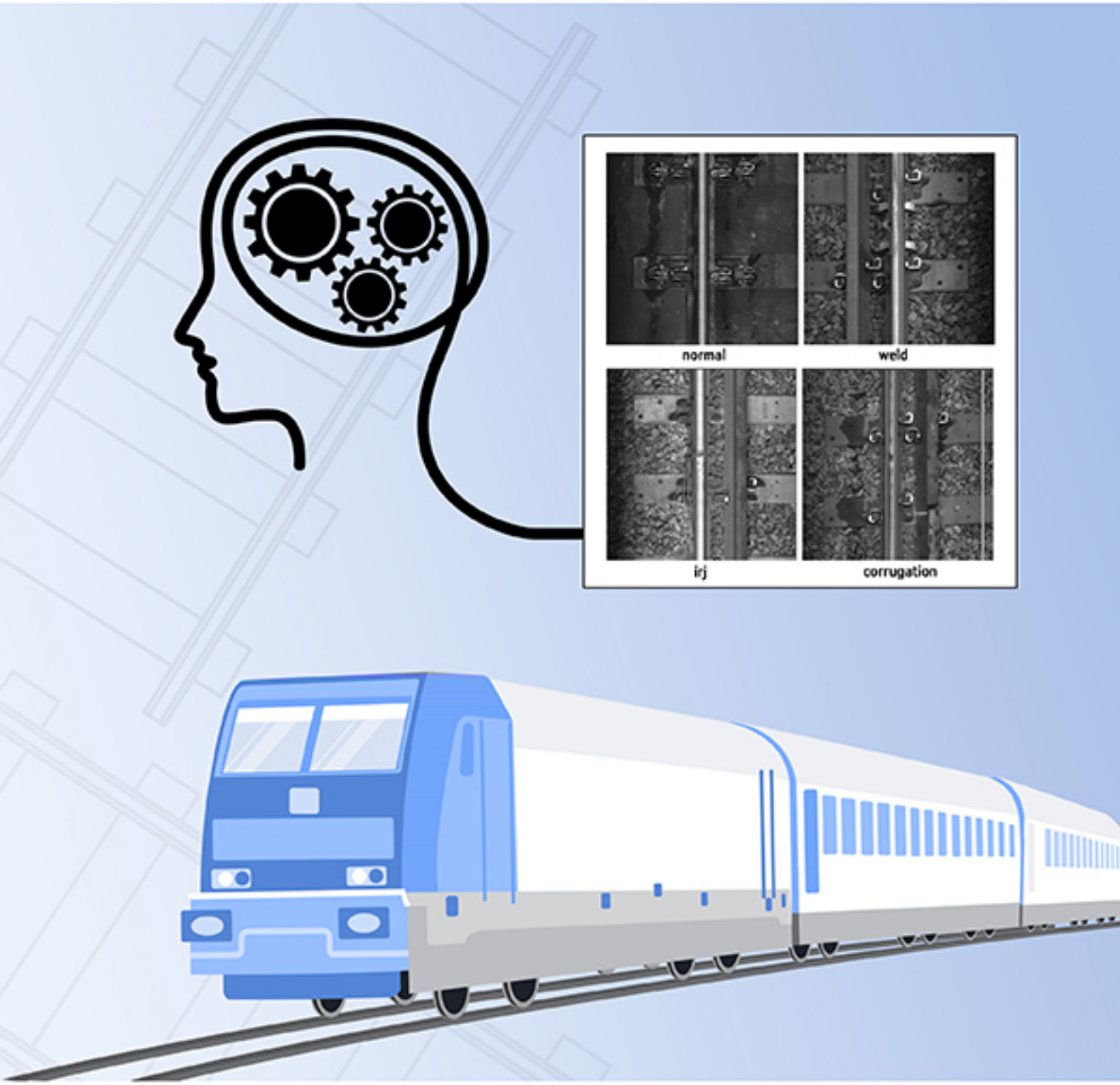
ISSN 2770-3541 (Online)
Volume 1 | Issue 2 | December 2021

IR Intelligence & Robotics



OPEN ACCESS

www.intellrobot.com

Editor-in-Chief

Simon X. Yang

Prof. Simon X. Yang is currently the Head of the Advanced Robotics and Intelligent Systems Laboratory at the University of Guelph. His research interests include artificial intelligent, robotics, sensors and multi-sensor fusion, wireless sensor networks, control systems, bio-inspired intelligence, machine learning, neural networks, fuzzy systems, and computational neuroscience.

Our Features

(1) Gold Open Access
(2) Strong Editorial Board
(3) Rigorous Peer-review
(4) Free English Ianguage Editing Service
(5) Online First Once Accepted
(6) Free Publication Before 31 Dec 2024
(7) Wide Promotion (Twitter\LinkedIn\WeChat\Facebook)

Editorial Board

- 1 Editor-in-Chief
- 5 Advisory Editorial Members
- 24 Associate Editors
- 15 Youth Editorial Board Members

Scope

Top-quality unpublished original technical and non-technical application-focused articles are welcome from intelligence and robotics, particularly on the interdisciplinary areas of intelligence and robotics, including but not limited to the following areas:

- biological, bio-inspired, and artificial intelligence;
- neural networks, fuzzy systems, and evolutionary algorithms;
- sensing, multi-sensor fusion, localization, data analysis, modeling, planning, and control for various mobile, aerial, and underwater robotic systems;
- robot cooperation, teleoperation, and human-machine interactions.
- development and maintenance of real-world intelligent and robotic systems by multidisciplinary teams of scientists and engineers.

Journal Home

https://intellrobot.com/

Submission Link

https://oaemesas.com/login?JournalId=ir

EDITORIAL BOARD

Editor-in-Chief

Simon X. Yang
University of Guelph, Canada

Advisory Board Members

Tianyou Chai
Northeastern University, China

Clarence W. De Silva
University of British Columbia, Canada

Toshio Fukuda
Nagoya University, Japan

Aike Guo
University of Chinese Academy of Sciences, China

Deyi Li
Chinese Academy of Engineering, China

Associate Editors

Mohammad Biglarbegian
University of Guelph, Canada

Hicham Chaoui
Carleton University, Canada

Guang Chen
Tongji University, China

Abdelghani Chibani
University of Paris-Est Creteil (UPEC), France

Carlos Renato Lisboa Francês
Federal University of Para, Brazil

Paulo Gonçalves
Polytechnic Institute of Castelo Branco, Portugal

Nallappan Gunasekaran
Toyoto Technological Institute, Japan

Shaidah Jusoh
Princess Sumaya University for Technology, Jordan

Fakhri Karray
University of Waterloo, Canada

Lei Lei
University of Guelph, Canada

Ming Liu
The Hong Kong University of Science and Technology, China

Chaomin Luo
Mississippi State University, USA

Jianjun Ni
Hohai University, China

Danilo Pelusi
University of Teramo, Italy

Tao Ren
Chengdu University of Technology, China

Gerasimos Rigatos
Industrial Systems Institute, Greece

Ricardo Sanz
Universidad Politécnica de Madrid, Spain

Jinhua She
Tokyo University of Technology, Japan

Farhad Soleimanian Gharehchopogh
Islamic Azad University, Iran

Jindong Tan
University of Tennessee, USA

Ying Wang
Kennesaw State University, USA

Xin Xu
National University of Defense Technology, China

Wen Yu
National Polytechnic Institute, Mexico

Anmin Zhu
Shenzhen University, China

Daqi Zhu
Shanghai Maritime University, China

Hao Zhang
Tongji University, China

Youth Editorial Board Members

Laith Abualigah
Amman Arab University, Jordan

Sawal Hamid Md Ali
Universiti Kebangsaan Malaysia, Malaysia

Hongtian Chen
University of Alberta, Canada

Changxin Gao
Huazhong University of Science and Technology, China

Jianye Hao
Tianjin University, China

Manju Khari
Jawaharlal Nehru University, India

Haitao Liu
Tianjin University, China

Anh-Tu Nguyen
Université Polytechnique Hauts-de-France, France

Farhad Pourpanah
Shenzhen University, China

Sangram Redkar
Arizona State University, USA

Bing Sun
Shanghai Maritime University, China

Xiaoqiang Sun
Jiangsu University, China

Yuxiang Sun
The Hong Kong Polytechnic University, China

Donglin Wang
Westlake University, China

Zhongkui Wang
Ritsumeikan University, Japan

Guanglei Wu
Dalian University of Technology, China

Yu Xue
Nanjing University of Information Science and Technology, China

Guoxian Yu
Shandong University, China

Zhiwei Yu
Nanjing University of Aeronautics and Astronautics, China

GENERAL INFORMATION

About the Journal

Intelligence & Robotics (*IR*), ISSN 2770-3541 (Online), publishes top-quality unpublished original technical and non-technical application-focused articles on intelligence and robotics, particularly on the interdisciplinary areas of intelligence and robotics. The Journal seeks to publish articles that deal with the theory, design, and applications of intelligence and robotics, ranging from software to hardware. The scope of the Journal includes, but is not limited to, biological, bio-inspired, and artificial intelligence; neural networks, fuzzy systems, and evolutionary algorithms; sensing, multi-sensor fusion, localization, data analysis, modeling, planning, and control for various mobile, aerial, and underwater robotic systems; and robot cooperation, teleoperation and human-machine interactions. The Journal would be interested in distributing development and maintenance of real-world intelligent and robotic systems by multidisciplinary teams of scientists and engineers.

Information for Authors

Manuscripts should be prepared in accordance with Author Instructions.
Please check https://intellrobot.com/pages/view/author_instructions for details.
All manuscripts should be submitted online at https://oaemesas.com/login?JournalId=ir.

Copyright

Articles in *IR* are published under a Creative Commons Attribution 4.0 International (CC BY 4.0). The CC BY 4.0 allows for maximum dissemination and re-use of open access materials and is preferred by many research funding bodies. Under this license users are free to share (copy, distribute and transmit) and remix (adapt) the contribution for any purposes, even commercially, provided that the users appropriately acknowledge the original authors and the source.
Copyright is reserved by © The Author(s) 2021.

Permissions

For information on how to request permissions to reproduce articles/information from this journal, please visit www.intellrobot.com.

Disclaimer

The information and opinions presented in the journal reflect the views of the authors and not of the journal or its Editorial Board or the Publisher. Publication does not constitute endorsement by the journal. Neither the *IR* nor its publishers nor anyone else involved in creating, producing or delivering the *IR* or the materials contained therein, assumes any liability or responsibility for the accuracy, completeness, or usefulness of any information provided in the *IR*, nor shall they be liable for any direct, indirect, incidental, special, consequential or punitive damages arising out of the use of the *IR*. *IR*, nor its publishers, nor any other party involved in the preparation of material contained in the *IR* represents or warrants that the information contained herein is in every respect accurate or complete, and they are not responsible for any errors or omissions or for the results obtained from the use of such material. Readers are encouraged to confirm the information contained herein with other sources.

Published by

OAE Publishing Inc.
245 E Main Street Ste 107, Alhambra CA 91801, USA
Website: www.oaepublish.com

Contacts

E-mail: editorial@intellrobot.com
Website: www.intellrobot.com

Intelligence & Robotics

CONTENTS

Volume 1 / Issue 2 /

Research Article

On the elastodynamics of a five-axis lightweight anthropomorphic robotic arm 99
*Guanglei Wu**

Planning robotic agent actions using semantic knowledge for a home environment 116
Rodrigo Bernardo, João M. C. Sousa, Paulo J. S. Gonçalves*

Autonomous navigation in unknown environment using sliding mode SLAM and genetic algorithm 131
*Salvador Ortiz , Wen Yu**

Review

Rail track condition monitoring: a review on deep learning approaches 151

Albert Ji, Wai Lok Woo, Eugene Wai Leong Wong, Yang Thee Quek*

Wu. *Intell Robot* 2021;1(2):99-115
DOI: 10.20517/ir.2021.11

Intelligence & Robotics

Research Article

Open Access

On the elastodynamics of a five-axis lightweight anthropomorphic robotic arm

Guanglei Wu

School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, Liaoning, China.

Correspondence to: Dr. Guanglei Wu, School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, Liaoning, China. E-mail: gwu@dlut.edu.cn

How to cite this article: Wu GL. On the elastodynamics of a five-axis lightweight anthropomorphic robotic arm. *Intell Robot* 2021;1(2):99-115. http://dx.doi.org/10.20517/ir.2021.11

Received: 17 Sep 2021 **First Decision:** 22 Oct 2021 **Revised:** 12 Nov 2021 **Accepted:** 18 Nov 2021 **Published:** 1 Dec 2021

Academic Editors: Simon X. Yang, Howard Li **Copy Editor:** Xi-Jun Chen **Production Editor:** Xi-Jun Chen

Abstract

This paper presents elastodynamic modeling and analysis for a five-axis lightweight robotic arm. Natural frequencies are derived and visualized within the dexterous workspace to show the overall performances and compare them to the frequencies when the robotics is with payload. The comparison shows that the payload has a relatively small influence to the first- and second-order frequencies. Sensitivity analysis is conducted, and the system's frequency is more sensitive to the second joint stiffness than the others. Moreover, observations from the displacement response analysis reveal that the robotics produces linear elastic displacements of the same level between the loaded and unloaded working modes but larger rotational deflections under the loaded working condition. The main contribution of this work lies in that a systematic approach of elastodynamic analysis for serial robotic manipulators is formulated, where the arm gravity and external load are taken into account to investigate the dynamic behaviors of the robotic arms, *i.e.*, frequencies, sensitivity analysis, and displacement responses, under the loaded mode.

Keywords: Lightweight robotic arm, elastodynamics, natural frequency, displacement response

1. INTRODUCTION

Lightweight robotic arms and anthropomorphic assistive robots with high payload capacity are desired for applications of industry and welfare, among other fields, such as assisted daily living[1–3], pick-and-place operations[4], *etc.* Pick-and-place robots are well suited for a static environment where the task is repeated and precise tolerances are demanded[5]. As a mechanical system, the dynamic characteristics of the robotic arm

© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

is of importance to account for the requirements of application, such as high precision, speed, and payload. Henceforth, higher natural frequencies and low elastic displacements of a robotic manipulator will allow higher operational speeds and working cycles for efficient productivity[6]. Natural frequencies indicate the condition in which a mechanism tends to vibrate[7,8]. Differing from a structure or element, the dynamic behavior of a mechanism usually heavily depends on its architecture and configurations[9]; thus, it is not a trivial task to characterize the robot dynamics throughout the workspace, which calls for the kineto-elastodynamic analysis to provide the fundamentals of the modeling, design and control.

The elastodynamic modeling and analysis of a robotic manipulator have been reported previously[10,11], and they are roughly grouped into two categories: lumped modeling[12–15] and distributed-flexibilities modeling[9,16–19]. In general, with lumped modeling it is simpler to model the elastodynamic equation with acceptable computational accuracy, while the latter provides a more accurate model but with the high-dimensional generalized coordinate space and more complex procedure[20]. The commonly used method to study the elasticity of the robotic manipulators is the virtual joint method (VJM) as it can provide acceptable computation accuracy that is close to that of finite element analysis (FEA)[21]. Besides, VJM can be time efficient. VJM is based on pseudo-rigid body models with "virtual joints"[22–25]. Generally, the link flexibilities and linear/torsional springs take into account the bending contributions to the mechanism[26–29]. The stiffness formulated in the above approaches is limited to a subspace defined by the degrees of freedom (dofs) of the manipulator end-effector. Pashkevich *et al.*[30] overcame this issue by introducing a full-mobility lumped-parameter model by localizing 6-dof virtual springs to the links' ends and/or joints. In these models, the stiffness matrix is calculated in an unloaded equilibrium configuration of a robotic manipulator. On the other hand, the external loads directly influence the manipulator equilibrium configuration and, consequently, may modify the static properties. The lightweight design of the robotics accordingly decreases the link structural stiffness; thus, the robot geometry change due to external loads should be considered[31–33]. Consequently, elastodynamics of the robotic manipulators is an important concern in their design and applications. Based on the matrix structural analysis, Cammarata *et al.*[9,34] proposed an algorithm to assemble the stiffness matrix to investigate the manipulators with lower kinematic pairs. In this manner, the overall robotic manipulator in-parallel architecture can be split into substructures for modeling the elastodynamics[35]. Wu *et al.*[36] analyzed and compared the stiffness and natural frequencies of a 3-dof parallel manipulator with/without a redundant leg, where the joint deformations are ignored in the stiffness modeling. The small-amplitude deformations of the active joints can be considered as parameter uncertainties in terms of small variations to be integrated into the dynamic model[37]. Briot and Khalil[14] used the Newton–Euler recursive approach to develop a general symbolic elastodynamic calculation model for flexible parallel robots. Taghvaeipour *et al.*[15] derived the posture-dependent stiffness matrix in the elastodynamic modeling by resorting to the generalized spring concept. The previous models were established in the nominal configurations; hence, the geometry changes of the manipulator in this work are considered in the elastodynamic modeling and analysis.

In this paper, the elastodynamic characteristics of a lightweight robotic arm are investigated. The arm gravity and external load are taken into account to derive the stiffness matrix. Isocontours of natural frequencies over the dexterous workspace are formulated and sensitivity analysis is conducted. The frequencies and displacement responses of the robotics with payload are analyzed and compared with the dynamic behaviors of the unloaded mode. The main contribution of this work lies in that a systematic approach of elastodynamic analysis for serial robotic manipulators is formulated, where the arm gravity and external load are taken into account to investigate the dynamic behaviors of the robotic arms, *i.e.*, frequencies, sensitivity analysis, and displacement responses, under the loaded mode.

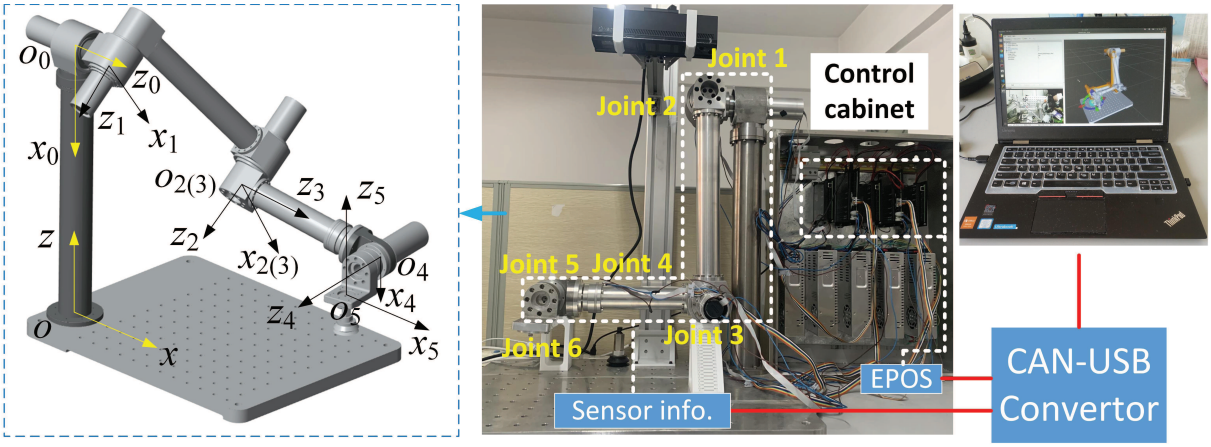


Figure 1. The 5-dof lightweight robotic arm and its coordinate systems [38].

Table 1. D–H parameters of the 5-dof robotic arm

Joint i	α_i	a_i [mm]	d_i [mm]	θ_i
1	$\pi/2$	0	250	θ_1
2	0	600	0	θ_2
3	$\pi/2$	0	0	θ_3
4	$-\pi/2$	0	600	θ_4
5	$\pi/2$	0	150	θ_5

2. KINEMATICS OF THE LIGHTWEIGHT ROBOTIC ARM

The lightweight robotic arm under study has five degrees of freedom (dof) [38], which adopts a modular design approach, as shown in Figure 1. The revolute joints are composed of CPU series gearboxes of Harmonic Drive and Maxon motor with gearhead to enhance the torque capabilities, except Joint 4 with geared motor. The actuators of joints are controlled by Maxon EPOS controllers. The Controller Area Network (CANopen) bus is adopted to build the communications between motors and controllers, and A CAN–USB interface is used to establish the communications between CANopen bus and the PC [38]. In accordance with the Denavit–Hartenberg (D–H) convention [39], the Cartesian coordinate systems are established accordingly.

2.1. Kinematics of robotic arm

Throughout this work, **i**, **j**, and **k** stand for the unit vectors of the x-axis, y-axis, and z-axis, respectively. The transformation matrix in forward kinematics of the end-effector in reference frame is expressed as

$$ {}^{0}\mathbf{A}_5 = \begin{bmatrix} \mathbf{R} & \mathbf{q} \\ \mathbf{0} & 1 \end{bmatrix} = \prod_{i=1}^{5} {}^{i-1}\mathbf{A}_i; \quad {}^{i-1}\mathbf{A}_i = \begin{bmatrix} {}^{i-1}\mathbf{R}_i & {}^{i-1}\mathbf{q}_i \\ \mathbf{0} & 1 \end{bmatrix} \tag{1} $$

with

$$ {}^{i-1}\mathbf{R}_i = \mathbf{R}(z_{i-1}, \theta_i)\mathbf{R}(x_i, \alpha_i) \tag{2a} $$

$$ {}^{i-1}\mathbf{q}_i = \begin{bmatrix} a_i \cos\alpha_i & a_i \sin\alpha_i & d_i \end{bmatrix}^T \tag{2b} $$

where D–H parameters are given in Table 1, and the inverse geometry problem for this robotics is well documented in the literature [8].

2.2. Kinematic jacobian matrix

The velocities between the joints and end-effector are mapped with the Kinematic Jacobian matrix

$$ \dot{\boldsymbol{\theta}} = \mathbf{J}^{-1}\mathbf{v}_{ef} \tag{3} $$

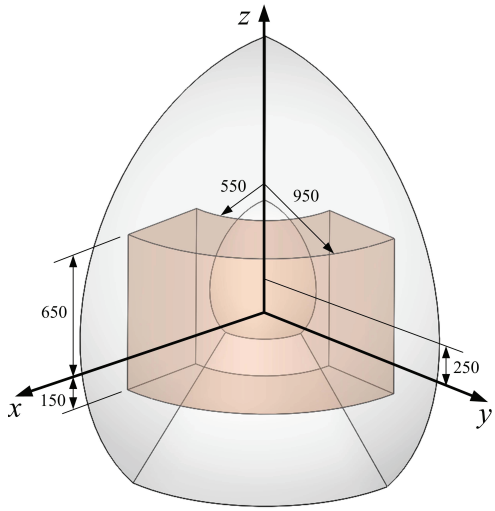


Figure 2. The quarter of the reachable and dexterous workspace (red volume) for the robotic arm.

where $\dot{\boldsymbol{\theta}} = \begin{bmatrix}\dot{\theta}_1 & \dot{\theta}_2 & \dot{\theta}_3 & \dot{\theta}_4 & \dot{\theta}_5\end{bmatrix}^T$ denotes the joint angular velocities and $\mathbf{v}_{ef} = \begin{bmatrix}\boldsymbol{\omega}^T & \dot{\mathbf{q}}^T\end{bmatrix}^T$ is the velocity of the end-effector. Moreover, $\mathbf{J}$ is the kinematic Jacobian matrix of the robotic arm[40], namely,

$$\mathbf{J} = \begin{bmatrix}\mathbf{j}_1 & \mathbf{j}_2 & \mathbf{j}_3 & \mathbf{j}_4 & \mathbf{j}_5\end{bmatrix} \quad \text{where} \quad \mathbf{j}_i = \begin{bmatrix}\mathbf{z}_{i-1} \\ \mathbf{p}_{i-1} \times \mathbf{z}_{i-1}\end{bmatrix} \tag{4}$$

with

$$\mathbf{z}_{i-1} = \mathbf{R}_{i-1}\mathbf{k}; \quad \mathbf{p}_{i-1} = \mathbf{q}_{i-1} - \mathbf{q} \tag{5}$$

where $\mathbf{R}_{i-1}$ and $\mathbf{q}_{i-1}$ denote the rotation matrix and position vector of the transformation matrix from the reference coordinate system to the $(i-1)$th coordinate system, respectively, which can be extracted from $\prod_{i=0}^{i-1} {}^{i-1}\mathbf{A}_i$ in Equation (1).

2.3. Dexterous workspace

The reachable workspace of the robotic arm can be visualized by considering the limitation of the joint displacements and link dimensions. To effectively perform the kinematic performance, a dexterous workspace is defined, throughout which the inverse of the condition number of the Jacobian matrix is greater than 0.2, namely $\kappa^{-1}(\mathbf{J}) \geq 0.2$. Since the Jacobian matrix of Equation (4) is not homogeneous, a characteristic length[41] is introduced to normalize the Jacobian matrix as follows:

$$\mathbf{j}_i' = \begin{bmatrix}\mathbf{z}_{i-1} \\ \mathbf{p}_{i-1} \times \mathbf{z}_{i-1}/L\end{bmatrix}; \quad L^2 = \frac{1}{5}\sum_{i=1}^{5} \|\mathbf{p}_{i-1} \times \mathbf{z}_{i-1}\|^2 \tag{6}$$

By constraining the condition number of the kinematic Jacobian matrix, a regular dexterous workspace is quarterly visualized in Figure 2.

3. ELASTODYNAMIC MODEL OF ROBOT

The elastodynamic modeling procedure pertains to the calculation of the stiffness and mass matrices of the manipulator, which is described in the following sections. Prior to the derivation of the elastodynamic model, the following assumptions are made:

- The actuator stiffness is considered as an 1-dof torsional spring, while the link is considered as cantilever with a 6-dof spatial spring located at the end but treated as rigid.

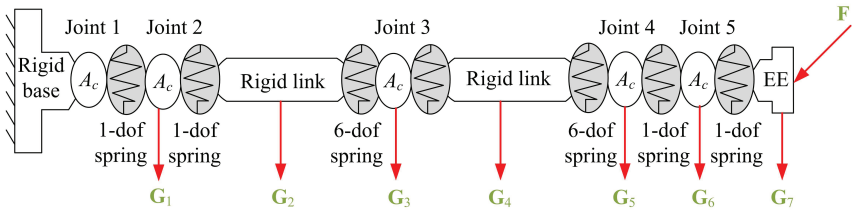


Figure 3. Virtual spring model of the 5-dof robotic arm with auxiliary loads, where A_c stands for the actuator and EE for end-effector.

- The centers of mass of the regular components are coincident with their geometric centers.
- The sum of moments of inertia of the actuators and Harmonic drivers are considered as lumped.

3.1. Stiffness matrix

To derive the elastodynamic equation of the robotic arm, the stiffness matrix is calculated with the virtual spring approach [42], based on the screw coordinates [43]. Hence, the component masses and external loads are taken into account to compute the Cartesian stiffness matrix. Figure 3 shows the VJM model of the robotic arm, where $\mathbf{G}_j$, $j = 1, 2, \ldots, 7$, stands for the gravity and $\mathbf{F}$ for the external loads.

Let $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ be the original and the deformed displacements of the virtual springs, respectively, following the principle of virtual work, *i.e.*, the work of the auxiliary forces is equal to the work of internal forces $\boldsymbol{\tau}_\theta$, namely,

$$\sum(\mathbf{G}_j^T \delta\mathbf{t}_j) + \mathbf{F}^T \delta\mathbf{t} = \boldsymbol{\tau}_\theta^T(\boldsymbol{\theta}' - \boldsymbol{\theta}) \tag{7}$$

where the virtual displacements $\delta\mathbf{t}_j$ and $\delta\mathbf{t}$ can be computed from the linearized geometrical model derived from $\delta\mathbf{t}_j = \mathbf{J}_j(\boldsymbol{\theta}' - \boldsymbol{\theta})$ and $\delta\mathbf{t} = \mathbf{J}_\theta(\boldsymbol{\theta}' - \boldsymbol{\theta})$, respectively, $\mathbf{J}_j$ and $\mathbf{J}_\theta$ being the Jacobians, namely,

$$\mathbf{J}_\theta = \begin{bmatrix} \mathbf{j}_1 & \mathbf{j}_2 & \mathbf{J}_u & \mathbf{j}_3 & \mathbf{J}_l & \mathbf{j}_4 & \mathbf{j}_5 \end{bmatrix} \in \mathbb{R}^{6\times 17} \tag{8}$$

$$\mathbf{J}_j = \mathbf{J}_\theta(:, 1:k) \tag{9}$$

where $\mathbf{J}_\theta(:, 1:k)$ stands for the first k columns in $\mathbf{J}_\theta$ and k stands for the total degrees of freedom of the virtual springs from the base to $\mathbf{G}_j$. Moreover,

$$\mathbf{J}_u = \begin{bmatrix} \mathbf{x}_1 & \mathbf{y}_1 & \mathbf{z}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{q}_2 \times \mathbf{x}_1 & \mathbf{q}_2 \times \mathbf{y}_1 & \mathbf{q}_2 \times \mathbf{z}_1 & \mathbf{x}_1 & \mathbf{y}_1 & \mathbf{z}_1 \end{bmatrix} \tag{10a}$$

$$\mathbf{J}_l = \begin{bmatrix} \mathbf{z}_3 & \mathbf{x}_3 & \mathbf{y}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{q}_4 \times \mathbf{z}_3 & \mathbf{q}_4 \times \mathbf{x}_3 & \mathbf{q}_4 \times \mathbf{y}_3 & \mathbf{z}_3 & \mathbf{x}_3 & \mathbf{y}_3 \end{bmatrix} \tag{10b}$$

Equation (7) is rewritten as

$$\sum(\mathbf{G}_j^T \mathbf{J}_j(\boldsymbol{\theta}' - \boldsymbol{\theta})) + \mathbf{F}^T \mathbf{J}_\theta(\boldsymbol{\theta}' - \boldsymbol{\theta}) = \boldsymbol{\tau}_\theta^T(\boldsymbol{\theta}' - \boldsymbol{\theta}) \tag{11}$$

consequently, the force equilibrium equation is derived as

$$\boldsymbol{\tau}_\theta = \sum(\mathbf{J}_j^T \mathbf{G}_j) + \mathbf{J}_\theta^T \mathbf{F} = \mathbf{J}_g^T \mathbf{G} + \mathbf{J}_\theta^T \mathbf{F} \tag{12}$$

with

$$\mathbf{J}_g = \begin{bmatrix} \mathbf{J}_1^T & \mathbf{J}_2^T & \ldots & \mathbf{J}_7^T \end{bmatrix}^T; \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_1^T & \mathbf{G}_2^T & \ldots & \mathbf{G}_7^T \end{bmatrix}^T \tag{13}$$

Assuming that $\mathbf{K}_\theta$ is the stiffness matrix in the joint space, with the linearized force–deflection relation, the equilibrium condition can be written as

$$\mathbf{J}_g^T \mathbf{G} + \mathbf{J}_\theta^T \mathbf{F} = \mathbf{K}_\theta(\boldsymbol{\theta}_i' - \boldsymbol{\theta}_i) \tag{14}$$

with

$$\mathbf{K}_\theta = \text{diag}\begin{bmatrix} K_{\text{act},1} & K_{\text{act},2} & \mathbf{K}_u & K_{\text{act},3} & \mathbf{K}_l & K_{\text{act},4} & K_{\text{act},5} \end{bmatrix} \tag{15}$$

where $K_{\text{act},i}$ is the actuation stiffness and $\mathbf{K}_u$ and $\mathbf{K}_l$ are the upper and lower link stiffness matrices, respectively.

To calculate the stiffness matrix of the loaded mode, a neighborhood in the loaded configuration in which the external loads and the joint location are supposed to be incremented by small values $\delta\mathbf{F}$ and $\delta\boldsymbol{\theta}$, which can still satisfy the equilibrium conditions, is considered, leading to

$$(\mathbf{J}_g + \delta\mathbf{J}_g)^T\mathbf{G} + (\mathbf{J}_\theta + \delta\mathbf{J}_\theta)^T(\mathbf{F} + \delta\mathbf{F}) = \mathbf{K}_\theta(\boldsymbol{\theta}' - \boldsymbol{\theta} + \delta\boldsymbol{\theta}) \tag{16}$$

and the linearized kinematic constraint

$$\delta\mathbf{t} = \mathbf{J}_\theta\delta\boldsymbol{\theta} \tag{17}$$

Based on Equation (14), expanding Equation (16) yields

$$\mathbf{H}_g^T \otimes \mathbf{G}\delta\boldsymbol{\theta} + \mathbf{J}_\theta^T\delta\mathbf{F} + \mathbf{H}_\theta^T \otimes \mathbf{F}\delta\boldsymbol{\theta} = \mathbf{K}_\theta\delta\boldsymbol{\theta} \tag{18}$$

where the symbol $\otimes$ represents the Kronecker product between matrices and $\mathbf{H}_g = \partial\mathbf{J}_g/\partial\boldsymbol{\theta}$, $\mathbf{H}_\theta = \partial\mathbf{J}_\theta/\partial\boldsymbol{\theta}$. Combining Equations (17) and (18), the stiffness model of the robotic manipulator is reduced to

$$\begin{bmatrix} \mathbf{0} & \mathbf{J}_\theta \\ \mathbf{J}_\theta^T & \mathbf{K}_F - \mathbf{K}_\theta \end{bmatrix} \begin{bmatrix} \delta\mathbf{F} \\ \delta\boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \delta\mathbf{t} \\ \mathbf{0} \end{bmatrix} \tag{19}$$

with

$$\mathbf{K}_F = \mathbf{H}_g^T \otimes \mathbf{G} + \mathbf{H}_\theta^T \otimes \mathbf{F} \tag{20}$$

From $\delta\mathbf{F} = \mathbf{K}\delta\mathbf{t}$, the Cartesian stiffness matrix $\mathbf{K}$ of the robotic arm is calculated as

$$\mathbf{K} = \left(\mathbf{J}_\theta\left(\mathbf{K}_\theta - \mathbf{K}_F\right)^{-1}\mathbf{J}_\theta^T\right)^{-1} \tag{21}$$

3.2. Mass matrix

The mass matrix can be derived from the expression of the system's kinetic energy, consisting of energies of the revolute joints, links, and end-effector. The energy of the five active joints are

$$E_J = \frac{1}{2}\left(\sum_{i=1}^{5} I_{\theta,i}\dot{\theta}^2 + \sum_{i=3}^{5} m_{\theta,i}\mathbf{v}_{\theta,i}^T\mathbf{v}_{\theta,i}\right) \tag{22}$$

with

$$\mathbf{v}_{\theta,3} = \mathbf{E}_3\dot{\boldsymbol{\theta}}; \quad \mathbf{v}_{\theta,i} = \mathbf{E}_{45}\dot{\boldsymbol{\theta}},\ i = 4,\ 5 \tag{23}$$

and

$$\mathbf{E}_3 = \begin{bmatrix} \mathbf{z}_0 \times \mathbf{q}_2 & \mathbf{z}_1 \times (\mathbf{q}_2 - \mathbf{q}_1) & \mathbf{0}_3 \end{bmatrix} \tag{24a}$$

$$\mathbf{E}_{45} = \begin{bmatrix} \mathbf{z}_0 \times \mathbf{q}_4 & \mathbf{z}_1 \times (\mathbf{q}_4 - \mathbf{q}_1) & \mathbf{z}_2 \times (\mathbf{q}_4 - \mathbf{q}_2) & \mathbf{0}_{3\times 2} \end{bmatrix} \tag{24b}$$

where $I_{\theta,i}$ is the moment of inertia of the ith joint, $m_{\theta,i}$ is the mass, and $\mathbf{v}_{\theta,i}$ is the velocity in the Cartesian space. Let $\mathbf{I}_\theta = \text{diag}[I_{\theta,1},\ I_{\theta,2},\ \ldots,\ I_{\theta,5}]$; then, Equation (22) can be written in a compact form, namely,

$$E_J = \frac{1}{2}\dot{\boldsymbol{\theta}}^T\mathbf{M}_J\dot{\boldsymbol{\theta}} \tag{25}$$

with

$$\mathbf{M}_J = \mathbf{I}_\theta + m_{\theta,3}\mathbf{E}_3^T\mathbf{E}_3 + (m_{\theta,4} + m_{\theta,5})\mathbf{E}_{45}^T\mathbf{E}_{45} \tag{26}$$

The kinetic energy of the upper/lower links and the wrist link can be expressed as

$$E_L = \frac{1}{2}\left(\mathbf{v}_u^T\mathbf{M}_u\mathbf{v}_u + \mathbf{v}_l^T\mathbf{M}_l\mathbf{v}_l + \mathbf{v}_w^T\mathbf{M}_w\mathbf{v}_w\right) \tag{27}$$

with

$$\mathbf{M}_u = \begin{bmatrix} \mathbf{R}_1\mathbf{I}_u\mathbf{R}_1^T & \mathbf{0} \\ \mathbf{0} & m_u\mathbf{1}_3 \end{bmatrix} \tag{28a}$$

$$\mathbf{M}_l = \begin{bmatrix} \mathbf{R}_3\mathbf{I}_l\mathbf{R}_3^T & \mathbf{0} \\ \mathbf{0} & m_l\mathbf{1}_3 \end{bmatrix} \tag{28b}$$

$$\mathbf{M}_w = \begin{bmatrix} \mathbf{R}_4\mathbf{I}_w\mathbf{R}_4^T & \mathbf{0} \\ \mathbf{0} & m_w\mathbf{1}_3 \end{bmatrix} \tag{28c}$$

where the subscripted $\mathbf{I}$, m, and $\mathbf{v}$ stand for the moment of inertia, mass, and velocities in the Cartesian space, respectively, and

$$\mathbf{v}_u = \mathbf{E}_u\dot{\boldsymbol{\theta}}; \quad \mathbf{v}_l = \mathbf{E}_l\dot{\boldsymbol{\theta}}; \quad \mathbf{v}_w = \mathbf{E}_w\dot{\boldsymbol{\theta}} \tag{29}$$

with

$$\mathbf{E}_u = \begin{bmatrix} \mathbf{z}_0 & \mathbf{z}_1 & \mathbf{0}_3 \\ \mathbf{z}_0 \times \mathbf{q}_u & \mathbf{z}_1 \times (\mathbf{q}_u - \mathbf{q}_1) & \mathbf{0}_3 \end{bmatrix} \tag{30a}$$

$$\mathbf{E}_l = \begin{bmatrix} \mathbf{z}_0 & \mathbf{z}_1 & \mathbf{z}_2 & \mathbf{0}_{3\times2} \\ \mathbf{z}_0 \times \mathbf{q}_l & \mathbf{z}_1 \times (\mathbf{q}_l - \mathbf{q}_1) & \mathbf{z}_2 \times (\mathbf{q}_l - \mathbf{q}_2) & \mathbf{0}_{3\times2} \end{bmatrix} \tag{30b}$$

$$\mathbf{E}_w = \begin{bmatrix} \mathbf{z}_0 & \mathbf{z}_1 & \mathbf{z}_2 & \mathbf{z}_3 & \mathbf{0}_{3\times1} \\ \mathbf{z}_0 \times \mathbf{q}_4 & \mathbf{z}_1 \times (\mathbf{q}_4 - \mathbf{q}_1) & \mathbf{z}_2 \times (\mathbf{q}_4 - \mathbf{q}_2) & \mathbf{z}_3 \times (\mathbf{q}_4 - \mathbf{q}_2) & \mathbf{0}_{3\times1} \end{bmatrix} \tag{30c}$$

where $\mathbf{q}_u$ and $\mathbf{q}_l$ are the position vector of the centers of the mass of the upper and lower links, respectively. Equation (27) can be cast in a matrix form as follows:

$$E_L = \frac{1}{2}\dot{\boldsymbol{\theta}}^T\mathbf{M}_L\dot{\boldsymbol{\theta}} \tag{31}$$

with

$$\mathbf{M}_L = \mathbf{E}_u^T\mathbf{M}_u\mathbf{E}_u + \mathbf{E}_l^T\mathbf{M}_l\mathbf{E}_l + \mathbf{E}_w^T\mathbf{M}_w\mathbf{E}_w \tag{32}$$

Similarly, the kinetic energy of the end-effector can be obtained as

$$E_E = \frac{1}{2}\mathbf{v}_{ef}^T\mathbf{M}_E\mathbf{v}_{ef}; \quad \mathbf{M}_E = \begin{bmatrix} \mathbf{R}\mathbf{I}_e\mathbf{R}^T & \mathbf{0}_3 \\ \mathbf{0}_3 & m_e\mathbf{1}_3 \end{bmatrix} \tag{33}$$

where $\mathbf{I}_e$ is the moment of inertia of the end-effector and m_e is the mass.

