Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105
Volume **10**, Number **3**, August **2025**

Swimming Pose Recognition Based on Inertial Sensor and CNN-SVM

Jian-Bin Du*

Ningbo Tech University
Ningbo 315100, P. R. China
djb0072024@163.com

Li-Chao Wei

Ningbo Tech University
Ningbo 315100, P. R. China
912672740@qq.com

Jin-Hui Zhu

Payap University
Chiang Mai 50000, Thailand
sf6906@163.com

*Corresponding author: Jian-Bin Du
Received May 11, 2024, revised September 7, 2024, accepted January 2, 2025.

Abstract. *Traditional swimming movement monitoring methods mainly rely on the subjective judgement of coaches and video playback, which limits the real-time assessment and analysis of swimmer movement details. In order to improve the accuracy and efficiency of swimming motion recognition, this paper proposes a swimming pose recognition method based on inertial sensors and Convolutional Neural Network-Support Vector Machines (CNN-SVM), which aims to improve the efficiency of motion analysis in swimming training and competitions. While traditional swimming motion capture relies on expensive optical systems or wearable devices, this study utilises low-cost Inertial Measurement Unit (IMU) sensors to capture and analyse swimmers' movements with high accuracy through data fusion algorithms and optimisation techniques. In the study, firstly, a mathematical model of swimming action was established and the swimmer's motion data was captured by the IMU sensor. Subsequently, Kalman filter and time alignment techniques were used to preprocess the data to improve the accuracy of the pose estimation. On this basis, CNN was used to automatically extract the motion features and classify them by SVM, and a combined CNN-SVM model was constructed to achieve the automatic identification of different swimming postures. The experimental results show that the model achieves an F1 score of more than 0.9 for the recognition of all four swimming strokes: freestyle, butterfly, breaststroke and backstroke, which verifies the effectiveness and generalisation ability of the proposed method. The method in this study not only improves the efficiency of swimming stroke analysis, but also provides a new technical tool for sports science and personal health management.*

Keywords: Swimming pose recognition; Inertial sensors; CNN-SVM; Motion capture

1. **Introduction.** With the development of sports science and personal health management, there is an increasing demand for the ability to remotely monitor, store large amounts of data and perform sophisticated analyses [1, 2, 3]. Swimming, as a whole-body

sport, the capture and analysis of its movements is important for athlete training, skill improvement, and health monitoring. However, traditional swimming motion capture methods rely on expensive optical systems [4] or manual recording [5], which have limitations such as high cost, poor real-time performance, and susceptibility to environmental influences. Therefore, the development of a low-cost sensor-based swimming pose recognition method that can effectively capture and analyse swimmers' movements is valuable for improving training efficiency, optimising athletes' performance as well as promoting scientific research on swimming.

The background of the study of inertial sensors in swimming posture recognition stems from the need for efficient and accurate monitoring of swimming sports. Swimming, as a sports activity with exercise effects on all muscle groups of the whole body, occupies an important position in competitive sports and public fitness. In order to enhance the training effect of athletes [6], optimise technical movements [7] and prevent sports injuries [8], it is particularly crucial to accurately identify and analyse swimming postures. While traditional video-based monitoring methods have limitations such as high cost, poor real-time performance, and susceptibility to environmental influences, inertial sensor-based pose recognition technology has become a hot research topic due to its portability, real-time performance, and ability to be used underwater. Inertial sensors are able to capture the acceleration and angular velocity changes generated during swimming, and the body posture and movement pattern of the swimmer can be inferred from these data. With the development of sensor technology and the advancement of data processing algorithms, the application of inertial sensors in the field of swimming posture recognition provides new technical means for swimming training and research, and helps to achieve a deeper understanding and analysis of swimming movements.

Convolutional Neural Networks (CNN) and Support Vector Machines (SVM), as the two main algorithms of deep learning, have demonstrated excellent performance in several fields. Therefore, the aim of this study is to apply these algorithms to swimming posture recognition, aiming to improve the accuracy and efficiency of swimming posture recognition through end-to-end co-training by taking advantage of the powerful feature extraction capability of CNN and the excellent classification performance of SVM.

1.1. **Related work.** The current state of the art in sports pose recognition is reflected in the capturing and analysing of sports movements through a variety of technological means in order to improve the training efficiency and technical level of athletes. Currently, researchers mainly use high-speed camera systems, Inertial Measurement Units (IMUs) and deep learning-based algorithms to identify and analyse sports postures. While high-speed camera systems can provide detailed visual information, they are costly and not easily portable. IMU sensors have become a hot research topic due to their low cost, portability, and ability to be used underwater. They are able to capture dynamic data during swimming, including acceleration and angular velocity, to infer changes in the swimmer's posture.

Stamm and Thiel [11] used two triaxial accelerometers and a Quadratic Discriminant Analysis (QDA) model for freestyle swimming. However, the model is simple and fails to recognise butterfly strokes and does not address the details of specific swimming movements. Kos and Umek [12] used smart accelerometers to recognise swimmers' strokes and stroke frequency in real time. Rusdiana et al. [13] used built-in accelerometers in an Android device and an SVM classification model to achieve autonomous recognition of multiple swimming strokes. However, the comprehensive recognition of movement details is insufficient, especially in the recognition of key features such as turning movements, movement switching points, and movement durations.

