IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 14, 2021, accepted January 27, 2022, date of publication February 1, 2022, date of current version February 14, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3148544

Semantic Hazard Labelling and Risk Assessment Mapping During Robot Exploration

REEM ASHOUR[1], MOHAMED ABDELKADER[2,3], JORGE DIAS[4], (Senior Member, IEEE), NAWAF I. ALMOOSA[5], AND TAREK TAHA[6]

[1]Autonomous Robotics Research Centre (ARRC), Technology Innovation Institute (TII), Masdar City, United Arab Emirates
[2]Robotics & Internet of Things Laboratory, Prince Sultan University, Riyadh 11586, Saudi Arabia
[3]Systemtrio Robotics, Abu Dhabi, United Arab Emirates
[4]Khalifa University Center for Autonomous Robotic Systems (KUCARS), Khalifa University of Science and Technology (KU), Abu Dhabi, United Arab Emirates
[5]Emirates ICT Innovation Centre (EBTIC), Khalifa University of Science and Technology (KU), Abu Dhabi, United Arab Emirates
[6]Robotics Laboratory, Dubai Future Foundation, Dubai, United Arab Emirates

Corresponding author: Reem Ashour (reemashour1@gmail.com)

This work was supported in part by the Khalifa University of Science and Technology under Award RC1-2018-KUCARS; and in part by the Autonomous Robotics Research Centre (ARRC), Technology Innovation Institute (TII).

ABSTRACT This paper proposes an innovative hazard identification and risk assessment mapping model for Urban Search and Rescue (USAR) environments, concentrating on a 3D mapping of the environment and performing grid-level semantic labeling to recognize all hazards types found in the scene and to distinguish their risk severity level. The introduced strategy employs a deep learning model to create semantic segments for hazard objects in 2D images and create semantically annotated point clouds that encapsulate occupancy and semantic annotations such as hazard type and risk severity level. After that, a 3D semantic map that provides situational awareness about the risk in the environment is built using the annotated point cloud. The proposed strategy is evaluated in a realistic simulated indoor environment, and the results show that the system successfully generates a risk assessment map. Further, an open-source package for the proposed approach is provided online for testing and reproducibility.

INDEX TERMS Hazard identification, mapping, object classification, risk assessment, risk mapping, semantic mapping.

I. INTRODUCTION

Consequences of real-world disasters create a need for safer Urban Search and Rescue (USAR) where victims are found. Disasters can be natural, such as earthquakes, floods, and hurricanes, or human-made catastrophes, such as wars, terrorism, and accidents. Most of the time, disasters lead to the collapse of buildings, leaving behind an unstructured environment with victims surrounded by sources of danger that both need to be located [1]–[3]. The unstructured disaster environment suffers from harsh conditions such as hazardous materials and blocked pathways that make the operation of the rescue team harder, where reaching services such as hospitals and power supply is restricted [3]. Finding survivors and rescuing them is the prime goal in search and rescue missions. According to [4], rescue workers have approximately 48 hours to find trapped survivors; otherwise, the likelihood of finding victims alive drops substantially. Thus, USAR response time should be fast enough to maximize saving as many survivors as possible and locate all sources of danger, yet adequate to respond without adding additional risk to the rescuers and victims [2], [3]. Rescue robots are required to explore the environment, detect and locate victims, locate the source of hazard, assess risk, create an informative map, and provide an optimal danger-free path for the first responders.

The work presented in this paper tackles the safety of the first responders, where the rescue robot provides situational awareness for the first responders before the intervention. The situational awareness is provided by assessing the risk types and severity levels by creating a 3D occupancy semantic map that exhibits the different levels of risks with various color labels.

In the current state of the art, various contributions studied the problem of object detection, including human detection [5]–[12] and semantic mapping [13]–[18]. However, most of the work focused on locating the victims inside the

The associate editor coordinating the review of this manuscript and approving it for publication was Emanuele Bellini.

USAR environment without conceding the assessment of the risk level and increasing the level of situational awareness of the first responders. Therefore, the primary objective of our research is to determine how to use current software and hardware to develop a system capable of autonomously recognizing, locating, and projecting the source of risk in 3D space in an indoor environment.

In this work, we present a hazard identification and risk assessment mapping system to locate the source of hazards in a simulated urban search and rescue environment, as shown in Figure 1. Commonly, The risk assessment in an indoor environment is performed manually by a human investigator that moves around the building, fills a checklist for the risk objects, and performs risk analysis. The risk analysis and quantification are usually performed using a risk matrix that explains the connection between the probability and severity of the risk, as shown in Figure 2. The quantitative risk analysis is a valuable method for determining the priority of dangers that first responders may consider. Risk analysis and assessment using drones for search and rescue missions were only performed in the literature for the outdoor environment.

The proposed hazard identification and risk assessment mapping system for the indoor environment uses a single onboard sensor, mainly RGBD, to generate a risk assessment map that shows three different types of risks with two severity levels of each. A deep learning model is applied to detect distinct sources of risks and indicate their severity level. Consequently, the information obtained from the deep learning model is used to create a 3D semantic map. The generated map labels the different risk types to speed up the search mission of locating victims and increase situational awareness.

The suggested system is modular for multiple independent software components. For example, alternative object detection algorithms may be evaluated without changing the entire system because the component is separate. A variety of deep neural networks might be investigated.

Our main contributions are:

- A risk assessment technique based on Deep Neural Network to identify and locate different sources of risk and categorize them to various levels of severity to increase the situational awareness of the indoor environment;
- A unique semantic 3D mapping model for risk assessment and situational awareness
- Dataset, provided for download, efficiently acquired on realistic sources of risk with their corresponding annotated labels, that was used to train and test the Deep Neural Network;
- A method implementation and presentation on a quadrotor for detecting, locating, and mapping previously unseen risk sources in an indoor environment.

The remainder of the paper is arranged as follows; Section II outlines the related work. The detailed proposed method is presented in Section III. Section IV shows the simulation process and the experimental results. Finally, section V draws the conclusion and future works.

II. RELATED WORK

Different robotics platforms are used in Search and Rescue missions to perform various tasks such as victim detection [12], exploration for mapping and 3D reconstruction [13], and risk assessment [19], [20]. The first responders should be aware of the environmental situation before intervention. Hence, risk assessment is a crucial task where the risks should be identified and located. In this work, state-of-the-art object detection, semantic mapping, and risk assessment methods are studied.

A. OBJECT DETECTION

Object detection in the search and rescue context includes the detection of victims and sources of danger/risk. The detection process can be performed using single or multiple sensors. For example, a single RGBD sensor can detect victims visually, or multiple sensors outputs can be fused to reduce detection uncertainty. Autonomous object detection is considered challenging due to object orientation, occlusions, and lighting conditions. Numerous approaches investigated object detection; the most common approach is visual detection using machine learning and deep learning models. The information captured from a visual sensor (camera) is used in visual detection. However, various features are extracted from the images in machine learning visual inspection. Examples of such feature extraction methods are Scale-invariant feature transform (SIFT) or Histogram of oriented gradients (HOG) features. Then, objects are classified using machine learning models such as support vector machines (SVM). The limitation of such approaches are highlighted below:

- Image training should be done with a nearly uniform background, which is unreliable in a real-world disaster scenario.
- Low detection accuracy as machine learning algorithms represents humans with pre-defined features.

Unlike machine learning methods, deep learning models can detect objects with higher detection accuracy since the algorithm finds the most efficient feature patterns in the captured images without needing a feature extraction step. Several deep learning techniques are proposed in the literature that either creates a bounding box surrounding the classified objects, such as Single Shot MultiBox Detector (SSD) [5], You Only Look Once (YOLO) [6]–[8] and others [9]–[11] or create semantic segmentation of the classified objects such as PSPNet [21]. The mentioned approaches are visual-based, where 2D images obtained from the camera classify the objects. In addition, depth-based information such as point cloud obtained from the depth camera are used to create point cloud segments and classify them to their corresponding class like PointNet, and PointNet++ [22]–[24].

