

WHAT IS CLAIMED IS:

1. A target matching method, comprising:

respectively extracting a feature vector of each frame in a query image sequence and a feature vector of each frame in a candidate image sequence, wherein the query image sequence comprises a target to be matched;

determining a self-expression feature vector of the query image sequence based on the feature vector of each frame in the query image sequence, and determining a self-expression feature vector of the candidate image sequence based on the feature vector of each frame in the candidate image sequence;

determining a collaborative expression feature vector of the query image sequence based on the feature vector of each frame in the query image sequence and the self-expression feature vector of the candidate image sequence, and determining a collaborative expression feature vector of the candidate image sequence based on the feature vector of each frame in the candidate image sequence and the self-expression feature vector of the query image sequence;

determining a similarity feature vector between the query image sequence and the candidate image sequence based on the self-expression feature vector of the query image sequence, the collaborative expression feature vector of the query image sequence, the self-expression feature vector of the candidate image sequence, and the collaborative expression feature vector of the candidate image sequence; and

determining a matching result between the query image sequence and the candidate image sequence based on the similarity feature vector.

2. The method according to claim 1, wherein the respectively extracting a feature vector of each frame in a query image sequence and a feature vector of each frame in a candidate image sequence comprises:

extracting the feature vector of each frame in the query image sequence and the feature vector of each frame in the candidate image sequence by a first sub-neural network.

3. The method according to claim 1 or 2, wherein after the respectively extracting a feature vector of each frame in a query image sequence and a feature vector of each frame in a candidate

image sequence, the method further comprises:

performing dimension reduction on the feature vector of each frame in the query image sequence and the feature vector of each frame in the candidate image sequence by a first fully connected layer of the first sub-neural network to obtain a first dimension-reduced feature vector of each frame in the query image sequence and a first dimension-reduced feature vector of each frame in the candidate image sequence.

4. The method according to claim 3, wherein the determining a self-expression feature vector of the query image sequence based on the feature vector of each frame in the query image sequence, and determining a self-expression feature vector of the candidate image sequence based on the feature vector of each frame in the candidate image sequence comprises:

inputting the feature vector of each frame in the query image sequence and the first dimension-reduced feature vector of each frame in the query image sequence into a second sub-neural network to determine the self-expression feature vector of the query image sequence; and

inputting the feature vector of each frame in the candidate image sequence and the first dimension-reduced feature vector of each frame in the candidate image sequence into the second sub-neural network to determine the self-expression feature vector of the candidate image sequence.

5. The method according to claim 4, wherein the inputting the feature vector of each frame in the query image sequence and the first dimension-reduced feature vector of each frame in the query image sequence into a second sub-neural network to determine the self-expression feature vector of the query image sequence comprises:

performing dimension reduction on the feature vector of each frame in the query image sequence by a second fully connected layer of the second sub-neural network to obtain a second dimension-reduced feature vector of each frame in the query image sequence;

subjecting the second dimension-reduced feature vector of each frame in the query image sequence to average pooling in a time dimension to obtain a global feature vector of the query image sequence; and

determining the self-expression feature vector of the query image sequence based on the second dimension-reduced feature vector of each frame in the query image sequence, the global feature vector of the query image sequence, and the first dimension-reduced feature vector of each frame in the query image sequence.

6. The method according to claim 4, wherein the inputting the feature vector of each frame in the candidate image sequence and the first dimension-reduced feature vector of each frame in the candidate image sequence into the second sub-neural network to obtain the self-expression feature vector of the candidate image sequence comprises:

performing dimension reduction on the feature vector of each frame in the candidate image sequence by the second fully connected layer of the second sub-neural network to obtain a second dimension-reduced feature vector of each frame in the candidate image sequence;

subjecting the second dimension-reduced feature vector of each frame in the candidate image sequence to average pooling in the time dimension to obtain a global feature vector of the candidate image sequence; and

determining the self-expression feature vector of the candidate image sequence based on the second dimension-reduced feature vector of each frame in the candidate image sequence, the global feature vector of the candidate image sequence, and the first dimension-reduced feature vector of each frame in the candidate image sequence.

7. The method according to claim 5, wherein the determining the self-expression feature vector of the query image sequence based on the second dimension-reduced feature vector of each frame in the query image sequence, the global feature vector of the query image sequence, and the first dimension-reduced feature vector of each frame in the query image sequence comprises:

calculating a correlation between the second dimension-reduced feature vector of each frame in the query image sequence and the global feature vector of the query image sequence through a parameter-less correlation function to obtain a first correlation weight of each frame in the query image sequence; and

weighting the first dimension-reduced feature vector of each frame in the query image sequence based on the first correlation weight of each frame in the query image sequence to

obtain the self-expression feature vector of the query image sequence.