In addition, deep learning algorithms, especially CNNs [14, 15, 16] and Support Vector Machines (SVMs) [17, 18], have been used to improve the accuracy and efficiency of motion pose recognition as they show strong performance in classification and recognition of image and sensor data. Iloga et al. [19] proposed a hybrid CNN and Hidden Markov Model (HMM) approach for human movement recognition, aiming to address the poor detection accuracy of current human movement recognition algorithms and the diversity of experimental scenarios. Although this hybrid model may improve recognition accuracy, it may suffer from high computational complexity and long training time. In addition, the generalisation ability of the model for different actions and scenarios needs further validation. Abdelbaky and Aly [20] proposed a human pose recognition method combining keypoint affinity fields and SVM, which solves the problem of traditional action pose recognition relying on physical data acquisition devices. However, the possible challenges of the method include high accuracy requirements for keypoint detection and robustness issues in complex background or occlusion situations. Vrskova et al. [21] proposed a 3D-CNN-based approach for human behaviour recognition, which allows features to be extracted directly from the raw input and in both temporal and spatial dimensions by 3D convolution. However, 3D CNN models may require a large amount of computational resources and data for training and may suffer from latency issues for real-time applications.

1.2. **Motivation and contribution.** These studies have shown that single sensors or algorithms may perform well on specific types of swimming strokes, but may not have sufficient generalisation capabilities when faced with different swimming strokes or individualised movements of different athletes. In addition, a single sensor or a simple feature extraction technique, which may not be accurate enough when facing complex swimming manoeuvres and variable underwater environments. To address the above issues, this paper improves the recognition accuracy and generalisation ability of swimming postures by combining the use of inertial sensors (which can provide rich dynamic information about acceleration, angular velocity) and CNN-SVM.

The main innovations and contributions of this work include:

(1) A method of action feature extraction based on multimodal sensing data is proposed, which not only considers traditional parameters such as acceleration and angular velocity, but also extracts richer action features such as joint angles, angular velocities, linear velocities and accelerations through time windowing processing and inverse kinematics problem solving, providing more detailed and comprehensive data support for swimming posture recognition. Special attention is paid to key steps such as data preprocessing, noise filtering, time alignment, and normalisation processing to improve data quality and model accuracy. In addition, an accurate estimation of the swimmer's real-time posture is achieved by a dynamic posture updating algorithm.

(2) The dynamic data collected by the IMU is innovatively combined with CNN-SVM for automatic swimming posture recognition. An end-to-end swimming posture recognition system is constructed by capturing the swimmer's motion data through the IMU, using CNN for deep feature extraction, and then SVM for efficient classification and recognition.

2. Inertial Sensor Based Swimming Motion Capture.

2.1. **The basic principle of human motion capture technology.** Human motion capture technology is a kind of technology that uses sensors, cameras and other devices to capture and record the trajectory of human movement. Its basic principle is to collect the data information generated when the human body moves, such as joint angle, speed,

acceleration, etc., and then use computer algorithms to analyse and process these data, and finally generate a model or animation of the human body's movement [22]. This technology can be widely used in film and television production, sports training, medical rehabilitation and other fields and can accurately capture and reproduce a variety of complex human body movement, providing people with a more intuitive and realistic sports experience and training effect.

The human body can be considered as consisting of multiple rigid bodies (limbs), each connected to the neighbouring rigid body through joints. Kinematic models usually use the following mathematical expressions to describe the motion of the limbs [23]:

$$T_i = T_{i-1} \cdot R(\theta_i) \cdot S(s_i) \tag{1}$$

where T_i denotes the transformation matrix of the i-th limb, T_{i-1} is the transformation matrix of its neighbouring limbs, $R(\theta_i)$ is the rotation matrix with respect to the angle of the joint θ_i, and $S(s_i)$ is the scaling matrix describing the length s_i of the limb.

Motion capture systems typically use multiple sensors, such as accelerometers, gyroscopes, and magnetometers, to capture the motion of the same limb. Data fusion algorithms, such as Kalman filters [24], can be used to integrate the data from these sensors and improve the accuracy of pose estimation.

$$x_{k|k} = F_k x_{k-1|k-1} + B_k u_k + w_k \tag{2}$$

$$P_{k|k} = F_k P_{k-1|k-1} F_k^T + Q_k \tag{3}$$

$$K_k = P_{k|k-1} H_k^T \left(H_k P_{k|k-1} H_k^T + R_k \right)^{-1} \tag{4}$$

$$x_{k|k} = x_{k|k-1} + K_k \left(z_k - H_k x_{k|k-1} \right) \tag{5}$$

$$P_{k|k} = \left(I - K_k H_k \right) P_{k|k-1} \tag{6}$$

where $x_{k|k}$ denotes the state estimate at time step k, z_k is the observation, K_k is the Kalman gain, $P_{k|k}$ is the estimated covariance matrix, and H_k is the observation matrix. $\mathbf{F}, \mathbf{B}, \mathbf{H}, \mathbf{Q}$, and $\mathbf{R}$ are the state transfer matrix, control input matrix, observation matrix, process noise covariance matrix, and observation noise covariance matrix, respectively.

To further improve the accuracy of pose estimation, optimisation algorithms can be used to minimise the discrepancy between measured data and model predictions [25]. A commonly used cost function can be expressed as

$$E(\mathbf{q}) = \frac{1}{2} \sum_{i=1}^{N} \left\| y_i - h(\mathbf{q}, x_i) \right\|^2 \tag{7}$$

where $E(\mathbf{q})$ is the cost function, $\mathbf{q}$ is the joint parameter to be optimised, y_i is the observation of the i-th sensor, $h(\mathbf{q}, x_i)$ is the model prediction for a given set of joint parameter and model predictions given the joint parameters and sensor positions, and N is the total number of sensor data points.

By minimising the cost function $E(\mathbf{q})$, the joint parameters $\mathbf{q}$ can be adjusted to best fit the observed data, resulting in a more accurate estimate of the human body posture.

