UDC 004.8

CLASSIFICATION OF RHYTHMIC GYMNASTICS SPORT ELEMENTS BY VIDEO

КЛАСИФІКАЦІЯ СПОРТИВНИХ ЕЛЕМЕНТІВ ХУДОЖНЬОЇ ГІМНАСТИКИ ЗА ВІДЕО

Anastasia Neskorodieva
Анастасія Нескородьєва

Vasyl' Stus Donetsk National University, Vinnytsia, Ukraine
ORCID: https://orcid.org/0000-0002-8591-085X
E-mail: neskorodieva.a@gmail.com

DOI: https://doi.org/10.15673/atbp.v16i2.2845

Abstract. *The work devoted to human posture recognition during rapid movements and complex non-standard poses due to the large number of limbs involved in the movement. Rhythmic gymnastics was chosen as the subject area and, accordingly, the specifics of the judge's assessment of the athlete's performance. Many synchronized videos with fast movements and sequences of complex poses from different angles allows us to form a data set necessary for further research and implementation of the results obtained both in socially important industries and in the market of commercial services using artificial intelligence technologies. A computer system has been developed that can be used to increase the objectivity of sports judging at rhythmic gymnastics competitions, as well as to become an alternative to the traditional judging system in the case of competitions held in a remote format. By scaling up the task, the system can also be used to diagnose problems with the human nervous system and musculoskeletal system. As a result of the research, a dataset depicting the performance of sports elements was collected and structured. The peculiarities of the mediapipe and ViTPose models were identified and the best solution for preprocessing the prepared set was chosen. The main result of this work is a built and trained model for classifying sports elements, which classifies 7 elements with an accuracy of 0.9048. The accuracy indicates that the model performs at a high level, correctly classifying sports elements in most cases. This level of accuracy indicates that the model has been effectively trained to classify these specific elements. In the future, to be able to fully evaluate the performances of female rhythmic gymnasts, it is necessary to add tracking of the object with which the athlete performs, to create a method for tracking interaction with it.*

Анотація. *Робота присвячена розпізнаванню пози людини під час швидких рухів і складних нестандартних поз за рахунок великої кількості задіяних у русі кінцівок. Тематичним напрямом було обрано художню гімнастику і, відповідно, специфіку суддівської оцінки виступу спортсменки. Велика кількість синхронізованих відео зі швидкими рухами та послідовністю складних поз з різних ракурсів дозволяє сформувати набір даних, необхідний для подальших досліджень та впровадження отриманих результатів як у соціально важливих галузях, так і на ринку комерційних послуг з використанням технологій штучного інтелекту. Розроблено комп'ютерну систему, яка може бути використана для підвищення об'єктивності спортивного суддівства на змаганнях з художньої гімнастики, а також стати альтернативою традиційній системі суддівства у разі проведення змагань у дистанційному форматі. Розширивши завдання, систему також можна використовувати для діагностики проблем з нервовою системою та опорно-руховим апаратом людини. У результаті дослідження було зібрано та структуровано набір даних, що відображає виконання спортивних елементів. Було визначено особливості моделей mediapipe та ViTPose та обрано найкраще рішення для попередньої обробки підготовленого набору. Основним результатом роботи є побудована та навчена модель класифікації спортивних елементів, яка класифікує 7 елементів з точністю 0,9048. Точність свідчить про те, що модель працює на високому рівні, у більшості випадків правильно класифікуючи спортивні елементи. Цей рівень точності вказує на те, що модель було ефективно навчено класифікувати ці конкретні елементи. У майбутньому, щоб мати можливість повноцінно оцінювати виступи художніх гімнасток, необхідно додати відстеження об'єкта, з яким виступає спортсменка, створити метод відстеження взаємодії з ним.*

Keywords: rhythmic gymnastics, machine learning, human pose recognition, classification, transformer.

Ключові слова: художня гімнастика, машинне навчання, розпізнавання пози людини, класифікація, трансформатор.

I. INTRODUCTION

The field of sports analytics is rapidly evolving due to advances in technology and data analysis. Intelligent video content analysis has become an important tool for sports coaches, athletes, and analysts to gain insight into the performance of individual players and teams. Professional sports are a promising area for the application of machine learning technologies. The sport of rhythmic gymnastics was chosen as a research subject for the following reasons.

Rhythmic gymnastics is a specific sport that combines elements of ballet, dance, and gymnastics, where athletes perform exercises using hand-held apparatus such as a rope, ball, hoop, clubs, or ribbon. It is a beautiful and elegant sport that requires strength, flexibility, coordination, and musicality. Therefore, judging Rhythmic Gymnastics is one of the most difficult and subjective tasks in sports judging in general.

The high degree of complexity of this sport also complicates the process of forming an objective assessment of Rhythmic Gymnastics by sports judges. Rhythmic gymnastics can involve intricate and complex movements that require great skill and precision. Athletes must perform a variety of body movements, jumps, balances, and throws while manipulating a variety of apparatus. The difficulty of the routines can vary greatly from athlete to athlete, and judges should award points based on the level of difficulty of the athletic elements.