8. The method according to claim 6, wherein the determining the self-expression feature vector of the candidate image sequence based on the second dimension-reduced feature vector of each frame in the candidate image sequence, the global feature vector of the candidate image sequence, and the first dimension-reduced feature vector of each frame in the candidate image sequence comprises:

calculating a correlation between the second dimension-reduced feature vector of each frame in the candidate image sequence and the global feature vector of the candidate image sequence through a parameter-less correlation function to obtain a first correlation weight of each frame in the candidate image sequence; and

weighting the first dimension-reduced feature vector of each frame in the candidate image sequence based on the first correlation weight of each frame in the candidate image sequence to obtain the self-expression feature vector of the candidate image sequence.

9. The method according to claim 7 or 8, wherein the first correlation weight comprises a first normalized correlation weight, the first normalized correlation weight being obtained by normalizing the first correlation weight.

10. The method according to any one of claims 3 to 9, wherein the determining a collaborative expression feature vector of the query image sequence based on the feature vector of each frame in the query image sequence and the self-expression feature vector of the candidate image sequence, and determining a collaborative expression feature vector of the candidate image sequence based on the feature vector of each frame in the candidate image sequence and the self-expression feature vector of the query image sequence comprises:

inputting the feature vector of each frame in the query image sequence, the first dimension-reduced feature vector of each frame in the query image sequence, and the self-expression feature vector of the candidate image sequence into a third sub-neural network to obtain the collaborative expression feature vector of the query image sequence; and

inputting the feature vector of each frame in the candidate image sequence, the first

dimension-reduced feature vector of each frame in the candidate image sequence, and the self-expression feature vector of the query image sequence into the third sub-neural network to obtain the collaborative expression feature vector of the candidate image sequence.

11. The method according to claim 10, wherein the inputting the feature vector of each frame in the query image sequence, the first dimension-reduced feature vector of each frame in the query image sequence, and the self-expression feature vector of the candidate image sequence into a third sub-neural network to obtain the collaborative expression feature vector of the query image sequence comprises:

performing dimension reduction on the feature vector of each frame in the query image sequence by a third fully connected layer of the third sub-neural network to obtain a third dimension-reduced feature vector of each frame in the query image sequence; and

obtaining the collaborative expression feature vector of the query image sequence based on the third dimension-reduced feature vector of each frame in the query image sequence, the self-expression feature vector of the candidate image sequence, and the first dimension-reduced feature vector of each frame in the query image sequence;

the inputting the feature vector of each frame in the candidate image sequence, the first dimension-reduced feature vector of each frame in the candidate image sequence, and the self-expression feature vector of the query image sequence into the third sub-neural network to obtain the collaborative expression feature vector of the candidate image sequence comprises:

performing dimension reduction on the feature vector of each frame in the candidate image sequence by the third fully connected layer of the third sub-neural network to obtain a third dimension-reduced feature vector of each frame in the candidate image sequence; and

obtaining the collaborative expression feature vector of the candidate image sequence based on the third dimension-reduced feature vector of each frame in the candidate image sequence, the self-expression feature vector of the query image sequence, and the first dimension-reduced feature vector of each frame in the candidate image sequence.

12. The method according to claim 11, wherein the obtaining the collaborative expression feature vector of the query image sequence based on the third dimension-reduced feature vector of each

frame in the query image sequence, the self-expression feature vector of the candidate image sequence, and the first dimension-reduced feature vector of each frame in the query image sequence comprises:

calculating a correlation between the third dimension-reduced feature vector of each frame in the query image sequence and the self-expression feature vector of the candidate image sequence through a parameter-less correlation function to obtain a second correlation weight of each frame in the query image sequence; and

weighting the first dimension-reduced feature vector of each frame in the query image sequence based on the second correlation weight of each frame in the query image sequence to obtain the collaborative expression feature vector of the query image sequence.

13. The method according to claim 11, wherein the obtaining the collaborative expression feature vector of the candidate image sequence based on the third dimension-reduced feature vector of each frame in the candidate image sequence, the self-expression feature vector of the query image sequence, and the first dimension-reduced feature vector of each frame in the candidate image sequence comprises:

calculating a correlation between the third dimension-reduced feature vector of each frame in the candidate image sequence and the self-expression feature vector of the query image sequence through a parameter-less correlation function to obtain a second correlation weight of each frame in the candidate image sequence; and

weighting the first dimension-reduced feature vector of each frame in the candidate image sequence based on the second correlation weight of each frame in the candidate image sequence to obtain the collaborative expression feature vector of the candidate image sequence.