2.2. **Definition of swimming manoeuvres.**

2.2.1. *Action modelling.* The definition of swimming action is the basis of swimming pose recognition research. In the field of machine learning and artificial intelligence, a well-defined swimming action model is crucial for designing effective feature extraction algorithms and recognition models. Swimming actions can be described by building a multi-stage action model, where each stage represents a specific part of the swimming cycle. For example, a simple swimming action model can include the following stages: (1) Starting position: the position in which the swimmer is ready to start paddling. (2) Grabbing the water: the movement in which the arm enters the water and begins to stroke backwards. (3) Pulling the water: the arm continues to stroke and propel the body forward. (4) Recovery phase: a movement in which the arms come out of the water and return to the starting position.

Each stage can be described by a specific set of joint angles and velocities. These parameters can be obtained from sensor data or directly measured by motion capture systems [26]. In order to mathematically define the swimming action, we can use the following expression:

Sequence of joint angles: $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$, where θ_i is the angle of the i-th joint at the particular point in time.

Velocity sequence: $\mathbf{V} = \{v_1, v_2, \ldots, v_n\}$, where v_i is the velocity of the i-th joint at the particular time point.

Acceleration sequence: $\mathbf{A} = \{a_1, a_2, \ldots, a_n\}$, where a_i is the acceleration of the i-th joint at a particular point in time.

Action phases: each action phase can be defined as a state vector S_t which contains information about the angles, velocities and accelerations of all joints at time point t.

Action Cycle: the entire swimming action can be defined as a series of state vectors $\mathbf{S} = \{S_1, S_2, \ldots, S_m\}$, where m is the number of state vectors in the action cycle.

2.2.2. *Extraction of action features.* In order to recognise different swimming actions, features need to be extracted from the action model defined above. These features should be able to reflect the key attributes of the swimming action, such as the amplitude, speed, rhythm and smoothness of the action.

The average joint angle is defined as shown below:

$$\bar{\theta} = \frac{1}{n} \sum_{i=1}^{n} \theta_i \tag{8}$$

The range of joint angle variation is $\Delta\theta_i = \max(\theta_i) - \min(\theta_i)$. The smoothness of the action can be measured by the consistency of the first order derivatives of the joint angles (velocities). The energy of the movement can be estimated by integrating over the sequence of velocities $\mathbf{v}$. The symmetry of the action can be assessed by comparing the angles of the same joints on both sides of the body. With these definitions and mathematical expressions, we can construct a comprehensive swimming action model that provides a solid foundation for subsequent data acquisition, posture estimation and action recognition.

2.3. **Data acquisition device design.** In order to accurately capture the swimmer's movement data, this study adopts an advanced inertial measurement unit (IMU) as the core device for data acquisition. The IMU integrates a three-axis accelerometer, a three-axis gyroscope, and a three-axis magnetometer, which is capable of comprehensively capturing the dynamic changes in the swimming movement. All sensors are waterproofed to meet the special requirements of the underwater environment. The structural composition of the sensor assembly is shown in Figure 1.

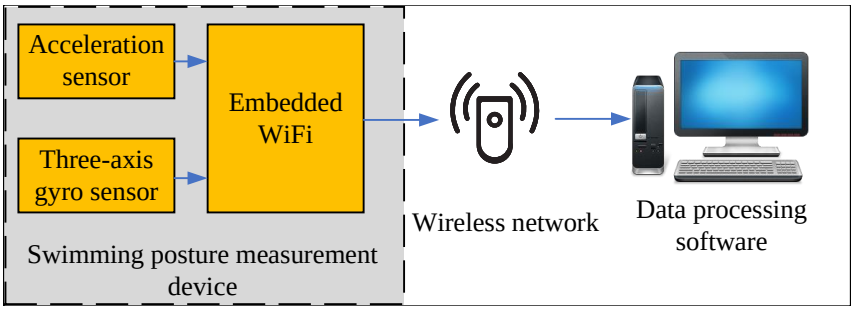

Figure 1. Swimming Pose Measurement Sensor Assembly Structure

Considering the full-body nature of swimming, this study chose to fix the IMU at the swimmer's waist, a position that balances the comprehensiveness of the captured movement information with the comfort of wearing the device. A special waterproof belt was used to securely fit the IMU to the swimmer's body, ensuring that the device would not shift or fall off during strenuous movements such as high-speed swimming or turning. The measurement device is worn schematically as shown in Figure 2.

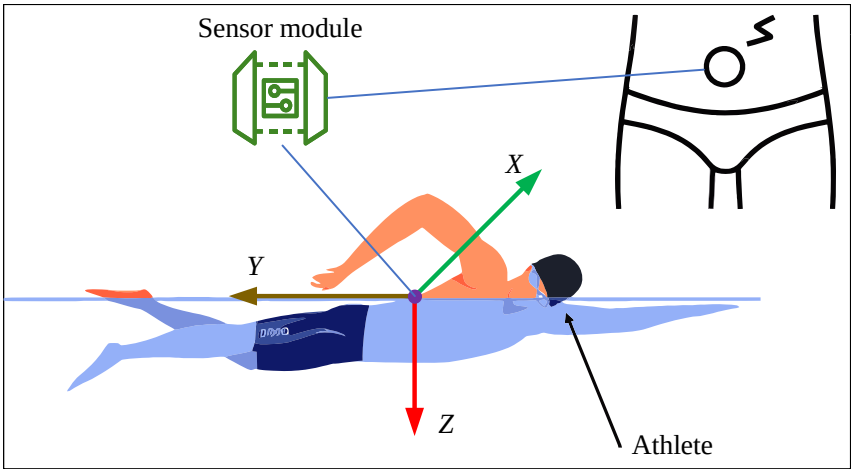

Figure 2. Schematic representation of the wearing of a measuring device

The acquired raw IMU data need to be preprocessed before they can be used for subsequent analyses. The pre-processing steps include noise removal, filtering and normalisation. In particular, in order to eliminate the problem of magnetometer data being highly affected by the swimming pool environment, this study mainly relies on accelerometer and gyroscope data for action recognition. With the design of the data acquisition device described above, this study is able to efficiently and accurately collect swimming action data for the subsequent implementation of inertial sensor and CNN-SVM based swimming pose recognition.

