Recognition Method Based on Visual Image

Husheng Zhou*

Abstract

Intelligent monitoring, life entertainment, medical rehabilitation, and other fields are only a few examples where visual image technology is becoming increasingly sophisticated and playing a significant role. Recognizing Wushu, or martial arts, movements through the use of visual image technology helps promote and develop Wushu. In order to segment and extract the signals of Wushu movements, this study analyzes the denoising of the original data using the wavelet transform and provides a sliding window data segmentation technique. Wushu movement The Wushu movement recognition model is built based on the hidden Markov model (HMM). The HMM model is trained and taught with the help of the Baum-Welch algorithm, which is then enhanced using the frequency weighted training approach and the mean training method. To identify the dynamic Wushu movement, the Viterbi algorithm is used to determine the probability of the optimal state sequence for each Wushu movement model. In light of the foregoing, an HMM-based martial arts movements recognition model is developed. The recognition accuracy of the HMM model increases to 99.60% when the number of samples is 4,000, which is greater than the accuracy of the SVM (by 0.94%), the CNN (by 1.12%), and the BP (by 1.14%). From what has been discussed, it appears that the suggested system for detecting martial arts acts is trustworthy and effective, and that it may contribute to the growth of martial arts.

Keywords

Action Recognition, Hidden Markov Model, Martial Art, Visual Image, Wushu

1. Introduction

Parallel to the development of computing technology, visual image technology has also advanced and matured to the point that it is now widely applicable across many different sectors [1]. Machine vision technology, also referred to as visual image technology, is a vast field that employs image processing, pattern recognition, and artificial intelligence (AI) technologies [2]. Wushu culture may continue to grow and be passed down by using visual image technology to identify Wushu movements. To increase the efficacy and accuracy of the recognition of Wushu movements, an updated version of the hidden Markov model (HMM) model is used. Using visual image technology, computers can simulate the human visual system, enabling the extraction of useful information from photos and movies prior to processing and application [3]. Visual image technology is more effective and can handle a lot of information compared to manual recognition [4]. Both the application of a wavelet transform to de-noise the data and the presentation of a sliding window data segmentation method to segment and extract the signals of Wushu

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received June 16, 2022; first revision September 1, 2022; accepted September 14, 2022.

* **Corresponding Author:** Husheng Zhou (hushengzhou0516@163.com)

School of Physical Education, Huaibei Normal University, Huaibei, China (hushengzhou0516@163.com)

movements improve the accuracy and efficiency of Wushu movements recognition. The research primarily introduces two breakthroughs. The first method involves modeling martial arts actions with HMM before identifying and classifying the input sample data using the Viterbi algorithm to further identify martial arts movements. The second is to enhance the Baum-Welch algorithm by employing the frequency weighted training method and the mean training method to boost the algorithm's training effect on the HMM model and boost the model's correctness. The study consists of four sections. The second section provides a brief history of and discussion on the latest advancements and uses of visual image technology. The third part depicts the preprocessing of noise reduction and signal extraction from martial arts movement data, and the improvement of HMM based on which the martial arts movement recognition model is built. The fourth section is to verify and analyzes the recognition effect of the Wushu movement recognition model. The final section provides a brief overview of the complete piece.

2. Related Works

In recent years, there has been a rise in academic interest in visual image recognition technology, since it has become an increasingly crucial component of AI. Using sequential trisectionite decision-making and a formal definition of granular computing, Savchenko [5] suggested a novel technique to address image recognition system's slow performance when processing a large number of image types. A near-infrared and visible wavelength (VW) iris image identification approach based on plaque statistical feature integration was presented by Umer et al. [6] to address potential issues that may arise during the process of acquiring near-infrared or visible images. Hu et al. [7] undertook considerable research to design a cutting-edge intelligence-based data improvement system for Internet of Things (IoT) image recognition generation to address the issue of existing image recognition methods' inferior accuracy and efficiency. Ling et al. [8] presented and empirically validated a network-layer recursive reduced-order model compression approach to image recognition as a means to enhance its accuracy and efficiency. By establishing an underwater target identification model according to the features of sonar images, Jin et al. [9] developed a sonar image recognition approach based on convolutional neural networks (CNNs) to enhance the precision of autonomous underwater target detection. In the discussion of the use of image recognition in fan vibration fault identification and diagnosis, Huang et al. [10] developed an image recognition method for vibration fault diagnosis that makes use of the vibration signal's spectrum picture. To optimize ResNet models with various layers, Jafar and Lee [11] developed a super parametric technique and tested the model's effectiveness.

