

N.V. HUNG, T.Q. LOI, N.H. BINH, N.T.T. NGA, T.T. HUONG, D.L. LUU **BUILDING AN ONLINE LEARNING MODEL THROUGH A DANCE RECOGNITION VIDEO BASED ON DEEP LEARNING**

Hung N.V., Loi T.Q., Binh N.H., Nga N.T.T., Huong T.T., Luu D.L. **Building an Online Learning Model Through a Dance Recognition Video Based on Deep Learning.**

Abstract. Jumping motion recognition via video is a significant contribution because it considerably impacts intelligent applications and will be widely adopted in life. This method can be used to train future dancers using innovative technology. Challenging poses will be repeated and improved over time, reducing the strain on the instructor when performing multiple times. Dancers can also be recreated by removing features from their images. To recognize the dancers' moves, check and correct their poses, and another important aspect is that our model can extract cognitive features for efficient evaluation and classification, and deep learning is currently one of the best ways to do this for short-form video features capabilities. In addition, evaluating the quality of the performance video, the accuracy of each dance step is a complex problem when the eyes of the judges cannot focus 100% on the dance on the stage. Moreover, dance on videos is of great interest to scientists today, as technology is increasingly developing and becoming useful to replace human beings. Based on actual conditions and needs in Vietnam. In this paper, we propose a method to replace manual evaluation, and our approach is used to evaluate dance through short videos. In addition, we conduct dance analysis through short-form videos, thereby applying techniques such as deep learning to assess and collect data from which to draw accurate conclusions. Experiments show that our assessment is relatively accurate when the accuracy and F1-score values are calculated. More than 92.38% accuracy and 91.18% F1-score, respectively. This demonstrates that our method performs well and accurately in dance evaluation analysis.

Keywords: online learning, deep learning, LSTM, GRU, RNN, Vietnam.

1. Introduction. Video technology research has been one of the hottest topics in recent years. Dance motion recognition, one of many video technologies, is critical for intelligent applications and is widely used in many aspects of life and education. Training and coaching a smart dance assistant takes significant time for the individual learner and the instructor. Some poses are difficult to perform repeatedly, impacting the psychology of students and teachers. The dancers' postures can be mapped using features extracted from their images. As a result, existing extraction techniques rely on this method to focus more on the video's spatial domain, that is, extract pixel information from video frames while ignoring changes in the motion state of the video motion in the time domain [1]. The process of transitioning from one movement to the next is critical because there are unrelated movements, making it difficult for learners to grasp the lessons.

Furthermore, all dance movements are based on human body movements [2]. The development of dance motion detection and recognition has been slow due to the complexity and breadth of dance movements. As

a result, technology assistance is required. In [3], dance motion recognition technology primarily needs to process activities that change in space and time to maintain stability when performing some dance movements, which is the primary goal of standard muscle training for general strength, strength, and bodybuilding.

Indeed, we can easily see two motion recognition methods: those based on manual features and those based on deep learning models. Although there are various aspects, deep learning is the primary technology to extract video features [4, 5]. Deep learning algorithms are used to recognize complex movements and capture them [6]. Through advanced features of many layers, deep learning has much higher generalizability than other methods [7]. Furthermore, classification [8, 9] and semantic segmentation [10, 11] outperform traditional methods. At the moment, depth neural networks are more commonly used by dividing them into many convolutional neural networks [12] that process both image data and string data. The cyclic neural network outperformed traditional classification methods and has gained widespread attention.

Many studies have shown that deep learning occurs primarily through learning and training a large amount of data [7, 13], with the goal of the parameters in the model exactly matching the rules given by the user to significantly reduce the gap between the original data, the target category, and perform the corresponding dance action recognition classification, through the use of the object's motion recognition technology. We can assess a dancer's jump position and suggest modifications by comparing the move with a standard action. Our dance motion recognition system will be significantly improved, especially today, as learning and teaching have become commonplace in the internet age as universities want to expand internships online with modern technology. Technology can evaluate teachers or judges in dance competitions or sports for performing complex moves. Figure 1 is an experimental case depicting how new research technology will replace humans. Instead of doing it manually, dance moves on stage will be marked automatically through technology. School needs are needed when teachers can only perform a few times per facility, and teachers can use this proposed method to improve assessment work.

Dance has become as famous worldwide as in India [14]. However, in Vietnam, when it comes to physical education and sports, generally referred to as physical education, most people immediately think of athletics, football, volleyball, badminton, etc. A few think about dancing because this subject has received little attention in Vietnamese universities based on the necessary conditions in Vietnam.

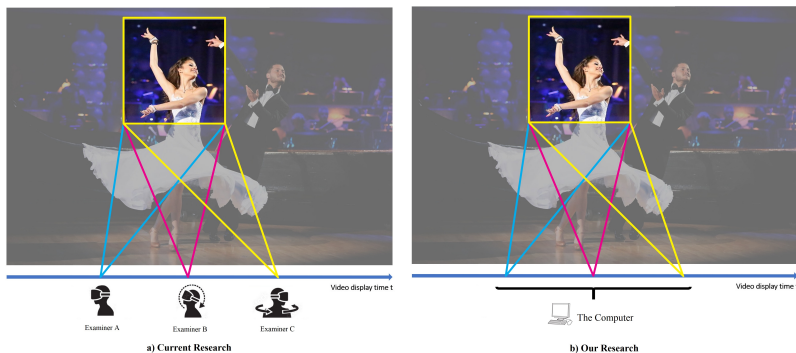


Fig. 1. a) **Current Research**: The examiner looks at the live dance, evaluates it, and gives it a live score; b) **Our Research**: We build and store the model on the computer; the computer will calculate and give the score directly

In Vietnam, processed videos are often analyzed and processed in manual evaluation, and some deep learning techniques have been widely used recently. However, contributions are still limited. Therefore, applying deep learning and video analysis techniques to classify dances from a computer perspective is necessary. Furthermore, because students at university in Vietnam need a realistic view of their experience, recorded videos help them describe the practice. On the one hand, there needs to be more experimentation in Vietnam with online video in teacher training and other professional training programs using web video applications. In our system, to support self-assessment, students must be familiar with the assessment standards and the criteria they represent. When trained, our system thus improves the quality and aids in evaluating each regular and memorable dance. Our system will help consider a teacher's comments critically and act more. On the other hand, subjective evaluations significantly impact the outcomes of dance participants. Therefore, safety and trust should be established during the assessment process, mainly when dealing with sensitive issues related to complex life. Consequently, we formulate the scientific novelty of the paper as follows: the proposed method can replace the current manual model; the proposed process will automatically evaluate and give direct results, avoiding waiting time from manual evaluation. The model is used for active online learning through trained videos. The experimental results are practical models of 3 models, RNN, LSTM, and GRU, with evaluation rates of 91.88%, 95.96%, and 96.98%, respectively.

This paper is organized as follows: Section 2 discusses the related work. Section 3 presents the concepts and features of identifying fake news on social

networks. Section 4 describes the suggested viewport estimation technique. Section 5 contains the performance assessment. Section 6 concludes with a discussion of our conclusions and open questions.

2. Related work. In [15], the algorithm provides a robust and trusted environment for video transactions by incorporating the FISCO alliance chain for secure transaction management and leveraging trusted computing techniques for video processing. It combines the benefits of blockchain technology, secure communication, and hardware-based security to enhance the video transaction process's overall security, integrity, and transparency without focusing on processing and dance analysis techniques in the dancer video.

