




# Image Super-Resolution Reconstruction in Sports Scenarios and Its Application in Motion Analysis

Zheng Yu 

College of Art and Media, Shangqiu University, Shangqiu 476000, China

Corresponding Author Email: 000322@sqxy.edu.cn

Copyright: ©2024 The author. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410249>

## ABSTRACT

**Received:** 21 October 2023

**Revised:** 19 January 2024

**Accepted:** 15 March 2024

**Available online:** 30 April 2024

### Keywords:

*image super-resolution reconstruction, motion scenarios, dynamic adaptive cascaded network, 3D motion scene imaging, motion analysis*

With the rapid development of sports technology, the demand for high-definition images in sports competition analysis has been increasing. Particularly in fast-paced sports such as basketball, traditional image capture technology often fails to provide sufficient detail resolution, limiting in-depth analysis of athletic techniques and tactical layouts. To address this, image super-resolution reconstruction technology has been extensively studied and applied to enhance image quality, thereby providing coaches and analysts with clearer visual materials. However, existing super-resolution methods mainly focus on static images and struggle to overcome the challenges of blurring and real-time processing demands in motion scenarios. This paper introduces a dynamic adaptive cascaded network-based method for super-resolution reconstruction of images in motion scenarios, combined with dynamic 3D motion scene imaging techniques, aimed at enhancing the accuracy and timeliness of motion analysis. Through these innovative methods, not only can image degradation caused by motion be effectively handled, but higher-dimensional data support can also be provided for motion analysis.

## 1. INTRODUCTION

In modern sports competitions and analysis, the demand for high-definition images is growing, especially in fast-paced and dynamic sports scenarios like basketball [1-3]. The rapid movements and complex scenes in basketball games pose high demands on image capture equipment, and conventional video capture technology often fails to provide sufficient resolution for detailed analysis of subtle motion skills and tactical layouts [4, 5]. Therefore, using image super-resolution reconstruction technology to process low-resolution images to restore high-resolution details has become a key technique to improve the quality and efficiency of basketball game analysis [6-8].

Image super-resolution reconstruction technology in motion scenarios not only can significantly enhance image quality, providing coaches and analysts with clearer visual materials, but also can assist in more precise analysis of athletic techniques and performance evaluation of players [9, 10]. Additionally, high-quality image reconstruction is of great value for various applications such as automatic video editing, athlete tracking, and replay of highlight moments in matches [11, 12]. The development and application of this technology contribute to the advancement of sports technology and promote innovation in sports training and competition strategies.

However, existing image super-resolution reconstruction methods face many challenges in motion scenarios. These methods often rely on static image processing techniques and struggle to effectively handle image degradation issues caused by motion blur and camera shake [13-15]. Moreover,

traditional algorithms usually have high computational demands and poor real-time performance when dealing with high-dynamic scenes, which are not suitable for sports event analysis that requires quick feedback [16-21]. Therefore, developing super-resolution techniques optimized for motion scenarios has become an important research direction in this field.

The main research contents of this paper include two parts: firstly, the motion scenario image super-resolution reconstruction technology based on the dynamic adaptive cascaded network, and secondly, the dynamic 3D motion scene imaging technology aimed at motion analysis. Firstly, through the dynamic adaptive cascaded network, this study aims to solve the image blur problem caused by rapid movement in motion scenarios, improving the capability of image detail restoration. Secondly, by constructing dynamic 3D motion scene images, this paper further enhances the dimensions and accuracy of motion analysis. These studies are expected not only to promote the application of super-resolution technology in the field of sports but also to provide more precise and real-time data support for motion analysis, thus playing an important role in formulating sports training and competition strategies.

## 2. IMAGE SUPER-RESOLUTION RECONSTRUCTION IN MOTION SCENARIOS

To adapt to the image blur and motion changes common in high-speed motion scenarios such as basketball games, this

study proposes a lightweight dynamic adaptive cascaded network specifically designed for image super-resolution reconstruction needs in motion scenarios. This network utilizes a dual-path residual learning mechanism, with one path responsible for deeply extracting image texture details, and the other filtering out redundant information to enhance the information interaction between the two paths, thereby effectively improving the restoration accuracy of textures and edges. Additionally, the network shares some convolutional parameters in the dual-path residual blocks through vertical parallelism and introduces learnable parameters to dynamically adjust the weights of the shared convolutions. This not only reduces the model's parameter count but also enables the convolution to better adapt to the nonlinear mapping between original features and target features, enhancing the ability to capture texture details in complex motion scenarios. This design is particularly suited for dealing with image blur problems caused by rapid movement and, compared to traditional static image super-resolution reconstruction methods, can more effectively handle image degradation caused by motion, making it more feasible and effective in real-time sports event analysis and high-dynamic environments.

## 2.1 Network structure

The lightweight dynamic adaptive cascaded network proposed in this study is composed of  $V$  dynamic adaptive cascaded modules, which are connected in series through residual connections to enhance learning depth and maintain the stability of the information flow. The specific architecture is shown in Figure 1. Each cascaded module contains  $L$  dual-path residual blocks and  $L-1$  dynamic adaptive module layers, a structural design that makes the network more efficient and precise in extracting mid-to-high-frequency information in complex motion scenarios. The dual-path residual blocks specifically handle the texture and edge details of the image, while the dynamic adaptive module layers adjust and optimize the expression of these features to better adapt to the image changes caused by motion. Additionally, the network introduces original low-resolution image features through global skip connections, ensuring the integrity and coherence of information throughout the network structure from input to output.

