ISSN: 3033-3512

Journal of Traditional Medicine and Chinese Medicine

SCIENTIFIC Research and Community

Research Article

Open Access

# A Heatmap Regression-Based HRNet Method for Hand Acupoint Recognition

**XU Mingluyu[1], HE Liuyi[1], ZHANG Yujie[1], LIN Wei[1], JIANG Tao[1], ZHANG Linshuai[1], ZHU Deliang[2] and XU Lin[1*]**

[1]School of Intelligent Medicine, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, P.R. China

[2]Department of Rehabilitation, Chengdu Fifth People's Hospital, Chengdu, 611130, PR. China

## ABSTRACT

**Objective:** This study investigates hand acupoint recognition using deep learning, leveraging image processing and deep learning technologies to automatically identify acupoint categories and locations on the hand. This approach aims to enhance the accuracy of acupoint recognition, providing a reference for young traditional Chinese medicine practitioners and advancing the modernization of acupuncture diagnosis and treatment.

**Methods:** This study developed an HRNet-based hand acupoint recognition method using heatmap regression, applied to the recognition task of six common dorsal hand acupoints, including Shaozhe (SI1) and Zhongzhu (SJ3). Standardization, normalization, and data augmentation techniques were employed during the image preprocessing stage. Model training employed either heatmap regression or coordinate regression strategies. The proposed method was compared with four deep learning models—Simple Baseline, ResNet50, and VGG16—on a self-created dataset of 1,102 hand acupoint images, and model performance was evaluated using multiple metrics including PCK, IoU, OKS, and precision.

**Results:** Experimental results show that the HRNet model based on heatmap regression performs best across all metrics (PCK: 95.37%, IoU: 80.04%, precision: 94.89%, OKS: 90.07%), outperforming Simple Baseline (PCK: 94.12%, etc.); among coordinate regression models, VGG16 outperforms ResNet50. The overall recognition accuracy of the heatmap regression model was significantly higher than that of the coordinate regression model.

**Conclusion:** This study validated the feasibility and effectiveness of deep learning models in hand acupoint recognition, particularly the HRNet model based on heatmap regression, which demonstrated significant advantages in terms of accuracy and stability. This method not only provides precise acupoint localization assistance for traditional Chinese medicine practitioners but also offers critical data support for precise diagnosis and treatment in intelligent medical devices such as acupuncture robots.

***Corresponding author**

XU Lin, School of Intelligent Medicine, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, P.R. China.

**Received:** August 06, 2025; **Accepted:** August 11, 2025; **Published:** August 25, 2025

**Keywords:** Deep Learning, Key Point Detection, Acupoint Recognition

## Introduction

Acupoint therapy, which stimulates acupoints using methods such as acupuncture and massage, is an essential component of traditional Chinese medicine (TCM), which can balance the flow of Qi and blood in the human body. Acupoint therapy has achieved promising effects in treating chronic diseases and alleviating pain. The hands, a key nerve-draining region in the human body, are not only home to numerous meridian-intersecting acupoints but are also commonly used for clinical treatment and healthcare. Acupoints such as Shaoze, Houxi, and Yanggu are commonly used to treat headaches, neck and shoulder pain, and gastrointestinal discomfort. With the international spread of acupuncture and the development of digital TCM, the precise identification and quantitative location of acupoints has become a key component in advancing the intelligentization of TCM [1-9].

In the early stages of acupoint recognition research, researchers often combined image processing with geometric models, extracting acupoint regions through traditional image processing techniques such as color segmentation and hand model fitting. Zhao et al. used corner detection and edge detection to locate multiple acupoints in facial thermal imaging and applied this method to the treatment of facial paralysis. Other studies have also applied edge detection, model building, and machine learning to localize acupoints in thermal imaging. However, deep learning methods in this field are relatively rare, leaving significant room for development. Wang et al. proposed a palm acupoint localization method based on a lightweight and efficient channel attention (LECA) network. This method uses MobileNetV2 as the backbone network and integrates an efficient channel attention (ECA) module to optimize feature extraction. It also employs the Huber loss function to address the varying difficulty of detecting different acupoints. However, this method is limited by the small number of acupoints it can locate. Some studies have also attempted to construct template matching-based localization algorithms,

estimating acupoint locations using preset point-to-line ratios or spatial structures. However, these methods are sensitive to image quality and pose variations, struggle with complex backgrounds and diverse hand shapes, and suffer from limited recognition accuracy and robustness. In addition, due to the close distribution and slight differences of acupuncture points on the hands, these traditional methods have limited promotion value in practical applications [10-18].

In recent years, with the rapid development of deep learning in medical image processing, image processing technology has been able to identify key points on the human body through feature extraction, thereby achieving precise acupoint location. Su et al. used Google's machine learning suite for facial detection, achieving real-time extraction of facial features and contours on mobile devices, providing technical support for facial acupoint recognition. proposed a facial acupoint localization model that significantly improved accuracy while maintaining recognition speed using the Faster PFLD-L network. Li et al. used the Open Pose library to implement hand acupoint recognition and applied it to massage assistance scenarios, offering advantages such as low hardware cost, strong real-time performance, and ease of integration. Masood proposed an indirect three-dimensional hand acupoint localization method for intelligent moxibustion robots to achieve more accurate positioning. This method combines a VGGNet-based RGB-CNN model with a deep CNN model. By continuously optimizing fusion parameters and learning to merge features, a multimodal fusion acupoint detection model was developed that directly outputs three-dimensional acupoint coordinates. Experimental results show that this method can locate five groups of acupoints with an average localization error of less than 0.09, demonstrating its potential application value [19-23].

Existing deep learning-based key point detection methods can be divided into two categories: coordinate regression-based and heatmap regression-based. The former directly extracts features from the input image through a neural network and then regresses to predict the two-dimensional coordinates of key points. This method features simple modeling and high computational efficiency, and is commonly used for tasks such as face and gesture recognition. It primarily includes network structures such as ResNet50, VGG16, and Deep Pose. The latter, on the other hand, outputs a probability heatmap corresponding to each key point through the network. The peak position in the heatmap represents the key point location, thereby achieving explicit modeling of spatial information and offering greater positioning accuracy and robustness. Key methods include SimpleBaseline, HRNet, and Hourglass Network. Among them, ResNet50 and VGG16, which are based on coordinate regression, are widely used in image classification and target localization tasks due to their classic structure. They can directly output acupoint coordinates, but are easily interfered with when faced with complex backgrounds and small targets. The Simple Baseline and HRNet models, based on heatmap regression, are more suitable for refined key point localization tasks. The former restores spatial resolution through a deconvolution module, while the latter achieves multi-scale fusion while maintaining high-resolution features, demonstrating superior performance in tasks such as human pose estimation. Although some studies have begun to explore the use of heatmap regression methods to improve the model's ability to recognize fine-grained acupoints under small sample conditions in recent years, due to the scarcity of public hand acupoint datasets, related research is still in its infancy and exploratory stage, and further technological breakthroughs and data resource support are still needed. To this end, this study focused on the automatic recognition of six common acupoints on the dorsal side of the hand (Shaoze, Zhongzhu, Shangyang, Wailaogong, Yaotongdian, and Hegu).

