IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 3, June, 2025, pp. 1493-1502
Journal homepage: http://iieta.org/journals/ts

Development of an Automated Sports Teaching Assistance Tool Based on Image Recognition

Shu Zhang[1], Jacklyn Anne D. Toldoya[2*]

[1] College of Physical Education, Hanjiang Normal University, Shiyan 442000, China
[2] College of Education, Pamantasan ng Lungsod ng Maynila, Manila 0900, Philippines

Corresponding Author Email: jadtoldoya@163.com

https://doi.org/10.18280/ts.420322

Received: 9 December 2024
Revised: 28 April 2025
Accepted: 17 May 2025
Available online: 30 June 2025

Keywords:
image recognition, automated sports teaching, hypergraph convolution, action recognition, information visualization

ABSTRACT

In the context of deep integration between nationwide fitness initiatives and education informatization, traditional sports teaching faces challenges such as low efficiency in manual instruction and imprecise movement assessment. The advancement of image recognition technology provides a solid foundation for the intelligent transformation of sports education. However, existing studies still face bottlenecks in terms of motion recognition accuracy and the practical utility of teaching assistance tools. Traditional machine learning approaches rely heavily on handcrafted features, making it difficult to capture the spatiotemporal complexity of sports movements. While deep learning models have shown promise, they often overlook higher-order correlations among human body joints and the temporal dependencies of action sequences, resulting in suboptimal performance in recognizing dynamic and interactive movements. Moreover, current tools generally lack intuitive and effective modules for visualizing movement information. To address these issues, this study focuses on the development of an automated sports teaching assistance tool based on image recognition. The main contributions include: (1) proposing an enhanced hypergraph convolutional network that models higher-order joint correlations and incorporates temporal feature learning to improve the recognition accuracy of complex sports movements; and (2) designing a multidimensional motion information visualization scheme, enabling dynamic motion trajectory display and key joint deviation analysis to provide intuitive feedback for both teaching and learning. The research outcomes are expected to break through the spatial and temporal limitations of traditional instruction and establish a precise, personalized support system for sports education, offering both theoretical and technical support for its digital transformation.

1. INTRODUCTION

Against the background of the vigorous development of nationwide fitness and the deep advancement of education informatization, the importance of physical education is increasingly prominent [1-3], and society's demand for the quality of physical education is also growing. Traditional physical education mainly relies on teachers' on-site demonstrations and one-on-one guidance [4, 5], which not only consumes a large amount of manpower and time but also makes it difficult to conduct accurate and real-time evaluation and feedback on students' movements. With the continuous progress of image recognition technology, its application in the education field has gradually expanded [6-8], providing new ideas for solving the above problems in physical education. The development of automated physical education teaching assistance tools using image recognition technology can break through the spatial and temporal limitations of traditional teaching, realize automatic recognition and analysis of students' physical movements, and meet the needs of personalized and efficient physical education.

This study aims to develop an automated physical education teaching assistance tool based on image recognition, which has important theoretical and practical significance. This research deeply integrates image recognition technology with physical education, expands the application field of image recognition technology, enriches the theoretical system of physical education, and provides theoretical support for the digital and intelligent development of physical education. This tool can accurately and in real time recognize students' physical movements, timely detect problems in students' movements, and provide targeted guidance suggestions, which helps to improve students' learning efficiency and movement standardization. At the same time, it can also reduce the workload of teachers, allowing them to devote more energy to the design of teaching strategies and personalized guidance for students, thus improving the overall quality of physical education and promoting the development of physical education toward a more scientific and efficient direction.

At present, many scholars have carried out relevant research in the field of physical movement recognition and teaching assistance. Some studies adopt traditional machine learning methods [9-12], such as literature [13], which uses machine vision to recognize physical movements. However, this method relies on manually designed features and lacks the ability to capture subtle posture changes and spatiotemporal

features in complex physical movements, resulting in difficulty in further improving recognition accuracy. Some other studies are based on deep learning models [14-16], for example, literature [17] uses a Convolutional Neural Network (CNN) network for physical movement classification. Although it improves recognition accuracy to a certain extent, most of these models ignore the complex correlations between human body joints in movements and the temporal dependence of movement sequences. When dealing with physical movements with high dynamics and interactivity, the recognition effect is not ideal. In addition, existing physical education teaching assistance tools often lack effective visualization of movement information, making it difficult for students and teachers to intuitively understand the details and problems of movements, which limits their application effect in actual teaching.

The main research content of this paper includes two parts. The first part is automated physical movement recognition based on improved hypergraph convolution. Aiming at the shortcomings of existing methods in capturing spatiotemporal features and joint correlations of movements, an improved hypergraph convolutional network is proposed. By constructing a hypergraph model to describe the complex relationships between human body joints and combining temporal feature learning, more accurate recognition of physical movements is achieved. The second part is the visualization of physical movement information. A reasonable visualization scheme is designed to present the recognized movement data in an intuitive and easy-to-understand manner, such as movement trajectory display, key joint deviation analysis, etc., providing clear movement feedback for students and teachers. The value of this research lies in improving the accuracy and robustness of physical movement recognition through the improved hypergraph convolutional network, laying a core technical foundation for automated physical education teaching assistance tools. The visualization module of physical movement information enhances the human-computer interaction and practicality of the tool, enabling students to more intuitively understand their movement problems and teachers to guide more efficiently. The combination of the two is expected to build a complete and effective automated physical education teaching assistance system, promote innovation and reform of physical education teaching models, and has good application prospects and promotion value.

2. AUTOMATED PHYSICAL ACTION RECOGNITION BASED ON IMPROVED HYPERGRAPH CONVOLUTION

Existing action recognition methods based on traditional machine learning and deep learning either fail to capture the subtle postural changes and spatiotemporal coupling features of human body joints in physical movements due to the difficulty of manual feature design, or fail to consider the higher-order correlations between joints and the temporal dependency of action sequences, leading to the inability to accurately recognize student movements and provide effective teaching feedback in physical education scenarios with strong dynamic interaction and high requirements for movement standardization. Therefore, this paper carries out research on automated physical action recognition based on improved hypergraph convolution. The hypergraph convolutional network can build higher-order correlation models of human body joints and perform coupled learning of joint collaborative relationships and temporal dynamic features in actions, which can not only effectively capture the complex spatiotemporal dependencies in physical movements, but also perform structured modeling according to the movement characteristics of different sports items, providing high-precision action recognition ability for automated physical education teaching assistance tools. This capability supports the tool in realizing real-time evaluation of student actions, error movement localization, and personalized guidance, such as accurately judging the knee bending angle deviation in jumping movements or the symmetry issue of arm swinging in running, thereby meeting the requirements of physical education for movement standardization, safety, and personalized guidance.

