CLAIMS

1. A retrieval method, comprising:

determining a first similarity between a text and at least one video, wherein the text is used for representing a retrieval condition;

determining a first character interaction graph of the text and a second character interaction graph of the at least one video;

determining a second similarity between the first character interaction graph and the second character interaction graph; and

determining, according to the first similarity and the second similarity, a video matching the retrieval condition from the at least one video.

2. The retrieval method of claim 1, wherein determining the first similarity between the text and the at least one video comprises:

determining a paragraph feature of the text;

determining a video feature of the at least one video; and

determining the first similarity between the text and the at least one video according to the paragraph feature of the text and the video feature of the at least one video.

3. The retrieval method of claim 2, wherein the paragraph feature comprises a sentence feature and a number of sentences; and the video feature comprises a shot feature and a number of shots.

4. The retrieval method of any one of claims 1 to 3, wherein determining the first character interaction graph of the text comprises:

detecting a person name included in the text;

retrieving, in a database, a portrait of a person corresponding to the person name, and extracting an image feature of the portrait to obtain a character node of the person;

determining a semantic tree of the text by parsing, and obtaining a motion feature of the person based on the semantic tree to obtain an action node of the person; and

linking a character node corresponding to each person with a respective action node,

wherein the character node of the person is represented by the image feature of the portrait, and the action node of the person is represented by the motion feature in the semantic tree.

5. The retrieval method of claim 4, further comprising:

linking character nodes linked with a same action node.

6. The retrieval method of claim 4 or 5, wherein detecting the person name included in the text comprises:

replacing a pronoun in the text with a person name represented by the pronoun.

7. The retrieval method of any one of claims 1 to 6, wherein determining the second character interaction graph of the at least one video comprises:

detecting a person in each shot of the at least one video;

extracting a human feature and a motion feature of the person;

attaching the human feature of the person to a character node of the person, and attaching the motion feature of the person to an action node of the person; and

linking a character node corresponding to each person with a respective action node.

8. The retrieval method of claim 7, wherein determining the second character interaction graph of the at least one video further comprises: taking a group of persons appearing in a same shot as a same group of persons, and linking the character nodes of the persons in the same group of persons two by two.

9. The retrieval method of claim 7 or 8, wherein determining the second character interaction graph of the at least one video further comprises:

linking one person in one shot with a character node of each person in an adjacent shot of the shot.

10. The retrieval method of any one of claims 1 to 9, wherein determining, according to the first similarity and the second similarity, the video matching the retrieval condition from the at least one video comprises:

performing weighted sum on the first similarity and the second similarity for each video, to obtain a similarity value for each video; and

determining a video with a highest similarity value as the video matching the retrieval conditions.

11. The retrieval method of any one of claims 1 to 10, wherein the retrieval method is implemented through a retrieval network, and the method further comprises:

determining a prediction value of the first similarity between the text and a video in a training sample set, wherein the text is used for representing the retrieval condition;

determining a prediction value of the second similarity between the first character interaction graph of the text and the second character interaction graph of the video in the training sample set;

determining a loss of the first similarity according to the prediction value of the first similarity and ground truth of the first similarity;

determining a loss of the second similarity according to the prediction value of the second similarity and ground truth of the second similarity;

determining an overall loss value according to the loss of the first similarity and the loss of the second similarity in combination with a loss function; and

adjusting weight parameters of the retrieval network according to the overall loss value.

12. The retrieval method of claim 11, wherein the retrieval network comprises a first sub-network and a second sub-network, the first sub-network being used for determining the first similarity between the text and the video, and the second sub-network being used for determining the similarity between the first character interaction graph of the text and the second character interaction graph of the video, and

wherein adjusting the weight parameters of the retrieval network according to the overall loss value comprises:

adjusting, based on the overall loss value, the weight parameters of the first sub-network and the second sub-network.

13. A retrieval device, comprising:

a first determining module, configured to determine a first similarity between a text and at least one video, wherein the text is used for representing a retrieval condition;

a second determining module, configured to determine a first character interaction graph of the text and a second character interaction graph of the at least one video, and determine a second similarity between the first character interaction graph and the second character interaction graph; and

a processing module, configured to determine, according to the first similarity and the second similarity, a video matching the retrieval condition from the at least one video.

