CLAIMS

What is claimed is:

1. A video data processing method, performed by a computer device, the method comprising:

obtaining video data of a target video requested by a target user, and performing video analysis on the video data to obtain a plurality of video segments, the video analysis comprising storyboard processing and attribute analysis based on a plurality of preset segment attribute tags, and each video segment in the plurality of video segments corresponding to one segment attribute tag and one storyboard segment;

determining a video template associated with the target user from a video template database based on a user portrait of the target user, and obtaining at least one template segment and a template tag sequence in the video template, the template tag sequence being based on a template attribute tag of the at least one template segment;

screening at least one video segment matching the template attribute tag of the at least one template segment from the plurality of video segments based on the template attribute tag of the at least one template segment and segment attribute tags corresponding to the plurality of video segments;

splicing the at least one matched video segment according to a position of a template attribute tag, of each template segment in the at least one template segment, in the template tag sequence, as a video material segment of the target video; and

pushing the video data and the video material segment to an application client corresponding to the target user to be output;

wherein the performing the video analysis comprises: performing the storyboard processing on a video sequence corresponding to the video data through a video partitioning component, to obtain a plurality of storyboard segments associated with the video sequence; inputting the plurality of storyboard segments into a network recognition model, and performing the attribute analysis on the plurality of storyboard segments through the network recognition model based on the plurality of preset segment attribute tags, to obtain segment attribute tags corresponding to the plurality of storyboard segments; and determining the plurality of storyboard segments comprising the segment attribute tags as the plurality of video segments of the video data,

wherein the performing the storyboard processing comprises: determining a first video frame serving as a cluster centroid in the video sequence through the video partitioning component, and generating storyboard cluster information of a storyboard cluster to which the first video frame belongs; determining video frames other than the first video frame in the video sequence as second video frames, sequentially obtaining each second video frame in the second video frames based on a pooling mechanism, and determining an image similarity between each second video frame and the first video frame; and determining a storyboard cluster to which each video frame in the video sequence belongs based on a result of the image similarity, and forming each video frame in the video sequence into the plurality of storyboard segments based on the storyboard cluster information of the storyboard cluster to which each video frame in the video sequence belongs, and

wherein the determining the storyboard cluster to which each video frame in the video sequence belongs comprises: based on the image similarity between the first video frame and a second video frame being greater than or equal to a clustering threshold, allocating the second video frame whose image similarity is greater than or equal to the clustering threshold to the storyboard cluster to which the first video frame belongs; and based on the image similarity between the first video frame and a second video frame being less than the clustering threshold, updating the first video frame using the second video frame whose image similarity is less than the clustering threshold, generating storyboard cluster information of a storyboard cluster to which the updated first video frame belongs, and sequentially performing image similarity matching between the updated first video frame and second video frames that were not previously matched until image similarity matching is performed on each video frame in the video sequence, to obtain storyboard cluster information of a storyboard cluster to which each second video frame in the video sequence belongs.

2. The method according to claim 1, further comprising, prior to the obtaining the video data of the target video: extracting, in response to receiving a video playing request for the target video from the application client, a video identifier of the target video from the video playing request; and obtaining the video data of the target video by searching a video service database based on the video identifier.

3. The method according to claim 1, wherein the network recognition model comprises a first network model comprising a first attribute tag extraction function, a second network model comprising a second attribute tag extraction function, and a third network model comprising a

third attribute tag extraction function; and

the inputting the plurality of storyboard segments into the network recognition model and the performing the attribute analysis comprises:

inputting the plurality of storyboard segments into the first network model, performing long shot and close shot analysis on each storyboard segment in the plurality of storyboard segments through the first network model to obtain long shot and close shot tags of the plurality of storyboard segments, using the long shot and close shot tags of the plurality of storyboard segments as a first attribute tag outputted by the first network model, and using storyboard segments comprising the first attribute tag as storyboard segments of a first type;

inputting the storyboard segments of the first type into the second network model, and performing face detection on each storyboard segment in the storyboard segments of the first type through the second network model to obtain a face detection result;

using, based on the face detection result indicating that a face of a target character exists in the storyboard segments of the first type, storyboard segments corresponding to the face of the target character existing in the storyboard segments of the first type as storyboard segments of a second type, determining a character tag to which the target character in the storyboard segments of the second type belongs through the second network model, and determining the character tag to which the target character belongs as a second attribute tag of the storyboard segments of the second type, wherein the target character is one or more characters in the target video;