2.4. **Multimodal swimming action pose estimation.**

2.4.1. *Data preprocessing.* Data preprocessing is a key step in swimming manoeuvre pose estimation, which involves extracting useful information from the raw data collected by the IMU and converting it into a format suitable for performing pose estimation.

Firstly, the data collected by the IMU needs to be filtered for noise. Due to the special characteristics of the underwater environment, the sensor data may be affected by various noises. We employed a Kalman filter to filter the acceleration and angular velocity data in real time to remove the noise and improve the signal-to-noise ratio of the data. The specific

methods of the state update method of the Kalman filter are shown in Equation (**??**) and Equation (**??**).

Due to the dynamic nature of swimming movements, data alignment is necessary. We synchronise the IMU data and video footage by time-only to ensure that the data at each moment corresponds to the actual action of the swimmer. The precise alignment of TimeDiva is the basis for subsequent action feature extraction and analysis. The method of data alignment is shown as follows:

$$t' = t + \Delta t \tag{9}$$

where t' is the aligned timestamp, t is the original time, and Δt is the corrected offset.

In order to eliminate the effect of physiological differences between different swimmers on posture estimation, we normalised the data. By scaling the data to zero mean and unit variance, we ensured that the data from different swimmers had the same scale, which improved the generalisation of the algorithm [27].

$$x' = \frac{x - \mu}{\sigma} \tag{10}$$

where x' is the normalised data, x is the original data, μ is the mean of the data, and σ is the standard deviation of the data.

In order to capture the dynamic characteristics of the swimming manoeuvre, we split the continuous sensor data into fixed-length time windows. Each window contains acceleration and angular velocity data for a certain time period, which will be used for subsequent pose feature extraction. The windowing process is shown as follows:

$$W_n = \{x_t, x_{t+1}, \ldots, x_{t+w-1}\} \tag{11}$$

where W_n is the data set in the n-th window, x_t is the data at time step t, and w is the width of the window.

2.4.2. *Human posture calibration.* Human body posture calibration is the process of mapping the data collected by the IMU to the actual human body movement posture. This step is crucial to ensure the accuracy of swimming movement recognition. Posture calibration involves the understanding of the human body model, the limitations of joint kinematics, and the real-time updating of dynamic postures.

Firstly, a human kinematic model is required, which consists of multiple rigid bodies (limbs) and joints. The motion of each limb can be described by the rotation of the joints. For each joint, we can use a rotation matrix or quaternion to represent its rotation with respect to the previous limb. For example, if R_i denotes the rotation matrix from limb i to limb $i-1$, then the overall transformation matrix T can be represented by the product of multiple rotation matrices:

$$T = T_0 R_1 T_1 R_2 \ldots T_{n-1} R_n \tag{12}$$

where T_i is the transformation matrix of the limb, which contains translation and rotation information.

With the IMU data, we can estimate the relative position and pose of each joint. Assuming p_i is the node position of limb i and q_i is a quaternion describing the rotation of the limb i with respect to limb $i-1$, the node positional pose P_i is calculated as:

$$P_i = T_{i-1} q_i p_i \tag{13}$$

Joint movements in the human body are biologically constrained, e.g., limited range of motion in the shoulder and hip joints. These limitations can be realised by an Inverse Kinematics (IK) problem, i.e., solving for the joint angles given the position of the end effectors (e.g., hand or foot). The inverse kinematics problem can be solved by the following optimisation problem:

$$\theta^* = \arg\min_{\theta} \left\{ \|A(\theta) - d\|^2 + \lambda g(\theta) \right\} \tag{14}$$

where $A(\theta)$ is the end position calculated from the joint angle, d is the observed end effector position, $g(\theta)$ is the joint kinematic constraint function, and λ is the regularisation parameter.

Swimming manoeuvres are dynamic and therefore require real-time updating of the human posture. This can be achieved by an iterative algorithm that continuously receives new data from the IMU and updates the pose estimate. The dynamic pose update can be expressed as:

$$\theta_{k+1} = \theta_k + \Delta\theta \tag{15}$$

where θ_k is the pose estimate for the current time step, and $\Delta\theta$ is the amount of pose update.

3. CNN-SVM based swimming pose recognition.

3.1. Principle of CNN.

Convolutional Neural Network (CNN) is a deep learning model that performs well in image and video recognition tasks. The core idea of CNN is to use convolutional layers to automatically learn hierarchical features of the input data. The convolutional layer extracts local features from the input data by means of a series of learnable filters (or convolution kernels). Each filter slides over the entire input and computes the dot product of the local region and the filter to generate a feature map. Mathematically, the convolution operation can be expressed as:

$$(f * g)(i) = \sum_m \sum_n f(m, n) \cdot g(i - m, i - n) \tag{16}$$

where f is the input image, g is the convolution kernel, $*$ denotes the convolution operation, and i is the position of the output feature map.

The output is typically passed through a nonlinear activation function, such as ReLU:

$$\text{ReLU}(x) = \max(0, x) \tag{17}$$

Pooling layers are then applied to reduce the spatial dimensionality and improve generalisation.

