Hindawi
Security and Communication Networks
Volume 2023, Article ID 9871714, 1 page
https://doi.org/10.1155/2023/9871714

Retraction

Retracted: Dance Movement Recognition Based on Modified GMM-Based Motion Target Detection Algorithm

Security and Communication Networks

Received 10 October 2023; Accepted 10 October 2023; Published 11 October 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Tian and X. Yang, "Dance Movement Recognition Based on Modified GMM-Based Motion Target Detection Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 6023784, 12 pages, 2022.

Hindawi
Security and Communication Networks
Volume 2022, Article ID 6023784, 12 pages
https://doi.org/10.1155/2022/6023784

Research Article

Dance Movement Recognition Based on Modified GMM-Based Motion Target Detection Algorithm

Jing Tian[1] and Xiaoqiang Yang[2]

[1]Zhengzhou Tourism College, College of Arts and Culture, Zhengzhou 451464, China
[2]Hainan University School of Music and Dance, HaiKou 570228, China

Correspondence should be addressed to Xiaoqiang Yang; 991593@hainanu.edu.cn

Received 21 June 2022; Revised 29 July 2022; Accepted 4 August 2022; Published 22 August 2022

Academic Editor: Hangjun Che

Under the synergistic development of social economy and science and technology, the intelligent teaching of dance has become more and more popular. This teaching method can not only decompose dance movements more specifically, which is easy for students to understand and master, but also get rid of the time and space limitation in traditional dance teaching and provide more independent learning opportunities for students. The problem of low accuracy of dance movement recognition due to complex gesture changes in dance movements is addressed. To this end, this paper proposes a modified motion target detection algorithm based on GMM. The dance movement recognition algorithm first extracts the features of dance movements through a feature pyramid network, then uses a multi-feature fusion module to fuse multiple features to improve the algorithm's estimation of complex postures, and finally completes the recognition of dance movements. Experiments show that our method can maintain a certain recognition rate in the case where the background and target are easily confused, and can effectively improve the dance action recognition accuracy, thus realizing the action correction function for dancers. This also verifies the effectiveness of the action recognition algorithm for dance movement recognition.

1. Introduction

Human pose estimation is a key technique in the field of human action recognition, which is based on the principle of recognizing human pose by extracting features in images [1]. This technique can be used in intelligent dance-assisted training to obtain a skeleton map of the dancer's posture by extracting features from the dancer's image. Thus, the dancer's dance movements are recognized, and the dancer's posture is evaluated and corrected [2].

As an aid to human eye vision and an important component of automated systems, computer vision is widely used in medical and transportation fields [3]. Compared with the human eye, the advantage of computer vision is that it has much higher computational power than the human brain and higher analysis capability for complex images [4]. Action recognition in dance video images is an important application area of computer vision technology, which can be applied to many scenarios, such as competition arbitration, introductory learning for dancers, and movement correction for professional dancers [5, 6].

Compared with low-level action recognition such as gesture recognition and simple limb action recognition, dance action recognition has penetrated into the level of motion recognition [7, 8]. Therefore, when simple limb localization algorithms are applied to dance movement recognition, it is usually difficult to obtain high recognition accuracy [9, 10]. The difficulties of dance movement recognition mainly include the following three points.

Dance movements are complex and variable. From the most basic action elements such as "lifting," "sinking," "rushing," and "leaning" to the coherent and complex actions such as "standing beat swallow," "pouncing step," "cloud step," and "turning over," there is a great degree of freedom [11, 12]. Therefore, it is more difficult to identify each movement accurately.

Obscuration in dance is a serious problem. If there is only one dancer, some of the dancer's limbs may be obscured by themselves, making it difficult to identify the position of certain limbs; if there are multiple dancers, the dancers will obscure each other [13, 14]. In particular, the dancers' clothes are loose, such as long dresses with skirt support, so the obscured area will be larger. In addition, the angle of the photo or video can also cause some obstacles to the recognition of dance movements [15].

The coherence of dance movements is strong. In simple body movements, the coherence of the movements is weak. Generally, everyday body movements change slowly and each body movement remains the same over a period of time. However, in dance, all movements are coherent and fluid, and fewer movements remain stationary. Therefore, it is more difficult to accurately detect the boundaries of each dance movement in time.

Early human posture estimation mainly focused on human contour features or part models. For example, the literature [16, 17] designed a human pose estimation algorithm based on part detection by extracting edge force field features through boosting classifier. Literature [18, 19], on the other hand, proposed an appearance model combining histogram of oriented gradients (HOG) and color features for human pose estimation. However, due to the complex variation of human pose, the traditional methods are difficult to achieve effective pose estimation [20, 21]. Therefore, deep learning-based methods are gradually used for human pose estimation. In 2015, deep learning-based human pose estimation algorithms started to return to the human skeleton heat map [22]. In 2016, a research team from the University of Michigan [23] designed an hourglass-like neural network structure for extracting multi-scale features for human pose estimation. In 2017, the literature [24] proposed an approach using partial affinity domain to obtain human skeleton maps. In addition, numerous deep learning-based algorithms for human pose estimation have been proposed, all of which can be used for dance movement recognition to assist dancers' training [25]. The rapid change of dancers' movements and the variability of their postures pose a challenge for dancer-assisted training intelligence.

To this end, a dance movement recognition algorithm based on multi-feature fusion is designed in the paper for learning complex and variable dancer movement recognition.

2. Motion Recognition System

The motion recognition system in this paper consists of a human detection module, a pose and feature detection module, and a motion recognition module. First, a modified GMM-based motion target detection algorithm is used to detect and segment the moving human body from the video. For the detected binarized human region, the pose, pose change rate, and human position change information are extracted, and the pose evaluation function is introduced to improve the accuracy of pose detection. In the process of action recognition, an action recognition algorithm based on multi-feature fusion is proposed in this paper. The algorithm not only analyzes the shape features of human appearance, but also fuses the motion features of human body, so the recognition results are more accurate. The algorithm is easy to understand and implement, with a small amount of operations and fast recognition speed. The flow of the whole algorithm is shown in Figure 1.

2.1. Human Body Detection. The first step in human motion recognition is to detect and segment the human body in motion or at rest. Due to the various colors and textures of human clothing, the uncertainty of human posture, and the uncertainty of the visual background, there is still no feasible method to detect the human body from static images. Therefore, this paper uses motion detection to extract human targets in video images.