B. SEMANTIC MAPPING

In literature, maps are represented in different formats; topological maps where the links between different points are identified; metric maps where the environment can be

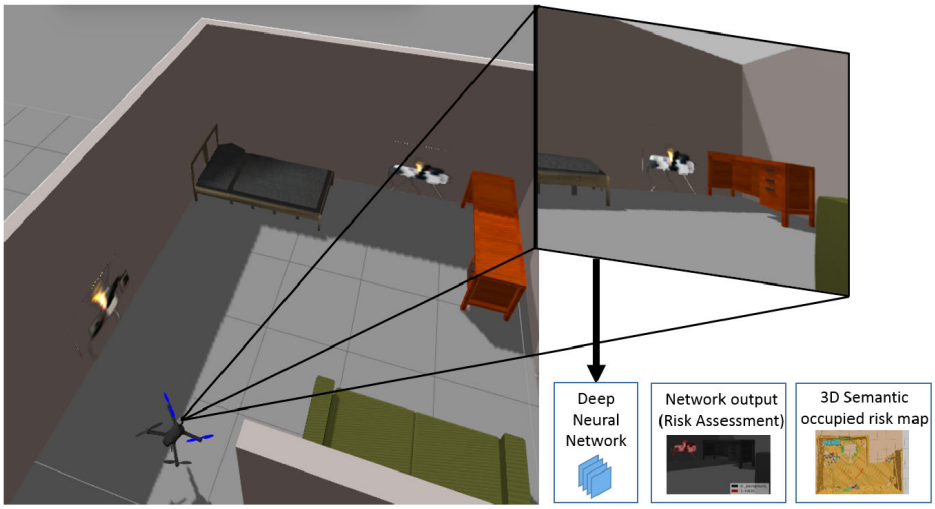


FIGURE 1. Our UAV captures the indoor environment images from a forward-looking camera; a Deep learning network identifies the risk sources and their corresponding level. The mapping module projects the risk in different annotation colors in a 3D semantic map.

presented in different dimensions, showing occupied and free spaces; and semantic maps that encapsulate characteristics compatible with human conception.

Metric maps are considered the most commonly used maps for robot exploration [14]–[18]. However, these maps do not have sufficient information that can guide the robot toward the detected objects (i.e., victims' location or sources in danger). Therefore, there is a need for more advanced skills than metric mapping, where the robot should understand the surrounding environment. Semantic maps provide such a solution by incorporating semantically labeled objects into the metric maps, which will provide a better understanding of the surrounding environment [25] and [26].

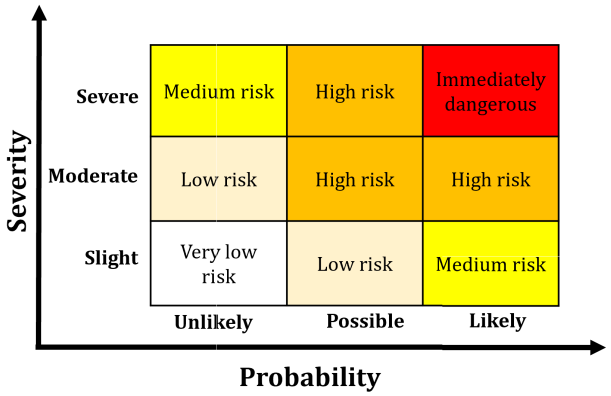


FIGURE 2. Risk matrix and risk quantification.

C. RISK ASSESSMENT

The risk assessment process is critical as it forms an essential component in search and rescue missions. The risk assessment is helpful since it increases the situational awareness of hazards and risks in the environment and prioritizes hazards for intervention decisions. Furthermore, the risk assessment is used to identify who may be at risk, determine if existing control models are sufficient or if more should be performed, and limit damages and injuries, mainly if the risk assessment is performed in the early stages.

Risk assessment in urban search and rescue (USAR) environments is a process that consists of hazard identification, risk analysis, risk evaluation, and risk control [27]. When multiple risks are presented in the environment, ranking or prioritizing hazards help to determine risk severity and, consequently, the required control used. There is no unique approach to define the level of risk. One of the popular approaches is using the risk matrix that explains the connection between probability and severity of the risk [27], as shown in Figure 2

UAVs have become a ubiquitous factor in indoor risk assessment. Their role is attributed to the significant technological advancements that resulted in highly sophisticated, economically affordable UAVs capable of carrying high-resolution cameras and other sensing equipment. Additionally, UAVs offer less hazardous alternatives than conventional risk assessment techniques as they eliminate human factors.

UAVs have been employed in real-world disaster scenarios for risk assessment. In [28], a drone equipped with LiDAR and uses photogrammetry technology is used to create high-resolution maps of impacted areas. The drone is called "Terra Drone Indonesia." In March 2019, Terra Drone Indonesia surveyed 750 hectares of land. The data and images collected from the drone are used to create maps that present the current conditions and the damages. In [19], manually controlled drones are deployed to capture aerial images, develop maps, and help in assessing the damage in some of the affected regions in Ecuador in April 2016. In [20], UAV has been used

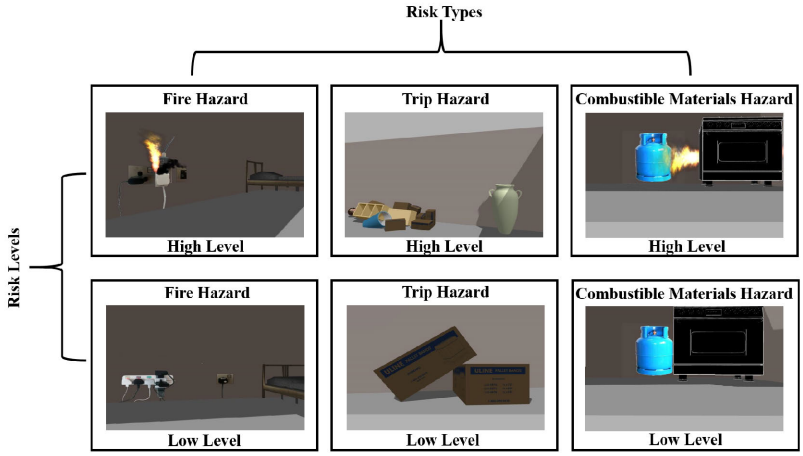


FIGURE 3. Risk definitions labels. Rows: Risk levels. Columns: Risk types.

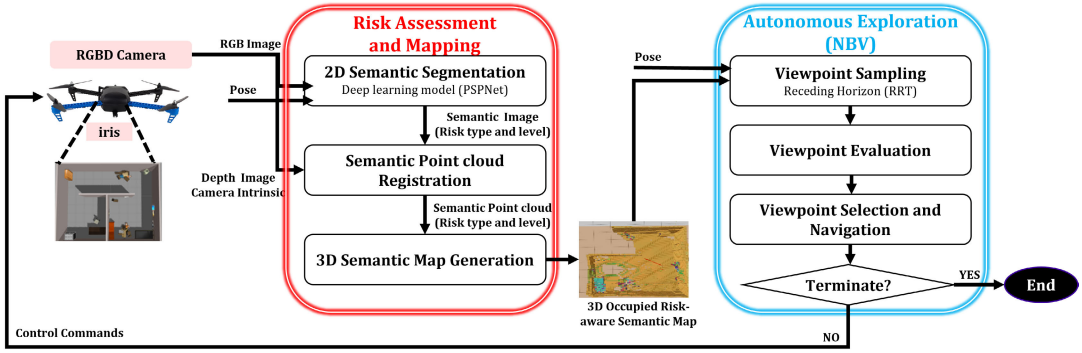


FIGURE 4. Proposed risk assessment model and mapping.

for landslide risk studies. The UAV was assigned to inspect rock cuts, cliffs, and rugged mountains. The UAV offered rapid and flexible data acquisition.

