Haifan P. (2025) USING COMPUTER IMAGE TECHNOLOGY TO MONITOR THE INTENSITY AND EFFECT OF EXERCISE IN REHABILITATION TRAINING. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 25 (100) pp. 496-517.
DOI: https://doi.org/10.15366/rimcafd2025.100.032

ORIGINAL

USING COMPUTER IMAGE TECHNOLOGY TO MONITOR THE INTENSITY AND EFFECT OF EXERCISE IN REHABILITATION TRAINING

Pang Haifan

Department of Physical Education, China University of Political Science and Law, Beijing 102249, China.
E-mail: haifan1209@163.com

Recibido 22 de junio de 2024 **Received** June 22, 2024
Aceptado 22 de diciembre de **2024 Accepted** December 22, 2024

ABSTRACT

This paper uses rehabilitation training monitoring technology based on computer image technology to monitor the exercise intensity and effect of patients in rehabilitation training in real-time. The dynamic video of the patient is captured by a high-definition camera, and the Open Pose algorithm is used to detect the key points of human movement to obtain movement data such as joint displacement and speed. The random forest model uses an integrated learning method to complete the classification of exercise intensity by constructing multiple decision trees. At the same time, ResNet (Residual Network) is applied to extract the spatiotemporal features of the movement pattern using multi-layer convolution operations, and classify and evaluate the exercise effect. Finally, the system feeds back the monitoring results to the rehabilitation trainer through a visual interface, uses metabolic equivalent of task (MET) and heart rate changes as the core indicators for exercise intensity evaluation, and comprehensively evaluates the recovery situation based on the range of joint movement, gait stability and cycle training results. The experimental results show that after 7 cycles of rehabilitation training, the patient's recovery rate increases to 93.2%.

KEYWORDS: Computer Vision Technology, Rehabilitation Training, Exercise Intensity, Exercise Effect, Deep Learning.

1. INTRODUCTION

With the intensification of the global aging trend, the importance of rehabilitation training in improving patients' quality of life and promoting the

recovery of physical functions has become increasingly prominent. According to statistics from the World Health Organization, more than 60% of the elderly who have undergone surgery, trauma or chronic diseases (Livneh, 2022) need rehabilitation training (Rutkowski et al., 2020). Most of the traditional rehabilitation training methods rely on manual monitoring and evaluation, which not only suffers from the short comings of high subjectivity and inaccurate data recording, but also makes it difficult to capture the intensity and effect of exercise in real-time. This limitation makes it impossible for many patients to receive timely and effective feedback, which in turn affects the effectiveness of rehabilitation. Developing a method that can monitor rehabilitation training in real-time is particularly important (Suchomel et al., 2021), which not only improves the accuracy of exercise monitoring, but also enhances patient engagement and training motivation. Benefiting from the rapid development of computer image technology (Zhao et al., 2022), people can capture the patient's movement status in real-time through the camera, and use image processing algorithms to analyze various parameters during the movement process. The application of this technology can not only provide more accurate motion data, but also integrate multi-dimensional motion information to achieve a comprehensive evaluation of the effect of rehabilitation training. With the continuous advancement of artificial intelligence and deep learning (Sejnowski, 2020)technology, more complex models are used to analyze and interpret motion data, improving the level of intelligent monitoring (Ageed et al., 2021). Many researchers have begun to focus on the application of computer image technology in motion monitoring in recent years. For example, scholars such as Wu Z and Dalal S (Dalal et al., 2023; Wu et al., 2021) proposed a motion detection method based on image processing. This method used a high-resolution camera to capture motion data in real-time and used computer vision technology to evaluate motion intensity. This method achieved good results in many sports and provided a new perspective and idea for motion monitoring. Scholars such as Tripathi M and Li Z (Li et al., 2021; Tripathi, 2021) proposed a new deep learning method in the field of rehabilitation training, using convolutional neural network (CNN) to analyze the patient's motion image data to monitor their motion quality. By optimizing the network structure and applying data enhancement technology, the accuracy and robustness of motion detection are effectively improved. Although these technologies have potential application value, they still face some challenges in practical applications. For example, factors such as lighting changes and background interference during image capture may affect the effect of motion monitoring. In the study of rehabilitation training, scholars such as Postolache O and Nascimento L (Nascimento et al., 2020; Postolache et al., 2020) emphasized the importance of combining motion monitoring systems with patients' personalized needs. They believed that traditional monitoring methods often ignore individual differences in patients, resulting in monitoring results that cannot effectively reflect their true rehabilitation progress. Therefore, how to incorporate

personalized design into monitoring technology can be a key direction for future research. In the study of the above problems, Cao X and Zhu Y C (Cao et al., 2020; Zhu et al., 2021) and other scholars used convolutional neural networks to improve the accuracy of image analysis and achieved good results. Their research showed that the application of deep learning models in motion monitoring can significantly improve the processing speed and accuracy of motion data. For example, in the recognition of key postures during patient rehabilitation training, deep learning technology has demonstrated powerful feature extraction capabilities and can effectively distinguish normal and abnormal motion patterns. Existing methods generally lack adaptability and generalization capabilities for different types of motion. Mao W and Gao Z (Gao et al., 2021; Mao et al., 2021) and other scholars pointed out in their research that traditional convolutional neural networks may suffer from a decrease in recognition rate due to overfitting when facing complex motion scenes. In order to overcome this problem, they proposed a multi-level convolution structure that effectively improved the temporal analysis capability of motion patterns. However, these methods are still insufficient in terms of multidimensional data integration. Therefore, this paper adopted the ResNet model, with its deep network structure and residual learning advantages, to overcome the limitations of existing methods in motion monitoring. The ResNet model showed strong adaptability and accuracy in processing complex data, which made it an ideal choice for solving motion monitoring problems. By applying more advanced algorithms, it is expected to effectively improve the monitoring effect and data processing capabilities of rehabilitation training. This paper proposed an innovative computer image technology monitoring system for evaluating the intensity and effect of patients in rehabilitation training. The system used advanced image processing technology to achieve real-time monitoring of rehabilitation training and integrate multi-dimensional motion data to provide more objective and accurate monitoring results. A high-definition camera was used to capture dynamic videos of patients in training to collect rich motion information. The collected video data was processed by multiple algorithms in the OpenCV library (Susim & Darujati, 2021) to denoise, grayscale and normalize the captured images to eliminate environmental interference and improve image quality. The Open Pose algorithm (D'Antonio et al., 2021) was used to detect motion key points and extract joint displacement and velocity information. The random forest algorithm (Boateng et al., 2020) was used to analyze the displacement and velocity data of key points of the movement, determine the intensity of the movement and classify it into mild, moderate and severe. The ResNet model (Wen et al., 2020) was applied to evaluate and classify the exercise effect, and the transfer learning technology (Wang et al., 2021) was used to enable the model to adapt to different movement patterns and further improve the accuracy of the evaluation. In order to improve the user experience, the system also built a real-time feedback mechanism to display the monitoring results to the rehabilitation trainees through a visual interface.