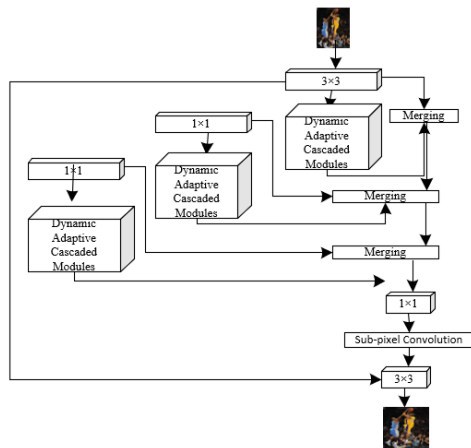


Figure 1. Dynamic adaptive cascaded network structure

In this study, a lightweight dynamic adaptive cascaded network is proposed, specifically designed for image super-resolution reconstruction in motion scenarios. The first step involves inputting a low-resolution image through a  $3 \times 3$  convolutional layer to extract shallow features. This step is basic but crucial as it sets the baseline for subsequent complex feature extraction. In motion scenarios, due to rapid movement and potential blur, shallow features include key initial visual information. Suppose the  $3 \times 3$  convolution operation is represented by  $d_i$ , the computation process for shallow features is given by the following equation:

$$D_0 = d_i(a) \quad (1)$$

The second step involves processing the extracted shallow features through  $N$  dynamic adaptive cascaded modules to obtain deep features. Each dynamic adaptive cascaded module includes multiple dual-path residual blocks and dynamic adaptive module layers, which are specifically designed to handle image changes and detail degradation caused by rapid movement. This structure allows the network to adapt to the constantly changing scene dynamics, optimizing the feature extraction process, thus enabling more accurate restoration of details in motion scenarios. Suppose the first dynamic adaptive cascaded module is represented by  $d_f^1$ , the computation process for deep features is given by the following equation:

$$D_1 = \text{CONCAT}(d_f^1(D_0), D_0) \quad (2)$$

The third step involves merging the deep features extracted from the dynamic adaptive cascaded modules with the original shallow features. Through this method, the model maintains the integrity of information during propagation, preventing the loss of important features. This step is particularly important as maintaining the coherence and integrity of images when processing high-speed motion is key to improving reconstruction outcomes. The merged features will then be further processed as input through  $V-1$   $1 \times 1$  convolution layers and more dynamic adaptive cascaded modules. Suppose the  $u$ -th  $1 \times 1$  convolution operation and dynamic adaptive cascaded module are represented by  $d_s^u$  and  $d_f^u$ , the output of the  $u$ -th layer can be calculated by the following equation:

$$D_u = \text{CONCAT}(d_f^u(d_s^{u-1}(D_{u-1})), D_{u-1}) \quad (3)$$

The final step involves converting the merged features into the final super-resolution image through a  $1 \times 1$  convolution layer. This step not only retains the rich information from input to output but also extracts more texture details through the network's powerful nonlinear mapping capability, laying the groundwork for the final image reconstruction. This process is particularly suitable for motion scenarios as it emphasizes restoring details lost from rapid movement, making the reconstructed image better suited for motion technique analysis and event replay. Suppose the final  $1 \times 1$  convolution operation is represented by  $d_s^v$ , the expression for the extracted deep features is given by the following equation:

$$D_d = d_s^v(D_v) \quad (4)$$

## 2.2 Dynamic adaptive cascaded modules

In the lightweight dynamic adaptive cascaded network designed in this study, the dynamic adaptive cascaded module is a core component, optimized for image super-resolution reconstruction in motion scenarios. The specific architecture is shown in Figure 2. This module combines dual-path residual blocks and dynamic adaptive modules, using a stacked arrangement to enhance the nonlinear mapping capability and feature extraction efficiency when processing motion images. The dual-path residual blocks focus on precisely capturing the texture details and edge information in the image, while the dynamic adaptive modules adjust and optimize the expression of these features, especially adapting better to image blur and deformation caused by motion. Additionally, the module uses a  $1 \times 1$  convolution as a transition layer after each feature enhancement, not only helping to reduce the number of parameters but also ensuring the effective transfer of important features, thus preventing a decrease in reconstruction quality due to feature loss in fast motion scenarios.

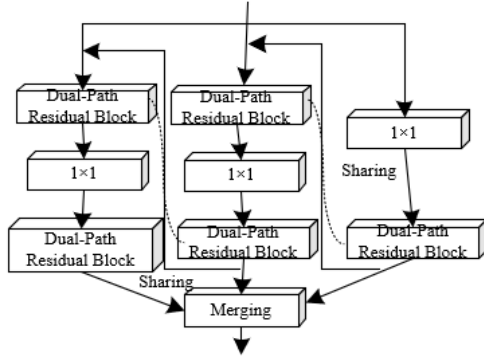


Figure 2. Dynamic adaptive cascaded module structure

The information flow transmission method of the dynamic adaptive cascaded module is specially designed to optimize the effect of image super-resolution reconstruction in motion scenarios. Information first passes through the initial  $3 \times 3$  convolutional layer for preliminary feature extraction, then enters multiple dynamic adaptive cascaded modules. Within each cascaded module, dual-path residual blocks are responsible for processing and refining features from two different perspectives; one path focuses on extracting texture and edge details, while the other filters and optimizes these features. The dynamic adaptive module adjusts the convolution parameters according to the specific features of the current image, making feature processing more precise. The information flow between cascaded modules is maintained through residual connections, ensuring that information is not lost from input to each layer's output, while also enhancing the network's nonlinear mapping capability of features. Let the input of the dynamic adaptive cascaded module be represented by  $D_v$ , the  $k$ -th  $1 \times 1$  convolution operation by  $d_z^s$ , the  $(k-1)$ -th dynamic adaptive module by  $d^{t-1}_e$ , and the  $k$ -th dual-path residual block by  $g_k$ . The information flow transmission of the dynamic adaptive cascaded module is given by the following equations:

$$D_1 = g_1(d_z^1(D_0)) \quad (5)$$

$$D_k = g_k(d_z^u(d_e^{k-1}(D_{k-2} + D_{k-1}))) \quad (6)$$

Finally, the results of the cascading are merged as the output of the dynamic adaptive cascaded module, represented by  $b$ .

