International Journal of Web Applications

DLINE JOURNALS

Print ISSN: 0974-7710
Online ISSN: 0974-7729

IJWA 2025: 17 (4)
https://doi.org/10.6025/ijwa/2025/17/4/163-170

Preserving Conservation Cultural Materials with AI: A Human Pose Estimation Approach

Zhijiang Lan
Transportation Management Department of Liuzhou Railway
Vocational Technical College, Liuzhou, Guangxi. China
fengyi@ltzy.edu.cn

ABSTRACT

This paper explores how deep learning can be leveraged to preserve and promote traditional Chinese martial arts, which are a vital part of China's intangible cultural heritage. Historically transmitted through master apprentice relationships, these arts now face challenges due to societal modernization and demographic shifts. The author proposes using deep learning models particularly convolutional neural networks (C-NNs) to capture, analyze, and reconstruct martial arts movements through human pose estimation. A novel model, IPN (Involution Pose Estimation Net), built on Simple Baselines and Involution mechanisms, is introduced to identify key body joints from video data with high accuracy. The study utilizes datasets like NTU-RGB+D and UTD-MHAD, though it acknowledges their limitations for martial arts specific actions, highlighting the need for a dedicated Chinese martial arts motion database. Evaluation metrics such as tMPJPE (time aware Mean Per Joint Position Error) are adapted to assess pose accuracy over time. Experimental results demonstrate effective movement recognition in "Youth Serial Boxing," with promising convergence and accuracy. The research underscores deep learning's potential not only in digitizing martial arts but also in enabling broader cultural preservation in the intelligent era. Future work includes expanding the model to other martial arts styles and improving dataset specificity to enhance training and generalization.

Keywords: Traditional Chinese Martial Arts, Deep Learning, Human Pose Estimation, Convolutional Neural Networks (CNNs), Cultural Heritage Preservation, IPN (InvolutionPose Estimation Net), MPJPE Evaluation Metric, Action Recognition

Received: 28 June 2025, Revised 27 September 2025, Accepted 19 October 2025

1. Introduction

Conservative cultural arts boast a rich history and cultural significance, embodying the essence of the nation.

They have consistently been viewed as a treasured asset and have played a crucial role throughout time. Their roots can be traced back to ancient agrarian societies and have gained further prominence in response to external challenges in modern times. [1] As modernization has progressed, traditional martial arts have transitioned from mere physical activity and recreation to a significant cultural pursuit. This shift suggests that, alongside the evolution of Chinese history, the social value and role of these arts have undergone considerable change. [2] Consequently, it is essential to align with contemporary realities, investigate and reflect on their historical foundations, and provide robust support for their future growth. In light of current circumstances, traditional Chinese martial arts are facing unprecedented obstacles. [3] Thus, to adapt to the advancements of modern society, these martial arts must acknowledge future developmental trends and strive to explore various viable models for growth, contributing to the flourishing of Chinese culture. To ensure a new skill is preserved for future generations, it is crucial to consider the environment in which it is practised thoughtfully. Traditional Chinese martial arts are an age old sporting endeavor, a precious heritage of the Chinese people, and a vital intangible cultural asset. Therefore, we must carefully contemplate the conditions surrounding it to facilitate its sustainable growth on the global stage. The ancient practices of Chinese martial arts carry a longstanding legacy, serving not only to strengthen the body but also to enrich the mind. [4] They symbolize not only time honored techniques but also a profound cultural legacy. [5] This paper aims to investigate how deep learning can be employed to safeguard traditional Chinese martial arts. [6] We will collect, integrate, and analyze dynamic data from various martial arts institutions and utilize this information to develop a model that can better identify and preserve the essence of these disciplines. [7]

