DANCE-MOVE GENERATION METHOD, COMPUTER DEVICE, AND STORAGE MEDIUM

CROSS-REFERENCE TO RELATED APPLICATION(S)

[0001] This application is the National Stage of International Application No. PCT/CN2023/090889, filed on April 26, 2023, which claims priority to Chinese Patent Application No. 202211441749.0, filed on November 17, 2022, both of which are incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] This disclosure relates to the field of artificial intelligence (AI), and in particular, to dance-move generation method, a computer device, and a storage medium.

BACKGROUND

[0003] Dancing is a high-level artwork consisting of an audio and dance moves that can deliver sentiment, and how an audio and dance moves are to be matched becomes the focus and difficulty for choreography. For professional dance actors, dance moves can be choreographed according to their own understanding of the emotion of the audio, however, such a way of manual choreography depends on professional dance actors, and ordinary dancers or users are unable to complete choreography on their own. Therefore, it is very important to efficiently implement automatic choreography of dance moves.

SUMMARY

[0004] Provided are a dance-move generation method, a computer device, and a storage medium, which can automatically generate dance moves, and improve the quality of dance move.

[0005] According to a first aspect, embodiments of the present disclosure provide a dance-move generation method. The method includes the following. An audio to be choreographed is obtained, and multiple audio segments are extracted from the audio to be choreographed. The multiple audio segments are input into a pre-trained encoding model to obtain a first move

feature of each audio segment in the multiple audio segments, where the encoding model is obtained by training sample audios and sample dance moves corresponding to the sample audios. According to the first move feature of each audio segment, a second move feature similar to the first move feature of each audio segment is determined from move features of multiple categories of dance moves pre-stored in a move library. The second move feature corresponding to each audio segment is input into a pre-trained decoding model to obtain third move features, and dance moves of the audio to be choreographed are determined according to the third move features. The decoding model is obtained by training sample move features corresponding to the sample audios and the sample dance moves corresponding to the sample audios. The sample move features corresponding to the sample audios are obtained by encoding the sample audios by using the pre-trained encoding model. The third move features are used to indicate move features of all of the multiple audio segments.

[0006] According to a second aspect, embodiments of the present disclosure provide a dance-move generation apparatus. The dance-move generation apparatus includes an obtaining unit, an encoding unit, a determining unit, and a decoding unit. The obtaining unit is configured to obtain an audio to be choreographed, and extract multiple audio segments from the audio to be choreographed. The encoding unit is configured to input the multiple audio segments into a pre-trained encoding model to obtain a first move feature of each audio segment in the multiple audio segments. The encoding model is obtained by training sample audios and sample dance moves corresponding to the sample audios. The determining unit is configured to determine, according to the first move feature of each audio segment from move features of multiple categories of dance moves pre-stored in a move library, a second move feature similar to the first move feature of each audio segment. The decoding unit is configured to input the second move feature corresponding to each audio segment into a pre-trained decoding model to obtain third move features, and determine dance moves of the audio to be choreographed according to the third move features. The decoding model is obtained by training sample move features corresponding to the sample audios and the sample dance moves corresponding to the sample audios. The sample move features corresponding to the sample audios are obtained by encoding the sample audios by using the pre-trained encoding model. The third move features are used to indicate move features of all of the multiple audio segments.

[0007] According to a third aspect, embodiments of the present disclosure provide a

computer device. The computer device includes a processor and a memory, and the processor is configured to execute the following operations. An audio to be choreographed is obtained, and multiple audio segments are extracted from the audio to be choreographed. The multiple audio segments are input into a pre-trained encoding model to obtain a first move feature of each audio segment in the multiple audio segments, where the encoding model is obtained by training sample audios and sample dance moves corresponding to the sample audios. According to the first move feature of each audio segment, a second move feature similar to the first move feature of each audio segment is determined from move features of multiple categories of dance moves pre-stored in a move library. The second move feature corresponding to each audio segment is input into a pre-trained decoding model to obtain third move features, and dance moves of the audio to be choreographed are determined according to the third move features. The decoding model is obtained by training sample move features corresponding to the sample audios and the sample dance moves corresponding to the sample audios. The sample move features corresponding to the sample audios are obtained by encoding the sample audios by using the pre-trained encoding model. The third move features are used to indicate move features of all of the multiple audio segments.

[0008] In a fourth aspect, embodiments of the present disclosure further provide a computer-readable storage medium. The computer-readable storage medium stores a program instruction which, when executed, is configured to implement the method according to the first aspect.

[0009] In embodiments of the present disclosure, an audio to be choreographed is obtained, and multiple audio segments are extracted from the audio to be choreographed. The multiple audio segments are input into a pre-trained encoding model to obtain a first move feature of each audio segment in the multiple audio segments, where the encoding model is obtained by training sample audios and sample dance moves corresponding to the sample audios. According to the first move feature of each audio segment, a second move feature similar to the first move feature of each audio segment is determined from move features of multiple categories of dance moves pre-stored in a move library. The second move feature corresponding to each audio segment is input into a pre-trained decoding model to obtain third move features, and dance moves of the audio to be choreographed are determined according to the third move features, The decoding model is obtained by training sample move features corresponding to the sample audios and the sample dance moves corresponding to the sample audios, the sample move features

corresponding to the sample audios are obtained by encoding the sample audios by using the pre-trained encoding model, and the third move features are used to indicate move features of all of the multiple audio segments. By means of this method, dance moves can be generated automatically, automation and intelligence requirements of a user for generating dance moves are satisfied, and the quality of the dance moves is improved.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] To describe the technical solutions in embodiments of the present disclosure or in the related art more clearly, the following briefly introduces accompanying drawings required for describing embodiments or the related art. Apparently, the accompanying drawings in the following description show merely some embodiments of the present disclosure, other drawings may also be obtained according to these drawings without creative efforts.

[0011] FIG. 1 is a schematic flowchart of a dance-move generation method provided in an embodiment of the present disclosure.

[0012] FIG. 2 is a schematic diagram of multi-frame dance moves.

[0013] FIG. 3 is a schematic diagram of key points of a human body.

[0014] FIG. 4 is a schematic flowchart of another dance-move generation method provided in an embodiment of the present disclosure.

[0015] FIG. 5 is a schematic flowchart of another dance-move generation method provided in an embodiment of the present disclosure.

[0016] FIG. 6 is a schematic flowchart of another dance-move generation method provided in an embodiment of the present disclosure.

[0017] FIG. 7 is a schematic structural diagram of a dance-move generation apparatus provided in an embodiment of the present disclosure.

[0018] FIG 8 is a schematic structural diagram of a computer device provided in an embodiment of the present disclosure.