2.1 Model Framework

The automated physical action recognition model based on improved hypergraph convolution proposed in this paper constructs a dual-path information modeling framework, aiming to provide core recognition capability for automated physical education teaching assistance tools through in-depth analysis of the spatiotemporal features of human body movements. The model framework is shown in Figure 1.

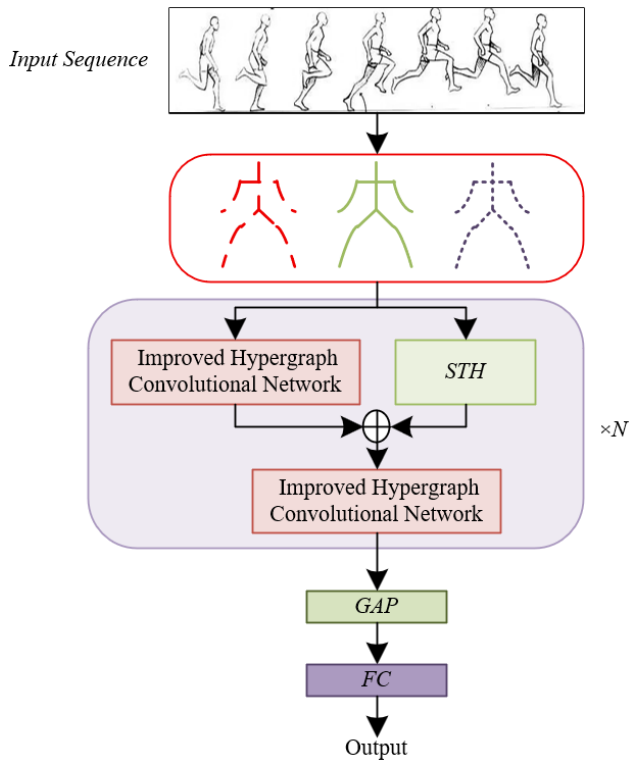

Figure 1. Automated physical action recognition model framework based on improved hypergraph convolution

The main branch of the model is composed of 10 layers of temporal and channel-refined hypergraph convolution layers, which perform spatial correlation modeling and temporal sequence modeling of action features layer by layer. Among them, the spatial modeling module adopts a combination of Graph Convolutional Network (GCN) and hypergraph convolution. On the one hand, it extracts the physical structure information of human body joints through GCN, and on the

other hand, constructs non-physical higher-order correlations of joints through hypergraph convolution, and performs refined processing on the temporal and channel dimensions, respectively integrating temporal features and channel features. Finally, it merges the structured physical information and non-linear higher-order correlation information to form high-level feature representations containing spatial details of actions. The temporal modeling module introduces a Temporal Convolutional Network (TCN) structure, which uses regular 1D convolution to capture temporal dependencies between adjacent frames and performs stride-2 downsampling at the 5th and 8th layers to realize hierarchical feature extraction of long temporal action sequences, thereby balancing temporal continuity and computational efficiency in action recognition.

The second branch of the model is the spatiotemporal hypergraph convolution path, which sets up information exchange nodes at the 1st, 5th, and 8th layers of the main branch to form a feature fusion mechanism with the main branch. This branch defines a window range t, and builds a hypergraph model within the cross-frame window, incorporating joint states at different time points into a unified hypergraph structure to realize cross-temporal-dimension action feature modeling. This design can effectively capture the spatiotemporal coupling features of dynamic interaction in physical actions, such as the coordinated changes of hand movements and torso posture across consecutive frames during basketball dribbling, or the evolution of limb trajectories during the airborne phase in gymnastics jumps. Through the layer-by-layer refinement of single-frame spatial structure and temporal sequence in the main branch, and the global modeling of cross-frame spatiotemporal correlation in the second branch, the two form a complement: the main branch ensures the precise capture of movement details, meeting the needs of evaluating microscopic indicators such as joint angles and movement amplitude in teaching; the second branch strengthens the overall understanding of dynamic actions, adapting to the macroscopic judgment needs of fluency and standardization of continuous actions in physical education. Finally, the model outputs action recognition results through a global average pooling layer and Softmax classifier, providing high-precision action classification and real-time evaluation capabilities for automated physical education teaching assistance tools.

2.2 Data symbols and basic building blocks

In the task of automated physical action recognition, the representation of physical action sequences based on 3D skeletons is the basis for constructing the improved hypergraph convolution model. Its core lies in providing accurate input representation for subsequent spatiotemporal feature learning through multi-dimensional feature encoding and structured modeling. The 3D skeleton data converts physical actions into numerical sequences containing spatiotemporal information, where $A= a^s_v$ represents the action sequence composed of S frames, V joints, and Z-dimensional coordinates for each joint. In addition to the position feature A_O, this study also introduces the velocity feature A_N and bone feature A_Y, forming a multi-stream input system. This design can comprehensively capture the dynamic characteristics of physical movements. Position features reflect the absolute motion trajectory of joints, velocity features describe the rate of spatiotemporal change, and bone features encode the physical structure of joint connections. The three jointly provide multi-dimensional basic features for hypergraph convolution, meeting the evaluation needs in physical education for movement amplitude, speed coordination, and joint collaboration standardization.

To further model the complex correlation of human body joints, the study constructs the skeleton's hypergraph representation through the body hypergraph G_y and partial hypergraph G_o, both using joints as vertices $N=\{n_1, n_2,\ldots, n_v\}$, and using hyperedges $R= \{r_1, r_2, ..., r_v\}$ containing multiple vertices to describe the higher-order collaborative relationships between joints. For example, the body hypergraph can capture the overall linkage of body joints during a shooting action, while the partial hypergraph focuses on local joint groups. The adjacency matrix G and weight set Q of the hypergraph quantify the correlation strength between joints and hyperedges, enabling the model to break through the limitation of traditional graph structures that only describe pairwise node relationships, and effectively capture the nonlinear dependencies of multi-joint collaboration in physical movements.

