SIN-CHN SCIENTIFIC PRESS PTE.LTD.

*Molecular & Cellular Biomechanics* 2024, 21(4), 675.
https://doi.org/10.62617/mcb675

Article

# Biomechanical analysis and application of an anti-fuzzy decomposition method for sports dance movement images based on multi-attention

**Tang Tang[1], Junjian Zheng[2], Lei Wang[3], Fei Yang[4,*]**

[1] College of Sports and Health, Guangdong Polytechnic of Science and Technology, Zhuhai 519090, China
[2] Police Physical Education and Research Department, The National Police University for Criminal Justice, Baoding 071000, China
[3] Department of Physical Education, Tangshan Normal University, Tangshan 063000, China
[4] Department of Sports Art, Hebei Institute of Physical Education, Shijiazhuang 050000, China
*** Corresponding author:** Fei Yang, 2008006@hepec.edu.cn

**CITATION**

Tang T, Zheng J, Wang L, Yang F. Biomechanical analysis and application of an anti-fuzzy decomposition method for sports dance movement images based on multi-attention. Molecular & Cellular Biomechanics. 2024; 21(4): 675. https://doi.org/10.62617/mcb675

**ARTICLE INFO**

Received: 30 October 2024
Accepted: 14 November 2024
Available online: 20 December 2024



**Abstract:** In the realm of biomechanics, accurate analysis of sports dance movements plays a crucial role in understanding human motion patterns and optimizing athletic performance. To address the challenge of analyzing sports dance error movement images that often suffer from high segmentation difficulty due to blurriness, we propose an anti-fuzzy decomposition method based on multi-head attention. Firstly, the Multi-scale Retinex automatic color enhancement algorithm is employed to correct the color of sports dance error action images, as color correction can enhance the visual clarity which is essential for subsequent biomechanical feature extraction. Subsequently, an anti-fuzzy decomposition model of sports dance error action images based on Diffusion Model-U-shaped network (DM-Unet) is constructed. The corrected image is divided into image blocks by a block embedding layer and then input into the encoder which is constructed by the confrontation generation network. The encoder selects the residual network not only to extract image features but also to capture biomechanically relevant details such as joint positions, limb orientations, and body postures. The multi-head attention module is utilized to suppress the dynamic blur of the image, which helps in precisely identifying the key movement elements during sports dance. Moreover, through down-sampling operations, the dimension of image features is reduced while retaining the essential biomechanical information. The decoder then uses the up-sampling module to restore the encoder output results to the original size. Through the global residual connection module, the features of each layer of the encoder and decoder are connected, enabling the retention of shallow features of the image that are significant for analyzing the fine-grained biomechanical aspects of sports dance movements. The comprehensive loss function is used to train the model, and the anti-fuzzy decomposition results of sports dance error action images are outputted. The experimental results show that this method can effectively decompose the wrong action image of sports dance into structure and texture parts. Importantly, the peak signal-to-noise ratio of the decomposed image being higher than 26dB indicates enhanced clarity for further biomechanical analysis. For example, these decomposed images can be used to study the impact of different movement errors on joint torques, muscle activations, and overall body balance in sports dance, providing valuable insights for coaches and biomechanics researchers to improve training programs and understand injury mechanisms.

**Keywords:** multiple attention; sports dance; biomechanical analysis; erroneous movements; anti-fuzzy image; decomposition methods; biomechanical application

## 1. Introduction

In the field of sport dance, the accuracy and aesthetics of dance movements are important criteria for judging dancers' skills [1]. In actual training and performance, due to rapid movement changes, insufficient light, jittery equipment and other factors, dance videos or images are often blurred, which not only affects the clarity of the dance movements [2], but also increases the difficulty of accurately analyzing and correcting the dance movements. Traditional dance video analysis methods mostly rely on the uncalibrated global visual feedback method, i.e., analyzing dance movements by video decomposition training [3]. This method is often difficult to accurately capture the visual changes of key nodes when dealing with fast-changing dance movements, resulting in problems such as large 3D visual motion tracking errors and unclear decomposition contours. Especially in the case of image blurring [4], these problems are more prominent, severely limiting the accuracy and effectiveness of dance movement analysis and correction. The development of a method that can effectively combat image blurring and improve the accuracy of dance movement decomposition is of great significance to enhance the effect of sports dance training and promote the skills of dancers [5]. As an effective feature filtering and enhancement technique, multi-attention mechanism has been widely used in many fields. The introduction of multi-attention mechanism realizes the effective extraction and enhancement of key action information in fuzzy images, suppresses the interference of non-key information, and thus improves the accuracy and efficiency of image decomposition [6]. Using the multi-attention model, the fuzzy dance image after preprocessing is subjected to anti-fuzzy decomposition to improve the accuracy and efficiency of image decomposition. Based on the clear image after decomposition, combined with the dance movement specification, the wrong movement of sports dance is automatically identified and analyzed. Combined with the professional guidance of the teacher, a targeted corrective program is formulated to help dancers quickly correct the wrong movements and improve their skills.

Chouksey et al. [7] studied a multivariate optimized color image decomposition method based on variable modulus decomposition. This method comprehensively utilizes the color, texture, shape and other feature information in color images to make the decomposition results more accurate and precise. Variational model allows to organically combine the feature information of the image itself, the location, shape and size of the target and other prior knowledge into one energy function, which can adapt to different types of image decomposition tasks. The energy function is minimized by graph cut optimization and other methods to improve the image decomposition effect. However, this method needs to process multiple feature information in color images. Although it can make full use of color information, it is sensitive to noise. In the noisy image, the segmentation effect may be affected. Harbi et al. [8] used a prior method of depth variation to decompose images. This method can automatically learn and extract the advanced features of the image from a large amount of data, more reflect the essential content of the image, and is conducive to more accurate map image decomposition. This method introduces flexible prior knowledge such as image statistical properties and structural information, which is helpful to guide the image decomposition process. But this method needs a lot of labeled data for training, and the generalization ability of the model depends on the diversity and representativeness of the training data. Although the method introduces prior knowledge to improve the image decomposition effect, the accuracy and integrity of prior knowledge also have an important impact on the decomposition results. If the prior knowledge is inaccurate or missing, the accuracy of image decomposition will be affected. Arco et al. [9] realized image decomposition through multimodal mode based on Siam neural network with self-attention mechanism. Self-attention mechanism can effectively fuse the complementary information in multimodality and improve the accuracy of image decomposition. Self-attention mechanism can capture long-distance dependencies in images and find global and local features in images. By dynamically adjusting the weight of different regions or features, the model pays more attention to the regions that have an important impact on the decomposition results, and improves the decomposition accuracy. But the self-attention mechanism needs to calculate the interaction between all positions in the image, especially when processing high-resolution neuroimaging data. Siam neural network needs to carefully design loss function, optimization algorithm and regularization technology to ensure that the model can converge stably and avoid over fitting. Baslamisli et al. [10] performs image decomposition through the fine-grained color decomposition method. The fine-grained color decomposition can capture the features inside the image more finely, which is suitable for application scenarios requiring high accuracy. Fine granularity decomposition can deeply understand the internal structure and composition of images, which is helpful for subsequent image processing and analysis tasks. However, this method needs to deal with a large number of fine features in the image, and the performance of fine-grained decomposition is highly dependent on the quality and feature distribution of the input image. If the image quality is poor or the feature distribution is uneven, the accuracy and reliability of decomposition will be directly affected.