DETAILED DESCRIPTION

[0019] The following clearly describes technical solutions in embodiments of the present disclosure with reference to accompanying drawings in the embodiments of the present

disclosure. Apparently, the described embodiments are merely a part rather than all of the embodiments of the present disclosure. All other embodiments obtained by a person of ordinary skill in the art based on the embodiments of the present disclosure without creative efforts shall belong to the protection scope of the present disclosure.

[0020] Artificial intelligence (AI) is a theory, method, technology, and application system that uses a digital computer or a machine controlled by the digital computer to simulate, extend, and expand human intelligence, sense environment, obtain knowledge, and use the knowledge to obtain an optimal result. In other words, artificial intelligence is a comprehensive technology of computer science, which attempts to understand the essence of intelligence and produce a new intelligent machine capable of reacting in a manner similar to human intelligence. Artificial intelligence aims to research on design principles and implementation methods of various intelligent machines to make the machines have perception, reasoning, and decision-making functions.

[0021] The artificial intelligence technology is a comprehensive discipline, relates to a wide field, and involves both hardware and software technologies. Artificial intelligence basic technologies generally include technologies such as sensors, dedicated artificial intelligence chips, cloud computing, distributed storage, large image processing technologies, operation/interaction systems, electromechanical integration, etc. The artificial intelligence software technologies mainly include a computer vision technology, a speech processing technology, a natural language processing technology, a machine learning/deep learning and other several directions.

[0022] Key technologies of the speech technology include an automatic speech recognition (ASR) technology, a speech synthesis (TTS) technology, and a voiceprint recognition technology. Enabling a computer to hear, see, talk, and feel, is a development direction of future human-machine interaction, and voice is one of the best human-machine interaction manners in the future.

[0023] Natural language processing (NLP) is an important direction in the field of computer science and artificial intelligence. It studies various theories and methods that enable effective communication between humans and computers using natural language. Natural language processing is a science that integrates linguistics, computer science, and mathematics. Therefore, research in this field relates to natural language, that is, a language used by people in daily life,

and therefore, it is closely related to research of linguistics. Natural language processing technologies generally include technologies such as text processing, semantic understanding, machine translation, robot question answering, and knowledge mapping.

[0024] Machine learning (ML) is cross-disciplinary, and relates to multiple disciplines such as probability theory, statistics, approximation theory, convex analysis, and algorithm complexity theory. Machine learning focuses on research on how a computer simulates or implements a learning behavior of a human, so as to obtain new knowledge or skills and reorganize an existing knowledge structure to continuously improve the performance of the computer. Machine learning is a core of artificial intelligence that enables a computer to have a basic approach of intelligence, and the application of machine learning is applied to various fields of artificial intelligence. Machine learning/deep learning typically includes techniques such as artificial neural networks, confidence networks, reinforcement learning, migration learning, inductive learning, teaching learning, and the like.

[0025] Based on the technologies such as machine learning mentioned in the described artificial intelligence technology, the present disclosure proposes a dance-move generation scheme. By encoding, with a pre-trained encoding model, an audio to be choreographed to obtain a first move feature of the audio to be choreographed, determining a second move feature similar to the first move feature from a move library obtained through model learning, and further decoding the second move feature with a decoding model, dance moves of the audio to be choreographed are generated. In this way, it is possible to achieve an automatic generation of dance movements and improve the quality of the generated dance movements.

[0026] The dance-move generation method provided in embodiments of the present disclosure can be configured as a dance-move generation apparatus. The dance-move generation apparatus can be provided in a computer device. In some embodiments, the computer device can include, but is not limited to, a smart phone, a tablet computer, a notebook computer, a desktop computer, a vehicle-mounted smart terminal, a smart watch, and other smart terminal devices.

[0027] In some embodiments, the dance-move generation method provided in embodiments of the present disclosure can be applicable in a choreography scenario, for example, generating dance moves matching the audio to be choreographed according to the audio to be choreographed, etc. Of course, the above application scenario is merely illustrative, and in other embodiments, the dance-move generation of embodiments of the present disclosure can be

applied to any scenario associated with the dance-move generation.

[0028] The dance-move generation method according to embodiments of the present disclosure will be schematically described below with reference to accompanying drawings.

[0029] Reference is made to FIG. 1 and FIG. 1 is a schematic flowchart of a dance-move generation method provided in an embodiment of the present disclosure. The dance-move generation method in the embodiment of the present disclosure can be executed by a dance-move generation apparatus. The dance-move generation apparatus is provided in a terminal or a computer device, and the specific explanation of the terminal or the computer device is as described above. Specifically, the method in the embodiment of the present disclosure includes the following operations.

[0030] At S101, an audio to be choreographed is obtained, and multiple audio segments are extracted from the audio to be choreographed.

[0031] In the embodiment of the present disclosure, when extracting the multiple audio segments from the audio to be choreographed, the computer device can obtain beat information of the audio to be choreographed, and extract the multiple audio segments from the audio to be choreographed according to the beat information.

[0032] In an embodiment, when the computer device extracts the multiple audio segments from the audio to be choreographed according to the beat information, the computer device may extract the multiple audio segments from the audio to be choreographed according to a specified beat. For example, the specified beat may be a half a beat, and the computer device may extract the multiple audio segments corresponding to the half a beat of each beat in the audio to be choreographed. In other embodiments, the specified beat may be another beat, which is not specifically limited in the present disclosure.

[0033] In the present disclosure, by extracting the multiple audio segments from the audio to be choreographed so as to generate matched dance moves for each audio segment, the quality of dance moves generated for the audio to be choreographed can be improved.

[0034] At S102, the multiple audio segments are input into a pre-trained encoding model to obtain a first move feature of each audio segment in the multiple audio segments.

[0035] In the embodiment of the present disclosure, the computer device can input the multiple audio segments into the pre-trained encoding model to obtain the first move feature of each audio segment in the multiple audio segments. The encoding model is obtained by training

sample audios and sample dance moves corresponding to the sample audios. In some embodiments, a data form of the first move feature may include, but is not limited to, a matrix, a polygonal mesh data (MMD), a three-dimensional generic model format (e.g., film box (FBX)), and so on.

[0036] In an embodiment, when the computer device inputs the multiple audio segments into the pre-trained encoding model and obtains the first move feature of each audio segment in the multiple audio segments, the computer device can input each audio segment in the multiple audio segments into the pre-trained encoding model, obtain move feature vectors corresponding to each audio segment, and determine the first move feature according to the move feature vectors corresponding to each audio segment.