From the total kinetic energy of the robotic arm $E = E_J + E_L + E_E$, the mass matrix $\mathbf{M}$ for the robotic arm can be expressed as

$$\mathbf{M} = \mathbf{M}_E + \mathbf{J}^{-T}(\mathbf{M}_J + \mathbf{M}_L)\mathbf{J}^{-1} \tag{34}$$

3.3. Dynamic equation and analysis

The dynamic equation of the robotic arm can be formulated as

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{f} - \mathbf{M}\dot{\mathbf{v}}_{ef} = \mathbf{F} \tag{35}$$

where $\mathbf{C}$ is the damping matrix, $\mathbf{F}$ is the resultant force, and $\mathbf{u}$ and $\ddot{\mathbf{u}}$ are the elastic displacement and acceleration, respectively. Since damping can only slightly influence the natural frequency and mode of free vibrations,

the damp can be ignored to determine the natural frequencies. Simplification of Equation (35) results in the linearized elastodynamic equation below

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{K}\mathbf{u} = 0 \tag{36}$$

The rigidity of the system may be represented by the natural frequency. The higher is the frequency, the higher is the stiffness. From Equation (36), we get

$$\det(-\omega^2\mathbf{M} + \mathbf{K}) = 0 \tag{37}$$

where $f = \omega/2\pi$ denotes the natural frequency.

The displacement response analysis can be carried out from Equation (35) based on the initial conditions

$$\mathbf{u}_0 = \mathbf{u}(0); \quad \dot{\mathbf{u}}_0 = \dot{\mathbf{u}}(0) \tag{38}$$

Here, the damping ratios are set to $\varsigma = 6\%$ according to the manipulator structure. From Equation (37), the displacement vector $\mathbf{u}$ can be represented in terms of the modal contributions, namely,

$$\mathbf{u} = \mathbf{Q}\boldsymbol{\eta} \tag{39}$$

where $\mathbf{Q}$ and $\boldsymbol{\eta}$ are the modal matrix and the vector of the displacements in each mode, respectively. Consequently, Equation (35) can be rewritten as

$$\ddot{\boldsymbol{\eta}} + \boldsymbol{\Phi}\dot{\boldsymbol{\eta}} + \boldsymbol{\Omega}\boldsymbol{\eta} = \mathbf{f}_d \tag{40}$$

with

$$\boldsymbol{\Phi} = \mathbf{Q}^T\mathbf{C}\mathbf{Q} = \mathrm{diag}\begin{bmatrix}2\varsigma\omega_1 & 2\varsigma\omega_2 & \dots & 2\varsigma\omega_6\end{bmatrix} \tag{41a}$$

$$\boldsymbol{\Omega} = \mathbf{Q}^T\mathbf{K}\mathbf{Q} = \mathrm{diag}\begin{bmatrix}\omega_1^2 & \omega_2^2 & \dots & \omega_6^2\end{bmatrix} \tag{41b}$$

$$\mathbf{f}_d = \mathbf{Q}^T\mathbf{F} \tag{41c}$$

Since the mass and stiffness matrices in Equation (35) are time-varying, the common way to solve such a problem is to divide the motion period into extremely short intervals, where the stiffness and mass matrices are considered as constant in each interval. Let T denote the complete motion period that is divided into N intervals, namely, $\Delta t = T/N$. In the nth time interval $\tau \in [t_{n-1}, t_n]$, the equation of motion in the ith mode is expressed as

$$\ddot{\eta}_i + 2\varsigma\omega_i\dot{\eta}_i + \omega_i^2\eta_i = f_{di} \tag{42}$$

Thus, the ith mode contributes to the displacement response[44] is

$$\begin{aligned}\eta_i(t_n) = {} & e^{\varsigma\omega_i\Delta t}\left(\cos\omega_{di}\Delta t + \frac{\varsigma}{\sqrt{1-\varsigma^2}}\sin\omega_{di}\Delta t\right)\eta_i(t_{n-1}) \\ & + \frac{1}{\omega_{di}}\int_{t_{n-1}}^{t_n} f_{di}(\tau)e^{-\varsigma\omega_i(t_n-\tau)}\sin\omega_{di}(t_n-\tau)\mathrm{d}\tau \\ & + \left(\frac{1}{\omega_{di}}e^{\varsigma\omega_i\Delta t}\sin\omega_{di}\Delta t\right)\dot{\eta}_i(t_{n-1})\end{aligned} \tag{43}$$

where

$$\omega_{di} = \omega_i\sqrt{1-\varsigma^2} \tag{44}$$

Table 2. Mass and moment of inertia of the active joints

Joint i	1	2	3	4	5
$I_{\theta,i}$ [kg · mm^2]	0.0210	0.0002	0.0001	0.0001	0.0002
$m_{\theta,i}$ [kg]	–	2.2272	1.8196	2.2442	2.0053

Table 3. The properties of the links and end-effector

Links	Mass [kg]	Moment of inerita [kg · cm^2]
upper link	$m_u = 4.7995$	$\mathbf{I}_u = \mathrm{diag}[1.1884,\ 25.0670,\ 24.4940]$
lower link	$m_l = 1.7795$	$\mathbf{I}_l = \mathrm{diag}[4.0802,\ 4.0861,\ 0.2345]$
wrist link	–	$\mathbf{I}_w = \mathrm{diag}[0.5556,\ 0.9154,\ 0.6119]$
end-effector	$m_e = 1.2961$	$\mathbf{I}_e = \mathrm{diag}[0.4563,\ 0.4382,\ 0.2347]$

Differentiating Equation (43) with respect to time leads to

$$\dot{\eta}_i(t_n) = \dot{\eta}_{i,1}(t_n) + \dot{\eta}_{i,2}(t_n) + \dot{\eta}_{i,3}(t_n) \tag{45}$$

with

$$\dot{\eta}_{i,1}(t_n) = e^{\varsigma\omega_i\Delta t}\left(\frac{2\varsigma^2-1}{\sqrt{1-\varsigma^2}}\omega_i \sin\omega_{di}\Delta t + \frac{2\varsigma}{\sqrt{1-\varsigma^2}}\omega_{di}\cos\omega_{di}\Delta t\right)\eta_i(t_{n-1}) \tag{46a}$$

$$\dot{\eta}_{i,2}(t_n) = \frac{1}{\omega_{di}}e^{\varsigma\omega_i\Delta t}\left(\varsigma\omega_i \sin\omega_{di}\Delta t + \omega_{di}\cos\omega_{di}\Delta t\right)\dot{\eta}_i(t_{n-1}) \tag{46b}$$

$$\dot{\eta}_{i,3}(t_n) = \frac{1}{\omega_{di}}\int_{t_{n-1}}^{t_n} f_{di}(\tau)e^{-\varsigma\omega_i(t_n-\tau)}\left(\varsigma\omega_i \sin\omega_{di}(t_n-\tau) - \omega_{di}\cos\omega_{di}(t_n-\tau)\right)\mathrm{d}\tau \tag{46c}$$

Hence, $\eta_i(t_n)$ and $\dot{\eta}_i(t_n)$ can be solved as long as $\eta_i(t_{n-1})$ and $\dot{\eta}_i(t_{n-1})$ are given, and

$$\eta_i(0) = \mathbf{e}_i^T\mathbf{M}\mathbf{u}(0); \qquad \dot{\eta}_i(0) = \mathbf{e}_i^T\mathbf{M}\dot{\mathbf{u}}(0) \tag{47}$$

where $\mathbf{e}_i$ is the ith column of the modal matrix. The total displacement response is calculated by the following addition

$$\mathbf{u}(t_n) = \sum_{i=1}^{6}\eta_i(t_n)\mathbf{e}_i(t_n) \tag{48}$$

Consequently, the natural frequency and displacement response can be obtained with numerical calculations.

4. NUMERICAL SIMULATION

Elastodynamic characteristics of the robotic arm are investigated in this section. The properties of the robotics components are listed in Tables 2 and 3, respectively. Moreover, according to the output shaft of the gearbox, the actuation stiffnesses are calculated and set to $K_{act,i} = 2 \cdot 10^4$ Nm/rad, $i = 1, \ldots, 5$, and the link stiffness matrices given in Appendix A are derived by means of FEA with ANSYS[45]. The numerical simulation was carried out with Matlab.

4.1. Natural frequency

To effectively measure the overall performance of the robotic arm, the distributions of natural frequencies over the dexterous workspace in Figure 2 are visualized, as displayed in Figures 4 and 5.

Let the end-effector orientation follow the ZXZ Euler convention; the distributions of the first- and second-order natural frequencies over workspace are displayed in Figures 4 and 5 when the end-effector remains

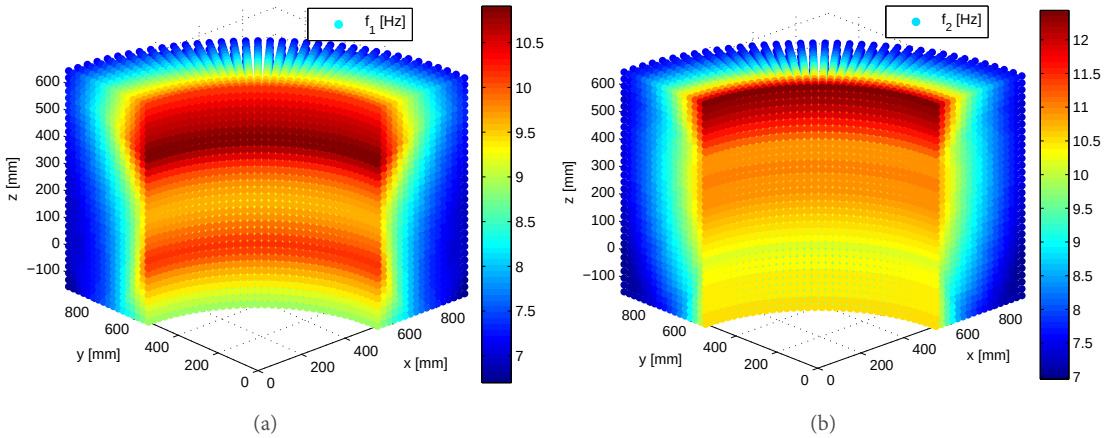


Figure 4. The natural frequency with constant-orientation $[0, 0, 0]$ (in unit of rad): (a) first order; (b) second order. (The color bar stands for the numerical value of the term in the legend, which is applicable to Figs. 5 to 7.)

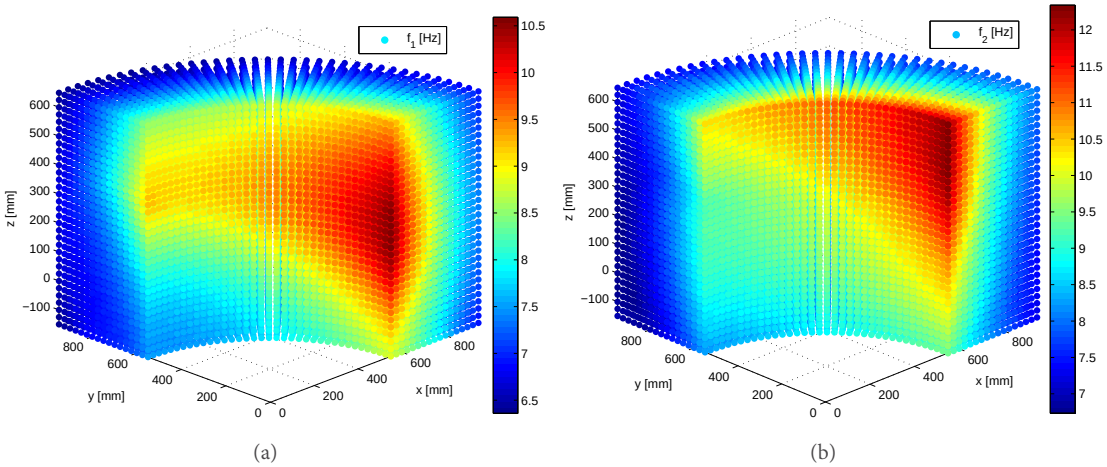


Figure 5. The natural frequency with constant-orientation $[0, \pi/2\, 0]$: (a) first order; and (b) second order.

vertical and horizontal, respectively. It can be observed that the nonsymmetric distributions of the natural frequencies in Figure 5 are different from the symmetric ones in Figure 4. This is because the robot configurations are not axisymmetric about the vertical direction with the vertical end-effector, leading to different inverse kinematic solutions of such a 5-dof robotic arm, which are different from the axisymmetric robot configurations with horizontal end-effector. As the mass and stiffness matrices of the robot are configuration dependent, non-symmetric distributions of natural frequencies in Figure 5 occur. These two figures show that the first two orders of natural frequencies increase with the increasing z coordinates but with decreased x and y coordinates, namely both the first and second frequencies increases from the workspace boundaries to the origin of the global coordinate systems. As displayed in Figure 4, when the end-effector remains vertical, the natural frequencies have the same varying trend in any vertical cross-section of the workspace. By contrast, the first- and second-order frequencies become smaller counterclockwise within the workspace when the end-effector is in the horizontal configuration, as shown in Figure 5. Moreover, it is found that the differences among the frequencies of the manipulator in different configurations are not so large, which means that the robotic arm has close frequencies inside the overall workspace.

4.2. Sensitivity analysis

Sensitivity analysis can be used to evaluate the influence of the geometric parameters and design variables to the manipulator performances. Based on the elastodynamic equation, there exists

$$(-\omega_i^2\mathbf{M}+\mathbf{K})\mathbf{e}_i = 0 \tag{49}$$

Upon differentiation of Equation (49), the derivative equation with respect to a variable δ is obtained as follows:

$$(-2\omega_i\frac{\partial\omega_i}{\partial\delta}\mathbf{M}-\omega_i^2\frac{\partial\mathbf{M}}{\partial\delta}+\frac{\partial\mathbf{K}}{\partial\delta})\mathbf{e}_i+(-\omega_i^2\mathbf{M}+\mathbf{K})\frac{\partial\mathbf{e}_i}{\partial\delta} = 0 \tag{50}$$

Taking the dot-product on both sides of Equation (50) yields

$$\mathbf{e}_i^T(-2\omega_i\frac{\partial\omega_i}{\partial\delta}\mathbf{M}-\omega_i^2\frac{\partial\mathbf{M}}{\partial\delta}+\frac{\partial\mathbf{K}}{\partial\delta})\mathbf{e}_i+\mathbf{e}_i^T(-\omega_i^2\mathbf{M}+\mathbf{K})\frac{\partial\mathbf{e}_i}{\partial\delta} = 0 \tag{51}$$

From

$$\mathbf{e}_i^T\mathbf{M}\mathbf{e}_i = 1; \quad \mathbf{e}_i^T(-\omega_i^2\mathbf{M}+\mathbf{K}) = \left((-\omega_i^2\mathbf{M}+\mathbf{K})\mathbf{e}_i\right)^T = 0 \tag{52}$$

we have

$$-2\omega_i\frac{\partial\omega_i}{\partial\delta}-\omega_i^2\mathbf{e}_i^T\frac{\partial\mathbf{M}}{\partial\delta}\mathbf{e}_i+\mathbf{e}_i^T\frac{\partial\mathbf{K}}{\partial\delta}\mathbf{e}_i = 0 \tag{53}$$

or

$$\frac{\partial\omega_i}{\partial\delta} = -\frac{1}{2\omega_i}\left(-\omega_i^2\mathbf{e}_i^T\frac{\partial\mathbf{M}}{\partial\delta}\mathbf{e}_i+\mathbf{e}_i^T\frac{\partial\mathbf{K}}{\partial\delta}\mathbf{e}_i\right) \tag{54}$$

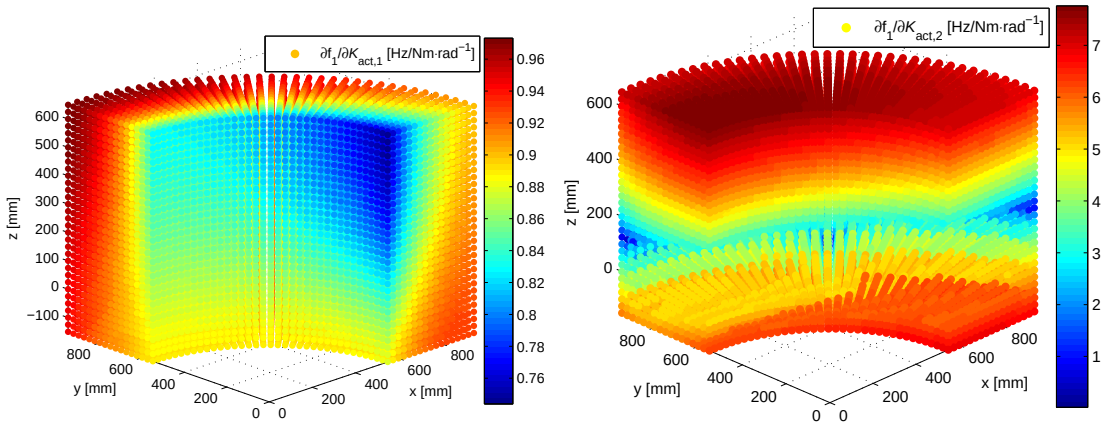
Figure 6 illustrates the sensitivity of the first-order natural frequency to the first two active joints with constant orientation $[0, \pi/2, 0]$. It is found that the first-order natural frequency is much more sensitive to the second joint, particularly in the upper and lower workspace regions, which implies that the robot's dynamic performance can be improved by replacing the second joint with a stiffer actuator. It is noted that the distributions of sensitivity coefficients are not symmetric, which is because the robot configurations are not axisymmetric about the vertical direction when the robot end-effector moves with some constant orientations, since the robot under study is a 5-dof robotic arm. Moreover, if a payload with more mass were exerted to the robot, it could be predicted that the sensitivity coefficients will be increased with very tiny varying trends, compared to the present results.

4.3. Dynamic analysis of loaded system

With the payload 5 kg applied to the end-effector of the robotic arm, they constitute a new dynamic system and the solved frequencies with constant-orientation $[0, \pi/2, 0]$ are illustrated in Figure 7, from which it is observed that the frequencies of the loaded robotic system decrease about 20% compared to Figure 5. Table 4 lists the average frequencies[46] within the constant-orientation workspace defined by

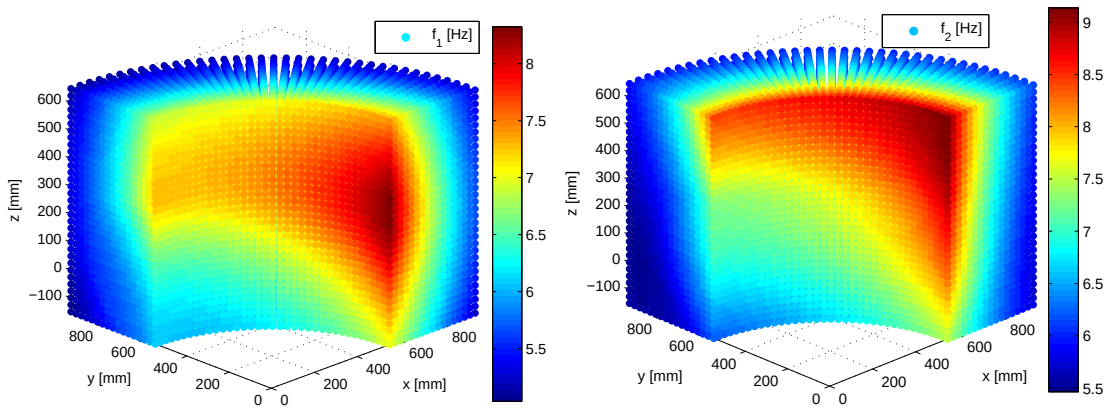
$$\bar{f}_i = \frac{\int f_i \mathrm{d}\Omega}{\int \mathrm{d}\Omega} \tag{55}$$

where Ω stands for the workspace volume. Different from the traditional industrial robots with low frequencies, the high order frequencies have large values to make the manipulator achieve high-speed motion. Compared to the average natural frequencies, the frequencies of the robotics with payload reduce 10%–40% for the six orders of frequencies. From the view of kineto-elastodynamic characteristics, the difference between the frequency of the loaded system and its natural frequency could be a consideration in the design of the mechanical system, where the smaller difference implies higher rigidity and higher payload capability.



(a) (b)

Figure 6. Sensitivities of the first-order natural frequency to the joint stiffness: (a) Joint 1; and (b) Joint 2.



(a) (b)

Figure 7. The frequencies with payload at constant-orientation $[0, \pi/2, 0]$: (a) first order; and (b) second order.

Table 4. The mean frequency (Hz) within the dexterous workspace

	$\bar{f}_1$	$\bar{f}_2$	$\bar{f}_3$	$\bar{f}_4$	$\bar{f}_5$	$\bar{f}_6$
Natural frequency	7.9404	8.9579	19.0263	84.9292	136.1418	300.6579
Frequency with payload	5.9538	6.5358	15.2772	55.8998	86.6291	276.5956

Assuming that the motion of the robotic arm follows the trajectory (unit: mm) defined by

$$\begin{aligned} x &= 750 + 750(\cos \pi\tau - 1) \\ y &= 750(1 - \cos \pi\tau) \\ z &= 600(1 - \cos \pi\tau) \end{aligned} \tag{56}$$

where the end-effector keeps constant-orientation $[0, \pi, 0]$ and the motion period $T = 0.5$ s is divided into 1024 intervals, Figure 8 shows the displacement responses of the end-effector, from which it is seen that the linear elastic displacement responses are close, whenever the robotic arm is under loaded and unloaded working modes. The angular displacements of the end-effector generate relatively large differences. The largest deformations appear around 0.3 s where the end-effector is located in the middle layer of the workspace, approximately $z = 250$ mm.

Figure 9 shows the comparison of the joint angular displacements between the numerical simulation and ex-

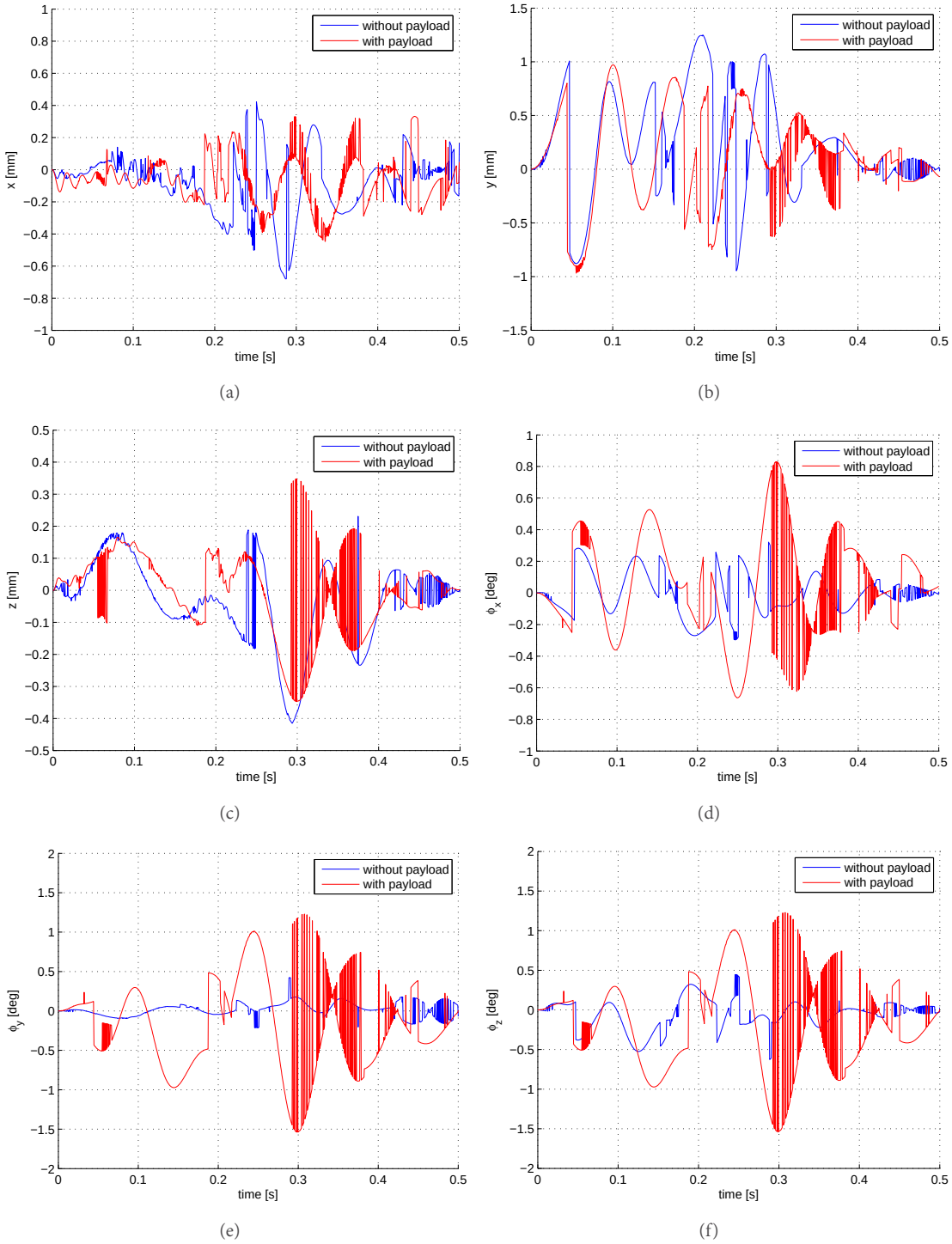


Figure 8. Displacement responses of the end-effector: (a) x direction; (b) y direction; (c) z direction; (d) ϕ_x direction; (e) ϕ_y direction; and (f) ϕ_z direction.

perimental measurements along previous trajectory, where the experimental data are read from the motor encoders. Due to the frictions and time-varying disturbance in the joints, the experimental curve profiles have more fluctuations and larger vibration amplitudes than the simulation ones. On the other hand, the comparison shows that the differences between these two curves are small, thus, the built analytical model can be acceptable for dynamic analysis of the robots.

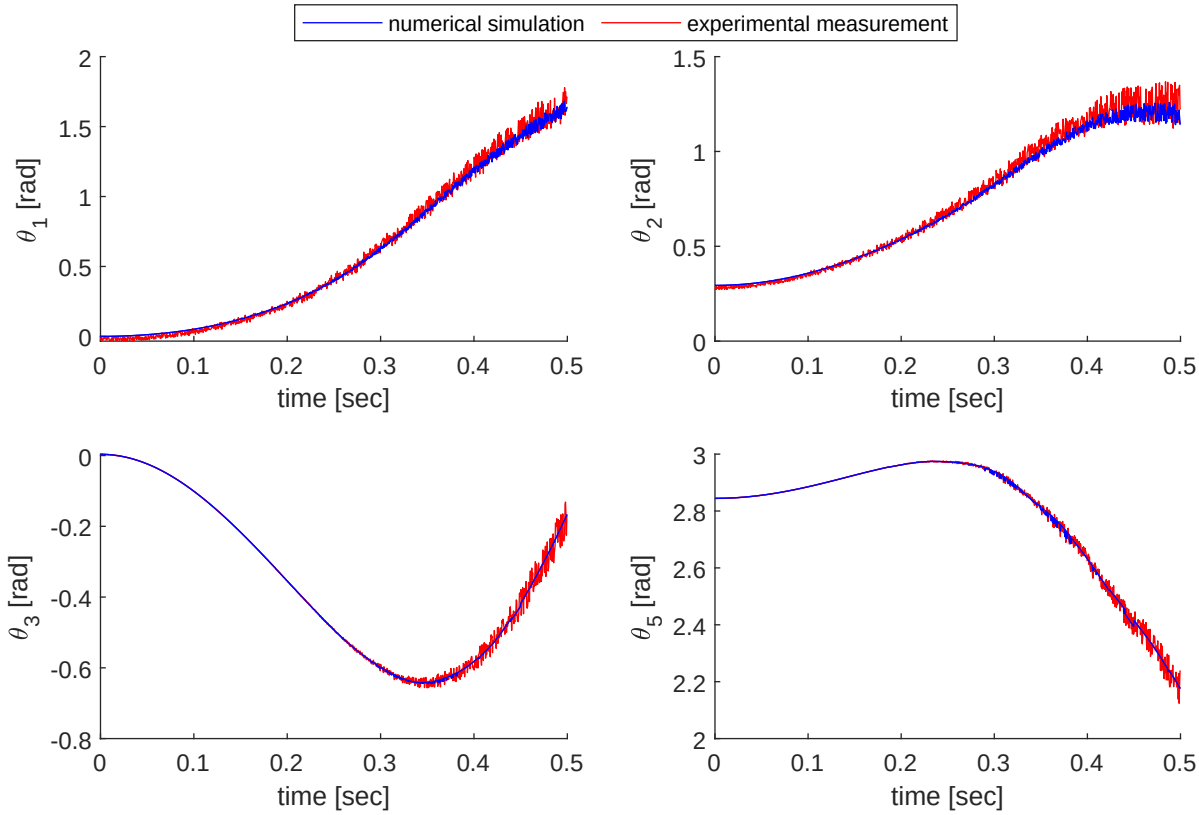


Figure 9. Comparison of the joint angular displacements under loaded mode.

5. CONCLUSION

This paper presents the elastodynamic characteristics of a 5-dof lightweight robotic arm. The main contribution is that a systematic approach of elastodynamic analysis for serial robotic manipulators is formulated, where the arm gravity and external load are taken into account to investigate the dynamic behaviors of the robotic arms, *i.e.*, frequencies, sensitivity analysis, and displacement responses, with auxiliary payloads exerted to the robot. The modeling in this work eases the evaluation of elastodynamics of the manipulator at a large number of postures as the elastodynamic aspect is usually time-consuming. As the mass and stiffness matrices are posture dependent, the proposed method can effectively provide a symbolic calculation and achieve the modal analysis along an operating trajectory. Moreover, such a model can compute the additional mass or evaluate the influence of an isolator to the system more precisely to eliminate/reduce vibration in the vibration control. The developed model can be used in either performance evaluation or design optimization.

The frequencies of the loaded robotics are visualized within the representative workspace regions to show the overall dynamic performance and compare them with the natural frequencies. The comparison reveals that the studied robot keeps relatively high rigidity with high payload ratio. It is found from sensitivity analysis that the natural frequency can effectively increase by improving the second joint stiffness. Based on the displacement responses analysis, the payload has a slight influence on the translational elastic displacements of this robotic system, although it leads to reduced frequencies, while the effect on the rotation deflections cannot be ignored. In the future, the developed model will be integrated into its control system and an optimum redesign of the robotics will be conducted.

DECLARATIONS

Authors' contributions
The author contributed solely to the article.

Availability of data and materials
Not applicable.

Financial support and sponsorship
This work was supported by Natural Science Foundation of Liaoning Province (Grant No. 20180520028).

Conflicts of interest
The author declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright
© The Author(s) 2021.

REFERENCES

1. Ivlev O, Martens C, Graeser A. Rehabilitation Robots FRIEND-I and FRIEND-II with the dexterous lightweight manipulator. *Technology and Disability* 2005;17:111–23.
2. Bien Z, Chung MJ, Chang PH, Kwon DS. Integration of a Rehabilitation Robotic System (KARES II) with Human-Friendly Man-Machine Interaction Units. *Auton Robot* 2004;16:165–91.
3. Mahoney RM. The Raptor wheelchair robot system. In: Mokhtari M, editor. Integration of Assistive Technology in the Information Age. IOS press; 2001. pp. 135–41.
4. Universal Robots. Available from https://www.universal-robots.com/GB/Cases.aspx.
5. Wu G, Shen H. In: Introduction. Singapore: Springer Singapore; 2021. pp. 1–15.
6. Hoevenaars AG, Krut S, Herder JL. Jacobian-based natural frequency analysis of parallel manipulators. *Mech Mach Theory* 2020;148:103775.
7. Briot S, Pashkevich A, Chablat D. On the optimal design of parallel robots taking into account their deformation and natural frequencies. In: ASME IDETC & CIE Conf.. vol. DETC2009-86230. San Diego, California, USA; 2009. pp. 367–76.
8. Siciliano B, Khatib O. Springer Handbook of Robotics. Springer; 2016.
9. Cammarata A, Condorelli D, Sinatra R. An algorithm to study the elastodynamics of parallel kinematic machines with lower kinematic pairs. *ASME J Mech Robot* 2013;5:011004.
10. Dwivedy SK, Eberhard P. Dynamic analysis of flexible manipulators, a literature review. *Mech Mach Theory* 2006;41:749–77.
11. Briot S, Khalil W. Dynamics of Parallel Robots. vol. 35 of Mechanisms and Machine Science. Ceccarelli M, editor. Springer International Publishing AG Switzerland; 2015.
12. Khalil W, Gautier M. Modeling of mechanical systems with lumped elasticity. In: Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065). vol. 4. San Francisco, CA, USA: IEEE; 2000. pp. 3964–69.
13. Wittbrodt E, Adamiec-Wójcik I, Wojciech S. Dynamics of Flexible Multibody Systems: Rigid Finite Element Method. Foundations of Engineering Mechanics. Springer Science & Business Media; 2007.
14. Briot S, Khalil W. Recursive and symbolic calculation of the elastodynamic model of flexible parallel robots. *Int J Robot Res* 2014;33:469–83.
15. Taghvaeipour A, Angeles J, Lessard L. Elastodynamics of a two-limb Schönflies motion generator. *Proc Ins Mech Eng Part C J Mech Eng Sci* 2015;229:751–64.
16. Boyer F, Coiffet P. Symbolic modeling of a flexible manipulator via assembling of its generalized Newton–Euler model. *Mech Mach Theory* 1996;31:45–56.
17. Rognant M, Courteille E, Maurine P. A systematic procedure for the elastodynamic modeling and identification of robot manipulators. *IEEE Trans Robot* 2010;26:1085–93.
18. Bauchau OA. Flexible Multibody Dynamics. vol. 176 of Solid Mechanics and Its Applications. Springer Science & Business Media;

2010.
19. de Jalon JG, Bayo E. Kinematic and Dynamic Simulation of Multibody Systems: the Real-time Challenge. Springer Science & Business Media; 2012.
20. Wu G, Shen H. Parallel PnP Robots. vol. 7 of Research on Intelligent Manufacturing. Ding H, Sun R, editors. Springer, Singapore; 2021.
21. Ku DM, Chen LW. Kineto-elastodynamic vibration analysis of robot manipulators by the finite element method. *Comput Struct* 1990;37:309–17.
22. Salisbury JK. Active stiffness control of a manipulator in cartesian coordinates. In: 1980 19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes. Albuquerque, NM, USA; 1980. pp. 95–100.
23. Gosselin C. Stiffness mapping for parallel manipulators. *IEEE Trans Robot Autom* 1990;6:377–82.
24. Wittbrodt E, Adamiec-Wójcik I, Wojciech S. Dynamics of Flexible Multibody Systems. Springer; 2006.
25. Quennouelle C, Gosselin C. Kinematostatic modeling of compliant parallel mechanisms. *Meccanica* 2011;46:155–69.
26. El-Khasawneh BS, Ferreira PM. Computation of stiffness and stiffness bounds for parallel link manipulators. *Int J Mach Tool Manuf* 1999;39:321–42.
27. Gosselin CM, Zhang D. Stiffness analysis of parallel mechanisms using a lumped model. *Int J Robot Autom* 2002;17:17–27.
28. Dai J, Ding X. Compliance analysis of a three-legged rigidly-connected platform device. *ASME J Mech Des* 2006;128:755–64.
29. Majou F, Gosselin C, Wenger P, Chablat D. Parametric stiffness analysis of the Orthoglide. *Mech Mach Theory* 2007;42:296–311.
30. Pashkevich A, Chablat D, Wenger P. Stiffness analysis of overconstrained parallel manipulators. *Mech Mach Theory* 2009;44:966–82.
31. Kövecses J, Angeles J. The stiffness matrix in elastically articulated rigid-body systems. *Multi Syst Dyn* 2007;18:169–84.
32. Quennouelle C, Gosselin CM. Stiffness Matrix of Compliant Parallel Mechanisms. In: Lenarčič J, Wenger P, editors. Advances in Robot Kinematics: Analysis and Design. Springer Netherlands; 2008. pp. 331–41.
33. Tyapin I, Hovland G. Kinematic and elastostatic design optimisation of the 3-DOF Gantry-Tau parallel kinematic manipulator. *Modeling, Identification and Control* 2009;30:39–56. DOI
34. Cammarata A, Caliò I, D'Urso D, et al. Dynamic stiffness model of spherical parallel robots. *J Sound Vib* 2016;384:312–24.
35. Wu L, Wang G, Liu H, Huang T. An approach for elastodynamic modeling of hybrid robots based on substructure synthesis technique. *Mech Mach Theory* 2018;123:124–36.
36. Wu J, Li T, Wang J, Wang L. Stiffness and natural frequency of a 3-DOF parallel manipulator with consideration of additional leg candidates. *Robot Auton Syst* 2013;61:868–75.
37. Lara-Molina FA, Koroishi EH, Costa TL. Elastodynamic Performance of a Planar Parallel Mechanism Under Uncertainties. In: International Symposiu on Multibody Systems and Mechatronics. Springer; 2017. pp. 183–92.
38. Wu G, Zhao W, Zhang X. Optimum time-energy-jerk trajectory planning for serial robotic manipulators by reparameterized quintic NURBS curves. *Proc Ins Mech Eng Part C J Mech Eng Sci* 2021;235:4382-93. DOI
39. Denavit J, Hartenberg RS. A kinematic notation for lower-pair mechanisms based on matrices. *ASME J Appl Mech* 1955;22:215–21.
40. Tsai LW. Robot Analysis: The Mechanics of Serial and Parallel Manipulators. John Wiley & Sons; 1999.
41. Ranjbaran F, Angeles J, González-Palacios MA, Patel RV. The mechanical design of a seven-axes manipulator with kinematic isotropy. *J Intell Robot Syst* 1995;14:21–41.
42. Pashkevich A, Klimchik A, Chablat D. Enhanced stiffness modeling of manipulators with passive joints. *Mech Mach Theory* 2011;46:662–79.
43. Wu G, Bai S, Kepler J. Mobile platform center shift in spherical parallel manipulators with flexible limbs. *Mech Mach Theory* 2014;75:12–26.
44. Rao SS. Mechanical Vibrations. 4th ed. Prentice Hall; 2003.
45. Dong C, Liu H, Huang T, Chetwynd DG. A screw theory-based semi-analytical approach for elastodynamics of the Tricept robot. *ASME J Mech Robot* 2019;11:031005.
46. Alessandro C, Rosario S. Elastodynamic optimization of a 3T1R parallel manipulator. *Mech Mach Theory* 2014;73:184–96.