A labeled dataset of 1,102 images was constructed. The HRNet method based on heatmap regression was used for acupoint recognition. Model performance was measured using metrics such as PCK, IoU, OKS, and precision. The model was compared with three current mainstream deep learning models: Simple Baseline, ResNet50, and VGG16. The advantages and disadvantages of each model were systematically analyzed for small-scale, high-density acupoint recognition. The purpose of this study was to explore the optimal deep learning architecture for automatic recognition of TCM hand acupoints and to provide high-precision positioning technology support for TCM intelligent devices [24-34].

## Materials and Methods

### General Data

Acupoints refer to specific locations along the body's meridians. Traditional Chinese Medicine practitioners use acupuncture and moxibustion to stimulate corresponding meridian points to treat illnesses. Some acupoints aren't located on meridians, but stimulation of them can still be therapeutic. Acupoints on the hands are typically located along specific meridians, which connect various organs and tissues and are closely related to physical health. Meridians are energy pathways that run throughout the body and are composed of 12 major meridians and eight extraordinary meridians. Each meridian has distinct acupoints corresponding to different organ functions. The three meridians on the back of the hand are the Small Intestine Channel of Hand-Taiyang (SI), the Triple Energizer Meridian of Hand-Shaoyang (SJ), and the Yangming Large Intestine Channel of Hand (LI); the three meridians on the palm are the Lung Meridian of Hand-Taiyin (LU), the Pericardium Meridian of Hand-Jueyin (PC), and the Heart Meridian of Hand-Shaoyin (HT) [35]. The distribution of acupuncture points on the hand meridians is shown in Figure 1.

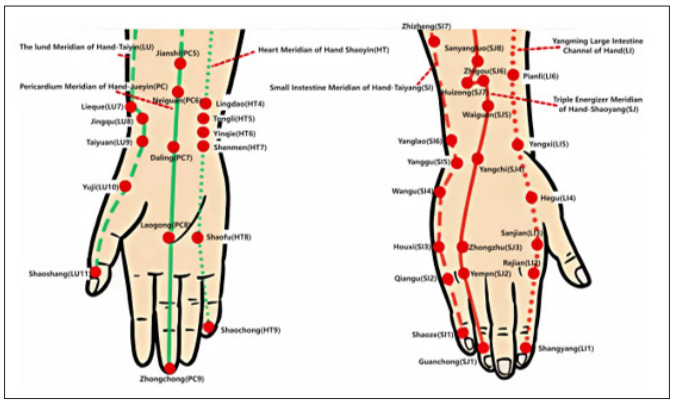



**Figure 1:** Distribution of Acupuncture Points on the Hand Meridians

Acupuncture points on the hands, such as Shaoze and Zhongzhu, are widely used in acupuncture treatment. Acupuncture at these points can regulate the flow of qi and blood in the meridians, thereby treating illnesses. The locations and specific functions of hand acupuncture points are shown in Table 1.

**Table 1: Functions of Hand Acupoints**

| Acupuncture points | Acupoint Location | Function of acupoints |
|---|---|---|
| Shaoze (SI1) | Located next to the nail corner at the end of the little finger, on the inner side of the nail root, at the slightly red part of the fingertip | It has the effects of clearing heart fire, calming the mind and promoting blood circulation. It is often used to treat insomnia, anxiety and other symptoms caused by excessive heart fire. |
| Zhongzhu (SJ3) | Located on the back of the hand, between the 4th and 5th metacarpal bones, in the depression near the metacarpophalangeal joint | It has the effects of dredging meridians, clearing away heat and detoxifying, relieving sore throat and reducing swelling. It is often used to treat symptoms such as sore throat, tinnitus, and headache. |
| Shangyang (LI1) | Located at the corner of the base of the nail at the end of the index finger | It has the effects of clearing away heat and detoxifying, clearing the bowels and detoxifying, and dredging the large intestine. It is often used to treat symptoms such as constipation, fever, and headache. |
| Wailaogong (EX-UE8) | On the back of the hand, about one-third of the way out from the center of the palm, usually at the wrist crease. This acupoint is usually located somewhere on the lower back, but its location varies depending on the individual. | It has the effect of relaxing meridians and relieving pain. It is mainly used to relieve shoulder, neck, wrist pain, relax nerves, etc. |
| Low Back Pain Points | On the back of the hand, about one-third of the way out from the center of the palm, usually at the wrist crease. This acupoint is usually located somewhere on the lower back, but its location varies depending on the individual. | Lumbar pain points are used to relieve lumbago and strengthen waist muscle function. They are often used to treat lumbago and waist stiffness caused by cold, dampness, strain, etc. |
| Hegu (LI4) | Located on the dorsum of the hand, between the 1st and 2nd metacarpal bones, close to the base of the 2nd metacarpal bone. | Hegu has the functions of harmonizing qi and blood, dredging meridians, relieving headaches and facial pain, and is often used to relieve symptoms such as abdominal pain, indigestion, and colds. |

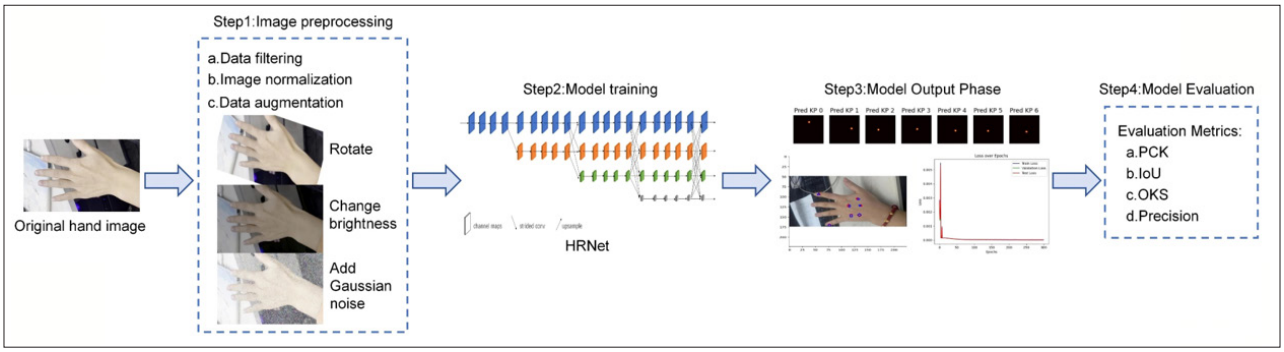



**Figure 2:** Flowchart

The research process is shown in Figure 2. First, the original hand image is preprocessed through standardization, normalization (its mathematical expression is shown in Equation (1)), and data augmentation (rotation, brightness adjustment, and Gaussian noise addition). Then, the HRNet model is constructed and trained. After training, the acupoint heat map, acupoint predicted and true position map, and loss map are output. Finally, the model performance is evaluated and compared using multi-dimensional evaluation indicators such as PCK, IoU, OKS, and precision. It is verified that the HRNet model has the best overall performance in the hand acupoint recognition task, providing key technical support for clinical diagnosis and treatment of traditional Chinese medicine and intelligent diagnosis and treatment equipment.

$$I_{norm} = \mathrm{I} \qquad (1)$$

**HRNet Network**

HRNet achieves the fusion of multi-scale information by maintaining the high-resolution features of the image. Its core is the parallel processing and fusion of features of different resolutions, rather than compressing the image to low resolution and then gradually restoring it. It breaks through the limitation of traditional networks that process image information by gradually reducing the resolution.