$$b = \text{CONCAT}(D_1, D_2, \dots, D_l) \quad (7)$$

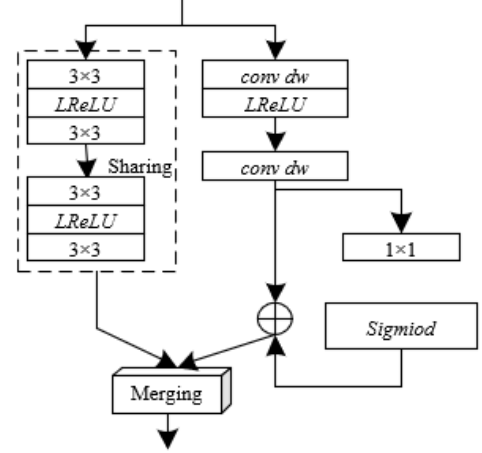


Figure 3. Dual-path residual block structure

In the lightweight dynamic adaptive cascaded network developed in this study, the dual-path residual block serves as a core module, effectively extracting and processing image features in motion scenarios through a carefully designed dual-path parallel strategy to achieve super-resolution reconstruction. The specific architecture is shown in Figure 3. In the upper path of the dual-path residual block, depthwise separable convolution is used to extract low-frequency features. Depthwise separable convolution splits traditional convolution into two independent operations: depth convolution and pointwise convolution, which not only significantly reduces the model's parameter count but also lowers computational complexity. When processing images in motion scenarios, this type of convolution can more efficiently handle large uniform areas in the image, which is very useful for fast-moving object backgrounds, thus maintaining the integrity of the image structure and reducing motion-induced blur. Let the input to the dual-path residual block be represented by  $a$ , the ReLU activation function by  $\omega$ , and the depthwise separable convolution by  $d_{fq}$ . The information transmission of the upper path of the dual-path residual block is given by the following equation:

$$a_1 = d_{fq}^2 \omega d_{fq}^1(a) \quad (8)$$

In terms of feature weight adjustment, depthwise separable convolution is followed by a pixel attention mechanism. Through the calculated attention map, feature weights for each pixel are dynamically adjusted. This step is particularly important in motion scenarios, as it highlights details of moving objects while suppressing background noise, enhancing the quality and clarity of the reconstructed image. Let the Sigmoid activation function be represented by  $\delta$ , then the computation is as follows:

$$b_1 = s_1 \times \delta(d_1(a_1)) \quad (9)$$

The lower path of the dual-path residual block is equipped with two serial residual blocks, mainly used to capture high-

frequency features and rich texture information. This pathway specifically processes details in the image, such as the movement of athletes' equipment and motion details, which are essential for subsequent motion analysis and technical assessment. Let the two residual blocks of the lower path be represented by  $d_e$ , then the computation is as follows:

$$b_2 = d_e(a) \quad (10)$$

Ultimately, the features from the upper and lower paths are merged, effectively integrating low-frequency and high-frequency features. The merged feature set not only contains basic structural information of the image but also includes detailed information about edges and textures, which is crucial for restoring image blur caused by motion. This merging mechanism ensures that whether in static or fast-moving scenes, images can be reconstructed with high quality, meeting the high demands of motion analysis in terms of accuracy and efficiency. Let the output of the dual-path residual block be represented by  $b$ , then the computation formula is as follows:

$$b = \text{CONCAT}(b_1, b_2) \quad (11)$$

This learning method allows the dual-path residual block to extract high-frequency features while preserving some necessary low-frequency information.

In the network proposed in this paper, the dynamic adaptive module is designed to address key challenges in image super-resolution reconstruction under motion scenarios. The specific architecture is shown in Figure 4. This module is based on the concept of dynamic convolution kernels, which can extract rich image features with a reduced number of network parameters. The dynamic adaptive module is particularly suitable for motion scenarios, as it uses an attention mechanism to dynamically adjust the weights of each convolution kernel, thereby accurately addressing image blurring and deformation caused by rapid movement. This attention mechanism is similar to traditional channel attention, but it differs in that it controls the weight distribution through the Softmax function, ensuring that the weights are between 0 and 1 and that the sum of the weights is 1. This allows the module to adjust the focus of the convolution operations based on the dynamic changes of each specific scene, enhancing the network's ability to capture motion details.

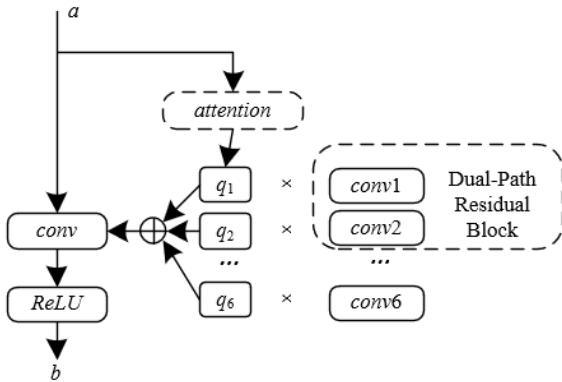


Figure 4. Dynamic adaptive module structure

Specifically, the module enhances the representational capability of the convolution kernels through a nonlinear attention mechanism, dynamically adjusting the weights of

each kernel. The module combines four existing convolution kernels from the lower path of the dual-path residual block and two newly added convolution kernels, which participate in the computation as sub-kernels. Each sub-kernel weights the features according to its weights, optimizing the feature extraction process. This design allows the kernels to learn "vertical" feature parameters on top of the existing horizontal feature extraction, greatly enhancing the utilization of the kernels and the overall efficiency of the model. In motion scenarios, this mechanism is particularly crucial as it can more accurately handle image changes caused by rapid movement, such as blurring and detail loss, ensuring the clarity and richness of detail in the reconstructed images.