Another major reason for the difficulty in judging Rhythmic Gymnastics performances is the subjective nature of the sport. The aesthetic appeal of rhythmic gymnastics plays a significant role in the scoring system. Athletes are judged not only on technical skill, but also on grace, artistic interpretation, and expression. Judges must consider factors such as music, choreography, costume, and use of equipment in their evaluation. These subjective factors make it difficult to standardize a scoring system and can lead to discrepancies in judges' scores. The speed and fluidity of a performance can make it difficult for judges to capture every movement and detail in real time. Video replays can help judges make more accurate judgments, but they can be time consuming.

This article is a continuation of the papers [1, 2, 3]. The aim of this article is to investigate models and methods for classifying rhythmic gymnastics sport elements from video using neural networks. The creation of a program to improve the objectivity of judging the performances of gymnasts is relevant. To accomplish this task, it was necessary to analyze the existing systems for classifying human postures in sports and to investigate the existing models and methods for classifying sports elements. After the theoretical study of the topic, it was necessary to form an appropriate dataset for training and analyze software implementations of systems for classifying sports elements in sport (rhythmic gymnastics). As a result, a classification model was developed taking into account the peculiarities of sports judging in rhythmic gymnastics.

II. LITERATURE ANALYSIS

The use of machine learning in modern sports is extremely diverse [4-9], and Sport tech is becoming an increasingly developed and commercially attractive industry. The first examples of application were predicting winners in sports competitions [10-13], i.e., in other words, sports betting. Initially, data mining was simply a classical approach to mathematical statistics using computer technology, and later developed into an independent scientific mathematical field [14] with powerful academic research tools.

Gymnastics is a competitive sport or to improve strength, agility, coordination, and physical conditioning [15]. The International Gymnastics Federation (FIG) recognizes the following disciplines among the areas of artistic gymnastics [16]:

- men's artistic gymnastics, women's artistic gymnastics;
- rhythmic gymnastics (rhythmic individual, rhythmic group);
- trampoline gymnastics, double-mini trampoline, tumbling;
- acrobatic gymnastics;
- aerobic gymnastics;
- parkour speed, parkour freestyle.

Parkour (freerunning) was added to this list in 2018. The organization of objective sports judging of rhythmic gymnastics among these sports is perhaps the most difficult task, as rhythmic gymnastics has certain characteristics that differ from the other mentioned disciplines.

Rhythmic gymnastics is a specific sport, the analysis of which is complicated by the need to assess the posture of a person during rapid movements and non-standard poses of a complex type due to the large number of limbs involved in the movement. [17-18].

Theoretically, the task presented in this paper is a special case of the general problem of recognizing human activity from graphical materials. This is a rather difficult task due to problems such as background clutter, partial occlusion, changes in scale, viewpoint, lighting, and appearance. To summarize [19], the classification of methods for recognizing human activity from video fragments or still images is shown in Figure 1.

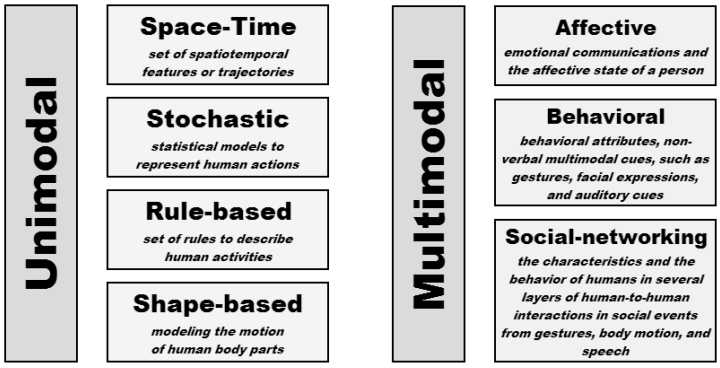


Fig.1 - human activity recognition methods

Given the specific aspects of rhythmic gymnastics, it is worth focusing on some of the existing specialized assessment systems.

The paper [18] describes an algorithm for assigning judges' scores to rhythmic gymnastics movements. The algorithm is implemented as real-time computer vision software. As input, it takes a video image or video stream of a live performance and extracts detailed information about the velocity field from the body movements and transforms it into specialized spatio-temporal image patterns. By comparing individual performances of the same atomic gymnastics routine, the method assigns a quality score that is related to the distance between the corresponding spatio-temporal trajectories. For several standard gymnastic movements, the method assigns scores that are comparable to those assigned by expert judges. The disadvantage of this solution is that the method has to collect a new set of data on how judges evaluate athletes' performances under the new rules and train new models after each rule change. The method does not have a variable to quickly change the scoring of elements whose "value" has changed.

Another study [20] notes that most existing work focuses only on the dynamic information of the video (i.e., motion information), but ignores the specific poses that the athlete performs in the video, which is important for action estimation in long videos. In this paper, we present a novel hybrid dynamic-static context-aware attention network for action estimation in long videos. To obtain more discriminative images for videos, not only the dynamic information of the video is studied, but also the static poses performed by the athletes in certain frames, which represent the quality of the action at certain moments, are paid attention to along with the proposed hybrid dynamic-static architecture. In addition, a context-sensitive attention module consisting of a temporal instance-based convolutional network unit and an attention unit for both streams is used to obtain more robust stream characteristics, where the former is designed to learn the relationships between instances and the latter to assign the correct weight to each instance. The result of the ablation study, which demonstrates the contribution of the context-sensitive attention module to the method developed in this paper, is shown in Table 1.

