

UNIVERSITY OF BIRMINGHAM

SEMANTIC INFORMATION IN HUMAN-ROBOT TEAMS FOR EXTREME ENVIRONMENTS

by

Tianshu Ruan

A thesis submitted to

the University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

Extreme Robotics Lab

School of Metallurgy and Materials

College of Engineering and Physical Sciences

University of Birmingham

July 2025

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

© Copyright by TIANSHU RUAN, 2025

All Rights Reserved

Abstract

This thesis explores the role of semantic-level information in Situational Awareness (SA) during robot deployments into hazardous environments. In such applications, robots are supervised remotely by a human operator who must remain in a safe zone, distant from the robot and its hazardous surroundings. In particular, we consider Human–Robot Teams (HRT) and Human–Robot Interaction (HRI) in the context of a Variable Autonomy (VA) framework, where the Level of Autonomy (LoA) of the robot can be changed dynamically during missions, varying between direct teleoperation by the human, and fully autonomous behaviours.

We begin with a discussion of different kinds of real-world deployments of robots for First Responder (FR) and Extreme Environment (EE) industries, and examine the types of semantic information that enable the operator to develop SA and understand the remote situation. We propose a taxonomy framework for semantics, describing low-level semantics, high-level semantics, and the relationship between these semantics and raw sensor data, mission tasks, and context.

Building on the taxonomy, we propose a semantic indicator metric and a framework that enables the robot (and its autonomous agent) to reason about the Situational Semantic Richness (SSR) of a scene. The framework shows how different types of semantic information (e.g. signs of human presence in Search and Rescue (SAR) missions, or Light Detection and Ranging (LiDAR) noise during navigation tasks) can each generate a semantic indicator that might be important for SA. These different indicators can then be combined, in a systematic way, to generate an overall SSR score for the current scene, in real-time during missions. This SSR score can then be used to flag

semantically-rich situations that may require the high-level reasoning and attention of a human operator. This can help in e.g. guiding LoA changes in a VA paradigm.

Subsequently, we apply the semantic richness framework in a SAR mission in a mock-up disaster environment, where we incorporate the real-time semantic indicators into a Graphical User Interface (GUI) on the human operator's control screen and the SSR score processing in the background. We investigate the impact of these semantic indicators on HRI, and on the overall mission performance of the HRT, the usefulness of the semantics-based SA framework, by successfully lowering the human's cognitive workload and increasing human's trust in the VA system. Moreover, some additional interesting findings were noticed in the experiments, e.g. a correlation between the LoA preference and trust in the robotic system.

Finally, we survey expert FRs and operators in EE industries to understand their perceptions, needs, and views with respect to semantic information. Their feedback is consistent with the findings of our research, and confirms the importance of semantic information in future robot developments and deployments.

In summary, the main contributions of the thesis include: a taxonomy framework for systematically reasoning about semantics; metrics and methods for autonomously computing the SSR of different scenes in real-time; an analysis of the impact of incorporating the SSR indicators into VA HRI; and insights and findings of interest for incorporating semantic information into future HRT/HRI developments.

Publication and author contributions

Part of the technical material in this thesis has now been published, and other parts are in preparation for publication as papers, at the time of writing. In this section, we list these publications and potential publications and explain the contributions of each author. The work in Chapter 3 has been published in: 2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR) (**P.1**). The work in Chapter 4 has been accepted by 2025 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR) (**P.4**) and will be published in the near future. The work in Chapter 5 has been accepted by the journal Frontiers in Robotics and AI (**P.2**) and will be published in the near future. The work in Chapter 6 has been accepted by the 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (**P.3**) and will be published in the near future. I am the first author and the major contributor to these papers. My supervisors, Dr. Manolis Chiou, and Prof. Rustam Stolkin provided high-level supervision, including discussing my work and ideas with me, making suggestions for how I might modify my experimental designs, and provided proof-reading and suggestions about writing. The exact contributions of each co-author on each paper are provided in more detail below.

P.1 Ruan, Tianshu, Hao Wang, Rustam Stolkin, Manolis Chiou. (2022). "A Taxonomy of Semantic Information in Robot-Assisted Disaster Response". In: 2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), pp. 285–292. I investigated and organized the literature and proposed the taxonomy. Our previous internship student, Hao Wang, helped me with the editing. Dr. Manolis Chiou, and Prof. Rustam Stolkin provided supervision, including discussing my ideas for the taxonomy, proof-reading, and suggestions about writing and presentation.

P.2 Ruan, Tianshu, Aniketh Ramesh, Hao Wang, Alix Johnstone-Morfoisse, Gokcenur Altindal, Paul Norman, Grigoris Nikolaou, Rustam Stolkin, and Manolis Chiou. (2025). "A Framework for Semantics-based Situational Awareness during Mobile Robot Deployments." Accepted by the journal Frontiers in Robotics and AI. In preparation, pre-print draft available at arXiv: 2502.13677 [cs.RO]. URL: https://arxiv.org/abs/2502.13677. I designed the framework, conducted the experiments, analyzed the results, and wrote the paper. Aniketh Ramesh helped by discussing my ideas with me. Aniketh had previously published some papers on how to convert raw sensor data into metrics relating to a robot's real-time task performance. By analogy, these discussions helped me to clarify my ideas on choosing appropriate mathematical approaches for converting raw sensor data into indicators and metrics for representing the "semantic richness" of scenes in which the robot is operating. Hao Wang helped me to train a computer vision tool for automatic detection of Hazardous Material (HAZMAT) signs. I used this tool to help me design a semantic indicator relating to hazards. Alix Johnstone-Morfoisse, Gokcenur Altindal, and Prof. Paul Norman are experts in radiation from the university's physics department. They assisted me with experiments in which our robot used a radiation detector to detect radiation hazards. They provided the radiation sources for my experiments, helped us ensure that our experiments were conducted in accordance with radiation safety rules, and helped me edit the radiation-related content in the paper. Prof. Grigoris Nikolaou helped me edit my draft paper. Dr. Manolis Chiou, and Prof. Rustam Stolkin provided high-level supervision, including discussing my overall work and ideas with me, making suggestions for how I might improve my experimental designs, and providing proof-reading, editing, and suggestions about writing and presentation.

P.3 Ruan, Tianshu, Aniketh Ramesh, Rustam Stolkin, Manolis Chiou. "An Exploratory Study on Human-Robot Interaction using Semantics-based Situational Awareness". Accepted by the 2025 International Conference on Systems, Man and Cybernetics (SMC). In preparation, pre-print draft available at arXiv: 2507.17376v2 [cs.RO]. URL: https://arxiv.org/abs/2507.17376. I designed the experiments, analyzed the data, and wrote the paper. Aniketh Ramesh gave me help and advice on setting up the mock-up environment scenes in which a robot would be op-

erated by human trial participants during experiments. Dr. Manolis Chiou, and Prof. Rustam Stolkin provided overall supervision. In particular, Dr. Chiou provided discussions and advice on designing effective experiments with human participants. Both supervisors also contributed to proof-reading, editing, and suggestions about writing and presentation.

P.4 Ruan, Tianshu, Zoe Betta, Georgios Tzoumas, Rustam Stolkin, Manolis Chiou. (2025). "A Study about First Responder User Requirements for Semantic Information". Accepted by the 2025 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). Pre-print draft available at arXiv: 2510.16692 [cs.RO]. URL: https://arxiv.org/abs/2510.16692. This paper is based on a survey conducted with expert FRs. I designed the questionnaires, disseminated the questionnaires, and analyzed the feedback. Zoe Betta and Georgios Tzoumas helped me look for FRs who I could approach to participate in my survey, and provided advice about designing survey questionnaires. Dr. Manolis Chiou, and Prof. Rustam Stolkin provided overall supervision, including: discussing my ideas for the survey of end-users, helping me make contact with some important FRs and industry experts to participate in the survey, proof-reading and editing the questions in my questionnaire document, and proof-reading the final paper.

Dedicated to my parents, **Huafen He** and **Derong Ruan**.

Acknowledgments

Pursuing a PhD. can be a solitary but fulfilling endeavor. It is only four years, but definitely it becomes one of my most rewarding and exhilarating experiences so far. I greatly cherish what I experienced, what I learned, and what I missed on this journey. I truly feel my growth and differences compared to what I was four years ago. Many people accompanied me, long or short, on this journey. At the very beginning of the thesis, I would like to thank for the academic support from my supervisors, **Dr. Manolis Chiou** and **Professor Rustam Stolkin**. They provided essential supervision and ideas on my research topics and writing. As a foreign student who does not speak English as a mother language, my supervisors devoted lots of effort to improving my writing, communication, and presentation skills. Apart from my supervisors in the lab, I would like to appreciate the advice and help from my colleagues **Aniketh Ramesh**, **Hao Wang**, and **Cesar Conteras**. I solved technical issues with them. There is no doubt that I would encounter a lot more problems to overcome if they were not there. Moreover, my **friends and housemates** gave me huge physical and mental support during the four years. My PhD. program started in the middle of the pandemic. At each period of time, I was taken good care of by them. They led me to experience the activities that I would never have thought of. Together, we went to bars, we went clubbing, we went to concerts, we went kart racing, we enjoyed the fancy meals... We explored Birmingham, UK, and even the European continent. These experiences will stay with me for life. It is a pity that friends cannot always participate in my journey. I hope we can reunite somewhere in the future. Last but not least, none of this would have been possible without the unwavering encouragement of my parents. They provided huge support to my academic life and every decision I made. They were my sturdy shield and cozy shelter when I was low. This

thesis is for them and all the people I love.

Almost forgot appreciating the following electronic opium companions in the relaxing nights: Bilibili, Steam, QQ music, and entertaining political shows from Taiwan.

Science is a way of trying not to fool yourself. The first principle is that you must not fool yourself, and you are the easiest person to fool.

– Richard Feynman

Contents

Abstract i

Publication and author contributions iii

Acknowledgments ix

Contents xiii

List of Figures xvii

List of Tables xxi

List of Acronyms xxiii

1 Introduction 1

1.1 Background 1

1.2 Scenario 3

1.3 Robot deployments in Extreme Environments 3

1.4 From raw information to high-level semantics 6

1.5 Challenges of addressing high-level semantics 8

1.6 Contributions of this thesis 10

1.7 Thesis structure 11

2 Background and Literature Survey 13

2.1 Human-Robot Teams 13

2.1.1 Human and robot perception . . . 14
2.1.2 Human-Robot Interaction . . . 16
2.1.3 Variable Autonomy . . . 17
2.2 Situational Awareness . . . 20
2.2.1 Framework . . . 21
2.2.2 Methods to deliver and enhance Situational Awareness for Human-Robot Teams . . . 22
2.3 Semantic information for Human Robot Teams . . . 23
2.3.1 RGB-based semantics . . . 24
2.3.2 Multi-sensor-based semantics . . . 26
2.4 Measurements and metrics . . . 28
2.4.1 Human factors . . . 28
2.4.2 Data analysis . . . 29

3 A Taxonomy for Semantics in High-hazard Applications 33
3.1 Common tasks for robots in disaster response . . . 35
3.2 Semantic information taxonomy . . . 37
3.2.1 Low-level semantics . . . 37
3.2.2 High-level semantics . . . 39
3.2.3 Relationship among low-level semantics, high-level semantics, and context 40
3.3 From the taxonomy to implementation . . . 42
3.3.1 Insight I . . . 42
3.3.2 Insight II . . . 47
3.4 Limitations . . . 50
3.5 Conclusions . . . 50

4 First Responder User Requirements for Semantic Information 53
4.1 Methodology . . . 55
4.2 FR background and demographics . . . 56

4.3 General attitudes towards robots . . . 57
4.4 Views about SA and use of semantic information . . . 60
4.5 Analysis and discussion . . . 63
4.6 Conclusion . . . 69

5 A Situational Semantic Richness Framework for Mobile Robot Deployment 71
5.1 Individual semantic indicators . . . 73
5.1.1 Laser noise intensity . . . 73
5.1.2 Risk to robots . . . 74
5.1.3 Sign of human activity . . . 75
5.1.4 Radiation . . . 76
5.2 Overall Situational Semantic Richness score: aggregating semantic indicators . 77
5.3 Experimental validation . . . 80
5.3.1 Vision system design . . . 81
5.4 Results Analysis . . . 87
5.4.1 Experiment I . . . 87
5.4.2 Experiment II . . . 89
5.5 Discussion . . . 93
5.6 Conclusion . . . 95

6 Impact of High-level Semantics on Variable Autonomy Human-Robot Teaming 103
6.1 Aims and Hypotheses . . . 105
6.2 Experimental setup and apparatus . . . 106
6.2.1 Arena . . . 106
6.2.2 Mission and Tasks . . . 108
6.3 Experimental design and procedure . . . 109
6.4 Evaluation with a high-fidelity simulated robot . . . 112
6.4.1 Significance analysis . . . 112
6.4.2 Correlation analysis . . . 113

6.4.3 Discussion and insights 115
6.5 Limitations 119
6.6 Conclusion 120

7 Conclusion and Future Work **121**
7.1 Contributions 122
7.2 Future work 124
7.3 Closing thoughts 126

References **129**

A Questionnaires for FR about Robotic Mission **149**
A.1 Introduction 149
A.2 Background Experience and Demographics 150
A.3 General attitudes to robots 152
A.4 Situational Awareness (SA) and use of semantic information 153

B Participants' Feedback from Experiments in Chapter 6 **157**

List of Figures

2.1 Relationship between different modes of VA and some example applications: adjustable autonomy and sliding autonomy can be seen as types of MI, which is distinct from shared control. 18

2.2 Samples of image classification, object detection, semantic segmentation, and instance segmentation (X. Wu, Sahoo, and Hoi, 2020). 25

3.1 The relationship between the low-level semantics and raw data. 38

3.2 The proposed high-level semantics in relation to the common disaster robotic tasks. 39

3.3 Different levels of information abstraction: each level of information is supported by the lower levels of information. 41

3.4 Two aspects of RQ3: the blue circle refers to existing techniques (See Insight I) and the red circle refers to useful semantics (See Insight II). The intersection region illustrates the current semantics understanding technique developments on useful disaster response semantics. 42

4.1 Distributions of role, field experience, education level, and robot deployment level by age group. Robot deployment level refers to the frequency of using robots in the mission. Higher level means higher frequency. 57

4.2 Distribution of general attitudes towards robots on personal and societal levels. Most points are under the diagonal line (green zone). This indicates that most FRs have positive attitudes towards robots in both personal and societal contexts. 59

4.3 Score distributions of numerical questions: Q43, Q47, Q49, Q50, Q53, and Q54. 61

4.4 Accuracy threshold values for achieving trust and usefulness: light blue bars refer to the accuracy threshold for achieving trust (Q55), and dark blue bars refer to the accuracy threshold for achieving usefulness (Q56). 63

4.5 Score distribution by role: crosses refer to the median in the group. In each question, the roles displayed from left to right are operator, supervisor, commander, and domain specialist. 64

4.6 Score distribution by experience level. 64

4.7 Score distribution by education level. In each question, the education level displayed from left to right is high school, bachelor, master, and PhD. 64

4.8 Important types of semantics for FRs during missions according to the FR's responses. 65

5.1 Mobile robot semantic SA with two levels. 72

5.2 Semantics-based SA framework: light blue box refers to low-level semantics, cornflower blue box refers to high-level semantics, and dark blue box refers to context. Black dashed lines indicate the potential connections among different levels. 78

5.3 The structure of the vision system. The RGB-D camera could provide both colored and depth images. A color image is fed to YOLACT to generate a detected mask and a segmented instance. The depth image is aligned with the colored image to generate the estimated depth of the object. 81

5.4 13 classes of HAZAMATs signs (Sharifi, Zibaei, and Rezaei, 2020). 82

5.5 Training results when using ResNet-50 backbone. (H. Wang, 2022) 84

5.6 Validation results when using ResNet-50 backbone. (H. Wang, 2022) 84

5.7 Samples of YOLACT personal belongings detection, e.g. hat, laptop, watch, mask, wallet, and others. Large objects can be detected from a long distance, e.g. backpack. Small objects like watches are detectable in short distances. . . . 86

5.8 Example of the raw image (a) and the mask image (b) generated by using YOLACT to detect common objects. (H. Wang, 2022) 87

5.9 (a): Layout for Experiment I. The dark blue box area (scenario 1) on the ground is used for laser noise or the radiation source (uranium rock). The yellow box area (scenario 2) is used for hazmat signs or personal belongings; (b): Layout for two environment features scenarios of Experiment II; (c): Layout for three environment features scenarios of Experiment II. In the picture, some environment features are covered by red barriers. 88

5.10 SSR and environment semantics intensity timeline. The lines refer to the average SSR environment semantics intensity in 5 runs. The shade areas show the minimum and maximum range. Zones between dashed lines refer to the corresponding semantics detected from the environment. 90

5.11 Radiation-noise combination semantics intensity timeline. Zones between dashed lines indicate detected semantic signals. 96

5.12 Radiation-risk combination semantics intensity timeline. Zones between dashed lines indicate detected semantic signals. 97

5.13 Risk-noise combination semantics intensity timeline. Zones between dashed lines indicate detected semantic signals. 98

5.14 SHA-noise combination semantics intensity timeline. Zones between dashed lines indicate detected semantic signals. 99

5.15 SHA-radiation combination semantics intensity timeline. Zones between dashed lines indicate detected semantic signals. 100

5.16 Three semantics combination intensity timeline. Zones between dashed lines indicate detected semantic signals. 101

6.1 Test arena with semantic indicators. Colored boxes denote zones where different semantic situations will arise, resulting in different indicators being displayed to the human operator (trial participant). Red circles mark the location of the victims. The purple dotted line is the robot's pre-set autonomous navigation trajectory. 106

6.2 Test arena for trials without semantic indicators. Note that we need to use two different arena layouts for the two different trials, to minimise the confounding factor of learning effects between trials. 107

6.3 Parallel task example. Are these objects identical or mirror images? 109

6.4 GUI for trials without semantic indicators: participants obtain SA from a pre-scanned LiDAR map, camera feed, real-time localisation, radiation readings, current robot LoA, and countdown of remaining mission time. The bubble is used for clearly presenting the semantic indicator scores in the figure. It does not show up in the GUI during the experiments. 110

6.5 GUI for trials with semantic indicators: three semantic indicator scores are displayed on the interface screen, in addition to the mapping and video data. . . . 111

6.6 Spearman's ρ correlation heatmaps of experimental trials with and without semantics. 114

6.7 Comparison of task performance metrics between trials with and without semantic indicators (Part 1). 117

6.8 Comparison of cognitive and interaction metrics between trials with and without semantic indicators (Part 2). 118

6.9 Teleoperation ratio vs. trust: red dashed curve is no-semantics condition. Blue dashed curve is with-semantics condition. 119

List of Tables

2.1 Comparison of different perception-based semantics 27

3.1 Indicative Summary of Advanced Semantic Processing Techniques (Part 1) . . 45
3.2 Indicative Summary of Advanced Semantic Processing Techniques (Part 2) . . 46
3.3 Current status of processing important semantics in the scope of common disaster response tasks 48

5.1 Performance comparison among different training setups and backbones. (H. Wang, 2022) 85
5.2 Sensitivity analysis ($S1$, St) of environment semantics with the SSR score in Experiment I 89
5.3 Combinations of environment semantics in different scenarios in Experiment II 91
5.4 Spearman's rank correlation coefficient (ρ) of environment semantics with the SSR score and bootstrap 95% Cl in Experiment II 92
5.5 Kendall's rank correlation coefficient (τ) of environment semantics with the SSR score and bootstrap 95% Cl in Experiment II 93

6.1 Participants' gender and background 110
6.2 Significance analysis using Wilcoxon signed-rank test (n = 14) 113

List of Acronyms

AI Artificial Intelligence. 2, 3, 8, 13–17, 19, 20, 22, 33–35, 42, 43, 50, 67–69, 72, 95, 121, 122

AR Augmented Reality. 22, 23

CNN Convolutional Neural Network. 15, 26

CV Computer Vision. 8, 10, 23, 24, 33–35, 50, 72, 95, 104, 121, 122

EE Extreme Environment. i, ii, 2–6, 10, 11, 14, 33, 34, 53, 54, 58, 121, 122, 124, 127

ER Exploration and reconnaissance. 36, 38, 45–47, 49

FD-NPP Fukushima Daiichi Nuclear Power Plant. 1, 2, 4, 50

FPS Frame Per Second. 25, 85

FR First Responder. i, ii, v, 6, 10, 11, 50, 53–61, 63, 65–69, 74, 121, 122, 124

GUI Graphical User Interface. ii, 110

HAZMAT Hazardous Material. iv, 5, 36, 37, 62, 74, 82, 85, 88, 89, 91, 122–124

HRC Human–Robot Collaboration. 17, 18, 20, 103

HRI Human–Robot Interaction. i, ii, 2, 8, 11, 12, 14, 16, 17, 20, 22, 29, 34, 69, 104–106, 120, 121, 123

HRT Human–Robot Teams. i, ii, 2–14, 16, 17, 19, 21, 23, 24, 28, 29, 69, 71, 73, 94, 103–106, 108, 116, 119–123, 125, 126

IoU Intersection of Union. 83, 85

LiDAR Light Detection and Ranging. i, 9, 15, 22, 27, 28, 38, 43, 45–47, 80, 105, 111, 124

LoA Level of Autonomy. i, ii, 3, 11, 16–18, 105, 106, 108, 109, 112, 113, 115–118, 120, 122–125, 127

mAP mean Average Precision. 25, 83, 85

MI Mixed Initiative. 17, 18, 125, 126

NASA TLX NASA Task Load Index. 29, 112, 113, 115, 116

NLP Natural Language Processing. 23, 27, 33, 43, 122

ROS Robot Operating System. 80, 85

SA Situational Awareness. i, ii, 2–7, 10, 11, 14–16, 20–24, 28, 29, 33, 34, 38, 51, 53–55, 60, 63, 65–69, 71–73, 77, 78, 93–95, 103–105, 112, 113, 116, 118–123, 126

SAR Search and Rescue. i, ii, 5, 8, 11, 12, 14, 54, 103, 105, 108, 121, 123

SHA Signs of Human Activity. 40, 49, 73, 75, 85, 87, 104, 105, 109, 111

SLAM Simultaneous Localization and Mapping. 1, 2, 4, 26, 27, 35, 43, 81, 109

SOTA State-Of-The-Art. 6, 24, 42, 47, 85, 123

SSR Situational Semantic Richness. i, ii, 11, 69, 71, 78–80, 88, 89, 91–95, 124

TSA Teams SA. 21–23

UAV Unmanned Aerial Vehicles. 3–5, 36, 94

UGV Unmanned Ground Vehicles. 3–5, 36, 43, 44, 73, 94

USV Unmanned Surface Vehicles. 4, 36

UUV Unmanned Underwater Vehicles. 3, 36

VA Variable Autonomy. i, ii, 3, 6, 11, 14, 17–20, 103–105, 111, 125

YOLACT You Only Look At CoefficienTs. 24, 25, 80, 82, 85, 86, 122

Chapter 1

Introduction

1.1 Background

At the time of writing this thesis, nearly fifteen years have elapsed since the tsunami and major nuclear accident at Fukushima Daiichi Nuclear Power Plant (FD-NPP) in 2011. Although the exclusion zones around FD-NPP have been reduced in recent years, the barren and abandoned farmland and houses in the area still remind people that the huge impact of the accident is continuing. During this doctoral research, we had the opportunity to join a collaborative project led by the Japan Atomic Energy Agency (JAEA) and the Tokyo Electric Power Company (TEPCO), developing a networked team of crawler robots to explore damaged buildings at the FD-NPP accident site. Our role was to assist the Japanese team with developing Simultaneous Localization and Mapping (SLAM) and autonomous navigation approaches, to enable the robots to capture accurately located radiation measurements, and also to autonomously navigate back out of the deployment site, if teleoperative communications broke down between the robot and the remote human operator. These robots were initially tested in FD-NPP Unit 5 (relatively undamaged) in 2024, and will progress to a deployment in Unit 2 (the site of an explosion) during 2025-2026.

The above opportunity provided valuable insights into the history of robot deployments at FD-NPP, working directly with Professor Shinji Kawatsuma, who led the emergency response robotics team during the 2011 accident (Kawatsuma, Fukushima, and T. Okada, 2012; Nagatani

et al., 2013). During the events of 2011, several HRT entered this hazardous zone with a variety of robots. Many of them were not specially customized for nuclear accidents (e.g. US military robots and remote control construction machines). A number of robots became stuck and had to be abandoned inside the damaged reactor buildings, because of snagging of umbilical tethers, interrupted communications, or electronic equipment being penetrated by high-dose gamma rays. Some of the more successful deployments were of remotely operated construction machines, outside of the reactor buildings.

In the years since the FD-NPP accident, robot deployments have continued to extend into the more difficult areas of the accident site. More robots have been designed for EEs (dark, irradiated, uneven terrain, communication restriction, high temperature, leaking chemicals). These have participated in a wide range of tasks, including radiation monitoring, wastewater air sampling, debris cleaning, decontamination, etc. During the same period of time, the capabilities of autonomous robots have developed very rapidly. Autonomous robots are now becoming able to robustly perform a variety of complex navigation and manipulation tasks in non-extreme environments, which were not feasible (or were relatively unreliable) in 2011, e.g. SLAM, navigation, cutting, grasping, and manipulation, and multi-robot tasks.

However, with a few exceptions, these developments have still not significantly penetrated into EE applications and high-consequence industries, which are understandably conservative and risk-averse. Hence, robotic missions in EEs predominantly remain directly teleoperated, in ways which have not changed fundamentally in several decades, i.e. joystick control with CCTV camera feed.

Robots (to avoid verbosity we sometimes use the term "robot" synonymously with the Artificial Intelligence (AI) or autonomous agents used to control the robot) are deployed as human-controlled tools or sensory instruments, rather than as intelligent partners in a HRT, i.e. the HRI is restricted to humans sending basic teleoperation commands, while receiving video images. Meanwhile, these mission tasks and environments are exceptionally complex and challenging, presenting severe difficulties for SA, scene understanding, decision-making, and control by human operators. There is great potential to alleviate these difficulties by leveraging the

past ten years of advances in autonomous robotics. However, there is still comparatively little research that addresses how to combine such autonomous functionalities with a human in-the-loop paradigm (required by such industries), i.e. in HRT.

1.2 Scenario

In this thesis, we envisage an HRT, comprising a human using control interfaces to interact with one or more remotely located robots equipped with various functionalities, tools, and capabilities. The mission involves remote access, exploration, and inspection in an unstructured environment. The complementary strengths of human reasoning and AI competencies must be combined optimally to understand scenes, make decisions, plan, and execute actions. Additionally, the team must decide on roles dynamically, i.e. whether a human or an autonomous agent will control a particular task in a VA paradigm, where the system can switch between different LoA dynamically throughout the mission. A human might take control (teleoperation) for a task that is too critical or complex for the agent. Conversely, control might be assigned to the agent if the human is stressed, busy with a secondary cognitive task, or feels that the agent can control the robot best during a particular part of the mission. Hence, the goals and actions of humans and autonomous agents within the HRT must be aligned. More specifically, the humans and robots (i.e. the HRT) should share a common SA, including sensory data, low-level and higher-level "semantic" understanding of the situation. This thesis explores the role of semantic information in such HRT contexts.

1.3 Robot deployments in Extreme Environments

Robots have been used in special missions, e.g. disaster response missions for decades. These missions require the HRT to have the capability to reach EEs and efficiently complete a variety of tasks. The wide range of robot deployments, including Unmanned Aerial Vehicles (UAV), Unmanned Ground Vehicles (UGV), and Unmanned Underwater Vehicles (UUV), extensively mitigates the accessibility problem. However, the efficient collaboration between humans and

robots remains challenging in the field.

The first deployment of robots in a real urban disaster response was the 9.11 event (Robin R Murphy, 2004a). Tracked robots (UGV) were used to find victims under the bricks/debris and provide additional SA via camera feed for task planning. This deployment, and the subsequent study by (Robin R Murphy, 2004a), was a seminal moment in the study of HRT in EEs. In the 2011 FD-NPP accident, a variety of UGVs were deployed to inspect the situation in damaged reactor buildings and assess damage (Kawatsuma, Fukushima, and T. Okada, 2012). With many abandoned robots in the power unit, researchers learned lessons of how to deploy robots in such difficult scenarios, e.g. they emphasized the importance of communication quality and radiation shielding to protect the electronic devices in these EEs. Moreover, Defense Advanced Research Projects Agency (DARPA) organized a robotics challenge in 2015 (Karumanchi et al., 2017), which involved multiple tasks in a mock-up EE, e.g. SLAM, locating objects, climbing stairs, opening doors etc. New models of robots, e.g, quadruped robots, participated in the challenge. The challenge explored the capability of robots and overall HRT performance in complex tasks and missions.

In addition to these high profile events, robots have participated in missions after water floods, earthquakes, hurricanes, landslides, forest fires, and other special missions, e.g. mine detection (Schneider et al., 2015).

More specifically, UAVs have been widely employed after floods and tsunamis. They conduct operations including surveying environmental impacts, mapping floods (R. Murphy et al., 2016), ground search, and damage estimation using cameras (Fernandes, R. Murphy, Merrick, et al., 2019). On some occasions, they helped to re-locate mines and provide a 3D map of the environment when a minefield shifted after a secondary disaster (e.g. mudslides) happened (De Cubber et al., 2014; Balta et al., 2015). In these cases, the robot's working space is not as extreme as in other cases, e.g. firefighting. Autonomous navigation can be employed. However, humans still tended to choose human control throughout such missions. Moreover, Unmanned Surface Vehicles (USV) have been employed to provide water SA, damage assessment, and deliver supplies after floods and tsunamis (Scerri et al., 2011).

Considerable research has been devoted to robot applications in earthquakes and hurricanes. UAV and UGV deployments have become increasingly frequent. Apart from large-scale survey tasks (e.g. mapping), UAVs help to provide illumination and communication signal coverage at the SAR team base or survivor resettlement camp. UGVs, including snakebots and other quadruped robots, could go through the narrow terrain and access places that humans find difficult or dangerous to reach. They can help to find secure paths for human rescuers, identify signs of lives, and sample the environment.

Most robots that deploy in fire disasters are custom-designed, especially UGVs. Common commercial or research robots do not have the capabilities to survive in these EEs, e.g. high temperatures. However, the past several years have seen a growing number of firefighting robot developments. They are capable of projecting water and carrying heavy hoses into places that are too dangerous for human firefighters (Peskoe-Yang, 2019). If a building collapses because of the high temperature, UGVs can be deployed to search for victims, search for paths through the rubble for quicker evacuation, inspect structures, and detect HAZMAT (Robin R Murphy, 2004b). UAVs that carry a water hose can be used to project water accurately into regions where traditional water cannons cannot reach easily (Ausonio, Bagnerini, and Ghio, 2021). UAVs can also map 3D environments in large scale fires to build an initial SA (Surmann et al., 2021).

After mudslides, UAVs were widely employed to collect remote sensing images (C.-C. Liu et al., 2015) and monitor the potential risk of the next mudslide (Robin R Murphy, Duncan, et al., 2016). This information is vital to keep the SAR team away from a secondary disaster. For UGVs deployment, they concentrate on the areas that have potential collateral damage. These tasks are similar to the tasks required in earthquakes and hurricanes because of the similar nature of the post-disaster site. In contrast, in directly affected areas, everything may be buried beneath the mud and soil. In such situations, there may be little value in deploying a mobile robot to survey unless the mud is removed.

Based on the reports of these deployments in the EEs, many lessons were learned about HRT, and we summarize and highlight the following insights:

- Customized robots are necessary to avoid risks and problems. Off-shelf commercialized

robots can take simple tasks, e.g. providing illumination, but they are not suitable for many emergency response tasks.

- Communication among HRT members should be transparent and explainable (Verhagen, Neerincx, and Tielman, 2022). It can be useful to use wired communication, rather than wireless communication, to keep low latency and secure the connection, especially it is often the case that wireless communication is not available due to the complex environments and limited infrastructure.

- Robots in the field mostly run in teleoperation mode. A robot's autonomy capability is limited by many factors, including humans' trust in the system (Israelsen and Ahmed, 2019), or even the law/regulations (H. He et al., 2021).

- Human cognitive workload deeply affects the overall performance of the mission (Sam et al., 2024), especially in long-term deployments with predominantly teleoperated dominant.

After reviewing such real-world robot deployments over recent decades, and comparing robotic techniques emerging from research labs over similar timescales, the gap between them does not seem to be narrowing as rapidly as one might expect. There may be a variety of reasons for this. From the scope of this thesis, we explore the feasibility of making efforts to bridge the gap by adapting the State-Of-The-Art (SOTA) solutions from the research community, which can align with the FRs' needs and expectations. One way of helping to bridge this gap, is to adopt a VA framework, where control can be traded back and forth between human teleoperation and an autonomous agent, during missions. This thesis explores the role of semantic information in HRTs within a VA paradigm.

1.4 From raw information to high-level semantics

In most EE robot deployments, raw sensor information, e.g. camera feed or radiation readings, is the main input for building SA. FRs have been predominantly relying on such raw informa-

tion for decades. This raw information can be processed to generate "low-level semantics". For the purposes of this thesis, we consider low-level semantic information to mean basic, salient features of the environment, which can be easily noticed and understood by a human without requiring complex thinking and interpretation. This can comprise many types of information, including radiation strength and degree of radiation hazard, degree of noise in navigation sensors. Such information can be provided to the HRT by remote robots, or by other human team members, or it can be derived from prior knowledge such plans and technical schematics of a building or facility.

The HRT can typically develop a certain level of SA and understanding before the deployment, based on raw information and low-level semantics. However, as the mission progresses, increasing amounts of information continue to flow into the HRT. At this stage, human operators can become increasingly fatigued and face heavy cognitive workloads to continuously update their SA (Carissoli et al., 2024), which can impact the mission's success. Meanwhile, operators must constantly be mindful of whichever type of semantic information is currently most important (and this can change over time). The HRT can easily run into situations of information overload (Arnold, Goldschmitt, and Rigotti, 2023), which could lead to overlooking some vital information.

Not every human is as sensitive to different types of raw information (Paivio, 1991). Some people are sensitive to digit numbers, while other people are sensitive to images. However, after we transform this raw information into an "understanding", or an "awareness", humans become better attuned to detecting important aspects or changes in the situation (Klein, 2005). Understanding and awareness can be considered as types of high-level semantics, which are brief, conclusive interpretations and understandings of raw information. From the perspective of neuroscience (Kosslyn and Koenig, 1992), these abstract understandings require more human effort and analysis to obtain, but they are easier to store and recall in the brain, compared with raw information (Craik and Tulving, 1975). We can regard this as a trade-off between employing raw information and high-level semantics. Raw information can provide specific and accurate data about the situation. Meanwhile understanding and awareness are overall descriptions which

are easy for human cognition to store, recall and use, but which lack detail.

The HRT must flexibly utilize all types of information, with different levels of abstraction according to different situations. However, much of the robotics and AI literature has focused on processing low-level semantics, e.g. vision-based object detection. Comparatively little research has attempted to differentiate low-level and high-level semantics, and how these relate to the mission tasks and context as well as the raw sensor data.

Low-level semantics are vital for autonomous robot perception. However, from the scope of HRT, a continual stream of low-level semantics might increase the human cognitive workload to process and understand the situation. Some researchers have noted this problem, and proposed methods to process high-level semantics. However, their work tends to be closely tied to a specific application, e.g. fire-fighting, and lack a systematic way to generalize the ideas to other applications.

In contrast, in this thesis, we attempt to: clearly define semantics in the context of robotics; differentiate the different types of semantics; and explore the relationship between raw data, semantics, and context. Based on these ideas, we propose a framework for computing semantic indicators and semantic richness. We evaluate the impact when this is incorporated into an HRI, and used for HRT in example SAR tasks.

1.5 Challenges of addressing high-level semantics

The concept of semantics originates in the community of linguistics. More recently, the Computer Vision (CV) community applied this term to tasks such as recognizing objects in images and understanding what is happening in the images. Hence, in the context of robotics and AI, the term semantics can comprise two distinct levels of meaning: I) an objective detection or measurement; II) a subjective inference or understanding based on the first level. Level I) carries direct or straightforward information, but is usually not organized and easy to understand intuitively. Level II) comprises individual understanding, which relies on human experiences, knowledge, and cognition.

The connection between Level I (low-level semantics) and Level II (high-level semantics) implies that it is difficult to precisely understand the meaning of high-level semantics without reference to low-level semantics. This is especially the case when exchanging high-level semantics between different HRT members. Information might be missed during the translation from Level I) to Level II), e.g. high-level semantics "this area is dangerous" does not clearly refer to what makes it dangerous, which can be identified from low-level semantics. In addition, contradictory information from Level I) might lead to an ambiguous meaning of high-level semantics. For instance, LiDAR might show random noise feedback in the sandy environments, which indicates the many obstacles that are not passable. However, from the camera, it might be clear that there are no big obstacles to stop the robot, and the robot could go ahead with the task. Hence, if the robot only relies on these low-level semantics, the robot might not be able to understand the real situations. How to filter out the outliers, conflicts, or errors from low-level semantics, to generate correct high-level semantics, is not well-explored.

Moreover, due to the subjective features of high-level semantics, different people have different understandings of the same statement, e.g. "The air temperature is high". Is 30 degrees Celsius classified as "high"? Different people might feel hot at different temperatures. Hence, expert knowledge from professional people needs to be incorporated in high-level semantics. They might be the end-users, expert robot operators, or experts in the type of environment being explored, and their experience and knowledge could help to reduce gaps in understanding between individuals.

For similar reasons of subjective features, high-level semantics strongly relies on the context for processing and sharing. Different mission backgrounds (context) could lead to different criteria for generating high-level semantics. For instance, 38 degrees Celsius is not a high air temperature in a firefighting mission. But it could be a high air temperature in an earthquake response mission. Hence, high-level semantics are difficult to directly share across different contexts.

Finally, thanks to the wide applications of deep learning techniques, processing high-level semantics is becoming easier. However, researchers are tending to become more and more de-

pendent on computing power. Many applications, e.g. CV-based object detections, are hard to realize without moderately configured computers. It brings the problem that the hardware is difficult to miniaturize and can consume excessive energy from the robot's batteries. This can become an obstacle to long-term flexible field robot deployments, and limits the additional capabilities of the robot. In our work, we cautiously apply some deep learning techniques for certain tasks, particularly image understanding, where it is necessary for extracting semantic-level information. However, our use of deep neural nets is quite limited, which can partially mitigate some of the above challenges.

1.6 Contributions of this thesis

High-level semantic understanding is a topic that has been under-explored, or mixed with other concepts, especially with respect to applied robotics. Its application in robotics requires more attention, especially with respect to the human element in remote robotic operations, with HRT. To begin addressing some of these challenges, the contributions of this thesis can be introduced as follows:

- This work systematically surveys robot deployments in the context of disaster response. It identifies the trend of robot applications and the potential synergy of high-level semantics and tasks in EEs.
- This work surveyed the FRs from different countries about their attitudes towards robot, views about SA and semantics in emergency response missions, and their specific requirements and interests. Their feedback affirmed the value and direction of our work.
- This work proposes a taxonomy of semantic information that clearly defines semantics in the context of robot deployments. We differentiate semantics with different levels of abstraction, and discuss their relationships with raw sensor data and context.
- This work proposes a semantics-based SA framework. It provides a systematic method for an autonomous agent to compute high-level semantic indicators of different modalities,

and overall SSR of scenes, in real-time online during missions. The limited use of deep learning techniques makes the framework lightweight, explainable, and hence suitable for use by robots deployed in EE.

- This work explores the impact of these semantic indicators, and the SSR. The overall performance of the combined HRT is evaluated in example SAR tasks, when semantic indicators are incorporated into the HRI with a VA paradigm. The benefits of applying high-level semantics are supported by our experimental results, including decreased reaction time when switching LoA, reduced cognitive workload, and increased trust in SA.

1.7 Thesis structure

This thesis is constructed of seven chapters.

Chapter 1 has provided an introduction to robot deployments in safety-critical and EEs. It analyzes the challenges and requirements of real-world robot deployments, to help motivate this work. We briefly introduced the characteristics of raw information, low-level, and high-level semantics. The difference between them and the challenges to identifying them were discussed. Finally, we summarized the main contributions of this thesis.

Chapter 2 introduces the related literature in detail. It reviews the literature from three areas: HRT, SA, and semantic information. Moreover, the measures and analysis techniques involved in the experiments are introduced.

Chapter 3 firstly summarizes the tasks that humans and robots conducted in the field. With this context, we propose a taxonomy of semantics in a high-hazard environment by clearly defining semantics and their different levels of abstraction. Moreover, we discuss its features and other related concepts, relating low-level semantics, high-level semantics, raw information, the mission tasks, and context. Finally, the gaps between current real-world robot deployments, and technologies emerging from the research community, are highlighted and discussed.

