Thesis for The Degree of Licentiate of Engineering

Toward Enabling Robotic Visual Perception for Assembly Tasks

An Application in Wire Harness Assembly onto Electric Vehicles

Hao Wang

Department of Industrial and Materials Science
Chalmers University of Technology
Gothenburg, Sweden, 2024

Toward Enabling Robotic Visual Perception for Assembly Tasks
An Application in Wire Harness Assembly onto Electric Vehicles

HAO WANG

This thesis has been prepared using LaTeX.

Licentiate Thesis at Chalmers University of Technology
Technical Report No IMS-2024-5

Department of Industrial and Materials Science
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone: +46 (0)31 772 1000

Printed by Chalmers Digital Printing,
Gothenburg, Sweden 2024

To Mom and Dad.

Contents

Toward Enabling Robotic Visual Perception for Assembly Tasks
An Application in Wire Harness Assembly onto Electric Vehicles

Hao Wang
Department of Industrial and Materials Science
Chalmers University of Technology

Abstract

Industry faces an urgent need for prospective solutions to scale up assembly automation, a challenge that requires immediate attention. In contemporary manufacturing, industrial robots need more intelligence to qualify for increasingly demanding flexible automation tasks. Research in artificial intelligence, computer vision, and robotics paints a promising picture of the future, where intelligent robots play a significant role in fostering sustainable and resilient manufacturing. However, academia and industry have yet to realize the potential of intelligent robots in production fully.

This thesis plays a pivotal role in advancing the development of intelligent robots for flexible automation tasks, a crucial area of research in automation and robotics. Toward this goal, this thesis investigates perception, a prerequisite of intelligence, and mainly focuses on visual perception, a critical contactless perception approach. A multi-method research approach, comprising a qualitative literature study and a quantitative experimental study, was adopted to explore the challenges and prospective technical solutions to enabling robotic visual perception for assembly tasks.

The research has identified four key challenges in enabling robotic visual perception for assembly tasks, particularly in developing and integrating vision systems in practical production. Additionally, the research has proposed six prospective directions for developing technical solutions, focusing on computer vision algorithms, dataset and benchmark, practical evaluation, human-robot collaboration, and product design.

Keywords

Robotic visual perception, Computer vision, Artificial intelligence, AI, Human-robot collaboration, HRC, Flexible automation, Assembly, Automotive industry

Acknowledgment

This work would only have been realized with much help and support from different people and organizations. I want to acknowledge those who have contributed to bringing this work to fruition.

Firstly, I am deeply grateful to my examiner, Johan Stahre, my main supervisor, Björn Johansson, and my co-supervisor, Anders Skoogh. Their guidance and support have been instrumental in instilling in me a passion for my study and research and in helping me realize this work.

I also appreciate the rest of the Production Systems division for the many beneficial discussions and friendly chats. A relaxed and friendly working environment is genuinely an incubator of inspiration and collaboration toward success. Likewise, I thank all other people in academia or industry with whom we have talked, collaborated, and exchanged ideas, especially those who co-authored the papers appended to this thesis.

Besides, I greatly appreciate the significant impact of all the partners we have collaborated with on research projects. The invaluable support from the funding organizations, Vinnova, Produktion2030, the European Commission, and Chalmers Production Area of Advance, has also been crucial and greatly acknowledged.

Lastly, my eternally special gratitude goes to my most loved ones. To my mom and dad, for bringing me the courage to explore the unknowns. To my fiancée, Shan, for casting the brightest light on my way forward.

Acronyms and Abbreviations

2D	Two-Dimensional
2DOD	2D Object Detection
3D	Three-Dimensional
3DOD	3D Object Detection
6D	Six-Degree-of-Freedom
6DOPE	6D Object Pose Estimation
AI	Artificial Intelligence
BDLO	Branched Deformable Linear Object
CAD	Computer-Aided Design
CIRP	The International Institution of Production Engineering Research
CNN	Convolutional Neural Network
DARE	The Database of Abstracts of Reviews of Effects
DETR	Detection Transformer
DINO	DETR with Improved Denoising Anchor Boxes
DLO	Deformable Linear Object
DLON	Deformable Linear Object Network
DOO	Deformable One-dimensional Object
DPM	Deformable Part Model
DSR	Design Science Research
DSRM	Design Science Research Methodology
EV	Electric Vehicle
FPN	Feature Pyramid Network
HOG	Histogram of Oriented Gradients
HRC	Human-Robot Collaboration
LiDAR	Light Detection And Ranging
mAP	mean Average Precision
NHS	The National Health Service
PASCAL	Pattern Analysis, Statistical Modelling and Computational Learning
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
R-CNN	Region-based CNN
RGB	Red-Green-Blue
RGB-D	Red-Green-Blue-Depth
ROI	Region of Interest
RQ	Research Question
SDLO	Semi-Deformable Linear Object
SPPNet	Spatial Pyramid Pooling Network
SSD	Single-Shot Detector
TRL	Technology Readiness Level
VOC	Visual Object Classes
YOLO	You Only Look Once

List of Publications

Appended Publications

This thesis is based on the following publications:

[**Paper 1**] **Overview of Computer Vision Techniques in Robotized Wire Harness Assembly: Current State and Future Opportunities**

Hao Wang, Omkar Salunkhe, Walter Quadrini, Dan Lämkull, Fredrik Ore, Björn Johansson, Johan Stahre

Presented at the 56th CIRP Conference on Manufacturing Systems (CIRP CMS 2023), Cape Town, South Africa, 24-26 October 2023.

Published in: *Procedia CIRP*, 2023.

`https://doi.org/10.1016/j.procir.2023.09.127`

Hao Wang initiated and wrote the paper with other authors' contributions and reviews. The research design as well as data collection and analysis were conducted together with Omkar Salunkhe and Walter Quadrini.

[**Paper 2**] **A Systematic Literature Review of Computer Vision Applications in Robotized Wire Harness Assembly**

Hao Wang, Omkar Salunkhe, Walter Quadrini, Dan Lämkull, Fredrik Ore, Mélanie Despeisse, Luca Fumagalli, Johan Stahre, Björn Johansson

Under the second revision by *Advanced Engineering Informatics*, April 2024.

Hao Wang initiated and wrote the paper with other authors' contributions and reviews. The research design as well as data collection and analysis were conducted together with Omkar Salunkhe and Walter Quadrini.

[**Paper 3**] **Deep Learning-Based Connector Detection for Robotized Assembly of Automotive Wire Harnesses**

Hao Wang, Björn Johansson

Presented at IEEE 19th International Conference on Automation Science and Engineering (CASE 2023), Auckland, New Zealand, 26-30 August 2023.

Published in *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*, 2023.

`https://doi.org/10.1109/CASE56687.2023.10260619`

Hao Wang initiated and wrote the paper, collected and analyzed the dataset, and designed and conducted the experiments, with Björn Johansson's contribution and review.

Other Publications

The following publications were published during my PhD studies, or are currently in submission/under revision. However, they are not appended to this thesis due to contents not sufficiently related to the thesis.

[Paper a] **Manufacturing Challenges and Opportunities for Sustainable Battery Life Cycles**

Björn Johansson, Mélanie Despeisse, Jon Bokrantz, Greta Braun, Huizhong Cao, Arpita Chari, Qi Fang, Clarissa A. González Chávez, Anders Skoogh, Henrik Söderlund, **Hao Wang**, Kristina Wärmefjord, Lars Nyborg, Jinhua Sun, Roland Örtengren, Kelsea Schumacher, Laura Espinal, Katherine Morris, Jason Nunley, Yusuke Kishita, Yasushi Umeda, Federica Acerbi, Marta Pinzone, Hanna Persson, Sophie Charpentier, Kristina Edström, Daniel Brandell, Maheshwaran Gopalakrishnan, Hossein Rahnama, Lena Abrahamsson, Anna Öhrwall Rönnbäck, Johan Stahre

Under review by a journal, March 2024.

Hao Wang contributed to the content related to computer vision techniques for automation and human-robot collaboration in battery manufacturing with other authors' contributions, content, and reviews.

[Paper b] **Review of Current Status and Future Directions for Collaborative and Semi-Automated Automotive Wire Harnesses Assembly**

Omkar Salunkhe, Walter Quadrini, **Hao Wang**, Johan Stahre, David Romero, Luca Fumagalli, Dan Lämkull

Presented at the 56th CIRP Conference on Manufacturing Systems (CIRP CMS 2023), Cape Town, South Africa, 24-26 October 2023.

Published in *Procedia CIRP*, 2023.

`https://doi.org/10.1016/j.procir.2023.09.061`

Hao Wang contributed to the content related to computer vision applications in automated automotive wire harness assembly and conducted data collection and analysis.

[Paper c] **Battery Production Systems: State of the Art and Future Developments**

Mélanie Despeisse, Björn Johansson, Jon Bokrantz, Greta Braun, Arpita Chari, Xiaoxia Chen, Qi Fang, Clarissa A. González Chávez, Anders Skoogh, Johan Stahre, Ninan Theradapuzha Mathew, Ebru Turanoglu Bekar, **Hao Wang**, Roland Örtengren

Presented at IFIP International Conference on Advances in Production Management Systems (APMS 2023), Trondheim, Norway, 17-21 September 2023.

Published in *Advances in Production Management Systems. Production Management Systems for Responsible Manufacturing, Service, and Logistics Futures*, 2023.

`https://doi.org/10.1007/978-3-031-43688-8_36`

Hao Wang contributed to the content related to computer vision techniques for automation and human-robot collaboration in battery manufacturing with other authors' contributions, content, and reviews.

Part I

Summary

Chapter 1

Introduction

This chapter begins with the research background and the core problem investigated. Then, the vision and aim of the research are elaborated, followed by the formulation of the research questions. Further, this chapter sets the scope and delimitation of the research. Lastly, the thesis outline is delineated.

1.1 Background

The advent of automation and robotics revolutionized the overall industry. In the third industrial revolution, the development in electronics and information technology promoted the massive adaptation of automation in different sectors of the modern industry significantly [1]. Implementing automation can better quality [2], promote productivity [3], optimize resource allocation [2], and improve working conditions [4]. The development in electronics and information technology also laid the foundation of robotics and has propelled robots to become a critical role in human life [5]. Specifically, industry has witnessed a remarkable expansion of robotic automation applications over the years [6]. Since its birth in the 1950s, industrial robots have become a significant enabler of automation and have changed the manufacturing industry radically [7].

Despite the extensive usage of industrial robots, there are still places in the manufacturing industry where a higher degree of robotic automation is desired but still needs to be achieved. As a representative area, the final assembly has long been anticipated to reach the automation rate between 35% to 75% of overall final assembly operations [8]. However, a high degree of manual operations remains common in contemporary assembly [9], albeit the degree of automation is rising [4]. The high complexity of assembly processes constrains the automation of the final assembly [10]. Human-based teamwork organizations have demonstrated superiority over machines in flexibility [11] and efficiency [12] for assembly tasks. Nonetheless, the global demographic change urges industry to pursue solutions to compensate for the potential shortage of workforce [12]. On the other hand, this high degree of manual operations causes production problems concerning the business aspect (e.g., quality and productivity [12]) and the human aspect (e.g., safety and ergonomics [13]). Hence, more automation is desired, especially considering the consistent and increasing pursuit of improvement in quality, efficiency, and sustainability in society [14], [15].

Currently, industry is exploring more automation solutions to address these remaining issues in production [16]. More industrial robots are expected to be deployed to enable and facilitate automation [6]. Meanwhile, industry needs industrial robots with more autonomy to handle assembly tasks that are even more challenging [17]. Particularly in the final assembly, the ongoing industrial paradigm shift from mass production to mass customization and personalization leads to small batch size and considerable variations in the assembly line [18]. This shift stimulates the transformation toward more flexible automation to fulfill increasingly complex manufacturing tasks and the enlarging market of customized products [19]. Numerous tasks require robots to generalize their skills to adapt to specific task scenarios and handle various activities [7]. However, it is infeasible for conventional industrial robots in current production to fulfill this requirement on generalization due to their lack of autonomy [10]. Recent robotics research has investigated human-robot collaboration (HRC) and its applications in industrial tasks to exploit the combination of robots' advantages on repeatability, accuracy, and physical strength and humans' superiority in cognitive abilities and flexibility [20]. Nevertheless, intelligent robots are needed to understand the surroundings in unstructured environments [16], especially for human-centered robot applications [21]. Therefore, industrial robots need advanced cognitive abilities to be competent at flexible automation tasks in the new era [1].

Perception is a prerequisite to enabling intelligent robots with cognitive abilities [22]. It consists of obtaining sensory input and interpreting it meaningfully [23]. Robotics has long been considered a discipline that studies "the intelligent connection of perception to action" [24]. A robot with advanced cognitive abilities can be regarded as a specific intelligent robotic agent that can perceive its environment through sensors and react to that environment through actuators [25]. As one of the fundamental components in robotic manipulation, knowing the position and orientation of an object is the premise of accomplishing the following manipulation operations upon the object. The autonomous recognition of positions and orientations of objects of interest is especially critical when product variants are involved in assembly tasks and/or the positions and orientations of objects of interest are difficult or impossible to define by humans in advance [26]. Therefore, enabling industrial robots to identify the parts to be assembled autonomously is significant to facilitating the robotization of complex assembly tasks [27].

Among other sensory inputs, vision is instrumental for recognizing objects and obtaining their positions and orientations [5]. Visual machine perception is one of the significant tasks studied in artificial intelligence (AI) and robotics, where many computer vision techniques have been adapted [22]. Previous research in computer vision, AI, and robotics has discussed various solutions for enabling robotic visual perception [28] and indicated the potential to promote intelligent robotic automation toward enabling flexible automation [1] and intelligent manufacturing [29]. Nevertheless, research on robotic visual perception for assembly tasks remains primary in laboratory scenarios [30]. Further research is required to reveal the challenges of enabling robotic visual perception for assembly tasks and prospective technical solutions to bring intelligent robotic assembly to fruition.

1.2 Problem Description

This thesis investigates the problem of enabling robotic visual perception to automate wire harness assembly in automobiles' final assembly. Specifically, this thesis explores the challenges and potential technical solutions to enabling robotic visual perception. Enabling visual perception capabilities will contribute to increasing industrial robots' autonomy and making them competent at wire harness assembly. With this, this thesis can provide insights into problems that should be addressed in academia. This thesis may also help industry decision-makers understand the potential challenges and opportunities of promoting robotization in the final assembly. In the long term, theoretical research can be realized satisfactorily in production.

Robotizing all or part of the assembly operations of wire harnesses in the final assembly is desired to address problems in production due to quality, efficiency, safety, ergonomics, and demographic change. It is essential to guarantee a high-quality installation of wire harnesses onto vehicles because: 1) they are fundamental elements within an automotive electronic system; 2) they are widely distributed in automobiles; and 3) they are responsible for quality-essential and safety-critical functions of automobiles [31]. The manual assembly in the current production of automobiles makes it challenging to guarantee consistent assembly quality [32]. It is also fundamental to assure the efficiency of wire harness assembly. On the one hand, the automotive industry persists in a continuous demand for productivity to promote competitiveness and acquire market share. On the other hand, the usage of wire harnesses in modern vehicles has been enlarging remarkably over the years, and industry expects the continued growth of wire harnesses installed in future automobiles [33]. The current manual assembly has been identified as one of the significant bottlenecks of automobile production promotion [32], [34], [35]. Moreover, it is crucial to ensure safety and improve ergonomics for human operators when assembling wire harnesses. Some manual assembly procedures could be more ergonomic for human operators, such as heavy lifting, high-pressure pressing, far-reaching operation, and repetitive movements [36]. These operations can cause severe musculoskeletal disorders and occupational safety and health issues in the workforce [37]. There are also high-voltage wire harnesses installed in automobiles, especially in electric vehicles (EVs), which demands more careful object handling regarding safety, assembly quality, and reliability [38], [39]. In addition, previous research has indicated a potential shortage of either skilled or unskilled workforce willing to work in automotive factories [12]. Therefore, assuring assembly quality and safety and promoting productivity while improving ergonomics and optimizing resource utilization is desired [36]. Implementing robotic assembly automation is one of the prominent approaches [7].

Automotive wire harness assembly has remained manual over the years and is challenging to automate mainly due to the high complexity of assembly processes [12]. This high complexity stems from various reasons, e.g., the considerable product variants due to the shift toward mass customization and personalization [14], the mix of rigid and non-rigid wire harness components and the deformation of wire harnesses [40], the limited process time in practical production [41], and safety concerns on robot deployment [42]–[44]. Among them, the considerable product variants and the deformation of wire harnesses make it unwieldy to program industrial robots in advance by hand. Industrial robots need to be able to perceive the whole or part of the assembly task and figure out their movements autonomously to be competent at either fully or semi-automated wire harness assembly [36], [40]. Visual perception is a fundamental contactless approach for robots to extract information from the surrounding environment [5]. However, in industrial applications, vision-based robotic assembly of wire harnesses has yet to succeed [40]. Enabling robotic visual perception is, thus, an important aspect to investigate when automating wire harness assembly.

1.3 Vision and Aim

This thesis envisions a sustainable manufacturing industry where robots are intelligent and cognizant of their tasks, the surrounding environment, and the humans nearby. With such intelligence, robots can handle all tasks that are either non-value-adding or not ergonomic to human operators. Robots can also adapt and react flexibly to tasks and situations with minimum human intervention requirements. Realizing and applying such intelligent robots will contribute to the symbiosis of humans and robots toward highly efficient production without problems regarding quality, safety, and ergonomics.

Toward such a vision, this thesis aims to facilitate enabling robotic visual perception to promote the degree of autonomy of industrial robots. With visual machine perception enabled, industrial robots can be improved to achieve higher levels of autonomy to handle more robotic manipulation required in flexible automation applications. With a better robotic perception, a robot can adapt and react to non-predefined situations and accomplish new tasks under unstructured physical configurations in the final assembly.

1.4 Research Questions

Although the significance of visual machine perception for increasing robotic autonomy has been recognized in previous research and by industry, the extensive application of vision-based robotic assembly has yet to succeed in practice [45]. To promote industrial robots' autonomy and competence in robotic assembly, this thesis investigates the aspect of visual machine perception and explores potential solutions to enabling robotic visual perception for assembly tasks.

The ultimate goal of this thesis is to suggest technical solutions to enabling robotic visual perception for assembly tasks. Nevertheless, it is necessary to understand the challenges of enabling robotic visual perception before solutions can be suggested and tested. Hence, the first research question (RQ1) is formulated to discover the challenges and indicate prospective directions for the following studies on technical solutions:

RQ1: What are the challenges of enabling robotic visual perception for assembly tasks?

This thesis intends to answer this research question by providing an overview of the challenges of adapting computer vision techniques for the robotic assembly of deformable objects in the final assembly.

With the challenges understood, the next step is to explore the opportunities for potential research and identify the prospective solutions to enabling robotic visual perception toward more autonomy on industrial robots. Research is needed to reveal how the identified challenges can be addressed and to study prospective technical solutions for enabling robotic visual perception, which would increase industrial robots' autonomy and competence in robotic assembly tasks. This leads to the second research question (RQ2):

RQ2: How can robotic visual perception be enabled for assembly tasks?

This research question is formulated to identify potential research opportunities and explore prospective vision-based approaches for promoting industrial robots' autonomy and competence in handling deformable objects in assembly tasks.

1.5 Scope and Delimitation

This thesis's topic lies at the intersection of automation, robotics, computer vision, and AI, particularly in the context of the final assembly. The scope of this thesis is meticulously defined, focusing on understanding the challenges and exploring potential opportunities and technical solutions for enabling robotic visual perception. This research aims to enhance the autonomy of industrial robots in robotic assembly. Though targeting to facilitate the overall manufacturing industry, the application area that this thesis researched is the final assembly in the automotive industry. This thesis explores robotic visual perception. It delves into potential technical solutions based on existing theoretical research in computer vision, AI, and robotics. In addition, this thesis only considers the technical aspect on the experimental level under simplified laboratory configurations. Thus, the delimitation of this thesis is defined as follows.

This thesis does not evaluate the extent of the improvement in robotic autonomy nor discusses the metrics for such an evaluation. This thesis does not analyze the practicality of potential technical solutions for enabling robotic visual perception for applications in actual production. This thesis does not investigate human-robot collaboration. Though critical for practical application, the human factors that may affect the application of technologies, such as the organizational culture and employees' attitudes, are also not studied in this thesis.

1.6 Thesis Outline

This thesis is organized in the following structure.

Chapter 1 Introduction delineates the background of the research in this thesis and the core problem focused in this research. The vision and aim are described, followed by the research questions of this research. Lastly, the scope and delimitation of this research and the thesis outline are explained.

Chapter 2 Theoretical Framework describes the theoretical framework of this thesis, including the background knowledge on robotic assembly and robotic perception and the state of the art of robotized wire harness assembly.

Chapter 3 Research Approach elaborates on the design of the research approach of this research, including the adopted philosophical worldview, research design, and research methods, followed by the methods adopted for guaranteeing the research quality.

Chapter 4 Summary of Appended Papers briefly summarizes each of the three appended papers, including each paper's core problem, methodology, and contribution. This chapter concludes with a summary of each appended paper's contribution to each research question.

Chapter 5 Discussion discusses the main findings in each study of this research and the answers to the research questions formulated in Section 1.4 in Chapter 1. Then, this chapter analyzes the contribution of this thesis to academia and industry. Reflections upon the research follow regarding the limitations of this thesis, the research methodology, ethics, and sustainability. Prospective research for the future is envisaged at the end of this chapter.

Chapter 6 Conclusion recapitulates the findings and contributions of this thesis.

Chapter 2

Theoretical Framework

This chapter describes the theoretical framework of this thesis. First, this chapter discusses robotic assembly, the background information and key concepts relevant to the context of this research, concerning assembly automation and robotic automation. Then, this chapter introduces robotic visual perception by elaborating the background information and related research on robotic perception, computer vision, and artificial intelligence. Lastly, this chapter summarizes the state-of-the-art research on robotized wire harness assembly.

2.1 Robotic Assembly

Robotic assembly is an essential application of robotic automation in assembly toward assembly automation. It is also the context of the research in this thesis. Hence, this section introduces the background knowledge and specifies terms relevant to robotic assembly.

2.1.1 Assembly Automation

Automation

Automation, a concept that has undergone diverse definitions in various contexts, holds immense significance in the field of production engineering. Linguistically, *automation* is defined as "the action or process of introducing automatic equipment or devices into a manufacturing or other process or facility; (also) the fact of making something (as a system, device, etc.) automatic" in Oxford English Dictionary [46]. Practically for research and industrial applications, *automation* is defined as "the conversion of a procedure, a process, or equipment to an automatic operation without intervention by a human operator" by the International Institution for Production Engineering Research (CIRP) [47] and as "the creation and application of technology to monitor and control the production and delivery of products and services" by International Society of Automation[1]. These definitions of automation underscore its potential to liberate the workforce from specific production scenarios by significantly reducing the need for human control or intervention.

Automation research in production has been conducted regarding physical and cognitive automation [48]. While physical automation aims more to automate physical operations [49], [50], cognitive automation targets more to automate the cognitive processes that are currently performed by human operators [51]. This thesis is dedicated to exploring technical solutions that can revolutionize manual assembly operations in current production, thereby emphasizing the practical application of physical automation.

[1] `https://www.isa.org/about-isa/what-is-automation`

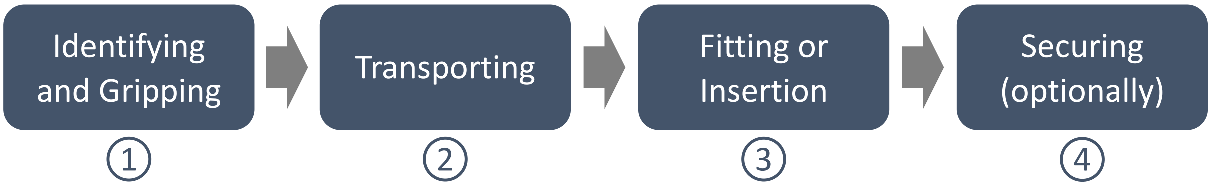


Figure 2.1: Sub-procedures consisted in assembly tasks, adapted from Lien [26].

Assembly

Assembly is a critical constituent part of production and has been researched from two major perspectives: 1) assembly as process and 2) assembly as product [52]. This thesis investigated problems in assembly regarding the former point of view, and particularly the final assembly in the automotive industry, where all sub-parts or units (e.g., engines, body frames, wire harnesses, glasses, and wheels) are fitted together to form the final product (automobiles) of production (the automotive production). In this scenario, as shown in Figure 2.1, the workflow of assembly tasks can be divided into several sub-procedures: 1) part identification and gripping; 2) part transportation to target assembly position; 3) part fitting or insertion; and 4) part securing (optionally) [26].

As the process where manufactured parts are fitted together into a complete product of any kind, the assembly can be performed manually [26], semi-automated [53], or fully automated [54]. The choice of assembly method is affected by multiple factors, such as the product design, the required production rate, the availability of labor, and the market life of the product [55].

Automation in Assembly

In assembly research, assembly automation describes introducing automatic machines to convert manual assembly operations into ones without needing human controls [54]. Assembly automation is desired and of benefit to both business (e.g., higher quality, more optimized resource allocation, and higher efficiency [2], [3]) and human operators (e.g., safer working space and better ergonomics [4], [56]). However, the scale of automation in assembly remains limited, especially in the final assembly [9]. From the perspective of assembly tasks, the complexity of assembly operations inhibits the automation of assembly [57]. From the perspective of technical solutions, assembly automation is constrained due to the inability of automatic systems to handle complex assembly tasks and the insufficient flexibility to handle product variants [12].

2.1.2 Robotic Automation

Robotics

Robotics is one of the subjects boosted by the research and application of automation [58]. Through research over decades, robotics has become an interdisciplinary subject studying the science and technology related to robots and similar automatic devices, including the design, construction, operation, and usage [59]. Already in the early stage of robotics research, Brady [24] had indicated the significance of concerning robotics as “the intelligent connection of perception to action”, which persists in the core of contemporary robotics research [60].

Industrial Robots

Industrial robots have been one of the core interests of robotics since the beginning of robotics research [61]. A robot, in general, can be seen as a physical agent that manipulates the physical world with equipped effectors based on the environmental information perceived through equipped sensors [62]. Notably, an *industrial robot* can be defined as "an automatically controlled, reprogrammable multipurpose manipulator programmable in three or more axes" [63]. In the context of production, "industrial robot" and "robot" are used interchangeably by convention [63], which is also adopted in this thesis hereinafter except where otherwise stated.

The first industrial robot is widely acknowledged to be the Unimate from Unimation, founded by George Devol and Joseph Engelberger in the late 1950s [64]. In 1961, the Unimate was first deployed industrially for unloading the finished castings in a General Motors plant in Trenton, New Jersey, the United States of America [64]. Since then, industrial robots have become significant in factory automation and radically changed the manufacturing industry [60]. Among other automation technologies, robots are widely adopted in the modern manufacturing industry due to their superiority in conducting repetitive and unergonomic tasks fast and precisely [7], [20]. Commonplace industrial application scenarios of industrial robots include spot welding, spray painting, part handling, packaging, and palletizing [63]. Industrial robots are also applied to automate assembly tasks, which, though highly favored [17], remains a small portion of the robotic automation application [10].