The purpose of this study is to improve the intelligence and objectivity of rehabilitation training monitoring, provide a scientific basis for the rehabilitation of patients, and thus promote the progress and development of the field of rehabilitation training.

2. Monitoring System Based on Computer Image Technology

2.1. Data Acquisition

This system uses a high-definition camera to collect data to ensure that high-quality dynamic video is obtained during rehabilitation training. The high-definition camera is installed in the patient's motion area to ensure that the training scene can be fully covered. This system chooses to use the Logitech C920 HD Pro Webcam, which supports 1080p high-definition resolution and can record the motion process at a speed of 30 frames per second to ensure that every detail of the movement is captured. As shown in Figure 1, through this high-definition camera device, the system can capture the trajectory and posture of key movements in the movement, providing a basis for subsequent motion analysis.

Figure 1: Device Capture Image

During the data acquisition process, the system uses camera calibration technology (Huang et al., 2020) to ensure that the camera data is aligned with the real-world coordinate system. The mathematical model of camera calibration can be expressed as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathrm{K} \cdot \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

Among them, $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ is the image coordinate; K is the camera intrinsic

parameter matrix; R is the rotation matrix; T is the translation vector; $\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$ is the world coordinate. By using a checkerboard pattern for calibration and using the corresponding functions in the OpenCV library for calculation, the influence caused by device errors can be effectively eliminated, thereby improving the accuracy and consistency of data. To achieve real-time monitoring, the data acquisition process adopts multi-threaded processing technology (Samy et al., 2021). One thread is responsible for the acquisition of high-definition video data to ensure that the data is recorded synchronously in real-time. Through this design, the system can effectively improve the efficiency of data acquisition and avoid the problem of inaccurate motion analysis caused by data delay. Assuming that thread A processes video data and the time delay is Δt_A, it is required that:

$$\Delta t_A \approx 0 \rightarrow \text{ensure data synchronization} \tag{2}$$

The collected dynamic video and depth information are transmitted to the data processing unit via Wi-Fi or Bluetooth. To ensure the stability of the signal and the smoothness of data transmission, this paper selects a high-bandwidth transmission protocol to avoid data loss due to signal interference. The bandwidth of data transmission can be expressed as follows:

$$B = \frac{D}{T} \tag{3}$$

Among them, B is the bandwidth; D is the data size (bits); T is the transmission time (seconds). The system also provides a local storage function so that data can still be recorded and uploaded later when the network is unstable. To improve the flexibility and convenience of monitoring, the system also considers the application of mobile devices. Users can use smartphones or tablets to remotely monitor and collect data, and use the camera and sensors of mobile devices to build temporary monitoring points. In this mode, the system still maintains data consistency and transmits motion data to the server in real-time through mobile applications, which facilitates the monitoring and adjustment of rehabilitation training. Through the two-dimensional dynamic video captured by the high-definition camera, this system can provide a detailed motion data set to assist in subsequent motion analysis. Combined with motion trajectory analysis technology, the system can accurately extract key points and action paths in motion. The extraction and trajectory analysis of motion key points can be described by the following formula:

$$\mathrm{P}_{2D} = \mathrm{K}^{-1} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{4}$$

Among them, P_{2D} is the coordinate of the motion key point in the two-dimensional image plane, and a complete motion trajectory is constructed by tracking these key points. This system effectively solves the problems of accuracy, real-time and flexibility of data acquisition during rehabilitation training by combining the Logitech C920 high-definition camera with multi-threaded processing technology, and provides reliable data support for the monitoring and analysis of subsequent exercise intensity and effect.

2.2. Image Preprocessing

After data acquisition is completed, the image preprocessing stage aims to provide high-quality and stable image data for subsequent motion analysis. Through strict preprocessing steps, the impact of external factors such as device noise and illumination changes on the accuracy of motion analysis can be significantly reduced. In order to simplify the amount of calculation and reduce the interference of color information on motion detection, the image needs to be grayed after denoising. This paper uses the cv2.cvtColor() function in OpenCV to convert the RGB image into a grayscale image Y=0.299R+0.587G+0.114B. R, G, and B represent the red, green, and blue channel pixel values in the image, respectively, and Y represents the brightness value in the grayscale image. Through this weighted average method, graying retains the brightness information of the image while eliminating the interference of color on subsequent analysis. Since the lighting conditions in the rehabilitation training scene may change, graying helps to reduce fluctuations under different lighting environments, making the motion features in the image more consistent. To further reduce the impact of lighting changes on the image, the grayed image also needs to be normalized. The purpose of normalization is to standardize the pixel values in an image to a certain range to enhance the contrast and improve the robustness of feature extraction. The following formula is used to linearly normalize the image pixel values:

$$I' = \frac{I - I_{min}}{I_{max} - I_{min}} \tag{5}$$

Among them, I is the original grayscale value; I_{min} and I_{max} are the minimum and maximum grayscale values in the image respectively; I' is the normalized pixel value, which is usually in the range of [0, 1]. This normalization method can effectively improve the contrast of the image, especially when the image brightness range is small, by enhancing the details of the darker area, ensuring that the motion key points can still be accurately detected under different lighting environments. The original video data is usually interfered by environmental noise, camera device jitter, etc., so image denoising is required. This paper uses Gaussian Blur (Liu et al., 2020) to smooth the image and remove high-frequency noise. This method reduces detail noise by weighted averaging each pixel in the image with its surrounding pixels. The calculation

formula is as follows:

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (6)$$

Among them, σ is the standard deviation parameter, which is used to control the degree of filtering. By selecting an appropriate standard deviation, small noise is smoothed out while retaining key features in motion. During the implementation, the standard deviation σ is experimentally determined to eliminate noise without blurring motion details. Gaussian filtering can effectively reduce the impact of device jitter and uneven ambient light on image quality, making subsequent key point detection more stable, as shown in Figure 2.