Chapter 4 presents a survey about the FRs' and EE robot operators' attitudes towards semantic information in their duties. We analyze FR's views and their background. Insights,

requirements, and challenges for these end-users are summarized from Likert score (numerical rating) questions, and also open questions.

Chapter 5 proposes a framework for automatically computing semantic information in a principled and systematic way, by building on the taxonomy in Chapter 3. The chapter gives examples of how to apply the proposed framework. Semantic indicators are tested for multiple modalities of semantic information, including tests on a real robot in a variety of scenarios with multiple semantics combinations. The experiments indicate the feasibility of the framework.

Chapter 6 explores the impact of the previous chapter's semantic indicators on HRT and HRI. The experiments verify the usefulness of the proposed framework in a mock-up SAR mission with human participants. The impact of high-level semantics on the overall performance of the HRT is summarized. Moreover, some additional interesting patterns of interactions between humans and robots in different scenarios are discussed.

Chapter 7 summarizes the contributions of this thesis. According to the problems encountered in our experiments, we suggest potential directions that need to be improved or addressed in future research. Finally, we present concluding thoughts and remarks.

Chapter 2

Background and Literature Survey

2.1 Human-Robot Teams

HRT research is attracting increasing attention, from the lab to real applications. However, so far there has been comparatively little focus on HRT involving AI and advanced autonomous robotics techniques. Across a variety of application areas, including disaster response, large-scale manufacturing, service and caring robots, an HRT approach can help mitigate the impact of tasks which exceed the autonomous robot's capabilities, while enhancing the contributions from robust autonomous functionalities, thus maximising the overall performance of the team (Langås, Zafar, and Sanfilippo, 2025).

An HRT is often deployed in scenarios where remotely located robots interact with each other and keep humans in-the-loop to manage the deployment, interpret subtle environment information (e.g. assessing risks), and make critical decisions e.g. changing to a new task. In such missions, unlike traditional automation systems, HRT emphasizes "collaboration" between team members, rather than using robots to substitute humans. Humans and robots need to employ their complementary competencies to plan and decide what area is worth exploring, which task is suitable for the robot, and what situation needs human intervention. Especially during critical periods of the mission, humans might prefer full control or shared control of the system. Meanwhile, robot autonomy could take over control when humans make unsafe decisions or face

a high cognitive workload. Therefore, the coordination of control between humans and robots is a major challenge in HRT. If a control conflict happens in the HRT (human and autonomous agents fail to understand the intentions of others, or fight each other for control of the robot), the mission might halt and even become a catastrophic failure. E.g. in a SAR mission, when a robot is executing an exploration task, the human operator might receive messages from other sources about the exact location of victims. In this context, if the robot is not aware of the human's new SA and not prepared for control switching, the control conflict might delay the task of searching for the victims. Essentially, robots and humans need to be aligned on the same page by understanding the intention.

Specifically, humans need to stay in-the-loop during the tasks, and robots need to have strong capabilities of perception, planning to achieve a transparent HRI, and effective decision-making. These three capabilities are one of the major goals of current robotics research. Among these dimensions, SA is a critical factor. And semantic information is an important component of SA. Our work focuses on how to obtain accurate perception (semantics) and post-processing (SA) for future planning and decision-making. Hence, this section will briefly introduce related work from the scope of i) perception; ii) HRI; and iii) VA involved in the HRT. In subsequent sections of this chapter, the SA and semantic information are introduced in detail.

2.1.1 Human and robot perception

Perception of the HRT is one of the most critical capabilities that ensures a robot can interact with the environment. Essentially, it is the foundation of the robot system. The community sometimes confuses or conflates human perception and robot perception. With rapid advances in the AI technologies over the past decade, in the scope of accuracy and complexity, robot perception has become mainstream in research. In contrast, "human perception" appears much more rarely in robotics-related literature. In EE scenarios, humans, as members of the HRT, are usually remote from the robot's working area. Direct perception of the human is not available in these cases. The remote perception from the robots becomes the major source of perceiving the on-site environment and situation. Hence, "human perception" is a synergy of robot perception

and other information in the base station, e.g. the analysis or historical data of the site. Due to the nature of human perception, it is more complex than robot perception and involves more psychological factors. In this context, these two concepts have a significant overlapping scope. Therefore, we introduce robot perception in this subsection, and focus on human SA in the next section.

Traditionally, robots rely on raw data from different sensors to perceive the environment. The Kalman filter (Kalman, 1960), and particle filter (Gustafsson, 2010) are widely used to obtain a basic type of perception, i.e. mobile robot localization. Compared with more recent AI-based methods, these methods cost less computational power. However, they show comparatively limited robustness in dynamic or complex environments that are not well pre-defined. Hence, they are suitable for perceiving structured environments with limited uncertainties, e.g. robot localization in warehouses.

In the deep learning era, techniques, such as Convolutional Neural Network (CNN) (Z. Li et al., 2021) and Transformer (Vaswani et al., 2017), have rapidly expanded to perception tasks, including semantic instance segmentation, 3D reconstruction, object detection and tracking, and pose estimation. Moreover, deep learning facilitates multimodal data fusion to combine a wider range of sensors, e.g. LiDAR, Inertial Measurement Unit (IMU), haptics, and acoustics. Such data fusion can substantially enrich the breadth, depth and robustness of robotic perception.

Rapid advances in high-performance hardware, e.g. high-resolution LiDAR (replacing traditional ultrasonic sensors) and powerful graphics cards, are providing greater possibilities for sophisticated perception. Current LiDAR systems can provide a 3D scan of the environment, with depth information at the imaging level of resolution. This offers the potential to replace RGB images for object detection and scene understanding in some applications, e.g. forest remote sensing (Gao et al., 2025). Moreover, modern graphics cards boost the progress of training deep networks. Miniaturisation of powerful computers is enabling robots to process complex computations onboard, rather than relying on wireless communication to a powerful computer at a remote base station.

2.1.2 Human-Robot Interaction

HRI is an interdisciplinary topic that combines knowledge from robotics, psychology, cognitive science, AI, and other fields. (Goodrich, Schultz, et al., 2008) defined HRI as "a field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans". The interaction could be remote (a human controlling a robot in a remote hazardous zone) or proximate (a human working closely with a co-bot arm on a production line). HRI research focuses on how humans and robots can clearly and effectively interact via multiple interface modalities, e.g. vision, language, haptics, and others. Specifically, HRI aims to enable robots to assist humans in completing tasks, collaborating with humans, and socializing with people. Within this scope, the HRT should achieve a shared understanding of the intention of both robots and humans; achieve transparent, smooth communication; increase human acceptance of robots; and improve mutual trust and collaboration efficiency. In our work, we do not involve all of these dimensions, but we focus on the factors that could potentially affect the HRI in the context of semantics as follows:

- **Situational awareness:** There is overlap among the concepts, including robot perception, human perception, and HRI. We are interested in what kind of SA is being preferred in the HRT and how SA could affect the HRT and HRI.

- **Level of autonomy:** The change of LoA is an objective, indirect measure which can be useful for investigating the effects of SA. Moreover, different choices of LoA may indicate different human understanding and emotions (Chiou, McCabe, et al., 2021). Therefore, LoA is an important factor to observe in our experiments.

- **Trust:** Trust in the system and SA is one of the important factors in an effective HRT. Too much trust (e.g. automation bias) (Goddard, Roudsari, and Wyatt, 2012) can lead to automation misuse, whereas too little trust can lead to unwillingness to make use of useful autonomous functionalities. Hence, an appropriate level of trust from humans can be an indicator of a successful HRI system design.

Furthermore, HRI researchers have explored a variety of other factors e.g. intention, and proposed frameworks for normative understanding. (Lemaignan et al., 2017) introduced a social HRI framework that enables robots to understand roles with respect to human intention. This framework has been widely used in social AI agents. (Nardelli, Sgorbissa, and Recchiuto, 2024) proposed cross-modal memory networks that give robots the capability to change their interaction pattern based on human personality. (Robin R Murphy and Schreckenghost, 2013) systematically surveyed the metrics to evaluate HRI and a structure to categorize HRI metrics from the scope of system productivity, efficiency, reliability, safety, and co-activity.

2.1.3 Variable Autonomy

Humans and robots possess different strengths and weaknesses when facing dynamic situations. VA will strengthen the Human–Robot Collaboration (HRC) in many dimensions. It is an important topic under HRT and HRI. VA refers to a system in which control of the robot is dynamically traded back and forth between the human and an autonomous agent. Such LoA switching decisions are based on the demand of tasks, situation changes, and human conditions. VA can be categorized into four major types: shared control, Mixed Initiative (MI), adjustable autonomy, and sliding autonomy.

Shared control refers to humans and robots both having control over parts of the system at the same time. Humans and robots could each operate a different part of the system simultaneously to complete the task. MI means that both human and autonomous agent can control the robot at different times, and only one member has full control of the system at any one time, i.e. full autonomy or full teleoperation. It emphasizes which entity could take the full control. In adjustable autonomy, both humans and autonomous agents can adjust the LoA. The research focus lies on which level of LoA is given to the HRT. Sliding autonomy means the autonomous system can smoothly automatically adjust the robot's LoA. E.g. in the precision components assembly scenario, multiple robots could perform grasping, positioning and locking operations with full automation. When the robot detects that the component is out of the tolerance, the system automatically downgrades the LoA of faulty robots to 50% (e.g. pause and wait for instructions). At

the same time, autonomous agent increase the LoA of the adjacent robots to 80% (e.g. taking over some collaborative tasks, by temporarily adjusting the assembly sequence). When humans notice the error and inform the system, the LoA of faulty robots would drop to manual operation for fine-tuning movement. Afterwards, the autonomous agent could gradually resume the LoA of the faulty robots if the problem is solved with the assistance of humans.

Based on this categorization, sliding autonomy and adjustable autonomy can be viewed as types of MI that humans and robots could take full control of the entire system. The relationship between these VA paradigms can be seen in Fig. 2.1.

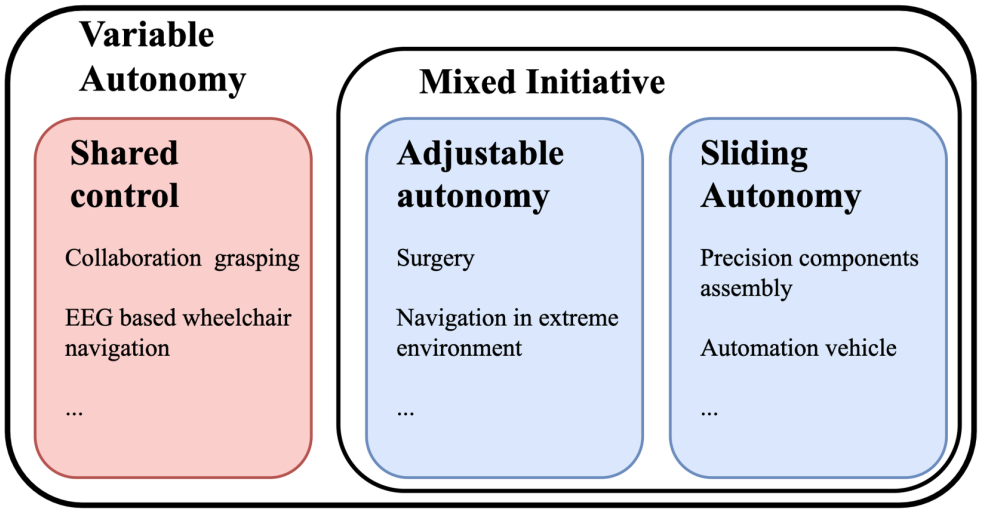


Figure 2.1: Relationship between different modes of VA and some example applications: adjustable autonomy and sliding autonomy can be seen as types of MI, which is distinct from shared control.

Given the autonomous capabilities of the robot, most works, in the context of VA, focus on shared control and adjustable autonomy. Sliding autonomy is more complex than the other two modes, and requires more intelligence of the robot. Shared control and adjustable autonomy can take advantage of both robot perception and human logical reasoning capabilities, to authorize robots with different LoA at appropriate times during the mission, thereby improving the efficiency of HRC and task safety. Research contributions range from general challenges and frameworks to specific control policies and related impacts.

(Contreras et al., 2025) summarized the application of mobile robots in unstructured environments and pointed out that cognitive workload and communication delays are major challenges

to adapting VA. (Tamura, Amano, and Ota, 2020) found that human control (teleoperation) was welcomed and necessary in complex missions like firefighting. Similar viewpoints are supported by (Verhagen, Neerincx, and Tielman, 2024), who believed that humans should be responsible for all ethical guideline-related decisions. The robot should determine if the task is ethically sensitive, and then decide if it should be handled by a human decision maker. Moreover, (Verhagen, Neerincx, and Tielman, 2024) emphasized that accountability, responsibility, and transparency play important roles in such a context. Hence, humans and teleoperation still play essential roles in such missions. At the same time, (Methnani et al., 2024) shared the same views and proposed a framework that helps to clearly exchange information between HRT members. Building on similar principles, (Hatanaka et al., 2020) proposed a control architecture based on Gaussian Process (GP) regression in agricultural applications. The system is trained on senior expert operators' data, and helps to improve the performance of junior human operators in moving a manipulator to target positions.

Previous shared control systems typically relied on static parameters, often inadequately handling dynamic changes. (M. Li et al., 2018) and (Bustamante et al., 2022) tackled this issue. (M. Li et al., 2018) claimed that most previous shared control systems focus on switch design that lacks dynamic control allocation. Game theory and situation assessment usually rely on static parameters that do not always meet the changing situation. (Bustamante et al., 2022) also proposed a dynamic controller with an action representation and planning framework, that is able to generate plans which can be executed with both shared control and supervised autonomy, even switching between them during task execution. The action representation combines the advantages of "Action Templates" and "Shared Control Templates".

In autonomous driving, AI needs to handle both steering and velocity ideally. Currently, in most jurisdictions, humans are still required to stay in-the-loop for on-road driving, for both regulatory and practical reasons. Some researchers suggest that many control systems focus on autonomy itself while overlooking human control, which leads to instability (Sentouh et al., 2018). Hence, they propose a collaboration strategy that enables control trading between humans and autonomous agents, with two tactic controllers to execute the strategy. (Benloucif

et al., 2019) proposed a trajectory-planning-based shared driving control. It reduces the chances that conflicts happen between the driver and AI. And it gives more flexibility to drivers when the car stays in a preset lane. Similarly, in airplane navigation, (Eraslan, Yildiz, and Annaswamy, 2020) found that current human-machine collaboration frameworks rely on human experts supplementing flight control automation. Such systems are not smooth switching, and cannot tackle situation changes quickly. The authors proposed two shared control architectures: trading action and supervisory action. They are proven effective when a serious emergency happens, and can switch control smoothly.

Besides autonomous driving, manipulation requires HRC as well. Considering the power and inertia of manipulators, poor collaborations (e.g. misunderstanding of intention, lack of SA) not only result in performance degradation, but also can pose significant risks in high-consequence environments, e.g. nuclear waste handling. A variety of design strategies have been proposed (Robinson, Wijayasinghe, and Popa, 2019; Abi-Farraj, Pedemonte, and Giordano, 2016). (Abi-Farraj, Pedemonte, and Giordano, 2016) proposed a vision-based shared control architecture for tele-manipulation to enhance the SA. It uses both visual and haptic feedback to detect if a human has deviated from a correct manipulative action and guide the manipulation to a correct path or region. Similarly, (Adjigble et al., 2019) proposed a assisted telemanipulation grasp planning approach and explored how human performance affected by the visual and haptic cues. In medical applications, manipulators are used to assist surgeons. (Su et al., 2022) designed adaptive control policies based on the understanding of surgeons' movements and activities.

VA is a framework which we use in our work, but is not itself the main focus of this thesis. Hence, our work refers to some of the factors mentioned above, e.g. explainability in our system design.

2.2 Situational Awareness

SA is another core concept in HRI. The study of SA has its origins in research to support military fighter-pilots (Endsley, 1988). It originally referred to knowledge and understanding of the

disposition of friendly and enemy forces in the airspace (*AMRAAM oue final report* 1982). (Endsley, 1995b) extended the definition to more general applications, and this definition has become widely adopted in robotics, autonomous driving, and other topics that require fast understanding and response to changing situations.

2.2.1 Framework

The most popular and widely used framework for SA was proposed by (Endsley, 1995b). The framework consists of three levels: 1) perception of elements in the current situation, 2) comprehension of the current situation, and 3) projection of future status. This framework systematically differentiated between raw perception, understanding and inference from perception, and estimation of future situations. This aligns with the robotics engineering structure, which traditionally involves perception, planning, and decision-making. Based on Endsley's model, researchers proposed enhanced frameworks to adapt to specific application scenarios. SA is not only adapted to humans, but to robots, and to the entire HRT. In (Salmon et al., 2006; Endsley, 1995b; Gary Klein, 2000), the authors discussed the concept of Teams SA (TSA). In contrast to individual SA, TSA considers the shared knowledge among team members. It is defined as "the degree to which every team member possesses the SA required for his or her job" by (Endsley, 1995a). Three major types of TSA are defined as follows:

- **Shared SA:** HRT members have the same information and have a consistent understanding of the SA.
- **Distributed SA:** HRT members have different local SA, but these SA together can form a complete picture of the overall SA.
- **Complementary SA:** SA among HRT members can be complementary, but each team member does not necessarily have to understand everything, or have the same types of understanding or access to the SA of other team members.

In current robot deployments in the field, distributed SA and complementary SA are more common. Besides the information from the robot, humans usually obtain information from other

sources, e.g. talking with local experts with knowledge of the site. Moreover, the robot may not understand all the information that humans have obtained. The study of TSA in the context of robotics is still in relatively early stages. The research community has no clear consensus on the definition or framework of TSA. Researchers have proposed their individual understandings in the context of specific applications. For instance, (Noriega, 1999) integrated human factors, including belief, desire, and intention, into a framework that improved complex understanding and inference of the SA.

2.2.2 Methods to deliver and enhance Situational Awareness for Human-Robot Teams

SA can be viewed as a type of information. In the context of HRI, information needs to be exchanged in a clear, explainable, and transparent way. Vision, acoustics, and haptics are considered major sensory modalities of humans. Researchers have developed different techniques and interface devices for conveying SA information, based on these modalities.

Traditional SA is obtained by collecting opinions from humans. Humans are asked about their understanding of the situation during tasks (Endsley, 1988), or after task completion (R. M. Taylor, 2017). If multiple team members are involved, SA conflicting opinions between different team members may arise, and must not be overlooked (McNeese et al., 2021). AI-assisted TSA could help to reduce such conflicts in processing accurate TSA.

Besides methods and techniques for obtaining SA directly from humans, researchers have turned their attention to other types of perception. For instance, Augmented Reality (AR) combines real-time camera images of a scene, with task-related information displayed simultaneously, enabling humans to receive more SA with less cognitive effort. Within this context, the traditional Heads Up Display (HUD) system can be categorized as AR also. (Arvanitis et al., 2023) proposed a scheme to render detected potholes on to vehicle windshields. Moreover, this distributed scheme allows multiple vehicles to share information, and thus enables humans to be notified about hazards which their vehicle has not yet encountered. Similarly, (Reardon et al., 2020) outlines environmental changes after comparing them with baseline LiDAR data.

This system helps humans maintain SA visually, even in cluttered data environments. Apart from rendering detections, displaying raw information on a AR headset can also enhance SA. (Zolotas, Elsdon, and Demiris, 2018) projected the rear-view display of a wheelchair, successfully reducing the number of human head rotations of the wheelchair user.

When vision is limited or humans are facing visual information overload, acoustics and haptics-based SA can be used to increase the responsive capability. (Fostick and Fink, 2021) found that spoken words were less sensitive than pink noise or gunshot for sound localization. It indicates that speech stimuli is not the best carrier to deliver location information acoustically. (Bennie et al., 2024) applied both natural language and haptics as an alternative interface in the HRT, which reduced the SA variance in understanding between human team members, and reduced human cognitive workload.

Training and expert knowledge play important roles in enhancing SA (Lu et al., 2022). Humans can raise their cognitive overload threshold if they are familiar with the system, interface, and environment (Kalyuga, 2009). (Kapellmann-Zafra et al., 2016) found that humans with a certain amount of training outperformed others with minimal training under limited SA conditions. This highlights the importance of training and indicates that solid proficiency can enhance SA. In addition, (Senaratne et al., 2025) systematically discussed the dynamic nature of TSA and the factors that could affect SA.

2.3 Semantic information for Human Robot Teams

Semantics, a concept originating in linguistics, has now been extended to other communities, especially CV. However, researchers do not have a widely accepted definition of semantics that could cover extended applications.

(Jurafsky, 2000) defined semantics as "the study of meaning communicated through language. In Natural Language Processing (NLP), semantics involves the representation of the meaning of a sentence or discourse in a form that a machine can process." (Chen et al., 2017) stated that "Semantic understanding of an image requires structured prediction where the output

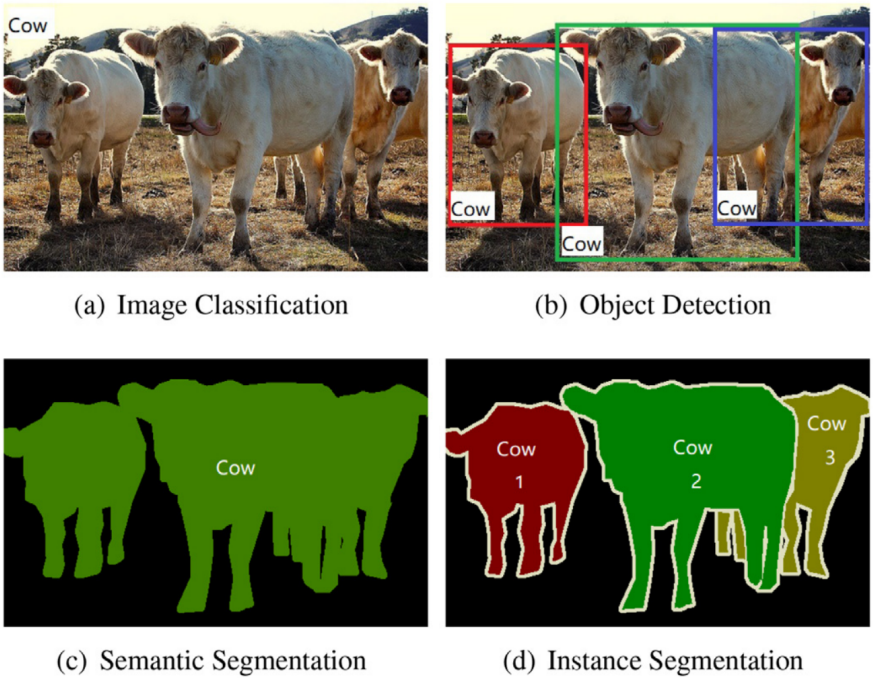
consists of interrelated components such as object categories, spatial locations, and scene labels" in the context of CV. Kirillov defined "Semantic information in CV systems enables reasoning about what objects are present and how they relate to each other in a meaningful way" (Kirillov et al., 2019). The above definitions focus on specific applications and lack broader generalization with a universal understanding. In this thesis research, we proposed and published a definition of semantics in the context of robotics - *Semantics extract and convey an understanding of environments, to both robots and humans, in terms of information about entities and both explicit and implicit relationships between entities*. It explains these concepts in the context of applied robotics deployments (Ruan, H. Wang, et al., 2022). This taxonomy will be introduced in more detail in Chapter 3. We focuse on semantics that are useful for HRT. I.e. semantics is something which helps to align human and robotic perceptions and SA, increase HRT collaboration efficiency, improve decision-making quality, and help to ensure safety.

2.3.1 RGB-based semantics

As introduced above, semantics is widely used in CV. Semantics can mean recognising objects, scenes, behaviours, positions and other types of image understanding. Semantic information can refer to specific objects, or to an abstract understanding of a scene. Thanks to the development of deep learning techniques, many of these semantics can now be processed directly from RGB images. Deep learning architectures, including Alex-net (Krizhevsky, Sutskever, and Hinton, 2012), Fully Convolutional Networks (Long, Shelhamer, and Darrell, 2015), U-net (Ronneberger, Fischer, and Brox, 2015), Transformer, SSD (W. Liu et al., 2016) and others, enable robots to autonomously perform object detection (Girshick, 2015) and tracking (C.-Y. Wang, Bochkovskiy, and Liao, 2023), pixel-wise segmentation (Long, Shelhamer, and Darrell, 2015), body frame detection (Zheng et al., 2023), scene understanding (Naseer, Khan, and Porikli, 2018), and other types of image understanding.

(Bolya et al., 2019) proposed the You Only Look At CoefficienTs (YOLACT) techniques that combine object detection and segmentation. It is one of the SOTA real-time instance segmentation algorithms. It achieves an excellent trade-off between inference time and accuracy with

31.2 masks mean Average Precision (mAP) and 23.4 Frame Per Second (FPS) tested on MS COCO dataset (Lin et al., 2014). For comparison, another well-known instance segmentation algorithm Mask-RCNN (K. He, Gkioxari, et al., 2017) has achieved 35.7 mask mAP and only 8.6 FPS on the same dataset. Later, based on the YOLOv8 (Varghese and Sambath, 2024) model and YOLACT segmentation method, FastSAM was proposed to achieve low computational cost segmentation (Xu Zhao et al., 2023). FastSAM is computationally lighter than YOLACT, however, YOLACT showed higher robustness on specific instance segmentation. In our experiments, we applied this technique and combined depth images with RGB images to generate instance detections with distance information. (See Chapter 3.)



(a) Image Classification (b) Object Detection

(c) Semantic Segmentation (d) Instance Segmentation

Figure 2.2: Samples of image classification, object detection, semantic segmentation, and instance segmentation (X. Wu, Sahoo, and Hoi, 2020).

In addition to object detection in 2D, RGB images can be used to generate 3D object bounding boxes. The differences of object detection, semantic segmentation, and instance segmentation can be seen in Fig. 2.2. Dai proposed a structure that comprises a 2D network and a 3D network (Dai and Nießner, 2018). It projects multiple RGB images to a voxel grid, and aligns these projections to generate an accurate 3D semantic label. This feature leads to an increasing demand to handle multiple RGB images. Hence, it is computationally intensive and can be hard

to deploy for real-time applications.

Compared with other information-based semantics, RGB-based semantics has the advantages of relatively low cost for processing, and high adaptability to different applications. It is closer to human's natural perception and inference, and has become the major semantics foundation in robotics.

2.3.2 Multi-sensor-based semantics

Apart from the image-based semantics, deep learning networks can process semantics from other sensors, or fuse them to generate a complex understanding to compensate for the problems of image-based semantics. For instance, RGB methods can struggle in scenes with strong light, plain texture, or mirrors. It is also highly restricted in low-visibility scenarios. Hence, multiple-sensor based and sensor-fusion semantics are necessary in situations where RGB-based semantics is not possible.

The combination of an RGB image and a depth image is one of the popular solutions. (Hazirbas et al., 2016) proposed a dual-channel convolutional network, which combined the features of the depth image with the RGB image. This early-stage RGB-D fusion method is sensitive to the accuracy of the depth image, and does not have a good performance on complex surfaces.

(McCormac et al., 2017) applied RGB-D SLAM with the combination of CNN network to enable the system to generate a pixel-wise segmented map by labeling objects including bed, table, floor, wall, and TV. (Jiang et al., 2018) employed ResNet (K. He, X. Zhang, et al., 2016) in RGB-D semantic segmentation, and outperformed the full convolutional network structure. However, this network is computationally heavy and is therefore difficult to implement in practical mobile robot deployments and real-time scenarios. Besides feeding the 2D images directly to a deep learning network, some researchers worked on the 3D perspective. (Luo et al., 2020) proposed a method that is based on the SSD structure to detect objects in the 3D environment with a 3D anchor box using an RGB-D camera. The network is able to accurately detect objects with occlusions, and overcome the poor detection problems caused by a sparse pointcloud.

Moreover, 3D imaging and reconstruction of pointclouds is widely applied to data from

LiDAR sensors, which have been rapidly improving in resolution while decreasing in cost. Researchers focused on the combination of RGB images and point clouds. In (Peng et al., 2022), the authors achieved pointcloud semantic segmentation by employing an attention mechanism. (Y. Zhang et al., 2020) improved real-time performance in pointcloud understanding with lower computational cost. Moreover, (X. Li et al., 2017; Rosinol et al., 2021) employed RGB images and LiDAR data for semantic SLAM.

Table 2.1: Comparison of different perception-based semantics

Perception	**Accuracy**	**Adaptability**	**Cost***	**Application**
RGB	★★★	★★★★	★★	Object detection, damage detection, navigation
RGB-D	★★★★	★★★	★★★	Mapping, grasping, interaction
LiDAR-based	★★★★★	★★	★★★★	Mapping, navigation, environment assessment
Sound	★★★★	★★★	★★	Social AI agent, live detection and localization, scene understanding
Radar-based	★★★	★★	★★★★	Navigation, localization

* More stars refer to higher cost, fewer stars refer to lower cost.

Acoustic semantic information is useful when image-based semantics is not available or limited. Acoustic semantics can be categorized into NLP and basic sounds and noises from nature. NLP focuses on understanding the context and meaning of a conversation (Vaswani et al., 2017; Devlin et al., 2019). In the scope of natural sounds and noises, researchers applied spectrogram analysis to discover semantic information from sounds, including underwater threats (Stolkin, Sutin, et al., 2006), signals of underwater sound-producing fish (Stolkin, Radhakrishnan, et al., 2007), scenes understanding (Bae, Choi, and N. S. Kim, 2016; Battaglino et al., 2016), and location of sounds (Okutani et al., 2012).

Similarly, deep infrared thermal imaging cameras can be an alternative to RGB cameras in poor illumination situations. Researchers can identify humans in complex environments (Jeon et al., 2015), detect damage on concrete structures (Pozzer et al., 2021), and process image-based semantics in dark environments (Y. Sun et al., 2020). Research on radiation-based semantics is limited. (Bird et al., 2018) generated a radiometric map of a radioactive facility by combining

a radiation detector and a LiDAR. This demonstrated the potential for processing radiation heat maps, for characterising and assessing hazardous areas in nuclear facilities.

In the end, we summarize and compare the perception techniques from the scope of accuracy, adaptability, cost, and major applications in Table 2.1.

2.4 Measurements and metrics

In this section, we discuss the background of evaluation methodologies and metrics used in our human-centered robotics work, including human factors assessments and data analysis.

2.4.1 Human factors

Human factor is an important topic in the context of HRT. Human factors can be categorized into the following: cognition (e.g. workload, SA, attention), emotion (e.g. trust, stress, confidence, anxiety), behavior (reaction time, error rate, task performance), and physiology (e.g. eye movement, heart rate, electroencephalography (EEG)). Our work comprises the evaluation of cognition, emotion, and behavior. A large number of studies have contributed to these topics.

Cognition mostly relies on subjective self-reports. Humans are asked to score or answer questions, during or after the task. For instance, Subjective Workload Assessment Technique (SWAT) (G. B. Reid and Nygren, 1988) assesses human workload from the aspects of time, mental effort, and amount of work. The Subjective Mental Effort Questionnaire (SMEQ) measures human efforts to complete the task, but can be easily biased when involving multiple dimensions, including mental demand, cognitive processing load, and working memory load. The questions in the above dimensions are professional, and humans likely substitute the questions with easier questions that they are more familiar with (Moore and Picou, 2018). Instantaneous Self-Assessment (ISA) requires the human to respond to the assessment during the tasks. This helps avoid some problems of doing assessments after the tasks, e.g. memory bias (Cain, 2007) and difficulty capturing real-time fluctuations. However, this assessment interrupts the experimental process and is not suitable for complex tasks (Galy, Paxion, and Berthelon, 2018). Among

these measures, NASA Task Load Index (NASA TLX) is widely applied (Hart, 2006). The original version comprises six dimensions, including mental demand, physical demand, temporal demand, performance, effort, and frustration, which covers both cognition and emotion. The weighted NASA TLX can alleviate some deviations caused by memory or emotion bias, which is helpful for our work.

Similarly, subjective measures are popular for assessing emotions. Muir & Moray proposed a systematic scale of trust in the automation system, with dimensions including competence, predictability, responsibility, and willingness of the human to rely on the system (Muir and Moray, 1996). This scale covers most dimensions regarding trust, but misses certain factors, e.g. explainability of the system in the context of HRT. Jian's scale (Jian, Bisantz, and Drury, 2000) is widely used in HRI experiments with marking six positive statements and six negative statements. It helps balance and reduce the effects of the bias, but cannot reveal the trust changes during the mission. Lee proposed a dynamic model that introduced the process of building and changing of the trust specific to the automation system (Lee and See, 2004). Some researchers have attempted to assess trust via indirect methods, e.g. analyzing the human's behavior, including intervention time (Riley, 2018). However, the connections between humans' behaviors and trust are not simple or static. They can be affected and vary between different tasks, and different automation systems. Not only trust in the entire system, but also trust in the SA is becoming important to track. (Holthausen et al., 2020) proposed a situational trust scale specific to autonomous car applications. The scale comprises six dimensions, i.e. system trust, self-efficacy, non-driving related task, perceived risk, judgment, and reaction. This scale was shown to distinctly differentiate the different levels of trust in experiments with 300 participants.

2.4.2 Data analysis

It is not reasonable to assume the data distribution in most exploratory experiments. Hence, nonparametric statistics (no assumptions on the overall distribution and overall parameters) are a popular tool in HRT system analysis. The background of the analysis tools used in our work is introduced here.

The Wilcoxon signed-rank test is used to compare two groups of paired data (Wilcoxon, 1992). The output W value is calculated by the rank sum, and p-value refers to the variance between two groups. The details are as follows:

If x_i and y_i is a sample pair, the difference $d_i = x_i - y_i$. Then, $r_i = rank(|d_i|)$ is the ranked result of absolute d_i. The singed rank sums is defined as $W^+ = \sum_{d_i>0} r_i, \quad W^- = \sum_{d_i<0} r_i$, then the Wilcoxon statistic is:

$$W = \min(W^+, W^-), \tag{2.1}$$

The p-value is computed from the distribution of the test statistic W (Wasserstein and Lazar, 2016). A higher W value refers to two groups that are not similar in the shape of the distribution. And a higher p-value refers to the higher similarity between two groups (p-value < 0.05 means two groups are significantly different).

Moreover, the effect size r measures the actual size or strength of the difference between two groups. It can be calculated based on the nonparametric effect size equation:

$$r = \frac{Z}{\sqrt{N}}, \tag{2.2}$$

where N is the number of the sample pairs, and Z score is calculated by the following:

$$Z = \frac{W - \mu_W}{\sigma_W}, \tag{2.3}$$

where μ_W is the expectation value of W and σ_W is the standard deviation of W.

Spearman's rank correlation (ρ) coefficient is used to compare the monotonicity between two sets of data (Spearman, 1961). When Spearman's rank correlation coefficient $-0.05 < \rho < 0.05$, two groups show a big variance in the changing trend.

Kendall's rank correlation (τ) coefficient analyzes the association between two groups (X and Y), especially adapts to the small size of samples (<20) (Kendall, 1938). The details are shown as follows:

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}, \tag{2.4}$$

where C is the concordant pairs (e.g. if $X_i < X_j$, then $Y_i < Y_j$, i and j refer to the different elements' tags in the group), D is the discordant pairs (e.g. if $X_i < X_j$, then $Y_i > Y_j$), n is the number of samples, $\frac{1}{2}n(n-1)$ is the number of pairs that is always above 0 at any time. If ties exist, Kendall's τ is used:

$$\tau = \frac{C - D}{\sqrt{(C + D + T_x)(C + D + T_y)}}, \tag{2.5}$$

where T_x is the number of pairs with the equal-rank in x, T_y is the number of pairs with the equal-rank in y.

The output τ reveals the overall correlations ranging from 0 to 1. If τ is close to 0, the correlation is weaker. If it is over 0.7, the correlation is strong. Kendall's rank correlation coefficient is more robust than Spearman's correlation when encountering outliers.

Sensitivity analysis helps to identify which factors contribute more to the system. The Sobel index is one of the global sensitivity analysis methods (Sobol, 2001). It can have three possible outputs, first-order sensitivity index S_i (individual impact to the output), second-order sensitivity index S_{ij} (impact from the interaction between factors), and total sensitivity index S_{Ti} (impact from the interactions between all factors). When the output $S_i > 0.5$, it refers to changes in this factor that could bring a large impact on the final outcome.

In summary, in Chapter 5, we use Spearman's rank correlation coefficient and Kendall's rank correlation coefficient to explore the correlations among factors in the experiments. In addition, we apply the Sobel global sensitivity analysis to identify which factors contribute more to the system. In Chapter 6, the Wilcoxon signed-rank test is applied in analyse the significance of variables in the experiments. And Spearman's rank correlation coefficient is calculated to explore the pattern among different factors.

Chapter 3

A Taxonomy for Semantics in High-hazard Applications

The work which forms the basis of this chapter has been published in the 2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR) (Ruan, H. Wang, et al., 2022).

The concept of semantics has been used in a variety of different contexts, from NLP to CV. The term semantics originated in the field of linguistics, but the term has become widespread in fields such as AI and CV in recent years. Detecting and understanding semantic information could enhance robots' abilities to perform complex tasks. Humans can intuitively understand semantics, infer context, and incorporate it into SA. In contrast, semantic understanding (i.e. the process of understanding semantic information) in robotics remains a challenging and open research problem. In this chapter, we discuss the definition of semantics and categorize it in the context of robot deployments in EE.

The term semantics is derived from the Greek verb [sēmainō] meaning "to signify" or "to mean" [1]. Typically in robotics and AI, semantics refers to the understanding of things, environments, and situations. Garg et al. define "semantics in a robotic context to be about the meaning of things: the meaning of places, objects, other entities occupying the environment, or even the

[1] https://www.britannica.com/science/semantics/Historical-and-contemporary-theories-of-meaning

language used in communicating between robots and humans or between robots themselves" (Garg et al., 2020). Additionally, in CV, the Australian Institute for Machine Learning (AIML) states that "semantic vision seeks to understand not only what objects are present in an image but, perhaps even more importantly, the relationship between those objects" (AIML, 2021).

In the context of this chapter, we define semantics as: *extracting and conveying an understanding of environments, to both robots and humans, in terms of information about entities and both explicit and implicit relationships between entities*. For utility, our working definition aims to be somewhat explicit and restrictive in terms of understanding of the environment. It comprises two key features of the environment: entities and the relationships between them. Note that this implies two levels of information and abstraction, i.e. relationships are abstract concepts that link two or more physical entities.

In this chapter, we focus on semantics in disaster and emergency response, and tasks in hazardous environments. Robots are increasingly being used in hazardous or EE industries and have become increasingly embedded in emergency response teams. Robots perform a wide range of tasks, including reconnaissance and remote inspection in hard-to-reach or dangerous zones, as well as heavy-duty manipulation or transporting tasks. For safety reasons, humans should be located remotely from hazardous zones, but must rapidly assess the circumstances where a robot is operating. With currently deployed robots being mostly teleoperated and lacking advanced AI capabilities, it is notoriously difficult for human operators to gain SA from limited video and sensor data (Chiou, Epsimos, et al., 2022).

As a foundational component, semantic understanding is significant for establishing real-time SA, which supports both robot autonomy (Robin R Murphy and Burke, 2005) and HRI (P. Smith et al., 2013). Given onboard semantic understanding capabilities (e.g. object detection), a remote robot could: i) make better and more complex autonomous decisions; and/or ii) provide enhanced information to assist human operators with clear and rapid SA. Currently, much of semantics-related AI research is focused on semantic image understanding (Garg et al., 2020), typically using deep neural networks, such as "semantic segmentation" (Long, Shelhamer, and Darrell, 2015; Ronneberger, Fischer, and Brox, 2015), or object detection and recognition (Gir-

shick, 2015; C.-Y. Wang, Bochkovskiy, and Liao, 2023). Related research also includes SLAM (Rosinol et al., 2021) and 3D reconstruction of a scene while identifying and labeling types of materials (C. Zhao, L. Sun, and Stolkin, 2017).

However, despite advances in disaster response robotics, and parallel growth in AI research on semantic understanding, there has been comparatively little transfer from one domain to the other. Partly this is because rapid AI advances in semantics are recent developments, and the technology remains predominantly "low-TRL" (Technology Readiness Level) (Mankins et al., 1995). In contrast, disaster environments are profoundly unstructured and difficult. "High-consequence" rescue tasks demand extremely robust, high-TRL equipment that can be easily adapted into missions (Chiou, Epsimos, et al., 2022). Additionally, applications of e.g. semantic segmentation in CV have tended to focus on domains such as driverless cars, which offer a large market and capture more attention than the specialized first responder domain.