Based on the above analysis, it can be concluded that existing technologies have significant limitations in handling image decomposition tasks with fast motion blur and low lighting conditions. For example, although the method based on variational mode decomposition can comprehensively utilize various feature information of images, it is sensitive to noise and difficult to maintain segmentation effects in blurry images with high noise; Although the prior method of deep mutation can learn advanced features of images, its generalization ability is limited by the training data, and the accuracy and completeness of prior knowledge are crucial to the decomposition results; Although Siamese neural networks with self-attention mechanisms can improve decomposition accuracy, they require a large amount of computation and careful design for model stability and convergence; Although fine-grained color decomposition methods can capture image features finely, they are highly dependent on input image quality and feature distribution. Aiming at the problems of the above methods in image decomposition, a multi head attention based anti blur decomposition method for sports dance erroneous action images is proposed. This article effectively suppresses dynamic blur and improves decomposition accuracy through strategies such as multi-scale color correction, DM Unet model construction, and global residual connection. It is particularly suitable for image decomposition under fast motion and low light conditions, demonstrating higher robustness and practicality. This provides strong support for the recognition and correction of erroneous movements in sports dance training, highlighting the necessity and innovation of research.

# 2. Anti-fuzzy decomposition method for sport dance error movement images

## 2.1. Color correction of sports dance error movement images

The input sports dance error movement image is pre-processed with denoising and contrast enhancement to improve the quality of the sports dance error movement image [11], which lays the foundation for the subsequent image segmentation. The color correction algorithm realizes the color balance of the sports dance error movement image. The original sports dance error movement image has the situation that the red light decays too fast, and the color deviation phenomenon is serious. In order to solve this problem [12], the red channel of the blurred sports dance error action image is compensated, so that the green channel can preserve more visual information, and part of the green channel is added to the red channel to compensate for the attenuation of the red light, so as to restore the spectral information in the original sports dance error action image. The expression of the red channel compensation of the sports dance wrong movement image is as follows:

$$I_R(x) = I_R(x) + \left(\bar{I}_G - \bar{I}_R\right)[1 - I_R(x)]I_G(x) \quad (1)$$

Among them, $I_R$ and $I_G$ are red and green channels for the sports dance error movement image; the $\bar{I}_R$ and $\bar{I}_G$ are the pixel mean of $I_R$ and $I_G$.

The multi-scale Retinex automatic color enhancement algorithm is used to correct the color of the compensated sports dance error action image. This method is based on the spatial position relationship between the image color and the light and dark areas [13], and carries out local adaptive filtering to achieve the brightness and color balance of the sports dance error action image. Automatic color enhancement algorithm based on Retinex theory by estimating the incident light part $L$ of sports dance error action image, the expression for the reflected portion of the image $R$ is as follows:

$$R_i(x,y) = lg\, I_i\,(x,y) - lg[F(x,y) * I_i(x,y)] \quad (2)$$

Among them, $R_i(x,y)$ indicates the reflective portion of the $i$-th color component; $Ii(x,y)$ indicates the image of the $i$-th color component. * denotes the convolution; $F(x,y)$ is the wrap-around function.

The expression for the wrap-around function in Equation (2) is as follows:

$$F(x,y) = Ke^{-(x^2+y^2)/g^2} \quad (3)$$

Among them, $g$ is the Gaussian parameter; $K$ is the normalization function.

The multiscale Retinex algorithm weights multiple different Gaussian parameters $g$ at different scales to obtain the sum of different reflected brightness of sports dance error action image [14]. The output expression of the multi-scale Retinex algorithm is as follows:

$$R_i = \sum_{n=1}^{N} w_n\, R_{n.i} \quad (4)$$

Among them, $N$ denotes the number of scales. $R_{n.i}$ represents the reflection brightness of the $i$-th channel $n$-th scale; $R_i$ is the reflection brightness obtained by the multiscale weighting of the $i$-th channel; $w_n$ is the weight of the $n$-th scale.

Multiscale Retinex algorithm with color restoration, which uses the reflected brightness $R_i$ calculated by multi-scale Retinex multiplied by the color recovery factor $\xi_i$, adjusting the color ratio of the 3 channels of the sports dance error movement image [15]. The expression of this process is as follows:

$$R = \xi_i R_i \tag{5}$$

$$\xi_i = lg[\alpha \cdot I_i(x, y)] - lg\left[\sum_{i=1}^{3} I_i(x, y)\right] \tag{6}$$

Among them, $a$ is a constant factor between 0.5 and 2.

For the resulting $R$, find for its mean $u$ and standard deviation $\sigma$, $R$ is normalized by $R^*$, is expressed as follows:

$$R^* = \frac{1}{255(\mu - \alpha\sigma)} \tag{7}$$

The physical significance of this process is that in the logarithmic space, the original sports dance error movement image is subtracted from the Gaussian function and the convolution result of the original image, and the smooth part of the original image is subtracted. The smaller the Gaussian function, the more low-frequency components in the sports dance error action image are removed, and the high-frequency components of the remaining image [16], which can better highlight the detailed information in the original sports dance error action image. For the serious blurred region in the sports dance error movement image, this process can realize the enhancement of sports dance error movement image, and effectively highlight the details in the dark region of the image.

## 2.2. DM-Unet based anti-fuzzy decomposition model of sports dance wrong action image

In the current field of image processing, Transformer based models perform well in natural language processing tasks, but they may lack accurate modeling of local features in image processing tasks because Transformers do not have explicit spatial information encoding. The anti-blur decomposition task of sports dance erroneous action images studied in this article requires both global contextual information and fine segmentation of local features. Therefore, using Transformer based models may not be sufficient to meet accuracy requirements. The DM-Unet model, also known as the diffusion model based on the U-Net structure, combines the advantages of diffusion models and U-Net network structures. The diffusion model decomposes the generation process in sequence and is implemented based on denoising autoencoders to generate high-quality results on sports dance erroneous action images. As a classic convolutional neural network structure, U-Net is known for its powerful feature extraction and fusion capabilities. By combining the two, the DM-Unet model has demonstrated excellent performance in image decomposition tasks. Therefore, in DM

Unet, a multi head attention mechanism is introduced to effectively screen and enhance key action information, suppress the interference of non key information, achieve anti blur decomposition of sports dance erroneous action images, and improve the accuracy of sports dance erroneous action image decomposition.

The overall structure of DM-Unet model for anti-fuzzy decomposition of sports dance error action images is shown in **Figure 1**.