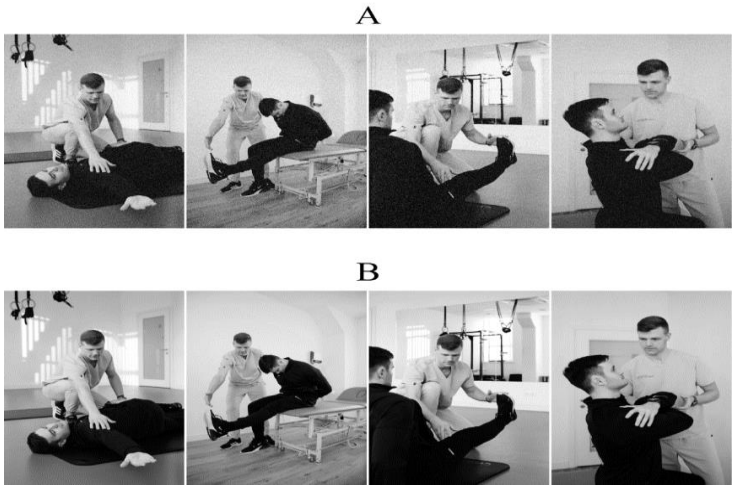

Figure 2: Grayscale and Gaussian Filter Denoising Results. (Figure 2 (A): Grayscale image; Figure 2 (B): Gaussian filter denoising image).

To further improve the contrast of the image, this paper also uses adaptive histogram equalization (AHE) (Mehdizadeh et al., 2023). Compared with the traditional global histogram equalization, adaptive histogram equalization divides the image into several small blocks and performs histogram equalization on each small block independently, thereby enhancing the contrast of the local area of the image. The specific implementation adopts the CLAHE (Contrast Limited Adaptive Histogram Equalization) algorithm (Singh et al., 2020), which applies a contrast limiting factor to avoid the noise amplification problem that may occur during the equalization process. The image is divided into several non-overlapping small grid areas, and the histogram is calculated in each grid area. The histogram is equalized. For the grids at the boundary of the image, the bilinear interpolation method is used to smooth the pixel values. The contrast limiting factor is used to suppress the excessive peaks in the histogram to avoid noise diffusion. Through CLAHE, the contrast of local areas in the image is enhanced, especially the details of key parts in motion become clearer after contrast enhancement, ensuring the accuracy of key point detection and motion trajectory analysis, as shown in

Figure 3. In Figure 3 (C) and Figure 3 (D), the horizontal axis represents the gray level of each pixel in the image, and the vertical axis represents the number of pixels with this gray level.

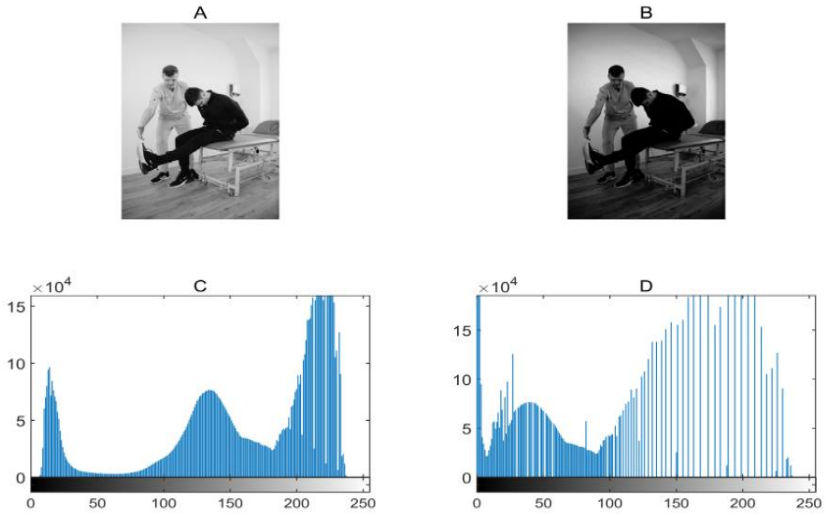

Figure 3: Adaptive Histogram Equalization. (Figure 3 (A): Original image; Figure 3 (B): Equalized image; Figure 3 (C): Histogram of original image; Figure 3 (D): Histogram of equalized image)

To ensure the accurate extraction of motion key points, Canny edge detection (Sekehravani et al., 2020) is used in the last step of preprocessing. Canny edge detection is a multi-step edge detection algorithm that detects the edge of an object by calculating the image gradient. A Gaussian filter is applied for noise smoothing, and the image gradient is calculated. The edge strength in the horizontal and vertical directions of the image is obtained, and the edge is refined and non-edge points are suppressed by non-maximum suppression (Guo et al., 2022) (NMS). The double threshold method (Yu et al., 2023) is used to retain strong edges and connect weak edges, as shown in Figure 4.

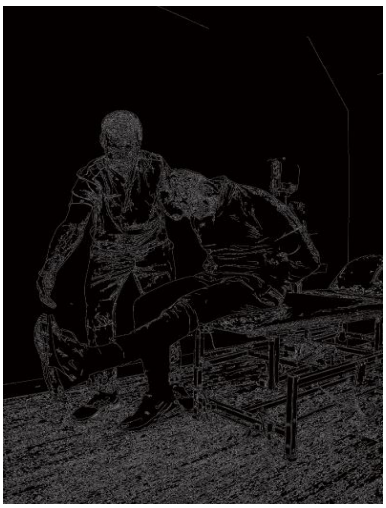

Figure 4: Canny Edge Detection Results

Edge detection further enhances the features of the human body contour during motion to ensure that the subsequent key point detection algorithm can accurately locate the position of each joint. The threshold of Canny edge detection is adjusted through multiple experiments to ensure that it can effectively detect moving bodies in different motion scenes.