Parallel to the upsurge of industrial robot applications promoting industry automation, academia is deepening research to continuously improve robots' capabilities. While previous research on industrial robots focused more on the aspect of robotic action and mainly investigated kinematic calibration, motion planning, and control laws, the primary research interest has veered toward the aspect of intelligence to improve the flexibility and enhance the autonomy of industrial robots [61]. Siciliano and Khatib [60] projects that "the new generation of robots is expected to safely and dependably co-habitat with humans in homes, workplaces, and communities, providing support in services, entertainment, education, healthcare, manufacturing, and assistance". Such new-generation robots need to not only succeed in their actions but also be capable of perceiving the environment, learning, and reasoning their choices of actions [5].

Strengths and Weaknesses of Conventional Industrial Robots for Assembly

Compared to entirely manual operations, robotic assembly's better precision, repeatability, transparency, and comprehensibility can enable more rigorous, safer, and more ergonomic-friendly manufacturing with better quality and higher productivity [20]. The increase in their degrees of freedom and payload promotes industrial robots to take over tasks harmful to humans from human operators, e.g., operations in dirty and dangerous work environment [65], repetitive operations [5], demanding and tedious operations [66], and unergonomic operations [20].

However, conventional industrial robots are only superior in repetitive and familiar industrial configurations but brittle when the assembly process involves increasing product variants and/or requires more flexibility in unstructured environments [5]. There are also limitations to expanding the usage of conventional industrial robots in assembly, such as the complexity and flexibility of assembly [10], [67] and the safety of humans [68]. Typically based on scripted trajectory planning, conventional industrial robots can only accomplish simple tasks with monotonous operations in structured scenarios, where robot programmers must specify the positions and orientations of objects in advance [69]. However, with the increasing complexity of assembly tasks and the industrial paradigm shift toward mass customization, contemporary assembly systems handle increasingly more product variants [18]. Particular assembly tasks also involve manipulating non-rigid objects, whose positions and orientations are impossible to pre-define due to their deformation with almost infinite degrees of freedom [70]. The increasingly demanding production requirements demand industrial robots with higher flexibility potentials and a higher degree of autonomy [14]. Thus, this situation has steered the robotics research toward developing adaptive and intelligent systems [61].

2.2 Robotic Visual Perception

The manufacturing industry has witnessed the significant promotion of the application of automation and the adoption of industrial robots since the third industrial revolution [1]. Nevertheless, the contemporary manufacturing industry calls for intelligent industrial robots to fulfill emerging production requirements of flexible robotic automation [61]. A critical task is to improve the capability of robots to reach a higher degree of autonomy so that robots can handle unstructured tasks in an unstructured environment [60].

Russell and Norvig [25] defines an agent as “anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators”. An industrial robot able to adapt to variations can be considered an intelligent agent. It perceives the environment through equipped sensors, interprets the perceived information with intelligent systems, and then interacts with the surrounding environment via end effectors. Generally, an intelligent industrial robot can only select its actions at any time based on its embedded knowledge and the information perceived to date [25]. Robots can acquire knowledge through either hard coding by humans or learning based on self-perceived information [71].

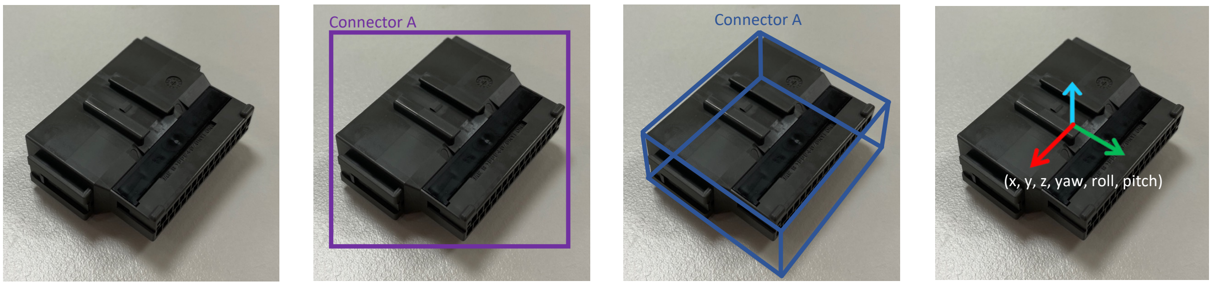
2.2.1 Robotic Perception

Perception is a premise of intelligent agents of any kind [25]. Russell and Norvig [62] defines *robotic perception* as “the process by which robots map sensor measurements into internal representations of the environment”. Regarding different types of sensory data input, robot systems can perceive the outer environment typically through visual, range, force/torque, and tactile perception [59]. Among approaches based on sensory data input, force/torque-based approaches are critical for robotic control at a low level, while recognition, measurement, and learning of robots at a higher level heavily rely on visual and range data-based perception [72].

Visual perception indicates the organization, identification, and interpretation of information from visual inputs, which has been extensively attractive in the academic field and industrial applications [73]. Specifically for robotic assembly, a robot needs to recognize *what* the object to be manipulated is and localize *where* the object is, among acquiring other physical properties of the object, so that the robot can reach, grasp, and manipulate the object to accomplish the assembly task. Vision is a fundamental source of information for object recognition and localization [5]. The vision system is one of the most common sensory systems integrated into robots to enable automatic operations [63].

2.2.2 Computer Vision

Visual perception is one of the significant topics studied within computer vision research [74]. Computer vision is also extensively related to automation, robotics, and manufacturing [45].



(a) Raw input (b) 2DOD (c) 3DOD (d) 6DOPE

Figure 2.2: Expected results of 2D object detection (2DOD), 3D object detection (3DOD), and 6D object pose estimation (6DOPE) regarding an RGB image of a connector.

Computer vision has a dual goal: 1) from the perspective of biological science, developing computational models of the human visual system; 2) from the perspective of engineering, developing autonomous systems capable of tasks performed by human visual system or even superior [75].

In alignment with the engineering perspective, this thesis investigates the challenges and solutions to developing robotic visual perception systems to extract useful information from visual inputs [76]. The robot system with a vision system will be able to convert the perceived visual information of a scene into a symbolic description [77]. With this description, the robot can understand the scene and decide the next operations [45]. As a thriving field, computer vision research encompasses diverse topics, where two-dimensional (2D) object detection, three-dimensional (3D) object detection, and six-degree-of-freedom (6D) object pose estimation are closely related to robotics and robotic assembly [72]. Figure 2.2 illustrates the expected results of these three tasks regarding a visual input.

2D Object Detection

In 2D space, object recognition implies image classification, object localization, and object detection. Image classification deals with classifying the principal object in an image, involving assigning a class label to an image [78], [79]. Object localization deals with locating object instance(s) of a given category in an image, involving drawing a bounding box around one or more objects in an image [80]. Object detection is a process of image classification and object localization [81]. As shown in Figure 2.2(b), object detection involves drawing a bounding box around each object instance of interest in the image (localization) and assigning each localized object instance a class label (classification) [82], i.e., recognizing what objects are where [83]. The year of 2014 marked a milestone in the research of 2D object detection with the advent of R-CNN [84]. It symbolized the division of research and breakthroughs into two historical periods: the traditional object detection period and the deep learning-based object detection period [85].

Traditional methods were designed based on handcrafted features and various feature descriptors [83]. The pipeline of traditional object detection methods generally comprises three steps: 1) informative region proposal, 2) feature extraction, and 3) classification and bounding box regression [82]. Traditional methods required the design of sophisticated feature representations and diverse speedup skills due to the lack of effective image representation at that time [85]. Classic traditional detectors include Viola–Jones detector [86], [87], Histogram of Oriented Gradients (HOG) detector [88], and deformable part model (DPM) [89]. However, traditional detection methods have significant flaws, e.g., slow speed, low accuracy, arduous manual feature engineering, and poor generalizability, which have gradually been replaced by deep learning-based methods [81].

Deep learning-based methods with the structure of deep convolutional neural networks (CNNs) dominate the latest research on 2D object detection [81]. Deep learning-based detectors can be classified into two major groups: two-stage detectors and one-stage detectors [85]. Motivated by the attentional mechanism of the human brain, two-stage detectors first scan the whole scenario coarsely and then focus on regions of interest (ROIs) to distinguish the object [83]. Quintessential two-stage detectors include the R-CNN family (R-CNN [84], [90], Fast R-CNN [91], and Faster R-CNN [92]), SPPNet [93], and FPN [94]. One-stage detectors were initiated to address the constraint of two-stage detectors on speed and computation [85]. Localization and classification are accomplished all at once through the backbone network. Typical one-stage detectors include YOLO [95] and its successors [96]–[99], SSD [100], RetinaNet [101], CornerNet [102], and CenterNet [103]. Recent research on Transformer models [104] also initiated the design of new detectors [105]. Typical Transformer-based detectors include DETR [106], Deformable DETR [107], DINO [108], and Mask DINO [109].

3D Object Detection

3D object detection is more challenging than 2D object detection. Different from 2D object detection, 3D object detection emphasizes the importance of recovering the amodal bounding box of the exact object instance, i.e., the minimum 3D bounding box enclosing the object of interest [110], as shown in Figure 2.2(c). Therefore, 3D object detection needs to obtain the size and direction of the object of interest in 3D space in addition to its position [111].

Previous research on 3D object detection explored diverse solutions regarding various types of input data, including RGB images [81], point clouds [112], RGB-D data [110], and multi-modal data [113]. Approaches based on RGB images can be divided into four categories: monocular-based, stereo-based, pseudo-LiDAR-based, and multi-view-based [81]. Approaches based on point clouds can be divided into two categories: region proposal-based and single-stage [112]. Approaches based on RGB-D data can be divided into two categories: 2D object detection-driven and data fusion-driven [110]. Approaches based on multi-modal data can be divided regarding the adopted data fusion method, e.g., early fusion, late fusion, and deep fusion [113].

6D Object Pose Estimation

Six-degree-of-freedom (6D) object pose estimation refers to determining the 6D pose of an object in 3D space. The 6D pose of an object is the combination of position, (x, y, z), and orientation, $(yaw, roll, pitch)$, of an object in 3D space [114]. Figure 2.2(d) illustrates an expected result of 6D pose estimation on a connector. Previous studies on 6D object pose estimation can be clustered into two primary settings: **instance-level 6D pose estimation** and **category-level 6D pose estimation**, regarding whether the computer-aided design (CAD) model of each object instance is a prerequisite [115].

Instance-level 6D pose estimation mainly addresses the pose estimation of objects whose 3D models are available [110]. Similar to research in object detection, research in instance-level 6D pose estimation can be divided into traditional methods and deep learning-based methods. Traditional methods address the problem based on CAD models [116] or 2D images synthesized from CAD models [117]. Regarding the modality of the input data, previous deep learning-based methods can be divided into three sub-groups: RGB-based, point cloud/depth-based, and RGB-D-based [115]. Research on RGB-based methods is extensive, thanks to the widespread use and affordability of RGB cameras. However, these RGB-based methods suffer from occlusion, illumination variations, objects without distinctive visual features, real-time performance, and generalizability [115]. Point cloud/depth-based methods process inputs in the format of point clouds or depth data acquired by 3D scanners or depth cameras. Point clouds or depth data include spatial information compared to RGB images, which facilitates the recovery of 3D information, especially for objects without distinctive texture features. However, this group of methods suffers from the expensive manual labeling of training data and heavy computing consumption, especially for point cloud-based methods. RGB-D-based methods consider RGB inputs and depth data jointly, which can promote the pose estimation performance in complex configurations with features extracted from both modalities of data [115]. However, it requires further investigation on efficient approaches for fusing features extracted from RGB data and depth data [115].

Category-level 6D pose estimation aims to achieve generalization to unseen instances when recovering the poses of objects [110]. Previous research efforts on category-level 6D pose estimation can also be divided into traditional methods and deep learning-based methods, in alignment with instance-level 6D pose estimation. Traditional methods address the problem based on images of natural objects but hardly remain in recent research efforts due to the time-consuming and challenging data collection [117]. Previous deep learning-based methods for category-level 6D pose estimation can be summarized into two major sub-groups: regression-based and prior-based [115]. Category-level pose estimation methods do not require accurate CAD models but are challenging due to the unavailability of ground truth data. Category-level pose estimation methods are also susceptible to the prominent appearance and/or shape variance across instances.

Computer Vision in Manufacturing

Computer vision techniques were already applied to industrial applications in the early 1970s [118] but remained scarcely commercialized in practical manufacturing until the 1990s due to the limitation of computing resource [119]. The research on industrial applications of computer vision techniques is also categorized as research in the field of **machine vision** [120] or industrial vision[2] [121]. While computer vision research is more methodology-oriented, machine vision research is more application-oriented and represents the particular implementation of computer vision for industrial purposes [120]. There are two major approaches to addressing industrial vision problems [122]–[124]. One is to address all problems with a general-purpose system, and the other is to design an ad-hoc system for each application scenario [121].

Throughout decades of development, computer vision techniques have become a vital booster of industrial manufacturing systems toward a higher level of informatization, digitalization, and intelligence [45]. In the manufacturing industry, the application of computer vision techniques can be classified regarding different criteria, e.g., application tasks [125] or stages of the product life cycle in the entire manufacturing process, including product design, modeling and simulation, planning and scheduling, production process, inspection and quality control, assembly, transportation, and disassembly [45].

As a critical stage in manufacturing, the assembly has drawn long-lasting attention to applying computer vision techniques [118]. Previous research mainly investigated applying computer vision techniques for automatic assembly, assembly quality control, and other assembly applications [45]. Researchers explored improving the performance of industrial robots in assembly tasks in unstructured environments using visual perception and learning techniques [27]. There are also research efforts on making industrial robots more adaptive to unknown scenarios using machine vision [126]. Nevertheless, machine

[2] Diverse industrial vision applications can be found on `https://www.cs.ubc.ca/~lowe/vision.html`

vision is mainly employed for quality-related purposes but is advocated to be expanded into assembly scenarios [127]. Promoting the adaptation of computer vision techniques in industrial applications remains challenging regarding computer vision algorithms, data, and benchmarks [45].

Particularly for assembly tasks, the accuracy of identifying the position and orientation of objects to be assembled is crucial to robotic grasping and the following manipulation operations. However, it is difficult to fulfill the demanding requirement with traditional computer vision techniques currently used in actual manufacturing systems [45]. On the other hand, it also remains challenging to identify objects' accurate position and orientation with existing deep learning-based methods [72]. The performance of visual recognition can also suffer from diverse problems in actual production environments, e.g., occlusions [128], illumination conditions [129], and the camera movement [130].

The dataset is essential for learning-based computer vision techniques [131] and scalable learning-based computer vision applications in manufacturing [132]. However, it is challenging to collect high-quality data in practical manufacturing scenarios and arduous to effectively preprocess and efficiently label the collected data [45].

Using benchmarks is critical to evaluating performance across different computer vision and robotic systems [133]. However, specific manufacturing cases require particular benchmarks for evaluating computer vision techniques in different scenarios [45].

There are also other concerns for the implementation of AI-driven computer vision systems in industry, e.g., the cost of implementing new systems in existing systems [45], industry's lack of trust in AI systems due to the lack of interpretability and explainability of the decision-making of AI [70] and the safety of human-robot collaboration [45].

2.2.3 Artificial Intelligence

The term *artificial intelligence* (AI) was first introduced in 1955 for proposing a workshop at Dartmouth College[3] to "proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" [134]. Throughout decades of research and development, AI has been evolving into a flourishing field encompassing diverse, active research topics and practical applications, ranging from general topics, e.g., learning and perception, to specific application scenarios, e.g., object detection, machine translation, weather forecasting, and robotics [22].

Machine Learning

One significant task in enabling machine intelligence is to equip machines with knowledge. Generally, machines acquire knowledge via either knowledge hard-coding by humans or self-learning [71]. Approaches based on hard-coded knowledge previously gained limited breakthroughs due to the inability to enumerate all scenarios or describe specific scenarios in formal languages [135]. This limitation suggested the significance of the machine's capability of self-learning. This capability is known as machine learning, with which machines observe data and extract patterns from the observed data [135].

Deep Learning

Deep learning is a broad family of techniques for machine learning and has significantly boosted the latest surge of public interest in AI [136]. The advent of deep learning radically

[3]https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth

reshaped the research and development in AI-relevant fields, e.g., computer vision, natural language processing, and robotics [137]. Deep learning enables computers to learn data representations with a hierarchy of abstraction via computational models consisting of multiple processing layers [71]. Adapting deep learning techniques can relieve the humans' burden of formally specifying all the knowledge required by computers in advance [71]. The genuine reasons for the success of deep learning remained obscured [137]. Nonetheless, deep learning-based approaches demonstrate their superiority over traditional approaches based on manually designed features, especially for tasks with high-dimensional input data, e.g., images, videos, and speech signals [136].

Convolutional Neural Network

Deep learning stemmed from early trials on mathematically modeling networks of neurons in the brain by McCulloch and Pitts [138], thus naming the networks trained by deep learning approaches neural networks [137]. Convolutional neural network (CNN) is "a specialized kind of neural network for processing data that has a known grid-like topology" [71]. CNN is characterized by using convolution operations instead of general matrix multiplication in at least one of their layers [139]. A typical CNN architecture usually consists of layers alternating between convolutional and pooling layers [140]. Since the early 1990s, CNN has achieved numerous practical successes and has been widely adopted in computer vision research, e.g., object detection, time series prediction, and human action recognition [141].

Artificial Intelligence in Manufacturing

Academia and industry have promisingly anticipated that the achievement in AI research will remarkably improve manufacturing systems in terms of productivity, quality, and profitability [142]. However, regardless of recent rapid advances in AI, previous research uncovered that "the application of AI technology in industry lags far behind the development of the AI technology" [143]. This gap promotes the research on AI applications, more specifically in industrial scenarios, also known as Industrial AI [143], aiming to address the problems and needs coming from industry. AI is applied or envisaged to be applied in diverse industrial application areas, e.g., supply chain management and production planning on the systems level, safe human-robot collaboration and robotic motion planning on the workstation level, and quality monitoring, tool wear prediction on the manufacturing process level [144]. For robotic assembly, AI has the potential to support different tasks, e.g., robotic perception, robotic motion, manipulation planning, and safety assurance [145].

2.3 Robotized Wire Harness Assembly

Wire harnesses can be theoretically generalized as deformable linear objects (DLOs) [70], [146], [147]. Wire harness assembly, thus, can be regarded as a specific task of DLO manipulation [148]. Challenges for automating wire harness assembly arise from product and production aspects. The development of robotic assembly of wire harnesses has attracted research over the years but has yet to succeed [36], [40]. This section elaborates on the research on robotic manipulation of DLO in general, the challenges for automating wire harness assembly, and the state-of-the-art research on robotic perception in the robotic assembly of wire harnesses.

2.3.1 Robotic Manipulation of Deformable Linear Objects

Objects manipulated by robots can be classified as rigid or non-rigid objects depending on shape-changing caused by an external physical force applied to objects [149]. Previous research also called non-rigid objects [70], [150] as deformable objects [149], [151], [152] or flexible materials [148]. Thus, in alignment with the existing research, this thesis uses these three terms interchangeably hereinafter except where otherwise stated.

Enabling robots to manipulate deformable objects can greatly benefit various application scenarios, e.g., flexible printed circuit board handling in manufacturing industry [153], tomato grasping in food industry [154], wound suturing in medical surgery [155], and garment manipulation in daily activity [156]. However, throughout decades of efforts in the robotics community, the study of non-rigid objects has yet to reach the compatible level of maturity of rigid object-centered research [149]. Notably, robotic manipulation of rigid objects primarily concerns the change of objects' pose (position and orientation) and the avoidance of collisions [157]. Differently, manipulation of non-rigid objects must further consider the shape-changing, which usually causes the changes in the objects' geometry and/or topology and lead to potential robotic failures [158]. Therefore, strategies designed for manipulating regular rigid objects cannot be adapted for deformable object manipulation directly [149].

Deformable objects can be more specifically classified into three groups: one-dimensional (or linear), two-dimensional, and three-dimensional (or volumetric) [148]. Further considering the physical properties, deformable two-dimensional objects can be separated into planar objects and cloth-like objects [149]. In line with the existing literature, wire harnesses can be categorized as deformable linear objects (DLOs). DLOs are also called deformable one-dimensional objects (DOOs) [159], [160]. Hence, wire harness assembly is a specific industrial application of DLO manipulation [148].

The DLO manipulation has also been a significant concern in industry over the years [70], [148], [149], e.g., the wire insertion in the electrical industry [161], [162] and the assembly of cables in the automotive industry [163], [164]. Various robotic tasks are involved in the robotic manipulation of DLOs, including modeling, perception, and manipulation [149]–[151]. Research has been conducted in investigating diverse deformable object models and integrating different sensors and AI into robots to endow robots with fast, accurate, and multi-modal perception capabilities [165] and adaptive modeling and control capabilities [151]. Regardless of the remarkable advancement of robotics, robotic manipulation of DLOs remains challenging in robotic flexible automation [11], [166], [167]. Specific challenges of accomplishing robotic DLO manipulation exist in object detection, deformation state estimation, object modeling, robotic motion planning, and robotic manipulation [149], [151], [152], [157], [167], [168].

Robotic perception is a prerequisite for accomplishing complex robotic manipulation tasks [165]. Particularly for robotic manipulation of DLOs, perceiving DLOs' physical properties, e.g., geometry, topology, deformation, and strain, before and during the robotic manipulation is required for modeling, motion planning, and manipulation planning [149], [152], [157]. Robotic perception for robotic manipulation of DLOs essentially involves perception based on independent or multi-modal sensing data, including visual, sound, force, tactile, and range data [70], [150]–[152]. Tactile perception is often adopted to obtain shape and contact information on the local level [152], while the global information of DLOs on a large scale, e.g., geometry, topology, and deformation, is often obtained via visual perception [169], [170].

2.3.2 Challenges for Automating Wire Harness Assembly

The automotive industry has implemented automation in assembly over the years to fulfill the increasingly demanding production requirements [14], [171]. While the body shop and the final assembly gather the majority of assembly operations, a remarkably higher level of automation has been achieved in the assembly of the body in white in the body shop than other assembly operations in the final assembly line [14]. Specifically, the installation of wire harnesses in the final assembly stage remains manual extensively and laborious to automate due to obstacles regarding the product and the production [36].

The customization and deformation of wire harnesses demand intelligent robots to handle flexible and agile automation tasks in the robotized assembly of wire harnesses. The automotive industry is consistently seeking solutions to robotize the overall or part of the assembly operations of wire harnesses in final assembly [36], [148]. However, no practical solution has been witnessed in actual production yet as equipment and technologies in the current production of automobiles are inadequate to accomplish the demanded flexible and agile automation tasks [7], [14], [54], [172]. Industrial robots deployed in current production are good at specialized tasks but weak at handling variants flexibly due to the lack of cognitive abilities [7]. Besides reducing the assembly complexity by simplifying the harness architecture [173], the robotization of wire harness assembly can be achieved by improving robots, i.e., making industrial robots more intelligent with more autonomy.

Industrial robots cannot handle variations due to limited perception and cognitive capabilities. For wire harness assembly, the deformability of wire harnesses further challenges the robotic perception [32], [34], [174]. This indicates that industrial robots need to have more autonomy and be more intelligent.

Challenges Regarding the Product

As a specific industrial application of DLO manipulation, the robotic assembly of wire harnesses inherits the challenges of robotic perception, modeling, and control in DLO manipulation. Robotizing wire harness assembly requires robotic systems to recognize the geometry and topology of wire harnesses, estimate the state of manipulation, and track the deformation so that they can model the wire harnesses and adapt their control strategies to handle the flexibility [174], [175]. Even if the robot achieves successful perception and modeling, the deformability of wire harnesses makes it complex to plan the robotic motion [175] and manipulation [157].

On the other hand, the robotic assembly of wire harnesses is more challenging than the generic robotic manipulation of DLOs. A bunch of wire harnesses is more complex than generic DLOs, considering its tree-like structure [176]. With multiple DLOs bound into bundles, it is necessary to address the interaction and constraints among different branches of wire harnesses while manipulating them. Besides the deformable cables, wire harnesses consist of rigid objects, e.g., connectors and clamps. This difference has also been identified in previous research, where wire harnesses were categorized more specifically as semi-deformable linear objects (SDLOs) [166], branched deformable linear objects (BDLOs) [32], [34], [177] or DLO networks (DLONs) [178]. Even with the mature research in robotic manipulation of rigid objects, the physical properties of some rigid wire harness components, e.g., the small sizes and the complex structures, exacerbate the arduous robotization of wire harness assembly [179].

Additionally, in mass customization, multiple variants of products are commonly produced on the same production line. This situation causes the wire harnesses installed onto each vehicle to be different. This increases the complexity of automation system design and challenges its adaptiveness and agility regarding different product variants.

Challenges Regarding the Production

To be deployed in actual production, technologies need to address challenges stemming from the actual production environments.

First, the proposed automation solution needs to be effective. However, Jiang, Nagaoka, Ishii *et al.* [180] indicated the challenge of the effectiveness of many proposals in actual production due to the required extremely tight position accuracy in assembly operations and the lack of precise contactless measurement to the state of the target wire in real-time. The actual production environment also challenges the effectiveness of proposals. The proposed system also needs to be reliable and robust in industrial environments. Extremely tight position accuracy in some assembly operations requires precise robotic perception of the object to be manipulated in real-time, which challenges the reliability of robotic perception capability in actual production [180]. The diverse and dynamic physical environments in actual production further challenge the robustness of the automation system. The automation systems also need to function reliably and efficiently in actual production to fulfill the demanded product rate and maintain the manufacturer's competitiveness.

Introducing new robotic systems brings challenges to safety and risk management. Physical equipment, such as steel fences and laser curtains, is typically required in industrial robotic applications to safeguard human operators [42], [43]. The growing applications of human-robot collaboration also require careful consideration of the safety aspect [44]. With new robotic systems introduced in actual production, a systematic re-design of the workspace and the human-robot interaction may be necessary, which poses challenges to safety and risk management within the existing system.

Robotic systems deployed in the final assembly stage need to deal with moving assembly lines. The final assembly lines in the automotive industry are typically non-stop, which requires robots to move in synchronization with the moving assembly line while executing assembly operations [10]. The mobility of robots and the synchronization between robots and assembly lines pose a challenge to the development of robotic assembly solutions.

Additionally, developing a universal solution is challenging due to the diverse production requirements across different productions and sectors, even within the same industry. In the automotive industry, for example, the production requirements of passenger, heavy, and special vehicles are different in terms of the physical production environments, required production quality, and production rate. These different production requirements set different criteria for developing automation solutions in terms of effectiveness, efficiency, reliability, and robustness. They also demand heterogeneous solutions, which increases the workload of automation solution development.

2.3.3 Robotic Perception in Robotic Assembly of Wire Harnesses

Robotic perception is pivotal to the robotization of wire harness assembly. Preliminary to accomplishing robotized wire harness assembly, robots demand advanced perception capability to recognize different wire harnesses and obtain their position and movement. Previous research efforts studied different types of perception based on diverse sensing data to perceive the physical properties of wire harnesses [36], [40], [148]. Visual perception is a critical contactless measurement to perceive the global information of wire harnesses on a large scale [169], [170].

Vision is instrumental for object localization, classification, and tracking [5]. The vast amount of information embedded in visual input [181], [182], the significance of computer vision on the perception for robotic manipulation [183], and numerous applications in different manufacturing scenarios [45] demonstrate the potential of applying computer vision techniques to facilitate robotized wire harness assembly. Previous research suggested the promising performance of vision-based approaches in robotic manipulation of DLOs [169], [170], [184] However, the application of vision-based robotized assembly of wire harnesses in actual production has not succeeded yet [45], [185], [186]. The vision-based robotization of wire harness assembly has drawn enduring attention and research effort [163], [178], but the task remains challenging to accomplish in actual production [34].