HRNet, first proposed in 2019, has demonstrated excellent performance on multiple standard datasets and boasts very high accuracy in human pose estimation tasks. proposed a method for segmenting cherry tomatoes by maturity in greenhouse environments, Y-HRNet. This method utilizes the YOLOv7 model for object detection, uses ROI for segmentation, and introduces ECA and DR-ASPP modules. Experimental results show that Y-HRNet can accurately segment cherry tomatoes of varying maturity.

Most traditional convolutional neural networks typically gradually reduce the image resolution, first down sampling to obtain strong semantic information and then up sampling to restore high-resolution positional information. This compresses information and reduces computational effort, but also results in a loss of detail. HRNet does not compress information by gradually reducing resolution. Instead, it maintains high-resolution feature maps throughout the network to avoid excessive information loss. It also enhances the network's expressive power through cross-resolution information exchange. High-resolution feature maps help capture details more accurately, while low-resolution feature maps help grasp global information. HRNet employs parallel multi-resolution branches, running high-resolution and low-resolution feature branches in parallel. A cross-resolution information exchange module enables multi-resolution feature fusion, fusing feature maps of different resolutions. This allows each branch to benefit from the features of other branches, enabling better processing of complex scenes or details. HRNet typically consists of multiple stages, each of which gradually increases the depth and width of the network by adding more parameters. Initially, the network is shallow and relies more on high-resolution features. As the network deepens, more low-resolution features are introduced, allowing the network to capture a wider range of contextual information. This improves the ability to capture contextual information while preserving high-resolution information.

Compared to traditional neural networks, HRNet preserves high-resolution features, can process more details, and provide more accurate predictions. Compared to more complex networks, HRNet maintains high accuracy while maintaining good computational efficiency. HRNet first down samples by a factor of 4 through two convolutional layers with a kernel size of 3×3 and a stride of 2. It then repeatedly stacks Bottleneck layers, followed by a series of Transition and Stage structures, with each Transition adding a new scale branch [36]. For example, Transition 1 builds on the output of the repeatedly stacked Bottleneck layers by applying two parallel convolutional layers with a kernel size of 3×3 to produce two different scale branches: one down sampled by a factor of 4 and one down sampled by a factor of 8. In Transition 2, a new scale branch is added to the two existing scale branches, down sampling by a factor of 16. Each scale branch in the Stage first passes through four Basic Blocks and then fuses information from different scales. The network structure is shown in Figure 3.

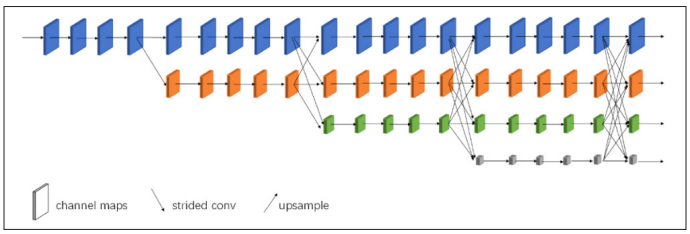



**Figure 3:** HRNet Network Architecture

HRNet exchanges information between layers of the same resolution by directly copying them, unifying the dimensions of layers requiring resolution increases through up sampling, and using convolution to reduce the resolution of layers of different resolutions. The output of each scale branch is obtained by fusing the outputs of all branches, as shown in Figure 4. Taking the output of the 4x down sampled branch as an example, the output of the 4x down sampled branch is left unprocessed, the output of the 8x down sampled branch is up sampled by a factor of 2, and the output of the 16x down sampled branch is up sampled by a factor of 4. Finally, a ReLU is used to obtain the fused output of the 4x down sampled branch [37]. In this experiment, Dropout is added after the second convolutional layer in each BottleNeck module with a dropout probability of 0.2. In each Stage Module branch, Dropout is added after the first Basic Block with a dropout probability of 0.1.

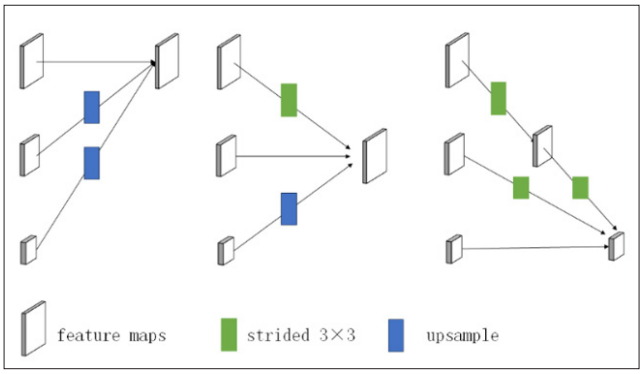



**Figure 4:** Feature Fusion at Different Resolution Layers

**Dataset Construction and Image Preprocessing**

The hand acupoint dataset is a research project from Chengdu University of Traditional Chinese Medicine. Acupoints were annotated by a Traditional Chinese Medicine practitioner with five years of acupuncture experience at the affiliated hospital. The dataset contains 1,102 hand images with labeled acupoint categories and locations. The Shaoze, Zhongzhu, Shangyang, Wailaogong, Yaotongdian, and Hegu acupoints are annotated. Taking the first hand image as an example, the corresponding JSON file is shown in Figure 5. Point 2 is Shaoze, Point 5 is Zhongzhu, Point 8 is Shangyang, Point 4 is Wailaogong, Points 6 and 7 are Yaotongdian, and Point 3 is Hegu. The dataset includes hand images from different locations and angles, as shown in Figure 6. Images missing the Shaoze, Zhongzhu, Shangyang, Wailaogong, Yaotongdian, and Hegu acupoints were removed, resulting in a total of 784 images. This filtered dataset was then rotated, brightness and contrast adjusted, and Gaussian noise added to generate diverse data. Among them, the rotation operation must ensure that the acupoints are still in the image after preprocessing, and the positions of each acupoint in the JSON file are rotated accordingly. If the key points are not in the image after rotation, they are rotated again. The size of the rotated image is saved as the original image size and the rotated area is filled with white. The data set processed by data enhancement technology has a total of 3136 images.

Image preprocessing involves normalizing the image's pixel values to a range of 0 to 1. Based on the original image's aspect ratio, the image is scaled to 224×224, while maintaining its proportions to ensure the hand image is not distorted. After scaling, the image is padded to 224×224 with zeros on the shorter sides. The image dimensions are converted from High Width Convolution (HWC) to High Width Convolution (CHW) to meet PyTorch input requirements. To improve model performance, the image is normalized by subtracting the ImageNet mean and dividing by the standard deviation. Key point data is extracted from a JSON file and adjusted based on the scaling factor. The preprocessed image, key points, and labels are then returned as training data, ensuring data consistency and adaptability, effectively improving the stability and accuracy of model training.


