CLAIMS

What is claimed is:

1. A method for processing a video stream, wherein the method comprises:

obtaining a video stream to be processed;

extracting features of S image frames in the video stream, and generating S N-dimensional feature vectors;

arranging the S N-dimensional feature vectors in a chronological order corresponding to the image frames to generate a feature matrix with N rows and M columns, wherein S is smaller than M, wherein arranging comprises supplementing M—S feature vectors to the S N-dimensional feature vectors by using an interpolation method to generate the feature matrix , and wherein, for a first interpolation position, the interpolation method comprises one of:

i) using a feature vector nearest to the first interpolation position as a feature vector to be interpolated into the first interpolation position,

ii) performing bilinear interpolation on elements corresponding to r feature vectors neighboring to the first interpolation position, and

iii) determining, based on an interpolation result, a feature vector to be interpolated into the first interpolation position; and

processing the feature matrix by using a pre-trained convolutional neural network, to obtain a processing result of the video stream , wherein the processing result includes an output video stream in a time range t of an extracted key frame corresponding to an image frame, wherein the convolutional neural network is pre-trained in the following way:

obtaining a plurality of sample video streams, wherein each sample video stream corresponds to a corresponding sample feature matrix with N rows and at least S columns that is generated by extracting features from S image frames in the video stream,

obtaining a time period label that labels a position of a key frame, wherein the time period label comprises a two-dimensional array used to describe a start moment and an end moment of a time period, wherein the start moment and the end moment are relative moments that are calculated starting from a 1st image frame of the video stream, and

using all the sample feature matrices and time period labels to train the convolutional neural network so that a result of a loss prediction function of the trained model corresponding to the samples decreases compared to a result of a loss prediction function used before the training.

2. The method according to claim 1, wherein the S image frames comprise a first image frame, and extracting features of S image frames in the video stream comprises:

determining a change feature of the first image frame based on a comparison between the current image frame and a previous image frame, wherein the change feature describes a change between the current image frame and the previous image frame

3. The method according to claim 2, wherein the S image frames comprise the current image frame, and extracting features of S image frames in the video stream further comprises:

using an optical flow change from the previous image frame to the current image frame as the change feature.

4. The method according to claim 2, wherein the S image frames comprise the current image frame, and extracting features of S image frames in the video stream further comprises:

determining, as the change feature of the current image frame, a structural similarity index measurement, SSIM index between the current image frame and the previous image frame that is determined by using a structural similarity model.

5. The method according to claim 2, wherein the S image frames comprise the current image frame, and extracting features of S image frames in the video stream further comprises:

determining a projection matrix for converting the previous image frame into the current image frame; and

using the projection matrix as the change feature.

6. The method according to any one of claims 2 to 5, wherein when the current

image frame is the 1st image frame in the video stream, the method further comprises: determining that the change feature comprises a predetermined value, or determining that the previous image frame and the current image frame are a same image frame.

7. The method according to claim 1, wherein the extracting features of S image frames in the video stream comprises:

performing target identification on the current image frame in the S image frames, to determine a target feature of the current image frame.

8. The method according to claim 7, wherein the video stream is a video stream for a damaged vehicle, and the target feature comprises one or more of a damaged part, a damaged material, and a damage category.

9. The method according to claim 1, wherein the extracting features of S image frames in the video stream comprises:

detecting one or more of the following items for the current image frame in the S image frames: whether a target is contained, whether illumination is sufficient, and whether a shooting angle is a predetermined angle; and

using a detection result as a corresponding feature of the current image frame.

10. The method according to claim 1, wherein when a plurality of features are extracted from the first image frame in the S image frames, the generating S N-dimensional feature vectors comprises:

combining feature values corresponding to the plurality of features of the first image frame, to generate one N-dimensional feature vector.

11. The method according to any one of claims 1 to 9, further including post-processing the output video stream.

12. The method according to claim 1, wherein the convolutional neural network performs convolution processing on the feature matrix by using at least one convolution kernel, wherein the convolution kernel sequentially processes a plurality of feature values of each time dimension in the feature matrix.

13. A computer-readable storage medium, wherein the computer-readable storage medium stores a computer program, and when the computer program is run on a computer, the computer is enabled to perform the method according to any one of claims 1 to 12.

14. A computing device, comprising a memory and a processor, wherein the memory stores executable code, and when the processor executes the executable code, the method according to any one of claims 1 to 12 is implemented.