Besides, previous research has explored obtaining the physical properties and contact information of wire harnesses on the local level based on tactile [180], [187] and sound data [188]. Though also critical, this part of the research is out of the scope of this thesis.

Chapter 3

Research Approach

This chapter elaborates on the design of the research approach of this thesis. The philosophical worldview of the author of this thesis is analyzed first, followed by the research design and methods adapted in this thesis. The method to guarantee the research quality is elaborated in the end.

3.1 Philosophical Worldview

Though concealed mainly in research, philosophical worldviews influence the research practice and need to be identified [189], [190]. Representing researchers' fundamental beliefs about the nature of knowledge, reality, and human behavior, philosophical worldviews guide researchers' actions in determining research approaches and conducting studies [190]–[192]. Revealing the espoused philosophical worldviews can not only facilitate researchers elaborating the reason behind choices on specific research approaches but also help readers better interpret the research with a clearer mind on the biases and the researcher's particular stance [190]. Multiple factors contribute to developing individuals' worldviews, e.g., externally, discipline orientations, research communities, advisors, mentors, and internally, personal education background and experiences in culture and research [193].

The education and research experiences in electrical engineering, computer science, artificial intelligence, and computer vision formulated the author of this thesis's gravitation on the empirical postpositivist worldview. Researchers with positivist worldviews deem the existence of the absolute truth of knowledge [194]. Positivists aim to determine the connection between cause and effect [195]. They coincide with empiricists when believing sense experience is indispensable with the inquiry of knowledge [196]. The postpositivist worldview complements the positivist worldview by stressing the reflection upon the potential personal bias of a researcher's claim of knowledge [197]. With the empirical postpositivist worldview, the author of this thesis focused more on studying problems related to the identification and evaluation of causes of particular effects through empirical observation [198]. Specifically, the author initiated the research in this thesis from the hypothesis that applying computer vision techniques will enable robotic visual perception on industrial robots for assembly tasks. The author needed to identify and assess, based on empirical evidence, the causes that may influence the effectiveness of applying computer vision techniques.

Nonetheless, researchers are seldom limited to one worldview and may involve different worldviews in parts of the research regarding the discipline orientations and research communities [190]. Before beginning the research, the author of this thesis realized the necessity of identifying remaining research problems and specifying research objectives based on an understanding of the state of the art of research. Constructivists aim to forge a theory inductively based on individual participants' view of a specific scenario being studied [199]. Thus, a constructivist worldview was also possessed to gain such knowledge from scientific literature.

3.2 Research Design

Within research approaches, research designs indicate specific directions for conducting procedures of inquiries [190]. The specific research design in a study is also a guideline for executing specific research methods to collect, analyze, and interpret data [200]. As elaborated in Section 3.1, the research in this thesis intended to identify problems, define objectives, and investigate technical solutions in sequence. Therefore, the author of this thesis recognized design science research methodology [201], [202] as the guideline for the high-level design of this research. This research also adopted a multiple-method design [203] encompassing two studies. The first study was a qualitative literature study with a qualitative descriptive design, mainly aiming to identify problems and define objectives. The second study was a quantitative experimental study with a quantitative experimental design, mainly aiming to investigate technical solutions.

Design Science Research Methodology

As defined in Hevner and Chatterjee [202], design science research (DSR) is a research paradigm where researchers, as designers, contribute scientific knowledge by creating innovative artifacts that are both useful and fundamental to understanding and addressing human problems. An artifact can be anything that can contribute to the transformation from the current state to the desired one, e.g., constructs, models, methods, and instantiations [204], [205]. Peffers, Tuunanen, Rothenberger *et al.* [201] proposed the design science research methodology (DSRM) comprising the following six steps:

1. Problem identification and motivation (Defining a problem and justifying the need)
2. Objective definition (Defining objectives for a solution)
3. Design and development (Designing and developing a solution)
4. Demonstration (Demonstrating the effectiveness of the solution)
5. Evaluation (Evaluating the solution)
6. Communication (Disseminating the findings to relevant audiences)

The framework of design science research methodology proposed by Peffers, Tuunanen, Rothenberger *et al.* [201] inspired the research design in this thesis, as delineated in Figure 3.1. This thesis initiated the research with a study (study A) focusing on understanding the current state of research and the challenges of enabling robotic visual perception on industrial robots for assembly tasks. This study was closely related to problem identification and motivation, as well as defining the objectives of a solution. Then, the research continued with a study (study B) exploring potential technical solutions for enabling robotic visual perception on industrial robots for assembly tasks. This study was to design and develop computer vision-based solutions to address specific challenges and objectives identified in the first study. Quantitative experiments were needed to demonstrate the developed solutions' effectiveness and evaluate the developed solutions' performance. Lastly, the research findings of each study were disseminated through publications (Paper 1, Paper 2, and Paper 3 appended to this thesis).

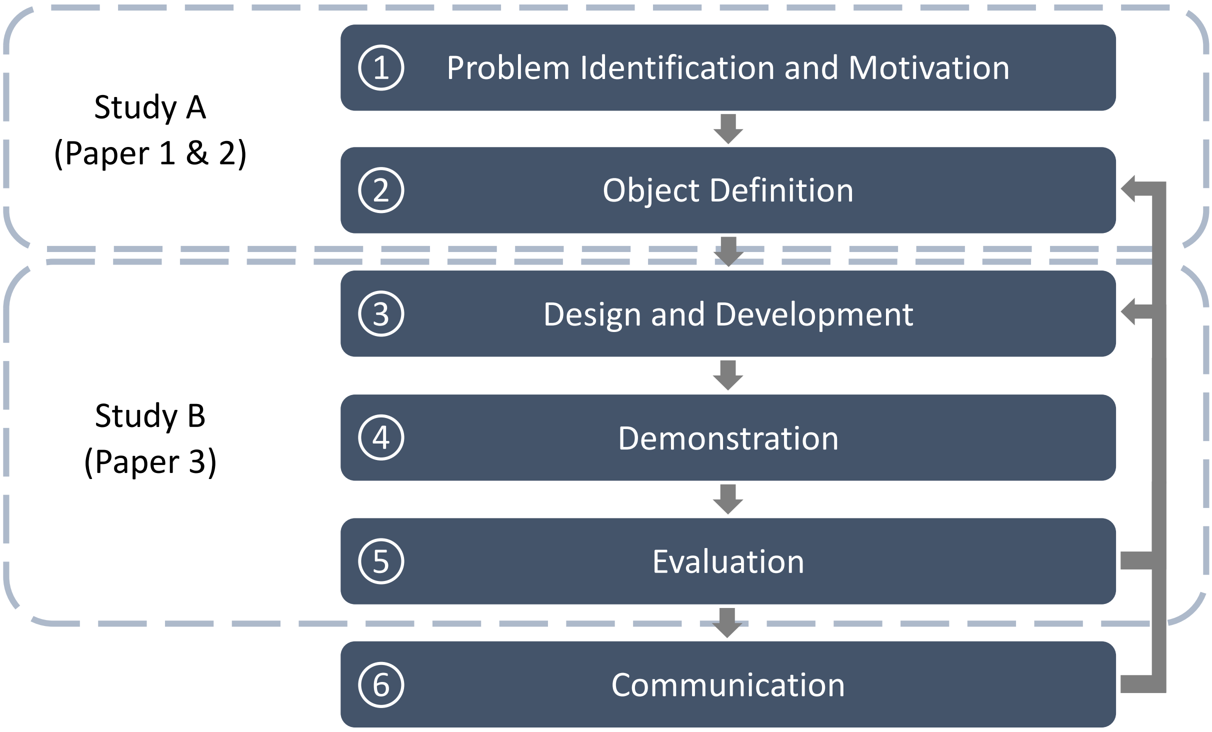


Figure 3.1: The framework of DSRM adapted from Peffers, Tuunanen, Rothenberger *et al.* [201]. The research in this thesis comprised two studies. Study A was designed to identify and motivate the problem of robotic visual perception in the robotic assembly of wire harnesses and define the objective for a vision-based solution. Study A led to the appended Paper 1 and Paper 2. Study B was designed to design and develop a technical solution, demonstrate its effectiveness, and evaluate its performance. Study B led to the appended Paper 3.

Multiple-Method Design

As interpreted in Morse [203], research with a multiple-method design consists of multiple self-contained and complete studies using different research methods to "address the same research question or different parts of the same research question or programmatic goal". This research adopted a multiple-method design, including a literature study (study A) following a qualitative descriptive design and an experimental study (study B) following a quantitative experimental design.

First, a qualitative descriptive design was adopted for the literature study to analyze the current state of research and clarify the prospective research directions. This literature study aimed to address the RQ1 (on challenges) by recognizing the challenges for developing technical solutions based on the analysis of the current state of research and to contribute to answering the RQ2 (on technical solutions) by discussing opportunities for further research and defining objectives for potential technical solutions. The research findings of this literature study were included in Paper 1 and Paper 2.

Then, a quantitative experimental study was conducted to assess the performance of designed technical solutions in addressing the challenges and objectives recognized in the previous literature study. Thus, this experimental study contributed to addressing the RQ2 (on technical solutions) by providing insights on potential solutions supported by quantitative experimental results. The potential positive results would demonstrate the effectiveness of the designed solutions. The potential negative results would, on the other hand, reveal constraints on the performance of the designed solutions, which would also contribute to addressing the RQ1 (on challenges). Considering the derivation of

Figure 3.2: A framework of positivist research design, adapted from Williamson, Burstein and McKemmish [195].

postpositivism from positivism, a framework of positivist research design [195], as shown in Figure 3.2, was referred to specify the stages of this experimental study. The research findings of this literature study were included in Paper 3.

3.3 Research Methods

Various research methods were implemented in each studies leading to the appended papers, as outlined in Table 3.1. The summary of the appended papers stemmed from the studies, as well as their contributions, will be elaborated in Chapter 4. Nonetheless, this section briefly summarizes the research methods implemented in each study to provide readers a quick reference.

The literature study was designed first to acknowledge the state-of-the-art research on enabling robotic visual perception of industrial robots for assembly tasks, particularly for wire harness assembly in the final assembly of automobiles. Based on the knowledge of the current state of research, this study intended to identify and motivate the problem of interest, i.e., to identify and motivate existing challenges demanding further inquiries. With challenges identified, this study would explore potential opportunities to enable robotic visual perception on industrial robots for the robotic assembly of wire harnesses in the final assembly of automobiles. With a constructivist worldview, a literature study was selected in this study through a qualitative approach with a descriptive design. The literature study was conducted following a systematic literature review methodology, considering systematic literature review as an instrumental methodology for comprehensively understanding the state of the art of a subject and identifying the gaps requiring future research [211]–[213]. A review protocol, including literature search and selection strategies, was determined first by a group of researchers. Then, literature data was collected by searching literature databases using a pre-defined search string. The collected literature was further selected with the consent of multiple researchers regarding pre-defined criteria. Additionally, the strategy of "snowballing" [209], including reference tracking and citation tracking, was conducted on the selected articles to identify potentially missing studies in the literature searching process. With the identified literature, this literature study adopted text coding based on a pre-defined scheme, followed by theme and pattern interpretation. More detailed research methodology of this study can be found in Section 4.1 and Section 4.2, as well as Paper 1 and Paper 2.

Based on the knowledge acquired from the literature study, an experimental study was designed to examine vision-based object detection solutions that may contribute to enabling robotic perception on industrial robots for assembly tasks. A quantitative

Table 3.1: The research design, research methods, and measures to ensure research quality of each appended paper.

Paper	Research design	Research method	Research quality measures
1	Qualitative approach Descriptive design Literature study	Systematic literature review - Database searching using a pre-defined string - Literature selection based on pre-defined criteria - Text coding using a pre-defined scheme - Inductive reasoning	Adopting consistent methods DARE criteria [206] Investigator triangulation [207] Peer debriefing [208]
2	Qualitative approach Descriptive design Literature study	Systematic literature review - Database searching using a pre-defined string - Literature selection based on pre-defined criteria - "Snowballing" [209] on selected literature - Text coding using a pre-defined scheme - Inductive reasoning	Adopting consistent methods DARE criteria [206] Investigator triangulation [207] Peer debriefing [208] Expert review [210]
3	Quantitative approach Experimental design Experimental study	Empirical measurement Statistical analysis Deductive reasoning	Adopting consistent methods Stratified sampling Statistical metrics Peer debriefing [208]

approach with an empirical postpositivist worldview was selected for this experimental study. Though positive that computer vision techniques are adequate, the performance of technical solutions needs to be examined and analyzed empirically [121]. Besides verifying the effectiveness of technical solutions based on quantitative evaluation, this study intended to discuss why the solutions failed on specific samples. A quantitative experimental study was selected to accomplish this objective. In this quantitative experimental study, numeric data was collected and analyzed based on statistical metrics, such as the rate of precision of detection. More detailed research methodology of this study can be found in Section 4.3 as well as Paper 3.

3.4 Research Quality

Validity and reliability are two significant criteria for research quality evaluation [193], [214]–[217]. Validity, including internal validity and external validity, suggests to what extent the study findings represent the truth among similar population outside the study [193], [211], [217]. Internal validity is the cornerstone of external validity and is defined as the extent to which systematic errors can be prevented in the design and conduct of the research, while external validity reflects the generalizability and applicability of the research outcomes outside the study [211], [217]. Reliability indicates the extent of consistency of the research approach in a study with the one in other studies by other researchers [218]. As shown in Table 3.1, each study adopted different methods to guarantee the quality of the research.

Qualitative research should concern both validity and reliability [193]. Researchers should adopt the research approach consistent across different researchers and projects to guarantee the reliability of the qualitative study and examine the accuracy of the research findings through specific procedures to ensure the validity [218]. A methodology for planning and conducting systematic literature reviews was suggested in Kitchenham [211], which has been adopted in various systematic literature reviews in computer science and engineering [219]–[227]. Thus, the systematic literature review in the qualitative literature study of this thesis was conducted following the methodology suggested in Kitchenham [211] to strengthen the reliability of the qualitative literature study. Besides, various methods were considered to reinforce the validity of the literature study. Funded by the Department of Health and the National Institute for Health Research of the United Kingdom, the Database of Abstracts of Reviews of Effects (DARE) and the NHS Economic Evaluation Database provide access to over 35000 systematic reviews in the field of health and social care interventions, whose quality was assessed based on publicly available criteria (shortened as "DARE criteria" hereinafter in this thesis) [206]. DARE criteria [206] qualify a systematic review considering[1]: 1) whether inclusion/exclusion criteria were reported; 2) whether the search was adequate; 3) whether the included studies were synthesized; 4) whether the quality of the included studies was assessed; and 5) whether sufficient details about the individual included studies were reported. According to DARE criteria [206], a systematic review is qualified only if it fulfills the first three and at least one of the fourth and fifth criteria. Following Kitchenham, Pearl Brereton, Budgen *et al.* [228] and Saleem, Khan, Zafar *et al.* [229], this thesis adopts DARE criteria [206] to evaluate the quality of the systematic literature review in the qualitative study. Besides, to reduce the subjective bias on data collection, analysis, and interpretation in the qualitative literature study, investigator triangulation [207] was adopted to strengthen the impartiality and mitigate the personal bias on the design of the review protocol and the judgment on literature selection and interpretation. Peer debriefing is also a helpful technique to strengthen the research quality by improving the study with feedback from colleagues and other researchers who are familiar with the research topic [208]. Multiple researchers were involved in conducting this qualitative literature study and reviewing different parts of the study. Additionally, expert review [210] were also adopted in the systematic literature review. Specifically, experts in relevant subjects from academia and industry (some of them as co-authors) were involved in calibrating the research methods and cross-validate the findings and interpretation.

For quantitative research, there are potential threats to the research validity embedded in, for example, for internal validity, experimental procedures, treatments, or participants' experiences, and for external validity, the generalization of research findings to the outside of the study [193], [230]. The quantitative experimental study in this thesis adopted the assessment suggested by Hammersley [230] to evaluate the reliability and validity of the quantitative experimental study concerning three aspects of the process of research: 1) whether measurement procedures were reliable and valid; 2) whether the findings can be generalized to a larger populations; and 3) whether variables were controlled effectively and sufficiently. In the quantitative experimental study in this research, the computer was the measure to collect data, which, with an experiment plan designed referring to relevant studies, strengthened the reliability and validity of the measurement. Then, stratified sampling was planned to separate the dataset regarding the distribution of data and the ratio among different data samples to ensure the generalizability of the treatment. Further, to guarantee the quality of the control of variables, the research intended to allocate sampled data randomly to treatment and control groups. In addition, statistical analysis was implemented to mitigate potential researchers' subjective bias in evaluating the treatment's performance. This study also adopted peer debriefing [208] to enhance the quality of this study.

[1]https://www.crd.york.ac.uk/CRDWeb/

Chapter 4

Summary of Appended Papers

This chapter briefly summarizes the three appended papers, including the core problem, the research methodology, and the contribution of each paper. This chapter also summarizes each appended paper's contribution to the research questions inquired in this thesis.

4.1 Paper 1

Title: Overview of Computer Vision Techniques in Robotized Wire Harness Assembly: Current State and Future Opportunities

Problem

Improving ergonomics and optimizing resource utilization while improving the assembly quality and assuring safety is desired for the installation of wire harnesses in the final assembly of automobiles [33], [176], [231]. Robotic assembly is a significant enabler and facilitator for achieving this goal, considering its superiority in replicability, transparency, and explainability over manual operations [7]. However, robotizing the assembly of wire harnesses remains laborious in actual production [129], [179]. The high degree of customization on wire harnesses and their deformation exacerbate the complexity of the assembly task, requiring robots to perceive and react to the surrounding environment and manipulate the object adaptively.

Vision is a fundamental source of information for object localization and recognition [5]. Previous research explored vision-based robotized assembly in different sectors by enabling the adaptive robotic visual perception [45], [185], [232]–[234]. However, the automotive industry has yet to identify any practical solution to robotize the assembly of wire harnesses in actual production. Therefore, further research is needed to understand the challenges of enabling robotic visual perception for robotizing wire harness assembly. Moreover, future research opportunities should also be identified to promote technical solutions for vision-based robotic assembly of wire harnesses.

Methodology

This paper's primary objective was to provide a comprehensive overview of the existing research on vision-based robotized wire harness assembly. Additionally, it aimed to identify crucial opportunities for future research in enabling visual perception of industrial robots to assemble wire harnesses. A qualitative literature study with a constructivist worldview was conducted to achieve these objectives.

The study initiated an inquiry on the Scopus database. The search string was TITLE-ABS-KEY((wir* OR cabl*) AND (harness* OR bundl*) AND assembl*). Then, three researchers examined the search results thoroughly and selected articles focused on vision-based robotized wire harness assembly in final assembly. The study only included articles in English for the analysis. Besides, the study excluded secondary studies, i.e., review articles and conference reviews. The selected articles were then grouped, considering multiple attributes of each study, such as the task of the operation, the object of interest, the type and location of vision systems, and the number of cameras. Lastly, the grouped articles were analyzed and interpreted through inductive reasoning to understand the current state of research and identify future research directions.

Contribution

This paper provides an overview of the existing research on vision-based robotized wire harness assembly. Table 1 and Table 2 in the appended Paper 1 summarized existing studies on vision-based robotic manipulation of wire harness components and visual machine perception of wire harness structures, respectively. This paper also discussed future research opportunities toward a more practical vision-based robotized wire harness assembly, including:

- Investigating the use of learning-based computer vision algorithms
- Evaluating proposed vision systems in actual production scenarios regarding the practicality and reliability
- Exploring new product designs of wire harnesses to facilitate robotic visual perception

4.2 Paper 2

Title: A Systematic Literature Review of Computer Vision Applications in Robotized Wire Harness Assembly

Problem

This article extended the study in the appended Paper 1 to summarize the state-of-the-art research and systematically discuss the challenges and future research directions.

Methodology

A systematic literature review was conducted in this study to answer the following three research questions:

1. What computer vision-based solutions have been proposed for robotized wire harness assembly?
2. What are the challenges for computer vision applications in robotized wire harness assembly?
3. What are the required future research activities and fields for developing more efficient and practical computer vision-based robotized wire harness assembly?

This study followed the methodology for planning and conducting a systematic literature review suggested by Kitchenham [211]. Following the methodology of investigator triangulation [207], three co-authors of this article collaborated continuously through all aspects of this study to ensure the quality of different stages of this literature study.

A review protocol was developed first to ensure a systematic and reproducible review method, as shown in Table 1 in the appended Paper 2. The following string was defined for the literature search within the field of *Article title*, *Abstract*, *Keywords* on Scopus on September 6, 2023: (wir* OR cabl*) AND (harness* OR bundl*) AND assembl*. The subject area was limited to *Engineering*, *Computer Science*, *Decision Sciences*, *Multidisciplinary*, and *Business, Management and Accounting*. The language of the article was limited to English. Finally, the literature search returned 662 articles for literature selection.

Next, three researchers conducted a two-step screening jointly regarding the inclusion and exclusion criteria shown in Table 1 in the appended Paper 2 to select the qualified literature for data synthesis and analysis. The article selection process was reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [235], as shown in Figure 4 in the appended Paper 2. The first screening was based on the title and abstract of each article, which selected 22 articles for the second screening. The second screening was based on the full text of each article, which further sifted out 13 articles. Moreover, "snowballing" [209], including *reference tracking* and *citation tracking*, was implemented on the selected 13 articles for analysis to identify other relevant articles missed in the original search, which returned 2 more articles. Therefore, 15 peer-reviewed scientific articles were included for further data analysis and interpretation.

The selected articles were then grouped, considering multiple attributes of each study, such as the task of the operation, the object of interest, the type and location of vision systems, and the number of cameras. Lastly, the grouped articles were analyzed and interpreted through inductive reasoning to understand the current state of research and identify future research directions.

In addition, this systematic literature review adopted DARE criteria [206] to evaluate the research quality.

Contribution

This study reviewed the state-of-the-art research on vision-based robotized wire harness assembly. This systematic literature review identified 15 relevant studies that discussed vision-based robotized wire harness assembly. Table 2 in the appended Paper 2 summarizes the contribution of each relevant study from the perspective of computer vision applications.

The identified 15 studies indicated the existing research on enabling robotic visual perception on different levels of the constituent structure of wire harnesses to facilitate the robotization of wire harness assembly [35], [129], [166], [174], [175], [179], [180], [186], [236]–[242].

Regarding the components of wire harnesses the study focused on, the identified 15 studies were categorized into four groups (section 4.1 in Paper 2):

- Four studies on clamps [129], [174], [180], [237]
- Seven studies on connectors [166], [179], [236], [238]–[241]
- Three studies on cables [175], [186], [242]
- One study on wire harness bags [35]

Regarding the sub-tasks of the assembly, the identified 15 studies were categorized into three groups:

- Robotic manipulation of wire harness components [35], [129], [166], [174], [179], [180], [236]–[241]

- Monitoring sub-processes of the assembly [175], [186], [236], [238], [240], [242]
- Fault detection during the assembly [236]

The proposed vision systems in the identified 15 studies also contributed to different operations in the current installation of wire harnesses in the final assembly of electric passenger vehicles (section 4.2 in Paper 2), including preparation [35], untanglement [175], routing [32], [34], [129], [174], [186], [237], [242], and assembly [129], [166], [174], [179], [180], [236]–[241]. Nevertheless, the automotive industry had yet to identify adequate vision-based solutions to robotize the assembly of wire harnesses in the final assembly of automobiles.

The study in Paper 2 identified challenges for computer vision applications in robotized wire harness assembly, including:

- Achieving compatible robustness compared with the human vision system, especially considering the demanding production rate and intricate production environments
- Accomplishing visual recognition based on intrinsic physical properties of different components of wire harnesses

The study in Paper 2 identified future research opportunities for introducing computer vision applications in robotized wire harness assembly more efficiently and effectively, including:

- Adapting learning-based visual recognition algorithms to exploit intrinsic features and multi-modality data of wire harnesses
- Investigating the adaptation of vision-based solutions proposed for robotized manufacture of wire harnesses
- Assessing vision-based solutions under practical production conditions in terms of practicality, robustness, reliability, and sustainability
- Considering vision-based HRC and exploring solutions to address different assembly operations
- Exploring new product designs to facilitate visual recognition

4.3 Paper 3

Title: Deep Learning-Based Connector Detection for Robotized Assembly of Automotive Wire Harnesses

Problem

It is challenging to design and implement rule-based connector detection as the manual feature engineering can be unwieldy due to the significantly various and complex structures of wire harness connectors. Extensive experimental results have demonstrated the superiority of learning-based object detection over the rule-based methods, especially in this task scenario [85], [136]. However, as identified in Paper 1 and Paper 2, there is a lack of research on deep learning-based solutions for recognizing wire harness components, e.g., connectors, in previous studies.

Paper 3 aimed to verify the effectiveness of deep learning-based object detection on wire harness components. Wire harness connectors were focused in particular, considering connectors as essential components for connecting wire harnesses and transmitting signals and power. Datasets are essential to training and evaluating learning-based object detection [243]–[246]. However, no publicly available dataset fulfilled the need of the study. Thus, creating a dataset of connectors was fundamental before investigating learning-based connector detection for robotized wire harness assembly in this study.

Methodology

The objectives of this study were to verify the effectiveness of deep learning-based object detection on wire harness connectors and identify constraints on its effectiveness. With a postpositivist worldview, a quantitative experimental study was conducted to achieve these objectives.

Given the absence of a publicly available connector dataset, this study first created a dataset comprising 360 images of single and multiple connectors. This was achieved by utilizing 20 different automotive wire harness connectors, a process detailed in Fig. 2 in the appended Paper 3. This study captured 60 images of mixed connectors and 300 images of each connector. In particular, this study captured 6 profile images for each connector, covering the front, back, left, right, top, and bottom views. Additionally, 9 images were captured from other random views of each connector, ensuring a comprehensive dataset. Fig. 3, Fig. 4, and Fig. 5 in the appended Paper 3 present example images in the collected dataset. The image annotation procedure of the collected connector dataset followed the methodology implemented in the PASCAL visual object classes (VOC) challenge 2007 [243]. The number of annotated object instances in the collected connector dataset is presented in TABLE I and Fig. 6 in the appended Paper 3.

Then, this study trained and evaluated a two-stage object detector, Faster R-CNN [92], [247], and a one-stage object detector, YOLOv5 [248], to investigate the effectiveness of deep learning-based object detection on connector detection using the collected dataset. Data augmentation methods were implemented to inflate the created dataset with artificially generated images. The experimental results were analyzed and interpreted based on statistical metrics, including the rate of precision and mean Average Precision (mAP), as reported in TABLE II and TABLE III in the appended Paper 3.

Contribution

This study investigated deep learning-based object detection on automotive wire harness connectors to facilitate the robotized wire harness assembly. A dataset of automotive wire harness connectors was collected for training and evaluation of a two-stage object detector, Faster R-CNN [92], [247], and a one-stage object detector, YOLOv5 [248], respectively. The experiment results verified the effectiveness of deep learning-based connector detection for automotive wire harness assembly.