On the other hand, the methods recently, the author will use deep learning, the procedures generally looking for ways to differentiate and predict and provide better prospects, such as [16, 17]; the author focuses on analyzing user behavior and then focuses on a series of behavioral recognition algorithms based on image and bone data; deeply examine their theories and performance, and finally offer further perspectives. However, this application still has many limitations, such as the fact that each person's behavior changes over time, and in the coming years, as people and society change, attitudes toward behavior assessment will also change. In [17], the author focuses on its construction as a dance figure classification problem using three-dimensional body joints and wearable sensors based on long short-term memory (LSTM), and it includes time- and orbit-wise structure using orbital information in timestamps and time-mask modules. However, as stated by the author, the author relies on three-dimensional body joints and sensors, which has the limitation that the research will heavily depend on the sensor. According to a recent study [18], the author recommends short-form videos on TikTok and songs and dances for users to repeat. However, the author only investigates the social spread of TikTok challenges by predicting user engagement and combining potential users and challenge representatives from previous videos to perform this user challenge prediction task. As a result, the author still needs to address the training issue and has limited his analysis to the dances in the video.

Furthermore, the advancement of Internet technology has brought opportunities and challenges for dance education in colleges and universities. Traditional dance instruction in universities still needs to meet students' needs [19]. Effective reform measures must be implemented to strengthen the reform of dance instruction in colleges and universities. The author proposes using dance teaching in universities as a research breakthrough while also researching and analyzing the methods of teaching dance in the new era in depth. Besides, researchers and technical staff have taken notice of a recent new

study that used virtual reality technology in multimedia technology to create a teaching system that simulates dance sports in [20, 21]. However, the author is primarily divided into two aspects to explore the inevitable dance process of editing the actions of virtual humanoid technology dancers. Finally, regarding the three movements included in the final sport dance teaching, the author focuses on fundamentally solving the shortcomings of contemporary dance sports teaching while offering sports-specific solutions and measures. In [21], the author also explains the related concepts of virtual reality technology dance teaching in colleges and universities. In general, the studies only focused on teaching work but did not focus on video analysis for processing, thereby making a final assessment of video analysis and processing.

Dance movement is becoming an increasingly common research topic in the broader field of human motion analysis [6]. Recent approaches primarily employ recurrent neural networks (RNNs), which have been shown to accumulate prediction errors, limiting models that synthesize short choreography to less than 100 poses. The author [22] also proposes a multimodal convolutional autoencoder capable of generating novel dance motion sequences of arbitrary length by combining 2D bone and audio information using an attention-based feature fusion mechanism. However, the author must consider the skeleton's characteristics as input in this method according to the previous model. This is only true for passive data sets, so our approach will carry out automatically when the data is transferred to the model for monitoring.

3. Theory background. In this section, we discuss some concepts, thereby discussing and recommending models for analysis and performance evaluation for future use models.

3.1. Dance identification. Counseling identification is an important area that many scientists focus on researching to serve life. Martial arts is considered the most exciting field in recent times. In [19 – 21, 23, 24], Sports dance is a sport that combines music with flexible, graceful, beautiful, and appealing movements. Currently, this sport is gradually becoming popular. It is practiced regularly, helping people exercise, improving exchanges and solidarity, and improving spiritual life quality. On the other hand, sports dance does not require high artistic and technical skills in daily practice, so in recent years, this subject has gradually become popular and developed rapidly and widely, with people of all ages, genders, and professions practicing together.

This subject not only brings health but also mental comfort to those who participate in the practice, assisting them in becoming more confident in themselves, meeting new friends, having a great time immersed in melodic music, and dispelling life's stress and depression. With the same starting point

as everyone's passion for sports, today's dance clubs are contributing to the development of the sport and physical training movement worldwide to improve everyone's health and spiritual life.

3.2. Discussion. In general, with the development of information technology, scientists have expanded and researched a lot in the evaluation, analysis, and comment models. Based on practical conditions, this paper analyzes three basic and widely used models below: RNN – Recurrent Neural Network, GRU – Gated Recurrent Unit, and LSTM – Long Short-Term Memory.

3.2.1. RNN – Recurrent Neural Network. RNN is a special machine learning model designed to process sequential data such as time series or time-dependent data [25]. An important feature of an RNN is its ability to maintain a hidden state during input data processing and reuse this information when processing subsequent components of sequential data. This allows the RNN to understand and preserve the context and temporal relationships between elements in the sequential data.

In dance recognition problems [26], RNNs can classify dances based on previous temporal information. To do this, we create an RNN with an LSTM or GRU layer to understand and model the material characteristics of the dances.

Suppose we have a data set of dance patterns recorded in a time series. Each dance pattern can be represented as feature vectors, and we need to feed them into the RNN to classify them as corresponding to different dances.

The RNN model is built with several layers, in which the first layer is an LSTM layer, then Dropout layers to avoid overfitting, and finally, a Dense layer with softmax activation function to perform classification. Once the model trains on the dataset, it recognizes new dances that have not been seen before by relying on the knowledge of dance patterns learned from the training set. While RNN has advantages in sequential data processing, it also faces some limitations, such as the problem of vanishing gradients and the difficulty of long-term information retention.

RNN is a powerful tool for dance recognition and other sequential data processing [27]. Combining LSTM layers helps the model understand and mimic complex time patterns and predict and classify dances accurately. First, an embedding layer is pushed in. Next, ignore the already stored space to remove the standalone feature that dropout keeps. The SimpleRNN class uses 50 units in Figure 2. Each type identifies with five topics. Next, the Dense class is used to classify feature features based on the outputs from the composite courses.

On the other hand, we perform classification using the softmax activation function commonly used for the output layer. Softmax ensures

that the predicted probabilities across all classes sum up to 1, providing a probability distribution over the classes. Besides, the appropriate loss function for multi-class classification with softmax activation is the categorical cross-entropy loss. This loss function measures the dissimilarity between the predicted probability distribution and the actual class labels.

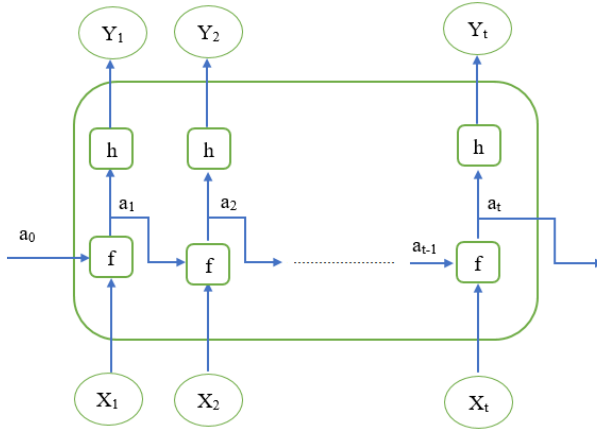


Fig. 2. Algorithm for RNN model

For the RNN model, for the input value x_t , there will be a corresponding y_t value. The process of using our RNN model is as follows:

$$a_t = f(T_a * x_t + W_a * a_{t-1}), \quad (1)$$

$$y_t = h(V_a * a_t), \quad (2)$$

where T_a , V_a , and W_a are learnable weight matrices; x is the input at t that is the one-hot vector corresponding to the size of $n - 1$; a_t is the hidden state at t that is calculated based on both the presiding state and input. y_t is the output at t , a probability vector of predicting words by learning information from all previous inputs.

3.2.2. GRU – Gated Recurrent Unit. GRU is a type of RNN developed to address limitations found in LSTM networks and enhance the performance of processing sequential data [28]. Like LSTM, the GRU maintains a hidden state to handle sequential information and learn intricate timing patterns. However, the GRU simplifies the LSTM architecture by eliminating sure gates and

replacing them with a single gate mechanism. This modification results in a more straightforward structure for the GRU to understand and work with.

The gate mechanism in the GRU allows the model to decide what information should be ignored and hidden. This helps the model focus on the important factors in the sequential data and reduces problems such as disappearing gradients during training.

In the dance recognition problem, GRU builds a dance classification model based on sequential data of dance features. Each dance pattern can be represented as feature vectors and fed into the GRU network to learn the designs and the time dependence between dances.

The GRU is built with a simple GRU layer, followed by Dropout layers, to avoid overfitting – finally, a Dense layer with a softmax activation function to perform multiclass classification. Thanks to the unique gate mechanism and the smaller number of parameters compared to LSTM, GRUs typically train faster and require fewer resources. GRU is popular for sequential data processing problems, including dance recognition.