APPENDIX A: STIFFNESS MATRICES OF ARM LINKS

The stiffness matrices of the upper and lower links $\mathbf{K}_u$ and $\mathbf{K}_l$ for the 5-dof robotic arm, computed by means of finite element analysis (FEA) with ANSYS, are given as

$$\mathbf{K}_u = \begin{bmatrix} 0.0309 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.2675 & 0 & 0 & 0 & 0.4176 \\ 0 & 0 & 0.3574 & 0 & -0.5957 & 0 \\ 0 & 0 & 0 & 15.3676 & 0 & 0 \\ 0 & 0 & -0.5957 & 0 & 1.6919 & 0 \\ 0 & 0.4176 & 0 & 0 & 0 & 1.7505 \end{bmatrix} \cdot 10^6 \tag{A-1a}$$

$$\mathbf{K}_l = \begin{bmatrix} 0.0417 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.0452 & 0 & 0 & 0 & 2.5493 \\ 0 & 0 & 1.1631 & 0 & -2.8369 & 0 \\ 0 & 0 & 0 & 17.2304 & 0 & 0 \\ 0 & 0 & -2.8369 & 0 & 8.3105 & 0 \\ 0 & 2.5493 & 0 & 0 & 0 & 8.2351 \end{bmatrix} \cdot 10^6 \tag{A-1b}$$

where the blocks corresponding to rotation, translation, and coupling terms are given in Nm/rad, N/rad, and N/m, respectively.

Bernardo *et al*. *Intell Robot* 2021;1(2):116-30
DOI: 10.20517/ir.2021.10

Intelligence & Robotics

Research Article

Open Access

Planning robotic agent actions using semantic knowledge for a home environment

Rodrigo Bernardo[1,2], João M. C. Sousa[1], Paulo J. S. Gonçalves[2]

[1]IDMEC, Mechanical Engineering Department, Instituto Superior Técnico, University of Lisbon, Lisbon, Lisbon 1049-001, Portugal.
[2]Industrial Engineering Department, Polytechnic Institute of Castelo Branco, Castelo Branco, AV. Empresário, Castelo Branco 6000-767, Portugal.

Correspondence to: Dr. Rodrigo Bernardo, IDMEC, Mechanical Engineering Department, Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais, 1 1049-001 Lisbon, Portugal. E-mail: rodrigo.f.bernardo@tecnico.ulisboa.pt

How to cite this article: Bernardo R, Sousa JMC, Gonçalves PJS. Planning robotic agent actions using semantic knowledge for a home environment. *Intell Robot* 2021;1(2):116-30. http://dx.doi.org/10.20517/ir.2021.10

Received: 14 Sep 2021 **First Decision:** 28 Oct 2021 **Revised:** 11 Nov 2021 **Accepted:** 11 Nov 2021 **Published:** 15 Dec 2021

Academic Editors: Shaidah Jusoh, Simon X. Yang **Copy Editor:** Huan-Liang Wu **Production Editor:** Huan-Liang Wu

Abstract

Autonomous mobile robotic agents are increasingly present in highly dynamic environments, thus making the planning and execution of their tasks challenging. Task planning is vital in directing the actions of a robotic agent in domains where a causal chain could lock the agent into a dead-end state. This paper proposes a framework that integrates a domain ontology (home environment ontology) with a task planner (ROSPlan) to translate the objectives coming from a given agent (robot or human) into executable actions by a robotic agent.

Keywords: Ontologies, autonomous robotics, planning, knowledge representation,semantic maps

1. INTRODUCTION

Robots are increasingly present in environments shared with humans and highly dynamic environments[1,2]. It is therefore imperative to study and develop new techniques so that robots can effectively move, locate themselves, detect objects, perform tasks, *etc.* in places that can change rapidly, in an autonomous way. More complex methodologies require systems capable of deliberating quickly and effectively. Several studies point to the need for knowledge as way to address this challenges[3,4]. Therefore the formal conceptualization of the robotics domain is an essential requirement for the future of robotics, to design robots that can autonomously perform a wide variety of tasks in a wide variety of environments[5].

© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Robots need to efficiently create semantic models of their environment (semantic maps). One way that has proved to have shown great value in representing the information of the environment where robots work, is through semantic maps. These combine semantic, topological, and geometric information into a compact representation[6,7]. Existing semantic maps need to evolve from task-specific representations to models that can be dynamically updated and reused in different tasks. This is one of the major limitations of these approaches. Moreover, the ontologies developed to date are not reusable, being one of the major limitations in this strategy. Ontologies should move towards a more homogeneous structure and easy interchangeability between different structures in order to be reusable[8]. Recently, robots are pouring into home environments, and thus the need to communicate with humans is increasing. The tasks of robots are not only to navigate in an accurate geometrical space but also to understand the indoor environment and share common semantic knowledge with people. Consider the task of fetching a cup of coffee. If a robot had only a representation of the environment through a metric map, it would have to search in a crude way all over the environment until it found the cup. If more knowledge were added through semantics to the robot, such as the probability of the cup being in each room, the search could be guided from locations with high probability to locations with lower probability. In short, with the evolution of systems and artificial intelligence (AI), ontologies become a great solution to make domain knowledge explicit and remove ambiguities, enable machines to reason, and facilitate knowledge sharing between machines and humans, focusing on a new generation of intelligent and integrated technologies for smart manufacturing.

Currently, in robotics, the most used middleware is the ROS (https://www.ros.org). This is the standard middleware for the development of robotic software, allowing the design of modular and scalable robotic architectures. There is a framework in ROS called ROSPlan (http://kcl-planning.github.io/ROSPlan/) that provides a collection of tools for AI planning, namely ROSPlan. It has a variety of nodes which encapsulate planning, problem generation, and plan execution. It possesses a simple interface and links to common ROS packages. To date, this framework does not yet have an adequate interface for semantic queries, thus lacking a general standardized framework for working with ontologies, natively supporting symbolic logic and advanced reasoning paradigms. In this sense, the paper proposes a framework that integrates a domain specific home environment ontology with a task planner (ROSPlan), translating the objectives coming from another agent (robot or human) into executable actions by the robotic agent. Two reasoning systems for task planning were developed, which are based on ontologies The first system uses the MongoDB (https://www.mongodb.com/) database, while the second system uses the domain specific ontology home environment that is proposed in this paper.

The paper is structured as follows. In the next section, the related work is reviewed. The design methodology section introduces the reasoning systems presented in this work. In the results section, the structure of the ontology is presented, along with the results obtained from the different proposed reasoning systems. In the validation and discussion section, the developed ontology is validated and discussed. Finally, the conclusions section presents the conclusions of the work and the future way forward.

2. RELATED WORK

Ontologies are a powerful solution for acquiring and sharing common knowledge. Ontologies represent a common understanding in a given domain, promoting semantic interoperability among stakeholders, because "sharing a common ontology is equivalent to sharing a common world view"[3]. All concepts in an ontology must be rigorously specified so that humans and machines can use them unambiguously, empowering robots to autonomously perform a wide variety of tasks in a wide variety of environments.

Depending on their level of generality, different types of ontologies can be identified[9]; among many types, we can identify as the main ones:

- An upper or general ontology (upper ontology or foundation ontology) is a model of the common objects that are generally applicable to a wide variety of domain ontologies. There are several higher ontologies standardized for use, such as SUMO (suggested upper merged ontology)[10], Cyc ontology[11], BFO (basic formal ontology)[12], and DOLCE (descriptive ontology for linguistic and cognitive engineering)[13].
- Domain ontologies (domain ontology or domain-specific ontology) model a specific domain or part of the world (*e.g.*, robotic[14], electronic, medical, mechanical, or digital domain).
- Task ontologies describe generic tasks or activities[15].
- Application ontologies are strictly related to a specific application and used to describe concepts of a particular domain and task.

In the next subsections, the related work is reviewed about semantic maps, ontologies for semantic maps, and some applications of knowledge representation for robotic systems.

2.1. Semantic maps

The daily challenges have drive the research for automated and autonomous solutions to enable mobile robots to operate in highly dynamic environments. For this purpose, mobile robots need to create and maintain an internal representation of their environment, commonly referred to as a map. Robotic systems rely on different types of maps depending on their goals. Different map typology's have been developed such as metric and topological maps, which are generally 2D representations of the environment[16], or hybrids (a combination of the previous two)[17,18]. There are also maps with 3D representation (sparse map, semi-dense map, and dense map). Metric and topological maps only contain spatial information[19]. A fundamental requirement for the successful construction of maps is to deal with uncertainty arising, from errors in robot perception (limited field of view and sensor range, noisy measurements, *etc.*), from inaccurate models and algorithms, *etc.*

To get around this limitation, semantic maps were developed to add additional information, such as instances, categories, and attributes of various constituent elements of the environment (objects, rooms, *etc.*)[5–7]. These provide robots with the ability to understand beyond the spatial aspects of the environment, the meaning of each element, and how humans interact with them (features, events, relationships, *etc.*). Semantic maps deal with meta information that models the properties and relationships of relevant concepts in the domain in question, encoded in a knowledge base (KB).

2.2. Ontologies for semantic maps

One of the tasks to be solved in mobile robot navigation is the acquisition of information from the environment. In the field of semantic navigation, information includes concepts such as objects, utilities, or room types. The robot needs to learn the relationships that exist between the concepts included in the knowledge representation model. Semantic maps add to classical robotic maps spatially grounded object instances anchored in a suitable way for knowledge representation and reasoning. The classification of instances through the analysis of the data collected by sensors is one of the biggest challenges in the creation of semantic maps (*i.e.*, to give a richer semantic meaning to the sensor data)[20,21].

In the last decade, several papers have appeared in the literature contributing different representations of semantic maps. Kostavelis *et al.*[5] summarized the significant progress made on a broad range of mapping approaches and applications for semantic maps, including task planning, localization, navigation, and human–robot interaction. Semantics has been used in a diverse range of applications. Lim *et al.*[22] presented an approach for unified robot knowledge for service robots in indoor environments. Rusu *et al.*[23] developed a map called Semantic Object Maps (SOM), which encodes spatial information about indoor household environments, in particular kitchens, but in addition it also enriches the information content with encyclopedic and common sense knowledge about objects, as well as includes knowledge derived from observations. Galindo *et al.*[24] proposed an approach for robotic agents to correct situations in the world that do not conform to the semantic model by generating appropriate goals for the robot. In short, it combines the use of a semantic

map with planning techniques in Planning Domain Description Language (PDDL) that converts the goals into actions (moving the robot, picking and dropping an object, *etc.*). Wang *et al.*[25] employed the relationships among objects to represent the spatial layout. Object recognition and region inference are implemented by using stereo image data. Vasudevan *et al.*[26] created a hierarchical probabilistic concept-oriented representation of space, based on objects. Diab M *et al.*[27] introduced an interpretation ontology to identify possible failures which occur during automatic planning and the execution phase; this ontology aims to improve planning and allow automatic replanning after error. Balakirsky *et al.*[28] proposed an ontology-based framework that allows a robotic system to automatically recognize and adapt to changes that occur in its workflow and dynamically change the details of task assignment, increasing process flexibility by allowing plans to adapt to production errors and task changes.

In short, semantic maps enable a robot to solve reasoning problems of geometric, topological, ontological, and logical nature, in addition to localization and path planning[29]. Formal conceptualization of the robotics domain is an essential requirement for the future of robotics, in order to be able to design robots that can autonomously perform a wide variety of tasks in a wide variety of environments.

2.3. Applications in robotic systems

Different research groups have used semantic knowledge in the area of robotics. Semantic knowledge allows a clear dialog between all stakeholders involved in the life cycle of a robotic system and enables the efficient integration and communication of heterogeneous robotic systems. These facilitate communication and knowledge exchange between groups from different fields, without actually forcing them to align their research with the particular view of a particular research group[30].

One of the most recent advances in the field of robotics can be denoted by analyzing the KnowRob project, where researchers aimed to enable a robot to answer different types of questions about possible interactions with its environment, using semantic knowledge[31,32]. For example, they developed an ontology that allows the robot to start an assembly activity, with incomplete knowledge. It identifies the missing parts, having the ability to reason about how the missing information can be obtained[32]. KnowRob employs the DUL foundational ontology, which is a slim version of the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE). DUL and DOLCE have a clear cognitive bias, and they are both well established in the knowledge engineering community as foundational ontologies. However, DUL does not define very specific concepts such as fork or dish. These concepts are needed for our robots that do everyday activities[8]. There are also other relevant works that aim at the standardization of knowledge representation in the robotics domain, such as IEEE-ORA[33], ROSETTA[34], CARESSES[35], RoboEarth[36], RoboBrain[37], RehabRobo-Onto[38], and OROSU[39].

All of the above work already represents promising advances in the use of semantics in robotic systems; however, it lacks the ability to perform advanced reasoning and relies heavily on *ad hoc* reasoning solutions, significantly limiting its scope. A general standardized framework for working with ontologies is needed, natively supporting symbolic logic and advanced reasoning paradigms.

The next sections present the reasoning frameworks, with particular emphasis on the domain specific ontology, home environment. The proposed ontology is designed to be easily reusable in different environments of a house, as well as by different robotic agents.

3. DESIGN METHODOLOGY

The focus of the developed ontology is to enable robotic agents to interact with elderly people within a home environment. The robots are to assist the elderly people to manage and better perform their daily lives, and

thus to longer live an independent life in their known surroundings. They are also there to help better maintain social contacts, which is known to have a very positive influence on the mental and physical health of elderly people. An initial and extensible list of tasks that these robots are eventually supposed to perform includes the following activities:

- Help elderly people out of bed and or the couch.
- Serve the breakfast.
- Supply elderly people medicine.
- Bring books or operate media (entertainment).
- Make up the bedroom.
- Play games.
- Adjust settings: shades, light, and heat.
- Serve drinks.
- Assist during bathing.
- Clean the rooms.

When developing the ontology, several concepts were searched in databases such as dictionaries on the web. The search was carried out on specific sections on the different concepts of a house (https://www.enchantedlearning.com/wordlist/house.shtml, https://dictionary.cambridge.org/pt/topics/buildings/houses-and-homes/). It was also extracted from the documentation of the project RoCKIn@Home challenge (http://rockinrobotchallenge.eu/home.php, http://rockinrobotchallenge.eu/RoCKIn_D2.1.1.pdf), concepts and typical tasks for a domestic robot. The concepts were chosen in order to characterize the simplified environment of a house, based on a smart-home environment, built on the robotics laboratory of Instituto Politécnico de Castelo Branco. A challenge arose in the choice of concepts due to the great complexity of objects that can be found in a given room. To simplify the process, the concepts were defined for characterizing the simplified environment that is found in the laboratory.

3.1. Knowledge Engine

Figure 1 depicts the global knowledge engine conceptual framework, designed to achieve the main objective of the paper: the integration of a domain specific home environment ontology, with a task planner (ROSPlan), transforming the goals coming from the reasoning into executable actions by the robotic agent. The framework have three main parts: reasoning (which includes the ontologies), planning, and the robot. These parts are presented in the remainder of the paper.

A domain specific home environment ontology, aligned with a MongoDB database, encapsulate the important concepts of the domain to be considered (space of a house, objects, *etc.*). Indeed, the main benefit of a domain ontology is to set standard definitions of shared concepts identified in the requirement phase and to define appropriate relations between the concepts and their properties[40]. The ontology contain concepts of Core Ontology for Robotics and Automation (CORA), with the representation of fundamental concepts of robotics and automation[41].

The ontology model is based on the concepts and relationships between different entities, and then aligned with the MongoDB database. The basic concept of the reasoning process is based on the premises that: a relational database contains both the entities of the conceptual hierarchy and the instances of the physical hierarchy, this information is stored in lists, and these lists are related to each other, as in the entity–relationship model of the environment[42,43].

The domain specific ontology home environment is defined with the Protégé software. Protégé version 5.5.0 was used[44]. The the domain ontology was verified through version 1.4.3 of HermiT Reasoner to ensure that it is free of inconsistencies[45]. Protégé is a free, open-source editor for developing the ontologies produced by

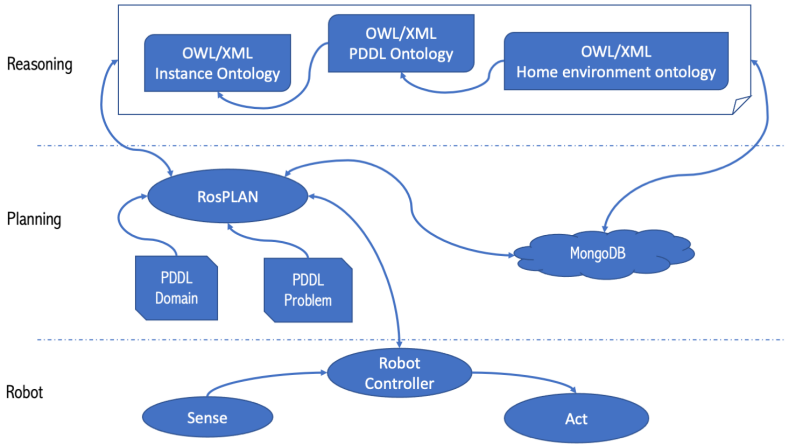


Figure 1. Global System, containing a Reasoning section that is based on an ontology, a Planning section that is based on the Planning Domain Definition Language (PDDL), a relational database (MongoDB) that is queried using the ontology and ROSPlan, and a Robot section that is based on robot controller, as well as its sensing and acting devices.

Stanford university. It is a java-based application (multi-platform), with plugins such as ontoViz to visualize the ontologies. The backbone of protégé is that it supports the tool builders, domain specialists, and knowledge engineers.

3.1.1. Domain specific ontology home environment

Figure 2 shows the main relationships of the developed domain-specific home environment ontology, which was designed for an agent to interpret and interact with its surrounding environment. In this case, the environment is a house, with special focus on the internal environment.

The developed model is subdivided into three main classes: *Home_lab*, *Information*, and *Objects*. The class *Home_lab* is subdivided into class *Environment*, which is subdivided into internal and external environment. The *Indoor_environment* is further subdivided into *Hall*, *Corridor*, and *Rooms*. The class *Rooms* is subdivided into the possible rooms types of an house (bathroom, bedroom, kitchen, *etc.*). The class *Information* is subdivided into *Information_Object*. The *Objects* class is subdivided into *Types_Devices* that are subdivided into *Sensors* and *Actuators*. The main class *Objects* is further subdivided into internal and external objects that contemplate the objects that can be found in a home environment. Finally, the *Objects* class is subdivided into *PhysicalAgent*, which is subdivided into the agents that can appear in the environment as *human* or *robot*, the latter further subdivided into the different types of robots.

Figure 2 presents the hierarchical class where the main concepts defined in the ontology are visible as:

- **Environment:** The surroundings or conditions in which an agent, person, animal, or plant lives or operates.
- **Indoor environment:** Environment situated inside of a house or other building.
- **Corridor:** A long passage in a building from which doors lead into rooms.
- **Hall:** The room or space just inside the front entrance of a house or flat.
- **Rooms:** Space that can be occupied or where something can be done (kitchen, bedroom, etc).
- **Outdoor Objects:** Used to describe objects that exist or appear outside a home.
- **Objects:** Any physical, social, or mental object, or a substance. Following DOLCE, objects are always participating in some event (at least their own life), and are spatially located (defined by: http://www.ontologydesignpatterns.org/ont/dul/DUL.owl).
- **Outdoor environment:** Environment situated outside of a house or other building.

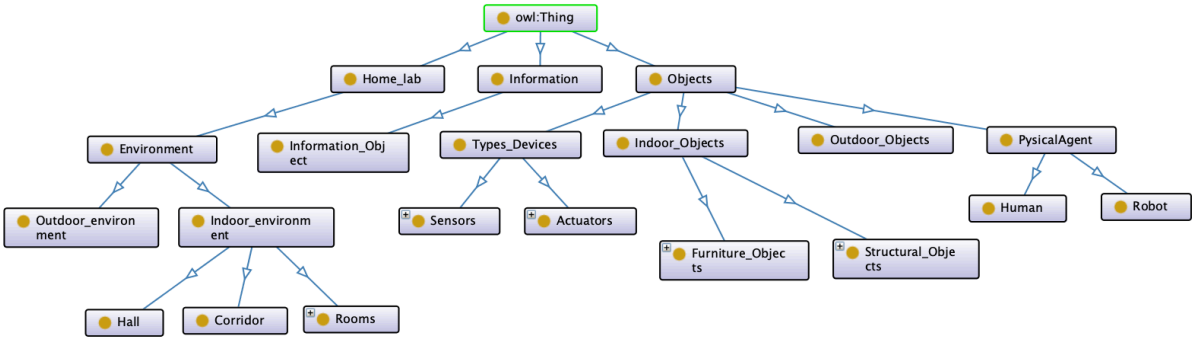


Figure 2. Snapshot of the main ontology (relations *is-a*).

- **Information:** Facts provided or learned about something or someone.
- **Information_Object:** They are messages performed by some entity [46]. They are ordered (expressed in accordance with) by some information encoding system (*e.g.*, sensors present in the agent). They can express a description (the ontological equivalent of a meaning/conceptualization), they can be about any entity, and they can be interpreted by an agent.
- **Indoor Objects:** Used to describe objects that exist or appear inside a home.
- **Outdoor Objects:** Used to describe objects that exist or appear outside a home.
- **Physical Agent:** Any agentive Object, either physical (*e.g.*, a whale, a robot, or an oak tree) or social (*e.g.*, a corporation, an institution, or a community) (defined by: http://www.ontologydesignpatterns.org/ont/dul/DUL.owl).
- **Mobile robot:** Robot that is able to move in the surrounding (locomotion) (*i.e.*, autonomous mobile robot and autonomous mobile and manipulator robot).
- **Not mobile robot:** Robot that is not able to move in its surroundings (*i.e.*, robot arm).
- **Types Devices:** A collection of properties that define different components and behaviors of a type of device (actuators, sensors, *etc.*).

The properties of objects and the spatial relationships between them represent the characteristics of the environment and the spatial arrangement, respectively. Several *Object properties* were created to relate the different concepts, so that an agent can characterize its surrounding environment:

- *ObjectProperty*:
  - *belowOf* ;
  - *isMember*;
  - *isObjectOf*;
  - *isPartOf*;
  - *LeftOf*;
  - *onTopOf*;
  - *RightOf*.
- *LocationProperty*:
  - *isConnectedTo*.
- *AgentProperty*:
  - *isGoingTo*;
  - *isIn*.

The object properties were defined to make explicit the relationships between concepts. The properties *belowOf*, *LeftOf*, *onTopOf*, and *RightOf* were created to define the relationships between the different concepts of *Indoor_Objects* and *Outdoor_Objects*. Through these, the agent can identify the disposition of objects in the environment, creating relationships about them (*e.g.*, based on Figure 3, if an agent has to guide an elderly person to the chair that is in the room, it knows that it is on the right side of the bed). The property *isConnectedTo*

is a transitive and symmetric property, which correlates the different concepts of the *Environment* according to the environment in which the agent is inserted. The object properties *isGoingTo* and *isIn* are defined in order to correlate the *AgentProperty* concept with the *Environment*. (*e.g.*, the robot *isIn* the living room, but it *isGoingTo* the bedroom). The object property *isPartOf* is a symmetric property that is used for the agent to link a given instance of the *Objects* class with an *Environment*. This allows the agent to know which objects are in a room. It differs from the object property *isObjectOf* because in this property the agent is sure that the object exists in the environment. Finally, the *isObjectOf* concept was defined to relate the concepts: Objects to the Environment. Through this, the objects that can be found in each zone of the environment are defined. (*i.e.*, in a room there can be an object of the type bed, chair, television, carpet, *etc.*). Thus, an agent can search for an object by the place with the highest probability of it being found (*i.e.*, if the agent has to find a frying pan, it knows that this object is commonly in a kitchen).

3.2. PDDL Planning Agent

The Planning Domain Definition Language (PDDL) describes problems through the use of predicates and actions. The problems in PDDL are defined in two parts, a domain and a problem file. This language has undergone different modifications in order to make it capable of dealing with more complex tasks [47–49]. The ROSPlan framework was used to perform the planning tasks [50]. ROSPlan is a high-level tool that provides planning in the ROS environment; it generates the PDDL problem, the plan, the action dispatch, the replanning, *etc.* Different action interfaces have been written in C++ to control the Autonomous Manipulator Mobile Robot (AMMR) (*i.e.*, base, arm, and gripper). These interfaces are constantly listening for action PDDL messages. In addition, the MongoDB database was used for semantic memory storage (locations, robots, home objects, goal parameters, *etc.*).

The POPF planner (https://nms.kcl.ac.uk/planning/software/popf.html), a forwards-chaining temporal planner, was used. After the plan was generated, the interface actions interconnect the plan with the lower level control actions, allowing the robotic agent (AMMR) to complete the plan. During execution, if an action fails due to changes in the environment, the planning agent reformulates the PDDL problem by re-planning.

4. RESULTS

This section discusses the main results obtained by applying the proposed framework. For a better understanding, the results are divided into two subsections: The first subsection refers to the validation results of the home environment ontology, where the main reasoning techniques were presented and how they can be used. In the second subsection, the results of the reasoning system through MongoDB are presented. The way the robot performs a set of tasks, in a real environment, is also presented.

4.1. Validation home environment ontology

The home environment ontology contains a vast number of concepts regarding the home environment as well as the different objects that may be present in a given room. For example, if a given agent is in a room for the first time, based on the objects it observes, through its sensors, it can categorize the space based on the knowledge represented on the ontology. Based on the ontology, if the agent sees objects such as knives, pots, and pans, then it infers that it must be in a kitchen. In this situation, the agent will identify the room and create all the relations of the objects it detects in the environment.

Knowledge reasoning techniques can infer new conclusions and thus help to plan dynamically in a non-deterministic environment. In the presented application, spatial reasoning and reasoning based on relations are used [51]. Spatial reasoning is mostly used, when it is done through reasoning on the ontology hierarchy and spatial relations therein, allowing to predict the exact spatial location of an object in the environment. This prediction is obtained using a set of asserted facts and axioms on the ontology. Reasoning over the ontology

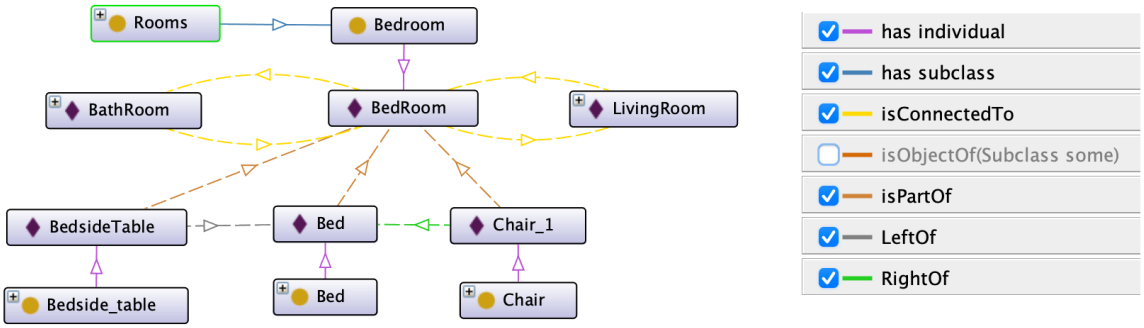


Figure 3. Example of the ontology being used to instantiate the house objects belonging to the instance *"Bedroom"*, present in the home environment.

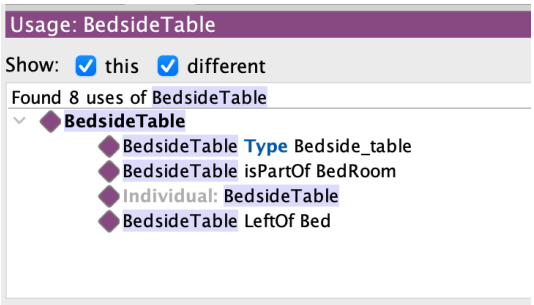


Figure 4. Example of the knowledge representation of an object in the environment of a house that has been instantiated.

relations is used to help with inferring new conclusions (*e.g.*, If B is a subclass of A and C is a subclass of B, then C is a subclass of A can be inferred as transitivity holds for the subclass property).

The object property hierarchy view displays the asserted and inferred object property hierarchies. From the knowledge base built, all existing components from the environment are instantiated. That is, the individuals are based on their type (Figure 3). This allowed testing the functionality of the ontology. Figure 3 depicts that each room was instantiated to the corresponding *Rooms* subclass (*e.g.*, the *BedRoom* instance is related to the *Bedroom* class). Several relations can be drawn from the instances that represent the environment, such as which objects belong to the *BedRoom* instance. Although the relations of the *BathRoom* and *LivingRoom* instances do not appear in Figure 3, these can be obtained, based on the created instances present in the ontology, in order to represent both rooms.

As previously stated, having the environment of a house in an ontological knowledge oriented database is of special interest, for example, to know where each object belongs. In fact, as depicted in Figure 4, the information about the BedsideTable is completely available to the user, using a simple logic description. A query to the ontology will retrieve useful information, for example where the object is attached. This issue is further discussed in the next subsection.

Performing logical description reasoning actions to obtain valuable data for the robot's reasoning can be done, as presented in Figure 5, namely information about objects, rooms, and their relations defined in the ontology. For example, what type of object is instantiated as Bed? In which rooms is it present? What is on its left? What is on its right? Using the ontology and description logic queries, it is straightforward to obtain the following

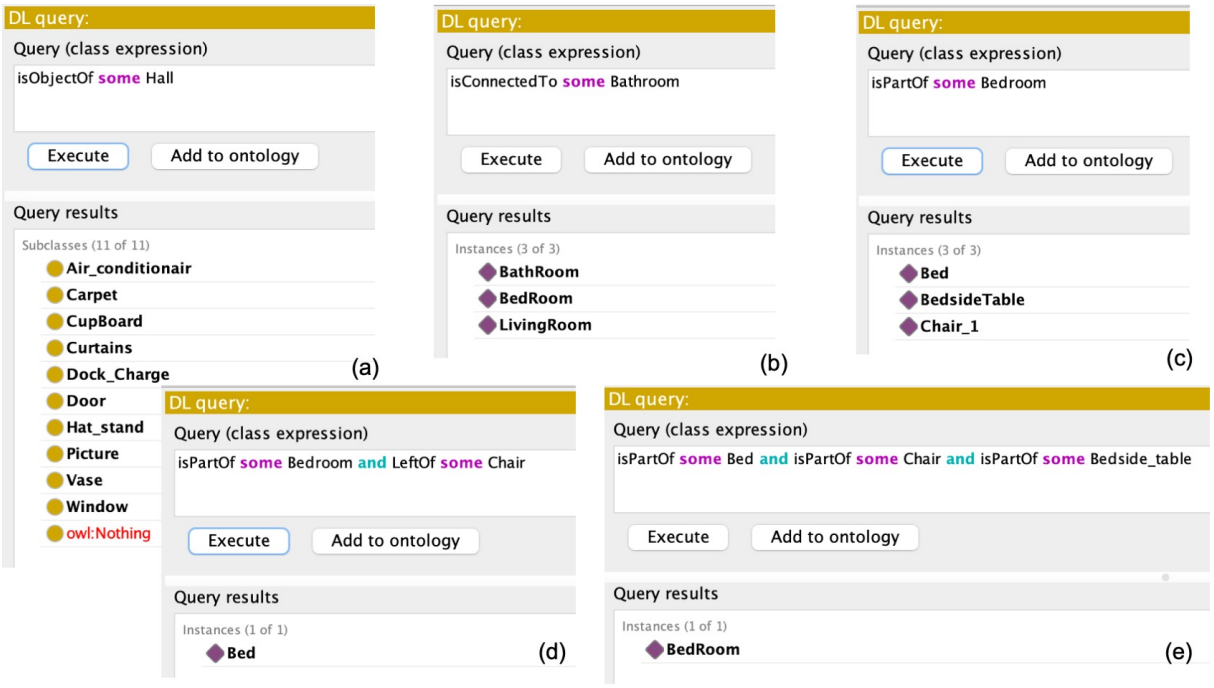


Figure 5. Reasoning using the ontology. (a) Query about the classes of objects that can be found in a class Hall. (b) Query about the instance(s) connected to instance of class Bathroom. (c) Query on the instance(s) that are part of the class Bedroom. (d) Query about the instance(s) that are part of class BedRoom and that are on the left of instances of class Chair. (e) Query to discover the location (room where the agent is in) based on what the agent observes.

information:

- Classes of *Objects* that can be found in a certain instance of room (*e.g.*, in the Hall in Figure 5a).
- Connectivity relationship between instances of the *Rooms* class referring to an environment (Figure 5b).
- Instance of objects present in an instance of room (Figure 5c).
- Recognize which instance(s) of the *Objects* class belong to a particular instance of a *Rooms* class and are to the left of an instance of the *Chair* class (Figure 5d).
- Recognize which instance of the *Rooms* class belong to a particular instance(s) of a *Objects* class (*e.g.*, the robotic agent can locate itself (know in which room it is), based on the objects it observes) (Figure 5e).

Through the ontology developed, one or more agents are able to locate themselves more efficiently in the environment. When the agent is lost, it can identify the room where it is, based on the objects it observes (Figure 5e). Observing Figure 3, if the robot recognizes a *Bed*, a *BedsideTable*, and a *Chair_1*, it knows that it is in a *Bedroom*. The robotic agent will be able to perform a search in an optimized way for an object. It does not need to perform a massive search for all the rooms; *e.g.*, it knows which are the rooms in which there is a higher probability of finding a fridge, teapot, *etc.*

4.2. Validation the reasoning system with MongoDB

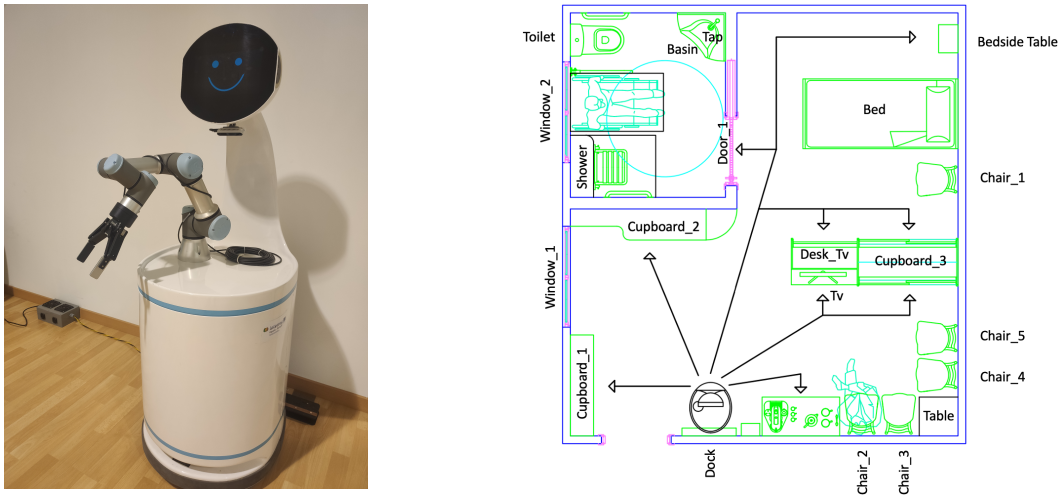
To test the system based on the MongoDB database, a problem was outlined, for the agent to execute/solve (Figure 6). An AMMR is used, composed of a mobile base and a robotic arm of the Universal Robotics UR3, equipped with a RobotIQ 2f-140 gripper, the whole system runs with the middleware: ROS (Figure 7a). Figure 7b depicts the layout of a simple home environment, an apartment for elderly people, created under the EUROAGE project[52], which is in the robotics laboratory of the Polytechnic Institute of Castelo Branco.

Initial conditions were established, such as the location of the AMMR (dock), the world coordinate at which the robot arm is located (p0), and the location of the object in the environment (LivingRoom), as well as its

```
(define (problem task)
(:domain AMMR)
(:objects
 LivingRoom BedRoom - Rooms
 robot_base - robot
 p0 p1 p2 p3 - p_goal
 obj1 dock - Objects
 robotiq - gripper
)
(:init
 (at robot_base dock)
 (at-arm p0)
 (gripper_is_free robotiq)
 (all obj1 p3 robot_base LivingRoom)
 (docked robot_base)
 (dock_at dock)

)
(:goal (and
 (at obj1 p2 robot_base BedRoom )
 (docked robot_base)
))
)
```

Figure 6. Task definition in PDDL.



(a) Robotic agent (AMMR) **(b)** Environment

Figure 7. Robotic agent (AMMR) and simplified environment of a house, created in the framework of the EUROAGE project.

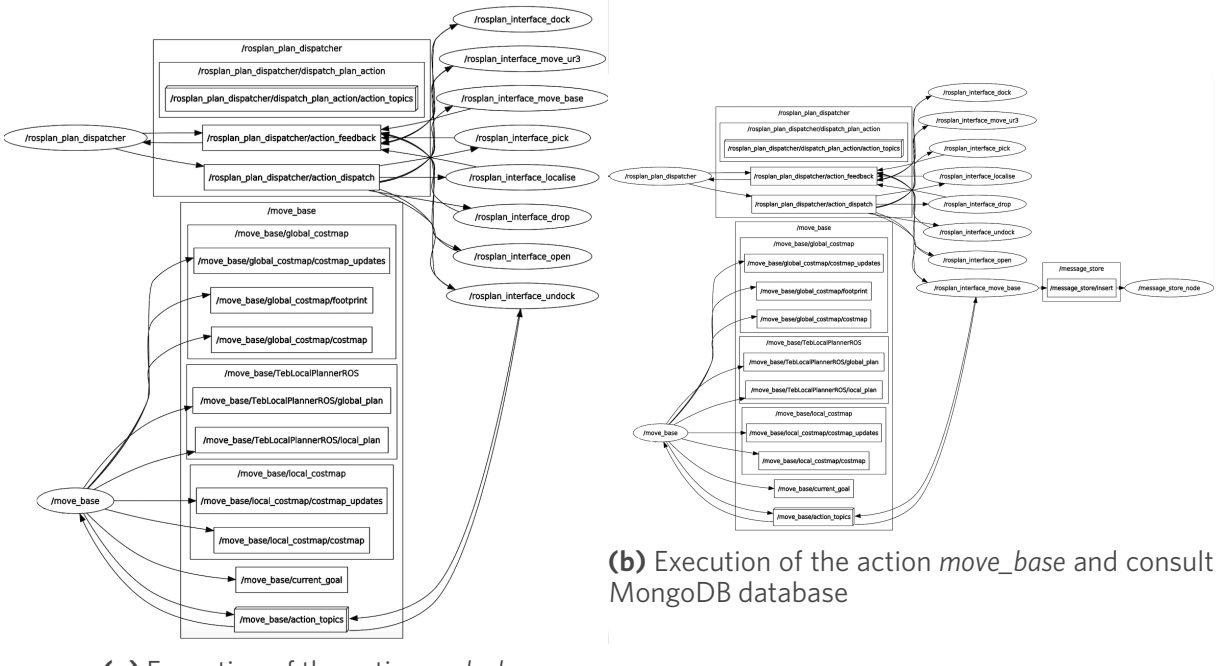
world coordinate (p3). The AMMR aims to leave the dock and pick up an object (obj1) that is in the living room, at the coordinates of world (p3). After the object is grabbed, the AMMR should take it to the bedroom and drop the object at the coordinates of world (p2). After the pick and place tasks are completed, the AMMR should return to the dock. This task definition is depicted in Figure 6.