Besides, this study indicated two potential hindrances to learning-based visual recognition:

- Insufficient amount of training data
- Confusing features due to product designs of connectors

The experiment results encouraged future research on collecting a benchmark dataset for better training and more consistent and rigorous evaluation across various object detectors. The results also encouraged three approaches to promote the detection performance:

- Designing new object detectors to extract nuance features on objects
- Conducting multi-view or video-based detection to extract more distinguishable features
- Optimizing the product design, especially the appearance, to facilitate the vision-based object detection

4.4 Contributions of Appended Papers

This section amalgamates each appended paper's contributions to this thesis's research questions formulated in Chapter 1. As shown in Table 4.1, all three appended papers contribute to both research questions of this thesis to different extents.

Table 4.1: Summary of each appended paper's contribution to each research question.

Paper	Contribution to RQ1 (Challenge)	Contribution to RQ2 (Solution)
1	**Minor contribution** 1) Visually recognizing and tracking objects of interest without using additional artificial fiducial markers 2) Obtaining spatial information in 3D space visually 3) Ensuring the practicality and reliability of vision systems in actual production	**Major contribution** 1) Adopting learning-based approaches 2) Evaluating vision systems under practical production conditions 3) Exploring new product designs that can facilitate visual recognition
2	**Major contribution** 1) Visual recognition exploiting intrinsic features of objects of interest instead of additional artificial fiducial markers 2) Recognizing and tracking the structure and topology of deformable linear objects 3) Obtaining spatial information in 3D space for robotic operations 4) Guaranteeing the practicality, reliability, robustness, and sustainability in actual production	**Major contribution** 1) Adopting learning-based approaches 2) Learning from intrinsic features and multi-modality data 3) Collecting benchmark dataset for training and evaluating learning-based solutions 4) Evaluating vision systems under practical production conditions 5) Developing vision-based HRC 6) Investigating new product designs to facilitate visual recognition
3	**Minor contribution** 1) High-precision position and orientation acquisition 2) Visual recognition addressing occluded features and high-similarity features	**Major contribution** 1) Adopting learning-based approaches 2) Collecting benchmark dataset for training and evaluating vision-based solutions 3) Investigating multi-view visual recognition 4) Exploring new product designs to facilitate visual recognition

Chapter 5

Discussion

This chapter discusses the main research findings toward answering the research questions of this thesis and analyzes this thesis's contribution to academia and industry. This chapter also reflects on the limitations, the research quality, and the aspects of ethics and sustainability of the research. Lastly, this chapter envisions future research.

5.1 Answers to Research Questions

5.1.1 The Answer to RQ1

RQ1: What are the challenges of enabling robotic visual perception for assembly tasks?

The Answer in a Nutshell

On the level of objects, robotic visual perception needs to address the challenge of visually recognizing objects of interest by exploiting their intrinsic features instead of being assisted by affixed artificial fiducial markers. It is also challenging to visually recognize and track non-rigid objects' structure and topology during robotic manipulation. Moreover, there are challenges for robotic visual perception caused by product designs, such as small sizes, complex structures, and highly similar appearances of objects of interest. The need for precise positions and orientations of objects of interest in 3D space is not just a requirement but a constant necessity for the success of robotic operations. However, this task remains a challenge for up-to-date computer vision techniques. Translating theoretical results into practical industrial applications is a task that demands meticulous consideration of physical production environments and practical business requirements. This consideration is crucial to ensure technology's efficient, effective, and safe application. This requirement poses challenges in enabling robotic visual perception, which would enhance industrial robots' autonomy and make them adequate for demanding flexible automation tasks.

The Implication of the Answer

In essence, enabling the visual perception capability of industrial robots for robotizing assembly operations is to make industrial robots capable of recognizing objects of interest visually. Given the objects recognized, industrial robots can recognize the task and adapt their actions to accomplish required assembly tasks. Hence, recognizing objects of interest first is of utmost importance. Beyond recognizing objects, industrial robots demand precise positions and orientations of objects before accomplishing required sub-tasks, such as reaching, grasping, and manipulating. Additionally, as a practical application in final assembly, physical production environments and practical business issues should be considered to guarantee an efficient, effective, and safe implementation of technologies in actual production.

On the object level, industrial robots handle both rigid and non-rigid objects in final assembly. Specifically, as elaborated in section 2.3, non-rigid objects involved in the robotized wire harness assembly are branched deformable linear objects, i.e., the part of wire harness cables. Clamps for fixing wire harnesses and connectors for connecting wires are examples of rigid objects involved in the robotized assembly of wire harnesses. This thesis identify that the intrinsic physical properties of these two types of objects lead to different hindrances to industrial robots' robotic visual perception. Non-rigid objects in final assembly create the challenge of the visual recognition and tracking of the structure and the topology of non-rigid objects due to their deformation during robotic manipulation, as recognized in Paper 1 and Paper 2. Paper 1 and Paper 2 also unearthed that rigid objects create the challenge for industrial robots to visually recognize objects of interest by exploiting intrinsic features of the objects instead of being assisted by affixed artificial fiducial markers. Moreover, rigid objects with small sizes, complex structures, confusing colors, and family designs exacerbate these challenges. The experimental results in Paper 3 indicated the challenge on differentiating products with family designs from particular views due to their highly similar appearances or occluded distinguishable features.

For robotic operations, obtaining precise positions and orientations in 3D space is critical for designing the robotic control strategy but unwieldy to achieve, as elaborated in section 2.2. Robots require more than just recognizing the position and orientation of an object to accomplish their operations. The recognition must be accurate within tolerances. Otherwise, problems can arise with robotic grasping, leading to failed manipulation and assembly. Paper 1 and Paper 2 revealed that obtaining positions and orientations of objects with enough precision is arduous, especially for objects with small sizes and complex structures. Previous research discussed visual recognition of objects of interest using traditional rule-based computer vision techniques. However, objects with complex structures may cause extreme difficulty in feature engineering when manually designing rule-based vision systems. Previous research exploited artificial fiducial markers with specific patterns to facilitate feature engineering. However, the effectiveness of this kind of rule-based visual recognition solution can be impaired, and the quality of visual recognition will be affected if these artificial fiducial markers are occluded or have low visibility in actual production. The limited space for assembly onto the final products and the large number of objects demanding affixing artificial fiducial markers make this approach more troublesome to implement in industrial applications. Thus, accomplishing visual recognition by exploiting the intrinsic features of objects of interest without attaching artificial fiducial markers is desired.

Further considering industrial applications in practice elaborated in section 2.3, Paper 1 and Paper 2 identified the challenge of guaranteeing a successful integration of visual machine perception into industrial robots for practical applications, regarding the practicality, reliability, robustness, and sustainability of vision systems. Plenty of studies suggested the potential of computer vision techniques for enabling robotic visual perception and the prospective success of vision systems in practical industrial applications [45]. Nevertheless, the practicality and reliability of vision systems in practical applications still need to be discovered. The robustness of vision systems in actual production has yet to be compatible with humans, especially considering the demanding production rate and intricate production environments, such as the background and illumination conditions of visual inputs and the moving production line. In addition, more consideration should be given to the sustainability aspect of vision systems, which is increasingly critical in the current industry and society.

5.1.2 The Answer to RQ2

RQ2: How can robotic visual perception be enabled for assembly tasks?

The Answer in a Nutshell

Learning-based computer vision techniques should be considered and adopted to enable robotic visual perception, increase industrial robots' autonomy, and make them adequate for demanding flexible automation tasks. Beyond 2D RGB images, multi-modality data, such as depth images and point clouds, can be integrated into learning and acquiring the necessary spatial information for robotic operations through multimodal learning. Multi-view visual inputs can improve visual recognition performance when distinguishable features are invisible from particular views. Meanwhile, the benchmark dataset is fundamental for training and evaluating different learning-based solutions, not only for comparing different algorithms' performance but also for evaluating vision systems under practical production configurations. Moreover, evaluating the developed vision systems in practical production configuration is necessary to examine the practicality, reliability, robustness, and sustainability of the developed vision-based solutions. In addition to developing vision systems, new product designs are worth investing in to facilitate robotic visual perception.

The Implication of the Answer

Recent research in computer vision and deep learning has remarkably promoted the performance of visual recognition with learning-based techniques [136]. This indicates the potential of learning-based computer vision techniques to exploiting intrinsic features of objects of interest. Section 2.2 elaborated the superiority of learning-based techniques over rule-based techniques in handling vision tasks around objects with complex features. Paper 1 and Paper 2 advocated to develop and implement learning-based visual recognition algorithms to enable robotic visual perception to utilize intrinsic features of objects of interest. The outcome of Paper 3 further verified the effectiveness of deep learning-based computer vision techniques on rigid object detection in wire harness assembly.

Paper 3 also discussed the potential invisibility of distinguishable features on objects of interest from particular views. Multi-view or video-based visual recognition can address this challenge. Notably, multi-view visual inputs can capture multiple views of an object besides the specific view where distinguishable features are invisible. The same idea is shared with video-based approaches as multi-view information will be captured in videos shooting around objects. The visual recognition algorithm can better recognize the object with more features observed. In addition to processing 2D RGB visual inputs, research in multimodal learning indicates the prospective solution for strengthening visual recognition performance by exploiting various modalities of data [113]. Advancing sensor technology has made it much more convenient to obtain data other than 2D RGB images, such as depth images and point clouds. As elaborated in Paper 1 and Paper 2, these modalities of data have spatial information embedded compared to 2D RGB images, which can facilitate the robotic visual perception of objects of interest in 3D space.

In parallel to the research on learning-based vision systems, research on benchmark datasets is also instrumental, considering the significance of datasets for learning-based computer vision techniques [243]–[246] and the application of computer vision techniques in industry [132]. However, as indicated in Paper 3, there is a lack of datasets for specific scenarios in enabling visual perception of industrial robots for assembly tasks. Therefore, in the experimental study in Paper 3, a dataset of connectors was created initially by taking pictures of connectors and annotating the positions of connectors manually. More studies investigating the aspect of the dataset are desired to improve the visual recognition performance further and lay the foundation for consistent evaluations of various technical solutions for enabling robotic visual perception for assembly tasks.

Moreover, as suggested in Paper 1 and Paper 2, evaluation under practical production configurations is desired and necessary to validate the practicality, reliability, robustness, and sustainability. This evaluation is a prerequisite for guaranteeing that the developed vision systems fulfill the practical requirements of actual production. Additionally, Paper 1, Paper 2, and Paper 3 identified various challenges due to product designs. Inspired by the philosophy of "Design for X" [249], novel designs on objects of interest are desired to facilitate visual recognition.

5.2 Contributions of This Thesis

As elaborated in Chapter 1, more research is required to understand why visual perception capabilities have yet to be enabled on industrial robots for assembly tasks in production and explore prospective technical solutions to achieve it. This thesis ventures into wire harness assembly operations for electric vehicles and the untapped potential of enabling visual perception of industrial robots for such assembly tasks. Through the designed studies, this thesis identified challenges for enabling robotic visual perception for assembly tasks and initiated exploring potential vision-based approaches to address the identified challenges. This thesis provides a practical lens for industry decision-makers, illuminating the potential challenges and opportunities of promoting the vision-based robotic assembly, thereby offering tangible insights for real-world applications. This thesis also provides academia with a road map to address the problems, thereby facilitating the translation of research on robotic visual perception into practical industrial applications.

To Academia

This thesis contributes to the applied research in computer vision, artificial intelligence, robotics, and automation. This thesis's outcome also contributes to research on the robotic assembly of wire harnesses and the robotic manipulation of deformable linear objects in production elaborated in section 2.3. The research conducted in Paper 1 and Paper 2 revealed the increasing research effort across countries. It also underscored the crucial role of vision systems in facilitating the automation of manual wire harness assembly in production. Despite research over a decade, industry has yet to implement practical solutions to address the problem [36]. Therefore, the outcome of Paper 1 and Paper 2 deepened the understanding of the existing research and challenges of enabling robotic visual perception for assembly tasks through a systematic literature review. These insights helped lay the foundation for designing and developing more efficient and effective solutions to enable robotic visual perception for assembly tasks. Afterward, the research conducted in Paper 3 dug deeper into deep learning-based computer vision techniques to verify the effectiveness of deep learning-based object detection for vision-based robotic assembly of wire harnesses through a quantitative experimental study. The computer vision, artificial intelligence, and robotics community discussed the potential applicability of deep learning-based vision systems in industrial applications [22]. The outcome of Paper 3 provided quantitative and qualitative verification of the potential applicability of deep learning-based vision systems in industrial applications. It also hinted at the transformative impact they could have on assembly tasks. On the other hand, the experimental results unearthed more challenges demanding further inquiries into enabling industrial robots' visual perception capabilities for assembly tasks.

To Industry

The technology readiness level (TRL) scale is widely adopted in complex system development to assess the maturity of technologies [250]. The research outcome of this thesis promotes the development of robot vision systems for complex assembly tasks toward higher TRLs. By the beginning of this thesis's research, industry had yet to implement automation solutions for complex assembly tasks that demand flexible automation. The outcome of Paper 1 and Paper 2 indicated that research in relevant fields had been conducted but led to a minor impact on industry. Therefore, the TRL was estimated to be 2 to 3. Afterward, the outcome of Paper 3 verified the effectiveness of deep learning-based techniques to sub-tasks of assembly. Hence, it is estimated that these results contribute to promoting the TRL to 3 to 4.

Specifically, the outcome of Paper 1, Paper 2, and Paper 3 revealed that, beyond research on computer vision algorithms, enabling robotic visual perception for assembly tasks can also be constrained by product designs, practical assembly environment, and production requirements. These insights have indicated the negative effect of several assembly environment conditions on robotic visual perception, which contributes to the assembly sector in the manufacturing industry regarding, for example, designing assembly lines that facilitate robotic visual perception. The research also identified the necessity of optimizing product designs to facilitate visual machine perception, which encourages product designers and manufacturers to optimize designs to facilitate robotic visual perception. Additionally, the experimental study in Paper 3 indicates the effectiveness of deep learning-based computer vision techniques. It suggests a general workflow for deploying a deep learning-based solution to enable robotic visual perception in practice, including dataset collection, data processing, deep learning model training, and evaluation.

5.3 Limitations of This Thesis

Though targeting to make the research findings as generalizable as possible, the author of this thesis foresees potential limitations to the outcome of this thesis, particularly regarding the object and application context.

The realm of robotic perception of the external environment is complex, encompassing not only objects to be manipulated but also objects in the surrounding environment and the presence of humans nearby. However, this thesis generalized all the things a robot needs to perceive as objects of interest and simplified the scope to the objects to be manipulated, i.e., wire harnesses. Before beginning the research, it was premised that wire harnesses could be seen as representative objects in assembly tasks because they consist of rigid and non-rigid objects. However, as discussed in section 2.3, research can generalize wire harnesses as specific DLOs, which is a sub-group of non-rigid object manipulation [70]. Additionally, wire harnesses typically used in industry are made of opaque materials. Given the complexity of the research area, further studies are crucial to examine the generalizability of this thesis's conclusions to other types of objects involved in practical assembly tasks.

Context-wise, the research in this thesis heavily relies on the Swedish automotive industry. Though anticipating the potential of generalizing the research findings to other industrial contexts, the outcome of this thesis related to practical production requirements, e.g., productivity, quality, ergonomics, and safety, may be interpreted differently in other manufacturing industries. Therefore, the research findings may need to be calibrated considering specific application scenarios.

5.4 Reflections on Research Quality

Validity and reliability are two significant criteria for research quality evaluation [193], [214]–[217]. This research consists of a qualitative literature study and a quantitative experimental study. As explained in section 3.4, various methods were adopted in each study of this thesis intended to guarantee the quality of the research, whereas there is a necessity to reflect on the assessment of the research quality of each study.

Following [228], the qualitative literature study in this thesis adopted DARE criteria [206] to evaluate the quality of the systematic literature review, as explained in section 3.4. Correspondingly, the review protocol and the literature searching and selection processes were documented in the appended Paper 2. Specifically, Paper 2 reported the inclusion/exclusion criteria, justified the adequacy of the search, and synthesized the included studies. The quality of each included study was assessed during the literature selection. Details about the included studies were also presented by summarizing each study's purpose, method, and result. The literature study also adopted investigator triangulation [207] to reduce the negative impact of personal bias on data selection, analysis, and interpretation. Specifically, three researchers collaborated through research planning, literature searching, and literature selection. Other researchers were also involved in this research to provide comments and reviews to improve the research quality through peer debriefing [208] and expert review [210].

Following [230], the reliability and validity of the quantitative experimental study in this thesis can be assessed concerning three aspects of the research process: measurement, generalization, and the control of variables. Regarding measurement, the research design and experiment settings were documented in Paper 3. Besides, the dataset was created following the methodology in peer-reviewed research to assure the reproducibility of the study on different occasions by different researchers. Adopting these methods contributed to the reliability of the measurement technique and strategy employed and the validity of the findings of this measurement process. Regarding generalization, the quantitative experimental study in this thesis did not involve sampling from a larger population nor demand inquiry on the generalization of the deep learning-based computer vision techniques. Such theoretical analysis on the generalizability of deep learning techniques can be found in other research [251]–[253]. In addition, the experimental study randomly allocated the dataset for training and testing following stratified sampling and conducted statistical analysis on the experimental results. Adopting both approaches reduced the potential impact of personal bias and enhanced the variable control's reliability and validity. Nevertheless, the author of this thesis foresaw a potential threat to the validity of this quantitative experimental study in the dataset collected for experiments. Remarkably, the research quality can be affected by the quality of images and the annotation quality. The camera and the image-capturing strategy can affect different aspects of the image quality, e.g., the image resolution, the noise, the illumination condition, and the background. This problem may be relieved by using a more sophisticated camera to shoot images using a well-designed strategy. The annotation quality may be affected by the strategy and accuracy of drawing the bounding boxes around objects, especially when the bounding box does not fit the object perfectly and includes parts of other objects. This problem may be relieved by increasing the number of images in the dataset to provide more data covering more scenarios or choosing different annotation strategies in future research. Hence, the author advocated a follow-up study on the benchmark dataset. The quantitative experimental study also adopted peer debriefing [208] to further enhance the quality of this study with reviews and comments from other researchers.

5.5 The Aspect of Ethics

Researchers should anticipate the potential ethical issues and endeavor to address them during their studies [254]–[257]. This section reflects on this thesis's ethics aspect with the intention of introspection.

Prior to beginning the studies, authorship for publication was first negotiated. The risk of breaching privacy is low for data collection, analysis, and interpretation, as this research only collected publicly available literature and images of products with the providers' consent, and this thesis did not collect data from or about people. In the literature study, papers were selected based on their quality and content to avoid potential bias against any author or organization of any author. In the experimental study, both good and bad results were analyzed and presented to avoid confirmation bias [258] or disclosing only positive results. Besides, the details of the research with the study design were released in each appended paper to make it possible for readers to assess the credibility of each study by themselves, following suggestions from [259]. Each study's raw data and other materials were also archived for retracing.

5.6 The Aspect of Sustainability

Attention to sustainability has been expanding immensely globally. Reflecting on this aspect is necessary to strengthen the inclusiveness of research. This section analyzes the impact of the output of this research regarding the aspect of sustainability, considering the triple bottom line of sustainability: economic, social, and environmental [260].

Economic Sustainability

Currently, some assembly tasks are desired to be automated with industrial robots but remain accomplished manually due to the high complexity of tasks. These manual operations need to be improved in terms of quality and productivity. By harnessing the power of robotic visual perception, industrial robots can achieve higher levels of autonomy and become more capable of handling assembly tasks that are currently accomplished manually due to their high complexity. This holds the promise of significant improvements in assembly quality and productivity. In this manner, production can achieve better assembly quality and higher productivity, strengthening economic sustainability.

Social Sustainability

The outcome of this research also has an impact on social sustainability. As explained in Chapter 1, some manual operations in assembly tasks in current production cause ergonomic problems and safety concerns to human operators. With robotic autonomy improved by enabling robotic visual perception, industrial robots will become more intelligent and capable of handling complex assembly tasks, especially those involving ergonomic problems or dangerous operations. In this manner, realizing this research will contribute to the transformation toward safer and more ergonomic-friendly workspace. Nevertheless, it is noteworthy that industrial robots will become more intelligent, with better perception capabilities, and increasingly capable of performing more tasks. Intelligent industrial robots may take over some current manual tasks from human operators. On the other hand, they will also create new challenges and responsibilities for human operators to address. Hence, workforce development through upskilling and reskilling, though out of the scope of this research, should be considered seriously, involving researchers and practitioners from different fields to prepare the organization and business for the forthcoming new industry.

Environmental Sustainability

Enabling the visual perception of industrial robots can indirectly strengthen environmental sustainability. For example, robotic automation in the final assembly may improve assembly quality. Better assembly quality can reduce the waste from the assembly process regarding material and rework. The reduced waste will result in better resource efficiency and less negative environmental impact.

5.7 Future Research

Collecting Better Benchmark Datasets

Although a connector dataset was collected to train and evaluate learning-based object detectors in Paper 3, the visual recognition performance remains unsatisfactory. One cause identified was the insufficient amount of data in the dataset. Therefore, the author of this thesis intends to collect better benchmark datasets for better training and evaluation. Besides 2D RGB images collected in the connector dataset in Paper 3, other modalities of data can also be considered, such as the depth images, to enable industrial robots' capabilities to perceive and infer 3D spatial information.

Investigating Other Vision Tasks

The literature study identifies the need to acquire positions and orientations of objects of interest to enable visual perception in industrial robots for assembly tasks. Moreover, the experimental study verified the effectiveness of learning-based object detectors in acquiring the positions of objects. Future research will include more studies on acquiring orientations, such as 3D object detection and 6D object pose estimation. Additionally, semi-automation based on human-robot collaboration is a direction worth further research to promote robotic automation for assembly tasks. Human-robot collaboration also involves diverse vision tasks for developing efficient, effective, and safe human-robot interaction.

Conducting Evaluation Under Industrial Configurations

The practical aspect of vision-based solutions for enabling robotic visual perception is also instrumental, considering the ultimate goal of integrating the developed vision systems into practical applications in production. Assessing vision-based solutions under practical manufacturing scenarios is critical to validating the proposed vision systems' practicality, reliability, robustness, and sustainability. In the future, the author of this thesis intends to explore how such evaluations can be accomplished.

Chapter 6

Conclusion

Inspired by design science research methodology, this thesis delved into the challenges of enabling robotic visual perception for assembly tasks while exploring potential opportunities and technical solutions through a multiple-method research approach.

This thesis identified four challenges hampering the enabling of robotic visual perception for assembly tasks:

- Visual recognition based on intrinsic features of objects of interest
- Recognizing and tracking the structure and topology of non-rigid objects
- Acquiring high-precision positions and orientations of objects of interest
- Ensuring effective, efficient, and safe applications in practical production

This thesis also identified six prospective directions for developing technical solutions to enable robotic visual perception for assembly tasks:

- Adopting learning-based computer vision techniques
- Developing vision systems based on multi-view and/or multi-modality visual inputs
- Collecting benchmark datasets to facilitate algorithm development and assessment
- Evaluating proposed vision systems under practical manufacturing scenarios
- Developing vision-based human-robot collaboration
- Exploring new product designs to facilitate visual recognition

This thesis has theoretical and practical implications. Theoretically, the results of this thesis are not just an endpoint but a stepping stone for further research. Notably, this thesis provides empirical evidence of challenges and opportunities for enabling visual perception of industrial robots for assembly tasks. These outcomes lay the foundation for developing intelligent robots for industrial applications, opening up new avenues for exploration and innovation in the field. Practically, practitioners can directly adapt the results of this thesis to analyze the specific challenges and opportunities of enabling more intelligent robotic assembly or robotic automation in their task scenarios. Notably, the findings of this thesis can serve as a solid foundation for developing specific machine vision-based robot systems in industry, thereby enhancing the intelligence and adaptability of industrial robots and contributing to more intelligent and sustainable production.

Bibliography

[1] X. Xu, Y. Lu, B. Vogel-Heuser and L. Wang, 'Industry 4.0 and industry 5.0-inception, conception and perception,' *Journal of Manufacturing Systems*, vol. 61, pp. 530–535, 2021. DOI: `10.1016/j.jmsy.2021.10.006` (cit. on pp. 3, 4, 12).

[2] F. J. Riley, *Assembly automation: a management handbook*, 2nd ed. New York, NY, USA: Industrial Press Inc., 1996 (cit. on pp. 3, 10).

[3] J. Krüger, T. Lien and A. Verl, 'Cooperation of human and machines in assembly lines,' *CIRP Annals*, vol. 58, no. 2, pp. 628–646, 2009. DOI: `10.1016/j.cirp.2009.09.009` (cit. on pp. 3, 10).

[4] S. J. Hu, J. Ko, L. Weyand *et al.*, 'Assembly system design and operations for product variety,' *CIRP Annals*, vol. 60, no. 2, pp. 715–733, 2011. DOI: `10.1016/j.cirp.2011.05.004` (cit. on pp. 3, 10).

[5] A. Billard and D. Kragic, 'Trends and challenges in robot manipulation,' *Science*, vol. 364, no. 6446, eaat8414, 2019. DOI: `10.1126/science.aat8414` (cit. on pp. 3–5, 11, 12, 20, 29).

[6] C. Müller, 'World robotics 2023 - industrial robots,' IFR Statistical Department, Tech. Rep., 2023 (cit. on p. 3).

[7] M. T. Mason, 'Toward robotic manipulation,' *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 1–28, 2018. DOI: `10.1146/annurev-control-060117-104848` (cit. on pp. 3–5, 11, 19, 29).

[8] T. Arai, 'Forecast of assembly automation in the automobile industry: Technological progress in robotics,' *Technological Forecasting and Social Change*, vol. 35, no. 2, pp. 133–148, 1989. DOI: `10.1016/0040-1625(89)90051-6` (cit. on p. 3).

[9] Å. Fast-Berglund, F. Palmkvist, P. Nyqvist, S. Ekered and M. Åkerman, 'Evaluating cobots for final assembly,' *Procedia CIRP*, vol. 44, pp. 175–180, 2016. DOI: `10.1016/j.procir.2016.02.114` (cit. on pp. 3, 10).

[10] H. Chen, B. Zhang and G. Zhang, 'Robotic assembly,' in *Handbook of Manufacturing Engineering and Technology*, A. Y. C. Nee, Ed., London, UK: Springer, 2015, pp. 2347–2401. DOI: `10.1007/978-1-4471-4670-4_105` (cit. on pp. 3, 4, 11, 20).

[11] K. E. Stecke and R. P. Parker, 'Flexible automation,' in *Encyclopedia of Production and Manufacturing Management*, P. M. Swamidass, Ed., Boston, MA, USA: Springer, 2000, pp. 213–217. DOI: `10.1007/1-4020-0612-8_343` (cit. on pp. 3, 18).

[12] T. Pardi, 'Fourth industrial revolution concepts in the automotive sector: Performativity, work and employment,' *Journal of Industrial and Business Economics*, vol. 46, no. 3, pp. 379–389, 2019. DOI: `10.1007/s40812-019-00119-9` (cit. on pp. 3, 5, 10).

[13] V. D. Pasquale, S. Miranda, W. P. Neumann and A. Setayesh, 'Human reliability in manual assembly systems: A systematic literature review,' *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 675–680, 2018. DOI: 10.1016/j.ifacol.2018.08.396 (cit. on p. 3).

[14] G. Michalos, S. Makris, N. Papakostas, D. Mourtzis and G. Chryssolouris, 'Automotive assembly technologies review: Challenges and outlook for a flexible and adaptive approach,' *CIRP Journal of Manufacturing Science and Technology*, vol. 2, no. 2, pp. 81–91, 2010. DOI: 10.1016/j.cirpj.2009.12.001 (cit. on pp. 3, 5, 12, 19).