This chapter offers a step towards linking AI semantics research with disaster robotics R&D. Contributions include: i) identifying end-user semantic information needs and potential synergies; ii) identifying semantics research gaps to be bridged; iii) framing a coherent structure for thinking about these issues, in terms of a taxonomy that connects the work of these two research communities. Specifically, we address the following three research questions:

> **RQ1)** What semantic information is useful in the robotic disaster response operation?
> **RQ2)** How is semantic information structured in the robotic disaster response domain?
> **RQ3)** What AI-extracted semantic information can be feasibly used in disaster robotics field deployments, and how can we obtain it?

3.1 Common tasks for robots in disaster response

We first survey robot deployments in the response and recovery phases (Robin R Murphy, 2014; Schneider et al., 2015) of disasters from the perspective of commonly executed tasks. A variety of different robotic platforms have been deployed after disasters, e.g. floods, earthquakes, hurricanes, mudslides, or nuclear incidents. This discussion here is not meant to be exhaustive, but

rather indicative, to facilitate identifying useful semantics linked to required robot tasks.

UAVs can perform tasks that include mapping floods, additional flood risk estimation (R. Murphy et al., 2016), damage, or ground search (Fernandes, R. Murphy, Merrick, et al., 2019). Furthermore, UAVs can offer high-resolution remote imagery when satellites or manned aircraft are unavailable (Fernandes, R. Murphy, Adams, et al., 2018) and help inspect dams, bridges, buildings, or monitor the risk potential of consecutive disasters (Robin R Murphy, Duncan, et al., 2016).

UGVs are often used to gain access to environments that are dangerous for humans to reach (Casper and Robin R. Murphy, 2003). UGVs can be assigned to build 3D maps of the interior of buildings (Kruijff et al., 2012), assess damage to buildings (Kruijff-Korbayová et al., 2016), monitor temperature (Nagatani et al., 2013), radiation levels (Connor et al., 2020), retrieve samples (e.g. contamination) (Chiou, Epsimos, et al., 2022), or search for victims (Micire, 2008). In addition to reconnaissance tasks, UGVs are deployed to install devices, serve as communications relay stations, or provide camera views to control other robots (Chiou, Epsimos, et al., 2022). In addition, heavy-duty UGVs can be used for cutting or moving structures and debris, or, e.g. removing inflammable gas cylinders from industrial buildings during fires.

USV and UUV are deployed after floods and hurricanes (Robin R Murphy, 2014), oil spills, or shipwrecks. They can deploy cameras and sonars to detect damage to structures, locate submerged debris (Micire, 2008), estimate the volume of water in rivers and streams (Fernandes, R. Murphy, Adams, et al., 2018), and deliver supplies (Scerri et al., 2011).

We summarize key tasks in the response and recovery phases into four main categories:

- **Exploration and reconnaissance (ER) tasks** such as mapping, HAZMAT detection, sampling, and structural inspection (Kruijff et al., 2012; Micire, 2008; Fernandes, R. Murphy, Adams, et al., 2018; Robin R Murphy, Duncan, et al., 2016; Chiou, Epsimos, et al., 2022).

- **Auxiliary tasks** such as acting as a mobile beacon or communications relay/repeater (Agha et al., 2021; Chiou, Epsimos, et al., 2022).

- **First-aid tasks** such as in situ medical assessment and medically sensitive extrication (Robin R Murphy, 2014).

- **Heavy-duty tasks** such as rubble removal, shoring unstable rubble, providing logistic support, fire extinguishing, switching the valves and breaching obstacles (Robin R Murphy, 2004b; Nagatani et al., 2013).

3.2 Semantic information taxonomy

Here, we present our proposed taxonomy, which is based on the semantic information needed in the common tasks (See Section 3.1). We group the corresponding semantics into two main categories from the perspective of the human's understanding of different abstractions: low-level semantics and high-level semantics. Low-level semantics refers to information that is obvious and readily available with less effort to humans (e.g. object detection). High-level semantics refers to an abstract understanding of the environment that humans need further evaluation to infer (e.g. risk of hazardous situations). Lastly, we discuss the relationship between low-level semantics, high-level semantics, context, and the methods to obtain them.

3.2.1 Low-level semantics

Low-level semantics reveal the most salient features of the environment. Salient features mean that they can be noticed and understood by humans' direct intuition without further thinking, rather than mental effort for inference. We categorize low-level semantics into the following categories:

- **Objects:** detection of HAZMAT, victims, and other significant features of the environment (Sharifi, Zibaei, and Rezaei, 2020). Some detailed sub-category low-level semantics can also be used to describe the objects (e.g. distance and material).

- **Status:** a more detailed condition of the detected objects, e.g. cracks or delamination of concrete (Pozzer et al., 2021), or an object on fire.

- **Scene understanding:** an identification of the types of room or space which is significant for high-level semantics processing, e.g. if the space is an office or warehouse (Shi and Han, 2021).

- **Human behaviour:** detection of humans' movements in the post-disaster scenario, e.g. dangerous movements of victims (Jang, Battulga, and Nasridinov, 2020).

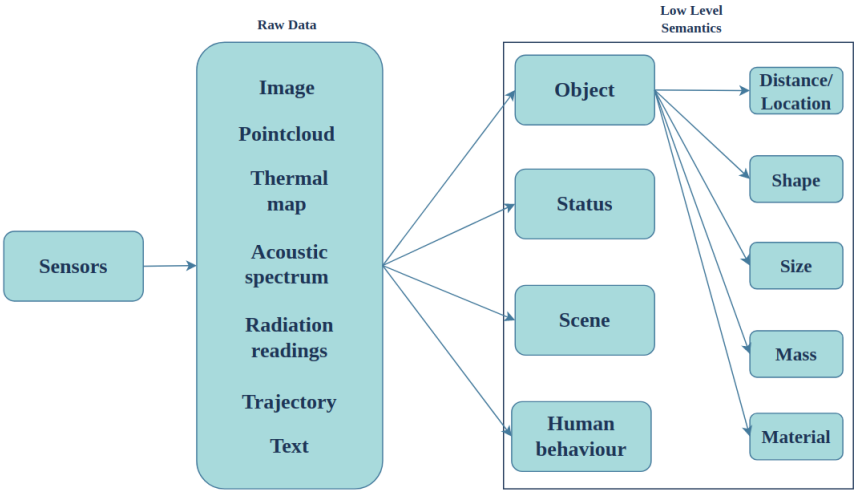


Figure 3.1: The relationship between the low-level semantics and raw data.

All the above categories help robots and humans to build an accurate SA and support tasks, including ER tasks, first aid tasks, and heavy-duty tasks (See Section 3.1).

We summarize the low-level semantics in Fig. 3.1. Note that a single low-level semantic feature can be potentially processed from different raw data sources. We can obtain the same low-level semantics from different sensors. For instance, we can obtain scenes from cameras, microphones, and LiDAR. Additionally, there is detailed sub-category low-level semantics (e.g. distance, material, and mass) constituting the understanding of objects. The synthesis among diverse types of raw data can extend the types of low-level semantics.

3.2.2 High-level semantics

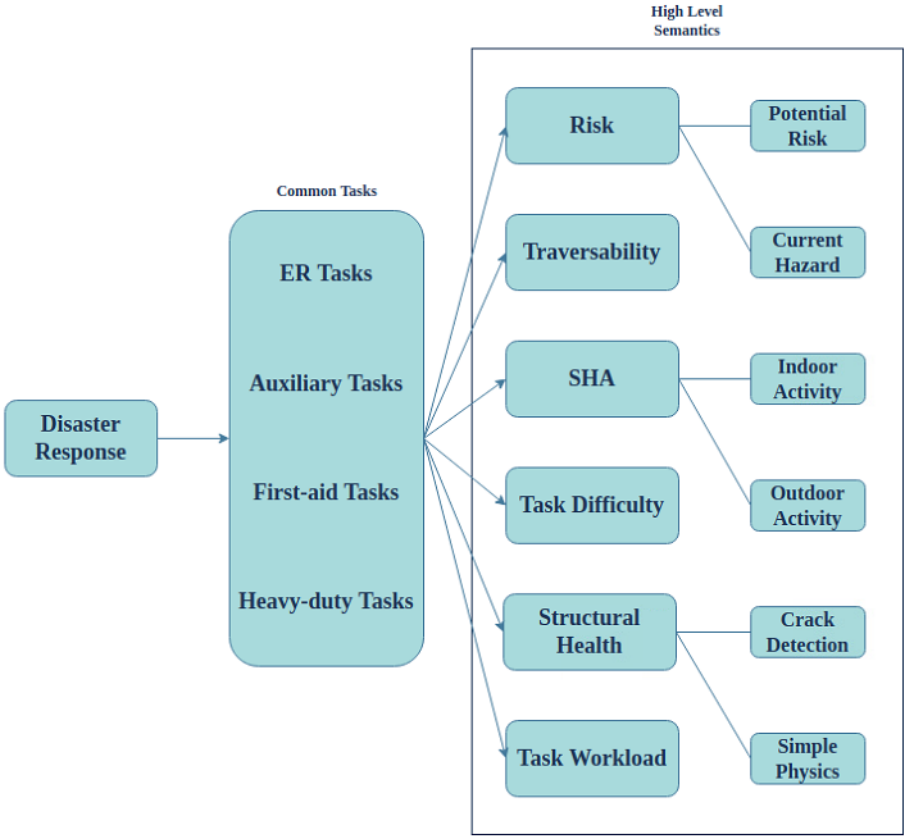


Figure 3.2: The proposed high-level semantics in relation to the common disaster robotic tasks.

High-level semantics is more complex than low-level semantics. It usually comes from the evaluation, estimation, or inference of a human's perspective about the onsite situation for robots and humans, rather than direct intuition. We group high-level semantics into the following main categories that are useful for disaster robotic operations (See Fig. 3.2):

- **Risk:** an assessment of the existing and/or the potential danger of the environment to humans and the robots.
- **Traversability:** the difficulty for humans or robots to travel through an area.
- **SHA (Sign of Human Activity):** the inference or estimation that victims or survivors can be potentially found in the scene.

- **Task difficulty:** the prediction of the difficulty for robots or humans to complete the given task.
- **Structural health:** the indication or assessment of how stable a structure is, e.g. the structure of debris or a building.
- **Task workload:** the estimation of how much work is required by the robots and first responders to complete the scheduled tasks.

We further divide risks, Signs of Human Activity (SHA), and structural health into more detailed subcategories. The category risk consists of potential risks and current hazards. The SHA is divided into indoor and outdoor activities because of the huge environmental differences. Indoor human activities can be inferred from personal belonging detection and indoor scene understanding, while outdoor human activities can be inferred from human trajectories and history population heatmap, e.g. from prior data in the cloud. Structural health can be determined by crack detection and "simple physics", which refers to the visual inference to mechanical structure. Task difficulty is divided in terms of semantics fusion, e.g. the extent of task completion, traversability of environment, expected time cost, and workload.

High-level semantics require information from low-level semantics and raw data as well. For instance, risks and SHA can be affected by detected objects, distance readings, and scenes. Moreover, high-level semantics might need additional information from other high-level semantics. For instance, task difficulty due to its complexity might not be accurately estimated just by low-level semantics. Other high-level semantics (e.g. workload and traversability) can also be considered as factors. Hence, high-level semantics require information from multiple inputs, including raw data, low-level semantics, and other high-level semantics.

3.2.3 Relationship among low-level semantics, high-level semantics, and context

Low-level semantics are processed from raw data. In contrast, high-level semantics are processed from the synergy of different low-level semantics, additional high-level semantics, and raw data.

High-level semantics processing relies on more different inputs than low-level semantics processing. Hence, low-level semantics is one of the bases of high-level semantics processing.

Context is often confused with semantics, as they have closely related meanings. Context is explained as "context now most commonly refers to the environment or setting in which something (whether words or events) exists" in Marriam-Webster dictionary [2]. From our perspective, context is the top level of abstraction, i.e. a larger concept that covers all the contents from low-level semantics to high-level semantics, from historical data to up-to-date information (See Fig. 3.3). The context describes a scenario or situation by using its components and the connections among the different components. It emphasizes the relationship among all the components, which may be semantic information, and sometimes historical information about the environment. In contrast, semantics denotes only part of the information that constructs the global understanding of the environment.

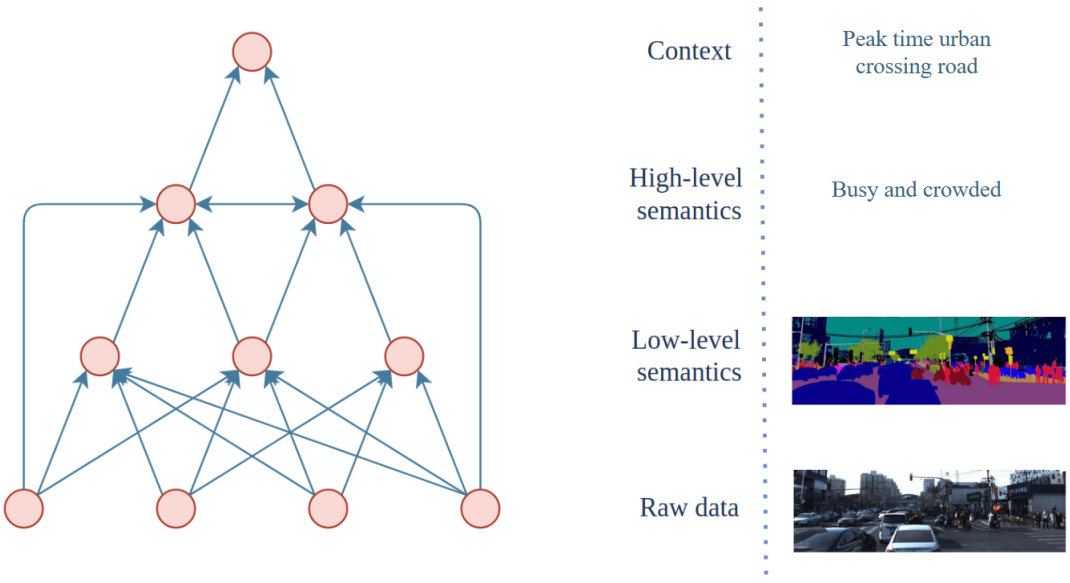


Figure 3.3: Different levels of information abstraction: each level of information is supported by the lower levels of information.

[2] https://www.merriam-webster.com/dictionary/context

3.3 From the taxonomy to implementation

In Section 3.2, we introduce the features of semantics and construct the taxonomy of useful semantics in the disaster response. Here, we aim to answer RQ3 from two aspects: **1)** SOTA semantic understanding and processing techniques and **2)** useful semantic information implementation for disaster response mission. We investigate RQ3 in Fig. 3.4 and obtain two corresponding insights:

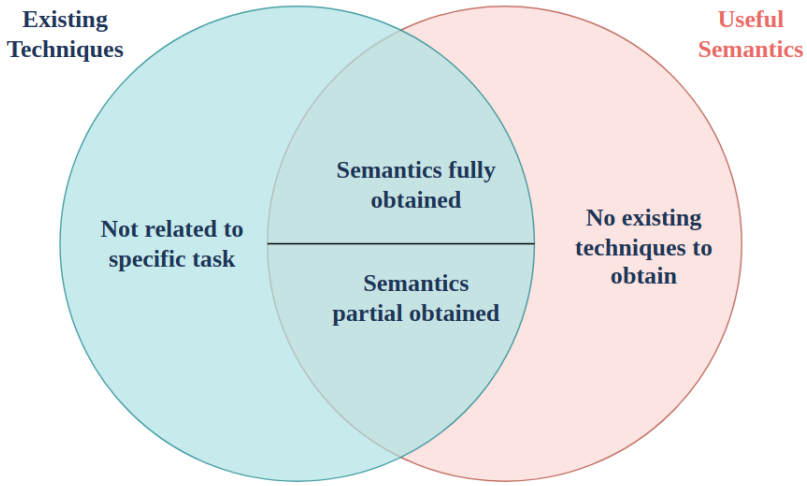


Figure 3.4: Two aspects of RQ3: the blue circle refers to existing techniques (See Insight I) and the red circle refers to useful semantics (See Insight II). The intersection region illustrates the current semantics understanding technique developments on useful disaster response semantics.

- **Insight I**: The AI community has developed large numbers of semantic understanding techniques. However, few of them have been deployed in the robotic disaster response operations.
- **Insight II**: We identified many useful semantics in disaster response based on the taxonomy, but there is a limited number of existing techniques to obtain these semantics.

3.3.1 Insight I

Here, we indicatively investigate the existing semantic understanding techniques from the perspective of disaster robot deployment.

Low-level semantics processing

We look into the low-level semantic processing based on the sensors that are popularly used, such as RGB cameras, LiDARs, microphones, thermal cameras, and radiation detectors.

RGB cameras provide colored images, and RGB-D cameras provide additional depth data. The AI community applies object detection (Girshick, 2015) and semantic segmentation (Long, Shelhamer, and Darrell, 2015) techniques to images to categorize and recognize objects. Besides, images provide more semantics e.g. human body keypoints (Bourdev and Malik, 2009).

LiDARs are widely used in SLAM and other robotic domains. They provide accurate 2D or 3D point cloud maps. (Peng et al., 2022) used an attention mechanism (Vaswani et al., 2017) to conduct a 3D LiDAR pointcloud semantic segmentation. Additionally, some researchers are focusing on LiDAR-only solutions without the help of RGB images (Milioto et al., 2019).

Acoustic semantic understanding using microphones has two domains. The first domain is NLP, which figures out the context and meaning of a dialogue or a conversation (Devlin et al., 2019). Another domain processes sound and noise from nature. It usually uses spectrogram representation to find embedded semantics like signal of lives underwater (Stolkin, Radhakrishnan, et al., 2007), scene (Battaglino et al., 2016), and location of sound (Okutani et al., 2012).

Thermal cameras outperform RGB cameras in poor illumination environments. Thermal images improve the performance of image semantic processing in dark environments where RGB cameras hardly capture clear images (Y. Sun et al., 2020). In addition to that, people can be easily identified in a complex environment (Jeon et al., 2015), and also damage to concrete structures can be detected (Pozzer et al., 2021).

Current research on radiation semantics is limited. Because the radiation readings can not provide further information, if they are not fused with other information. (Bird et al., 2018) used a UGV to generate a geometric and radiometric map of a radioactive facility with the assistance of a LiDAR sensor. It has exciting potential to process semantic information, e.g. identifying radiation sources, drawing heat maps, and evaluating dangerous areas.

High-level semantics processing

Since most semantic understanding techniques are focusing on low-level semantic processing, there are only two approaches to obtaining high-level semantics based on current techniques, which are traditional machine learning methods and deep learning methods.

The traditional machine learning methods are mathematical model-based and typically have two steps. First, building a mathematical model. We can quantify high-level semantics by using low-level semantics, additional high-level semantics, or raw data to build metrics. Second, applying traditional machine learning techniques to identify or classify different high-level semantics. To get different extents of one specific high-level semantics, we can apply rules using a mathematical toolbox, e.g. fuzzy logic, based on human cognition, evaluation, and inference to the designed metrics.

The second approach applies deep learning techniques. It is possible to build an end-to-end deep learning network with all the required input and high-level semantics output. However, the deep learning approach has two major defects: First, although a disaster semantic dataset from remote sensing imagery exists (Leach et al., 2021), we still lack specific semantic datasets supported by expert knowledge from first responders. We do not have any disaster semantic dataset based on the First-Person View (FPV) of UGV. Moreover, the creation of a new dataset is difficult because it requires expert knowledge and a large amount of handcrafted work. Second, deep learning networks increase the computation cost when compared with traditional approaches. This might lead to the degradation of real-time performance on robots with limited computational power deployed in the field.

Table 3.1: Indicative Summary of Advanced Semantic Processing Techniques (Part 1)

Reference	Sensors	Type of Raw Data	Semantics	Real Time Performance	Environments	Possible Tasks
Jang, Battulga, and Nasridinov, 2020; Ma, Marturi, Y. Li, Leonardis, et al., 2018	RGB camera	Colored image	Human behaviour	No	Indoor, outdoor	ER tasks, first-aid tasks
Simon et al., 2017	RGB camera	Colored image	Gesture	Yes	Indoor, outdoor	First-aid tasks
Ma, Marturi, Y. Li, Stolkin, et al., 2016	RGB camera	Colored image	Human pose	No	Indoor, outdoor	First-aid tasks
Bosch, Zisserman, and Munoz, 2007	RGB camera	Colored image	Objects	No	Indoor, outdoor	ER tasks, heavy-duty tasks
Yu et al., 2020	RGB camera	Remote sensing image	Object, scene	No	Outdoor	ER tasks
C. Zhao, L. Sun, and Stolkin, 2017	RGB camera	Colored image	Material	Yes	Indoor	ER tasks, heavy-duty tasks
Fei et al., 2021	LiDAR	Pointcloud	Object	No	Outdoor	ER tasks
Christie et al., 2017	LiDAR, RGB camera, radiation detector	Pointcloud, colored image, radiation readings	Object	No	Outdoor	ER tasks
Z. Zhuang et al., 2021	LiDAR, RGB camera	Pointcloud, colored image	Object	Yes	Outdoor	ER tasks, heavy-duty tasks
Zang et al., 2020	Thermal camera	Far-infrared image	Object	Yes	Outdoor	ER tasks
Pozzer et al., 2021	Thermal camera	Thermal image	Concrete condition	No	Outdoor	ER tasks, heavy-duty tasks

Table 3.2: Indicative Summary of Advanced Semantic Processing Techniques (Part 2)

Reference	Sensors	Type of Raw Data	Semantics	Real Time Performance	Environments	Possible Tasks
Masouleh and Shah-Hosseini, 2019	Thermal camera	Remote sensing thermal image	Object	Yes	Outdoor	ER tasks, first-aid tasks
J. Zhu et al., 2020	Infrared camera	Infrared image	Object and fire status	Yes	Outdoor	Heavy-duty task
Shi and Han, 2021	Microphone	Audio spectrogram	Scene	No	Indoor, outdoor	ER tasks
Ren et al., 2018	Microphone	Audio spectrogram, scalogram	Scene	No	Indoor, outdoor	ER tasks
Okutani et al., 2012	Microphone, RGBD camera	Audio spectrogram, colored image with depth	Location	Yes	Outdoor	ER tasks
C.-H. Wu and Liang, 2010	Microphone	Audio spectrogram	Human mood	No	Indoor, outdoor	First-aid tasks, heavy-duty tasks
Groves et al., 2021	LiDAR, radiation detector	Pointcloud, radiation readings	Terrain, radiation location	Yes	Indoor	ER tasks, heavy-duty tasks
Hellfeld et al., 2019	LiDAR, RGB camera, IMU, radiation detector	Pointcloud, radiation readings	Radiation location	Yes	Indoor, outdoor	ER tasks, heavy-duty tasks
Vetter et al., 2018	RGBD camera, radiation detector	Colored image with depth, radiation readings	Radiation location	No	Indoor, outdoor	ER tasks, heavy-duty tasks

We summarize the SOTA semantic understanding techniques in Table 3.1 and Table 3.2, and we contribute the following findings:

- RGB cameras and LiDAR are the most widely used sensors. They are good options to fuse.
- Most semantic understanding techniques concentrate on object detection and scene understanding. This leads them to be very useful in ER tasks, and heavy-duty tasks, but limited use in other tasks.
- Real-time performance is not considered by some of the researchers. However, for a disaster robotic operation, real-time performance is quite significant and needs more attention. Some of the computationally cheap techniques can be developed into real-time versions that are compatible with disaster robotics requirements.
- Due to the limitations of RGB cameras in dusty and poor lighting environments, many researchers focus on multiple sensor fusion to overcome the disadvantages of single-sensor solutions. However, these solutions stay on the laboratory level. Few solutions are applied, and they are limited in mapping the environments.

3.3.2 Insight II

The deployment of useful semantic understanding techniques in disaster response operations is still an open problem. Thus, we elaborate on Insight II by comparing both common tasks in the disaster operation and what the current semantic understanding techniques can achieve in Table 3.3, to indicate the open opportunities for deploying semantics in disaster response operations.

Table 3.3: Current status of processing important semantics in the scope of common disaster response tasks

Semantics	Object	Distance	Status	Scene	Human Behaviour	Risk	SHA	Structural Health	Task Difficulty	Traversability	Task Workload
ER Tasks	T[1] (C.-Y. Wang, Bochkovskiy, and Liao, 2023; Leach et al., 2021)	T	P (Pozzer et al., 2021)	T (C. Zhao, L. Sun, and Stolkin, 2017)	P (Ma, Marturi, Y. Li, Leonardis, et al., 2018; Jang, Battulga, and Nasridinov, 2020)	P (Sharifi, Zibaei, and Rezaei, 2021)	N[3]	P (Pozzer et al., 2021)	-[4]	T (Sevastopoulos and Konstantopoulos, 2022)	N
Auxiliary Tasks	T	T	P[2]	-	-	-	-	-	-	-	-
First-aid Tasks	T	T	P	T	P (Ma, Marturi, Y. Li, Stolkin, et al., 2016)	P	N	-	-	-	-
Heavy-duty Tasks	T	T	P	T	P (M. Zhang, M. Zhu, and Xuefeng Zhao, 2020)	P	-	P	P (Sevastopoulos and Konstantopoulos, 2022)	T	N

[1] T: technique exist.
[2] P: partial technique achieved.
[3] N: no technique.
[4] -: not related.

Based on the robot field deployments literature after disasters, a few semantic-related techniques are deployed. That is to say, the application of semantic information is limited. Hence, we define four levels to reveal the semantics applications status: technique exists, not applied in the field (T), partial techniques achieved (P), techniques do not exist (N), and not related to the tasks (-). We provide some references and make examples to illustrate the labels. Roof, ventilator, AC unit, and other building structures or components can be detected (Leach et al., 2021). They are useful in reconnaissance. Hence, we label the object "technique exists, not applied (T)". People are now able to detect the health of concrete via thermal and optical images (Pozzer et al., 2021). However, the real-time performance is not verified. Therefore, we considered it a "partial technique achieved (P)". For SHA, even though people can obtain human belongings semantics from the environment, we do not have any solutions to evaluate the intensity of human presence. In this case, we label it "N". The Table 3.3 reflects the following:

- Considering the features of auxiliary tasks (discussed in Section 3.1), semantics provides limited support to auxiliary tasks.
- Compared with low-level semantics, high-level semantics requires more focus from researchers. There are few high-level semantic applications in the field.
- Some of the semantics are close to being practically used (e.g. object, distance, and scene) while others are far from obtaining semantics (e.g. SHA, human behaviour, and task workload).
- It's clear that the semantic applications on ER tasks are more mature to be deployed in the field. For other tasks, because of their complexities, it's hard to rank their deployment possibility in the future. However, they all have big potential to involve semantic information in the sub-tasks and benefit the operation.
- For first-aid tasks, there are techniques to obtain the basic required semantics (e.g. objects, distance, and scene). However, we still lack techniques to obtain expert knowledge (e.g. identify injuries) related semantics.

The highlighted insights aim to encourage researchers to engage in grounding the advanced semantic understanding techniques on the practical disaster robotic field deployment and discover more semantic information that is useful for robots and humans.

3.4 Limitations

We note that:

- Our coverage of semantics has been somewhat skewed towards image-based methods, as much of the semantics literature is based on CV. This is partly because images carry more information than other carriers, and semantics is a rapidly growing topic in CV, with large advances in deep learning techniques in recent years. However, we have also attempted coverage of other sensing modalities as multi-sensor fusion seems likely to play an increasingly important role in overcoming the shortcomings of CV or single sensor solutions.

- This taxonomy is based on reports of robot deployments in disaster response. Such literature is limited because disasters are themselves rare events, and robot use remains limited. As researchers, we do not have first-hand experience in disaster response. However, our team members have participated in training and field exercises with Kerntechnische Hilfsdienst GmbH (KHG), Germany's nuclear emergency robot response team, and documented these experiences in (Chiou, Epsimos, et al., 2022). We have also undertaken extensive discussions, over several years, with key personnel who deployed robots at the 2011 FD-NPP accident (Kawatsuma, Mimura, and Asama, 2017). We encourage more direct interaction between robotics researchers and FRs.

3.5 Conclusions

This chapter aims to identify areas of potential synergy, as well as current gaps and challenges, between the AI semantic understanding community and the applied robotics disaster response

domain. Our taxonomy of semantic information provides an explicit understanding of semantics' intrinsic features, and their relationships to disaster response applications. These insights suggest directions for future research in semantics understanding techniques, to robustify, adapt, and apply these complex, high-level technologies to practical robot deployments in high-consequence, unstructured environments such as disaster response. Moreover, a clear taxonomy of semantics in the context of robot deployments provides a solid basis for our semantics-based SA framework design, presented in Chapter 5.

Chapter 4

First Responder User Requirements for Semantic Information

At the time of writing, the work in this chapter has been accepted by 2025 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR) (Ruan, Betta, et al., 2025).

This chapter explores the role of semantic information in practical field robotics deployments. Having designed a taxonomy of semantics in the previous chapter, we share these insights with FRs and seek their opinions on the role of semantics in their missions via structured questionnaires. Identifying the types of semantics used in the EE and FR robotics communities in this way is valuable. It helps us to understand if there are commonalities in the views of different EE and FR experts across various backgrounds and experiences, and also helps us to see a correlation between the views of these field experts and the ideas we developed in the previous chapter. Some researchers have collaborated with FRs to enhance SA through semantics on social media (Snyder et al., 2019). However, to the best of our knowledge, this is the first time that a questionnaire regarding semantic-based SA has been given to FRs and explicitly focuses on semantics.

This chapter explores the demands and challenges experienced by robot operators in the field, and attempts to learn lessons that help put this thesis into perspective. It is important that

robotics researchers carefully consider the end-users of the robot system. FRs' voices should not be ignored. We therefore attempted to capture and understand the valuable knowledge and experience of FRs and robot operators, by carrying out a survey, which is presented in this chapter.

Thanks to the fast iteration of new robotics techniques, commercial robots e.g. drones, or delivery vehicles, are becoming increasingly common in our daily lives. However, these developments in terms of robotic perception, planning, and control have not yet made significant inroads in practical FRs field robotics deployments. A significant gap between the lab and the field remains, especially for field robotics in EEs. Most autonomous robots are not designed for uneven and muddy terrain, radiation shielding, heat shielding, and other EE needs. From the perspective of researchers, the gap is not caused by the absence of specific techniques, but by the large-scale mismatching of the real environments and lab experiments (Huang et al., 2023). Shortcomings include low robustness of research technologies, low adaptivity to EEs, and limited communications bandwidth and intermittency which often occurs in the field. We therefore need to listen to the ideas and demands from the final end-users of these robots.

Relatively few FRs have directly collaborated with researchers to develop customized robots for specific missions. Researchers focus on the engineering side of the solution more and not on how the solution improves the SA of the end users. The effect on the SA is not measured, and the opinion of the end users has not been formally collected. Other works, such as (Betta et al., 2024) and (J. P. Murphy, Bergström, and Gyllencruetz, 2025), focus more on the human aspect of the task. The work detailed in (Betta et al., 2024) focuses on the requirements of FRs for robotic deployment in a SAR scenario. On the other hand, the work from Murphy et al. (J. P. Murphy, Bergström, and Gyllencruetz, 2025) focuses on the collection of the challenges in rural areas for what concerns disaster management operations. Both of these works propose a series of interviews with experts in the field to gather information from FR experts. These are efforts to narrow the gap between researchers and FRs. However, the knowledge sharing between FR and researchers is still limited and may often be superficial (Doherty et al., 2021). Hence, the adoption of robot deployments in EE missions is still limited. More efforts are needed to

better connect FRs and robotics researchers. Partly, this needs greater interest from the research community. However, partly we should understand that it is a difficult task, since FRs are often overstretched with very little spare time and resources to engage with researchers.

In this context, we survey FRs from different backgrounds and different missions to explore how they view current robotics technologies in their fields, and learn about the role of semantic information in their robot deployments. Moreover, the survey helps to verify if our work is valuable, and how far it is from being able to assist with real-world demands. The English version of our questionnaire can be seen in Appendix A.

4.1 Methodology

In this survey, insights and demands were collected via online questionnaires and paper questionnaire sheets. To reach industry experts in different countries, the questionnaires were translated from English into Italian, Greek, and Chinese by members of our multi-national research group at the University of Birmingham, particularly given the abstract concepts in the questionnaire. This helps to minimize any misunderstandings caused by second language users. The questionnaire consists of three sections: **Background and basic demographic information**, **General attitudes towards robots**, and **Situational awareness and use of semantic information**.

In the questionnaires, we investigate the FRs' backgrounds from two key dimensions: personal information and experiences with robots. The demographic characteristics include the following features: age, gender, country, education level, roles in the team/field, experience in the field, and experience with robots. Next, we examine general attitudes towards robots, as they may affect how FRs perceive domain specific robotics. The part of our assessment on attitudes draws on ideas from (Koverola et al., 2022), and the details are presented in Section 4.3. In the questionnaire section **Situational awareness and use of semantic information**, we survey FRs about their usage, experience, expectations, and overall opinions on SA and semantics.

By leveraging our professional networks and contacts, we successfully invited 22 FRs from eight countries to participate in the survey. All FRs are safety and emergency professionals

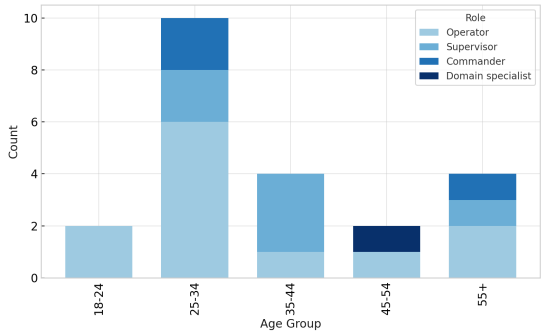
with relevant experience, covering multiple countries and emergency roles, e.g. firefighters, emergency medical personnel, technical operators, on-site commanders, operational scientists, and vehicle drivers.

In the data display and analysis section, the line plot, scatter chart, box plot, and violin plot are used to analyse data and compare distributions of responses. In addition, we explore connections between FR's views and their backgrounds.

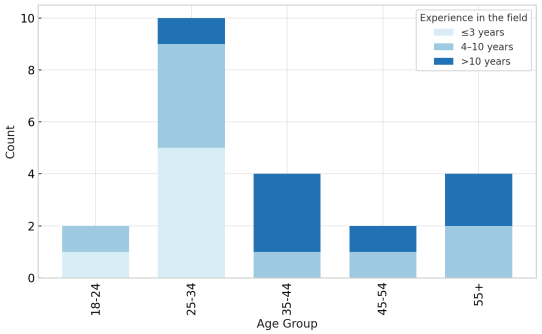
4.2 FR background and demographics

The recruited FRs are from eight regions: China, Japan, Germany, Iraq, Italy, Greece, the UK, and the USA. We understand there might be bias caused by the numbers of FRs from different regions. However, it is difficult to reach the FRs from different regions, let alone equal numbers. Fortunately, we notice they shared a similar standard of process and training in the field. We believe the bias caused by the unequal number would not affect the findings much. From the demographics, we notice females are much less represented than males. This is likely due to the heavy physical demands involved in many FR roles, which still tend to be male dominated in most countries. Only three female respondents participated in our study. Similarly, the age distribution shows a certain pattern. FRs below 34 years old make up the majority in the survey. They mostly work on the front line of missions, while older FRs are more senior in the field, serving as commanders or scientists who give expert instructions and guidance. For the same reason, FRs do not typically require a very high level of academic education. More than 60% of FRs in the survey achieved a degree below master's or equivalent. Even some senior commanders do not have a high level of education. However, they do require special training, skills, experience, and knowledge to tackle different types of emergencies. Hence, education level is not the most important factor in the mission. When talking about their expertise and experience, FRs show a balanced distribution. The junior and senior FRs take 29% and 33% respectively. In addition, we inquired about their robotics experience and the current robot deployment status from their perspective in the field. Important distributions can be seen in Fig. 4.1.

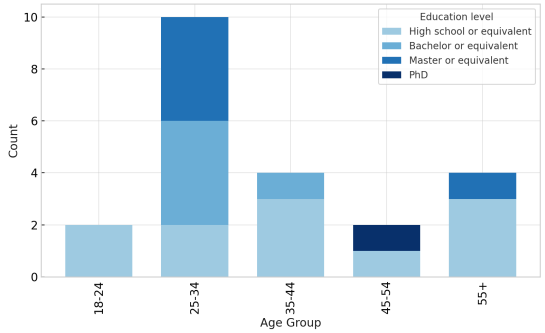
From the distributions in Fig. 4.1, we can see that there are relationships between certain background factors, such as field experience and age. Even though there are some unusual cases, e.g. senior FR personnel with only a high-school education level, the distribution reveals some patterns. Older FRs are more experienced, tend to be better educated, and typically take leadership roles when compared with younger FRs. However, the opinions of both older and younger FRs on current robot deployments in the field converge. Most FRs feel that robot deployments are still not very advanced, and few of the deployed robots can operate fully autonomously.



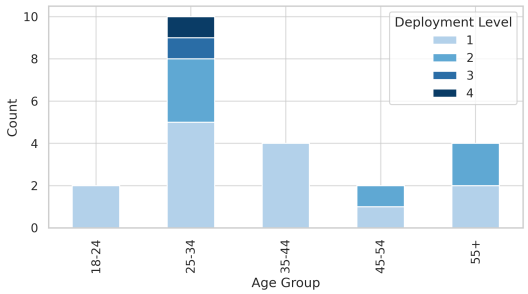
(a) Role distribution by age.



(b) Field experience distribution by age.



(c) Education distribution by age.



(d) Robot deployment level distribution by age.

Figure 4.1: Distributions of role, field experience, education level, and robot deployment level by age group. Robot deployment level refers to the frequency of using robots in the mission. Higher level means higher frequency.

4.3 General attitudes towards robots

As robots continue to grow in prevalence in society, people have developed different perceptions and attitudes towards them. Such background attitudes might potentially become a confounding factor in our survey, when we ask about more specific opinions on the deployment of robots

in FR and EE applications. We therefore first investigate and analyze participants' general attitudes towards robots as a reference, and then explore their attitudes towards robots in their own specialized application areas.

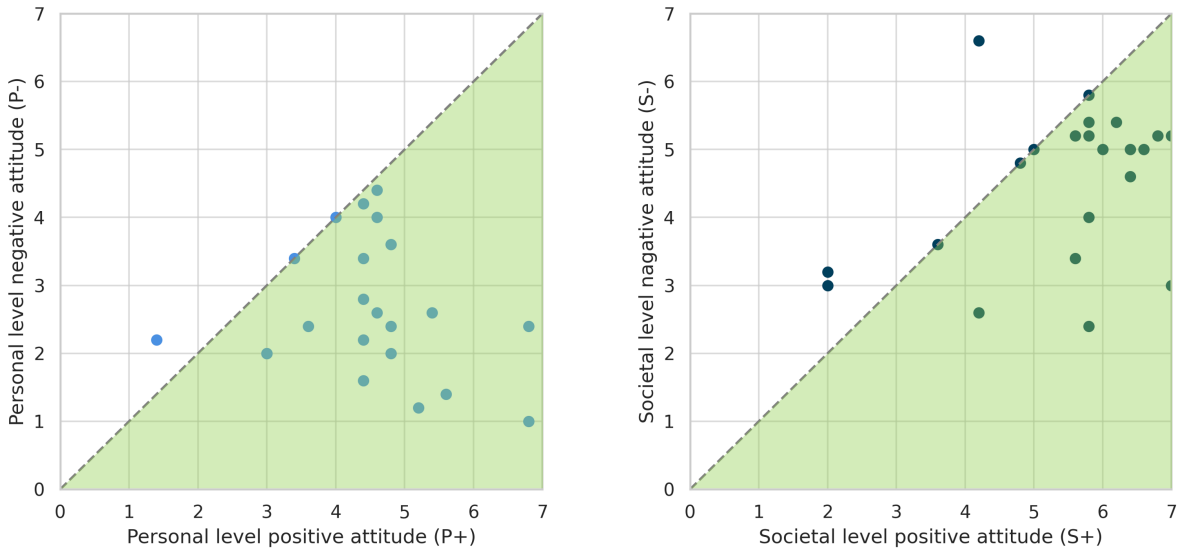
The general attitude section is designed based on (Koverola et al., 2022). The validated questionnaire is the General Attitudes Towards Robots Scale (GAToRS). The scale differentiates i) comfort and enjoyment around robots, (ii) unease and anxiety around robots, iii) rational hopes about robots in general (at societal level), and iv) rational worries about robots in general (at societal level) (Koverola et al., 2022). Hence, the questionnaire is divided into four dimensions:

- **Personal level positive attitude (P+)** evaluates personal positive feelings and emotions toward robots, e.g. relaxed, trust, and friendly interaction.
- **Personal level negative attitude (P-)** evaluates personal negative feelings and emotions toward robots, e.g. anxiety, upset, and fear.
- **Societal level positive attitude (S+)** evaluates expectations for the positive role of robots at the social level, e.g. assisting humans and removing humans from hazardous environments.
- **Societal level negative attitude (S-)** evaluates social concerns about the potential impact of robots, e.g. unemployment, regulations, and changes in social structure.

