CLAIMS:

1. A computer-implemented system for real-time video processing, comprising:

at least one memory configured to store instructions; and

at least one processor configured to execute the instructions to perform operations comprising:

receive real-time video generated by a medical image system, the real-time video including a plurality of image frames; and

while receiving the real-time video generated by the medical image system:

process the plurality of image frames of the real-time video to obtain context information indicating an interaction of a user with the medical image system;

perform an object detection to detect at least one object in the plurality of image frames;

perform a classification to generate classification information for the at least one detected object in the plurality of image frames; and

perform an image modification to modify the received real-time video based on at least one of the object detection and the classification; and

generate a display of the real-time video with the image modification on a video display device;

wherein the at least one processor is further configured to invoke at least one of the object detection and the classification based on the interaction of the user with the medical image system indicated by the context information.

2. The system of claim 1, wherein at least one of the object detection and the classification is performed by applying at least one neural network trained to process frames received from the medical image system.

3. The system of claim 1 or 2, wherein the at least one processor is configured to invoke the object detection when the context information indicates that the user is interacting with the medical image system to identify objects.

4. The system of claim 3, wherein the at least one processor is further configured to deactivate the object detection when the context information indicates that the user is no longer interacting with the medical image system to identify objects.

5. The system of any preceding claim, wherein the at least one processor is configured to invoke the classification when the context information indicates that the user is interacting with the medical image system to examine the at least one object in the plurality of image frames.

6. The system of claim 5, wherein the at least one processor is further configured to deactivate the classification when the context information indicates that the user is no longer interacting with the medical image system to examine the at least one object in the plurality of image frames.

7. The system of any preceding claim, wherein the at least one processor is further configured to invoke the object detection when context information indicates that the user is interested in an area in the plurality of image frames containing at least one object, and wherein the at least one processor is further configured to invoke classification when context information indicates that the user is interested in the at least one object.

8. The system of any preceding claim, wherein the at least one processor is further configured to perform an aggregation of two or more frames containing the at least one object, and wherein the at least one processor is further configured to invoke the aggregation based on the context information.

9. The system of any preceding claim, wherein the image modification comprises at least one of an overlay including at least one border indicating a location of the at least one detected object, classification information for the at least one object, a zoomed image of the at least one object, or a modified image color distribution.

10. The system of any preceding claim, wherein the at least one processor is configured to generate the context information based on an Intersection over Union (IoU) value for the location of the at least one detected object in two or more image frames over time.

11. The system of any preceding claim, wherein the at least one processor is configured to generate the context information based on an image similarity value in two or more image frames.

12. The system of any preceding claim, wherein the at least one processor is configured to generate the context information based on a detection or a classification of one or more objects in the plurality of image frames.

13. The system of any preceding claim, wherein the at least one processor is configured to generate the context information based on an input received by the medical image system from the user.

14. The system of any preceding claim, wherein the at least one processor is further configured to generate the context information based on the classification information.

15. The system of any preceding claim, wherein the plurality of image frames include image frames of a gastro-intestinal organ.

16. The system of any preceding claim, wherein the plurality of image frames comprise images from a medical image device used during at least one of an endoscopy, a gastroscopy, a colonoscopy, an enteroscopy, a laparoscopy, or a surgical endoscopy.

17. The system of any preceding claim, wherein the at least one detected object is an abnormality.

18. The system of claim 17, wherein the abnormality comprises at least one of a formation on or of human tissue, a change in human tissue from one type of cell to another type of cell, an absence of human tissue from a location where the human tissue is expected, or a lesion.

19. A method for real-time video processing, comprising:

receiving a real-time video generated by a medical image system, the real-time video including a plurality of image frames;

providing at least one neural network, the at least one neural network being trained to process image frames from the medical image system;

processing the plurality of image frames of the real-time video to obtain context information indicating an interaction of a user with the medical image system;

identifying a type of interaction based on the context information; and

performing real-time processing on the plurality of image frames based on the identified type of the interaction by applying the at least one trained neural network.

20. The method of claim 19, wherein performing real-time processing includes performing at least one of an object detection to detect at least one object in the plurality of image frames, a classification to generate classification information for the at least one detected object, and an image modification to modify the received real-time video.

21. The method of claim 20, wherein the object detection is invoked when the identified interaction is the user interacting with the medical image system to navigate to identify objects.

22. The method of claim 21, wherein the object detection is deactivated when the context information indicates that the user no longer interacting with the medical image system to navigate to identify objects.

23. The method of any one of claims 20 to 22, wherein the classification is invoked when the identified interaction is the user interacting with the medical image system to examine the at least one detected object in the plurality of image frames.

24. The method of any one of claims 20 to 23, wherein the classification is deactivated when the context information indicates that the user no longer interacting with the medical image system to examine at least one detected object in the plurality of image frames.

25. The method of any one of claims 20 to 24, wherein the object detection is invoked when context information indicates that the user is interested in an area in the plurality of image frames containing at least one object, and wherein classification is invoked when context information indicates that the user is interested in the at least one object.

26. The method of any one of claims 20 to 25, wherein at least one of the object detection and the classification is performed by applying at least one neural network trained to process frames received from the medical image system.

27. The method of any one of claims 20 to 26, wherein the image modification comprises

at least one of an overlay including at least one border indicating a location of the at least one detected object, classification information for the at least one detected object, a zoomed image of the at least one detected object, or a modified image color distribution.

28. The method of any one of claims 20 to 27, wherein the at least one detected object is an abnormality.

29. The method of claim 28, wherein the abnormality comprises at least one of a formation on or of human tissue, a change in human tissue from one type of cell to another type of cell, an absence of human tissue from a location where the human tissue is expected, or a lesion.

30. The method of any one of claims 19 to 29, further comprising the step of performing an aggregation of two or more frames containing at least one object based on the context information.

31. The method of any one of claims 19 to 30, wherein the plurality of image frames include image frames of a gastro-intestinal organ.

32. The method of any one of claims 19 to 31, wherein the frames comprise images from a medical image device used during at least one of an endoscopy, a gastroscopy, a colonoscopy, an enteroscopy, a laparoscopy, or a surgical endoscopy.