

WHAT IS CLAIMED IS:

1. A computer-implemented method, comprising:
determining an encoding state of a video encoder, the encoding state associated with
5 encoding a first video unit by the video encoder;
determining, by a reinforcement learning model and based on the encoding state of
the video encoder, an encoding parameter associated with rate control for the video encoder;
and
encoding a second video unit different from the first video unit based on the
10 encoding parameter.
2. The method of claim 1, wherein determining the encoding parameter
comprises:
determining, by the reinforcement learning model, an action based on the encoding
15 state of the video encoder; and
mapping the action to the encoding parameter.
3. The method of claim 1, wherein the encoding state associated with encoding
the first video unit comprises:
20 a state representing an outcome for encoding at least the first video unit;
a state of a buffer configured to buffer video units encoded by the video encoder
before transmission; and
a state associated with a status of a network for transmitting the encoded video
units.
25
4. The method of claim 3, wherein the outcome for encoding at least the first
video unit comprises the encoding parameter of the encoded first video unit and a size of
the encoded first video unit, the state of the buffer comprises the usage of the buffer, and
the state associated with the status of the network comprises a target bits per pixel.
30
5. The method of claim 4, wherein the usage of the buffer comprises at least one
of:
a ratio of an occupied space to maximum space of the buffer; and

remaining space of the buffer measured in video units.

6. The method of claim 1, wherein the reinforcement learning model is trained by:

5 determining a reward for the encoding parameter based on the encoding of the second video unit, wherein the reward is configured to penalize buffer overshooting and to increase as the encoding parameter results in a higher visual quality.

7. The method of claim 6, wherein determining the reward comprises:

10 determining a base reward in such a way that the base reward has a negative value if buffer overshooting occurs and the base reward is proportional to the encoding parameter by a negative coefficient if the buffer overshooting does not occur; and

scaling the base reward by a scaling factor to obtain the reward, wherein the scaling factor is based on a ratio of a bandwidth associated with encoding the second video unit to
15 maximum bandwidth of a transmission channel.

8. The method of claim 6, wherein the reinforcement learning model is further trained by:

determining an action associated with the encoding parameter based on the
20 encoding state of the video encoder;

determining an evaluation value for the encoding state for encoding the second video unit;

determining a value loss based on the reward and the evaluation value;

determining a policy loss based on the action; and

25 updating the reinforcement learning model based on the value loss and the policy loss.

9. The method of claim 1, wherein the reinforcement learning model comprises a neural network of an agent, wherein the neural network comprises:

30 at least one input fully connected layer configured to extract features from the encoding state;

at least one recurrent neural network coupled to receive the extracted features; and

at least one output fully connected layer configured to decide an action for the agent.

10. The method of claim 9, wherein the neural network is trained based on an actor-critic architecture, the actor configured to generate the action based on the encoding state and the critic configured to generate an evaluation value for the encoding state; and

5 wherein the actor and the critic share a common portion of the neural network comprising the at least one input fully connected layer and the at least one recurrent neural network.

11. The method of claim 1, wherein the encoding parameter comprises at least one
10 of a quantization parameter and a lambda parameter.

12. The method of claim 1, wherein the video encoder is configured to encode screen content for real-time communication.

13. A device comprising:
a processor; and
a memory having instructions stored thereon for execution by the processor, the instructions, when executed by the processor, cause the device to perform acts including:
determining an encoding state of a video encoder, the encoding state associated with
20 encoding a first video unit by the video encoder;
determining, by a reinforcement learning model and based on the encoding state of the video encoder, an encoding parameter associated with rate control in the video encoder;
and
encoding a second video unit different from the first video unit based on the
25 encoding parameter.

14. The device of claim 13, wherein the encoding state associated with encoding the first video unit comprises:

a state representing an outcome for encoding at least the first video unit;
30 a state of a buffer configured to buffer video units encoded by the video encoder before transmission; and
a state associated with a status of a network for transmitting the encoded video units.

15. The device of claim 14, wherein the outcome for encoding at least the first video unit comprises the encoding parameter of the encoded first video unit and a size of the encoded first video unit, the state of the buffer comprises the usage of the buffer, and the state associated with the status of the network comprises a target bits per pixel.

5

16. The device of claim 13, wherein the reinforcement learning model is trained by:

determining a reward for the encoding parameter based on the encoding of the second video unit, wherein the reward is configured to penalize buffer overshooting and to increase as the encoding parameter results in a higher visual quality.

10

17. The device of claim 16, wherein determining the reward comprises:

determining a base reward in such a way that the base reward has a negative value if buffer overshooting occurs and the base reward is proportional to the encoding parameter by a negative coefficient if the buffer overshooting does not occur; and

15

scaling the base reward by a scaling factor to obtain the reward, wherein the scaling factor is based on a ratio of a bandwidth associated with encoding the second video unit to maximum bandwidth of a transmission channel.

20

18. The device of claim 15, wherein the reinforcement learning model comprises a neural network of an agent, wherein the neural network comprises:

at least one input fully connected layer configured to extract features from the encoding state;

at least one recurrent neural network coupled to receive the extracted features; and

25

at least one output fully connected layer configured to decide an action for the agent.

19. The device of claim 13, wherein the video encoder is configured to encode screen content for real-time communication.

30

20. A computer program product, comprising a computer readable storage medium having program instructions embodied therewith, the program instructions executable by a processor to cause the processor to perform acts comprising:

determining an encoding state of a video encoder, the encoding state associated with encoding a first video unit by the video encoder,

determining, by a reinforcement learning model and based on the encoding state of the video encoder, an encoding parameter associated with rate control in the video encoder; and

5 encoding a second video unit different from the first video unit based on the encoding parameter.