First, the dynamic adaptive module receives input features and compresses them through global average pooling to obtain the global features of each channel. This operation is designed to reduce the number of parameters and increase computational efficiency, while also capturing global information that is often lost in motion scenarios, as described below. The channel-level features extracted by global average pooling provide the basis for the next step of feature activation, ensuring that features are transmitted and optimized throughout the network without interference from rapid movements.

$$i = D_{tw}(a) = \frac{1}{G \times Q} \sum_{u=1}^G \sum_{k=1}^Q a(u, k) \quad (12)$$

Next, the Excitation operation is performed, which is achieved through two fully connected layers that learn and adjust the global features. The first fully connected layer compresses and reduces the dimensionality of the global features to reduce computational complexity and prevent overfitting; the second fully connected layer restores the features to their original dimensions, allowing the model to learn and enhance the dependencies among the channels. Assuming the Softmax function is represented by  $\delta$ , and the two fully connected layers by  $n_1$  and  $n_2$ , the weights obtained after the Softmax are represented by  $q_1, q_2, \dots, q_6$ :

$$q_1, q_2, \dots, q_6 = \delta(n_2 \text{ReLU}(n_1 i)) \quad (13)$$

Finally, the convolution kernels shared in the lower path of the dual-path residual block and the two newly added kernels are used for the weighted multiplication operation. The combination and adjustment of these kernels allow the module to dynamically adjust its parameters in a nonlinear manner, optimizing for specific features in motion scenarios. This step is crucial for timely adjusting the convolution operations based on the dynamic changes in image content, effectively extracting texture and edge information of objects in motion, significantly enhancing the detail restoration and overall visual quality of the image. Assuming the  $u$ -th shared convolution kernel in the lower path of the dual-path residual block is represented by  $z_f^u$ , and the  $u$ -th new convolution kernel by  $z_v^u$ , then the computation formula is:

$$b = \text{ReLU}((q_1 z_f^1 + q_2 z_f^2 + \dots + q_6 z_v^6) a) \quad (14)$$

## 2.3 Reconstruction module

In the proposed network for image super-resolution

reconstruction in motion scenarios, the reconstruction module employs sub-pixel convolution technology, which is effective in handling image blur and information loss caused by rapid movement. Traditional upsampling methods like bilinear or bicubic interpolation often introduce irrelevant information, thus reducing the quality of the reconstructed image. Conversely, sub-pixel convolution achieves upsampling by rearranging the output feature maps of convolutions, effectively reducing the introduction of unnecessary information and increasing the precision of upsampling. Assuming deep features are represented by  $D_d$ , and the final reconstructed image by  $b$ , the reconstruction module encompassing sub-pixel convolution and  $3 \times 3$  convolution operations is represented by  $\psi$ . The process involving the addition of  $D_d$  to the shallow features  $D_0$ , and processing through the reconstruction module is as follows:

$$b = \psi(D_d + D_0) \quad (15)$$

## 2.4 Loss function

In the context of image super-resolution reconstruction for motion scenarios, choosing the appropriate loss function is crucial for the effectiveness of model training. The network proposed in this article employs the  $L1$  loss function because, compared to the  $L2$  loss function,  $L1$  loss typically converges faster in image reconstruction tasks and more effectively handles anomalous pixel values caused by motion, such as sharp edges in motion blur. The  $L1$  loss function is less sensitive to outliers, which helps better preserve details and reduce blurring when reconstructing fast-moving objects. This choice is particularly suitable for motion scenarios, as these often contain rapidly changing image features, and  $L1$  loss encourages the model to focus on accurately restoring these dynamic details. Assuming the total number of images in the training set is represented by  $V$ , the set of model parameters to be optimized by  $\phi$ , the  $u$ -th low-resolution and high-resolution images by  $U_{ME}^u$  and  $U_{GE}^u$  respectively, and the network model proposed by  $G$ , the functional expression is as follows:

$$M(\phi) = \frac{1}{V} \sum_{u=1}^V \|G(U_{ME}^u) - U_{GE}^u\| \quad (16)$$

## 3. DYNAMIC 3D MOTION SCENE IMAGING FOR MOTION ANALYSIS

In motion analysis, especially in dynamic 3D motion scenarios, traditional methods of 3D mapping such as SLAM face many challenges. These conventional approaches typically assume a static scene, thus struggling to accommodate moving objects, leading to significant inaccuracies in pose estimation and scene tracking. Moreover, when objects in the scene move in a specific direction or plane, traditional methods of dynamic feature extraction, like optical flow and depth uncertainty methods, often fail to effectively identify these dynamic objects. These issues not only limit the application of SLAM technology in dynamic environments but also significantly reduce the accuracy of constructions and the practicality of the system. Therefore, developing an algorithm that can accurately identify and handle dynamic objects has become key to enhancing the precision of 3D image construction in dynamic scenes.

Addressing this issue, this study proposes a new algorithm for dynamic 3D motion scene imaging, which adds a semantic segmentation module and a dynamic extraction module to the foundation of 2D motion scene image super-resolution reconstruction. By incorporating semantic segmentation, the algorithm can identify specific objects in the image and dynamically extract moving objects in the scene using a newly designed geometric method, combining the position and depth information of the objects. This strategy not only improves the accuracy of scene recognition but also allows the remaining static feature points to be more reliably used for scene reconstruction. The system processes dynamic and static elements in this manner, significantly enhancing the accuracy and practicality of 3D motion scene imaging, especially in motion analysis and tracking, providing more stable and detailed scene information.

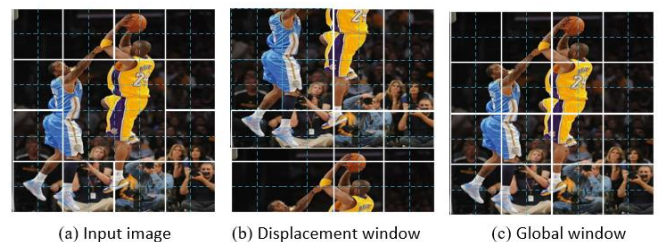
## 3.1 System framework

The system framework of the dynamic 3D motion scene imaging algorithm proposed in this paper is designed to enhance the precision and practicality of motion analysis by effectively addressing the problem of 3D scene reconstruction in dynamic environments. Compared to traditional methods of dynamic 3D map construction, this system specifically focuses on the identification and extraction of dynamic objects within motion scenes, thus providing more accurate static scene information for motion analysis. The implementation steps of this system framework are as follows:

(1) Firstly, the collected RGB images of the motion scene are semantically segmented using the MulAttenNet network to identify various objects in the environment. This step utilizes prior knowledge to identify and extract objects with a high probability of movement, thereby reducing the occlusion and interference these dynamic objects may cause.