In order to achieve the quick and precise detection of fan blade damage, minimize loss, and increase efficiency, Yang et al. [12] developed an in-depth learning model for fan blade damage image recognition based on transfer learning and an integrated learning classifier in order to achieve the rapid and accurate detection of fan blade damage, minimize loss, and enhance efficiency. A novel facial image recognition method based on the cerebella basal ganglia mechanism was put out by Tang and Shabaz [13] to increase recognition accuracy. Tan and Liu [14] integrated CNNs and least weighted random search techniques to address the domain shift problem of the image recognition model in order to enhance the effect of domain shift on the recognition model. With the intention of mitigating visual attacks on neural networks, Andriyanov et al. [15] examined a variety of related techniques and proposed different preventative strategies. Meng et al. [16] realized the feature recognition and categorization of power line photos in

power company-owned forests by combining the Naive Bayesian classifier with an automated classification (NBC-AC) algorithm to properly categorize the observed inspection probability variables. Following the improved grey wolf optimization method, Jeya Christy and Dhanalakshmi [17] utilized the novel CNN ResNet-50 as hash technology and classifier to achieve content-based image recognition and deep learning marking. Wang et al. [18] categorized the characteristics recovered by the convolution network with a random forest model, thus resolving the issue of the poor recognition rate of gesture photos.

From the information provided above, it is clear that contemporary image recognition technology is essentially mature and has been extensively and significantly used in a variety of industries. Recognizing and fixing mistakes in martial arts technique is crucial. As a result, a martial arts movements identification technique based on visual picture is provided in order to detect martial arts movements properly and effectively, offer a basis for move mistake correction of martial arts performers, and increase the training effectiveness of martial arts performers.

3. Construction of Wushu Movement Recognition Model

3.1 Preprocessing and Segmentation of Martial Arts Movement Image Data

Technology based on the analysis of visual images is used to process Wushu videos in order to enhance the training impact and spread Wushu culture. The identification accuracy of Wushu movements can be enhanced by preprocessing the data collection. A data signal is denoised using a wavelet transform technique. Formula denotes the wavelet transform (1).

$$WT(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \cdot \psi\left(\frac{t - \tau}{a}\right) dt \quad (1)$$

In formula (1), a represents the variable scale; τ represents the variable translation. After multiple wavelet transforms, the obtained signal components are shown in formula (2).

$$S = A_n + D_1 + D_2 + \dots + D_n \quad (2)$$

In formula (2), S represents the original signal; D_n represents the noise signal obtained after n times of wavelet transform; A_n represents the effective signal obtained after n times of wavelet transform. Signal-to-noise ratio (SNR) and mean square error (MSE) are two metrics that may be used to assess the denoising performance of the wavelet transform (MSE). The best denoising impact may be found when SNR and MSE are the biggest and smallest, respectively. Support vector machine (SVM) is commonly employed in the segmentation and classification of video image sequences. However, overfitting phenomena and local optimization are easily produced by SVM, leading to inadequate classification and identification accuracy of Wushu movements. Consequently, the research utilizes window extraction and segmentation tools to segment data containing distinct martial arts moves. Sliding window extraction and event window extraction are the two primary categories of window extraction technology. The sliding window adopts a window with a predetermined size, slides along the time axis, and extracts the martial arts movements, in other words the collected martial arts movements signal is divided into n windows of the same length, which may overlap or not overlap. This formula is used to determine the net acceleration of Wushu movements (3).