[0037] Further, when the data form of the first move feature is a matrix, the computer device determines the first move feature according to the move feature vectors corresponding to each audio segment as follows. The computer device can take a move feature vector corresponding to each move in each audio segment as a row vector, so that the first move feature is formed according to the move feature vector corresponding to each move in each audio segment in the multiple audio segments. One audio segment may include one or more moves, and each move corresponds to one move feature vectors. For example, the first move feature may be a t*512 matrix, with t being the number of audio segments.

[0038] When the data form of the first action characteristic is a matrix, the number of columns of the matrix may be determined according to the number of moves in the move library. For example, if the number of moves in the move library is 64, a first move feature matrix can be determined as a t*64 matrix.

[0039] At S103, according to the first move feature of each audio segment, a second move feature similar to the first move feature of each audio segment is determined from move features of multiple categories of dance moves pre-stored in the move library.

[0040] In the embodiment of the present disclosure, the move library pre-stores move features of dance moves of multiple dance categories, and the data form of the move feature includes, but is not limited to, a matrix. For example, the move library may be a T*24*3 matrix, where T is used to indicate T frames of dance moves, 24 is used to indicate 24 key points of a dance move in a human body, and 3 is used to indicate a coordinate position of each dance move having three dimensions. FIG. 2 is a schematic diagram of multi-frame dance moves. As

illustrated in FIG. 2, each human body moves corresponds to one frame of dance move, and FIG. 2 illustrates multi-frame dance moves. As illustrated in FIG. 3, FIG. 3 is a schematic diagram of key points of a human body. Numbers 0-23 marked in FIG. 3 are used to indicate multiple key points of a human body, and a dance move is determined according to the position of each key point, and key points at different positions can constitute multiple dance moves.

[0041] In an embodiment, when determining, according to the first move feature, the second move feature similar to the first move feature from move features of multiple categories of dance moves pre-stored in the move library, the computer device may obtain multiple first move feature vectors included in the first move feature of each audio segment, where each first move feature vector is used to indicate an move; according to each first move feature vector of each audio segment, determine a second move feature vector corresponding to each of the multiple first move feature vectors from move features of multiple categories of dance moves pre-stored in the move library; and determine the second move feature of each audio segment according to each second move feature vector of each audio segment.

[0042] In an embodiment, the computer device determines, according to each first move feature vector of each audio segment and from the move features of the multiple categories of dance moves pre-stored in the move library, the second move feature vector corresponding to the first move feature vector as follows. A distance between each first move feature vector of each audio segment and each move feature vector in the move library is obtained, where each move feature vector in the move library is used to indicate a pre-stored dance move. A move feature vector with the shortest distance to each first move feature vector of each audio segment is obtained from the move library as the second move feature vector of each audio segment.

[0043] In an embodiment, when obtaining the distance between each first move feature vector of each audio segment and each move feature vector in the move library, the computer device may calculate the distance between each first move feature vector of each audio segment and each move feature vector in the move library by using a Euclidean algorithm.

[0044] In an embodiment, the first move feature is specifically a first move feature matrix, and the computer device obtains, for each audio segment, various row vectors in the first move feature matrix corresponding to the audio segment to obtain multiple first move feature vectors corresponding to the audio segment, where each row vector is used to indicate a move. For each first move feature vector of each audio segment, the computer device determines a second move

feature vector corresponding to the first move feature vector from the move features of multiple categories of dance moves pre-stored in the move library, so as to obtain multiple second move feature vectors corresponding to each audio segment. For each audio segment, the computer device combines multiple second move feature vectors corresponding to the audio segment to obtain a second move feature matrix corresponding to the audio segment, where the second move feature matrix is used to indicate the second move feature.

[0045] In the present disclosure, the second move feature corresponding to the first move feature is determined from the move library, which is helpful to more effectively decode the second move feature of each audio segment according to the decoding model subsequently, so as to obtain third move features, and further determine dance moves of the audio to be choreographed according to the third move features.

[0046] At S104, the second move feature corresponding to each audio segment is input into a pre-trained decoding model to obtain third move features, and dance moves of audio to be choreographed are determined according to the third move features.

[0047] In embodiments of the present disclosure, the computer device inputs the second move feature corresponding to each audio segment into the pre-trained decoding model to obtain the third move features, and determine the dance moves of the audio to be choreographed according to the third move features. The decoding model is obtained by training an initial decoding model according to sample move features corresponding to the sample audios and sample dance moves corresponding to the sample audios. The sample move features corresponding to the sample audios are obtained by using the pre-trained encoding model to encode the sample audios. The third move features are used to indicate move features of all of the multiple audio segments. The third move features include the number of key points of a human body and a position of each key point in all audio segments. In some embodiments, the third move features include, but are not limited to, a third move features matrix. For example, the third move features may be a matrix of T*24*3, where T is used to indicate the number of audio segments, 24 represents key points of a human body, and 3 is used to indicate a three-dimensional coordinate of each key point.

[0048] When determining the dance moves of the audio to be choreographed according to the third move features, the computer device can determine the number of key points of a human body and a position of each key point in all the audio segments according to the third move

features; and determine the dance moves of the audio to be choreographed according to the number of key points of a human body and the position of each key point in all the audio segments. For example, assuming that the third move features are a T*24*3 matrix, the computer device may determine the number of key points of the human body and the position of each key point in all the audio segments according to the T*24*3 matrix, determine human moves in all audio segments according to the quantity of key points of a human body in all the audio segments and the position of each key point, and further determine the human moves of all the audio segments as dance moves of the audio to be choreographed.

[0049] In embodiments of the present disclosure, an audio segment of the audio to be choreographed is encoded by using the pre-trained encoding model to obtain a corresponding first move feature, and a second move feature corresponding to the first move feature is determined from move features of multiple categories of dance moves pre-stored in the move library, facilitating generating more accurate dance moves of high quality for the audio to be choreographed by decoding the second move feature using the pre-trained decoding model.

[0050] Reference is made to FIG. 4, which is a schematic flowchart of another dance-move generation method provided in an embodiment of the present disclosure. The dance-move generation method according to the embodiment of the present disclosure can be executed by a dance-move generation apparatus, where the dance-move generation apparatus is provided in a terminal or a computer device, and the specific explanation of the terminal or the computer device is as described above. Specifically, the embodiment of the present disclosure mainly describes a training process of an encoding model, and specifically includes the following.

[0051] At S401, an audio to be choreographed is obtained, and multiple audio segments are extracted from the audio to be choreographed.

[0052] At S402, a sample data set is obtained, where the sample data set includes multiple pieces of sample dance data, and each piece of sample dance data includes a sample audio and sample dance moves.

[0053] At S403, an initial encoding model is trained according to the sample audio and sample dance moves of each piece of sample dance data to obtain a pre-trained encoding model.