determining storyboard segments other than the storyboard segments of the second type in the storyboard segments of the first type as storyboard segments of a third type, inputting the storyboard segments of the third type into the third network model, and performing scene detection on each storyboard segment in the storyboard segments of the first type through the third network model to obtain a third attribute tag of the storyboard segments of the third type; and

determining a segment attribute tag corresponding to each storyboard segment in the plurality of storyboard segments according to the first attribute tag of the storyboard segments of the first type, the second attribute tag of the storyboard segments of the second type, and the third attribute tag of the storyboard segments of the third type.

4. The method according to claim 1, wherein the determining the video template and the obtaining the at least one template segment and the template tag sequence in the video template

comprises:

obtaining a behavior log table of the target user, and extracting behavior data information associated with the target user from the behavior log table;

performing user portrait analysis on the behavior data information to obtain the user portrait used for representing the target user; and

determining the video template associated with the target user from the video template database based on the user portrait of the target user, and obtaining the at least one template segment and the template tag sequence in the video template.

5. The method according to claim 1, wherein a number of the at least one template segment is N, and N is a positive integer greater than 1; the template tag sequence comprises N sequence positions, one sequence position corresponds to one template attribute tag, and one template attribute tag corresponds to one template segment; and

wherein the screening the at least one video segment matching the template attribute tag of the at least one template segment comprises:

using each of N template segments as a target template segment, determining a queue position of the target template segment in the template tag sequence as a target queue position, and determining a template attribute tag corresponding to the target queue position as a target template attribute tag;

screening a segment attribute tag matching the target template attribute tag from the segment attribute tags corresponding to the plurality of video segments, and determining one or more video segments corresponding to the screened segment attribute tag as candidate video segments; and

performing similarity analysis on each candidate video segment in the candidate video segments and the target template segment to obtain a similarity threshold between each candidate video segment and the target template segment, determining a maximum similarity threshold in the similarity thresholds, and determining a candidate video segment corresponding to the maximum similarity threshold as a target candidate video segment matching the target template segment.

6. The method according to claim 5, wherein the splicing the at least one matched video segment comprises:

determining a target tag sequence formed by a segment attribute tag corresponding to the target candidate video segment based on the target queue position of the target template segment in the template tag sequence, and performing splicing processing on all target candidate video segments associated with the target tag sequence to obtain the video material segment.

7. The method according to claim 6, wherein the performing the splicing processing on all the target candidate video segments comprises:

performing video splicing processing on all the target candidate video segments associated with the target tag sequence to obtain spliced video data associated with the N template segments; and

obtaining template audio data associated with the N template segments, and performing audio and video combination processing on the template audio data and the spliced video data through an audio and video combination component to obtain the video material segment.

8. A video data processing method, comprising:

obtaining, in response to a play operation performed by a target user on a target video in an application client, video data of the target video and a video material segment associated with the target video from a server, wherein the video material segment is a plurality of video segments obtained by video analysis performed by the server on the video data, the video analysis comprises storyboard processing and attribute analysis based on a plurality of preset segment attribute tags, and each video segment in the plurality of video segments corresponds to one segment attribute tag and one storyboard segment; a video template associated with the target user is determined from a video template database based on a user portrait of the target user, and at least one template segment and a template tag sequence in the video template are obtained, wherein the template tag sequence is formed by a template attribute tag of the at least one template segment; at least one video segment matching the template attribute tag of the at least one template segment is screened from the plurality of video segments based on the template attribute tag of the at least one template segment and segment attribute tags corresponding to the plurality of video segments; and the video material segment is obtained by splicing the at least one matched video segment according to a position of a template attribute tag, of each template segment in the at least one template segment, in the template tag sequence; and outputting the video data and the video material segment in an application display interface of the application client,

wherein performing the video analysis comprises: performing the storyboard processing on a video sequence corresponding to the video data through a video partitioning component, to obtain a plurality of storyboard segments associated with the video sequence; inputting the plurality of storyboard segments into a network recognition model, and performing the attribute analysis on the plurality of storyboard segments through the network recognition model based on the plurality of preset segment attribute tags, to obtain segment attribute tags corresponding to the plurality of storyboard segments; and determining the plurality of storyboard segments comprising the segment attribute tags as the plurality of video segments of the video data,

