CLAIMS

1. A pose generation method, comprising:

acquiring a temporally associated predicted pose feature sequence, wherein the predicted pose feature sequence is generated at least according to an audio feature;

determining a quantized pose feature matching each predicted pose feature in the predicted pose feature sequence from a preset pose feature library;

determining a spatial association relationship of respective pose features in the predicted pose feature sequence; and

combining the quantized pose feature matching each predicted pose feature according to the spatial association relationship of the respective pose features and a temporal association relationship of the respective pose features, to obtain a target pose feature sequence,

wherein the method further comprises:

acquiring a sample audio feature and a sample starting pose feature of a sample role;

splitting the sample starting pose feature according to part information of the sample role to obtain at least two sample starting pose sub-features;

determining a predicted sample pose feature sequence according to each sample audio sub-feature in the sample audio feature and each sample starting pose sub-feature; and

training a starting pose prediction model based on each predicted sample pose feature in the predicted sample pose feature sequence and each preset sample pose feature, to obtain a target pose prediction model,

wherein acquiring the temporally associated predicted pose feature sequence comprises:

acquiring the temporally associated predicted pose feature sequence using the target pose prediction model,

wherein training the starting pose prediction model based on each predicted sample pose feature in the predicted sample pose feature sequence and each preset sample pose feature, to obtain the target pose prediction model comprises:

obtaining candidate pose features and a probability value corresponding to each candidate pose feature according to each sample audio sub-feature and each sample starting pose sub-feature, wherein the probability value is used for determining the predicted sample pose feature from the candidate pose features;

determining a third loss value according to a number of the sample audio sub-features, the preset sample pose feature and the probability value;

determining a fourth loss value according to the number of the sample audio sub-features, the predicted sample pose feature and the probability value;

obtaining a second target loss value according to the third loss value and the fourth loss value; and

updating a model parameter of the starting pose prediction model using the second target loss value, so as to obtain the target pose prediction model.

2. The method of claim 1, wherein determining the quantized pose feature matching each predicted pose feature in the predicted pose feature sequence from the preset pose feature library comprises:

respectively performing dimension reduction processing on each predicted pose feature to obtain an encoded feature, wherein a dimension of the encoded feature is the same as a dimension of each quantized pose feature in the preset pose feature library; and

determining the quantized pose feature matching each predicted pose feature from the preset pose feature library according to a distance between the encoded feature and each quantized pose feature in the preset pose feature library.

3. The method of claim 2, wherein determining the quantized pose feature matching each predicted pose feature from the preset pose feature library according to the distance between the encoded feature and each quantized pose feature in the preset pose feature library comprises:

determining a distance between a to-be-processed encoded feature and each quantized pose feature in the preset pose feature library;

sorting respective distances corresponding to the to-be-processed encoded feature to obtain a sorting result;

determining the quantized pose feature corresponding to the to-be-processed encoded feature according to the sorting result; and

determining the quantized pose feature matching each predicted pose feature according to the quantized pose feature corresponding to each to-be-processed encoded feature.

4. The method of any of claims 1 to 3, wherein determining the spatial association relationship of respective pose features in the predicted pose feature sequence comprises:

determining relative position information between each key point of a target role and a preset standard point and displacement information of the target role according to each quantized pose feature, wherein the predicted pose feature sequence is generated according to the audio feature and a starting pose feature of the target role; and

determining the spatial association relationship of the respective pose features according to the relative position information and the displacement information.

5. The method of any of claims 1 to 4, further comprising:

acquiring a sample pose feature sequence;

updating a model parameter of a starting feature processing model and a quantized pose feature in a starting pose feature library of the starting feature processing model according to respective sample pose features in the sample pose feature sequence and a preset spatial association relationship of the respective sample pose features, to obtain an updating result; and

obtaining a target feature processing model having the preset pose feature library based on the updating result,

wherein determining the quantized pose feature matching each predicted pose feature in the predicted pose feature sequence from the preset pose feature library comprises:

determining the quantized pose feature matching each predicted pose feature in the predicted pose feature sequence from the preset pose feature library of the target feature processing model.

6. The method of claim 5, wherein updating the model parameter of the starting feature processing model and the quantized pose feature in the starting pose feature library of the starting feature processing model according to respective sample pose features in the sample pose feature sequence and the preset spatial association relationship of the respective sample pose features, to obtain the updating result comprises:

obtaining a predicated spatial association relationship of the respective sample pose

features using the starting feature processing model;

determining a first loss value according to the predicted spatial association relationship and the preset spatial association relationship;

performing dimension reduction processing on each sample pose feature to obtain a sample encoded feature;

determining a second loss value according to the sample encoded feature;

obtaining a first target loss value according to the first loss value and the second loss value; and

updating the model parameter of the starting feature processing model and the quantized pose feature in the starting pose feature library of the starting feature processing model using the first target loss value, to obtain the updating result.

7. The method of claim 6, wherein determining the second loss value according to the sample encoded feature comprises:

determining a quantized pose feature matching the sample pose feature from the starting pose feature library according to the sample encoded feature; and

determining the second loss value according to the sample encoded feature and the quantized pose feature matching the sample pose feature.

