CLAIMS

1. A computer-implemented method, comprising:

identifying an object in a first video frame of a video by applying a first AI-assisted object detection algorithm that outputs a mask defining location of the object in the first video frame, wherein the first AI-assisted object detection algorithm includes a promptable image segmentation model;

tracking the mask and defining its location in at least one temporally subsequent frame of the video using a second AI-assisted object detection algorithm, wherein the second AI-assisted object detection algorithm includes a diffusion-based generative model; and

determining that a further frame of the video is a keyframe and, on that basis, re-applying the first AI-assisted objected detection algorithm to identify the object in the further frame and to output a next mask defining the location of the object in the further frame before repeating the tracking and defining using the second AI-assisted object detection algorithm for one or more frames subsequent to the keyframe.

2. The method of claim 1, wherein the first AI-assisted object detection algorithm determines the mask defining location of the object in the first video frame without reference to preceding frames of the video or previously-determined masks for the video, and wherein the second AI-assisted object detection algorithm tracks the mask and defines its location in the at least one temporally subsequent frame of the video based on a location of the mask in a preceding frame and a determination of correspondence between the preceding frame and the at least one temporally subsequent frame.

3. The method of claim 1, wherein determining that the further frame is a keyframe includes receiving keyframe identification data from at least one of a video encoder and a video decoder regarding the video.

4. The method of claim 3, wherein the keyframe identification data includes data identifying frames within decoded video that are keyframes.

5. The method of claim 3, wherein receiving keyframe identification data includes extracting the keyframe identification data from decoded video.

6. The method of claim 1, wherein determining that the further frame is a keyframe in based on determining that the further frame was intra-coded by an encoder.

7. The method of claim 1, wherein determining that the subsequent frame is a keyframe includes performing scene change analysis on the video and determining that the subsequent frame is a scene change.

8. The method of claim 1, wherein determining that the further frame is a keyframe includes tracking the mask and defining its location in the further frame of the video using the second AI-assisted object detection algorithm, applying the first AI-assisted object detection algorithm to identify a new mask defining a location of the object in the further frame, comparing the new mask to the location of the mask determined by the second AI-assisted object detection algorithm, and determining that an error measurement exceeds a threshold value.

9. The method of claim 8, wherein the operations of tracking the mask and defining its location in the further frame, applying the first AI-assisted object detection algorithm to identify the new mask, comparing the new mask to the location of the mask determined by the second AI-assisted object detection algorithm, and determining the error measurement are carried out only on frames of the video that were intra-coded.

10. The method of claim 1, further comprising displaying the video on a display screen with a visual overlay indicating the object based on the mask and the next mask.

11. A non-transitory processor-readable medium storing processor-executable instructions that, when executed by one or more processors, are to cause the one or more processors to:

identify an object in a first video frame of a video by applying a first AI-assisted object detection algorithm that outputs a mask defining location of the object in the first video frame, wherein the first AI-assisted object detection algorithm includes a promptable image segmentation model;

track the mask and define its location in at least one temporally subsequent frame of the video using a second AI-assisted object detection algorithm, wherein the second AI-assisted object detection algorithm includes a diffusion-based generative model; and

determine that a further frame of the video is a keyframe and, on that basis, re-apply the first AI-assisted objected detection algorithm to identify the object in the further frame and to output a next mask defining the location of the object in the further frame before repeating the tracking and defining using the second AI-assisted object detection algorithm for one or more frames subsequent to the keyframe.

12. The non-transitory processor-readable medium of claim 11, wherein the first AI-assisted object detection algorithm determines the mask defining location of the object in the first video frame without reference to preceding frames of the video or previously-determined masks for the video, and wherein the second AI-assisted object detection algorithm tracks the mask and defines its location in the at least one temporally subsequent frame of the video based on a location of the mask in a preceding frame and a determination of correspondence between the preceding frame and the at least one temporally subsequent frame.

13. The non-transitory processor-readable medium of claim 11, wherein the instructions, when executed, are to cause the one or more processors to determine that the further frame is a keyframe by at least receiving keyframe identification data from at least one of a video encoder and a video decoder regarding the video.

14. The non-transitory processor-readable medium of claim 13, wherein the keyframe identification data includes data identifying frames within decoded video that are keyframes.

15. The non-transitory processor-readable medium of claim 13, wherein receiving keyframe identification data includes extracting the keyframe identification data from decoded video.

16. The non-transitory processor-readable medium of claim 11, wherein the instructions, when executed, are to cause the one or more processors to determine that the further frame is a keyframe by determining that the further frame was intra-coded by an encoder.

17. The non-transitory processor-readable medium of claim 11, wherein the instructions, when executed, are to cause the one or more processors to determine that the subsequent frame is a keyframe by performing scene change analysis on the video and determining that the subsequent frame is a scene change.

18. The non-transitory processor-readable medium of claim 11, wherein the instructions, when executed, are to cause the one or more processors to determine that the further frame is a keyframe by tracking the mask and defining its location in the further frame of the video using the second AI-assisted object detection algorithm, applying the first AI-assisted object detection algorithm to identify a new mask defining a location of the object in the further frame, comparing the new mask to the location of the mask determined by the second AI-assisted object detection algorithm, and determining that an error measurement exceeds a threshold value.

19. The non-transitory processor-readable medium of claim 18, wherein the operations of tracking the mask and defining its location in the further frame, applying the first AI-assisted object detection algorithm to identify the new mask, comparing the new mask to the location of the mask determined by the second AI-assisted object detection algorithm, and determining the error measurement are carried out only on frames of the video that were intra-coded.

20. The non-transitory processor-readable medium of claim 11, wherein the instructions, when executed, are to cause the one or more processors to display the video on a display screen with a visual overlay indicating the object based on the mask and the next mask.

21. A computing system, comprising: one or more processors; and memory coupled to at least one of the one or more processors, the memory storing computer-executable instructions that, when executed by the one or more processors, are to cause the one or more processors to:

identify an object in a first video frame of a video by applying a first AI-assisted object detection algorithm that outputs a mask defining location of the object in the first video frame, wherein the first AI-assisted object detection algorithm includes a promptable image segmentation model;

track the mask and define its location in at least one temporally subsequent frame of the video using a second AI-assisted object detection algorithm, wherein the second AI-assisted object detection algorithm includes a diffusion-based generative model; and

determine that a further frame of the video is a keyframe and, on that basis, re-apply the first AI-assisted objected detection algorithm to identify the object in the further frame and to output a next mask defining the location of the object in the further frame before repeating the tracking and defining using the second AI-assisted object detection algorithm for one or more frames subsequent to the keyframe.