14. The retrieval device of claim 13, wherein the first determining module is configured to:

determine a paragraph feature of the text;

determine a video feature of the at least one video; and

determine the first similarity between the text and the at least one video according to the paragraph feature of the text and the video feature of the at least one video.

15. The retrieval device of claim 14, wherein the paragraph feature comprises a sentence feature and a number of sentences; and the video feature comprises a shot feature and a number of shots.

16. The retrieval device of any one of claims 13 to 15, wherein the second determining module is configured to:

detect a person name included in the text;

retrieve, in a database, a portrait of a person corresponding to the person name, and extract an image feature of the portrait to obtain a character node of the person;

determine a semantic tree of the text by parsing, and obtain a motion feature of the person based on the semantic tree to obtain an action node of the person; and

link a character node corresponding to each person with a respective action node,

wherein the character node of the person is represented by the image feature of the portrait, and the action node of the person is represented by the motion feature in the semantic tree.

17. The retrieval device of claim 16, wherein the second determining module is further configured to:

link character nodes linked with a same action node.

~~18. The retrieval device of claim 16 or 17, wherein the second determining module is configured to:~~

~~replace a pronoun in the text with a person name represented by the pronoun.~~

~~19. The retrieval device of any one of claims 13 to 18, wherein the second determining~~

module is configured to:

detect a person in each shot of the at least one video;

extract a human feature and a motion feature of the person;

attach the human feature of the person to a character node of the person, and attach the motion feature of the person to an action node of the person; and

link a character node corresponding to each person with a respective action node.

20. The retrieval device of claim 19, wherein the second determining module is further configured to take a group of persons appearing in a same shot as a same group of persons, and link the character nodes of the persons in the same group of persons two by two.

21. The retrieval device of claim 19 or 20, wherein the second determining module is further configured to:

link one person in one shot with a character node of each person in an adjacent shot of the shot.

22. The retrieval device of any one of claims 13 to 21, wherein the processing module is configured to:

perform weighted sum on the first similarity and the second similarity for each video, to obtain a similarity value for each video; and

determine a video with a highest similarity value as the video matching the retrieval condition.

~~23. The retrieval device of any one of claims 13 to 22, wherein the retrieval device is implemented through a retrieval network, the apparatus further comprises:~~

~~a training module, configured to:~~

~~determine a prediction value of the first similarity between the text and a video in a training sample set, wherein the text is used for representing the retrieval condition;~~

~~determine a prediction value of the second similarity between the first character interaction graph of the text and the second character interaction graph of the video in the training sample set;~~

~~determine a loss of the first similarity according to the prediction value of the first similarity and ground truth of the first similarity;~~

~~determine a loss of the second similarity according to the prediction value of the second similarity and ground truth of the second similarity;~~

~~determine an overall loss value according to the loss of the first similarity and the loss of the second similarity in combination with a loss function; and~~

~~adjust weight parameters of the retrieval network according to the overall loss value.~~

~~24. The retrieval device of claim 23, wherein the retrieval network comprises a first sub-network and a second sub-network, the first sub-network being used for determining the first similarity between the text and the video, and the second sub-network being used for determining the similarity between the first character interaction graph of the text and the second character interaction graph of the video, and~~

~~wherein the training module is configured to:~~

~~adjust, based on the overall loss value, the weight parameters of the first sub-network and the second sub-network.~~

~~25.~~18.A retrieval device, comprising a memory, a processor, and computer programs stored in the memory and executable on the processor, wherein the processor implements the retrieval method of any one of claims 1 to 12 when executing the programs.

~~26.~~19.A storage medium having stored thereon computer programs that, when executed by a processor, cause the processor to perform the retrieval method according to any one of claims 1 to 12.

~~27.~~20.A computer program, comprising computer-readable codes that, when run on an electronic device, cause a processor in the electronic device to perform the retrieval method according to any one of claims 1 to 12.