3.2. Principle of SVM.

Support Vector Machine (SVM) is a supervised learning algorithm that excels in small sample learning, nonlinear problems, and high dimensional data processing. The core idea of SVM is to find the optimal decision boundary that maximises the margin between samples. For linearly divisible data, the hyperplane can be expressed as:

$$w \cdot x + b = 0 \tag{18}$$

where w is the normal vector to the hyperplane and b is the bias term.

The support vectors are the closest sample points to the decision boundary, and they determine the location of the hyperplane, as shown in Figure 3. For non-linearly differentiable data, SVM maps the original feature space to a high-dimensional space by

means of a kernel function, making the data linearly differentiable in the new space. Commonly used kernel functions include linear kernel, polynomial kernel, Radial Basis Function (RBF) kernel, etc. The training process of SVM can be transformed into a convex optimisation problem, where the optimal model parameters are determined by solving the Lagrange multipliers. In order to improve the generalisation ability, SVM introduces the concept of Soft Margin, which allows some sample points to violate the interval constraints, this is achieved by introducing regularisation parameters.

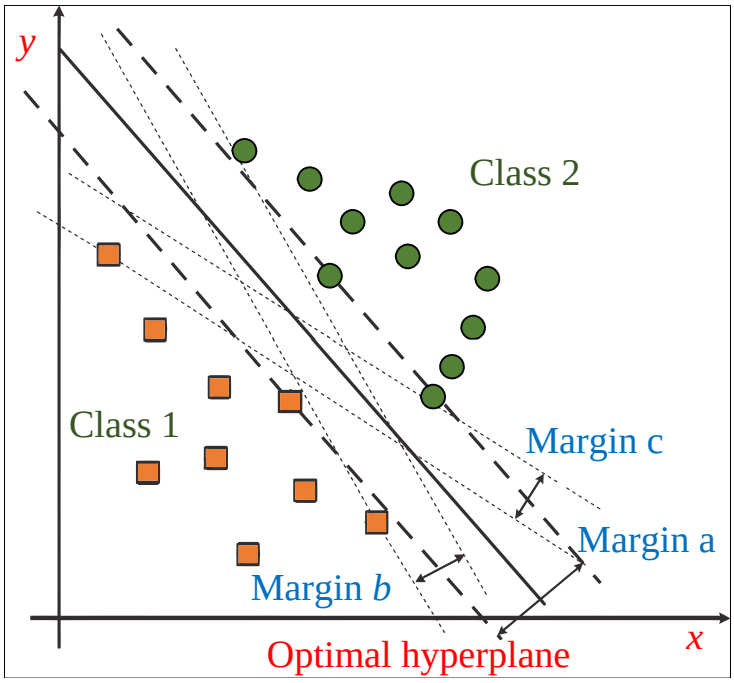

Figure 3. Principles of SVM

3.3. **Multimodal swimming movement characterisation.** When using model-based features for swim stroke action analysis, we typically rely on an understanding of human kinematics and dynamics, as well as dynamic information extracted from IMU data.

The joint angle is one of the key features describing the swimming posture and can be obtained by solving the IK problem. If p_i is the position of the i-th marker in the Olive Vendor system of the world, and p_{i-1} is the position of its neighbouring marker, then the joint angle θ can be computed as:

$$\theta = \arccos\left(\frac{p_i \cdot p_{i-1}}{\|p_i\|\|p_{i-1}\|}\right) \tag{19}$$

where arccos is the inverse cosine function used to calculate the angle between two vectors.

In addition to this, the angular velocity is a feature that describes the rotational speed of the joint and can be obtained by integrating the gyroscope data. If ω_{t-1} is the angular velocity at time $t-1$ and Δt is the time interval, then the angular velocity ω_t at time t can be approximated as:

$$\omega_t \approx \omega_{t-1} + \frac{1}{\Delta t}\left(\text{gyro}_t - \text{gyro}_{t-1}\right) \tag{20}$$

where gyro_t is the gyroscopic measurement at time t [28].

Linear velocity and acceleration can be obtained by integrating and filtering the accelerometer data. If acc_{t-1} is the acceleration measurement at time $t-1$ and v_{t-1} is

the corresponding velocity, then the velocity v_t and acceleration acc_t at time t can be calculated as:

$$v_t = v_{t-1} + acc_{t-1}\Delta t \tag{21}$$

$$acc_t = \frac{acc_t - acc_{t-1}}{\Delta t} \tag{22}$$

3.4. **Combined CNN-SVM model.** The CNN-SVM combined model proposed in this paper is an approach that uses CNN for feature extraction and then SVM for classification, as shown in Figure 4.