8. The method of claim 7, wherein determining the second loss value according to the sample encoded feature and the quantized pose feature matching the sample pose feature comprises:

determining a first loss sub-value according to a difference value between the sample encoded feature and a sample quantized pose feature;

determining a second loss sub-value according to an opposite value of the difference value between the sample encoded feature and the sample quantized pose feature; and

determining the second loss value according to the first loss sub-value, the second loss sub-value, and their respective weight coefficients.

9. The method of any of claims 1 to 8, wherein acquiring the temporally associated predicted pose feature sequence comprises:

determining the predicted pose feature temporally aligned with each audio sub-feature in the audio feature according to the each audio sub-feature and a starting pose feature of a target role, wherein respective audio sub-features in the audio feature are temporally associated; and

obtaining the predicted pose feature sequence according to each predicted pose feature.

10. The method of claim 9, wherein determining the predicted pose feature temporally aligned with each audio sub-feature in the audio feature according to the each audio sub-feature and the starting pose feature of the target role comprises:

splitting the starting pose feature according to part information of the target role to obtain at least two starting pose sub-features;

determining an association degree between each audio sub-feature and each starting pose sub-feature, wherein the association degree at least characterizes an association degree between each audio sub-feature and each starting pose sub-feature in terms of content and time; and

determining the predicted pose feature according to the association degree between

each audio sub-feature and each starting pose sub-feature.

11. The method of claim 1, wherein determining the third loss value according to the number of the sample audio sub-features, the preset sample pose feature and the probability value comprises:

determining a difference value between a vector feature of the probability value and the preset sample pose feature; and

determining the third loss value according to the difference value between the vector feature of the probability value and the preset sample pose feature, and the number of the sample audio sub-features.

12. The method of claim 1, wherein determining the fourth loss value according to the number of the sample audio sub-features, the predicted sample pose feature and the probability value comprises:

determining a difference value between the predicted sample pose feature and a vector feature of the probability value;

determining a direction feature of the sample role according to each sample starting pose sub-feature; and

determining the fourth loss value according to the number of the sample audio sub-features, the difference value between the predicted sample pose feature and the vector feature of the probability value, and the direction feature.

13. The method of any of claims 1 to 12, further comprising:

acquiring an audio feature of a target music and a starting pose feature of a target role;

generating the temporally associated predicted pose feature sequence according to the audio feature of the target music and the starting pose feature of the target role, wherein the predicted pose feature sequence comprises predicted dance pose features;

determining a quantized pose feature matching each predicted dance pose feature from the preset pose feature library;

determining a spatial association relationship of respective predicted dance pose features in the predicted pose feature sequence;

combining the quantized pose feature matching each predicted dance pose feature according to the spatial association relationship of respective predicted dance pose features and a temporal association relationship of the respective predicted dance pose features, to obtain the target pose feature sequence corresponding to the target role; and

controlling the target role to perform a target dance pose corresponding to the target pose feature sequence.

14. A pose generation apparatus, comprising:

an acquisition module, configured to acquire a temporally associated predicted pose feature sequence, wherein the predicted pose feature sequence is generated at least according to an audio feature;

a first determination module, configured to determine a quantized pose feature matching each predicted pose feature in the predicted pose feature sequence from a preset pose feature library;

a second determination module, configured to determine a spatial association relationship of respective pose features in the predicted pose feature sequence; and

a combination module, configured to combine the quantized pose feature matching each predicted pose feature according to the spatial association relationship of the respective pose features and a temporal association relationship of the respective pose features, to obtain a target pose feature sequence,

wherein the apparatus further comprises:

a fifth determination module, configured to acquire a sample audio feature and a sample starting pose feature of a sample role;

a splitting module, configured to split the sample starting pose feature according to part information of the sample role to obtain at least two sample starting pose sub-features;

a sixth determination module, configured to determine a predicted sample pose feature sequence according to each sample audio sub-feature in the sample audio feature and each sample starting pose sub-feature; and

a seventh determination module, configured to train a starting pose prediction model based on each predicted sample pose feature in the predicted sample pose feature sequence and each preset sample pose feature, to obtain a target pose prediction model,

wherein the acquisition module is configured to: acquire the temporally associated predicted pose feature sequence using the target pose prediction model,

wherein the seventh determination module is configured to:

obtain candidate pose features and a probability value corresponding to each candidate pose feature according to each sample audio sub-feature and each sample starting pose sub-feature, wherein the probability value is used for determining the predicted sample pose feature from the candidate pose features;

determine a third loss value according to a number of the sample audio sub-features, the preset sample pose feature and the probability value;

determine a fourth loss value according to the number of the sample audio sub-features, the predicted sample pose feature and the probability value;

obtain a second target loss value according to the third loss value and the fourth loss value; and

update a model parameter of the starting pose prediction model using the second target loss value, so as to obtain the target pose prediction model.

15. An electronic device, comprising a processor and a memory configured to store a computer program executable on the processor, wherein

the processor is configured to operate the computer program to perform the pose generation method of any of claims 1 to 13.

16. A computer storage medium, having a computer program stored thereon, wherein the pose generation method of any of claims 1 to 13 is implemented when the computer program is performed by a processor.