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2} - 1 \tag{3}$$

In formula (3), a_x , a_y , and a_z represent the triaxial acceleration signal, -1 is the component of gravity acceleration and exists from beginning to end, and a represents the net resultant acceleration of martial arts move. The extraction of Wushu movements is shown in formula (4).

$$A_m = arg \max_m \left[\frac{1}{k} \sum_m^{m+k-1} a_m^2 \right] \tag{4}$$

In formula (6), a_m is the net resultant acceleration of the m -th sample point; k is the window size set in advance; A_m indicates that when the window size is k , a^2 takes the maximum value of the average. Set a minimum threshold and reset A_m to extract several martial arts movements continuously. Fig. 1 depicts the schematic diagram of sliding window segmentation and extraction of side leg pressing, which can be seen as an example of the side leg pressing in martial arts.

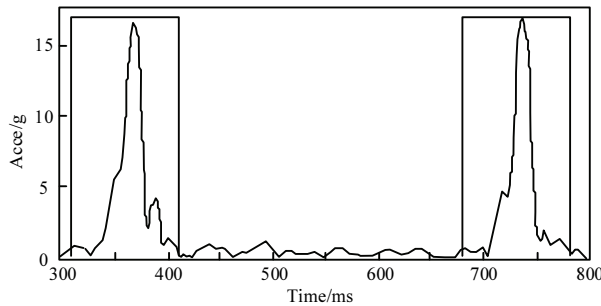


Fig. 1. Segmentation and extraction of side leg pressing movement.

As shown in Fig. 1, the sliding window extraction method can identify and extract the acceleration characteristic signal of Wushu movements, thereby facilitating their further recognition and categorization. With the use of event window extraction technology, a continuous martial arts movement signal is divided into n windows of varying widths, where the beginning and end of each window correspond to the beginning and end of a martial arts move. After extracting the martial arts movement, it contains multiple data sample points, and any sample point contains the acceleration component data of x-, y-, and z-axis.

3.2 Martial Arts Movement Recognition based on HMM

Martial arts techniques are broken down into frames in order to compress the quantity of data. Each martial arts move's data volume is set to 100 sample points. All actions are then broken down into meta actions. The waveform eigenvalues of the meta actions are then evaluated for their clustering properties, and the two eigenvalues with the highest clustering properties are used as the evaluation parameters. All feature vectors are quantified using the k-means algorithm and then transformed into one-dimensional vectors related to time series to facilitate future martial arts movement recognition. An HMM is a time series-related probability model. To recognize Wushu movements, the converted one-dimensional vector is then fed into an HMM for recognition and classification. Each martial arts movement may be modeled using HMM since martial arts movements have time. Each martial arts movement is broken down into N

meta actions, each of which is performed sequentially. To discover the optimal values for the HMM parameters of each martial arts movement model, we treat each technique as an observation sequence of length N and utilize the Baum-Welch method for training and learning. Wushu movement recognition involves observation sequence data extracted to input into HMM before Viterbi algorithm to get the probability of optimal sequence of such moves made by each Wushu model. The martial arts movements corresponding to the model with the highest output probability have been recognized as a result of the most current observation sequence. HMM can be expressed by formula (5).

$$\lambda = (A, B, \pi) \quad (5)$$

In formula (5), A is the state probability distribution; B is the observation probability distribution; π is the initial probability distribution. The forward algorithm and backward algorithm are used to calculate the probability of observation sequence in HMM; Baum-Welch algorithm is used to train various parameters of HMM to maximize the probability of observation sequence in the model. Baum-Welch algorithm regards the state sequence data as unobservable hidden data, so here is formula (6).

$$P(O|\lambda) = \sum_I P(O|I, \lambda) P(I|\lambda) \quad (6)$$

In formula (6), O is the observation sequence data. The parameter learning of HMM can be realized by maximum expectation algorithm (EM). At the E step, the Q function is solved, such as formula (7).