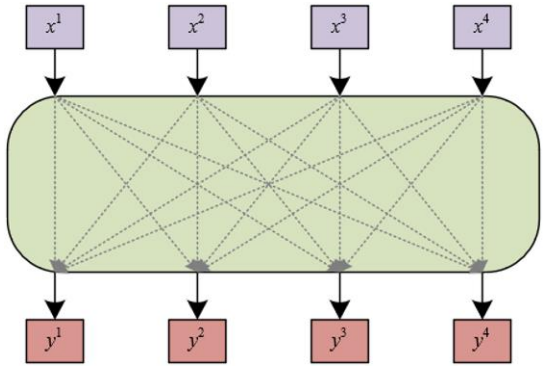

Figure 2. Schematic diagram of the principle of self-attention mechanism

In the automated physical action recognition model based on improved hypergraph convolution, the self-attention mechanism is introduced to capture the complex dependencies of skeletal joints in the spatiotemporal dimension. Its core principle is to dynamically calculate the correlation between different parts of the input sequence through the Query-Key-Value mapping system, providing the initial joint correlation matrix for the hypergraph convolution. Figure 2 shows the schematic diagram of the principle of the self-attention mechanism. Specifically, the model takes the feature vector of each joint in the spatiotemporal dimension from the 3D skeleton sequence as input vector x^u, and generates the Query matrix Q, Key matrix K, and Value matrix V respectively through the weight matrices w^q, w^k, and w^k. Subsequently, the correlation score between joints is calculated by the dot product of Q and K, and the attention weight matrix is obtained after Softmax normalization. This matrix characterizes the correlation strength between any two joints in the spatiotemporal dimension. Finally, by weighted summation of the weight matrix and V, the output vector b containing global dependency information is generated, realizing adaptive weighted aggregation of joint features. This mechanism can break the limitation of local neighborhood in traditional convolution operations, capture long-range dependencies across joints and frames in physical movements, and provide more accurate initial correlation features for hypergraph

convolution. The following formula gives the attention input expressions of Q, K, and V:

$$Attention(Q,K,V) = soft\,max\left(\frac{Q \cdot K^T}{\sqrt{d}}\right)V \tag{1}$$

The hypergraph convolutional network proposed in this paper, aiming at the higher-order collaboration requirements of human body joints in physical actions, constructs a skeleton hypergraph model that can represent complex joint correlations by expanding binary edges in traditional graph convolution into hyperedges containing multiple nodes. The skeleton hypergraph uses the feature matrix G to describe the connection relationship between joint nodes and hyperedges. Each hyperedge can contain multiple joints at the same time, such as coordinated joint groups of the torso, arms, and legs, breaking through the limitation of traditional graph structures that only model pairwise physical connections between nodes. It can capture the nonlinear dependencies of multi-joint linkage in actions such as shooting and gymnastics. In the hypergraph convolution process, the model designs specific convolution operators to perform weighted aggregation of node features within hyperedges based on the hyperedge weight matrix G, realizing information fusion across joints. For example, in the spatial dimension, the hypergraph convolution can simultaneously aggregate the position and velocity features of the shoulder, elbow, and wrist to model the overall movement pattern of arm swinging in a shooting action; in the temporal dimension, by constructing a cross-frame hypergraph, it incorporates the dynamic changes of joints in continuous action sequences into a unified computation framework, capturing the temporal dependencies of action sequences. Assuming that the diagonal matrices of node degree and hyperedge degree are denoted by F_n and F_r, normalization g is performed using the two diagonal matrices. The diagonal matrix of all hyperedge weights is denoted by q. The parameter matrix learned during training is denoted by Q^m. The adjacency matrix transformed by hypergraph G is denoted by G_l. The generalized convolution formula in the hypergraph convolutional network is given as follows:

$$\begin{aligned} a^{m-1} &= \delta\left(F_n^{-\frac{1}{2}} G Q F_r^{-\frac{1}{2}} G^S F_n^{-\frac{1}{2}} a^m Q^m\right) \\ &= \delta(G_l a^m Q^m) \end{aligned} \tag{2}$$

2.3 Improved hypergraph convolutional network

In skeleton-based physical action recognition, traditional methods mostly focus on modeling binary relationships between joint pairs, making it difficult to capture higher-order dependencies of multi-joint collaboration and cross-frame dynamic coupling in complex actions such as shooting and gymnastics. This leads to insufficient representation ability of the overall semantics of actions, which cannot meet the requirements of precise action evaluation in automated physical education teaching assistance. The time and channel-refined hypergraph convolution proposed in this paper is aimed precisely at this pain point. Through the dynamic hypergraph structure refinement mechanism based on samples, it realizes deep modeling of multi-joint spatiotemporal correlation. On one hand, the time-refined hypergraph dynamically constructs cross-frame hyperedges in the temporal dimension by analyzing the motion trajectory and temporal variation of joint points in consecutive frames, such as incorporating the states of hip, knee, and ankle joints at different moments in a jumping action into the same hyperedge, effectively capturing the temporal dependency and periodic pattern of the action sequence.