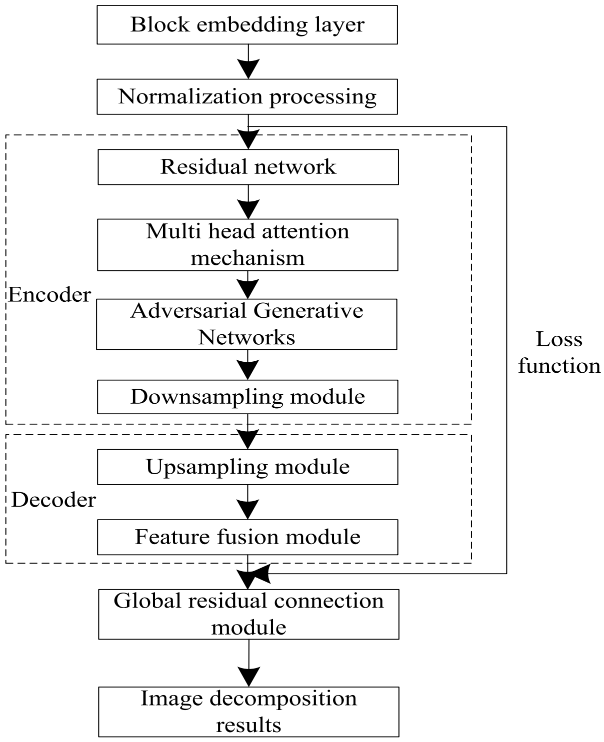



**Figure 1.** Overall structure diagram of the model.

It can be seen from the overall structure diagram of anti-fuzzy decomposition model of sports dance error action image in **Figure 1** that the anti-fuzzy decomposition model of sports dance error action image based on DM Unet mainly includes block embedding layer, confrontation generation network, multi head attention mechanism, encoder composed of down sampling module, decoder composed of up sampling module and feature fusion module The global residual connection module is composed. Use the block embedding layer to enhance the input sports dance error action image $x$ divide into non-overlapping image blocks of size 2 × 2, the sports dance error movement images were dimensionally mapped to obtain the image blocks $x'$ and normalized. The sports dance error movement image block is inputted into the encoder, which processes the sports dance error movement image using an adversarial generative network, and utilizes the multi-attention module to suppress the dynamic blurring that may exist in the sports dance error movement image, and to highlight the sports dance error movement. The height and width of the input features are reduced by downsampling operation, and the number of channels is multiplied at the same time. In the decoder section, the corresponding size of the dance error movement image is restored by the up-sampling module. The global residual linkage connects the layers of encoder and decoder to reduce the gradient vanishing and network degradation problems, and retains the organization details of the shallow features better, and outputs the result of anti-blurring decomposition of the sports dance error movement image.

## 2.3. Adversarial generative network-based encoder construction

### 2.3.1. Residual network based feature extraction of sports dance error movement images

DM-Unet model uses residual network to extract the features of sports dance error action images. Residual network solves the problem of gradient explosion and disappearance in the depth deepening of convolutional neural network. The residual network structure based on ResNet 101 is shown in **Figure 2**.

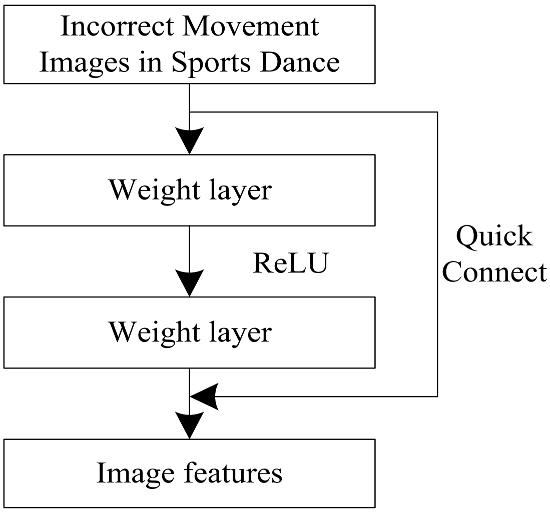



**Figure 2.** Residual network structure diagram.

The residual network draws a fast connection from the input dance error action image *X*. The link is used to add the original image and the features processed by two layers of convolution. Finally, the image features *H*(*X*) of sports dance error actions are obtained through ReLU function. Use the ResNet 101 model trained on the training dataset, modify the full connection layer parameters of the ResNet 101 model, and output the number of types of sports dance error action image tags [17]. Keep the input parameters unchanged, modify the output parameters, and obtain the feature extractor of multi label sports dance error action images. After cutting the size of the wrong action image of sports dance $I_i$ into 224 × 224, input it to ResNet101 feature extractor to obtain the feature tensor of multi tag sports dance error action image $f_i$. The equation for the above process is as follows:

$$f_i = f_{Re\,sNet}(I_i; \theta_{Re\,sNet}) \tag{8}$$

Among them, $f_{Re\,sNet}$ represents ResNet101 image feature extractor, $I_i$ indicates the wrong movements in physical education and dance of the $i$-th images. $\theta_{Re\,sNet}$ indicates the parameters of ResNet101 image feature extractor.

In order to fusion the attention weights of sports dance error action image features and labels when sports dance error action image is anti-fuzzy decomposition, and to match the dimension of the matrix of sports dance error action image features and label attention weights [18], it is necessary to make dimensionality reduction to the feature tensor of the multi-label sports dance error action image $f_i$. In order to maximize the retention of the original information and simplify the model, the length and width of the feature tensor of the wrong movement image of sports dance that set in the feature extraction module $f_i$ are both 1, and only for the number of channels $D$ to perform dimensionality reduction. The feature vector $x_i$ after dimensionality reduction is obtained by a convolution operation with one convolutional layer, the expression is as follows:

$$x_i = f_{conv1}(f_i; \theta_{conv1}) \tag{9}$$

Among them, $f_{conv1}$ represents the convolution layer conv1, $\theta_{conv1}$ indicates the parameters of the convolution layer conv1.

The above process is used to complete the feature extraction and dimensionality reduction processing of sports dance error movement images, and the processing results are used as the input of the multi-head self-attention mechanism.

### 2.3.2. Multi self-attention mechanisms

The multi-head attention module is utilized to suppress the dynamic blurring that may exist in the sports dance error movement image and to highlight the sports dance error movement. Multi-head self-attention infers the correlation between each pixel of the sports dance error action image from each subspace separately, and obtains the feature map with global vision, which makes the semantic representation of the sports dance error action image richer by stacking the resultant feature maps corresponding to different self-attention heads [19]. Assuming that the number of attention heads of the multi-head self-attention mechanism is $N$, such that the input sports dance error movement characterization map is $X \in R^{D\times h\times w}$, of which $D$ is the number of channels, $h$ and $w$ are the height and width of the feature map, respectively. By dimensionality adjustment and linear transformation to make $X$ transform into 3 matrices, which are the query matrix $Q$, key matrix $K$, value matrix $U$, $Q$, $K$and $U$ are divided by columns into $N$ copies, each of which are denoted, respectively, as $Q^n$, $K^n$ and $U^n$, and are the query matrix, key matrix and value matrix of the $n$-th attention head. The $n$-th output head $O^n$ is:

$$O^n = SA(Q^n, K^n, U^n) = soft\,max\left(\frac{Q^n(K^n)^T}{\sqrt{k}}\right)U^n \tag{10}$$

In the equation, $SA$ is the self-attention function.