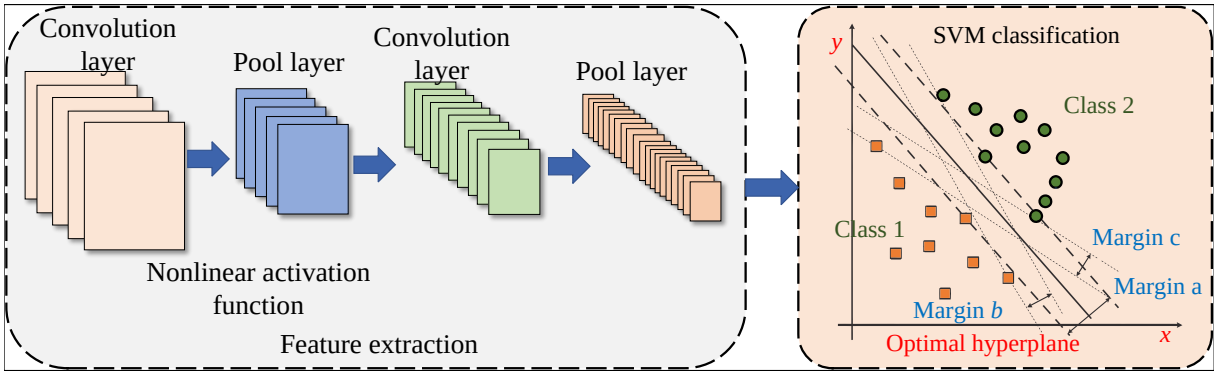

Figure 4. Combined CNN-SVM model

The model aims to exploit the powerful feature extraction capability of CNNs and the excellent classification performance of SVMs to obtain higher recognition accuracy. CNNs perform well in feature extraction of image and time-series data and are able to automatically learn hierarchical data representations. In this model, CNN is used to automatically extract motion features from the original inertial sensor data. Specifically, the original data is first passed through a 1D convolutional layer and a pooling layer for feature extraction and dimensionality reduction, and then a feature vector is output through a fully connected layer.

The obtained feature vector is then fed into the SVM classifier, which is a supervised machine learning method that constructs an optimal hyperplane in a high-dimensional space to separate the samples of different categories. SVM has good generalisation ability, and can obtain good performance even in the case of small training data. In this model, SVM maps the feature vectors extracted from CNN to different swimming pose categories.

The two modules, CNN and SVM, are jointly trained in an end-to-end manner. During the training process, the CNN weights and SVM parameters are optimally updated by a combination of backpropagation and convex optimisation to minimise the recognition error. This end-to-end joint training method is conducive to the mutual adaptation of CNN feature extraction and SVM classification, thus improving the overall model performance.

The advantage of this CNN-SVM hybrid model is that it integrates the powerful feature learning capability of CNN with the excellent classification performance of SVM; the CNN part automatically learns the high-level motion feature representations from the original inertial data, while the SVM part maps these features to the final pose categories. Compared with a single CNN or SVM model, the hybrid CNN-SVM model is expected to achieve better performance in the swimming pose recognition task.

4. **Experimental results and analyses.**

4.1. **Experimental method.** Twenty swimmers, including athletes of different levels and ordinary swimmers, were selected for the experiment to ensure the diversity and representativeness of the data. The experimental design followed the principles of scientificity and practicality. During the experiment, the swimmers were asked to perform a series of standard swimming manoeuvres, including different strokes and turns. Each swimmer completed at least two laps in a 25-metre pool to ensure that sufficient data were collected for analysis. A high-speed video camera was used to make simultaneous video recordings during the experiments so that they could be compared and analysed with the IMU data.

IMU sensors are installed at the swimmer's waist to capture dynamic data of swimming manoeuvres. IMU data including acceleration, angular velocity and magnetic field data were recorded as the swimmer executed different strokes. At the same time, the swimmer's movements were recorded using video for subsequent synchronisation with the IMU data. Video analysis was used to annotate the start and end time of each swimming movement and the corresponding stroke category. Synchronising the IMU data with the video data ensured that the data segments for each movement corresponded to the actual movements. The main parameters of the IMU used in the experiment are shown in Table 1.

Table 1. Main parameters of the IMU

Parameter name	Accelerometer	Gyros	Magnetometer
Dimension	Triaxial	Triaxial	Triaxial
Range	$\pm 16g$	$\pm 1000°/s$	$\pm 1200\,\mu T$
Sensitivity (LSB/g)	1	0.1	0.3
Bandwidth (Hz)	100	100	100

Accelerometers are used to measure linear acceleration and can detect the acceleration of a swimmer's movement. The triaxial accelerometer is able to provide acceleration data in X, Y and Z directions. Gyroscopes are used to measure angular velocity and can detect rotational movements of the swimmer's body. The 3-axis gyroscope provides angular velocity data around the X, Y, and Z axes. Magnetometers are used to measure the magnetic field and assist in determining the spatial orientation of the swimmer. The triaxial magnetometer provides data on the strength of the magnetic field in the X, Y, and Z directions.

4.2. **Swimming stance capture results.** The acceleration signals continuously monitored for the four swimming strokes of low intensity butterfly, backstroke, breaststroke and freestyle are shown in Figure 5. It is found that the Z-axis acceleration information is only negative for the data corresponding to backstroke, which is due to the fact that the face faces upwards in the backstroke swimming posture. In addition, from the gyroscope signals, it can be seen that the three-axis gyroscope data of backstroke and freestyle are quite different, and the Y-axis data are much larger than the X-axis and Z-axis data, which is because the swimming postures of backstroke and freestyle need to rotate the body left and right. The results of the above motion signal characteristics are fully consistent with the corresponding action characteristics of various swimming strokes, which verifies the feasibility of the proposed inertial sensor-based swimming motion capture.

4.3. **Comparison of swimming pose recognition results.** In cross validation, set the turning action as Positive and the paddling action as Negative, then the classifier classification results can be counted.

TP_k denotes the number of data correctly classified as Turning manoeuvres (actually Turning manoeuvres),