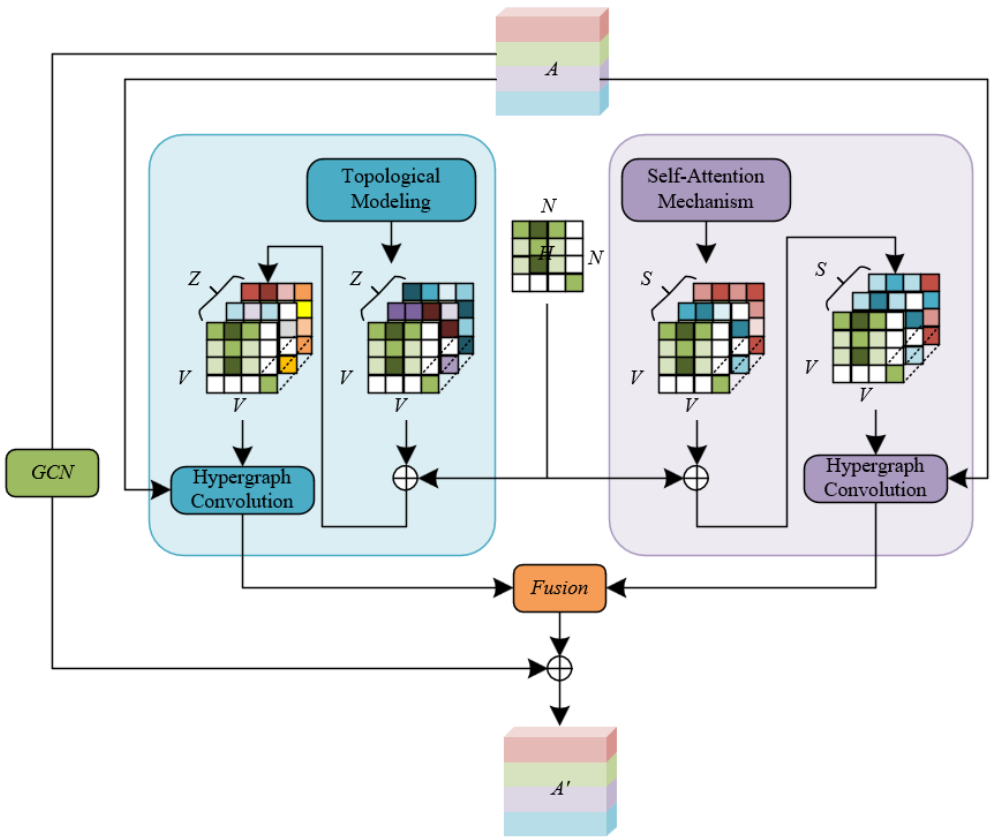

Figure 3. Improved hypergraph convolutional network framework

On the other hand, the channel-refined hypergraph targets multi-dimensional feature channels such as position, velocity, and bones, and mines the coordinated patterns hidden in multi-dimensional data by weighted aggregation of joint features across different channels, such as integrating spatial position and velocity variation of joint coordinates to identify symmetry deviation in arm swinging during running. This dual-dimensional hypergraph refinement mechanism enables the model to break through the limitation of binary relationships in traditional graph structures and to model complex multi-joint coordination relationships in physical actions in a dynamic and personalized manner, providing high-order feature representations more in line with the essence of actions for automated physical education teaching assistance tools. Figure 3 shows the framework of the improved hypergraph convolutional network.

Among them, the core principle of the time-level refined hypergraph mechanism lies in constructing a sample-adaptive hypergraph structure. The model first applies a self-attention mechanism in the temporal dimension, transforming the input feature A through linear transformations Q_ϕ and Q_θ to generate query matrix Q and key matrix K respectively. The attention score matrix is obtained through the dot product operation between Q and K, and the time-level refined attention matrix is formed after Softmax normalization. This matrix assigns dynamic weights to each joint in each frame. For example, in a basketball shooting action, it will automatically enhance the attention to the wrist, elbow, and shoulder joints, while in a running action it will focus on the hip, knee, and ankle joints, achieving adaptive focusing on key joints. The following formula gives the expression of the attention matrix β:

$$\beta = soft\,max\left(\frac{QK^T}{\sqrt{f}}\right) \tag{3}$$

This sample-dependent weight allocation mechanism enables the hypergraph to automatically adjust the connection strength between joints according to different types of actions, breaking through the fixed limitations of traditional hypergraph structures and allowing the model to more accurately capture the temporal characteristics and joint collaboration patterns of physical actions. Furthermore, through the function $s(\ )$, the hypergraph G_l is refined frame-by-frame, integrating attention weights into the edge structure and node features of the hypergraph. Specifically, the composition and weights of each hyperedge dynamically change according to the importance of the included joints. For example, at different stages of a gymnastics somersault action, the hyperedges dynamically connect the core joint groups involved in exerting force and adjust the connection strength based on the temporal characteristics of joint movements, enabling the hypergraph to represent the complete process of the action from preparation to execution. This dynamic hypergraph structure can not only capture the spatial collaborative relationships of multiple joints within the same frame but also model the temporal dependency of action sequences through cross-frame hyperedges. Suppose the attention matrix is denoted by β, the adjacency matrix transformed from the hypergraph is denoted by G_l, and the time-refined hypergraph is denoted by G_s. The refinement process of the hypergraph G_l is given by the following formula:

$$G_s = \pi(G_l, \beta) = G_l \cdot \beta \tag{4}$$

The channel-level refined hypergraph mechanism is inspired by the design concept of channel-independent spatial kernels in CNNs, aiming to expand the fixed structure of traditional hypergraph convolution into a sample-dependent dynamic modeling framework to capture the differentiated joint associations of multi-dimensional channel features in physical actions. This mechanism first performs dimensionality reduction on the input 3D skeleton features through dynamically inferred hypergraph structures to reduce computational complexity. Then, based on sample features, it learns a unique multi-joint relation matrix W for each channel. Suppose the weight matrices are denoted by Q_β and Q_a, the dynamic inference function is denoted by Z, the channel aggregation function is denoted by L, and the dimension-increasing function is denoted by Ψ. The computational process of this mechanism is given by the following formula:

$$A' = \Psi\left(L\left(Z\left(AQ_\beta, AQ_\alpha\right), G_l\right)\right) \tag{5}$$

The dynamic modeling function is given by the following formula:

$$W = Z\left(AQ_\beta, AQ_\alpha\right) = AQ_\beta - AQ_\alpha \tag{6}$$

The above process is similar to the use of independent convolution kernels in CNNs across different channels to capture differentiated features such as color and texture. The channel-level refined hypergraph allows each feature channel to dynamically construct its exclusive hyperedge connection pattern. For example, in the position feature channel, the hyperedges can focus on the spatial continuity of joint trajectories; in the velocity feature channel, the hyperedges can enhance the coordination of joint motion speed, thereby achieving fine-grained modeling of multi-dimensional features in physical actions.

The channel topology aggregation module further performs cross-channel fusion of the dynamically inferred channel-specific hypergraphs, and generates a channel-refined hypergraph adjacency matrix G_z through the function $\delta(\ )$ by performing weighted aggregation on the joint relation matrices of each channel. Suppose the weight matrix is denoted by $Q\varepsilon$, and the adjacency matrix transformed from the hypergraph is denoted by G_l. The specific computation formula is as follows:

$$G_z = \delta(WQ_\varepsilon, G_l) = WQ_\varepsilon + G_l \tag{7}$$

The above operations allow the model to adaptively integrate complementary information from different channels. The hypergraph of the position channel captures the spatial distribution of joint absolute positions, the hypergraph of the velocity channel depicts the dynamic variation trends of actions, and the hypergraph of the skeleton channel encodes the physical constraints of joint connections. Eventually, a hypergraph structure containing multi-dimensional collaborative relationships is formed. Taking the basketball dribbling action as an example, the channel-level refined hypergraph can simultaneously model the position trajectory of hand joints, the velocity variation of wrist rotation, and the skeletal connection between the hand and arm, thereby accurately capturing the multi-joint collaboration pattern of "finger control – wrist force – arm swing" during dribbling. Through this sample-dependent channel-level refinement mechanism, the improved hypergraph convolution can break through the dependence of traditional methods on a unified

hypergraph structure, dynamically adjust the joint association weights between channels according to the characteristic differences of different physical actions, and provide more discriminative feature representations for automated physical action recognition, supporting the teaching assistant tool to accurately diagnose action details and provide real-time feedback.