Equation (10) consists of two parts, namely similarity matrix calculation and output head calculation. The similarity matrix is normalized by SoftMax, and the output head is obtained by multiplying the similarity matrix and the value matrix [20]. Stack the sports dance error action feature map of each output head to generate the output of multi head self-attention as follows:

$$MHSA(Q, K, U) = Concat([O^1, \cdots, O^i, \cdots, O^N]) \tag{11}$$

In the equation, $MHSA$ is the multinomial self-attention function, $Concat$ is the stacking function.

In multi headed self-attention, the inner product operation between elements makes self-attention have a global receptive field in essence, and is good at capturing long-distance dependence. Since the self-attention mechanism only calculates the correlation between different pixel features in the wrong action image of sports dance, it does not consider the location information of each pixel, lacks the spatial perception ability, and the output will lose structural information. The commonly used absolute position coding will lose important relative position relationship when calculating SoftMax. Therefore, the relative position coding is used to retain the structure information of sports dance error action images. The relative position encoding method sets the $(2h-1) \times (2w-1)$ trainable relative position information matrix $P^n$ for each attention head to represent the $(2h-1) \times (2w-1)$ relative position relationship, and integrates the relative position information by rewriting Equation (10). The expression for the above process is as follows:

$$O^n = soft\,max\left(\frac{Q^n(K^n)^T}{\sqrt{k}} + Z^n\right)U^n \tag{12}$$

In the equation, $Z^n$ is the positional correlation between individual pixels. The weight values of each relative position of the sports dance error movement image are learned by gradient backpropagation. Due to the integration of relative position information, the multi-attention mechanism is able to capture the long-distance dependencies while preserving the structural information in the input feature map.

#### 2.3.3. Adversarial generative network construction based on multi-residual blocks

The encoder of the DM-Unet model is composed of multiple residual blocks. Each residual block contains a group of parallel convolutions to increase the network depth. The 4-way parallel design with different convolution sizes enables the residual block to extract more abundant image features of wrong movements in sports dance from the receptive field of the same size. The Inception structure is introduced to increase the depth of the network, while taking into account the optimization of the width of the model, to obtain the optimal image decomposition results of sports dance error actions. The confrontation generation network relies on the learning of a small number of samples, which is different from the solution of super-resolution of sports dance wrong action images first, and then decomposition of sports dance wrong action images. During training, when the confrontation generation network synthesizes high-resolution sports dance error action images, it generates more realistic super-resolution sports dance error action image decomposition results through the judgment of the discriminator. Train a generating function $G$, based on the input images of incorrect movements in sport dancing $I_n$, output the corresponding sports dance error movement decomposition image. Denote the parameters of the generator feed-forward convolutional neural network as $G_\psi$, of which $\psi = \{W_L; b_L\}$ denotes the weights and biases of network for the $L$-th layers. With the set loss function $I_G$, the model is optimized to learn and obtain the corresponding $I_n^F$. Optimize $\psi$ through Equation (13), bringing the image decomposition results closer to the real labels $I_n^{Mask}$.

$$\hat{\psi} = argmin\frac{1}{N}\sum_{n=1}^{N} l_G\left(G_\psi(I_n), I_n^{Mask}\right) \tag{13}$$

In order to make the generator of the confrontation generation network cheat the discriminator better, the parameters are continuously optimized through the confrontation to generate accurate decomposition results of the wrong action image of sports dance. The generator of the anti-generation network uses the design of residual blocks. Inception contains four 1 × 1 convolution blocks, two 3 × 3 convolution blocks, and one 5 × 5 convolution block. By using convolution kernels of different sizes, the generator can obtain more information about wrong action images of sports dance in the same receptive field, normalize the generated wrong action images of sports dance, activate them with ReLU function, and output the decomposition results of wrong action images of sports dance.

## 2.4. Comprehensive loss function for anti-fuzzy decomposition modeling of sports dance error movement images

Based on the above analysis, it can be concluded that the DM Unet model segments the input image into non overlapping small blocks through a block embedding layer, and performs dimension mapping and normalization processing. Subsequently, these image blocks are fed into an encoder consisting of an adversarial generation network, multi head attention mechanism, and downsampling module. Among them, the multi head attention mechanism divides the input feature map into multiple subspaces and infers the correlation between pixels in the image from each subspace, thereby capturing feature maps with a global view. This process enriches the semantic representation of incorrect action images in sports dance by stacking feature maps of different self-attention heads. More importantly, the multi head attention mechanism can suppress the interference of non-critical information, such as background noise or irrelevant actions, while enhancing critical action information, such as details and features of erroneous actions. In addition, by introducing relative position encoding, the mechanism captures long-range dependencies while preserving the structural information in the input feature map, ensuring the accuracy and reliability of the decomposition results. The encoder utilizes residual networks to deeply extract image features and effectively suppresses dynamic blur through multi head attention modules, highlighting key action information. The downsampling operation further reduces the feature dimension, making it easier for subsequent processing. In the decoder section, the upsampling module restores the image size, the feature fusion module integrates the features of each layer, and the global residual connection reduces the problems of gradient vanishing and network degradation, while preserving the shallow features of the image. Finally, the model outputs sports dance error action images after anti fuzzy decomposition.

Considering the particularity of anti-fuzzy decomposition of sports dance wrong action images, the comprehensive loss function of anti-fuzzy decomposition of sports dance wrong action images was constructed with the goal of accurately reconstructing the brightness range of sports dance wrong action images, improving color authenticity and eliminating noise. The comprehensive loss function is mainly composed of $L_1$ loss function, structure similarity loss function, perception loss function and perception color loss function. Its expression is as follows:

$$L = (1 - \varsigma_s - \varsigma_p - \varsigma_c)L_1 + \varsigma_s L_s + \varsigma_p L_p + \varsigma_c L_c \tag{14}$$

Among them, $L_1$ denotes pixel-level paradigm loss; $L_s$ denotes the loss of structural similarity; $L_p$ denotes perceived loss. $L_c$ denotes the perceived color loss; $\varsigma_s$, $\varsigma_p$ and $\varsigma_c$ indicate the corresponding loss factor.

Use loss function $L_1$ reduce the difference between the decomposition result of the sports dance error movement image and the original sports dance error movement image, improves the contrast of the contour and the smoothness of the uniform region in the sports dance error movement image, and improves the image decomposition performance of the model. The expression for the loss function $L_1$ is as follows:

$$L_1 = \sqrt{\left\|I_g - I_h\right\|^2} \tag{15}$$

Among them, $I_{gt}$ represents real sports dance error movement images; $I_h$ indicates the result of decomposition of sports dance error movement images.

The structural similarity loss function measures the structural loss of the original sports dance error movement image and the decomposed sports dance error movement image in terms of brightness, contrast and image structure, and restores the structure and local details of the sports dance error movement image. The structural similarity value is set to be in the range of 0–1, and the higher the value is, the better the similarity is. The expression of the structural similarity loss function is as follows:

$$L_s = 1 - \frac{1}{N}\sum_{img}\left(\frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2} \cdot \frac{2\sigma_{xy}}{\sigma_x^2 + \sigma_y^2}\right) \tag{16}$$

Among them, $u$ and $\sigma^2$ denote the mean and standard deviation of the pixel mean and standard deviation of the sports dance error movement images, respectively.