GRU is an efficient variant of recurrent neural networks in recognizing and classifying dances based on sequential data [29]. This makes processing temporal data in dance applications easier and is highly efficient in identifying and classifying different dance patterns. In Figure 3 simulates our calculation algorithm; we recalculate the values using the formulas for the gates Reset gate (r_t), Update gate (z_t), Candidate hidden state (h'_t), and Hidden state (h_t) as follows:

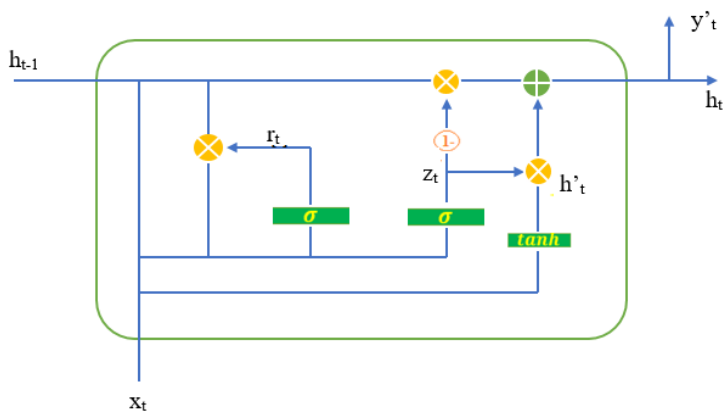


Fig. 3. Algorithm for GRU unit

$$r_t = \sigma(W_r * [h_{t-1}, x_t]), \quad (3)$$

$$z_t = \sigma(W_z * [h_{t-1}, x_t]), \quad (4)$$

$$h'_t = \tanh(W_h * [r_t * h_{t-1}, x_t]), \quad (5)$$

$$h_t = (1 - r_t) * h_{t-1} + z_t * h'_t, \quad (6)$$

where W_r , W_z , and W_h are learnable weight matrices, x_t is the input x_t time step t , h_{t-1} is the previous hidden state, and h_t is the current hidden state. Besides, y'_t is used for passing information to future time steps and computing the output.

3.2.3. LSTM – Long Short Term Memory. LSTM is an RNN architecture that processes sequential data with complex long-term and short-term characteristics [28, 30]. LSTM helps to solve the problem of gradient vanishing and the difficulty of long-term information retention in conventional RNNs. This makes LSTM a powerful tool for handling complex sequential data such as time series or natural language.

In the dance recognition problem, LSTM builds a dance classification model based on the temporal information of the dance features [14]. Each dance pattern is represented as feature vectors and fed into the LSTM network to learn the designs and the time dependence between dances. The LSTM model is built with an LSTM layer, then Dropout layers to avoid overfitting, and finally, a Dense layer with a softmax activation function to perform multiclass classification.

The ability to store long-term information in long-term memory is an important feature of LSTM. This allows the model to capture the dances' complex relationships and temporal structure. Long-term memory helps the model recognize intricate dance patterns based on information learned from the training set and then applies this knowledge to recognize new dances that have not been seen before. Although LSTM has outstanding advantages in handling complex sequential data, it also requires more computational resources than other models, such as GRU. This can sometimes make it difficult to train and deploy the model.

LSTM is a recurrent neural network architecture powerful in recognizing and classifying dances based on temporal information of dance features. This makes time data processing in dance applications efficient

and reliable and allows identifying complex and varied dance patterns. In Figure 4 simulates our calculation algorithm; we recalculate the values using the formulas for the gates Forget gate (z_t), Input gate (y_t), Output gate (n_t) as follows:

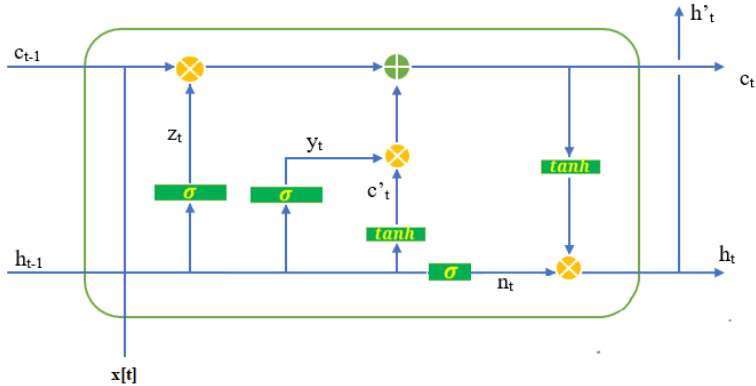


Fig. 4. Algorithm for LSTM cell

$$z_t = \sigma(U_z * x_t + W_z * h_{t-1} + b_z), \quad (7)$$

$$y_t = \sigma(U_y * x_t + W_y * h_{t-1} + b_y), \quad (8)$$

$$n_t = \sigma(U_n * x_t + W_n * h_{t-1} + b_n), \quad (9)$$

$$c'_t = \tanh(U_c * x_t + W_c * h_{t-1} + b_c), \quad (10)$$

$$c_t = z_t * c_{t-1} + y_t * c'_t, \quad (11)$$

$$h_t = h'_t = n_t * \tanh(c_t), \quad (12)$$

where U_z , U_y , U_n , U_c , W_z , W_y , W_n , and W_n are learnable weight matrices, b_z , b_y , and b_n are bias coefficients, x_t is the input x_t time step t , h_{t-1} is the previous hidden state, and h_t is the current hidden state. c_t is the forget gate that decides how much to get from the cell state first, and the input gate decides how much

to take from the input of the state and a hidden layer of the previous layer. Besides, h_t is the output gate that decides how much to take from the cell state to be the output of the hidden state. In addition, h_t is also used to calculate the output h'_t for state t .

3.2.4. Loss function. A loss function is a function that measures the difference between the model's predicted value and the ground truth. We calculate the loss function using the following formula:

$$Loss = -\frac{1}{N} \sum_{i=1}^N \log(\rho(y_i|x_i)), \quad (13)$$

where N is the number of training samples,

x_i is the input data of the i sample,

y_i : is the actual label of the i sample,

$\rho(y_i|x_i)$: is the probability that the model correctly predicts label y_i for data x_i .

4. Proposed model. In this section, we present the proposed method and the model we use. The article is analyzed and proposed with three separate models: RNN, GRU, and LSTM. Detailed descriptions are provided in the following sections.

4.1. Problem Formulation. Dancing is a type of sport that most students are interested in when they want to exercise their health and can solve boring problems in life, with the desire to change the lifestyle of society and change the daily life of Vietnamese people. Besides that, the aspect changes, and towards the competition expands beyond the country.

In addition, higher education in Vietnam aims to be more comprehensive. Therefore, this study also seeks to build a quality assessment team for university dance. Besides, analyze dances to cluster and evaluate dance classification for each object, thereby extracting information. In this section, we identified five popular dance types in Vietnam corresponding to 5 trends on social networks: "Heyhey," "Kynqucfan," "Thuyen Quyen," "Trong Hoa," and "Mua Bai Vietnam" dance. Each type of dance consists of a set of basic movements represented by corresponding landmarks on the body.