(2) Then, ORB features are extracted from the RGB images and matched to roughly compute the camera's pose. ORB feature extraction and matching lay the foundation for subsequent dynamic feature point extraction and pose estimation.

(3) Next, combining the calculated camera pose and depth information, a global window divides the window into different blocks for recombination to establish long-range dependencies that were not established in the shifted window, thereby enhancing the acquisition of global information. This also estimates the motion transformation and depth changes from the reference frame to the current frame. Using this information, the dynamic probability of each feature point can be calculated to determine whether these feature points are in motion and to extract those with high dynamic probabilities. Figure 5 shows examples of input images, shifted windows, and global windows.



**Figure 5.** Input image, displacement window, and global window



(4) Additionally, the system calculates the proportion of dynamic feature points within the semantic segmentation area to determine whether an object area is dynamic. Based on this determination, further extraction of identified dynamic moving objects is performed to reduce their negative impact on the construction of the 3D map.

(5) Finally, the system uses the remaining static feature points for pose estimation and based on these static feature points, builds a local map and performs pose optimization to achieve precise reconstruction of static object scenes.

### 3.2 Dynamic extraction with semantic segmentation

In the construction of dynamic 3D motion scene images for motion analysis, dynamic extraction is a crucial step, directly affecting the accuracy and reliability of the entire scene reconstruction. Existing dynamic SLAM algorithms attempt to extract dynamic feature points using optical flow and depth uncertainty methods, but these techniques have limitations in practical applications, especially in the complex environments of dynamic 3D motion scene image construction. Optical flow utilizes the two-dimensional plane movement of feature points between consecutive frames to identify dynamic feature points. This method judges whether feature points are dynamic by calculating the motion difference of feature points in consecutive images. If an object containing a 3D point is stationary, then the position calculated based on the camera pose should coincide with its actual position; otherwise, these feature points are considered dynamic. However, the effectiveness of optical flow depends on significant movement of the object in the two-dimensional plane, making it difficult to accurately identify objects that move only along the z-axis or move minimally.

Depth uncertainty identifies dynamic points by projecting feature points from keyframes and calculating their depth changes in adjacent frames. This method requires accurate depth information and stable feature point matching but can be unreliable in dynamic scenes, especially when feature point depths are difficult to obtain or the depth difference between consecutive frames is minimal. This situation is common in motion scenarios, such as with fast-moving objects or scenes with subtle depth changes.

Therefore, in the process of constructing dynamic 3D motion scene images for motion analysis, it is necessary to further optimize these dynamic extraction techniques to improve the recognition of various motion patterns and ensure accurate extraction of all dynamic elements from 3D scenes. This requires not only improving existing technologies but also possibly integrating data from multiple sensors or employing more advanced machine learning techniques for a more precise and robust dynamic extraction strategy. Such improvements will directly enhance the quality of 3D motion scene reconstruction, providing more reliable data support for motion analysis.

To effectively address the impact of dynamic objects and optimize the dynamic extraction capabilities of semantic segmentation in the algorithm for constructing dynamic 3D motion scene images for motion analysis, this paper proposes an improved geometric extraction method to precisely identify and extract dynamic feature points in the scene. This method focuses on enhancing the accuracy of scene reconstruction in complex dynamic environments, a common challenge encountered by traditional dynamic SLAM methods such as

ORB-SLAM2 when dealing with dynamic scenes. The core of this method lies in calculating the positional and depth differences of feature points between different frames. Specifically, feature points  $A_{GJ}$  in the current keyframe are projected into the current frame based on previously calculated relative camera poses to obtain their anticipated positions ( $a_{TY}$ ,  $b_{TY}$ ) and depth  $c_{TY}$  in the current frame. By comparing these anticipated values with the actual positions ( $a_{SC}$ ,  $b_{SC}$ ) and depth  $c_{SC}$  measured using the current frame's RGB and depth maps, this method can effectively identify which feature points exhibit significant changes in position or depth due to object movement. Furthermore, by integrating semantic segmentation results, this algorithm can more precisely identify and extract those dynamic areas. Through semantic segmentation, the algorithm first determines the static or dynamic nature of various areas in the scene, then applies the aforementioned geometric extraction method to further confirm and extract dynamic feature points. Assuming the difference in projected position and actual measured position of feature points in the current frame from the keyframe is represented by  $\Delta f$ , and the depth difference by  $\Delta c$ , the computation formula is as follows:

$$\begin{cases} \Delta f = \sqrt{(a_{TY} - a_{SC})^2 + (b_{TY} - b_{SC})^2} \\ \Delta c = |c_{TY} - c_{SC}| \end{cases} \quad (17)$$

If  $\Delta f$  or  $\Delta c$  exceeds a threshold, the feature point is considered to be in motion between the two frames of the motion scene images and is identified as a dynamic feature point. If  $\Delta f$  or  $\Delta c$  is below the threshold, it is identified as a static feature point. Taking into account both the calculated results of whether the feature points are dynamic and the results of semantic segmentation, if 30% of the feature points in each segmented area are determined to be dynamic, the object is considered to be a moving object. In subsequent calculations, this dynamic area is extracted and used for motion analysis.