[15] European Commission, Directorate-General for Research and Innovation and J Müller, *Enabling Technologies for Industry 5.0: results of a workshop with Europe's technology leaders*. Luxembourg: Publications Office of the European Union, 2020. DOI: 10.2777/082634 (cit. on p. 3).

[16] J. Leng, W. Sha, B. Wang *et al.*, 'Industry 5.0: Prospect and retrospect,' *Journal of Manufacturing Systems*, vol. 65, pp. 279–295, 2022. DOI: 10.1016/j.jmsy.2022.09.017 (cit. on pp. 3, 4).

[17] L. Rozo, A. G. Kupcsik, P. Schillinger *et al.*, 'The e-bike motor assembly: Towards advanced robotic manipulation for flexible manufacturing,' *Robotics and Computer-Integrated Manufacturing*, vol. 85, p. 102 637, 2024. DOI: 10.1016/j.rcim.2023.102637 (cit. on pp. 3, 11).

[18] H. ElMaraghy, L. Monostori, G. Schuh and W. ElMaraghy, 'Evolution and future of manufacturing systems,' *CIRP Annals*, vol. 70, no. 2, pp. 635–658, 2021. DOI: 10.1016/j.cirp.2021.05.008 (cit. on pp. 4, 12).

[19] F. Jovane, Y. Koren and C. Boër, 'Present and future of flexible automation: Towards new paradigms,' *CIRP Annals*, vol. 52, no. 2, pp. 543–560, 2003. DOI: 10.1016/S0007-8506(07)60203-0 (cit. on p. 4).

[20] L. Wang, S. Liu, H. Liu and X. V. Wang, 'Overview of human-robot collaboration in manufacturing,' in *Proceedings of 5th International Conference on the Industry 4.0 Model for Advanced Manufacturing*, L. Wang, V. D. Majstorovic, D. Mourtzis, E. Carpanzano, G. Moroni and L. M. Galantucci, Eds., Cham, Switzerland: Springer, 2020, pp. 15–58. DOI: 10.1007/978-3-030-46212-3_2 (cit. on pp. 4, 11).

[21] W. He, Z. Li and C. L. P. Chen, 'A survey of human-centered intelligent robots: Issues and challenges,' *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 602–609, 2017. DOI: 10.1109/JAS.2017.7510604 (cit. on p. 4).

[22] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach* (Pearson series in artificial intelligence), 4th ed. Hoboken, NJ, USA: Pearson, 2021 (cit. on pp. 4, 16, 38).

[23] K. M. Galotti, 'Perception: Recognizing patterns and objects,' in *Cognitive Psychology In and Out of the Laboratory*, 5th ed., SAGE Publications, Inc., 2014, ch. 3, pp. 38–64 (cit. on p. 4).

[24] M. Brady, 'Artificial intelligence and robotics,' *Artificial Intelligence*, vol. 26, no. 1, pp. 79–121, 1985. DOI: 10.1016/0004-3702(85)90013-X (cit. on pp. 4, 10).

[25] S. J. Russell and P. Norvig, 'Intelligent agents,' in *Artificial intelligence: a modern approach*, ser. Pearson series in artificial intelligence, 4th ed., Hoboken, NJ, USA: Pearson, 2021, ch. 2, pp. 36–62 (cit. on pp. 4, 12).

[26] T. K. Lien, 'Manual assembly,' in *CIRP Encyclopedia of Production Engineering*, L. Laperrière and G. Reinhart, Eds., Berlin, Heidelberg, Germany: Springer, 2014, pp. 825–828. DOI: 10.1007/978-3-642-20617-7_6624 (cit. on pp. 4, 10).

[27] M. Peña-Cabrera, I. Lopez-Juarez, R. Rios-Cabrera and J. Corona-Castuera, 'Machine vision approach for robotic assembly,' *Assembly Automation*, vol. 25, no. 3, pp. 204–216, 2005. DOI: 10.1108/01445150510610926 (cit. on pp. 4, 15).

[28] H. A. Pierson and M. S. Gashler, 'Deep learning in robotics: A review of recent research,' *Advanced Robotics*, vol. 31, no. 16, pp. 821–835, 2017. DOI: 10.1080/01691864.2017.1365009 (cit. on p. 4).

[29] S. W. Kim, J. H. Kong, S. W. Lee and S. Lee, 'Recent advances of artificial intelligence in manufacturing industrial sectors: A review,' *International Journal of Precision Engineering and Manufacturing*, pp. 1–19, 2022. DOI: 10.1007/s12541-021-00600-3 (cit. on p. 4).

[30] C.-Y. Lee, H. Lee, I. Hwang and B.-T. Zhang, 'Visual perception framework for an intelligent mobile robot,' in *2020 17th International Conference on Ubiquitous Robots (UR)*, 2020, pp. 612–616. DOI: 10.1109/UR49135.2020.9144932 (cit. on p. 4).

[31] H. G. Nguyen, M. Kuhn and J. Franke, 'Manufacturing automation for automotive wiring harnesses,' *Procedia CIRP*, vol. 97, pp. 379–384, 2021. DOI: 10.1016/j.procir.2020.05.254 (cit. on p. 5).

[32] M. Wnuk, C. Hinze, M. Zürn, Q. Pan, A. Lechler and A. Verl, 'Tracking branched deformable linear objects with structure preserved registration by branch-wise probability modification,' in *2021 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, 2021, pp. 101–108. DOI: 10.1109/M2VIP49856.2021.9665147 (cit. on pp. 5, 19, 32).

[33] S. M. Research, 'Automotive wiring harness market: By material (copper, aluminum, others), by vehicle type (passenger cars, two-wheelers, commercial vehicles), by application (chassis harness, engine harness, body & lighting harness, dashboard/cabin harness, sunroof harness, battery harness, hvac harness, seat harness, door harness), by geography, size, global share, forecast, 2021-2030,' Strategic Market Research, Tech. Rep., 2022 (cit. on pp. 5, 29).

[34] M. Zürn, M. Wnuk, A. Lechler and A. Verl, 'Topology matching of branched deformable linear objects,' in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 7097–7103. DOI: 10.1109/ICRA48891.2023.10161483 (cit. on pp. 5, 19, 21, 32).

[35] B. L. Žagar, A. Caporali, A. Szymko *et al.*, 'Copy and paste augmentation for deformable wiring harness bags segmentation,' in *2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, 2023, pp. 721–726. DOI: 10.1109/AIM46323.2023.10196168 (cit. on pp. 5, 31, 32).

[36] O. Salunkhe, W. Quadrini, H. Wang *et al.*, 'Review of current status and future directions for collaborative and semi-automated automotive wire harnesses assembly,' *Procedia CIRP*, vol. 120, pp. 696–701, 2023. DOI: 10.1016/j.procir.2023.09.061 (cit. on pp. 5, 17, 19, 20, 38).

[37] S. A. Pascual and S. Naqvi, 'An investigation of ergonomics analysis tools used in industry in the identification of work-related musculoskeletal disorders,' *International Journal of Occupational Safety and Ergonomics*, vol. 14, no. 2, pp. 237–245, 2008. DOI: 10.1080/10803548.2008.11076755 (cit. on p. 5).

[38] C. Fischer, J. Bönig, J. Franke, M. Lušić and R. Hornfeck, 'Worker information system to support during complex and exhausting assembly of high-voltage harness,' in *2015 5th International Electric Drives Production Conference (EDPC)*, 2015, pp. 1–7. DOI: 10.1109/EDPC.2015.7323211 (cit. on p. 5).

[39] S. Olbrich and J. Lackinger, 'Manufacturing processes of automotive high-voltage wire harnesses: State of the art, current challenges and fields of action to reach a higher level of automation,' *Procedia CIRP*, vol. 107, pp. 653–660, 2022. DOI: 10.1016/j.procir.2022.05.041 (cit. on p. 5).

[40] H. Wang, O. Salunkhe, W. Quadrini *et al.*, 'Overview of computer vision techniques in robotized wire harness assembly: Current state and future opportunities,' *Procedia CIRP*, vol. 120, pp. 1071–1076, 2023. DOI: 10.1016/j.procir.2023.09.127 (cit. on pp. 5, 17, 20).

[41] J. Shi, B. Hamner, R. Simmons and S. Singh, 'Mobile robotic assembly on a moving vehicle,' in *International Symposium on Flexible Automation*, vol. 45110, 2012, pp. 117–123. DOI: 10.1115/ISFA2012-7193 (cit. on p. 5).

[42] International Organization for Standardization, *Iso 10218-1:2011 robots and robotic devices — safety requirements for industrial robots — part 1: Robots*, 2011 (cit. on pp. 5, 20).

[43] International Organization for Standardization, *Iso 10218-2:2011 robots and robotic devices — safety requirements for industrial robots — part 2: Robot systems and integration*, 2011 (cit. on pp. 5, 20).

[44] International Organization for Standardization, *Iso/ts 15066:2016 robots and robotic devices — collaborative robots*, 2016 (cit. on pp. 5, 20).

[45] L. Zhou, L. Zhang and N. Konz, 'Computer vision techniques in manufacturing,' *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 1, pp. 105–117, 2023. DOI: 10.1109/TSMC.2022.3166397 (cit. on pp. 6, 12, 13, 15, 16, 20, 21, 29, 36).

[46] O. E. Dictionary, *Automation (n.), sense 2*, 2023. DOI: 10.1093/OED/5111105470 (cit. on p. 9).

[47] CIRP, 'Fundamental terms of manufacturing/grundlegende begriffe der produktion/termini fondamentali della produzione,' in *Dictionary of Production Engineering III – Manufacturing Systems Wörterbuch der Fertigungstechnik III – Produktionssysteme Dizionario di Ingegneria della Produzione III – Sistemi di produzione: Trilingual Edition Dreisprachige Ausgabe Edizione completa trilingue*, Berlin, Heidelberg, Germany: Springer, 2020, ch. 1, pp. 1–59. DOI: 10.1007/978-3-662-53334-5_1 (cit. on p. 9).

[48] Å. Fasth-Berglund and J. Stahre, 'Cognitive automation strategy for reconfigurable and sustainable assembly systems,' *Assembly automation*, vol. 33, no. 3, pp. 294–303, 2013. DOI: 10.1108/AA-12-2013-036 (cit. on p. 9).

[49] Å. Fasth, 'Quantifying levels of automation - to enable competitive assembly systems,' Ph.D. dissertation, 2012 (cit. on p. 9).

[50] D. Romero, J. Stahre and M. Taisch, 'The operator 4.0: Towards socially sustainable factories of the future,' *Computers & Industrial Engineering*, vol. 139, p. 106 128, 2020. DOI: 10.1016/j.cie.2019.106128 (cit. on p. 9).

[51] D. Thurman, D. Brann and C. Mitchell, 'An architecture to support incremental automation of complex systems,' in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, vol. 2, 1997, pp. 1174–1179. DOI: 10.1109/ICSMC.1997.638109 (cit. on p. 9).

[52] S. J. Hu, 'Assembly,' in *CIRP Encyclopedia of Production Engineering*, L. Laperrière and G. Reinhart, Eds., Berlin, Heidelberg, Germany: Springer, 2014, pp. 50–52. DOI: 10.1007/978-3-642-20617-7_6616 (cit. on p. 10).

[53] T. Lien and F. Rasch, 'Hybrid automatic-manual assembly systems,' *CIRP Annals*, vol. 50, no. 1, pp. 21–24, 2001. DOI: 10.1016/S0007-8506(07)62062-9 (cit. on p. 10).

[54] G. Reinhart, 'Assembly automation,' in *CIRP Encyclopedia of Production Engineering*, L. Laperrière and G. Reinhart, Eds., Berlin, Heidelberg, Germany: Springer, 2014, pp. 52–54. DOI: 10.1007/978-3-642-20617-7_6617 (cit. on pp. 10, 19).

[55] G. Reinhart, 'Assembly line,' in *CIRP Encyclopedia of Production Engineering*, L. Laperrière and G. Reinhart, Eds., Berlin, Heidelberg, Germany: Springer, 2014, pp. 55–60. DOI: 10.1007/978-3-642-20617-7_8 (cit. on p. 10).

[56] G. Boothroyd, *Assembly automation and product design*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2005. DOI: 10.1201/9781420027358 (cit. on p. 10).

[57] H. Bley, G. Reinhart, G. Seliger, M. Bernardi and T. Korne, 'Appropriate human involvement in assembly and disassembly,' *CIRP Annals*, vol. 53, no. 2, pp. 487–509, 2004. DOI: 10.1016/S0007-8506(07)60026-2 (cit. on p. 10).

[58] M. T. Mason, 'Creation myths: The beginnings of robotics research,' *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 72–77, 2012. DOI: 10.1109/MRA.2012.2191437 (cit. on p. 10).

[59] B. Siciliano and O. Khatib, Eds., *Springer Handbook of Robotics*, 2nd ed. Cham, Switzerland: Springer, 2016. DOI: 10.1007/978-3-319-32552-1 (cit. on pp. 10, 12).

[60] B. Siciliano and O. Khatib, 'Robotics and the handbook,' in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds., 2nd ed., Cham, Switzerland: Springer, 2016, ch. 1, pp. 1–6. DOI: 10.1007/978-3-319-32552-1_1 (cit. on pp. 10–12).

[61] E. Garcia, M. A. Jimenez, P. G. De Santos and M. Armada, 'The evolution of robotics research,' *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 90–103, 2007. DOI: 10.1109/MRA.2007.339608 (cit. on pp. 11, 12).

[62] S. J. Russell and P. Norvig, 'Robotics,' in *Artificial intelligence: a modern approach*, ser. Pearson series in artificial intelligence, 4th ed., Hoboken, NJ, USA: Pearson, 2021, ch. 26, pp. 925–980 (cit. on pp. 11, 12).

[63] T. K. Lien, 'Robot,' in *CIRP Encyclopedia of Production Engineering*, L. Laperrière and G. Reinhart, Eds., Berlin, Heidelberg, Germany: Springer, 2014, pp. 1068–1076. DOI: 10.1007/978-3-642-20617-7_6628 (cit. on pp. 11, 12).

[64] L. A. Ballard, S. Sabanovic, J. Kaur and S. Milojevic, 'George charles devol, jr. [history],' *IEEE Robotics & Automation Magazine*, vol. 19, no. 3, pp. 114–119, 2012. DOI: 10.1109/MRA.2012.2206672 (cit. on p. 11).

[65] S. Yun, M.-S. Choi, M.-Y. Cho *et al.*, 'Next-generation furniture assembly by ai and robots: Team sk2y: A winner of the furniture assembly competition at ai-robot challenge 2021,' *IEEE Robotics & Automation Magazine*, vol. 30, no. 2, pp. 96–108, 2023. DOI: 10.1109/MRA.2022.3188214 (cit. on p. 11).

[66] D. Jiang, H. Wang and Y. Lu, 'Mastering the complex assembly task with a dual-arm robot: A novel reinforcement learning method,' *IEEE Robotics & Automation Magazine*, vol. 30, no. 2, pp. 57–66, 2023. DOI: 10.1109/MRA.2023.3262461 (cit. on p. 11).

[67] C. Heyer, 'Human-robot interaction and future industrial robotics applications,' in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 4749–4754. DOI: 10.1109/IROS.2010.5651294 (cit. on p. 11).

[68] R. R. Galin and R. V. Meshcheryakov, 'Human-robot interaction efficiency and human-robot collaboration,' in *Robotics: Industry 4.0 Issues & New Intelligent Control Paradigms*, ser. Studies in Systems, Decision and Control, A. G. Kravets, Ed., vol. 272, Cham, Switzerland: Springer, 2020, pp. 55–63. DOI: 10.1007/978-3-030-37841-7_5 (cit. on p. 11).

[69] A. Grau, M. Indri, L. L. Bello and T. Sauter, 'Industrial robotics in factory automation: From the early stage to the internet of things,' in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, 2017, pp. 6159–6164. DOI: 10.1109/IECON.2017.8217070 (cit. on p. 11).

[70] S. Makris, F. Dietrich, K. Kellens and S. J. Hu, 'Automated assembly of non-rigid objects,' *CIRP Annals*, vol. 72, no. 2, pp. 513–539, 2023. DOI: 10.1016/j.cirp.2023.05.003 (cit. on pp. 12, 16–18, 39).

[71] I. Goodfellow, Y. Bengio and A. Courville, *Deep learning*. Cambridge, MA, USA: MIT press, 2016 (cit. on pp. 12, 16, 17).

[72] R. Li and H. Qiao, 'A survey of methods and strategies for high-precision robotic grasping and assembly tasks—some new trends,' *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 6, pp. 2718–2732, 2019. DOI: 10.1109/TMECH.2019.2945135 (cit. on pp. 12, 13, 16).

[73] J. Yang, C. Wang, B. Jiang, H. Song and Q. Meng, 'Visual perception enabled industry intelligence: State of the art, challenges and prospects,' *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2204–2219, 2021. DOI: 10.1109/TII.2020.2998818 (cit. on p. 12).

[74] R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed. Cham, Switzerland: Springer, 2022. DOI: 10.1007/978-3-030-34372-9 (cit. on p. 12).

[75] T. S. Huang, 'Computer vision: Evolution and promise,' in *1996 CERN School of Computing*, C. E. Vandoni, Ed., 1996, pp. 21–25. DOI: 10.5170/CERN-1996-008.21 (cit. on p. 13).

[76] S. J. D. Prince, *Computer Vision: Models, Learning, and Inference*. Cambridge, UK: Cambridge University Press, 2012. DOI: 10.1017/CBO9780511996504 (cit. on p. 13).

[77] B. K. P. Horn, *Robot vision*. Cambridge, MA, USA: MIT press, 1986 (cit. on p. 13).

[78] A. Krizhevsky, I. Sutskever and G. E. Hinton, 'Imagenet classification with deep convolutional neural networks,' in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou and K. Weinberger, Eds., vol. 25, Curran Associates, Inc., 2012 (cit. on p. 13).

[79] A. Krizhevsky, I. Sutskever and G. E. Hinton, 'Imagenet classification with deep convolutional neural networks,' *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. DOI: 10.1145/3065386 (cit. on p. 13).

[80] H. Harzallah, F. Jurie and C. Schmid, ‘Combining efficient object localization and image classification,’ in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 237–244. DOI: 10.1109/ICCV.2009.5459257 (cit. on p. 13).

[81] W. Chen, Y. Li, Z. Tian and F. Zhang, ‘2d and 3d object detection algorithms from images: A survey,’ *Array*, vol. 19, p. 100 305, 2023. DOI: 10.1016/j.array.2023.100305 (cit. on pp. 13, 14).

[82] A. B. Amjoud and M. Amrouch, ‘Object detection using deep learning, cnns and vision transformers: A review,’ *IEEE Access*, vol. 11, pp. 35 479–35 516, 2023. DOI: 10.1109/ACCESS.2023.3266093 (cit. on p. 13).

[83] Z.-Q. Zhao, P. Zheng, S.-T. Xu and X. Wu, ‘Object detection with deep learning: A review,’ *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019. DOI: 10.1109/TNNLS.2018.2876865 (cit. on pp. 13, 14).

[84] R. Girshick, J. Donahue, T. Darrell and J. Malik, ‘Rich feature hierarchies for accurate object detection and semantic segmentation,’ in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587. DOI: 10.1109/CVPR.2014.81 (cit. on pp. 13, 14).

[85] Z. Zou, K. Chen, Z. Shi, Y. Guo and J. Ye, ‘Object detection in 20 years: A survey,’ *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023. DOI: 10.1109/JPROC.2023.3238524 (cit. on pp. 13, 14, 32).

[86] P. Viola and M. Jones, ‘Rapid object detection using a boosted cascade of simple features,’ in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. 511–518. DOI: 10.1109/CVPR.2001.990517 (cit. on p. 13).

[87] P. Viola and M. J. Jones, ‘Robust real-time face detection,’ *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004. DOI: 10.1023/B:VISI.0000013087.49260.fb (cit. on p. 13).

[88] N. Dalal and B. Triggs, ‘Histograms of oriented gradients for human detection,’ in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 1, 2005, pp. 886–893. DOI: 10.1109/CVPR.2005.177 (cit. on p. 13).

[89] P. Felzenszwalb, D. McAllester and D. Ramanan, ‘A discriminatively trained, multiscale, deformable part model,’ in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. DOI: 10.1109/CVPR.2008.4587597 (cit. on p. 13).

[90] R. Girshick, J. Donahue, T. Darrell and J. Malik, ‘Region-based convolutional networks for accurate object detection and segmentation,’ *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, 2016. DOI: 10.1109/TPAMI.2015.2437384 (cit. on p. 14).

[91] R. Girshick, ‘Fast r-cnn,’ in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448. DOI: 10.1109/ICCV.2015.169 (cit. on p. 14).

[92] S. Ren, K. He, R. Girshick and J. Sun, ‘Faster r-cnn: Towards real-time object detection with region proposal networks,’ in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett, Eds., vol. 28, Curran Associates, Inc., 2015, pp. 91–99 (cit. on pp. 14, 33).

[93] K. He, X. Zhang, S. Ren and J. Sun, 'Spatial pyramid pooling in deep convolutional networks for visual recognition,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015. DOI: 10.1109/TPAMI.2015.2389824 (cit. on p. 14).

[94] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, 'Feature pyramid networks for object detection,' in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 936–944. DOI: 10.1109/CVPR.2017.106 (cit. on p. 14).

[95] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, 'You only look once: Unified, real-time object detection,' in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91 (cit. on p. 14).

[96] J. Redmon and A. Farhadi, *Yolov3: An incremental improvement*, 2018. DOI: 10.48550/arXiv.1804.02767 (cit. on p. 14).

[97] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, *Yolov4: Optimal speed and accuracy of object detection*, 2020. DOI: 10.48550/arXiv.2004.10934 (cit. on p. 14).

[98] J. Redmon and A. Farhadi, 'Yolo9000: Better, faster, stronger,' in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6517–6525. DOI: 10.1109/CVPR.2017.690 (cit. on p. 14).

[99] C.-Y. Wang, A. Bochkovskiy and H.-Y. M. Liao, 'Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,' in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7464–7475. DOI: 10.1109/CVPR52729.2023.00721 (cit. on p. 14).

[100] W. Liu, D. Anguelov, D. Erhan *et al.*, 'Ssd: Single shot multibox detector,' in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe and M. Welling, Eds., Cham, Switzerland: Springer, 2016, pp. 21–37. DOI: 10.1007/978-3-319-46448-0_2 (cit. on p. 14).

[101] T.-Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, 'Focal loss for dense object detection,' in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2999–3007. DOI: 10.1109/ICCV.2017.324 (cit. on p. 14).

[102] H. Law and J. Deng, 'Cornernet: Detecting objects as paired keypoints,' in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss, Eds., Cham, Switzerland: Springer, 2018, pp. 765–781. DOI: 10.1007/978-3-030-01264-9_45 (cit. on p. 14).

[103] X. Zhou, D. Wang and P. Krähenbühl, *Objects as points*, 2019. DOI: 10.48550/arXiv.1904.07850 (cit. on p. 14).

[104] A. Vaswani, N. Shazeer, N. Parmar *et al.*, 'Attention is all you need,' in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio *et al.*, Eds., vol. 30, Curran Associates, Inc., 2017 (cit. on p. 14).

[105] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan and M. Shah, 'Transformers in vision: A survey,' *ACM Computing Surveys*, vol. 54, no. 10s, 2022. DOI: 10.1145/3505244 (cit. on p. 14).

[106] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov and S. Zagoruyko, 'End-to-end object detection with transformers,' in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox and J.-M. Frahm, Eds., Cham, Switzerland: Springer, 2020, pp. 213–229. DOI: 10.1007/978-3-030-58452-8_13 (cit. on p. 14).

[107] X. Zhu, W. Su, L. Lu, B. Li, X. Wang and J. Dai, 'Deformable detr: Deformable transformers for end-to-end object detection,' in *ICLR2021 - 9th International Conference on Learning Representations*, 2021 (cit. on p. 14).

[108] H. Zhang, F. Li, S. Liu *et al.*, 'Dino: Detr with improved denoising anchor boxes for end-to-end object detection,' in *ICLR2023 - 11th International Conference on Learning Representations*, 2023 (cit. on p. 14).

[109] F. Li, H. Zhang, H. Xu *et al.*, 'Mask dino: Towards a unified transformer-based framework for object detection and segmentation,' in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 3041–3050. DOI: 10.1109/CVPR52729.2023.00297 (cit. on p. 14).

[110] C. Sahin, G. Garcia-Hernando, J. Sock and T.-K. Kim, 'A review on object pose recovery: From 3d bounding box detectors to full 6d pose estimators,' *Image and Vision Computing*, vol. 96, p. 103 898, 2020. DOI: 10.1016/j.imavis.2020.103898 (cit. on pp. 14, 15).

[111] S. Hoque, M. Y. Arafat, S. Xu, A. Maiti and Y. Wei, 'A comprehensive review on 3d object detection and 6d pose estimation with deep learning,' *IEEE Access*, vol. 9, pp. 143 746–143 770, 2021. DOI: 10.1109/ACCESS.2021.3114399 (cit. on p. 14).

[112] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu and M. Bennamoun, 'Deep learning for 3d point clouds: A survey,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4338–4364, 2021. DOI: 10.1109/TPAMI.2020.3005434 (cit. on p. 14).

[113] W. Liang, P. Xu, L. Guo, H. Bai, Y. Zhou and F. Chen, 'A survey of 3d object detection,' *Multimedia Tools and Applications*, vol. 80, no. 19, pp. 29 617–29 641, 2021. DOI: 10.1007/s11042-021-11137-y (cit. on pp. 14, 37).

[114] Y. Zhu, M. Li, W. Yao and C. Chen, 'A review of 6d object pose estimation,' in *2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, vol. 10, 2022, pp. 1647–1655. DOI: 10.1109/ITAIC54216.2022.9836663 (cit. on p. 14).

[115] J. Guan, Y. Hao, Q. Wu, S. Li and Y. Fang, 'A survey of 6dof object pose estimation methods for different application scenarios,' *Sensors*, vol. 24, no. 4, p. 1076, 2024. DOI: 10.3390/s24041076 (cit. on pp. 14, 15).

[116] F. Gorschlüter, P. Rojtberg and T. Pöllabauer, 'A survey of 6d object detection based on 3d models for industrial applications,' *Journal of Imaging*, vol. 8, no. 3, p. 53, 2022. DOI: 10.3390/jimaging8030053 (cit. on p. 14).

[117] Z. He, W. Feng, X. Zhao and Y. Lv, '6d pose estimation of objects: Recent technologies and challenges,' *Applied Sciences*, vol. 11, no. 1, p. 228, 2021. DOI: 10.3390/app11010228 (cit. on pp. 14, 15).

[118] G. J. Agin, 'Computer vision systems for industrial inspection and assembly,' *Computer*, vol. 13, no. 5, pp. 11–20, 1980. DOI: 10.1109/MC.1980.1653613 (cit. on p. 15).

[119] A. Soini, 'Machine vision technology take-up in industrial applications,' in *ISPA 2001. Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis. In conjunction with 23rd International Conference on Information Technology Interfaces (IEEE Cat.*, 2001, pp. 332–338. DOI: 10.1109/ISPA.2001.938651 (cit. on p. 15).

[120] M. Ejiri, 'Machine vision technology: Past, present and future,' in *EEE International Workshop on Intelligent Robots and Systems, Towards a New Frontier of Applications*, vol. 1, 1990, pp. XXIX–XXXX. DOI: 10.1109/IROS.1990.262354 (cit. on p. 15).