Structural perceptual loss constraints are used to constrain the difference between the original image and the decomposition result of the sports dance error movement image to maintain the perceptual and detailed realism of the sports dance error movement image, while maintaining the perceptual and semantic fidelity. The expression of structure-aware loss constraint is as follows:

$$L_p(I_g, I_h) = \frac{\varphi}{H_j W_j}\left\|(I_g) - (I_h)\right\| \tag{17}$$

Among them, $I_g$ represents an image of the original sports dance error movement; $I_h$ denotes the result of decomposition of sports dance error movement images; $H_j$ and $W_j$ denote respectively the height and width of the $j$-th feature map. $\varphi$ represents the feature maps of incorrect movement images in sports dance.

The perceptual color loss constraint is used to constrain the color difference in Euclidean space between the image decomposition result and the original sports dance error action image, so as to make the decomposition result have a similar color with the reference sports dance error action image, and to ensure the color consistency of the decomposition result. The expression of the perceptual color loss constraint is as follows:

$$L_c = \Delta E(I_g, I_h) \tag{18}$$

Among them, $I_g$ represents an image of the original sports dance error movement; $I_h$ represents the image of the decomposed sports dance error movement; $\Delta E$ indicates color chromatic aberration.

Using the set loss function, the constructed anti-blur decomposition model of sports dance error movement images is trained to improve the anti-blur decomposition performance of sports dance error movement images.

## 3. Experimental analysis

In order to verify the anti-fuzzy decomposition performance of the researched method on sports dance error movement images, the method was applied to the teaching of sports dance majors in a sports college, and the results of the decomposition of sports dance error movement images were utilized to analyze the sports dance error movement images in detail, and to guide the teaching of sports dance majors.

The experiments set the parameters of this paper's method as follows:

Image resolution: All input sports dance error movement images are uniformly adjusted to a resolution of 512 × 512 pixels.

Color correction: Use multi-scale Retinex automatic color enhancement algorithm, set the number of scales to 3 (that is, use three different scales of Retinex processing), and the weight of each scale is 0.33 (average distribution).

Block embedding layer: The corrected image is divided into image blocks of 64 × 64 pixels with a step size of 32 pixels (the overlapping part helps to retain more contextual information).

Residual network layers: ResNet-18 is used as the basic network, including 4 residual blocks, and each residual block contains 2 convolution layers.

Multi head attention module: Set the number of heads to 8, and the dimension of each head to 64 (the total feature dimension is 512, matching with the output of the last convolution layer of ResNet-18).

Downsampling operation: Use the largest pooling layer, pooling window size is 2 × 2, step size is 2.

Up-sampling module: Up-sampling is performed using a transposed convolutional layer, each up-sampling doubles the size of the feature map until it is restored to the original image size.

Feature fusion: After each upsampling step of the decoder, the feature map of the corresponding layer of the encoder is spliced with the feature map of the current layer of the decoder through jump connection (Concat operation).

Connection method: The input image (before processing through the block embedding layer) is directly connected to the last layer of the decoder, and the output of the decoder is summed to preserve the shallow features of the image.

Content loss (L1 loss): Used to measure the pixel level difference between the output image and the clear image, with a weight of 0.8.

Adversarial loss (cross-entropy loss): Used to train the discriminator in the adversarial generative network to ensure that the generated image is visually indistinguishable from the real image, with a weight of 0.1.

Perception loss (VGG network based on pre training): Used to measure the similarity between the output image and the clear image in high-level features, with a weight of 0.1.

Optimizer: With the Adam optimizer, the learning rate is initialized to 0.0001. After every 10 epochs, the learning rate decreases to 0.9 times of the original.

Batch Size: Set to 8.

Training rounds: 50 epochs in total.

Data Enhancement: Random horizontal flip, random rotation (angle range −10° to 10°), random cutting (cut to 480 × 480 pixels and then resize to 512 × 512).

A randomly selected image of an original sport dance error movement of a sport dance student is shown in **Figure 3**.

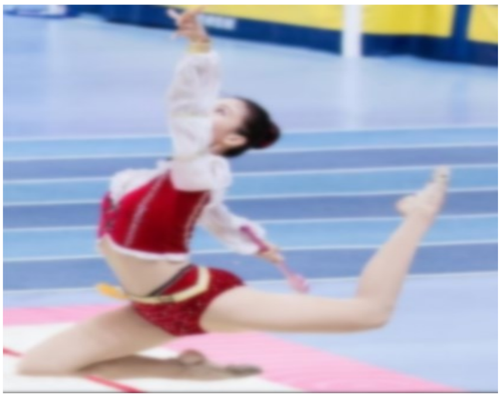


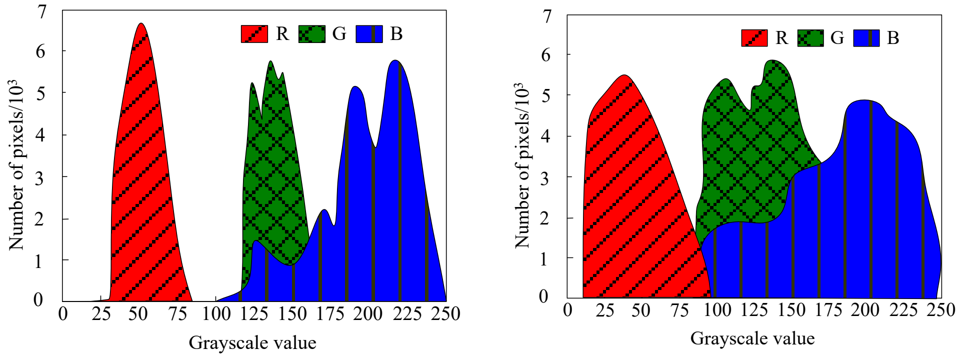
**Figure 3.** Incorrect movement image of original sports dance.

The method of this paper is used to enhance the color balance of the original sports dance wrong movement images collected. The results of the enhancement process of the sports dance error movement image are shown in **Figure 4**.

**Figure 4.** Image enhancement results of incorrect actions in sports dance.

Using this paper's method of sports dance error movement image enhancement processing, the image does not have obvious texture and other details of the loss of features, **Figure 4** verifies that this paper's method of sports dance error movement enhancement processing has a high degree of effectiveness, and can effectively improve the quality of sports dance error movement image. Through the image enhancement process to improve the quality of sports dance error movement image, so that it is more suitable for the subsequent image decomposition task.

Before and after the enhancement process, the histogram changes of the wrong movement images of sports dance were counted, and the statistical results are shown in **Figure 5a,b**.