14. The method according to claim 12 or 13, wherein the second correlation weight comprises a second normalized correlation weight, the second normalized correlation weight being obtained by normalizing the second correlation weight.

15. The method according to any one of claims 1 to 14, wherein the determining a similarity feature vector between the query image sequence and the candidate image sequence based on the

self-expression feature vector of the query image sequence, the collaborative expression feature vector of the query image sequence, the self-expression feature vector of the candidate image sequence, and the collaborative expression feature vector of the candidate image sequence comprises:

calculating a difference between the self-expression feature vector of the query image sequence and the collaborative expression feature vector of the candidate image sequence to obtain a first difference vector;

calculating a difference between the self-expression feature vector of the candidate image sequence and the collaborative expression feature vector of the query image sequence to obtain a second difference vector; and

obtaining the similarity feature vector between the query image sequence and the candidate image sequence based on the first difference vector and the second difference vector.

16. The method according to claim 15, wherein the obtaining the similarity feature vector between the query image sequence and the candidate image sequence based on the first difference vector and the second difference vector comprises:

calculating a sum of the first difference vector and the second difference vector to obtain the similarity feature vector between the query image sequence and the candidate image sequence; or,

calculating a product of elements of corresponding bits of the first difference vector and the second difference vector to obtain the similarity feature vector between the query image sequence and the candidate image sequence.

17. The method according to any one of claims 1 to 16, wherein the determining a matching result between the query image sequence and the candidate image sequence based on the similarity feature vector comprises:

inputting the similarity feature vector between the query image sequence and the candidate image sequence into a fourth fully connected layer to obtain a matching score between the query image sequence and the candidate image sequence; and

determining the matching result between the query image sequence and the candidate

image sequence based on the matching score between the query image sequence and the candidate image sequence.

18. The method according to claim 17, wherein after the obtaining a matching score between the query image sequence and the candidate image sequence, the method further comprises:

optimizing, based on the matching score between the query image sequence and the candidate image sequence, network parameters by using a same pair of labeled data and a binary cross-entropy loss function.

19. The method according to any one of claims 1 to 18, wherein before extracting the feature vector of each frame in the query image sequence, the method further comprises:

segmenting a query video into multiple query image sequences; and

segmenting a candidate video into multiple candidate image sequences;

after the determining a matching result between the query image sequence and the candidate image sequence, the method further comprises:

determining a matching result between the query video and the candidate video based on matching results between the query image sequences of the query video and the candidate image sequences of the candidate video.

20. The method according to claim 19, wherein the segmenting a query video into multiple query image sequences comprises:

segmenting the query video into the multiple query image sequences according to a preset sequence length and a preset step size, wherein a length of each of the multiple query image sequences is equal to the preset sequence length, and a number of overlapping images between adjacent query image sequences is equal to a difference between the preset sequence length and the preset step size;

the segmenting a candidate video into multiple candidate image sequences comprises:

segmenting the candidate video into the multiple candidate image sequences according to a preset sequence length and a preset step size, wherein a length of each of the multiple candidate image sequence is equal to the preset sequence length, and a number of overlapping

images between adjacent candidate image sequences is equal to the difference between the preset sequence length and the preset step size.

21. The method according to claim 19 or 20, wherein the determining a matching result between the query video and the candidate video based on matching results between the query image sequences of the query video and the candidate image sequences of the candidate video comprises:

- determining matching scores, each being a matching score between the query image sequences of the query video and a respective one of the candidate image sequences of the candidate video;

- calculating an average value of top N matching scores among the matching scores between the query image sequences of the query video and the candidate image sequences of the candidate video to obtain a matching score between the query video and the candidate video, wherein N is a positive integer; and

- determining the matching result between the query video and the candidate video based on the matching score between the query video and the candidate video.

22. A target matching apparatus, wherein modules in the target matching apparatus are configured to perform the method as claimed in any one of the claims 1 to 21.

23. An electronic device, comprising:

- a processor; and

- a memory configured to store processor-executable instructions;

- wherein the processor is configured to execute the method according to any one of claims 1 to 21.

24. A computer-readable storage medium, having computer program instructions stored thereon, wherein when the computer program instructions are executed by a processor, the method according to any one of claims 1 to 21 is implemented.