Table 1 - contribution of the context-sensitive attention module

Method	MIT-Skating	Rhythmic Gymnastics			
		Ball	Clubs	Hoop	Ribbon
Avg Pooling	0.605	0.518	0.650	0.650	0.552
SAU	0.590	0.501	0.610	0.703	0.528
LSTM + SAU	0.596	0.487	0.651	0.690	0.570
Bi-LSTM + SAU	0.572	0.522	0.568	0.674	0.600
RTA	0.611	0.522	0.634	0.713	0.565
CAA	0.615	0.528	0.657	0.708	0.578

Finally, the features of the two streams are combined to obtain a regression-based final video score, which is controlled by the real-world information scores provided by the experts. Experimental results confirm the effectiveness of the proposed method, which outperforms similar approaches.

III. OBJECT, SUBJECT AND METHODS OF RESEARCH

3.1 Analysis of models based on convolutional neural networks and transformers.

In machine learning, a perceptron is an algorithm for supervised training of binary classifiers [21]. A binary classifier is a function that can decide whether an input represented by a vector of numbers belongs to a particular class. It is a type of linear classifier, i.e., a classification algorithm that makes its predictions based on a linear predictor function that combines a set of weighting coefficients with a feature vector.

According to modern terminology, perceptrons can be classified as artificial neural networks: with one hidden layer, with a threshold transfer function, and with direct signal propagation.

Convolutional Neural Networks (CNNs) in machine learning are a class of deep artificial neural networks with direct propagation that have been successfully applied to visual image analysis [22].

CNNs use a type of multilayer perceptron designed to require a minimal amount of preprocessing. They have a shared-weight architecture and characteristics of invariance with respect to parallel transfer.

Convolutional networks are based on a biological process, namely the connection pattern of neurons in the visual cortex of animals. Individual cortical neurons respond to stimuli only in a limited area of the visual field, known as a receptive field. The receptive fields of different neurons partially overlap so that they cover the entire visual field.

CNNs use relatively little preprocessing compared to other algorithms for image classification. This means that the network learns filters that in traditional algorithms were constructed manually. This independence in feature construction from a priori knowledge and human effort is a great advantage. They have applications in image and video recognition, recommender systems, and natural language processing.

Transformer is a deep learning model. It is distinguished by differential weighting of the importance of each part of the input data (which includes recursive output).

Like recurrent neural networks (RNNs) [23], transformers are designed to process sequential input data, such as natural language, with applications in tasks such as translation and text summarization. However, unlike RNNs, transformers process the entire input signal simultaneously. The attention mechanism provides context for any position in the input sequence. For example, if the input is a natural language sentence, the transformer does not need to process one word at a time. This allows for more parallelization than RNNs, and therefore reduces training time and allows for training models on larger data sets.

In [24], the authors compared well-known projects that used this structure to build models. The study showed that the use of transformers for CV tasks has become a new area of research to reduce the complexity of the architecture and study the scalability and efficiency of training.

Transformer becomes a more general framework for exploring sequential data, including text, images, and time series data.

3.2. Dataset

Human pose estimation is a well-known problem in computer vision for determining joint positions. Existing datasets for pose learning have proven to be insufficiently complex in terms of pose variety, object occlusion, and viewpoints. This makes the process of pose annotation relatively simple and limits the use of models trained on them. In [25], the authors study existing datasets for human pose classification and propose a new dataset. For a greater variety of human poses, the authors propose the concept of fine-grained hierarchical pose classification, in which they formulate pose estimation as a classification task and propose the Yoga-82 dataset for recognizing large-scale yoga poses using 82 classes. Yoga-82 consists of complex poses where accurate annotation may not be possible. To solve this problem, the authors provide hierarchical labels for yoga poses based on the body configuration of the pose. The dataset contains a three-level hierarchy, including body positions, variations of body positions, and actual pose names. The paper also presents the classification accuracy of state-of-the-art convolutional neural network architectures on Yoga-82. And it also presents several hierarchical variants of DenseNet for the use of hierarchical labels.

For the study [17], the authors created their own dataset consisting of 1000 videos of female athletes performing with the objects hoop, ball, clubs, and ribbon, 250 videos with each of these objects. The length of each recording is on average one minute and 30 seconds. The video deletes the parts of the athlete's entrance to the carpet, preparation for the start of the performance, and the athlete leaving the carpet after the performance. Each video is described by the following 5 parameters: difficulty score, execution score, overall score, penalties, and number of frames.

The disadvantage is that after each change in the rules, you will have to collect and prepare a new dataset to build methods for the updated rules. This dataset is designed only for the 2016-2020 rules and scoring system.

The proposed dataset consists of a test and training data set for further training of the sports elements classification model. The five most popular elements in terms of performance were selected for classification. When testing the first trained models on this dataset, many false positive classifications were found. Namely, the method often assigned an element class to an image, even though the athlete does not perform any of the elements in the dataset. Therefore, it was decided to try to add two additional classes that characterize the most frequent positions of athletes that are not evaluated elements. Figure 2 shows how the five scored elements, classes 1 through 5, and the two non-scored positions, classes 6 and 7, look like. Examples of the elements are taken from the official FIG rules.

