

## CLAIMS

### 1. A pedestrian re-identification method, comprising:

obtaining at least one candidate video and a target video containing a target pedestrian;

encoding each target video segment in the target video and each candidate video segment in the at least one candidate segment respectively;

determining, according to encoding results, a score of similarity between the each target video segment and the each candidate video segment, the score of similarity being used for representing a degree of similarity between pedestrian features in the target video segment and pedestrian features in the candidate video segment; and

performing, according to the score of similarity, pedestrian re-identification on the at least one candidate video;

wherein the encoding each target video segment in the target video and each candidate video segment in the at least one candidate segment respectively comprises:

obtaining a first target feature vector and a second target feature vector of each target video frame in the each target video segment as well as an index feature vector of the each target video segment;

obtaining a first candidate feature vector and a second candidate feature vector of each candidate video frame in the each candidate video segment;

generating, according to the index feature vector, the first target feature vector, and the first candidate feature vector, an attention weight vector; and

obtaining, according to the attention weight vector, the second target feature vector, and the second candidate feature vector, an encoding result of the each target video segment and an encoding result of the each candidate video segment.

2. The method according to claim 1, wherein the obtaining a first target feature vector and a second target feature vector of each target video frame in the each target video segment as well as an index feature vector of the each target video segment, and obtaining a first candidate feature vector and a second candidate feature vector of each candidate video frame in the each candidate video segment comprises:

extracting an image feature vector of the each target video frame and an image feature vector of the each candidate video frame respectively;

generating, according to the image feature vector of the each target video frame, the first target feature vector and the second target feature vector of the each target video frame as well as the index feature vector of the each target video frame; and

generating, according to the image feature vector of the each candidate video frame, the first candidate feature vector and the second candidate feature vector of the each candidate video frame.

3. The method according to claim 1 or 2, wherein the generating, according to the index feature vector, the first target feature vector, and the first candidate feature vector, an attention weight vector comprises:

generating, according to the index feature vector and the first target feature vector, a target attention weight vector of the each target video frame; and

generating, according to the index feature vector and the first candidate feature vector, a candidate attention weight vector of the each candidate video frame.

4. The method according to claim 3, wherein the generating, according to the index feature vector and the first target feature vector, a target attention weight vector of the each target video frame comprises:

generating, according to the index feature vector and the first target feature vector of

the each target video frame, a target heat map of the each target video frame; and

performing normalization processing on the target heat map to obtain the target attention weight vector of the each target video frame;

and/or

the generating, according to the index feature vector and the first candidate feature vector, a candidate attention weight vector of the each candidate video frame comprises:

generating, according to the index feature vector and the first candidate feature vector of the each candidate video frame, a candidate heat map of the each candidate video frame; and

performing normalization processing on the candidate heat map to obtain the candidate attention weight vector of the each candidate video frame.

5. The method according to any one of claims 1 to 4, wherein the obtaining, according to the attention weight vector, the second target feature vector, and the second candidate feature vector, an encoding result of the each target video segment and an encoding result of the each candidate video segment comprises:

obtaining, according to the target attention weight vector and the second target feature vector of the target video frame, the encoding result of the each target video segment; and

obtaining, according to the candidate attention weight vector and the second candidate feature vector of the candidate video frame, the encoding result of the each candidate video segment.

6. The method according to claim 5, wherein the obtaining, according to the target attention weight vector and the second target feature vector of the each target video frame, the encoding result of the each target video segment comprises:

multiplying the target attention weight vector of the each target video frame by the second target feature vector of the each target video frame;

adding, in time dimension, multiplication result of the each target video frame; and

obtain the encoding result of the each target video segment;

and/or

the obtaining, according to the candidate attention weight vector and the second candidate feature vector of the each candidate video frame, the encoding result of the each candidate video segment comprises:

multiplying the candidate attention weight vector of the each candidate video frame by the second candidate feature vector of the each candidate video frame;

adding, in time dimension, multiplication result of the each candidate video frame; and

obtaining the encoding result of the each candidate video segment.

7. The method according to any one of claims 1 to 6, wherein the determining, according to encoding results, a score of similarity between the each target video segment and the each candidate video segment comprises:

performing a subtraction operation on an encoding result of the each target video segment and an encoding result of the each candidate video segment in sequence;

performing, in each dimension, a square operation on a result of the subtraction operation;

performing a full connection operation on a feature vector obtained by the square operation to obtain a two-dimensional feature vector;

performing a normalization operation on the two-dimensional feature vector; and

obtaining the score of similarity between the each target video segment and the each

candidate video segment.

8. The method according to any one of claims 1 to 7, wherein the performing, according to the score of similarity, pedestrian re-identification on the at least one candidate video comprises:

for each candidate video in the at least one candidate video, taking a sum of a preset proportion of top scores in scores of similarity between each candidate video segment of the candidate video and the each target video segment as a score of similarity of the candidate video;

ranking the score of similarity of the each candidate video in a descending order; and

determining one or several candidate videos ranked top as videos containing a same target pedestrian as the target video.