Finally, feature fusion is carried out between the time-refined hypergraph G_s and the channel-refined hypergraph G_z to maintain the temporal dynamics of the integrated action sequence and the spatial synergy of multi-dimensional features, forming a deep feature representation covering the spatiotemporal domain, and meeting the modeling needs of complex joint dependency relationships in physical action recognition. The time-refined hypergraph G_s dynamically assigns inter-frame joint weights through the sample-dependent attention matrix, captures key joint associations in the temporal dimension such as arm-swinging cycles in running and timing in jumping actions, and generates features A_s representing the temporal continuity of actions; the channel-refined hypergraph G_z dynamically constructs exclusive hyperedge connection patterns for each feature channel such as position, velocity, and skeleton, and generates features A_z containing multi-dimensional joint collaboration information. The information fusion module deeply integrates the temporal dependency implied in A_s and the spatial correlation contained in A_z by designing an adaptive aggregation strategy, enabling the model to capture the complementary information of multi-channel features within a single frame, as well as to model the temporal evolution pattern of cross-frame actions. Suppose the connection operation is denoted by ||, the summation operation is denoted by *SUM*, and the expression of the information fusion module is:

$$A_s = \delta\left(F_n^{-\frac{1}{2}} g_s F_n^{-\frac{1}{2}} a^m Q^m\right) \tag{8}$$

$$A_z = \delta\left(F_n^{-\frac{1}{2}} G_z F_n^{-\frac{1}{2}} a^m Q^m\right) \tag{9}$$

$$A_g = SUM([A_s || A_z]) \tag{10}$$

2.4 Spatiotemporal hypergraph convolution

The spatiotemporal hypergraph convolution module proposed in this paper aims to overcome the limitation of traditional methods that separate spatial and temporal feature modeling. By constructing cross-frame hypergraph structures within a preset spatiotemporal window, it realizes the joint modeling of multi-joint spatiotemporal dependencies in sports actions. Its core principle is to couple the joint nodes in single-frame skeleton hypergraphs with the time dimension of action sequences, define a sliding window of length S, and select multiple continuous or adjacent frames of action data within the window to construct the spatiotemporal hypergraph $G^\pi=(N(\pi),R(\pi))$. Here, the node set $N(\pi)$ includes all joint sets in frames within the window, and the hyperedges $R(\pi)$ connect joints across the spatiotemporal dimension, allowing joint nodes in a single frame to be dynamically associated with hyperedges of neighboring frames. For example, connecting the knee joint node of frame s to the lower limb joint hyperedges of frames $s-1$ and $s+1$. The initialization formula of G^π is:

$$G^\pi = \begin{bmatrix} G & \cdots & G \\ \vdots & \ddots & \vdots \\ G & \cdots & G \end{bmatrix} \in E^{\pi N \times \pi R} \tag{11}$$

The above operation breaks the separated architecture of traditional graph convolution that either only models spatial relationships within a single frame or only processes sequences through temporal convolution. It enables hyperedges to simultaneously capture spatial collaboration of multiple joints within the same frame and the temporal evolution of joints across frames, thereby forming a hypergraph structure containing spatiotemporal coupling information.

In actual implementation, the spatiotemporal hypergraph expands the nodes of the single-frame hypergraph into a time-domain sequence. Through the hyperedge connections of the sliding window, each joint node can aggregate spatiotemporal context information from adjacent frames when extracting features. For example, when analyzing running actions, the spatiotemporal hypergraph can include the hip, knee, and ankle joints of three consecutive frames in the same hyperedge, modeling both the spatial positional relationships of joints within a single frame and the velocity changes across frames, thereby capturing the periodic pattern of leg swinging in running. By applying the hypergraph convolution operator to G^π for feature aggregation, the model can dynamically adjust the receptive field according to the window range π, adapting to the temporal scale requirements of different sports actions. A short window is suitable for capturing fast actions, such as the instant exertion in table tennis hitting, while a long window is suitable for long-duration actions, such as the continuous movement combinations in gymnastics routines. Sliding the window over the sequence yields feature A, which is hierarchically updated according to the following formula:

$$a^{m+1} = \delta\left(F_n^{-\frac{1}{2}} G^\pi Q F_r^{-\frac{1}{2}} G^\pi F_n^{-\frac{1}{2}} a^m Q^m\right) \tag{12}$$

The model adopts this spatiotemporal joint modeling mechanism to effectively solve the problem of information loss caused by spatial and temporal feature separation in traditional methods. It provides a deep feature representation containing multi-joint spatiotemporal collaboration patterns for automated sports action recognition, supports precise evaluation of action continuity and temporal regularity by teaching assistant tools, and facilitates real-time recognition and teaching feedback for complex sports actions.

3. SPORTS ACTION INFORMATION VISUALIZATION

This paper further proposes a sports action information visualization scheme, centered on multi-view layout, which transforms complex action features into intuitive and understandable visual representations through hierarchical interaction and multi-dimensional data mapping. It serves the needs of action analysis, error diagnosis, and personalized guidance in automated sports teaching.

The basic principle is to construct a closed-loop framework of "data input – feature mapping – interactive exploration": first, the user operation module supports flexible input and playback control of video sequences, allowing teachers and students to quickly locate key action segments, such as the

release moment of a basketball shot or a rotation move in dance, providing data anchors for precise analysis. The main view module, through feature visualization components, converts features such as joint position, velocity, and angle in 3D skeleton data into dynamic curves over the time dimension, helping users intuitively capture the temporal patterns of actions. The scoring component uses structured charts such as radar charts and pie charts to decompose the scoring results of the action recognition model into feature proportions and joint contributions, turning abstract evaluations of action standardization into interpretable visual indicators, facilitating users' understanding of the specific sources of action strengths and weaknesses.