Using ROSPlan, the generated plan is visible in Table 1. The right column presents the time of each durative action.

Figure 8, presents an excerpt of the global view of the ROS nodes and topics used by the system. It is possible to verify how the connections between them occur. It is visible in Figure 8a that, when the planner (ROSPlan)

Table 1. Example: generated plan

Global Time	Action and respective objects used	Time of action
0.000	(undock robot_base dock)	[5.000]
5.001	(localise robot_base)	[10.000]
15.002	(open robotiq robot_base)	[2.000]
17.002	(move_base dock LivingRoom robot_base)	[5.000]
22.003	(move_ur3 p0 p3 LivingRoom robot_base)	[5.000]
27.003	(pick obj1 p3 robotiq robot_base LivingRoom)	[2.000]
29.003	(move_base LivingRoom BedRoom robot_base)	[5.000]
34.003	(move_ur3 p3 p2 BedRoom robot_base)	[5.000]
39.003	(drop obj1 p2 robotiq robot_base BedRoom)	[2.000]
41.003	(dock robot_base dock)	[5.000]



(a) Execution of the action *undock*

(b) Execution of the action *move_base* and consult MongoDB database

Figure 8. Execution of a plan by the robotic agent (AMMR).

requests that the agent should leave the dock, the action responsible for the task, */rosplan_interface_undock*, is triggered. This action in turn communicates with the action */move_base* that communicates with the actuators (motors) to move the robot. Figure 8b shows that, when the robot is requested to move to a certain room, */rosplan_interface_move_base* is activated, which in turn consults the MongoDB database to know where the room is located in the map. Then, the robot moves in the environment, after calling */move_base action*.

5. CONCLUSION AND FUTURE WORK

The use of ontologies has become a great solution, and one of the paths to follow in the future to make domain knowledge explicit and eliminate ambiguities, allow machines to reason, and facilitate knowledge sharing between machines and humans. In this work, a structured ontology is presented to be used by robotic agents in order to assist them in their deliberation tasks (interaction with the environment and robot movement). It is imperative to endow robotic agents with semantic knowledge. Several approaches in the literature show advantages in systems using databases such as MongoDB, pointing to their speed of response compared to ontology-based systems. This paper introduces a framework that combines both, in terms of concepts and their implementation in real robotic systems.

The proposed framework improves the problem specification in PDDL based on the updated information coming from the ontology, making the generated plan more efficient. For example, if an external agent (robot or human) launches a task for the robotic agent to collect a certain object and transport it to a specific location, the ontology will be queried. As such, the problem in PDDL is written with specific information, such as the relations between objects, relations between objects and the environment, the current location of the robot, and so on.

The developed home environment ontology was validated by performing successful queries to it, using a standard reasoner in Protégé. The concepts within the home environment ontology, used to define the MongoDB database, were experimentally validated, for semantic reasoning in the the home environment of the laboratory. Moreover, the ROSPlan, together with the developed interface actions, was shown to be a very efficient approach, in interfacing the low-level control with the semantic reasoning of the robot agent.

In future work, the developed domain ontology will be aligned with upper ontologies, *e.g.*, DOLCE. Further developments will be pursued to speed up the ontology-based approach, by exploring ways to make querying more efficient by solving the limitation presented in other works[43], where it is pointed out that these solutions are slower than the pure database approaches.

DECLARATIONS

Authors' contributions

Implemented the methodologies presented and wrote the paper: Rodrigo Bernardo
Developed the idea of the proposed framework: Rodrigo Bernardo, Paulo J. S.Gonçalves
Managed and supervised the research project: João M. C. Sousa , Paulo J. S.Gonçalves
All authors have revised the text and agreed to the published version of the manuscript.

Availability of data and materials

The main data supporting the results in this study are available within the paper. The raw datasets here reported will be available upon request.

Financial support and sponsorship

This work is financed by national funds through FCT - Foundation for Science and Technology, I.P., through IDMEC, under LAETA, project UIDB/50022/2020. The work of Rodrigo Bernardo was supported by the PhD Scholarship BD\6841\2020 from FCT. This work has indirectly received funding from the European Union's Horizon 2020 programme under StandICT.eu 2023 (under Grant Agreement No.: 951972).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2021.

REFERENCES

1. Burke JL, Murphy RR, Rogers E, Lumelsky VJ, Scholtz J. Final report for the DARPA/NSF interdisciplinary study on human-robot interaction. *IEEE Trans Syst , Man, Cybern C* 2004;34:103–12.
2. Thrun S. Toward a framework for human-robot interaction. *Human–Computer Interaction* 2004;19:9–24. [DOI: 10.5898/JHRI.3.2.Beer]
3. Olszewska JI, Barreto M, Bermejo-Alonso J, et al. Ontology for autonomous robotics. In: 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE; 2017. pp. 189–94.
4. de Freitas EP, Olszewska JI, Carbonera JL, et al. Ontological concepts for information sharing in cloud robotics. *J Ambient Intell Human Comput* 2020:1–12.
5. Kostavelis I, Gasteratos A. Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems* 2015;66:86–103.
6. Hanheide M, Göbelbecker M, Horn GS, et al. Robot task planning and explanation in open and uncertain worlds. *Artificial Intelligence* 2017;247:119–50.
7. Toscano C, Arrais R, Veiga G. Enhancement of industrial logistic systems with semantic 3D representations for mobile manipulators. In: Iberian Robotics conference. Cham: Springer International Publishing; 2018. pp. 617-28.
8. Olivares-Alarcos A, Beßler D, Khamis A, Goncalves P, Habib MK, et al. A review and comparison of ontology-based approaches to robot autonomy. *The Knowledge Engineering Review* 2019;34.
9. Guarino N. Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy. vol. 46. IOS press; 1998.
10. Niles I, Pease A. Towards a standard upper ontology. In: Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001; 2001. pp. 2–9.
11. Lenat D, Guha R. Building large knowledge-based systems: Representation and inference in the CYC project. *Artificial Intelligence* 1993;61:53-63.
12. Arp R, Smith B, Spear AD. Building ontologies with basic formal ontology. Mit Press; 2015.
13. Masolo C, Borgo S, Gangemi A, Guarino N, Oltramari A. Wonderweb deliverable d18: Ontology library. Technical report, ISTC-CNR; 2003.
14. Prestes E, Carbonera JL, Rama Fiorini S, et al. Towards a core ontology for robotics and automation. *Robotics and Autonomous Systems* 2013;61:1193–204.
15. Balakirsky S, Schlenoff C, Rama Fiorini S, et al. Towards a robot task ontology standard. In: International Manufacturing Science and Engineering Conference. vol. 50749. American Society of Mechanical Engineers; 2017. p. V003T04A049.
16. Efficient integration of metric and topological maps for directed exploration of unknown environments. *Robotics and Autonomous Systems* 2002;41:21–39.
17. Sim R, Little JJ. Autonomous vision-based robotic exploration and mapping using hybrid maps and particle filters. *Image and Vision Computing* 2009;27:167–77.
18. He Z, Sun H, Hou J, Ha Y, Schwertfeger S. Hierarchical topometric representation of 3D robotic maps. *Auton Robot* 2021;45:755-71.
19. Niloy A, Shama A, Chakrabortty RK, et al. Critical design and control issues of indoor autonomous mobile robots: A Review. *IEEE Access* 2021;9:35338-70.
20. Liu J, Li Y, Tian X, Sangaiah AK, Wang J. Towards semantic sensor data: an ontology approach. *Sensors (Basel)* 2019;19:1193.
21. Günther M, Wiemann T, Albrecht S, Hertzberg J. Model-based furniture recognition for building semantic object maps. *Artificial Intelli gence* 2017;247:336–51.
22. Lim GH, Suh IH, Suh H. Ontology-based unified robot knowledge for service robots in indoor environments. *IEEE Trans Syst , Man, Cybern A* 2011;41:492-509.
23. Rusu RB, Marton ZC, Blodow N, Holzbach A, Beetz M. Model-based and learned semantic object labeling in 3D point cloud maps of kitchen environments. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE; 2009. pp. 3601–8.
24. Galindo C, Saffiotti A. Inferring robot goals from violations of semantic knowledge. *Robotics and Autonomous Systems* 2013;61:1131–43.
25. Wang T, Chen Q. Object semantic map representation for indoor mobile robots. In: Proceedings 2011 International Conference on System Science and Engineering. IEEE; 2011. pp. 309–13.
26. Vasudevan S, Siegwart R. Bayesian space conceptualization and place classification for semantic maps in mobile robotics. *Robotics and Autonomous Systems* 2008;56:522–37.
27. Diab M, Pomarlan M, Beßler D, et al. An ontology for failure interpretation in automated planning and execution. In: Iberian Robotics conference. Springer; 2019. pp. 381–90.
28. Balakirsky S. Ontology based action planning and verification for agile manufacturing. *Robotics and ComputerIntegrated Manufacturing*

2015;33:21–28.
29. Garg S, Sünderhauf N, Dayoub F, et al. Semantics for robotic mapping, perception and interaction: A Survey. *FNT in Robotics* 2020;8: 1-224.
30. Manzoor S, Rocha YG, Joo SH, et al. Ontology-based knowledge representation in robotic systems: a survey oriented toward applications. *Applied Sciences* 2021;11:4324.
31. Tenorth M, Beetz M. KnowRob: A knowledge processing infrastructure for cognition-enabled robots. *The International Journal of Robotics Research* 2013;32:566–90.
32. Beßler D, Pomarlan M, Beetz M. Owl-enabled assembly planning for robotic agents. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems; 2018. pp. 1684–92.
33. Schlenoff C, Prestes E, Madhavan R, et al. An IEEE standard ontology for robotics and automation. In: 2012 IEEE/ RSJ International Conference on Intelligent Robots and Systems. IEEE; 2012. pp. 1337–42.
34. Stenmark M, Malec J. Knowledge-Based Industrial Robotics. In: SCAI; 2013. pp. 265–74.
35. Bruno B, Chong NY, Kamide H, et al. The CARESSES EU-Japan project: making assistive robots culturally competent. In: Italian Forum of Ambient Assisted Living. Springer; 2017. pp. 151–69.
36. Waibel M, Beetz M, Civera J, et al. Roboearth. *IEEE Robot Automat Mag* 2011;18:69-82.
37. Saxena A, Jain A, Sener O, et al. Robobrain: Large-scale knowledge engine for robots. arXiv preprint arXiv:14120691 2014.
38. Dogmus Z, Erdem E, Patoglu V. RehabRobo-Onto: Design, development and maintenance of a rehabilitation robotics ontology on the cloud. *Robotics and Computer-Integrated Manufacturing* 2015;33:100-9.
39. Gonçalves PJS, Torres PMB. Knowledge representation applied to robotic orthopedic surgery. *Robotics and Computer-Integrated Manufacturing* 2015;33:90-9.
40. Bayat B, Bermejo-Alonso J, Carbonera J, et al. Requirements for building an ontology for autonomous robots. *IR* 2016;43:469-80.
41. IEEE standard ontologies for robotics and automation. *IEEE Std 18722015* 2015:1–60.
42. Crespo J, Barber R, Mozos O. Relational model for robotic semantic navigation in indoor environments. *J Intell Robot Syst* 2017;86:617 -39.
43. Crespo J, Barber R, Mozos O, BeBler D, Beetz M. Reasoning Systems for Semantic Navigation in Mobile Robots. In: 2018 IEEE/ RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE; 2018. pp. 5654–59.
44. Protégé. Protégé; Access May, 2020. Available from: https://protege.stanford.edu.
45. Reasoner H. Hermit Reasoner; Access May, 2020. Available from: http://www.hermit-reasoner.com.
46. Gangemi A, Borgo S, Catenacci C, Lehmann J. Task taxonomies for knowledge content. *Metokis Deliverable D* 2004;7:2004.
47. Fox M, Long D. PDDL2. 1: An extension to PDDL for expressing temporal planning domains. *jair* 2003;20:61-124.
48. Edelkamp S, Hoffmann J. PDDL2. 2: The language for the classical part of the 4th international planning competition. Technical Report 195, University of Freiburg; 2004.
49. Gerevini A, Long D. Plan constraints and preferences in PDDL3. Technical Report 2005-08-07, Department of Electronics for Automa- tion …; 2005.
50. Cashmore M, Fox M, Long D, et al. Rosplan: Planning in the robot operating system. In: Proceedings of the International Conference on Automated Planning and Scheduling. vol. 25; 2015.
51. Gayathri R, Uma V. Ontology based knowledge representation technique, domain modeling languages and planners for robotic path planning: A survey. *ICT Express* 2018;4:69–74.
52. Gonçalves PJS, Lourenço B, Santos S, Barlogis R, Misson A. Computer vision intelligent approaches to extract human pose and its activity from image sequences. *Electronics* 2020;9:159.

Ortiz *et al*. *Intell Robot* 2021;1(2):131-50
DOI: 10.20517/ir.2021.09

Intelligence & Robotics

Research Article

Open Access

Autonomous navigation in unknown environment using sliding mode SLAM and genetic algorithm

Salvador Ortiz, Wen Yu

Departamento de Control Automatico, National Polytechnic Institute, Mexico City 07360, Mexico.

Correspondence to: Prof. Wen Yu, Departamento de Control Automatico, National Polytechnic Institute, Mexico City 07360, Mexico. E-mail: yuw@ctrl.cinvestav.mx

How to cite this article: Ortiz S, Yu W. Autonomous navigation in unknown environment using sliding mode SLAM and genetic algorithm. *Intell Robot* 2021;1(2):131-50. http://dx.doi.org/10.20517/ir.2021.09

Received: 9 Sep 2021 **First Decision:** 28 Oct 2021 **Revised:** 15 Nov 2021 **Accepted:** 2 Dec 2021 **Published:** 15 Dec 2021

Academic Editors: Simon X. Yang, Hao Zhang **Copy Editor:** Yue-Yue Zhang **Production Editor:** Yue-Yue Zhang

Abstract

In this paper, sliding mode control is combined with the classical simultaneous localization and mapping (SLAM) method. This combination can overcome the problem of bounded uncertainties in SLAM. With the help of genetic algorithm, our novel path planning method shows many advantages compared with other popular methods.

Keywords: Autonomous navigation, sliding mode, SLAM, genetic algorithm

1. INTRODUCTION

1.1. Autonomous navigation in unknown environment

Autonomous navigation (AN) has three jobs [1].

(1) Perception: Mapping from signal to information is the perception of AN [2]. Its algorithms can use human thought [3], intelligent methods [4], optimization [5], probability methods [6], and genetic algorithms [7].

(2) Motion planning: It has three classes, namely graph methods such as a roadmap [8], random sampling [9], and grid [10].

(3) Localization and mapping: In unknown environments, sensors, actuators, and maps may have big uncertainties.

© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Path planning (PP) can be performed under the following conditions:

(1) The environment is known. PP is an optimization problem [11–13].

(2) The environment is partially known. PP can find new objects during navigation [14,15].

(3) The environment is totally unknown. PP depends on the navigation and has a recursive solution [16–18].

Simultaneous localization and mapping (SLAM) can be used in unknown environments [19] or in partially unknown environments [20]. SLAM [21] uses the current position to construct a map, and it can be classified into feature-based [22], pose-based [23], appearance-based [24], and variants [25].

The most popular SLAM uses Kalman filter [21] for Gaussian noise. Nonlinear SLAM uses extended Kalman filter (EKF) [26], where the noise assumptions are not satisfied [27]. EKF-SLAM applies linearization [28].

1.2. Related work

Few AN uses SLAM. Visual SLAM uses several cameras [29]. AN can use both SLAM and GPS signals [30]. Robots can avoid moving obstacles using neural networks [31]. Swarm optimization helps robots follow an object [32]. Neural networks help robots construct the navigation path [33]. The optimal path is considered in the sense of trajectory length, execution time, or energy consumption.

Genetic algorithms (GA) have been developed recently [34,35]. They are easy to use for optimization in non-deterministic cases [36], uncertainty models [37], and robust cases [38]. GA can be in form of ant-based GA [39,40], cell decomposition GA [41], potential field GA [42], ant colony [43], and particle swarm optimization [44]. Finite Markov chain is a theory tool for GA [45,46].

1.3. Our work

In this paper, we try to design AN in an unknown environment in real time. The contributions are as follows:

(1) Sliding mode SLAM: The robustness of this SLAM is better than other SLAM models in bounded noise.

(2) GA SLAM: We use roadmap PP and GA to generate the local optimal map.

(3) Comparisons and simulations with other SLAM models were made by using a mobile robot [47].

2. SLIDING MODE SLAM

SLAM gives the robot position and environment map at the same time. At time k, the state is $\mathbf{x}_k^r = (x_k, y_k, \theta_k)$, where (x_k, y_k) is the position and θ_k is the orientation of the robot. $\mathbf{x}_k^m = \left(\mathbf{m}_k^1, \mathbf{m}_k^2, \ldots, \mathbf{m}_k^L\right)^T$ are landmarks, with $\mathbf{m}_k^i = (x_k^i, y_k^i)^T$ the ith landmark. We assume the true location is time-invariant.

$\mathbf{x}_k$ has two parts: the robot $\mathbf{x}_k^r$ and the landmarks $\mathbf{x}_k^m$. The state equation is

$$\mathbf{x}_{k+1} = \begin{pmatrix} \mathbf{x}_{k+1}^r \\ \mathbf{x}_{k+1}^m \end{pmatrix} = \begin{pmatrix} \mathbf{f}(\mathbf{x}_k^r, \mathbf{u}_k) + \mathbf{w}_k \\ \mathbf{x}_k^m \end{pmatrix} = \mathbf{F}(\mathbf{x}_k, \mathbf{u}_k) + [\mathbf{w}_k, 0]^T \tag{1}$$

where $\mathbf{f}()$ is the robot dynamics, $\mathbf{w}_k$ is the noise, and $\mathbf{u}_k$ is the robot control. Since $\mathbf{x}_k^m$ is not influenced by motion noise, the noise is $[\mathbf{w}_k, 0]^T$.

$\mathbf{z}_k$ is defined as the position between the robot and the landmark, whose model is

$$\mathbf{z}_k^i = \mathbf{h}(\mathbf{x}_k^r, \mathbf{m}_k^i) + \mathbf{v}_k^i \tag{2}$$

where $\mathbf{h}()$ is the geometry and $\mathbf{v}_k^i$ is the noise. Here, $\mathbf{w}_k$ and $\mathbf{v}_k^i$ are not Gaussian noises. We assume $\mathbf{w}_k$ and $\mathbf{v}_k^i$ are bounded.

To estimate $\mathbf{x}_k$ in Equations (1) and (2), EKF is needed. We linearize the state model in Equation (1) and the observation model in Equation (2) as

$$\begin{aligned} &\mathbf{x}_{k+1} = \mathbf{F}(\hat{\mathbf{x}}_k, \mathbf{u}_k) + \nabla\mathbf{F}_k \cdot \left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right) \\ &+O_1\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k)^2\right] + [\mathbf{w}_k, 0]^T \\ &\mathbf{z}_k^i = \mathbf{h}(\hat{\mathbf{x}}_k) + \nabla\mathbf{h}_k \cdot \left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right) + O_2\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k)^2\right] + \mathbf{v}_k^i \end{aligned} \tag{3}$$

where $\nabla\mathbf{F}_k = \frac{\partial\mathbf{F}}{\partial\mathbf{x}_k}|_{\mathbf{x}_k=\hat{\mathbf{x}}_k}$, $\nabla\mathbf{h}_k = \frac{\partial\mathbf{h}}{\partial\mathbf{x}_k}|_{\mathbf{x}_k=\hat{\mathbf{x}}_k}$, $O_1\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k)^2\right]$, $\hat{\mathbf{x}}_k$ is the estimation of $\mathbf{x}_k$.

Prediction. The estimation $\hat{\mathbf{x}}_{k+1}$ is based on past states, control, and landmarks:

$$\begin{aligned} &\hat{\mathbf{x}}_{k+1} = \mathbf{F}(\hat{\mathbf{x}}_k, \mathbf{u}_k) \\ &\mathbf{P}_{k+1} = \nabla\mathbf{F}_k\mathbf{P}_k\nabla\mathbf{F}_k^T + R_1 \end{aligned} \tag{4}$$

where R_1 is the covariance of $\mathbf{w}_k$, $R_1 = E\left\{[\mathbf{w}_k - E(\mathbf{w}_k)]\,[\mathbf{w}_k - E(\mathbf{w}_k)]^T\right\}$.

Correction. The new state is based on predicted states, landmarks, and current observations:

$$\begin{aligned} &\hat{\mathbf{x}}_{k+2} = \hat{\mathbf{x}}_{k+1} + \mathbf{K}_{k+1}\left[\mathbf{z}_{k+1}^i - \mathbf{h}(\hat{\mathbf{x}}_{k+1})\right] \\ &\mathbf{K}_{k+1} = \mathbf{P}_{k+1}\nabla\mathbf{h}_{k+1}\left[\nabla\mathbf{h}_{k+1}\mathbf{P}_{k+1}\nabla\mathbf{h}_{k+1}^T + R_2\right]^{-1} \\ &\mathbf{P}_{k+2} = \left[I - \mathbf{K}_{k+1}\nabla\mathbf{h}_{k+1}\right]\mathbf{P}_{k+1} \end{aligned} \tag{5}$$

The motivations of using sliding mode modification to the EKF bases SLAM based are the following:

(1) The noises $\mathbf{w}_k$ and $\mathbf{v}_k^i$ in Equations (1) and (2) are not Gaussian.

(2) There are linearization error terms, $O_1\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k)^2\right]$ and $O_2\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k)^2\right]$, in Equation (3), and the traditional EKF-based methods do not work well for these errors.

We use the sliding mode method to estimate the robot state $\mathbf{x}_k^r$ and the landmark $\mathbf{x}_k^m$.

Sliding modes have a number of attractive features, and thus have long been in use for solving various control problems. The basic idea behind design of system with sliding mode is the following two steps: (1) a sliding motion in a certain sense is obtained by an appropriate choice of discontinuity surfaces; and (2) a control is chosen so that the sliding modes on the intersection of those discontinuity surface would be stable. A general class of discontinuous control $u(x,t)$ is defined by the following relationships:

$$u(x,t) = \begin{cases} u^+(x,t) & \text{with } \mathbf{s}(\mathbf{x}) > 0 \\ u^-(x,t) & \text{with } \mathbf{s}(\mathbf{x}) < 0 \end{cases} \tag{6}$$

where the functions $u^+(x,t)$ and $u^-(x,t)$ are continuous.

The function $\mathbf{s}(\mathbf{x})$ is the discontinuity surface (subspace). The objective of the sliding mode control is to design some switching strategy of the continuous control $u^+(x,t)$ and $u^-(x,t)$, such that

$$\mathbf{s}(\mathbf{x}) = 0 \tag{7}$$

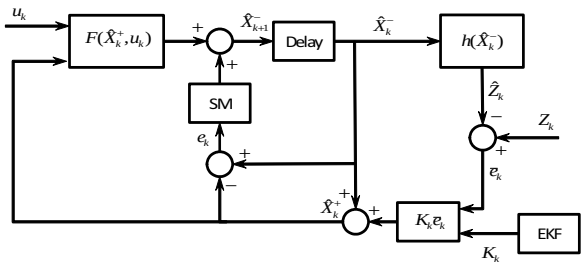


Figure 1. Sliding mode simultaneous localization and mapping.

In this paper, the sliding surface is defined by the SLAM estimation error as

$$e(k) = \mathbf{x}_k - \hat{\mathbf{x}}_k \tag{8}$$

Here, the discontinuity surface is $e(k) = [e_1 \cdots e_n]$. We consider the following positive definite function,

$$V = \frac{1}{2} e^T(k) P e(k) \tag{9}$$

where P is diagonal positive definite matrix, $P = P^T > 0$. The derivative of V is

$$\dot{V} = e^T(k) P \dot{e}(k) \tag{10}$$

The motion $e(k)$ satisfies

$$\dot{e}(k) = -\rho \times sgn[e(k)], \quad \rho > 0 \tag{11}$$

where $sgn[e(k)] = [sgn(e_1), \ldots, sign(e_n)]^T$, $sgn(e_i) = \begin{cases} 1 \text{ with } e_i(x) > 0 \\ -1 \text{ with } e_i(x) < 0 \end{cases}$, $sgn(0) = 0$, then (10) is

$$\dot{V} = e^T(k) P \{-\rho \times sgn[e(k)]\} = -\rho e^T(k) P sgn[e(k)]$$

because $P = diag\{p_i\}$, $p_i > 0$, and $e_i \times sgn(e_i) = |e_i|$

$$\dot{V} = -\rho \sum_{i=1}^{n} p_i |e_i| \tag{12}$$

Thus, $\dot{V} \leq 0$. By Barbalat's lemma[48], the estimation error is $e(k) \to 0$.

The classical SLAM in Equations (4) and (5) is modified by the sliding surface in Equation (11). The sliding mode control can be regarded as a compensator for Equation (4):

$$\hat{\mathbf{x}}_{k+1} = \mathbf{F}(\hat{\mathbf{x}}_k, \mathbf{u}_k) - \rho \times sgn[e(k)] \tag{13}$$

where ρ is a positive constant. The correction step is the same as EKF in Equation (5). The sliding mode SLAM is shown in Figure 1. Here, the estimation error, $e(k)$, is applied to the sliding surface to enhance the robustness in the prediction step with respect to the noise and disturbances.

It is the discrete-time version of Equation (6). We give the stability analysis of this discrete-time sliding mode SLAM at the end of this section.

For the mobile robot, the sliding mode SLAM can be specified as follows. We define a critical distance $d_{\min}$ to limit the maximal landmark density. It can reduce false positives in data association and avoid overload with

useless landmarks. If the new landmark is far from the other landmarks on the map, then the landmark is added; otherwise, it is ignored. If the distance between the new landmark $\mathbf{x}_{k+1} = [x_{m+1}, y_{m+1}]$ and the others is bigger than $d_{\min}$, it should be added into $\mathbf{x}_k$, *i.e.*,

$$\mathbf{x}_{k+1} = g(\mathbf{x}_k^r, \mathbf{z}_k) \tag{14}$$

It can be transformed into an absolute framework as

$$\mathbf{x}_{k+1} = \begin{pmatrix} \mathbf{x}_k \\ g(\mathbf{x}_k^r, \mathbf{z}_k) \end{pmatrix} = \mathbf{T}(\mathbf{x}_k, \mathbf{z}_k) \tag{15}$$

The nonlinear transformation function $\mathbf{T}$ also applies to the uncertainties. We approximate the transformation $\mathbf{T}$ by the linearization. $\mathbf{P}_k$ can be expressed as

$$\mathbf{P}_k = \begin{pmatrix} \mathbf{P}_k^r & \mathbf{P}_k^{rm} & \mathbf{0} \\ \left(\mathbf{P}_k^{rm}\right)^T & \mathbf{P}_k^m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_k \end{pmatrix} \tag{16}$$

where

$$\mathbf{P}_k = \nabla\mathbf{T}\mathbf{P}_k\nabla\mathbf{T}^T$$

with $\nabla\mathbf{T} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m & \mathbf{0} \\ \nabla\mathbf{g}_x & \mathbf{0} & \nabla\mathbf{g}_z \end{pmatrix}$, $\nabla\mathbf{g}_x := \frac{\partial\mathbf{g}}{\partial\mathbf{x}_k^r}(\mathbf{x}_k, \mathbf{z}_k)$, $\nabla\mathbf{g}_z := \frac{\partial\mathbf{g}}{\partial\mathbf{z}}(\mathbf{x}_k, \mathbf{z}_k)$.

For the motion part, we use the Ackerman vehicle model[49]

$$\begin{pmatrix} x_k^r \\ y_k^r \\ \theta_k^r \end{pmatrix} = \begin{pmatrix} x_{k-1}^r + T_{k-1}v_{k-1}\cos\theta_{k-1}^r \\ y_{k-1}^r + T_{k-1}v_{k-1}\sin\theta_{k-1}^r \\ \theta_{k-1}^r + T_{k-1}\frac{v_{k-1}}{b_a}\tan\alpha_{k-1} \end{pmatrix} + \mathbf{w}_k \tag{17}$$

where $\mathbf{w}_k$ is the process noise, v_k is the linear velocity, α_k is the steering angle, T_k is the sample time, and b_a is the distance between the front and the rear wheels.

At the beginning of map building, the vector $\hat{\mathbf{x}}_k$ only contains the robot states without landmarks. As exploration increases, the robot detects landmarks and decides if it should add these new landmarks to the state.

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{T}(\mathbf{x}_k, \mathbf{z}_k)\begin{pmatrix} \mathbf{x}_{k,x}^r \\ \mathbf{x}_{k,y}^r \end{pmatrix} + r_k^j \begin{pmatrix} \cos(\theta_k^i + \mathbf{x}_{k,\phi}^r) \\ \sin(\theta_k^i + \mathbf{x}_{k,\phi}^r) \end{pmatrix} \\ \mathbf{z}_k &= \begin{pmatrix} \sqrt{(m_x^i - x_k)^2 + (m_y^i - y_k)^2} \\ \arctan\left(\dfrac{m_y^i - y_k}{m_x^i - x_k)}\right) - \phi_k \end{pmatrix}_i + \mathbf{V}_k \end{aligned} \tag{18}$$

where $\mathbf{x}$, y, $\mathbf{z}$, and m are defined in Equations (1) and (2),

We exploit the same property in the sliding SLAM. The landmarks with fewer corrections are removed from the state vector.

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + G_x^T \begin{bmatrix} u_{k,v}\delta_t\cos(\mathbf{x}_{k,\phi}^r) \\ u_{k,v}\delta_t\sin(\mathbf{x}_{k,\phi}^r) \\ u_{k,\gamma}\delta_t \end{bmatrix} + \sigma_k \tag{19}$$

where $G_x = \begin{pmatrix} 1 & 0 & 0 & 0\cdots 0 \\ 0 & 1 & 0 & 0\cdots 0 \\ 0 & 0 & 1 & \underbrace{0\cdots 0}_{2N} \end{pmatrix}$, σ_k is the compensator, and

$$\sigma_k = -\rho \times sgn\left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right) \tag{20}$$

This sliding SLAM algorithm is given in the following algorithm.

Sliding mode SLAM. $\hat{x}_1 = 0, P_{1|1} = \alpha I, k = 1, \alpha \gg 1$ u_1 =get_controls, z_1 =get_observations; $k_z = 1$ $\left[\hat{\mathbf{x}}_1, \mathbf{P}_1\right]$ =add_features$\left(\hat{\mathbf{x}}_1, \mathbf{P}_1, \mathbf{z}_1\right)$ (1) While not_stop if controls_are_available $\left[\hat{\mathbf{x}}_{k+1}, \mathbf{P}_{k+1}\right]$ =prediction$\left(\hat{\mathbf{x}}_k, \mathbf{P}_k, \mathbf{u}_k\right)$ (2) u_k =get_controls end if if observations_are_available get_observations z_k data_association$\left(\mathbf{z}_k, \hat{\mathbf{x}}_{k+1}, \mathbf{P}_{k+1}\right)$ $\left[\hat{\mathbf{x}}_{k+2}, \mathbf{P}_{k+2}, \mathbf{c}_k\right] = \left(\hat{\mathbf{x}}_{k+1}, \mathbf{P}_{k+1}, \mathbf{z}_k\right)$ (5) $\left[\hat{\mathbf{x}}_{k+2}, \mathbf{P}_{k+2}\right] = \left(\hat{\mathbf{x}}_{k+2}, \mathbf{P}_{k+2}, \mathbf{z}_k\right)$ (1) $k_z = k_z + 1$ end if if $\mathrm{mod}(k_z, K_z) = 0$ $\left[\hat{\mathbf{x}}_{k+2}, \mathbf{P}_{k+2}\right]$ =pruning$\left(\hat{\mathbf{x}}_{k+2}, \mathbf{P}_{k+2}, \mathbf{c}_k, \mathbf{a}_k\right)$ end if $k = k + 1$ end While

---

The discrete-time sliding mode SLAM in Equation (19) can be written as

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \hat{F}(\hat{\mathbf{x}}_k, \mathbf{u}_k) + \sigma_k$$

where $\hat{F} = G_x^T \begin{bmatrix} u_{k,v}\delta_t \cos(\mathbf{x}^r_{k,\phi}) \\ u_{k,v}\delta_t \sin(\mathbf{x}^r_{k,\phi}) \\ u_{k,\gamma}\delta_t \end{bmatrix}$, $e(k) = \mathbf{x}_k - \hat{\mathbf{x}}_k$, $\sigma_k = \rho \times sgn\left[e(k)\right]$

The correction step for $\hat{\mathbf{x}}_{k+2}$ is the same as EKF:

$$\begin{aligned} \hat{\mathbf{x}}_{k+2} &= \hat{\mathbf{x}}_{k+1} + \mathbf{K}_{k+1}\left[\mathbf{z}^i_{k+1} - \mathbf{h}(\hat{\mathbf{x}}_{k+1})\right] \\ \mathbf{K}_{k+1} &= \mathbf{P}_{k+2}C_{k+1}\left[C_{k+1}\mathbf{P}_{k+2}C^T_{k+1} + R_2\right]^{-1} \\ \mathbf{P}_{k+2} &= \left[I - \mathbf{K}_{k+1}C_{k+1}\right]\mathbf{P}_{k+2} \end{aligned} \tag{21}$$

where $C_k = \nabla\mathbf{h}_k = \frac{\partial \mathbf{h}}{\partial \mathbf{x}_k}|_{\mathbf{x}_k = \hat{\mathbf{x}}_k}$.

The error dynamic of this discrete-time sliding mode observer is

$$e(k+1) = A_k e(k) - A_k\mathbf{K}_k C_k e(k) + \sigma_k + d_k \tag{22}$$

where $d_k = \hat{F}(\hat{\mathbf{x}}_k, \mathbf{u}_k) + \xi_k$ is bounded uncertainty, $\|d_k\|^2 \le \bar{d}$, $A_k = \nabla\mathbf{F}_k = \frac{\partial \mathbf{F}}{\partial \mathbf{x}_k}|_{\mathbf{x}_k = \hat{\mathbf{x}}_k}$, and $\mathbf{K}_k$ is the gain of EKF in Equation (21).

The next theorem gives the stability of the discrete-time sliding mode SLAM.

Theorem 1 *If the gain of the sliding mode SLAM is positive, then the estimation error is stable, and the estimation error converges to*

$$\|e(k)\|^2 \le \frac{\lambda_{\max}\left[\mathbf{P}^{-1}_{k+1}\right](\bar{\rho} + \bar{s}) + \bar{\rho}}{\alpha\lambda_{\min}\left[\mathbf{P}^{-1}_{k+2}\right]} \tag{23}$$

where $\|\sigma_k\|^2 \le \bar{\rho}, \sigma_k$ $\|d_k\|^2 \le \bar{s}$, $\mathbf{P}_{k+2}$ *is the gain of EKF in Equation (21),* $0 < \alpha = \frac{1}{(1+\bar{p}\bar{a}^2/\underline{q})(1+\bar{k}\bar{c}+\lambda)} < 1$, $\underline{p}I \le \mathbf{P}_{k+2} \le \bar{p}I$, *and* $\underline{q}I \le R_1$.

Proof 1 *Consider the Lyapunov function as*

$$V_k = e(k)\,\mathbf{P}_k^{-1} e(k) \tag{24}$$

where $\mathbf{P}_{k+2}$ *is the prior covariance matrix in Equation (21), and* $\mathbf{P}_{k+2} > 0$. *From Equation (22),*

$$\begin{aligned} &V_{k+1} = e(k+1)\,\mathbf{P}_{k+1}^{-1} e(k+1) \\ &= e(k)\,(I - \mathbf{K}_k C_k)^T A_k^T \left(\mathbf{P}_{k+1}\right)^{-1} A_k (I - \mathbf{K}_k C_k) e(k) \\ &+2(d_k + \sigma_k^T)\mathbf{P}_{k+1}^{-1}\left[A_k (I - \mathbf{K}_k C_k) e(k)\right] \\ &+(d_k + \sigma_k^T)\mathbf{P}_{k+1}^{-1}(d_k + \sigma_k) \end{aligned} \tag{25}$$

Because $\|\sigma_k\|^2 \le \bar{\rho}$, $\|d_k\|^2 \le \bar{s}$, *the last term on the right side of Equation (25) is*

$$(d_k + \sigma_k^T)\mathbf{P}_{k+1}^{-1}(d_k + \sigma_k) \le \lambda_{\max}\left[\mathbf{P}_{k+1}^{-1}\right](\bar{\rho} + \bar{s}) \tag{26}$$

where $\lambda_{\max}\left[\mathbf{P}_{k+1}^{-1}\right]$ *is the maximum eigenvalue of* $\mathbf{P}_{k+1}^{-1}$.