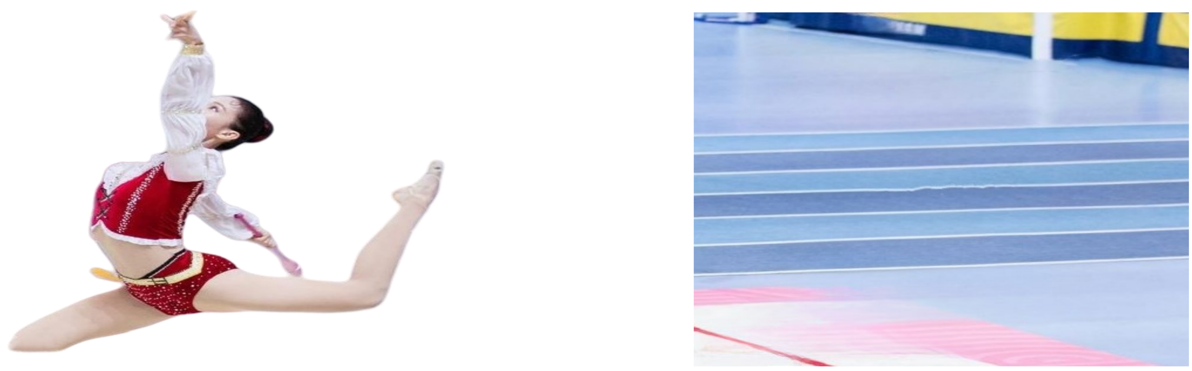



(**a**) Before enhancement. (**b**) After enhancement.

**Figure 5.** Histogram changes of images before and after enhancement processing.

The experimental results of **Figure 5** show that the overall gray level of the image R, G and B channels has been significantly enhanced after the image enhancement processing of sports dance error actions. Before image enhancement processing of sports dance error actions, the red channel of the image is concentrated in the low gray value area of 0–100. After the enhancement processing, the gray values of the red channel are evenly distributed to different gray levels. The enhanced image has high color fidelity, good visual sense, higher background discrimination, and can highlight the wrong action information of sports dance in the image.

The anti-fuzzy decomposition of sports dance error movements is carried out by using the method of this paper, and the anti-fuzzy decomposition results of sports dance error movement images are shown in **Figure 6a,b**.



(**a**) Structural image. (**b**) Texture image.

**Figure 6.** Anti fuzzy decomposition results of sports dance movements.

As can be seen from the experimental results in **Figure 6**, the method in this paper can effectively distinguish the structure and texture of the sports dance error movement image, and realize the effective decomposition of the sports dance error movement image. The decomposition of the sports dance error movement image by this method can effectively reduce the "ladder effect" in the sports dance error movement image, and the texture part of the sports dance error movement image can be maintained better.

Through the objective evaluation method, the decomposition performance of the method in this paper on the wrong action image of sports dance is verified. The image quality is evaluated through the underwater color image quality evaluation (UCIQE) index of the image decomposition of wrong movements in sports dance. UCIQE uses the weighted combination of chromaticity, saturation and clarity in the Lab color space to evaluate the quality of the decomposed sports dance error action image. The value range of the UCIQE index is [0,1], which is proportional to the quality of the decomposed dance error action image. The higher the evaluation value, the better the balance performance of the result between chroma, saturation and clarity. The information entropy can describe the information richness of the decomposed sports dance error action image. The clear sports dance error action image has a large information entropy value. The evaluation value can be obtained by averaging the information entropy values of different channels. The UCIQE value and information entropy evaluation results of the decomposition of sports dance error action images using this method are shown in **Table 1**.

**Table 1.** UCIQE value and information entropy.

| Image Number | Before decomposition | | After decomposition | |
|---|---|---|---|---|
| | UCIQE | Information entropy | UCIQE | Information entropy |
| 1 | 0.56 | 4.85 | 0.86 | 7.15 |
| 2 | 0.61 | 5.61 | 0.89 | 7.26 |
| 3 | 0.49 | 4.19 | 0.91 | 7.35 |
| 4 | 0.61 | 4.64 | 0.87 | 7.29 |
| 5 | 0.59 | 4.58 | 0.88 | 7.46 |
| 6 | 0.72 | 5.16 | 0.81 | 7.29 |
| 7 | 0.64 | 5.74 | 0.92 | 7.41 |
| 8 | 0.29 | 5.34 | 0.87 | 7.37 |
| 9 | 0.46 | 5.19 | 0.82 | 7.85 |
| 10 | 0.49 | 4.28 | 0.93 | 7.23 |

**Table 1** shows the UCIQE evaluation index and information entropy evaluation value of sports dance error action images. It can be seen from the data in **Table 1** that the comparison results of the two evaluation indicators verify the effectiveness of the algorithm applied to underwater sports dance error action images. In 10 image samples, the UCIQE value after decomposition is significantly higher than that before image decomposition. This shows that the image decomposition process has significantly improved the image quality, and the image has been significantly improved in terms of color, contrast and clarity. The information entropy value after

image decomposition is generally higher than that before decomposition, which indicates that the decomposition process increases the amount of information in the image, making the image contain more details and changes.

Using the method of this paper to decompose the sports dance error movement image, the peak signal-to-noise ratio of the decomposed image is counted, and the peak signal-to-noise ratio is used to verify the anti-blurring performance of this paper's method for the decomposition of the sports dance error movement image, and the statistical results are shown in **Figure 7**.

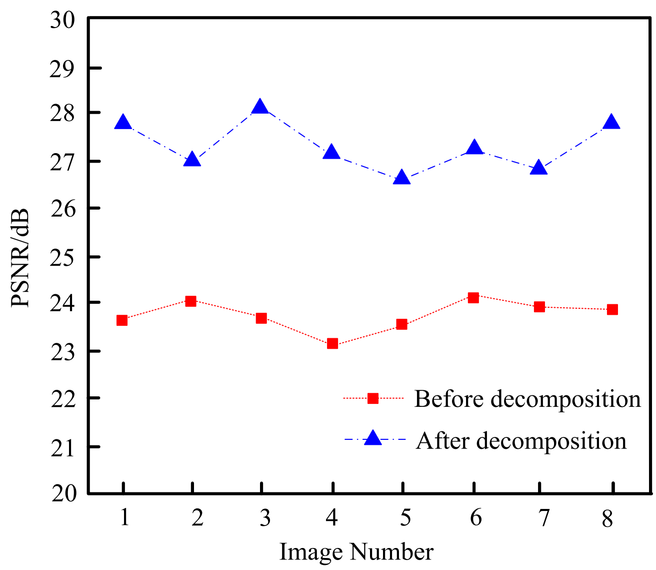



**Figure 7.** Peak signal to noise ratio.

By comparing the original sports dance error action image in **Figure 7** with the image decomposed using the method proposed in this paper, it can be clearly seen that the peak signal-to-noise ratio of the image has been significantly improved. Specifically, the peak signal-to-noise ratio of the decomposed image is generally higher than 26dB, even reaching the highest value of 28dB. This result not only verifies the high anti blur performance of the proposed method, but also effectively improves the image quality, significantly improving the blurring problem that is prone to occur in the decomposition process of sports dance erroneous action images. The experimental data in **Figure 7** fully demonstrates that after adopting the method proposed in this paper, the corrected image is divided into image blocks through block embedding layers and input into the encoder constructed by the adversarial generative network. The encoder uses a residual network to extract image features and utilizes a multi head attention module to suppress dynamic blur in the image. This step can effectively remove the fuzzy components in the image, and the clarity of the image has been significantly improved, providing a more accurate and reliable basis for subsequent image analysis and error action recognition.