The way we do it is as follows. We used landmarks from the Mediapipe Pose to show the moves. Each landmark is represented by a vector by:

$$\Phi_i = (x_i, x_{i+1}, \dots, x_{i+n-1}), \quad (14)$$

where x_i ($1 < i < n$) are the x coordinates of the i landmark. Next, we use geometric analysis and linear algebra to determine the degree of correlation between landmarks by calculating the Euclidean distance between landmark points. Finally, Building a Dance Recognition Model After extracting landmarks and performing analysis, we make a machine learning model to recognize the dances. The model includes hidden layers to learn complex relationships between landmarks and correlations. To do this, we calculate according to the following formula:

Step 1: Calculate the expected vector of all data, with N data points represented by the column vectors x_1, x_2, \dots, x_n , then the expectation vector and the covariance matrix of the entire data are defined as:

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n. \quad (15)$$

Step 2: Subtract each data point from the expected vector of all data by:

$$\hat{x} = x_n - \bar{x}. \quad (16)$$

Step 3: Calculate the covariance matrix by:

$$S = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})(x_n - \bar{x})^T = \frac{1}{N} \hat{x} \hat{x}^T. \quad (17)$$

Thus, to calculate the eigenvalues and eigenvectors with a norm equal to 1 of this matrix, we arrange them in descending order of eigenvalues, with K eigenvectors corresponding to K largest eigenvalues to build a matrix U_K with columns forming an orthogonal system. Furthermore, these K vectors, also called principal components, include a subspace close to the distribution of the normalized initial data. Projecting the original data normalized \hat{x} down the subspace finds new data, which is the coordinates of the data points on the new space as follows:

$$Z = U_K^T * \hat{x}. \quad (18)$$

Thus, the original data can be approximated according to the new data as follows:

$$x = U_K * Z + \bar{x}. \tag{19}$$

After determining the correlation of the data, we analyze and put in the models presented above for training.

4.2. Design and problem solve. In this section, we design a model to evaluate the dance in Figure 5. We created this model to expect the model to consider each dancer automatically for each video. We pulled some images from selected videos, as shown in Figure 6.

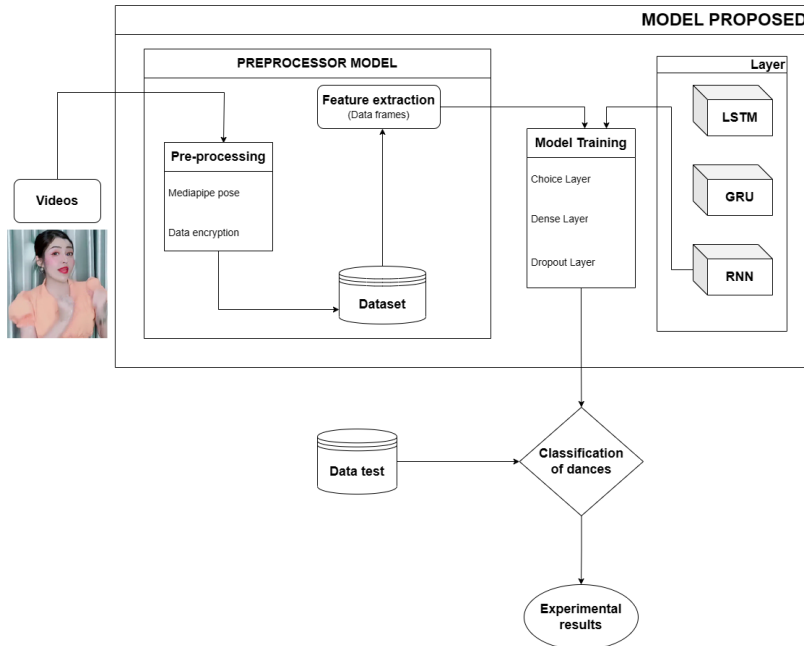
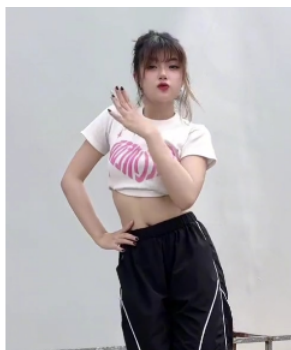


Fig. 5. Procedures for receiving and handling dancer information of our



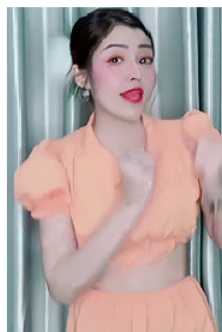
a) Video #1



b) Video #2



c) Video #3



d) Video #4

Fig. 6. Some images represent some videos extracted from the experiment

After being trained and trained, the specific model is analyzed as follows:

– **Step 1:** The first data is the videos to be injected and preprocessed through the Mediapipe. Mediapipe is a very accurate and lightweight body gesture detection library. So, we used Mediapipe to assign body gestures to identify objects and categorize them later. We can see that Figure 7 shows the process of validating dance gestures using Mediapipe. Besides, we also encode the data to know the data into the data with the middle sentence and save the data set for training in the following steps.

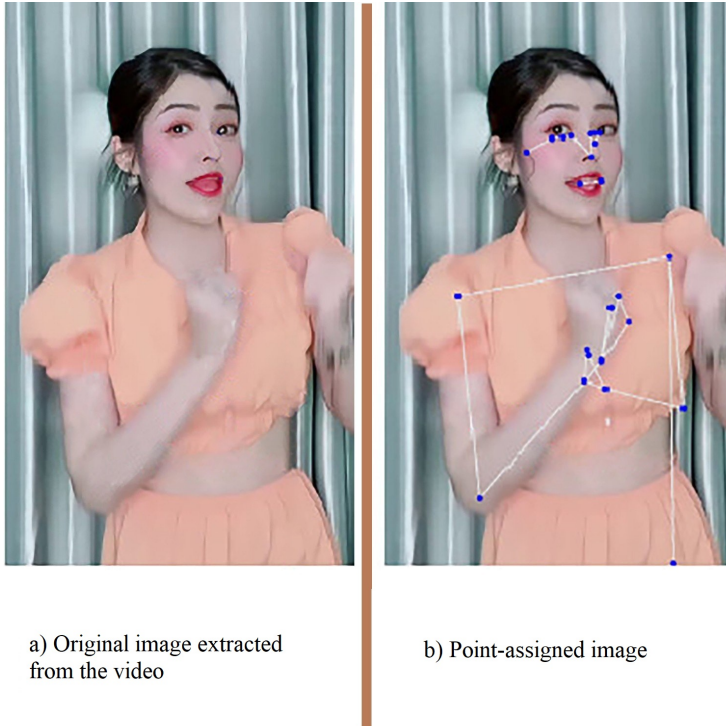


Fig. 7. a) Original image extracted from the video; b) Point-assigned image

– **Step 2:** After the data is normalized, the data will be saved to the Dataset and the area partition as shown in Figure 8; the data will be used based on the features to extract into the models. Thus, steps one and two, which are our proposed system, these steps perform data preprocessing and feature extraction to get into training models. Their recommendation system implements this process to improve the direct processing of the data before it is fed to the training models, thus reducing the training load.

Furthermore, the Extracting Feature will be the analysis content of the video will be converted to a vector to create a folder to use the features in the video. Each dance is labeled, analyzed, and given the corresponding numbers. Each dance is converted to an array with fixed-assigned natural numbers.

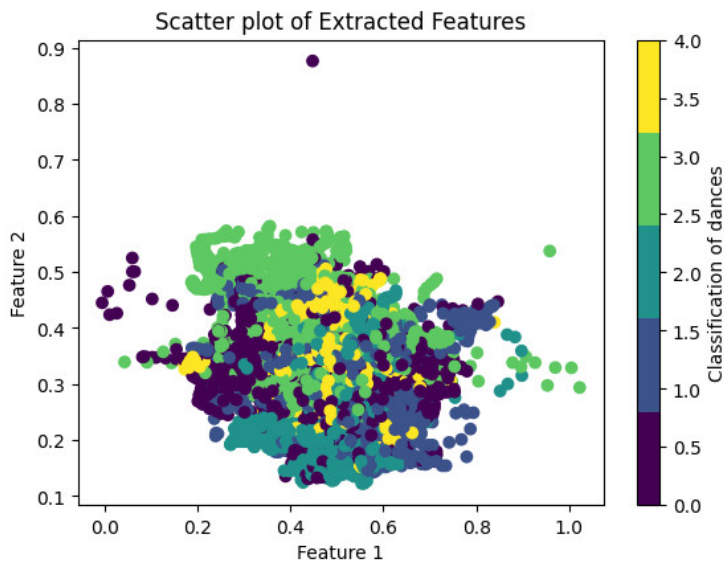


Fig. 8. Data Partition of Model

– **Step 3:** As shared in steps 1 and 2, these are two essential steps that support the implementation and training of models. In this step, after the data is preprocessed and extracted features are sent to the model, the Modes will drop unnecessary features, traversing the classes to train and storing the parts for later testing.