The second term of Equation (25) is

$$\begin{aligned} &2(d_k + \sigma_k^T)\mathbf{P}_{k+1}^{-1}\left[A_k (I - \mathbf{K}_k C_k) e(k)\right] \\ &= 2(d_k +)\mathbf{P}_{k+1}^{-1}\left[A_k (I - \mathbf{K}_k C_k) e(k)\right] \\ &+\sigma_k^T \mathbf{P}_{k+1}^{-1}\left[A_k (I - \mathbf{K}_k C_k) e(k)\right] \end{aligned} \tag{27}$$

where $\mathbf{K}_k$ *is the gain of EKF in Equation (5). In view of the matrix inequality*

$$X^T Y + \left(X^T Y\right)^T \le X^T \Lambda^{-1} X + Y^T \Lambda Y \tag{28}$$

which is valid for any $X, Y \in \Re^{n\times k}$ *and for any positive definite matrix* $0 < \Lambda = \Lambda^T \in \Re^{n\times n}$, *the first term of Equation (27) is*

$$\begin{aligned} &2d_k \mathbf{P}_{k+1}^{-1}\left[A_k (I - \mathbf{K}_k C_k) e(k)\right] \\ &\le d_k \Lambda (d_k +) + e(k)\,\mathbf{P}_{k+1}^{-1}\left[A_k (I - \mathbf{K}_k C_k)\right]\Lambda^{-1} e(k) \\ &\le \bar{s}\lambda_{\max}[\Lambda] + \left\|\left(\mathbf{P}_{k+1}\right)^{-1}\left[A_k (I - \mathbf{K}_k C_k)\right]\Lambda^{-1}\right\| \|e(k)\|^2 \\ &\le \bar{s}\lambda_{\max}[\Lambda] + e(k)\left[\mathbf{P}_{k+1}^{-1}\left[A_k (I - \mathbf{K}_k C_k)\right]\Lambda^{-1}\right] e(k) \end{aligned} \tag{29}$$

We apply the sliding mode compensation in Equation (20) to the second term of Equation (27):

$$\begin{aligned} &\rho \times sgn\left[e(k)\right]_k^T \Upsilon_k e(k) \\ &= -\rho \sum_{k=1}^{m} |e(k)| \left(l_{kk} + \sum_{i=1, i\neq k}^{m} l_{ki} sign\left(\left[e(k)\right] e_i(k)\right)\right) \end{aligned} \tag{30}$$

where l_{ij} *are the elements of the matrix* Υ, $\Upsilon_k = \mathbf{P}_{k+1}^{-1}\left[A_k (I - \mathbf{K}_k C_k)\right]$. *When the the orientation* θ_k *is not big,* $\sin\theta_k \approx 0$, $\cos\theta_k \approx 0$,

$$l_{kk} \gg \sum_{i=1, i\neq k}^{m} |l_{ki}|, \quad l_{kk} > 0, \quad k = 1, \ldots m, \tag{31}$$

Thus, the second term of Equation (27) is negative.

The first term on the right side of Equation (25) has the following properties:

$$\mathbf{P}_{k+2} \ge (I - \mathbf{K}_k C_k)\mathbf{P}_{k+2}(I - \mathbf{K}_k C_k)^T$$

$(I - \mathbf{K}_k C_k)$ is invertible, and we have

$$\left(\mathbf{P}_{k+2}\right)^{-1} \leq (I - \mathbf{K}_k C_k)_k^{-T} \left(\mathbf{P}_{k+2}\right)^{-1} (I - \mathbf{K}_k C_k)^{-1} \tag{32}$$

According to EKF,

$$\mathbf{P}_{k+1} = A_k \mathbf{P}_{k+2} A_k^T + R_1 = A_k (\mathbf{P}_{k+2} + A_k^{-1} R_1 A_k^{-T}) A_k^T$$

Thus,

$$\left(\mathbf{P}_{k+1}\right)^{-1} = A_k^{-T} (\mathbf{P}_{k+2} + A_k^{-1} R_1 A_k^{-T})^{-1} A_k^{-1}$$

By the following matrix inversion lemma,

$$(\Gamma^{-1} + \Omega)^{-1} = \Gamma - \Gamma(\Gamma + \Omega^{-1})^{-1}\Gamma$$

where Γ and Ω are two non-singular. matrices,

$$\mathbf{P}_{k+1}^{-1} = A_k^{-T} [\mathbf{P}_{k+2}^{-1} - \mathbf{P}_{k+2}^{-1} (\mathbf{P}_{k+2}^{-1} + A_k^T Q^{-1} A_k)^{-1} \mathbf{P}_{k+2}^{-1}] A_k^{-1}$$

Using Equation (32) and defining $L = (I - \mathbf{K}_k C_k)$,

$$\begin{aligned} \mathbf{P}_{k+1}^{-1} &\leq A_k^{-T} L^{-T} [\mathbf{P}_{k+2}^{-1} \\ &- \left(\mathbf{P}_{k+2}\right)^{-1} L^{-1} (\mathbf{P}_{k+2}^{-1} + A_k^T R_1^{-1} A_k)^{-1} L^{-T} \mathbf{P}_{k+2}^{-1}] L^{-1} A_k^{-1} \end{aligned} \tag{33}$$

Now,

$$\mathbf{P}_{k+2}^{-1} = \mathbf{P}_{k+2}^{-1} (I - \mathbf{K}_k C_k)^{-1} = \mathbf{P}_{k+2}^{-1} L^{-1}$$

Hence,

$$L^T A_k^T \mathbf{P}_{k+1}^{-1} A_k L \leq (I - (I + \mathbf{P}_{k+2}^{-1} A_k^T Q^{-1} A_k)^{-1} L^{-T}) \mathbf{P}_{k+2}^{-1}$$

Combining the last term of Equation (29) with the first term on the right side of Equation (25),

$$\begin{aligned} & e(k) (I - \mathbf{K}_k C_k)^T A_k^T \mathbf{P}_{k+1}^{-1} A_k (I - \mathbf{K}_k C_k) e(k) \\ & \leq e(k) (1 - (1 + \bar{p}\bar{a}^2/\underline{q})^{-1} (1 + \bar{k}\bar{c} + \lambda)^{-1}) \mathbf{P}_{k+2}^{-1} e(k) \\ & \leq (1 - \alpha) \left\|\mathbf{P}_{k+2}^{-1}\right\| \|e(k)\|^2 \end{aligned} \tag{34}$$

where $\|A_k\| = \sqrt{tr(A_k A_k)} \leq \bar{a}$, $\|C_k\| = \sqrt{tr(C_k C_k)} \leq \bar{c}$, $\|\mathbf{K}_k\| = \sqrt{tr(\mathbf{K}_k \mathbf{K}_k)} \leq \bar{k}$, $\lambda = \left\|\Lambda^{-1}\right\|$, $\underline{p} I \leq \mathbf{P}_{k+2} \leq \bar{p} I$, $\underline{q} I \leq R_1$, and

$$\alpha = \frac{1}{(1 + \bar{p}\bar{a}^2/\underline{q})(1 + \bar{k}\bar{c} + \lambda)} < 1$$

Combining Equation (26), the first term of Equation (29), and Equation (34),

$$\begin{aligned} V_{k+1} &= (1 - \alpha) \left\|\mathbf{P}_{k+2}^{-1}\right\| \|e(k)\|^2 \\ &+ \lambda_{\max} [\Lambda] \bar{\rho} + \lambda_{\max} \left[\mathbf{P}_{k+1}^{-1}\right] (\bar{\rho} + \bar{s}) \\ &\leq (1 - \alpha) e(k) \mathbf{P}_{k+2}^{-1} e(k) \\ &+ \lambda_{\max} [\Lambda] \bar{\rho} + \lambda_{\max} \left[\mathbf{P}_{k+1}^{-1}\right] (\bar{\rho} + \bar{s}) \end{aligned} \tag{35}$$

Thus,

$$V_{k+1} - V_k \leq -\alpha V_k + \kappa$$

where $\kappa = \lambda_{\max}\left[\mathbf{P}_{k+1}^{-1}\right](\bar{\rho} + \bar{s}) + \lambda_{\max}[\Lambda]\,\bar{\rho}$. *If*

$$\alpha\lambda_{\min}\left[\mathbf{P}_{k+2}^{-1}\right]\|e(k)\|^2 \geq \kappa$$

then $V_{k+1} - V_k \leq 0$, $\|e(k)\|$ *decreases. Thus,* $\|e(k)\|$ *converges to Equation (23).*

3. GENETIC ALGORITHM AND SLAM FOR PATH PLANNING

Path planning is one key problem of autonomous robots. Here, the map is built by the sliding mode SLAM:

- The obstacle set is defined by $B_{obs}(t)$.
- The position is $x_r(t)$, $B_{free}(t) = B \backslash B_{obs}(t)$.
- The path planning is $f(x(t), x_S, x_T)$, $x_S = x_r(t)$.

The previous map is $B_{free}(t)$, which requires the path $f(x(t), x_S, x_T)$.

We assume the previous map is obstacle-free, the initial point is x_S, the target point is $x_T \in B$,

$$B_{free} = \{z_r \in B \mid A(z_r) \cap B_{obs} = \emptyset\}$$

the obstacle is $B_{obs} = \frac{B}{B_{f\ \ e}}$, z_r is the shape of the robot, and $A(z_r)$ is the area of the robot. The objective of the path planning is to find a path $f(x, x_S, x_T) \in B_{free}$ that allows the robot to navigate.

D is defined as the search space. We use the GA to find an optimal trajectory $f(x, x_S, x_T)$, such that

$$\min_{x \in D} f(x, x_S, x_T), \text{ where } f : D \to R \tag{36}$$

Here, we use stochastic search for GA, and each iteration includes: reproduction or selection, crossing or combination, and mutation. The population is $P(k) = \{S_1^k, S_2^k, .., S_m^k\}$ with m being the size of the population that represents the possible solutions:

(1) Every chromosome S_i^t has a solution in D

$$S_i^k = [\varsigma_l, \varsigma_{l-1}, \ldots, \varsigma_2, \varsigma_1] \text{ with } \varsigma_i \in D \ \ \forall i = 1, 2, \ldots, l$$

(2) Crossing the chromosomes. An intersection in $S_a^k = [\varsigma_l^a, \varsigma_{l-1}^a, \ldots, \varsigma_2^a, \varsigma_1^a]$ and $S_b^k = [\varsigma_l^b, \varsigma_{l-1}^b, \ldots, \varsigma_2^b, \varsigma_1^b]$ belongs to D, such that $S_a^t \cap S_b^t \neq \emptyset$; then,

$$S_{a'}^k = [\varsigma_l^a, \varsigma_{l-1}^a, \ldots \varsigma_i^{ab}, \ldots, \varsigma_2^b, \varsigma_1^b]$$
$$S_{b'}^k = [\varsigma_l^b, \varsigma_{l-1}^b, \ldots \varsigma_j^{ab}, \ldots, \varsigma_2^a, \varsigma_1^a]$$

where $S_{a'}^t$ and $S_{b'}^t$ are the next generation from two compatible chromosomes by crossing.

(3) Mutation. It replace a number of chromosomes by chromosomes in D.

The mutation operation is calculated by the fitness of each chromosome,
$P(Mut) = [fit(S_1^{Mut}), fit(S_2^{Mut}), \ldots, fit(S_n^{Mut})]$, where n is the number of mutations and fit uses the

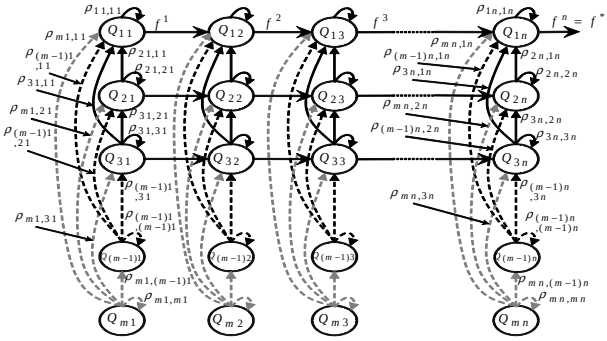


Figure 2. Roadmap genetic algorithm model as a finite Markov chain.

Euclidean distance. The total fitness is $Fit = \sum_{i=1}^{n} fit(S_i^{Mut})$. Therefore, the probability of selection p_i of a chromosome S_i for $i = 1, 2, \ldots, n$ is

$$p_i = \frac{fit(S_i^{Mut})}{Fit} \tag{37}$$

An optimal solution p^M in D is mutated by

$$p^M = \lim_{n\to\infty} (1 - p_i) = \lim_{n\to\infty} \left(1 - \frac{fit(S_i^M)}{\sum_{i=1}^{n} fit(S_i^M)} \right) = 1 \tag{38}$$

In the mutation operation, an optimal solution with $p^M = 1$ is a global solution if $n \to \infty$. To prove the convergence, we use a Markov chain, as shown in Figure 2. Each chromosome can move from Q_{ij} to the state $Q_{i(j+1)}$. The moving probability is $\rho_{ji,ik} > 0, i = 1, 2, \ldots, n, k, j = 1, 2, \ldots, m$.

The operators, selection, crossing, and mutation create $P(k)$ with p^k. It preserves the best chromosomes of $P(k-1)$. $P(k+1)$ in the population $P(k)$ can be regarded as the Markov transition:

$$H\{Q_{k+1} = p^{k+1} | Q_k = p^k\} = H(p^{k+1}, p^k) \tag{39}$$

Theorem 2 *If GA for the roadmap is an elitist process, then the probability of p^* in D is exponential.*

Proof 2 *The iteration Q_1 is changed with the chromosomes when genetic process is elitist,*

$$H(Q_1 = p^* \forall 0 < \tau \le n) = \sum_{i=2}^{n} \rho_{i1,11} = \frac{n-1}{n} \tag{40}$$

where n is the size of the population. If for all $\alpha, \beta \in D$, there is $0 < \tau \le m$ such that $H^\tau(\alpha, \beta) \ge \epsilon > 0$, then

$$\epsilon = \min\{H^\tau(\alpha, \beta) \forall 0 < \tau \le n\} \le 1 \tag{41}$$

This implies that, given certain state Q_t, the probability of transition in time t between t and $t + m$ is at least ϵ,

$$H(Q_t \ne p^* \forall t < \tau \le t + n) \ge 1 - \epsilon \tag{42}$$

Without loss of generality, the transition in the iteration $k + 1$ is

$$H(Q_{k+1}) = H(Q_t \ne p^* \forall 0 < t \le (k+1)n)$$
$$H(Q_t \ne p^* \forall 0 < t \le kmn) H(Q_t \ne p^* \forall kn < t \le (k+1)n)$$

Using Equation (42), we have

$$\begin{aligned}
&H(Q_{k+1}) \leq H(Q_t \neq f^* \forall 0 < t \leq km)(1-\epsilon) \\
&\leq H(Q_t \neq f^* \forall 0 < t \leq (k-1)m)(1-\epsilon)^2 \\
&\leq H(Q_t \neq f^* \forall 0 < t \leq 0) H(Q_t \neq f^* \forall 0 < t \leq m)(1-\epsilon)^k \\
&= \tfrac{1}{m}(1-\epsilon)^k
\end{aligned}$$

where $H(P_t \neq p^* \forall 0 < t \leq 0) = 1$, *then*

$$\begin{aligned}
&\lim_{k\to\infty} H(Q_{k+1}) \leq \lim_{k\to\infty} \tfrac{1}{n}(1-\epsilon)^k \\
&= \tfrac{1}{n} \lim_{k\to\infty} (1-\epsilon)^k = 0
\end{aligned}$$

Since $0 < \epsilon \leq 1$, *the algorithm converges exponentially to* p^* *in: population size* n *and iteration number* k.

The algorithm of the SLAM-based roadmap GA for the path planning is as follows.

SLAM based roadmap GA. (1) Initiate population randomly $P(k)$ of size n that belong in set D. (2) The fitness value is fit with Euclidean distance for each chromosome S_i. (3) The population is from lower to higher fitness: $fit(S_1) \geq fit(S_2) \geq \cdots \geq fit(S_M)$. (4) Crossing set in the chromosomes, $S_i \cap S_j \rightarrow S_{ij}, S_{ji}$. (5) Next population $P(k+1)$ is replaced by the chromosomes with poor skills. (6) Random mutation with poor skills. (7) Go to Step (2)

4. AUTONOMOUS NAVIGATION

Our AN uses both sliding mode SLAM (Algorithm 1) and the roadmap GA method (Algorithm 2). The autonomous navigation algorithm is:

Autonomous navigation. The initial state S_I, the target S_T $P^- = covariance(R_1)$, $P^+ = covariance(R_2)$ ρ = gain sliding mode, r_O = search radius Map =search_obstacles(S_I, r_O) S_n = Path_Planning(Map, S_I, S_T) U_0 = Controller(S_I, S_n) while $S_n \neq S_T$

$$\left[Map_k, \hat{X}_k\right] = SM_SLAM\left(S_k, S_n, U_k, P_k^-, P_k^+, \rho, z_k\right)$$

S_i = $\hat{X}_k$(end state) $Map\,(end : length(Map_k))$ = Map_k $[S_n$ $\theta_i]$ =Path_Planning(Map, S_i, S_T) if $\theta_i > \pi$ $S_n = ComputeOutsideLocal(Map, S_i, S_T, \theta_i)$ end if; U_{k+1} = Controller(S_i, S_n) end while return $S_n, \hat{X}_k$

The PP needs the map, robot position, and target. This information is given by the sliding mode SLAM algorithm. When the algorithm falls into a local solution, we use the $ComputeOutsideLocal$ function to provide another S_n which is outside the local zone.

5. COMPARISONS

In this section, we use several examples to compare our method with the three other recent methods: the polar histogram method for path planning[50], the grid method for path planning[51], and SLAM with extended Kalman filter[52].

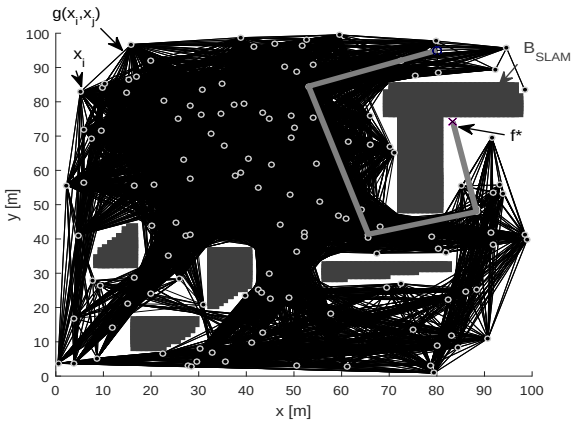


Figure 3. Sliding mode simultaneous localization and mapping.

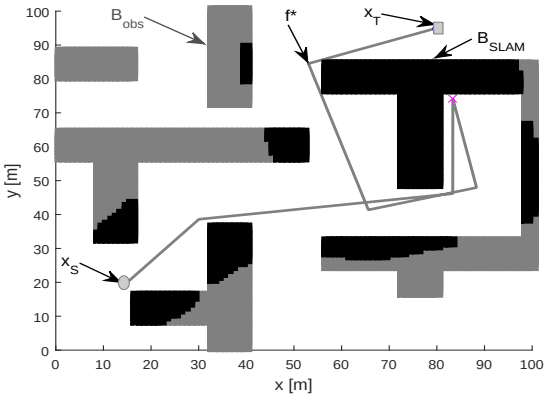


Figure 4. Autonomous navigation using sliding mode simultaneous localization and mapping and genetic algorithm method.

5.1. Simulations

The following simulations were implemented in partially unknown and completely unknown environments. The size of the environments was 100 m × 100 m, in which a solution was sought to find a trajectory from the initial point x_S to the target point x_T. The sliding mode gains were selected as $\rho = diag([0.1, \cdots, 0.1])$.

In the partially unknown environments, $B_{obs}(0) \neq \emptyset$. The path planning solution p^* was partial because the environment $B_{obs}(t)$ was variant in time. Figure 3 shows a partial solution p^* from an initial point x_S to the objective point x_T for the partially unknown environment. Figure 4 shows the overall result of the robot navigation from point x_S to point x_T with the robust SLAM algorithm combined with the GA.

Here, the SLAM algorithm was used to construct the environment and find the position of the robot. At the beginning of navigation in the partially unknown environment, there was a planned trajectory of navigation through the GA algorithm; however, if an obstacle was found in the planned trajectory, the GA algorithm needed to be used to search for a new trajectory within the built environment by the SLAM, B_{SLAM}. The planned trajectory belonged to the set of obstacles that prevent reaching the goal, $p^* \subset B_{obs}(t)$; therefore, it was necessary to look for a new trajectory using RGA that allowed reaching the goal.

For the completely unknown environments, $B_{obs}(0) = \emptyset$. In these environments, the SLAM algorithm was required to know the environment B_{SLAM} and the position of the robot; in this way, when an obstacle was found that contained the planned trajectory, a new trajectory with the GA algorithm was searched on the map

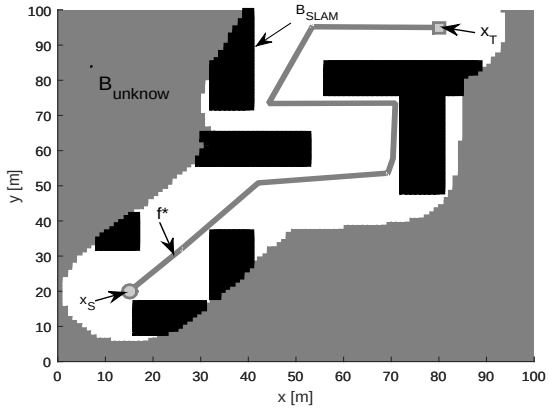


Figure 5. Sliding mode simultaneous localization and mapping and genetic algorithm in complete unknown environment.

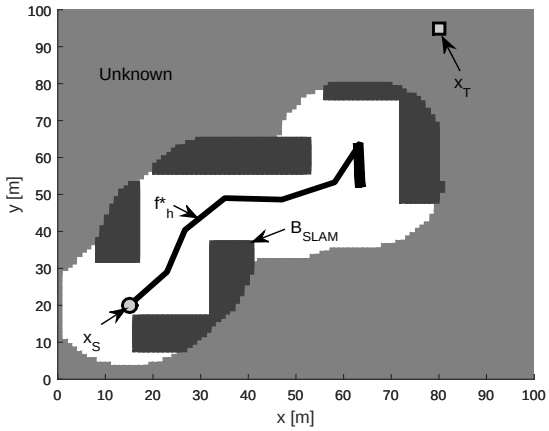


Figure 6. Polar histogram method in complete unknown environments.

B_{SLAM} until the target point was reached the results obtained are shown in Figure 5. When we used the polar histogram method for path planning[50], only the local solutions could be found [Figure 6].

Now, we compare the path lengths with the polar histogram method. The following density of the obstacles give the navigation complexity. The environment is free of obstacles when $d_{obs} = 0$. The whole environment is occupied by the obstacles when $d_{obs} = 1$. The index for the trajectory error is

$$E_{PP} = \frac{effective\ path - optimal\ path}{100} \tag{43}$$

We use the averages of the path length. The obstacles density is defined as

$$d_{obs} = \frac{\sum_{n \in G_O} S_{obs}(n)}{\|G_E\|} \tag{44}$$

The path length is defined as

$$l_{sub} = \frac{effective\ path}{optimal\ path} \tag{45}$$

The averages of the path lengths of our RA and the polar histogram are shown in Figure 7. When the density

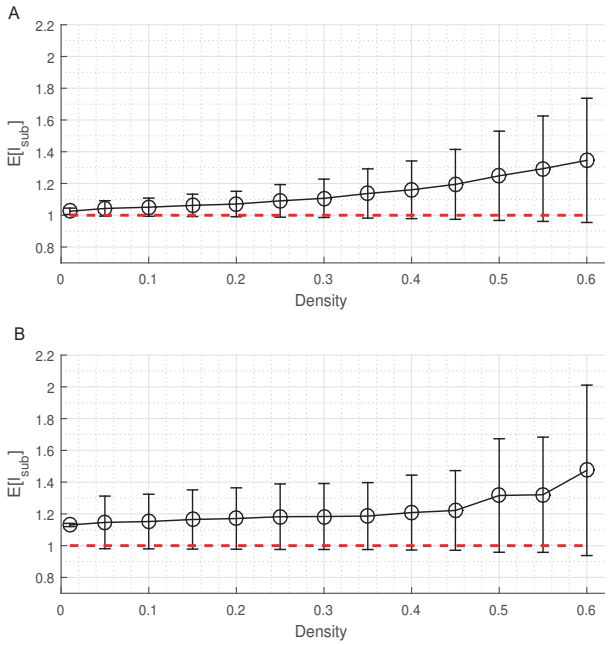


Figure 7. The average path length: (A) sliding mode simultaneous localization and mapping and genetic algorithm; and (B) polar histogram.

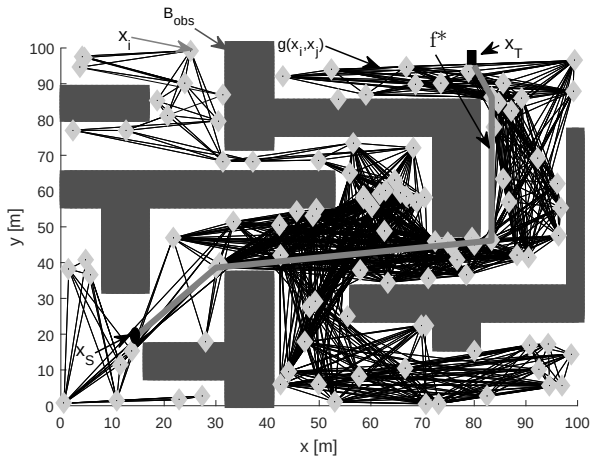


Figure 8. Sliding mode simultaneous localization and mapping (gray) and grid method (black)

of obstacles was bigger, the path length of the polar histogram grew more quickly than that of ours. When the obstacle density d_{obs} was 0.3, $E[l_{sub}, Navigation1] = 1.053$, $E[l_{sub}, Navigation2] = 1.152$.

Next, we compare our method with the grid method[51]. The comparison results are shown in Figure 8. For the task of navigating the robot or system in partially unknown or completely unknown environments, the SLAM algorithm was used to construct the environment and know the position of the robot. At the beginning of navigation in the partially unknown environment, there was a planned trajectory of navigation through the GA algorithm; however, if an obstacle were found in the planned trajectory, the GA algorithm needed to be used to search for a new trajectory within the built environment by the SLAM, B_{SLAM}.

The size of the environments was 100 m × 100 m, in which a solution was sought to find a trajectory from the initial point x_S to the target point x_T. Figure 9 shows a path planning based on the proposed methods to find

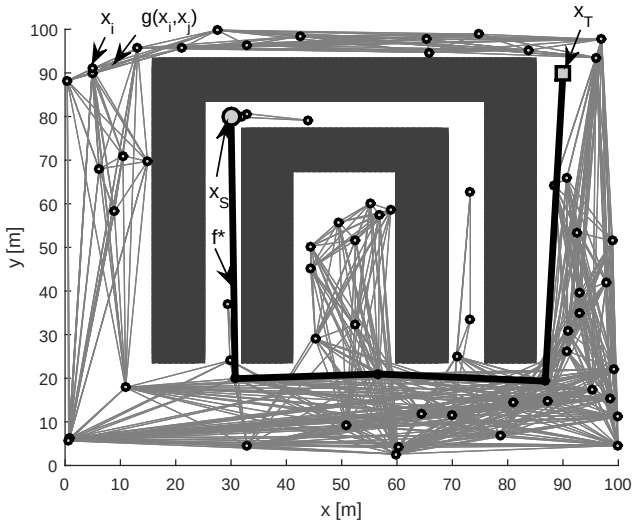


Figure 9. Sliding mode simultaneous localization and mapping based path planning (bold) and grid method (gray).

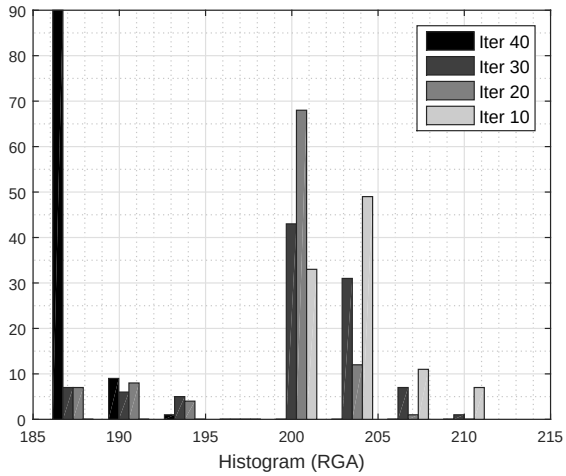


Figure 10. Performance of the genetic algorithm.

a solution f^*. Here, 60 local targets were generated x_i with the search space D generated by the trajectories of the local targets that do not intersect with the set of obstacles; thus, it became an optimization problem to find an optimal path.

In an environment with previously generated obstacles of 100 m × 100 m, 120 possible targets were randomly generated x_i; therefore, the search space D would be all trajectories $g(x_i, x_j)$ that do not intersect with the set of obstacles, where the roadmap genetic algorithm solved the problem of optimization to find a solution to the problem of path planning. For the problem presented above, we found that the proposed algorithm converged in 40 iterations. For these results, 100 tests were performed for each number of iterations and, as shown in Figure 10, the roadmap genetic algorithm converged with greater probability within 40 iterations.

5.2. Application

The Koala mobile robot by K-team Corporation 2013 was used to validate our sliding mode SLAM. This mobile robot has encoders and one laser range finder. The position precision is less than 0.1 m.

Figure 11. The environment of the autonomous navigation.

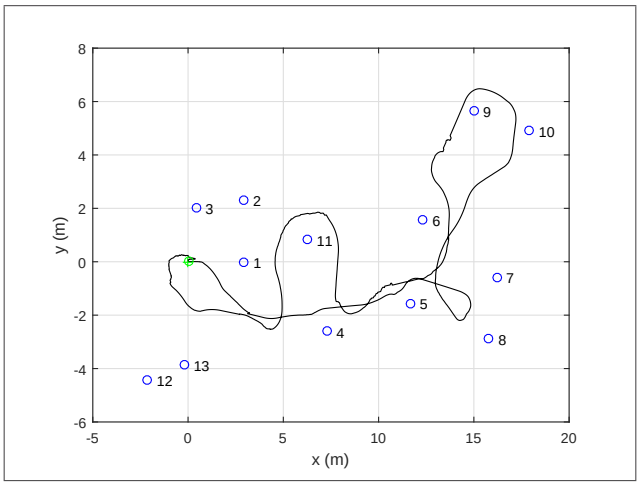


Figure 12. Results of extended Kalman filter simultaneous localization and mapping and sliding mode simultaneous localization and mapping with small noises.

The objective of this autonomous navigation is to force the robot to return to the starting point. The sliding mode SLAM was compared with SLAM with extended Kalman filter (EKF-SLAM) [52].

The initial covariance matrices are zero. The parameters of the algorithm are

$$\rho = diag([1e^{-3}, 1e^{-3}, 4e^{-3}, 2e^{-4}, \cdots, 2e^{-4}])$$
$$R_1 = diag([0.05, 0.05, 0.005]), R_2 = diag([6e^{-4}, 1e^{-5}])$$

Since the robot moves in the environment with bounded noise (see Figure 11), the noises are not Gaussian. Two different conditions are considered: (1) Koala robot pre-processes off-line the sensors data to reduce 90% noises; and (2) the computer uses sliding mode SLAM on-line.

For the first case, Figure 12 shows that sliding mode SLAM and EKF-SLAM are similar. Both sliding mode SLAM and EKF-SLAM work well for the case with less noise. The robot can return to the starting point, and the map is constructed correctly.

For the second case, Figure 13 gives the results of EKF-SLAM. As can be seen, the robot cannot return to the starting point and the map is not exactly the same as the real one with EKF-SLAM because EKF-SLAM is sensitive to non-Gaussian noises.

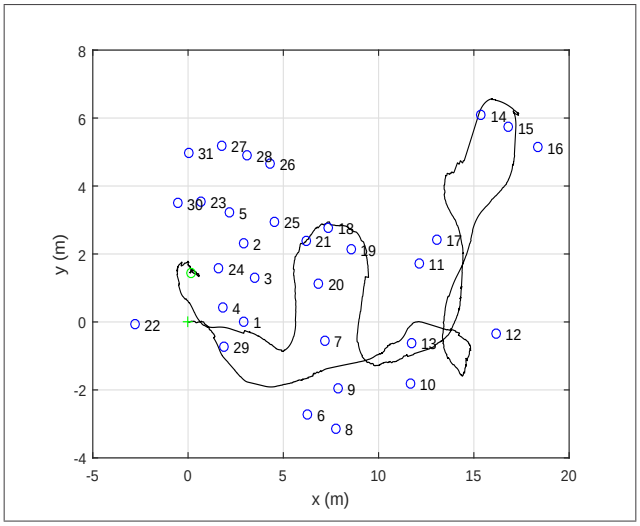


Figure 13. Results of extended Kalman filter simultaneous localization and mapping with noises.

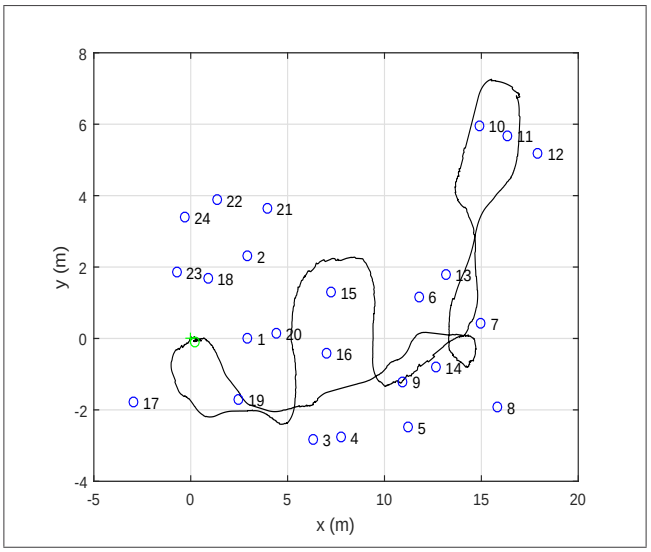


Figure 14. Results of sliding mode simultaneous localization and mapping with noise.

Figure 14 shows the results with SM-SLAM. Under the same bounded noises, SM-SLAM works very well, because of the sliding mode technique.

To compare the errors, we define the average of the Euclidean errors as

$$E_d = \frac{1}{N_T}\sum_{k=1}^{N_T}\sqrt{(x_k - x_k^*)^2 + (y_k - y_k^*)^2}, E_a = \frac{1}{N_T}\sum_{k=1}^{N_T}\left|\phi_k - \phi_k^*\right| \tag{46}$$

where N_T is the data number; x_k^*, y_k^*, and ϕ_k^* are real values for robot position and orientation; and x_k, y_k, and ϕ_k are estimations of them. Figure 15 shows the errors of EKF-SLAM and sliding mode SLAM. Obviously, the errors of EKF-SLAM increase quickly. Robots have better estimation in long distances with sliding mode SLAM.

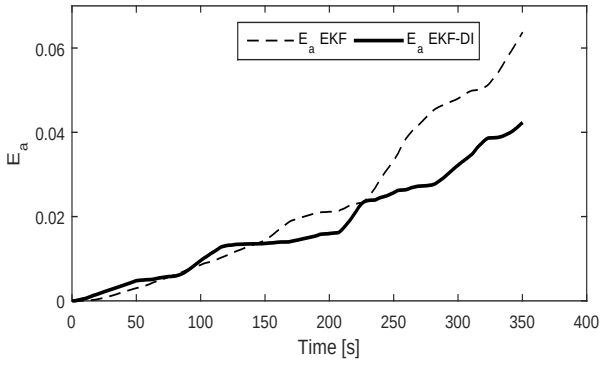


Figure 15. Direction estimation errors of sliding mode simultaneous localization and mapping (SLAM) and EKF-SLAM. EKF: Extended Kalman filter.

6. CONCLUSION

Navigation in unknown environments is a big challenge. In this paper, we propose sliding mode SLAM with genetic algorithm for path planning. Both sliding mode and GA can work in unknown environments. Convergence analysis is given. Two examples were applied to compare our model with other models, and the results show that our algorithm is much better in unknown environments.

DECLARATIONS

Authors' contributions

Revised the text and agreed to the published version of the manuscript: Ortiz S, Yu W

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

Both authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2021.

REFERENCES

1. Luettel T, Himmelsbach M, Wuensche HJ. Autonomous ground vehicles—concepts and a path to the future. *Proceedings of the IEEE* 2012;100:1831-39.
2. Galceran E, Carreras M. A survey on coverage path planning for robotics. *Rob Auton Syst* 2013;61:1258-76.
3. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 2004;60:91-110.

4. Arel I, Rose DC, Karnowski TP. Deep machine learning-a new frontier in artificial intelligence research. *IEEE Comput Intell Mag* 2010;5:13-8.
5. Scheirer WJ, de Rezende Rocha A, Sapkota A, et al. Toward open set recognition. *IEEE Trans Pattern Anal Mach Intell* 2012;35:1757-72.
6. Ramos FT, Kumar S, Upcroft B, et al. A natural feature representation for unstructured environments. *IEEE Trans Robot* 2008;24:1329-40.
7. Lillywhite K, Lee DJ, Tippetts B, et al. A feature construction method for general object recognition. *Pattern Recognition* 2013;46:3300-14.
8. Chabini I, Lan S. Adaptation of the A* algorithm for the computation of fastest paths in deterministic discrete-time dynamic networks. *IEEE trans Intell Transp Syst* 2002;3:60-74.
9. LaValle SM, Kuffner JJ Jr. Randomized kinodynamic planning. *Int J Rob Res* 2001;20:378–400.
10. Valencia R, Andrade-Cetto J. Mapping, planning and exploration with Pose SLAM. Berlin: Springer; 2018. p. 60-84.
11. Wang X, Shi Y, Ding D, et al. Double global optimum genetic algorithm–particle swarm optimization-based welding robot path planning. *Eng Optim* 2016;48:299-316.
12. Jones M, Peet MM. A generalization of Bellman's equation with application to path planning, obstacle avoidance and invariant set estimation. *Automatica* 2021;127:109510.
13. Rehman NU, Kumar K, Abro GeM. Implementation of an autonomous path plan-ning & obstacle avoidance UGV using SLAM. 2018 International Conference on Engineer-ing and Emerging Technologies (ICEET); 2018 Feb 22-23; Lahore, Pakistan. IEEE; 2018. p. 1-5.
14. de Moura Souza G, Toledo CFM. Genetic algorithm applied in UAV's path planning. 2020 IEEE Congress on Evolutionary Computation (CEC); 2020 Jul 19-24; Glasgow, UK. IEEE; 2020. p. 1-8.
15. Zhang X, Zhao Y, Deng N, et al. Dynamic path planning algorithm for a mobile robot based on visible space and an improved genetic algorithm. *Int J Adv Robot Syst* 2016;13:91.
16. Clemens J, Reineking T, Kluth T. An evidential approach to SLAM, path planning, and active exploration. *Int J Approx Reason* 2016;73:1-26.
17. Chen Y, Huang S, Fitch R. Active SLAM for mobile robots with area coverage and obstacle avoidance. *IEEE ASME Trans Mechatron* 2020;25:1182-92.
18. da Silva Arantes M, Toledo C F M, Williams B C, et al. Collision-free encoding for chance-constrained nonconvex path planning. *IEEE Trans Robot* 2019;35:433-48.
19. Yu W, Zamora E, Soria A. Ellipsoid SLAM: a novel set membership method for simultaneous localization and mapping. *Autonomous Robots* 2016;40:125-37.
20. Williams H, Browne WN, Carnegie DA. Learned action slam: sharing slam through learned path planning information between heterogeneous robotic platforms. *Appl Soft Comput* 2017;50:313-26.
21. Dissanayake MWMG, Newman P, Clark S, et al A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Trans Rob Autom* 2001;17:229-41.
22. Thrun S, Liu Y, Koller D, et al. Simultaneous localization and mapping with sparse extended information filters. *Int J Robot Res* 2004;23:693-716.
23. Folkesson J, Christensen HI. Closing the loop with graphical SLAM. *IEEE Trans Robot* 2007;23:731-41.
24. Ho KL, Newman P. Loop closure detection in SLAM by combining visual and spatial appearance. *Rob Auton Syst* 2006;54:740-9.
25. Nieto J, Guivant J, Nebot E. Denseslam: simultaneous localization and dense mapping. *Int J Robot Res* 2006;25:711-44
26. Chen SY. Kalman filter for robot vision: a survey. *IEEE Trans Ind Electron* 2011;59:4409-20.
27. Sibley G, Matthies L, Sukhatme G. Sliding window filter with application to planetary landing. *J Field Robot* 2010;27:587-608.
28. Kaess M, Johannsson H, Roberts R, et al. iSAM2: Incremental smoothing and mapping using the Bayes tree. *Int J Robot Res* 2012;31:216-35.
29. Yang S, Scherer SA, Yi X, et al. Multi-camera visual SLAM for autonomous navigation of micro aerial vehicles. *Rob Auton Syst* 2017;93:116-34
30. Weiss S, Scaramuzza D, Siegwart R. Monocular-SLAM–based navigation for autonomous micro helicopters in GPS-denied environments. *J Field Robot* 2011;28:854-874
31. Zhu A, Yang SX. Neurofuzzy-based approach to mobile robot navigation in unknown environments. *IEEE Trans Syst Man Cybern C Appl Rev* 2007;37:610-21.
32. Juang CF, Chang YC. Evolutionary-group-based particle-swarm-optimized fuzzy controller with application to mobile-robot navigation in unknown environments. *IEEE Trans Fuzzy Syst* 2011;19:379-92.
33. Villacorta-Atienza JA, Makarov VA. Neural network architecture for cognitive navigation in dynamic environments. *IEEE Trans Neural Netw Learn Syst* 2013;24:2075-87.
34. Song B, Wang Z, Sheng L. A new genetic algorithm approach to smooth path planning for mobile robots. *Assem Autom* 2016;36:138-45.
35. Zhang Y, Gong D, Zhang J. Robot path planning in uncertain environment using multi-objective particle swarm optimization. *Neurocomputing* 2013;103:172-85.
36. Karami AH, Hasanzadeh M. An adaptive genetic algorithm for robot motion planning in 2D complex environments. *Comput Electr Eng* 2015;43:317-29.
37. Pereira AGC, de Andrade BB. On the genetic algorithm with adaptive mutation rate and selected statistical applications. *Comput Stat* 2015;30:131-50.
38. Tsai CC, Huang HC, Chan CK. Parallel elite genetic algorithm and its application to global path planning for autonomous robot navigation. *IEEE Trans Ind Electron* 2011;58:4813-21.