Background subtraction (BS) is a general and widely used technique for generating foreground masks (i.e., binary images containing pixels belonging to moving objects in the scene) by using a static camera. BS is the most commonly used method for detecting motion targets, and it can detect motion targets in indoor environments very well. It is found that the GMM background model adopts a global uniform update strategy, which highlights its shortcomings in dealing with complex target motion forms. The main manifestation is that the suspended motion targets are absorbed as part of the background, resulting in incomplete extracted motion targets such as people. This absorption phenomenon is unavoidable due to the need for an adaptive background model to handle slow background changes (e.g., illumination).Therefore, the results of motion segmentation and the recognition of the target can be used to guide the background update. For example, if the motion target $O_{(ij)}$ is a human object, the corresponding background update confidence f_{Bg} takes the value of 0, and the pixels at that point do not participate in the background update; otherwise, the corresponding background update confidence f_{Bg} takes the value of 1, and the pixels at that point participate in the background update. The region-based background update strategy avoids pose false detection caused by local human motion and improves the accuracy of action recognition using pose changes.

2.2. Motion State Characterization. Different features reflect the characteristics of human motion states from different perspectives. When selecting features, we should consider not only their distinguishability, but also the difficulty of their extraction. The object of this paper is the whole human body, including the limbs, and the goal of the study is to identify the typical daily movements (standing up, lying down, etc.) and sudden abnormal movements (falling down) in a complex environment. Therefore, the key features related to the shape and movement of the human body as a whole are considered in the human motion state characterization. In this paper, we adopt the idea of feature fusion to characterize the human motion state by fusing multiple features, because there are shortcomings in using appearance-based shape features alone or motion features alone.

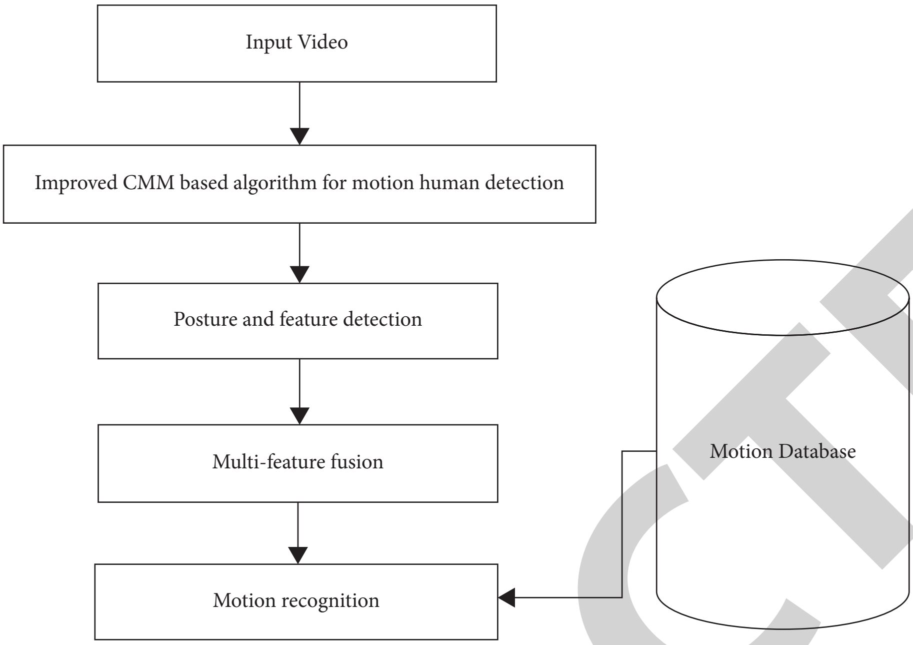

Input Video
Improved CMM based algorithm for motion human detection
Posture and feature detection
Multi-feature fusion
Motion recognition
Motion Database

Figure 1: Action recognition system based on multi-features fusion.

2.2.1. Posture Features. Human motion in the home environment consists mainly of several key pose transformations, so this paper selects the overall human pose information to describe the human motion state based on appearance and shape. Model-based pose acquisition can describe complex poses, but the model is difficult to initialize, computationally intensive, and prone to local minima, making it difficult to find globally optimal and robust parameters. Therefore, this paper chooses the human body width and height ratio, which is invariant to the target size and distance and has little influence on the viewpoint, to describe the human body's pose characteristics.

2.2.2. Pose Change Rate Feature. The posture of the human body always changes smoothly in normal daily movements, but the rate of posture change can be dramatic when a sudden abnormal (fall) situation occurs. In this paper, the motion feature of the rate of change of posture is introduced to detect the fall of a person. In the two actions of falling and lying down, the critical posture change process is similar and the posture change rate is different.

2.2.3. Position Change Feature. It is impossible to determine whether the human body is walking or standing by only relying on the selected posture features and their posture change rates. In these two kinds of movements, the human posture changes from the standing posture to the standing posture, and the introduction of the motion information of position change can solve this problem. Then, how to determine whether the position of the "human target" changes? After region segmentation, we can obtain the position of the moving human target in the image coordinate system (coordinates of the top left vertex of the smallest rectangle containing the foreground target) and compare the positions of the two frames before and after to determine whether the target position has changed.

2.3. Attitude and Feature Detection

2.3.1. Pose Feature Detection. The basic postures of human daily actions are defined as stand, sit, and lay, and the set of postures P.

$$P = \{s \tan d, \text{sit}, \text{lay}\}. \tag{1}$$

The currently detected pose $p(t) \in P$ is considered to be detected only when the human body is in motion, so the human body is considered to remain in the same pose when there is no motion information.

Since the motion body detection is a high-dimensional signal, it is not easy to recognize the pose in this high-dimensional space, and feature transformation is needed for later classification. The body posture ratio k is calculated for different postures in 900 daily actions, and the threshold value of k is set to distinguish between standing, sitting, and lying postures based on the minimum probability of misjudgment criterion, as described in Table 1.

One of the problems found in the experiments that affects the accuracy of posture is the false detection of posture. For example, when the human arm is unfolded, the standing posture is mistakenly detected as a sitting posture, as shown in Figure 2.

By constructing a posture evaluation function to eliminate the misjudgment of posture due to "unusual" human movements, define the posture evaluation function S as in equation.

Table 1: Thresholds for different actions.

Gesture	Stand	Sit	Lay
Threshold k	$k \geq 1.8$	$1.8 > k \geq 0.7$	$k < 0.7$

Figure 2: Posture detection.

$$S = \frac{\sum F_g}{S(T)}, \tag{2}$$

where $\sum F_g = \sum_{(x,y)\in T} I(x, y)$. is the area of the foreground image of the moving human body and $S(T) = W * H$. is the area of the smallest external rectangle of the foreground image. From the expression of the evaluation function, we can see that the value of the function varies in the range [0, 1]. The value of the evaluation function is highest when $\sum F_g = S(T)$, and becomes very low when the human arm is expanded. If the pose evaluation function is low, it is considered as an undefined pose and is not involved in action recognition.