$$Q(\lambda, \bar{\lambda}) = \sum_I \log P(O, I|\lambda) P(O, I|\bar{\lambda}) \quad (7)$$

In formula (7), $\bar{\lambda}$ represents the current estimated value of model parameters; λ is the maximized model parameter. At the M step, EM maximizes the Q function and combines the Lagrange multiplier method to obtain the A , B , and π parameters of HMM. There may be several data sets that correspond to the same martial arts movements recognized, but the HMM parameter model is trained and learned using just one data set. Therefore, it is easier to optimize local data during the training and learning process, resulting in a low recognition rate of martial arts movements in the model. As a result, the Baum-Welch algorithm is improved using the mean training technique and the frequency weighted training method. In a martial arts movement model, the frequency weighted training technique linearly weights the occurrence frequency of the sample observation sequence. The mean value training rule fuses the data samples of the input model to generate the mean value of numerous data samples in the same Wushu movement model, and then trains the Wushu movement model. The Viterbi algorithm is used to find the best HMM solution after the Wushu movement model has been trained. For a given HMM model and observation sequence data, the optimal path $I^* = (i_1^*, i_2^*, \dots, i_T^*)$ is solved, and T represents the length of sequence I .

On the basis of the preceding information, a Wushu movement recognition model may be developed. The training sample data consists of 100 actions representing eight different martial arts movements, such as the flat horse, lunge, punch, split, palm push, palm split, kick, and spring. The observation sequences for each martial arts movement are generated by processing data. The Viterbi algorithm is then used to determine the probability of the optimal state sequence of the observation sequence in these martial arts movement models, allowing the observation sequence and model to be aligned and the recognition results to be output. The flow of Wushu movement recognition model is shown in Fig. 2.

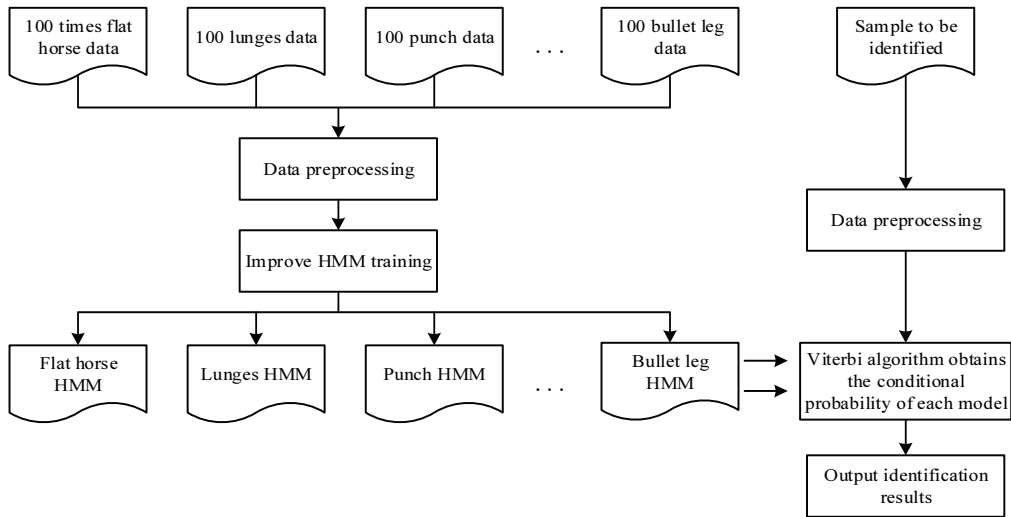


Fig. 2. Process of Wushu movement recognition model.

4. Performance Analysis of Wushu Movement Recognition Model

To ensure recognition accuracy, the wavelet transform denoising approach is used to preprocess the martial arts movements data, which is verified by the x-axis of Wushu punch. Three-layer wavelet denoising and 2-layer wavelet transform denoising are employed, respectively, to test the denoising impact of 3-layer wavelet transform using the sample data mentioned above. SNR and MSE values before and after noise reduction are shown in Fig. 3.