In four dimensions, different statements are given to participants. Participants need to assign a score that indicates to what extent they agree with the statements, e.g. "I can trust a robot.", "Widespread use of robots is going to take away jobs from people." The scale is from one to seven, with one denoting "totally disagree" and seven denoting "totally agree".

Apart from these four dimensions, additional four statements are used to obtain a clearer picture of participants' views. The additional statements are used to help detect whether the analysis drawn from the four dimensions is conflicting. The details of all the statements are listed in Appendix A.

The distributions of FR's attitudes are presented in Fig. 4.2:



(a) Personal Level Attitude

(b) Societal Level Attitude

Figure 4.2: Distribution of general attitudes towards robots on personal and societal levels. Most points are under the diagonal line (green zone). This indicates that most FRs have positive attitudes towards robots in both personal and societal contexts.

In Fig. 4.2a, all the points are close to or below the diagonal line (the diagonal line demarcates where positive and negative views are equal). Therefore all the FRs demonstrated more positive personal views than negative personal views. They indicate a certain degree of trust, relaxation, and welcome towards robots. At the same time, they have some concerns and reservations, however, these are less than their positive views. In contrast, in Fig. 4.2b, the participants show greater concern about robots at a societal level. In the figure, one can notice that most points cluster in the upper right area below the diagonal line. This suggests that most participants are confident about the positive influence of robots on society. However, they are more concerned about the negative impacts of robots on society, than about the negative impact on individuals. Moreover, there are two points above the diagonal line, i.e. two participants indicated strong concerns about the impact of robots on society in general. In the end, the additional four statements about participants' general views towards robot converge with the above four dimensions. In summary, most FRs are willing to trust and use robots without having very strong concerns about societal impact. It is apparent that the surveyed FRs are mostly pro-robots, which suggests that they might be willing to err on the side of giving the benefit of the doubt, and showing more

tolerance and patience to the introduction of robots in their missions. It is natural that they have a positive view of SA and the use of semantic information. However, if many of them express negative views about some aspects of SA and the use of semantic information, this could imply a serious level of concern.

4.4 Views about SA and use of semantic information

In this section, we would like to know more about FR's views on SA and semantics based on their understanding and experience. We give some examples and briefly explain the definition of SA and semantics to all FRs before collecting the data.

In this section, numerical questions with a marking scale and open-ended questions are provided to obtain more specific ideas on different occasions, e.g. different task backgrounds. Specifically, the numerical questions are listed below:

- **Q43** How much do you have an estimation of the Situational Awareness before the mission?
- **Q47** How often do you experience loss of concentration, fatigue, and cognitive overload?
- **Q49** To what extent do you think having semantic information is useful in building Situational Awareness?
- **Q50** To what extent is semantic understanding and reasoning used during missions?
- **Q53** Can semantic information and semantic scene understanding help predict or mitigate an emergency that was not predicted or foreseen before the mission?
- **Q54** Imagine you have some AI agent assisting in your mission (e.g. in a robot, in a computer, others). How much do you agree with the statement "It is helpful if the AI agent can process an understanding of the situation"?

These questions are answered on a scale from one to five, where one refers to a low level (e.g. totally disagree, low frequency etc.), and five refers to a high level (totally agree, high frequency etc.). The results of the numerical questions are shown in Fig. 4.3.

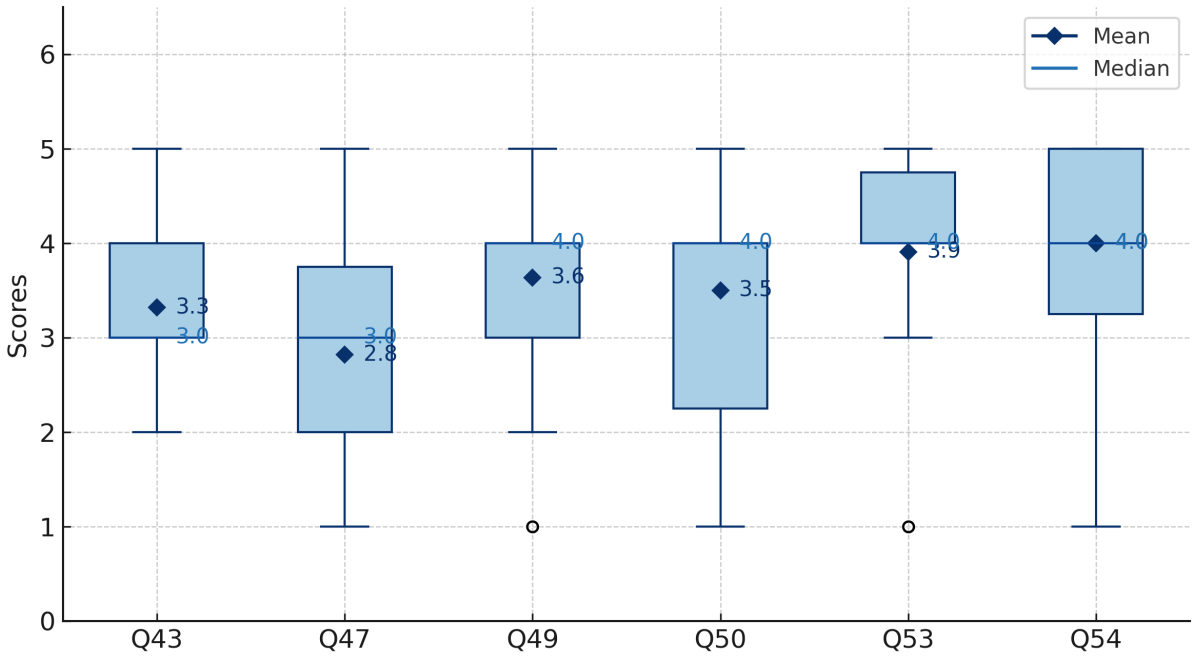


Figure 4.3: Score distributions of numerical questions: Q43, Q47, Q49, Q50, Q53, and Q54.

The open-ended questions covered a wide range of topics. They have two types of formats: fill-in questions and multiple choice (additional choices not listed are supported by adding after the last choice option). Below, we summarise the answers provided by participants. Where digits are given in brackets, these refer to the number of FRs who aligned with certain choices and options in their answers. Moreover, the answers to fill-in questions are presented with no preference, i.e. we do not rank the importance of the information from participants. All the details and open-ended questions are shown in Appendix A.

- **Q42 Information known before mission:** Type of event (location, hazardous material), extent, people/animal (distribution) involved; type of intervention required; specific task goals; weather forecasts; information from emergency call center that collected from people onsite; availability of resources (nearby medical center, airlifting services).
- **Q44 Methods to obtain SA:** Camera/images (15); sensor readings (13); communica-

tions with other people from the site, e.g. civilians, other team members (16); in-person exploration (16); comparison of current surveying data with previous status (e.g. Google maps) (1).

- **Q45 Challenges to obtain SA:** Unstable and low speed communication with the sensors; communication delay among team members; malfunction of perception devices; rapid changes of environments (seemed a safe area but rapidly changed to dangerous); data integration; incomplete data; complex terrains and debris; low visibility; darkness; complex air traffic control and conflicts for drone deployment (e.g. drones interfering with helicopters); lack of sound; lack of 360 vision; false or inaccurate information.

- **Q46 Top 3 important semantics:** Type of event (location, HAZMAT), extent, people/animal (distribution) involved; entities in the area (vehicles, victims, obstacles), impacts caused by interactions among entities (an ambulance blocked by an abandoned vehicle, risks from the HAZMATs); entities positioning in space and time (estimate path and schedule of an active ambulance); availability of resources (nearby medical center, airlifting services); object and hazard identification; task-relevant location information; distance to the object; available path.

- **Q48 Reason of fatigue, and cognitive overload:** 24 hours work shift rotations; chaotic situations; difficulties in collaboration and communication within the team; simultaneous presence of heterogeneous and fragmented information; time pressure to make decisions in the circumstance of having incomplete or uncertain information; high responsibility and fear of error; complex environments (high temperature, potential hazards); infrastructure and technological shortage; physical fatigue after intensive non-stop work; heat stress due to wearing protective gear like Tyvek suits; victims; situation (injury, death); overload information; communication with many stakeholders; lose control of the robot.

- **Q51 Current method to deliver semantics:** Vision (19); sound (15); haptics (4); vibration (4).

- **Q52 Ideal method to deliver semantics:** Vision (20); sound (19); haptics (7); vibration (11).

Finally, two questions about semantics accuracy were surveyed: the accuracy threshold for having trust in semantics (**Q55**); and the accuracy threshold for usefulness of semantics (**Q56**) (See Fig. 4.4). Only 19 participants gave a specific accuracy. The other two participants provided individual insights about this comparison. One participant thinks any accuracy is fine, as long as FRs are informed about the reliability of the accuracy. Another participant feels that the accuracy requirements vary depending on different occasions. But this participant feels that such sensors should show better performance than human perception and inference.

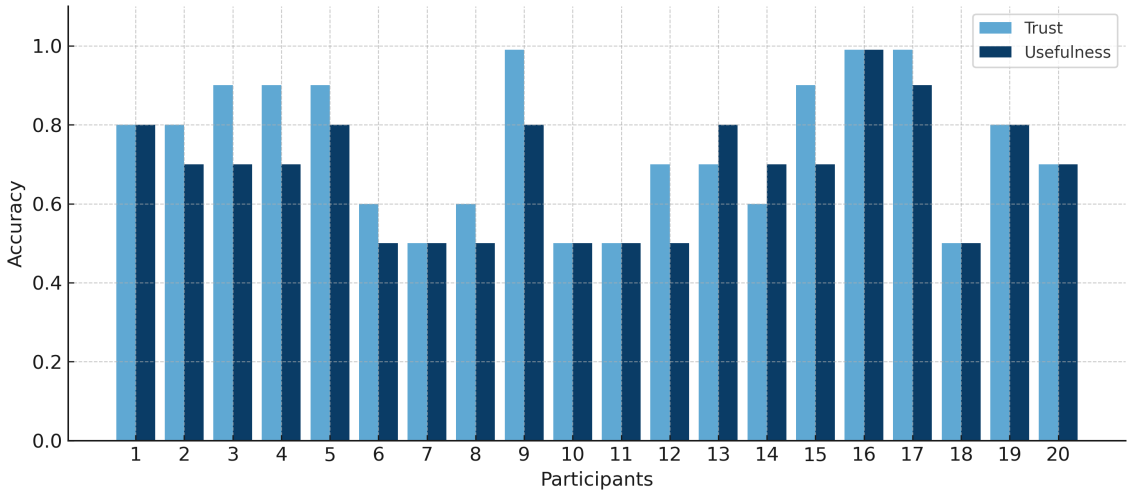


Figure 4.4: Accuracy threshold values for achieving trust and usefulness: light blue bars refer to the accuracy threshold for achieving trust (Q55), and dark blue bars refer to the accuracy threshold for achieving usefulness (Q56).

4.5 Analysis and discussion

Due to the number of FRs in this study, we briefly analyze their feedback by discussing their mean, median, and distribution. Moreover, combined with their background, we discuss the insights and findings from their reported field experiences.

- **Q43 Estimation of SA before an incident:** From the responses, the mean SA was 3.3, and the median was 3.0 (See Fig. 4.3). In Fig. 4.5, it is seen that FRs with a higher position

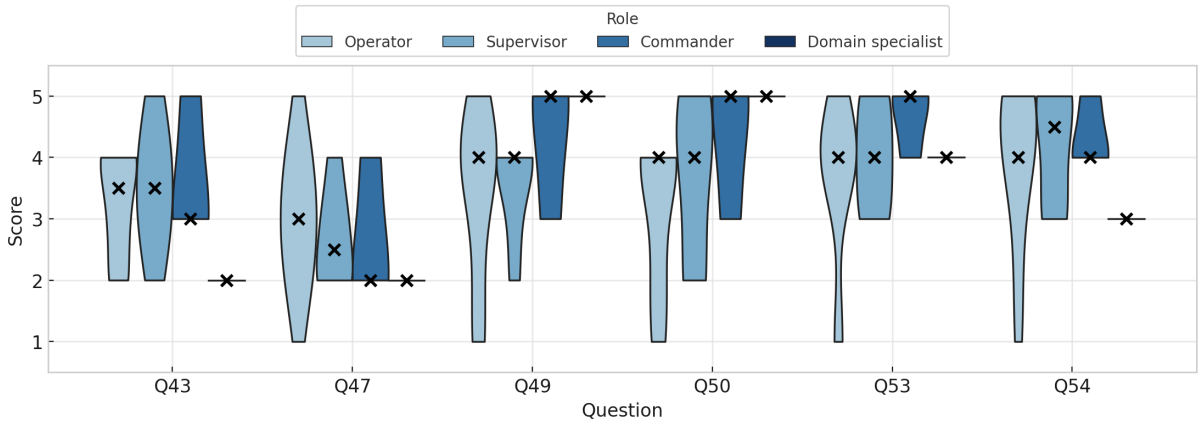


Figure 4.5: Score distribution by role: crosses refer to the median in the group. In each question, the roles displayed from left to right are operator, supervisor, commander, and domain specialist.

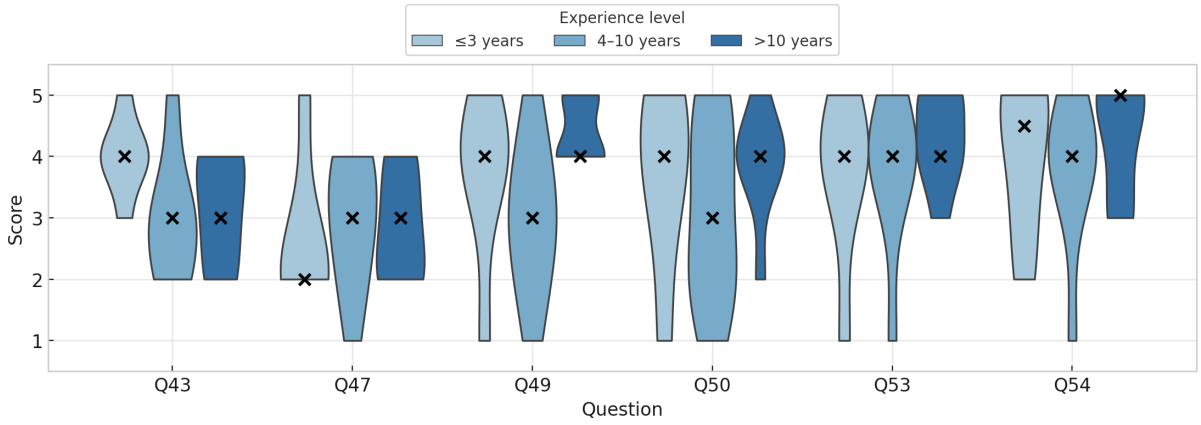


Figure 4.6: Score distribution by experience level.

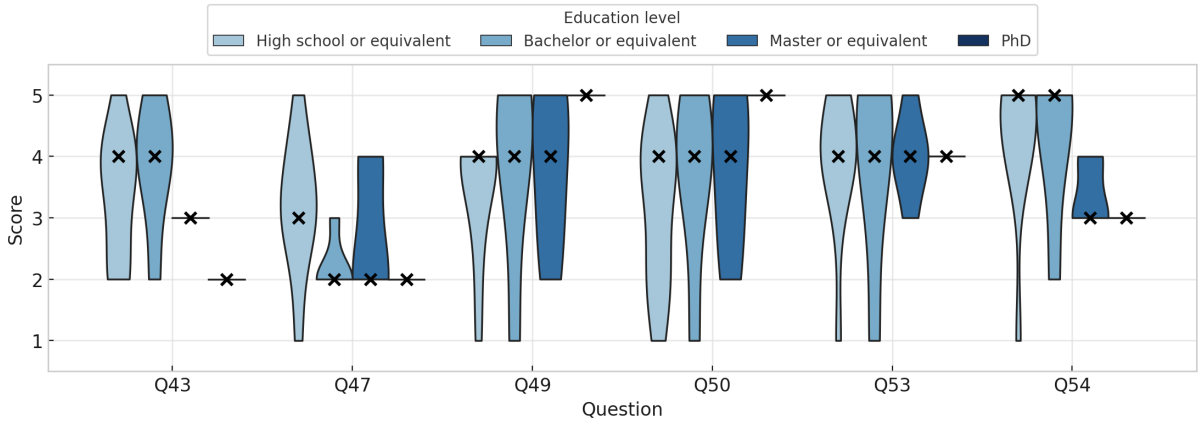


Figure 4.7: Score distribution by education level. In each question, the education level displayed from left to right is high school, bachelor, master, and PhD.

are usually more informed about the incident's SA compared to the operators. Additionally, in Fig. 4.6, a higher score for SA is seen by less experienced operators compared to more years of experience. A potential reason for this difference is that the senior FRs (leaders) need to think more about the overall mission, whereas junior FRs handle specific lower-level tasks that require less SA. I.e. senior FRs might require higher levels of abstract information e.g., high-level semantics and context.

- **Q44 Methods to obtain SA:** FRs use a variety of methods to obtain SA. Communication with other people from the site, and in-person exploration, are the most popular ways to obtain information, suggesting that human perception still dominates in obtaining SA. FRs have not yet deeply depended on robot sensors such as cameras. However, such sensors are still welcomed in many deployments and gain 15 votes out of 21. Moreover, a comparison of the current scene against historical data is mentioned. Since this type of information requires support from other departments or individuals and is not always available, it works as a supplement to first-hand information and observations.

- **Q45 Challenges to obtain SA:** The challenges come from different aspects of field deployments, including hardware and device malfunction (Ramachandran et al., 2021), complex environments, lacking adequate perception, disorganized data, and communication problems with team members.

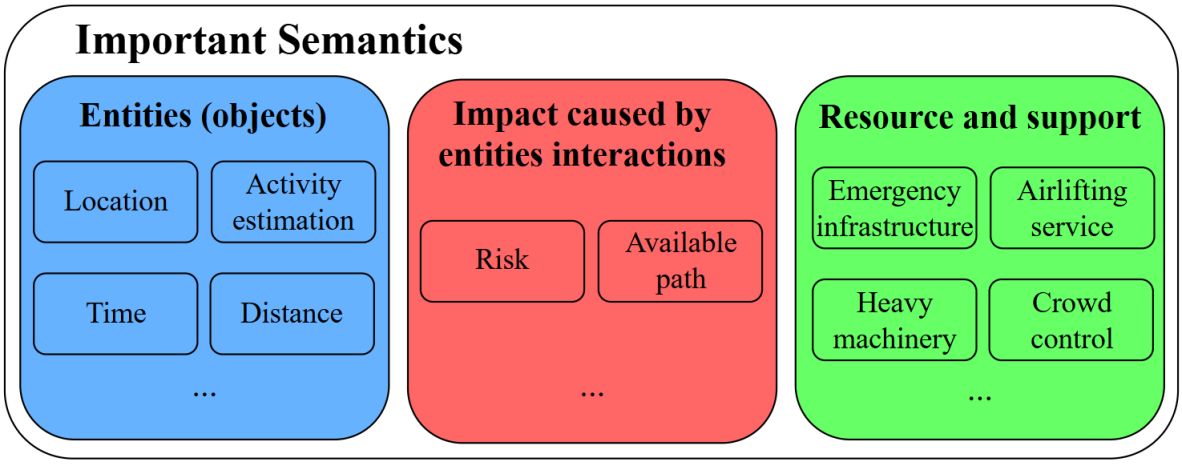


Figure 4.8: Important types of semantics for FRs during missions according to the FR's responses.

- **Q46 Top three important semantics:** Based on the feedback of the FRs, we summarize the important semantics in the mission in Fig. 4.8. If we fit these semantics into the proposed semantic taxonomy from Chapter 3, we find many of them are low-level semantics that can be obtained by robots e.g. location, distance in the blue box. Some of them are high-level semantics e.g. risk in the red box. And many of them are context e.g. medical center, emergency infrastructure in the green box. According to the answers, Junior FRs, who work on the front line and handle the specific operations, focus on the semantics in the blue and red boxes. Senior FRs, who are often the team leader or commander, need to do overall management of the mission. They focus on the semantics and context in the red and green boxes more often.

- **Q47 and Q48 Loss of concentration, fatigue, and cognitive overload:** Results show most FRs experienced a slight loss of concentration, fatigue, and cognitive overload (mean 2.8 on a score from one to five, median 3, see Fig. 4.3). The sources of fatigue and cognitive overload are tasks, environments, and ethical professional constraints. This problem will not be solved by the application of a specific technology, but can be partially alleviated by the use of semantics (See Section 6.4.3).

- **Q49, Q50, Q51, Q52 and Q53:** These five questions investigate FRs' views on semantic information. Most FRs feel semantics has a positive effect on obtaining SA (mean 3.6 on a score from one to five, median 4, see Fig. 4.3). FRs with a higher education background and experience believe semantics help SA more (See Fig. 4.7). Similarly, FRs with higher education level and experience use semantics more in the mission (mean 3.5 on a score from one to five, median 4, see Fig. 4.3). Most FRs have the consensus that semantics help to predict the situation (mean 3.9 on a score from one to five, median 4, see Fig. 4.3). In terms of the delivery methods of semantics, current methods are mostly a combination of vision and sound (from fixed or robot cameras at the base camp). Haptics and vibrations are used in a few special tasks from the joystick, e.g. the deployment of drones. Hence, these answers denote that multimodal is the best option to build SA. FRs who operate

the drones are expecting to see an increase in haptics and vibration in other robotic tasks, along with the increasing application of semantics in the future.

- **Q54 If AI understands the situation, helpful?:** FRs have the consensus that AI understanding is valuable in the mission (mean 4 on a score from 1-5, median 4). They probably expect the AI-generated SA to help them reduce the cognitive workload and avoid potential risks.

- **Q55 and Q56 Accuracy threshold changes between trust and usefulness:** The mean accuracy threshold for achieving trust in semantics is 74.6% (i.e. participants will trust the semantics if it has over 75% accuracy), and the mean accuracy threshold for usefulness is 67.8% (i.e. participants will use the semantics if it has over 68% accuracy). In Fig. 4.4, it can be seen that the trust threshold is generally higher than the usefulness threshold for most participants. It indicates that having semantics can be useful even if participants do not feel they can completely trust them. In other words, FRs are expecting to see increasing amounts of semantic information becoming used in a useful way in the field, but it will take longer before these are fully trusted. We further compare the background of FRs who have a higher accuracy threshold for usefulness than for trust. These tend to be younger and more junior FRs with a high school education or equivalent level. Possibly this reflects difficulties in understanding the difference between our use of the terms trust and usefulness in the survey. Moreover, some FRs believe the threshold of trust and usefulness in this context is the same concept. That is, if they are willing to use a certain type of semantic information, it means that they already trust it. We assume the difference might come from variation in the different types of tasks that each participant performs in their work. Some tasks and decisions might have much stricter requirements on data accuracy than others. Hence, these FRs might emphasize the accuracy of the semantics more highly.

We now specifically discuss the connections between the FR's views and their backgrounds, roles in the mission, experience in the field, and education levels.

From Fig. 4.5, we find that different roles in the mission do not cause much variation in opinions. Hence, the role might not be the main reason that differentiates the FR's views. However, different levels of experience in the field do correlate with differences in opinions. In Fig. 4.6, FRs with more than 10 years of field experience have an overall higher rating on Q49, Q50, Q53, and Q54. These questions are related to new technologies for semantic-based SA in future deployments. This suggests that senior FRs are more positive and have higher expectations on this compared with other groups. FR with four to ten years of experience have a more even distribution on these questions. And junior FRs (less than three years of experience) show much greater uncertainty about their thinking. In the context of education from Fig. 4.7, it's interesting that the score on Q43 shows a decline when education levels increase. This suggests that FRs with higher education tend to be cautious before the mission. They might take more factors into account that are overlooked by FRs with lower education backgrounds. Moreover, there is a decline in distributions on Q54, suggesting FR's with higher education levels do not show the same level of confidence in AI applications for building SA. They show more independence in AI assistance than other groups.

We gained some additional insights from FR comments in the open-ended questions. For instance, one FR highlights the importance of distance information, which could help reduce potential risks. One senior FR with experience in robot deployments, especially drone deployment, feels that the acceptance accuracy for trustworthy and useful semantics varies according to different missions and scenarios. By analysing his background, we found that he was the leader of the squad and participated in many different types of response missions.

FRs' background of role, experience level, and education level are inevitably have correlations. However, the major goal of this study is not exploring the hidden pattern between these factors and their views about SA and semantic information. Thus, we did not discuss this topic in detail.

In summary, most FRs across a variety of backgrounds are willing to apply semantics to help them build SA. On one hand, senior FRs have more experience in using semantics to obtain SA, and show higher expectations for using semantics-based SA in the future. On the other hand,

they do not feel so dependent on assisted SA compared with other groups. They have enough expert knowledge and experience to leverage information from different sources and mitigate or overcome the impact of errors in the system or incomplete information.

4.6 Conclusion

This survey has explored views about SA and semantics in the context of robot deployments of FRs. It reveals that **a)** most surveyed FRs' have positive attitudes towards robots and expectations of more robot deployments. Moreover, this survey summarizes **b)** the important semantics in the mission, and some of them could be obtained with the assistance of the AI and robotic techniques, e.g. using images or sensor fusion techniques to obtain semantics such as object identification and positioning, and more abstract situational assessments such as risk. The analysis provides **c)** insights into their expectations regarding semantics-based SA. Most FRs are expecting and willing to apply AI-based semantics in their missions, even though it does not achieve a high accuracy and robustness of SA. I.e. they could deploy a robot that does not meet their detection accuracy expectation. And such semantic information could be delivered by a well-organized multi-modal way e.g. visual display, and sound notifications.

In summary, this survey explores FRs' understanding of robots, SA, and semantics. And Chapter 3 outlines the picture of robot deployments in the field. Together, we see the gap between FRs' requirement and current robot deployments. Specifically, the application of different levels of semantics is expected to build a comprehensive SA with less human effort by FRs, which might help to reduce their workload, and avoid potential risks. Hence, in the following chapters, we propose a SSR framework that provides low-level and high-level semantics (See Chapter 5), helping HRT build a comprehensive SA. Furthermore, in Chapter 6, we explore how this semantic information would impact the HRI and HRT.

Chapter 5

A Situational Semantic Richness Framework for Mobile Robot Deployment

At the time of writing this thesis, the work in this chapter has been accepted by the journal Frontiers in Robotics and AI (Ruan, Ramesh, H. Wang, et al., 2025).

Building on our definition and taxonomy of semantics for field applications in Chapter 3, this chapter proposes a semantics-based SA framework. The framework synthesizes a variety of low-level semantic indicators, and then computes an overall evaluation of the SSR. Moreover, this chapter presents the results of implementing this framework on a real robot.

SA is vital for robots deployed in the field to function with sufficient autonomy, resiliency, and robustness. This is especially true for HRT in safety-critical applications such as disaster response, remote inspection in unstructured environments, or nuclear operations (Chiou, Epsimos, et al., 2022; Ruan, H. Wang, et al., 2022; Stolkin, Molitor, et al., 2023). In all cases, humans and robots require SA to make plans or decisions in the context of HRT (e.g. identifying a proper timing to switch control between human operators and robots). Hence, humans and robots need to know and share what is happening in the environment to plan and act in a safe and coordinated manner. Humans and robots have different and distinct strengths and weaknesses in terms of perception, ability to interpret sensory data, and ability to plan and execute decisions in response to that data in real time.

Building on low-level signals from multiple modalities of on-board cameras and sensors, higher-level "semantic" understanding (Ruan, H. Wang, et al., 2022) of scenes, environments, and situations must be developed. Often, this higher-level semantic knowledge will be critical for determining subsequent decisions and actions. Recent advances, especially from the CV community (Long, Shelhamer, and Darrell, 2015; X. Li et al., 2017), have begun to provide autonomous agents with some elements of semantic-level perception. Meanwhile, in real-world robotic systems at present, the intelligence of human operators may often be necessary to correctly interpret and act upon semantically rich situations. In this chapter, we propose a framework for robots to acquire semantically enhanced SA that combines with human understanding in an explainable and intuitive way.

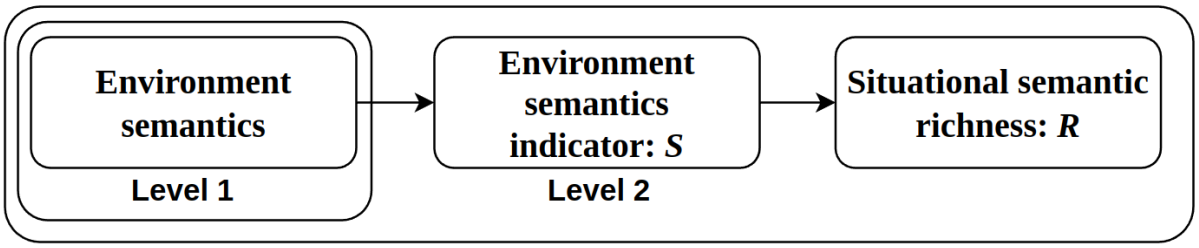


Figure 5.1: Mobile robot semantic SA with two levels.

SA can be modeled as three levels of awareness (Endsley, 1995b): Level 1) perception of elements in the current situation; Level 2) comprehension of the current situation; Level 3) projection of future status. In the robotics and AI research literature, it is common to use terms such as sensing, perception, scene understanding, semantics, and context (Bavle et al., 2023) instead of SA. There are connections among these related concepts, e.g. the concept of perception "elements" in SA can be linked to the "semantics" concept in AI. Hence, although the conventional SA model is designed to represent the awareness of human operators, the SA of an autonomous or semi-autonomous robot can be structured similarly in the scope of semantics. Elements of Level 1 SA can be objects, sensor readings, and other low-level semantics (Ruan, H. Wang, et al., 2022). The comprehension of the current situation at Level 2 corresponds to high-level semantics (Ruan, H. Wang, et al., 2022) (See Fig. 5.1). Prediction, planning, or decision-making, based on these, are the main focus in Level 3.

This chapter aims to build a systematic framework and concepts that: **a)** make SA sharing from robot to human easier, practical, and intuitive; and **b)** facilitate the use of semantically enhanced SA in HRT planning and decision-making frameworks. We build upon our previous work that proposed a taxonomy of semantic information which discussed in Chapter 3 and definitions of low-level semantics, high-level semantics, and the context in robot-assisted disaster response. We explore Level 2 SA and propose the concept of an "environment semantic indicator" along with an example realization. Each indicator captures the understanding of an environment's semantics intensity (e.g. SHA). Furthermore, we build upon those indicators to develop an aggregated metric, the "Situational Semantics Richness" (SSR), which expresses the overall intensity and plethora of semantic information in an environment.

5.1 Individual semantic indicators

5.1.1 Laser noise intensity

Many UGVs rely on lasers for autonomous navigation. However, laser noise potentially affects the navigation. We adapt the method to obtain laser noise variance (σ^2_{noise}) in the previous work from our lab (Ramesh, Stolkin, and Chiou, 2022). It is calculated by convolving the laser map image with a 3x3 mask (computation cheap with low latency) and applying summations on the resultant matrix. Then, we adapt the noise variance into a sigmoid function (See Eq. (5.1) and (Ramesh, Stolkin, and Chiou, 2022)) to obtain the laser noise intensity $S(noise)$. The noise variance thresholds ($\sigma^2_{noise} \approx 1.4$) is obtained by our test in the environment of halting the robot navigation based on our hardware setup. Combining the preliminary test results, the laser noise intensity is designed as:

$$S(noise) = \frac{1}{1 + \exp(-a \cdot \sigma^2_{noise} + a \cdot b)}, \tag{5.1}$$

where $a = 5$ and $b = 1$ to have a curve that responds to medium inputs but is not oversensitive to low and high inputs.

5.1.2 Risk to robots

Risk to robots can be quantified based on hazard level and hazard length (Soltani and Fernando, 2004). Hazard level refers to how dangerous the hazard is to the robot, and hazard distance refers to the distance to the object. Furthermore, time may also affect the risk level, e.g. the accumulated dose received from the radiation sources.

In this chapter, we assume that the robot can detect potential risks by detecting HAZMAT signs that commonly exist in hazardous environments. In general, humans and robots face similar risks. However, considering the slight difference between risk to humans and risk to robots, the categorization might differ but can be trivially adapted to reflect human risks, expert knowledge (e.g. by FRs), or different scenarios. We make a categorization of HAZMAT signs into three levels heuristically: low risk or no risk for the robot (e.g. poison, infectious substance, nonflammable gas, inhalation hazard), medium risk (delayed hazard to a robot or they can be high risks under certain occasions, e.g. corrosive, radioactive, dangerous when wet, oxygen, organic peroxide), and high risk (immediate hazard to the robot, e.g. explosives, flammable solid, flammable gas, spontaneously combustible material). Other risks not included can also be added trivially in different scenarios.

Intuitively, distance is a factor that relates to the risk intensity. Referring to the relationship between radiation strength and distance (Voudoukis and Oikonomidis, 2017), we apply a similar model to risk to robots $S(risk)$ by using the inverse square law:

$$S(risk) = \sum_{i=1}^{n} \frac{H_j}{d_i^2}, \tag{5.2}$$

where n is the number of detected signs; i is the label of each detected sign; d refers to the distance to the robot; j refers to the level of the corresponding signs; H_j refers to the risk intensity of each level of HAZMATs sign.

Then, we normalize the risk score by applying the sigmoid function. The reason for using the sigmoid function is that when the x-axis approaches infinity, the slope of risk is low and accords with human common understanding, i.e. the environment which has 6 high-risk objects has a similar $S(risk)$ to the environment containing 5 high-risk objects. The normalized score is:

$$S(risk)_{norm} = \frac{1}{1 + e^{-aS(risk)-b}}, \tag{5.3}$$

where a and b are used for tuning the functions. In our experiments, we set them as $a = 0.09$ and $b = -4.8$ heuristically to get a meaningful and usable curve. Experts can tune a and b for different tasks.

5.1.3 Sign of human activity

Robots might not always directly detect human victims in the environment (e.g. trapped under debris or occluded by objects). Hence, robots must identify clues to find victims. SHA are considered a potential factor in finding people (Yang et al., 2018). We use human belongings, e.g,. mobile phones, keys, and watches, as indicators of human activity. Intuitively, the dispersion of personal belongings makes a difference to the SHA. Thus, the SHA is modeled from two aspects: the class of objects and the dispersion of the objects.

We design three classes of objects to differentiate the impact of different human belongings: i) high impact means there is a high chance to be found on the human body (e.g. glasses, key, cell phone, watch); ii) medium impact means there are chances to be found in proximity to the human body (e.g. cap, mask, wallet); iii) low impact means there is a high chance not be carried on the human body (e.g. laptop, backpack). This heuristic classification is an example of realization for our framework, and experts can adjust it.

Based on the above, we propose the following model to estimate the SHA score $S(SHA)$:

$$dispersion = \sum_{i=1}^{n} |d_i - d_{average}|,$$
$$S(SHA) = \begin{cases} \sum_{i=1}^{n} \frac{P_j}{dispersion} & dispersion \; != 0, \\ +\infty & dispersion = 0, \end{cases} \tag{5.4}$$

where n is the number of detected objects ($n > 1$), i is the label for each object, d_i is the distance from the corresponding objects to the robot, $d_{average}$ is the average distance of all the objects, j is the label of class for objects, P_j refers to the impact of corresponding objects. In the special case that the $dispersion = 0$, which means objects are at the same distance from the robot. Considering the limited view of the range of the camera, we believe this case deserves human attention ($+\infty$ is given to this case).

Similarly, we normalize the $S(SHA)$ using the sigmoid function:

$$S(SHA)_{norm} = \frac{1}{1 + e^{-aS(SHA)-b}}, \tag{5.5}$$

where a and b are used for tuning the function. In our experiments, we used $a = 10$ and $b = -0.5$ to have a curve that has similar sensitivity characteristics to Eq. (5.1). The parameters can be adjusted when expert knowledge is involved.

5.1.4 Radiation

Not only humans but robots may also be affected by radiation (Nagatani et al., 2013). The highest risk will be that of the onboard electronics, as radiation can cause disruptions, malfunctions, or even complete failure of electronic components. These issues can be addressed by applying appropriate shielding and mitigation techniques or designing radiation-hardened robots. Alternatively, we might monitor the radiation strength in deployment so that robots and humans can avoid exposure.

The risk associated with radiation depends on the distance from the source and the radiation's

type, strength, and energy. It is important to distinguish between the dose rate and the total integrated dose. Dose rate is commonly measured in sieverts per hour (Sv/h) and microsieverts per hour ($\mu Sv/h$), which is the rate at which the radiation is received at a given moment. The total integrated dose measured in sieverts which results from the accumulation of radiation over time.

We designed a mapping $I : \mathcal{G} \mapsto S(radiation)$. It translates the raw readings, which require radiation backgrounds for understanding, to metrics ranging from 0 to 1. $\mathcal{G}$ is a set of gamma radiation dose rates G in $\mu Sv/h$ from the sensor, and $S(radiation)$ denotes the radiation strength. Specifically, the sigmoid function is applied in the mapping I to normalize the $S(radiation)$:

$$S(radiation)_{norm} = \frac{1}{1 + e^{-aG-b}}, \tag{5.6}$$

where $a = 8$ and $b = -8$. Specifically, $S(radiation) = 0$ refers to no radiation, and $S(radiation) = 1$ refers to the radiation that can instantly damage the robot. This tuning setting takes into account the effects of background radiation (under 0.4 $\mu Sv/h$ usually) and decreases its impact on the final output.

5.2 Overall Situational Semantic Richness score: aggregating semantic indicators

With these semantics in Section 5.1, we propose a framework (shown in Fig. 5.2) that fuses semantics of the environment on a higher level, i.e. a metric that describes the overall status of the environment in an aggregate representation.

In real-world situations, if we do not have a dataset, applying data-driven approaches is impractical or intractable. To the best of our knowledge, no dataset involves all the environment semantics, i.e. it is not feasible to build a parametric model and train an end-to-end network to assess the situation. Moreover, SA is a subjective understanding and needs to be intuitive and explainable, especially in safety-critical and hazardous applications. Thus, we must capture

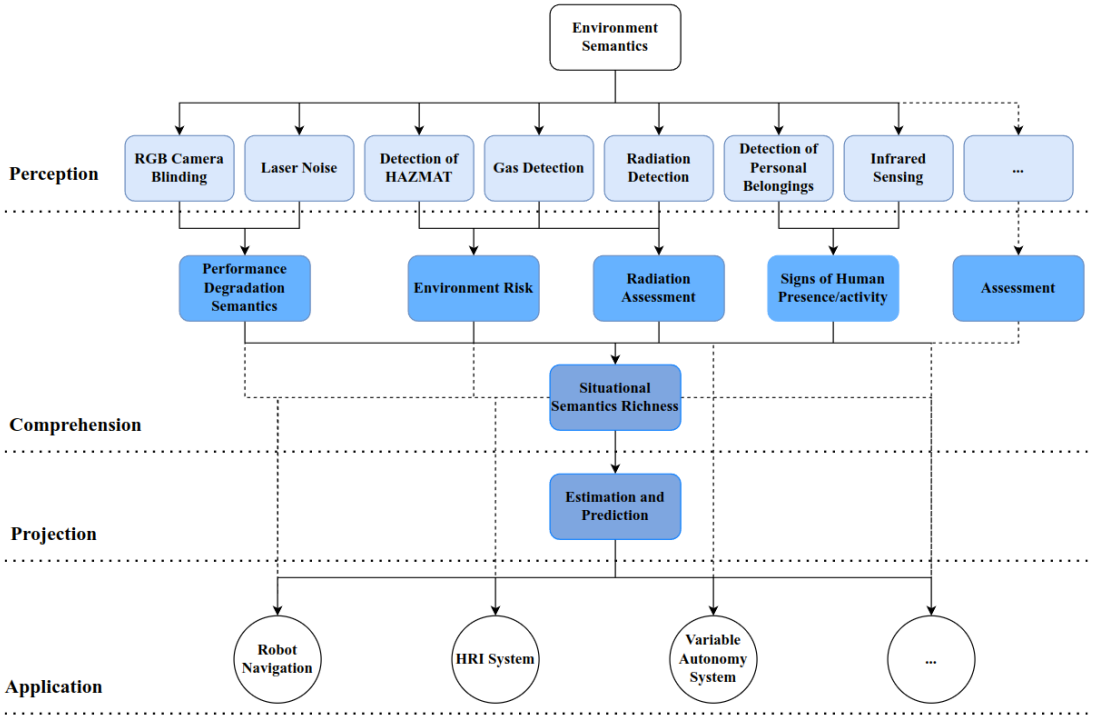


Figure 5.2: Semantics-based SA framework: light blue box refers to low-level semantics, cornflower blue box refers to high-level semantics, and dark blue box refers to context. Black dashed lines indicate the potential connections among different levels.

human understanding into our framework. It is common practice to build a heuristic-based system that comprises important factors and expert knowledge to obtain SA. It is straightforward to adjust. The SSR is proposed based on this idea. We obtain the score of each environment semantic indicator and adapt it into the proposed framework by developing the SSR score, which expresses the overall intensity and plethora of semantic information from the environment.