All risk assessment methods provided in the literature are performed on large outdoor areas or the inspection for building from outside. Additionally, the risk assessment using UAVs requires trained pilots. However, in this work, we tackle the risk assessment for an unknown indoor environment using a UAV that explores the environment autonomously utilizing a deep learning model. The proposed system creates a 3D semantic map that increases situational awareness about the indoor environment by identifying the risk types, severity, and accurate risk location.

FIGURE 5. Trip hazard, high level objects placed in the environments for dataset collection.

III. MATERIALS AND METHODS

A. PROBLEM FORMULATION

Consider a USAR environment -a generic scene- with various types of risk placed in different locations with different severity levels. In this work, only three types of risk are considered, fire from electric plugs, trip hazards, and combustible materials. Two risk levels are considered for each type; high and low levels, as illustrated in Figure 3.

Our input is an image acquired from a camera mounted on the drone. Given an image, our goal is to identify the risk source from the 2D image and project it in a 3D semantic map to facilitate the first responders' work and find the safest path for survivors. The risk is identified by semantically

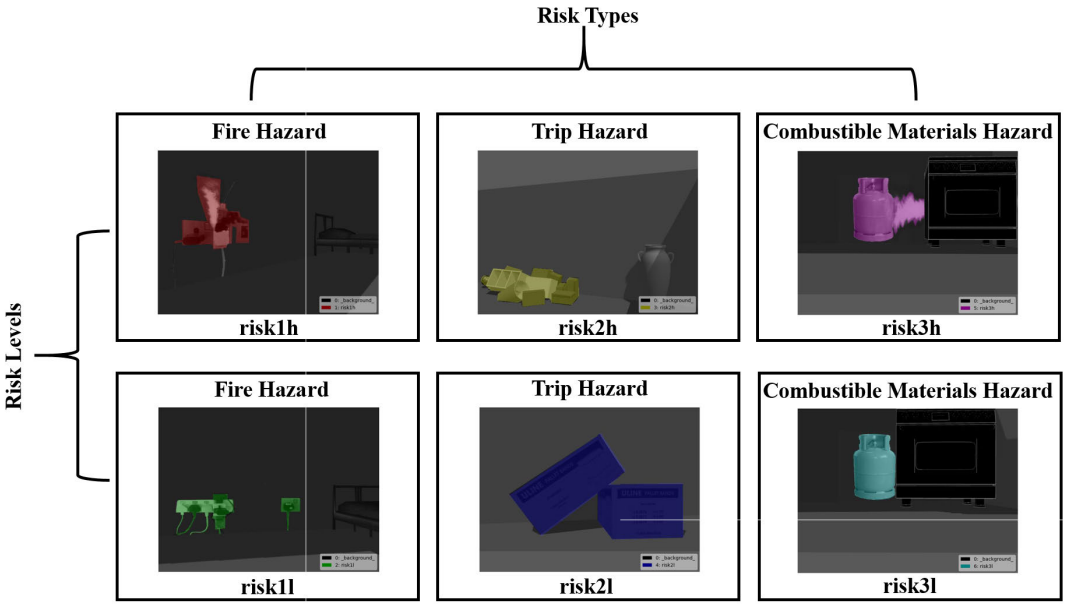


FIGURE 6. Dataset annotation.

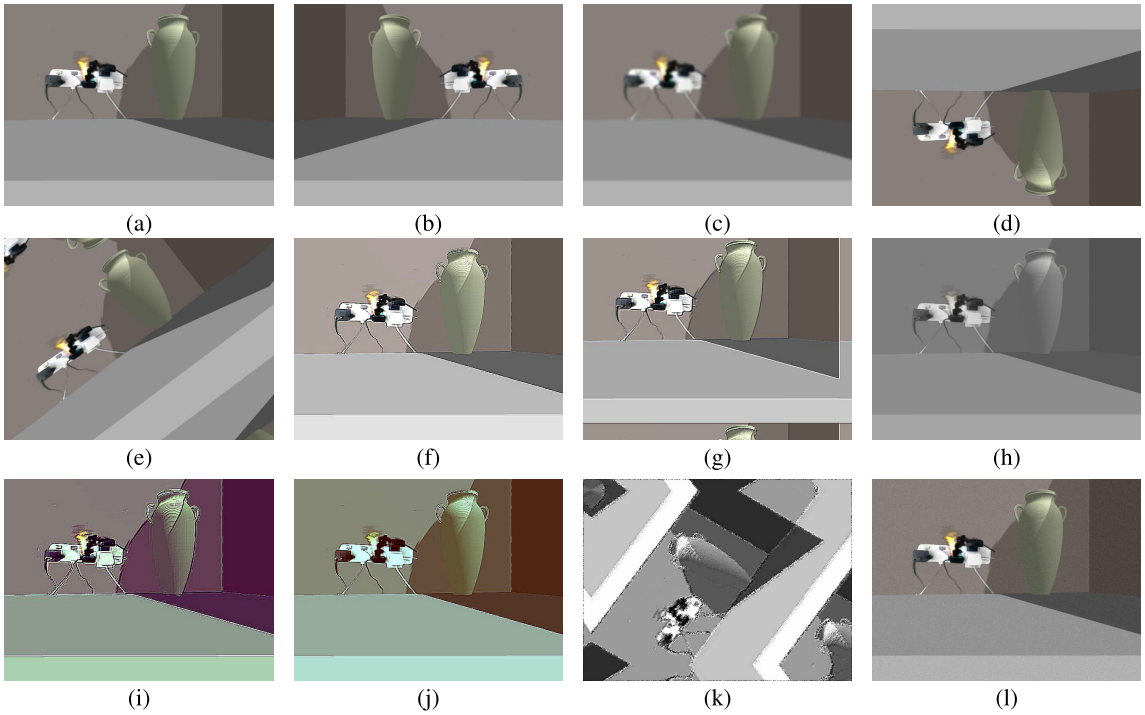


FIGURE 7. Examples of dataset augmentation. Each images shows the method used for augmentation. (a) Original image. (b) Horizontal flip (c) Gaussian blur (blur images with a sigma of 0 to 3.0) (d) Vertical flip (e) Scale images to 80-120% of their size, individually per axis, translate by −20 to +20 percent (per axis), rotate by −45 to +45 degrees, and shear by −16 to +16 degrees (f) Sharpen and emboss images (g) e + f + Change brightness of images (by -10 to 10 of original value), change hue and change saturation (h) Improve or worsen the contrast, convert it to gray scale, move pixels locally around (with random strengths) and move parts of the image around (i) Improve or worsen the contrast, convert it to gray scale (j) f + i (k) j + e + Change brightness of images (by −10 to 10 of original value), change hue and saturation, and crop pixels (l) k + Gaussian noise to images, move pixels locally around (with random strengths), and sometimes move parts of the image around.

segmenting the source of danger from 2D colored images and assigning a level risk. After that, the point cloud corresponding to the risk segments is annotated. Finally, a mapping module is used to project the annotated point cloud and create a 3D semantic occupancy map. The map presents the risk locations with a unique color for each type.

B. PROPOSED APPROACH

A new approach for risk assessment and mapping, shown in Figure 4, is proposed.

In this work, bounded volume $V \in R^3$ is initially unknown and must be semantically mapped. Our strategy attempts to create a 3D map M that locates the risk source and identifies their severity level. The 3D semantic map for different risk types and levels is generated by identifying the source of risks and classifying them into different labels with their corresponding level. The source of risk is identified and localized in 2D images obtained from a camera by performing semantic segmentation through a deep learning model to locate the risk and estimate its level. After that, the depth information obtained from a depth camera is used along with the output of the deep learning model to register an annotated point cloud. Finally, the annotated point cloud is used to create a 3D semantic occupancy map for risk assessment and situational awareness.