2.3. Motion Key Point Detection

During the motion key point detection process, the Open Pose algorithm is used to analyze the patient's motion posture to obtain the precise coordinates of the joint points. Open Pose is a multi-body posture estimation method based on convolutional neural network (CNN) that can detect multiple joint positions in the image in real-time, including joints such as shoulders, elbows, and knees. The multi-stage convolutional network structure of Open Pose is used to process the input dynamic video. The input dynamic video is converted into a series of frame images and processed by the multi-stage convolutional network structure of Open Pose. Assuming the input image is I and the key point heatmap output by the network is H, each heat map represents a specific joint point k. The generation of the heatmap can be expressed as:

$$H_k(x, y) = \sigma(W_k * I(x, y) + b_k) \tag{7}$$

Among them, W_k is the convolution kernel; b_k is the bias term; σ is the activation function ReLU (Strong et al., 2023); $*$ represents the convolution operation. In this way, the model generates heatmaps for the main joints of the human body (such as shoulders, elbows, knees, etc.). In order to extract the key point coordinates from the heatmap, the maximum value method is used:

$$(p_x, p_y) = \underset{x,y}{\operatorname{argmax}}(H_k(x, y)) \tag{8}$$

In this formula, (p_x, p_y) represents the best position of the joint point k in the heatmap H_k. This process uses NMS technology to filter out redundant key points to ensure that the final output key point set is optimal. By setting a certain threshold T, only key points with confidence $H_k(x, y)$ greater than T are retained, thereby improving the stability and reliability of the detection results. In the key point detection stage, the heatmap generated by Open Pose contains the distribution of each joint point, indicating the confidence of the specific key point at that position. In order to enhance the detection accuracy, the new key point detection strategy is combined with the Alpha Pose technology (Fang et al., 2022) and the Feature Pyramid Network (FPN) (Zhou & Zhang, 2022) in deep learning, which effectively improves the detection ability of small objects and overlapping objects. The algorithm is expressed as follows through multi-scale feature fusion:

$$F_{pyramid}(x) = \sum_{i=1}^{N} \alpha_i \cdot F_i(x) \tag{9}$$

Among them, $F_i(x)$ is the feature map of the i-th layer; α_i is the corresponding weight coefficient; N is the number of feature layers. AlphaPose processes multiple input images in parallel and utilizes the multi-scale capability of deep feature extraction to enhance the ability to capture key points of multiple people in dynamic scenes, as shown in Figure 5.

Figure 5: Open Pose Key Point Detection

2.4. Exercise Intensity Calculation

During the exercise intensity calculation process, this system uses the random forest algorithm to classify and evaluate exercise intensity based on the detected exercise key point data and combined with individual physiological characteristics, aiming to precisely measure and quantify the exercise intensity in rehabilitation training and provide data support for personalized training programs. The whole process includes the displacement and velocity calculation of the exercise key points, feature extraction and physiological data fusion, and the training and optimization of the classification model, ultimately achieving accurate classification of exercise intensity. Exercise intensity calculation relies on quantitative analysis of the displacement and velocity information of the exercise key points. The key points detected by the Open Pose algorithm (such as shoulders, elbows, knees, etc.) generate a series of time series coordinate data. Assuming that the positions of the two key points k1 and k2 are (x1, y1) and (x2, y2), respectively, the displacement d between the key points can be expressed as:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{10}$$

Next, the velocity v is calculated based on the displacement d. The

velocity is defined as the displacement per unit time:

$$v = \frac{d}{\Delta t} \tag{11}$$

By calculating the displacement and velocity of each key point during the entire exercise process, a set of time series feature data containing the key point movement pattern is constructed. These data reflect the dynamic changes of each part during the exercise process and provide a basis for subsequent intensity calculation. To further improve the accuracy of exercise intensity calculation, the system not only considers the displacement and velocity information of the key points of the exercise, but also integrates physiological data, especially heart rate changes. The heart rate data HR obtained by the heart rate monitoring device is integrated with the displacement and velocity characteristics during exercise. The individual's resting heart rate is set to HR_{rest} and the maximum heart rate is set to HR_{max}. The heart rate reserve (Kelshiker et al., 2022)(HRR) during exercise is calculated using the following formula:

$$HRR = \frac{HR - HR_{rest}}{HR_{max} - HR_{rest}} \tag{12}$$

According to the characteristics of displacement, velocity, and HRR, the system constructs a multidimensional feature vector, including displacement features, velocity features, heart rate features, and time series features, to describe the characteristic performance under different exercise intensities. These features not only reflect the individual's exercise pattern, but also comprehensively consider the individual's physiological differences, thereby achieving personalized processing in exercise intensity assessment. After normalization, these features are input into the random forest model for classification. In this system, the random forest model is used to classify exercise intensity. Random forest is an algorithm based on ensemble learning, which has strong nonlinear fitting ability and anti-overfitting ability. The algorithm completes the classification task by constructing multiple decision trees, and generates the final strong classifier by integrating multiple weak classifiers to improve the accuracy and stability of the model. In the model training stage, the training data set is first labeled, and the exercise intensity categories are set to be mild, moderate, and severe, corresponding to different exercise intensity levels. The training data of each category contains the displacement, speed, heart rate and other characteristics of the key points of the exercise. The construction of the decision tree uses the Gini Index (Daniya et al., 2020) as the node splitting standard. The Gini Index is defined as:

$$Gini(p) = 1 - \sum_{i=1}^{n} p_i^2 \tag{13}$$

Among them, p_i is the probability of the class, and n is the number of

classification categories. The smaller the Gini index, the purer the classification. In the random forest training process, the system generates different training subsets through multiple resampling and constructs multiple decision trees. Each decision tree performs exercise intensity classification independently, and the final classification result is determined by a voting mechanism. This voting mechanism can reduce the bias caused by a single decision tree and improve the generalization ability of the model. Figure 6 depicts a decision tree model in a random forest for distinguishing three different activity intensity levels: light, moderate, and heavy. The decision tree consists of several nodes, each representing a feature and its corresponding threshold. There are two main features: speed (v) and heart rate reserve (HRR). Root node: The decision tree starts from the root node, and its condition is whether the speed (v) is less than or equal to 10.0. If this condition is met, all three samples are classified as light activity intensity with a Gini coefficient of 0.667. Left branch: The first branch from the root node is the case when the speed (v) is less than or equal to 10.0. The Gini coefficient of this branch is 0, which means that all samples belong to the same category - light activity intensity. Right branch: The second branch from the root node is the case when the speed is greater than 10.0. Next, the decision tree checks whether the heart rate reserve (HRR) is less than or equal to 55.0. If HRR is less than or equal to 55.0, then two samples are classified as moderate activity intensity and one sample is classified as light activity intensity, with a gini coefficient of 0.5. If HRR is greater than 55.0, then only one sample is classified as heavy activity intensity, and the Gini coefficient is also 0. The goal of a decision tree is to minimize the Gini coefficient and improve classification accuracy by recursively dividing the dataset into purer subsets. The decision tree uses two features, speed and heart rate reserve, to determine a person's activity intensity level.