We obtain a set of normalized metrics S from 0 to 1 in Section 5.1. In different scenarios, different semantics might have different importance. To address this, we assign an importance weight W_i for each S_i. Tuning W_i gives the framework extensibility to different deployment cases and tasks while enabling leveraging expert knowledge. In our experiment, we adopt the exponential weight $W_i = e^{\frac{S_i}{0.99}}$ to emphasize the environment semantics with a higher score, where S_i is the score of each environment semantic, i is the label of environment semantics. The exponential weight lets high-score semantics impact more in the final SSR score. Then, we define the situational semantics richness R as:

$$R = \sum_{i=1}^{n} W_i \times S_i, \tag{5.7}$$

where n denotes the number of environment semantics. We normalize the R using a sigmoid function to obtain R_{norm}, which can enable a better understanding of the SSR intensity by humans:

$$R_{norm} = \frac{1}{1 + e^{-aR-b}}, \tag{5.8}$$

where a and b can be set as 10 and -0.5 respectively, to fit the range from (0,1).

To address the effects of unreliable scores caused by noise or false detection, we process the normalized SSR score by involving historical data. We apply the attention mechanism regression to comprise the past SSR scores and emphasize the impact of the latest score. The attention mechanism was first proposed by (Nadaraya, 1964) and (Watson, 1964), and has been widely used for non-parametric estimation and deep learning (Vaswani et al., 2017). It runs like the human's attention to indicate which value or factor deserves more focus among the rest of the data.

In our case, we can obtain a set of R in time sequence $\mathcal{R}_t = \{R_1, R_2, ..., R_t\}, t > 0$, t refers to the current timestamp. Then, the estimated situational semantics richness $\hat{R}_t$ at time t can be defined as:

$$\hat{R}_t = \sum_{i=1}^{n} \frac{K(t - t_i)}{\sum_{j=1}^{n} K(t - t_j)} \times \mathcal{R}_t, \tag{5.9}$$

where K is the kernel function, t_i is the timestamp of different R_t. Hence, $t - t_i$ refers to their individual time gap to the current time. If we apply the Gaussian kernel, which is mostly used

in Eq. (5.10), the estimated situational semantics richness $\hat{R}_t$ is:

$$
\begin{aligned}
K(u) &= \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}), \\
\hat{R}_t &= \sum_{i=1}^{n} softmax(-\frac{1}{2}(t-t_i)^2) \times \mathcal{R}_t,
\end{aligned}
\tag{5.10}
$$

where u refers to $t - t_i$ in this case.

In the experiment, we selected the time window of the five latest R_t, which means $n = 5$. According to the features of the applied attention mechanism, the older of the sample, the less impact on the final score. Hence, the 5 latest samples are enough to refer to. Considering the sampling rate limitation from the radiation sensor (1 Hz), the updating rate of the SSR score is synchronized to 1 Hz. Hence, the tuned SSR $\hat{R}_t$ refers to the R_t in five seconds, and we apply the $\hat{R}_t$ as the final SSR score. The time complexity of the whole process is $O(n)$, and the space complexity is $O(n)$ as well, which means that it is an efficient algorithm in the scope of computation.

5.3 Experimental validation

We test our framework intending to evaluate: i) if the framework can accurately perceive each semantics and their changes separately, and ii) if the framework is robust and can adapt in an environment with multiple levels and types of semantic indicators.

We use a Jackal mobile robot with an Intel I5 CPU and a GTX 1650 Ti GPU onboard. The framework is built based on the Robot Operating System (ROS) Noetic system. We run the framework directly on the Jackal to avoid image transferring to the offsite computer. Additionally, sensors, including a real-sense D435i camera, a Velodyne vlp-16 LiDAR, and a Hamamatsu Gamma Sensor C12137 are mounted on the Jackal. We apply the YOLACT (Bolya et al., 2019) as our vision model, providing object detection and instance segmentation results. We make some modifications to the system to enable us to attach depth data to each detected object by aligning the RGB image and depth image. The Hamamatsu Gamma Sensor C12137 is specif-

ically designed to measure gamma radiation in the range 0.03 to 2 MeV, and dose rate up to 100 $\mu Sv/h$. Even though the robot would be able to detect high-strength radiation sources from a distance, constrained by regulations from the university, we had to use a low-strength radiation source (uranium rock) that cannot be detected from long distances (over 10 centimeters). Stronger sources are not available in our project.

We assume the robot has a prior map from the SLAM, but SLAM is outside the scope of this chapter. The proposed environment features are placed in the environment after the mapping. Then, we predefine a set of waypoints that the robot has to navigate. The reason for applying autonomous navigation is that we would like to keep the similar trajectories of the robot in corresponding trials.

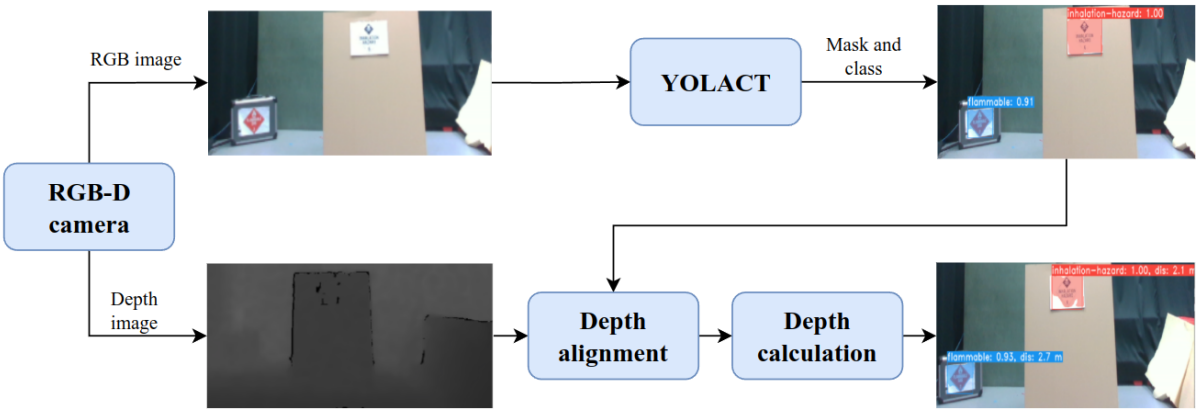


Figure 5.3: The structure of the vision system. The RGB-D camera could provide both colored and depth images. A color image is fed to YOLACT to generate a detected mask and a segmented instance. The depth image is aligned with the colored image to generate the estimated depth of the object.

5.3.1 Vision system design

Vision system design is collaborated with the previous internship student in our lab. Some of the work has been documented in (H. Wang, 2022). Here, we refer to some figures from his work to present the entire design.

The vision system is designed to detect low-level semantics, e.g. objects, and identify depth. Specifically, we train two neural networks separately for personal belongings detection and

HAZMAT sign detection. The methodology is similar to training both neural networks. Hence, we take HAZMAT sign detection as an example and introduce the process in detail.

The vision system comprises two modules for object detection and depth calculation. The structure is shown in Fig. 5.3. Specifically, we apply YOLACT for object detection and align the depth image with the mask and bounding box to generate low-level semantics (HAZMAT type, distance to the object).

Poison		Organic Peroxide	
Oxygen		Expolsive	
Flammable Gas		Radioactive	
Flammable Solid		Inhalation Hazard	
Corrosive		Spontaneously Combustible	
Dangerous		Infectious Substance	
Non-flammable Gas			

Figure 5.4: 13 classes of HAZAMATs signs (Sharifi, Zibaei, and Rezaei, 2020).

Training

Given the objectives of the project, it is not necessary to obtain the best detection accuracy, which requires a large amount of resources e.g. big dataset. Hence, transfer learning (F. Zhuang et al., 2020) is used to avoid collecting huge numbers of images and training the network from scratch.

For HAZMAT detection, we apply the public dataset [1] that includes 500 images with 13 classes of HAZMAT signs (See Fig. 5.4). Open-source software Labelme [2] is used to label images with the exact shape and classes. All the annotated images are converted into the MS COCO (Lin et al., 2014) format for training.

[1] https://github.com/mrl-amrl/DeepHAZMAT
[2] https://github.com/wkentaro/labelme

In transfer learning workflows for instance segmentation, ResNet (K. He, X. Zhang, et al., 2016) is often used as a backbone. To gauge the impact of backbone depth and architecture on detection accuracy, we test with three pre-trained networks: ResNet-50, ResNet-101 (the numerals indicate layer counts), and DarkNet (Redmon, 2016). The pre-defined loss function is suggested by (W. Liu et al., 2016) and (Bolya et al., 2019) as shown in Eq. (5.11):

$$L_{total} = 1 \times L_{class} + 1.5 \times L_{box} + 6.125 \times L_{mask}, \tag{5.11}$$

where L_{mask} refers to mask cross-entropy loss, L_{box} refers to bounding box loss, and L_{class} refers classification loss. Loss functions are used to determine how close the predicted value is to the label value specifically. When the loss is close to 0, the prediction is more accurate. When predictions show errors, the loss will boost.

The mAP and Intersection of Union (IoU) are used to assess the overall accuracy. The mAP is defined as :

$$mAP = \sum_{i=1}^{N} \frac{AP(i)}{N}, \tag{5.12}$$

where N is the number of classes, and $AP(i)$ denotes the average precision of i_{th} class (Padilla, Netto, and Da Silva, 2020).

And IoU refers to the degree of overlap between the predicted area and the labeled area. For instance, IoU50 means that when the overlap is over 50% of the total area, the prediction can be considered as a true positive. It is a threshold for assessing accuracy. Therefore, the IoU is defined as:

$$\text{IoU} = \frac{|\text{Prediction} \cap \text{GroundTruth}|}{|\text{Prediction} \cup \text{GroundTruth}|}. \tag{5.13}$$

The dataset was randomly divided into training and validation portions at an 8:2 ratio. During training, we evaluate the network on the validation subset every two epochs. Among three tested backbones, the ResNet-50 outperforms the rest two (See Table 5.1) and exhibits the expected learning behaviour (See Fig. 5.5): the horizontal axis denotes the number of training steps, while

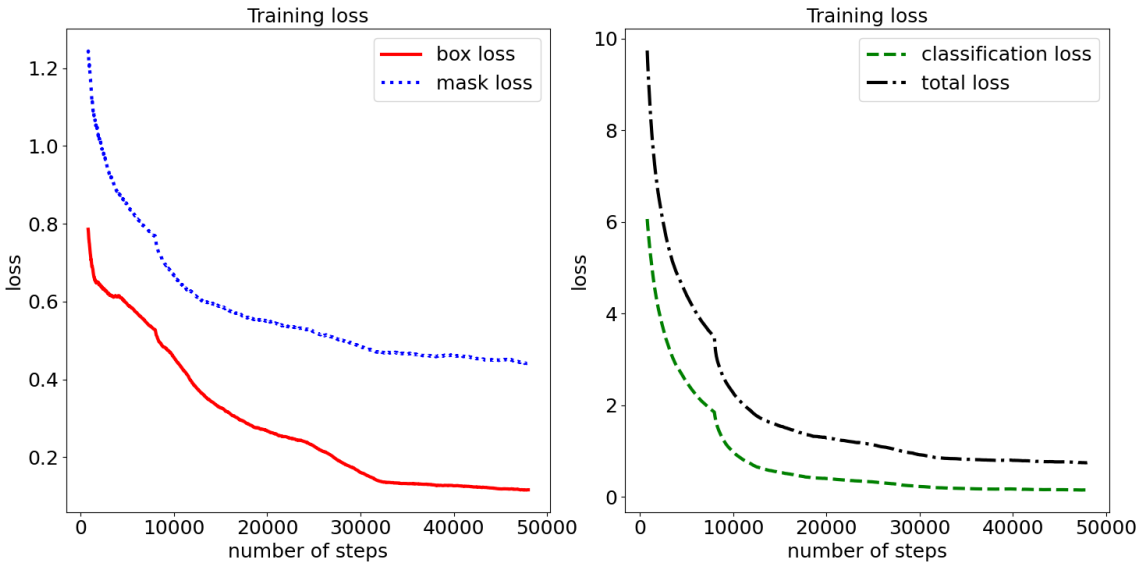


Figure 5.5: Training results when using ResNet-50 backbone. (H. Wang, 2022)

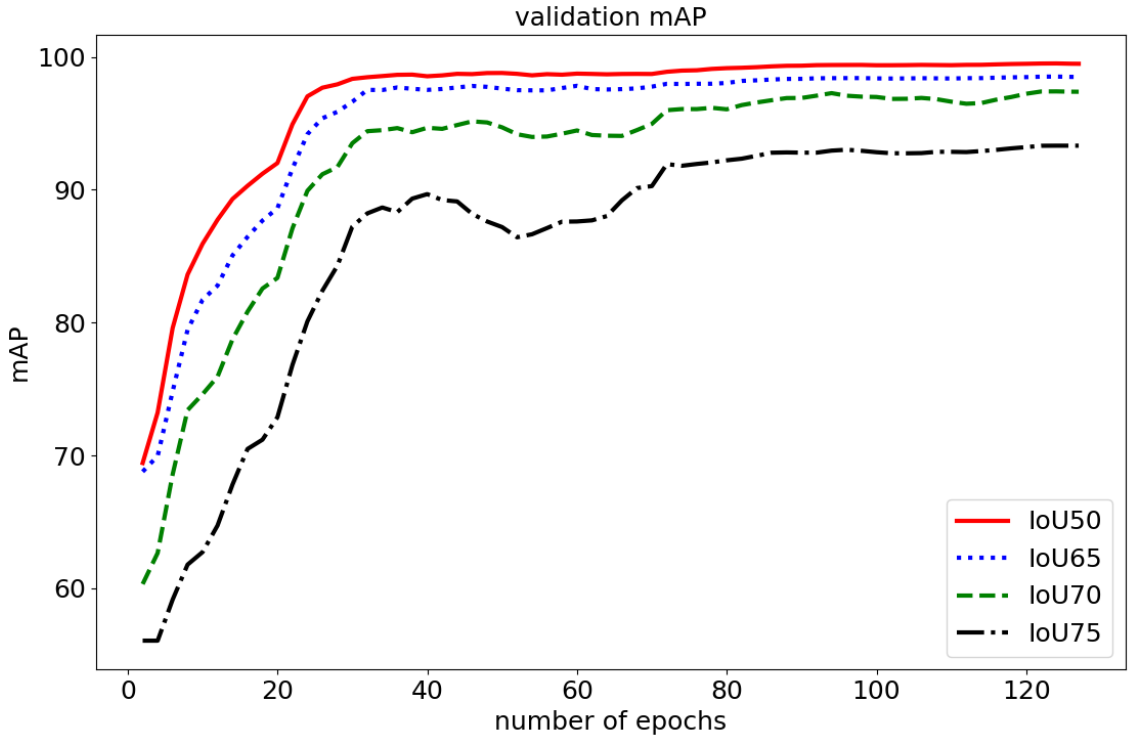


Figure 5.6: Validation results when using ResNet-50 backbone. (H. Wang, 2022)

the vertical axis plots the individual losses, all of which decline steadily as training progresses. Validation performance is summarised in Fig. 5.6. Here, the horizontal axis indicates epochs and the vertical axis reports mAP value. The four curves correspond to mAP computed at different IoU thresholds (Padilla, Netto, and Da Silva, 2020). Higher thresholds (e.g., IoU75) impose a stricter overlap requirement, therefore, their mAP values are consistently lower than those obtained with looser criteria (e.g., IoU50).

Table 5.1: Performance comparison among different training setups and backbones. (H. Wang, 2022)

		YOLACT-ResNet50		YOLACT-ResNet101		YOLACT-DarkNet		DeepHAZMAT
Hyper parameters	Learning rate	45	35	45	35	45	35	Non
	Batch size	4	8	4	8	4	8	Non
	Epoch	106	127	106	127	106	127	Non
mAP@IoU^{50}		**99.93%**	99.87%	98.87%	99.35%	99.04%	98.90%	99.03%
FPS		41.66	**42.33**	39.43	38.68	39.80	41.44	32

By adjusting the learning rate, batch size, the number of training epochs, and different backbones, we totally train eight models. Their performance was benchmarked against DeepHAZMAT, a SOTA solution for HAZMAT sign detection. The best network is accurate enough for our work, as reported in Table 5.1. The table indicates that the YOLACT-based network can reach a similar SOTA performance (99.93% accuracy on the validation set), and the FPS is around 8 frames higher than DeepHAZMAT, demonstrating that our network attains comparable accuracy with noticeably higher throughput.

Similarly, 627 images are taken and labeled in our lab for personal belongings detection (the detection network for SHA). 501 are used for training and 126 are used for validation (Samples see Fig. 5.7).

Depth fusion

With the detected masks and bounding boxes, distance information is required for each object. The RGB-D camera Intel RealSense D435i is used to obtain both RGB images and depth images. With the ROS driver from Intel, the camera can automatically align RGB images and depth

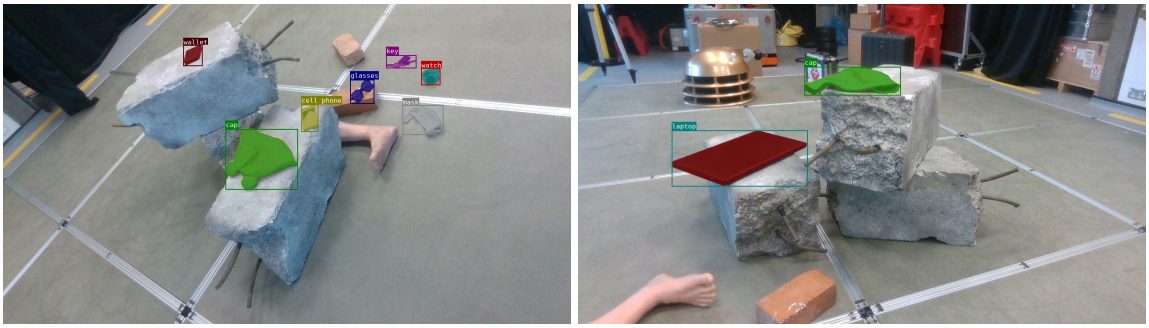


Figure 5.7: Samples of YOLACT personal belongings detection, e.g. hat, laptop, watch, mask, wallet, and others. Large objects can be detected from a long distance, e.g. backpack. Small objects like watches are detectable in short distances.

images under different resolutions.

We process the distance information with the segmentation results in the following steps:

- **Mask preparation:** YOLACT predicts the mask image for each object, and it is stored in a binary matrix. The size of the matrix is determined by the resolution of the camera. Each value in the matrix aligns with a pixel. If the pixel belongs to the detected object, it is marked as 1. If the pixel refers to the background, it is marked as 0. (See Fig. 5.8).

- **Crop:** The camera calibration enables the RGB images and the depth images to be aligned. Therefore, the location of the detected object in the mask image should be the same in the depth image. We crop out the matrix on the depth image based on the location information from the mask image.

- **Filter:** All depth pixels inside the object mask are first sorted in ascending order. We then estimate the object's range by taking the mean of only the deepest 10% of those values. It discards background outliers introduced by imperfect segmentation, giving a robust distance tag while adding virtually no computational overhead. Compared with averaging over the entire mask, the approach strikes a better balance between accuracy and speed and is intrinsically less sensitive to noise.

Apart from images from the validation set, we test some images around our lab and find that the vision system struggles to give a precise segmentation result when facing occlusion.



(a) raw image (b) mask image

Figure 5.8: Example of the raw image (a) and the mask image (b) generated by using YOLACT to detect common objects. (H. Wang, 2022)

However, addressing this potential problem of the vision system is not the focus of this chapter. We try to avoid the occlusions in the following experiments to overcome this issue.

5.4 Results Analysis

5.4.1 Experiment I

Implementation

We test the framework in the scenarios with single environment feature in Experiment I. We set two scenarios separately in the area (See Fig. 5.9 (a)). In each scenario, only one environment feature is added, i.e. each environment feature is independent, and only one feature with a big impact can be perceived at any time.

Specifically, we define three cases to differentiate the intensity of the environment semantics: LOW (radiation and noise), MEDIUM (radiation and SHA), and HIGH (risk and noise), which correspond to the levels of environment semantics. For instance, the HIGH case refers to risk and noise that can be detected at a high level; the MEDIUM case refers to radiation, and SHA detected is medium level. Due to the nature of the uranium rock, in the high-strength radiation scenario, the robot needs to move close to the radiation rock to mock the situations in which the sensor receives a high dose rate. And we apply longer distances to mock MEDIUM and LOW cases. We run the robots 10 times in three cases each.

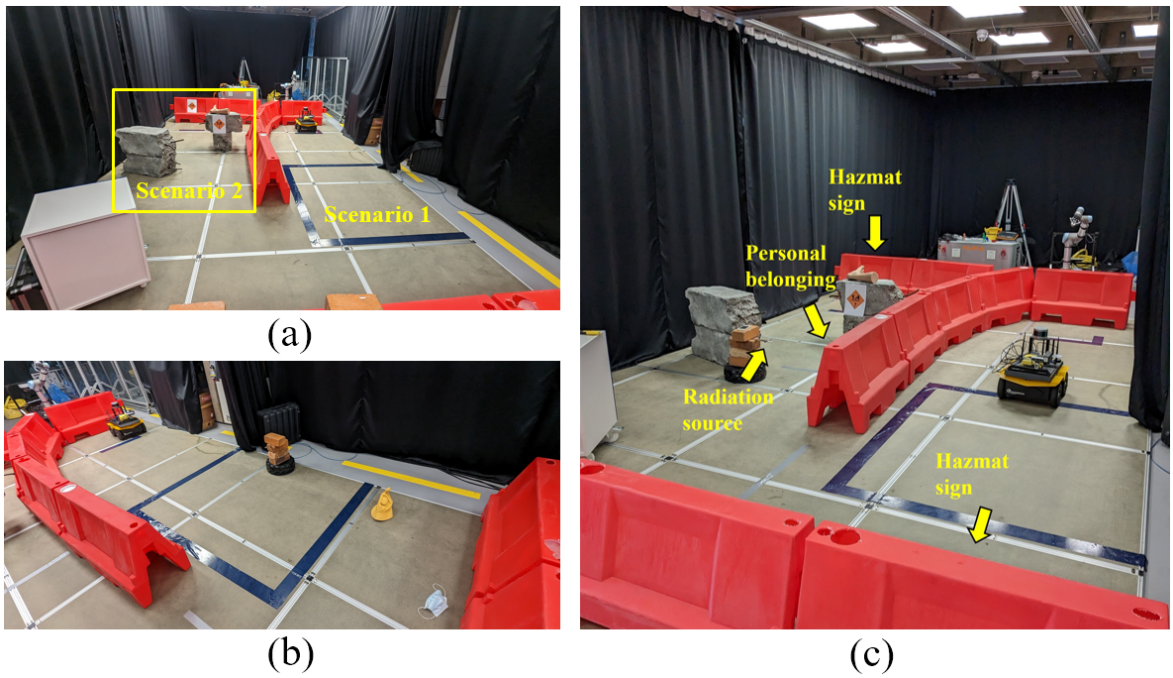


Figure 5.9: (a): Layout for Experiment I. The dark blue box area (scenario 1) on the ground is used for laser noise or the radiation source (uranium rock). The yellow box area (scenario 2) is used for hazmat signs or personal belongings; (b): Layout for two environment features scenarios of Experiment II; (c): Layout for three environment features scenarios of Experiment II. In the picture, some environment features are covered by red barriers.

Results

We collect data, including the scores of all environment semantic indicators and timestamps in seconds. To examine if the framework can differentiate environments with different levels of semantics, we generate the timeline of the SSR score and environment semantics in Fig. 5.10. Note that, in Fig. 5.10 (c), there is a deep flat between 60-70 seconds. It is caused by the occasion that the robot turns momentarily and faces the black curtain on the left. At that point, the camera, which is constrained by the view of the field, is unable to see the last HAZMAT sign on the red barrier until it moves forward and faces the sign. To the best of our knowledge, there is no prior baseline or dataset to compare with. The figures from the different cases reveal that the framework is capable of correctly outputting the corresponding SSR scores in the individual environment feature scenario. When the robot reaches the scenario, the environment semantic indicator responds in time, and the SSR score is more affected by the semantics with higher scores as designed, i.e. the SSR score tracks the environment semantics with the high-

Table 5.2: Sensitivity analysis ($S1$, St) of environment semantics with the SSR score in Experiment I

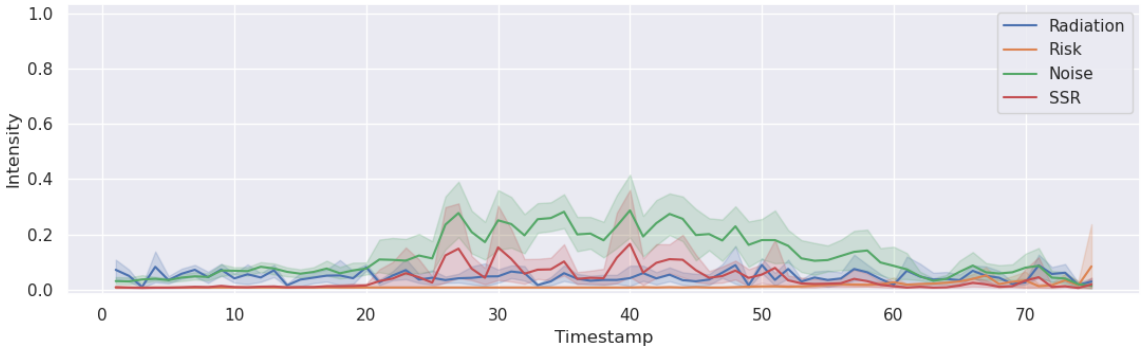
Case	Semantic Indicator	$S1$	St
MEDIUM	Radiation (red zone)	0.997	0.997
	SHA (blue zone)	0.977	0.978
HIGH	Noise (blue zone)	0.997	0.997
	Risk (red zone)	0.766	0.769

est score and magnitude. When zooming in specific environment feature zones, we notice that low-intensity semantics do not affect the SSR score much if medium or high-intensity semantics exist. We apply variance-based sensitivity analysis (Sobol, 2001) in the cases MEDIUM and HIGH (See Fig. 5.10 (b) (c)). Specifically, we calculate the global sensitivity (first-order index $S1$, total-effect index S_t) of each environment semantic indicator to the SSR score in Table 5.2. We skip analyzing the LOW case, as the LOW case is not that important in real deployments as long as no significant error has been found. The analysis shows similar $S1$ and S_t results. The individual semantics of the corresponding zones show strong sensitivity (over 0.7) to the SSR score. It indicates that the SSR score can respond quickly and accurately to single semantics changes in the environment. The medium or high intensity semantics dominate the impact and lead the changes in the SSR score. Our framework is tested to capture the changes in individual semantics and reflect correctly on the SSR score.

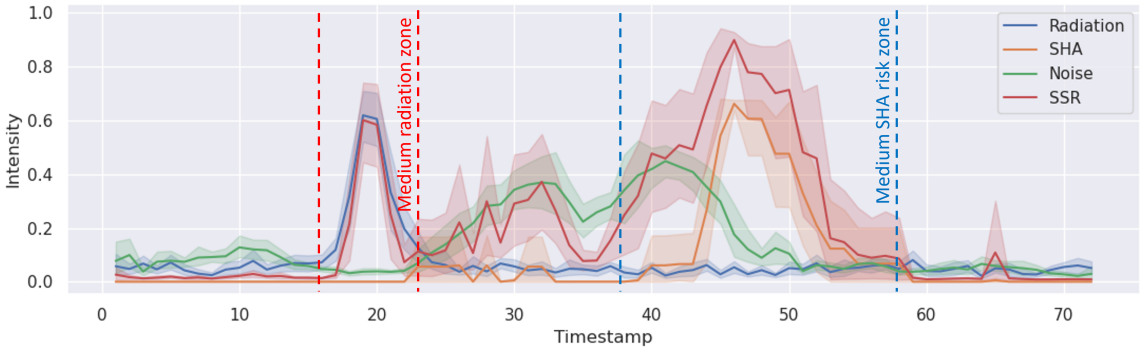
5.4.2 Experiment II

Implementation

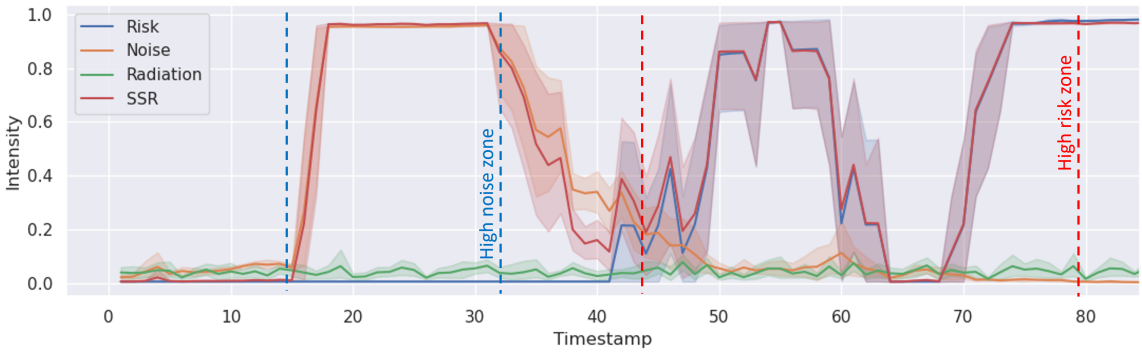
We test the framework's performance in scenarios with concurrent multi-environment features in Experiment II. We design 12 scenarios in this experiment (See Table 5.3). Specifically, we pick all the possible two-semantics combinations with high and medium level intensity and two three-semantics combinations. The scenarios cover a wide spectrum of situations that robots may encounter. We do not test with all of the semantics concurrently, as we are constrained by our vision system (train the HAZMAT and personal belonging detection separately). Additionally,



(a) SSR and environment semantics (radiation and noise) intensity timeline in LOW case.



(b) SSR and environment semantics (radiation and SHA) intensity timeline in MEDIUM case.



(c) SSR and environment semantics (risk and noise) intensity timeline in HIGH case.

Figure 5.10: SSR and environment semantics intensity timeline. The lines refer to the average SSR environment semantics intensity in 5 runs. The shade areas show the minimum and maximum range. Zones between dashed lines refer to the corresponding semantics detected from the environment.

Table 5.3: Combinations of environment semantics in different scenarios in Experiment II

Scenario	Radiation	Risk	SHA	Noise
1	High	/	/	High
2	Medium	/	/	Medium
3	High	High	/	/
4	Medium	Medium	/	/
5	/	High	/	High
6	/	Medium	/	Medium
7	/	/	High	High
8	/	/	Medium	Medium
9	High	/	High	/
10	Medium	/	Medium	/
11	High	/	High	High
12	Medium	Medium	/	Medium

the noise generation design (Ramesh, C. A. Braun, et al., 2023) has only two levels of intensity available. High refers to deliberately adding laser noise into the scenario. Medium refers to no artificial laser noise added (e.g. normal noise caused by turning the robot). Low refers to background noise. In the scenarios with two environment features, we use the corresponding area in Fig. 5.9 (b). The dark blue box on the ground is used for laser noise or the radiation source. HAZMAT signs are put on the right red barrier, and personal belongings are scattered in front of the right red barrier. In the scenarios with three environment features, we apply laser noise in the dark blue box, hazmat signs, and radiation source as Fig. 5.9 (c) shows, and personal belongings are scattered around the gray bricks. The robot runs 5 times in each scenario.

Results

We collect data on all environment semantic indicators, SSR scores, and timestamps in Experiment II. We align the data in each scenario with the timestamp to reduce the error caused by mismatching. The processed results are shown in Fig. 5.11, Fig. 5.12, Fig. 5.13, Fig. 5.14, Fig. 5.15, and Fig. 5.15. The shade zone reveals the range of semantic indicators and the SSR score. The solid lines refer to the mean from five trials.

Because of the limitation of the sensitivity analysis (unable to reflect the sensitivity in the complex system with multiple factors), we analyze individual environment semantics correla-

Table 5.4: Spearman's rank correlation coefficient (ρ) of environment semantics with the SSR score and bootstrap 95% Cl in Experiment II

Scenario	Radiation	Risk	SHA	Noise
1	0.75 [0.65, 0.82]	/	/	0.81 [0.74, 0.85]
2	0.86 [0.80, 0.91]	/	/	0.22 [0.07, 0.37]
3	0.67 [0.57, 0.75]	0.80 [0.77, 0.82]	/	/
4	0.56 [0.45, 0.67]	0.88 [0.85, 0.90]	/	/
5	/	0.73 [0.68, 0.75]	/	0.67 [0.56, 0.75]
6	/	0.53 [0.41, 0.62]	/	0.76 [0.72, 0.79]
7	/	/	0.83 [0.77, 0.86]	0.79 [0.75, 0.82]
8	/	/	0.78 [0.71, 0.84]	0.46 [0.35, 0.56]
9	0.68 [0.58, 0.76]	/	0.87 [0.83, 0.89]	/
10	0.67 [0.56, 0.75]	/	0.79 [0.74, 0.84]	/
11	0.53 [0.47, 0.60]	/	0.44 [0.36, 0.54]	0.91 [0.88, 0.92]
12	0.42 [0.34, 0.50]	0.63 [0.57, 0.68]	/	0.50 [0.40, 0.53]

tions to the SSR score. Our framework is nonlinear. Hence, we calculate Spearman's rank correlation coefficient (ρ) in Table 5.4 and Kendall's rank correlation coefficient (τ) in Table 5.5 to examine the relations among them.

In Table 5.4, Spearman's rank correlation coefficient demonstrates that in most scenarios, environment semantic indicators show at least a "weak" correlation (0.1 - 0.39) to the SSR score. In most scenarios with high-intensity semantics, the correlation index is above 0.4 and can be considered a "moderate" (0.4 - 0.69). If the semantics have high intensity over a long time, tables show "strong" (0.7 - 0.89) correlation (Schober, Boer, and Schwarte, 2018). The table indicates that different semantics show at least moderate impacts on the SSR score, i.e. the SSR score can reflect the changes in multiple semantics changes accordingly. If we connect to Experiment I, we find the results are consistent in the scope of responding to the situation changes correctly. It means our framework shows the generality ability when adding or removing semantic indicators.

Kendall's rank correlation coefficient (Chok, 2010) is robust to outliers. Similarly, most results in Table 5.5 align the monotonic relationship (at least moderate positive correlation) shown in Table 5.4, which is expected. However, Scenario 2 shows a "weak positive correlation" with noise. When we check Fig. 5.11, the weak correlation is reasonable. We did not deploy artificial noise in that scenario. Hence, the noise score is mainly affected by the robot's movement, which

Table 5.5: Kendall's rank correlation coefficient (τ) of environment semantics with the SSR score and bootstrap 95% Cl in Experiment II

Scenario	Radiation	Risk	SHA	Noise
1	0.62 [0.54, 0.69]	/	/	0.60 [0.53, 0.65]
2	0.70 [0.64, 0.76]	/	/	0.18 [0.08, 0.28]
3	0.52 [0.45, 0.60]	0.57 [0.53, 0.61]	/	/
4	0.43 [0.34, 0.51]	0.73 [0.69, 0.77]	/	/
5	/	0.53 [0.50, 0.55]	/	0.51 [0.43, 0.58]
6	/	0.42 [0.36, 0.52]	/	0.52 [0.48, 0.57]
7	/	/	0.67 [0.62, 0.71]	0.57 [0.52, 0.61]
8	/	/	0.63 [0.57, 0.68]	0.32 [0.24, 0.39]
9	0.55 [0.46, 0.61]	/	0.70 [0.65, 0.74]	/
10	0.52 [0.44, 0.59]	/	0.63 [0.57, 0.67]	/
11	0.39 [0.35, 0.42]	/	0.34 [0.26, 0.40]	0.74 [0.70, 0.77]
12	0.31 [0.25, 0.37]	0.52 [0.47, 0.56]	/	0.40 [0.29, 0.42]

stays at the bottom of the graph. It does not impact the changes in SSR score much as we expected. A similar situation happens in Scenario 8 and results in a weak correlation. In Scenario 11, there is no environment semantics showing a dominant impact on the SSR score from the Fig. 5.16. Hence, the correlation coefficients indicate weak or moderate correlations only.

Both Spearman's rank correlation coefficient and Kendall's rank correlation coefficient demonstrate that our framework reveals the situation changes and can adapt the complex scenarios with multiple semantics. These environment semantics show considerable impacts on the SSR score that enable SSR as a trustworthy representative for warning situation changes.

5.5 Discussion

SA is one of the prerequisites of prediction, planning, and decision-making. Our framework provides a way to obtain real-time SA that is intuitive and explainable to humans and easily usable for robots. From the scope of the experiments, the sensitivity analysis of Experiment I and the Spearman's and Kendal's rank coefficient of Experiment II, along with the analysis of the figures, reveal that our framework is sensitive enough for individual semantics situations and responds correctly in semantics-rich situations.

Compared with deep learning methods, our framework is designed to apply expert knowledge instead of data-driven training, which avoids the issue of lacking large annotated training datasets. It can be potentially explainable to a human, contrary to black box models. This enables humans to understand what exactly happens in the framework, and makes it more intuitive for experts tuning the framework. Human operators and robots can obtain shared SA not only from the SSR score but also from the changes in individual semantic indicators, helping them to identify the semantic characteristics of the situation. This is useful for real-world deployments in safety-critical applications.

Note that context and semantics are concepts that are arguably infinite in potential breadth and depth. Robots may not be able to understand all situations with a complex combination of semantics. The SSR metric can be used to make robots aware of whether they are in a semantics-rich situation, which might be beyond their capability to understand, and helps robots to decide whether they need help from human intelligence. Hence, both the semantic indicators and the aggregated metric SSR can be used in a prediction, planning, and decision-making framework, especially in the context of HRT.

Regarding the flexibility and generality of our proposed approach, Experiment II indicates that the framework is flexible enough to comprise multiple environment semantics. It will remain robust, easily explainable, and intuitive if additional environment semantics are added or removed. Depending on the applications and mission, experts can directly adjust parameters to generalize the framework into a more reasonable representation of the given scenario. For instance, experts can highlight the weights of those semantics that are most important to the goal of the mission, which makes the framework more sensitive and responsive to these semantics. The framework is not restricted to UGV deployments and can be adopted for different robotics platforms and sensors, such as UAVs, remote manipulators, or heterogeneous multi-robot teams.

This work have some limitations also. The vision system constrains the implementation of the framework. Considering the training process, it may be possible to fuse personal belonging detection and hazmat detection to simplify the deployment process. Alternatively, we could adapt other state-of-the-art perception algorithms to improve the accuracy and real-time per-

formance. However, our overall framework is scalable, and will continue to be useful as the semantics capabilities of AI and CV continue to grow more powerful over time.

5.6 Conclusion

In this chapter, we proposed a semantics-based SA framework to represent and quantify a variety of semantic information, as well as the overall situational richness, via the concepts of semantic indicators and the aggregated SSR metric. We also provided an example implementation on a real robot, to process high-level environment semantic indicators, and we showed how these correspond to the robot's environment. Semantic perception capabilities of AI are still in an early stage of development. For those reasons, we have chosen some relatively simple and robust examples in these experiments. However, the experiments demonstrate our framework is capable of obtaining SA and indicate its extensibility to semantic-rich environments, and potential to involve multiple types of environment semantics. The modularized design increases the flexibility of this framework, and it should adapt and scale well as these AI capabilities grow in the future.