2. Earlier Work

In the context of the Web 6.0 era, there has been a significant leap forward in information technology, with closed loop systems exemplified by the gradual expansion of artificial intelligence and the complete integration of the Internet with the Internet of Things. Merging traditional martial arts with artificial intelligence through data mining and organization can enhance and render the dissemination of content more genuine. The foundational structure of the AI + traditional martial arts communication ecosystem is established using artificial intelligence techniques such as expert systems, deep learning, intelligent design, and human computer interaction. [8] For instance, the LeNet-5 model, now essential for many beginners in AI, is utilized to accomplish the MNIST data recognition task. Despite substantial advancements in neural network development, they continue to lack the attention they deserve, potentially due to the disconnect between technological progress and theoretical application, along with the absence of robust mathematical theory support, leading to a lack of acknowledgment of their developmental successes. Since AlexNet triumphed over the second-place SVM in the ImageNet image classification competition in 2012, the significance of deep learning has increasingly come to the forefront. [9, 10]. In order to attain effective deep learning, numerous established architectures such as Caffe, TensorFlow, Pytorch, Keras, and MXNet have undergone continuous enhancements and refinements, becoming widely utilized in computer programming technologies to address evolving requirements. With the swift advancement of Python technology, it has emerged as a standard in numerous domains, particularly in gaming, e-commerce, and entertainment. [11, 12] Its robust parallel processing abilities render it the favoured option for many organisations, and deep learning applications, including cost-efficient training and extensive clusters, are gaining substantial attention. The emergence of ResNet and DenseNet enables developers to use them as reliable foundations for constructing more intricate network architectures. [13, 14] This facilitates the creation of new network designs and greatly aids in the development of numerous innovative applications.

3. The Evolution of Conservative Cultural Arts Tracing

3.1 How Neural Networks assist ?

As technology progresses, the latest generation of artificial intelligence has been implemented across various sectors and is widely acknowledged as a significant research achievement. Through investigation and study, it has been discovered that by leveraging the interactions and mechanisms of animal neurons, it is possible to develop a new generation of artificial intelligence technology networks that boast high efficiency and precision. [15, 16] Owing to technological advancements, artificial intelligence has evolved into a critical tool for helping individuals transform data stored in human cognition, thereby enhancing recognition, analysis, prediction, and decision making capabilities. In its initial phases, the new generation of artificial intelligence technology employed a three tier framework, consisting of an input layer, hidden layer, and output layer, to convert data stored in human thought into recognizable, analyzable, and predictable formats. [17] In this section, we will examine a novel method for disseminating trends in martial arts. This approach is grounded in 12 convolutional layers: the first and second layers each contain 32 filters, and the remaining layers contain 64 filters. All layers share identical kernel sizes and strides, and are subjected to continuous batch processing for normalization. Using these twelve convolutional layers, we combine two outputs and channel them into the subsequent network model. By fine tuning the two convolutional layers, we effectively balance the model's complexity and accuracy to achieve the final reconstruction. Specifically, each layer functions independently, as illustrated in Formula 1.

$$C_i(X) = W_i * X + B_i \qquad (1)$$

Here, X represents the input, i signifies the ith layer, where i ∈{1,2,...11}; Ci(X) indicates the convolution operation, while W and B denote the filters and biases, respectively; $Bi(X)$ represents batch normalization, with the parameters to be learned being γ and β and $max(0,)$ represents the ReLU activation function. Initially, this research contemplated utilizing MPJPE (Mean Per Joint Position Error) as the evaluation metric for the experiments. This metric is frequently employed in 3D pose estimation tasks, where the objective is to ascertain the 3D pose (x, y, z coordinates) of each joint from RGBD images. MPJPE computes the Euclidean distance between the actual values and the predictions, then averages the positional errors across all joints. The new task proposed in this study is also related to 3D pose estimation and requires measuring the error between the generated results and the ground truth. Therefore, evaluation metrics from 3D pose estimation tasks can be explored. However, considering the temporal dimension of the generated results in this task, a time dimension is added to the calculation based on MPJPE, as shown in formula 2:

$$tMPJPE = \frac{d(w, v)}{k \times t} \qquad (2)$$

where w is the output, v is the target, J is the sample dimension, $d(w,v)$ represents the Euclidean distance, k is the number of key points, and t is the time parameter. The *tMPJPE* metric in this section is the MPJPE per unit time, facilitating comparisons over the same time interval. It should be noted that the evaluation metric proposed in this section focuses on joint position accuracy and applies only to short time intervals. The gap between samples and the generated results can be accurately evaluated using mean squared error to determine whether the experimental results meet expectations.