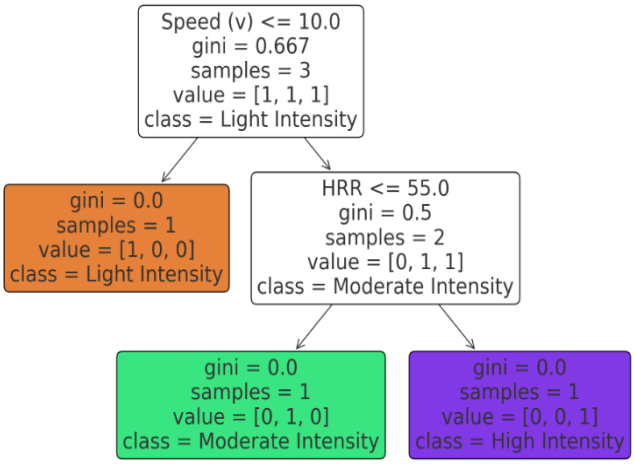

Figure 6: Decision Tree Structure

2.5. Effect Evaluation Model Construction

In the process of motion effect evaluation, this system uses the deep

learning model ResNet to classify and evaluate the motion effects in rehabilitation training, including data processing, model training, transfer learning application and implementation of optimization process. Through these technical means, the system can accurately distinguish the normal and abnormal motion patterns of patients in rehabilitation training, and further improve the accuracy of evaluation by combining time series models.

To build a motion effect evaluation model, this system chooses ResNet as the core architecture and applies residual blocks to solve the common gradient vanishing problem in deep neural networks, which can effectively improve the model depth and thus improve the recognition ability of complex motion patterns. The input motion feature vector is mapped to a high-dimensional space, and the spatiotemporal features in the motion pattern are extracted through multi-layer convolution operations. The output of each convolutional layer is transformed nonlinearly through the activation function ReLU (Rectified Linear Unit):

$$f(x) = max(0, x) \tag{14}$$

In each residual block of ResNet, the input directly skips multiple convolutional layers and is added to the output. This form of residual connection ensures that information can be smoothly transmitted in the network and solves the problem of information attenuation that may exist in deep networks.

$$y = F(x, \{W_i\}) + x \tag{15}$$

Among them, x is the input, and $F(x, \{W_i\})$ is the feature map calculated by the convolution kernel $\{W_i\}$. The residual connection effectively improves the training stability and convergence speed of the model. This system uses transfer learning technology to improve training efficiency and reduce the amount of data required. A pre-trained ResNet model is used for fine-tuning. The pre-trained model is trained on large-scale datasets such as ImageNet (Mettes et al., 2020). The model has rich general feature representation capabilities. Through transfer learning, the system only needs to perform a small amount of training on a dataset specific to rehabilitation training to adapt to the task of evaluating exercise effects. During the fine-tuning process, the parameters of the first few layers of the pre-trained model are fixed, and only the last few layers are retrained to adapt to the data distribution of the new task.

The specific steps are as follows: The pre-trained ResNet model is loaded and the weight parameters of the first few convolutional layers are frozen. The rehabilitation training data is used to perform backpropagation and gradient updates on the last few layers to learn feature representations specific to motion effect evaluation. The validation set is used to evaluate the model

performance and continuously adjust hyperparameters such as the learning rate to ensure a balance between model convergence speed and effect. Transfer learning greatly reduces the time of model training and effectively solves the problem of insufficient training data. Since the motion data in rehabilitation training has significant time dependence, relying solely on ResNet to classify the motion effect of a single frame may not be enough to capture the overall trend of motion changes.

This system integrates the Long Short-Term Memory Network (Van Houdt et al., 2020) (LSTM) with ResNet to enhance the processing capability of time series data. LSTM is a special recurrent neural network that solves the problem of information loss in long sequence data by applying memory units. Time series data is selectively memorized and updated through the forget gate, input gate, and output gate. Based on the features output by ResNet, the system passes the feature vectors of multiple frames into the LSTM network:

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b) \tag{16}$$

Among them, x_t is the current frame feature extracted by ResNet; h_t is the hidden state at the current moment; h_{t-1} is the hidden state at the previous moment. LSTM can effectively integrate historical information at the current moment and output the evaluation results of the motion effect.

3. Effect Evaluation Indicators

3.1. Exercise Intensity Evaluation Indicators

In rehabilitation training, precise evaluation of exercise intensity is particularly important for guiding the training process, adjusting training intensity, and ensuring patient safety(Gorle & Padala). This system uses MET (Leal-Martin et al., 2022) and heart rate changes as core indicators for exercise intensity evaluation. By combining these indicators, the system can quantify the patient's exercise intensity level in rehabilitation training and help trainers optimize rehabilitation plans.

As shown in Table 1, the table lists the types of exercise performed by patients in rehabilitation training and their corresponding metabolic equivalent of task (MET), heart rate reserve (HRR), and exercise intensity levels. MET is a key indicator for evaluating exercise intensity and indicates the level of energy consumption; HRR reflects the change in heart rate during exercise. Exercise intensity is divided into three levels: mild, moderate, and severe (mild is 30% to 50% HRR; moderate is 50% to 70% HRR; severe is more than 70% HRR). By monitoring these indicators, the patient's exercise effect can be effectively evaluated and a basis for personalized rehabilitation training can be provided.