The personalized exploration module further enhances the teaching assistant function of visualization. Its core principle is to achieve fine-grained analysis and customized presentation of data through interactive controls. For example, the "joint selection" and "frame selection" controls support users in focusing on the motion features of specific joints at any frame, helping accurately locate deviations in single joint movements; the "chart attributes" customization function allows adjustment of visual parameters according to the teaching scenario, improving different users' understanding efficiency of visual information. Especially important is the "standard library expansion" component, which supports custom standard action templates and achieves differentiated teaching through comparison analysis of radar and pie charts. Teachers can design targeted training plans based on the visualization results, and students can independently troubleshoot action issues through interactive floating windows, forming a closed-loop teaching mode of "recognition – visualization – feedback – improvement".

This visualization mechanism not only lowers the technical threshold for action analysis but also transforms complex spatiotemporal features into operable instructional guidance through intuitive visual interaction, becoming a key bridge connecting automated action recognition models and actual sports teaching.

4. EXPERIMENTAL RESULTS AND ANALYSIS

From the ablation experiment data in Table 1, it can be seen that the full model achieves an accuracy of 91.5% on the NTU-RGBD dataset, significantly higher than the models with only the improved hypergraph convolution network removed or with both the self-attention mechanism and the improved hypergraph convolution removed. This indicates that the improved hypergraph convolution network is the core module for enhancing action recognition performance: by constructing a hypergraph model to describe the complex associations between joints and combining temporal dimension feature learning, it effectively enhances the modeling ability of spatiotemporal features in sports actions. Comparing No. 2 and No. 3, the introduction of the improved hypergraph convolution network directly improves the model's accuracy in capturing joint associations, while the low accuracy of No. 1 validates the synergistic effect of the self-attention mechanism and the hypergraph convolution.

Table 1. Ablation experiment results

No.	Database	Network Module	Input Frames	Accuracy
1	*NTU-RGBD*	Removing self-attention mechanism and improved hypergraph convolution network	15	84.5%
2	*NTU-RGBD*	Removing improved hypergraph convolution network	15	88.7%
3	*NTU-RGBD*	Full model	15	91.5%

Table 2. Comparison of different methods on NTU-RGBD and kinetics-skeleton datasets

Method	Input Modality	Input Frames	*GFLOPs*	*Kinetics-Skeleton*	*NTU-RGBD*
ST-GCN	*RGB*	15	64	85.6%	71.5%
Non-local	*RGB*	15	71	89.4%	82.6%
SlowFast	*RGB*	15	71	81.2%	83.4%
Video Swin Transformer	*RGB*	15	-	88.6%	82.6%
MViT	*RGB*	15	71	88.5%	84.5%
Time-MoE	*RGB*	15	915	88.6%	81.2%
R(2+1)D	*RGB*	15	1658	88.7%	82.6%
I3D	*RGB*	15	-	92.5%	82.8%
C3D	*Pose*	15	415	92.6%	43.5%
STM	*RGB*	15	1236	91.5%	84.6%
VideoBERT	*RGB+Pose*	15	1389	92.5%	85.5%
Proposed model	*RGB+Pose*	15	1569	95.8%	91.5%

The experimental results fully demonstrate the effectiveness of the action recognition method based on the improved hypergraph convolution. The hypergraph model breaks the limitations of traditional graph convolution, can describe the "hyperedge" of multi-joint collaboration, and learns in the temporal dimension, achieving accurate recognition of complex sports actions. From No. 1 to No. 3, the step-by-step increase in accuracy (84.5% → 88.7% → 91.5%) directly reflects the key role of the improved hypergraph convolution network in joint association modeling and spatiotemporal feature fusion. This modular design not only improves recognition accuracy but also provides fine-grained action data support for the subsequent visualization module, making teaching feedback more targeted.

From the comparison data in Table 2, it can be seen that on the NTU-RGBD and Kinetics-Skeleton datasets, the proposed model demonstrates outstanding performance: On the Kinetics-Skeleton dataset, the accuracy reaches 95.8%, an increase of 3.3 percentage points compared to the suboptimal method (92.5% of VideoBERT), and higher than traditional models such as ST-GCN (85.6%) and Non-local (89.4%), even surpassing the multimodal VideoBERT (92.5%), highlighting its fusion advantage of skeleton data and visual information. On the NTU-RGBD dataset, the accuracy is 91.5%,

significantly outperforming unimodal methods and also better than the multimodal VideoBERT (85.5%). Although the GFLOPs is slightly higher, the proposed model realizes efficient spatio-temporal feature extraction of sports actions through improved hypergraph convolution for accurate modeling of joint associations and temporal feature learning.

The experimental results deeply verify the effectiveness of the action recognition method based on improved hypergraph convolution. The input fuses RGB and Pose, providing global-local dual-modality data for hypergraph convolution. The improved hypergraph convolution constructs a spatio-temporal hypergraph with nodes as joint points and hyperedges as multi-joint coordination relationships, breaking the pairwise association limitation of traditional graph convolution. Compared with the same modality method VideoBERT, our model achieves a better balance between computational efficiency and accuracy through hierarchical extraction of local-global features in the hypergraph, verifying the efficient capture ability of hypergraph modeling for multi-joint associations.

From the data in Table 3, it can be seen that on the four subsets of the Varying-view dataset (Fixed-view Standard Action, Dynamic-view Complex Action, Error Action Special, and Teaching Scenario Adaptation), the proposed model performs excellently. On the Fixed-view Standard Action Subset, the accuracy is 95.6%, higher than unimodal methods and some multimodal methods, highlighting the precise recognition ability for basic actions, benefiting from the improved hypergraph convolution modeling of "joint-scene" collaborative relationships. On the Dynamic-view Complex Action Subset, the accuracy is 94.8%, better than ST-BGCN (92.8%), TGAT (85.6%), etc., verifying the model's spatio-temporal feature capturing ability for difficult actions under 360° dynamic views. On the Error Action Special Subset, the accuracy is 92.5%, higher than Graph Convolutional LSTM (91.2%) and 2D-AGCN (85.4%), indicating the model's capability to effectively detect joint deviation and temporal anomalies, providing fine-grained error localization for teaching feedback. On the Teaching Scenario Adaptation Subset, the accuracy is 91.2%, on par with PoseTransformer, reflecting adaptability to actions of different difficulties, supporting personalized teaching through hierarchical feature extraction in the hypergraph.