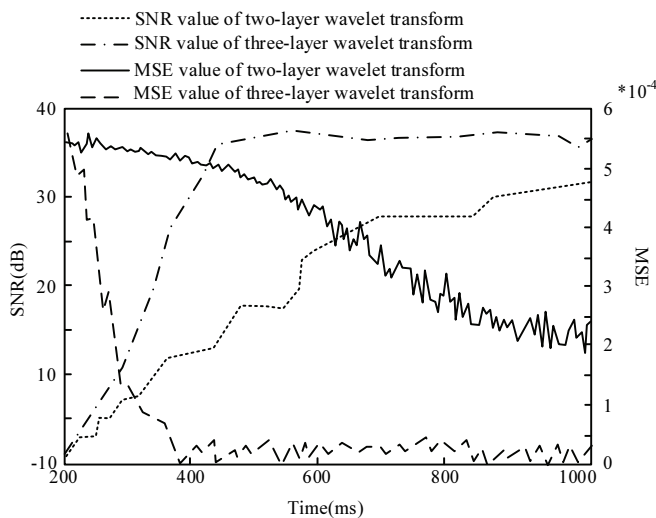


Fig. 3. SNR and MSE values before and after noise reduction.

As demonstrated in Fig. 3, the SNR value of the x-axis acceleration data of the Wushu movement punching is much lower than the data following the noise reduction by the 3-layer wavelet transform

following the noise reduction by the 2-layer wavelet transform. The 3-layer wavelet transform reduces noise, but this makes MSE much bigger than the data. It can be demonstrated that the wavelet transform denoising approach can successfully denoise and filter the martial arts movement data, making it easier to extract and recognize the moves, later on. Additionally, the 3-layer wavelet transform performs denoising on the martial arts movement data better than the 2-layer wavelet transform. The Baum-Welch single training algorithm (algorithm 1), the Baum-Welch average training algorithm (algorithm 2), and the Baum-Welch frequency weighted training algorithm (algorithm 3) are each developed in order to test the effectiveness of the two HMM training algorithm optimization approaches proposed in the study on Baum-Welch algorithm. The above three algorithms are used to train the HMM model of martial arts movement, and the training effect is shown in Fig. 4.

The convergence of the Baum-Welch algorithm with frequency-weighted training is evidently superior to that of the other two methods. To validate the recognition influence on Wushu movement, the proposed Wushu movement recognition model is compared to commonly used image recognition algorithms. CNN, BP neural network, and support vector machine are examples of common image recognition methods. There are four different types of martial arts movement recognition models: HMM, SVM, CNN, and BP neural network. As shown in Fig. 5, the aforementioned four martial arts movements recognition models are trained and evaluated using sample data from the martial arts movements images in the Baidu gallery and the martial arts videos on the Youku website.

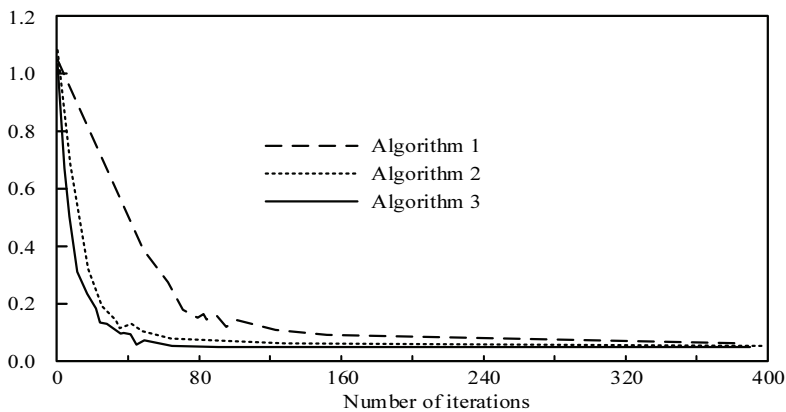


Fig. 4. Training effect of three algorithms on HMM model of Wushu movement.