**Figure 5:** Dataset Details

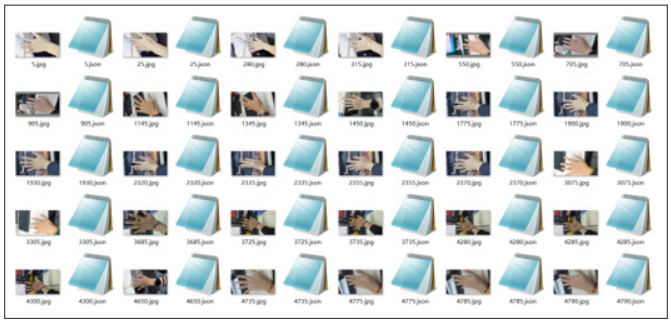


**Figure 6:** Some Different Hand Images and Hand Images Taken from Different Angles

**Experimental Design**

Experiments for this algorithm were conducted in the PyTorch framework, including network training and testing. The experimental environment was Ubuntu 18.04, using the Miniconda environment manager, Python 3.8, and CUDA 11.1. The hardware configuration consisted of an RTX 4090 (24GB GPU), a 16-core Intel Xeon Platinum 8352V processor (2.10GHz), and 120GB of RAM. During ResNet and VGGNet network training, the input images were resized to 224×224. The MSE loss was used, and the networks were optimized using the AdamW optimizer with a batch size of 32, a learning rate of 0.000, and 300 iterations. During HRNet and Simple Baseline network training, the heatmap output size was fixed at 224×224.

During training, the dataset was divided into a training set, a test set, and a validation set, accounting for 70%, 20%, and 10%, respectively. Dropout and early stopping strategies were implemented to prevent overfitting. Dropout randomly drops neurons in each training iteration to reduce model complexity. These dropped neurons do not participate in forward and backward propagation. This method can be considered a method for training multiple subnetworks, enhancing the model's generalization ability. Early stopping involves monitoring the validation set loss during training. After training for 200 epochs to ensure complete training, if the validation set loss does not decrease after 20 consecutive epochs, training is stopped early to prevent the model from over-learning patterns and features from the training set, potentially leading to overfitting.

In this experiment, the training set was used for training and parameter updates during forward propagation. The validation set was used for performance evaluation during training, with early stopping implemented to prevent overfitting. The test set was used to evaluate model performance. A model preservation mechanism was also introduced: if the validation set loss for the current epoch is lower than the global validation set loss, the model parameters for that epoch are saved. A visualization library is used to intuitively display model evaluation indicators such as PCK and precision in the form of a line graph, which makes it easier to visually display whether there is overfitting and adjust hyperparameters.

**Model Evaluation**

To comprehensively evaluate the performance of the constructed deep learning model in the hand acupoint recognition task, this study selected several commonly used evaluation metrics for key point detection and image recognition, including mean squared error (MSE), correct key point ratio (PCK), intersection over union (IoU), precision, and target key point similarity (OKS). A detailed description of each metric is as follows:

Mean Squared Error (MSE) is used to evaluate the performance of the prediction model. MSE measures the difference between the predicted value and the true value by calculating the square of the difference between the predicted value and the true value and dividing it by the number of samples. The smaller the MSE value, the better the model's prediction performance and the closer the prediction result is to the true value. The formula for the mean squared error is shown in Equation (2).

$$MSE=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2 \qquad (2)$$

Among them, is the i-th true value, is the i-th predicted value, and n is the total number of samples.

The Percentage of Current Key points (PCK) is a metric used to evaluate the performance of key point detection models. It represents the proportion of correctly estimated key points and measures whether the distance between the predicted key points and the true key points is within a certain threshold. If so, the prediction is considered correct. The core idea is to calculate the proportion of detected key points whose normalized distance to their corresponding ground truth is less than a set threshold. The calculation formula is shown in Equation (3). The proportion of correct detections is used to measure model accuracy.

$$PCK_{\sigma}^{p}(d_0)=\frac{1}{|T|}\sum_{T}\delta(\|x_p^f-y_p^f\|_2<\sigma) \qquad (3)$$

Where, represents the detector, the threshold indicating whether the key point matches the ground truth, and the correct estimated key point ratio. In this experiment, the threshold is 0.03 of the image size.

Intersection over Union (IoU) measures the degree of overlap between the predicted box and the ground-truth box. Model performance is evaluated by comparing the ratio of the intersection to the union between the model's predicted area and the actual annotated area. In acupoint recognition, IoU reflects the degree of spatial overlap between the predicted acupoint location and the ground-truth location. A higher value indicates more accurate positioning. The formula for calculating IoU is shown in Equation (4).

$$IoU=\frac{The\ intersection\ of\ the\ predicted\ box\ and\ the\ true\ box}{The\ union\ of\ the\ predicted\ box\ and\ the\ true\ box} \qquad (4)$$

**Precision** indicates how many of the samples predicted as positive by the model are actually positive, that is, the ratio of true positive samples to predicted positive samples. For acupoint recognition tasks, precision reflects the accuracy of the model in identifying the target acupoint category. The calculation formula is shown in Equation (5).

$$Precision=\frac{TP}{TP+FP} \quad (5)$$

TP represents the number of acupoints correctly predicted by the model, TN represents the number of acupoints not correctly predicted by the model, FP represents the number of acupoints incorrectly predicted by the model as non-acupoints, and FN represents the number of acupoints that the model failed to identify as actually existing. In this experiment, a correct prediction was considered when the Euclidean distance between the true and predicted acupoint coordinates was less than 2 pixels.

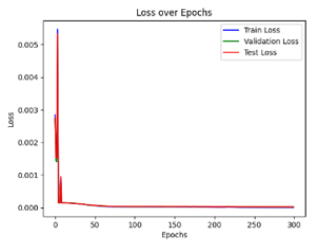
Object Key Point Similarity (OKS) is an evaluation metric used to assess the performance of keypoint detection models. Similar to IoU, it takes into account key point prediction errors and human body scale, and calculates the similarity between predicted key points and true key points. Higher OKS values indicate closer key point predictions are to the true values, leading to higher recognition accuracy [38]. The OKS calculation formula is shown in Equation (6).

$$OKS=\frac{\sum_i \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right)\delta(v_i>0)}{\sum_i \delta(v_i>0)} \quad (6)$$

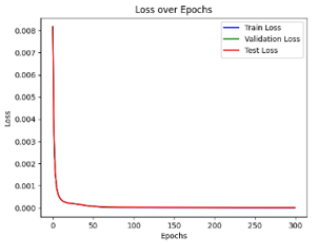
$$d_i=\sqrt{\left(x_{pre}-x_g\right)^2+\left(y_{pre}-y_g\right)^2} \quad (7)$$

Among them, represents the Euclidean distance between the predicted key point and the true key point of the i-th person, and the calculation formula is shown in formula (7), () is the predicted key point, () is the true key point, s represents the target scale, and is a constant that controls attenuation. In this experiment, all settings in this article are a constant of 0.3. is the visibility indicator function, which indicates whether the key point is visible in the image. When >0, it means that the position of the key point can be observed, and ≤0 means that the position of the key point cannot be observed.