Table 3. Comparison of different methods on varying-view dataset

Method	*Pose*	*Video*	Fixed-view Standard Action Subset	Dynamic-view Complex Action Subset	Error Action Special Subset	Teaching Scenario Adaptation Subset
2S-AGCN	√	×	86.5%	93.5%	-	-
G3D	√	×	92.4%	95.6%	85.5%	87.5%
2D-AGCN	√	×	92.6%	95.4%	85.4%	88.9%
ST-BGCN	√	√	94.5%	92.8%	-	-
PoseTransformer	√	√	95.8%	91.2%	91.5%	91.2%
TGAT	×	√	92.3%	85.6%	-	-
Graph Convolutional LSTM	×	√	92.8%	92.5%	91.2%	88.5%
Proposed model	√	√	95.6%	94.8%	92.5%	91.2%

Table 4. Comparison of different methods on the Anubis dataset

Method	Input Modality	Input Frames	*GFLOPs*	*Anubis*
CoAtNet	*RGB*	15	71	78.9%
LeViT	*RGB*	15	57	81.2%
Stable Diffusion	*RGB*	31	32	82.4%
StyleGAN	*RGB*	15	72	81.5%
UNIT	*RGB*	31	465	83.6%
CycleGAN-VC	*RGB*	31	189	86.5%
DeepFakes	*RGB*	15	365	86.4%
NeuralStyle	*RGB*	15	119	86.5%
Proposed model	*RGB+Pose*	15	448	88.9%

The experimental results deeply reveal the core advantages of the improved hypergraph convolution method. The input fuses Pose and Video to construct a spatio-temporal hypergraph, where nodes include joint points and visual key areas, and hyperedges describe the spatio-temporal associations of multiple nodes. This modeling method breaks the limitation of traditional graph convolution. Under dynamic views, it can simultaneously capture "joint motion trajectories" (Pose) and "scene visual changes" (Video), improving recognition accuracy for complex actions. Compared with the same modality method PoseTransformer, the proposed model achieves performance optimization in subsets such as Fixed-view (+0.8%) and Teaching Scenario (0% equal), verifying the efficient modeling ability of hypergraph for multimodal collaborative relationships.

As shown in Table 4 on the Anubis dataset, the proposed model achieved an accuracy of 88.9%, significantly surpassing all single RGB modality methods. In comparison, the combination of multimodal fusion and improved hypergraph convolution enables the model to simultaneously capture "scene visual information" and "joint motion information," breaking the limitation of single RGB modality in describing human body structure. Experimental data deeply validate the core advantages of the improved hypergraph convolution method. The improved hypergraph convolution combined with time dimension learning can capture dynamic temporal features of actions. In the action recognition on the Anubis dataset, the temporal module can locate the "force peak frame," enhancing the ability to distinguish different action phases. Although the GFLOPs are slightly higher than single RGB methods, the improvement in accuracy indicates that spatiotemporal hypergraph modeling + temporal feature fusion provides a higher computational cost-effectiveness in complex action recognition, especially suitable for sports teaching scenarios.

Figure 4 visualizes the spatiotemporal motion features of sports actions through joint trajectory curves in three-dimensional space. The effectiveness of this visualization scheme is reflected in: (1) Accurate presentation of action trajectories: The three-dimensional coordinate system clearly

depicts the position changes of joints in space, and the shape of the curve reflects the dynamic stages of the action. For example, the spatial distribution and temporal variation of the red and green trajectories can help teachers quickly identify the "start-execution-end" phases of the action. (2) Intuitive display of multi-joint coordination: The joint associations modeled by hypergraph convolution are reflected through spatial associations of the trajectories. In teaching, comparing trajectory deviations between standard and student actions can locate joint motion errors and achieve fine-grained error diagnosis.

The sports action information visualization method proposed in this paper transforms fine-grained action data recognized by improved hypergraph convolution into intuitive teaching feedback through 3D trajectory display, joint deviation annotation, and multimodal fusion presentation. The combination of examples in Figure 4 and experimental data validates its effectiveness in automated sports teaching: it not only enhances students' understanding of action structure but also provides teachers with quantitative error analysis tools, realizing a "recognition–visualization–feedback" closed-loop teaching assistant mechanism, strongly supporting the core research objective of the paper and providing key technical support for the intelligent upgrade of sports teaching.

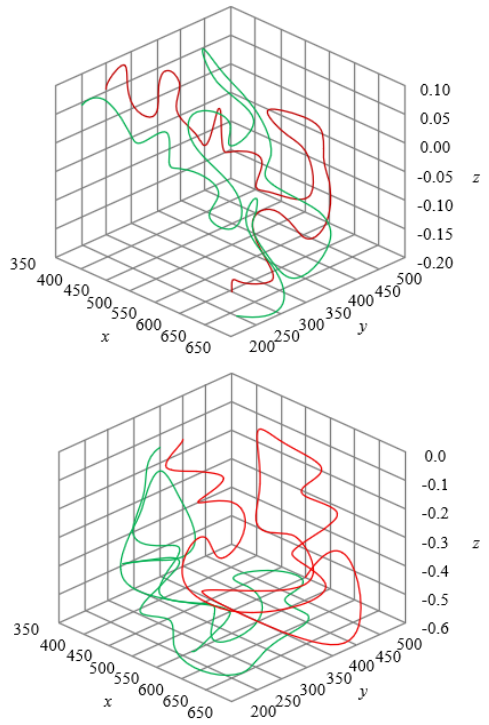

Figure 4. Examples of joint trajectory visualization of different sports actions

5. CONCLUSION

This paper, focusing on "the development of an automated sports teaching assistant tool based on image recognition," constructed a closed-loop system of "action recognition–information visualization." In the part of automated sports action recognition based on improved hypergraph convolution, aiming at the problem that traditional methods are insufficient in modeling high-order joint collaboration and spatiotemporal features, a time and channel refined hypergraph convolution network is proposed: by constructing a hypergraph model to describe dynamic associations among multiple joints, combining spatiotemporal hypergraph convolution to capture inter-frame dependencies in action sequences, and enhancing feature expression through multimodal data fusion. In the part of sports action information visualization, a multi-view interactive framework is designed. Through 3D joint trajectory display, keyframe deviation annotation, and multi-dimensional feature charts, the fine-grained action data output from hypergraph convolution are transformed into intuitive teaching feedback. This scheme supports user-defined chart parameters, querying specific joint/frame features, and achieves action comparison analysis through "standard library expansion," significantly lowering the technical threshold of action analysis, providing teachers with quantitative error diagnosis tools, and offering students visual guidance for action correction.