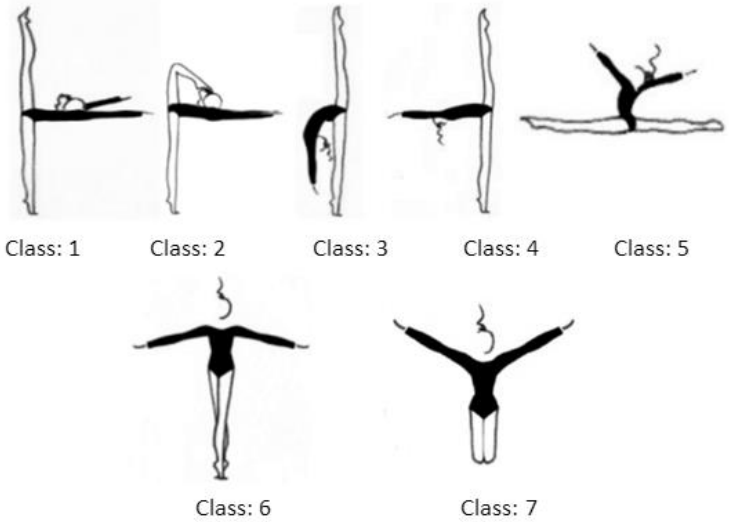


Fig.2 - visualization of classes
Рис.2 - Візуалізація класів

Dataset requirements: images of individual female athletes, which should occupy at least 50% of the image length in full growth. All parts of the body should be included in the image. The minimum image quality can be from 240 by 360 pixels and should be sufficient to accurately determine the position of the athlete's skeleton.

Open resources were used to collect the dataset: [26-30].

IV. RESULTS

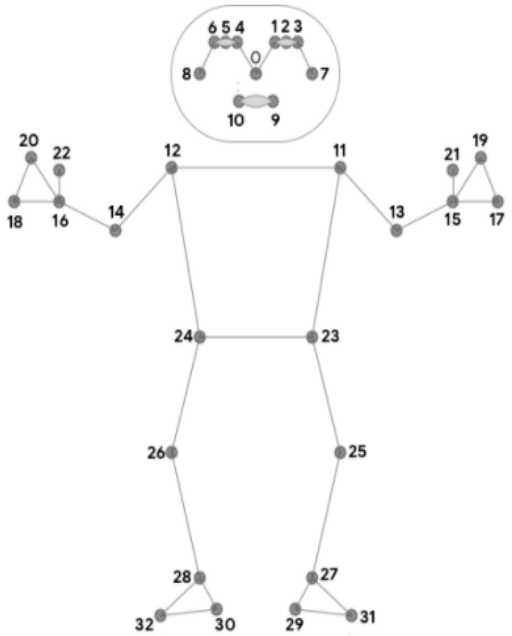
4.1 Practical comparison of different methods of pose assessment

The MediaPipe library offers cross-platform customizable machine learning solutions for live and streaming media [31]. This set of flexible tools is built on top of the TensorFlow Lite library to maximize machine learning adoption and hardware performance.

MediaPipe Pose is a machine learning solution for body pose tracking that identifies 33 3D landmarks, shown in Figure 3, and a full-body background segmentation mask from RGB video footage.

The result of MediaPipe Pose processing of a single image is a vector of 132 values, where the x, y, z coordinates in space and the quality of visibility are recorded in the sequence from the first key point to the 33rd. Visibility means whether it is possible to see the body part to which the key point belongs in the frame. If the value is close to 1, then the body part is visible in a high-quality, clear and unambiguous way. If the value is closer to 0.5, then the body part to which the point belongs may be overlapped by another object (body part), but the points adjacent to the skeleton allow you to determine the likely location of this point. If the value is close to 0, then the body part is not visible at all or does not even appear in the image.

Having processed our own dataset with the MediaPipe Pose method, we have a set of vectors describing 7 classes, 96 vectors on average for each class. This will be used to train the element classification model.



0. nose
1. left_eye_inner
2. left_eye
3. left_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. left_ear
8. right_ear
9. mouth_left
10. mouth_right
11. left_shoulder
12. right_shoulder
13. left_elbow
14. right_elbow
15. left_wrist
16. right_wrist
17. left_pinky
18. right_pinky
19. left_index
20. right_index
21. left_thumb
22. right_thumb
23. left_hip
24. right_hip
25. left_knee
26. right_knee
27. left_ankle
28. right_ankle
29. left_heel
30. right_heel
31. left_foot_index
32. right_foot_index

Fig.3 - marking key points in the mediapipe model

Рис.3 - що позначає ключові точки в моделі медіаканалу

A solution based on transformers [32] was chosen. It demonstrates the good capabilities of simple vision transformers for pose estimation from several aspects, namely the simplicity of the model structure, the scalability of the model size, the flexibility of the learning paradigm, and the ability to transfer knowledge between models using a simple base model called ViTPose. Specifically, ViTPose uses conventional and non-hierarchical vision transformers as a backbone to extract features for a given human instance and a lightweight decoder to estimate pose. It can be scaled from 100M to 1B parameters by exploiting the scalable model capacity and high parallelism of the transformers, establishing a new Pareto frontier between throughput and performance. Furthermore, ViTPose is highly flexible in terms of attention type, input resolution, pre-training and fine-tuning strategies, and handling of multiple pose tasks. The authors also empirically demonstrate that knowledge from large ViTPose models can be easily transferred to small ones using a simple knowledge marker. Experimental results show that the basic ViTPose model outperforms representative methods on the challenging MS COCO Keypoint Detection benchmark, while the largest model sets a new state-of-the-art. To run this method in real time, a minimum-performance graphics card such as the Nvidia RTX 2080 is required. This model uses a different annotation to describe the detected human skeleton, namely it has only 17 key points. As a result of image processing using this method, a vector describing the x, y coordinates and detection accuracy for each of the 17 key points is generated for each constructed skeleton. The order of the key points is shown in Figure 4.