Moreover, the poses in training will be relevant if there is a wrong posture, and our system will find and analyze the mistakes that the dancer often encounters. In this section, the system will clearly show some typical features that dance encounters in the training data set. After being stored and analyzed, data will be saved as feature vectors. During training, the model can extract this feature to give warnings and guide the dancer through challenging poses.

5. Performance Evaluation

5.1. Experimental Settings. In this paper, we use five dances, each with ten videos. These are short-format videos, less than 20 seconds in length. The experiment was conducted on a Windows 10 computer using an ASUS Rog Strix G15 G513IC Laptop with a Ryzen 7- 4800H processor, 16 GB RAM, and an RTX 3050 4GB graphics card. Moreover, to test our model, we conduct the video extraction evaluation analysis into two parts from the dynamic; part 1 consists of 80% for the training and the remaining 20% for testing. The

language we use is the python programming language with the libraries used including:

- Mediapipe: we use it to detect body gestures;
- Pandas: We use it to build data structures;
- Opencv-python-headless: We use it for video processing;
- Tensorflow: We use to compute, train, and infer deep neural networks;
- Scikit-learn: We use it to handle classification problems.

In addition, we set up each model to 50 units, Dropout (0.2), Dense(units = 1, activation = "sigmoid"), and each model saves on a file with the extension h5.

5.2. Performance evaluation. The precision, recall, accuracy, and F1-score criteria assessed the model performance. Table 1 [6, 9] shows the parameters, which are as follows:

- TP: The model predicts 1 while actually, it is 1;
- TN: The model predicts 0 while actually, it is 0;
- FN: The model predicts 0, but the truth is 1;
- FP: The model predicts 1, but the truth is 0.

Table 1. Matrix of confusion

Labels	Positive prediction	Negative prediction
<i>Positive action</i>	False positive (FP)	True negative (TN)
<i>Negative action</i>	True positive (TP)	False negative (FN)

On the other hand, the Precision is calculated as the number of positive points divided by the total points of TP and FP as follows:

$$Pre = \frac{TP}{TP + FP}. \quad (20)$$

Recall is calculated as the number of positive points divided by the total points of TP and FN as follows:

$$Rec = \frac{TP}{TP + FN}. \quad (21)$$

Accuracy is determined by the sum of TP and TN overall points on the sum of TP, TN, FP, and FN as follows:

$$Ac = \frac{TP + TN}{TP + TN + FP + FN}. \quad (22)$$

Finally, the F1-score is calculated as follows:

$$\frac{F1}{2} = \frac{Pre}{Pre + Rec} * \frac{Rec}{Pre + Rec}. \quad (23)$$

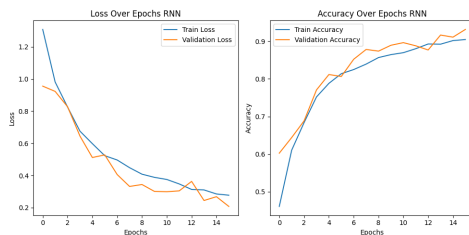
5.3. Results. Experiments show that all three methods show stable results. In five independent runs, the GRU model method gives the best results, and the RNN gives the most limited results in Table 2.

Table 2. Experimental results of 3 models RNN, GRU, LSTM

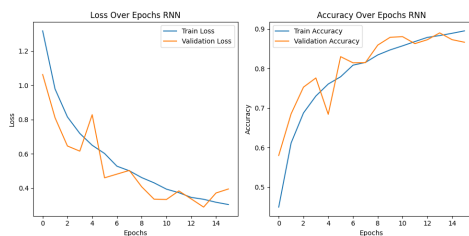
Value	Nu. of times for testing	LSTM	RNN	GRU
F1-score (%)	1	98.42	92.7	94.87
	2	95.31	86.93	96.65
	3	96.58	92.22	97.58
	4	95.19	91.03	98.24
	5	98.36	92.3	98.22
Average		96.77	91.04	97.11
Accuracy (%)	1	98.48	92.81	94.94
	2	95.41	87.43	96.84
	3	96.73	92.43	97.66
	4	95.48	91.08	98.29
	5	98.46	92.58	98.27
Average		96.91	91.27	97.20
Loss (%)	1	4.57	21.93	15.63
	2	13.69	36.19	8.83
	3	9.4	23.27	7.05
	4	14.23	25.78	4.97
	5	4.87	20.83	5.12
Average		9.35	25.60	8.32

Experimentally, we can see that using the model of GRU is the best choice when the F1-score value reaches about 97.11%, followed by LSTM and RNN, respectively. Besides, the Loss or Accuracy values also show promising results when Accuracy offers is 97.20%, while the loss value is 8.32%. For the RNN model, the loss value of 25.60% is relatively high and three times higher than that of the LSTM model at 9.35%.

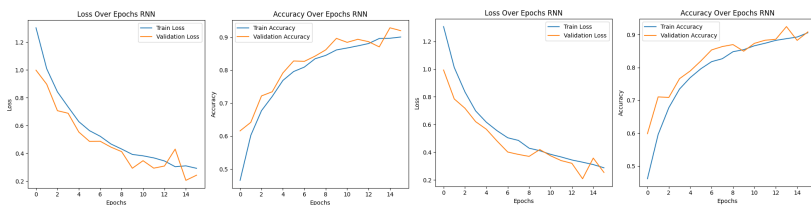
Overall, Model GRU showed better results with three models than the LSTM and RNN models. However, the experimental results are processed in short video formats. Our future direction will be to conduct more reviews with longer videos. Table 2, Figures 9, 10, and 11 show the results for each time to experiment.



a) Random experimental model RNN - time 1

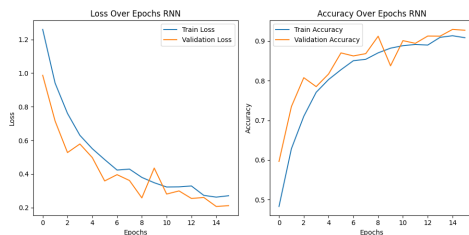


b) Random experimental model RNN - time 2



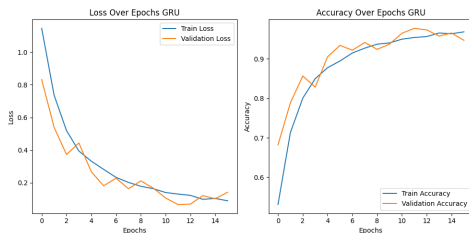
c) Random experimental model RNN - time 3

d) Random experimental model RNN - time 4

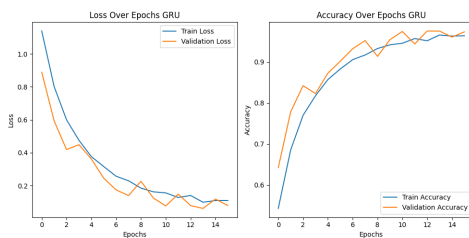


e) Random experimental model RNN - time 5

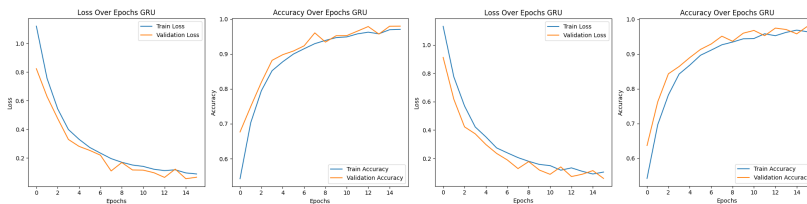
Fig. 9. Accuracy and Loss value of RNN model