## 4. EXPERIMENTAL RESULTS AND ANALYSIS

In Table 1, the performance of various super-resolution reconstruction methods for sports scene images is assessed through several key metrics: PSNR- $\mu$  (Mean Peak Signal-to-Noise Ratio), SSIM- $\mu$  (Mean Structural Similarity Index), PSNR-L (Peak Signal-to-Noise Ratio under Low Illumination), SSIM-L (Structural Similarity Index under Low Illumination), and HDR-VDP-2 (High Dynamic Range Visual Difference Predictor). Our model demonstrates superior performance, especially in the PSNR- $\mu$  and HDR-VDP-2 metrics. The proposed model reached 43.6985 dB in PSNR- $\mu$ , which is higher by 1.386 dB and 1.3829 dB compared to other methods like RCAN and RDN, showing better average signal-to-noise ratio. In HDR-VDP-2 scoring, our model leads with a score of 65.1235, indicating significant advantages in handling high dynamic range content. These experimental results fully demonstrate that by dynamically adjusting and optimizing the network structure, our model not only enhances the detail recovery capability, particularly in fast-moving sports scenes, but also effectively improves the overall visual quality of images.

**Table 1.** Evaluation metrics results for super-resolution reconstruction methods in various sports scene images

Method	PSNR- $\mu$ / dB	SSIM- $\mu$	PSNR-L / dB	SSIM-L	HDR - VDP-2
ESPCN	42.1325	0.9854	41.2314	0.9854	60.1241
FSRCNN	12.3256	0.7854	12.3265	0.4562	51.2348
SRCNN	12.4758	0.8412	9.6521	0.3321	53.1247
VDSR	41.2368	0.9752	41.2147	0.9784	62.3562
EDSR	41.2589	0.9762	40.2314	0.9625	63.5487
RCAN	42.3125	0.9862	41.2589	0.9754	63.2145
RDN	42.3156	0.9751	38.2631	0.9785	61.2358
The Proposed Model	43.6985	0.9862	41.2358	0.9856	65.1235

**Table 2.** Ablation study results of different modules in our model

Method	PSNR- $\mu$ / dB	SSIM- $\mu$	PSNR-L / dB	SSIM-L	HDR - VDP-2
Baseline Model	42.1256	0.9785	41.2145	0.9856	63.2145
Our Model - Dual-Path Residual Block	42.1485	0.9854	40.1254	0.9854	63.2569
Our Model - Dynamic Adaptive Module	43.2658	0.9785	40.2356	0.9862	65.1245
Our Model	43.2315	0.9921	41.2569	0.9874	65.5892

Table 2 presents the ablation study results of different modules in our model, evaluating the contribution of each component to super-resolution performance. From the five key metrics of PSNR- $\mu$ , SSIM- $\mu$ , PSNR-L, SSIM-L, and HDR-VDP-2, it is evident that each module impacts performance differently. Compared to the baseline model, the full version of our model shows significant performance improvements, especially in PSNR- $\mu$  and HDR-VDP-2 scores, reaching 43.2315 dB and 65.5892 dB respectively, indicating overall improvements in image quality and visual perception. In ablation experiments, removing the dual-path residual block slightly decreases PSNR-L, highlighting its importance in processing low-light images. Removing the dynamic adaptive module, while slightly reducing PSNR- $\mu$ , has a larger impact on HDR-VDP-2, indicating the significant contribution of the dynamic adaptive module to high dynamic range vision. These results conclude that our dynamic adaptive cascaded network greatly enhances the super-resolution reconstruction effect in sports scene images. The dynamic adaptive module is crucial for handling high dynamic range content while maintaining detail and structure, as evidenced by its significant contribution to HDR-VDP-2. The design of the entire network allows the modules to work together, providing superior performance not only in solving the image blur caused by rapid movement but also in improving image detail recovery capability and visual quality.

The data in Figure 6 shows that with an increasing number of dynamic adaptive cascaded modules, PSNR gradually improves, rising from 32.175 dB with two modules to 32.378 dB with six modules. This incremental increase reflects the cascaded modules' effectiveness in enhancing image quality, especially in terms of detail recovery and noise reduction. Additionally, the parameter count and the number of floating-

point operations increase with the number of modules, from 50,000 parameters and 2 billion operations with two modules to 150,000 parameters and 65 billion operations with six modules. This growth indicates that while adding more modules can yield better image reconstruction quality, it also requires more computational resources and storage space, which may impose higher demands on deployment in practical application environments. The experimental data demonstrates that the dynamic adaptive cascaded network excels in handling image super-resolution reconstruction in sports scenes, achieving progressively higher PSNR with the addition of modules, thus proving its effectiveness. However, the performance improvement also comes with increased computational complexity and parameter count, necessitating a trade-off in practical applications to balance reconstruction quality with resource consumption.

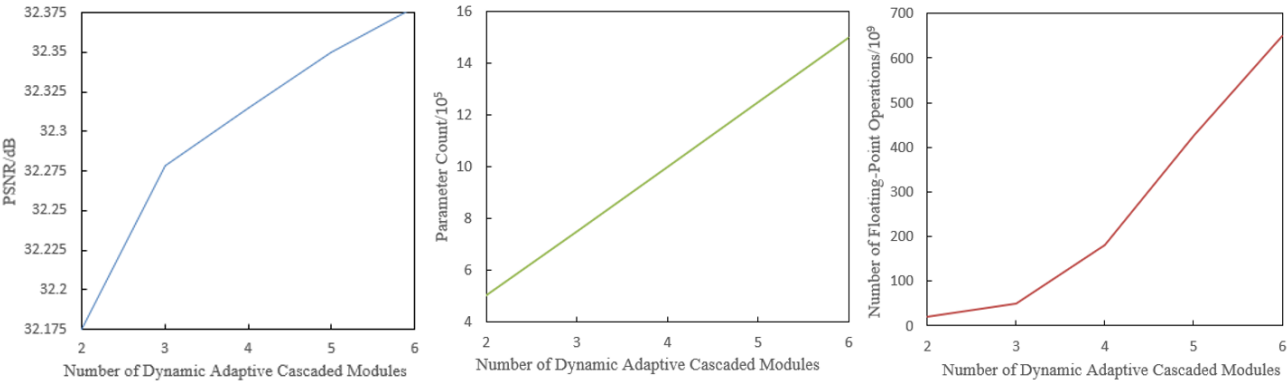
This study compares the performance of various super-resolution reconstruction methods in sports scene images, including ESPCN, FSRCNN, SRCNN, VDSR, EDSR, RCAN, RDN, and the dynamic adaptive cascaded network proposed in this paper. The data in Figure 7 shows that while high-performance models like EDSR and RCAN provide higher PSNR values, these models typically involve a substantial amount of parameters and computational resource consumption. In contrast, the dynamic adaptive cascaded network proposed in this paper maintains lower amounts of parameters and computational load while still achieving competitive PSNR performance with high-end models. Compared to other models, our model reaches nearly top-tier model levels in PSNR with only a fraction of the parameter count, effectively demonstrating an optimized network architecture that achieves a good balance between resource efficiency and image quality. Moreover, our model's superiority is also evident in its ability to handle dynamic scenes, effectively reducing blur caused by fast motion and enhancing image detail and clarity without significantly increasing the computational burden. This is particularly important for sports image analysis, as it not only improves the visual quality of images but also enhances the accuracy and reliability of subsequent analyses.