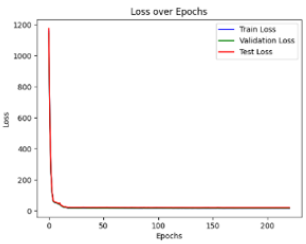
## Results



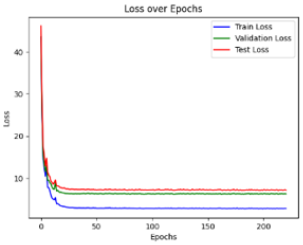

(A) HRNet loss graph




(B)Simple Baseline loss graph




C) ResNet50 loss graph




(D)VGG16 loss graph

**Figure 7:** Loss Graph of Each Model

Figure 7 compares the loss changes of four models, HRNet, Simple Baseline, ResNet50, and VGG16, over 300 rounds of training. The loss calculation is the mean squared error between the predicted and true heatmaps, analyzing the fitting performance and convergence speed during training.

Figure 7(A) shows the loss plot of the HRNet model. It can be seen that the loss value of the HRNet model is high at the beginning of training. As the number of training rounds increases, the loss continues to decrease, eventually reaching $1.92\times10^{-5}$ on the test set, demonstrating good convergence and stability. Figure 7(B) shows the loss plot of the Simple Baseline model. The Simple Baseline model also demonstrates good fitting ability, with a final loss of $1.89\times10^{-5}$. The loss curve generally shows a downward trend with increasing training rounds. Figure 7(C) shows the loss plot of the ResNet50 model. It can be seen that the ResNet50 model triggers early stopping at the 221st round, with a test set loss as high as 20.64. The loss decreases slowly, and the final convergence value is high, indicating that the acupoint locations predicted by the model after training are far from the actual acupoints, making it difficult to effectively fit the target distribution. Figure 7(D) shows the loss plot of the VGG16 model. The VGG16 model also triggers early stopping at the 221st round, with a test set loss of 7.17, indicating that its positioning accuracy has improved, but there is still a significant gap compared to the model based on heatmap regression.

Overall, the loss of HRNet and Simple Baseline is below $2\times10^{-5}$, and the decline is steady. VGG16 has a loss of 7.17, and ResNet50 has the highest loss of 20.64. The loss of the key point detection model based on heatmap regression is lower than that based on coordinate regression.