The 3D map is divided into small cubical voxels $m \in M$ with an edge of length res that indicates the map resolution. Moreover, a single voxel on the map carries the occupancy information and class type (risk type and severity level) represented by a unique color. Each voxel, in the map, stores three prime information i) the occupancy, ii) semantic color, and iii) risk level that corresponds to the i) voxel's vacancy, ii) risk type, and iii) severity level, respectively.

Initially, all the voxels are assigned to be unknown. After the initial data gathering from the camera, the 3D semantic occupancy map is updated to contain $V_{free} \in V$, $V_{occ} \in V$, as well as the updated information in each voxel about the class type and severity level. V_{free} and V_{occ} are the free and occupied volume, respectively.

C. VISUAL PERCEPTION FOR USAR ENVIRONMENTS

We solve the problem of locating risk sources and identifying their severity level as a supervised machine learning task, which is exceptionally challenging because of the nature of the USAR environment. USAR environment suffers from a wide variability of structures. Risk perception is profoundly affected by the vast amount of risk types and levels, the dynamic nature of the environment, and many other factors. We only chose three different types of risks, and we dealt with their different formation by collecting large numbers of shapes orientations.

1) DATASET

To acquire such a dataset, firstly, we created objects for each type of risk. For the first type (fire hazard), the objects are images of electric plugs with a transparent background. For this type, two-level of risks have been identified, high and low. Figure 3 shows an example of the first type of risk with two levels, high and low, respectively. These objects will be placed in the simulation environment as images on the wall. For the second type (Trip Hazard), the objects are presented as rappels. These objects are presented as multiple objects on top of each other as if they are fallen. Figure 3 shows an example of type two risk with two levels high and low, respectively. These objects will be placed in the simulation environment as rubble objects, not as images as in the first type. For the third type (combustible materials hazard), the objects are images of a gas cylinder with a transparent background. Figure 3 shows an example of the third type of risk with two levels, high and low, respectively. These objects will be placed in the simulation environment as images on the wall.

In order to acquire a reliable dataset, data has been obtained for each type separately. For each type and level of risk, a unique environment is created in the gazebo simulator [29]. For example, as the environment is shown in Figure 5, it consists of objects of trip hazard risk with a high severity level. All risk objects placed in this environment are of the same type and level. A drone was used to explore every single environment and capture image frames. For each environment, a rosbag file was recorded for 5 min. A rosbag, often known as a bag, is a file type for storing ROS message data in ROS. Subscribing to one or more ROS topics and storing the received message data in an efficient file format are the primary ways to create these bags. These files may be read in various ways, including MATLAB, which can be used to

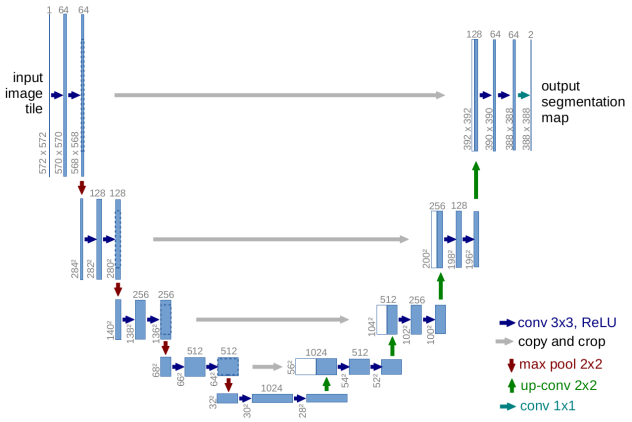


FIGURE 8. U-Net architecture. Extracted from [33].

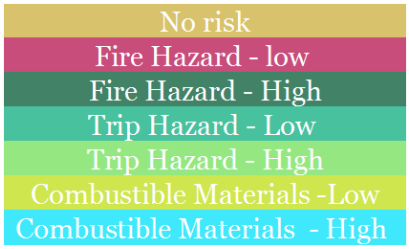


FIGURE 9. Semantic labels.

TABLE 1. Deep neural network models used for semantic segmentation.

model_name	Base Model	Seg Model	Accuracy	Loss
U-Net	Vanilla CNN	U-Net	0.9731	0.07131

TABLE 2. Deep learning models - evaluation metrics.

model name	loss	acc	v_loss	v_acc	m_IU	class_wise_IU							fre_w_IU
						1	2	3	4	5	6	7	
fcn_8	0.1380	0.9542	0.1709	0.9471	0.3154	0.9608	0.2365	0.0028	0.4086	0.0784	0.2862	0.2343	0.9131
fcn_8_resnet	**0.0573**	**0.9818**	**0.1623**	**0.9634**	**0.6199**	**0.9657**	**0.6736**	**0.2510**	**0.7378**	**0.3856**	**0.6923**	**0.6330**	**0.9402**
fcn_32	0.1686	0.9451	0.1982	0.9412	0.2620	0.9588	0.0886	0.02003	0.3376	0.0004	0.2050	0.2238	0.9070
unet	**0.1764**	**0.9452**	**0.1874**	**0.9436**	**0.2796**	**0.9595**	**0.2280**	**0.0296**	**0.2224**	**0.0279**	**0.2189**	**0.2707**	**0.9089**
unet_mini	0.2033	0.9399	0.1930	0.9415	0.2312	0.9591	0.0700	0.01203	0.1115	0.0013	0.2228	0.2417	0.9053
vgg_unet	0.2087	0.9406	0.3295	0.9125	0.1920	0.9301	0.0645	0.0024	0.1317	0.0172	0.1659	0.0324	0.87577
segnet	0.1881	0.9423	0.2094	0.9427	0.2256	0.9525	0.1133	0.0000	0.1526	0.0000	0.1499	0.2113	0.8987

acc: accuracy - v_acc: value accuracy - m_IU: mean IU - class wise IU: class wise score - fre_w_IU: frequency weighted IU

FIGURE 10. Risk images that has not been shown in the training set and their corresponding prediction.

filter and retrieve message data. Image frames were extracted and filtered to the different risk types and levels. The dataset is composed of the images acquired by a single camera. For each environment, each image is labeled, i.e., it is associated with its ground-truth class. Because of the definition of our classes, all images acquired in env 1. are of class ''risk1h''. Conversely, images acquired in env 2. are of class ''risk1l'', images acquired in env 3. are of class ''risk2h'', images acquired in env 4. are of class ''risk2l'', images acquired in env 5. are of class ''risk3l'', images acquired in env 6. are of class ''risk3h''. Images that do not contain risk are of class ''background.''

The ground truth data have been annotated using *labelme* tool [30]. Figure 6 shows an example of the annotated images using *labelme* package.The labels/classes were manually annotated using the graphical user interface provided by the package. The dataset has been augmented by [31]. The augmentation is applied to increase the size of the dataset. The dataset is augmented by using random transformations on the images and changing the color properties of the input images. The transformations can be rotation, scale, and flipping. However, changing the color can be hue, saturation, brightness, etc. The transformation is applied to both the original and the segmented labeled images. The augmentation