The image decomposition performance of this paper's method is measured by detecting the number of corner points in an image. The neighborhood where the corner points are located is usually a stable and information-rich region in the image. By comparing the number of corner points of different algorithms for the decomposition of sports dance error images, it can effectively reflect the enhancement of contrast and clarity of different methods for the decomposition of sports dance error images. The number of corner points decomposed by different methods for sports dance error movement images is counted, and the statistical results are shown in **Table 2**.

**Table 2.** Number of corner points in image decomposition.

| Image Number | Before decomposition | Rotate | Translation |
| --- | --- | --- | --- |
| 1 | 1568 | 1685 | 1705 |
| 2 | 1238 | 1362 | 1405 |
| 3 | 1384 | 1405 | 1469 |
| 4 | 1405 | 1459 | 1497 |
| 5 | 1364 | 1376 | 1405 |
| 6 | 1185 | 1205 | 1285 |
| 7 | 1267 | 1364 | 1395 |
| 8 | 1294 | 1359 | 1387 |
| 9 | 1305 | 1345 | 1395 |
| 10 | 1375 | 1405 | 1439 |

As can be seen from the experimental results in **Table 2**, the number of corner points in the decomposition result of the anti-fuzzy decomposition of the sports dance error movement image using the method of this paper is significantly higher than that of the image before decomposition. The method of this paper can avoid the influence of image changes such as rotation and translation on the decomposition results. The experimental results in **Table 2** verify that the method in this paper can improve the decomposition performance of sports dance error movement images with high anti-fuzzy performance.

In order to further verify the effectiveness of the method proposed in this paper, the variational mode decomposition method [7], deep mutation prior method [8], and Siamese neural network [9] mentioned in the introduction were used as comparison methods. The anti blur decomposition performance of different methods for sports dance incorrect action images was analyzed using image sharpness, contour extraction accuracy, and decomposition success rate as indicators. The results are shown in **Table 3**.

**Table 3.** Analysis of image decomposition performance of different methods.

| Method name | Sharpness | Contour extraction accuracy | Decomposition success rate |
| --- | --- | --- | --- |
| This paper's method | 0.925 | 0.943 | 0.968 |
| Variational Mode Decomposition [7] | 0.853 | 0.881 | 0.895 |
| Depth variation prior [8] | 0.891 | 0.907 | 0.922 |
| Siamese Neural Network [9] | 0.876 | 0.912 | 0.934 |

According to **Table 3**, the image clarity of our method reached 92.5%, which is higher than the other three methods. This indicates that the method proposed in this article can more effectively restore the detailed information of the image and improve the clarity of the image in the anti-blur decomposition process. In terms of contour extraction accuracy, this method achieved 94.3%. This means that the method proposed in this article can more accurately extract the contour information of incorrect motion images in sports dance, providing strong support for subsequent motion analysis and correction. This method also achieved the highest success rate in decomposition, reaching 96.8%. This indicates that the method proposed in this article has higher stability and reliability in processing blurry sports dance erroneous action images, and can successfully decompose more key action information. In summary, through comparative experiments with the other three methods, it can be concluded that the proposed multi head attention based anti-lur decomposition method for sports dance erroneous action images performs well in terms of image clarity, contour extraction accuracy, and decomposition success rate, outperforming other methods. This experimental result validates the effectiveness and superiority of our method in processing blurry sports dance erroneous action images.

In order to evaluate the practical usability of decomposed images by dance professionals, this article invited 10 experienced dance professionals to evaluate the decomposed images. The main evaluation content is the clarity and accuracy of the decomposed sports dance erroneous movement images, as well as whether they contribute to the correction and improvement of dance movements. During the evaluation process, a 5-point scale is used, where 5 points indicate very usable, 4 points indicate relatively usable, 3 points indicate generally usable, 2 points indicate not very usable, and 1 point indicates completely unusable. The results are shown in **Table 4**.

**Table 4.** Evaluation of actual usability of decomposed images.

| Evaluator ID | Clarity rating | Action accuracy rating | Helps correct dance movements and improve ratings |
|---|---|---|---|
| 1 | 4.5 | 4.8 | 4.7 |
| 2 | 4.2 | 4.6 | 4.5 |
| 3 | 4.7 | 4.9 | 4.8 |
| 4 | 4.0 | 4.3 | 4.2 |
| 5 | 4.6 | 4.7 | 4.6 |
| 6 | 4.3 | 4.5 | 4.4 |
| 7 | 4.8 | 5.0 | 4.9 |
| 8 | 4.1 | 4.4 | 4.3 |
| 9 | 4.4 | 4.7 | 4.6 |
| 10 | 4.2 | 4.6 | 4.5 |

According to **Table 4**, all evaluators rated the clarity of the decomposed images as above 4 points, with an average score of 4.35 points. This indicates that the decomposed image performs well in terms of clarity, and can clearly present the key details of dance movements. In terms of action accuracy, evaluators generally score high, with an average score of 4.6. This indicates that decomposing images can accurately reflect the characteristics and key points of dance movements, which is helpful for dance professionals to conduct accurate movement analysis and correction. When evaluating whether decomposed images contribute to the correction and improvement of dance movements, the evaluator also gave a high rating, with an average score of 4.55. This indicates that decomposing images has practical application value in dance teaching and training, which can help dance professionals better understand and correct errors and deficiencies in dance movements. In summary, it can be concluded from this experiment that the decomposed sports dance error motion images have been highly recognized by dance professionals in terms of clarity, motion accuracy, and practical application value. This experimental result validates the practical usability of decomposed images in dance teaching and training, providing strong support for the correction and improvement of dance movements.

In order to further verify the deblurring effect of the proposed method, experiments were conducted under different lighting conditions and action complexities, with the recognition accuracy of erroneous actions and the tracking accuracy of action key points as indicators, using Wiener filtering and Lucy Richardson deconvolution as comparative methods. The results are shown in **Table 5**.

**Table 5.** Deblurring effect under different conditions.

| **Conditions** | | **Recognition accuracy** | | | **Tracking accuracy/pixel** | | |
|---|---|---|---|---|---|---|---|
| | | **Wiener filtering** | **Lucy-Richardson deconvolution** | **This paper's method** | **Wiener filtering** | **Lucy-Richardson deconvolution** | **This paper's method** |
| Lighting conditions | Strong light | 0.82 | 0.87 | **0.93** | 2.5 | 2.2 | **1.0** |
| | Weak light | 0.78 | 0.82 | **0.90** | 2.8 | 2.5 | **1.4** |
| Action complexity | Simple actions | 0.88 | 0.90 | **0.95** | 1.8 | 1.6 | **0.8** |
| | Complex actions | 0.80 | 0.84 | **0.88** | 2.4 | 2.1 | **1.6** |

According to **Table 5**, under different lighting conditions and motion complexity, our method exhibits significant advantages in deblurring performance compared to Wiener filtering and Lucy Richardson deconvolution. In strong light environments, the recognition accuracy of our method is as high as 93.0%, which is 11 percentage points higher than Wiener filtering. The tracking accuracy also reaches 1.0 pixel, far lower than Wiener filtering's 2.5 pixels and Lucy Richardson deconvolution's 2.2 pixels. Even under low light conditions, the recognition accuracy of our method remains at 89.5%, higher than the other two methods, and the tracking accuracy is 1.4 pixels, better than Wiener filtering's 2.8 pixels and Lucy Richardson deconvolution's 2.5 pixels. In addition, when dealing with simple and complex actions, the recognition accuracy of our method reached 95.0% and 88.0%, respectively, both higher than Wiener filtering and Lucy Richardson deconvolution, and the tracking accuracy was also more accurate. These results indicate that the proposed method can effectively improve the recognition and tracking accuracy of images under various conditions, and its deblurring performance is significantly better than the other two methods.