39. Zhu Q, Hu J, Cai W, et al. A new robot navigation algorithm for dynamic unknown environments based ondynamic path re-computation and an improved scout ant algorithm. *Appl Soft Comput* 2011;11:4667-76.
40. Arantes MS, Arantes JS, Toledo CFM, et al. A hybrid multi-population genetic algorithm for uav path planning. Proceedings of the Genetic and Evolutionary Computation Conference 2016; 2016 Jul 20. New York, NY: Association for Computing Machinery; 2016. p. 853-60.
41. Tuncer A, Yildirim M. Dynamic path planning of mobile robots with improved genetic algorithm. *Comput Electr Eng* 2012;38:1564-72.
42. Raja R, Dutta A, Venkatesh KS. New potential field method for rough terrain path planning using genetic algorithm for a 6-wheel rover. *Rob Auton Syst* 2015;72:295-306.
43. Arora T, Gigras Y, Arora V. Robotic path planning using genetic algorithm in dynamic environment. *Int J Comput Appl* 2014;89:9-12.
44. Roberge V, Tarbouchi M, Labonté G. Comparison of parallel genetic algorithm and particle swarm optimization for real-time UAV path planning. *IEEE Trans Industr Inform* 2012;9:132-41.
45. Bhandari D, Murthy CA, Pal SK. Genetic algorithm with elitist model and its convergence. *Intern J Pattern Recognit Artif Intell* 1996;10:731-47.
46. Hu ZB, Xiong SW, Su QH, et al. Finite Markov chain analysis of classical differential evolution algorithm. *J Comput Appl Math* 2014;268:121-34.
47. K-Team Corporation. Available from: http://www.k-team.com/ [Last accessed on 10 Dec 2021]
48. Souahi A, Naifar O, Makhlouf AB, et al. Discussion on Barbalat Lemma extensions for conformable fractional integrals. *Int J Control* 2019;92:234-41.
49. Erdman AG, Sandor GN, Kota S. Mechanism design: analysis and synthesis. Upper Saddle River, NJ: Prentice hall; 2001.
50. Wilde N, Kulić D, Smith SL. Learning user preferences in robot motion planning through interaction. 2018 IEEE International Conference on Robotics and Automation (ICRA); 2018 May 21-25; Brisbane, QLD, Australia. IEEE; 2018. p. 619-26.
51. Farzan S, DeSouza GN. Path planning in dynamic environments using time-warped grids and a parallel implementation. *arXiv:1903.07441* 2019.
52. Saeedi S, Nardi L, Johns E, et al. Application-oriented design space exploration for SLAM algorithms. 2017 IEEE International Conference on Robotics and Automation (ICRA); 2017 Jul 24; Singapore. IEEE; 2017. p. 5716-23.

Ji *et al. Intell Robot* 2021;1(2):151-75
DOI: 10.20517/ir.2021.14

Intelligence & Robotics

Review

Open Access

Rail track condition monitoring: a review on deep learning approaches

Albert Ji[1], Wai Lok Woo[2], Eugene Wai Leong Wong[1], Yang Thee Quek[3]

[1]NewRIIS, Newcastle University, Singapore 609607, Singapore.
[2]Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, UK.
[3]School of Engineering, Republic Polytechnic Singapore, Singapore 738964, Singapore.

Correspondence to: Prof. Wai Lok Woo, Department of Computer and Information Sciences, Northumbria University, Ellison Place, Newcastle upon Tyne NE1 8ST, UK. E-mail: wailok.woo@northumbria.ac.uk

How to cite this article: Ji A, Woo WL, Wong EWL, Quek YT. Rail track condition monitoring: a review on deep learning approaches. *Intell Robot* 2021;1(2):151-75. https://dx.doi.org/10.20517/ir.2021.14

Received: 21 Oct 2021 **First Decision:** 23 Nov 2021 **Revised:** 12 Dec 2021 **Accepted:** 29 Dec 2021 **Published:** 31 Dec 2021

Academic Editors: Simon X. Yang, Xin Xu **Copy Editor:** Xi-Jun Chen **Production Editor:** Xi-Jun Chen

Abstract

Rail track is a critical component of rail systems. Accidents or interruptions caused by rail track anomalies usually possess severe outcomes. Therefore, rail track condition monitoring is an important task. Over the past decade, deep learning techniques have been rapidly developed and deployed. In the paper, we review the existing literature on applying deep learning to rail track condition monitoring. Potential challenges and opportunities are discussed for the research community to decide on possible directions. Two application cases are presented to illustrate the implementation of deep learning to rail track condition monitoring in practice before we conclude the paper.

Keywords: Rail track maintenance, condition monitoring, anomaly detection and classification, deep learning

1. INTRODUCTION

The rail industry plays an important role in a nation's economy and development and directly affects the lifestyle of the residents. Hence, there is a low tolerance level by the public to any accidents or negative events happening to the rail operations as the economy, the livelihood, and the country's reputation would be brought down and the social and political risk level will rise. The rail systems around the world operate under different environments with their most critical infrastructure, the steel rail track, including rails, sleepers, ballast, fastener, and subgrade. Undesirable consequences such as derailment, death, injury,

© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

economic burden and loss of public confidence could be caused by the defects on the rail track and the failure of rail tracks[1]. It was also reported that rail maintenance workers were injured or lost life during rail inspection and maintenance operations. Thus, safe railway operations demand effective maintenances which rely heavily on inspection and monitoring of rail track conditions. The condition monitoring is fundamentally critical to the safety, reliability and cost-efficiency of the rail operations[2]. There are also regulations by governments on the frequency of regular track inspections which traditionally requires a large amount of personnel and equipment resources. Therefore, the rail track condition monitoring is of great importance due to the safety, economic and regulatory factors.

Rail defects are normally initiated by loads and stresses applied to rails in longitudinal, transverse, or vertical directions. The vehicle wheels can apply vertical, lateral, and creep loads to the rails, while bulk stress such as bending stress, thermal stress, and residual stresses can be applied to the rails as well. Different rail defects such as rail corrugations, rolling contact fatigue defects, squat defects, shatter cracking, split head, and wheel burns have their own causes and characteristics, lead to different effects, and thus require corresponding treatments[3]. Therefore, proper detection and classification of rail defects play an important role in effective rail maintenance operations. In practice and research, some common rail defects may be selected and targeted according to Pareto's principle as some types of rail track defects are commonly found such as rail corrugation, transverse cracks, shelling, and wheel burns. The train wheels contact with the rail track and frictions in between will gradually cause rail corrugation over a period of time where crests, troughs, and waves remain on the track and will become worse and worse. The rail tracks will have concave deform on the top due to the corrugations, which also cause the rail track lifespan to shorten and, therefore, a replacement will be needed. Around the faulty weld joint areas, which could be caused by the difference in weld material or a manufacturing flaw in the rail, transverse cracks may form. Another cause could be the welding processes, such as arc welding, pores, inclusions, and misalignments. Around the gauge corners of the rail tracks, there could be subsurface fatigue, which causes the loss of materials and then shelling defects. Later on, shelling cracks will develop inside the area which can usually be seen as dark spots on the outside of the gauge corners. However, in the beginning, the cracks may grow so fast that unforeseen failures appear before the crack defects are detected. The train wheels may slide quite often on the railway tracks, which will raise the rail surface temperature to be very high and therefore cause the wheel burn defects, which are usually found in pairs, opposite of one another on the two rail tracks. The high temperature will drop down quickly, which makes the rail track in the brittle martensite phase according to material science. Wheel burns could be found on the surface of rail tracks and might appear similar to squat defects. Wheel burns are usually found in pairs, opposite of one another on two rails.

Defects and deteriorated conditions on the rail track can normally be seen by human eyes; therefore, manual inspection by patrollers can identify and locate the defects and monitor the condition. However, such inspections are labor-intensive and can only be arranged during non-operating hours in order not to disrupt the regular service operations. Ultrasound testing, magnetic particle testing, radiographic testing, eddy current testing, and penetrating testing are the common non-destructive testing (NDT) methods to measure the surface and internal parameters or performance of the tested object without destroying it. Among these testing methods, ultrasound testing and eddy current testing are more suitable for the train tracks. The ultrasound testing method is the most effective rail track NDT inspection method. It utilizes the propagation and attenuation characteristics of sound waves in the medium and the reflection and refraction characteristics on the interface. It can detect the internal defects and find cracks on bolt holes, head, and web and the longitudinal crack at the bottom of the rail. Due to rail steel being ferromagnetic, the eddy currents do not penetrate into the material. Therefore, they flow along the crack side so that the pocket length of the cracks can be determined in the railhead by eddy current testing. The eddy current testing

method is particularly suitable to inspect steel rails even under high temperatures[4]. Due to the harsh environment, late hours, and tired patrollers, the inspection accuracy might be affected[5]. Rail inspection vehicle and sensor technologies are being deployed as an efficient and cost-effective data collection technology solution to support rail maintenance operations and can capture vast amounts of data. Data-driven automatic condition monitoring and detection and classification of rail track anomalies have been attracting attention from researchers at universities and railway institutes.

Deep learning methods are producing successes in various applications with the recent advances of the techniques. Deep learning has neural networks as its functional unit to mimic how the human brain solves complex problems based on data. Methods such as long-short-term memory (LSTM) as a type of recurrent neural network (RNN) and convolutional neural network (CNN) propelled the development of deep learning and the field of artificial intelligence and have been reported with convincing performances in monitoring conditions for tools, machines, and turbines. The performances in prediction and learning of these methods are improving with the increasing amount of data available[6-8]. Deep learning methods have been adopted for rail track condition monitoring and anomaly detection and classification.

Some research questions are formulated to guide our review with clear purposes. Subsequently, we select the publications for the detailed review.

• What types of deep learning models are available for rail track condition monitoring?

• What deep learning techniques can be useful for applications in rail track condition monitoring?

• What types of rail track anomalies are more commonly chosen to identify?

• What types of data are collated for the deep learning applications?

• Where are the specific objectives of applying deep learning models to rail track condition monitoring?

• What are the deep learning data pre-processing methods adopted?

• What are the challenges that researchers face?

• How does the application in rail track condition monitoring correspond to the evolution of deep learning techniques?

• What are the trend and insights for future directions of research and practice?

Figure 1 shows the review framework that is proposed by this paper to address the list of research questions. The importance, the types of defects, and the existing manual inspection techniques of rail track monitoring are presented to give the context and introduction of this study. The shortcomings of manual inspection techniques partially provide the need to adopt deep learning methods. We then review the deep learning methods available and their relevance by briefly discussing the evolution of the deep learning field and describing the deep learning models for the ease of selecting suitable models for tasks. The studies applying deep learning methods to rail track condition monitoring are then reviewed where summaries are made according to the trend over time, the region of study, the raw data type, the pre-processing data, the purpose

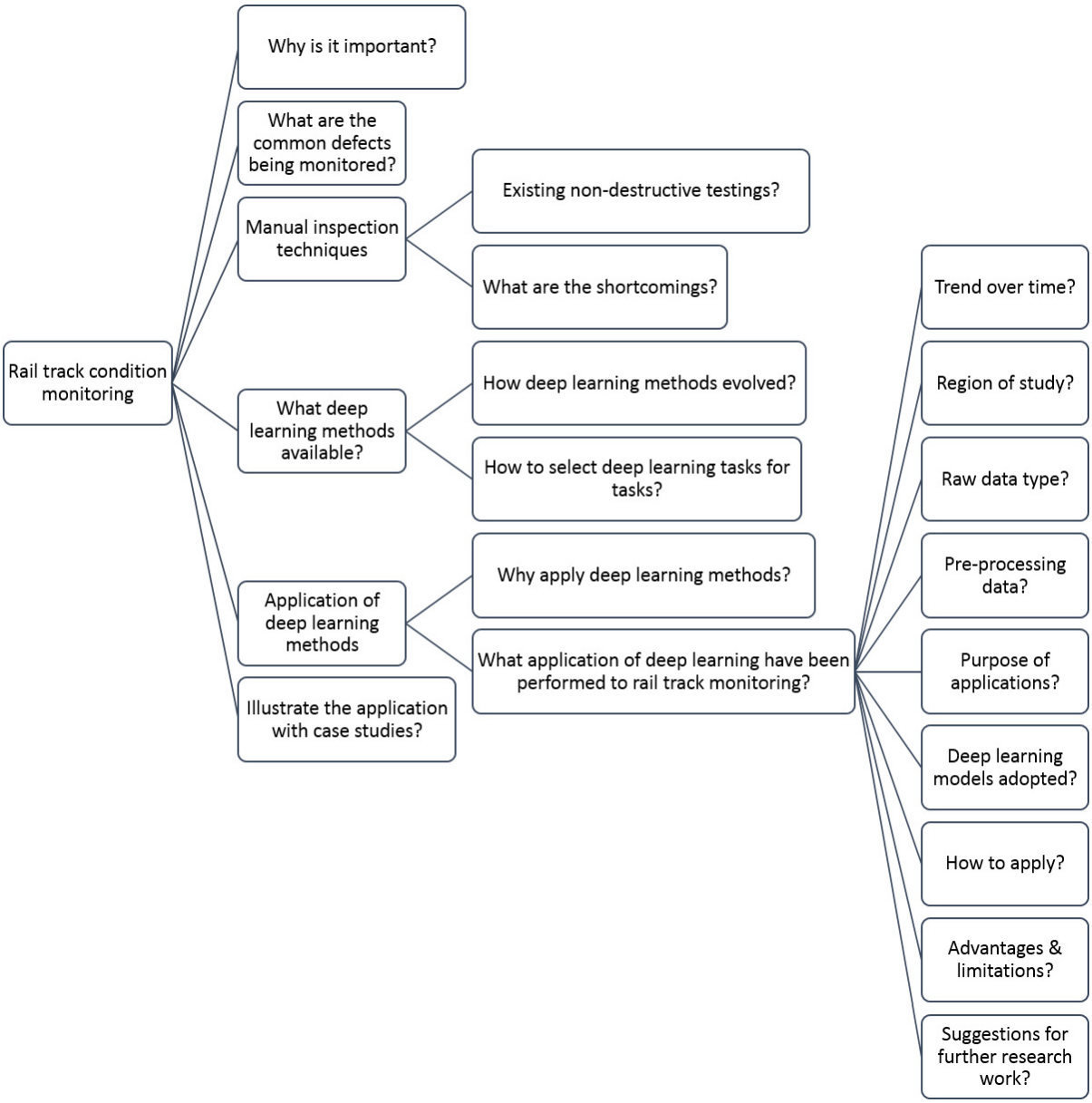


Figure 1. Review framework.

of applications, and the deep learning models adopted. How to apply these deep learning techniques for applications and the advantage and limitations of deep learning methods are discussed before some possible further research areas are outlined as well. Illustrative case studies are also included to show the practical considerations of applying deep learning methods to rail track condition monitoring. This review framework provides a balance between deep learning methods and their application to rail track condition monitoring. The scope is only about condition monitoring of rail tracks; we do not review other rail components such as rolling stocks or pantographs. Our studies also focus on the deep learning methods instead of more broadly artificial intelligence or machine learning approaches. The review framework caters to the needs of both practitioners who need to solve operational issues and researchers who might have more interests in the methodologies.

Through this paper, the authors intend to provide a useful guide to researchers and practitioners who are interested in applying deep learning methods to rail track condition monitoring tasks. There are three major contributions of this paper. First, the evolution of deep learning and a collection of relevant deep learning methods provide clear coverage of the relevance, usefulness, and applicability of deep learning methods, which enable fellow researchers to navigate through the usually rather daunting deep learning domain with confidence. Second, a systematic search and review of the application publications proves the relevance of deep learning methods to rail track condition monitoring tasks and provides insights into how such research works are carried out and what potential further studies can be followed up. Third, two illustrative case studies demonstrate practical considerations and aim to motivate wider and more creative adoption of deep learning methods to rail industries. This paper is organized as follows. Section 2 describes a historical overview of deep learning and briefly introduces common deep learning models. Section 3 reviews research adopting deep learning methods for rail track condition monitoring and anomaly detection. Section 4 discusses challenges and opportunities. Section 5 presents case studies applying deep learning to rail anomaly detection and classification while Section 6 concludes the paper.

2. DEEP LEARNING MODELS

2.1. Historical overview of deep learning

We provide a simplified timeline for deep learning and its evolution. Important issues and development at critical junctures are highlighted. A modern definition of deep learning describes a current understanding of the topic. Multiple layers of a deep learning model learn to represent the data with abstractions at multiple levels. The intricate structure of the large input data is discovered through the computations at each layer. Each layer computes its own representation from the representation of its previous layer according to the deep learning model's internal parameters which are updated using the backpropagation algorithm. Images, video, speech, and audio data can be processed by deep convolutional nets while sequential data such as text and speech by recurrent nets[9]. In the following paragraphs, we examine the journey of deep learning from a single neuron to the current status and hence determine the scope of the following review work.

The McCulloch-Pitts (MCP) neuron proposed in 1943 was the first computational model mimicking the functionality of a biological neuron which marks the start of the era of artificial neural networks. An aggregation of Boolean inputs determines the output through a threshold parameter[10]. The classical perceptron model[11] proposed in 1958 was further refined and analyzed[12] in 1969. The perceptron model brought in the concept of numerical weights to measure the importance of inputs and a mechanism for learning those weights. The model is similar to but more generalized than the MCP neuron as it takes weighted real inputs and the threshold value is learnable. As a single artificial neuron is incapable of implementing some functions such as the XOR logical function, larger networks also have similar limitations which cooled down the artificial neural network development.

The multi-layer perceptron (MLP) was proposed in 1986 where node outputs of hidden layers are calculated using sigmoid function and biogeography based optimization is used to find the weights of the network model[13]. The universal approximation theorem of MLP, proved in 1989, states that, for any given function f(x), there is a backpropagation neural network that can approximately approach the result[14]. The LeNet network was proposed in 1989 to recognize handwritten digits with good performances[15]. In 1991, with the backpropagation neural network, the vanishing gradient problem was discovered, that is back-propagated error signals either shrink rapidly or grow out of bounds in typical deep or recurrent networks because certain activation functions, such as the sigmoid function, take a large input space but have a small output space between 0 and 1[16]. The LSTM model was proposed in 1997[17] and performs well in predicting sequential data. However, since then, neural networks had not been progressing well until 2006. It is worth mentioning that statistical learning theory, a framework for machine learning, blossomed between 1986 and

2006. Methods and models such as decision trees[18], support vector machines (SVM)[19], AdaBoost[20], kernel SVM[21], and random forests[22] were proposed. Graphical models were proposed in 2001 to provide a description framework for various machine learning methods such as SVM, naïve Bayes, and hidden Markov model[23].

Complementary priors were introduced in 2006 to eliminate the vanishing gradient problem that makes inference difficult in densely connected belief nets that have many hidden layers[24]. The rectified linear activation function (ReLU) was introduced in 2011 and became the default activation function of many neural networks. ReLU outputs the input directly if it is positive; otherwise, it is zero. ReLU is effective in tackling the vanishing gradient problems[25].

In 2012, AlexNet, a large deep convolutional neural network, was developed and trained to participate in the ImageNet large scale visual recognition challenge (ILSVRC) for the first time and delivered state-of-the-art results which drew attention from researchers[26]. AlexNet was the pioneer of using the graphics processing unit (GPU) to train the neural network. In the following years, deep learning became more and more popular, the architecture and training methods were improved rapidly, and the hardware advanced quickly and became more powerful. Deep learning has since been adopted by more industries including the railway industry and delivered more meaningful results. This paper focuses on reviewing research works of deep learning applications to rail track condition monitoring since 2013.

2.2. Common deep learning models

Artificial intelligence, machine learning, and deep learning have developed rapidly in recent years. There are many more deep learning networks than one can practically remember. As the resources to learn a particular deep learning method are abundant, we only list some deep learning methods in this section to provide an overview of the techniques available for practitioners and researchers to select and provide brief introductions about the methods. More detailed guides on implementations can be found from the abundance of references available.

The most commonly heard neural network names are probably CNN and RNN. CNN might be noted as ConvNet. The architecture of a CNN was inspired by the organization of the visual cortex and is analogous to that of the connectivity pattern of neurons in the human brain. Individual neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. A collection of such fields overlaps to cover the entire visual area. CNN takes in an input image, assigns importance (learnable weights and biases) to various aspects/objects in the image, and can differentiate one from the other. In RNN, which was derived from feedforward neural networks, nodes are connected to form a directed graph along a temporal sequence to exhibit temporal dynamic behavior. RNN's internal states (memory) are utilized to process variable-length sequences of inputs. A typical RNN architecture is LSTM which has feedback connections and can process both single data points (such as images) and entire sequences of data (such as speech or video). Applications of CNN and RNN to rail maintenance operations are commonly available, but CNN has been more widely adopted.

There are also some neural network architectures based on CNN with a novel configuration and supporting specific functions and tasks which might give inspirations for the rail maintenance operations. A Siamese neural network, also called a twin neural network, is an artificial neural network that uses the same weights while working in tandem on two different input vectors to calculate similarity scores of output vectors[27]. Figure 2 shows how the CNN layers are positioned to form the architecture of the Siamese neural network.

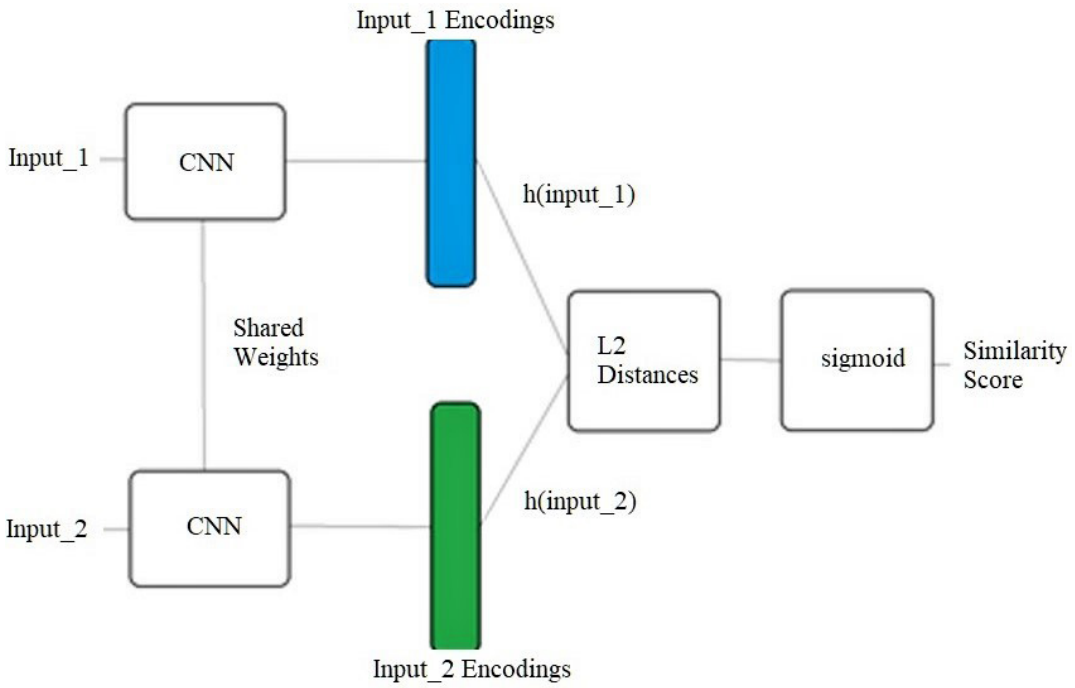


Figure 2. Siamese neural network. CNN: Convolutional neural network

U-Net is a CNN that was developed for biomedical image segmentation. It supplements the usual contracting network by successive layers to increase the output resolutions, where up-sampling operators replace pooling operations[28]. Figure 3 shows how the CNN layers are positioned to form the architecture of the U-Net.

Transfer learning and generative adversarial networks (GANs) are exciting and rapidly changing fields that have been drawing attention from researchers and practitioners in and out of the rail industry. The idea of transfer learning [Figure 4] is that a model developed for a task can be reused as the starting point for a model on another task[29]. Pre-trained models are used as the starting point as transfer learning on both computer vision and natural language processing tasks so that computing and human resources can be preserved and provide a big jump for new deep learning tasks.

Generative modeling is performed to auto-learn and discover the regularities or patterns in input data, and then the model can generate new examples that are plausibly the same as the original dataset[30]. GANs frame the problem with two sub-models: the generator model that is trained to generate new examples, and the discriminator model that tries to classify examples as either real (from the domain) or fake (generated). The two models are adversarially trained together with an objective that the discriminator model cannot distinguish between real and generated inputs. Figure 5 illustrates the main ideas of transfer learning and GANs.

There are different deep learning methods suitable for different tasks. The most important problems that humans have been interested in solving with computer vision are image classification, object detection, and segmentation in the increasing order of their difficulty. Rail track anomalies might need to be classified so that appropriate actions can be taken, thus it is an image classification task. A foreign object might need to be located from a rail track image taken, thus it is an object detection task. Sometimes both the types of

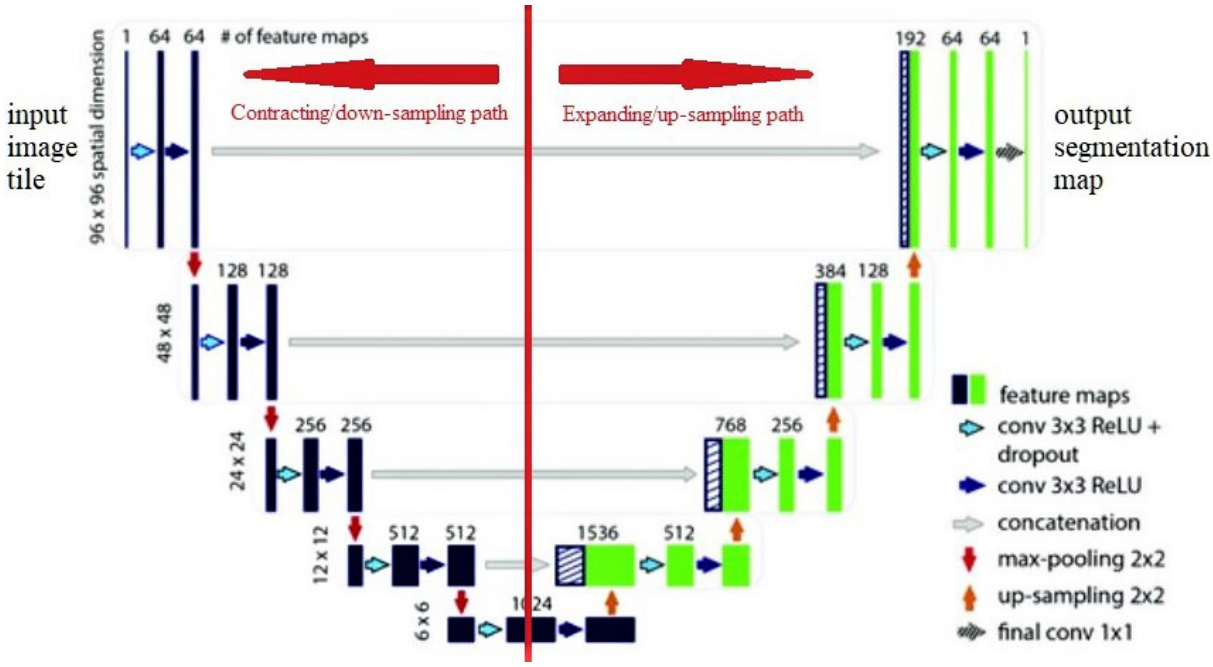


Figure 3. A sample U-Net architecture.

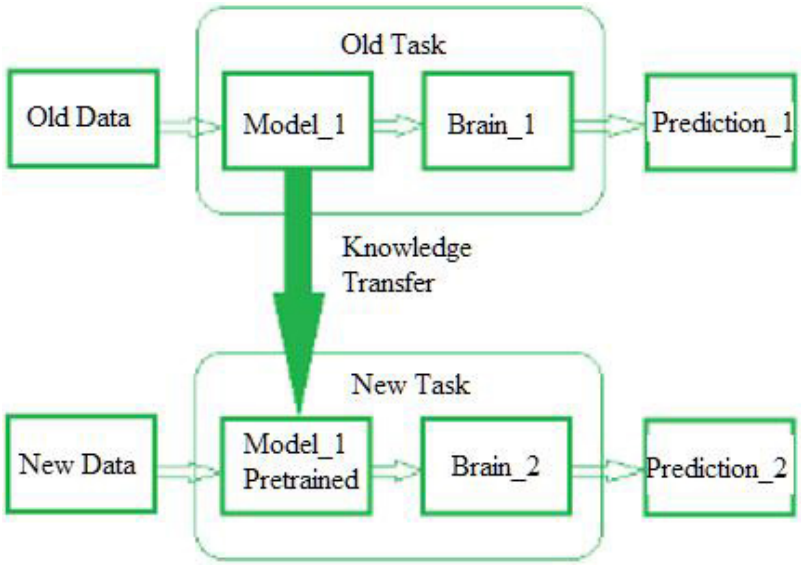


Figure 4. Transfer learning.

anomalies and the location of the anomalies need to be identified. It means classification tasks and localization tasks need to be performed concurrently, which is semantic segmentation. Classification networks are created to be invariant to translation and rotation, thus giving no importance to location information, whereas localization involves getting accurate details with respect to the location. Thus, these two tasks are inherently contradictory. Most segmentation algorithms give more importance to localization and thus lose sight of the global context.

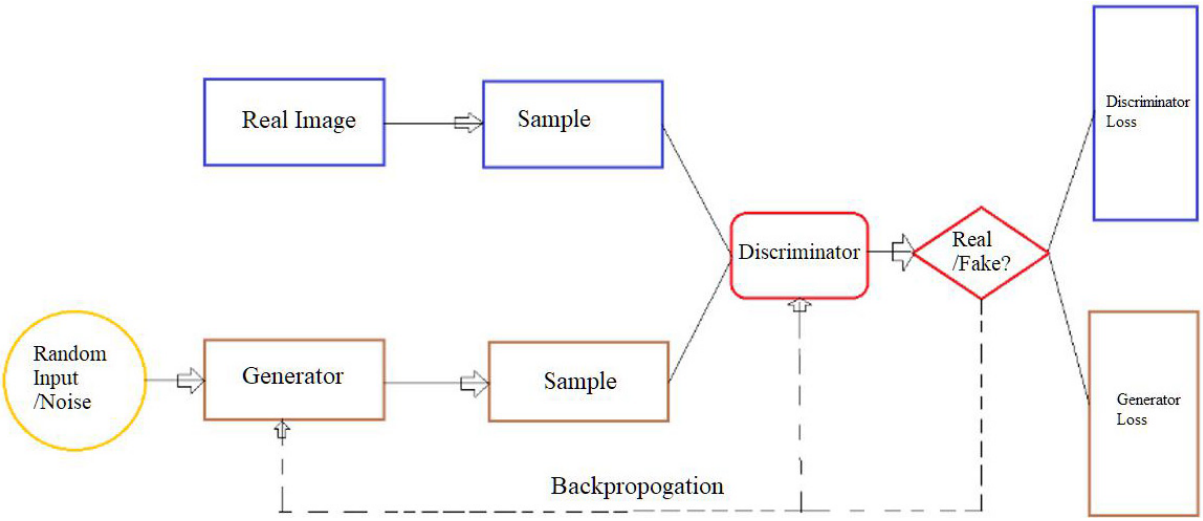


Figure 5. Generative adversarial networks.

For image classification tasks, the following deep learning methods could be adopted:

• LeNet is the earliest pre-trained model used for recognizing handwritten and machine-printed characters and has a simple and straightforward architecture.

• AlexNet consists of eight layers, five convolutional layers and three fully connected layers, features ReLU and overlapping techniques, and allows multiple GPU. The dropout technique is used to prevent overfitting problems while suffering from longer training time. The dropout technique is that, at every training step, the number of interconnecting neurons of a neural network is randomly reduced by a percentage. ZFNet is a classic CNN and was motivated by visualizing intermediate feature layers and the operation of the classifier[31]. It has smaller filters and convolution stride than AlexNet.

• Inception network differs from the CNN classifiers in that it has filters with multiple sizes operating on the same level and concatenated outputs are sent to the next inception module which makes the neural network wider[32].

• GoogLeNet is a 27-layer architecture including nine inception modules that reduce the input images while retaining important spatial information to achieve efficiency. Users can utilize a GoogLeNet network trained on Imagenet with transfer learning instead of implementing or training the network from the scratch.

• ResNet introduces, as shown in Figure 6, an identity shortcut connection that skips one or more layers. The identity mapping layers do nothing to avoid producing higher training error[33]. Pre-activation ResNet makes the optimization easier and reduces the overfitting. RiR (ResNet in ResNet) makes the input with residual stream and transient stream for better accuracy attempts in order to generalize the ResNet block for residual network[34]. Residual networks of residual networks (RoR) proposes to have shortcut connections across a group of residual blocks[35]. On top of this, another level of shortcut connection can exist across a group of "groups of residual blocks". Wide residual network (WRN) reduces training time but has more parameters as the network widens and it tests plenty of parameters such as the design of the ResNet block including the depth and the widening factor[36].

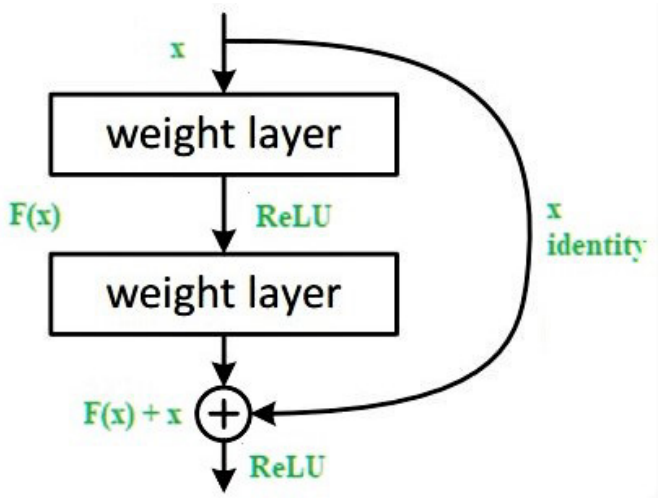


Figure 6. A residual block.

• ResNeXt is a variant to ResNet and looks similar to the inception network as they both follow the split-transform-merge paradigm, but the outputs of different paths are merged by adding them together for ResNeXt instead of depth-concatenated for inception network[37]. ResNeXt architecture's paths share the same topology. The number of independent paths is introduced as a hyper-parameter cardinality to provide a new way of adjusting the model capacity.

• DenseNet further connects all layers directly with each other for the benefit of shortcut connections[38]. All earlier layers' feature maps are aggregated with depth-concatenation and passed to subsequent layers. DenseNet is highly parameter-efficient due to feature reuse.

• Network with stochastic depth randomly drops layers during training but uses the full network during testing. The ResNet neural network takes less time in training and is thus more useful for real-world applications[39]. Each layer is randomly dropped with a survival probability during training, and all layers are active and recalibrated according to their survival probabilities during testing time. During training, the input of a residual block flows through both the identity shortcut and the weight layers when it is enabled; otherwise, only it only flows through the identity shortcut.

• VGGNet is a standard deep CNN architecture with 16 and 19 convolutional layers for VGG-16 and VGG-19[40]. The VGG architecture is a well-performing image recognition architecture on many tasks and datasets beyond ImageNet.

• SPPNet is a type of CNN that employs spatial pyramid pooling to remove the fixed-size constraint of the network[41]. An SPP layer is added on top of the last convolutional layer to pool the features and generate fixed-length outputs which are then fed into the fully connected layers or other classifiers with the aim to avoid the need for cropping or warping at the beginning.

• PReLU-Net is a kind of CNN using parameterized ReLUs for activation function and a robust Kaiming initialization scheme to account for non-linear activation functions[42].

• Xception is a 71-layer CNN with an input image size of 299 × 299[43]. The network was trained on more than a million images from the ImageNet database and learned rich feature representations for a wide range of images. Users can load a pre-trained version of the network that can classify images into 1000 object categories.

• MobileNet is a lightweight deep neural network designed for mobile applications of computer vision tasks[44]. As a filter's depth and spatial dimension can be separated, MobileNet uses depthwise separable convolutions to significantly reduce the number of parameters A depthwise separable convolution is made from depthwise convolution, the channel-wise DK × DK spatial convolution, and pointwise, the 1 × 1 convolution to change the dimension.