[121] J. Jolion, 'Computer vision methodologies,' *CVGIP: Image Understanding*, vol. 59, no. 1, pp. 53–71, 1994. DOI: 10.1006/ciun.1994.1004 (cit. on pp. 15, 27).

[122] J. M. Jolion, 'Methodology for design of image analysis systems: An application to electron microscopy,' Ph.D. dissertation, Lyon, France, 1987 (cit. on p. 15).

[123] H. Freeman, Ed., *Machine vision: algorithms, architectures, and systems* (Perspectives in Computing). San Diego, CA, USA: Academic Press, Inc., 1988, vol. 22. DOI: 10.1016/B978-0-12-266720-6.X5001-3 (cit. on p. 15).

[124] H. Freeman, Ed., *Machine vision for inspection and measurement* (Perspectives in Computing). San Diego, CA, USA: Academic Press, Inc., 1989, vol. 24. DOI: 10.1016/B978-0-122-66719-0.X5001-5 (cit. on p. 15).

[125] H. Golnabi and A. Asadpour, 'Design and application of industrial machine vision systems,' *Robotics and Computer-Integrated Manufacturing*, vol. 23, no. 6, pp. 630–637, 2007. DOI: 10.1016/j.rcim.2007.02.005 (cit. on p. 15).

[126] V. Kuts, T. Otto, T. Tähemaa, K. Bukhari and T. Pataraia, 'Adaptive industrial robots using machine vision,' in *Proceedings of the ASME 2018 International Mechanical Engineering Congress and Exposition*, ser. ASME International Mechanical Engineering Congress and Exposition, vol. 2, Pittsburgh, PA, USA: ASME, 2018, V002T02A093. DOI: 10.1115/IMECE2018-86720 (cit. on p. 15).

[127] F. K. Konstantinidis, S. G. Mouroutsos and A. Gasteratos, 'The role of machine vision in industry 4.0: An automotive manufacturing perspective,' in *2021 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2021, pp. 1–6. DOI: 10.1109/IST50367.2021.9651453 (cit. on p. 16).

[128] Y. Guo, M. Bennamoun, F. Sohel, M. Lu and J. Wan, '3d object recognition in cluttered scenes with local surface features: A survey,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2270–2287, 2014. DOI: 10.1109/TPAMI.2014.2316828 (cit. on p. 16).

[129] X. Jiang, K. mo Koo, K. Kikuchi, A. Konno and M. Uchiyama, 'Robotized assembly of a wire harness in a car production line,' *Advanced Robotics*, vol. 25, no. 3-4, pp. 473–489, 2011. DOI: 10.1163/016918610X551782 (cit. on pp. 16, 29, 31, 32).

[130] S. Ojha and S. Sakhare, 'Image processing techniques for object tracking in video surveillance- a survey,' in *2015 International Conference on Pervasive Computing (ICPC)*, 2015, pp. 1–6. DOI: 10.1109/PERVASIVE.2015.7087180 (cit. on p. 16).

[131] L. Alzubaidi, J. Zhang, A. J. Humaidi *et al.*, 'Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions,' *Journal of Big Data*, vol. 8, no. 1, p. 53, 2021. DOI: 10.1186/s40537-021-00444-8 (cit. on p. 16).

[132] H. G. Nguyen, R. Habiboglu and J. Franke, 'Enabling deep learning using synthetic data: A case study for the automotive wiring harness manufacturing,' *Procedia CIRP*, vol. 107, pp. 1263–1268, 2022. DOI: 10.1016/j.procir.2022.05.142 (cit. on pp. 16, 37).

[133] K. Kimble, J. Albrecht, M. Zimmerman and J. Falco, 'Performance measures to benchmark the grasping, manipulation, and assembly of deformable objects typical to manufacturing applications,' *Frontiers in Robotics and AI*, vol. 9, 2022. DOI: 10.3389/frobt.2022.999348 (cit. on p. 16).

[134] J. McCarthy, M. L. Minsky, N. Rochester and C. E. Shannon, 'A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955,' *AI Magazine*, vol. 27, no. 4, p. 12, 2006. DOI: 10.1609/aimag.v27i4.1904 (cit. on p. 16).

[135] S. J. Russell and P. Norvig, 'Learning from examples,' in *Artificial intelligence: a modern approach*, ser. Pearson series in artificial intelligence, 4th ed., Hoboken, NJ, USA: Pearson, 2021, ch. 19, pp. 651–720 (cit. on p. 16).

[136] Y. LeCun, Y. Bengio and G. Hinton, 'Deep learning,' *nature*, vol. 521, no. 7553, pp. 436–444, 2015. DOI: 10.1038/nature14539 (cit. on pp. 16, 17, 32, 37).

[137] S. J. Russell and P. Norvig, 'Deep learning,' in *Artificial intelligence: a modern approach*, ser. Pearson series in artificial intelligence, 4th ed., Hoboken, NJ, USA: Pearson, 2021, ch. 21, pp. 750–788 (cit. on p. 17).

[138] W. S. McCulloch and W. Pitts, 'A logical calculus of the ideas immanent in nervous activity,' *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943. DOI: 10.1007/BF02478259 (cit. on p. 17).

[139] X. Zhang, X. Zhang and W. Wang, 'Convolutional neural network,' in *Intelligent Information Processing with Matlab*. Singapore: Springer, 2023, pp. 39–71. DOI: 10.1007/978-981-99-6449-9_2 (cit. on p. 17).

[140] S. Cong and Y. Zhou, 'A review of convolutional neural network architectures and their optimizations,' *Artificial Intelligence Review*, vol. 56, no. 3, pp. 1905–1969, 2023. DOI: 10.1007/s10462-022-10213-5 (cit. on p. 17).

[141] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, 'A survey of convolutional neural networks: Analysis, applications, and prospects,' *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2022. DOI: 10.1109/TNNLS.2021.3084827 (cit. on p. 17).

[142] I. Kovalenko, K. Barton, J. Moyne and D. M. Tilbury, 'Opportunities and challenges to integrate artificial intelligence into manufacturing systems: Thoughts from a panel discussion [opinion],' *IEEE Robotics & Automation Magazine*, vol. 30, no. 2, pp. 109–112, 2023. DOI: 10.1109/MRA.2023.3262464 (cit. on p. 17).

[143] J. Lee, 'Definition and meaning of industrial ai,' in *Industrial AI: Applications with Sustainable Performance*. Singapore: Springer, 2020, pp. 33–61. DOI: 10.1007/978-981-15-2144-7_3 (cit. on p. 17).

[144] G. Chryssolouris, K. Alexopoulos and Z. Arkouli, *A Perspective on Artificial Intelligence in Manufacturing*, 1st ed. Cham, Switzerland: Springer, 2023. DOI: 10.1007/978-3-031-21828-6 (cit. on p. 17).

[145] A. Simeth and P. Plapper, 'Artificial intelligence based robotic automation of manual assembly tasks for intelligent manufacturing,' in *Smart, Sustainable Manufacturing in an Ever-Changing World*, K. von Leipzig, N. Sacks and M. Mc Clelland, Eds., Cham, Switzerland: Springer, 2023, pp. 137–148. DOI: 10.1007/978-3-031-15602-1_11 (cit. on p. 17).

[146] K. Galassi and G. Palli, 'Robotic wires manipulation for switchgear cabling and wiring harness manufacturing,' in *2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS)*, 2021, pp. 531–536. DOI: 10.1109/ICPS49255.2021.9468128 (cit. on p. 17).

[147] N. Lv, J. Liu and Y. Jia, 'Dynamic modeling and control of deformable linear objects for single-arm and dual-arm robot manipulations,' *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2341–2353, 2022. DOI: 10.1109/TRO.2021.3139838 (cit. on p. 17).

[148] M. Saadat and P. Nan, 'Industrial applications of automatic manipulation of flexible materials,' *Industrial Robot: An International Journal*, vol. 29, no. 5, pp. 434–442, 2002. DOI: 10.1108/01439910210440255 (cit. on pp. 17–20).

[149] J. Sanchez, J.-A. Corrales, B.-C. Bouzgarrou and Y. Mezouar, 'Robotic manipulation and sensing of deformable objects in domestic and industrial applications: A survey,' *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 688–716, 2018. DOI: 10.1177/0278364918779698 (cit. on p. 18).

[150] F. Nadon, A. J. Valencia and P. Payeur, 'Multi-modal sensing and robotic manipulation of non-rigid objects: A survey,' *Robotics*, vol. 7, no. 4, p. 74, 2018. DOI: 10.3390/robotics7040074 (cit. on p. 18).

[151] H. Yin, A. Varava and D. Kragic, 'Modeling, learning, perception, and control methods for deformable object manipulation,' *Science Robotics*, vol. 6, no. 54, eabd8803, 2021. DOI: 10.1126/scirobotics.abd8803 (cit. on p. 18).

[152] J. Zhu, A. Cherubini, C. Dune *et al.*, 'Challenges and outlook in robotic manipulation of deformable objects,' *IEEE Robotics & Automation Magazine*, vol. 29, no. 3, pp. 67–77, 2022. DOI: 10.1109/MRA.2022.3147415 (cit. on p. 18).

[153] X. Li, X. Su and Y.-H. Liu, 'Vision-based robotic manipulation of flexible pcbs,' *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 6, pp. 2739–2749, 2018. DOI: 10.1109/TMECH.2018.2869147 (cit. on p. 18).

[154] M. C. Gemici and A. Saxena, 'Learning haptic representation for manipulating deformable food objects,' in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 638–645. DOI: 10.1109/IROS.2014.6942626 (cit. on p. 18).

[155] L. Cao, X. Li, P. T. Phan *et al.*, 'Sewing up the wounds: A robotic suturing system for flexible endoscopy,' *IEEE Robotics & Automation Magazine*, vol. 27, no. 3, pp. 45–54, 2020. DOI: 10.1109/MRA.2019.2963161 (cit. on p. 18).

[156] F. Zhang and Y. Demiris, 'Learning garment manipulation policies toward robot-assisted dressing,' *Science Robotics*, vol. 7, no. 65, eabm6010, 2022. DOI: 10.1126/scirobotics.abm6010 (cit. on p. 18).

[157] P. Jiménez, 'Survey on model-based manipulation planning of deformable objects,' *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 2, pp. 154–163, 2012. DOI: 10.1016/j.rcim.2011.08.002 (cit. on pp. 18, 19).

[158] F. Guo, H. Lin and Y.-B. Jia, 'Squeeze grasping of deformable planar objects with segment contacts and stick/slip transitions,' in *2013 IEEE International Conference on Robotics and Automation*, 2013, pp. 3736–3741. DOI: 10.1109/ICRA.2013.6631102 (cit. on p. 18).

[159] S. Javdani, S. Tandon, J. Tang, J. F. O'Brien and P. Abbeel, 'Modeling and perception of deformable one-dimensional objects,' in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 1607–1614. DOI: 10.1109/ICRA.2011.5980431 (cit. on p. 18).

[160] A. Keipour, M. Bandari and S. Schaal, 'Deformable one-dimensional object detection for routing and manipulation,' *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4329–4336, 2022. DOI: 10.1109/LRA.2022.3146920 (cit. on p. 18).

[161] C. Chen and Y. Zheng, 'Deformation identification and estimation of one-dimensional objects by using vision sensors,' in *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, vol. 3, 1991, pp. 2306–2311. DOI: 10.1109/ROBOT.1991.131538 (cit. on p. 18).

[162] W. Wang, D. Berenson and D. Balkcom, 'An online method for tight-tolerance insertion tasks for string and rope,' in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 2488–2495. DOI: 10.1109/ICRA.2015.7139532 (cit. on p. 18).

[163] J.-E. Byun and T. i Nagata, 'Determining the 3-d pose of a flexible object by stereo matching of curvature representations,' *Pattern Recognition*, vol. 29, no. 8, pp. 1297–1307, 1996. DOI: 10.1016/0031-3203(95)00165-4 (cit. on pp. 18, 21).

[164] G. E. Navas-Reascos, D. Romero, J. Stahre and A. Caballero-Ruiz, 'Wire harness assembly process supported by collaborative robots: Literature review and call for r&d,' *Robotics*, vol. 11, no. 3, p. 65, 2022. DOI: 10.3390/robotics11030065 (cit. on p. 18).

[165] M. A. Lee, Y. Zhu, P. Zachares *et al.*, 'Making sense of vision and touch: Learning multimodal representations for contact-rich tasks,' *IEEE Transactions on Robotics*, vol. 36, no. 3, pp. 582–596, 2020. DOI: 10.1109/TRO.2019.2959445 (cit. on p. 18).

[166] H. Zhou, S. Li, Q. Lu and J. Qian, 'A practical solution to deformable linear object manipulation: A case study on cable harness connection,' in *2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM)*, 2020, pp. 329–333. DOI: 10.1109/ICARM49381.2020.9195380 (cit. on pp. 18, 19, 31, 32).

[167] D. Andronas, Z. Arkouli, N. Zacharaki *et al.*, 'On the perception and handling of deformable objects - a robotic cell for white goods industry,' *Robotics and Computer-Integrated Manufacturing*, vol. 77, p. 102 358, 2022. DOI: 10.1016/j.rcim.2022.102358 (cit. on p. 18).

[168] T. Hermansson, R. Bohlin, J. S. Carlson and R. Söderberg, 'Automatic assembly path planning for wiring harness installations,' *Journal of Manufacturing Systems*, vol. 32, no. 3, pp. 417–422, 2013. DOI: 10.1016/j.jmsy.2013.04.006 (cit. on p. 18).

[169] A. Nair, D. Chen, P. Agrawal *et al.*, 'Combining self-supervised learning and imitation for vision-based rope manipulation,' in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 2146–2153. DOI: 10.1109/ICRA.2017.7989247 (cit. on pp. 18, 20, 21).

[170] M. Yan, Y. Zhu, N. Jin and J. Bohg, 'Self-supervised learning of state estimation for manipulating deformable linear objects,' *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2372–2379, 2020. DOI: 10.1109/LRA.2020.2969931 (cit. on pp. 18, 20, 21).

[171] S. Heydaryan, J. Suaza Bedolla and G. Belingardi, 'Safety design and development of a human-robot collaboration assembly process in the automotive industry,' *Applied Sciences*, vol. 8, no. 3, p. 344, 2018. DOI: 10.3390/app8030344 (cit. on p. 19).

[172] M. Despeisse, B. Johansson, J. Bokrantz *et al.*, 'Battery production systems: State of the art and future developments,' in *Advances in Production Management Systems. Production Management Systems for Responsible Manufacturing, Service, and Logistics Futures*, E. Alfnes, A. Romsdal, J. O. Strandhagen, G. von Cieminski and D. Romero, Eds., Cham, Switzerland: Springer, 2023, pp. 521–535. DOI: 10.1007/978-3-031-43688-8_36 (cit. on p. 19).

[173] S. Chandra, I. J. Chung, A. Esmail, M. Blum and R. Bhandari, *Wiring system architecture*, 2022 (cit. on p. 19).

[174] K. mo Koo, X. Jiang, K. Kikuchi, A. Konno and M. Uchiyama, 'Development of a robot car wiring system,' in *2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2008, pp. 862–867. DOI: 10.1109/AIM.2008.4601774 (cit. on pp. 19, 31, 32).

[175] X. Zhang, Y. Domae, W. Wan and K. Harada, 'Learning efficient policies for picking entangled wire harnesses: An approach to industrial bin picking,' *IEEE Robotics and Automation Letters*, vol. 8, no. 1, pp. 73–80, 2023. DOI: 10.1109/LRA.2022.3222995 (cit. on pp. 19, 31, 32).

[176] E. Aguirre and B. Raucent, 'Economic comparison of wire harness assembly systems,' *Journal of Manufacturing Systems*, vol. 13, no. 4, pp. 276–288, 1994. DOI: 10.1016/0278-6125(94)90035-3 (cit. on pp. 19, 29).

[177] M. Zürn, A. Kienzlen, L. Klingel *et al.*, 'Deep learning-based instance segmentation for feature extraction of branched deformable linear objects for robotic manipulation,' in *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*, 2023, pp. 1–6. DOI: 10.1109/CASE56687.2023.10260646 (cit. on p. 19).

[178] T. Toner, V. Molazadeh, M. Saez, D. M. Tilbury and K. Barton, *Sequential manipulation of deformable linear object networks with endpoint pose measurements using adaptive model predictive control*, 2024. DOI: 10.48550/arXiv.2402.10372 (cit. on pp. 19, 21).

[179] B. Sun, F. Chen, H. Sasaki and T. Fukuda, 'Robotic wiring harness assembly system for fault-tolerant electric connectors mating,' in *2010 International Symposium on Micro-NanoMechatronics and Human Science*, 2010, pp. 202–205. DOI: 10.1109/MHS.2010.5669533 (cit. on pp. 19, 29, 31, 32).

[180] X. Jiang, Y. Nagaoka, K. Ishii, S. Abiko, T. Tsujita and M. Uchiyama, 'Robotized recognition of a wire harness utilizing tracing operation,' *Robotics and Computer-Integrated Manufacturing*, vol. 34, pp. 52–61, 2015. DOI: 10.1016/j.rcim.2014.12.002 (cit. on pp. 20, 21, 31, 32).

[181] X. Wang, X. L. Wang and D. M. Wilkes, 'An automated vision based on-line novel percept detection method for a mobile robot,' *Robotics and Autonomous Systems*, vol. 60, no. 10, pp. 1279–1294, 2012. DOI: 10.1016/j.robot.2012.06.005 (cit. on p. 20).

[182] E. Martinez-Martin and A. P. del Pobil, 'Vision for robust robot manipulation,' *Sensors*, vol. 19, no. 7, p. 1648, 2019. DOI: 10.3390/s19071648 (cit. on p. 20).

[183] M. T. Shahria, M. S. H. Sunny, M. I. I. Zarif, J. Ghommam, S. I. Ahamed and M. H. Rahman, 'A comprehensive review of vision-based robotic applications: Current state, components, approaches, barriers, and potential solutions,' *Robotics*, vol. 11, no. 6, p. 139, 2022. DOI: 10.3390/robotics11060139 (cit. on p. 20).

[184] S. Zhaole, H. Zhou, L. Nanbo, L. Chen, J. Zhu and R. B. Fisher, 'A robust deformable linear object perception pipeline in 3d: From segmentation to reconstruction,' *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 843–850, 2024. DOI: 10.1109/LRA.2023.3337695 (cit. on p. 21).

[185] K. Nottensteiner, A. Sachtler and A. Albu-Schäffer, 'Towards autonomous robotic assembly: Using combined visual and tactile sensing for adaptive task execution,' *Journal of Intelligent & Robotic Systems*, vol. 101, no. 3, p. 49, 2021. DOI: 10.1007/s10846-020-01303-z (cit. on pp. 21, 29).

[186] J. Guo, J. Zhang, Y. Gai, D. Wu and K. Chen, 'Visual recognition method for deformable wires in aircrafts assembly based on sequential segmentation and probabilisitic estimation,' in *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, vol. 6, 2022, pp. 598–603. DOI: 10.1109/ITOEC53115.2022.9734432 (cit. on pp. 21, 31, 32).

[187] F. Duan, J. T. C. Tan and T. Arai, 'A new human-robot collaboration assembly system for cellular manufacturing,' in *Proceedings of the 30th Chinese Control Conference*, 2011, pp. 5468–5473 (cit. on p. 21).

[188] N. S. Joshi, S. Singh, M. Krugh and L. Mears, 'Background noise mitigation of dual microphone system for defect detection in electrical cable connection,' *Procedia Manufacturing*, vol. 26, pp. 1287–1295, 2018. DOI: 10.1016/j.promfg.2018.07.139 (cit. on p. 21).

[189] B. D. Slife and R. N. Williams, *What's Behind the Research? Discovering Hidden Assumptions in the Behavioral Sciences.* Thousand Oaks, CA, USA: SAGE Publications, Inc., 1995. DOI: 10.4135/9781483327372 (cit. on p. 23).

[190] J. W. Creswell and J. D. Creswell, *Research design: qualitative, quantitative, and mixed methods approaches*, 6th ed. Thousand Oaks, CA, USA: SAGE Publications, Inc., 2022 (cit. on pp. 23, 24).

[191] R. Hirschheim, 'Information systems epistemology: An historical perspective,' *Research methods in information systems*, pp. 9–33, 1985 (cit. on p. 23).

[192] E. G. Guba, 'The alternative paradigm dialog,' in *The Paradigm Dialog*, SAGE Publications, Inc, 1990, pp. 17–30 (cit. on p. 23).

[193] J. W. Creswell, *Research design: qualitative, quantitative, and mixed methods approaches*, 4th ed. Thousand Oaks, CA, USA: SAGE Publications, Inc., 2014 (cit. on pp. 23, 27, 28, 40).

[194] J. Paley, 'Positivism,' in *The SAGE Encyclopedia of Qualitative Research Methods*, L. M. Given, Ed., vol. 2, Thousand Oaks, CA, USA: SAGE Publications, Inc., 2008, pp. 647–650. DOI: 10.4135/9781412963909 (cit. on p. 23).

[195] K. Williamson, F. Burstein and S. McKemmish, 'The two major traditions of research,' in *Research Methods for Students, Academics and Professionals*, ser. Topics in Australasian Library and Information Studies, K. Williamson, A. Bow, F. Burstein *et al.*, Eds., 2nd ed., Chandos Publishing, 2002, ch. 2, pp. 25–47. DOI: 10.1016/B978-1-876938-42-0.50009-5 (cit. on pp. 23, 26).

[196] J. Laird, 'Xi.-positivism, empiricism, and metaphysics,' *Proceedings of the Aristotelian Society*, vol. 39, no. 1, pp. 207–224, 1939. DOI: 10.1093/aristotelian/39.1.207 (cit. on p. 23).

[197] D. C. Phillips and N. C. Burbules, *Postpositivism and educational research.* Lanham, MD, USA: Rowman & Littlefield Publishers, Inc., 2000 (cit. on p. 23).

[198] N. J. Fox, 'Postpositivism,' in *The SAGE Encyclopedia of Qualitative Research Methods*, L. M. Given, Ed., vol. 2, Thousand Oaks, CA, USA: SAGE Publications, Inc., 2008, pp. 660–664. DOI: 10.4135/9781412963909 (cit. on p. 23).

[199] T. E. Costantino, 'Constructivism,' in *The SAGE Encyclopedia of Qualitative Research Methods*, L. M. Given, Ed., vol. 2, Thousand Oaks, CA, USA: SAGE Publications, Inc., 2008, pp. 116–120. DOI: 10.4135/9781412963909 (cit. on p. 24).

[200] E. Bell, A. Bryman and B. Harley, *Business research methods*, 6th ed. Oxford, UK: Oxford University Press, 2022 (cit. on p. 24).

[201] K. Peffers, T. Tuunanen, M. A. Rothenberger and S. Chatterjee, 'A design science research methodology for information systems research,' *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007. DOI: 10.2753/MIS0742-1222240302 (cit. on pp. 24, 25).

[202] A. Hevner and S. Chatterjee, 'Design science research in information systems,' in *Design Research in Information Systems: Theory and Practice*, ser. Integrated Series in Information Systems, vol. 22, Boston, MA, USA: Springer, 2010, ch. 2, pp. 9–22. DOI: 10.1007/978-1-4419-5653-8_2 (cit. on p. 24).

[203] J. Morse, 'Procedures and practice of mixed method design: Maintaining control, rigor, and complexity,' in *SAGE Handbook of Mixed Methods in Social & Behavioral Research*, A. Tashakkori and C. Teddlie, Eds., 2nd ed., SAGE Publications, Inc., 2010, pp. 339–352. DOI: 10.4135/9781506335193 (cit. on pp. 24, 25).

[204] S. Gregor and A. R. Hevner, 'Positioning and presenting design science research for maximum impact,' *MIS Quarterly*, vol. 37, no. 2, pp. 337–355, 2013. DOI: 10.25300/MISQ/2013/37.2.01 (cit. on p. 24).

[205] A. R. Hevner, S. T. March, J. Park and S. Ram, 'Design science in information systems research,' *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004. DOI: 10.2307/25148625 (cit. on p. 24).

[206] M. Petticrew, F. Song, P. Wilson and K. Wright, 'Quality-assessed reviews of health care interventions and the database of abstracts of reviews of effectiveness (dare),' *International Journal of Technology Assessment in Health Care*, vol. 15, no. 4, pp. 671–678, 1999. DOI: 10.1017/S0266462399015469 (cit. on pp. 27, 28, 31, 40).

[207] N. K. Denzin, *The research act: A theoretical introduction to sociological methods*, 1st ed. New York, NY, USA: Routledge, 2009. DOI: 10.4324/9781315134543 (cit. on pp. 27, 28, 30, 40).

[208] V. J. Janesick, 'Peer debriefing,' in *The Blackwell Encyclopedia of Sociology*, G. Ritzer, Ed., John Wiley & Sons, Ltd, 2015. DOI: 10.1002/9781405165518.wbeosp014.pub2 (cit. on pp. 27, 28, 40).

[209] T. Greenhalgh and R. Peacock, 'Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources,' *BMJ*, vol. 331, no. 7524, pp. 1064–1065, 2005. DOI: 10.1136/bmj.38636.593461.68 (cit. on pp. 26, 27, 31).

[210] M. W. Tracey, 'Design and development research: A model validation case,' *Educational Technology Research and Development*, vol. 57, no. 4, pp. 553–571, 2009. DOI: 10.1007/s11423-007-9075-0 (cit. on pp. 27, 28, 40).

[211] B. Kitchenham, 'Procedures for performing systematic reviews,' Keele University, Keele, Staffs, UK, Tech. Rep. TR/SE-0401, 2004 (cit. on pp. 26–28, 30).

[212] J. Rowley and F. Slack, 'Conducting a literature review,' *Management research news*, vol. 27, no. 6, pp. 31–39, 2004. DOI: 10.1108/01409170410784185 (cit. on p. 26).

[213] D. Denyer and D. Tranfield, 'Producing a systematic review,' in *The SAGE Handbook of Organizational Research Methods*, D. A. Buchanan and A. Bryman, Eds., Sage Publications Ltd., 2009, pp. 671–689 (cit. on p. 26).

[214] H. K. Mohajan, 'Two criteria for good measurements in research: Validity and reliability,' *Annals of Spiru Haret University*, vol. 17, no. 3, pp. 58–82, 2017. DOI: 10.26458/1746 (cit. on pp. 27, 40).

[215] L. Sürücü and A. Maslakci, 'Validity and reliability in quantitative research,' *Business & Management Studies: An International Journal*, vol. 8, no. 3, pp. 2694–2726, 2020. DOI: 10.15295/bmij.v8i3.1540 (cit. on pp. 27, 40).

[216] R. Adcock and D. Collier, 'Measurement validity: A shared standard for qualitative and quantitative research,' *American Political Science Review*, vol. 95, no. 3, pp. 529–546, 2001. DOI: 10.1017/S0003055401003100 (cit. on pp. 27, 40).

[217] C. M. Patino and J. C. Ferreira, 'Internal and external validity: Can you apply research study results to your patients?' *Jornal brasileiro de pneumologia*, vol. 44, no. 3, p. 183, 2018. DOI: 10.1590/S1806-37562018000000164 (cit. on pp. 27, 40).

[218] G. R. Gibbs, *Analyzing Qualitative Data*. London, UK: SAGE Publications, Ltd, 2007. DOI: 10.4135/9781849208574 (cit. on pp. 27, 28).