[0054] In an embodiment, the computer device trains the initial encoding model according to the sample audio and the sample dance moves of each piece of sample dance data to obtain the pre-trained encoding model as follows. A sample move feature can be extracted out from the

sample dance moves of each piece of sample dance data; and the sample move feature and the sample audio are input into an initial encoding model to obtain a pre-trained encoding model. In some embodiments, the data form of the sample move feature includes, but is not limited to, a matrix.

[0055] In an embodiment, when the computer device extracts out the sample move feature from the sample dance moves of each piece of sample dance data, the computer device can extract the number of key points and positions of the key points of a human body corresponding to each sample dance move of each piece of sample dance data, where the positions of the key points include coordinates of each key point. The computer device inputs the number of key points and the positions of the key points of the human body corresponding to each sample dance move into an initial encoding model to obtain a sample move feature by extracting. In some embodiments, the key points of the human body for each sample dance move may include 24 key points of the human body. In some embodiments, the positions of the key points for each sample dance move may include three-dimensional coordinate data for each sample dance move.

[0056] In an embodiment, the computer device inputs the sample move feature and the sample audio into an initial encoding model to obtain the pre-trained encoding model as follows. The computer device inputs the sample move feature and the sample audio into the initial encoding model to obtain a first sample move feature of the sample audio, and determines, from move features of multiple categories of dance moves pre-stored in the move library and according to the first sample move feature, a second sample move feature similar to the first sample move feature. The computer device inputs the second sample move feature into an initial decoding model to obtain a third sample move feature. The computer device adjusts a model parameter of the initial decoding model according to the third sample move feature, and inputs the second sample move feature into an adjusted decoding model for training to obtain the pre-trained decoding model. In some embodiments, the third sample move feature is used to indicate the number of key points of a human body and a position of each key point in the sample audio.

[0057] In an embodiment, the computer device adjusts the model parameter of the initial decoding model according to the third sample move feature, and inputs the second sample move feature into the adjusted decoding model for training to obtain the pre-trained decoding model as follows. The computer device determines dance moves of the sample audio according to the third sample move feature. The computer device compares the determined dance moves of the sample

audio with the sample dance moves of the sample audio, and adjusts the model parameter of the initial decoding model according to a comparison result. The computer device inputs the second sample move feature into the decoding model with the model parameter being adjusted for retraining of the decoding model, so as to obtain the pre-trained decoding model.

[0058] The computer device determines the dance moves of the sample audio according to the third sample move feature as follows. The computer device determines the number of key points of a human body and a position of each key point in the sample audio according to the third sample move feature; and determines the dance moves of the sample audio according to the number of key points of the human body and the position of each key point in the sample audio.

[0059] Further, when comparing the determined dance moves of the sample audio with the sample dance moves of the sample audio, the computer device can compare a move feature matrix of the dance moves of the sample audio with a move feature of the sample dance moves.

[0060] Further, the computer device adjusts the model parameter of the initial encoding model according to the comparison result as follows. A vector distance can be calculated for each vector of the move feature of the dance moves of the sample audio and each vector of the move feature of the sample dance moves. When a vector distance between a vector of the move feature of dance moves of the sample audio and a vector of the move feature of the sample dance moves is greater than a first distance threshold, the model parameter of the initial encoding model is adjusted according to the vector distance between the vector of the move feature of dance moves of the sample audio and the vector in the move feature matrix of the sample dance moves.

[0061] In the present disclosure, a mapping relationship between various sample dance moves and a sample move feature can be obtained by using the initial encoding model, which helps to obtain the pre-trained encoding model through training of the initial encoding model according to a sample move feature and a sample audio, and obtain a mapping relationship among a sample move feature, a sample audio, and a move feature matrix of dance moves of the sample audio.

[0062] At S404, the multiple audio segments are input into the pre-trained encoding model to obtain a first move feature of each audio segment in the multiple audio segments.

[0063] At S405, a second move feature which is similar to the first move feature of each audio segment is determined according to the first move feature of each audio segment from move features of multiple categories of dance moves pre-stored in the move library.

[0064] At S406, the second move feature of each audio segment is input into a pre-trained decoding model to obtain third move features, and dance moves of the audio to be choreographed are determined according to the third move features.

[0065] In embodiments of the present disclosure, a mapping relationship between various sample dance moves and a sample move feature is obtained by using the initial encoding model, which helps to obtain a pre-trained encoding model through training of the initial encoding model according to the sample move feature and the sample audio, and obtain a mapping relationship among a sample move feature, a sample audio, and a first sample move feature. In this way, the first move feature of the audio to be choreographed is generated by using the pre-trained encoding model during the test.

[0066] Reference is made to FIG. 5, which is a schematic flowchart of still another dance-move generation method provided in an embodiment of the present disclosure. The dance-move generation method according to the embodiment of the present disclosure can be executed by a dance-move generation apparatus, where the dance-move generation apparatus is provided in a terminal or a computer device, and the specific explanation of the terminal or the computer device is as described above. Specifically, the embodiment of the present disclosure mainly describes a training process of a decoding model, and specifically includes the following steps.

[0067] At S501, an audio to be choreographed is obtained, and multiple audio segments are extracted from the audio to be choreographed.

[0068] At S502, the multiple audio segments are input into a pre-trained encoding model to obtain a first move feature of each audio segment in the multiple audio segments.

[0069] At S503, a second move feature which is similar to the first move feature of each audio segment is determined according to the first move feature of each audio segment from move features of multiple categories of dance moves pre-stored in the move library.

[0070] At S504, a sample audio is input into the pre-trained encoding model to obtain a first sample move feature corresponding to the sample audio.

[0071] At S505, an initial decoding model is trained according to the first sample move feature to obtain a pre-trained decoding model.

[0072] In an embodiment, the computer device trains the initial decoding model according to the first sample move feature to obtain the pre-trained decoding model as follows. A second sample move feature similar to the first sample move feature can be determined from move

features of multiple categories of dance moves pre-stored in the move library according to the first sample move feature. The second sample move feature is input into the initial decoding model for training to obtain the pre-trained decoding model.

[0073] In an embodiment, the computer device inputs the second sample move feature into the initial decoding model for training to obtain the pre-trained decoding model as follows. The computer device inputs the second sample move feature into the initial decoding model to obtain a third sample move feature. The computer device adjusts a model parameter of the initial decoding model according to the third sample move feature. The computer device inputs the second sample move feature into the adjusted decoding model for training to obtain the pre-trained decoding model.

[0074] In an embodiment, the computer device determines dance moves of the sample audio according to the third sample move feature, compares the determined dance moves of the sample audio with the sample dance moves of the sample audio, and adjusts a model parameter of the initial decoding model according to a comparison result. The computer device inputs the second sample move feature into the decoding model with the model parameter being adjusted for retraining of decoding model, so as to obtain the pre-trained decoding model.