The pose recognition algorithm in Haritaoglu compares the pose recognition algorithm in this paper with the algorithm in Haritaoglu. The pose recognition algorithm in Haritaoglu projects the foreground image of a moving human body in the x-axis and y-axis directions, and matches the projected contour lines with the training contour line templates of different poses to obtain the human pose. The pose recognition rates of the two algorithms are comparable, but in terms of complexity, Haritaoglu's algorithm is more complex than the one in this paper.

2.3.2. Pose Change Rate Detection. Define the ratio of the body posture ratio of the previous frame to the body posture ratio of the current frame as the inter-frame change rate of human posture, denoted by Q, i.e.,

$$Q = \frac{k(t-1)}{k(t)}. \tag{3}$$

Q characterizes how quickly a person's posture changes: when a person maintains the same posture, Q is close to 1; when a person sits or lies down normally, Q increases slowly; when a person falls, Q increases rapidly. Thus, the rate of change of the human posture ratio can be used to detect the falling action. The rate of change of human posture during normal movement $Q < 1.5$ is obtained statistically.

2.3.3. Position Detection. The position of the human body in the image coordinate system is defined as $P(t, i)$, and the position of the human body changes using the Euclidean distance metric, i.e., $k(t, i) = \|P(t+1, i) - P(t, i)\|$; when $k(t, i)$ is greater than the threshold value K_s, the human body position is in motion. The discriminant method is

$$\text{if } k(t, i) n K_s, \text{then Action} = 1; \text{else Action} = 0; \tag{4}$$

In order to accurately determine whether the position of the human body has changed and to improve the robustness of the algorithm, the concept of confidence is introduced. The confidence level is used to measure the degree to which the human target is in motion and ranges from 0 to 40. A confidence level of 0 means that the human body is definitely in motion and a confidence level of 40 means that the target is definitely not in motion. If Action = 0, then the confidence

Table 2: Condition setting of different actions.

Action			Conditions	
$a(t)$	$p(t)$	$p(t-1)$	Q	UAction
Walk	Stand	Stand	< 1.5	≤ 0.5
Sit down	Stand	Sit	< 1.5	—
Stand up	Sit	Stand	< 1.5	—
Lay down	Sit, stand	Lay	< 1.5	—
Get up	Lay	Sit, stand	< 1.5	—
Fall down	Stand, sit	Sit, lay	≥ 1.5	—
Stand still	Stand	Stand	< 1.5	> 0.5

level of the target is increased by 1; otherwise, the confidence level is zero. Given a confidence threshold, the target is considered to be in a nonmotion state when the confidence level is greater than the threshold, which is set to 20 in the text. The confidence level is defined as CAction, the initial value of CAction is 0, and the confidence level is normalized to UAction. If UAction is greater than 0.5, the human position changes and vice versa, and there is no change. The specific method is

$$\begin{aligned} &\text{if Action} = 0, \text{then}\, C\text{Action} \\ &\quad + +; \text{else}\, C\text{Action} = 0; \\ &\text{if}\, C\text{Actionn40nthen}\, C\text{Action} = 40 \\ &U\text{Action} = C\text{Action}/40. \end{aligned} \tag{5}$$

2.4. Action Recognition. Define the daily actions of a person: walk, sit down, stand up, lay down, get up, fall down, and standstill, which form the action set A.

$$\text{A} = \{\text{walk, sit down, stand up, lay down, get up, fall down, stand still}\}. \tag{6}$$

The current detected action $a(t) \in A$. According to the regularity that different human actions are composed of different postures, this paper detects human actions by using the posture change combined with the frame-to-frame change rate feature and the position change feature, as shown in Table 2. $p(t)$ denotes the posture at moment t, and $p(t-1)$ denotes the posture at moment $t-1$.

In order to filter out meaningless or undefined actions, this paper introduces a threshold model of the minimum number of frames of pose duration. The threshold model gives the bottom line for performing action judgments, and action judgments are performed only when the number of pose duration frames of the observed sequence is greater than the threshold; otherwise, the observed sequence is considered as a meaningless or undefined action. According to this criterion, the minimum number of frames that can be statistically obtained to describe the pose of each action is 10; i.e., the threshold value in the threshold model is 10 frames (the video image acquisition rate is 30 frames/s). This method is able to eliminate the false detection of motion due to noise.

2.5. Dance Video Image Motion Pose Extraction and Joint Modeling. With the development of human behavior recognition field and the depth of research tasks, from the initial recognition of simple single actions under restricted conditions to the complex group behavior recognition in real natural scenes nowadays, both the information acquisition equipment and algorithm capability have posed serious challenges. As an important part of the behavior recognition process, the result of feature extraction largely affects the real time and accuracy of the behavior recognition effect. As a classical problem in the field of computer vision and machine learning, feature extraction is different from feature extraction in image space, and the feature representation of human action in video not only describes the human form in image space, but also must extract the human appearance and posture changes, which extends the feature extraction problem from two-dimensional space to three-dimensional space-time, which greatly increases the complexity of behavior mode expression and subsequent recognition tasks. At the same time, it also broadens the thinking of vision researchers in terms of solution ideas and technical methods. Human features are the information that can be extracted from the underlying video sequence to characterize the target behavior, such as color, contour, texture, depth, or human motion direction, speed, trajectory, as well as spatiotemporal interest points and spatiotemporal context.

2.6. Identifying Action Features Using Pose Feature Extraction Method

2.6.1. Dancer's Action Recognition Feature Classification. There are great differences between the dance movements of dancers and the daily movements of ordinary people, and many movements require dancers to use their arms and legs to complete, so when selecting the target area for background recognition, it is necessary to grasp the whole body movement information of dancers to accurately identify their movements. Dancer's movement recognition can be divided into several categories: static features, mainly in the form of dancer's human target size, color, body contour, depth, etc., which can convey the overall information of dancer's movement, such as the current basic shape that can be derived from the dancer's contour features; dynamic features, mainly in the form of dancer's movement speed, direction, and trajectory, which can reflect the dancer's movement path. The identification of these features can calculate the movement direction characteristics of the dancer and create conditions for modeling; spatiotemporal features are mainly manifested as spatiotemporal shapes, points of interest, etc.; descriptive features include the scene the dancer is in, surrounding objects, posture, etc.