## 4. Conclusion

An anti-fuzzy decomposition method based on multi-head attention is proposed for sports dance error movement images, aiming at solving the problem of difficult movement analysis due to fuzzy sports dance images. The multi-attention mechanism can automatically learn and recognize the key action regions in the image, focus on these regions, effectively suppress the interference of blurring and non-critical information, and accurately capture the details of dance movements in the blurred image, which provides a strong support for the subsequent recognition of erroneous movements. This method not only improves the image clarity, but also preserves the integrity of the key movement information, which provides high-quality image data for the subsequent analysis and correction of dance movements. Through the proposed image decomposition method, dancers can see their own dance movements more clearly during the training process and find and correct the errors in time, which helps to improve the dancers' self-correcting ability and promotes the rapid improvement of their skills. This method also provides coaches with a more objective and accurate assessment tool, which helps to develop a more scientific and reasonable training program.

Although the anti blur decomposition method for sports dance erroneous action images based on multi head attention has shown significant decomposition effects in theory, accurately identifying and decomposing the structure and texture parts in blurry images, and the peak signal-to-noise ratio of the decomposed image is as high as 26dB or more, it still faces some potential challenges in practical applications. Especially in real-time sports dance training scenarios, the real-time performance and computational resource consumption of this model have become urgent issues to be addressed. Due to the complex multi head attention mechanism and deep neural network structure involved in the model, its computational complexity is high, which may result in slow processing speed and difficulty in meeting the requirements of real-time feedback. In addition, the consumption of high-performance computing resources may also limit the widespread application of this method in resource limited environments. Therefore, in order to fully utilize the potential of this model in practical sports dance training scenarios, we will further optimize the model structure, reduce computational complexity, improve processing speed, and explore effective ways to maintain high performance under limited computing resources in the future. These optimizations and adjustments will be of great significance in promoting the implementation and popularization of this method in practical applications.

**Author contributions:** Conceptualization, FY and TT; methodology, TT; formal analysis, TT, JZ and LW; investigation, JZ; resources, TT; data curation, TT and JZ; writing—original draft preparation, TT; writing—review and editing, TT and JZ; supervision, FY; project administration, FY. All authors have read and agreed to the published version of the manuscript.

**Ethical approval:** Not applicable.

**Conflict of interest:** The authors declare no conflict of interest.

## References

1. Jain C, Kumar A, Chugh A. Efficient image deblurring application using combination of blind deconvolution method and blur parameters estimation method. ECS transactions. 2022; 107(1): 3695–3704.
2. IvanovB, Milovanovi GV, Stanimirovi PS. Accelerated dai-liao projection method for solving systems of monotone nonlinear equations with application to image deblurring. Journal of Global Optimization. 2022; 85(2): 377–420.

3. Chairatsiripong C, Yambangwai D, Thianwan T. New iterative methods for nonlinear operators as concerns convex programming applicable in differential problems, image deblurring, and signal recovering problems. Mathematical Methods in the Applied Sciences. 2023; 46(3): 3332–3355.
4. Rioux G, Choksi R, Hoheisel T, Maréchal P, Scarvelis C. The maximum entropy on the mean method for image deblurring. Inverse Problems. 2021; 37(1): 1015–1022.
5. Filippozzi R, Rabelo JC, Boiger R, Leito A. A range-relaxed criteria for choosing the lagrange multipliers in the iterated tikhonov kaczmarz method for solving systems of linear ill-posed equations. Inverse Problems. 2021; 37(4): 045005 (20pp).
6. Li XX, Li XJ. An Adaptive Residual Dense Network Image Deblurring Method Based on GAN. Computer Simulation. 2022; 39(11): 457–462.
7. Chouksey M, Jha RK. A multiverse optimization based colour image segmentation using variational mode decomposition. Expert Systems with Applications. 2021; 171: 114587.1–114587.36.
8. Harbi JS, Najm ZN. Image decomposition using deep variation priors system. Nonlinear Optics, Quantum Optics. 2022; 56(3/4): 193–204.
9. Arco JE, Ortiz A, Gallego M, Nicolás J, Górriz JM, Ramírez J. Enhancing multimodal patterns in neuroimaging by siamese neural networks with self-attention mechanism. International Journal of Neural Systems. 2023; 33(4): 2350019–2350027.
10. Baslamisli AS, Das P, Le HA, Karaoglu S, Gevers T. Shadingnet: image intrinsics by fine-grained shading decomposition. International Journal of Computer Vision. 2021; 129(8): 2445–2473.
11. Parvaz, R. Point spread function estimation for blind image deblurring problems based on framelet transform. The Visual Computer. 2023; 39(7): 2653–2669.
12. Iqbal J, Sidhu MS. Acceptance of dance training system based on augmented reality and technology acceptance model. Virtual Reality. 2021; 26(1): 33–54.
13. Li W, Resmerita E, Vese LA. Multiscale hierarchical image decomposition and refinements: qualitative and quantitative results. SIAM Journal on Imaging Sciences. 2021; 14(2): 844–877.
14. Pereyra M, Vargas-Mieles LA, Zygalakis KC. The split gibbs sampler revisited: improvements to its algorithmic structure and augmented target distribution. SIAM Journal on Imaging Sciences. 2023; 16(4): 2040–2071.
15. Verma V, Kumar S, Rani N. Novel image encryption algorithm using hybrid 3d-icpcm and hessenberg decomposition. Nonlinear Dynamics. 2024; 112(12): 10583–10609.
16. Dogani J, Khunjush F, Mahmoudi MR, Seydali M. Multivariate workload and resource prediction in cloud computing using cnn and gru by attention mechanism. Journal of supercomputing. 2023; 79(3): 3437–3470.
17. Das S, Ghosal S. Unmixing aware compression of hyperspectral image by rank aware orthogonal parallel factorization decomposition. Journal of Applied Remote Sensing. 2023; 17(4): 046509.1–046509.21.
18. Tiwari A, Srivastava VK. Novel schemes for the improvement of lifting wavelet transform-based image watermarking using schur decomposition. Journal of supercomputing. 2023; 79(12): 13142–13179.
19. Agrawal M, Jalal AS, Sharma H. Enhancing visual question answering with a two-way co-attention mechanism and integrated multimodal features. Computational intelligence. 2024; 40(1): e12624.1–e12624.22.
20. Dhalia SJ, Bhuvaneshwari R, Bhagya N, Bavya DN. Robust fingerprint reconstruction using attention mechanism based autoencoders and multi-kernel autoencoders. Applied Intelligence. 2024; 54(17–18): 8262–8277.