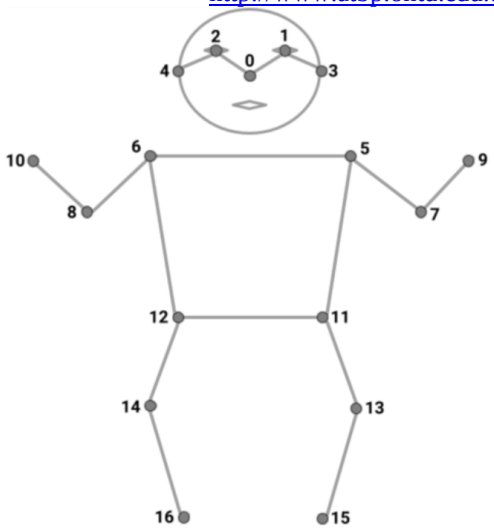


Fig.4 - marking key points in the ViTPose model
Рис.4 позначення ключових точок у моделі ViTPose

Using the methods implemented in these libraries, two algorithms were developed. The first algorithm is for processing our own dataset, and the second is for processing the video stream of athletes' performances.

For comparison, we selected four images that were processed using both methods. The processing results are shown in Figure 5.

In the first row, the images were processed using the mediapipe method, and in the second row, the images were processed using the ViTPose method. In the first image processed with the mediapipe method, the key points of the hips are shifted and the points on the legs are misidentified. In the second image, processed with the same method, the right knee point is marked closer to the hip. In the third image, only the toes of the right foot are misidentified. In the fourth image, where the athlete is performing an element that belongs to the 3rd class, the key points of the torso were confused during the mediapipe method. This resulted in the following confusions:

- the right leg as the right arm;
- the right arm as if it were the left leg;
- the left foot as right foot.

The disadvantage of the Mediapipe Pose method is that the quality of tracking the human skeleton during elements related to rotation and grouping in space is low. For example, when performing rotations, the model starts to confuse the right and left parts of the body. This complicates further development based on this method.

The result of image processing, shown in Figure 5, using the ViTPose method is better than the mediapipe method and has no critical errors. However, this method requires more powerful hardware and takes up more memory.

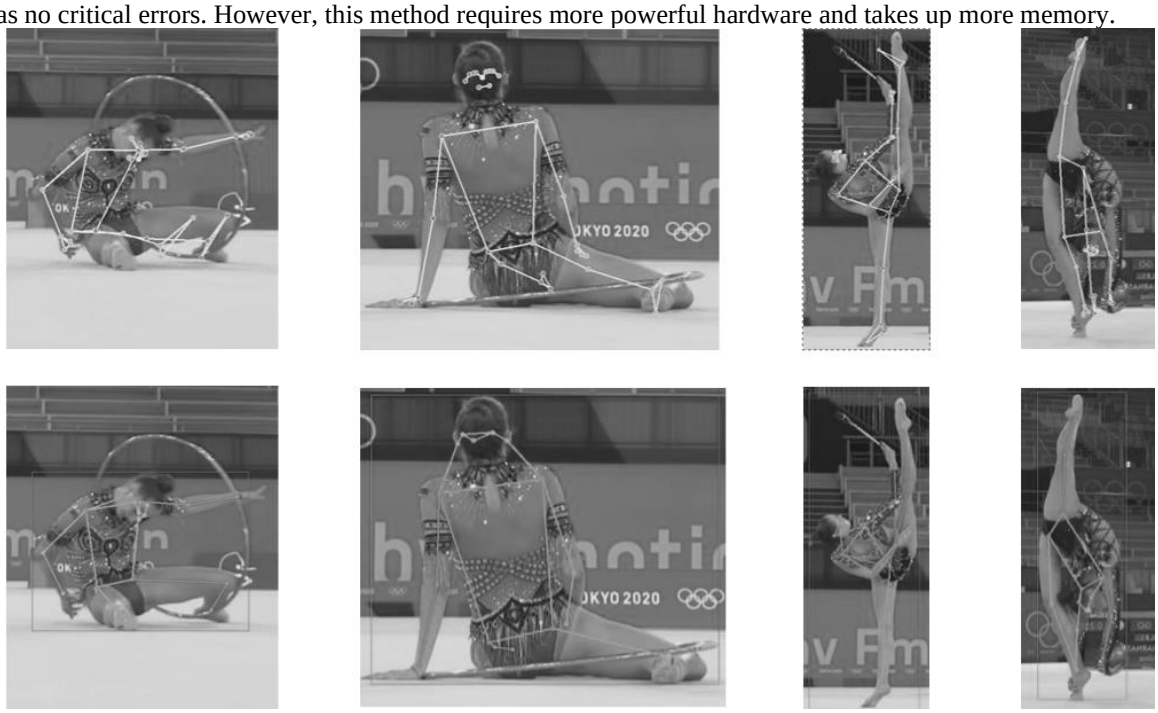


Fig. 5 - first row - mediapipe, second row – ViTPose
Рис. 5 - перший ряд – медіапайп, другий ряд – ViTPose