9. A pedestrian re-identification apparatus, comprising:

an obtaining module, configured to obtain at least one candidate video and a target video containing a target pedestrian;

an encoding module, configured to encode each target video segment in the target video and each candidate video segment in the at least one candidate segment respectively;

a determining module, configured to determine, according to encoding results, a score of similarity between the each target video segment and the each candidate video segment, the score of similarity being used for representing a degree of similarity between pedestrian features in the target video segment and pedestrian features in the candidate video segment; and

an identifying module, configured to perform, according to the score of similarity, pedestrian re-identification on the at least one candidate video;

wherein the encoding module comprises:

a feature vector obtaining module, configured to obtain a first target feature vector and a second target feature vector of each target video frame in the each target video segment as well as an index feature vector of the each target video segment, and obtain a first candidate feature vector and a second candidate feature vector of each candidate video frame in the each candidate video segment;

a weight vector generating module, configured to generate, according to the index feature vector, the first target feature vector, and the first candidate feature vector, an attention weight vector; and

an encoding result obtaining module, configured to obtain, according to the attention weight vector, the second target feature vector, and the second candidate feature vector, an encoding result of the each target video segment and an encoding result of the each candidate video segment.

10. The apparatus according to claim 9, wherein the feature vector obtaining module is configured to extract an image feature vector of the each target video frame and an image feature vector of the each candidate video frame respectively; and generate, according to the image feature vector of the each target video frame, the first target feature vector and the second target feature vector of the each target video frame as well as an index feature vector of the each target video frame, and generate, according to the image feature vector of the each candidate video frame, the first candidate feature vector and the second candidate feature vector of the each candidate video frame.

11. The apparatus according to claim 9 or 10, wherein the weight vector generating module is configured to generate, according to the index feature vector and the first target feature vector, a target attention weight vector of the each target video frame, and generate, according to the index feature vector and the first candidate feature vector, a candidate attention weight vector of the each candidate video frame.

12. The apparatus according to claim 11, wherein the weight vector generating module is configured to generate, according to the index feature vector and the first target feature vector of the each target video frame, a target heat map of the each target video frame; and perform normalization processing on the target heat map to obtain the target attention weight vector of the each target video frame; and/or generate, according to the index feature vector and the first candidate feature vector of the each candidate video frame, a candidate heat map of the each candidate video frame; and perform normalization processing on the candidate heat map to obtain the candidate attention weight vector of the each candidate video frame.

13. The apparatus according to any one of claims 9 to 12, wherein the encoding result obtaining module is configured to obtain, according to the target attention weight vector and the second target feature vector of the each target video frame, the encoding result of the each target video segment, and obtain, according to the candidate attention weight vector and the second candidate feature vector of the each candidate video frame, the encoding result of the each candidate video segment.

14. The apparatus according to claim 13, wherein the encoding result obtaining module is configured to multiply the target attention weight vector of the each target video frame by the second target feature vector of the each target video frame; and add, in time dimension, multiplication result of the each target video frame; and obtain the encoding result of the each target video segment; and/or multiply the candidate attention weight vector of the each candidate video frame by the second candidate feature vector of the each candidate video frame; and add, in time dimension, multiplication result of the each candidate video frame; and obtain the encoding result of the each candidate video segment.

15. The apparatus according to any one of claims 9 to 14, wherein the determining module is configured to perform a subtraction operation on an encoding result of the each target video segment and an encoding result of the each candidate video segment in sequence; perform, in each dimension, a square operation on a result of the subtraction operation; perform a full

connection operation on a feature vector obtained by the square operation to obtain a two-dimensional feature vector; and perform a normalization operation on the two-dimensional feature vector; and obtain the score of similarity between the each target video segment and the each candidate video segment.

16. The apparatus according to any one of claims 9 to 15, wherein the identifying module is configured to, for each candidate video in the at least one candidate video, take a sum of a preset proportion of top scores in scores of similarity between each candidate video segment of the candidate video and the each target video segment as a score of similarity of the candidate video; rank the score of similarity of the each candidate video in a descending order; and determine one or several candidate videos ranked top as videos containing a same target pedestrian as the target video.

17. An electronic device, comprising: a processor and a memory,

wherein the memory is configured to store at least one executable instruction, and the executable instruction causes the processor to perform the pedestrian re-identification method according to any one of claims 1 to 8.

18. A computer-readable storage medium having computer programs stored thereon, wherein when the computer programs are executed by a processor, the pedestrian re-identification method according to any one of claims 1 to 8 is implemented.

19. A computer program product, comprising at least one executable instruction, wherein when the executable instruction is executed, the pedestrian re-identification method according to any one of claims 1 to 8 is implemented.