[219] K. Petersen, S. Vakkalanka and L. Kuzniarz, 'Guidelines for conducting systematic mapping studies in software engineering: An update,' *Information and Software Technology*, vol. 64, pp. 1–18, 2015. DOI: 10.1016/j.infsof.2015.03.007 (cit. on p. 28).

[220] A. Botta, W. de Donato, V. Persico and A. Pescapé, 'Integration of cloud computing and internet of things: A survey,' *Future Generation Computer Systems*, vol. 56, pp. 684–700, 2016. DOI: 10.1016/j.future.2015.09.021 (cit. on p. 28).

[221] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto and S. G. S. Alcalá, 'A systematic literature review of machine learning methods applied to predictive maintenance,' *Computers & Industrial Engineering*, vol. 137, p. 106 024, 2019. DOI: 10.1016/j.cie.2019.106024 (cit. on p. 28).

[222] J. Dalzochio, R. Kunst, E. Pignaton *et al.*, 'Machine learning and reasoning for predictive maintenance in industry 4.0: Current status and challenges,' *Computers in Industry*, vol. 123, p. 103 298, 2020. DOI: 10.1016/j.compind.2020.103298 (cit. on p. 28).

[223] O. Ali, M. Ally, Clutterbuck and Y. Dwivedi, 'The state of play of blockchain technology in the financial services sector: A systematic literature review,' *International Journal of Information Management*, vol. 54, p. 102 199, 2020. DOI: 10.1016/j.ijinfomgt.2020.102199 (cit. on p. 28).

[224] A. Zuiderwijk, Y.-C. Chen and F. Salem, 'Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda,' *Government Information Quarterly*, vol. 38, no. 3, p. 101 577, 2021. DOI: 10.1016/j.giq.2021.101577 (cit. on p. 28).

[225] A. F. Borges, F. J. Laurindo, M. M. Spínola, R. F. Gonçalves and C. A. Mattos, 'The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions,' *International Journal of Information Management*, vol. 57, p. 102 225, 2021. DOI: 10.1016/j.ijinfomgt.2020.102225 (cit. on p. 28).

[226] C. Collins, D. Dennehy, K. Conboy and P. Mikalef, 'Artificial intelligence in information systems research: A systematic literature review and research agenda,' *International Journal of Information Management*, vol. 60, p. 102 383, 2021. DOI: 10.1016/j.ijinfomgt.2021.102383 (cit. on p. 28).

[227] M. Haghi Kashani, M. Madanipour, M. Nikravan, P. Asghari and E. Mahdipour, 'A systematic review of iot in healthcare: Applications, techniques, and trends,' *Journal of Network and Computer Applications*, vol. 192, p. 103 164, 2021. DOI: 10.1016/j.jnca.2021.103164 (cit. on p. 28).

[228] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey and S. Linkman, 'Systematic literature reviews in software engineering - a systematic literature review,' *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, 2009. DOI: 10.1016/j.infsof.2008.09.009 (cit. on pp. 28, 40).

[229] S. Saleem, N. F. Khan, S. Zafar and N. Raza, 'Systematic literature reviews in cyberbullying/cyber harassment: A tertiary study,' *Technology in Society*, vol. 70, p. 102 055, 2022. DOI: 10.1016/j.techsoc.2022.102055 (cit. on p. 28).

[230] M. Hammersley, 'Assessing validity in social research,' in *The SAGE Handbook of Social Research Methods*, P. Alasuutari, L. Bickman and J. Brannen, Eds., London, England: SAGE Publications Ltd, 2008, ch. 4, pp. 42–53. DOI: 10.4135/9781446212165 (cit. on pp. 28, 40).

[231] J. Tilindis and V. Kleiza, 'The effect of learning factors due to low volume order fluctuations in the automotive wiring harness production,' *Procedia CIRP*, vol. 19, pp. 129–134, 2014. DOI: 10.1016/j.procir.2014.05.019 (cit. on p. 29).

[232] A. Hormann and U. Rembold, 'Development of an advanced robot for autonomous assembly,' in *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, vol. 3, 1991, pp. 2452–2457. DOI: 10.1109/ROBOT.1991.131992 (cit. on p. 29).

[233] C. Feng, Y. Xiao, A. Willette, W. McGee and V. R. Kamat, 'Vision guided autonomous robotic assembly and as-built scanning on unstructured construction sites,' *Automation in Construction*, vol. 59, pp. 128–138, 2015. DOI: 10.1016/j.autcon.2015.06.002 (cit. on p. 29).

[234] R. Ahmad and P. Plapper, 'Safe and automated assembly process using vision assisted robot manipulator,' *Procedia CIRP*, vol. 41, pp. 771–776, 2016. DOI: 10.1016/j.procir.2015.12.129 (cit. on p. 29).

[235] M. J. Page, J. E. McKenzie, P. M. Bossuyt *et al.*, 'The prisma 2020 statement: An updated guideline for reporting systematic reviews,' *BMJ*, vol. 372, no. n71, 2021. DOI: 10.1136/bmj.n71 (cit. on p. 31).

[236] P. Di, J. Huang, F. Chen, H. Sasaki and T. Fukuda, 'Hybrid vision-force guided fault tolerant robotic assembly for electric connectors,' in *2009 International Symposium on Micro-NanoMechatronics and Human Science*, 2009, pp. 86–91. DOI: 10.1109/MHS.2009.5352078 (cit. on pp. 31, 32).

[237] X. Jiang, K.-m. Koo, K. Kikuchi, A. Konno and M. Uchiyama, 'Robotized assembly of a wire harness in car production line,' in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 490–495. DOI: 10.1109/IROS.2010.5653133 (cit. on pp. 31, 32).

[238] P. Di, F. Chen, H. Sasaki, J. Huang, T. Fukuda and T. Matsuno, 'Vision-force guided monitoring for mating connectors in wiring harness assembly systems,' *Journal of Robotics and Mechatronics*, vol. 24, no. 4, pp. 666–676, 2012. DOI: 10.20965/jrm.2012.p0666 (cit. on pp. 31, 32).

[239] T. Tamada, Y. Yamakawa, T. Senoo and M. Ishikawa, 'High-speed manipulation of cable connector using a high-speed robot hand,' in *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2013, pp. 1598–1604. DOI: 10.1109/ROBIO.2013.6739695 (cit. on pp. 31, 32).

[240] H.-C. Song, Y.-L. Kim, D.-H. Lee and J.-B. Song, 'Electric connector assembly based on vision and impedance control using cable connector-feeding system,' *Journal of Mechanical Science and Technology*, vol. 31, no. 12, pp. 5997–6003, 2017. DOI: 10.1007/s12206-017-1144-7 (cit. on pp. 31, 32).

[241] F. Yumbla, M. Abeyabas, T. Luong, J.-S. Yi and H. Moon, 'Preliminary connector recognition system based on image processing for wire harness assembly tasks,' in *2020 20th International Conference on Control, Automation and Systems (ICCAS)*, 2020, pp. 1146–1150. DOI: 10.23919/ICCAS50221.2020.9268291 (cit. on pp. 31, 32).

[242] P. Kicki, M. Bednarek, P. Lembicz *et al.*, 'Tell me, what do you see?-interpretable classification of wiring harness branches with deep neural networks,' *Sensors*, vol. 21, no. 13, p. 4327, 2021. DOI: 10.3390/s21134327 (cit. on pp. 31, 32).

[243] M. Everingham, L. Van Gool, C. K. Williams, J. Winn and A. Zisserman, 'The pascal visual object classes (voc) challenge,' *International Journal of Computer Vision*, vol. 88, pp. 303–308, 2009. DOI: 10.1007/s11263-009-0275-4 (cit. on pp. 33, 37).

[244] T.-Y. Lin, M. Maire, S. Belongie *et al.*, 'Microsoft coco: Common objects in context,' in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele and T. Tuytelaars, Eds., Cham, Switzerland: Springer, 2014, pp. 740–755. DOI: 10.1007/978-3-319-10602-1_48 (cit. on pp. 33, 37).

[245] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn and A. Zisserman, 'The pascal visual object classes challenge: A retrospective,' *International Journal of Computer Vision*, vol. 111, pp. 98–136, 2015. DOI: 10.1007/s11263-014-0733-5 (cit. on pp. 33, 37).

[246] O. Russakovsky, J. Deng, H. Su *et al.*, 'Imagenet large scale visual recognition challenge,' *International Journal of Computer Vision*, vol. 115, pp. 211–252, 2015. DOI: 10.1007/s11263-015-0816-y (cit. on pp. 33, 37).

[247] S. Ren, K. He, R. Girshick and J. Sun, 'Faster r-cnn: Towards real-time object detection with region proposal networks,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017. DOI: 10.1109/TPAMI.2016.2577031 (cit. on p. 33).

[248] G. Jocher, *Yolov5 by ultralytics*, version 7.0, 2020. DOI: 10.5281/zenodo.3908559 (cit. on p. 33).

[249] M.-C. Chiu and G. l. E. Okudan, 'Evolution of design for x tools applicable to design stages: A literature review,' in *Proceedings of the ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 6, 2010, pp. 171–182. DOI: 10.1115/DETC2010-29091 (cit. on p. 38).

[250] J. C. Mankins, 'Technology readiness assessments: A retrospective,' *Acta Astronautica*, vol. 65, no. 9, pp. 1216–1223, 2009. DOI: 10.1016/j.actaastro.2009.03.058 (cit. on p. 39).

[251] B. Neyshabur, S. Bhojanapalli, D. Mcallester and N. Srebro, 'Exploring generalization in deep learning,' in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio *et al.*, Eds., vol. 30, Curran Associates, Inc., 2017 (cit. on p. 40).

[252] C. Zhang, S. Bengio, M. Hardt, B. Recht and O. Vinyals, 'Understanding deep learning (still) requires rethinking generalization,' *Communications of the ACM*, vol. 64, no. 3, pp. 107–115, 2021. DOI: 10.1145/3446776 (cit. on p. 40).

[253] K. Kawaguchi, Y. Bengio and L. Kaelbling, 'Generalization in deep learning,' in *Mathematical Aspects of Deep Learning*, P. Grohs and G. Kutyniok, Eds., Cambridge, UK: Cambridge University Press, 2022, ch. 2, pp. 112–148. DOI: 10.1017/9781009025096.003 (cit. on p. 40).

[254] B. L. Berg, *Qualitative research methods for the social sciences*, 7th ed. Boston, MA, USA: Allyn & Bacon, 2009 (cit. on p. 41).

[255] S. N. Hesse-Biber, *The Practice of Qualitative Research*, 3rd ed. Thousand Oaks, CA, USA: SAGE Publications, Inc, 2016 (cit. on p. 41).

[256] K. F. Punch, *Introduction to social research: quantitative and qualitative approaches*, 2nd ed. Thousand Oaks, CA, USA: SAGE Publications, Inc, 2005 (cit. on p. 41).

[257] J. E. Sieber, 'Planning ethically responsible research,' in *The SAGE Handbook of Applied Social Research Methods*, L. Bickman and D. J. Rog, Eds., 2nd ed., Thousand Oaks, CA, USA: SAGE Publications, Inc., 2009, ch. 4, pp. 106–142. DOI: 10.4135/9781483348858 (cit. on p. 41).

[258] R. S. Nickerson, 'Confirmation bias: A ubiquitous phenomenon in many guises,' *Review of General Psychology*, vol. 2, no. 2, pp. 175–220, 1998. DOI: 10.1037/1089-2680.2.2.175 (cit. on p. 41).

[259] W. L. Neuman, *Social research methods: qualitative and quantitative approaches*, 7th ed. Boston, MA, USA: Allyn & Bacon, 2011 (cit. on p. 41).

[260] J. Elkington, 'The triple bottom line,' in *Environmental management: Readings and cases*, M. V. Russo, Ed., 2nd ed., SAGE Publications, Inc., 1997, pp. 49–66 (cit. on p. 41).

Part II

Appended Papers

Paper 1

Overview of Computer Vision Techniques in Robotized Wire Harness Assembly: Current State and Future Opportunities

Hao Wang, Omkar Salunkhe, Walter Quadrini, Dan Lämkull, Fredrik Ore, Björn Johansson, Johan Stahre

Presented at the 56th CIRP Conference on Manufacturing Systems (CIRP CMS 2023), Cape Town, South Africa, 24-26 October 2023

Published in: *Procedia CIRP*

DOI: 10.1016/j.procir.2023.09.127

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 120 (2023) 1071–1076

www.elsevier.com/locate/procedia

56th CIRP Conference on Manufacturing Systems, CIRP CMS '23, South Africa

Overview of Computer Vision Techniques in Robotized Wire Harness Assembly: Current State and Future Opportunities

Hao Wang*[a], Omkar Salunkhe[a], Walter Quadrini[b], Dan Lämkull[c], Fredrik Ore[d], Björn Johansson[a], Johan Stahre[a]

[a]Department of Industrial and Materials Science, Chalmers University of Technology, Hörsalsvägen 7A, SE-412 96 Gothenburg, Sweden
[b]Department of Management, Economics and Industrial Engineering, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milan, Italy
[c]Department of Manufacturing Technology, Volvo Car Corporation, SE-418 78 Gothenburg, Sweden
[d]Department of Global Industrial Development, Scania CV AB, SE-151 87 Södertälje, Sweden

* Corresponding author. Tel.: +46-(0)31-772-1202. *E-mail address:* haowang@chalmers.se

Abstract

Wire harnesses are essential hardware for electronic systems in modern automotive vehicles. With a shift in the automotive industry towards electrification and autonomous driving, more and more automotive electronics are responsible for energy transmission and safety-critical functions such as maneuvering, driver assistance, and safety system. This paradigm shift places more demand on automotive wire harnesses from the safety perspective and stresses the greater importance of high-quality wire harness assembly in vehicles. However, most of the current operations of wire harness assembly are still performed manually by skilled workers, and some of the manual processes are problematic in terms of quality control and ergonomics. There is also a persistent demand in the industry to increase competitiveness and gain market share. Hence, assuring assembly quality while improving ergonomics and optimizing labor costs is desired. Robotized assembly, accomplished by robots or in human-robot collaboration, is a key enabler for fulfilling the increasingly demanding quality and safety as it enables more replicable, transparent, and comprehensible processes than completely manual operations. However, robotized assembly of wire harnesses is challenging in practical environments due to the flexibility of the deformable objects, though many preliminary automation solutions have been proposed under simplified industrial configurations. Previous research efforts have proposed the use of computer vision technology to facilitate robotized automation of wire harness assembly, enabling the robots to better perceive and manipulate the flexible wire harness. This article presents an overview of computer vision technology proposed for robotized wire harness assembly and derives research gaps that require further study to facilitate a more practical robotized assembly of wire harnesses.

Keywords: wire harness assembly; robotized assembly; computer vision; deformable linear object; collaborative robot applications

1. Introduction

The shift towards electrification and autonomous driving in the automotive industry places increasing importance on the automotive electronic system, where more and more wire harnesses are installed as they are essential for connecting automotive electronics and supporting signal transmission in the electronic system. This shift, in turn, makes the efficient, safe, and high-quality assembly of wire harnesses in vehicles critical for manufacturers to increase their competitiveness and gain more market share. Fig. 1 illustrates an example of an automotive wire harness.

However, a large proportion of current automotive wire harness assembly operations remains manual and skill-demanding, leading to potential quality problems. Some of the manual assembly processes also lead to severe ergonomic issues. Meanwhile, the automotive industry persists in a consistent requirement for productivity. Hence, it is desired to ensure assembly quality and improve ergonomics while keeping the utilization to optimal levels.

Robotized assembly has been implemented to facilitate automation in various industries. Although different automation solutions have been proposed under simplified industrial configurations [26, 7], robotized assembly of wire harnesses re-

10.1016/j.procir.2023.09.127

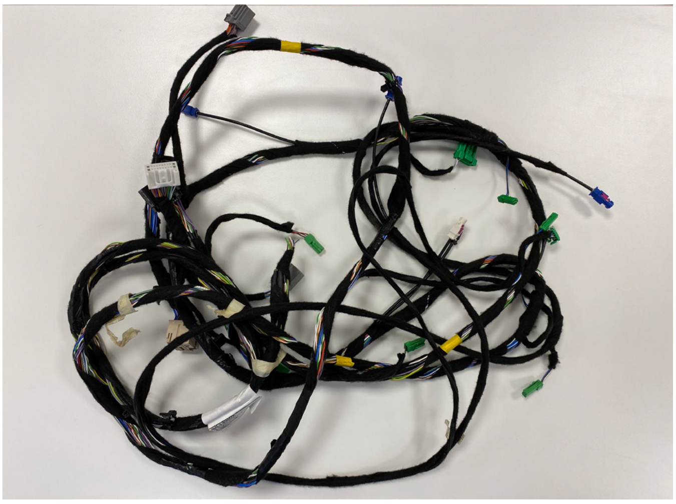


Fig. 1. An example of an automotive wire harness.

Fig. 2. The heavily-increasing length of wire harnesses in passenger cars over time. (Courtesy of Volvo Car Corporation).

mains challenging in production due to the deformability of wire harnesses and short takt time. In theory, as a special Deformable Linear Object (DLO) manipulation, robotized assembly of wire harnesses is challenging in modeling, state estimation, and operation [32, 17]. The high flexibility of DLOs also limits transplanting methods for manipulating rigid objects [23, 32] and makes it complex to design robot programs in production applications. Other practical issues on general automation, such as risk management between robots and human operators, production management, and generalization of automation for multiple product variants on the same production line, also limit the robotization of wire harness assembly.

Computer vision presents a potential to the automotive industry and computer vision combined with robots can be the game changer in solving ergonomic issues whilst increasing quality and productivity. Increasing research efforts in robotics have explored the complex DLO manipulation, addressing both theoretical research [22, 24, 28] and engineering practices [3, 7, 8, 9, 23, 4]. Although various proposals under laboratory configurations have been discussed in previous studies [3, 26, 7, 8, 2, 27, 25, 32], vision-based robotized wire harness assembly remains challenging in actual production, considering extracting image features from an intricate background [13] and fast recognition of wire harnesses for following robot operations [4]. Besides, vision-based automation solutions have not reached the robustness levels of human capabilities, especially under dynamic configurations in practical production.

This article aims at providing an overview of what type of vision systems have been proposed for robotized assembly of wire harnesses and exploring the future research directions toward a more practical vision system for robotized wire harness assembly. Based on the previous literature, this study identifies three major trends for future research, including 1) the adaptation of more advanced object detection and recognition algorithms to facilitate the robotized wire harness assembly; 2) new product design on the components of wire harnesses to facilitate the visual machine perception; and 3) the evaluation of proposed vision systems in actual production to examine the practicality and reliability.

In the rest of this article, Section 2 introduces the basics of automotive wire harnesses and the task for automation of the final assembly of wire harnesses according to current assembly workflow in production, followed by the methodology of this study in Section 3. Then, Section 4 and Section 5 describe vision systems proposed in previous research for manipulating components of wire harnesses and perceiving the structure of wire harnesses, respectively, followed by discussions in Section 6. Section 7 concludes the article with summarized findings and outlooks on future research direction. The follow-up research of this study is briefly introduced in Section 8.

2. Basics of automotive wire harnesses

A wire harness has a tree-like structure, consisting of a bundle of routed cables with various components, such as clamps and connectors, which are used to transmit current or signals within electrical equipment [1, 29]. The usage of wire harnesses keeps enlarging in modern vehicles due to the increasing number of electronic devices installed for various functions, which can be reflected in the heavily-increasing length of automotive wire harnesses installed in automobiles over time, as illustrated in Fig. 2.

Wire harnesses can be categorized according to the location of installation, for example, engine harness, instrument panel harness, floor harness, and bumper harness. Based on the observation in a car manufacturing plant, the current wire harness assembly in the passenger cabin of a passenger vehicle can be divided into (1) automated preprocessing on the packed wire harness; (2) transfer of the preprocessed wire harness into the car body using lifting equipment; (3) disentangle and route the wire harness manually; and (4) fix and connect the wire harness manually. The various heavyweights of wire harnesses and several far-reaching assembly positions pose challenges to human operators to manipulate and assemble them manually in the final assembly in terms of quality and ergonomics. In addition, several high-voltage wire harnesses, especially in an electric car, need to be handled more carefully regarding assembly quality, safety, and reliability, making automation desired and in urgent need.

Robots are widely implemented as automated solutions in manufacturing and have demonstrated tremendous potential for the automation of wire harness assembly. Besides mechanical control, a robot needs to know the position of picking and placing the target objects before operation. Considering the complexity of wire harness assembly, computer vision may facilitate a more flexible positioning and prove considerable aid in pick-n-place and assembly operations.

3. Methodology

To acquire an overview of what vision-based technology has been discussed for robotized wire harness assembly, this study first implemented an inquiry on the Scopus database with the search string, *TITLE-ABS-KEY((wir* OR cabl*) AND (harness* OR bundl*) AND assembl*)*. According to the main purpose of this study, the searching results were examined thoroughly by the co-authors to select the articles focusing on the final assembly of wire harnesses onto other products and proposing vision systems for the robotized assembly. The selected articles were then analyzed to identify the current state and future research needs. Besides, only articles in English were included in the analysis, and secondary studies, i.e., review and conference review, were excluded from the analysis.

4. Component manipulation

Some previous studies have focused on the manipulation of different components of wire harnesses, including clamps [13, 7, 8, 9] and connectors [3, 26, 2, 27, 25, 30, 32], for achieving robotized assembly of wire harnesses, as listed in Table 1.

4.1. Clamp insertion

Clamps are used for fixing wire harnesses on target locations. Four articles discussed clamp insertion onto an automobile instrument panel frame with different vision systems [13, 7, 8, 9], where CCD cameras were implemented to recognize clamps so that the end-effector on a robot arm could reach, grasp, and manipulate the detected clamps. All four studies implemented hand-eye vision systems by mounting different numbers of cameras on the end-effectors of robot arms [13, 7, 8, 9]. Two of them also adopted global vision systems with multiple cameras fixed around the operation area to support recognition, for example, avoiding occlusion [7, 8].

Koo et al. [13] first proposed to facilitate the clamp recognition and manipulation by installing cubic clamp covers with markers, whose poses were recognized by identifying the markers with SIFT [15, 16] using stereo vision systems consisting of two CCD cameras with different focal lengths mounted on the end-effectors of two robot arms. Later, Jiang et al. [7] and Jiang et al. [8] improved the clamp covers to a cylinder-like shape with more markers from ARToolKit [10, 11] and implemented a global vision system comprising ten fixed cameras with different angles surrounding the work-frame alongside hand-eye cameras to relief the occlusion problem. Furthermore, Jiang et al. [9] proposed to replace visual clamp detection with a tracing operation. Nonetheless, one wrist CMOS camera on the right robot arm remained to estimate the pose of the cover later following the similar design of pattern recognition in Jiang et al. [7] and Jiang et al. [8].

4.2. Connector mating

Mating of connectors is critical in wire harness assembly to ensure quality and functionality, which is complicated with demanding manipulation accuracy and intricate structures and non-rigid materials of connectors [26]. There are seven articles proposing various vision systems for different tasks of connector mating, including state estimation [27, 30, 32], vision-guided mating [2, 25], and fault detection [3, 26].

State estimation of the 3D geometric information is essential for robot control in vision-based robotized assembly. Tamada et al. [27] implemented a global-fixed high-speed camera to recognize the types and poses of connectors and monitor the mating process by detecting the corners of connectors with image processing at high speed. Zhou et al. [32] used three cameras, one fixed global camera and one hand-eye camera per arm of a dual-arm robot, to achieve a rough locating-then-fine positioning detection based on MobileNet-SSD [6] and CAD model registration. Instead of using 2D cameras, Yumbla et al. [30] adopted a RealSense D435 depth camera, where RGB images were first processed for connector detection with image processing methods and then the depth image was integrated to obtain the 3D geometry.

Dedicated to the vision-guided mating process, Di et al. [2] designed a scheme for monitoring relative motions between two connectors with two mutually perpendicular cameras based on basic pattern matching. Song et al. [25] proposed a pattern matching-based visual servoing for locating connector headers using a hand-eye camera and markers on connector headers.

Fault detection has also been discussed to ensure the quality of grasp and insertion so that the control system could react to problematic manipulations earlier. Di et al. [3] adopted an In-Sight 5100 camera to detect faults in grasp and insertion by checking the relative translational and rotational displacements between the gripper and the connector, which was further improved by Sun et al. [26] by adding one camera perpendicular to the first one to supplement the displacement detection.

5. Structure perception

Recently, three studies have discussed different tasks in perceiving the structure of wire harnesses, including interpretable classification of branches [12], 3D profile extraction [19], and wire recognition [4], as listed in Table 2.

Kicki et al. [12] focused on the interpretable classification of wire harness branches and proposed an RGB-D image dataset of four branches of an automotive wire harness captured by a global-fixed RealSense D435 depth camera. This research compared different networks sharing the same Downsample layer from ERFNet [21] for different data modalities and evaluated the impact of elastic transformation and pre-training with inpainting task. Saliency maps based on class activation mapping

Table 1. Vision systems in articles for manipulation on components of wire harnesses.

Component	Article	Type of cameras	Location of cameras	Number of cameras
Clamp	[13]	-	Hand-eye	4
	[7, 8]	CCD cameras	Global-fixed + Hand-eye	10 fixed + 6 on end-effectors
	[9]	Point Grey Firefly MV	Hand-eye	1
Connector	[27]	MC1362, Mikrotron	Global-fixed	1
	[30]	RealSense D435, Intel	Hand-eye	1
	[32]	Industrial cameras	Global-fixed + Hand-eye	1 fixed + 2 on robot arms
	[3]	In-Sight 5100	Global-fixed	1
	[26]	CCD cameras	Global-fixed	2
	[2]	CCD cameras	Global-fixed	2
	[25]	FL2G-13S2C-C, PGR	Hand-eye	1

Table 2. Vision systems in articles for perceiving the structure of a wire harness.

Article	Purpose	Type of cameras	Location of cameras	Number of cameras
[12]	Interpretable classification	RealSense D435, Intel	Global-fixed	1
[19]	3D profile extraction	Helios Time-of-Flight camera	Hand-eye	1
[4]	Visual recognition	RGB-D	-	-

(CAM) [31] were adopted to visualize the interpretability of the classification results.

Nguyen and Yoon [19] proposed to obtain the 3D geometry of a wire harness through clamp detection and profile tracking and correction using a hand-eye depth camera (Helios Time-of-Flight Camera) mounted on the right UR3 robot to guide the selection of next robot picking point.

Guo et al. [4] proposed an RGB-D-based segmentation and estimation for aircraft wire harness recognition, where the complete segmented wires were obtained through supervoxel over-segmentation, segmentation based on Cartesian distance, color similarity, and bending continuity, and estimation with Gaussian Mixture Model [20] on the raw point cloud data acquired by an RGB-D camera.

6. Discussion

In general, previous research efforts designed different vision systems to recognize different components of wire harnesses [13, 3, 7, 26, 8, 2, 27, 9, 25, 30, 32], estimate the state of sub-processes [3, 2, 25, 12, 4], and detect errors in assembly [3, 26] so that the control system of robotized wire harness assembly could conduct manipulations on wire harnesses and monitor the operation process according to the real-time configurations.