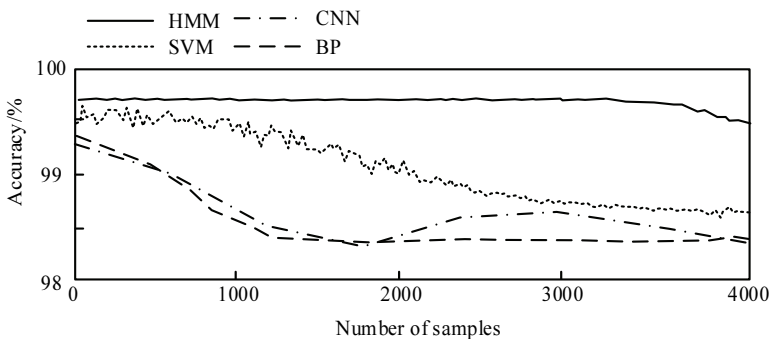


Fig. 5. Recognition accuracy of four Wushu movement recognition models.

As shown in Fig. 5, the recognition rates of the HMM-, SVM-, CNN-, and BP-based model all declined marginally as the number of samples increased. The HMM-based has a recognition accuracy of 99.75% for 1,000 samples, while the SVM-based model has a recognition accuracy of 99.50%, which is 0.25%p lower than the HMM-based model. The CNN-based model has a recognition accuracy of 98.63%, lower than that of the HMM-based model by 1.42%p, while the BP-based model's recognition accuracy is 98.55%, 1.50%p lower than that of HMM-based model. When there are 4,000 samples, the HMM-based model has a recognition accuracy of 99.60%, whereas the SVM-based model has a recognition accuracy of 98.66%, which is 0.94%p lower than the HMM-based model. The recognition accuracy of CNN-based model is 98.48%, 1.12%p lower than the HMM-based model. BP-based model's recognition accuracy is 1.14%p less than that of HMM Wushu-based model, reaching 98.46%. In conclusion, the HMM-based model for identifying martial arts movements is highly accurate and useful.

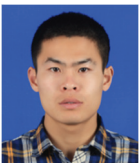
5. Conclusion

The rapid development of artificial intelligence technology promotes the advancement of visual image technology, which plays a crucial role in human movement and facial recognition. This research presents a method for the recognition of martial arts movements based on visual image technology in order to enhance the inheritance and growth of martial arts in China. The data is first preprocessed using the three-layer wavelet transform. The experimental results suggest that the 3-layer wavelet transform may successfully eliminate noise from martial arts movement data. The 3-layer wavelet transform's SNR is 17.5 dB higher and its MSE is 4.48 dB lower than that of the 2-layer wavelet transform. The Viterbi method is used to determine the HMM's best solution, the HMM is used to generate the martial arts movements model, the modified Baum-Welch algorithm is used to train the martial arts movements HMM, and the recognition results are produced. According to the results, the HMM Wushu movement recognition model has a recognition accuracy of 99.75% when there are 1,000 samples in total, which 0.25%p higher than the SVM Wushu movement recognition model, 98.63% higher than the CNN Wushu movement recognition model, and 1.50%p higher than the BP Wushu movement recognition model. The aforementioned findings demonstrate that the martial arts movements recognition model based on HMM has a high identification accuracy and can detect martial arts movements accurately and efficiently, hence fostering the inheritance and growth of martial arts. The absence of prepared sample data for the study increases the possibility that the experimental results are flawed and need to be refined.