[0075] When determining the dance moves of the sample audio according to the third sample move feature, the computer device can determine the number of key points of a human body and the position of each key point in the sample audio according to the third sample move feature; and determine the dance moves of the sample audio according to the number of key points of the human body and the position of each key point in the sample audio.

[0076] In an embodiment, when the computer device adjusts the model parameter according to the third sample move feature, the computer device may adjust the model parameter of the initial decoding model according to the third sample move feature and the sample move feature.

[0077] Further, when the computer device adjusts the model parameter of the initial decoding model according to the third sample move feature and the sample move feature, the computer device may adjust the model parameter of the initial decoding model according to a vector distance between a vector of the third sample move feature matrix and a vector in the sample move feature matrix.

[0078] Further, when the vector distance between the vector of the third sample move feature and the vector in the sample move feature is greater than a second distance threshold, the

computer device may adjust the model parameter of the initial decoding model according to the vector distance between the vector of the third sample move feature and the vector in the sample move feature.

[0079] At S506, a second move feature corresponding to each audio segment is input into the pre-trained decoding model to obtain third move features, and dance moves of the audio to be choreographed are determined according to the third move features.

[0080] In embodiments of the present disclosure, a first sample move feature of a sample audio is generated by using a pre-trained encoding model, and an initial decoding model is trained according to the first sample move feature, so as to obtain a pre-trained decoding model, which helps to generate third move features of an audio to be choreographed more accurately during testing, thereby generating dance moves with more accuracy and higher quality.

[0081] Reference is made to FIG. 6, which is a schematic flowchart of another dance-move generation method provided in an embodiment of the present disclosure. An audio 61 to be choreographed is obtained. Multiple audio segments 62 are extracted from the audio to be choreographed. The multiple audio segments are input into a pre-trained encoding model to obtain a first move feature 63 of each audio segment in the multiple audio segments. A second move feature similar to the first move feature of each audio segment is determined from move features of multiple categories of dance moves pre-stored in a move library, and the second move feature of each audio segment is input into a pre-trained decoding model to obtain third move features 64 of all of the multiple audio segments, thereby determining dance moves of the audio to be choreographed according to the third move features.

[0082] Reference is made to FIG. 7, which is a schematic structural diagram of a dance-move generation apparatus provided in an embodiment of the present disclosure. The dance-move generation apparatus is provided in a computer device. The dance-move generation apparatus includes an obtaining unit 701, an encoding unit 702, a determining unit 703, and a decoding unit 704.

[0083] The obtaining unit 701 is configured to obtain an audio to be choreographed, and extract multiple audio segments from the audio to be choreographed.

[0084] The encoding unit 702 is configured to input the multiple audio segments into a pre-trained encoding model to obtain a first move feature of each audio segment in the multiple audio segments, where the encoding model is obtained by training sample audios and sample

dance moves corresponding to the sample audios.

[0085] The determining unit 703 is configured to determine, according to the first move feature of each audio segment from move features of multiple categories of dance moves pre-stored in a move library, a second move feature similar to the first move feature of each audio segment.

[0086] The decoding unit 704 is configured to input the second move feature corresponding to each audio segment into a pre-trained decoding model to obtain third move features, and determine dance moves of the audio to be choreographed according to the third move features, where the decoding model is obtained by training sample move features corresponding to the sample audios and the sample dance moves corresponding to the sample audios, the sample move features corresponding to the sample audios are obtained by encoding the sample audios by using the pre-trained encoding model, and the third move features are used to indicate move features of all of the multiple audio segments.

[0087] Further, in terms of extracting the multiple audio segments from the audio to be choreographed, the obtaining unit 701 is configured to obtain beat information of the audio to be choreographed; and extract the multiple audio segments from the audio to be choreographed according to the beat information.

[0088] Further, the first move feature is a first move feature matrix. In terms of determining, according to the first move feature of each audio segment from the move features of the multiple categories of dance moves pre-stored in the move library, the second move feature similar to the first move feature of each audio segment, the determining unit 703 is configured to: for each audio segment, obtain various row vectors in the first move feature matrix corresponding to the audio segment, to obtain multiple first move feature vectors corresponding to the audio segment, where each of the row vectors is used to indicate a move; for each of the multiple first move feature vectors of each audio segment, determine, from the move features of the multiple categories of dance moves pre-stored in the move library, a second move feature vector corresponding to the first move feature vector, to obtain multiple second move feature vectors corresponding to each audio segment; and for each audio segment, combine the multiple second move feature vectors corresponding to the audio segment to obtain a second move feature matrix corresponding to the audio segment, where the second move feature matrix is used to indicate the second move feature.

[0089] Further, in terms of determining, from the move features of the multiple categories of dance moves pre-stored in the move library, the second move feature vector corresponding to the first move feature vector, the determining unit 703 is configured to: obtain a distance between the first move feature vector and each move feature vector in the move library, where each move feature vector in the move library is used to indicate a pre-stored dance move; and obtain from the move library a move feature vector with the shortest distance to the first move feature vector as the second move feature vector corresponding to the first move feature vector.

[0090] Further, before the multiple audio segments are input into the pre-trained encoding model to obtain the first move feature of each audio segment in the multiple audio segments, the encoding unit 702 is further configured to obtain a sample data set, where the sample data set includes multiple pieces of sample dance data, and each piece of sample dance data includes a sample audio and sample dance moves; extract out a sample move feature from the sample dance moves of each piece of sample dance data; and input the sample move feature and the sample audio into an initial encoding model for training to obtain the pre-trained encoding model.

[0091] Further, in terms of extracting out the sample move feature from the sample dance moves for each piece of sample dance data, the encoding unit 702 is configured to: obtain a number of key points of a human body and poisons of the key points corresponding to each of the sample dance moves; and input the number of key points of the human body and the positions of the key points corresponding to each of the sample dance moves into the initial encoding model, and extract out the sample move feature.

[0092] Further, before the second move feature corresponding to each audio segment is input into the pre-trained decoding model to obtain the third move features, the decoding unit 704 is further configured to: input a sample audio into the pre-trained encoding model to obtain a first sample move feature of the sample audio; and determine, from the move library and according to the first sample move feature, a second sample move feature similar to the first sample move feature; input the second sample move feature into an initial decoding model to obtain a third sample move feature; and determine dance moves of the sample audio according to the third sample move feature; and compare the determined dance moves of the sample audio with sample dance moves of the sample audio, and adjust a model parameter of the initial decoding model according to a comparison result, to obtain the pre-trained decoding model.