There are certain limitations in this research. The action coverage for niche sports is insufficient; error action annotation relies on manual work, and the annotation efficiency and consistency need improvement. The dynamic edge computation of hypergraph convolution increases model parameters, requiring further optimization of computational efficiency for deployment on mobile devices. Currently, only skeleton and RGB data are fused, and other modalities such as inertial sensors and pressure sensors are not fully utilized, leading to insufficient modeling of mechanical characteristics of actions.

Future research directions include the following:

(1) Dataset expansion and automatic annotation: Introduce generative adversarial networks to synthesize data for niche actions, combine weak supervision learning to reduce manual annotation costs; build cross-modal annotation tools to automatically align skeleton data with mechanical sensor signals.

(2) Model optimization and lightweight design: Study sparse representation of hypergraph structures, compress models using knowledge distillation technology, and realize real-time inference on mobile devices.

(3) Deep multimodal fusion and cross-scenario application: Fuse IMU data to construct a "skeleton-mechanics" joint hypergraph, expand to fields such as sports injury prevention and competitive sports technical analysis, forming a more comprehensive action analysis system.

Through the innovative combination of improved hypergraph convolution and visual interaction, this paper lays a key technical foundation for automated sports teaching assistant tools. The research results not only improve the recognition accuracy of complex sports actions but also, through the dataset design adapted to teaching scenarios and visual feedback, realize the technical implementation. Future research can continue to deepen along the dimensions of data, models, and applications, promoting the development of sports teaching towards intelligence and personalization, and contributing to the popularization of national fitness and the improvement of professional training under the background of "integration of sports and education."

ACKNOWLEDGMENT

This paper was supported by 2024 Hubei Provincial Higher Education Teaching Research Project (Grant No.: 2024509).

REFERENCES

[1] Yell, M.L., McNamara, S., Prince, A.M. (2021). Adapted physical education: Meeting the requirements of the individuals with disabilities education act. Teaching Exceptional Children, 54(1): 70-78. https://doi.org/10.1177/00400599211038380

[2] Wang, G.Y., Pereira, B., Mota, J. (2005). Indoor physical education measured by heart rate monitor: A case study in Portugal. Journal of Sports Medicine and Physical Fitness, 45(2): 171-177.

[3] Mackenzie, T., Sallis, J., Beets, M., Beighle, A., Erwin, H., Lee, S. (2012). Physical education's role in public health: Steps forward and backward over 20 years and HOPE for the future. Research Quarterly for Exercise and Sport, 83(2): 125-135.

[4] Pate, R.R., O'Neill, J.R., McIver, K.L. (2011). Physical activity and health: Does physical education matter?. Quest, 63(1): 19-35. https://doi.org/10.1080/00336297.2011.10483660

[5] Koh, Y. (2021). Combining adapted physical education with individualized education programs: Building Korean pre-service teachers' self-efficacy for inclusive physical education. Sustainability, 13(5): 2879. https://doi.org/10.3390/su13052879

[6] Fang, Q., Zhang, Y.W. (2024). Optimizing remote teaching interaction platforms through multimodal image recognition technology. Traitement du Signal, 41(1): 225-235. https://doi.org/10.18280/ts.410118

[7] Tran, T., Ternov, N.K., Weber, J., Barata, C., et al. (2022). Theory-based approaches to support dermoscopic image Interpretation Education: A review of the literature. Dermatology Practical & Conceptual, 12(4): e2022188. https://doi.org/10.5826/dpc.1204a188

[8] Morris, B. (2014). Image processing and pattern recognition research center of the technical university of Cluj-Napoca, Romania [ITS Research Lab]. IEEE Intelligent Transportation Systems Magazine, 6(2): 70-73. https://doi.org/10.1109/MITS.2014.2309315

[9] Zheng, D., Yuan, Y. (2022). Time series data prediction and feature analysis of sports dance movements based on machine learning. Computational Intelligence and Neuroscience, 2022(1): 5611829. https://doi.org/10.1155/2022/5611829

[10] Koshio, T., Haraguchi, N., Takahashi, T., Hara, Y., Hase, K. (2024). Estimation of Ground Reaction Forces during Sports Movements by sensor fusion from inertial measurement units with 3D forward dynamics model. Sensors, 24(9): 2706. https://doi.org/10.3390/s24092706

[11] Wesely, S., Hofer, E., Curth, R., Paryani, S., Mills, N., Ueberschär, O., Westermayr, J. (2025). Artificial intelligence for objective assessment of acrobatic movements: applying machine learning for identifying tumbling elements in cheer sports. Sensors, 25(7): 2260. https://doi.org/10.3390/s25072260

[12] Alshardan, A., Mahgoub, H., Alahmari, S., Alonazi, M., Marzouk, R., Mohamed, A. (2025). Cloud-to-Thing continuum-based sports monitoring system using machine learning and deep learning model. PeerJ Computer Science, 11: e2539. https://doi.org/10.7717/peerj-cs.2539

[13] Host, K., Ivašić-Kos, M. (2022). An overview of Human Action Recognition in sports based on Computer Vision. Heliyon, 8(6): e09633. https://doi.org/10.1016/j.heliyon.2022.e09633

[14] Blythman, R., Saxena, M., Tierney, G.J., Richter, C., Smolic, A., Simms, C. (2022). Assessment of deep learning pose estimates for sports collision tracking. Journal of Sports Sciences, 40(17): 1885-1900. https://doi.org/10.1080/02640414.2022.2117474

[15] Ait-Bennacer, F.E., Aaroud, A., Akodadi, K., Cherradi, B. (2022). Applying deep learning and computer vision techniques for an e-sport and smart coaching system using a multiview dataset: Case of shotokan karate. International Journal of Online & Biomedical Engineering, 18(12): 35-53. https://doi.org/10.3991/ijoe.v18i12.30893

[16] Sun, Y., Li, Y. (2022). A deep learning method for intelligent analysis of sports training postures. Computational Intelligence and Neuroscience, 2022(1): 2442606. https://doi.org/10.1155/2022/2442606

[17] Lee, H., Kim, Y.S., Kim, M., Lee, Y. (2021). Low-cost network scheduling of 3D-CNN processing for embedded action recognition. IEEE Access, 9: 83901-83912. https://doi.org/10.1109/ACCESS.2021.3087509