wherein the performing the storyboard processing comprises: determining a first video frame serving as a cluster centroid in the video sequence through the video partitioning component, and generating storyboard cluster information of a storyboard cluster to which the first video frame belongs; determining video frames other than the first video frame in the video sequence as second video frames, sequentially obtaining each second video frame in the second video frames based on a pooling mechanism, and determining an image similarity between each second video frame and the first video frame; and determining a storyboard cluster to which each video frame in the video sequence belongs based on a result of the image similarity, and forming each video frame in the video sequence into the plurality of storyboard segments based on the storyboard cluster information of the storyboard cluster to which each video frame in the video sequence belongs, and

wherein the determining the storyboard cluster to which each video frame in the video sequence belongs comprises: based on the image similarity between the first video frame and a second video frame being greater than or equal to a clustering threshold, allocating the second video frame whose image similarity is greater than or equal to the clustering threshold to the storyboard cluster to which the first video frame belongs; and based on the image similarity between the first video frame and a second video frame being less than the clustering threshold, updating the first video frame using the second video frame whose image similarity is less than the clustering threshold, generating storyboard cluster information of a storyboard cluster to which the updated first video frame belongs, and sequentially performing image similarity matching between the updated first video frame and second video frames that were not previously matched until image similarity matching is performed on each video frame in the video sequence, to obtain storyboard cluster information of a storyboard cluster to which each second video frame in the video sequence belongs.

9. The method according to claim 8, wherein the obtaining the video data of the target video and the video material segment associated with the target video from the server comprises:

generating, in response to the play operation performed by the target user on the target video in the application client, a video playing request used for requesting to play the target video, and transmitting the video playing request to the server, wherein the video playing request includes a video identifier of the target video; and

receiving the video data and the video material segment associated with the target video returned by the server based on the video playing request, wherein the video material segment is obtained after video analysis and video matching are performed on the video data by the server according to the video template based on the video template that is determined according to the user portrait of the target user, and the user portrait is determined according to user behavior information of the target user in the application client.

10. The method according to claim 8, wherein the outputting the video data and the video material segment comprises:

determining a video playing interface used for playing the video data in the application display interface of the application client, and playing the video data in the video playing interface; and

playing the video material segment in the application display interface in response to a trigger operation on the application display interface.

11. A video data processing apparatus, comprising:

a segment generation module, configured to obtain video data of a target video requested by a target user, and perform video analysis on the video data to obtain a plurality of video segments, the video analysis comprising storyboard processing and attribute analysis based on a plurality of preset segment attribute tags, and each video segment in the plurality of video segments corresponding to one segment attribute tag and one storyboard segment;

a template obtaining module, configured to determine a video template associated with the target user from a video template database based on a user portrait of the target user, and obtain at least one template segment and a template tag sequence in the video template, the template tag sequence based on a template attribute tag of the at least one template segment;

a material determining module, configured to screen at least one video segment matching

the template attribute tag of the at least one template segment from the plurality of video segments based on the template attribute tag of the at least one template segment and segment attribute tags corresponding to the plurality of video segments, and splice the at least one matched video segment according to a position of a template attribute tag of each template segment in the at least one template segment in the template tag sequence as a video material segment of the target video; and

a data transmission module, configured to push the video data and the video material segment to an application client corresponding to the target user to be output;

wherein performing the video analysis comprises: performing the storyboard processing on a video sequence corresponding to the video data through a video partitioning component, to obtain a plurality of storyboard segments associated with the video sequence; inputting the plurality of storyboard segments into a network recognition model, and performing the attribute analysis on the plurality of storyboard segments through the network recognition model based on the plurality of preset segment attribute tags, to obtain segment attribute tags corresponding to the plurality of storyboard segments; and determining the plurality of storyboard segments comprising the segment attribute tags as the plurality of video segments of the video data,

wherein the performing the storyboard processing comprises: determining a first video frame serving as a cluster centroid in the video sequence through the video partitioning component, and generating storyboard cluster information of a storyboard cluster to which the first video frame belongs; determining video frames other than the first video frame in the video sequence as second video frames, sequentially obtaining each second video frame in the second video frames based on a pooling mechanism, and determining an image similarity between each second video frame and the first video frame; and determining a storyboard cluster to which each video frame in the video sequence belongs based on a result of the image similarity, and forming each video frame in the video sequence into the plurality of storyboard segments based on the storyboard cluster information of the storyboard cluster to which each video frame in the video sequence belongs, and