The data shown in Figure 8 reflects the PSNR and SSIM results of the dynamic 3D sports motion scene imaging methods tested across three sample sets with different numbers of iterations. From a PSNR perspective, all three sample sets show a gradual increase in PSNR values as the number of iterations increases, indicating that image quality proportionally improves with more iterations. Particularly, in the process from 5000 to 50000 iterations, Sample Set 1's PSNR increased from 30.85 to 31.6, while Sample Sets 2 and 3 also showed steady growth from lower starting values to 31.4 and 31.5, respectively. For SSIM, this metric also improved across all sample sets with increasing iterations, with Sample Set 1's SSIM rising from 0.885 to 0.894, showing a stable improvement in structural similarity. These results clearly demonstrate that the proposed method progressively optimizes image perceptual quality and structural details through iterations.

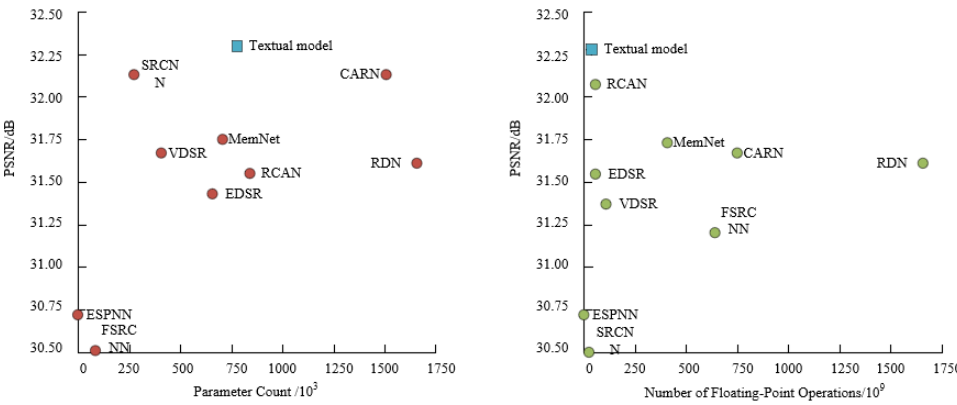
From these test results, it can be concluded that the dynamic 3D sports motion scene imaging technology proposed for sports motion analysis effectively enhances the quality of super-resolution image reconstruction. As the number of iterations increases, the model's ability to capture and reconstruct details significantly strengthens, as evidenced by the continuous improvement in the key performance metrics

of PSNR and SSIM. Overall, this research successfully demonstrates the effectiveness of dynamic 3D imaging technology in improving the quality and accuracy of sports

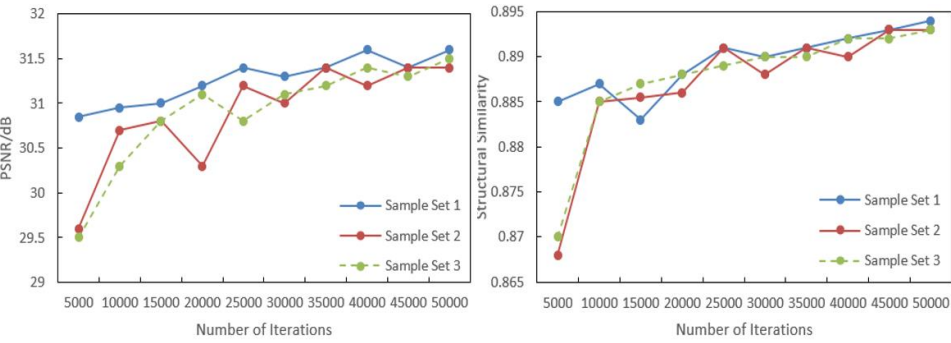
image analysis, proving its substantial potential and value in applications.



**Figure 6.** The impact of the number of dynamic adaptive cascaded modules on PSNR, parameter count, and computational load



**Figure 7.** Relationship between parameter count, computational load, and PSNR in different networks



**Figure 8.** Test results comparison of dynamic 3D motion scene imaging methods across three sample sets

## 5. CONCLUSION

This paper's research revolves around two core technologies: first, using a dynamic adaptive cascaded network for super-resolution reconstruction of images in sports scenes, and second, constructing dynamic 3D images of sports scenes to enhance the dimensions and accuracy of sports analysis. Through a series of experiments, this study not only clearly demonstrates the effectiveness of the dynamic adaptive cascaded network in handling image blur caused by rapid motion but also proves its significant advantages in detail recovery. Additionally, for dynamic 3D image construction technology, this research validated its effectiveness in

improving image quality (PSNR) and structural similarity (SSIM) across different sample sets.

The comprehensive experimental results show that the model proposed in this paper exhibits outstanding performance in key evaluation metrics such as PSNR and SSIM, particularly excelling in the HDR-VDP-2 visual quality assessment. Through module ablation studies, the contribution of each module to overall performance was further validated, especially the critical role of the dynamic adaptive module in enhancing computational efficiency and image quality. Moreover, although an increase in the number of dynamic adaptive cascaded modules leads to a rise in parameter count and computational load, appropriate optimization has allowed



this model to successfully balance performance with computational resources.