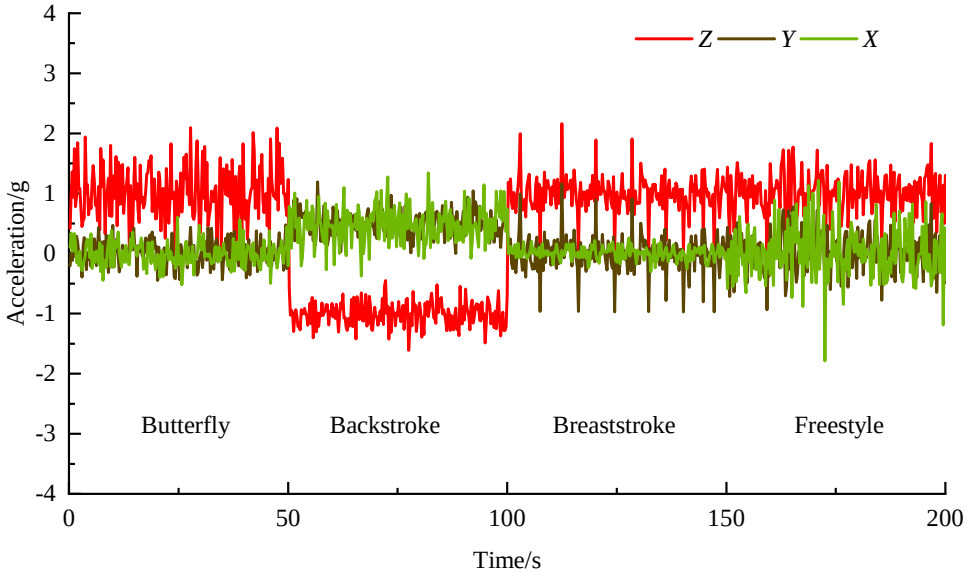

Figure 5. Motion signals corresponding to butterfly, backstroke, breaststroke and freestyle.

TN_k denotes the number of data correctly classified as Paddling manoeuvres (actually Paddling manoeuvres),

FP_k denotes the number of data incorrectly classified as Paddling manoeuvres (actually Turning manoeuvres),

and FN_k denotes the number of data misclassified as a turning action (actually a paddling action).

Then, the evaluation metrics of the classifier: Accuracy, Precision, Recall and F1-score were calculated according to Equation (23)–(26).

$$Acc_k = \frac{\sum_i TP_k + TN_k}{\sum_i TP_k + TN_k + FN_k + FP_k} \tag{23}$$

$$Prec_k = \frac{\sum_i TP_k}{\sum_i TP_k + FP_k} \tag{24}$$

$$Rec_k = \frac{\sum_i TP_k}{\sum_i TP_k + FN_k} \tag{25}$$

$$F1_k = \frac{2 \cdot Prec_k \cdot Rec_k}{Prec_k + Rec_k} \tag{26}$$

Correct rate is the proportion of swimming manoeuvres that can be correctly identified, and a higher correct rate usually indicates a better classifier performance. However, it cannot be used as the only evaluation criterion when the number of samples is different for each type of instances. Therefore, accuracy and recall are added as basic evaluation indexes, as shown in Table 2.

It can be seen that the recognition accuracy of the four strokes is very high on both training and test data. Freestyle and Backstroke achieved accuracies of 0.982 and 0.951 on both training and test data, and Butterfly and Breaststroke also achieved accuracies of 0.939 and 0.924 (training data) and 0.931 and 0.915 (test data), respectively. This indicates that the CNN-SVM model has good performance on the swim stroke recognition task.

Table 2. Swimming Posture Recognition Results for Four Swimming Styles

Swimming position	Training data				Test data			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
Freestyle swimming	0.982	1.004	0.959	0.981	0.951	0.988	0.911	0.948
Butterfly stroke	0.939	0.973	0.902	0.936	0.931	0.966	0.889	0.917
Breaststroke	0.924	0.981	0.862	0.917	0.915	0.954	0.871	0.912
Backstroke	0.982	1.004	0.959	0.981	0.951	0.988	0.911	0.948

Precision and Recall are two important indicators of the performance of a classification model. It can be seen that both Precision and Recall are very high for Freestyle and Backstroke, which indicates that the model has high accuracy and coverage for the recognition of these two strokes. The butterfly and breaststroke are slightly lower but still maintain a high level, indicating that the model also has a good performance for the recognition of these strokes. However, the recall of breaststroke is relatively low.

The F1-score is the reconciled mean of precision and recall, providing a performance metric that combines precision and recall. The F1-scores for all the swims in the table exceed 0.9 (training data) and 0.91 (test data), which further demonstrates the effectiveness of the CNN-SVM model on the swim recognition task. The recognition results on the test data are comparable to the training data, indicating that the model has good generalisation ability. This means that the model not only performs well on the training data, but is also able to adapt to new and unseen data. The recognition results for all four swim strokes show a high degree of consistency on both the training and test sets, which further demonstrates the robustness of the CNN-SVM model.

5. **Conclusions.** In this study, we proposed a swimming pose recognition method based on inertial sensor and CNN-SVM combined model, which achieves high accuracy recognition of swimming action by fusing deep learning algorithm and sensor data. With a well-designed IMU data acquisition device and scientific preprocessing steps including noise filtering, time alignment, normalisation processing and windowing, we ensure the quality and consistency of the input data, which provides a solid foundation for pose estimation and action recognition. Multi-dimensional features such as joint angles, angular velocities, linear velocities and accelerations obtained by solving the inverse kinematics problem provide richer and more detailed information for action recognition. The comprehensive analysis of these features significantly improves the model's ability to recognise the nuances of swimming movements.

Experimental results show that the proposed combined CNN-SVM model exhibits excellent performance on the action recognition task for different swimming strokes. The model makes full use of the advantages of CNN in feature extraction and the powerful ability of SVM in classification to achieve accurate classification of swimming actions. The high consistency of the model on both the training and test sets demonstrates its good generalisation ability. In addition, the model gives accurate recognition results for data from different levels of swimmers, proving its potential and reliability in practical applications. Despite the overall high performance of the CNN-SVM model, the relatively low recall for breaststroke may indicate that the model has some challenges in recognising breaststroke movements. Future work could focus on improving the recognition performance of this stroke.

REFERENCES

[1] R. Jain, V. B. Semwal, and P. Kaushik, "Stride segmentation of inertial sensor data using statistical methods for different walking activities," *Robotica*, vol. 40, no. 8, pp. 2567–2580, 2022.