2.6.2. Dancer Pose Feature Extraction. The pose feature extraction method can be used in conjunction with a pose estimation sensor, which is commonly used in the field of motion tracking and robot vision to determine the directional points of a dancer's motion. It can use the optical flow value in the pose estimator to filter the background information in the image to obtain the dancer's joint coordinate region, and can eliminate the occlusion and influence of factors such as the dancer's clothes on the dancer's motion as shown in Figure 3.

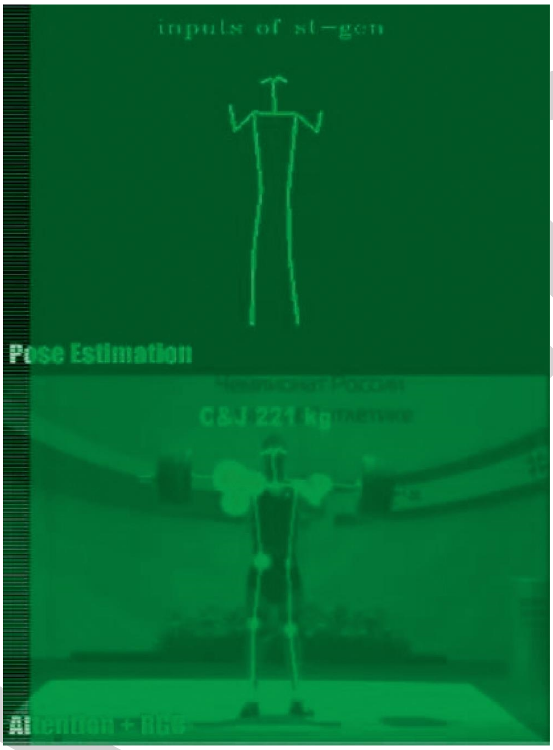


Figure 3: Dancer's posture characteristics.

2.7. Dancer Joint Point Recognition Modeling Using Kinect. The Kinect method is used to consider the human body as an axis composed of 25 joint point coordinates, and the human skeletal structure of the dancer is built with these joint points to obtain the human skeletal model of the dancer as shown in Figure 4.

From the model, it can be seen that the dancer's joint points are mainly distributed in the extremities. There is one joint point in the center of the head, neck, spine, and shoulder, and the most concentrated joint points are located in the upper limbs. The left upper limb has joint points. The principle of using these joint points to build the model is to accurately record the movement of each joint point during the dancers doing various movements, accurately identify each movement of the dancer, and thus output the correct skeleton of the dancer's movements. With the skeleton model of dancers' movements, the recognition accuracy and efficiency of the computer vision system for dancers' movements can be significantly improved, and the whole recognition process is shown in Figure 5.

3. Dance Movement Recognition

The new algorithm uses a feature pyramid network (FPN) for feature extraction, then deepens the extraction for different scale features, and finally upsamples each feature to the original image size for feature fusion, as shown in Figure 6. The residual block in the figure indicates the residual module as shown in Figure 7.

3.1. Feature Pyramid-Based Backbone Network. The shallow features in C_1, C_2, C_3 have a high spatial resolution. However, the semantic information contained is not sufficient, while the opposite is true for the deeper features in C_4, C_5.

Based on the FPN backbone network, as shown in Figure 8, it is difficult to identify human pose key points in complex environments, such as occluded hidden key points. The localization of such complex key points usually requires richer feature information, for which a multi-feature fusion module is designed in the paper.

3.2. Multi-Feature Fusion Module. The FPN-based backbone network is used to identify the estimation of simple key points, and the multi-feature fusion module is used to handle the estimation of more complex key points, whose structure is shown in Figure 9. To obtain better local features, this paper enhances the feature resolution at each stage by upsampling operation. Finally, the individual features from the FPN are fused into the CONCAT operation.

During the training process, the FPN extracts features and returns the human skeleton key points, and simple key points will be basically completed in the FPN stage. For complex key points, such as occlusion and hiding, the multi-feature fusion module will further deepen the learning of features from each layer of the FPN and fuse them, and finally return to the human skeleton heat map.

3.3. Loss Function. Human pose estimation is a regression problem, and the common loss functions in regression problems are L1 loss function and L2 loss function. The dancer movement recognition in this paper adopts the regression of the key points of the dancer's skeleton, so the algorithm in this paper adopts the loss function of L2 parametric optimized Euclidean distance, as shown in (1).

$$L(\theta) = \frac{1}{2N} \sum_{i=1}^{N} \left\| F(X_i; \theta) - F_i \right\|_2^2, \tag{7}$$

where θ denotes the dancer movement recognition network parameters to be optimized; N is the total number of dancer images involved in the learning training; X_i denotes the current learning dancer image sample i; F_i denotes the heat map of the i-th dancer image; and $F(X_i; \theta)-$ denotes the key points of the dancer skeleton regressed by the model heat map of the key points of the dancer's skeleton regressed by the model.

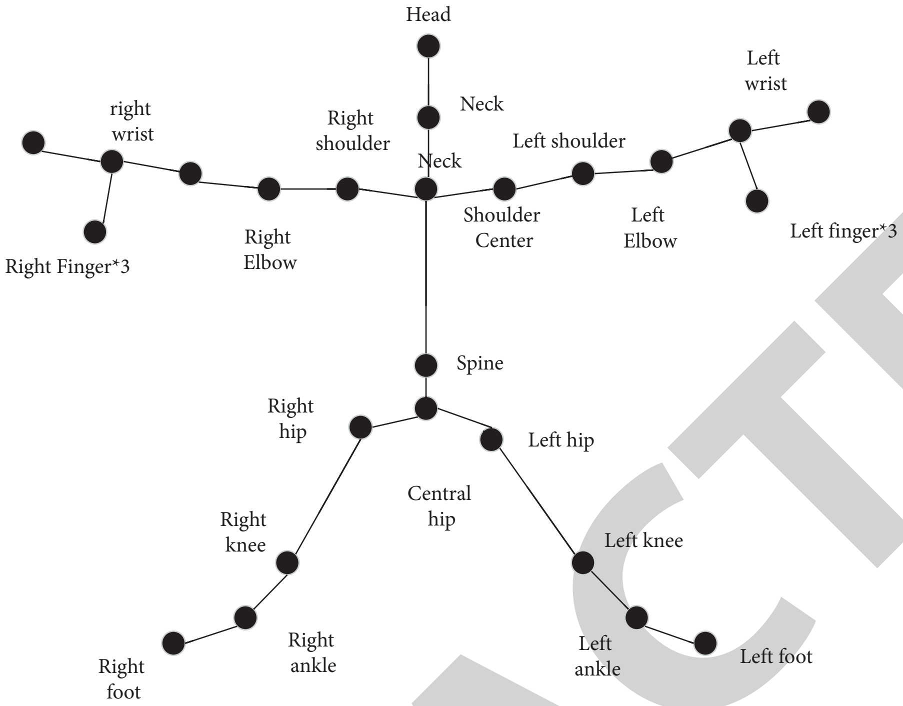


Figure 4: Dancer's joint point recognition model by 2 Kinect method.