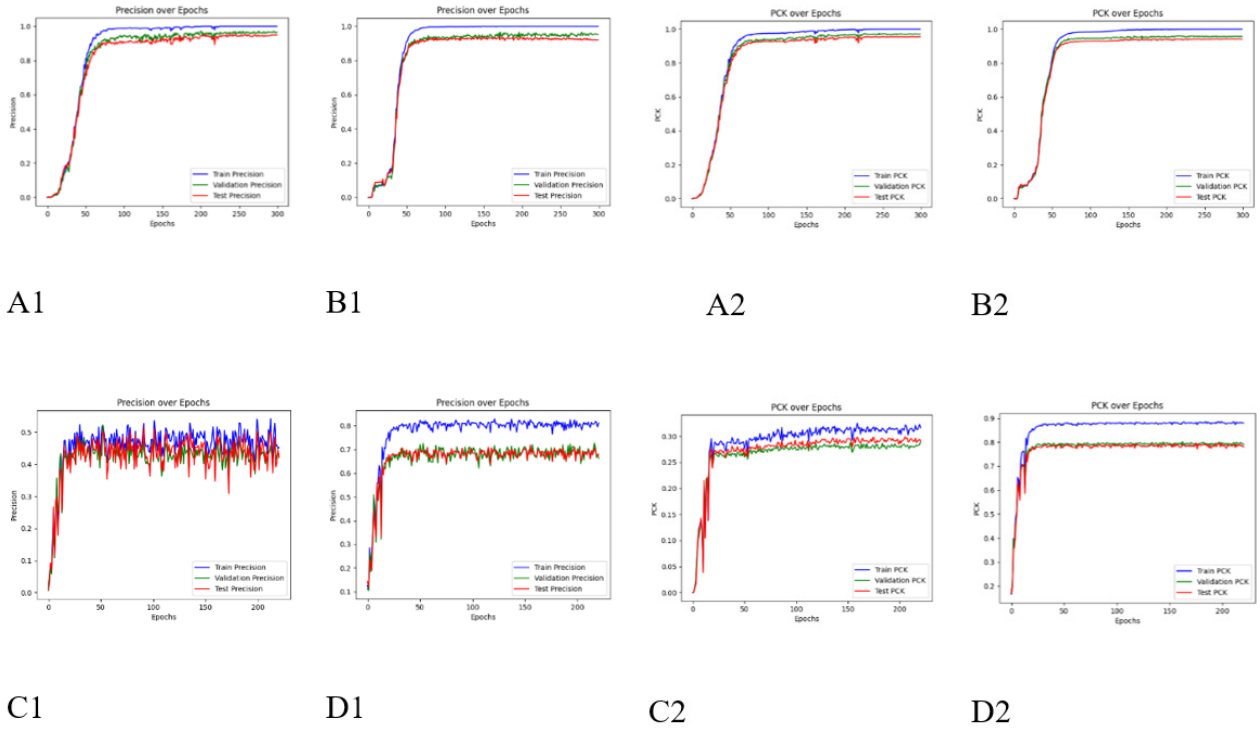



**Figure 8:** Visualization of Accuracy

**Figure 9:** Visualization of PCK

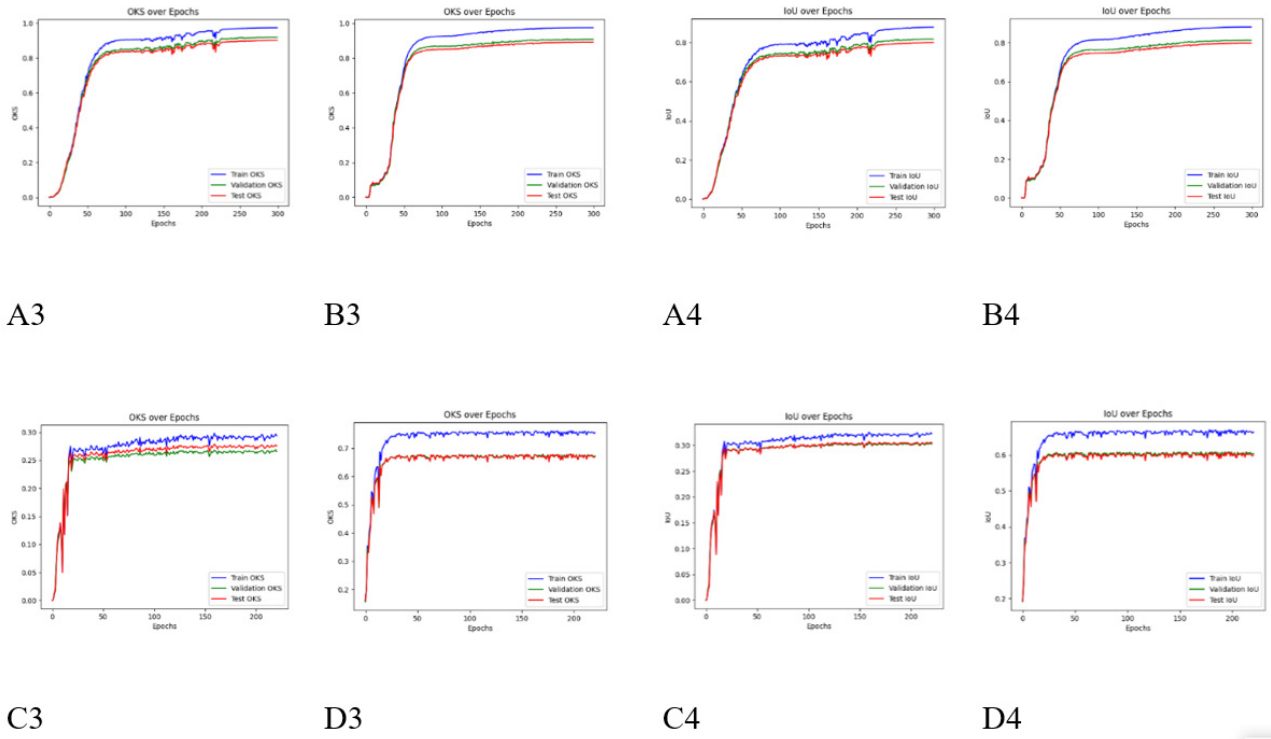



**Figure 10:** Visualization of OKS

**Figure 11:** Visualization of IoU

**Note:** A represents the HRNet model, B represents the Simple Baseline model, C represents the ResNet50 model, and D represents the VGG16 model.

Figure 8: compares the accuracy of the four models on the test set: HRNet, Simple Baseline, ResNet50, and VGG16. The HRNet model achieved an accuracy of 94.89%, Simple Baseline achieved 92.34%, VGG16 achieved 69.37%, and ResNet50 achieved 23.28%.

Figure 9: compares the PCK (key point correctness) of the four models on the test set. HRNet achieved a PCK of 95.37%, Simple Baseline achieved 94.12%, VGG16 achieved 78.94%, and ResNet50 achieved 28.86%.

Figure 10: compares the OKS (key point similarity) values of the four models on the test set. HRNet achieved an OKS of 90.07%, Simple Baseline achieved 89.19%, VGG16 achieved 67.60%, and ResNet50 achieved 27.45%.

Figure 11: compares the Intersection over Union (IoU) values of the four models on the test set. The IoU for the HRNet model is 80.04%, Simple Baseline is 79.78%, VGG16 is 60.47%, and ResNet50 is 30.40%.

Overall, HRNet outperforms Simple Baseline, ResNet50, and VGG16 in all four metrics: PCK, OKS, IoU, and Precision. Simple Baseline is slightly lower than HRNet in all metrics, but higher than ResNet50 and VGG16. VGG16 outperforms ResNet50 in all four metrics: PCK, OKS, IoU, and Precision. ResNet50 achieves the lowest overall scores for all four metrics.

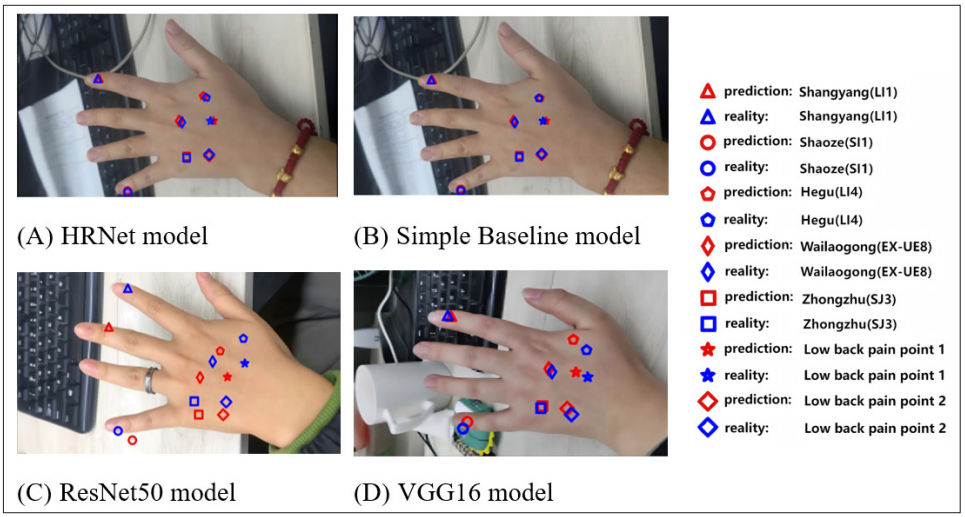



**Figure 12:** Predicted and actual locations of acupoints by each model

Figure 12 compares the results of four models, HRNet, Simple Baseline, ResNet50, and VGG16, after converting the heatmaps of each acupoint into coordinates. The predicted and true locations of each acupoint are displayed in different colors and shapes within the original image. Figure 13(A) shows the predicted and true locations of the acupoints by the HRNet model; Figure 13(B) shows the predicted and true locations of the acupoints by the Simple Baseline model; Figure 13(C) shows the predicted and true locations of the acupoints by the ResNet50 model; and Figure 13(D) shows the predicted and true locations of the acupoints by the VGG16 model. It can be seen that the HRNet and Simple Baseline models are relatively accurate in predicting acupoint locations within a certain error range, but the degree of overlap is low. The ResNet50 and VGG16 models have low acupoint localization accuracy, with the true acupoint locations being far from the model-predicted locations and the degree of overlap being low.

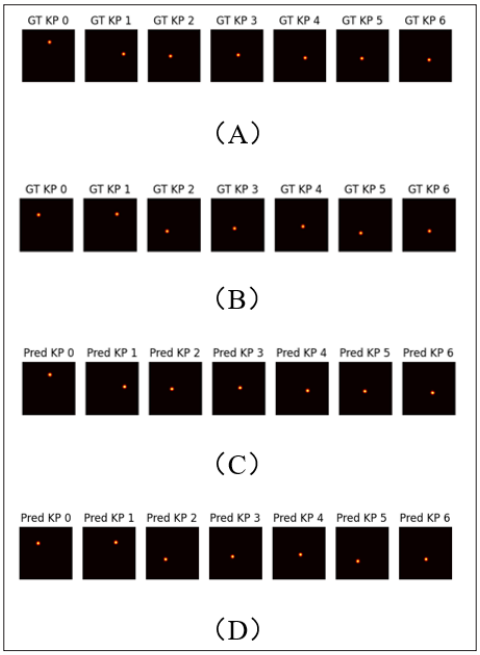



**Figure 13:** Heatmap of the actual and predicted acupoints based on the HRNet and Simple Baseline models

**Note**
(A) is the real heat map of each acupoint based on the HRNet model, (B) is the real heat map of each acupoint based on the Simple Baseline model, (C) is the predicted heat map of each acupoint based on the HRNet model, and (D) is the predicted heat map of each acupoint based on the Simple Baseline model.