[0093] Further, in terms of determining the dance moves for the audio to be choreographed

according to the third move features, the decoding unit 704 is configured to: determine a number of key points of a human body and a position of each key point in the audio to be choreographed according to the third move features; and determine the dance moves of the audio to be choreographed according to the number of key points of the human body and the position of each key point in the audio to be choreographed.

[0094] In embodiments of the present disclosure, the audio to be choreographed is encoded by using the pre-trained encoding model to obtain the first move feature of each audio segment, and a second move feature similar to the first move feature is determined from the move features of the multiple categories of dance moves pre-stored in the move library, which facilitates generating more accurate dance moves of high quality for the audio to be choreographed by decoding the second move feature using the pre-trained decoding model.

[0095] Reference is made to FIG. 8, which is a schematic structural diagram of a computer device provided in an embodiment of the present disclosure. Specifically, the computer device includes a memory 801 and a processor 802.

[0096] In an embodiment, the computer device further includes a data interface 803, and the data interface 803 is configured to transmit data information between the computer device and another device.

[0097] The memory 801 may include a volatile memory; the memory 801 may also include a non-volatile memory; and the memory 801 may also include a combination of memories of those kinds described above. The processor 802 may be a central processing unit (CPU). The processor 802 can further include a hardware chip. The foregoing hardware chip may be an application-specific integrated circuit (ASIC), a programmable logic device (PLD), or a combination thereof. The PLD may be a complex programmable logic device (CPLD), a field-programmable gate array (FPGA), or any combination thereof.

[0098] The memory 801 is configured to store a program, and the processor 802 is configured to invoke the program stored in the memory 801 to perform the following. An audio to be choreographed is obtained, and multiple audio segments are extracted from the audio to be choreographed. The multiple audio segments are input into a pre-trained encoding model to obtain a first move feature of each audio segment in the multiple audio segments, where the encoding model is obtained by training sample audios and sample dance moves corresponding to the sample audios. According to the first move feature of each audio segment, a second move

feature similar to the first move feature of each audio segment is determined from move features of multiple categories of dance moves pre-stored in a move library. The second move feature corresponding to each audio segment is input into a pre-trained decoding model to obtain third move features, and dance moves of the audio to be choreographed are determined according to the third move features, where the decoding model is obtained by training sample move features corresponding to the sample audios and the sample dance moves corresponding to the sample audios, the sample move features corresponding to the sample audios are obtained by encoding the sample audios by using the pre-trained encoding model, the third move features are used to indicate move features of all of the multiple audio segments.

[0099] Further, in terms of extracting the multiple audio segments from the audio to be choreographed, the processor 802 is configured to obtain beat information of the audio to be choreographed; and extract the multiple audio segments from the audio to be choreographed according to the beat information.

[0100] Further, the first move feature is a first move feature matrix. In terms of determining, according to the first move feature of each audio segment from the move features of the multiple categories of dance moves pre-stored in the move library, the second move feature similar to the first move feature of each audio segment, the processor 802 is configured to: for each audio segment, obtain various row vectors in the first move feature matrix corresponding to the audio segment, to obtain multiple first move feature vectors corresponding to the audio segment, where each of the row vectors is used to indicate a move; for each of the multiple first move feature vectors of each audio segment, determine, from the move features of the multiple categories of dance moves pre-stored in the move library, a second move feature vector corresponding to the first move feature vector, to obtain multiple second move feature vectors corresponding to each audio segment; and for each audio segment, combine the multiple second move feature vectors corresponding to the audio segment to obtain a second move feature matrix corresponding to the audio segment, where the second move feature matrix is used to indicate the second move feature.

[0101] Further, in terms of determining, from the move features of the multiple categories of dance moves pre-stored in the move library, the second move feature vector corresponding to the first move feature vector, the processor 802 is configured to: obtain a distance between the first move feature vector and each move feature vector in the move library, where each move feature

vector in the move library is used to indicate a pre-stored dance move; and obtain from the move library a move feature vector with the shortest distance to the first move feature vector as the second move feature vector corresponding to the first move feature vector.

[0102] Further, before the multiple audio segments are input into the pre-trained encoding model to obtain the first move feature of each audio segment in the multiple audio segments, the processor 802 is further configured to obtain a sample data set, where the sample data set includes multiple pieces of sample dance data, and each piece of sample dance data includes a sample audio and sample dance moves; extract out a sample move feature from the sample dance moves of each piece of sample dance data; and input the sample move feature and the sample audio into an initial encoding model for training to obtain the pre-trained encoding model.

[0103] Further, in terms of extracting out the sample move feature from the sample dance moves for each piece of sample dance data, the processor 802 is configured to: obtain a number of key points of a human body and poisons of the key points corresponding to each of the sample dance moves; and input the number of key points of the human body and the positions of the key points corresponding to each of the sample dance moves into the initial encoding model, and extract out the sample move feature.

[0104] Further, before the second move feature corresponding to each audio segment is input into the pre-trained decoding model to obtain the third move features, the processor 802 is further configured to: input a sample audio into the pre-trained encoding model to obtain a first sample move feature of the sample audio; and determining, from the move library and according to the first sample move feature, a second sample move feature similar to the first sample move feature; input the second sample move feature into an initial decoding model to obtain a third sample move feature; and determine dance moves of the sample audio according to the third sample move feature; and compare the determined dance moves of the sample audio with sample dance moves of the sample audio, and adjust a model parameter of the initial decoding model according to a comparison result, to obtain the pre-trained decoding model.

[0105] Further, in terms of determining the dance moves for the audio to be choreographed according to the third move features, the processor 802 is configured to: determine a number of key points of a human body and a position of each key point in the audio to be choreographed according to the third move features; and determine the dance moves of the audio to be choreographed according to the number of key points of the human body and the position of each

key point in the audio to be choreographed.

[0106] In embodiments of the present disclosure, the audio to be choreographed is encoded by using the pre-trained encoding model to obtain the first move feature of each audio segment, and a second move feature similar to the first move feature of each audios segment is determined from the move features of the multiple categories of dance moves pre-stored in the move library, which facilitates generating more accurate dance moves of high quality for the audio to be choreographed by decoding the second move feature of each audio segment using the pre-trained decoding model.

[0107] Embodiments of the present disclosure further provide a computer-readable storage medium. The computer-readable storage medium stores a computer program which when executed by a processor, is configured to implement the method described in the embodiments corresponding to FIG. 1, FIG. 4, or FIG. 5 of the present disclosure, and is also configured to implement the apparatus of the embodiment corresponding to FIG. 7 of the present disclosure, which will not be repeatedly described in detail herein.

