CLAIMS

1. A video behavior recognition method, performed by a computer device, comprising:

extracting a video image feature from at least two target video images;

performing contribution adjustment on a spatial feature of the video image feature to obtain an intermediate image feature;

fusing, based on priori information, a temporal feature of the intermediate image feature and a cohesive feature corresponding to the temporal feature to obtain a fused feature, the priori information being obtained according to change information of the intermediate image feature in a temporal dimension, and the cohesive feature being obtained by performing attention processing on the temporal feature;

performing temporal feature contribution adjustment on the fused feature to obtain a behavior recognition feature; and

performing video behavior recognition based on the behavior recognition feature,

wherein the fusing, based on priori information, a temporal feature of the intermediate image feature and a cohesive feature corresponding to the temporal feature to obtain a fused feature comprises:

determining the priori information comprising a first priori parameter and a second priori parameter;

performing temporal feature extraction on the intermediate image feature to obtain the temporal feature of the intermediate image feature, and performing attention processing on the temporal feature based on an attention mechanism algorithm to obtain the cohesive feature corresponding to the temporal feature;

performing weighting on the temporal feature based on the first priori parameter to obtain the weighted temporal feature;

performing, based on the second priori parameter, weighting on the cohesive feature corresponding to the temporal feature to obtain the weighted cohesive feature; and

fusing the weighted temporal feature and the weighted cohesive feature to obtain the fused feature.

2. The method according to claim 1, wherein the performing contribution adjustment on a spatial feature of the video image feature to obtain an intermediate image feature comprises:

performing spatial feature extraction on the video image feature to obtain the spatial feature of the video image feature, and

performing contribution adjustment on the spatial feature based on a spatial structural parameter of a structural parameter to obtain the intermediate image feature, the structural parameter being obtained by training based on a video image sample carrying a behavior label, and wherein

the performing temporal feature contribution adjustment on the fused feature to obtain a behavior recognition feature comprises:

performing contribution adjustment on the fused feature based on a temporal structural parameter of the structural parameter to obtain the behavior recognition feature.

3. The method according to claim 2, further comprising:

determining a to-be-trained structural parameter;

performing contribution adjustment on a spatial sample feature of a video image sample feature based on the spatial structural parameter of the to-be-trained structural parameter to obtain an intermediate sample feature, the video image sample feature being extracted from the video image sample;

fusing, based on priori sample information, a temporal sample feature of the intermediate sample feature and a cohesive sample feature corresponding to the temporal sample feature to obtain a fused sample feature, the cohesive sample feature being obtained by performing attention processing on the temporal sample feature, and the priori sample information being obtained according to change information of the intermediate sample feature in a temporal dimension;

performing contribution adjustment on the fused sample feature based on the temporal structural parameter of the to-be-trained structural parameter to obtain the behavior recognition sample feature; and

performing video behavior recognition based on the behavior recognition sample feature, updating the to-be-trained structural parameter according to the behavior recognition result and the behavior label corresponding to the video image sample, and performing training continually until the training is ended to obtain the structural parameter.

4. The method according to claim 3, wherein the method is implemented by a video behavior recognition model, and the to-be-trained structural parameter is a parameter of the video behavior recognition model during training, and wherein

the updating the to-be-trained structural parameter according to the behavior recognition result and the behavior label corresponding to the video image sample, and performing training continually until the training is ended to obtain the structural parameter comprises:

obtaining a behavior recognition result outputted from the video behavior recognition model;

determining a difference between the behavior recognition result and the behavior label corresponding to the video image sample;

updating a model parameter of the video behavior recognition model and the to-be-trained structural parameter according to the difference; and

performing training continually based on the updated video behavior recognition model until the training is ended, and obtaining the structural parameter of the trained video behavior recognition model.

5. The method according to claim 3, wherein the updating the to-be-trained structural parameter according to the behavior recognition result and the behavior label corresponding to the video image sample, and performing training continually until the training is ended to obtain the structural parameter comprises:

determining a behavior recognition loss between the behavior recognition result and the behavior label corresponding to the video image sample;

obtaining a reward according to the behavior recognition loss and a previous behavior recognition loss; and

updating the to-be-trained structural parameter according to the reward, and performing training continually based on the updated to-be-trained structural parameter to obtain the structural parameter until a target function satisfies an ending condition, the target function being obtained based on the reward in a training process.

6. The method according to claim 5, wherein the updating the to-be-trained structural parameter according to the reward comprises:

updating a model parameter of a policy gradient network model according to the reward; and

updating the to-be-trained structural parameter based on the updated policy gradient network model.

7. The method according to claim 6, wherein the updating the to-be-trained structural parameter based on the updated policy gradient network model comprises:

performing, by the updated policy gradient network model, structural parameter prediction based on the updated model parameter and the to-be-trained structural parameter to obtain a predicted structural parameter; and

obtaining, according to the predicted structural parameter, the structural parameter after the to-be-trained structural parameter is updated.

8. The method according to claim 1, further comprising:

determining a similarity of intermediate image features in the temporal dimension; and

correcting initial priori information based on the similarity to obtain the priori information.

9. The method according to claim 8, wherein the correcting initial priori information based on the similarity to obtain the priori information comprises:

dynamically adjusting the similarity according to the first initial priori parameter, the second initial priori parameter, and a preset threshold;

correcting the first initial priori parameter and the second initial priori parameter based on the dynamically adjusted similarity to obtain a first priori parameter and a second priori parameter; and

obtaining the priori information according to the first priori parameter and the second priori parameter.