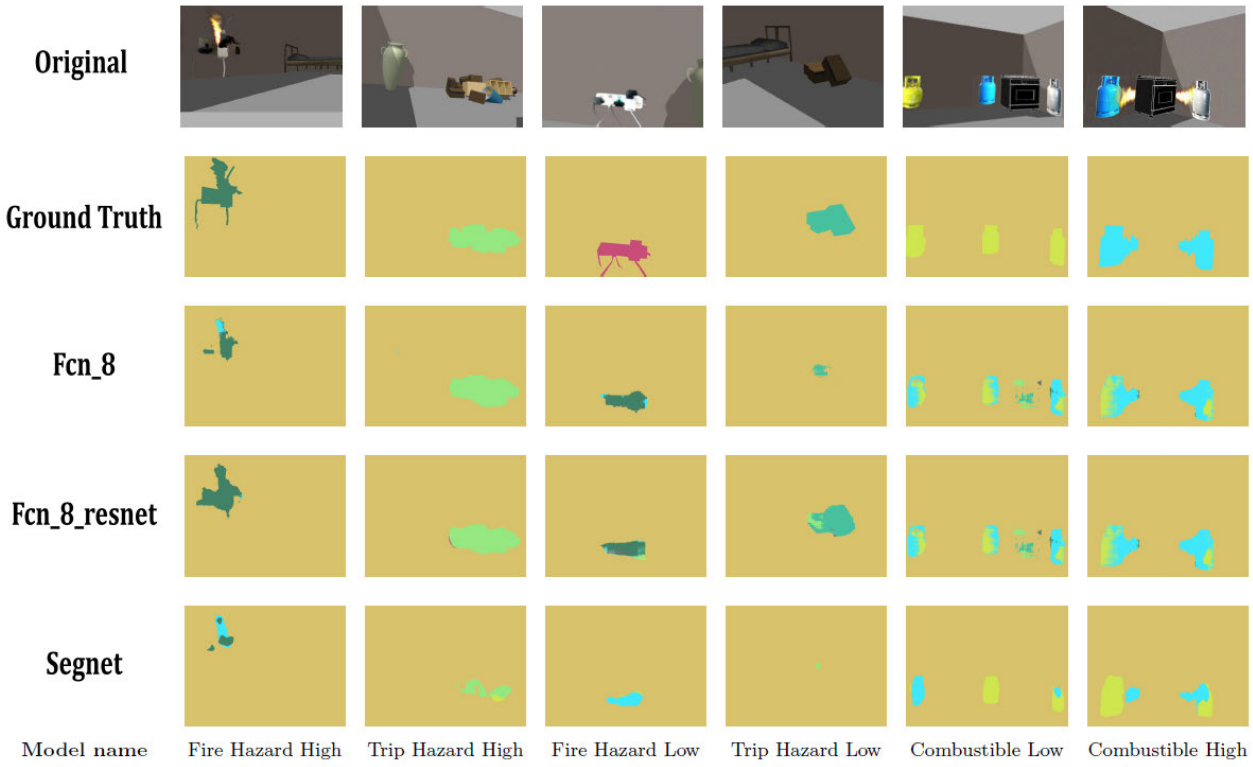


FIGURE 11. Risk images that has not been shown in the training set and their corresponding prediction.

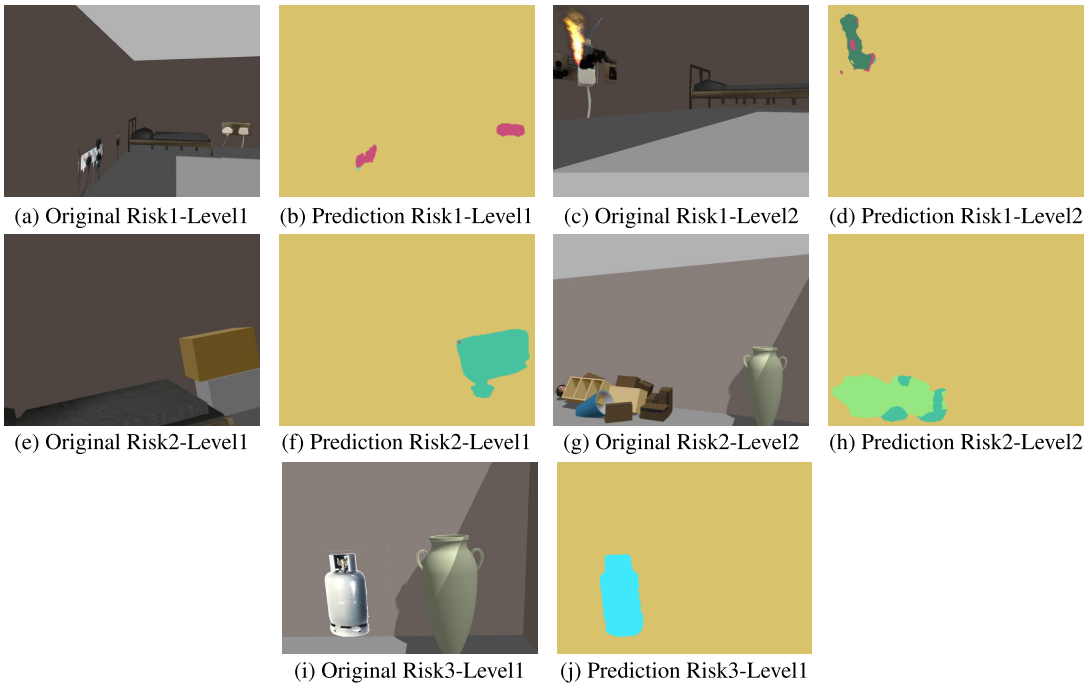


FIGURE 12. Risk images that has not been shown in the training set and their corresponding prediction.

has been performed in both the original images and their corresponding annotations. An example of dataset augmentation is presented in Figure 7. The augmentation increased the dataset from 580 images to 12180 images for both colored and corresponding annotated images.

2) EVALUATION METRICS

The deep learning model has been evaluated using classification and computer vision metrics. Classification accuracy is the most simplistic metric that is defined as the number of correct predictions divided by the total number

of predictions. However, if the objects that need to be classified in the image are small compared to the image size, then the accuracy metric does not appear to coincide with labeling objects correctly. Also, image segmentation is different from image classification. Therefore, there is a label for every individual pixel in this image rather than having one label for an image. Hence, different evaluation metric is required to evaluate the deep learning model. As an alternative to accuracy, the Jaccard index also called the IoU score (Intersection over Union), is defined as the intersection of two sets defined by their union. The fundamental idea is to see the image masks as sets. These sets can overlap in the image. If both sets are the same size, both masks are identical and overlap to 100%. Then the intersection is equal to the union. In this scenario, the IoU equals one, and it is optimal. However, in contrast, the IoU score decreases if the union gets bigger than the intersection, which happens when the predicted mask is moved or changed in size compared to the original mask.

3) DEEP LEARNING MODEL

Different deep learning models for image segmentation have been trained for five epochs, and the evaluation metrics are recorded to select the best choice for our problem (see Table 2). In addition, the results of the prediction for the different deep learning models are compared to the ground truth presented in Figure 11 and Figure 10 respectively. As the table implies, the best performance is provided by fcn_8_resnet. However, the U-Net was chosen for testing and 3D mapping because it is a more lightweight computation than the fcn_8_resnet. The various deep Image segmentation models were derived from [32] and implemented in Keras. The network parameters were adapted as well. The loss function utilized was based on [32]. Furthermore, the same loss function was employed for a fair comparison between the different networks, and it was not modified. However, to improve performance, we contemplate evaluating alternative loss functions in future work.

After creating a dataset that contains both the images and their corresponding annotated/masked images, we use the U-Net deep learning model for the semantic segmentation of 2D images. The U-Net architecture shown in Figure 8 uses an encoder-decoder structure with skip connections with symmetrical encoder and decoder layers [33].

The evaluation metrics of the U-NET model are recorded after 20 epochs, as presented in Table 1. The U-NET model is trained on a customized ''Semantic-Hazard'' dataset. Semantic-Hazard dataset includes six classes from risk type-level in an indoor environment. Semantic labels of the Semantic-Hazard dataset are shown in Figure 9.

4) TRAINING AND EVALUATION

The dataset is currently composed of 12180 frames acquired using a single camera in the Gazebo simulator in different configurations. To train the model, 20% of the dataset is used for testing; hence, the dataset has been divided into disjoint training(9744 frames) and testing (2436 frames) sets. The division was defined by carefully avoiding the same images from the same risk type appearing in both sets. The six classes are equally embodied within each set. The images and their corresponding annotations are used for dataset training and validation. The deep neural network is trained for 20 epochs using the U-Net model using an ASUS laptop (Intel Core i7 @ 2.8 GHz x 8, 16 GB RAM). The training was conducted using the package from [32].