[2] P. Picerno, M. Iosa, C. D'Souza, M. G. Benedetti, S. Paolucci, and G. Morone, "Wearable inertial sensors for human movement analysis: A five-year update," *Expert Review of Medical Devices*, vol. 18, no. sup1, pp. 79–94, 2021.
[3] X. Ru, N. Gu, H. Shang, and H. Zhang, "MEMS inertial sensor calibration technology: Current status and future trends," *Micromachines*, vol. 13, no. 6, 879, 2022.
[4] D. C. Luvizon, D. Picard, and H. Tabia, "Multi-task deep learning for real-time 3D human pose estimation and action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2752–2764, 2020.
[5] V. Mazzia, S. Angarano, F. Salvetti, F. Angelini, and M. Chiaberge, "Action transformer: A self-attention model for short-time pose-based human action recognition," *Pattern Recognition*, vol. 124, 108487, 2022.
[6] W. An, S. Yu, Y. Makihara, X. Wu, C. Xu, Y. Yu, R. Liao, and Y. Yagi, "Performance evaluation of model-based gait on multi-view very large population database with pose sequences," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 4, pp. 421–430, 2020.
[7] C. L. Reed, E. J. Moody, K. Mgrublian, S. Assaad, A. Schey, and D. N. McIntosh, "Body matters in emotion: restricted body movement and posture affect expression and recognition of status-related emotions," *Frontiers in Psychology*, vol. 11, 1961, 2020.
[8] V. Adithya, and R. Rajesh, "A deep convolutional neural network approach for static hand pose recognition," *Procedia Computer Science*, vol. 171, pp. 2353–2361, 2020.
[9] C. Chen, W. Zhu, and T. Norton, "Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning," *Computers and Electronics in Agriculture*, vol. 187, 106255, 2021.
[10] M. Oudah, A. Al-Naji, and J. Chahl, "Hand pose recognition based on computer vision: a review of techniques," *Journal of Imaging*, vol. 6, no. 8, pp. 73, 2020.
[11] A. Stamm, and D. V. Thiel, "Investigating forward velocity and symmetry in freestyle swimming using inertial sensors," *Procedia Engineering*, vol. 112, pp. 522–527, 2015.
[12] A. Kos, and A. Umek, "Reliable communication protocol for coach based augmented biofeedback applications in swimming," *Procedia Computer Science*, vol. 174, pp. 351–357, 2020.
[13] A. Rusdiana, B. Mulyana, D. R. Nurjaya, I. I. Badruzaman, E. Fauziah, and A. M. Syahid, "3D Biomechanical analysis of swimming start movements using a portable smart platform with android pie," *Journal of Engineering Science & Technology*, vol. 16, no. 1, pp. 571–585, 2021.
[14] S. Qiu, H. Zhao, N. Jiang, D. Wu, G. Song, H. Zhao, and Z. Wang, "Sensor network oriented human motion capture via wearable intelligent system," *International Journal of Intelligent Systems*, vol. 37, no. 2, pp. 1646–1673, 2022.
[15] Y. Desmarais, D. Mottet, P. Slangen, and P. Montesinos, "A review of 3D human pose estimation algorithms for markerless motion capture," *Computer Vision and Image Understanding*, vol. 212, 103275, 2021.
[16] T. B. Moeslund, and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.
[17] E. Van der Kruk, and M. M. Reijne, "Accuracy of human motion capture systems for sport applications; state-of-the-art review," *European Journal of Sport Science*, vol. 18, no. 6, pp. 806–819, 2018.
[18] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2–3, pp. 90–126, 2006.
[19] S. Iloga, A. Bordat, J. Le Kernec, and O. Romain, "Human activity recognition based on acceleration data from smartphones using HMMs," *IEEE Access*, vol. 9, pp. 139336–139351, 2021.
[20] A. Abdelbaky, and S. Aly, "Two-stream spatiotemporal feature fusion for human action recognition," *The Visual Computer*, vol. 37, no. 7, pp. 1821–1835, 2021.
[21] R. Vrskova, R. Hudec, P. Kamencay, and P. Sykora, "Human activity classification using the 3DCNN architecture," *Applied Sciences*, vol. 12, no. 2, 931, 2022.
[22] T.-Y. Wu, H. Li, S. Kumari, and C.-M. Chen, "A Spectral Convolutional Neural Network Model Based on Adaptive Fick's Law for Hyperspectral Image Classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19–46, 2024.
[23] F. Zhang, T.-Y. Wu, J.-S. Pan, G. Ding, and Z. Li, "Human motion recognition based on SVM in VR art media interaction environment," *Human-centric Computing and Information Sciences*, vol. 9, 40, 2019.

[24] T.-Y. Wu, L. Yang, Z. Lee, S.-C. Chu, S. Kumari, and S. Kumar, "A Provably Secure Three-Factor Authentication Protocol for Wireless Sensor Networks," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1–15, 2021.

[25] T. Roska, and L. O. Chua, "The CNN universal machine: an analogic array computer," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 40, no. 3, pp. 163–173, 1993.

[26] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, pp. 1–74, 2021.

[27] Y. Ren, C. Zhao, Y. He, P. Cong, H. Liang, J. Yu, L. Xu, and Y. Ma, "Lidar-aid inertial poser: Large-scale human motion capture by sparse inertial and lidar sensors," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2337–2347, 2023.

[28] W. Hu, X. Zhu, T. Wang, Y. Yi, and G. Yu, "Discrete subspace structure constrained human motion capture data recovery," *Applied Soft Computing*, vol. 129, 109617, 2022.