What is claimed is:

1. A system for operating a remotely operated vehicle (ROV) using simultaneous localization and mapping (SLAM) comprising:

a ROV with (i) a video camera operable to output real video and (ii) a positional sensor operable to output position data;

a SLAM engine comprising:

a video dataset operable to store video data and real images coming from the ROV;

a depth dataset operable to store depth maps, each of the depth maps being a two-dimensional matrix having a same size as a width and height of a target frame from the video camera, wherein each element of the two-dimensional matrix represents a distance from the video camera to a scene structure;

a 3D model dataset operable to store 3D model data of a scene where an ROV may operate;

a depth map simulator with access to the 3D model dataset and a set of camera parameters, wherein the depth map simulator is operable to synthesize a depth map for storage in the depth dataset;

a model’s weights dataset operable to store weights of the SLAM engine;

a SLAM trainer module with access to the video dataset and the depth dataset, wherein the SLAM trainer module is operable to run a SLAM-Net architecture; and

an application module communicatively coupled to the ROV and operable to receive the real video, the position data, and the model’s weights dataset,

wherein the application module is operable to smooth the position data, reconstruct the scene, and display the scene on a graphical user interface.

2. The system according to claim 1, wherein the SLAM-Net architecture comprises a set of input frames.

3. A system according to any of claims 1 - 2, wherein the SLAM-Net architecture comprises a depth map, a set of camera motions represented as transformation matrices, segmentation masks, and a plurality of convolutional neural networks.

4. The system of claim 3, wherein the SLAM-Net architecture comprises at least one skip connection.

5. A system according to any of claims 1 - 4, further comprising:

a set of unlabeled videos stored in the video dataset;

wherein the SLAM engine receives the set of unlabeled videos from the video dataset and minimizes photometric error between a target frame and a set of remaining frames.

6. A system according to any of claims 1 - 5, wherein the SLAM engine segments a plurality of pixels from the video data.

7. The system of claim 1, wherein the SLAM engine is operable to perform bilinear sampling by linearly interpolating an intensity value of four discrete pixel neighbors of a homogeneous pixel coordinate projection.

8. A system according to any of claims 1 - 7, wherein the SLAM engine tracks at least one point across a plurality of frames.

9. A system according to any of claims 1 - 8, wherein the SLAM engine uses a generative adversarial network (GAN) to learn a depth prior to improve a depth map.

10. The system according of claim 9, wherein the GAN comprises a generator network operable to output at least one fake example and a discriminator network operable to distinguish between the at least one fake example and a real example.

11. A system according to any of claims 1 – 10, wherein the SLAM engine synthesizes depth maps using a 3D model depiction of a real scene.

12. A system for undersea exploration comprising:

a networked operating system comprising a computer and computer executable software comprising a simultaneous localization and mapping (SLAM) engine;

a ROV communicatively coupled with the operating system and comprising (i) a video camera operable to output real video and (ii) a positional sensor operable to output position data;

wherein the SLAM engine comprises:

a video dataset operable to store video data and real images coming from the ROV;

a depth dataset operable to store depth maps, each of the depth maps being a two-dimensional matrix having a same size as a width and height of a target frame from the video

camera, wherein each element of the two-dimensional matrix represents a distance from the video camera to a scene structure;

a 3D model dataset operable to store 3D model data of a scene where an ROV may operate;

a depth map simulator with access to the 3D model dataset and a set of camera parameters, wherein the depth map simulator is operable to synthesize a depth map for storage in the depth dataset;

a model's weights dataset operable to store weights of the SLAM engine;

a SLAM trainer module with access to the video dataset and the depth dataset, wherein the SLAM trainer module is operable to run a SLAM-Net architecture; and

an application module communicatively coupled to the ROV and operable to receive the real video, the position data, and the model's weights dataset, wherein the application module is operable to smooth the position data and reconstruct the scene; and

a navigation interface configured to display the scene, the navigation interface comprising at least one networked monitor.

13. The system according to claim 12, wherein the SLAM-Net architecture comprises a set of input frames.

14. A system according to any of claims 12 - 13, wherein the SLAM-Net architecture comprises a depth map, a set of camera motions represented as transformation matrices, segmentation masks, and a plurality of convolutional neural networks.

15. A system according to any of claims 12 - 14, further comprising:

a set of unlabeled videos stored in the video dataset;

wherein the SLAM engine receives the set of unlabeled videos from the video dataset and minimizes photometric error between a target frame and a set of remaining frames.

16. A system according to any of claims 12 - 15, wherein the SLAM engine segments a plurality of pixels from the video data.

17. The system of claim 16, wherein the SLAM engine is operable to perform bilinear sampling by linearly interpolating an intensity value of four discrete pixel neighbors of a homogeneous pixel coordinate projection.

18. A system according to any of claims 12 - 17, wherein the SLAM engine tracks at least one point across a plurality of frames.

19. A system according to any of claims 12 - 18, wherein the SLAM engine uses a generative adversarial network (GAN) to learn a depth prior to improve a depth map; and

wherein the GAN comprises a generator network operable to output at least one fake example and a discriminator network operable to distinguish between the at least one fake example and a real example.

20. A method of simultaneous localization and mapping (SLAM) for remotely operated vehicles (ROV) comprising:

obtaining video data, real images, and position data

from an ROV;

obtaining depth maps, each of the depth maps being a two-dimensional matrix having a same size as a width and height of a target frame from the video camera, wherein each element of the two-dimensional matrix represents a distance from the video camera to a scene structure;

smoothing the position data using a SLAM-Net convolutional neural network (CNN) architecture and outputting smoothed position data;

reconstructing a 3D scene based at least in part on the smoothed position data; and

displaying the 3D scene on a graphical user interface.

21. A computer program product comprising instructions which, when the program is executed by a computer, cause the computer to carry out the steps of the method of claim 20.