Despite the significant academic value and practical application potential of this research in the field of sports image processing, it still has certain limitations. For instance, as model complexity increases, so does the demand for computational resources, which may limit its application in resource-constrained environments. Future research could explore more model light-weighting techniques and optimization algorithms to reduce computational burden while maintaining image reconstruction quality. Further studies might also consider extending this technology to other types of dynamic scenes, such as natural environments or urban traffic scenarios, to verify its universality and adaptability.

## REFERENCES

- [1] Han, M. (2022). Sports video image analysis system based on SIFT algorithm. In *International Conference on Big Data Analytics for Cyber-Physical System in Smart City*, Singapore: Springer Nature Singapore, pp. 364-371. [https://doi.org/10.1007/978-981-99-1157-8\\_44](https://doi.org/10.1007/978-981-99-1157-8_44)
- [2] Li, Z., Ye, X., Liang, H. (2023). Sports video analysis system based on dynamic image analysis. *Neural Computing and Applications*, 35(6): 4409-4420. <https://doi.org/10.1007/s00521-022-07131-6>
- [3] Tan, L. (2024). A method for identifying sports behaviors in sports adversarial project training based on image block classification under the Internet of Things. *Journal of Testing and Evaluation*, 1-3. <https://doi.org/10.1520/JTE20230025>
- [4] Qin, Y.J. (2023). Virtual reality-assisted deep learning for support vector machine-based analysis of mass sports fitness medical images. *Computer-Aided Design and Applications*, 20(S14): 208-215. <https://doi.org/10.14733/cadaps.2023.S14.208-2015>
- [5] Li, L. (2022). Multi-view block Fusion Algorithm for Data Mining and Intelligent Sports Training. In *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, pp. 1616-1619. <https://doi.org/10.1109/ICESC54411.2022.9885361>
- [6] Li, G. (2022). Construction of sports training performance prediction model based on a generative adversarial deep neural network algorithm. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/1211238>
- [7] Rodriguez-Lozano, F.J., Gámez-Granados, J.C., Martínez, H., Palomares, J.M., Olivares, J. (2023). 3D reconstruction system and multiobject local tracking algorithm designed for billiards. *Applied Intelligence*, 53(19): 21543-21575. <https://doi.org/10.1007/s10489-023-04542-3>
- [8] Lu, Y., An, S. (2020). Research on sports video detection technology motion 3D reconstruction based on hidden Markov model. *Cluster Computing*, 23(3): 1899-1909. <https://doi.org/10.1007/s10586-020-03097-z>
- [9] Jin, Z., Zheng, Y., Liu, J., Yu, Y. (2024). A semantic web-based approach for bat trajectory reconstruction with human Keypoint information. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 20(1): 1-22. <https://doi.org/10.4018/IJSWIS.338999>
- [10] Mustafa, A., Russell, C., Hilton, A. (2022). 4D Temporally coherent multi-person semantic reconstruction and segmentation. *International Journal of Computer Vision*, 130(6): 1583-1606. <https://doi.org/10.1007/s11263-022-01599-4>
- [11] Wang, J., Li, M., Dziatkovskii, A., Hryneuski, U., Krylova, A. (2022). Research on contour feature extraction method of multiple sports images based on nonlinear mechanics. *Nonlinear Engineering*, 11(1): 347-354. <https://doi.org/10.1515/nleng-2022-0037>
- [12] Qu, C. (2020). Virtual reconstruction of random moving image capturing points based on chaos embedded particle swarm optimization algorithm. *Microprocessors and Microsystems*, 75: 103069.
- [13] Xu, Y., Shi, Y., Huang, C. (2022). Improving reconstruction-based coding methods for image classification: a visual dictionary refining method. *Journal of Electronic Imaging*, 31(4): 043048-043048.
- [14] Chen, L.H., Su, C.W., Hsiao, H.A. (2018). Player trajectory reconstruction for tactical analysis. *Multimedia Tools and Applications*, 77(23): 30475-30486. <https://doi.org/10.1117/1.JEL.31.4.043048>
- [15] Li, P., Ge, H., Geng, P. (2023). Signal and image reconstruction with tight frames via unconstrained  $\ell_1$ - $\alpha\ell_2$ -analysis minimizations. *Signal Processing*, 203: 108755. <https://doi.org/10.1016/j.sigpro.2022.108755>
- [16] Zhang, Z., Li, S. (2024). A Study on the Development of Professional Training in Physical Education in Colleges and Universities under Information Technology. *Applied Mathematics and Nonlinear Sciences*.
- [17] Deng, L., Pu, Y. (2021). Analysis of college martial arts teaching posture based on 3D image reconstruction and wavelet transform. *Displays*, 69: 102044. <https://doi.org/10.1016/j.displa.2021.102044>
- [18] Zhang, L., Liu, S. (2020). Quality evaluation of multi-frame sports image fusion based on gradient domain. *International Journal of Performability Engineering*, 16(3): 411-418.
- [19] Huo, L., Chen, W., Ge, H., Ng, M.K. (2022). Stable image reconstruction using transformed total variation minimization. *SIAM Journal on Imaging Sciences*, 15(3): 1104-1139. <https://doi.org/10.1137/21M1438566>
- [20] Shao, Y., Wu, Y. (2021). Analysis of 3D Modeling and Detection of Wrong Action in Sports Training. In *International Conference on Forthcoming Networks and Sustainability in the IoT Era*, Cham: Springer International Publishing, pp. 175-179. [https://doi.org/10.1007/978-3-030-99581-2\\_24](https://doi.org/10.1007/978-3-030-99581-2_24)
- [21] Wei, W., Lu, L. (2024). Optical super-resolution imaging based on improved planning algorithm in sports teaching simulation. *Optical and Quantum Electronics*, 56(3): 300. <https://doi.org/10.1007/s11082-023-06009-8>