Table 1: Patient Metabolic Equivalents and Exercise Intensity

PATIENT ID	EXERCISE TYPE	MET	HRR	EXERCISE INTENSITY
1	Walking	3.5	40%	Mild
2	Jogging	7	60%	Moderate
3	Cycling	6	55%	Moderate
4	Swimming	8	65%	Moderate
5	Yoga	2.5	30%	Mild
6	Aerobics	5	50%	Mild
7	High-Intensity Interval Training	9.5	80%	Severe
8	Climbing	6.5	70%	Moderate
9	Strength Training	4	45%	Mild
10	Jump Rope	10	85%	Severe

Figure 7 shows the heart rate monitoring of patients, showing a line graph of the heart rate changes of patients. The horizontal axis represents the recording time, and the vertical axis represents the heart rate (bpm). Each curve represents the heart rate of a patient at different time points. The heart rates of two patients fluctuate between 80 and 110 bpm, and their heart rates are relatively stable under moderate light exercise. The heart rates of three patients fluctuate between 110 and 150 bpm. In particular, patients whose heart rates are continuously above 110 bpm should pay attention to their exercise load and possible fatigue. The heart rate of one of them even remains above 130 bpm, and the relevant monitoring personnel should reduce the intensity of his training. If the patient's heart rate is continuously higher than 150 bpm or lower than 60 bpm, intervention measures should be taken immediately, and the exercise should be terminated to ensure that it is within the safe range and abnormal heart rate is discovered in time.

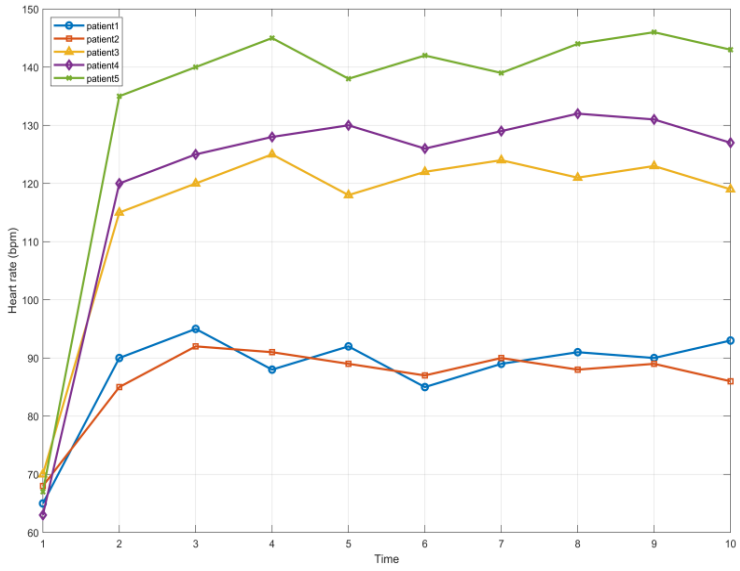

Figure 7: Patient Heart Rate Monitoring

3.2. Recovery of Joint Range of Motion

In the evaluation of the range of motion of joints in rehabilitation training, the main focus is on the range of motion (ROM), and the rehabilitation effect of the patient is evaluated by measuring the maximum motion angle of each joint. Table 2 shows the patient's joint motion data.

According to the criterion of significance level 0.01, the flexion of the hip, knee and ankle joints shows significant improvement, but the training effect of knee and ankle extension is relatively limited. The caregiver should adjust or strengthen the training method according to the evaluation results, especially in the extension training of the knee and ankle joints, appropriately increase the training intensity or apply new training methods.

Table 2: Patient's Joint Motion Range

JOINT	TYPE OF MOVEMENT	PRE-TRAINING RANGE OF MOTION (°)	POST-TRAINING RANGE OF MOTION (°)	CHANGE (°)	STATISTICAL TEST RESULT (P-VALUE)
HIP JOINT	Flexion	100	130	30	0.003
HIP JOINT	Extension	0	10	10	0.01
KNEE JOINT	Flexion	90	120	30	0.005
KNEE JOINT	Extension	10	15	5	0.02
ANKLE JOINT	Flexion	45	75	30	0.001
ANKLE JOINT	Extension	15	25	10	0.015

3.3. Gait Stability Evaluation Indicators

In the gait stability evaluation, it is necessary to obtain the dynamic video data of the patient during rehabilitation training, combine the key points of movement detected by Open Pose, and focus on extracting the gait-related joint coordinates (hip joint, knee joint and ankle joint) to form a series of time series data. For these time series data, gait stability is measured by calculating the standard deviation (SD) of the motion trajectory.

Table 3: Motion Stability within the Gait Cycle

GAIT CYCLE NUMBER	KEY POINTS (JOINT)	AVERAGE DISPLACEMENT (MM)	STANDARD DEVIATION (SD)	LOF	DTW (%)	MOTION STABILITY
1	Hip Joint	32.5	2.3	1.05	92	88.5
1	Knee Joint	28.7	1.8	0.98	95	91
1	Ankle Joint	36.4	2.6	1.1	90	87
2	Hip Joint	31.8	2.1	1.08	93	89.5
2	Knee Joint	29	1.7	1.02	94	91.5
2	Ankle Joint	35.9	2.5	1.12	91	87.5
3	Hip Joint	34	2.5	1.15	89	85.5
3	Knee Joint	30.2	2	1.1	92	89
3	Ankle Joint	38	2.7	1.18	87	84.5

The gait cycle stability is calculated based on the joint motion range of the previous patient. From the data in Table 3, it can be seen that there are differences in the motion stability of each key joint in different gait cycles. The average displacement of the patient's knee joint is small and the standard deviation is low, with relatively stable gait; the standard deviation of the hip and ankle joints is slightly higher. Especially in the third gait cycle, the LOF (Local Outlier Factor) value of the ankle joint reaches 1.18, indicating that there are large motion fluctuations in this gait cycle, and unstable gait movements may have occurred. The DTW (Dynamic Time Warping) matching degree shows the matching degree of the hip, knee and ankle joints in each gait cycle. The matching degree of the knee joint is generally high, indicating that its motion trajectory is closer to the standard gait. The DTW matching degree of the ankle joint is low, and its matching degree in the third cycle is only 87%, which further confirms that the gait stability in this gait cycle is poor and needs to be focused on in rehabilitation training. Through the analysis of these data, there is a clearer understanding of the stability of the patient's gait and provide data support for personalized rehabilitation training.

3.4. Evaluation Indicators of Cycle Training Results

This experiment is divided into seven training cycles, each cycle representing stage of rehabilitation training for patients.