a) Random experimental model GRU - time 1

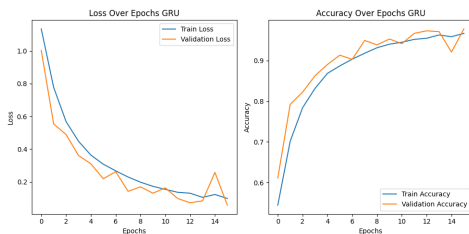


b) Random experimental model GRU - time 2



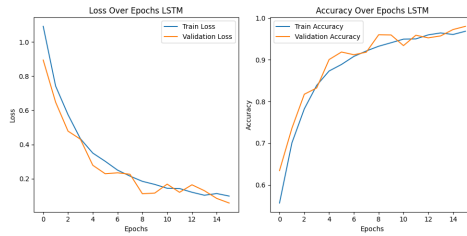
c) Random experimental model GRU - time 3

d) Random experimental model GRU - time 4

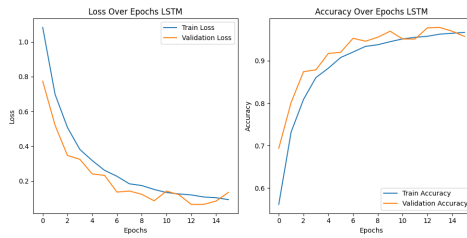


e) Random experimental model GRU - time 5

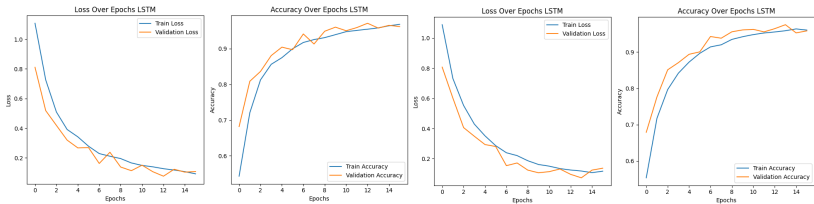
Fig. 10. Accuracy and Loss value of GRU model



a) Random experimental model LSTM - time 1

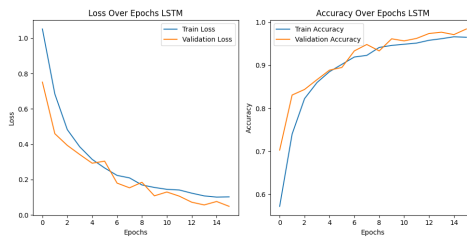


b) Random experimental model LSTM - time 2



c) Random experimental model LSTM - time 3

d) Random experimental model LSTM - time 4



e) Random experimental model LSTM - time 5

Fig. 11. Accuracy and Loss value of LSTM model

6. Conclusions. The dance classification plays the most important role in online learning through video tutorials. This paper proposes a model to classify and evaluate dance based on RNN, GRU, and LSTM models. The GRU algorithm showed better results for this study in short-form videos. However, our experiments show that our proposed models achieve a reasonable f1-score rate of over 91%. In which GRU is approximately 97%. The article's contribution is processing videos and thereby improving online learning tools for students at universities in Vietnam.

In addition, our method can evaluate learners without requiring a study coach or a panel of judges to comment. We can learn continuously and repeat many times for complex dance.

However, our method also has a few limitations, as follows:

- First: the experimental model for short-form video.
- Second: the current model has limited data, so it only works with trained videos; in case the model has not been trained, it needs to be trained, so it takes time to train.

Therefore, in the future, we plan to build a reinforcement learning model to limit the need to retrain the mode and expand the study of longer videos.

References

1. Zhai X. Dance movement recognition based on feature expression and attribute mining. *Complexity*. 2021. vol. 2021. pp. 1–12.
2. Krishna V.B. Ballroom dance movement recognition using a smart watch. arXiv preprint. 2020. arXiv:2008.10122.
3. Sun Y., Chen J. Human movement recognition in dancesport video images based on chaotic system equations. *Advances in Mathematical Physics*. 2021. vol. 2021. pp. 1–12.
4. Zhang J., Sun J., Wang J., Yue X.-G. Visual object tracking based on residual network and cascaded correlation filters. *Journal of ambient intelligence and humanized computing*. 2021. vol. 12. pp. 8427–8440.
5. Zhou Q., Wang J., Wu P., Qi Y. Application development of dance pose recognition based on embedded artificial intelligence equipment. *Journal of Physics: Conference Series*. 2021. vol. 1757(1). no. 012011.
6. Nguyen H., Dao T.N., Pham N.S., Dang T.L., Nguyen T.D., Truong T.H. An accurate viewport estimation method for 360 video streaming using deep learning. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*. 2022. vol. 9. no. 4. DOI: 10.4108/eetinis.v9i4.2218.
7. Wang S., Li J., Cao T., Wang H., Tu P., Li Y. Dance emotion recognition based on laban motion analysis using convolutional neural network and long short-term memory. *IEEE Access*. 2020. vol. 8. pp. 124928–124938.
8. Ding Y., Zhang Z., Zhao X., Cai Y., Li S., Deng B., Cai W. Self-supervised locality preserving low-pass graph convolutional embedding for large-scale hyperspectral image clustering. *IEEE Transactions on Geoscience and Remote Sensing*. 2022. vol. 60. pp. 1–16.
9. Hung N., Loi T., Huong N., Hang T.T., Huong T. Aafndl – an accurate fake information recognition model using deep learning for the vietnamese language. *Informatics and Automation*. 2023. vol. 22. no. 4. pp. 795–825.

10. Hao S., Zhou Y., Guo Y. A brief survey on semantic segmentation with deep learning. *Neurocomputing*. 2020. vol. 406. pp. 302–321.
11. Cai W., Song Y., Duan H., Xia Z., Wei Z. Multi-feature fusion-guided multiscale bidirectional attention networks for logistics pallet segmentation. *Computer Modeling in Engineering and Sciences*. 2022. vol. 131. no. 3. pp. 1539–1555.
12. Zhao M., Chang C.H., Xie W., Xie Z., Hu J. Cloud shape classification system based on multi-channel cnn and improved fdm. *IEEE Access*. 2020. vol. 8. pp. 44111–44124.
13. Bakalos N., Rallis I., Doulamis N., Doulamis A., Protopapadakis E., Voulodimos A. Choreographic pose identification using convolutional neural networks. 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games). IEEE, 2019. pp. 1–7.
14. Rani C.J., Devarakonda N. An effectual classical dance pose estimation and classification system employing convolution neural network–long shortterm memory (CNN-LSTM) network for video sequences. *Microprocessors and Microsystems*. 2022. vol. 95. no. 104651.
15. Yang Y., Yu D., Yang C. Video transaction algorithm considering fisco alliance chain and improved trusted computing. *PeerJ Computer Science*. 2021. vol. 7. no. e594.
16. Hu K., Jin J., Zheng F., Weng L., Ding Y. Overview of behavior recognition based on deep learning. *Artificial Intelligence Review*. 2023. vol. 56. no. 3. pp. 1833–1865.
17. Matsuyama H., Aoki S., Yonezawa T., Hiroi K., Kaji K., Kawaguchi N. Deep learning for ballroom dance recognition: A temporal and trajectory-aware classification model with three-dimensional pose estimation and wearable sensing. *IEEE Sensors Journal*. 2021. vol. 21. no. 22. pp. 25437–25448.
18. Ng L.H.X., Tan J.Y.H., Tan D.J.H., Lee R.K.-W. Will you dance to the challenge? predicting user participation of TikTok challenges. *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 2021. pp. 356–360.
19. He H., Luo Q. Online teaching mode of college sports dance course under the background of internet plus. *International Conference on Information Technology and Contemporary Sports (TCS)*. 2021. pp. 160–164.
20. Zhu X. Research on the application of digital media technology in sports dance teaching. *International Conference on Education, Information Management and Service Science (EIMSS)*. 2021. pp. 22–26.
21. Hu Z. Research on the application of virtual reality technology in the teaching of sports dance in colleges and universities. *2nd International Conference on Artificial Intelligence and Education (ICAIE)*. 2021. pp. 414–418.
22. Kritsis K., Gkiokas A., Pikrakis A., Katsouros V. Danceconv: Dance motion generation with convolutional networks. *IEEE Access*. 2022. vol. 10. pp. 44982–45000.
23. Chen Y., Li X. Research on the application of flipped classroom model in college sports dance teaching. *International Conference on Information Technology and Contemporary Sports (TCS)*. 2021. pp. 508–511.
24. Li Y., Xu K. Online sports dance body contour extraction and training algorithm based on dsp chip intelligent high-definition camera image processing. *7th International Conference on Communication and Electronics Systems (ICCES)*. 2022. pp. 908–911.
25. Kaur M., Mohta A. A review of deep learning with recurrent neural network. *International Conference on Smart Systems and Inventive Technology (ICSSIT)*. 2019. pp. 460–465.
26. Rallis I., Voulodimos A., Bakalos N., Protopapadakis E., Doulamis N., Doulamis A. Machine learning for intangible cultural heritage: a review of techniques on dance analysis. *Visual Computing for Cultural Heritage*. 2020. pp. 103–119.