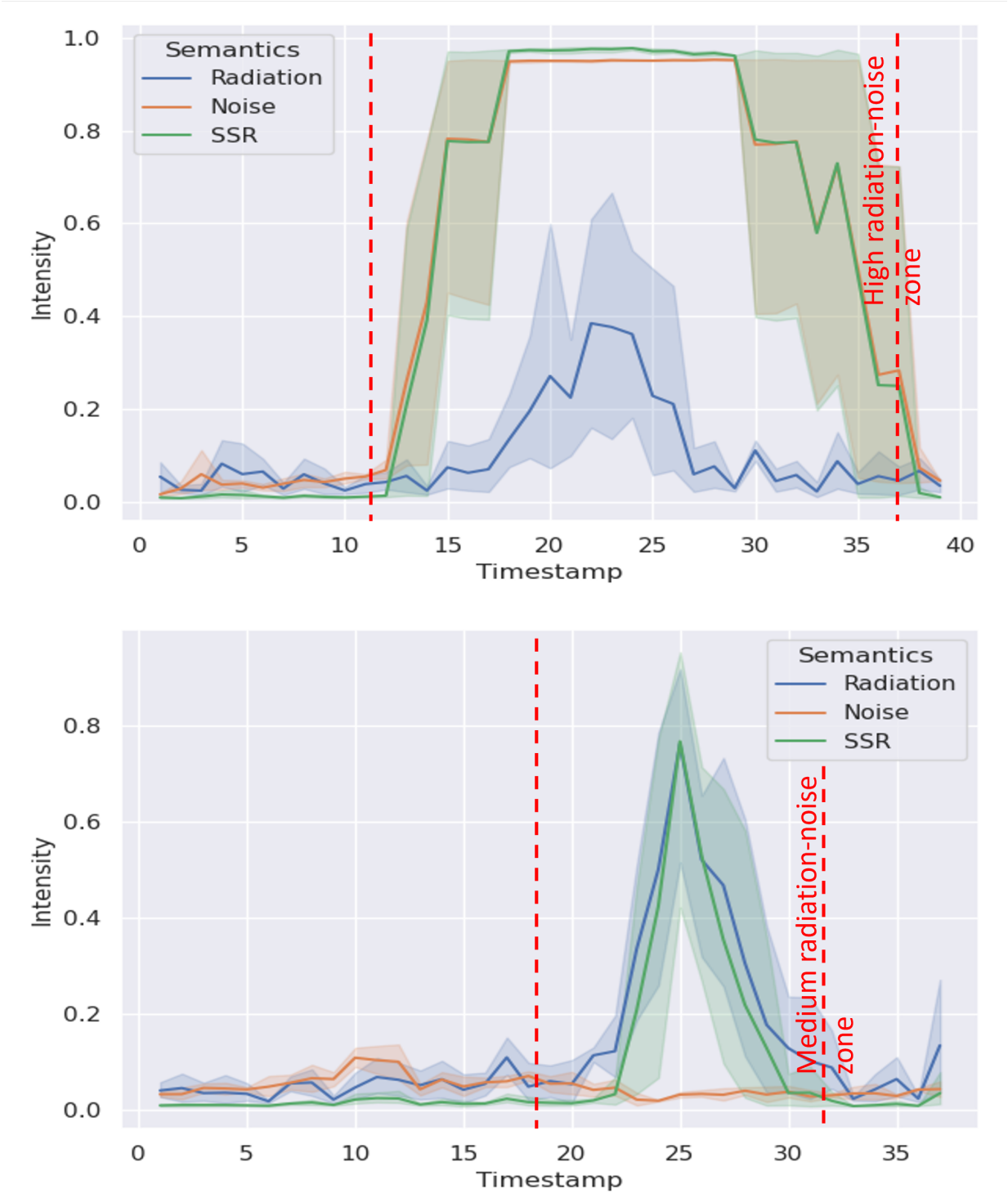
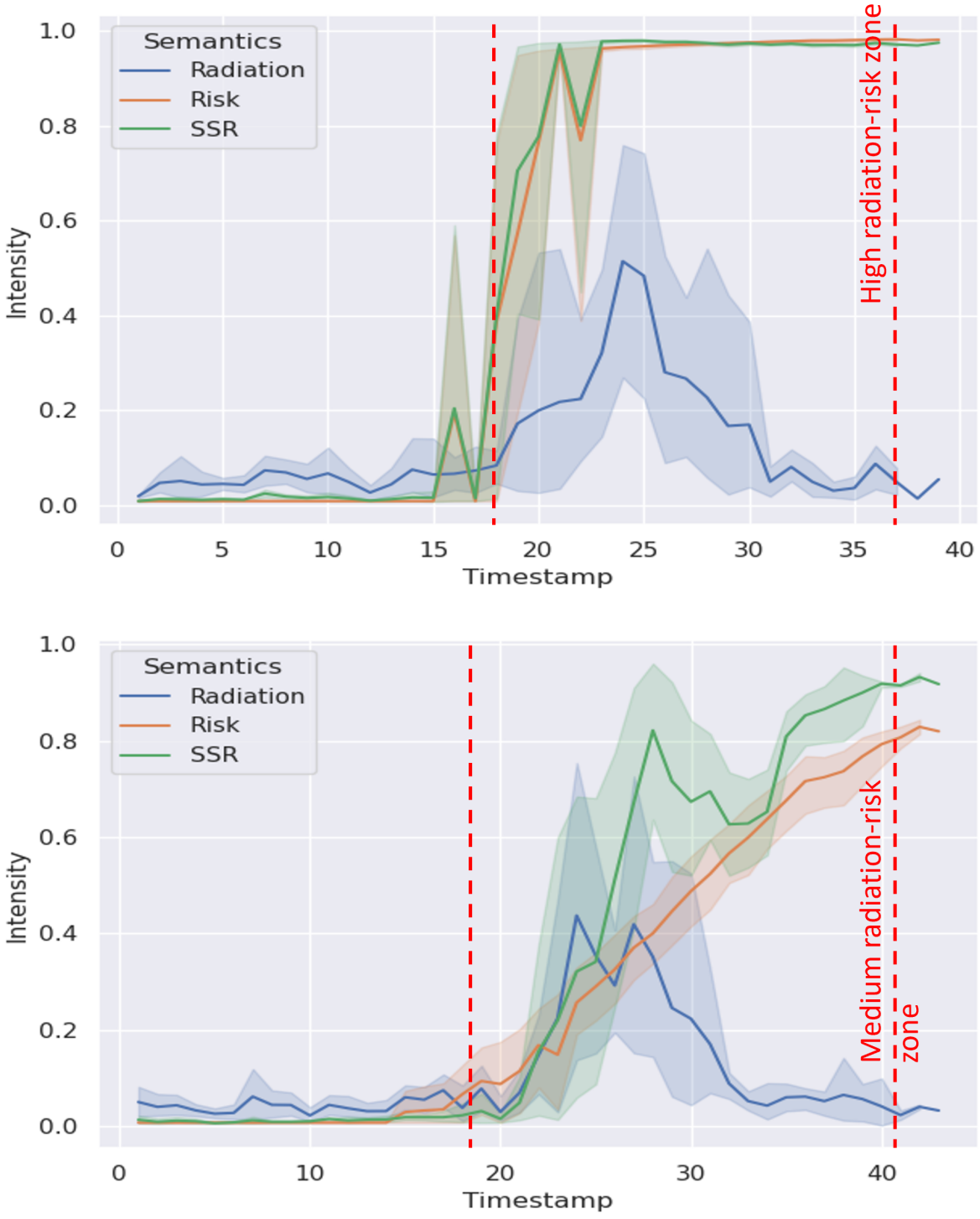


Figure 5.11: Radiation-noise combination semantics intensity timeline. Zones between dashed lines indicate detected semantic signals.



(a) Radiation-risk combination

Figure 5.12: Radiation-risk combination semantics intensity timeline. Zones between dashed lines indicate detected semantic signals.

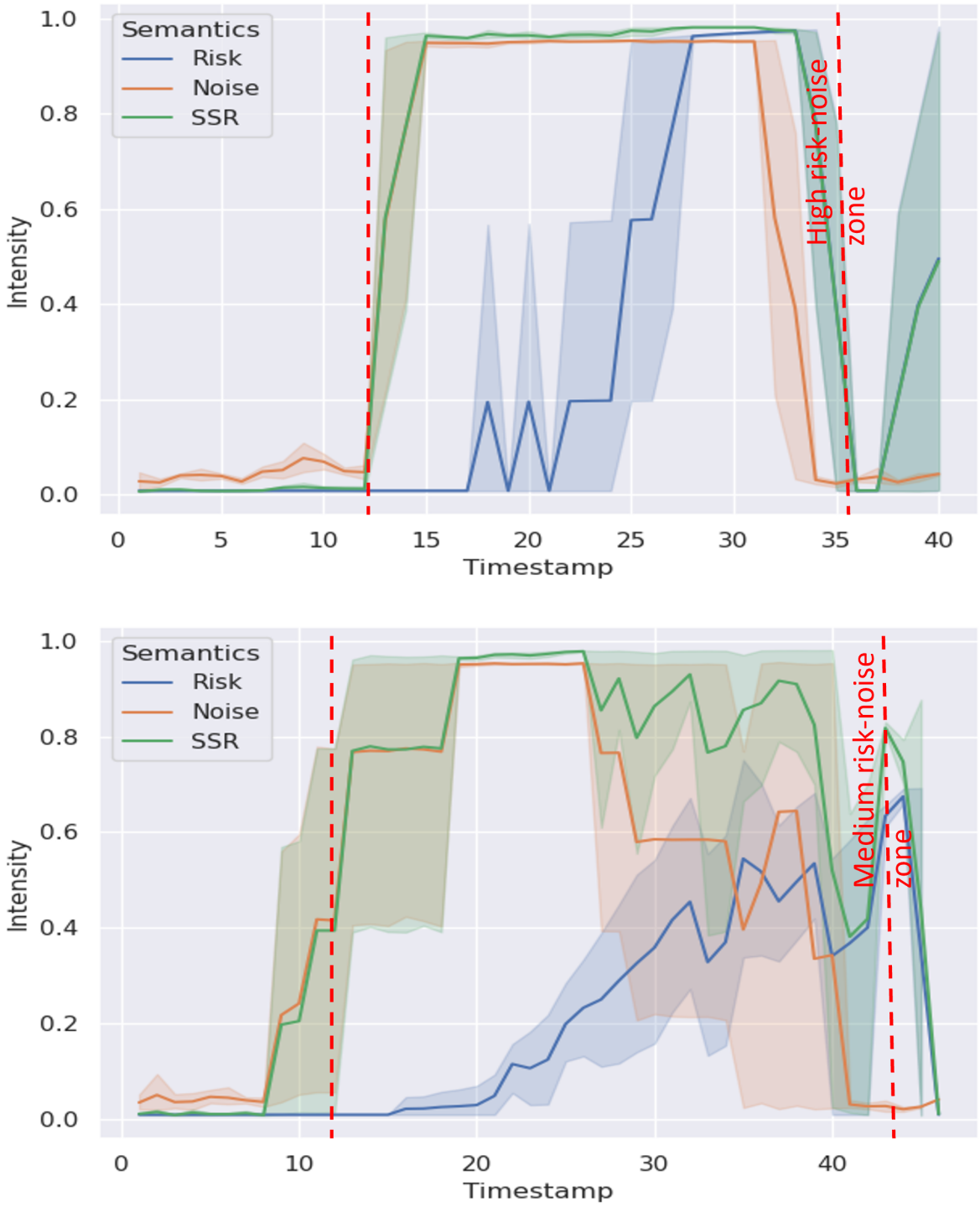


Figure 5.13: Risk-noise combination semantics intensity timeline. Zones between dashed lines indicate detected semantic signals.

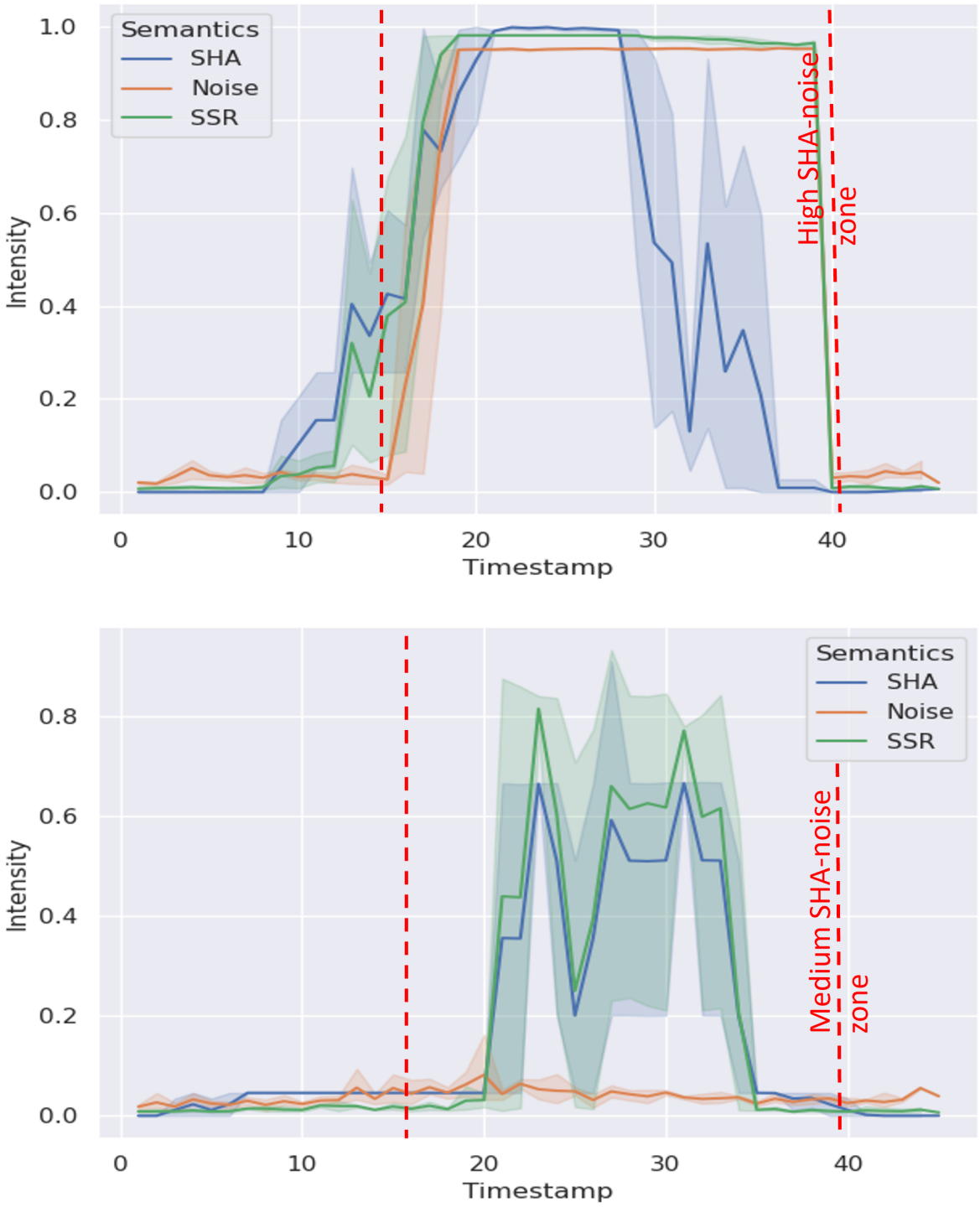


Figure 5.14: SHA-noise combination semantics intensity timeline. Zones between dashed lines indicate detected semantic signals.

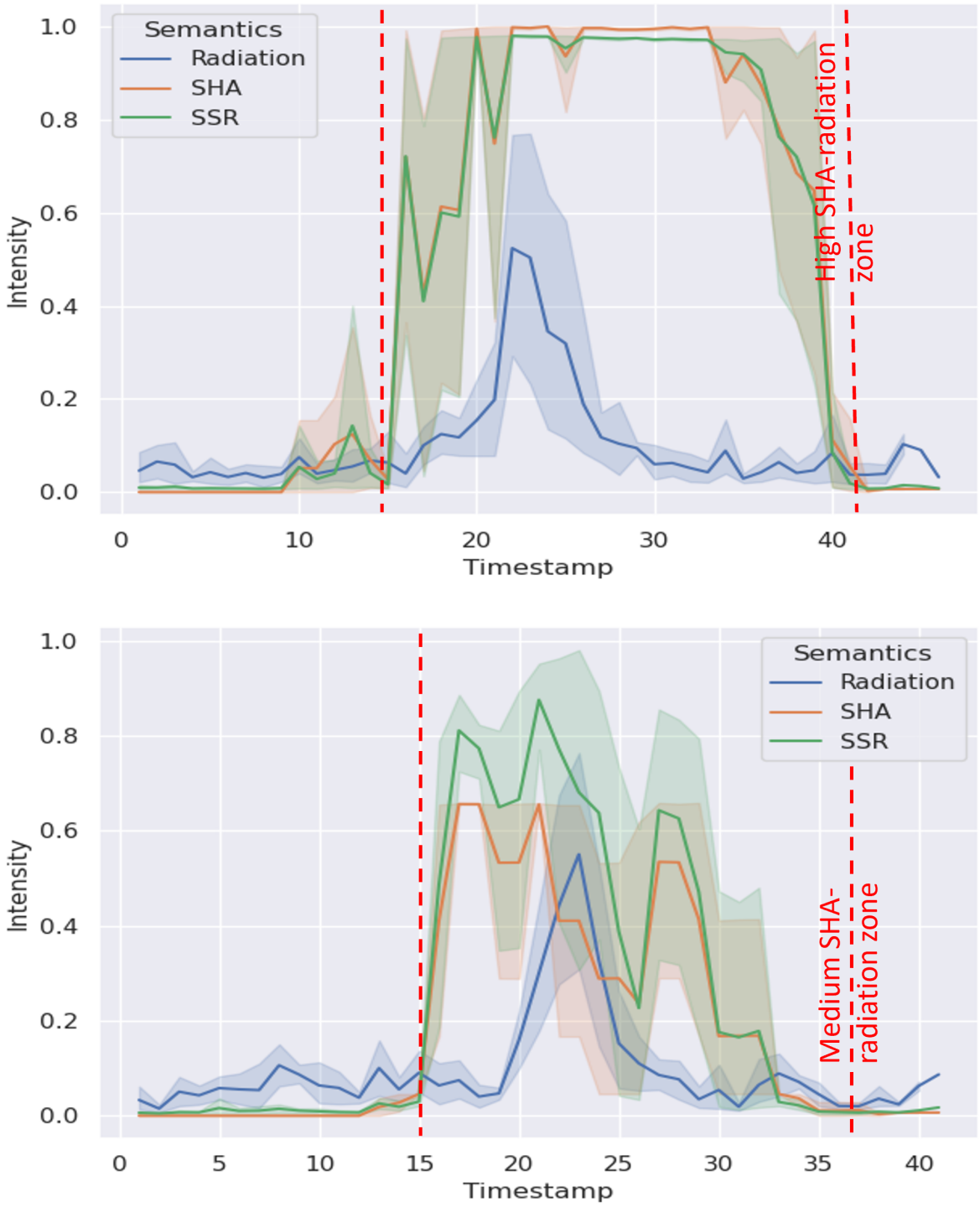


Figure 5.15: SHA-radiation combination semantics intensity timeline. Zones between dashed lines indicate detected semantic signals.

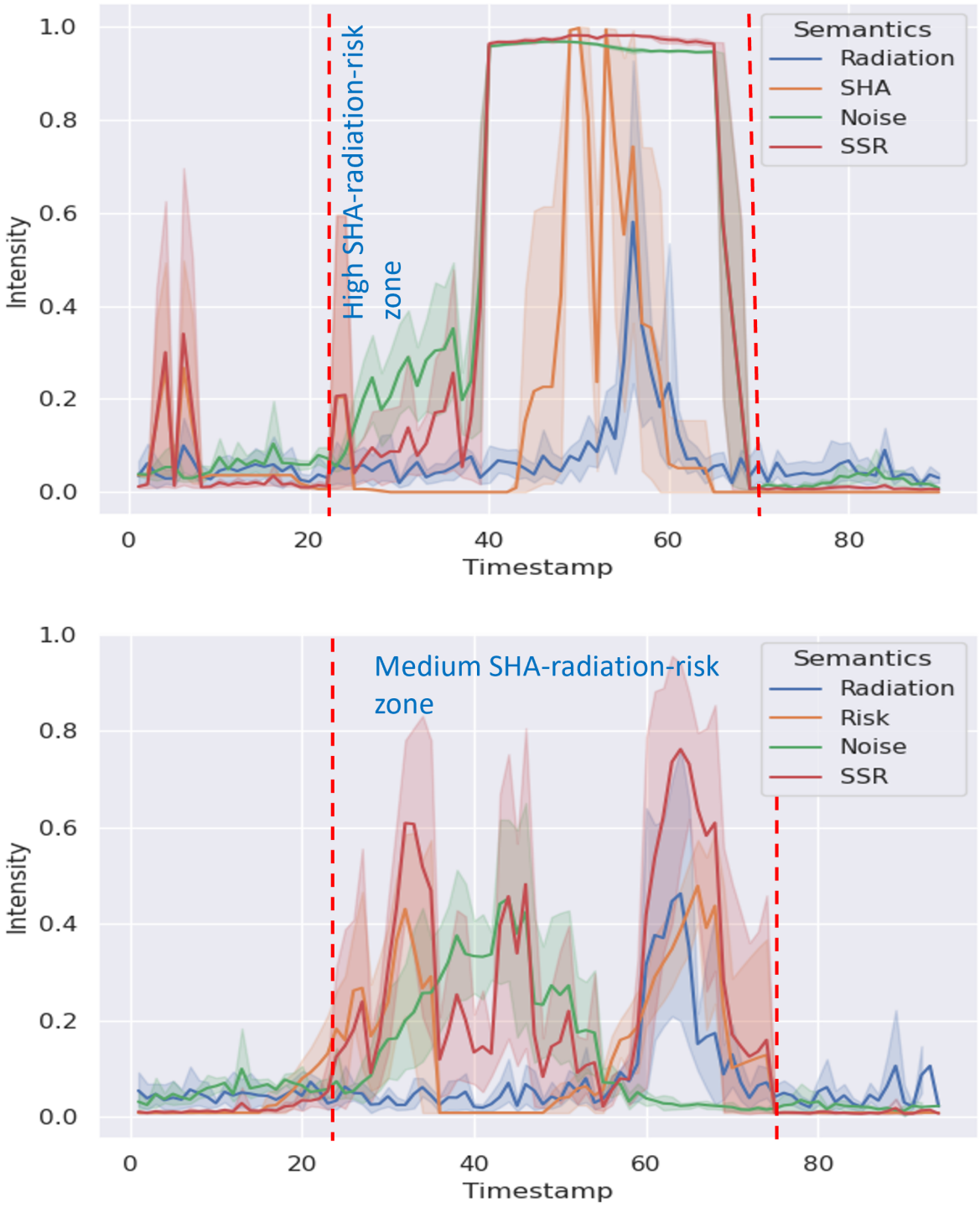


Figure 5.16: Three semantics combination intensity timeline. Zones between dashed lines indicate detected semantic signals.

Chapter 6

Impact of High-level Semantics on Variable Autonomy Human-Robot Teaming

At the time of writing this thesis, the work in this chapter has been accepted by the 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (Ruan, Ramesh, Stolkin, et al., 2025).

The proposed semantics-based SA framework has been verified for its feasibility and generalization in the lab environment in Chapter 5. However, its effectiveness as a tool to help the HRT, and the consequent impacts of semantics on HRT remain unexplored. In Chapter 6, we address this by carrying out human participant experiments, in which the semantics-based SA framework is incorporated into a VA HRT. We summarize some insights about how semantics affect human trust in SA, response time, and cognitive workload by conducting a simulated SAR mission.

The increasing use of robots is engendering challenges and higher demands for HRC. HRT needs humans and robots to work together and leverage their complementary strengths, to achieve the shared mission goal by combining perception capabilities (remote sensing, object detection) and cognition (e.g. scene understanding, decision-making, task planning). SA is one of the sig-

nificant factors in bridging cognition and perception. Explainable and comprehensive SA can directly benefit robots and humans in many ways, e.g. task planning (Fraune et al., 2021), VA (Wagner, Borenstein, and A. Howard, 2018), and HRI.

However, obtaining SA remains a challenging problem in HRT. Robots are often considered extensions of humans to perceive a remote environment using sensors through raw or low-level perception. Humans typically struggle to develop precise SA remotely, compared with the SA that we obtain when physically present in a scene (Mutzenich et al., 2021). The problem is worse when humans shift their attention back to the robot from other tasks, e.g. when humans are out of the context/loop for a while. It takes much more effort and attention to restore the awareness of what the robot is doing and what the situation is like (Hergeth et al., 2016), especially when humans have a limited background in the mission (Stanton et al., 2017). Moreover, SA often relies on human operators' prior experience and knowledge, which means that there is no comprehensive standard to quantify SA.

Semantic information, which is one of the components of SA, is suggested to alleviate the above difficulties (Roldán et al., 2017). Semantic information is widely discussed in the artificial intelligence community, e.g. in CV. However, this literature is predominantly limited to relatively low-level semantics (e.g. image segmentation). The impact of high-level semantics, which requires more understanding and inference, has been less well-explored. In the context of this work (except where specifically stated), we use the term "semantics" to refer to high-level semantic understanding. This can include a variety of indicators for environment understanding (Ruan, H. Wang, et al., 2022), e.g. concepts such as risk to robots, or SHA.

Our previous chapters (Chapter 3) categorize environmental semantics in the scope of disaster response missions and propose a framework (Chapter 5) that applies the above categorization and quantifies different understandings of the situation by semantic indicators. Moreover, the framework enables robots to provide an indicator of the overall semantic richness of a scene. This method translates sensor readings, measurements, and detections into a semantics richness score, which can inform remote human operators if the situation requires more attention.

In this chapter, we explore several high-level semantic indicators from the above framework

and investigate their impact on HRT and HRI. We compare robot deployments in different experiment setup, with and without such semantic indicators (high-level semantics), in the scope of an SAR mission.

This chapter provides new insights into HRT and HRI with respect to high-level semantics. We present three main contributions: **i)** identifying the usefulness of the proposed semantics-based SA framework (Ruan, Ramesh, H. Wang, et al., 2025); **ii)** identifying how high-level semantics affect HRT and HRI in the context of a SAR mission; **iii)** insights about patterns and relationships among HRT factors and high-level semantics, e.g. LoA preference versus trust in a VA system (Chiou, McCabe, et al., 2021).

6.1 Aims and Hypotheses

This chapter explores the effects of presenting high-level semantics to a human operator during HRT deployments. Specifically, we explore the performance variance of HRT, with and without the display of semantic indicators (high-level semantics). We also explore the practical utility of our previously proposed framework for real-time computing of such indicators in Chapter 5.

Our semantic richness framework in Chapter 5 quantifies the richness of low-level semantics and high-level semantics (Ruan, H. Wang, et al., 2022) in the current scene. These high-level semantic indicators can be used to help a human operator achieve SA. In our experiments, high-level semantics include: SHA; radiation levels, and navigational laser noise. SHA is detected as a distribution of personal belongings. Radiation level is a mapping from radiation sensor readings that enables human operators to visualise radiation intensity. The laser noise refers to the robot's LiDAR. It helps robots and humans understand whether noise is becoming significant enough to disable the robot's navigation. Each of these metrics carries a normalized score between 0 and 1, helping assess potential risks or challenges in real time.

In this exploratory study, we use a simulated SAR mission with human participants. Specifically, we simulate the first round of HRT deployment in a first phase disaster response mission. The HRT needs to explore the environment and find victims while avoiding risks to robots.

The robot can operate in two different LoAs: teleoperation and autonomous navigation modes. Humans can switch the LoA dynamically in a Human-Initiative paradigm (Chiou, Stolkin, Bieksaite, Hawes, K. Shapiro, et al., 2016) and can take over with teleoperation control when they feel it is necessary. Moreover, parallel cognitive tasks are assigned to human operators to simulate non-robotic emergencies that may require human attention. We have two major research questions in this experiment: i) how do high-level semantics affect the HRT and HRI, and ii) what is the correlation among HRT factors when affected by high-level semantics?

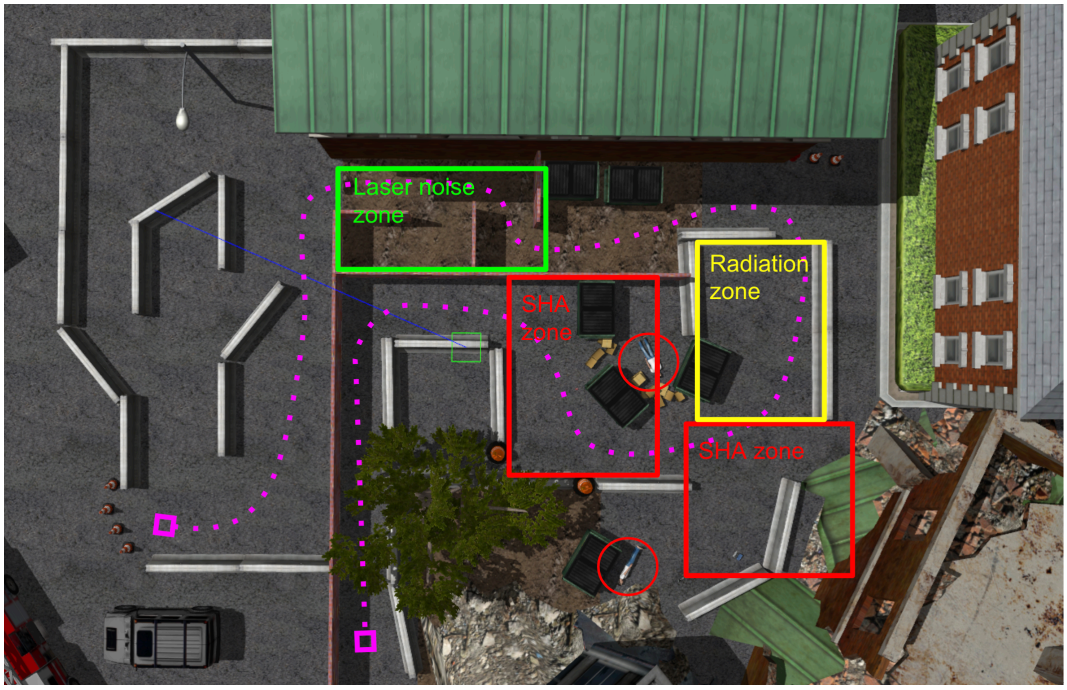


Figure 6.1: Test arena with semantic indicators. Colored boxes denote zones where different semantic situations will arise, resulting in different indicators being displayed to the human operator (trial participant). Red circles mark the location of the victims. The purple dotted line is the robot's pre-set autonomous navigation trajectory.

6.2 Experimental setup and apparatus

6.2.1 Arena

The experimental arena is an outdoor post-disaster environment (See Fig. 6.1 and Fig. 6.2). The robot is set to autonomously navigate along a pre-planned route, but the human operator can

switch to teleoperation mode at any time, and explore other regions away from this route (Chiou, Stolkin, Bieksaite, Hawes, K. L. Shapiro, et al., 2016). Two human victims, one radiation source, and one laser noise zone are set up in the arena. Specifically, one human victim can be found by the camera on the scheduled path (i.e. the robot does not need to deviate) by locating the personal belongings. Another victim is hidden behind obstacles and away from the scheduled path. In this case, the robot can barely see the victim unless it deviates from the path and explores the nearby area. The radiation source and laser noise zones are set to occur at pre-set regions along the planned path. The starting point and endpoint for robots are pre-defined. Note that we use a "within groups" experimental design, where each participant undertakes both experimental conditions (with and without semantic indicators). Therefore, we need to use a different test arena layout for each condition, to minimise learning effects as a confounding factor. However, the length and complexity of the tasks, and the types of semantics encountered, are kept very similar to ensure meaningful comparability.

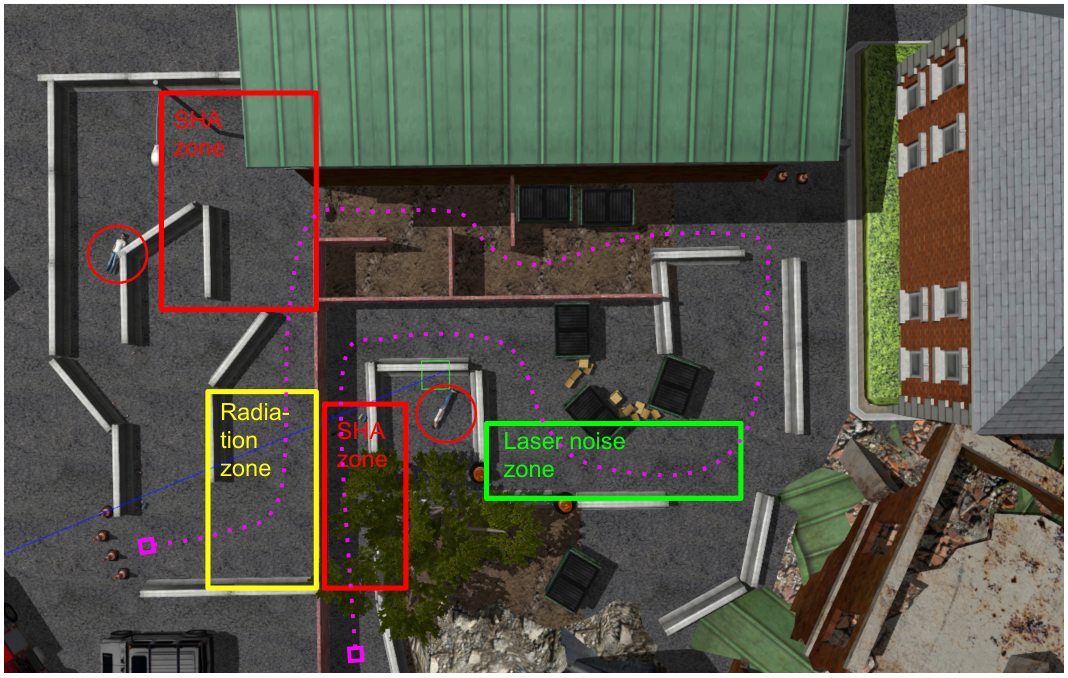


Figure 6.2: Test arena for trials without semantic indicators. Note that we need to use two different arena layouts for the two different trials, to minimise the confounding factor of learning effects between trials.

6.2.2 Mission and Tasks

We assume that the HRT needs to conduct its first deployment to investigate the site in a SAR mission. The participant can switch back and forth between the two LoAs (teleoperation and autonomy) whenever they choose. In autonomy mode, the robot will try to follow the pre-set route. Three tasks are assigned to the HRT: **A)** finding victims, **B)** navigating the robot, and **C)** a parallel to A and B tasks. Specifically, task A requires the HRT to find as many victims as they can in the arena. Task B requires the HRT to navigate the robots from a pre-set starting point to the endpoint within set time limits. Task C is a non-robotics task, used to simulate additional types of cognitive loading. This additional cognitive loading would distract humans' attention from the robotics tasks. In this case, participants are encouraged to make use of the LoA switching functionality, and helps elucidate situations where the additional semantic information becomes useful. Participants are instructed to do their best on all tasks (equal importance). They conduct two trials comparing two conditions: semantics presented Fig. 6.1; semantics not presented Fig. 6.2. The sequence of two trials is counterbalanced among participants, to minimize the learning effects. Both trials share the same arena. However, victims' locations are different, and starting/end points are swapped (clockwise/anti-clockwise path) between participants in every two trials.

Participants have five minutes for each trial. If they fail to bring the robot back to the endpoint within the 5-minute limit, the mission will be marked as a failure. Participants are not informed of the number of victims. They need to maximize their utilization of the five minutes. Hence, a time-to-completion metrics is not meaningful. Instead, we examine success rates on each task.

The parallel task (task C) takes place continuously (an endless task) during the mission. Participants need to identify whether a pair of 3D objects is identical, or is a mirrored image pair, see Fig. 6.3. This task has been validated from the scope of mocking up human cognitive workload in (Ganis and Kievit, 2015). Standardized training (Chiou, Talha, and Stolkinl, 2019) was provided on this task before the experiments to ensure all participants were aware of the effort required to manage this alongside the two robotics tasks.

Participants' performances on this task (accuracy, completion, and completion speed), along

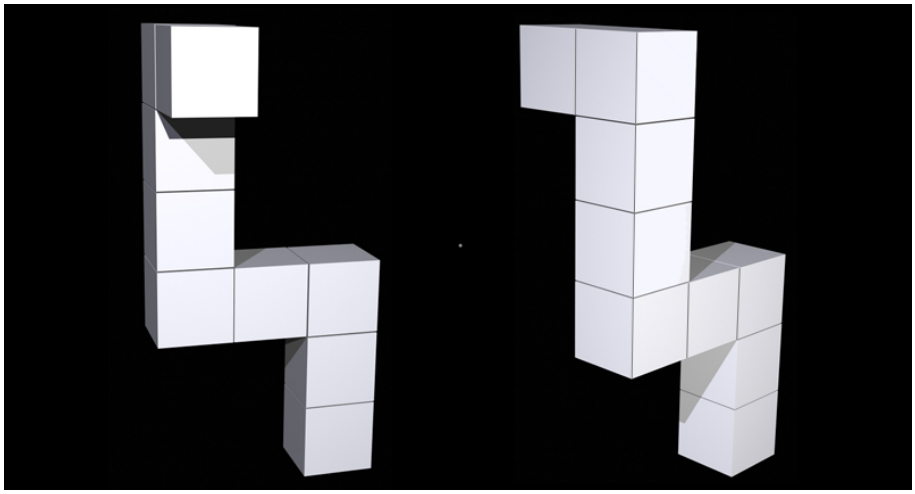


Figure 6.3: Parallel task example. Are these objects identical or mirror images?

with the robot-related tasks, are included in the overall evaluation. In addition, we recorded the baseline (accuracy of ten questions) of each participant before the test and compared the baseline to the parallel task accuracy during the experiments. By comparing baseline performance with the performance during the robot-assisted missions, we can evaluate the impact of high-level semantics on participants' parallel task execution quality.

6.3 Experimental design and procedure

A total of 16 participants (8 male / 8 female) were recruited from university staff and students aged 22-50, see Table 6.1. Most had a driver's license or experience with 3D video games that correlate with controlling remote vehicles (Chiou, Hawes, et al., 2015). Participants undertook standardised training to understand the system functionalities and control interface needed for the mission, including navigation in teleoperation and autonomy modes, robot's SLAM map, and collision avoidance. Participants were introduced to high-level semantics, including SHA, radiation, and laser noise. When a semantic indicator score increases, the system emits a warning sound to attract human attention. Participants were trained on how to react to situations, e.g. switch LoA forth and back when they felt necessary. Training is designed to avoid very large individual performance differences (e.g. the differences in skills and experience in driving robots), especially if participants become lost in the arena.

Table 6.1: Participants' gender and background

Gender	Participants	Background	Participants
Male	8/7	Driver only	5/5
Female	8/7	Gamer only	2/2
		Gamer + Driver	7/6
		Neither both	2/1

Red digits refer to the number of participants after data cleaning, which are involved in the future analysis.

To minimize the effects of potential confounding factors, we surveyed participants' background information Table 6.1 and their feedback on arena and mission tasks. We removed two participants' records from the database. One of them had a pre-existing bias toward the autonomy system, with no use of autonomy mode. She explained she had "zero trust in any robot system". Another participant failed to follow the instructions and training given, which led to incomplete data. The second digit in Table 6.1 refers to the number of participants after removal.

The GUIs of the two experimental conditions are presented in Fig. 6.4 and Fig. 6.5.

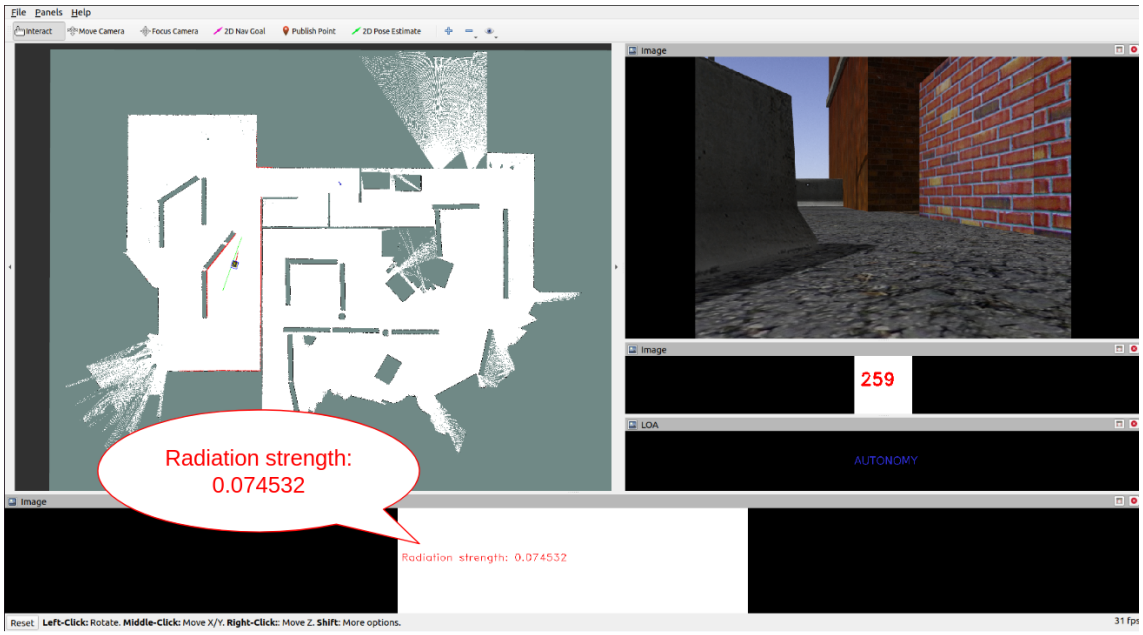


Figure 6.4: GUI for trials without semantic indicators: participants obtain SA from a pre-scanned LiDAR map, camera feed, real-time localisation, radiation readings, current robot LoA, and countdown of remaining mission time. The bubble is used for clearly presenting the semantic indicator scores in the figure. It does not show up in the GUI during the experiments.