3.2 Convolutional Neural Networks

LeCun and other sub investigators made a breakthrough by applying convolutional neural networks to more complex scenarios [18]. By adjusting the network structure and parameters, they achieved globalized neural network training from input to output, significantly improving the effectiveness of deep learning. By dividing complex hierarchical information, such as convolutional neural networks, pooling layers, or fully connected layers, into multiple layers, the differences between these layers can be reflected in comparisons. Although these differences can be represented through contrast for example, in the case of perceptrons, their hidden layers are defined by comparison this is not the case for convolutional neural networks, whose results are not entirely the same. Through convolutional operations, information can be collected and processed across multiple receptive fields to yield more valuable insights. This operation may need to be repeated numerous times and may require various convolution kernels to better capture and process the diversity of information. The parameters and values can be calculated using mathematical methods by processing the receptive field of the convolution kernel. By inputting the extracted short video data into the human pose recognition network, the body key points of the martial arts instructor can be accurately recognised and recorded. Furthermore, we need to extract coordinate information from the original images to estimate human pose better. To ensure the reliability of the data, the obtained human skeleton data needs to be processed. Firstly, all key point data is reduced to 1280*720 centered around the head joint, and the coordinate points are standardized and normalized.

4. Experimental Architecture and Results

4.1 Experimental Architecture

In the history of computer vision, especially in image processing, evaluating a neural network's performance often requires considering its feature extraction capabilities. Therefore, building a model with good feature extraction performance and effectively utilizing this information is crucial. The development of deep learning technology has become an essential tool that helps us better predict and recognize complex physical phenomena. For example, DeepPose, OpenPose, Hourglass, and HRNet are typical deep learning models that predict physical phenomena. Researchers are actively exploring new techniques, such as expanding model boundaries, combining models, and leveraging model characteristics to create more complex human pose models. However, the success of these techniques depends on their effective application in practical scenarios to achieve optimal model accuracy.

After years of in depth exploration, we ultimately decided to use Simple Baselines as the base network and adopt it as the new feature extraction unit, replacing ResNet, to better evaluate recent human pose estimation algorithms. Building on the Involution foundation, we constructed an IPN (Involution Pose Estimation Net) for

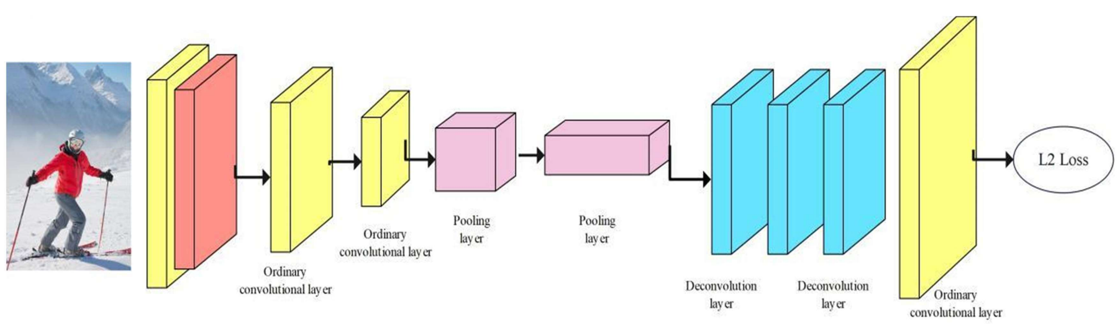

Figure 1. IPN Network Architecture

human identity recognition, as shown in Figure 1. *In Net* is a method based on convolutional neural networks that extracts 17 key information. In this method, we first expand the convolutional layers, then process them, then process three deconvolution layers, and finally apply L2 Loss. By using *In Net* as the base grid, the parameter complexity can be greatly reduced, and the advantages of Involution can significantly improve accuracy.

4.2 Performance Analysis and Results

To improve the identification of martial arts techniques in "少年连环拳" (Youth Serial Boxing), it is essential to utilize two of the most prevalent and respected datasets, specifically NTU-RGB+D and UTD-MHAD. At present, these datasets are not being fully leveraged, so we must extract information from them using automated methods. The NTU-RGB+D dataset is a large scale, publicly accessible dataset encompassing over 60 action categories and more than 56,880 samples. Within this collection, 40 are everyday activities, such as drinking water and sitting; 9 are health related actions, such as dancing and waving; and 11 are interactive actions, such as handshaking and high fiving. A group of 40 individuals from various age groups participated in the experiments, with their behavioral data recorded using Kinect v2 sensors. During this experiment, participants were instructed to perform two actions, captured by three cameras at angles of 45 degrees, 0 degrees, and 45 degrees. The footage collected from these cameras can be divided into four categories: depth, 3D skeletal data, RGB frames, and infrared sequences. The RGB frames have a resolution of up to 1920 1080, while the 3D skeletal data includes 25 body joints, all visible to participants. This system possesses extensive observational capabilities, documenting a range of dynamic alterations. While these two datasets encompass interactive actions found in daily life, the movements in martial arts are often distinct. Moreover, most of these datasets originate from foreign countries, indicating a need for further investigation. Even though a specialized action database for a specific martial art does not currently exist in China, establishing a dedicated martial arts action database remains essential.