As the cycle increases, the effect of rehabilitation training gradually improves(González Martínez et al., 2023). As shown in Figure 8, the horizontal axis is the cycle and the vertical axis is the value. The root mean square error (RMSE) reflects the patient's movement accuracy during the training process. The RMSE of cycle 1 is 12.6%. As the training progresses, the RMSE error gradually decreases, and it drops to 4.9% in cycle 7. The patient's movement stability is significantly improved during the rehabilitation process, and the accuracy of movement execution is improved. The patient's abnormal movement is detected, and the proportion of abnormal movement detected is expressed in percentage. In cycle 1, the proportion of abnormal movement detection is 45.60%, and by cycle 7, the proportion drops to 10.30%. After multiple trainings, the patient's movement pattern gradually tends to be normal, and the incidence of abnormal movement is significantly reduced, showing an improvement in rehabilitation effect. Through the combination of features of multiple modalities, the fusion result of cycle 1 is 0.48, and the final result is improved to 0.92. The value range is from 0 to 1, and the closer to 1, the better the rehabilitation effect. Combining the above evaluation results to obtain a percentage value, the recovery degree in cycle 1 is 52.1%. After systematic training, the recovery degree increases to 93.2% in cycle 7. From the above data, it can be seen that after systematic rehabilitation training, the patient's motor ability and stability are significantly improved; the incidence of abnormal movement is significantly reduced; the overall recovery degree is greatly improved. These results support the effectiveness of using computer imaging technology to monitor rehabilitation training, indicating that the monitoring system can provide a reliable quantitative evaluation basis for clinical rehabilitation.

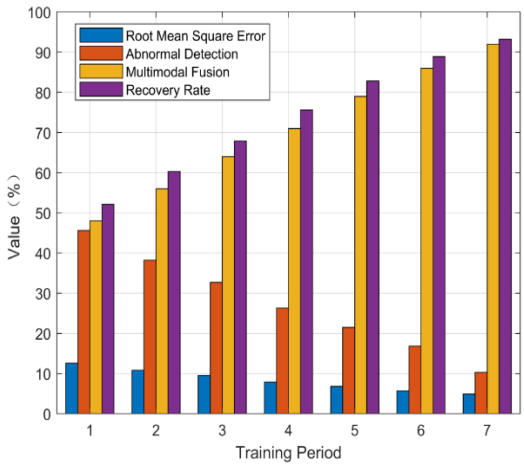

Figure 8: Cycle Training Results

4. Conclusions

This paper applies computer image technology, combines OpenPose and ResNet models, and builds a rehabilitation training monitoring system to

monitor the patient's exercise intensity and effect in real-time during training. The system integrates a high-definition camera to capture dynamic video and uses the OpenCV library for image preprocessing to remove noise and improve image stability; key points of movement are extracted through the OpenPose algorithm, and the random forest classification algorithm is combined to achieve personalized evaluation of exercise intensity. The ResNet model is used to classify the exercise effect, and the evaluation accuracy is improved through transfer learning. The processing of time series data is optimized by combining the LSTM model. In the experiment, the system successfully achieves efficient and accurate exercise intensity monitoring and exercise effect evaluation, solving the problems of strong subjectivity and inaccurate data in traditional manual monitoring. By analyzing the movement trajectory and movement frequency, the system provides more detailed feedback, which helps rehabilitation trainers to adjust their training plans in time. The method in this paper still has certain limitations. For example, the accuracy of the system may decrease in environments with large lighting changes or complex backgrounds; the training time of the model is long, and there are certain performance bottlenecks when processing multi-dimensional data. Future research directions include further optimizing the real-time performance of the system, improving its adaptability to complex environments, and exploring how to combine virtual reality technology to provide a more interactive rehabilitation training experience, which can help provide patients with more intelligent and personalized rehabilitation training programs.

REFERENCES

Ageed, Z. S., Zeebaree, S., Sadeeq, M., Abdulrazzaq, M. B., Salim, B. W., Salih, A. A., Yasin, H. M., & Ahmed, A. M. (2021). A state of art survey for intelligent energy monitoring systems. *Asian Journal of Research in Computer Science*, *8*(1), 46-61.

Boateng, E. Y., Otoo, J., & Abaye, D. A. (2020). Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review. *Journal of Data Analysis and Information Processing*, *8*(4), 341-357.

Cao, X., Yao, J., Xu, Z., & Meng, D. (2020). Hyperspectral image classification with convolutional neural network and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(7), 4604-4616.

D'Antonio, E., Taborri, J., Mileti, I., Rossi, S., & Patané, F. (2021). Validation of a 3D markerless system for gait analysis based on OpenPose and two RGB webcams. *IEEE Sensors Journal*, *21*(15), 17064-17075.

Dalal, S., Lilhore, U. K., Manoharan, P., Rani, U., Dahan, F., Hajjej, F., Keshta, I., Sharma, A., Simaiya, S., & Raahemifar, K. (2023). An Efficient Brain Tumor Segmentation Method Based on Adaptive Moving Self-Organizing Map and Fuzzy K-Mean Clustering. *Sensors*, *23*(18), 7816.

Daniya, T., Geetha, M., & Kumar, K. S. (2020). Classification and regression

trees with gini index. *Advances in Mathematics: Scientific Journal*, *9*(10), 8237-8247.

Fang, H.-S., Li, J., Tang, H., Xu, C., Zhu, H., Xiu, Y., Li, Y.-L., & Lu, C. (2022). Alphapose: Whole-body regional multi-person pose estimation and tracking in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(6), 7157-7173.

Gao, Z., Zhang, Y., Zhang, H., Guan, W., Feng, D., & Chen, S. (2021). Multi-level view associative convolution network for view-based 3D model retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(4), 2264-2278.

González Martínez, A., Martínez Rodríguez, A., & Águila, G. (2023). Mechanic-mathematical model of javelin flight. *Rev. int. med. cienc. act. fis. deporte*, 247-270.

Gorle, D. L., & Padala, A. The Impact of COVID-19 Deaths, Medical Analysis & Visualization Using Plotly. *International journal of health sciences*, *6*(S3), 11957-11971.

Guo, C., Cai, M., Ying, N., Chen, H., Zhang, J., & Zhou, D. (2022). ANMS: attention-based non-maximum suppression. *Multimedia Tools and Applications*, *81*(8), 11205-11219.