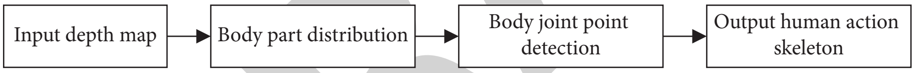


Figure 5: Dancer's joint point recognition process.

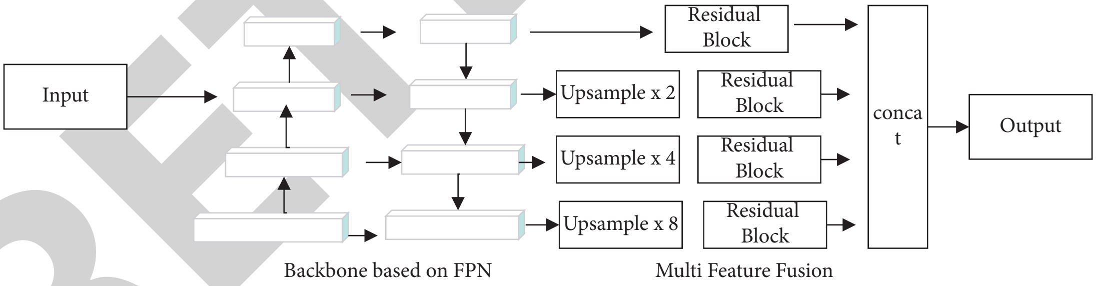


Figure 6: Dance movement recognition.

4. Long-Time Target Tracking Algorithm

The extraction of features directly affects the accuracy and efficiency of target tracking. Given a new image frame, two filtering templates, target position and scale prediction, are learned based on HOG and texture features, respectively. The filtering output is calculated using (2).

$$f(z) = \gamma_{HOG} f_{HOG}(z) + \gamma_{\text{tex}} f_{\text{tex}}(z). \tag{8}$$

The contribution of the two feature responses is $\gamma_{\text{HOG}}, \gamma_{\text{tex}}$, and the calculation rule is shown in (3) and satisfies $\gamma_{\text{HOG}} + \gamma_{\text{tex}} = 1$.

$$\begin{cases} \gamma_{\text{HOG}} = \dfrac{f_{\text{HOG}}(z)}{f_{\text{HOG}}(z) + f_{\text{tex}}(z)}, \\ \gamma_{\text{tex}} = \dfrac{f_{\text{tex}}(z)}{f_{\text{HOG}}(z)/\left(f_{\text{HOG}}(z) + f_{\text{tex}}(z)\right)}. \end{cases} \tag{9}$$

The filtered response values of the two features are linearly weighted and fused using (3), and the maximum response value after fusion is used to determine the target region.

In this paper, a simple and effective region suggestion generation scheme, EdgeBox, is chosen to generate candidate regions for the whole image and calculate their

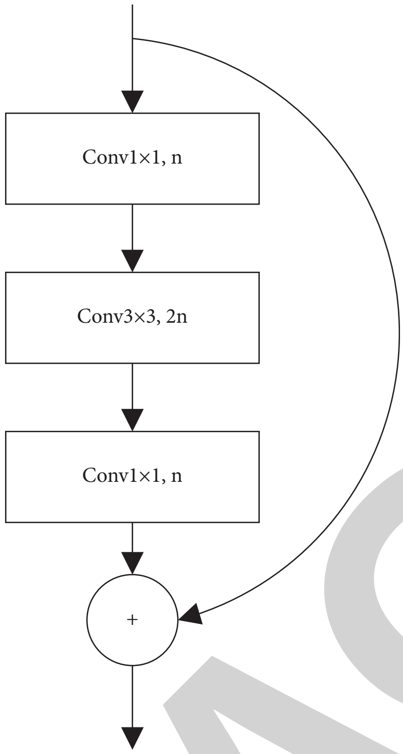
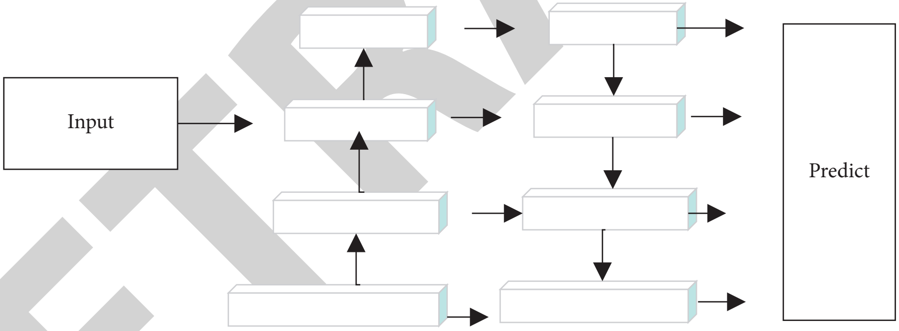


Figure 7: Dance movement recognition residual module.



Backbone base d on FPN

Figure 8: Backbone network.