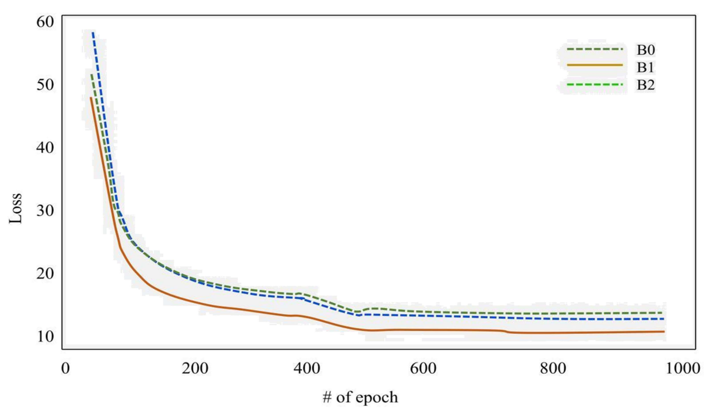

Figure 2. Model B1 Data

Based on the data in Figure 2, Model B1 shows excellent training performance, with a much higher convergence rate than B2, but it remains relatively low. However, B1's performance is still outstanding, especially in the 1s and 3s experiments, where it performs even better. B2 relies on residual blocks, while B0 uses only one channel. We found that this model demonstrates remarkable efficiency through systematic training and experiments. After 200 epochs of training, some network parameters have changed, while others remain unchanged; the final training results are shown in Figure 3. According to Figure 3, through training, we have transformed the received information into recognizable martial arts actions. We used the videos from multiple schools, and each performance had different martial arts movements. By observing each performance, we could accurately identify which actions the audience saw and the positions of the audience members. Through in depth research, we found that the proposed martial arts action recognition method, which considers both human body postures and corresponding key information, achieves high accuracy and holds significant academic value.

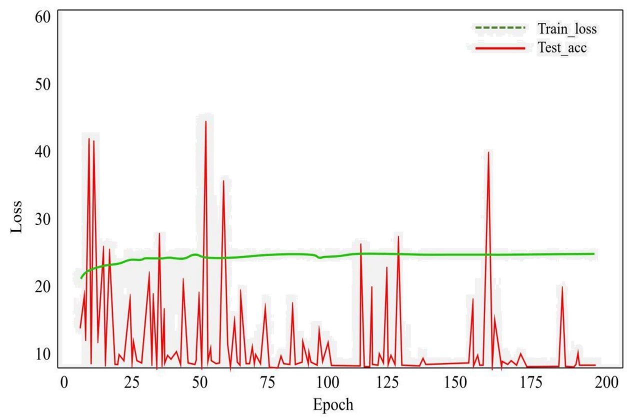

Figure 3. Training Results 3

5. Conclusion

Overall, deep learning has introduced fresh perspectives and techniques for preserving and advancing traditional Chinese martial arts. Conventional martial arts transmission typically depends on the interaction and instruction between masters and apprentices, yet this method has become increasingly complex due to societal shifts and demographic mobility. With the rise of the intelligent era, deep learning has emerged as a pivotal force. It plays a crucial role in artificial intelligence and expedites AI development. Deep learning enables machines to replicate human analysis, particularly in speech and image processing. Numerous research findings grounded in deep learning have already been implemented in various facets of everyday life, including semantic analysis and facial recognition. Its effectiveness in digital recognition is noteworthy as well. The advent of deep learning models presents new opportunities for preserving martial arts. Through our research, we have found that the proposed method for recognizing martial arts actions, which takes into account both human body postures and relevant key information, achieves high accuracy and possesses considerable academic significance. In the future, we can extend the application of deep learning models across a broader range of martial arts disciplines to sustain traditional Chinese martial arts.