4.2. Problem mathematical model

The model of a multilayer perceptron is shown in the formula.

$$y_j^{(k)} = f^{(k)}(s_j^{(k)}),\ \ s_j^{(k)} = b_j^{(k)} + \sum_{i=1}^{N^{(k-1)}} w_{ij}^{(k)} y_i^{(k-1)},\ \ j \in \overline{1, N^{(k)}}, k \in \overline{1, L},$$

where $b_j^{(k)}(n)$ is the offset, $w_{ij}^{(k)}(n)$ is the weighting coefficients, $N^{(k)}$ is the number of neurons in k layer, and L is the number of layers.

The mathematical methods used in the code snippet are primarily related to the architecture and training of the neural network. List of methods used.

The dense layers created with keras.layers.Dense perform a matrix multiplication between the input data and the weight matrix, followed by the addition of a bias term. This uses the ReLU6 activation function (tf.nn.relu6), which applies element-wise nonlinearity to the matrix multiplication result.

$$f(s) = \max\{0, s\},\ \ f(s) \in R_+.$$

Neuronal dropout: The keras.layers.dropout layer applies random dropout, which randomly sets a fraction of the input to 0 during training. This method helps prevent overfitting. At each training iteration, a binary mask is created with elements set to 0 with a certain probability.

Softmax activation: the last layer of keras.layers.Dense uses the softmax activation function in multiclass classification tasks. The softmax function converts the output of the previous layer into a probability distribution by class. Mathematically, softmax activation can be defined as:

$$f(\mathbf{s})_i = \text{softmax}(\mathbf{s})_i = \frac{\exp(s_i)}{\sum_j \exp(s_j)},\ \ f(\mathbf{s})_i \in (0,1).$$

The loss function is the categorical cross-entropy. This loss function is commonly used in multi-class classification tasks. It quantifies the difference between the predicted probability distribution and the true distribution of class labels.

$$E = -\frac{1}{P}\sum_{m=1}^{P}\sum_{j=1}^{K} d_{mj} \ln y_{mj},\ \ d_{mj} \in \{0,1\},\ \ y_{mj} \in [0,1].$$

For optimization, we chose the Adam method (optimizer='adam') [33], which is a popular gradient-based optimization algorithm. Adam combines the ideas of adaptive learning rates and momentum methods to efficiently update network weights during training.

These mathematical methods and operations are the main components of building and training neural networks to solve various machine learning tasks.

4.3. Project programmatic implementation

The model building starts with defining the input data, which has the size of 51 neurons. tf.keras. Input is a function for creating the input layer of the neural network. Then, the input data is passed to the landmarks_to_embedding function, which converts the key points of the human skeleton into a vector with a fixed number of elements. After that, two Dense layers are created with 128 and 64 neurons, respectively, with the activation function tf.nn.relu6. Dense is a type of layer that connects all the neurons of the previous layer to each neuron of the current layer. Between these two layers, a Dropout layer is inserted with a 0.5 probability of randomly turning off neurons to prevent overtraining.

At the output of the last Dropout layer, the final Dense layer is created with the number of neurons corresponding to the number of classes. The activation function of the last layer is "softmax" to obtain the probability of each class. During the first iteration of model training, the model weights are initialized with random values. Then the model is trained on the training set with random weights. During this process, the model will adjust its weights using the Adam algorithm to reduce the value of the loss function on the training set.

During training, the model uses gradient descent to optimize the weights. The goal is to find the global minimum of the loss function where the model weights provide optimal predictions on the new data.

At the end of each epoch, the model is evaluated on the validation set to determine its accuracy and decide whether to stop training by calling the early stopping callback and saving the checkpoint callback that were set during compilation and training.

The accuracy of the trained model on the test sample is 0.9048.

Python version 3.7.9 was chosen to write the computer program. The technical requirements for software and hardware are as follows: operating system: Windows 11, processor: AMD Ryzen 5 3600, RAM: DDR4 16GB 3600hz, internal memory: 512Gb, Video card: Nvidia RTX 2080 Super, which is necessary for launching a real-time solution based on Transformers.

The work with this program is divided into three stages: data markup, training of the pose classification model, and processing of video performances of athletes. To mark up images, you need to properly organize the data set:

1. The images of all elements (classes) that are planned to be identified are sorted into separate folders, where the folder name is the name of the element whose images are in it.

2. Divide the image data into two folders: train and test without repetition.
3. The train and test folders should be located in the dataset folder.

After executing the data_markup.py file, two files train_data.csv and test_data.csv will appear in the dataset folder. These files are needed to train the pose classification model.

To train the classification model, run the train.py file. After training, the file rg_pose_classifier.tflite will appear in the models folder. This is the model for classifying sports elements.