[0108] The above-mentioned computer-readable storage medium may be an internal storage unit of an apparatus provided in any of the aforementioned embodiments, such as a hard disk or memory of an apparatus. The computer-readable storage medium may further be an external storage device of the apparatus, such as a plug-in hard disk, a smart media card (SMC), a secure digital (SD) card, a flash card, and so on equipped on the apparatus. Furthermore, the computer-readable storage medium may further include both the internal storage unit and the external storage device of the apparatus. The computer-readable storage medium is configured to store the computer program and other programs and data required by the apparatus. The computer-readable storage medium may further be configured to temporarily store data that has been output or is to be output.

[0109] Those of ordinary skill in the art can understand that all or some operations of methods in the foregoing embodiments can be implemented by instructing related hardware by a computer program. The program can be stored in a computer-readable storage medium. The program is operable to perform operations including the embodiments of the above-mentioned methods. The computer-readable storage medium can be a magnetic disk, an optical disk, a read-only memory (ROM), a random access memory (RAM), etc.

[0110] The above embodiments are only part of embodiments of the disclosure, which

certainly cannot be used to limit the scope of the disclosure. The ordinary skill in the art can understand all or a part of the process to realize the above embodiments of the disclosure, and the equivalent changes made in accordance with the claims of this disclosure, still belong to the scope of the disclosure covered.

What is claimed is:

1. A dance-move generation method, comprising:

obtaining an audio to be choreographed, and extracting a plurality of audio segments from the audio to be choreographed;

inputting the plurality of audio segments into a pre-trained encoding model to obtain a first move feature of each audio segment in the plurality of audio segments, wherein the encoding model is obtained by training sample audios and sample dance moves corresponding to the sample audios;

according to the first move feature of each audio segment, determining, from move features of a plurality of categories of dance moves pre-stored in a move library, a second move feature similar to the first move feature of each audio segment; and

inputting the second move feature corresponding to each audio segment into a pre-trained decoding model to obtain third move features, and determining dance moves of the audio to be choreographed according to the third move features, wherein the decoding model is obtained by training sample move features corresponding to the sample audios and the sample dance moves corresponding to the sample audios, the sample move features corresponding to the sample audios are obtained by encoding the sample audios by using the pre-trained encoding model, and the third move features are used to indicate move features of all of the plurality of audio segments.

2. The method of claim 1, wherein extracting the plurality of audio segments from the audio to be choreographed comprises:

obtaining beat information of the audio to be choreographed; and

extracting the plurality of audio segments from the audio to be choreographed according to the beat information.

3. The method of claim 1, wherein the first move feature is a first move feature matrix; and according to the first move feature of each audio segment, determining, from the move features of the plurality of categories of dance moves pre-stored in the move library, the second move feature similar to the first move feature of each audio segment, comprises:

for each audio segment, obtaining various row vectors in the first move feature matrix

corresponding to the audio segment, to obtain a plurality of first move feature vectors corresponding to the audio segment, wherein each of the row vectors is used to indicate a move;

for each of the plurality of first move feature vectors of each audio segment, determining, from the move features of the plurality of categories of dance moves pre-stored in the move library, a second move feature vector corresponding to the first move feature vector, to obtain a plurality of second move feature vectors corresponding to each audio segment; and

for each audio segment, combining the plurality of second move feature vectors corresponding to the audio segment to obtain a second move feature matrix corresponding to the audio segment, wherein the second move feature matrix is used to indicate the second move feature.

4. The method of claim 3, wherein determining, from the move features of the plurality of categories of dance moves pre-stored in the move library, the second move feature vector corresponding to the first move feature vector comprises:

obtaining a distance between the first move feature vector and each move feature vector in the move library, wherein each move feature vector in the move library is used to indicate a pre-stored dance move; and

obtaining from the move library a move feature vector with the shortest distance to the first move feature vector as the second move feature vector corresponding to the first move feature vector.

5. The method of claim 1, wherein before inputting the plurality of audio segments into the pre-trained encoding model to obtain the first move feature of each audio segment in the plurality of audio segments, the method further comprises:

obtaining a sample data set, wherein the sample data set comprises a plurality of pieces of sample dance data, and each piece of sample dance data comprises a sample audio and sample dance moves;

extracting out a sample move feature from the sample dance moves of each piece of sample dance data; and

inputting the sample move feature and the sample audio into an initial encoding model for training to obtain the pre-trained encoding model.

6. The method of claim 5, wherein extracting out the sample move feature from the sample dance moves for each piece of sample dance data comprises:

obtaining a number of key points of a human body and poisons of the key points corresponding to each of the sample dance moves; and

inputting the number of the key points of the human body and the positions of the key points corresponding to each of the sample dance moves into the initial encoding model, and extracting out the sample move feature.

7. The method of claim 1, wherein before inputting the second move feature corresponding to each audio segment into the pre-trained decoding model to obtain the third move features, the method further comprises:

inputting a sample audio into the pre-trained encoding model to obtain a first sample move feature of the sample audio; and determining, from the move library and according to the first sample move feature, a second sample move feature similar to the first sample move feature;

inputting the second sample move feature into an initial decoding model to obtain a third sample move feature; and determining dance moves of the sample audio according to the third sample move feature; and

comparing the determined dance moves of the sample audio with the sample dance moves of the sample audio, and adjusting a model parameter of the initial decoding model according to a comparison result, to obtain the pre-trained decoding model.

8. The method of claim 1, wherein determining the dance moves of the audio to be choreographed according to the third move features comprises:

determining a number of key points of a human body and a position of each key point in the audio to be choreographed according to the third move features; and

determining the dance moves of the audio to be choreographed according to the number of the key points of the human body and the position of each key point in the audio to be choreographed.

9. A computer device, comprising a processor, an input device, an output device, and a

memory, wherein the processor, the input device, the output device, and the memory are connected to one another, the memory is configured to store a computer program, the computer program comprises a program, and the processor is configured to invoke the program to execute;

obtaining an audio to be choreographed, and extracting a plurality of audio segments from the audio to be choreographed;

inputting the plurality of audio segments into a pre-trained encoding model to obtain a first move feature of each audio segment in the plurality of audio segments, wherein the encoding model is obtained by training sample audios and sample dance moves corresponding to the sample audios;

according to the first move feature of each audio segment, determining, from move features of a plurality of categories of dance moves pre-stored in a move library, a second move feature similar to the first move feature of each audio segment; and

inputting the second move feature corresponding to each audio segment into a pre-trained decoding model to obtain third move features, and determining dance moves of the audio to be choreographed according to the third move features, wherein the decoding model is obtained by training sample move features corresponding to the sample audios and the sample dance moves corresponding to the sample audios, the sample move features corresponding to the sample audios are obtained by encoding the sample audios by using the pre-trained encoding model, and the third move features are used to indicate move features of all of the plurality of audio segments.