References

[1] Chude, C. I., Schaaf, S. (2021). A Unique Cause of Sciatica in a Masters Martial Arts Athlete: 1264. *Medicine Science in Sports Exercise,* 53 (8S), 416-416.

[2] Selitrenikova, T., Ageev, E., Kolokoltsev, M., et al. (2022). Òranscranial electrical stimulation to increase psychophysiological stability, technical and tactical readiness of MMA fighters. *Journal of Physical Education and Sport,* 22 (6), 1419-1425.

[3] Qiuntero, Monterrosa A., Rosa, DeLa, A., Chagnaud, Arc C., et al. (2023). Morphology, lower limbs performance and baropodometric characteristics of elite Brazilian Ji jitsu athletes. *Ido Movement for Culture. Journal of Martial Arts Anthropology,* 23 (2), 58-69.

[4] Starke, S., Zhao, Y., Zinno, F., et al. (2021). Neural animation layering for synthesizing martial arts movements. *ACM Transactions on Graphics,* 40 (4), 1-16.

[5] Liu, M., Zhang, J. (2022). Gesture estimation for 3D martial arts based on neural network. *Displays: Technology and Applications,* (72) 72.

[6] Envelope, V. L. A. A., Envelope, Y. F. F. B. P., Envelope, P. B. A. C., et al. (2022). Validity and reliability of a specific anaerobic test for mixed martial arts. *Science Sports* 37 (5–6).

[7] Zhao, C., Yang, H., Li, X., et al. (2021). Analysis and application of martial arts video image based on fuzzy clustering algorithm. *Journal of Intelligent Fuzzy Systems Applications in Engineering and Technology,* (4), 40.

[8] Peng, X. B., Abbeel, P., Levine, S., et al. (2018). Deep Mimic: Example Guided Deep Reinforcement Learning of Physics *Based Character Skills*. *ACM Transactions on Graphics,* 37 (4CD), 143.1-143.

[9] Theeboom, M., Dong, Z., Vertonghen, J. (2012). Traditional Asian martial arts and youth: Experiences of young Chinese wushu athletes. *Archives of Budo,* 8 (1), 27-35.

[10] Mawardi, S., Firdaus, F. (2019). The Character Values in Minangkabau Traditional Martial Arts. *International Journal of Scientific Technology Research,* 8 (10), 846-850.

[11] Jain, A., Awan, A. A., Subramoni, H., et al. (2019). Scaling tensorflow, pytorch, and mxnet using mvapich2 for high performance deep learning on frontera, 2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS). *IEEE,* 76-83.

[12] Tabani, H., Pujol, R., Abella, J., et al. (2020). A cross layer review of deep learning frameworks to ease their optimization and reuse, *2020 IEEE 23rd International Symposium on Real Time Distributed Computing (ISORC). IEEE,* 144-145.

[13] Zhang, J., Zhang, Y., Lu, Y. (2021). A Comparative Study of Deep Learning Frameworks Based on Short-term Power Load Forecasting Experiments, *Journal of Physics: Conference Series. IOP Publishing,* 2005 (1), 012070.

[14] Haixia, Y., Chuandong, X., Wenbo, F., et al. (2022). Intelligent mask recognition and voice dialogue system based on Raspberry Pi control, 2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA). *IEEE,* 103-105.

[15] Guaman, Barba., L., Naranjo, Eugenio., J, Ortiz, A. (2020). Deep learning framework for vehicle and pedestrian detection in rural roads on an embedded GPU. *Electronics,* 9 (4), 589.

[16] Demidovskij, A., Tugaryov, A., Suvorov, A., et al. (2020). Openvino deep learning workbench: A platform for model optimization, analysis and deployment, 2020 IEEE 32nd international conference on tools with artificial intelligence (ICTAI). *IEEE,* 661-668.

[17] Long, G., Chen, T. (2022). On reporting performance and accuracy bugs for deep learning frameworks: An

exploratory study from github, *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering*. 90-99.

[18] LeCun, Y., Bottou, L., Bengio, Y., et al. (1998). Gradient based learning applied to document recognition. *Proceedings of the IEEE,* 86 (11), 2278-2324.