27. Biswal A. Recurrent neural network (RNN) tutorial: Types, examples, LSTM and more. Simplilearn. Com. Retrieved. 2022. Available at: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn> (accessed: 05.10.2023).
28. Zargar S. Introduction to sequence learning models: RNN, LSTM, GRU. Department of Mechanical and Aerospace Engineering, North Carolina State University. 2021. DOI: 10.13140/RG.2.2.36370.99522.
29. Zhang W., Li H., Tang L., Gu X., Wang L., Wang L. Displacement prediction of jiuxianping landslide using gated recurrent unit (gru) networks. Acta Geotechnica. 2022. vol. 17. no. 4. pp. 1367–1382.
30. Li W., Wei Y., An D., Jiao Y., Wei Q. LSTM-TCN: Dissolved oxygen prediction in aquaculture, based on combined model of long short-term memory network and temporal convolutional network. Environmental Science and Pollution Research. 2022. vol. 29. no. 26. pp. 39545–39556.

Nguyen Viet Hung — Ph.D., Lecturer, East Asia University of Technology; Student, Faculty of telecommunications engineering, Hanoi University of Science and Technology. Research interests: multimedia communications, network security, artificial intelligence, traffic engineering in next-generation networks, QoE/QoS guarantee for network services, green networking, applications. The number of publications — 16. hungnv@eaut.edu.vn; Ky Phu - Ky Anh, Ha Tinh, Viet Nam; office phone: +84(098)911-2079.

Thang Quang Loi — Research assistant, East Asia University of Technology. Research interests: applications, networks. The number of publications — 2. thangquangloi21@gmail.com; Xuan Long, Yen Binh, Yen Bai, Viet Nam; office phone: +84(084)6203-0902.

Nguyen Hai Binh — Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: multimedia communications, database management systems, artificial intelligence, applications. The number of publications — 1. binhnh@eaut.edu.vn; Chien Thang - Thi Tran Xuan Mai, Chuong My, Hanoi, Viet Nam; office phone: +84(090)627-8669.

Nguyen Thi Thuy Nga — Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: multimedia communications, database management systems, artificial intelligence, applications. The number of publications — 1. ngantt@eaut.edu.vn; Ta Thanh Oai - Thanh Tri, Hanoi, Viet Nam; office phone: +84(097)529-4770.

Truong Thu Huong — Ph.D., Dr.Sci., Deputy head of the department, School of electrical and electronic engineering, Hanoi University of Science and Technology. Research interests: network security, artificial intelligence, traffic engineering in next-generation networks, QoE/QoS guarantee for network services, green networking, and development of the Internet of Things ecosystems, applications. The number of publications — 90. huong.truongthu@hust.edu.vn; 1, Dai Co Viet St., Hanoi, Viet Nam; office phone: +84(243)869-2242.

Duc Lich Luu — Lecturer, University of Science and Technology. Research interests: vehicle dynamics control, cooperative driving, model-based predictive control, networked control systems. The number of publications — 30. ldlich@dut.udn.vn; 54, Nguyen Luong Bang St., 550000, Da Nang, Viet Nam; office phone: +84(0236)384-2308.

Н.В. Хунг, Т.К. Лои, Н.Х. Бинь, Н.Т. Нга, Т.Т. Хьонг, Д.Л. Луу ПОСТРОЕНИЕ МОДЕЛИ ОНЛАЙН-ОБУЧЕНИЯ С ПОМОЩЬЮ ВИДЕО ПО РАСПОЗНАВАНИЮ ТАНЦЕВ, ОСНОВАННОГО НА ГЛУБОКОМ ОБУЧЕНИИ

Хунг Н.В., Лои Т.К., Бинь Н.Х., Нга Н.Т., Хьонг Т.Т., Луу Д.Л. Построение модели онлайн-обучения с помощью видео по распознаванию танцев, основанного на глубоком обучении.

Аннотация. Распознавание движений при прыжках с помощью видео является значительным вкладом, поскольку оно значительно влияет на интеллектуальные приложения и будет широко применяться в жизни. Этот метод может быть использован для обучения будущих танцоров с использованием инновационных технологий. Сложные позы будут повторяться и совершенствоваться с течением времени, что снизит нагрузку на инструктора при многократном выполнении. Танцоров также можно воссоздать, удалив элементы из их изображений. Распознавать движения танцоров, проверять и корректировать их позы, и еще одним важным аспектом является то, что наша модель может извлекать когнитивные функции для эффективной оценки и классификации, а глубокое обучение в настоящее время является одним из лучших способов сделать это для возможностей коротких видеороликов. Кроме того, при оценке качества видеозаписи выступления точность выполнения каждого танцевального шага является сложной проблемой, когда глаза судей не могут на 100% сфокусироваться на танце на сцене. Более того, танцы на видео сегодня представляют большой интерес для ученых, поскольку технологии все больше развиваются и становятся полезными для замены людей. Основываясь на реальных условиях и потребностях Вьетнама. В этой статье мы предлагаем метод, заменяющий ручную оценку, и наш подход используется для оценки танца с помощью коротких видеороликов. Кроме того, мы проводим танцевальный анализ с помощью коротких видеороликов, применяя таким образом такие методы, как глубокое обучение, для оценки и сбора данных, на основе которых можно делать точные выводы. Эксперименты показывают, что наша оценка является относительно точной, когда рассчитываются значения точности и F1-балла. Точность составляет более 92,38% и 91,18% F1-балла соответственно. Это демонстрирует, что наш метод хорошо и точно работает при анализе оценки танца.

Ключевые слова: планирование траектории, кинематическое планирование, примитивы движения, эвристический поиск.

Литература

1. Zhai X. Dance movement recognition based on feature expression and attribute mining. Complexity. 2021. vol. 2021. pp. 1–12.
2. Krishna V.B. Ballroom dance movement recognition using a smart watch. arXiv preprint. 2020. arXiv:2008.10122.
3. Sun Y., Chen J. Human movement recognition in dancesport video images based on chaotic system equations. Advances in Mathematical Physics. 2021. vol. 2021. pp. 1–12.
4. Zhang J., Sun J., Wang J., Yue X.-G. Visual object tracking based on residual network and cascaded correlation filters. Journal of ambient intelligence and humanized computing. 2021. vol. 12. pp. 8427–8440.