• FractalNet is a type of CNN that uses a fractal design instead of residual connections[45]. A simple expansion rule is repeatedly applied to generate deep networks. These networks have structures of truncated fractals and contain interacting subpaths of different lengths. There are no pass-through or residual connections, and every internal signal is transformed before flowing to subsequent layers.

• Both Trimps-Soushen and PolyNet performed very well in the ILSVRC image classification competition. Trimps-Soushen uses the pre-trained models from Inception-v3, Inception-v4, Inception-ResNet-v2, Pre-Activation ResNet-200, and Wide ResNet (WRN-68-2) for classification. PolyNet introduced a building block called PolyInception module formed by adding a polynomial second-order term to increase the accuracy. Then, a very deep PolyNet is composed based on the PolyInception module.

For object detection tasks, the following deep learning methods can be deployed:

• OverFeat is a classic type of CNN architecture, employing convolution, pooling, and fully connected layers[46].

• R-CNN extracts only 2000 regions from the image as region proposals to work with using the selective search algorithm[47]. The CNN extracts the features from the image. The extracted features at the output dense layer are fed into an SVM to classify the presence of the object within that candidate region proposal. For Fast R-CNN, the region proposals are identified from the convolutional feature map generated by the CNN with the input image[48]. The region proposals are then warped into squares and reshaped into a fixed size using a RoI pooling layer before being fed into a fully connected layer. From the RoI feature vector, the class of the proposed region and the offset values for the bounding box are predicted with a softmax layer. Fast R-CNN is faster than R-CNN because the convolution operation is performed only once per image to generate a feature map. Faster R-CNN is similar to Fast R-CNN but much faster. It uses a separate network to predict the region proposals instead of using a selective search algorithm to identify the region proposals on the feature map generated by CNN[49]. An RoI pooling layer then reshapes the predicted region proposals for classifying the image within the proposed region and predicting the offset values for the bounding boxes. Real-time object detection tasks can adopt faster R-CNN.

• DeepID-Net introduces a deformable part-based CNN[50]. A new deformable constrained pooling layer models the deformation of the object parts with geometric constraint and penalty. Besides directly detecting the entire object, it is also crucial to detect object parts which can then support detecting the entire object.

• R-FCN is similar to the logic of R-CNN-based detectors but reduces the amount of work needed for each region proposal to increase the speed[51]. The region-based feature maps can be computed outside each region proposal and are independent of region proposals.

• You only look once (YOLO) algorithm detects objects in real-time using a neural network[52]. Its architecture passes the nxn image once through the fully convolutional neural network and outputs mxm prediction. The YOLO architecture splits the input image into an mxm grid and generates two bounding boxes and associated class probabilities of the boxes for each grid. The bounding boxes could be larger than the grid itself. Differential weights of confidence predictions are adopted for boxes with or without objects during training. The square root of width and height are predicted differently for bounding boxes containing small or large objects. These changes to the loss function enable YOLO to produce better results. YOLOv3 (You Only Look Once, Version 3) was commonly adopted by the studies reviewed here.

• Single Shot multibox Detector (SSD) builds on the VGG-16 architecture while discarding its fully connected layers [Figure 7][53]. The original VGG fully connected layers are replaced with a set of auxiliary convolutional layers (from conv6 onwards) to extract features at multiple scales and progressively decrease the size of the input to each subsequent layer.

For semantic segmentation tasks, the following deep learning methods can be adopted:

• FCN uses a CNN to transform image pixels to pixel classes[54]. Instead of image classification or object detection, FCN transforms the height and width of intermediate feature maps back to those of the input image by using the transposed convolutional layer. Thus, the classification output and the input image have a one-to-one correspondence at the pixel level. Therefore, the classification results for the input pixel are held by the channel dimension at its output pixel at the same spatial position.

• DeconvNet gradually deconvolutes and un-pools to obtain its output label map, different from the conventional FCN with possible rough segmentation output label map[55].

• DeepLab applies atrous convolution for up-sampling[56]. Atrous convolution is a shorthand for convolution with up-sampled filters. Filter up-sampling amounts to inserting holes between nonzero filter taps. Atrous convolution allows effectively enlarging the field of view of filters without increasing the number of parameters or the amount of computation. Up-sampling the output of the last convolution layer and computing pixel-wise loss produce the dense prediction.

• ParseNet aggregates the values of each channel feature map's activations to declare contextual information[57]. These aggregations are then merged to be appended to the final features of the network. This approach is less tiresome than the proposal cum classification approach and avoids unrelated predictions for different pixels under FCN approach.

• DilatedNet uses dilated convolutions, filters with holes, to avoid losing resolution altogether[58]. In this way, the receptive field can grow exponentially while the number of parameters only grows linearly. The front end is based on VGG-16 by replacing the last two pooling layers with dilated convolutions. A context module and a plug-and-play structure are introduced for multi-scale reasoning using a stack of dilated convolutions on a feature map.

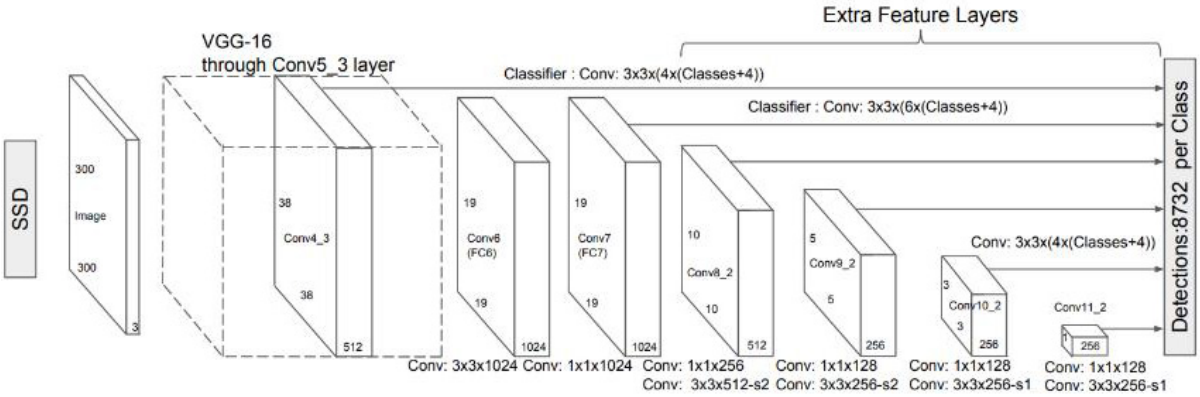


Figure 7. Single Shot multibox Detector architecture.

• PSPNet utilizes a pyramid parsing module to exploit global context information by aggregating different region-based contexts[59]. A pre-trained CNN with the dilated network strategy is used to extract the feature map, on top of which the pyramid pooling module gathers context information. The final feature map size is one-eighth of the input image. Using a four-level pyramid, the pooling kernels cover the whole, half of, and small portions of the image. They are fused as the global prior, which is then concatenated with the original feature map in the final part. It is followed by a convolution layer to generate the final prediction map. The local and global clues together make the final prediction more reliable.

3. REVIEW OF RAIL TRACK CONDITION MONITORING WITH DEEP LEARNING

The authors systematically searched published peer-reviewed journal articles and papers found in Google Scholar. Combinations of keywords such as "rail", "surface", "rail track", "defect", and "deep learning" were used as search keys to find research works published in the application of deep learning techniques to rail track condition monitoring and anomaly detection and classification. The review covers work from 2013 to 2021. In total, we identified 62 relevant research publications to review.

The trend over time: a clear increasing trend can be observed of the popularity of deep learning approaches in rail track condition monitoring applications. Table 1 summarizes the findings. The number of papers surged in 2018. Before 2018, machine learning techniques other than deep learning approaches were more widely adopted. The rail industries are adopting deep learning methods with growing interests. An upwards trend of publication number is observed. There is also a gap of a few years from the invention of a deep learning model to its adoption by the rail industry.

Regions of study: fourteen regions are represented by the papers identified. Among them, China has the highest number of papers, which indicates the popularity of rail-related research work corresponding to the expanding rail networks across the country. Papers from China surged in 2018 and kept a high number in the following years. Table 2 summarizes the distribution of papers over regions.

Raw data type: it is observed that 70% of studies used image-type raw data for the deep learning models. Nevertheless, acoustic emission signals[65,71,100,103,108], defectogram[96,109], speed accelerations[98], concatenated vector of curve and numbers[101], current signal[89], maintenance records[80,99], synthetic data from generative model[63], time-frequency measurement data[82], time-series[60], geometry data[87], and vibration signal[119] could all be possible input data sources as well.

Table 1. Number of papers over time

Year	Total number of papers	Ref.
2014	1	[60]
2015	1	[61]
2016	3	[62-64]
2017	5	[65-69]
2018	13	[70-82]
2019	16	[83-98]
2020	13	[99-111]
2021	10	[112-121]

Table 2. Paper distribution over regions

Region	Total number of papers	Ref.
Australia	1	[70]
Canada	1	[83]
China	35	[65,71-78,84-92,99-107,112-119]
Germany	1	[93]
Hong Kong	2	[94,108]
India	2	[66,67]
Korea	1	[95]
Netherlands	4	[62,63,78,80]
Russia	2	[96,109]
Singapore	3	[81,97,110]
Swiss	1	[60]
Turkey	3	[68,69,111]
UK	1	[120]
USA	5	[61,64,82,98,121]

Purpose of study: it is observed that detection, classification, and/or localizing rail surface defects including various components (rail, insulator, valves, fasteners, switches, track intrusions, *etc.*) are the most common purpose of the studies. There are also papers on predicting maintenance time[99] and detecting track geometry elevations[98]. Detection and classification tasks are more common than prediction tasks[60,80,87,99,119].

Adoption of deep learning models: many deep learning models are adopted by researchers. Table 3 summarizes the distribution of deep learning models. CNN is the most popular deep learning model being adopted; however, many researchers created their own structure or divided their tasks into a few stages. CNN has been popular for extracting features and RNN/LSTM has been used for the sequential data type.

From the summary in Table 3, there are various deep learning methods being adopted in different forms. The effectiveness and the results differ from each other depending on the tasks. It is observed that image is the most popular input data type used for deep learning applications. However, there is a consistent process flow for how to apply the deep learning methods to rail track condition monitoring. First, the image acquisition subsystem (cameras/recording devices) is usually installed on rail engineering maintenance vehicles to capture raw input data. Second, the raw input data are transferred to the image processing subsystem where optional data pre-processing could be performed. Images could be resized, enhanced, have noise removed, or cropped for target areas with image processing techniques. Third, the input data are prepared for the training and testing of deep learning models. Data are labeled accordingly and then

Table 3. Adoption of deep learning models

Deep learning model	Ref.
AlexNet, ResNet	[76]
Autoencoder	[89,101]
CNN	[61,62,65,68,69,74,75,79,85,94,95,102,108,109]
CNN (object detection), RNN (distance estimation)	[93]
CNN + YOLOv3	[117]
CNN based self-proposed DFF-Net	[118]
CNN based self-proposed FR-Net	[72]
CNN to extract feature, only 1 class	[82]
CNN, transfer learning	[64]
CNN, transfer learning, Bayesian optimization to tune hyperparameters	[100]
CNN-LSTM	[87,103,107]
Faster R-CNN	[78,88]
Faster R-CNN + CNN	[73]
FastNet, convolutional network-based	[120]
Fine-grained bilinear CNNs model	[70]
FCN	[119]
GAN for CNN	[115]
Inception-ResNet-v2 & CNN	[113]
LSTM-RNN	[63,71,99]
Mask R-CNN	[121]
ML Tree based methods	[80]
MobileNetV2, YOLOv3	[84]
Multilayer feedforward neural networks based on multi-valued neurons (MLMVN)	[60]
neural network	[96]
Point Cloud deep learning	[92]
ResNet classifier, DenseNet classifier	[81]
ResNet, FCN	[83]
Resnet50, transfer learning, Inception, Faster R-CNN	[67]
Self-proposed, 2 stage FaultyNet, CNN based	[97]
Self-proposed, segment U-Net (CNN based) then detect, progressive	[116]
Self-proposed, ShuffleNet-v2 extracts features from the track image, RPN predicts	[86]
Siamese convolutional neural network	[66,91]
Single Shot multibox Detector (SSD)	[90]
SqueezeNet, MobileNetV2	[111]
U-Net to segment	[105]
Variational autoencoder	[98]
YOLO V3	[77,104,106,110,114]
YOLOv5 detect object; mast R-CNN detect surface defect; ResNet classify fastener state	[112]

CNN: Convolutional neural network; RNN: recurrent neural network; YOLOv3: You Only Look Once, Version 3; LSTM: long-short-term memory; DFF-net: differential feature fusion convolution neural network; FR-net: feature fusion refine neural network; FCN: fully convolutional networks; RPN: region proposal network.

randomly assigned for training and testing purposes. Fourth, the selected deep learning model is trained with the training data and validated by the testing data. Depending on the purposes, the deep learning model could perform classification or localization tasks. It is also possible to perform classification and localization concurrently, which is the most common type of task. The feature representations of the input data are always extracted; however, the next steps to deal with the feature vectors differ. It is noted that researchers also proposed their own neural network architectures to replace or complement the existing

architectures[72,86,97,116,118]. Fifth, the trained deep learning model is put in production with the trained parameters for real-world applications. Due to the criticality of rail track condition monitoring, redundancy of inspections by human operators could be provided to double confirm the accuracy. Finally, the efficiency and effectiveness of the deep learning models are reviewed and enhanced for improved performances.

Various deep learning methods are reported to produce promising results. With more data available from the rail industry, breakthroughs of deep learning methods, and more advanced and cheaper hardware, deep learning methods will only become more popular and useful for rail track condition monitoring. Deep learning models performed well for feature extraction and data classification tasks. The image processing requirements and the man-made feature extraction efforts are low for deep learning methods, which make the application economical. However, the nature of rail operations causes the distribution of rail track image data to be uneven and extremely disproportional, which could cause class imbalance problems in deep learning applications. The extremely high safety requirement of rail operations and the considerably black-box nature of deep learning models contradict each other and might cause some trust issues, which is demonstrated by the redundancies applied to rail track inspections.

Data pre-processing: removing outliers, normalizing data, and applying image process techniques to enhance the images are common pre-processing techniques. Fourier transforms such as Fast Fourier Transforms and Short Time Fourier Transform have been applied to transform sequential data (e.g., acoustic emission) to 2D spectrograms, which can then be applied to CNN models[65,82,108].

4. DISCUSSIONS

Condition monitoring and anomaly detection and classification are important to a productive rail maintenance operation. There are four main types of maintenance strategies: corrective maintenance, preventive maintenance, proactive maintenance, and predictive maintenance. Deep learning methods can support the maintenance strategies depending on the tasks it performs. For example, detection of a certain type of defect will normally trigger corrective maintenance actions to rectify the rail track. The prediction tasks supported by deep learning methods can be used for predictive maintenance strategies, and they are becoming more and more popular.

Internet of Things (IoT) technologies can be implemented to support rail maintenance operations. Sensor, networking, and application layers in IoT can collect big data, which can then be analyzed by deep learning techniques for application services. In rail track condition monitoring, various types of sensors are being deployed across the rail network to collect data that need enablers such as deep learning techniques to unleash their full potential.

Deep learning methods can be more effective when they can be used in real time in the field by the technicians. Integration of deep learning, IoT, mobile technologies, and edge computing has the potential to develop useful applications that support the daily rail track maintenance operations.

Deep learning models are normally trained with high computing powers. In order for the deep learning models to be used in the field, light deep learning models need to be developed so that less powerful but more accessible devices such as mobile phones can be used with deep learning techniques to support the rail maintenance operations.

Image data have been widely used with useful results. However, they tend to be used in detection and classification, which normally correspond to corrective or preventive maintenance strategies. In order for the industry to advance to predictive maintenance, the causes of defects need to be investigated and corresponding signals can be used for deep learning models' training and testing. Thus, deep learning application in rail track condition monitoring need to cater to more varieties of data types; for example, the vibration signal is normally a sequential signal and vibration tends to cause wear on the rails over time.

The performance of deep learning models such as accuracy, response time, and precision tend to be influenced by the data used. Therefore, a common dataset of rail tracks will be helpful for researchers to use and validate the performance of deep learning models developed around the world. True performance comparisons could be made as well.

There are always more sections of normal rail tracks than defective sections; therefore, the number of normal track images tend to be much more than the defective ones. When considering the different classes of defects, the number of defect images will be even lower, which can cause class imbalance issues. GAN is a deep learning-based generative model. With the application of GAN to generate more images of defects, the training of deep learning models for detecting and classifying rail defects could be more efficient and accurate.

Transfer learning focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. As rail track condition monitoring task is shared by researchers around the world, a "more" related problem and the knowledge gained will be more useful when it is transferred to another problem but in the same industry. This could be a promising research area for future work.

Reinforcement learning is the training of machine learning models to make a sequence of decisions. The agent learns to achieve a goal in an uncertain, potentially complex environment. This trial-and-error approach suits well the complex situation of deciding when it is the right time to perform the maintenance operations. This research area could generate meaningful results for the industry.

5. ILLUSTRATIVE CASE STUDIES

In this section, we use two application examples to illustrate the implementation of deep learning models to support rail track condition monitoring and rail defect detection and classification.

5.1. Data acquisition and preparations

Data acquisition equipment or devices could be installed on the rail inspection vehicles or passenger trains at different positions. We use rail vision systems here to record videos of the rail tracks for both head and rail checks. Lights are usually required to further enhance the quality of images taken; we use halogen floodlights to support visibility. The train speed might affect the image quality; therefore, a maximum speed may be set. For us, it is 100 km/h. Figure 8 illustrates a typical setup.

Various types of input data (image, sound, vibration, *etc.*) could be adopted for deep learning tasks. For image data, different formats and sizes might be used. Original greyscale images are captured from the rail track videos and then used without segmenting the tracks from the ballast, the sleepers, or other background textures around the rail tracks so as to minimize the image pre-processing efforts and maximize the utility of deep learning models. Pre-processing of input images is optional. Figure 9 shows a sample image that was used for training and test purpose.

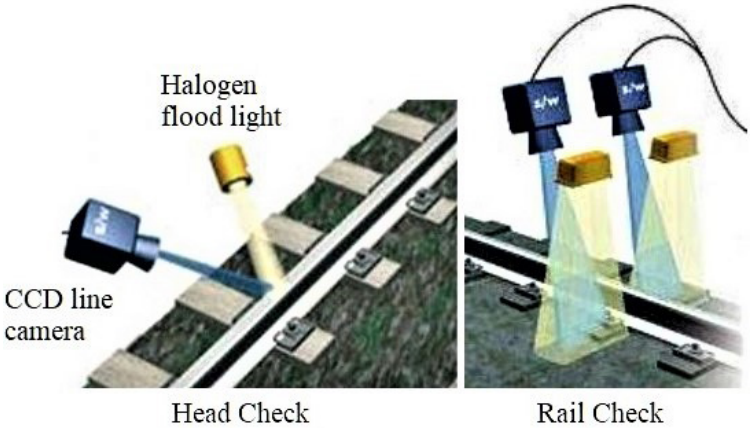


Figure 8. Image acquisition setup.

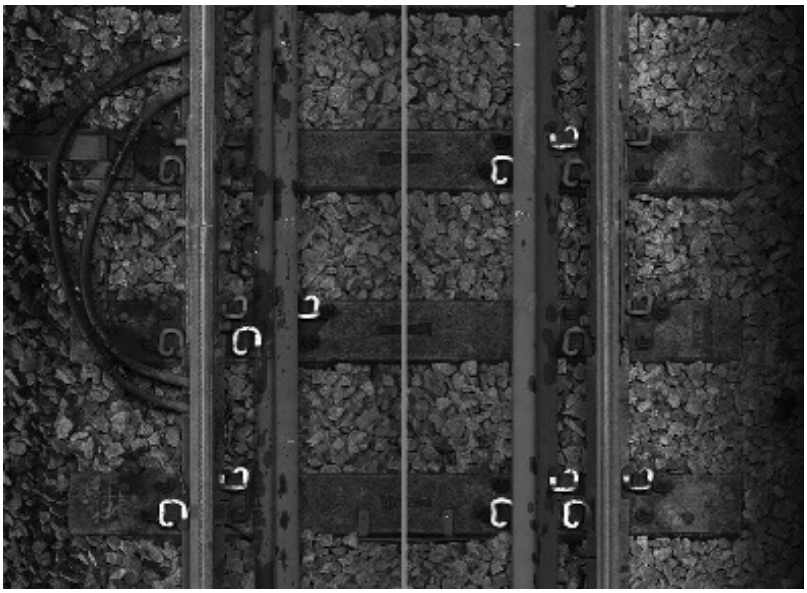


Figure 9. A sample image for training.

We worked with rail maintenance operations staff to describe and label the images according to their properties. Figure 10 shows some sample patterns of "normal", "corrugation", "insulated rail joint", and "weld" images.

5.2. Deep learning environment configurations

The deep learning experiments were performed on a platform with OS (Windows or Linux) and GPU. Our configurations include Intel Core i7-8700 CPU, Nvidia GeForce RTX 2080 Ti GPU, and 64 GB RAM. Software packages such as Python 3.6, Nvidia CUDA Toolkit, cuDNN, and TensorFlow with GPU support were installed on Windows 10 operating system for our experiments.

5.3. Application 1: CNN

We conducted training and prediction experiments for both anomaly detection and classification purpose. The input images were categorized by the neural network into two output classes for detection tasks and ten output classes for classification tasks. For both tasks, 90% of the image samples were randomly reserved for

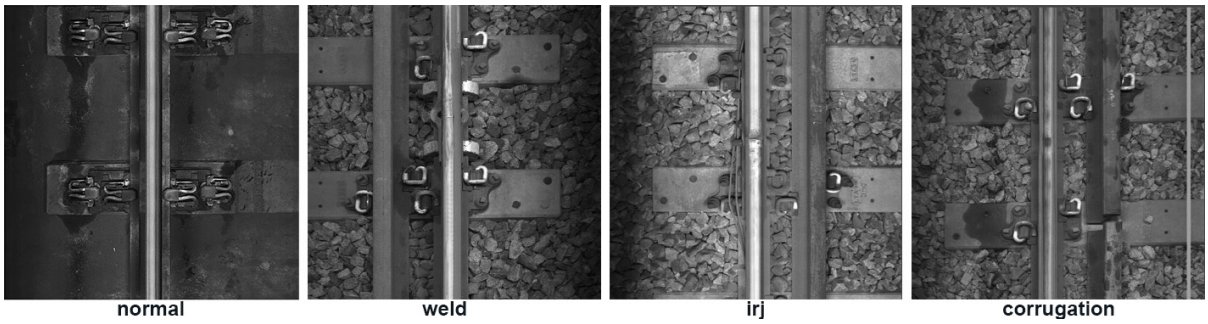


Figure 10. Sample patterns of different classes.

training while the remaining 10% for testing for every class. In anomaly detection, the distribution of images is comprised of 57% normal images and 43% images with various types of defects. In anomaly classification, there are ten classes in total that are to be classified; their distributions are: 57%, 7%, 2%, 4%, 1%, 1%, 7%, 2%, 17%, and 2%. Input greyscale images were resized to 186 × 256 pixels before being fed into the neural network for training and prediction.

Four convolutional layers, two max-pooling layers, and four fully connected layers were connected for the deep convolutional neural network, and the convolutional kernel size was set as 3 × 3 pixels. We used max-pooling units of size 2 × 2 pixels. ReLU was used as an activation function for convolutional layers. We added a dropout layer as an effective regularization method after max-pooling to reduce overfitting by randomly dropping out nodes during training. After the convolutional and max-pooling layers, we used fully connected layers to perform high-level reasoning in a convolution neural network.

We ran the deep convolutional neural network model for detection and classification tasks separately while keeping the training and testing images the same and only adjusting the number of output classes for the network classifiers at the end of the network. The learning rate of the Adam optimizer was set as 0.001 for training the model. For both the binary classification of detection tasks and the multi-class classification of classification tasks, we counted the number of true positives, true negatives, false positives, and false negatives. The binary classification accuracy was calculated as 87.45% and F1-score as 88.33%. The performance is acceptable and substantially improves the performance of the existing auto-detect method based on image process techniques and man-made feature representations in operation.

5.4. Application 2: Siamese neural network

We conducted training and prediction experiments for classification tasks with four classes of data which comprise normal images and three common types of defects. We used an equal number of images for each class. We created the training data samples for Siamese neural network, which is much easier than the classic convolutional neural network datasets that require images to be labeled. Image samples were randomly chosen from this dataset to form anchor-positive-negative trios. While sampling an image pair, the two images were chosen from the same category with a probability of 0.5 with a corresponding label of $y = 0$. Similarly, the images were chosen from two different categories with the remaining probability of 0.5 with the label $y = 1$.

Two identical four-layer convolutional neural networks were used to form the twin structure of the Siamese neural network to perform the identification of rail surface defects. The batch size was 128. The number of epochs was 50. The number of steps per epoch was 5. ReLU was used as an activation function. The neural network optimizer used was Adam. During testing, data of matching pairs and non-matching pairs were

formed by randomly selecting images from each of the categories. All combinations, for example, "Normal *vs.* Defect_1" or "Defect_2 *vs.* Defect_3", were tested with 10 pairs and the mean, minimum, and maximum values of test output distance were summarized for analysis. During the testing phase, the images were chosen randomly. These images belonged to a different set of classes that were never shown to the network during training. As described, all combinations of pairwise comparisons were tested with 10 different sample image pairs. The mean Euclidean (L2) distance was computed as the similarity score.

Our experiment results show that, when two test images belong to the same class, their dissimilarity score is smaller than those of images from two different classes. A threshold value can be chosen to determine whether two images are from the same or different classes based on the test similarity scores. We also notice that the threshold value has a decent margin to vary. From our experiments, 82.5% of test images actually from the same class were predicted to be "from the same class", while 80.8% of test images actually from different classes were predicted to be "from different classes". The binary classification accuracy was calculated as 81.67% and F1-score as 81.82%. The accuracy level is acceptable to the current rail maintenance operations with the potentials for further improvement.

Both cases can perform rail track condition monitoring and anomaly detection and classifications tasks with deep learning methods with the same dataset. The case studies deal with them in different ways with different deep learning models. Datasets generated from the maintenance operations are put into good use with the deep learning models to improve the rail track maintenance operations. Training of Siamese convolutional neural network was observed to take a shorter time than the classic convolutional neural network approach. The existing hardware setup in the rail operations did not require significant modification. The features extracted by the deep learning models performed better than the approaches of selected man-made features.

6. CONCLUSIONS

This paper presents the importance and criticality of rail track condition monitoring to safe rail operations. We give a brief overview of the historical development of deep learning and list common deep learning models. Deep learning came into the rapid development phase after 2012; therefore, we review the deep learning applications to rail track condition monitoring from 2013 to 2021. The applications are reviewed according to the temporal evolutions, the regional adoptions, the data type, and the deep learning models. We then discuss the potential challenges and research opportunities for applying deep learning to rail track condition monitoring. Two application case studies are shared to illustrate the implementation of deep learning methods in rail track condition monitoring.

DECLARATIONS

Acknowledgment

This research is part of the project supported by SMRT Corporation Ltd. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect those of the company.

Authors' contributions

Investigated the research area, reviewed and summarized the literature, wrote and edited the original draft: Ji A

Managed the research activity planning and execution, contributed to the development of ideas according to the research aims: Woo WL, Wong EWL

Performed critical review, commentary and revision, and provided administrative, technical, and material support: Quek YT

Availability of data and materials
Not applicable.

Financial support and sponsorship
None.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright
© The Author(s) 2021.

REFERENCES

1. Cannon DF, Edel K, Grassie SL, Sawley K. Rail defects: an overview. *Fatigue Fract Eng M* 2003;26:865-86. DOI
2. Track circuit monitoring tool: standardization and deployment at CTA. Available from: http://www.trb.org/Main/Blurbs/177054.aspx [Last accessed on 5 Jan 2022].
3. Rail Defects Handbook. Available from: https://extranet.artc.com.au/docs/eng/track-civil/guidelines/rail/RC2400.pdf [Last accessed on 5 Jan 2022].
4. Dey A, Kurz J, Tenczynski L. Detection and evaluation of rail defects with non-destructive testing methods. Available from: https://www.ndt.net/article/wcndt2016/papers/we1g4.pdf [Last accessed on 5 Jan 2022].
5. Min Y, Xiao B, Dang J, Yue B, Cheng T. Real time detection system for rail surface defects based on machine vision. *J Image Video Proc* 2018. DOI
6. Serin G, Sener B, Ozbayoglu AM, Unver HO. Review of tool condition monitoring in machining and opportunities for deep learning. *Int J Adv Manuf Technol* 2020;109:953-74. DOI
7. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX. Deep learning and its applications to machine health monitoring. *Mech Syst Signal Process* 2019;115:213-37. DOI
8. Fu J, Chu J, Guo P, Chen Z. Condition monitoring of wind turbine gearbox bearing based on deep learning model. *IEEE Access* 2019;7:57078-87. DOI
9. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44. DOI PubMed
10. Mcculloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biol* 1943;5:115-33. PubMed
11. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 1958;65:386-408. DOI PubMed
12. Newell A. A step toward the understanding of information processes. *Science* 1969;165:780-2. DOI
13. Rodan A, Faris H, Alqatawna J. Optimizing feedforward neural networks using biogeography based optimization for E-mail spam identification. *IJCNS* 2016;9:19-28. DOI
14. Robert HN. Theory of the backpropagation neural network. *Proc 1989 IEEE IJCNN* 1989;1:593-605.
15. Lecun Y, Boser B, Denker JS, et al. Backpropagation applied to handwritten zip code recognition. *Neural Comput* 1989;1:541-51. DOI
16. Hochreiter S. Untersuchungen zu dynamischen neuronalen Netzen. *Diploma: Technische Universität München* 1991. DOI
17. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735-80. DOI PubMed
18. Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1:81-106. DOI
19. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20:273-97. DOI
20. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 1997;55:119-39. DOI
21. Cristianini N, Scholkopf B. Support vector machines and kernel methods: the new generation of learning machines. *Ai Magazine* 2002;23:31. DOI
22. Breiman L. Random forests. *Mach Learn* 2001;45:5-32. DOI

23. Murphy K. An introduction to graphical models. *Rap tech* 2001;96:1-19.
24. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput* 2006;18:1527-54. DOI PubMed
25. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS); Fort Lauderdale, FL, USA. 2011. p. 315-23.
26. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60:84-90. DOI
27. Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. Proceedings of the 32nd International Conference on Machine Learning; Lille, France. 2015.
28. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical image computing and computer-assisted intervention - MICCAI 2015. Cham: Springer International Publishing; 2015. p. 234-41. DOI
29. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In: Kůrková V, Manolopoulos Y, Hammer B, Iliadis L, Maglogiannis I, editors. Artificial neural networks and machine learning - ICANN 2018. Cham: Springer; 2008. p. 270-9.
30. Goodfellow I, Pouget-abadie J, Mirza M, et al. Generative adversarial networks. *Commun ACM* 2020;63:139-44. DOI
31. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Computer Vision - ECCV 2014. Cham: Springer International Publishing; 2014. p. 818-33. DOI
32. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015 Jun 7-12; Boston, MA. IEEE; 2005. p. 1-9. DOI
33. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Las Vegas, NV, USA. IEEE; 2016. p. 770-8. DOI
34. Targ S, Almeida D, Lyman K. Resnet in resnet: generalizing residual architectures. arXiv preprint arXiv:1603.08029.
35. Zhang K, Sun M, Han TX, Yuan X, Guo L, Liu T. Residual networks of residual networks: multilevel residual networks. *IEEE Trans Circuits Syst Video Technol* 2018;28:1303-14. DOI
36. Zagoruyko S, Komodakis N. Wide residual networks. arXiv preprint arXiv:1605.07146.
37. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21-26; Honolulu, HI, USA. IEEE; 2017. p. 5987-95. DOI
38. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21-26; Honolulu, HI, USA. IEEE; 2017. p. 2261-9. DOI
39. Huang G, Sun Y, Liu Z, Sedra D, Weinberger KQ. Deep networks with stochastic depth. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer vision - ECCV 2016. Cham: Springer International Publishing; 2016. p. 646-61. DOI
40. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
41. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell* 2015;37:1904-16. DOI PubMed
42. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV); 2015 Dec 7-13; Santiago, Chile. IEEE; 2015. p. 1026-34. DOI
43. Chollet F. Xception: deep learning with depthwise separable convolutions. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21-26; Honolulu, HI, USA. IEEE; 2017. p. 1800-7. DOI
44. Howard AG, Zhu M, Chen B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
45. Larsson G, Maire M, Shakhnarovich G. Fractalnet: ultra-deep neural networks without residuals. arXiv preprint arXiv:1605.07648.
46. Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. Overfeat: integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229.
47. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition; 2014 Jun 23-28; Columbus, OH, USA. IEEE; 2014. p. 580-7. DOI
48. Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV); 2015 Dec 7-13; Santiago, Chile. IEEE; 2015. p. 1440-8. DOI
49. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2017;39:1137-49. DOI PubMed
50. Ouyang W, Zeng X, Wang X, et al. DeepID-Net: deformable deep convolutional neural networks for object detection. *IEEE Trans Pattern Anal Mach Intell* 2017;39:1320-34. DOI PubMed
51. Dai J, Li Y, He K, Sun J. R-fcn: object detection via region-based fully convolutional networks. Available from: https://arxiv.org/pdf/1605.06409.pdf [Last accessed on 5 Jan 2022].
52. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27-30; Las Vegas, NV, USA. IEEE; 2016. p. 779-88. DOI
53. Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer vision - ECCV 2016. Cham: Springer International Publishing; 2016. p. 21-37. DOI
54. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell*

2017;39:640-51. DOI PubMed
55. Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation. Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV); 2015 Dec 7-13; Santiago, Chile. IEEE; 2015. p. 1520-8. DOI
56. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell* 2018;40:834-48. DOI PubMed
57. Liu W, Rabinovich A, Berg AC. Parsenet: looking wider to see better. arXiv preprint arXiv:1506.04579.
58. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.
59. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21-26; Honolulu, HI, USA. IEEE; 2017. p. 6230-9. DOI
60. Fink O, Zio E, Weidmann U. Predicting component reliability and level of degradation with complex-valued neural networks. *Reliability Engineering & System Safety* 2014;121:198-206. DOI
61. Giben X, Patel VM, Chellappa R. Material classification and semantic segmentation of railway track images with deep convolutional neural networks. Proceedings of 2015 IEEE International Conference on Image Processing (ICIP); 2015 Sep 27-30; Quebec City, QC, Canada. IEEE; 2015. p. 621-5. DOI
62. Faghih-Roohi S, Hajizadeh S, Núnez A, Babuska R, De Schutter B. Deep convolutional neural networks for detection of rail surface defects. Proceedings of 2016 International joint conference on neural networks (IJCNN); 2016 Jul 24-29; Vancouver, BC, Canada. IEEE; 2016. p. 2584-9. DOI
63. Bruin T, Verbert K, Babuska R. Railway track circuit fault diagnosis using recurrent neural networks. *IEEE Trans Neural Netw Learn Syst* 2017;28:523-33. DOI PubMed
64. Gibert X, Patel VM, Chellappa R. Deep multitask learning for railway track inspection. *IEEE Trans Intell Transport Syst* 2017;18:153-64. DOI
65. Zhang X, Wang K, Wang Y, Shen Y, Hu H. An improved method of rail health monitoring based on CNN and multiple acoustic emission events. Proceedings of 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC); 2017 May 22-25; Turin, Italy. IEEE; 2017. p. 1-6. DOI
66. Rao DJ, Mittal S, Ritika S. Siamese neural networks for one-shot detection of railway track switches. arXiv preprint arXiv:1712.08036.
67. Mittal S, Rao D. Vision based railway track monitoring using deep learning. arXiv preprint arXiv:1711.06423. DOI
68. Santur Y, Karaköse M, Akin E. A new rail inspection method based on deep learning using laser cameras. Proceedings of 2017 International Artificial Intelligence and Data Processing Symposium (IDAP); 2017 Sep 16-17; Malatya, Turkey. IEEE; 2017. p. 1-6. DOI
69. Santur Y, Karaköse M, Akin E. An adaptive fault diagnosis approach using pipeline implementation for railway inspection. *Turk J Elec Eng & Comp Sci* 2018;26:987-98. DOI
70. Huang H, Xu J, Zhang J, Wu Q, Kirsch C. Railway infrastructure defects recognition using fine-grained deep convolutional neural networks. Proceedings of 2018 Digital Image Computing: Techniques and Applications (DICTA); 2018 Dec 10-13; Canberra, ACT, Australia. IEEE; 2018. p. 1-8. DOI
71. Zhang X, Zou Z, Wang K, et al. A new rail crack detection method using LSTM network for actual application based on AE technology. *Appl Acoust* 2018;142:78-86. DOI
72. Ye T, Wang B, Song P, Li J. Automatic railway traffic object detection system using feature fusion refine neural network under shunting mode. *Sensors (Basel)* 2018;18:1916. DOI
73. Kang G, Gao S, Yu L, Zhang D. Deep architecture for high-speed railway insulator surface defect detection: denoising autoencoder with multitask learning. *IEEE Trans Instrum Meas* 2019;68:2679-90. DOI
74. Liang Z, Zhang H, Liu L, He Z, Zheng K. Defect detection of rail surface with deep convolutional neural networks. Proceedings of 2018 13th World Congress on Intelligent Control and Automation (WCICA); 2018 Jul 4-8; Changsha, China. IEEE; 2018. p. 1317-22. DOI
75. Shang L, Yang Q, Wang J, Li S, Lei W. Detection of rail surface defects based on CNN image recognition and classification. Proceedings of 2018 20th International Conference on Advanced Communication Technology (ICACT); 2018 Feb 11-14; Chuncheon, Korea (South). IEEE; 2018. p. 45-51. DOI
76. Wang S, Dai P, Du X, Gu Z, Ma Y. Rail fastener automatic recognition method in complex background. Proceedings of Tenth International Conference on Digital Image Processing (ICDIP 2018); 2018 Aug 9; Shanghai, China. International Society for Optics and Photonics; 2018. p. 1080625. DOI
77. Yanan S, Hui Z, Li L, Hang Z. Rail surface defect detection method based on yolov3 deep learning networks. Proceedings of 2018 Chinese Automation Congress (CAC); 2018 Nov 30-Dec 2; Xi'an, China. IEEE; 2018. p. 1563-8.2. DOI
78. Xu X, Lei Y, Yang F. Railway subgrade defect automatic recognition method based on improved faster R-CNN. *Sci Programming* 2018;2018:1-12. DOI
79. Jamshidi A, Hajizadeh S, Su Z, et al. A decision support approach for condition-based maintenance of rails based on big data analysis. *Transp Res Part C Emerg Technol* 2018;95:185-206. DOI
80. Bukhsh Z, Saeed A, Stipanovic I, Doree AG. Predictive maintenance using tree-based classification techniques: a case of railway switches. *Transp Res Part C Emerg Technol* 2019;101:35-54. DOI
81. James A, Jie W, Xulei Y, et al. Tracknet - a deep learning based fault detection for railway track inspection. Proceedings of 2018 International Conference on Intelligent Rail Transportation (ICIRT); 2018 Dec 12-14; Singapore. IEEE; 2018. p. 1-5. DOI
82. Peng X, Jin X. Rail suspension system fault detection using deep semi-supervised feature extraction with one-class data.

