## CLAIMS

**What is claimed is:**

1. A method for detecting and tracking a target, comprising:

acquiring a target tracking result of each video frame of a plurality of video frames received from a video stream by inputting the each video frame of the plurality of video frames into a tracking network, wherein the tracking network is configured to track a detected target in a video frame;

in response to the each video frame being a key frame, acquiring a target detection result output by a detection network upon receiving a last delay frame in a specified number of delay frames by simultaneously inputting the key frame into the detection network and controlling the detection network to operate during a period of receiving the specified number of delay frames, wherein the key frame is a video frame for target detection in the video stream, and the specified number of delay frames are successive video frames following the key frame; and

generating a final target tracking result based on the target detection result and a target tracking result of the last delay frame, wherein a tracked target list in the final target tracking result is updated based on a detected target list in the target detection result, and target tracking for subsequent frames is performed based on the updated tracked target list.

2. The method according to claim 1, wherein acquiring the target detection result output by the detection network upon receiving the last delay frame in the specified number of delay frames by simultaneously inputting the key frame into the detection network and controlling the detection network to operate during the period of receiving the specified number of delay frames in response to the each video frame being the key frame comprises:

in response to the each video frame being the key frame, determining the specified number of delay frames and determining a plurality of computing nodes of the detection network based on the specified number, or, determining the plurality of computing nodes of the detection network and determining the specified number of delay frames based on the plurality of computing nodes of the detection network; and

upon inputting the key frame into the detection network, controlling the plurality computing nodes of the detection network to operate sequentially in response to sequentially receiving the specified number of delay frames, and acquiring the target detection result output by the detection network upon receiving the last delay frame.

3. The method according to claim 2, wherein

the detection network is a cascaded detection network comprising many levels of sub-networks; and

determining the plurality of computing nodes of the detection network in response to the each video frame being the key frame comprises:

acquiring a first number of the sub-networks of the cascaded detection network; and

setting an input node of each of the sub-networks in the first number of the sub-networks as the computing node.

4. The method according to claim 3, wherein determining the specified number of delay frames based on the plurality of computing nodes of the detection network comprises:

determining, based on the first number, a plurality of successive video frames following the key frame as the specified number of delay frames.

5. The method according to claim 4, wherein controlling the plurality computing nodes of the detection network to operate sequentially in response to sequentially receiving the specified number of delay frames, and acquiring the target detection result output by the detection network upon receiving the last delay frame comprises:

upon inputting the key frame into a first-level sub-network of the cascaded detection network, each time one of the delay frames is received, inputting a detection result output by a previous-level sub-network and the one of the delay frames into a current-level sub-network; and

upon receiving the last delay frame, determining a detection result output by the current-level sub-network as the target detection result of the cascaded detection network.

6. The method according to claim 2, wherein

the detection network is a non-cascaded detection network; and

determining the specified number of delay frames in response to the each video frame being the key frame comprises:

determining a specified number of successive video frames following the key frame as the specified number of delay frames.

7. The method according to claim 6, wherein determining the plurality of computing nodes of the detection network based on the specified number comprises:

predicting a total duration required for the target detection on the key frame by the non-cascaded detection network; and

calculating an operation suspend time of each of the computing nodes of the non-cascaded detection network based on the total duration and the specified number.

8. The method according to claim 7, wherein upon inputting the key frame into the detection network, controlling the plurality computing nodes of the detection network to operate sequentially in response to sequentially receiving the specified number of delay frames, and acquiring the target detection result output by the detection network upon receiving the last delay frame comprises:

calculating an operation duration of the non-cascaded detection network upon inputting the key frame into the non-cascaded detection network;

in response to the operation duration reaching the operation suspend time of the computing node, suspending an operation of the non-cascaded detection network;

resuming the operation of the non-cascaded detection network from a suspended computing node each time one of the delay frames is received; and

in response to the last delay frame being received, determining a detection result output, upon resumption of the operation of the non-cascaded detection network, by the non-cascaded detection network as the target detection result of the non-cascaded detection network.

9. The method according to any one of claims 1 to 8, wherein upon acquiring the target detection result output by the detection network upon receiving the last delay frame in the specified number of delay frames by simultaneously inputting the key frame into the detection network and controlling the detection network to operate during the period of receiving the specified number of delay frames, the method further comprises:

determining whether the each video frame is the key frame; and

in response to the each video frame being the key frame, performing the process of acquiring the target detection result output by the detection network upon receiving the last delay frame in the specified number of delay frames by inputting the key frame into the detection network and controlling the detection network to operate during the period of receiving the specified number of delay frames; and

the method further comprises:

in response to the each video frame being not the key frame, outputting the target tracking result of the each video frame.

10. The method according to any one of claims 1 to 8, wherein generating the final target tracking result based on the target detection result and the target tracking result of the last delay frame comprises:

updating a state of a target in the tracked target list.

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

determining a key frame in response to a target state of a target in the tracked target list being indicated as lost; and

acquiring a target detection result by inputting the determined key frame into the detection network.

12. The method according to claim 11, wherein determining the key frame in response to the target state of the target in the tracked target list being indicated as lost comprises:

in response to the target state of the target in the tracked target list being indicated as lost, determining an $N^{th}$ video frame received following the last delay frame as the key frame, wherein N is a natural number.

13. An apparatus for detecting and tracking a target, comprising:

a tracking module, configured to acquire a target tracking result of each video frame of a plurality of video frames received from a video stream by inputting the each video frame of the plurality of video frames into a tracking network, wherein the tracking network is configured to track a detected target in a video frame;

a detecting module, configured to, in response to the each video frame being a key frame, acquire a target detection result output by a detection network upon receiving a last delay frame in a specified number of delay frames by simultaneously inputting the key frame into the detection network and control the detection network to operate during a period of receiving the specified number of delay frames, wherein the key frame is a video frame for target detection in the video stream, and the specified number of delay frames are successive video frames following the key frame; and

a detection and tracking result generating module, configured to generate a final target tracking result based on the target detection result and a target tracking result of the last delay frame, wherein a tracked target list in the final target tracking result is updated based on a detected target list in the target detection result, and target tracking for subsequent frames is performed based on the updated tracked target list.

14. An electronic device, comprising:

at least one processor; and

a memory configured to store at least one program;

wherein the at least one processor, when loading and running the at least one program, is caused to perform the method for detecting and tracking the target as defined in any one of claims 1 to 12.

15. A computer-readable storage medium storing one or more computer programs therein, wherein the one or more computer programs, when loaded and run by a

processor, cause the processor to perform the method for detecting and tracking the target as defined in any one of claims 1 to 12.