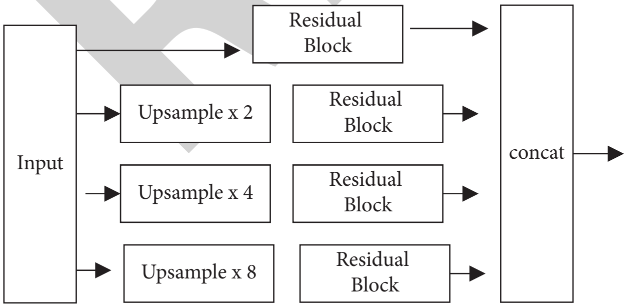
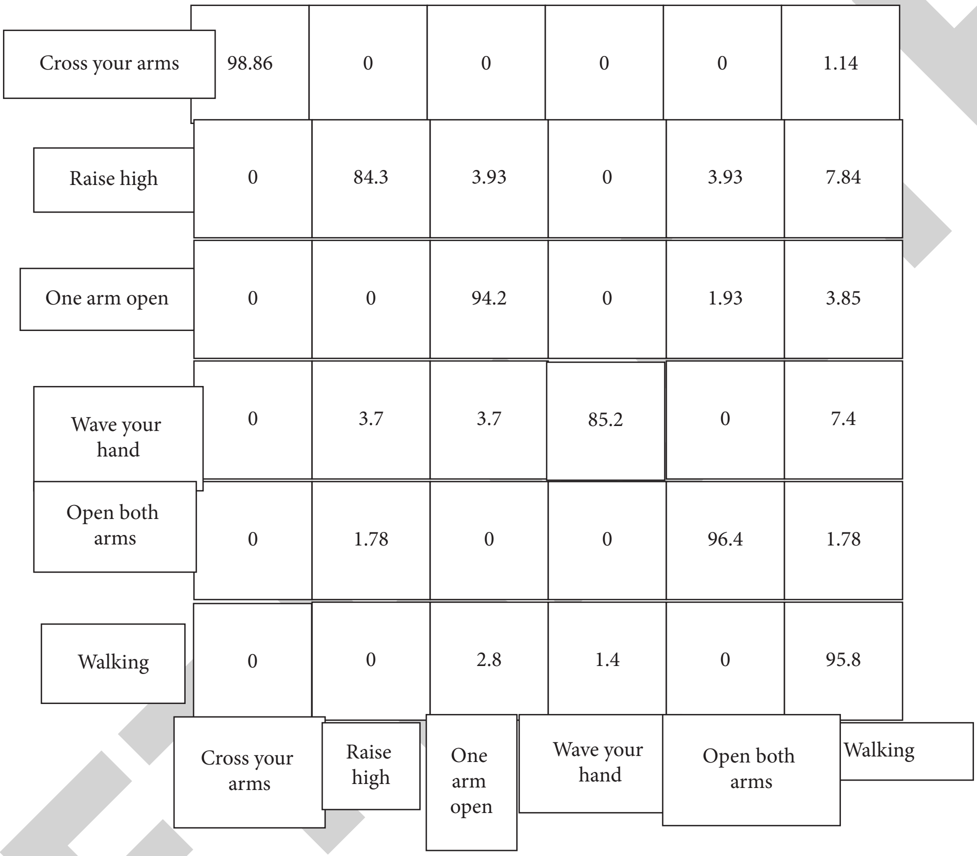


Figure 9: Multi-feature fusion structure diagram.

confidence scores, and the candidate region with the highest confidence is the retracing result. For an image, the edge information is used to determine the bounding box of the object. Based on the number of contours within the bounding box and the number of contours overlapping with the edges of the bounding box, the bounding box is scored and the candidate region information is determined according to the order of the scores. The candidate regions generated by EdgeBox include two types, one near the predicted target (denoted by B_s) and one for the whole image region (denoted by B_h). b is a candidate bounding box in B_s or B_h. Define $g(b)$ as the maximum filter response, and detect the tracking result by checking whether $g(b)$ is lower than the given threshold $T1$, which is less than the threshold value to indicate tracking failure and start the retracking procedure. In the normal tracking process, the filter template of the previous frame is generally used to find the target position of the current frame. However, during retracking, the tracking result of the previous frame

Table 3: Number and complexity of training set images for each movement.

Dance action name	Number of training set images	Action complexity
Lift	185	1
Sink	210	1
Charge	122	2
Cloud step	108	1
Turn over	150	2
Stand up and shoot the swallow	188	3



	Cross your arms	Raise high	One arm open	Wave your hand	Open both arms	Walking
Cross your arms	98.86	0	0	0	0	1.14
Raise high	0	84.3	3.93	0	3.93	7.84
One arm open	0	0	94.2	0	1.93	3.85
Wave your hand	0	3.7	3.7	85.2	0	7.4
Open both arms	0	1.78	0	0	96.4	1.78
Walking	0	0	2.8	1.4	0	95.8

Figure 10: Confusion matrix of 6 dance movements.

is no longer reliable, so it is necessary to select a reference image from the label library as the retracking head (the first frame is used as the retracking head by default). The confidence level of all images in the label library is read, and the Euclidean distance between a candidate frame (b_t^i) and these images $(b_{t-j}, j = 1 \longrightarrow t)$ is calculated for the current frame.

$$D\left(b_t^i, b_{t-j}\right) = \exp\left(-\frac{1}{2\sigma^2}\left\|\left(x_t^i, y_t^i\right) - \left(x_{t-j}, y_{t-j}\right)\right\|^2\right), \tag{10}$$

σ is the initial target size diagonal length. Based on the confidence, Euclidean distance, the best element is found as the retracking head and trained online to update the filtering model and regain the normal tracking pattern, so that the algorithm maintains high robustness and efficiency in long-time tracking.

$$\begin{gathered} \underset{i,j}{\arg\min}\, \beta g\left(b_t^i\right) + (1-\beta) D\left(b_t^i, b_{t-j}\right), \\ g\left(b_t^i\right) > T_1. \end{gathered} \tag{11}$$

The β in (5) is a weight parameter that adaptively adjusts the confidence, the contribution of the Euclidean distance. If $\{g(b) | b \in B_s\}$ is greater than $g(z)$ (the confidence level of the current template), the new target size (w_t, h_t) is defined as

$$(w_t, h_t) = \alpha\left(w_t^*, h_t^*\right) + (1-\alpha)\left(w_{t-1}, h_{t-1}\right), \tag{12}$$

It can also be expressed as

$$(w_t, h_t) = \left(w_{t-1}, h_{t-1}\right) + \alpha\left(\left(w_t^*, h_t^*\right) - \left(w_{t-1}, h_{t-1}\right)\right), \tag{13}$$

(w_t^*, h_t^*) indicates the width and height of the maximum confidence candidate region, and (w_{t-1}, h_{t-1}) is the width and height of the previous tracking target [26–28].

Table 4: Recognition results.

Movement type	Accuracy (%)
Arms crossed	98.9
Arms raised	85.3
One-hand wave	95.2
One arm open	85.1
Both arms open	96.9
Walking	95.8

5. Results and Analysis

5.1. Algorithm Validation. In order to better verify the accuracy of the dance movement recognition method designed in this paper, two data sets, PASCAL VOC2011-val set and Stanford 40 actions, are more commonly used, and the collected dance images were used to conduct the experiments. All experiments were performed on a computer with Intel Core i7-4790 CPU and 16 GB RAM and Windows 10 operating system based on Visual Studio 2010 development platform and OpenCV2.4.3 programming environment [29, 30]. The complexity of each dance movement and the number of images in the training set are shown in Table 3.