PHM_CONF 2018. DOI

83. Sun Y, Liu Y, Yang C. Railway joint detection using deep convolutional neural networks. Proceedings of 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE); 2019 Aug 22-26; Vancouver, BC, Canada. IEEE; 2019. p. 235-40. DOI
84. Yuan H, Chen H, Liu S, Lin J, Luo X. A deep convolutional neural network for detection of rail surface defect. Proceedings of 2019 IEEE Vehicle Power and Propulsion Conference (VPPC); 2019 Oct 14-17; Hanoi, Vietnam. IEEE; 2019. p. 1-4. DOI
85. Wang Y, Zhu L, Yu Z, Guo B. An adaptive track segmentation algorithm for a railway intrusion detection system. *Sensors (Basel)* 2019;19:2594. DOI PubMed PMC
86. Dong B, Li Q, Wang J, Huang W, Dai P, Wang S. An end-to-end abnormal fastener detection method based on data synthesis. Proceedings of 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI); 2019 Nov 4-6; Portland, OR, USA. IEEE; 2019. p. 149-56. DOI
87. Ma S, Gao L, Liu X, Lin J. Deep learning for track quality evaluation of high-speed railway based on vehicle-body vibration prediction. *IEEE Access* 2019;7:185099-107. DOI
88. Jin X, Wang Y, Zhang H, et al. DM-RIS: deep multimodel rail inspection system with improved MRF-GMM and CNN. *IEEE Trans Instrum Meas* 2020;69:1051-65. DOI
89. Li Z, Yin Z, Tang T, Gao C. Fault diagnosis of railway point machines using the locally connected autoencoder. *Appl Sci* 2019;9:5139. DOI
90. Guo B, Shi J, Zhu L, Yu Z. High-speed railway clearance intrusion detection with improved SSD network. *Appl Sci* 2019;9:2981. DOI
91. Liu J, Huang Y, Zou Q, et al. Learning visual similarity for inspecting defective railway fasteners. *IEEE Sensors J* 2019;19:6844-57. DOI
92. Cui H, Li J, Hu Q, Mao Q. Real-time inspection system for ballast railway fasteners based on point cloud deep learning. *IEEE Access* 2020;8:61604-14. DOI
93. Haseeb M, Ristić-Durrant D, Gräser A. A deep learning based autonomous distance estimation and tracking of multiple objects for improvement in safety and security in railways. Available from: https://www.bmvc2019.org/wp-content/ODRSS2019/ODRSS2019_P_5_Haseeb.pdf [Last accessed on 5 Jan 2022].
94. Chen SX, Ni YQ, Liu JC, Yao N. Deep learning-based data anomaly detection in rail track inspection. Proceedings of 12th International Workshop on Structural Health Monitoring: Enabling Intelligent Life-Cycle Health Management for Industry Internet of Things (IIOT); 2019; Stanford, USA. DEStech Publications Inc.; 2019. p. 3235-42. DOI
95. Jang J, Shin M, Lim S, Park J, Kim J, Paik J. Intelligent image-based railway inspection system using deep learning-based object detection and weber contrast-based image comparison. *Sensors* 2019;19:4738. DOI PubMed PMC
96. Kuzmin EV, Gorbunov OE, Plotnikov PO, Tyukin VA, Bashkin VA. Application of neural networks for recognizing rail structural elements in magnetic and eddy current defectograms. *Aut Control Comp Sci* 2019;53:628-37. DOI
97. Pahwa RS, Chao J, Paul J, et al. Faultnet: faulty rail-valves detection using deep learning and computer vision. Proceedings of 2019 IEEE Intelligent Transportation Systems Conference (ITSC); 2019 Oct 27-30; Auckland, New Zealand. IEEE; 2019. p. 559-66. DOI
98. Liu J, Wei Y, Bergés M, Bielak J, Garrett Jr JH, Noh H. Detecting anomalies in longitudinal elevation of track geometry using train dynamic responses via a variational autoencoder. Proceedings of Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2019; 2019 Mar 27; Denver, CO, USA. International Society for Optics and Photonics; 2019. p. 109701B. DOI
99. Wang Q, Bu S, He Z. Achieving predictive and proactive maintenance for high-speed railway power equipment with LSTM-RNN. *IEEE Trans Ind Inf* 2020;16:6509-17. DOI
100. Li D, Wang Y, Yan W, Ren W. Acoustic emission wave classification for rail crack monitoring based on synchrosqueezed wavelet transform and multi-branch convolutional neural network. *Struct Health Monit* 2021;20:1563-82. DOI
101. Guo Z, Wan Y, Ye H. An unsupervised fault-detection method for railway turnouts. *IEEE Trans Instrum Meas* 2020;69:8881-901. DOI
102. Zhan Y, Dai X, Yang E, Wang KC. Convolutional neural network for detecting railway fastener defects using a developed 3D laser system. *Int J Rail Transp* 2021;9:424-44. DOI
103. Li Z, Zhang J, Wang M, Zhong Y, Peng F. Fiber distributed acoustic sensing using convolutional long short-term memory network: a field test on high-speed railway intrusion detection. *Opt Express* 2020;28:2925-38. DOI PubMed
104. Wei X, Wei D, Suo D, Jia L, Li Y. Multi-target defect identification for railway track line based on image processing and improved YOLOv3 model. *IEEE Access* 2020;8:61973-88. DOI
105. Lu J, Liang B, Lei Q, et al. SCueU-Net: efficient damage detection method for railway rail. *IEEE Access* 2020;8:125109-20. DOI
106. Zheng Y, Wu S, Liu D, Wei R, Li S, Tu Z. Sleeper defect detection based on improved YOLO V3 algorithm. Proceedings of 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA); 2020 Nov 9-13; Kristiansand, Norway. IEEE; 2020. p. 955-60. DOI
107. Zhang D, Song K, Wang Q, He Y, Wen X, Yan Y. Two deep learning networks for rail surface defect inspection of limited samples with line-level label. *IEEE Trans Ind Inf* 2021;17:6731-41. DOI
108. Chen S, Zhou L, Ni Y, Liu X. An acoustic-homologous transfer learning approach for acoustic emission-based rail condition evaluation. *Struct Health Monit* 2021;20:2161-81. DOI
109. Kuzmin EV, Gorbunov OE, Plotnikov PO, Tyukin VA, Bashkin VA. Application of convolutional neural networks for recognizing long structural elements of rails in eddy-current defectograms. *Model anal inf sist* 2020;27:316-29. DOI

110. Liu Y, Sun X, Pang JHL. (, March). A YOLOv3-based deep learning application research for condition monitoring of rail thermite welded joints. Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing; 2020 Mar; New York, NY, USA. Association for Computing Machinery; 2020. p. 33-8. DOI
111. Aydin I, Akin E, Karakose M. Defect classification based on deep features for railway tracks in sustainable transportation. *Appl Soft Comput* 2021;111:107706. DOI
112. Zheng D, Li L, Zheng S, et al. A defect detection method for rail surface and fasteners based on deep convolutional neural network. *Comput Intell Neurosci* 2021;2021:2565500. DOI PubMed PMC
113. Wang W, Hu W, Wang W, et al. Automated crack severity level detection and classification for ballastless track slab using deep convolutional neural network. *Autom Constr* 2021;124:103484. DOI
114. Chen Z, Wang Q, Yang K, et al. Deep learning for the detection and recognition of rail defects in ultrasound B-scan images. *Transp Res Rec* 2021;2675:888-901. DOI
115. Liu J, Ma Z, Qiu Y, Ni X, Shi B, Liu H. Four discriminator cycle-consistent adversarial network for improving railway defective fastener inspection. *IEEE Trans Intell Transport Syst* 2021. DOI
116. Wu Y, Qin Y, Qian Y, Guo F, Wang Z, Jia L. Hybrid deep learning architecture for rail surface segmentation and surface defect detection. *Computer aided Civil Eng* 2022;37:227-44. DOI
117. Wan Z, Chen S. Railway tracks defects detection based on deep convolution neural networks. In: Liang Q, Wang W, Mu J, Liu X, Na Z, Cai X, editors. Artificial intelligence in China. Singapore: Springer; 2021. p. 119-29. DOI
118. Ye T, Zhang X, Zhang Y, Liu J. Railway traffic object detection using differential feature fusion convolution neural network. *IEEE Trans Intell Transport Syst* 2021;22:1375-87. DOI
119. Chen M, Zhai W, Zhu S, Xu L, Sun Y. Vibration-based damage detection of rail fastener using fully convolutional networks. *Veh Syst Dyn* 2021. DOI
120. Tai JJ, Innocente MS, Mehmood O. FasteNet: a fast railway fastener detector. In: Yang X, Sherratt S, Dey N, Joshi A, editors. Proceedings of Sixth International Congress on Information and Communication Technology. Singapore: Springer; 2022. p. 767-77. DOI
121. Guo F, Qian Y, Rizos D, Suo Z, Chen X. Automatic rail surface defects inspection based on mask R-CNN. *Transp Res Rec* 2021;2675:655-68. DOI

AUTHOR INSTRUCTIONS

1. Submission Overview

Before you decide to publish with *Intelligence & Robotics* (*IR*), please read the following items carefully and make sure that you are well aware of Editorial Policies and the following requirements.

1.1 Topic Suitability

The topic of the manuscript must fit the scope of the journal. Please refer to Aims and Scope for more information.

1.2 Open Access and Copyright

The journal adopts Gold Open Access publishing model and distributes content under the Creative Commons Attribution 4.0 International License. Copyright is retained by authors. Please make sure that you are well aware of these policies.

1.3 Publication Fees

Before December 31, 2024, there are no article processing charges for papers accepted for publication after peer review. OAE subsidizes and helps authors publish their manuscripts totally free. For more details, please refer to OAE Publication Fees.

1.4 Language Editing

All submissions are required to be presented clearly and cohesively in good English. Authors whose first language is not English are advised to have their manuscripts checked or edited by a native English speaker before submission to ensure the high quality of expression. A well-organized manuscript in good English would make the peer review even the whole Editorial handling more smoothly and efficiently.

If needed, authors are recommended to consider the language editing services provided by Charlesworth to ensure that the manuscript is written in correct scientific English before submission. Authors who publish with OAE journals enjoy a special discount for the services of Charlesworth via the following two ways.

Submit your manuscripts directly at http://www.charlesworthauthorservices.com/~OAE;

Open the link http://www.charlesworthauthorservices.com/, and enter Promotion Code "OAE" when you submit.

1.5 Work Funded by the National Institutes of Health

If an accepted manuscript was funded by National Institutes of Health (NIH), the author may inform editors of the NIH funding number. The editors are able to deposit the paper to the NIH Manuscript Submission System on behalf of the author.

2. Submission Preparation

2.1 Cover Letter

A cover letter is required to be submitted accompanying each manuscript. Here is a guideline of a cover letter for authors' consideration:

List the highlights of the current manuscript and no more than 5 short sentences;

All authors have read the final manuscript, have approved the submission to the journal, and have accepted full responsibilities pertaining to the manuscript's delivery and contents;

Clearly state that the manuscript is an original work on its own merit, that it has not been previously published in whole or in part, and that it is not being considered for publication elsewhere;

No materials are reproduced from another source (if there is material in your manuscript that has been reproduced from another source, please state whether you have obtained permission from the copyright holder to use them);

Conflicts of interest statement;

If the manuscript is contributed to a Special Issue, please also mention it in the cover letter;

If the manuscript was presented partly or entirely in a conference, the author should clearly state the background information of the event, including the conference name, time, and place in the cover letter.

2.2 Types of Manuscripts

There is no restriction on the length of manuscripts, number of figures, tables and references, provided that the manuscript is concise and comprehensive. The journal publishes Research Article, Review, Technical Note, etc. For more details about paper type, please refer to the following table.

Manuscript Type	Definition	Abstract	Keywords	Main Text Structure
Research Article	A Research Article is a seminal and insightful research study and showcases that often involves modern techniques or methodologies. Authors should justify that their work is of novel findings.	The abstract should state briefly the purpose of the research, the principal results and major conclusions. No more than 250 words.	3-8 keywords	The main content should include four sections: Introduction, Methods, Results and Discussion.
Review	A Review should be an authoritative, well balanced, and critical survey of recent progress in an attractive or a fundamental research field.	Unstructured abstract. No more than 250 words.	3-8 keywords	The main text may consist of several sections with unfixed section titles. We suggest that the author include an "Introduction" section at the beginning, several sections with unfixed titles in the middle part, and a "Conclusions" section at the end.
Technical Note	A Technical Note is a short article giving a brief description of a specific development, technique, or procedure, or it may describe a modification of an existing technique, procedure or device applied in research.	Unstructured abstract. No more than 250 words.	3-8 keywords	/
Editorial	An Editorial is a short article describing news about the journal or opinions of senior Editors or the publisher.	None required	None required	/
Commentary	A Commentary is to provide comments on a newly published article or an alternative viewpoint on a certain topic.	Unstructured abstract. No more than 250 words.	3-8 keywords	/
Perspective	A Perspective provides personal points of view on the state-of-the-art of a specific area of knowledge and its future prospects.	Unstructured abstract. No more than 250 words.	3-8 keywords	/

2.3 Manuscript Structure

2.3.1 Front Matter

2.3.1.1 Title

The title of the manuscript should be concise, specific and relevant, with no more than 16 words if possible.

2.3.1.2 Authors and Affiliations

Authors' full names should be listed. The initials of middle names can be provided. The affiliations and email addresses for all authors should be listed. At least one author should be designated as the corresponding author. In addition, corresponding authors are suggested to provide their Open Researcher and Contributor ID upon submission. Please note that any change to authorship is not allowed after manuscript acceptance. The authors' affiliations should be provided in this format: department, institution, city, postcode, country.

2.3.1.3 Abstract

The abstract should be a single paragraph with word limitation and specific structure requirements (for more details please refer to Types of Manuscripts). It usually describes the main objective(s) of the study, explains how the study was done, including any model organisms used, without methodological detail, and summarizes the most important results and their significance. The abstract must be an objective representation of the study: it is not allowed to contain results that are not presented and substantiated in the manuscript, or exaggerate the main conclusions. Citations should not be included in the abstract.

2.3.1.4 Graphical Abstract

The graphical abstract is essential as this can catch first view of your publication by readers. We recommend you submit an eye-catching figure. It should summarize the content of the article in a concise graphical form. It is recommended to use it because this can make online articles get more attention.

The graphical abstract should be submitted as a separate document in the online submission system. Please provide an image with a minimum of 531 × 1328 pixels (h × w) or proportionally more. The image should be readable at a size of 5 cm × 13 cm using a regular screen resolution of 96 dpi. Preferred file types: TIFF, PSD, AI, JPEG, and EPS files.

2.3.1.5 Keywords

Three to eight keywords should be provided, which are specific to the article, yet reasonably common within the subject discipline.

2.3.2 Main Text

Manuscripts of different types are structured with different sections of content. Please refer to Types of Manuscripts to make sure which sections should be included in the manuscripts.

2.3.2.1 Introduction

The introduction should contain background that puts the manuscript into context, allow readers to understand why the study is important, include a brief review of key literature, and conclude with a brief statement of the overall aim of the work and a comment about whether that aim was achieved. Relevant controversies or disagreements in the field should be introduced as well.

2.3.2.2 Methods

The methods should contain sufficient details to allow others to fully replicate the study. New methods and protocols should be described in detail while well-established methods can be briefly described or appropriately cited. Statistical terms, abbreviations, and all symbols used should be defined clearly. Protocol documents for clinical trials, observational studies, and other non-laboratory investigations may be uploaded as supplementary materials.

2.3.2.3 Results

This section contains the findings of the study. Results of statistical analysis should also be included either as text or as tables or figures if appropriate. Authors should emphasize and summarize only the most important observations. Data on all primary and secondary outcomes identified in the section Methods should also be provided. Extra or supplementary materials and technical details can be placed in supplementary documents.

2.3.2.4 Discussion

This section should discuss the implications of the findings in context of existing research and highlight limitations of the study. Future research directions may also be mentioned.

2.3.2.5 Conclusion

It should state clearly the main conclusions and include the explanation of their relevance or importance to the field.

2.3.3 Back Matter

2.3.3.1 Acknowledgments

Anyone who contributed towards the article but does not meet the criteria for authorship, including those who provided professional writing services or materials, should be acknowledged. Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgments section. This section is not added if the author does not have anyone to acknowledge.

2.3.3.2 Authors' Contributions

Each author is expected to have made substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data, or the creation of new software used in the work, or have drafted the work or substantively revised it.

Please use Surname and Initial of Forename to refer to an author's contribution. For example: made substantial contributions to conception and design of the study and performed data analysis and interpretation: Salas H, Castaneda WV; performed data acquisition, as well as providing administrative, technical, and material support: Castillo N, Young V.

If an article is single-authored, please include "The author contributed solely to the article." in this section.

2.3.3.3 Availability of Data and Materials

In order to maintain the integrity, transparency and reproducibility of research records, authors should include this section in their manuscripts, detailing where the data supporting their findings can be found. Data can be deposited into data repositories or published as supplementary information in the journal. Authors who cannot share their data should state that the data will not be shared and explain it. If a manuscript does not involve such issues, please state "Not applicable." in this section.

2.3.3.4 Financial Support and Sponsorship

All sources of funding for the study reported should be declared. The role of the funding body in the experiment design, collection, analysis and interpretation of data, and writing of the manuscript should be declared. Any relevant grant numbers and the link of funder's website should be provided if any. If the study is not involved with this issue, state "None." in this section.

2.3.3.5 Conflicts of Interest

Authors must declare any potential conflicts of interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. If there are no conflicts of interest, please state "All authors declared that there are no conflicts of interest." in this section. Some authors may be bound by confidentiality agreements. In such cases, in place of itemized disclosures, we will require authors to state "All authors declared that they are bound by confidentiality agreements that prevent them from disclosing their conflicts of interest in this work.". If authors are unsure whether conflicts of interest exist, please refer to the "Conflicts of Interest" of *IR* Editorial Policies for a full explanation.

2.3.3.6 Ethical Approval and Consent to Participate

Research involving human subjects, human material or human data must be performed in accordance with the Declaration of Helsinki and approved by an appropriate ethics committee. An informed consent to participate in the study should also be obtained from participants, or their parents or legal guardians for children under 16. A statement detailing the name of the ethics committee (including the reference number where appropriate) and the informed consent obtained must appear in the manuscripts reporting such research.

Studies involving animals and cell lines must include a statement on ethical approval. More information is available at Editorial Policies.

If the manuscript does not involve such issue, please state "Not applicable." in this section.

2.3.3.7 Consent for Publication

Manuscripts containing individual details, images or videos, must obtain consent for publication from that person, or in the case of children, their parents or legal guardians. If the person has died, consent for publication must be obtained from the next of kin of the participant. Manuscripts must include a statement that written informed consent for publication was obtained. Authors do not have to submit such content accompanying the manuscript. However, these documents must be available if requested. If the manuscript does not involve this issue, state "Not applicable." in this section.

2.3.3.8 Copyright

Authors retain copyright of their works through a Creative Commons Attribution 4.0 International License that clearly states how readers can copy, distribute, and use their attributed research, free of charge. A declaration "© The Author(s) 2021." will be added to each article. Authors are required to sign License to Publish before formal publication.

2.3.3.9 References

References should be numbered in order of appearance at the end of manuscripts. In the text, reference numbers should be placed in square brackets and the corresponding references are cited thereafter. If the number of authors is less than or equal to six, we require to list all authors' names. If the number of authors is more than six, only the first three authors' names are required to be listed in the references, other authors' names should be omitted and replaced with "et al.". Abbreviations of the journals should be provided on the basis of Index Medicus. Information from manuscripts accepted but not published should be cited in the text as "Unpublished material" with written permission from the source.

References should be described as follows, depending on the types of works:

Types	Examples
Journal articles by individual authors	Weaver DL, Ashikaga T, Krag DN, et al. Effect of occult metastases on survival in node-negative breast cancer. *N Engl J Med* 2011;364:412-21. [PMID: 21247310 DOI: 10.1056/NEJMoa1008108]
Organization as author	Diabetes Prevention Program Research Group. Hypertension, insulin, and proinsulin in participants with impaired glucose tolerance. *Hypertension* 2002;40:679-86. [DOI: 10.1161/01.HYP.0000035706.28494.09]
Both personal authors and organization as author	Vallancien G, Emberton M, Harving N, van Moorselaar RJ; Alf-One Study Group. Sexual dysfunction in 1,274 European men suffering from lower urinary tract symptoms. *J Urol* 2003;169:2257-61. [PMID: 12771764 DOI: 10.1097/01.ju.0000067940.76090.73]
Journal articles not in English	Zhang X, Xiong H, Ji TY, Zhang YH, Wang Y. Case report of anti-N-methyl-D-aspartate receptor encephalitis in child. *J Appl Clin Pediatr* 2012;27:1903-7. (in Chinese)
Journal articles ahead of print	Odibo AO. Falling stillbirth and neonatal mortality rates in twin gestation: not a reason for complacency. BJOG 2018; Epub ahead of print [PMID: 30461178 DOI: 10.1111/1471-0528.15541]
Books	Sherlock S, Dooley J. Diseases of the liver and billiary system. 9th ed. Oxford: Blackwell Sci Pub; 1993. pp. 258-96.
Book chapters	Meltzer PS, Kallioniemi A, Trent JM. Chromosome alterations in human solid tumors. In: Vogelstein B, Kinzler KW, editors. The genetic basis of human cancer. New York: McGraw-Hill; 2002. pp. 93-113.
Online resource	FDA News Release. FDA approval brings first gene therapy to the United States. Available from: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm574058.htm. [Last accessed on 30 Oct 2017]

Conference proceedings	Harnden P, Joffe JK, Jones WG, Editors. Germ cell tumours V. Proceedings of the 5th Germ Cell Tumour Conference; 2001 Sep 13-15; Leeds, UK. New York: Springer; 2002.
Conference paper	Christensen S, Oppacher F. An analysis of Koza's computational effort statistic for genetic programming. In: Foster JA, Lutton E, Miller J, Ryan C, Tettamanzi AG, editors. Genetic programming. EuroGP 2002: Proceedings of the 5th European Conference on Genetic Programming; 2002 Apr 3-5; Kinsdale, Ireland. Berlin: Springer; 2002. pp. 182-91.
Unpublished material	Tian D, Araki H, Stahl E, Bergelson J, Kreitman M. Signature of balancing selection in Arabidopsis. *Proc Natl Acad Sci U S A*. Forthcoming 2002.

The journal also recommends that authors prepare references with a bibliography software package, such as EndNote to avoid typing mistakes and duplicated references.

2.3.3.10 Supplementary Materials

Additional data and information can be uploaded as Supplementary Materials to accompany the manuscripts. The supplementary materials will also be available to the referees as part of the peer-review process. Any file format is acceptable, such as data sheet (word, excel, csv, cdx, fasta, pdf or zip files), presentation (powerpoint, pdf or zip files), image (cdx, eps, jpeg, pdf, png or tiff), table (word, excel, csv or pdf), audio (mp3, wav or wma) or video (avi, divx, flv, mov, mp4, mpeg, mpg or wmv). All information should be clearly presented. Supplementary materials should be cited in the main text in numeric order (e.g., Supplementary Figure 1, Supplementary Figure 2, Supplementary Table 1, Supplementary Table 2, *etc*.). The style of supplementary figures or tables complies with the same requirements on figures or tables in main text. Videos and audios should be prepared in English, and limited to a size of 500 MB.

2.4 Manuscript Format

2.4.1 File Format

Manuscript files can be in DOC and DOCX formats and should not be locked or protected.
Manuscript prepared in LaTex must be collated into one ZIP folder (including all source files and images, so that the Editorial Office can recompile the submitted PDF).
When preparing manuscripts in different file formats, please use the corresponding Manuscript Templates.

2.4.2 Length

There are no restrictions on paper length, number of figures, or number of supporting documents. Authors are encouraged to present and discuss their findings concisely.

2.4.3 Language

Manuscripts must be written in English.

2.4.4 Multimedia Files

The journal supports manuscripts with multimedia files. The requirements are listed as follows:

Video or audio files are only acceptable in English. The presentation and introduction should be easy to understand. The frames should be clear, and the speech speed should be moderate;

A brief overview of the video or audio files should be given in the manuscript text;

The video or audio files should be limited to a size of up to 500 MB;

Please use professional software to produce high-quality video files, to facilitate acceptance and publication along with the submitted article. Upload the videos in mp4, wmv, or rm format (preferably mp4) and audio files in mp3 or wav format.

2.4.5 Figures

Figures should be cited in numeric order (e.g., Figure 1, Figure 2) and placed after the paragraph where it is first cited;
Figures can be submitted in format of TIFF, PSD, AI, EPS or JPEG, with resolution of 300-600 dpi;
Figure caption is placed under the Figure;
Diagrams with describing words (including, flow chart, coordinate diagram, bar chart, line chart, and scatter diagram, *etc*.) should be editable in word, excel or powerpoint format. Non-English information should be avoided;
Labels, numbers, letters, arrows, and symbols in figure should be clear, of uniform size, and contrast with the background;
Symbols, arrows, numbers, or letters used to identify parts of the illustrations must be identified and explained in the legend;
Internal scale (magnification) should be explained and the staining method in photomicrographs should be identified;
All non-standard abbreviations should be explained in the legend;
Permission for use of copyrighted materials from other sources, including re-published, adapted, modified, or partial figures and images from the internet, must be obtained. It is authors' responsibility to acquire the licenses, to follow any citation instruction requested by third-party rights holders, and cover any supplementary charges.

2.4.6 Tables

Tables should be cited in numeric order and placed after the paragraph where it is first cited;
The table caption should be placed above the table and labeled sequentially (e.g., Table 1, Table 2);
Tables should be provided in editable form like DOC or DOCX format (picture is not allowed);
Abbreviations and symbols used in table should be explained in footnote;
Explanatory matter should also be placed in footnotes;
Permission for use of copyrighted materials from other sources, including re-published, adapted, modified, or partial tables from the internet, must be obtained. It is authors' responsibility to acquire the licenses, to follow any citation instruction requested by third-party rights holders, and cover any supplementary charges.

2.4.7 Abbreviations

Abbreviations should be defined upon first appearance in the abstract, main text, and in figure or table captions and used consistently thereafter. Non-standard abbreviations are not allowed unless they appear at least three times in the text. Commonly-used abbreviations, such as DNA, RNA, ATP, *etc*., can be used directly without definition. Abbreviations in titles and keywords should be avoided, except for the ones which are widely used.

2.4.8 Italics

General italic words like *vs*., *et al*., *etc*., *in vivo*, *in vitro*; *t* test, *F* test, *U* test; related coefficient as *r*, sample number as *n*, and probability as *P*; names of genes; names of bacteria and biology species in Latin.

2.4.9 Units

SI Units should be used. Imperial, US customary and other units should be converted to SI units whenever possible. There is a space between the number and the unit (i.e., 23 mL). Hour, minute, second should be written as h, min, s.

2.4.10 Numbers

Numbers appearing at the beginning of sentences should be expressed in English. When there are two or more numbers in a paragraph, they should be expressed as Arabic numerals; when there is only one number in a paragraph, number < 10 should be expressed in English and number > 10 should be expressed as Arabic numerals. 12345678 should be written as 12,345,678.

2.4.11 Equations

Equations should be editable and not appear in a picture format. Authors are advised to use either the Microsoft Equation Editor or the MathType for display and inline equations.
Display equations should be numbered consecutively, using Arabic numbers in parentheses;
Inline equations should not be numbered, with the same/similar size font used for the main text.

2.4.12 Headings

In the main body of the paper, three different levels of headings may be used.
Level one headings: they should be in bold, and numbered using Arabic numbers, such as **1. INTRODUCTION**, and **2. METHODS**, with all letters capitalized;
Level two headings: they should be in bold and numbered after the level one heading, such as **2.1 Statistical analyses**, **2.2** ..., **2.3**..., *etc*., with the first letter capitalized;
Level three headings: they should be italicized, and numbered after the level two heading, such as *2.1.1 Data distributions*,and *2.1.2 outliers and linear regression*, with the first letter capitalized.

2.4.13 Text Layout

As the electronic submission will provide the basic material for typesetting, it is important to prepare papers in the general editorial style of the journal.
The font is Times New Roman;
The font size is 12pt;
Single column, 1.5× line spacing;
Insert one line break (one Return) before the heading and paragraph, if the heading and paragraph are adjacent, insert a line break before the heading only;
No special indentation;
Alignment is left end;
Insert consecutive line numbers;
For other details please refer to the Manuscript Templates.

2.5 Submission Link

Submit an article via https://oaemesas.com/login?JournalId=ir.

3. Publication Ethics Statement

OAE is a member of the Committee on Publication Ethics (COPE). We fully adhere to its Code of Conduct and to its Best Practice Guidelines.

The Editors of this journal enforce a rigorous peer-review process together with strict ethical policies and standards to guarantee to add high-quality scientific works to the field of scholarly publication. Unfortunately, cases of plagiarism, data falsification, image manipulation, inappropriate authorship credit, and the like, do arise. The Editors of *IR* take such publishing ethics issues very seriously and are trained to proceed in such cases with zero tolerance policy.

Authors wishing to publish their papers in *IR* must abide by the following:

The author(s) must disclose any possibility of a conflict of interest in the paper prior to submission;
The authors should declare that there is no academic misconduct in their manuscript in the cover letter;
Authors should accurately present their research findings and include an objective discussion of the significance of their findings;
Data and methods used in the research need to be presented in sufficient detail in the manuscript so that other researchers can replicate the work;
Authors should provide raw data if referees and the Editors of the journal request;
Simultaneous submission of manuscripts to more than one journal is not tolerated;
Republishing content that is not novel is not tolerated (for example, an English translation of a paper that is already published in another language will not be accepted);
The manuscript should not contain any information that has already been published. If you include already published figures or images, please get the necessary permission from the copyright holder to publish under the CC-BY license;
Plagiarism, data fabrication and image manipulation are not tolerated;
Plagiarism is not acceptable in OAE journals.

Plagiarism involves the inclusion of large sections of unaltered or minimally altered text from an existing source without appropriate and unambiguous attribution, and/or an attempt to misattribute original authorship regarding ideas or results, and copying text, images, or data from another source, even from your own publications, without giving credit to the source.

As to reusing the text that is copied from another source, it must be between quotation marks and the source must be cited. If a study's design or the manuscript's structure or language has been inspired by previous studies, these studies must be cited explicitly.

If plagiarism is detected during the peer-review process, the manuscript may be rejected. If plagiarism is detected after publication, we may publish a Correction or retract the paper.

Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results so that the findings are not accurately represented in the research record.

Image files must not be manipulated or adjusted in any way that could lead to misinterpretation of the information provided by the original image.

Irregular manipulation includes: introduction, enhancement, moving, or removing features from the original image; the grouping of images that should be presented separately, or modifying the contrast, brightness, or color balance to obscure, eliminate, or enhance some information.

If irregular image manipulation is identified and confirmed during the peer-review process, we may reject the manuscript. If irregular image manipulation is identified and confirmed after publication, we may publish a Correction or retract the paper.

OAE reserves the right to contact the authors' institution(s) to investigate possible publication misconduct if the Editors find conclusive evidence of misconduct before or after publication. OAE has a partnership with iThenticate, which is the most trusted similarity checker. It is used to analyze received manuscripts to avoid plagiarism to the greatest extent possible. When plagiarism becomes evident after publication, we will retract the original publication or require modifications, depending on the degree of plagiarism, context within the published article, and its impact on the overall integrity of the published study. Journal Editors will act under the relevant COPE guidelines.

4. Authorship

Authorship credit of *IR* should be solely based on substantial contributions to a published study, as specified in the following four criteria:

1. Substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work;
2. Drafting the work or revising it critically for important intellectual content;
3. Final approval of the version to be published;
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

All those who meet these criteria should be identified as authors. Authors must specify their contributions in the section Authors' Contributions of their manuscripts. Contributors who do not meet all the four criteria (like only involved in acquisition of funding, general supervision of a research group, general administrative support, writing assistance, technical editing, language editing, proofreading, *etc*.) should be acknowledged in the section of Acknowledgement in the manuscript rather than being listed as authors.

If a large multiple-author group has conducted the work, the group ideally should decide who will be authors before the work starts and confirm authors before submission. All authors of the group named as authors must meet all the four criteria for authorship.

5. Reviewers Exclusions

You are welcome to exclude a limited number of researchers as potential Editors or reviewers of your manuscript. To ensure a fair and rigorous peer review process, we ask that you keep your exclusions to a maximum of three people. If you wish to exclude additional referees, please explain or justify your concerns—this information will be helpful for Editors when deciding whether to honor your request.

6. Editors and Journal Staff as Authors

Editorial independence is extremely important and OAE does not interfere with Editorial decisions. Editorial staff or Editors shall not be involved in processing their own academic work. Submissions authored by Editorial staff/Editors will be assigned to at least two independent outside reviewers. Decisions will be made by the Editor-in-Chief, including Special Issue papers. Journal staff are not involved in the processing of their own work submitted to any OAE journals.

7. Conflict of Interests

OAE journals require authors to declare any possible financial and/or non-financial conflicts of interest at the end of their manuscript and in the cover letter, as well as confirm this point when submitting their manuscript in the submission system. If no conflicts of interest exist, authors need to state "All authors declared that there are no conflicts of interest". We also recognize that some authors may be bound by confidentiality agreements, in which cases authors need to state "All authors declared that they are bound by confidentiality agreements that prevent them from disclosing their competing interests in this work".

8. Editorial Process

8.1. Pre-Check

New submissions are initially checked by the Managing Editor from the perspectives of originality, suitability, structure and formatting, conflicts of interest, background of authors, *etc*. Poorly prepared manuscripts may be rejected at this stage. If your manuscript does not meet one or more of these requirements, we will return it for further revisions.

Once your manuscript has passed the initial check, it will be assigned to the Assistant Editor, and then the Editor-in-Chief, or an Associate Editor in the case of a conflict of interest, will be notified of the submission and invited to review. Regarding Special Issue paper, after passing the initial check, the manuscript will be successively assigned to the Assistant Editor, and then to the Editor-in-Chief, or an Associate Editor in the case of conflict of interest for the Editor-in-Chief to review. The Editor-in-Chief, or the Associate Editor may reject manuscripts that they deem highly unlikely to pass peer review without further consultation. Once your manuscript has passed the Editorial assessment, the Associate Editor will start to organize peer-review.

All manuscripts submitted to *IR* are screened using CrossCheck powered by iThenticate to identify any plagiarized content. Your study must also meet all ethical requirements as outlined in our Editorial Policies. If the manuscript does not pass any of these checks, we may return it to you for further revisions or decline to consider your study for publication.

8.2. Peer Review

IR operates a single-blind review process, which means that reviewers know the names of authors, but the names of the reviewers are hidden from the authors. The scientific quality of the research described in the manuscript is assessed

by a minimum of two independent expert reviewers. The Editor-in-Chief is responsible for the final decision regarding acceptance or rejection of the manuscript.

All information contained in your manuscript and acquired during the review process will be held in the strictest confidence.

8.3. Decisions

Your research will be judged on scientific soundness only, not on its perceived impact as judged by Editors or referees. There are three possible decisions: Accept (your study satisfies all publication criteria), Invitation to Revise (more work is required to satisfy all criteria), and Reject (your study fails to satisfy key criteria and it is highly unlikely that further work can address its shortcomings). All of the following publication criteria must be fulfilled to enable your manuscript to be accepted for publication:

Originality
The study reports original research and conclusions.
Data availability
All data to support the conclusions either have been provided or are otherwise publicly available.
Statistics
All data have been analyzed through appropriate statistical tests and these are clearly defined.
Methods
The methods are described in sufficient detail to be replicated.
Citations
Previous work has been appropriately acknowledged.
Interpretation
The conclusions are a reasonable extension of the results.
Ethics
The study design, data presentation, and writing style comply with our Editorial Policies.

8.4. Revisions

Authors are required to submit the revised manuscript within one week if minor revision is recommended while two weeks if major revision recommended or one month if additional experiments are needed. If authors need more than one month to revise their manuscript, we usually require the authors to resubmit their paper. We request that a document of point-to-point response to all comments of reviewers and the Editor-in-Chief or the Associate Editor should be supplied along with the revised manuscript to allow quick assessment of your revised manuscript. This document should outline in detail how each of the comments was addressed in the revised manuscript or should provide a rebuttal to the criticism. Manuscripts may or may not be sent to reviewers after revision, dependent on whether the reviewer requested to see the revised version. Apart from in exceptional circumstances, *IR* only supports a round of major revision per manuscript.

9. Contact Us

Journal Contact

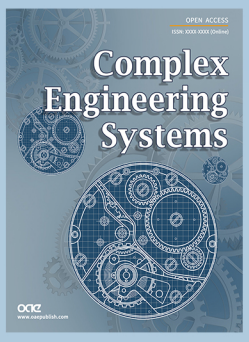
Intelligence & Robotics Editorial Office
Suite 1504, Plaza A, Xi'an National Digital Publishing Base,
No. 996 Tiangu 7th Road, Gaoxin District, Xi'an 710077, Shaanxi, China.

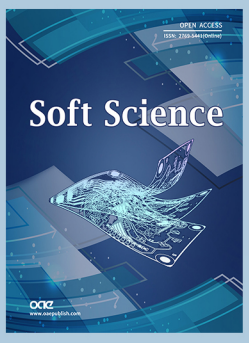
Managing Editor

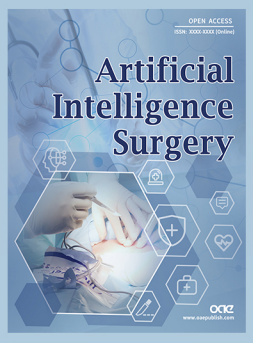
Lijun Jin
Email: editorial@intellrobot.com

OAE Publishing Inc. (https://oaepublish.com/) is a multidisciplinary open-access publishing company, founded in Los Angeles in 2015. Until now, OAE has been recognized by authoritative organizations in publishing industries, such as the ORCID, COPE, Scientific, Technical and Medical Publishers (STM), Crossref, and EASE.

As of July 2021, more than 1,200 outstanding scholars have joined OAE, who are from world-renowned universities and research institutions, including European Academy of Sciences, American Academy of Invention Sciences, Chinese Academy of Sciences, Royal Academy of Sciences of Belgium, British Academy of Medical Sciences, etc. There are more than 30 journals founded by OAE (https://oaepublish.com/about/journals), such as Intelligence & Robotics, Journal of Materials Informatics, Complex Engineering Systems, Journal of Smart Environments and Green Computing, and Soft Science, etc. Part of journals have been indexed by Scopus and CAS. We are currently working on database application including PubMed and ESCI. Up to July 2021, 2,354 articles have been published online, with 13,131,129 hits and 963,586 downloads. In the future, OAE Publishing Company will continue to found more quality journals with outstanding scholars, to promote the global academic development.







OAE Official Website