The above training ensured that participants understood the system features and operation

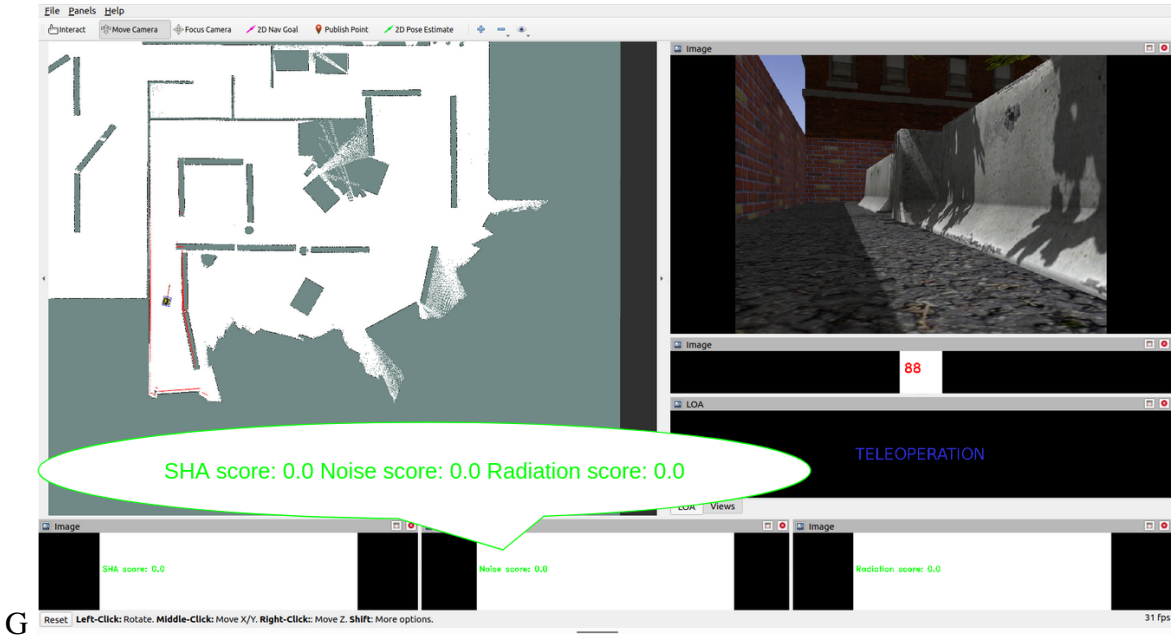


Figure 6.5: GUI for trials with semantic indicators: three semantic indicator scores are displayed on the interface screen, in addition to the mapping and video data.

modes, creating a foundation for effective decision-making and navigation. Moreover, training gave participants an understanding of the workloads, which helps them make decisions on how to utilise the VA and semantic indicator functionalities. To minimize the effects of various confounding factors, some additional information and tips were given to participants:

- Because of blind spots from the LiDAR, the autonomy mode might fail in some places. Participants may need to switch to teleoperation mode.
- The 5-minute time limitation is tight for most operators. Participants should carefully manage the time and areas to explore.
- Try to follow the scheduled path (the shortest navigation path) most of the time, and only deviate when participants feel it is necessary to achieve tasks.
- The three semantic indicators will not appear at the same time.
- When a situation changes (e.g. a SHA appears), participants are recommended to switch to teleoperation mode to check the real situation in the training.

6.4 Evaluation with a high-fidelity simulated robot

In the experiments, we collected the following data:

- Objective measurements: number of victims found, parallel task accuracy, parallel task completed, parallel task completion speed, parallel task accuracy degradation, reaction time to situation changes, number of LoA switches, percentage of teleoperation use (teleoperation ratio), navigation task completion.

- Subjective measurements: NASA TLX (Hancock and Meshkati, 1988), situational trust scale for automated driving (Reardon et al., 2020), and trust in the system.

Specifically, parallel task completion is the number of the completed questions in the parallel task; parallel task completion speed refers to the average time taken in seconds for each parallel task question; parallel task accuracy degradation is the negative accuracy variance compared with the baseline; reaction time to situation changes records the time interval between the time that environment features appear on the GUI and the time that participants switch the LoA accordingly; navigation task completion is the mark to label if the participant successfully navigate the robot to the endpoint within time limits ("0" is failure, "1" is success); NASA TLX applies a questionnaire to evaluate human workload; situational trust scale for automated driving is derived from a questionnaire which evaluates trust in the obtained SA; trust in the system is a metrics from the situational trust scale, rating from one (low) to seven (high).

6.4.1 Significance analysis

Given the number of participants and no assumption about the data, we conduct a significance analysis using the Wilcoxon signed-rank test, a non-parametric method. We calculate the W-statistic value W, the p-value P, and the effect size r for each measurement. We consider a result statistically significant when P is < 0.05, and a meaningful and strong practical difference when r is > 0.5 according to Cohen's standard (Cohen, 2013). We did not use Bonferorri or related p-value corrections as the study is exploratory.

Table 6.2: Significance analysis using Wilcoxon signed-rank test (n = 14)

Metric	W Statistic	P-value	Effect size (r)
Number of victims found	0.0	0.102	0.436
Parallel task accuracy	17.0	0.514	0.174
Parallel task accuracy degradation	17.0	0.515	0.174
Parallel task completion	42.0	0.542	0.163
Parallel task completion speed	31.0	0.194	0.347
Number of LoA switches	33.0	0.241	0.313
Teleoperation ratio	30.0	0.173	0.365
Trust in the system	10.5	0.080	0.468
Response time to switch LoA	**16.0**	**0.040**	**0.570**
NASA TLX	**11.0**	**0.007**	**0.724**
Situational trust scale	**11.5**	**0.031**	**0.577**
Navigation task completion	15.0	0.317	0.267

We observed statistically significant differences in the reaction time to situation changes, the NASA TLX, and the situational trust scale. Table 6.2 shows that response time to switch LoA is significantly lower and shows a strong practical difference (W = 16.0, P = 0.040, r = 0.570) with semantic indicators. It means high-level semantics significantly help to lower the reaction time to situation changes (e.g. a new detection of personal belongings). NASA TLX is significantly lower and shows a strong practical difference (W = 11.0, P = 0.007, r = 0.724). It refers that semantic indicators lower the perceived human workload. Moreover, the situational trust scale is significantly higher and shows a strong practical difference (W = 11.5, P = 0.031, r = 0.577) with the semantic indicators. It means participants have more trust in their SA with the assistance of semantic indicators.

No significant differences are observed in the metrics: parallel task accuracy, parallel task accuracy degradation, parallel task completion, parallel task completion speed, number of LoA switches, and navigation task completion. However, for the number of victims found and the trust in the system, they show a trend for finding more victims with the semantic indicators.

6.4.2 Correlation analysis

We employ Spearman's correlation analysis to find the hidden relationships/patterns among these factors. The correlation heat map of each measurement is presented in Fig. 6.6. With

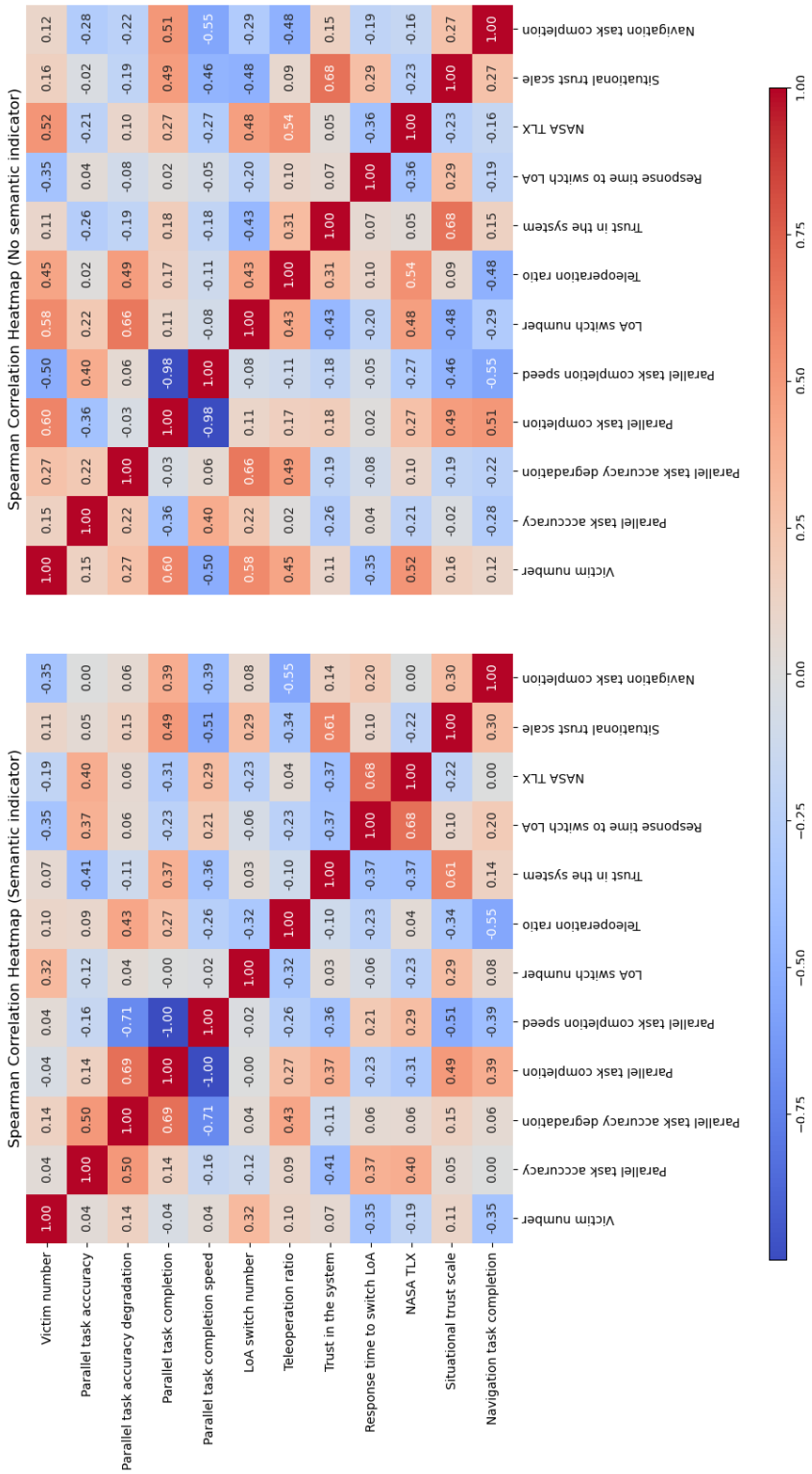


Figure 6.6: Spearman's ρ correlation heatmaps of experimental trials with and without semantics.

the comparison of the two heatmaps, we observe some interesting findings with strong correlations. They are:

- The parallel task accuracy has a positive mild correlation with the NASA TLX (ρ = 0.40) in the semantics trial, i.e. when the parallel task accuracy is high, the cognitive workload is high. However, this shows a negative correlation (ρ = -0.21) in the non-semantics condition. I.e., when the parallel task accuracy is high, the cognitive workload appears slightly decreased.

- The parallel task completion speed is significantly negatively correlated with parallel task accuracy degradation (ρ = -0.71) in the semantics trial. This suggests that, when the parallel task completion speed is high, the parallel task accuracy is higher than baseline. However, this becomes non-correlated (ρ = -0.06) in the non-semantics trial.

- The trust in the system has a mild negative correlation with the NASA TLX (ρ = -0.37) in the semantics trial. I.e. when trust in the system is high, cognitive workload is low. However, this shows non-correlated in the non-semantics trial (ρ = 0.05).

- Trust in the system has a mild negative correlation with the LoA switch number (ρ = -0.43) in the non-semantics trial. This suggests that, when trust in the system is high, the number of LoA switches is low. However, this shows little correlation (ρ = 0.03) in the semantics trial.

6.4.3 Discussion and insights

The significance analysis indicates that high-level semantics have impacts on response time to switch LoA, NASA TLX, and situational trust scale. Specifically, with the help of the semantic indicators, human operators could quickly switch LoA (W = 16.0, P = 0.04, r = 0.570) when the situation changes (e.g. new personal belongings detected). And, NASA TLX significantly decreases (W = 11.0, P = 0.007, r = 0.724), indicating that semantic indicators help to reduce overall cognitive load. Moreover, the situational trust scale significantly increases (W = 11.5, P

= 0.031, r = 0.577), which means human operators have more trust in their SA when assisted by semantic indicators.

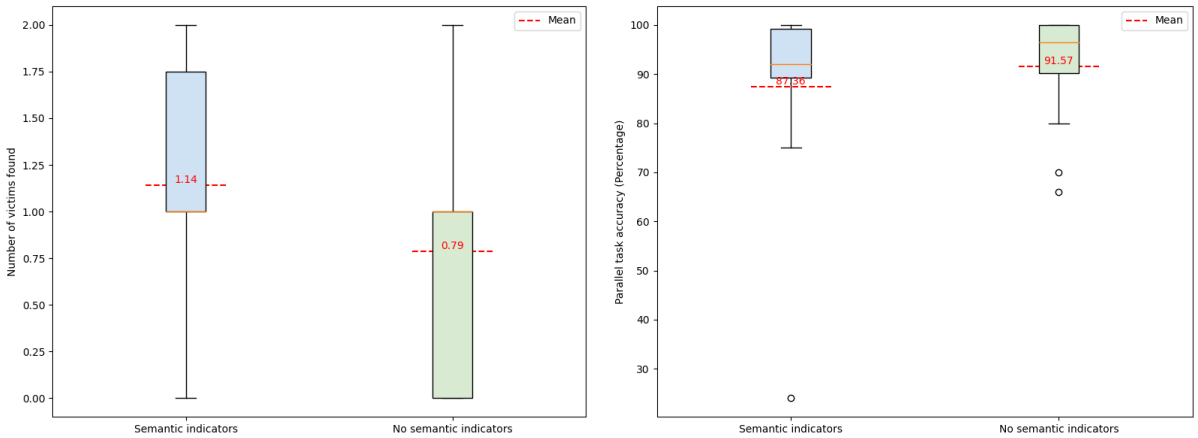
Besides the above significant changes, the analysis shows the trend that high-level semantics might have positive effects on the number of victims found (W = 0, P = 0.102, r = 0.436) and trust in the system (W = 10.5, P = 0.080, r = 0.468). These trends have the potential to become significant impacts if we involve more participants in the experiment. In the end, no significant differences were detected in rest indicators (parallel task accuracy, completion rate, navigation completion.), suggesting that simply adding high-level semantics does not improve these performances. They may still be affected by task design and other factors in the experiment.

In summary, the significance analysis suggests that using high-level semantic indicators, based on our framework, can significantly benefit HRT with respect to: response time to switch LoA; workload of the human operator; and the operator's trust in their SA. High-level semantic indicators show a trend that they help to increase the number of victims found and human's trust in the robotic system.

The correlation analysis reveals some additional interesting findings:

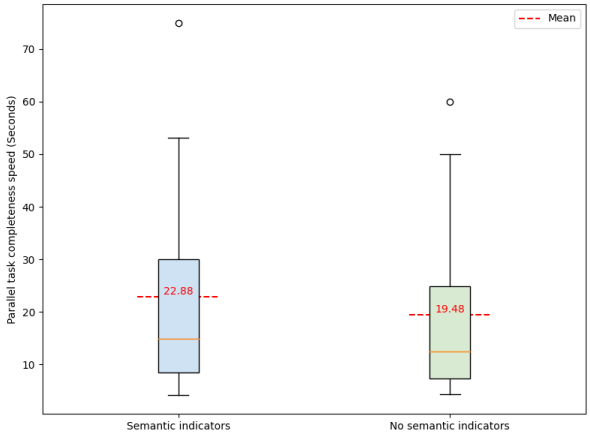
- The correlation between parallel task accuracy and the NASA TLX suggests that the high-level semantic indicators may be providing useful additional information, but this requires additional cognitive workload. If participants seek to improve the parallel task performance by exploiting the semantic indicator displays, this might cause additional workload (ρ = 0.40 suggests a moderate positive monotonic relationship). Hence, even though high-level semantics helps to alleviate cognitive workload, this may cause the human to exert additional effort to improve the parallel task accuracy, beyond what might otherwise have been possible without the semantic indicators. A similar finding is claimed by (Endsley, 2018) that "subjects may have been able to distribute their excess attention to other displays and tasks", if the automation improves SA by reducing excessive workload.

- Correlation between trust in the system and NASA TLX (with-semantics trial) suggests that, when participants have a lower workload, they may trust the system more. However,

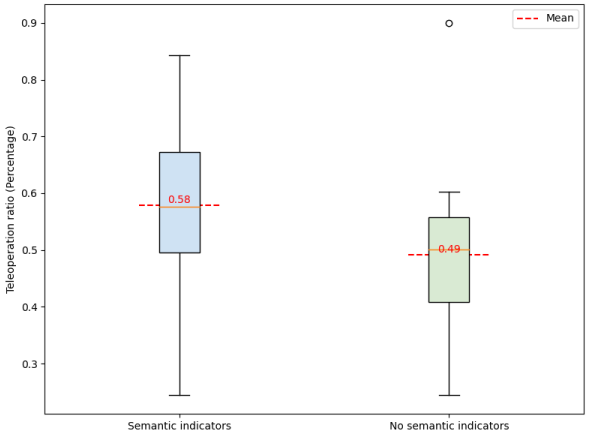


(a) Number of victims found

(b) Parallel task accuracy



(c) Parallel task completion speed

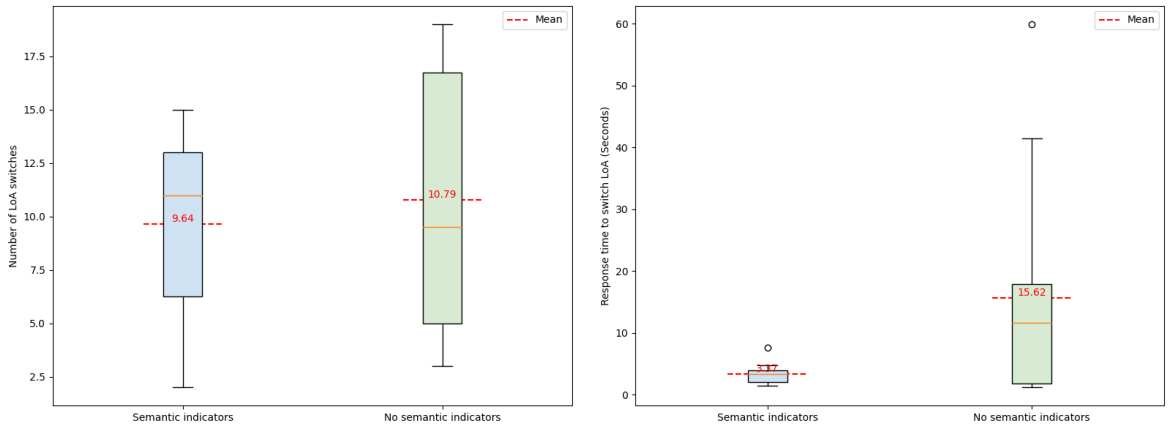


(d) Teleoperation ratio

Figure 6.7: Comparison of task performance metrics between trials with and without semantic indicators (Part 1).

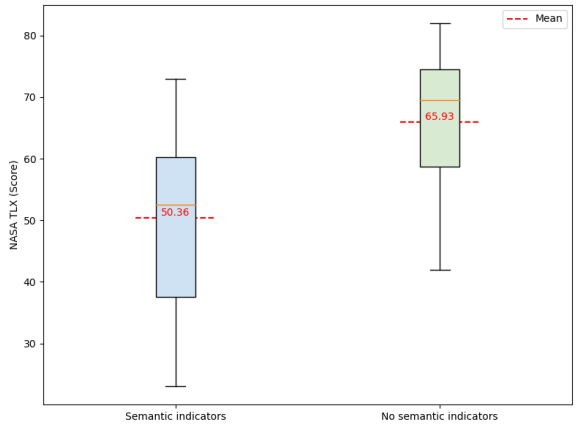
the non-semantics trial shows no such correlation. By providing the high-level semantics, human participants could strengthen their trust in the system from the saved "bandwidth" of cognitive load (Endsley, 2017).

- Trust shows negative correlation with the number of LoA switches (non-semantic trials), but little correlation in with-semantics trials. Some researchers (J. Wang and Kosowski, 2020) claim that "higher trust usually results in an operator accepting the automation's decisions." Human operators take more control when trust goes down. However, we believe that there are more complex reasons hidden behind. Participants with low system trust make many LoA switches (perhaps "noisy decision making"(Hilbert, 2012)) when semantic indicators are unavailable. Noisy decision making could easily happen when

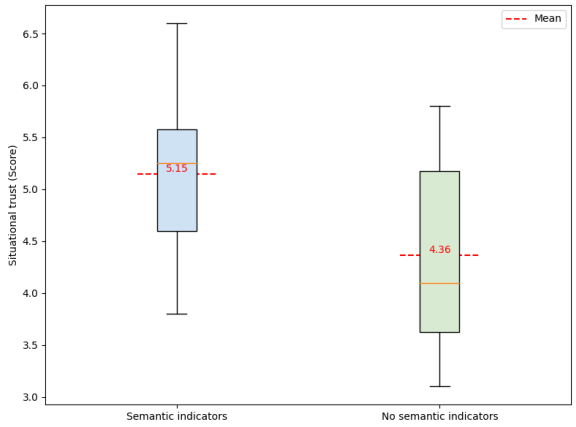


(a) Number of LoA switches

(b) Response time to switch LoA



(c) NASA TLX



(d) Situational trust scale

Figure 6.8: Comparison of cognitive and interaction metrics between trials with and without semantic indicators (Part 2).

humans are in a high-uncertainty situation (poor SA). By introducing semantic indicators (lowering uncertainty and improving SA), human operators might avoid dropping into that high-uncertainty situation.

The box plots of Fig. 6.7 and Fig. 6.8 suggest a correlation between high trust in the robotic system, and large amounts of driving the robot in teleoperation mode. Meanwhile, the correlation between trust and the number of LoA switches in different experiments suggests an interesting relationship, prompting us to explore these relationships further. From Fig. 6.9 we see that participants with low trust and no semantics strongly favour autonomy mode. In contrast, with semantic indicators, participants with low trust strongly favour teleoperation. This suggests that: **i)** the semantic indicator displays comprise a large part of participants' SA, i.e. displaying

these semantic indicators causes participants to feel they have a much better understanding of the remote scene surrounding the robot; **ii)** when participants trust in their SAs, they feel much more confident about using teleoperation mode; **iii)** they prefer teleoperation mode (provided they have sufficient SA) when they have low trust in the robotic "system" i.e. in the autonomous agent. Furthermore, we can see that, at high levels of trust, participants converge on equal use of autonomy and teleoperation, regardless of the amount of SA available to them.

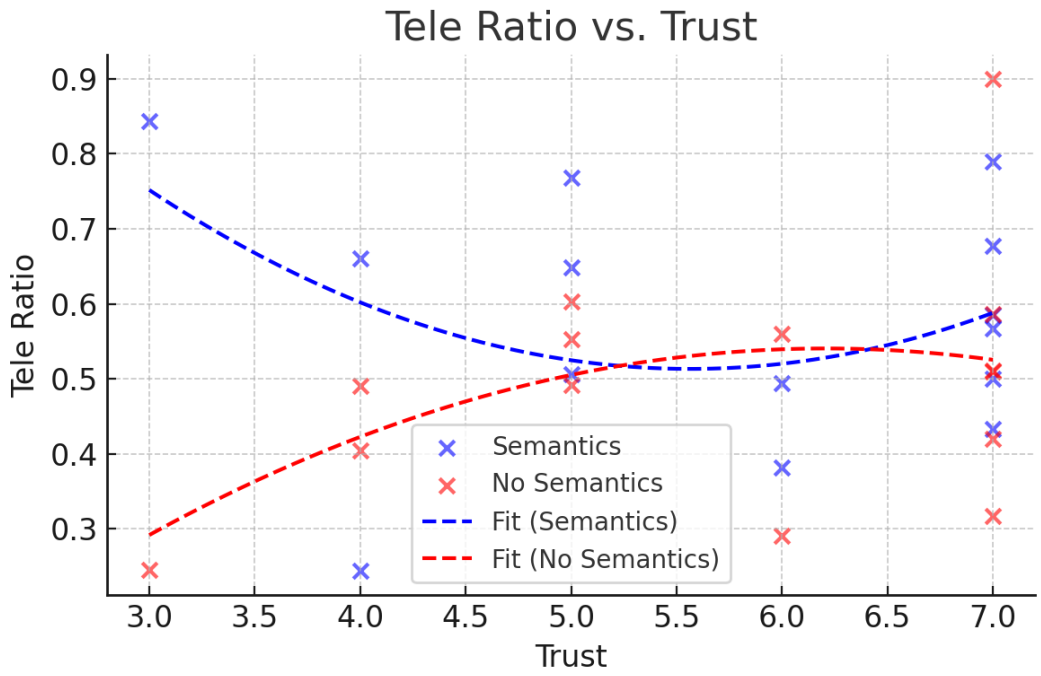


Figure 6.9: Teleoperation ratio vs. trust: red dashed curve is no-semantics condition. Blue dashed curve is with-semantics condition.

Moreover, we record the participants subjective feedback in Appendix B. This information is not strong enough to reveal some patterns, however, it should not be overlooked. We believe it is valuable for further HRT system design to refer to.

6.5 Limitations

To minimize learning effects as a confounder in our "within-groups" experimental design, we used (slightly) different arena layouts between the two trials of each participant. This might have introduced a difference in overall difficulty between the two trials, although we tried to design

the two arenas to have similar difficulties. To check this, we asked participants which arena they thought was more difficult. Five participants thought both arenas were equally easy. Three thought the with-semantics arena was more difficult. Six thought the without-semantics arena was more difficult. While not identical, these results suggest similar difficulties. Note also that the without-semantics condition might make the without-semantics arena seem more difficult, further reassuring that the arenas had similar overall difficulty.

Several participants felt that their performance would improve with a wider-angle camera, a finding supported in the literature (Chiou, Stolkin, Bieksaite, Hawes, K. Shapiro, et al., 2016). Participants with no experience in video games may struggle to understand the 2D bird's-eye view (map) and easily become lost in the arena, despite a substantial pre-experiment training session.

Some participants waited for the sound warning from the system (3 seconds) to end and before they switched LoA. This unnecessarily increased the human idle time in the experiment. Other methods, such as short vibrations on the joystick, might be a better option for delivering this semantic warning.

6.6 Conclusion

In this chapter, we validated the usefulness of our proposed semantics-based SA framework and investigated the effects of high-level semantic indicators in HRT and HRI. Moreover, we explored LoA and HRT patterns when high-level semantics are involved. Our experiments indicate that displaying high-level semantic indicators can help humans: decrease reaction time when switching LoA, reduce cognitive workload, and increase trust in their SA.

Chapter 7

Conclusion and Future Work

This thesis has explored the role of semantic information in HRT, especially for EEs. Semantic information has been explored in AI research, especially in the CV literature. However, there has been relatively little research on semantics in practical field robotics applications. For instance, for FRs in most cases, SA is synthesized by the human operator directly from raw sensory data and prior information about the site. Semantic information, processed by the robot's autonomous agent system, has not previously received much attention in the context of HRT.

In this thesis, we identified the user requirements and challenges from the field and clearly defined the concept of semantics in this context. With this knowledge, we surveyed the FRs about their views and attitudes towards robots, SA, and semantics, which helps us identify the real demands about semantics from the field. We then proposed a semantics-based SA framework, which provides a method to aggregate different semantics into an overall assessment of the semantic richness of environments. We presented experiments on real robots, which show the adaptability and generalizability of this framework to different tasks. Subsequently, we experimentally evaluated this semantic framework's impact on the performance of an HRT and HRI using a simulated SAR mission. The experiment highlighted the difficulty of searching for a victim in such a complex environment, given a limited deployment time. The semantic indicators and the semantic richness framework significantly assisted the human operator in a variety of ways, including reducing cognitive workload and reducing the reaction time of switching

LoA. Moreover, some additional interesting insights were gained, in terms of the relationship between human trust of the system and LoA preference. This further highlighted the complexity of human factors in HRT.

7.1 Contributions

The main contributions of the thesis lie in the scope of HRT and the practical implementation of remote robotic deployments in high consequence or EEs:

- **A systematic survey of requirements and challenges from the perspective of FRs and robots tasks:** The real demands from FRs about semantics and robot tasks requirements, are often overlooked by academic researchers. While the situation has been improving in recent years, the research community has sometimes lacked close collaboration and information sharing with stakeholders in the community. This thesis is an effort to help bridge this gap. From the survey of the FRs' requirements over semantics and SA, and by studying real robot deployments, we clearly identified the real demands expected by FRs, and the potential synergy of high-level semantics processing with practical tasks in EEs.

- **A taxonomy of semantic information in the context of robotics:** Semantics is an abstract concept, but is receiving increasing attention in the AI community, especially in the CV community. Its meaning has been extended from the original linguistic concept to additional emerging topics of study. In the context of robotics, which could be regarded as a platform of NLP, CV, and other AI techniques, we propose a taxonomy of semantic information that clearly defines semantics by categorising them into two distinct levels. Moreover, we clarify the relationships among concepts that are at different layers of abstraction, including raw information, low and high-levels of semantics, and context.

- **A visual detection solution with depth information:** Using the RGB-D camera, we train an image detection network for object detection, specifically covering HAZMAT signs and personal belongings. We apply the YOLACT algorithm with different backbones for ob-

ject detection (locating bounding boxes) and instance segmentation (pixel-wise labeling). Moreover, the depth image is aligned with the RGB image and allows for generation of the semantics, including object classification and depth information. To the best of our knowledge, the detection accuracy of HAZMAT is close to the SOTA.

- **A Situational Awareness framework for robot deployments and its evaluation:** Given the structured knowledge about semantics, we design a semantics-based SA framework. The framework consists of different semantic indicators that reflect different characteristics of the environment, and an aggregated semantic richness score that describes whether the situation needs human attention. In the experiment, we design some SAR-related semantic indicators e.g. signs of human activities. We try to avoid employing deep learning techniques except where necessary, to make the framework computationally lightweight. Moreover, this approach helps to make the framework explainable for tuning. Experiments on real robots demonstrate that the framework is flexible in adapting to different tasks that require different semantic indicator combinations.

- **An analysis of semantics impact on HRT and HRI:** Due to the lack of semantics applications in the field robotics, the semantics impacts in the context of HRT and HRI are not well explored. In a mock-up HRT SAR mission, we identify the impact of high-level semantics on the HRT and HRI across several different modalities of data and different characteristics of the environment. We evaluated the mission performance and human factors using both objective and subjective approaches. We show that the high-level semantics framework successfully reduces the human cognitive workload, reduces reaction time to switch LoA, and increases human trust in the SA. The experiments highlight the complexity of such missions. In addition, we observe a complex relationship, between human trust in the system and LoA preference, under the impact of the semantics framework. E.g. human with poor SA and low trust of the system tends to prefer the autonomy mode.

7.2 Future work

Given the findings from the surveys and our experiments, we here provide some insights and directions for future work.

- **Continuous collaboration with FRs:** As the end-user of the robot system in the field, FR's views and insights need more attention from the academic research community. Their expert knowledge and experience are critical to help researchers identify the specific requirements and narrow the gap between the lab and the field, especially regarding decision-making and planning topics. Researchers need end-user advice to tune algorithm parameters and thresholds, and assess the performance of the system. Hence, the connection with the FRs is necessary for engineering work that is going to be deployed in the field. Moreover, based on our experience engaging with the FRs, the information collected from them is often fragmented. There is a need to formalize the way to communicate with FRs for collecting organized and structured information.

- **An aggregated visual system that synergizes potential visual semantics:** Our work separately trains two networks (one for HAZMAT sign detection and another for personal belonging detection) for the reasons mentioned in Chapter 5. Therefore, if an aggregated network could detect all the related visual semantics, from HAZMAT signs to signs of human victims and other objects of interest, the framework will become more flexible for different missions. Apart from the detection accuracy, the robustness of the visual system is critical in such scenarios. Most cameras do not have good performance in dusty and dark environments with limited visibility. This issue could be addressed by leveraging additional types of sensors, e.g. LiDAR could assist with generating visual semantics. These solutions could help to overcome some of the limitations of RGB cameras in certain types of EEs.

- **A mixed control policy based on the semantic indicators and Situational Semantic Richness:** In this thesis, we have shown how semantic indicators and SSR, can be used to inform human-initiated LoA switches during missions. As a next step, we would like

to incorporate this semantic framework into an MI framework, where LoA switches can be initiated by both the human and the autonomous agent. Semantic indicators and semantic richness could provide additional valuable information for helping to inform LoA switching decisions of the autonomous agent. Previous work from our research group has developed MI controllers in VA HRTs based on fuzzy logic. This provides a promising framework for incorporating our semantic indicators into an MI VA system.

- **A dataset of robot deployment processes with logs of different decision factors, e.g. semantics and LoA switches:** Automated task planning or scheduling is rare in the context of practical mobile robot missions in the field. One of the reasons is that the community lacks an appropriate dataset to directly train a planning strategy using a Markov Decision Process (MDP) or Partially Observable Markov Decision Process (POMDP) model. In the future, it would be very valuable to build a dataset that records time-stamped critical decisions and mission factors in real-time during a large number of robot deployments. This data could include LoA status and switching decisions, human attention, and semantic indicators.

- **A framework to analyze the semantics impact on the HRT and other factors:** The impacts of semantics are not well-explored. Our work provides some insights and findings towards the impact of semantics in HRT. These findings can be entry points to study the semantics impact in this context. However, we still lack a structured and systematic way to assess these impacts. As semantic information and reasoning become increasingly prevalent in robot deployments, the need for an objective impact assessment tool will increase to help inform semantic indicator design.

- **Extended research about high-level semantics impact on the HRT:** Our research in Chapter 6 is exploratory and does not involve a large number of participants. We obtain the preliminary insights based on the results. If more participants could join the research in the future, we expect to have clearer and statistically significant results and insights about the high-level semantics impact.

7.3 Closing thoughts

The development of robotic systems resembles performing a symphony. Core technologies e.g. perception, control, planning, and decision-making form the orchestra's sections. The performance falters when any section is missing or out of tempo; likewise, a robot underperforms when one capability lags.

HRT plays the role of the conductor, coordinating these heterogeneous components into a coherent experience for end-users. Yet technological progress in each of these components is not proceeding at the same pace or with equal robustness. Perception algorithms, which effortlessly detect objects in a brightly lit warehouse, can rapidly fail in darkness, dust, smoke, or uneven and dynamic lighting e.g. spotlights mounted on and moving with the robot. Dynamic task planners may struggle to perform multi-task planning, or optimise decisions with multiple costs and benefits (multi-objective functions). In practice, the conductor, HRT techniques, must adjust the tempo so that every section can contribute what it can. HRT techniques thus bridge between robotic subsystems and users, but it cannot replace any individual subsystem's role.

This asynchronous development constrains where robots can be deployed, especially in hazardous environments. Most field robotics deployments still require human oversight, or direct human control. As autonomous capabilities increase, robots become less predictable, and performance benchmarking and safety verification become much more challenging. Meanwhile, unanswered questions of responsibility and liability, reinforced by regulation, can also limit advanced robotic deployments, especially autonomous systems. Within these tight boundaries, effective HRT maximises robotic potential by blending the efficiency and convenience of automation, with human strengths in high-level cognition, strategic planning, and complex judgment.

Looking ahead, once robots explicitly build the SA, we could arm both human operators and autonomous planners with cues that are immediately interpretable. Such distilled semantics shortens the cognitive distance between raw sensor values and comprehensive SA, allowing MI controllers to modulate autonomy with clearer intent. In practice, this translates to faster pivots when conditions change, better allocation of different tasks, enhancing safety for both humans

and robots, and ultimately a higher probability of mission success under the relentless pressures of an EE.

Progress now hinges on open, cross-domain datasets that fuse semantics, operator actions, and LoA transitions; on shared benchmarks that stress systems in the realistic conditions of real-world emergency response; and on regulatory sandboxes where engineers, first responders, ethicists, regulators, and lawmakers can iterate side by side. Only through this sustained, collaborative cadence can we transform today's promising frameworks into a mature discipline, one where humans and robots perform as a seamless ensemble, ready to confront the unforgiving scenarios in which their partnership matters the most.

References

Abi-Farraj, Firas, Nicolò Pedemonte, and Paolo Robuffo Giordano (2016). "A visual-based shared control architecture for remote telemanipulation". In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 4266–4273.

Adjigble, Maxime et al. (2019). "An assisted telemanipulation approach: combining autonomous grasp planning with haptic cues". In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 3164–3171.

Agha, Ali et al. (2021). "Nebula: Quest for robotic autonomy in challenging environments; team costar at the darpa subterranean challenge". In: *arXiv preprint arXiv:2103.11470*.

AIML (2021). *Semantic vision*. https://www.adelaide.edu.au/aiml/our-research/robotic-vision/semantic-vision.

AMRAAM oue final report (1982).

Arnold, Miriam, Mascha Goldschmitt, and Thomas Rigotti (2023). "Dealing with information overload: a comprehensive review". In: *Frontiers in psychology* 14, p. 1122200.

Arvanitis, Gerasimos et al. (2023). "Cooperative saliency-based pothole detection and ar rendering for increased situational awareness". In: *IEEE Transactions on Intelligent Transportation Systems* 25.5, pp. 3588–3604.

Ausonio, Elena, Patrizia Bagnerini, and Marco Ghio (2021). "Drone swarms in fire suppression activities: A conceptual framework". In: *Drones* 5.1, p. 17.

Bae, Soo Hyun, Inkyu Choi, and Nam Soo Kim (2016). "Acoustic scene classification using parallel combination of LSTM and CNN". In: *Proceedings of the detection and classification of acoustic scenes and events 2016 workshop (DCASE2016)*, pp. 11–15.

Balta, Haris et al. (2015). "UAS deployment and data processing during the Balkans flooding with the support to Mine Action". In: *8th IARP International Workshop on Robotics for Risky Environments; January; Lisbon. Portugal: IARP*. Citeseer.

Battaglino, Daniele et al. (2016). "Acoustic scene classification using convolutional neural networks". In: *IEEE AASP Challenge on Detec*.

Bavle, Hriday et al. (2023). "From slam to situational awareness: Challenges and survey". In: *Sensors* 23.10, p. 4849.

Benloucif, Amir et al. (2019). "Cooperative trajectory planning for haptic shared control between driver and automation in highway driving". In: *IEEE Transactions on Industrial Electronics* 66.12, pp. 9846–9857.

Bennie, Callum et al. (2024). "Alternative Interfaces for Human-initiated Natural Language Communication and Robot-initiated Haptic Feedback: Towards Better Situational Awareness in Human-Robot Collaboration". In: *arXiv preprint arXiv:2401.13903*.

Betta, Zoe et al. (2024). "Perceptions and opinions of rescuers about a quadruped robot in an earthquake scenario". In: *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. IEEE, pp. 1092–1099.

Bird, Benjamin et al. (2018). "A robot to monitor nuclear facilities: Using autonomous radiation-monitoring assistance to reduce risk and cost". In: *IEEE Robotics & Automation Magazine* 26.1, pp. 35–43.

Bolya, Daniel et al. (2019). "Yolact: Real-time instance segmentation". In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9157–9166.

Bosch, Anna, Andrew Zisserman, and Xavier Munoz (2007). "Image classification using random forests and ferns". In: *2007 IEEE 11th international conference on computer vision*. IEEE, pp. 1–8.

Bourdev, Lubomir and Jitendra Malik (2009). "Poselets: Body part detectors trained using 3d human pose annotations". In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE, pp. 1365–1372.