Huang, B., Tang, Y., Ozdemir, S., & Ling, H. (2020). A fast and flexible projector-camera calibration system. *IEEE Transactions on Automation Science and Engineering*, *18*(3), 1049-1063.

Kelshiker, M. A., Seligman, H., Howard, J. P., Rahman, H., Foley, M., Nowbar, A. N., Rajkumar, C. A., Shun-Shin, M. J., Ahmad, Y., & Sen, S. (2022). Coronary flow reserve and cardiovascular outcomes: a systematic review and meta-analysis. *European heart journal*, *43*(16), 1582-1593.

Leal-Martin, J., Munoz-Munoz, M., Keadle, S. K., Amaro-Gahete, F., Alegre, L. M., Mañas, A., & Ara, I. (2022). Resting oxygen uptake value of 1 metabolic equivalent of task in older adults: a systematic review and descriptive analysis. *Sports Medicine*, *52*(2), 331-348.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, *33*(12), 6999-7019.

Liu, Y.-Q., Du, X., Shen, H.-L., & Chen, S.-J. (2020). Estimating generalized gaussian blur kernels for out-of-focus image deblurring. *IEEE Transactions on Circuits and Systems for Video Technology*, *31*(3), 829-843.

Livneh, H. (2022). Psychosocial adaptation to chronic illness and disability: An updated and expanded conceptual framework. *Rehabilitation Counseling Bulletin*, *65*(3), 171-184.

Mao, W., Liu, M., Salzmann, M., & Li, H. (2021). Multi-level motion attention for human motion prediction. *International journal of computer vision*, *129*(9), 2513-2535.

Mehdizadeh, M., Tavakoli Tafti, K., & Soltani, P. (2023). Evaluation of histogram equalization and contrast limited adaptive histogram equalization effect on image quality and fractal dimensions of digital periapical radiographs. *Oral Radiology*, *39*(2), 418-424.

Mettes, P., Koelma, D. C., & Snoek, C. G. (2020). Shuffled imagenet banks for video event detection and search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *16*(2), 1-21.

Nascimento, L. M. S. d., Bonfati, L. V., Freitas, M. L. B., Mendes Junior, J. J. A., Siqueira, H. V., & Stevan Jr, S. L. (2020). Sensors and systems for physical rehabilitation and health monitoring—A review. *Sensors*, *20*(15), 4063.

Postolache, O., Hemanth, D. J., Alexandre, R., Gupta, D., Geman, O., & Khanna, A. (2020). Remote monitoring of physical rehabilitation of stroke patients using IoT and virtual reality. *IEEE Journal on Selected Areas in Communications*, *39*(2), 562-573.

Rutkowski, S., Kiper, P., Cacciante, L., Cieslik, B., Mazurek, J., Turolla, A., & Szczepanska-Gieracha, J. (2020). Use of virtual reality-based training in different fields of rehabilitation: A systematic review and meta-analysis. *Journal of Rehabilitation Medicine*, *52*(11), 1-16.

Samy, H., Tammam, A., Fahmy, A., & Hasan, B. (2021). Enhancing the performance of the blockchain consensus algorithm using multithreading technology. *Ain Shams Engineering Journal*, *12*(3), 2709-2716.

Sejnowski, T. J. (2020). The unreasonable effectiveness of deep learning in artificial intelligence. *Proceedings of the National Academy of Sciences*, *117*(48), 30033-30038.

Sekehravani, E. A., Babulak, E., & Masoodi, M. (2020). Implementing canny edge detection algorithm for noisy image. *Bulletin of Electrical Engineering and Informatics*, *9*(4), 1404-1410.

Singh, P., Mukundan, R., & De Ryke, R. (2020). Feature enhancement in medical ultrasound videos using contrast-limited adaptive histogram equalization. *Journal of digital imaging*, *33*(1), 273-285.

Strong, C. A., Wu, H., Zeljić, A., Julian, K. D., Katz, G., Barrett, C., & Kochenderfer, M. J. (2023). Global optimization of objective functions represented by ReLU networks. *Machine Learning*, *112*(10), 3685-3712.

Suchomel, T. J., Nimphius, S., Bellon, C. R., Hornsby, W. G., & Stone, M. H. (2021). Training for muscular strength: Methods for monitoring and adjusting training intensity. *Sports Medicine*, *51*(10), 2051-2066.

Susim, T., & Darujati, C. (2021). Pengolahan citra untuk pengenalan wajah (face recognition) menggunakan opencv. *Jurnal Syntax Admiration*, *2*(3), 534-545.

Tripathi, M. (2021). Analysis of convolutional neural network based image classification techniques. *Journal of Innovative Image Processing (JIIP)*, *3*(02), 100-117.

Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-

term memory model. *Artificial Intelligence Review*, *53*(8), 5929-5955.

Wang, M., Lin, Y., Tian, Q., & Si, G. (2021). Transfer learning promotes 6G wireless communications: Recent advances and future challenges. *IEEE Transactions on Reliability*, *70*(2), 790-807.

Wen, L., Li, X., & Gao, L. (2020). A transfer convolutional neural network for fault diagnosis based on ResNet-50. *Neural Computing and Applications*, *32*(10), 6111-6124.

Wu, Z., Chen, Y., Zhao, B., Kang, X., & Ding, Y. (2021). Review of weed detection methods based on computer vision. *Sensors*, *21*(11), 3647.

Yu, F., Cao, J., & Shi, Y. (2023). Double-threshold effect of technological innovation on environmental-responsibility fulfillment: Evidence from high-polluting SMEs in China. *Journal of Small Business Management*, *61*(4), 1871-1895.

Zhao, Y., Jiang, J., Chen, Y., Liu, R., Yang, Y., Xue, X., & Chen, S. (2022). Metaverse: Perspectives from graphics, interactions and visualization. *Visual Informatics*, *6*(1), 56-67.

Zhou, X., & Zhang, L. (2022). SA-FPN: An effective feature pyramid network for crowded human detection. *Applied Intelligence*, *52*(11), 12556-12568.

Zhu, Y.-C., Jin, P.-F., Bao, J., Jiang, Q., & Wang, X. (2021). Thyroid ultrasound image classification using a convolutional neural network. *Annals of translational medicine*, *9*(20).