Figure 13 compares the two key point detection models based on heat map regression, HRNet and Simple Baseline, which convert the heat map of each acupoint into coordinates and display the predicted and actual positions of each acupoint in different colors and shapes in the original image. Among them, Figure 8 (A) and Figure 8 (C) are the actual heat map and predicted heat map of each acupoint based on the HRNet model, respectively; Figure 8 (B) and Figure 8 (D) are the actual heat map and predicted heat map of each acupoint based on the Simple Baseline model, respectively. Overall, both models show high accuracy in predicting acupoint positions.

**Table 2: Quantitative Data of Each Model Evaluation**

| 10-5 | Dataset | Loss | PCK | IoU | Precision | OKs |
|---|---|---|---|---|---|---|
| HRNet | Training set | $1.68\times10^{-6}$ | 99.89% | 87.84% | 1.0 | 97.34% |
| | Validation set | $1.50\times10^{-5}$ | 97.04% | 81.72% | 96.49% | 91.79% |
| | Test set | $1.92\times10^{-5}$ | 95.37% | 80.04% | 94.89% | 90.07% |
| | Training set | $1.24\times10^{-6}$ | 99.97% | 88.11% | 1.0 | 97.52% |
| Simple Baseline | Validation set | $1.46\times10^{-5}$ | 95.67% | 81.28% | 95.85% | 90.87% |
| | Test set | $1.89\times10^{-5}$ | 94.12% | 79.78% | 92.34% | 89.19% |
| | Training set | 16.44 | 31.68% | 32.21% | 24.00% | 29.35% |
| ResNet50 | Validation set | 17.37 | 27.79% | 30.21% | 21.01% | 26.47% |
| | Test set | 20.64 | 28.86% | 30.40% | 23.28% | 27.45% |
| | Training set | 2.77 | 88.18% | 66.61% | 80.91% | 75.73% |
| VGG16 | Validation set | 6.14 | 80.20% | 60.66% | 70.70% | 67.67% |
| | Test set | 7.17 | 78.94% | 60.47% | 69.37% | 67.60% |

Table 2 shows that HRNet's test set loss is $1.92\times10^{-5}$, with a PCK value of 95.37%, an IoU value of 80.04%, a precision of 94.89%, and an OKS value of 90.07%. SimpleBaseline's test set loss is $1.89\times10^{-5}$, with a PCK value of 94.12%, an IoU value of 79.78%, a precision of 92.34%, and an OKS value of 89.19%. ResNet50's test set loss is 20.64, with a PCK value of 28.86%, an IoU value of 30.40%, a precision of 23.28%, and an OKS value of 27.45%. VGG16's test set loss is 7.17, with a PCK value of 78.94%, an IoU value of 60.47%, a precision of 69.37%, and an OKS value of 67.60%.

Overall, the evaluation metrics of the heatmap regression-based keypoint detection models HRNet and SimpleBaseline are significantly higher than those of the coordinate regression-based keypoint detection models ResNet50 and VGG16. HRNet's loss is slightly higher than SimpleBaseline's, while all other evaluation metrics are higher. VGG16's loss is lower than ResNet50's, and all other evaluation metrics are higher than ResNet50's.

**Discussion**
With the in-depth integration of deep learning technology with the medical field, deep learning has been widely applied in fields such as medical image segmentation and assisted diagnosis. Experimental results show that the proposed HRNet model, based on heatmap regression, achieves the best overall performance in the hand acupoint recognition task. On the test set, the HRNet model leads the way in all metrics, demonstrating its high consistency in recognition accuracy and spatial localization. Compared to the HRNet model, Simple Baseline, while performing better on the training set, performs slightly worse on the test set, indicating a degree of overfitting—overfitting on the training set leads to decreased generalization and an inability to maintain optimal performance on the test set.

Among coordinate regression models, ResNet50 has the worst overall performance, reflecting its weak ability to model the spatial distribution of acupoints. While VGG16 also performs worse in accuracy and spatial localization than the heatmap regression model, it shows significant improvements over ResNet50. This demonstrates that VGG16, despite its relatively shallow architecture, still possesses considerable feature extraction capabilities. While ResNet50, despite its deeper network, may not be sufficiently expressed in this task due to its residual structure, its training performance is poor. The combined training, validation, and test performance of the four models confirms the effectiveness and superiority of the proposed method for hand acupoint recognition. Furthermore, comparative analysis further emphasizes the technical advantages of key point detection methods based on heatmap regression in this task.

This study focuses on the task of hand acupoint recognition and proposes a deep learning-based hand acupoint recognition algorithm. This innovatively combines deep learning with hand acupoint recognition, aiming to utilize an annotated dataset and data augmentation techniques to train a deep learning model and achieve automated hand acupoint recognition.

This approach provides location support for automated diagnostic and treatment equipment such as acupuncture and moxibustion robots; provides hand acupoint recognition data for young Traditional Chinese Medicine practitioners to facilitate rapid mastery of the method; and provides data reference for TCM practitioners to avoid misidentification of acupoints due to lack of experience or fatigue, which can lead to missed or misdiagnosed cases.

Looking forward, the current model still has certain limitations. The dataset used in this study contains only images of the left hand's back and covers only six acupoints: Shaoze, Zhongzhu, Shangyang, Wailaogong, Yaotongdian, and Hegu. Therefore, the model's recognition capabilities are limited to a specific hand region. Future work could further expand the dataset to collect and annotate diverse hand images, including left palm, right back, and right palm. Incorporating a module for automatic hand type discrimination (e.g., left or right hand, palm/back) into model construction would enable intelligent recognition of acupoints across the entire hand. Furthermore, the trained model can be deployed using the Flask framework, creating a concise and practical web interface. This will promote the practical application and commercialization of hand acupoint recognition systems and further advance the digital and intelligent development of Traditional Chinese Medicine acupuncture.