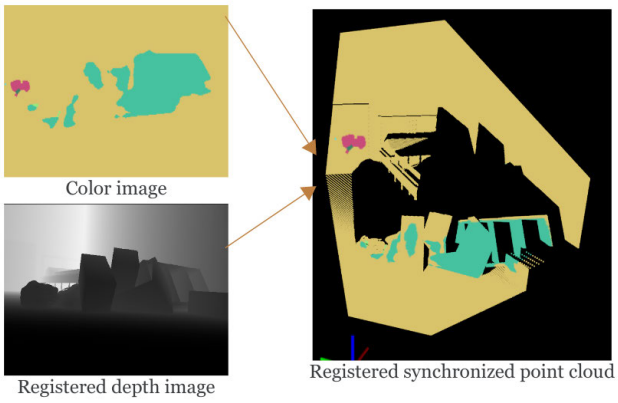


FIGURE 13. Semantic pointcloud registration.

5) TESTING

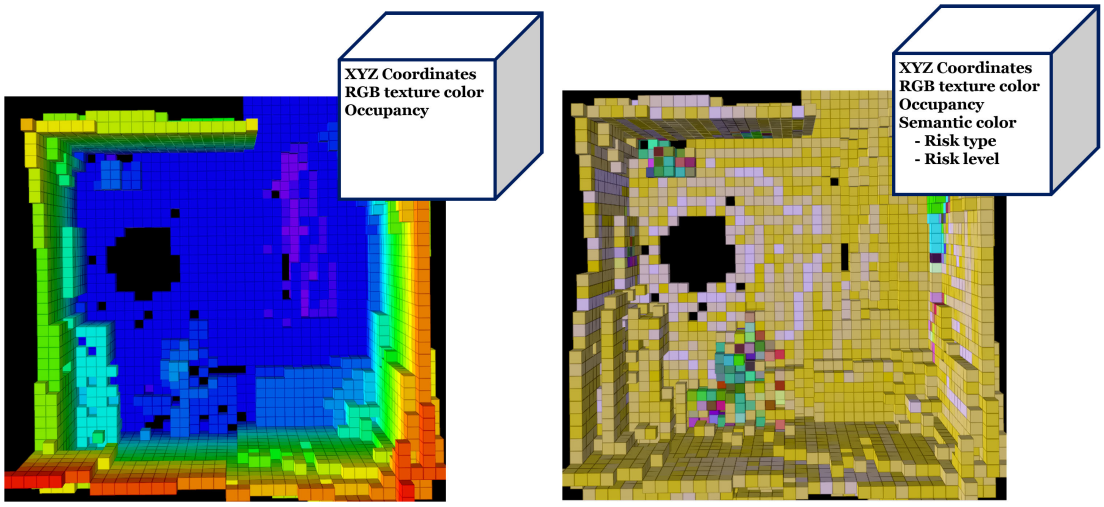
The model has been tested to predict the risk type and level using a new set of images that have not been provided earlier. An example of risk images that have not been shown in the training set and their corresponding prediction is shown in Figure 12.

D. SEMANTIC POINTCLOUD REGISTRATION

The point cloud is a set of points in an unordered structure where each point includes a coordinate in a particular reference frame. The point cloud is created by first registering a depth image with a color image with the same reference frame, generally the camera frame. Subsequently, the pixels from the color image are converted from the camera frame to the world frame to produce a point cloud using the image position, its depth, and the intrinsic parameters. In this case, the semantic image is projected in the 3D map. The colors of the 3D map represent the risk level and types. An illustration of the point cloud data structure when storing semantic information is shown in Figure 13.

E. 3D SEMANTIC MAPPING

To obtain a 3D semantic-occupied annotated risk map representation of the environment, object classes are detected and localized per depth and color image input. The proposed 3D semantic-occupied risk map structure is based



(a) 3D occupancy map (b) 3D semantic annotated risk map

FIGURE 14. Voxel's information in 3D occupancy map and 3D semantic-occupied annotated risk map.

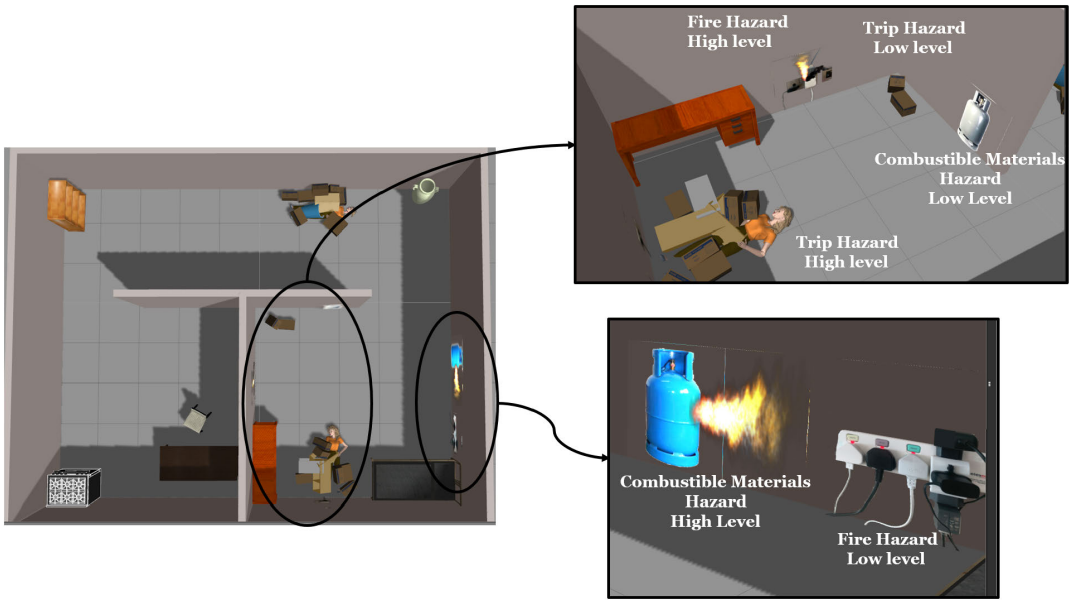


FIGURE 15. Simulation environment.

on an occupancy grid map called octomap [34] with $0.05m$ resolution. The original 3D occupancy map is presented in Fig.14a that considers the elevation of the voxels in the z-direction, while the proposed 3D semantic-occupied annotated risk map is presented in Fig.14b. The map $M = \{m_1, m_2, \ldots, m_i\}$ composed of the cubical element of the same size $m_i \leftarrow$ voxels for index i. Each voxel m_i in the proposed map accommodates volumetric and semantic information. The semantic information is the semantic color that represents the risk type and level obtained from the deep learning model.

Each voxel m_i holds the following information:

1) Position (x, y, z)
2) Probability of being occupied (value) P_o
3) Semantic color (r, g, b)

First, all voxels are assigned to be unknown. The mapping module takes both the 2D image and 3D annotated point cloud as input at every robot position and generates annotated colored voxels. The color indicates a particular class type and not the surface color. The point cloud creates the voxels in the semantic-occupied risk map. This information is provided to the map where the point cloud position and semantic color are attached to the voxels located within the camera field of view. Concurrently, the class type is assigned to the voxel using semantic color indication. The visualized built map is colored according to the object class.