To verify the effectiveness of the algorithm, all heat maps are visualized as shown in Figure 8. The left image is the input image, the middle image is the singer's skeletal key point heat map, the right image is the computed singer heat map, and the right image is the maximum probability key point and key point limb region obtained from the computed singer heat map. The algorithm is trained and recognized, the accuracy on the training set and test set is shown in Figure 10, and the recognition accuracy on the specific test set is shown in Table 4.

From Table 4, it can be seen that the average recognition accuracy of this research method on the test set is more than 92%, and the overall recognition accuracy is high, but the recognition accuracy of arm raising and one-hand waving does not reach 85%. The reason is that the amplitude of the arm is larger for the arm raise and one-handed wave compared with the other four actions, and the status of the other hand is uncertain when waving with one hand, thus reducing the accuracy of the algorithm [31, 32]. In addition, the arm raise and one-hand wave movements may have certain deviations due to the distance between the human body and the camera and the different shooting angles, resulting in recognition errors. Therefore, the recognition accuracy of these two actions is low. In addition, the difference of data set is also the reason for the low recognition rate of arm raising and one-hand waving by this algorithm. One arm open and arm raise movements are less frequent than other movements in dance, so they have some influence on the recognition accuracy of the algorithm.

The recognition accuracy of each dance movement is shown in Table 5. To avoid the influence of chance on the experimental results, the number of test set images for all movements is 100.

Table 5: Action recognition accuracy test results.

Dance action name	Accuracy (%)
Lift	88.62
Sink	92.35
Charge	86.35
Cloud step	72.22
Turn over	85.98
Stand up and shoot the swallow	73.15

Table 6: Comparison of recognition accuracy of different algorithms.

Method	Accuracy (%)
Residual network four-channel algorithm	79.2
Calculating H_u moment algorithm	89.9
Our method	92.7

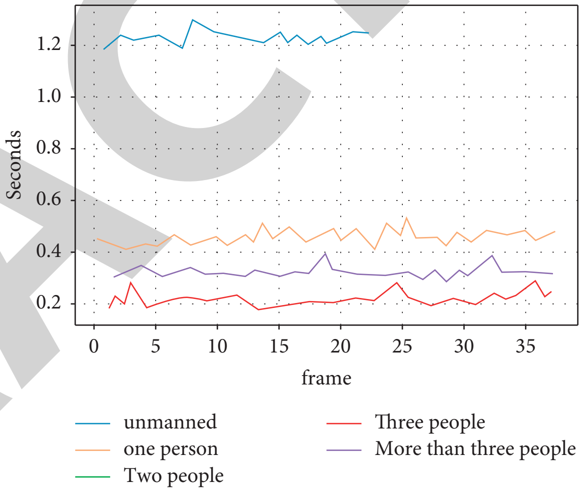


Figure 11: HOC algorithm efficiency.

From the experimental results, it can be seen that the accuracy of the dance action recognition method proposed in this paper is above 70% for all kinds of dance actions, and the highest can be above 90%. Under the condition of similar motion complexity, the accuracy rate of dance motion recognition is higher; under the condition of the same number of images in the training set, the lower the complexity, the higher the accuracy rate of motion recognition.

In order to solve the problem of low recognition accuracy caused by the differences in data sets, this study constructed confusion matrices for the above six dance movements in the data set processing, as shown in Figure 10, to ensure that the number of each movement data set is basically the same, and then used this algorithm for recognition. According to the recognition results, the classification accuracy of all six dance movements reached over 90%, indicating that increasing the number of data sets of arm raising and one-handed waving movements through the confusion matrix can effectively reduce the influence of human differences on the recognition results and improve the recognition accuracy.

5.2. Comparison of Algorithms. As can be seen from Table 6, the accuracy of our algorithm is higher, reaching more than 92%, which indicates that the present algorithm is more ideal for recognition of dance movements and has a higher accuracy rate. In addition, it is known by the experimental time that the present algorithm runs at 0.75 frames/s on the Tesla P4 graphics card and can recognize multi-person movements in a single picture.

To further verify the recognition efficiency of the algorithm, we tested it in scenes with 0 to multiple people, respectively, and found that the time spent by the algorithm gradually increases with the number of people in the image, but the magnitude is small. The running time of the Hoff orientation calculator (HOC) algorithm increases linearly with the number of people, as shown in Figure 11. In contrast, the running time of the algorithm in this study essentially did not increase significantly. This indicates that the present algorithm is more efficient and the algorithm performs better.

6. Conclusion

Dance video image recognition should take into account the influence of dance background, costume, etc., on action recognition, as well as the obscuration and self-obscuration in the dancer's own movements, and an action recognition technique that can accurately and completely record and reflect the dancer's action information should be used in order to obtain the dancer's body static information and action information. The practical test verifies the feasibility and high efficiency of the method, and the design can be widely applied to the visual perception interaction between human and service robots in the future, so that the robots can understand human actions better and faster, and engage in general service work according to human behaviors.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

References

[1] Y. Liu, M. Fan, and W. Xu, "Recognition method of dance rotation based on multi-feature fusion," *International Journal of Arts and Technology*, vol. 13, no. 2, pp. 91–107, 2021.

[2] G. Li, Z. Y. Wang, J. Luo, X. Chen, and H. B. Li, "Spatio-context-based target tracking with adaptive multi-feature fusion for real-world hazy scenes," *Cognitive Computation*, vol. 10, no. 4, pp. 545–557, 2018.

[3] A. Zhao, L. Qi, J. Dong, and H. Yu, "Dual channel LSTM based multi-feature extraction in gait for diagnosis of Neurodegenerative diseases," *Knowledge-Based Systems*, vol. 145, pp. 91–97, 2018.

[4] S. Koehne, A. Behrends, M. T. Fairhurst, and I. Dziobek, "Fostering social cognition through an imitation-and synchronization-based dance/movement intervention in adults with autism spectrum disorder: a controlled proof-of-concept study," *Psychotherapy and Psychosomatics*, vol. 85, no. 1, pp. 27–35, 2016.

[5] J. Young, "The therapeutic movement relationship in dance/movement therapy: a phenomenological study," *American Journal of Dance Therapy*, vol. 39, no. 1, pp. 93–112, 2017.