5. Zhou Q., Wang J., Wu P., Qi Y. Application development of dance pose recognition based on embedded artificial intelligence equipment. *Journal of Physics: Conference Series*. 2021. vol. 1757(1). no. 012011.
6. Nguyen H., Dao T.N., Pham N.S., Dang T.L., Nguyen T.D., Truong T.H. An accurate viewpoint estimation method for 360 video streaming using deep learning. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*. 2022. vol. 9. no. 4. DOI: 10.4108/eetinis.v9i4.2218.
7. Wang S., Li J., Cao T., Wang H., Tu P., Li Y. Dance emotion recognition based on laban motion analysis using convolutional neural network and long short-term memory. *IEEE Access*. 2020. vol. 8. pp. 124928–124938.
8. Ding Y., Zhang Z., Zhao X., Cai Y., Li S., Deng B., Cai W. Self-supervised locality preserving low-pass graph convolutional embedding for large-scale hyperspectral image clustering. *IEEE Transactions on Geoscience and Remote Sensing*. 2022. vol. 60. pp. 1–16.
9. Hung N., Loi T., Huong N., Hang T.T., Huong T. Aafndl – an accurate fake information recognition model using deep learning for the vietnamese language. *Informatics and Automation*. 2023. vol. 22. no. 4. pp. 795–825.
10. Hao S., Zhou Y., Guo Y. A brief survey on semantic segmentation with deep learning. *Neurocomputing*. 2020. vol. 406. pp. 302–321.
11. Cai W., Song Y., Duan H., Xia Z., Wei Z. Multi-feature fusion-guided multiscale bidirectional attention networks for logistics pallet segmentation. *Computer Modeling in Engineering and Sciences*. 2022. vol. 131. no. 3. pp. 1539–1555.
12. Zhao M., Chang C.H., Xie W., Xie Z., Hu J. Cloud shape classification system based on multi-channel cnn and improved fdm. *IEEE Access*. 2020. vol. 8. pp. 44111–44124.
13. Bakalos N., Rallis I., Doulamis N., Doulamis A., Protopapadakis E., Voulodimos A. Choreographic pose identification using convolutional neural networks. *11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. IEEE, 2019. pp. 1–7.
14. Rani C.J., Devarakonda N. An effectual classical dance pose estimation and classification system employing convolution neural network–long shortterm memory (CNN-LSTM) network for video sequences. *Microprocessors and Microsystems*. 2022. vol. 95. no. 104651.
15. Yang Y., Yu D., Yang C. Video transaction algorithm considering fisco alliance chain and improved trusted computing. *PeerJ Computer Science*. 2021. vol. 7. no. e594.
16. Hu K., Jin J., Zheng F., Weng L., Ding Y. Overview of behavior recognition based on deep learning. *Artificial Intelligence Review*. 2023. vol. 56. no. 3. pp. 1833–1865.
17. Matsuyama H., Aoki S., Yonezawa T., Hiroi K., Kaji K., Kawaguchi N. Deep learning for ballroom dance recognition: A temporal and trajectory-aware classification model with three-dimensional pose estimation and wearable sensing. *IEEE Sensors Journal*. 2021. vol. 21. no. 22. pp. 25437–25448.
18. Ng L.H.X., Tan J.Y.H., Tan D.J.H., Lee R.K.-W. Will you dance to the challenge? predicting user participation of TikTok challenges. *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 2021. pp. 356–360.
19. He H., Luo Q. Online teaching mode of college sports dance course under the background of internet plus. *International Conference on Information Technology and Contemporary Sports (TCS)*. 2021. pp. 160–164.
20. Zhu X. Research on the application of digital media technology in sports dance teaching. *International Conference on Education, Information Management and Service Science (EIMSS)*. 2021. pp. 22–26.

21. Hu Z. Research on the application of virtual reality technology in the teaching of sports dance in colleges and universities. 2nd International Conference on Artificial Intelligence and Education (ICAIE). 2021. pp. 414–418.
22. Kritsis K., Gkiokas A., Pikrakis A., Katsourou V. Danceconv: Dance motion generation with convolutional networks. IEEE Access. 2022. vol. 10. pp. 44982–45000.
23. Chen Y., Li X. Research on the application of flipped classroom model in college sports dance teaching. International Conference on Information Technology and Contemporary Sports (TCS). 2021. pp. 508–511.
24. Li Y., Xu K. Online sports dance body contour extraction and training algorithm based on dsp chip intelligent high-definition camera image processing. 7th International Conference on Communication and Electronics Systems (ICCES). 2022. pp. 908–911.
25. Kaur M., Mohta A. A review of deep learning with recurrent neural network. International Conference on Smart Systems and Inventive Technology (ICSSIT). 2019. pp. 460–465.
26. Rallis I., Voulodimos A., Bakalos N., Protopapadakis E., Doulamis N., Doulamis A. Machine learning for intangible cultural heritage: a review of techniques on dance analysis. Visual Computing for Cultural Heritage. 2020. pp. 103–119.
27. Biswal A. Recurrent neural network (RNN) tutorial: Types, examples, LSTM and more. Simplilearn. Com. Retrieved. 2022. Available at: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn> (accessed: 05.10.2023).
28. Zargar S. Introduction to sequence learning models: RNN, LSTM, GRU. Department of Mechanical and Aerospace Engineering, North Carolina State University. 2021. DOI: 10.13140/RG.2.2.36370.99522.
29. Zhang W., Li H., Tang L., Gu X., Wang L., Wang L. Displacement prediction of jiu-xian-ping landslide using gated recurrent unit (gru) networks. Acta Geotechnica. 2022. vol. 17. no. 4. pp. 1367–1382.
30. Li W., Wei Y., An D., Jiao Y., Wei Q. LSTM-TCN: Dissolved oxygen prediction in aquaculture, based on combined model of long short-term memory network and temporal convolutional network. Environmental Science and Pollution Research. 2022. vol. 29. no. 26. pp. 39545–39556.

Хунг Нгуен Вьет — Ph.D., преподаватель, Восточноазиатский технологический университет; студент, факультет телекоммуникаций, Ханойский университет науки и технологий. Область научных интересов: мультимедийные коммуникации, сетевая безопасность, искусственный интеллект, проектирование трафика в сетях следующего поколения, гарантия качества сетевых услуг, экологически чистые сети, приложения. Число научных публикаций — 16. hungnv@eaut.edu.vn; Ки Фу - Ки Ань, Хатинь, Вьетнам; р.т.: +84(098)911-2079.

Лонг Тханг Куанг — научный сотрудник, Восточноазиатский технологический университет. Область научных интересов: приложения, сети. Число научных публикаций — 2. thangquangloi21@gmail.com; Суан Лонг, Йен Бинь, Йенбай, Вьетнам; р.т.: +84(084)6203-0902.

Бинь Нгуен Хай — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, системы управления базами данных, искусственный интеллект, приложения. Число научных публикаций — 1. binhnh@eaut.edu.vn; Чьен Тханг - Тхи Тран Суан Май, Тьонгми, Ханой, Вьетнам; р.т.: +84(090)627-8669.

Нгуен Тхи Туй Нга — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, системы управления базами данных, искусственный

интеллект, приложения. Число научных публикаций — 1. ngantt@eaut.edu.vn; Та Тхань Оай, Тханьчи, Ханой, Вьетнам; р.т.: +84(097)529-4770.

Хьонг Труонг Ту — Ph.D., Dr.Sci., заместитель начальника отдела, школа электротехники и электронной инженерии, Ханойский университет науки и технологий. Область научных интересов: сетевая безопасность, искусственный интеллект, проектирование трафика в сетях следующего поколения, гарантия качества сетевых услуг, экологичные сети и развитие экосистем Интернета вещей, приложений. Число научных публикаций — 90. huong.truongthu@hust.edu.vn; улица Дай Ко Вьет, 1, Ханой, Вьетнам; р.т.: +84(243)869-2242.

Луу Дук Лич — преподаватель, Университет науки и техники. Область научных интересов: управление динамикой транспортного средства, совместное вождение, прогнозирующее управление на основе моделей, сетевые системы управления. Число научных публикаций — 30. ldlich@dut.udn.vn; улица Нгуен Луонг Банга, 54, 550000, Дананг, Вьетнам; р.т.: +84(0236)384-2308.