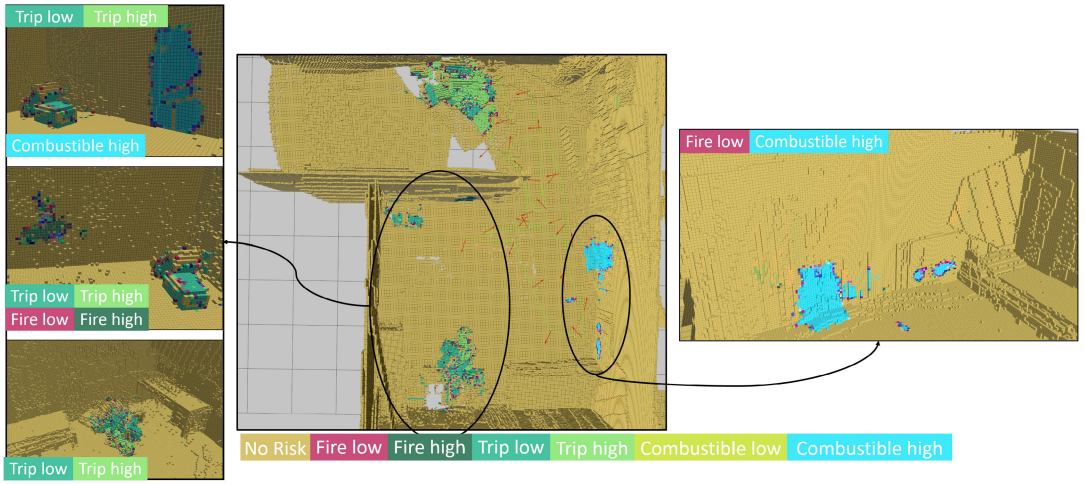


FIGURE 16. 3D semantic-occupied risk map.

F. NEXT BEST VIEW EXPLORATION

The next best view planner, a real-time exploration path planner, is employed in this work for the exploration process. It builds a tree from the current location to obtain the following position with a high exploration gain. The high gain indicates the part of the environment that has not been (entirely) explored yet. The tree is recomputed as the vehicle moves along the path, taking into account the updated information from the sensor. In this work, the next best view planner adapted from [35] is used to explore the environment.

IV. EXPERIMENTAL SETUP AND RESULTS

A. EXPERIMENTAL SETUP

The main goal of the proposed approach is to create a 3D semantic-occupied risk map that labels the types of risk and their corresponding level. An exploration technique [35] is used by the drone to navigate the environment and build the map simultaneously. The validation is performed after the exploration process terminates.

1) SIMULATION

Simulation experiments performed on an ASUS laptop (Intel Core i7 @ 2.8 GHz × 8, 16 GB RAM). The NBV framework was implemented on Ubuntu 16.04 using the Robot Operating System (ROS- kinetic) to manage message distribution and facilitate the shift to hardware. The gazebo was utilized to perform the simulation, with programming done in both C++ and Python. The simulation was performed using a UAV equipped with one RGB-D camera only. The specification of the camera is shown in Table 3. A collision box is constructed around the UAV to inspect for collisions with the environment and constrained within a work-space of size $0.25m, 0.25m, 0.25m$.

The gazebo environment, shown in Figure 15, is used as the unknown environment that the robot should create a semantic-occupied risk map for it. The simulation environment has the dimension of $9.2m$, $8.2m$, $2.5m$ of multi-connected rooms with a corridor with several risk hazards placed inside. The environment contains all types of risks and levels distributed randomly. The constructed maps are based on 3D occupancy grid using OctoMap library [34] with $res = 0.15m$ per pixel.

TABLE 3. Camera parameters.

Specification	Value		Specification	Value
Horizontal FOV	60°		Vertical FOV	45°
Resolution	0.1 m		Range	5 m
Mounting orientation (Roll,Pitch,Yaw)	(0,0,0)			

B. EXPERIMENT RESULTS AND ANALYSIS

Figure 16 shows that the proposed mapping model was able to create a 3D semantic map annotated with the risk level for risk assessment and increasing situational awareness of the environment. The different colors indicate the type of risk and its severity. The results presented in Figure 16 showed that all risk types and their severity levels were detected and located precisely in the constructed 3D semantic map. Moreover, the results showed that the system successfully distinguished all risk types and identified them with different colors than the background or the other objects presented in the environment.

For example, Figure 17 shows that the system was able to detect the risk object of type trip hazard with a high severity level. The majority of the object voxels are detected and assigned with the correct color annotation. However, the other dominant color is from the same type with a low severity level. This behavior was expected since both types share similar features for the deep learning model. In addition, the other colors that are not presented in Figure 9 are generated because of the technical development of the 3D semantic occupied map. In our development, the map resolution is 0.05, which indicates that each 5 point cloud could be located within the location of one voxel. Hence, if the point clouds in one voxel have different colors, the voxel color is the average

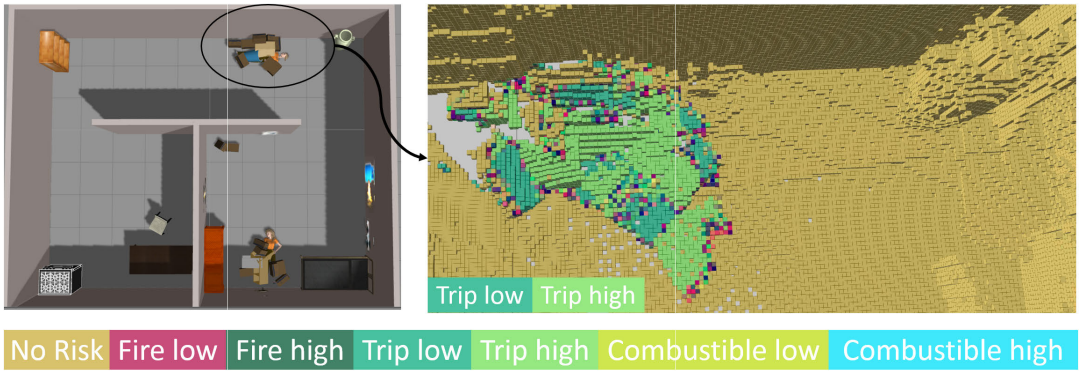


FIGURE 17. 3D semantic-occupied risk object.

color of the point cloud's colors. This behavior is presented in figure 16 where the hazards of the combustible materials are identified, and all the border voxels are with different colors than the colors presented in Figure 9.

V. CONCLUSION

This paper introduced a new approach for 3D semantic mapping of an unknown environment to increase situational awareness of the indoor environment by detecting, labeling, and mapping the different types and levels of hazards presented in USAR environments. In this work, we proposed a novel 3D semantic-occupied risk map data structure that encapsulates both occupancy and semantic risk annotations. The mapping approach uses a deep learning model to semantically segment risk types and levels, create an annotated 3D point cloud, and generate a 3D semantic-occupied risk map. Furthermore, we provided a new dataset for three different types of risks, each with two risk levels. The dataset was acquired efficiently on semi real-world sources of danger images with the corresponding annotated labels. The deep learning model was trained using the efficiently acquired dataset. Moreover, the created 3D semantic occupied risk map is used for risk assessment and evaluation. Experimental results in the simulation demonstrated that the system could create a semantic-occupied risk map that increases awareness about the hazards in the environment and facilitates the first responders' mission. The system will be tested in a real-time scenario for indoor experiments for a different neural network with multiple evaluation metrics in future work.