To process video, you need to assign the path to the file to be processed to the input_video variable in the app.py file. It is possible to interactively select other models of human detection and skeleton construction for image processing. Next, you need to run the app.py file. The result of its work will be a new video in which in the upper left corner there will be an image of the class of the element performed by the athlete and the accuracy of predicting this class by the classification model.

The program itself works like this:

1. The program reads the video to be processed frame by frame.
2. Each frame is passed to the first model for human detection. As a result, the selected method returns the coordinates of the bounding box and the detection accuracy of each box where a person was detected.
3. The next model receives a frame from the video and the corresponding annotation from the first model to determine the position of the skeleton. The results of the first model are needed to narrow the image processing area for the second model. In other words, the model will determine the position of the human skeleton only within the bounding box of the first model. This allows you to speed up the program in this place. As a result, this model provides the coordinates of key points of the human skeleton and the quality of visibility for each of the key points.
4. The element classification model receives a vector of values for each frame, and the result of processing is the probability for each of the 7 classes. Among the classes, the one with the highest probability is selected and satisfies the minimum desired value.

In the upper left corner of each frame, the program visualizes the corresponding image of the class that was classified by the corresponding method. Also, to the right of the class image, the program adds the accuracy with which the method has identified this class. If the classification accuracy is not sufficient, the program leaves the original frame unchanged (Figure 6).

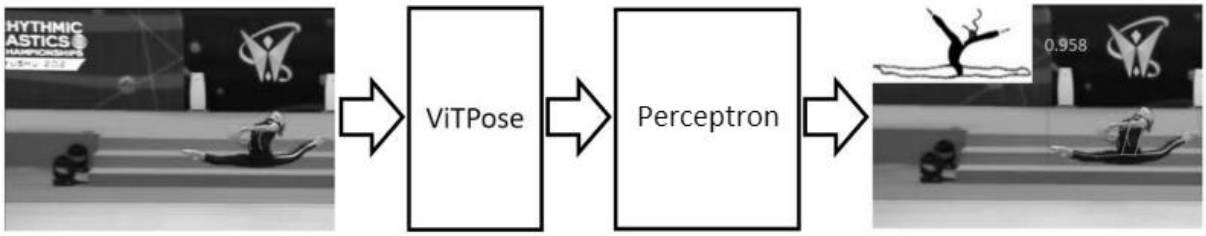


Fig.6 - frame processing
Рис.6 Обробка кадру

V. CONCLUSIONS

A computer system has been developed that can be used to increase the objectivity of sports judging in rhythmic gymnastics competitions, as well as an alternative to the traditional judging system in the case of competitions held in a remote format. By scaling up the task, the system can also be used to diagnose problems with the human nervous system and musculoskeletal system. As a result of the research, a dataset representing the performance of sports elements was collected and structured. The peculiarities of the MediaPipe and ViTPose models were identified and the best solution for preprocessing the prepared set was chosen. The main result of this work is a built and trained model for classification of sports elements, which classifies 7 elements with an accuracy of 0.9048. The accuracy indicates that the model performs at a high level, correctly classifying sports elements in most of the cases. This level of accuracy indicates that the model has been effectively trained to classify these specific elements.

To further develop this method, it is necessary to expand the dataset by increasing the number of classes and examples for them. Also, to improve the classification results, try to use LSTM and the attention mechanism. To maintain the relevance of the evaluation of the elements performed by the athlete, develop a database for storing and easily modifying the evaluation values. In the future, to be able to fully evaluate the performances of rhythmic gymnasts, it is necessary to add the tracking of the object with which the athlete performs, to create a method of tracking the interaction with it. In the future, a similar system can be built to detect problems such as scoliosis by analyzing a patient's posture, prescribe physical therapy, and study the cognitive development of young children's brains by monitoring their motor activity.

REFERENCES

[1.]Neskorodieva, A., Strutovskyi, M., Baiev, A., & Vietrov O. (2023). Real-time Classification, Localization and Tracking System (Based on Rhythmic Gymnastics). 2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT), 11-16. https://doi.org/10.1109/ELIT61488.2023.10310664

[2.]Neskorodieva, A. (2023). Neural network methods for automatic person pose estimation in rhythmic gymnastics exercises. Ukrainian Journal of Information Systems and Data Science, 1(1), 53-65. https://jujisds.donnu.edu.ua/article/view/14739

