CLAIMS

1. A cross-modal information retrieval method, comprising:

acquiring first modal information and second modal information;

performing feature fusion on a modal feature of the first modal information and a modal feature of the second modal information to determine a first fused feature corresponding to the first modal information and a second fused feature corresponding to the second modal information; and

determining a similarity between the first modal information and the second modal information based on the first fused feature and the second fused feature.

2. The method of claim 1, wherein performing feature fusion on the modal feature of the first modal information and the modal feature of the second modal information to determine the first fused feature corresponding to the first modal information and the second fused feature corresponding to the second modal information comprises:

determining, based on the modal feature of the first modal information and the modal feature of the second modal information, a fusion threshold parameter for feature fusion of the first modal information and the second modal information; and

performing feature fusion on the modal feature of the first modal information and the modal feature of the second modal information based on the fusion threshold parameter to determine the first fused feature corresponding to the first modal information and the second fused feature corresponding to the second modal information, wherein the fusion threshold parameter is configured for fused features obtained by feature fusion according to a matching degree between features, and the fusion threshold parameter becomes smaller as the matching degree between the features is lower.

3. The method of claim 2, wherein determining, based on the modal feature of the first modal information and the modal feature of the second modal information, the fusion threshold parameter for feature fusion of the first modal information and the second modal information comprises:

determining a second attention feature attended by the first modal information to the second modal information according to the modal feature of the first modal information and the modal feature of the second modal information; and

determining a first fusion threshold parameter corresponding to the first modal information according to the modal feature of the first modal information and the second attention feature.

4. The method of claim 3, wherein the first modal information comprises at least one information unit, and the second modal information comprises at least one information unit; and wherein determining the second attention feature attended by the first modal information to the second modal information comprises:

acquiring a first modal feature of each of the at least one information unit of the first modal information;

acquiring a second modal feature of each of the at least one information unit of the second modal information;

determine an attention weight between each information unit of the first modal information and each information unit of the second modal information according to the first modal feature and the second modal feature; and

determining a second attention feature attended by each information unit of the first modal information to the second modal information according to the attention weight and the second modal feature.

5. The method of claim 2, wherein determining, based on the modal feature of the first modal information and the modal feature of the second modal information, the fusion threshold parameter for feature fusion of the first modal information and the second modal information comprises:

determining a first attention feature attended by the second modal information to the first modal information according to the modal feature of the first modal information and the modal feature of the second modal information; and

determining a second fusion threshold parameter corresponding to the second modal information according to the modal feature of the second modal information and the first attention feature.

6. The method of claim 5, wherein the first modal information comprises at least one information unit, and the second modal information comprises at least one information unit; and wherein determining the first attention feature attended by the second modal information to the first modal information according to the modal feature of the first modal information and the modal feature of the second modal information comprises:

acquiring the first modal feature of each of the at least one information unit of the first modal information;

acquiring the second modal feature of each of the at least one information unit of the second modal information;

determining an attention weight between each information unit of the first modal information and each information unit of the second modal information according to the first modal feature and the second modal feature; and

determining a first attention feature attended by each information unit of the

second modal information to the first modal information according to the attention weight and the first modal feature.

7. The method of claim 2, wherein determining the first fused feature corresponding to the first modal information comprises:

determining a second attention feature attended by the first modal information to the second modal information according to the modal feature of the first modal information and the modal feature of the second modal information; and

performing feature fusion on the modal feature of the first modal information and the second attention feature by using the fusion threshold parameter to determine the first fused feature corresponding to the first modal information.

8. The method of claim 7, wherein performing feature fusion on the modal feature of the first modal information and the second attention feature by using the fusion threshold parameter to determine the first fused feature corresponding to the first modal information comprises:

performing feature fusion on the modal feature of the first modal information and the second attention feature to obtain a first fusion result;

processing, by using the fusion threshold parameter, the first fusion result to obtain a processed first fusion result; and

determining the first fused feature corresponding to the first modal information based on the processed first fusion result and the first modal feature.