REFERENCES

[1] Y. Liu and G. Nejat, "Robotic urban search and rescue: A survey from the control perspective," *J. Intell. Robotic Syst.*, vol. 72, no. 2, p. 147, 2013.
[2] B. Siciliano, O. Khatib, and T. Kröger, *Springer Handbook of Robotics*, vol. 200. Springer, 2008
[3] C. H. G. Horsch, "Urban search and rescue robots: The influence of team membership of robots on team performance," M.S. thesis, Eindhoven Univ. Technol., Eindhoven, The Netherlands, 2012.
[4] S. Grayson, "Search & rescue using multi-robot systems," 2014. [Online]. Available: http://www.maths.tcd.ie/~graysons/documents/COMP47130_SurveyPaper.pdf
[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Computer Vision—ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 21–37. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-46448-0_2
[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[7] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.
[8] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
[9] I. A. Sulistijono and A. Risnumawan, "From concrete to abstract: Multilayer neural networks for disaster victims detection," in *Proc. Int. Electron. Symp. (IES)*, Sep. 2016, pp. 93–98.
[10] M. B. Bejiga, A. Zeggada, and F. Melgani, "Convolutional neural networks for near real-time object detection from UAV imagery in avalanche search and rescue operations," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 693–696.
[11] L. Zhang, L. Lin, X. Liang, and K. He, "Is faster R-CNN doing well for pedestrian detection?" in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 443–457. [Online]. Available: https://link.springer.com/article/10.1007/s10489-021-02798-1?utm_source=xmol&utm_medium=affiliate&utm_content=meta&utm_campaign=DDCN_1_GL01_metadata
[12] A. Goian, R. Ashour, U. Ahmad, T. Taha, N. Almoosa, and L. Seneviratne, "Victim localization in USAR scenario exploiting multi-layer mapping structure," *Remote Sens.*, vol. 11, no. 22, p. 2704, Nov. 2019.
[13] R. Ashour, T. Taha, J. M. M. Dias, L. Seneviratne, and N. Almoosa, "Exploration for object mapping guided by environmental semantics using UAVs," *Remote Sens.*, vol. 12, no. 5, p. 891, Mar. 2020.
[14] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, "Kimera: An open-source library for real-time metric-semantic localization and mapping," 2019, *arXiv:1910.02490*.
[15] I. Kostavelis and A. Gasteratos, "Semantic mapping for mobile robotics tasks: A survey," *Robot. Auton. Syst.*, vol. 66, pp. 86–103, Apr. 2015.
[16] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.
[17] D. Lang and D. Paulus, "Semantic maps for robotics," in *Proc. Workshop Workshop AI Robot. (ICRA)*, 2014, pp. 1–6.
[18] I. Kostavelis, K. Charalampous, A. Gasteratos, and J. K. Tsotsos, "Robot navigation via spatial and temporal coherent semantic maps," *Eng. Appl. Artif. Intell.*, vol. 48, pp. 173–187, Feb. 2016.
[19] T. Luege. (2020). *Case Study, No. 14: Using Drones to Create Maps and Assess Building Damage in Ecuador Capacity4dev*. [Online]. Available: https://europa.eu/capacity4dev/innov-aid/blog/case-study-no-14-using-dr%ones-create-maps-and-assess-building-damage-ecuador
[20] I. Jaukovic and A. Hunter, "Unmanned Aerial Vehicles: A new tool for landslide risk assessment," Oct. 2016.
[21] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, Jul. 2017, pp. 2881–2890.
[22] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 652–660.
[23] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5099–5108.

[24] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum PointNets for 3D object detection from RGB-D data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 918–927.
[25] T. Dang, C. Papachristos, and K. Alexis, "Visual saliency-aware receding horizon autonomous exploration with application to aerial robotics," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2018, pp. 2526–2533.
[26] T. Dang, C. Papachristos, and K. Alexis, "Autonomous exploration and simultaneous object search using aerial robots," in *Proc. IEEE Aerosp. Conf.*, Mar. 2018, pp. 1–7.
[27] Ccohs.Ca. (2017). *Risk Assessment: Osh Answers*. Accessed: May 6, 2020. [Online]. Available: https://www.ccohs.ca/oshanswers/hsprograms/risk_assessment.html
[28] (2020). *In a First, Indonesia is Using Lidar Drones for Disaster Recovery Efforts 2020*. [Online]. Available: https://www.terra-drone.net/global/2019/05/15/terra-drone-indonesia-lid%ar-drones-for-disaster-recovery-palu/
[29] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2004, pp. 2149–2154.
[30] *Labelme: Image Polygonal Annotation With Python*. Accessed: Dec. 12, 2019. [Online]. Available: https://github.com/wkentaro/labelme
[31] A. B. Jung *et al.* (2020). *imgaug*. Accessed: Feb. 1, 2020. [Online]. Available: https://github.com/aleju/imgaug
[32] *Image-Segmentation-Keras*. Accessed: Dec. 21, 2019. [Online]. Available: https://github.com/divamgupta/image-segmentation-keras
[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28
[34] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Auton. Robot.*, vol. 34, no. 3, pp. 189–206, 2013.
[35] A. Bircher, M. Kamel, K. Alexis, H. Oleynikova, and R. Siegwart, "Receding horizon 'next-best-view' planner for 3D exploration," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 1462–1468.

REEM ASHOUR received the M.Sc. degree in haptic teleoperation of remotely operated vehicles (ROVs) in unknown environments and the Ph.D. degree in engineering concentration in robotics from the Khalifa University of Science and Technology, United Arab Emirates, in 2015 and 2020, respectively. Her Ph.D. research investigates semantic exploration and mapping in urban search and rescue environments. She was a Research Associate with the Electrical and Electronics Department, Khalifa University, from 2015 to 2020. Since 2020, she has been a Senior Researcher with the Technology Innovation Institute, Autonomous Robotics Research Centre, United Arab Emirates. She worked as a Postdoctoral Researcher with Khalifa University, from 2020 to 2021. Her research interests include autonomous exploration, semantic mapping, deep learning, and artificial intelligence.

MOHAMED ABDELKADER received the Ph.D. degree in mechanical engineering from the King Abdullah University of Science & Technology (KAUST). His research work focused on developing real-time distributed autonomous multi-robot systems for pursuit-evasion applications which was demonstrated by a real multi-drone system with KAUST. He is currently working as an Assistant Professor with Prince Sultan University, Riyadh, Saudi Arabia. Prior to joining PSU, he held a couple of positions as a senior unmanned aerial vehicles engineer and the technical team lead for a couple of startups (Algorythma and Systemtio) in Abu Dhabi, United Arab Emirates, where he led projects for developing autonomous aerial robots for security and surveillance. He also worked as a Lab Scientist with the Research & Development Center, Saudi Aramco, where he contributed to the development of intelligent industrial inspection robots supported by over 12 patent applications. His main research works are in the development of intelligent swarm robotic systems and autonomous navigation.

JORGE DIAS (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Coimbra, Portugal. He is currently a Full Professor with the Department of Electrical Engineering and Computers, Institute of Systems and Robotics (ISR-UC), University of Coimbra. He is also a Full Professor with Khalifa University, Abu Dhabi, United Arab Emirates. His expertise has been in the area of artificial perception (computer vision and robotic vision) and has contributions on the field, since 1984. He has been the principal investigator and a consortia coordinator of several research international projects, and coordinates a research group on artificial perception for intelligent systems and robotics with ISR-UC. He was the Director of the Laboratory of Systems and Automation, Instituto Pedro Nunes (IPN), of which he was the Vice-President, from June 2008 to June 2011. He published several articles in the area of computer vision and robotics that include more than 130 publications in international journals, one published book, 22 books chapters, and more than 360 articles in international conferences with referees.

NAWAF I. ALMOOSA is currently an Assistant Professor at the School of Electrical Engineering and Computer Science (EECS), Khalifa University of Science and Technology (KUST), with a joint appointment as the Acting Director of the Emirates ICT Innovation Center (EBTIC); an ICT-focused research center founded by KUST, Etisalat, and British Telecom. At EBTIC, he leads the Smart Infrastructure team, which tackles research challenges in network synchronization, dynamic management of radio access networks, and AI-based radio signal analysis. His research interests include distributed optimization and control applied for energy-efficient computing, as well as telecommunications networks and robotics.

TAREK TAHA received the M.Eng. degree in computer control from the University of Technology Sydney, Australia, in 2004, and the Ph.D. degree from the Centre of Excellence for Autonomous Systems (CAS), University of Technology Sydney, in 2012. He worked as a Senior Mechatronics Engineer with Sydney-based Engineering Research and Development Company, from 2008 to 2013, before joining Khalifa University, from 2014 to 2018. Then, he led the Algorythma's Autonomous Aerial Laboratory, before joining Dubai Future Foundation to lead the Robotics Laboratory. His research interests include autonomous exploration, navigation and mapping, machine vision, human–robot interaction, assistive robotics, and artificial intelligence.