[3.]Neskorodieva, A.R. (2023). Computer program "Pose estimation for sports (Rhythmic gymnastics)", UANIPIO, Ukraine, #116622, bul. no. 75. https://sis.nipo.gov.ua/en/search/detail/1739332/.
[4.]Rizzoli, A. (2021). 7 Game-Changing AI Applications in the Sports Industry. https://www.v7labs.com/blog/ai-in-sports (date of access: 30.01.2024).
[5.]Brefeld, U., Davis, J., Lames, M., & Little, J.J. (2021). Machine Learning in Sports. *Dagstuhl-Seminar*, 11 (9), 21411. https://doi.org/10.4230/DagRep.11.9.45
[6.]Chmait, N., & Westerbeek, H. (2021). Artificial Intelligence and Machine Learning in Sport Research: An Introduction for Non-data Scientists. *Front Sports Act Living*, 3, 682287. https://doi.org/10.3389/fspor.2021.682287
[7.]Richter, C., O'Reilly, M., & Delahunt, E. (2021). Machine learning in sports science: challenges and opportunities. *Sports Biomechanics*. https://doi.org/10.1080/14763141.2021.1910334
[8.]Musa, R.M., Taha, Z., Majeed, A.P.P.A., & Abdullah, M.R. (2019). Machine Learning in Sports. *Springer Singapore, SpringerBriefs in Applied Sciences and Technology*. https://doi.org/10.1007/978-981-13-2592-2
[9.]Pearson, A.W. (2019). *The A.I. Sports Book: How AI and Machine Learning can revolutionize the sports*. Independently published.
[10.] Bunker, R., & Susnjak, T. (2022). The Application of Machine Learning Techniques for Predicting Match Results in Team Sport: A Review. *Journal of Artificial Intelligence Research*, 73. https://doi.org/10.1613/jair.1.13509
[11.] Horvat, T., & Job, J. (2022). The use of machine learning in sport outcome prediction: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12 (2). https://doi.org/10.1002/widm.1380
[12.] Lotfi, S., & Rebbouj, M. (2021). Machine Learning for sport results prediction using algorithms. *International Journal of Information Technology*, 3 (3), *148-155*. https://doi.org/10.52502/ijitas.v3i3.114
[13.] Bunkera, R.P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics*, 15 (1), *27-33*. https://doi.org/10.1016/j.aci.2017.09.005
[14.] Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep Learning*. The MIT Press.
[15.] Frederick, A.B. (2023). Gymnastics. https://www.britannica.com/sports/gymnastics. (date of access: 30.01.2024).
[16.] Fédération Internationale de Gymnastique (2024). https://www.gymnastics.sport/site/. (date of access: 30.01.2024).
[17.] Mack, M., Bryan, M., Heyer, G., & Heinen, T. (2019). Modeling Judges' Scores in Artistic Gymnastics. *The Open Sports Sciences Journal*, 12 (1), 1-9. https://doi.org/10.2174/1875399X01912010001
[18.] Pino Díaz-Pereira, M., Gómez-Conde, I., Escalona, M., & Olivieri, D.N. (2014). Automatic recognition and scoring of Olympic rhythmic gymnastic movements. *Human Movement Science*, 34, 63-80. https://doi.org/10.1016/j.humov.2014.01.001
[19.] Vrigkas, M., Nikou, C., & Kakadiaris, I.A. (2015). A Review of Human Activity Recognition Methods. *Front. Robot. AI*, 2:28. https://doi.org/10.3389/frobt.2015.00028
[20.] Zeng, L.-A., Hong, F.-T., Zheng, W.-S., Yu, Q.-Z., Zeng, W., Wang, Y.-W., & Lai, J.-H. (2020). Hybrid Dynamic-static Context-aware Attention Network for Action Assessment in Long Videos. *arXiv.org*. 1-10. https://arxiv.org/abs/2008.05977.
[21.] Freund, Y., & Schapire, R. (1999) Large Margin Classification Using the Perceptron Algorithm. *Machine Learning*, 37, 277-296. https://doi.org/10.1023/A:1007662407062
[22.] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv.org*, 1-21. https://arxiv.org/abs/1311.2524.
[23.] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin I. (2017). Attention Is All You Need. *arXiv.org*, 1-15. https://arxiv.org/abs/1706.03762.
[24.] He, C. (2020). Transformer in CV. *Medium*. https://towardsdatascience.com/transformer-in-cv-bbdb58bf335e.
[25.] Verma, M., Kumawa t, S. Nakashima, Y., Raman, S. (2020). Yoga-82: A New Dataset for Fine-grained Classification of Human Poses. *Arxiv.org*, 1-9. https://arxiv.org/abs/2004.10362.
[26.] Olympics Gymnastics: Rhythmic Gymnastics - Individual All-Around-Qualification 1&2 | Tokyo 2020. *YouTube*. https://www.youtube.com/watch?v=uRzmkLF8MVI (date of access: 30.01.2024).
[27.] Olympics: FULL Rhythmic Gymnastics Individual All Around Final at Tokyo 2020. *YouTube*. URL: https://www.youtube.com/watch?v=v6ZuroWdLTs (date of access: 30.01.2024).
[28.] Albums from shooting at sports tournaments by photographer Maria Muzychenko. https://muzychenko.photos/our-services/sports-photography (date of access: 30.01.2024).
[29.] Igor Sakhatsky's portfolio. https://sakhatskyi.com/portfolio/ (date of access: 30.01.2024).
[30.] Ukrainian RG Federation: Victoria Onopriienko Ball Qual 26,200 - World Championships Kitakyushu 2021. *YouTube*. https://www.youtube.com/watch?v=IKzuWUIe8Rc (date of access: 30.01.2024).
[31.] GitHub - google/mediapipe: Cross-platform, customizable ML solutions for live and streaming media. *GitHub*. https://github.com/google/mediapipe (date of access: 30.01.2024).
[32.] Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. *Arxiv.org*, 1-16. https://arxiv.org/abs/2204.12484.
[33.] Kingma, D.P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *Arxiv.org*, 1-15. https://arxiv.org/abs/1412.6980.
Отримана в редакції 09.05.2024. Прийнята до друку 24.05.2024. Received 09 May 2024. Approved 24 May 2024. AvailableinInternet 30 July 2024