References

- [1] J. Liang, F. Xu, and S. Yu, "A multi-scale semantic attention representation for multi-label image recognition with graph networks," *Neurocomputing*, vol. 491, pp. 14-23, 2022.
- [2] Z. Wu, H. Li, X. Wang, Z. Wu, L. Zou, L. Xu, and M. Tan, "New benchmark for household garbage image recognition," *Tsinghua Science and Technology*, vol. 27, no. 5, pp. 793-803, 2022.
- [3] W. Tang and H. Chen, "Research on intelligent substation monitoring by image recognition method," *International Journal of Emerging Electric Power Systems*, vol. 22, no. 1, pp. 1-7, 2021.

- [4] A. V. Savchenko, "Sequential three-way decisions in multi-category image recognition with deep features based on distance factor," *Information Sciences*, vol. 489, pp. 18-36, 2019.
- [5] S. Umer, B. C. Dhara, and B. Chanda, "NIR and VW iris image recognition using ensemble of patch statistics features," *The Visual Computer*, vol. 35, no. 9, pp. 1327-1344, 2019.
- [6] W. J. Hu, T. Y. Xie, B. S. Li, Y. X. Du, and N. N. Xiong, "An edge intelligence-based generative data augmentation system for IoT image recognition tasks," *Journal of Internet Technology*, vol. 22, no. 4, pp. 765-778, 2021.
- [7] H. Ling, W. Zhang, Y. Tao, and M. Zhou, "Research on network layer recursive reduction model compression for image recognition," *Scientific Programming*, vol. 2021, article no. 4054435, 2021. <https://doi.org/10.1155/2021/4054435>
- [8] L. Jin, H. Liang, and C. Yang, "Sonar image recognition of underwater target based on convolutional neural network," *Journal of Northwestern Polytechnical University*, vol. 39, no. 2, pp. 285-291, 2021.
- [9] G. Huang, L. Qiao, S. Khanna, P. A. Pavlovich, and S. Tiwari, "Research on fan vibration fault diagnosis based on image recognition," *Journal of Vibroengineering*, vol. 23, no. 6, pp. 1366-1382, 2021.
- [10] A. Jafar and M. Lee, "High-speed hyperparameter optimization for deep ResNet models in image recognition," *Cluster Computing*, 2021. <https://doi.org/10.1007/s10586-021-03284-6>
- [11] X. Yang, Y. Zhang, W. Lv, and D. Wang, "Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier," *Renewable Energy*, vol. 163, pp. 386-397, 2021.
- [12] S. Tang and M. Shabaz, "A new face image recognition algorithm based on cerebellum-basal ganglia mechanism," *Journal of Healthcare Engineering*, vol. 2021, article no. 3688881, 2021. <https://doi.org/10.1155/2021/3688881>
- [13] Z. Tan and X. Liu, "ConvNet combined with minimum weighted random search algorithm for improving the domain shift problem of image recognition model," *Applied Intelligence*, vol. 52, no. 6, pp. 6889-6904, 2022.
- [14] N. A. Andriyanov, V. E. Dementiev, and Y. D. Kargashin, "Analysis of the impact of visual attacks on the characteristics of neural networks in image recognition," *Procedia Computer Science*, vol. 186, pp. 495-502, 2021.
- [15] F. Meng, B. Xu, T. Zhang, B. Muthu, and C. B. Sivaparthipan, "Application of AI in image recognition technology for power line inspection," *Energy Systems*, 2021. <https://doi.org/10.1007/s12667-020-00414-8>
- [16] A. Jeya Christy and K. Dhanalakshmi, "Content-based image recognition and tagging by deep learning methods," *Wireless Personal Communications*, vol. 123, pp. 813-838, 2022.
- [17] F. Wang, R. Hu, and Y. Jin, "Research on gesture image recognition method based on transfer learning," *Procedia Computer Science*, vol. 187, pp. 140-145, 2021.



Husheng Zhou <https://orcid.org/0000-0001-5865-892X>

He received B.S. and M.S. degrees. Now, he is a lecture from School of Physical Education in Huaibei Normal University. His current research interests include theory and practice of national traditional sports.