9. The method of claim 2, wherein determining the second fused feature corresponding to the second modal information comprises:

determining a first attention feature attended by the second modal information to the first modal information according to the modal feature of the first modal information and the modal feature of the second modal information; and

determining the second fused feature corresponding to the second modal information according to the modal feature of the second modal information and the first attention feature.

10. The method of claim 9, wherein determining the second fused feature corresponding to the second modal information according to the modal feature of the second modal information and the first attention feature comprises:

performing feature fusion on the modal feature of the second modal information and the first attention feature to obtain a second fusion result;

processing, by using the fusion threshold parameter, the second fusion result to obtain a processed second fusion result; and

determining the second fused feature corresponding to the second modal information based on the processed second fusion result and the second modal feature.

11. The method of claim 1, wherein determining the similarity between the first modal information and the second modal information based on the first fused feature and the second fused feature comprises:

determining the similarity between the first modal information and the second modal information based on first attention information of the first fused feature and second attention information of the second fused feature.

12. The method of claim 1, wherein the first modal information comprises information to be retrieved of a first modality, and the second modal information comprises pre-stored information of a second modality; and wherein the method further comprises:

determining the second modal information as a retrieval result of the first modal information in condition that the similarity meets a preset condition.

13. The method of claim 12, wherein the second modal information comprises multiple pieces of second modal information, and wherein determining the second modal information as the retrieval result of the first modal information in condition that the similarity meets the preset condition comprises:

sequencing the multiple pieces of second modal information according to a similarity between the first modal information and each of the multiple pieces of second modal information to obtain a sequencing result;

determining, according to the sequencing result, a second modal information that the similarity meets the preset condition; and

determining the second modal information that the similarity meets the preset condition as the retrieval result of the first modal information,

wherein the preset condition comprises any one of the following conditions that:

the similarity is greater than a preset value; and a rank of the similarity sequenced from low to high is higher than a preset rank.

14. The method of claim 1, wherein the first modal information comprises one piece of modal information in text information or image information; and the second modal information comprises the other piece of modal information in the text information or

the image information.

15. The method of claim 1, wherein the first modal information comprises training sample information of a first modality, the second modal information comprises training sample information of a second modality, and wherein each piece of the training sample information of the first modality and each piece of the training sample information of the second modality form a training sample pair.

16. The method of claim 15, wherein the training sample pair comprises a positive sample pair and a negative sample pair; and wherein the method further comprises:

acquiring a similarity of each training sample pair,

determining a loss in feature fusion of the first modal information and the second modal information according to a similarity of a positive sample pair with a highest matching degree of modal information in positive sample pairs and a similarity of a negative sample pair with a lowest matching degree in negative sample pairs, and

adjusting, according to the loss, a model parameter of a cross-modal information retrieval model that is adopted for the feature fusion of the first modal information and the second modal information.

17. A cross-modal information retrieval device, comprising:

an acquisition module, configured to acquire first modal information and second modal information;

a fusion module, configured to perform feature fusion on a modal feature of the first modal information and a modal feature of the second modal information to

determine a first fused feature corresponding to the first modal information and a second fused feature corresponding to the second modal information; and

a determination module, configured to determine a similarity between the first modal information and the second modal information based on the first fused feature and the second fused feature.

18. A cross-modal information retrieval device, comprising:

a processor; and

a memory, configured to store instructions executable by the processor,

wherein the processor is configured to execute the executable instructions stored in the memory to implement the method of any one of claims 1-16.

19. A non-transitory computer-readable storage medium, having stored therein computer program instructions that, when being executed by a processor, cause the processor to implement the method of any one of claims 1-16.

20. A computer program, comprising computer-executable instructions that, when being executed by a computer, cause the computer to implement the steps of the method of any one of claims 1-16.