**References**

1. Lin JG, Kotha P, Chen YH (2022) Understandings of Acupuncture Application and Mechanisms. Am J Transl Res 14: 1469.
2. Deng H, Shen X (2013) The Mechanism of Moxibustion: Ancient Theory and Modern Research. Evid Based Complement Alternat Med 2013: 379291.
3. Patel M, Urits I, Kaye AD, Omar Viswanath (2020) The Role of Acupuncture in the Treatment of Chronic Pain. Best Pract Res Clin Anaesthesiol 34: 603-616.
4. Mehta P, Dhapte V, Kadam S, Vividha Dhapte (2017) Contemporary Acupressure Therapy: Adroit Cure for Painless Recovery of Therapeutic Ailments. J Tradit Complement Med 7: 251-263.
5. Fang T, Li Q, Zhou F, Liu F, Liu Z, et al. (2020) Effect and Safety of Acupotomy in Treatment of Knee Osteoarthritis: A Systematic Review and Meta-Analysis. J Tradit Chin Med 40: 355-364.
6. Ying C, Jingqing S, Tianli L, Jiahui H, Yuhan L, et al. (2023) Effect of Acupuncture Treatment on Nonketotic Hyperglycemic Hemichorea-Hemiballismus: A Case Report. J Tradit Chin Med 43: 829-833.
7. Ling FY, Qi WC, Li X (2023) Exploration of Shaoze Acupoint's Main Indications and Compatibility Rules Based on Data Mining Technology. World Sci Technol Mod Tradit Chin Med 25: 1648-1655.
8. Fu W, Li S, Zhang W, Feng Gao, Zhongyu Zhou (2021) Rule of Acupoint Combination in Acupuncture Treatment of Xiang Bi Disease: A Study Based on Association Rules. In: Proceedings of the 2021 International Conference on Social Sciences and Big Data Application (ICSSBDA 2021) Atlantis Press 235-240.
9. Qin XH, Yuan SG (2019) Clinical Observation on Acupuncture at Yanggu and Hegu Points for the Treatment of Post-Stroke Finger Contracture. J Clin Res Tradit Chin Med 11: 86-89.
10. Sun Q, Ma J, Craig P, Dai L, Lim EG (2025) Acusim: A Synthetic Dataset for Cervicocranial Acupuncture Points Localisation. Sci Data 12: 625.
11. Zhang F, An Q, Song W, Bai M (2024) Research on Human Acupoint Detection by Integrating Key Point Information and Acupoint Theory. IEEE Access 12: 181889-181898.
12. Zhao Y, Zhang T, Lian QS (2012) Acupuncture Points Localization Algorithm for Bell's Palsy on Facial Infrared Images. Mini-Micro Syst 33: 1613-1619.
13. Shu J, Ding R, Jin A, Rongtao Ding, Shu Chen (2022) Acupoint Selection for Autonomous Massage Based on Infrared Thermography. Trait Signal 39: 355-362.
14. Xie S, Liu Z, Zhang L, Fengfeng Song (2023) A New Approach to the Position of Acupoints on the Back of Humans. Recent Pat Mech Eng 16: 150-162.
15. Lian Y, Wang Z, Yuan H, Lifang Gao (2020) Partial Occlusion Face Recognition Method Based on Acupoints Locating Through Infrared Thermal Imaging. in: 2020 International Wireless Communications and Mobile Computing (iwcmc). Limassol: IEEE 1394-1399.
16. Wang H, Liu L, Wang Y, Senhao Du (2023) Hand Acupuncture Point Localization Method Based on a Dual-Attention Mechanism and Cascade Network Model. Biomed Opt Express 14: 5965-5978.
17. Wang C, He Z, Wang X, Yang B, Guo J, et al. (2024) An Acupoint Location Method Based on Lightweight Efficient Channel Attention Network. In: 2024 6th International Conference on Electronic Engineering and Informatics (EEI). IEEE 1592-1595.
18. Li J, Fei Z, Xie Y, Da Deng, Xingcheng Ming, et al. (2025) A review of acupoint localization based on deep learning. Chin Med 20: 116.
19. Su MT, Chiang ML, Tsai CH, Chi-Wei Lin, Rong-Xuan Liu, et al. (2023) An Acupoint Health Care System with Real-Time Acupoint Localization and Visualization in Augmented Reality. Multimed Syst 29: 2217-2238.
20. Liu YB, Qin JH, Zeng GF (2023) Facial Acupoint Location Method Based on Faster Pfld. Signal Image Video Process 17: 1-9.
21. Lee YT, Sun CH, Chiu CS (2023) Design of a Hand Back Acupoint Massage Aid. in: 2023 International Conference on System Science and Engineering (ICSSE). IEEE 508-513.
22. Bohr A, Memarzadeh K (2020) The Rise of Artificial Intelligence in Healthcare Applications. in: Bohr a, Memarzadeh k, eds. Artificial Intelligence in Healthcare. Academic Press 25-60.
23. Masood D, Qi J (2022) 3D Localization of Hand Acupoints Using Hand Geometry and Landmark Points Based on Rgb-D Cnn Fusion. Ann Biomed Eng 50: 1103-1115.
24. He K, Zhang X, Ren S, Sun J (2016) Deep Residual Learning for Image Recognition. in: Proceedings of the Ieee Conference on Computer Vision and Pattern Recognition (CVPR). IEEE 770-778.
25. Lu H, Song H (2024) Research on Image Classification Based on ResNet. Acad J Comput Inf Sci 7: 9-14.
26. Gao G, Sun Z, Mu G, Hui Yin, Yuxuan Ren (2024) Research on Marine Fish Classification and Recognition Based on an Optimized Resnet50 Model. Mar Coast Fish 16: e10317.
27. Simonyan K, Zisserman A (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015) 1-14.
28. Singh PY, Chaurasia KB, Shukla MM (2024) Mango Fruit Variety Classification Using Lightweight Vggnet Model. SN Comput Sci 5: 1083.
29. Das AG, George SB, Paul VC (2022) Electrocardiogram Signal Classification Using Vggnet: A Neural Network Based Classification Model. Int J Inf Technol 15: 119-128.
30. Toshev A, Szegedy C (2014) DeepPose: Human Pose Estimation via Deep Neural Networks. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH: IEEE; 1653-1660.
31. Xiao B, Wu HP, Wei YC (2018) Simple Baselines for Human Pose Estimation and Tracking. In: European Conference on Computer Vision (ECCV). Springer 472-487.
32. Sun K, Xiao B, Liu D, Wang J (2019) Deep High-Resolution

Representation Learning for Human Pose Estimation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE 5686-5696.

33. Liu M, Chen W, Cheng J, Wang Y, Zhao C (2024) Y-HRNet: Research on Multi-Category Cherry Tomato Instance Segmentation Model Based on Improved Yolov7 and Hrnet Fusion. Comput Electron Agric 227: 109531.
34. Newell A, Yang K, Deng J (2016) Stacked Hourglass Networks for Human Pose Estimation. in Leibe b, Matas j, Sebe n, Welling m, Editors, Computer Vision - 14th european conference, ECCV 2016, Proceedings. Springer Verlag 483-499.
35. Chang ZQ, Zhang J, Peng K (2024) Research on Hand Acupoint Positioning Based on Deep Learning. J Sens Technol 37: 1893-1902.
36. Kuang RX (2022) Identification and Positioning of Acupuncture Points on the Human Back Based on Deep Learning [Dissertation]. Nanchang: Nanchang University.
37. Lv C (2024) Research and Application of Human Posture Estimation Algorithm Based on Improved Hrnet Model [Dissertation]. Taiyuan: Taiyuan University of Science and Technology.
38. Li R, Yuchen Wen, Shujin Zhang, Xingshi Xu, Baoling Ma, et al. (2024) Automatic Measurement Method of Beef Cattle Body Size Based on Key Point Detection and Monocular Depth Estimation [Dissertation]. Northwest Agriculture and Forestry University. https://www.sciencedirect.com/science/article/abs/pii/S0957417423035443?via%3Dihub

**Copyright:** ©2025 XU Lin, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.