Bustamante, Samuel et al. (2022). "Cats: Task planning for shared control of assistive robots with variable autonomy". In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 3775–3782.

Cain, Brad (2007). "A review of the mental workload literature". In: *Dtic Document*.

Carissoli, Claudia et al. (2024). "Mental workload and human-robot interaction in collaborative tasks: A scoping review". In: *International Journal of Human–Computer Interaction* 40.20, pp. 6458–6477.

Casper, Jennifer and Robin R. Murphy (2003). "Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 33.3, pp. 367–385.

Chen, Liang-Chieh et al. (2017). "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs". In: *IEEE transactions on pattern analysis and machine intelligence* 40.4, pp. 834–848.

Chiou, Manolis, Georgios-Theofanis Epsimos, et al. (2022). "Robot-assisted nuclear disaster response: Report and insights from a field exercise". In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 4545–4552.

Chiou, Manolis, Nick Hawes, et al. (2015). "Towards the principled study of variable autonomy in mobile robots". In: *2015 IEEE international conference on systems, man, and cybernetics*. IEEE, pp. 1053–1059.

Chiou, Manolis, Faye McCabe, et al. (2021). "Trust, shared understanding and locus of control in mixed-initiative robotic systems". In: *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, pp. 684–691.

Chiou, Manolis, Rustam Stolkin, Goda Bieksaite, Nick Hawes, Kimron Shapiro, et al. (Oct. 2016). "Experimental Analysis of a Variable Autonomy Framework for Controlling a Remotely Operating Mobile Robot". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems 2016*.

Chiou, Manolis, Rustam Stolkin, Goda Bieksaite, Nick Hawes, Kimron L Shapiro, et al. (2016). "Experimental analysis of a variable autonomy framework for controlling a remotely oper-

ating mobile robot". In: *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, pp. 3581–3588.

Chiou, Manolis, Mohammed Talha, and Rustam Stolkinl (2019). "Learning effects in variable autonomy human-robot systems: how much training is enough?" In: *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, pp. 720–727.

Chok, Nian Shong (2010). "Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data". PhD thesis. University of Pittsburgh.

Christie, Gordon et al. (2017). "Radiation search operations using scene understanding with autonomous UAV and UGV". In: *Journal of Field Robotics* 34.8, pp. 1450–1468.

Cohen, Jacob (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Connor, Dean T et al. (Mar. 2020). "A Data Processing Workflow for Drone-based Radiation Mapping in the Chernobyl Exclusion Zone (CEZ) - 20119". English. In: *Proceedings of the Waste Management Symposia 2020. Phoenix, Arizona, USA*.

Contreras, Cesar Alan et al. (2025). "A mini-review on mobile manipulators with Variable Autonomy". In: *Frontiers in Robotics and AI* 12, p. 1540476.

Craik, Fergus IM and Endel Tulving (1975). "Depth of processing and the retention of words in episodic memory." In: *Journal of experimental Psychology: general* 104.3, p. 268.

Dai, Angela and Matthias Nießner (2018). "3dmv: Joint 3d-multi-view prediction for 3d semantic scene segmentation". In: *Proceedings of the European conference on computer vision (ECCV)*, pp. 452–468.

De Cubber, Geert et al. (2014). "UAS deployment and data processing during the Balkans flooding". In: *2014 IEEE International Symposium on Safety, Security, and Rescue Robotics (2014)*. IEEE, pp. 1–4.

Devlin, Jacob et al. (2019). "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pp. 4171–4186.

Doherty, Patrick et al. (2021). “Hastily formed knowledge networks and distributed situation awareness for collaborative robotics”. In: *Autonomous Intelligent Systems* 1.1, p. 16.

Endsley, Mica R (1988). “Situation awareness global assessment technique (SAGAT)”. In: *Proceedings of the IEEE 1988 national aerospace and electronics conference*. IEEE, pp. 789–795.

— (1995a). “Measurement of situation awareness in dynamic systems”. In: *Human factors* 37.1, pp. 65–84.

— (1995b). “Toward a theory of situation awareness in dynamic systems”. In: *Human factors* 37.1, pp. 32–64.

— (2017). “From here to autonomy: lessons learned from human–automation research”. In: *Human factors* 59.1, pp. 5–27.

— (2018). “Automation and situation awareness”. In: *Automation and human performance*. CRC Press, pp. 163–181.

Eraslan, Emre, Yildiray Yildiz, and Anuradha M Annaswamy (2020). “Shared control between pilots and autopilots: An illustration of a cyberphysical human system”. In: *IEEE Control Systems Magazine* 40.6, pp. 77–97.

Fei, Juncong et al. (2021). “PillarSegNet: Pillar-based semantic grid map estimation using sparse LiDAR data”. In: *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, pp. 838–844.

Fernandes, Odair, Robin Murphy, Justin Adams, et al. (2018). “Quantitative data analysis: CRASAR small unmanned aerial systems at hurricane Harvey”. In: *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, pp. 1–6.

Fernandes, Odair, Robin Murphy, David Merrick, et al. (2019). “Quantitative data analysis: Small unmanned aerial systems at hurricane michael”. In: *2019 IEEE international symposium on safety, security, and rescue robotics (SSRR)*. IEEE, pp. 116–117.

Fostick, Leah and Nir Fink (2021). “Situational awareness: The effect of stimulus type and hearing protection on sound localization”. In: *Sensors* 21.21, p. 7044.

Fraune, Marlena R et al. (2021). "Developing future wearable interfaces for human-drone teams through a virtual drone search game". In: *International Journal of Human-Computer Studies* 147, p. 102573.

Galy, Edith, Julie Paxion, and Catherine Berthelon (2018). "Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: an example with driving". In: *Ergonomics* 61.4, pp. 517–527.

Ganis, Giorgio and Rogier Kievit (2015). "A new set of three-dimensional shapes for investigating mental rotation processes: validation data and stimulus set". In: *Journal of Open Psychology Data*.

Gao, Yuan et al. (2025). "Lidar remote sensing meets weak supervision: Concepts, methods, and perspectives". In: *arXiv preprint arXiv:2503.18384*.

Garg, Sourav et al. (2020). "Semantics for robotic mapping, perception and interaction: A survey". In: *Foundations and Trends® in Robotics* 8.1–2, pp. 1–224.

Girshick, Ross (2015). "Fast r-cnn". In: *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448.

Goddard, Kate, Abdul Roudsari, and Jeremy C Wyatt (2012). "Automation bias: a systematic review of frequency, effect mediators, and mitigators". In: *Journal of the American Medical Informatics Association* 19.1, pp. 121–127.

Goodrich, Michael A, Alan C Schultz, et al. (2008). "Human–robot interaction: a survey". In: *Foundations and trends® in human–computer interaction* 1.3, pp. 203–275.

Groves, Keir et al. (2021). "Robotic Exploration of an Unknown Nuclear Environment Using Radiation Informed Autonomous Navigation". In: *Robotics* 10.2, p. 78.

Gustafsson, Fredrik (2010). "Particle filter theory and practice with positioning applications". In: *IEEE Aerospace and Electronic Systems Magazine* 25.7, pp. 53–82.

"Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research" (1988). In: *Human Mental Workload*. Ed. by Peter A. Hancock and Najmedin Meshkati. Vol. 52. Advances in Psychology. North-Holland, pp. 139–183. DOI: https://doi.org/10.1016/S0166-4115(08)62386-9.

Hart, Sandra G (2006). "NASA-task load index (NASA-TLX); 20 years later". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. 9. Sage publications Sage CA: Los Angeles, CA, pp. 904–908.

Hatanaka, Takeshi et al. (2020). "Human-robot collaboration with variable autonomy via Gaussian process". In: *IFAC-PapersOnLine* 53.5, pp. 126–133.

Hazirbas, Caner et al. (2016). "Fusenet: Incorporating depth into semantic segmentation via fusion-based cnn architecture". In: *Asian conference on computer vision*. Springer, pp. 213–228.

He, Hongmei et al. (2021). "The challenges and opportunities of human-centered AI for trustworthy robots and autonomous systems". In: *IEEE Transactions on Cognitive and Developmental Systems* 14.4, pp. 1398–1412.

He, Kaiming, Georgia Gkioxari, et al. (2017). "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969.

He, Kaiming, Xiangyu Zhang, et al. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Hellfeld, Daniel et al. (2019). "Real-time free-moving active coded mask 3D gamma-ray imaging". In: *IEEE Transactions on Nuclear Science* 66.10, pp. 2252–2260.

Hergeth, Sebastian et al. (2016). "Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving". In: *Human factors* 58.3, pp. 509–519.

Hilbert, Martin (2012). "Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making." In: *Psychological bulletin* 138.2, p. 211.

Holthausen, Brittany E et al. (2020). "Situational trust scale for automated driving (STS-AD): Development and initial validation". In: *12th international conference on automotive user interfaces and interactive vehicular applications*, pp. 40–47.

Huang, Peide et al. (2023). "What went wrong? closing the sim-to-real gap via differentiable causal discovery". In: *Conference on Robot Learning*. PMLR, pp. 734–760.

Israelsen, Brett W and Nisar R Ahmed (2019). ""Dave... I can assure you... that it's going to be all right..." A definition, case for, and survey of algorithmic assurances in human-autonomy trust relationships". In: *ACM Computing Surveys (CSUR)* 51.6, pp. 1–37.

Jang, Sein, Lkhagvadorj Battulga, and Aziz Nasridinov (2020). "Detection of dangerous situations using deep learning model with relational inference". In: *Journal of Multimedia Information System* 7.3, pp. 205–214.

Jeon, Eun Som et al. (2015). "Human detection based on the generation of a background image by using a far-infrared light camera". In: *Sensors* 15.3, pp. 6763–6788.

Jian, Jiun-Yin, Ann M Bisantz, and Colin G Drury (2000). "Foundations for an empirically determined scale of trust in automated systems". In: *International journal of cognitive ergonomics* 4.1, pp. 53–71.

Jiang, Jindong et al. (2018). "Rednet: Residual encoder-decoder network for indoor rgb-d semantic segmentation". In: *arXiv preprint arXiv:1806.01054*.

Jurafsky, Dan (2000). *Speech & language processing*. Pearson Education India.

Kalman, Rudolph Emil (1960). "A new approach to linear filtering and prediction problems".

Kalyuga, Slava (2009). "The expertise reversal effect". In: *Managing cognitive load in adaptive multimedia learning*. IGI Global, pp. 58–80.

Kapellmann-Zafra, Gabriel et al. (2016). "Human-robot swarm interaction with limited situational awareness". In: *Swarm Intelligence: 10th International Conference, ANTS 2016, Brussels, Belgium, September 7-9, 2016, Proceedings 10*. Springer, pp. 125–136.

Karumanchi, Sisir et al. (2017). "Team RoboSimian: semi-autonomous mobile manipulation at the 2015 DARPA robotics challenge finals". In: *Journal of Field Robotics* 34.2, pp. 305–332.

Kawatsuma, Shinji, Mineo Fukushima, and Takashi Okada (2012). "Emergency response by robots to Fukushima-Daiichi accident: summary and lessons learned". In: *Industrial Robot: An International Journal* 39.5, pp. 428–435.

Kawatsuma, Shinji, Ryuji Mimura, and Hajime Asama (2017). "Unitization for portability of emergency response surveillance robot system: Experiences and lessons learned from the

deployment of the JAEA-3 emergency response robot at the Fukushima Daiichi Nuclear Power Plants". In: *Robomech Journal* 4.1, pp. 1–7.

Kendall, Maurice G (1938). "A new measure of rank correlation". In: *Biometrika* 30.1-2, pp. 81–93.

Kirillov, Alexander et al. (2019). "Panoptic segmentation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9404–9413.

Klein, G (2005). "Recognition-primed decisions". In: *Ergonomics: major writings* 271, pp. 271–307.

Klein, Gary (2000). "Cognitive task analysis of teams". In: *Cognitive task analysis*. Psychology Press, pp. 417–429.

Kosslyn, Stephen Michael and Olivier Koenig (1992). *Wet mind: The new cognitive neuroscience*. Simon and Schuster.

Koverola, Mika et al. (2022). "General attitudes towards robots scale (GAToRS): A new instrument for social surveys". In: *International Journal of Social Robotics* 14.7, pp. 1559–1581.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25, pp. 1097–1105.

Kruijff, Geert-Jan M et al. (2012). "Rescue robots at earthquake-hit Mirandola, Italy: A field report". In: *2012 IEEE international symposium on safety, security, and rescue robotics (SSRR)*. IEEE, pp. 1–8.

Kruijff-Korbayová, Ivana et al. (2016). "Deployment of ground and aerial robots in earthquake-struck amatrice in italy (brief report)". In: *2016 IEEE international symposium on safety, security, and rescue robotics (SSRR)*. IEEE, pp. 278–279.

Langås, Even Falkenberg, Muhammad Hamza Zafar, and Filippo Sanfilippo (2025). "Exploring the synergy of human-robot teaming, digital twins, and machine learning in Industry 5.0: a step towards sustainable manufacturing". In: *Journal of Intelligent Manufacturing*, pp. 1–24.

Leach, Samuel et al. (2021). "Data Augmentation for Improving Deep Learning Models in Building Inspections or Postdisaster Evaluation". In: *Journal of Performance of Constructed Facilities* 35.4, p. 04021029.

Lee, John D and Katrina A See (2004). "Trust in automation: Designing for appropriate reliance". In: *Human factors* 46.1, pp. 50–80.

Lemaignan, Séverin et al. (2017). "Artificial cognition for social human–robot interaction: An implementation". In: *Artificial Intelligence* 247, pp. 45–69.

Li, Mingjun et al. (2018). "Shared control driver assistance system based on driving intention and situation assessment". In: *IEEE Transactions on Industrial Informatics* 14.11, pp. 4982–4994.

Li, Xuanpeng et al. (2017). "Fast semi-dense 3D semantic mapping with monocular visual SLAM". In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 385–390. DOI: 10.1109/ITSC.2017.8317942.

Li, Zewen et al. (2021). "A survey of convolutional neural networks: analysis, applications, and prospects". In: *IEEE transactions on neural networks and learning systems* 33.12, pp. 6999–7019.

Lin, Tsung-Yi et al. (2014). "Microsoft coco: Common objects in context". In: *European conference on computer vision*. Springer, pp. 740–755.

Liu, Cheng-Chien et al. (2015). "Rapidly responding to landslides and debris flow events using a low-cost unmanned aerial vehicle". In: *Journal of Applied Remote Sensing* 9.1, p. 096016.

Liu, Wei et al. (2016). "Ssd: Single shot multibox detector". In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, pp. 21–37.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell (2015). "Fully convolutional networks for semantic segmentation". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

Lu, Lu et al. (2022). "Mental stress and safety awareness during human-robot collaboration-Review". In: *Applied ergonomics* 105, p. 103832.

Luo, Qianhui et al. (2020). "3d-ssd: Learning hierarchical features from rgb-d images for amodal 3d object detection". In: *Neurocomputing* 378, pp. 364–374.

Ma, Miao, Naresh Marturi, Yibin Li, Ales Leonardis, et al. (2018). "Region-sequence based six-stream CNN features for general and fine-grained human action recognition in videos". In: *Pattern Recognition* 76, pp. 506–521.

Ma, Miao, Naresh Marturi, Yibin Li, Rustam Stolkin, et al. (2016). "A local-global coupled-layer puppet model for robust online human pose tracking". In: *Computer Vision and Image Understanding* 153, pp. 163–178.

Mankins, John C et al. (1995). *Technology readiness levels*.

Masouleh, Mehdi Khoshboresh and Reza Shah-Hosseini (2019). "Development and evaluation of a deep learning model for real-time ground vehicle semantic segmentation from UAV-based thermal infrared imagery". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 155, pp. 172–186.

McCormac, John et al. (2017). "Semanticfusion: Dense 3d semantic mapping with convolutional neural networks". In: *2017 IEEE International Conference on Robotics and automation (ICRA)*. IEEE, pp. 4628–4635.

McNeese, Nathan J et al. (2021). "Team situation awareness and conflict: A study of human–machine teaming". In: *Journal of Cognitive Engineering and Decision Making* 15.2-3, pp. 83–96.

Methnani, Leila et al. (2024). "Who's in charge here? A survey on Trustworthy AI in Variable Autonomy Robotic Systems". In: *ACM computing surveys* 56.7, pp. 1–32.

Micire, Mark J (2008). "Evolution and field performance of a rescue robot". In: *Journal of Field Robotics* 25.1-2, pp. 17–30.

Milioto, Andres et al. (2019). "Rangenet++: Fast and accurate lidar semantic segmentation". In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 4213–4220.

Moore, Travis M and Erin M Picou (2018). "A potential bias in subjective ratings of mental effort". In: *Journal of Speech, Language, and Hearing Research* 61.9, pp. 2405–2421.

Muir, Bonnie M and Neville Moray (1996). "Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation". In: *Ergonomics* 39.3, pp. 429–460.

Murphy, Jason P, Clara Bergström, and Lina Gyllencruetz (2025). "Rural hospital incident command leaders' perceptions of disaster preparedness". In: *BMC Emergency Medicine* 25.1, p. 45.

Murphy, Robin et al. (2016). "Two case studies and gaps analysis of flood assessment for emergency management with small unmanned aerial systems". In: *2016 IEEE international symposium on safety, security, and rescue robotics (SSRR)*. IEEE, pp. 54–61.

Murphy, Robin R (2004a). "Activities of the rescue robots at the world trade center from 11-21 September 2001". In: *IEEE Robotics & Automation Magazine* 11.3, pp. 50–61.

— (2004b). "Trial by fire [rescue robots]". In: *IEEE Robotics & Automation Magazine* 11.3, pp. 50–61.

— (2014). *Disaster robotics*. MIT press.

Murphy, Robin R and Jennifer L Burke (2005). "Up from the rubble: Lessons learned about HRI from search and rescue". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 49. 3. SAGE Publications Sage CA: Los Angeles, CA, pp. 437–441.

Murphy, Robin R, Brittany A Duncan, et al. (2016). "Use of a Small Unmanned Aerial System for the SR-530 Mudslide Incident near Oso, Washington". In: *Journal of field Robotics* 33.4, pp. 476–488.

Murphy, Robin R and Debra Schreckenghost (2013). "Survey of metrics for human-robot interaction". In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 197–198.

Mutzenich, Clare et al. (2021). "Updating our understanding of situation awareness in relation to remote operators of autonomous vehicles". In: *Cognitive research: principles and implications* 6, pp. 1–17.

Nadaraya, Elizbar A (1964). "On estimating regression". In: *Theory of Probability & Its Applications* 9.1, pp. 141–142.

Nagatani, Keiji et al. (2013). "Emergency response to the nuclear accident at the Fukushima Daiichi Nuclear Power Plants using mobile rescue robots". In: *Journal of Field Robotics* 30.1, pp. 44–63.

Nardelli, Alice, Antonio Sgorbissa, and Carmine Tommaso Recchiuto (2024). "Personality-and memory-based software framework for human-robot interaction". In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 17388–17394.

Naseer, Muzammal, Salman Khan, and Fatih Porikli (2018). "Indoor scene understanding in 2.5/3d for autonomous agents: A survey". In: *IEEE access* 7, pp. 1859–1887.

Noriega, Pablo (1999). "Agent mediated auctions: the fishmarket metaphor". Citeseer.

Okutani, Keita et al. (2012). "Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter". In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 3288–3293.

Padilla, Rafael, Sergio L Netto, and Eduardo AB Da Silva (2020). "A survey on performance metrics for object-detection algorithms". In: *2020 international conference on systems, signals and image processing (IWSSIP)*. IEEE, pp. 237–242.

Paivio, Allan (1991). "Dual coding theory: Retrospect and current status." In: *Canadian Journal of Psychology/Revue canadienne de psychologie* 45.3, p. 255.

Peng, Kunyu et al. (2022). "MASS: Multi-attentional semantic segmentation of LiDAR data for dense top-view understanding". In: *IEEE Transactions on Intelligent Transportation Systems*.

Peskoe-Yang, Lynne (2019). *Paris Firefighters Used This Remote-Controlled Robot to Extinguish the Notre Dame Blaze*. https://spectrum.ieee.org/colossus-the-firefighting-robot-that-helped-save-notre-dame.

Pozzer, Sandra et al. (2021). "Semantic segmentation of defects in infrared thermographic images of highly damaged concrete structures". In: *Journal of Performance of Constructed Facilities* 35.1, p. 04020131.

Ramachandran, Ragesh Kumar et al. (2021). "Resilient monitoring in heterogeneous multi-robot systems through network reconfiguration". In: *IEEE Transactions on Robotics* 38.1, pp. 126–138.

Ramesh, Aniketh, Christian A. Braun, et al. (2023). "Experimental Evaluation of Model Predictive Mixed-Initiative Variable Autonomy Systems Applied to Human-Robot Teams". In: *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC), October 1-4, 2023, Hyatt Maui, Hawaii, USA*. IEEE International Conference on Systems, Man, and Cybernetics. SMC 2023 (Maui, Hawaii, Oct. 1–4, 2023). IEEE.

Ramesh, Aniketh, Rustam Stolkin, and Manolis Chiou (2022). "Robot vitals and robot health: Towards systematically quantifying runtime performance degradation in robots under adverse conditions". In: *IEEE Robotics and Automation Letters* 7.4, pp. 10729–10736.

Reardon, Christopher et al. (2020). "Enabling situational awareness via augmented reality of autonomous robot-based environmental change detection". In: *International Conference on Human-Computer Interaction*. Springer, pp. 611–628.

Redmon, Joseph (2016). *Darknet: Open Source Neural Networks in C*. http://pjreddie.com/darknet/.

Reid, Gary B and Thomas E Nygren (1988). "The subjective workload assessment technique: A scaling procedure for measuring mental workload". In: *Advances in psychology*. Vol. 52. Elsevier, pp. 185–218.

Ren, Zhao et al. (2018). "Deep scalogram representations for acoustic scene classification". In: *IEEE/CAA Journal of Automatica Sinica* 5.3, pp. 662–669.

Riley, Victor (2018). "Operator reliance on automation: Theory and data". In: *Automation and human performance*. CRC Press, pp. 19–35.

Robinson, Christopher, Indika B Wijayasinghe, and Dan O Popa (2019). "Quantitative variable autonomy levels for traded control in a pick-and-place task". In: *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE, pp. 697–702.

Roldán, Juan Jesús et al. (2017). "Multi-robot interfaces and operator situational awareness: Study of the impact of immersion and prediction". In: *Sensors* 17.8, p. 1720.

Ronneberger, O., P. Fischer, and T. Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *MICCAI*.

Rosinol, Antoni et al. (2021). "Kimera: From SLAM to spatial perception with 3D dynamic scene graphs". In: *The International Journal of Robotics Research* 40.12-14, pp. 1510–1546.

Ruan, Tianshu, Zoe Betta, et al. (2025). "First Responders' Perceptions of Semantic Information for Situational Awareness in Robot-Assisted Emergency Response". In: *arXiv preprint arXiv:2510.16692*.

Ruan, Tianshu, Aniketh Ramesh, Rustam Stolkin, et al. (2025). "An Exploratory Study on Human-Robot Interaction using Semantics-based Situational Awareness". In: *arXiv preprint arXiv:2507.17376*.

Ruan, Tianshu, Aniketh Ramesh, Hao Wang, et al. (2025). "A Framework for Semantics-based Situational Awareness during Mobile Robot Deployments". In: *arXiv preprint arXiv:2502.13677*.

Ruan, Tianshu, Hao Wang, et al. (2022). "A Taxonomy of Semantic Information in Robot-Assisted Disaster Response". In: *2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pp. 285–292. DOI: 10.1109/SSRR56537.2022.10018727.

Salmon, Paul et al. (2006). "Situation awareness measurement: A review of applicability for C4i environments". In: *Applied ergonomics* 37.2, pp. 225–238.

Sam, Yi Ting et al. (2024). "The Impact of Stress and Workload on Human Performance in Robot Teleoperation Tasks". In: *IEEE Transactions on Robotics*.

Scerri, Paul et al. (2011). "Flood disaster mitigation: A real-world challenge problem for multi-agent unmanned surface vehicles". In: *International Conference on Autonomous Agents and Multiagent Systems*. Springer, pp. 252–269.

Schneider, Frank E et al. (2015). *Possible scenarios for radiation measurementsand sampling using unmanned systems: ERNCIP thematic group Radiological and nuclear threatsto critical infrastructure Task 3 deliverable 2*. Publications Office of the European Union.

Schober, Patrick, Christa Boer, and Lothar A Schwarte (2018). "Correlation coefficients: appropriate use and interpretation". In: *Anesthesia & analgesia* 126.5, pp. 1763–1768.

Senaratne, Hashini et al. (2025). "A Framework for Dynamic Situational Awareness in Human Robot Teams: An Interview Study". In: *ACM Transactions on Human-Robot Interaction*.

Sentouh, Chouki et al. (2018). "Driver-automation cooperation oriented approach for shared control of lane keeping assist systems". In: *IEEE Transactions on Control Systems Technology* 27.5, pp. 1962–1978.

Sevastopoulos, Christos and Stasinos Konstantopoulos (2022). "A survey of traversability estimation for mobile robots". In: *IEEE Access*.

Sharifi, Amir, Ahmadreza Zibaei, and Mahdi Rezaei (2020). "DeepHAZMAT: Hazardous materials sign detection and segmentation with restricted computational resources". In: *Available at SSRN 3649600*.

— (2021). "A deep learning based hazardous materials (HAZMAT) sign detection robot with restricted computational resources". In: *Machine Learning with Applications* 6, p. 100104.

Shi, Qiuying and Jiqing Han (2021). "Semantic feature extraction based on subspace learning with temporal constraints for acoustic event recognition". In: *Digital Signal Processing* 110, p. 102947.

Simon, Tomas et al. (2017). "Hand keypoint detection in single images using multiview bootstrapping". In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1145–1153.

Smith, Patrick et al. (2013). "Human error analysis of the Macondo well blowout". In: *Process Safety Progress* 32.2, pp. 217–221.

Snyder, Luke S et al. (2019). "Situational awareness enhanced through social media analytics: A survey of first responders". In: *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*. IEEE, pp. 1–8.

Sobol, Ilya M (2001). "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates". In: *Mathematics and computers in simulation* 55.1-3, pp. 271–280.

Soltani, AR and T Fernando (2004). "A fuzzy based multi-objective path planning of construction sites". In: *Automation in construction* 13.6, pp. 717–734.

Spearman, Charles (1961). "The proof and measurement of association between two things." In.

Stanton, Neville A et al. (2017). *Human factors methods: a practical guide for engineering and design*. CRC Press.

Stolkin, Rustam, Norbert Molitor, et al. (2023). “Status, Barriers and Cost-Benefits of Robotic and Remote Systems Applications in Nuclear Decommissioning and Radioactive Waste Management”. In: NEA/RWM/R(2022)1.

Stolkin, Rustam, Sreeram Radhakrishnan, et al. (2007). “Passive acoustic detection of modulated underwater sounds from biological and anthropogenic sources”. In: *OCEANS 2007*. IEEE, pp. 1–8.

Stolkin, Rustam, Alexander Sutin, et al. (2006). “Feature based passive acoustic detection of underwater threats”. In: *Photonics for Port and Harbor Security II*. Vol. 6204. SPIE, pp. 40–49.

Su, Hang et al. (2022). “A human activity-aware shared control solution for medical human–robot interaction”. In: *Assembly Automation* 42.3, pp. 388–394.

Sun, Yuxiang et al. (2020). “FuseSeg: semantic segmentation of urban scenes based on RGB and thermal data fusion”. In: *IEEE Transactions on Automation Science and Engineering* 18.3, pp. 1000–1011.

Surmann, Hartmut et al. (2021). “Deployment of Aerial Robots after a major fire of an industrial hall with hazardous substances, a report”. In: *2021 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, pp. 40–47.

Tamura, Yoshihiro, Hisanori Amano, and Jun Ota (2020). “Analysis of firefighting skill with a teleoperated robot”. In: *Robomech Journal* 7, pp. 1–14.

Taylor, Richard M (2017). “Situational awareness rating technique (SART): The development of a tool for aircrew systems design”. In: *Situational awareness*. Routledge, pp. 111–128.

Varghese, Rejin and M Sambath (2024). “Yolov8: A novel object detection algorithm with enhanced performance and robustness”. In: *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*. IEEE, pp. 1–6.

Vaswani, Ashish et al. (2017). “Attention is all you need”. In: *Advances in neural information processing systems* 30.

Verhagen, Ruben S, Mark A Neerincx, and Myrthe L Tielman (2022). "The influence of interdependence and a transparent or explainable communication style on human-robot teamwork". In: *Frontiers in Robotics and AI* 9, p. 993997.

— (2024). "Meaningful human control and variable autonomy in human-robot teams for firefighting". In: *Frontiers in Robotics and AI* 11, p. 1323980.

Vetter, Kai et al. (2018). "Gamma-Ray imaging for nuclear security and safety: Towards 3-D gamma-ray vision". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 878, pp. 159–168.

Voudoukis, Nikolaos and Sarantos Oikonomidis (2017). "Inverse square law for light and radiation: A unifying educational approach". In: *European Journal of Engineering and Technology Research* 2.11, pp. 23–27.

Wagner, Alan R, Jason Borenstein, and Ayanna Howard (2018). "Overtrust in the robotic age". In: *Communications of the ACM* 61.9, pp. 22–24.

Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao (2023). "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7464–7475.

Wang, Hao (2022). "Risk score: a vision-based risk assessment method for autonomous robot navigation". Bachelor dissertation. University of Birmingham.

Wang, Jun and Les Kosowski (2020). "Modelling Human Attributes in Autonomous Systems Operations". Australian Government Department of Defence.

Wasserstein, Ronald L and Nicole A Lazar (2016). *The ASA statement on p-values: context, process, and purpose*.

Watson, Geoffrey S (1964). "Smooth regression analysis". In: *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 359–372.

Wilcoxon, Frank (1992). "Individual comparisons by ranking methods". In: *Breakthroughs in statistics: Methodology and distribution*. Springer, pp. 196–202.

Wu, Chung-Hsien and Wei-Bin Liang (2010). "Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels". In: *IEEE Transactions on Affective Computing* 2.1, pp. 10–21.

Wu, Xiongwei, Doyen Sahoo, and Steven CH Hoi (2020). "Recent advances in deep learning for object detection". In: *Neurocomputing* 396, pp. 39–64.

Yang, Wei et al. (2018). "Visual semantic navigation using scene priors". In: *arXiv preprint arXiv:1810.06543*.

Yu, Donghang et al. (2020). "An efficient and lightweight convolutional neural network for remote sensing image scene classification". In: *Sensors* 20.7, p. 1999.

Zang, Ying et al. (2020). "Far-infrared object segmentation focus on transmission of overall semantic information". In: *IEEE Access* 8, pp. 182564–182579.

Zhang, Mingyuan, Mi Zhu, and Xuefeng Zhao (2020). "Recognition of high-risk scenarios in building construction based on image semantics". In: *Journal of Computing in Civil Engineering* 34.4, p. 04020019.

Zhang, Yang et al. (2020). "Polarnet: An improved grid representation for online lidar point clouds semantic segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9601–9610.

Zhao, Cheng, Li Sun, and Rustam Stolkin (2017). "A fully end-to-end deep learning approach for real-time simultaneous 3D reconstruction and material recognition". In: *2017 18th International Conference on Advanced Robotics (ICAR)*. IEEE, pp. 75–82.

Zhao, Xu et al. (2023). "Fast segment anything". In: *arXiv preprint arXiv:2306.12156*.

Zheng, Ce et al. (2023). "Deep learning-based human pose estimation: A survey". In: *ACM Computing Surveys* 56.1, pp. 1–37.

Zhu, Jinsong et al. (2020). "Intelligent fire monitor for fire robot based on infrared image feedback control". In: *Fire Technology* 56.5, pp. 2089–2109.

Zhuang, Fuzhen et al. (2020). "A comprehensive survey on transfer learning". In: *Proceedings of the IEEE* 109.1, pp. 43–76.

Zhuang, Zhuangwei et al. (2021). “Perception-aware Multi-sensor Fusion for 3D LiDAR Semantic Segmentation”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16280–16290.

Zolotas, Mark, Joshua Elsdon, and Yiannis Demiris (2018). “Head-mounted augmented reality for explainable robotic wheelchair assistance”. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1823–1829.

Appendix A

Questionnaires for FR about Robotic Mission

The questionnaires has been translated into Italian, Greek, and Chinese for recruiting more participants based on the English version. Here, we present the original English version of the questionnaires.

A.1 Introduction

Dear survey participant,

Thank you very much for taking the time to participate in this study. We are looking forward to learning from your expert knowledge and experience. The estimated completion time is 15 minute.

We are robotics researchers from the University of Birmingham, University of Genoa, Queen Mary University of London, and University of Bristol. We are carrying out research to develop advanced robotics and AI technologies which could assist first responders, and workers in other hazardous environments. We aim to make robotic systems more effective, more reliable and easier to use in future.

In particular, we want to improve our understanding of how expert human workers and first responders currently carry out their tasks during safety-critical missions. We would like to also

understand the real-life experiences of missions which have involved the use of robots or might involve robots in the future. We are especially interested in issues around Situational Awareness, and also how humans make use of "Semantic Information" (understanding the "meaning" of objects, scenes and situations) during missions.

If you do not have direct experience with deploying robots, your answers can still be very helpful for us. Please answer all questions, and select the answers which feel most appropriate to you.

All data collected is anonymous. All collected data will be handled in an ethical way, and will only be used for research purposes.

This questionnaire includes 3 sections:

- **1)** Background and basic demographic information
- **2)** General attitudes to robots
- **3)** Situational Awareness (SA) and use of semantic information

We briefly explain the meaning of SA and semantic information here:

- Situational Awareness (SA) is **perception**, **comprehension**, and **prediction** of environment and how these change over time. It comprises the current on-site information, the understanding of the situation, and the prediction of the situation.
- Semantic Information (SI) refers to the understanding about the objects/measurements, and the relationships among them. It provides information on the **perception** and **comprehension** level of SA.

A.2 Background Experience and Demographics

In this section, we would like to assess your general attitude towards robots, e.g. the general acceptance of robot applications. You need to choose to which extent you agree with the statements.

- **1)** What is your age?
- **2)** What is your gender?
- **3)** What is your highest education level?
- **4)** Which country/state are you based in?
- **5)** What is your role in the organization?
- **6)** Which option best describes your role in the field during the mission/deployment?
- **7)** How much deployment/field experience do you have?
- **8)** During deployment, what are the tasks you conduct? Please answer briefly.
- **9)** What kind of training/knowledge is required for the mission? Please answer briefly.
- **10)** To what extent are robots used in field deployment within your organization?
- **11)** What kind of tasks do robots conduct within your organization? (choose all that apply)
- **12)** To what extent do the current deployments of robots achieve the mission/task objectives? (If not applicable, please skip this question)
- **13)** In the mid-term future (e.g. in around 5 years), do you expect robots to be deployed less frequently more frequently, or about the same
- **14)** What kinds of robots are deployed? (choose all that apply)
- **15)** What capabilities in terms of their level of autonomy do the deployed robots have? (choose all that apply)
- **16)** What capabilities in terms of autonomy or intelligence do the robots have?
- **17)** Is the robot allowed to take any initiative (e.g. actions, tasks, team role)? If yes, please briefly explain. If no use of robots, please enter ”not applicable”.

A.3 General attitudes to robots

In this section, we would like to assess your general attitude towards robots, e.g. the general acceptance of robot applications. You need to choose to which extent you agree with the statements.

- **18)** I would feel relaxed talking with a robot.
- **19)** I can trust persons and organizations related to development of robots.
- **20)** Persons and organizations related to development of robots will consider the needs, thoughts and feelings of their users.
- **21)** I can trust a robot.
- **22)** If robots had emotions, I would be able to befriend them.
- **23)** I would feel uneasy if I was given a job where I had to use robots.
- **24)** Robots scare me.
- **25)** I would feel very nervous just being around a robot.
- **26)** I fear that a robot would not understand my commands.
- **27)** I don't want a robot to touch me.
- **28)** Robots are necessary because they can do jobs that are too hard or too dangerous for people.
- **29)** Robots can make life easier.
- **30)** Assigning routine tasks to robots lets people do more meaningful tasks.
- **31)** Dangerous tasks should primarily be given to robots.
- **32)** Robots are a good thing for society, because they help people.

- **33)** Robots may make us even lazier.
- **34)** Widespread use of robots is going to take away jobs from people.
- **35)** I am afraid that robots will encourage less interaction between humans.
- **36)** Robotics is one of the areas of technology that needs to be closely monitored.
- **37)** Unregulated use of robotics can lead to societal upheavals.
- **39)** Generally speaking, I have a positive view of robots.
- **38)** I have personal experience using robots.
- **40)** I am interested in scientific discoveries and technological developments.
- **41)** Robotics is a familiar topic to me.

A.4 Situational Awareness (SA) and use of semantic information

In this section, we would like to know more about your views on Situational Awareness (SA) and semantics based on your understanding and experience. We define Situational Awareness and semantics here: Situational Awareness (SA) refers to perception, comprehension, and prediction in a given time and space, especially in dynamic and high-stakes situations. In other words, SA means knowing what's around you, understanding what it means, and being able to predict what might happen next. Semantics refers to extracting and conveying an understanding of environments, to both robots and humans, in terms of information about entities and both explicit and implicit relationships between entities. In other words, semantic information is about identifying objects/measurements, understanding their meanings, and how they relate to each other. It only provides perception and comprehension of the current situation. Example: A self-driving car detects a pedestrian stepping onto the road (perception/entities), understands the potential risk (comprehension/relationships), and analyzes the possibility of crashing (prediction).

Hence, SA is a more general concept than semantics. Semantics is a part of SA.

- **42)** Could you briefly explain what information is usually known before the mission?
- **43)** How much do you have an estimation of the Situational Awareness before the mission?
- **44)** How do you typically obtain real-time Situational Awareness during the mission? (choose all that apply)
- **45)** Could you briefly explain some of the challenges in obtaining Situational Awareness?
- **46)** Please list, in order, the top 3 most important types of semantic information based on your experience?
- **47)** How often do you experience loss of concentration, fatigue, and cognitive overload?
- **48)** From your experience, please list some of the main causes (e.g. 3 max) of the above fatigue and overload.
- **49)** To what extent do you think having semantic information is useful in building Situational Awareness?
- **50)** To what extent is semantic understanding and reasoning used during missions?
- **51)** In what ways are semantic information delivered to human workers during missions? (Choose all that apply – this question is not limited to only missions involving robots)
- **52)** In the ideal scenario, how would you prefer for the semantic-related information to be delivered/communicated to inform humans? (choose all that apply)
- **53)** Can semantic information and semantic scene understanding help predict or mitigate an emergency that was not predicted or foreseen before the mission?
- **54)** Imagine you have some AI agent assisting in your mission (e.g. in a robot, in a computer etc). How much do you agree with the statement ”It is helpful if the AI agent can process an understanding of the situation”?

- **55)** If AI-generated semantic information/understanding is not always correct or reliable, what is the minimum accuracy that would enable you to trust the AI?
- **56)** If AI-generated semantic information/understanding is not always correct or reliable, what is the minimum accuracy that you think would still be useful?

Appendix B

Participants' Feedback from Experiments in Chapter 6

Besides the measurements and questionnaires (e.g. NASA TLX and situational trust scale), we also collect the subjective feedback from each human participant. It is scattered and unorganized information, and hard to present in a scientific way. However, this information should not be overlooked, and it reveals some potential factors to affect the HRT and HRI. We list their suggestions and feedback below:

- Exploring mode, which is missing from the experiment, might help to reduce the workload.
- More training is required, especially for detecting objects in the scenarios.
- Try to follow the scheduled path (the shortest navigation path) most of the time, and only deviate when participants feel it is necessary to achieve tasks.
- As the human operator who has little disaster response knowledge, I expect more instructions about how to leverage the importance of different tasks before the mission.
- I misunderstood the ability of the autonomy mode.
- The current GUI design is clear enough for the given mission.

- I forgot to switch to the proper LoA when I felt comfortable with the current situation (e.g. turn to autonomy when SSR is low).
- The workload is so high that I forgot to react when the situation changed, even with the clear instructions before the mission.
- The warning sounds of high SSR delayed my response to switch LoA.
- The NASA TLX score is affected by personality. A stable mind will show less variance on NASA TLX score.
- Having the map rotate and keeping the robot pointing upwards will help the human navigate and locate the robot.
- The performance of the mission would increase if a higher top velocity is available.
- It is hard to identify the size of the robot from a first-person view of the camera feed.
- Learning effects still play important roles in robot navigation.

From their feedback, we noticed that even though we are aware of many confounding factors e.g. learning effects before the mission, we can not 100% eliminate all the potential confounding factors in the mission. However, the above feedback from human participants helps us balance these factors.