wherein the determining the storyboard cluster to which each video frame in the video sequence belongs comprises: based on the image similarity between the first video frame and a second video frame being greater than or equal to a clustering threshold, allocating the second video frame whose image similarity is greater than or equal to the clustering threshold to the storyboard cluster to which the first video frame belongs; and based on the image similarity

between the first video frame and a second video frame being less than the clustering threshold, updating the first video frame using the second video frame whose image similarity is less than the clustering threshold, generating storyboard cluster information of a storyboard cluster to which the updated first video frame belongs, and sequentially performing image similarity matching between the updated first video frame and second video frames that were not previously matched until image similarity matching is performed on each video frame in the video sequence, to obtain storyboard cluster information of a storyboard cluster to which each second video frame in the video sequence belongs.

12. A video data processing apparatus, comprising:

a data obtaining module, configured to obtain, in response to a play operation performed by a target user on a target video in an application client, video data of the target video and a video material segment associated with the target video from a server, wherein the video material segment is a plurality of video segments obtained by video analysis performed by the server on the video data, wherein the video analysis comprises storyboard processing and attribute analysis based on a plurality of preset segment attribute tags, and each video segment in the plurality of video segments corresponds to one segment attribute tag and one storyboard segment; a video template associated with the target user is determined from a video template database based on a user portrait of the target user, and at least one template segment and a template tag sequence in the video template are obtained, wherein the template tag sequence is formed by a template attribute tag of the at least one template segment; at least one video segment matching the template attribute tag of the at least one template segment is screened from the plurality of video segments based on the template attribute tag of the at least one template segment and segment attribute tags corresponding to the plurality of video segments; and the video material segment is obtained by splicing the at least one matched video segment according to a position of a template attribute tag, of each template segment in the at least one template segment, in the template tag sequence; and

a data output module, configured to output the video data and the video material segment in an application display interface of the application client,

wherein performing the video analysis comprises: performing the storyboard processing on a video sequence corresponding to the video data through a video partitioning component, to obtain a plurality of storyboard segments associated with the video sequence; inputting the plurality of storyboard segments into a network recognition model, and performing the attribute

analysis on the plurality of storyboard segments through the network recognition model based on the plurality of preset segment attribute tags, to obtain segment attribute tags corresponding to the plurality of storyboard segments; and determining the plurality of storyboard segments comprising the segment attribute tags as the plurality of video segments of the video data,

wherein the performing the storyboard processing comprises: determining a first video frame serving as a cluster centroid in the video sequence through the video partitioning component, and generating storyboard cluster information of a storyboard cluster to which the first video frame belongs; determining video frames other than the first video frame in the video sequence as second video frames, sequentially obtaining each second video frame in the second video frames based on a pooling mechanism, and determining an image similarity between each second video frame and the first video frame; and determining a storyboard cluster to which each video frame in the video sequence belongs based on a result of the image similarity, and forming each video frame in the video sequence into the plurality of storyboard segments based on the storyboard cluster information of the storyboard cluster to which each video frame in the video sequence belongs, and

wherein the determining the storyboard cluster to which each video frame in the video sequence belongs comprises: based on the image similarity between the first video frame and a second video frame being greater than or equal to a clustering threshold, allocating the second video frame whose image similarity is greater than or equal to the clustering threshold to the storyboard cluster to which the first video frame belongs; and based on the image similarity between the first video frame and a second video frame being less than the clustering threshold, updating the first video frame using the second video frame whose image similarity is less than the clustering threshold, generating storyboard cluster information of a storyboard cluster to which the updated first video frame belongs, and sequentially performing image similarity matching between the updated first video frame and second video frames that were not previously matched until image similarity matching is performed on each video frame in the video sequence, to obtain storyboard cluster information of a storyboard cluster to which each second video frame in the video sequence belongs.

13. A computing device, comprising a processor, a memory, and a network interface,

the processor being connected to the memory and the network interface, the network interface being configured to provide a data communication function, the memory being configured to store a computer program, the processor being configured to invoke the computer

program to perform the method according to any one of claims and 1 to 10.

14. A computer-readable storage medium, storing a computer program, the computer program being adapted to be loaded and executed to perform the method according to any one of claims 1 to 10.