10. The method according to claim 1, further comprising:

determining a current base vector;

performing feature reconstruction on the temporal feature of the intermediate image feature based on the current base vector to obtain a reconstructed feature;

generating a next base vector for attention processing according to the reconstructed feature and the temporal feature; and

obtaining the cohesive feature corresponding to the temporal feature according to the next base for attention processing, the base vector, and the temporal feature.

11. The method according to claim 10, wherein the generating a next base vector for attention processing according to the reconstructed feature and the temporal feature comprises:

fusing the reconstructed feature and the temporal feature to generate an attention feature;

performing regularization on the attention feature to obtain a regularized feature; and

performing moving average updating on the regularized feature to generate the next base vector for attention processing.

12. The method according to claim 10, wherein the current base vector comprises a data size in batch processing, a number of channels of the intermediate image feature, and a dimension of the base vector, and wherein

the performing feature reconstruction on the temporal feature of the intermediate image feature based on the current base vector to obtain a reconstructed feature comprises:

performing matrix multiplication and normalized mapping on the current base vector and the temporal feature of the intermediate image feature in sequence to obtain the reconstructed feature;

the generating a next base vector for attention processing according to the reconstructed feature and the temporal feature comprises:

performing matrix multiplication on the reconstructed feature and the temporal feature to obtain the next base vector for attention processing; and

the obtaining the cohesive feature corresponding to the temporal feature according to the next base vector for attention processing, the base vector, and the temporal feature comprises:

fusing the next base vector for attention processing, the base vector, and the temporal feature to obtain the cohesive feature corresponding to the temporal feature.

13. The method according to claim 1, wherein the priori information comprises the first priori parameter and the second priori parameter, and wherein

the performing, based on the priori information, weighting and fusion on the temporal feature and the cohesive feature corresponding to the temporal feature to obtain the fused feature comprises:

performing weighting on the temporal feature based on the first priori parameter to obtain the weighted temporal feature;

performing, based on the second priori parameter, weighting on the cohesive feature corresponding to the temporal feature to obtain the weighted cohesive feature; and

fusing the weighted temporal feature and the weighted cohesive feature to obtain the fused feature.

14. The method according to claim 1, wherein before the fusing, based on priori information, a temporal feature of the intermediate image feature and a cohesive feature corresponding to the temporal feature to obtain a fused feature, the method further comprises:

performing normalization on the intermediate image feature to obtain a normalized feature, and

performing nonlinear mapping according to the normalized feature to obtain a mapped intermediate image feature, and wherein

the fusing, based on priori information, a temporal feature of the intermediate image feature and a cohesive feature corresponding to the temporal feature to obtain a fused feature comprises:

fusing, based on the priori information, the temporal feature of the mapped intermediate image feature and the cohesive feature corresponding to the temporal feature to obtain the fused feature, the priori information being obtained according to change information of the mapped intermediate image feature in a temporal dimension.

15. The method according to claim 14, wherein the performing normalization on the intermediate image feature to obtain a normalized feature comprises:

performing normalization on the intermediate image feature through a batch normalization

layer structure to obtain the normalized feature, and wherein

the performing nonlinear mapping according to the normalized feature to obtain a mapped intermediate image feature comprises:

performing nonlinear mapping on the normalized feature through an activation function to obtain the mapped intermediate image feature.

16. A video behavior recognition apparatus, comprising:

a video image feature extraction module, configured to extract a video image feature from at least two target video images;

a spatial feature contribution adjustment module, configured to perform contribution adjustment on a spatial feature of the video image feature to obtain an intermediate image feature;

a feature fusion module, configured to fuse, based on priori information, a temporal feature of the intermediate image feature and a cohesive feature corresponding to the temporal feature to obtain a fused feature, the priori information being obtained according to change information of the intermediate image feature in a temporal dimension, and the cohesive feature being obtained by performing attention processing on the temporal feature;

a temporal feature contribution adjustment module, configured to perform temporal feature contribution adjustment on the fused feature to obtain a behavior recognition feature; and

a video behavior recognition module, configured to perform video behavior recognition based on the behavior recognition feature,

wherein the feature fusion module is further configured to:

determine the priori information comprising a first priori parameter and a second priori parameter;

perform temporal feature extraction on the intermediate image feature to obtain the temporal feature of the intermediate image feature, and perform attention processing on the temporal feature based on an attention mechanism algorithm to obtain a cohesive feature corresponding to the temporal feature;

perform weighting on the temporal feature based on the first priori parameter to obtain the weighted temporal feature;

perform, based on the second priori parameter, weighting on the cohesive feature

corresponding to the temporal feature to obtain the weighted cohesive feature; and

fuse the weighted temporal feature and the weighted cohesive feature to obtain the fused feature.

17. A computer device, comprising a memory and a processor, the memory storing computer-readable instructions, the processor, when executing the computer-readable instructions, implementing steps of the method according to any one of claims 1 to 15.

18. A computer-readable storage medium, storing computer-readable instructions, the computer-readable instructions, when executed by a processor, implementing steps of the method according to any one of claims 1 to 15.

19. A computer program product, comprising computer-readable instructions, the computer-readable instructions, when executed by a processor, implementing steps of the method according to any one of claims 1 to 15.