10. The computer device of claim 9, wherein in terms of extracting the plurality of audio segments from the audio to be choreographed, the processor is configured to invoke the program to execute:

obtaining beat information of the audio to be choreographed; and

extracting the plurality of audio segments from the audio to be choreographed according to the beat information.

11. The computer device of claim 9, wherein the first move feature is a first move feature matrix; and according to the first move feature of each audio segment, in terms of determining, from the move features of the plurality of categories of dance moves pre-stored in the move

library, the second move feature similar to the first move feature of each audio segment, the processor is configured to invoke the program to execute:

for each audio segment, obtaining various row vectors in the first move feature matrix corresponding to the audio segment, to obtain a plurality of first move feature vectors corresponding to the audio segment, wherein each of the row vectors is used to indicate a move;

for each of the plurality of first move feature vectors of each audio segment, determining, from the move features of the plurality of categories of dance moves pre-stored in the move library, a second move feature vector corresponding to the first move feature vector, to obtain a plurality of second move feature vectors corresponding to each audio segment; and

for each audio segment, combining the plurality of second move feature vectors corresponding to the audio segment to obtain a second move feature matrix corresponding to the audio segment, wherein the second move feature matrix is used to indicate the second move feature.

12. The computer device of claim 11, wherein in terms of determining, from the move features of the plurality of categories of dance moves pre-stored in the move library, the second move feature vector corresponding to the first move feature vector, the processor is configured to invoke the program to execute:

obtaining a distance between the first move feature vector and each move feature vector in the move library, wherein each move feature vector in the move library is used to indicate a pre-stored dance move; and

obtaining from the move library a move feature vector with the shortest distance to the first move feature vector as the second move feature vector corresponding to the first move feature vector.

13. The computer device of claim 9, wherein the processor is further configured to invoke the program to execute:

before inputting the plurality of audio segments into the pre-trained encoding model to obtain the first move feature of each audio segment in the plurality of audio segments,

obtaining a sample data set, wherein the sample data set comprises a plurality of pieces of sample dance data, and each piece of sample dance data comprises a sample

audio and sample dance moves;

extracting out a sample move feature from the sample dance moves of each piece of sample dance data; and

inputting the sample move feature and the sample audio into an initial encoding model for training to obtain the pre-trained encoding model.

14. The computer device of claim 13, wherein in terms of extracting out the sample move feature from the sample dance moves for each piece of sample dance data, the processor is configured to invoke the program to execute:

obtaining a number of key points of a human body and poisons of the key points corresponding to each of the sample dance moves; and

inputting the number of the key points of the human body and the positions of the key points corresponding to each of the sample dance moves into the initial encoding model, and extracting out the sample move feature.

15. The computer device of claim 9, wherein the processor is further configured to invoke the program to execute:

before inputting the second move feature corresponding to each audio segment into the pre-trained decoding model to obtain the third move features,

inputting a sample audio into the pre-trained encoding model to obtain a first sample move feature of the sample audio; and determining, from the move library and according to the first sample move feature, a second sample move feature similar to the first sample move feature;

inputting the second sample move feature into an initial decoding model to obtain a third sample move feature; and determining dance moves of the sample audio according to the third sample move feature; and

comparing the determined dance moves of the sample audio with the sample dance moves of the sample audio, and adjusting a model parameter of the initial decoding model according to a comparison result, to obtain the pre-trained decoding model.

16. The computer device of claim 9, wherein in terms of determining the dance moves of the

audio to be choreographed according to the third move features, the processor is configured to invoke the program to execute:

determining a number of key points of a human body and a position of each key point in the audio to be choreographed according to the third move features; and

determining the dance moves of the audio to be choreographed according to the number of the key points of the human body and the position of each key point in the audio to be choreographed.

17. A non-transitory computer-readable storage medium, wherein the computer-readable storage medium stores a program instruction which, when executed, is configured to implement:

obtaining an audio to be choreographed, and extracting a plurality of audio segments from the audio to be choreographed;

inputting the plurality of audio segments into a pre-trained encoding model to obtain a first move feature of each audio segment in the plurality of audio segments, wherein the encoding model is obtained by training sample audios and sample dance moves corresponding to the sample audios;

according to the first move feature of each audio segment, determining, from move features of a plurality of categories of dance moves pre-stored in a move library, a second move feature similar to the first move feature of each audio segment; and

inputting the second move feature corresponding to each audio segment into a pre-trained decoding model to obtain third move features, and determining dance moves of the audio to be choreographed according to the third move features, wherein the decoding model is obtained by training sample move features corresponding to the sample audios and the sample dance moves corresponding to the sample audios, the sample move features corresponding to the sample audios are obtained by encoding the sample audios by using the pre-trained encoding model, and the third move features are used to indicate move features of all of the plurality of audio segments.

18. The non-transitory computer-readable storage medium of claim 17, wherein in terms of extracting the plurality of audio segments from the audio to be choreographed, the program instruction, when executed, is configured to implement:

obtaining beat information of the audio to be choreographed; and

extracting the plurality of audio segments from the audio to be choreographed according to the beat information.

19. The non-transitory computer-readable storage medium of claim 17, wherein the first move feature is a first move feature matrix; and according to the first move feature of each audio segment, in terms of determining, from the move features of the plurality of categories of dance moves pre-stored in the move library, the second move feature similar to the first move feature of each audio segment, the program instruction, when executed, is configured to implement:

for each audio segment, obtaining various row vectors in the first move feature matrix corresponding to the audio segment, to obtain a plurality of first move feature vectors corresponding to the audio segment, wherein each of the row vectors is used to indicate a move;

for each of the plurality of first move feature vectors of each audio segment, determining, from the move features of the plurality of categories of dance moves pre-stored in the move library, a second move feature vector corresponding to the first move feature vector, to obtain a plurality of second move feature vectors corresponding to each audio segment; and

for each audio segment, combining the plurality of second move feature vectors corresponding to the audio segment to obtain a second move feature matrix corresponding to the audio segment, wherein the second move feature matrix is used to indicate the second move feature.

20. The non-transitory computer-readable storage medium of claim 19, wherein in terms of determining, from the move features of the plurality of categories of dance moves pre-stored in the move library, the second move feature vector corresponding to the first move feature vector, the program instruction, when executed, is configured to implement:

obtaining a distance between the first move feature vector and each move feature vector in the move library, wherein each move feature vector in the move library is used to indicate a pre-stored dance move; and

obtaining from the move library a move feature vector with the shortest distance to the first move feature vector as the second move feature vector corresponding to the first move feature vector.