[6] S. C. Koch, L. Mehl, E. Sobanski, M. Sieber, and T. Fuchs, "Fixing the mirrors: a feasibility study of the effects of dance movement therapy on young adults with autism spectrum disorder," *Autism*, vol. 19, no. 3, pp. 338–350, 2015.

[7] E. Shuper Engelhard, "Dance movement psychotherapy for couples (DMP-C): systematic treatment guidelines based on a wide-ranging study," *Body, Movement and Dance in Psychotherapy*, vol. 14, no. 4, pp. 204–217, 2019.

[8] R. Melhuish, C. Beuzeboc, and A. Guzmán, "Developing relationships between care staff and people with dementia through Music Therapy and Dance Movement Therapy: a preliminary phenomenological study," *Dementia*, vol. 16, no. 3, pp. 282–296, 2017.

[9] M. Shim, R. B. Johnson, S. Gasson, S. Goodill, R. Jermyn, and J. Bradt, "A model of dance/movement therapy for resilience-building in people living with chronic pain," *European Journal of Integrative Medicine*, vol. 9, pp. 27–40, 2017.

[10] R. T. H. Ho, J. K. K. Cheung, W. C. Chan, I. K. M. Cheung, and L. C. W. Lam, "A 3-arm randomized controlled trial on the effects of dance movement intervention and exercises on elderly with early dementia," *BMC Geriatrics*, vol. 15, no. 1, pp. 127-128, 2015.

[11] K. E. Raheb, M. Stergiou, A. Katifori, and Y. Ioannidis, "Dance interactive learning systems: a study on interaction workflow and teaching approaches," *ACM Computing Surveys*, vol. 52, no. 3, pp. 1–37, 2020.

[12] S. Lyons, V. Karkou, B. Roe, B. Meekums, and M. Richards, "What research evidence is there that dance movement therapy improves the health and wellbeing of older adults with dementia? A systematic review and descriptive narrative summary," *The Arts in Psychotherapy*, vol. 60, pp. 32–40, 2018.

[13] B. Levine and H. M. Land, "A meta-synthesis of qualitative findings about dance/movement therapy for individuals with trauma," *Qualitative Health Research*, vol. 26, no. 3, pp. 330–344, 2016.

[14] S. Wiedenhofer and P. D. S. C. Koch, "Active factors in dance/movement therapy: specifying health effects of non-goal-orientation in movement," *The Arts in Psychotherapy*, vol. 52, pp. 10–23, 2017.

[15] O. K. N. Streater, "Truth, justice and bodily accountability: dance movement therapy as an innovative trauma treatment modality," *Body, Movement and Dance in Psychotherapy*, vol. 17, no. 1, pp. 34–53, 2022.

[16] R. Preda, "Power dynamics in dance movement therapy," *Body, Movement and Dance in Psychotherapy*, vol. 17, no. 1, pp. 71–80, 2022.

[17] S. Lotan Mesika, H. Wengrower, and H. Maoz, "Waking up the bear: dance/movement therapy group model with depressed adult patients during Covid-19 2020," *Body, Movement and Dance in Psychotherapy*, vol. 16, no. 1, pp. 32–46, 2021.

[18] K. Michels, O. Dubaz, E. Hornthal, and D. Bega, "Dance Therapy" as a psychotherapeutic movement intervention in Parkinson's disease," *Complementary Therapies in Medicine*, vol. 40, pp. 248–252, 2018.

[19] O. Alemi, J. Françoise, and P. Pasquier, "GrooveNet: real-time music-driven dance movement generation using artificial neural networks," *Networks*, vol. 8, no. 17, p. 26, 2017.

[20] T. Hens and K. F. Dunphy, “Developing participants’ capacity for reflection and self-assessment in a dance movement therapy program for people with intellectual disability,” *Disability & Society*, vol. 37, no. 2, pp. 271–295, 2022.

[21] X. Ning, W. Li, B. Tang, and H. He, “BULDP: biomimetic uncorrelated locality discriminant projection for feature extraction in face recognition,” *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2575–2586, 2018.

[22] M. F. Leung and J. Wang, “A collaborative neurodynamic approach to multiobjective optimization,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5738–5748, 2018.

[23] Q. Liu, C. Liu, and Y. Wang, “etc. Integrating external dictionary knowledge in conference scenarios the field of personalized machine translation method [J],” *Journal of Chinese Informatics*, vol. 33, no. 10, pp. 31–37, 2019.

[24] P. An, Z. Wang, and C. Zhang, “Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection,” *Information Processing & Management*, vol. 59, no. 2, Article ID 102844, 2022.

[25] R. Ali, M. H. Siddiqi, and S. Lee, “Rough set-based approaches for discretization: a compact review,” *Artificial Intelligence Review*, vol. 44, no. 2, pp. 235–263, 2015.

[26] R. Ali, S. Lee, and T. C. Chung, “Accurate multi-criteria decision making methodology for recommending machine learning algorithm,” *Expert Systems with Applications*, vol. 71, pp. 257–278, 2017.

[27] G. Cai, Y. Fang, J. Wen, S. Mumtaz, Y. Song, and V. Frascolla, “Multi-carrier M-ary DCSK system with code index modulation: an efficient solution for chaotic communications,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 6, pp. 1375–1386, Oct, 2019.

[28] K. Chandra, A. S. Marcano, S. Mumtaz, R. V. Prasad, and H. L. Christiansen, “Unveiling capacity gains in ultradense networks: using mm-wave NOMA,” *IEEE Vehicular Technology Magazine*, vol. 13, no. 2, pp. 75–83, June 2018.

[29] S. Palanisamy, B. Thangaraju, O. I. Khalaf, Y. Alotaibi, S. Alghamdi, and F. Alassery, “A novel approach of design and analysis of a hexagonal fractal antenna array (HFAA) for next-generation wireless communication,” *Energies*, vol. 14, no. 19, p. 6204, 2021.

[30] S. Nagi Alsubari, S. N Deshmukh, A. Abdullah Alqarni et al., “Data analytics for the identification of fake reviews using supervised learning,” *Computers, Materials & Continua*, vol. 70, no. 2, pp. 3189–3204, 2022.

[31] A. Radwan, M. F. Domingues, and J. Rodriguez, “Mobile caching-enabled small-cells for delay-tolerant e-Health apps,” in *Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 103–108, IEEE, Paris, France, May 2017.

[32] F. B. Saghezchi, A. Radwan, J. Rodriguez, and T. Dagiuklas, “Coalition formation game toward green mobile terminals in heterogeneous wireless networks,” *IEEE Wireless Communications*, vol. 20, no. 5, pp. 85–91, 2013.