Previously, the most focused components of wire harnesses were clamps [13, 7, 8, 9] and connectors [3, 26, 2, 27, 30, 32, 25]. Earlier studies on clamp insertion took advantage of designed clamp covers with unique markers to facilitate detection and manipulation [13, 7, 8, 9]. However, these studies [13, 7, 8, 9] focused on the assembly of wire harnesses to an automobile instrument panel, which could be significantly different from other assembly locations that demand a different number of wire harnesses, for example, in the engine room or cabin. Clamp covers would occupy space in other installation areas with more compact installation, and post-assembly cover removal would lead to further challenges. Thus, new product designs to facilitate the visual machine perception and robotic manipulation are desired. Other studies on connectors discussed how to detect connectors [27, 30, 32] and monitor the assembly process [3, 26, 2, 25]. 2D image recognition was primarily adopted in previous research except for one that introduced depth information alongside RGB color information captured by an RGB-D camera [30]. However, the 2D RGB image was processed initially to detect connectors, and the depth information of detected connectors was simply extracted from the depth image [30].

Considering more recent studies on perceiving the structure of wire harnesses with RGB-D and ToF cameras [12, 19, 4] and the increasingly advanced and affordable imaging technology, it is promising to implement depth or 3D cameras to acquire spatial information and conduct detection by processing 3D information directly. The recent renaissance of convolutional neural networks (CNN) and the successful development of deep learning in computer vision research [14] also facilitate numerous research on learning-based 2D and 3D computer vision problems [5, 18, 33], which inspires the adaptation of latest object detection and recognition algorithms to the future computer vision-driven robotized assembly of wire harnesses to facilitate the detection, manipulation, and tracking of various components of wire harnesses.

In addition, although previous studies have demonstrated the potential of vision-based solutions for robotized wire harness assembly tasks, the practicality and reliability of the proposals in practical applications remain unknown. The actual production has a strict requirement on the production rate, which makes the operating speed and quality of the vision system critical and validation of vision systems in actual circumstances necessary. Thus, future studies on the evaluation of computer vision techniques for robotized wire harness assembly in practical manufacturing configurations are desired and necessary to assure the practicality and reliability as well as the successful integration of the vision systems in actual manufacturing.

7. Conclusion

In conclusion, previous studies have noticed and explored various computer vision-based solutions for different tasks in robotized assembly of wire harnesses, including vision-guided manipulation of different components and visual machine perception of the structure of wire harnesses. Yet, further development on robotized assembly and the enabling computer vision-based techniques are needed to promote a sound work environment, less ergonomic stress, and sustainable production, especially considering the increasing number of cables in high-tech products, e.g., cars, as depicted above. The state of the art of computer vision technology in robotized wire harness assembly was reviewed in this paper and some future research opportunities were discussed, including:

- Developing new learning-based computer vision algorithms to exploit 3D information captured by depth or 3D cameras to facilitate the detection, manipulation, and tracking in robotized wire harness assembly;
- Evaluating the practicality and reliability of vision systems in actual production to promote the integration of vision systems into practice;
- Exploring new product designs of wire harnesses to enable a more efficient component detection and manipulation without affixing additional parts.

8. Future work

In the future, vision-guided robotized manipulations on various components of wire harnesses will be proposed and evaluated in both laboratory and practical scenarios. A systematic literature review will also be conducted to better understand the state of the art of vision-guided robotized wire harness assembly and distinguish the challenges in the task.

Acknowledgements

This work was supported by the Swedish innovation agency, Vinnova, and the strategic innovation program, Produktion2030, under grant number 2022-01279. The work was carried out within Chalmers' Area of Advance Production. The support is gratefully acknowledged.

References

[1] Aguirre, E., Raucent, B., 1994. Economic comparison of wire harness assembly systems. Journal of Manufacturing Systems 13, 276–288. doi:10.1016/0278-6125(94)90035-3.

[2] Di, P., Chen, F., Sasaki, H., Huang, J., Fukuda, T., Matsuno, T., 2012. Vision-force guided monitoring for mating connectors in wiring harness assembly systems. Journal of Robotics and Mechatronics 24, 666–676. doi:10.20965/jrm.2012.p0666.

[3] Di, P., Huang, J., Chen, F., Sasaki, H., Fukuda, T., 2009. Hybrid vision-force guided fault tolerant robotic assembly for electric connectors, in: 2009 International Symposium on Micro-NanoMechatronics and Human Science, pp. 86–91. doi:10.1109/MHS.2009.5352078.

[4] Guo, J., Zhang, J., Gai, Y., Wu, D., Chen, K., 2022. Visual recognition method for deformable wires in aircrafts assembly based on sequential segmentation and probabilisitic estimation, in: 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC), pp. 598–603. doi:10.1109/ITOEC53115.2022.9734432.

[5] Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M., 2021. Deep learning for 3d point clouds: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 43, 4338–4364. doi:10.1109/TPAMI.2020.3005434.

[6] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 doi:10.48550/arXiv.1704.04861.

[7] Jiang, X., Koo, K.m., Kikuchi, K., Konno, A., Uchiyama, M., 2010. Robotized assembly of a wire harness in car production line, in: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 490–495. doi:10.1109/IROS.2010.5653133.

[8] Jiang, X., Koo, K.M., Kikuchi, K., Konno, A., Uchiyama, M., 2011. Robotized assembly of a wire harness in a car production line. Advanced Robotics 25, 473–489. URL: https://doi.org/10.1163/016918610X551782, doi:10.1163/016918610X551782, arXiv:https://doi.org/10.1163/016918610X551782.

[9] Jiang, X., Nagaoka, Y., Ishii, K., Abiko, S., Tsujita, T., Uchiyama, M., 2015. Robotized recognition of a wire harness utilizing tracing operation. Robotics and Computer-Integrated Manufacturing 34, 52–61. URL: https://www.sciencedirect.com/science/article/pii/S0736584514001069, doi:10.1016/j.rcim.2014.12.002.

[10] Kato, H., 1999. Artoolkit. URL: http://www.hitl.washington.edu/artoolkit.html. accessed: 2023-01-08.

[11] Kato, H., Billinghurst, M., 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system, in: Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99), pp. 85–94. doi:10.1109/IWAR.1999.803809.

[12] Kicki, P., Bednarek, M., Lembicz, P., Mierzwiak, G., Szymko, A., Kraft, M., Walas, K., 2021. Tell me, what do you see?—interpretable classification of wiring harness branches with deep neural networks. Sensors 21. URL: https://www.mdpi.com/1424-8220/21/13/4327, doi:10.3390/s21134327.

[13] mo Koo, K., Jiang, X., Kikuchi, K., Konno, A., Uchiyama, M., 2008. Development of a robot car wiring system, in: 2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 862–867. doi:10.1109/AIM.2008.4601774.

[14] LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. nature 521, 436–444. doi:10.1038/nature14539.

[15] Lowe, D., 1999. Object recognition from local scale-invariant features, in: Proceedings of the Seventh IEEE International Conference on Computer Vision, pp. 1150–1157 vol.2. doi:10.1109/ICCV.1999.790410.

[16] Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International journal of computer vision 60, 91–110. doi:10.1023/B:VISI.0000029664.99615.94.

[17] Lv, N., Liu, J., Jia, Y., 2022. Dynamic modeling and control of deformable linear objects for single-arm and dual-arm robot manipulations. IEEE Transactions on Robotics 38, 2341–2353. doi:10.1109/TRO.2021.3139838.

[18] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2022. Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 44, 3523–3542. doi:10.1109/TPAMI.2021.3059968.

[19] Nguyen, T.P., Yoon, J., 2021. A novel vision-based method for 3d profile extraction of wire harness in robotized assembly process. Journal of Manufacturing Systems 61, 365–374. URL: https://www.sciencedirect.com/science/article/pii/S0278612521002089, doi:10.1016/j.

jmsy.2021.10.003.

[20] Reynolds, D.A., 2009. Gaussian mixture models. Encyclopedia of biometrics 741. doi:10.1007/978-1-4899-7488-4_196.

[21] Romera, E., Álvarez, J.M., Bergasa, L.M., Arroyo, R., 2018. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. IEEE Transactions on Intelligent Transportation Systems 19, 263–272. doi:10.1109/TITS.2017.2750080.

[22] Saha, M., Isto, P., 2006. Motion planning for robotic manipulation of deformable linear objects, in: Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006., pp. 2478–2484. doi:10.1109/ROBOT.2006.1642074.

[23] Sanchez, J., Corrales, J.A., Bouzgarrou, B.C., Mezouar, Y., 2018. Robotic manipulation and sensing of deformable objects in domestic and industrial applications: a survey. The International Journal of Robotics Research 37, 688–716. doi:10.1177/0278364918779698.

[24] Schulman, J., Lee, A., Ho, J., Abbeel, P., 2013. Tracking deformable objects with point clouds, in: 2013 IEEE International Conference on Robotics and Automation, pp. 1130–1137. doi:10.1109/ICRA.2013.6630714.

[25] Song, H.C., Kim, Y.L., Lee, D.H., Song, J.B., 2017. Electric connector assembly based on vision and impedance control using cable connector-feeding system. Journal of Mechanical Science and Technology 31, 5997–6003. doi:10.1007/s12206-017-1144-7.

[26] Sun, B., Chen, F., Sasaki, H., Fukuda, T., 2010. Robotic wiring harness assembly system for fault-tolerant electric connectors mating, in: 2010 International Symposium on Micro-NanoMechatronics and Human Science, pp. 202–205. doi:10.1109/MHS.2010.5669533.

[27] Tamada, T., Yamakawa, Y., Senoo, T., Ishikawa, M., 2013. High-speed manipulation of cable connector using a high-speed robot hand, in: 2013 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1598–1604. doi:10.1109/ROBIO.2013.6739695.

[28] Tang, T., Wang, C., Tomizuka, M., 2018. A framework for manipulating deformable linear objects by coherent point drift. IEEE Robotics and Automation Letters 3, 3426–3433. doi:10.1109/LRA.2018.2852770.

[29] Tilindis, J., Kleiza, V., 2014. The effect of learning factors due to low volume order fluctuations in the automotive wiring harness production. Procedia CIRP 19, 129–134. doi:10.1016/j.procir.2014.05.019.

[30] Yumbla, F., Abeyabas, M., Luong, T., Yi, J.S., Moon, H., 2020. Preliminary connector recognition system based on image processing for wire harness assembly tasks, in: 2020 20th International Conference on Control, Automation and Systems (ICCAS), pp. 1146–1150. doi:10.23919/ICCAS50221.2020.9268291.

[31] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2921–2929. doi:10.1109/CVPR.2016.319.

[32] Zhou, H., Li, S., Lu, Q., Qian, J., 2020. A practical solution to deformable linear object manipulation: A case study on cable harness connection, in: 2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM), pp. 329–333. doi:10.1109/ICARM49381.2020.9195380.

[33] Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J., 2023. Object detection in 20 years: A survey. Proceedings of the IEEE 111, 257–276. doi:10.1109/JPROC.2023.3238524.

Paper 2

A Systematic Literature Review of Computer Vision Applications in Robotized Wire Harness Assembly

Hao Wang, Omkar Salunkhe, Walter Quadrini, Dan Lämkull, Fredrik Ore, Mélanie Despeisse, Luca Fumagalli, Johan Stahre, Björn Johansson

Under the second revision by *Advanced Engineering Informatics*, April 2024

A Systematic Literature Review of Computer Vision Applications in Robotized Wire Harness Assembly

Hao Wang[a,*], Omkar Salunkhe[a], Walter Quadrini[b], Dan Lämkull[c], Fredrik Ore[d], Mélanie Despeisse[a], Luca Fumagalli[b], Johan Stahre[a], Björn Johansson[a]

[a]*Department of Industrial and Materials Science, Chalmers University of Technology, Hörsalsvägen 7A, Gothenburg, SE-412 96, Sweden*
[b]*Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, 20133, Italy*
[c]*Department of Innovation and Strategy, Volvo Car Corporation, Gothenburg, SE-418 78, Sweden*
[d]*Department of Global Industrial Development, Scania CV AB, Södertälje, SE-151 87, Sweden*

Abstract

In the current automotive industry, human operators perform wire harness assembly manually, which causes significant quality, productivity, safety, and ergonomic problems. Robotic assembly is a critical facilitator in addressing these problems. However, it remains challenging to implement for robotizing the assembly of wire harnesses. Wire harness assembly is a specific scenario of deformable linear object manipulation. Robotizing this assembly task demands robots to flexibly adapt their actions to the dynamically changing industrial environment based on robotic perception results. Existing research suggested the significance of robotic visual perception in the robotic assembly of wire harnesses. Implementing computer vision techniques is fundamental to enabling robots' visual perception capabilities. Nonetheless, the industry has yet to introduce vision-based solutions to robotize wire harness assembly fully or partially. Through a systematic literature review, this article identifies 15 scientific publications in vision-based robotized wire harness assembly. The results show various computer vision applications regarding wire harness components and assembly operations studied in previous research. Nevertheless, this article recognizes two significant challenges for computer vision applications in robotized wire harness assembly: 1) fulfilling production requirements on robustness and practicality and 2) exploiting the intrinsic physical features of wire harnesses for visual recognition. This article also advocated five prospective research directions toward more efficient and practical vision-based robotized wire harness assembly: 1) developing learning-based vision systems to exploit intrinsic features and multi-modality data of wire harnesses; 2) adapting vision systems proposed for robotizing assembly operations in manufacturing wire harnesses; 3) assessing the practicality, robustness, reliability, and sustainability of vision systems; 4) inquiring vision-based human-robot collaboration; and 5) exploring new product designs for facilitating visual recognition.

Keywords:
Wire harness assembly, Robotic assembly, Computer vision, Human-robot collaboration, Deformable linear object, Electric vehicle

1. Introduction

The assembly of wire harnesses in the final assembly station in the automotive industry has been performed entirely manually by skilled human operators over time. Industry has identified this manual assembly as one of the bottlenecks constraining the promotion of automobile production remarkably [1]. The manual assembly operations also jeopardize assembly quality, safety, and ergonomics in the contemporary automotive industry [2, 3]. Significantly, the rapidly growing demand for electric vehicles (EV) exacerbates the production problems amid the massive industrial and social transformation toward autonomous driving, electrification, and green transportation [4–6]. Implementing automation, particularly robotic assembly, is one of the prominent approaches to address these problems in actual productions [7–9].

*Corresponding author. Tel.: +46-(0)31-772-1202. *E-mail address*: haowang@chalmers.se

The robotic assembly has been implemented to facilitate automation in various industries, accomplished by robots solely or human-robot collaboration (HRC) [10, 11]. Robotic assembly is favored to fulfill the increasingly demanding wire harness assembly in the automotive industry [3]. The robotic assembly enables more rigorous, safer, and more ergonomic-friendly manufacturing than manual operations due to its better replicability, transparency, and comprehensibility [12]. Various preliminary automation solutions for wire harness assembly were proposed under simplified industrial configurations decades ago [13]. The past few years have also witnessed studies proposing various robotized solutions for wire harness assembly in different industrial sectors, e.g., electronics, automobile, and aviation [3]. However, a low level and limited scale of automation on wire harness assembly in actual automobile production is observed, and robotic assembly of wire harnesses remains arduous to achieve in actual circumstances [14].

The conventional robotic automation solutions struggled with the task of wire harness assembly. The challenges stem from the high complexity of the assembly tasks, regarding, e.g., the numerous variants of wire harnesses to be assembled, the complex object deformation and dynamics [15], and the limited process time in actual production [16]. Before assembling wire harnesses robotically, robots require more advanced perception capability to recognize wire harness components' positions and orientations and track wire harnesses' movement and deformation. Vision is instrumental for object localization, classification, and tracking [17].

Research in computer vision and machine vision has demonstrated to the automotive industry the potential to facilitate robotized wire harness assembly. Visual inputs embed vast information of the outer environment [18]. Research on robotic manipulation reveals the significance of vision systems for robotic perception [19]. Moreover, numerous studies have investigated diverse computer vision applications in different manufacturing scenarios [20]. There has also been discussion on facilitating better robotic visual perception and manipulation of wire harnesses with computer vision techniques in previous research [21]. Nonetheless, no such practical automated solutions for the final assembly of wire harnesses have yet to be witnessed in the actual production [14]. The overall robotic assembly of automotive wire harnesses enabled by vision systems remains unsolved. The full potential of vision-based solutions in the robotized wire harness assembly remains unrevealed. Hence, it is necessary to recognize existing challenges and potential remedies for developing computer vision-enabled robotized wire harness assembly.

This study aimed to summarize the state-of-the-art research on applying computer vision techniques to facilitate the robotic assembly of wire harnesses, identify the challenges for vision systems, and propose future research directions for developing a more practical vision-based robotic assembly of wire harnesses. Regarding these aims, this study intended to answer the following research questions (RQ) through a systematic literature review:

1. What computer vision-based solutions have been proposed for robotized wire harness assembly?
2. What are the challenges for computer vision applications in robotized wire harness assembly?
3. What are the required future research activities and fields for developing more efficient and practical computer vision-based robotized wire harness assembly?

For clarification, the "wire harness assembly" discussed in this article is defined as the final installation of wire harnesses onto other products, e.g., installing wire harnesses onto electric vehicles in the final assembly of automobiles, instead of manufacturing wire harnesses, which has been reviewed in other studies [22–24].

This article is organized as follows: Section 1 introduces the systematic literature review's background and research questions. Section 2 describes the need for robotic assembly, the related research in robotic manipulation of deformable linear objects (DLO), the current assembly operations of automotive wire harnesses, and the challenges of automating the assembly. Section 3 describes the methodology implemented for the systematic literature review. Section 4 summarizes the latest advances in computer vision techniques implemented in robotized wire harness assembly, followed by discussions on current challenges and opportunities for future studies in Section 5. Section 6 concludes this article with an outlook of future trends and research.

2. Robotized automotive wire harnesses assembly

2.1. Why is robotic assembly needed for wire harness assembly?

It is essential to guarantee a high-quality assembly of wire harnesses in automobiles. A wire harness is a bundle of routed cables and wires with a tree-like structure, consisting of numerous components, e.g., wires, terminals, connectors, clamps, and wrapping materials [25]. Figure 1 presents an example of the floor harnesses to be assembled into passenger cabins. Wire harnesses are distributed extensively in modern automobiles, as the electrical infrastructure of a *Volvo XC40 Recharge* illustrated in Figure 2, They are fundamental elements within automotive electronic systems responsible for quality-essential functions (e.g., engine control unit and energy transmission system) and safety-critical processes (e.g., maneuvering, driver assistance, and safety system) [23]. Through correctly connected wire harnesses, signals and the current of electricity are transmitted among different electrical components scattered throughout electrical equipment to enable the overall system to function properly [25, 26]. However, the fully manual assembly of wire harnesses into vehicles causes problems regarding assembly quality due to the inevitable inconsistency of manual operation quality [2, 27].

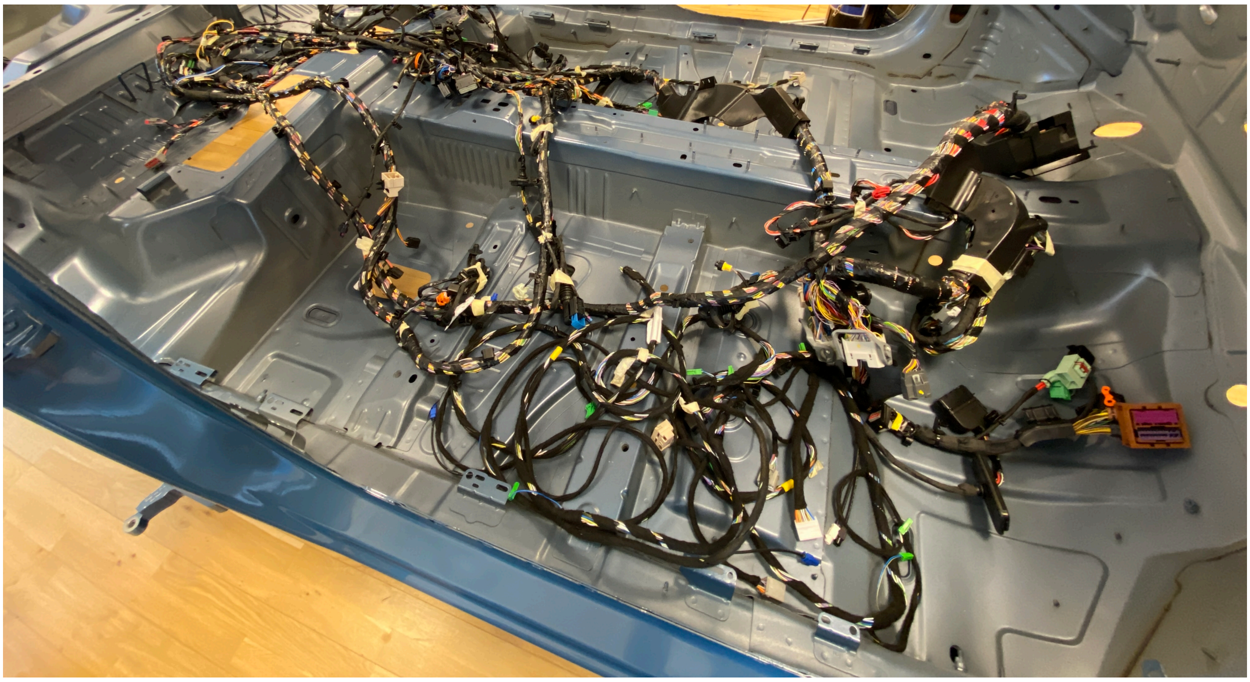


Figure 1: An example of the floor harnesses to be installed into passenger cabins of automobiles. (Courtesy of Volvo Car Corporation) **Color should be used for this figure in print and size should be calibrated for the camera-ready version (This sentence will be removed for the final submission).**

The efficiency of wire harness assembly is also instrumental in production, considering the increasing number of wire harnesses installed in modern vehicles within limited cycle time. The usage of wire harnesses in modern vehicles has been enlarging remarkably over time [1]. Figure 3 illustrates an example of increasing usage of wire harnesses from the sector of passenger vehicles. Industry also anticipates the continuous increment of wire harnesses installed in future automobiles [28], especially considering the growing number of electronic devices installed for various functions and the shift toward autonomous driving, electrification, and more sustainable mobility in the automotive industry [5, 6]. Meanwhile, the automotive industry persists in a continuous demand to promote competitiveness and acquire market share, which lays a consistent requirement on promoting productivity. The manual assembly, though, constrains the promotion of the overall productivity [7].

Moreover, it is crucial to ensure human operators' safety and improve the wire harness assembly's ergonomics. A large proportion of the manual operations in current wire harness assembly are skill-demanding and not ergonomic for human operators, e.g., heavy lifting (e.g., approximately 40 kg for some automotive wire harnesses), high-pressure pressing, and far-reaching operation [3]. These manual operations cause musculoskeletal disorders (MSD) and occupational safety and health (OSH) issues in the workforce [22]. Adapting exoskeleton or other powered mechanics can enhance the physical strength of human operators [29, 30]. However, the assembly problems due to manual operations cannot be addressed effectively. In addition, high-voltage wire harnesses are installed in automobiles, especially in electric vehicles. These components demand meticulous material handling regarding safety, assembly

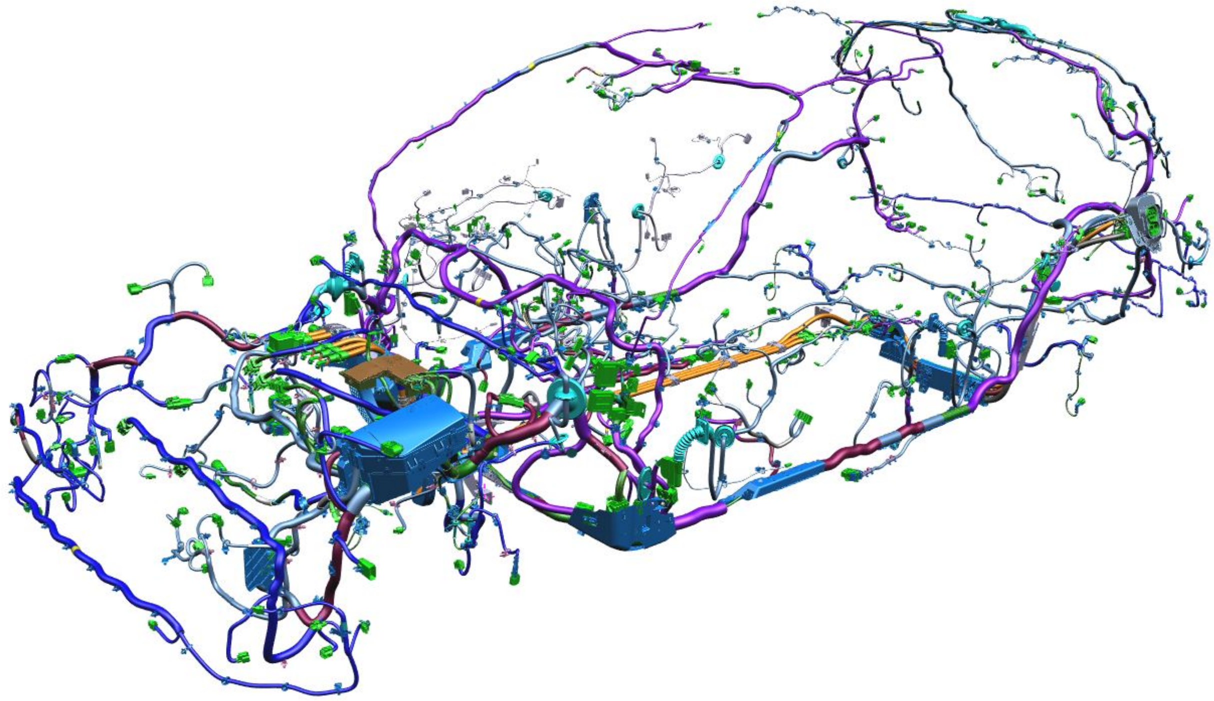


Figure 2: The electrical infrastructure of a Volvo XC40 Recharge, consisting of front, central, cabin, and rear cable systems. (Courtesy of Volvo Car Corporation) **Color should be used for this figure in print and size should be calibrated for the camera-ready version (This sentence will be removed for the final submission).**

quality, and reliability [31, 32]. Therefore, it is desired to ensure assembly quality and safety, promote productivity, improve ergonomics, and optimize resource utilization. Implementing automation and robotic assembly is a prominent approach [7–9].

2.2. Robotic manipulation of deformable linear objects

Wire harnesses can be theoretically generalized as deformable linear objects (DLO) [14, 33, 34]. Wire harness assembly, thus, can be regarded as a specific task of DLO manipulation [13].

The robotics community has extensively devoted to addressing robotic manipulation of deformable objects [35–38], also called non-rigid objects [14, 39] or flexible materials [13]. Deformable object manipulation widely exists in various application scenarios [13], e.g., manufacturing industry [40], food industry [41], medical surgery [42], and daily activity [43].

Deformable objects with “one dimension significantly larger than the other two” are defined as DLOs [35], e.g., ropes and cables. Deformable linear objects are also called deformable one-dimensional objects (DOO) [44, 45]. DLO manipulation has also been a significant concern in industry over time [13, 14, 35], e.g., the wire insertion in the electrical industry [46, 47] and the assembly of cables in the automotive industry [24, 48].

Robotic manipulation of DLOs involves various robotic tasks, e.g., modeling, perception, and manipulation [35, 36, 39]. The robotics community has been continuously investigating different models of deformable objects and the integration of various sensors and artificial intelligence (AI) into robots to equip robots with fast, accurate, and multi-modal perception abilities [49] and adaptive modeling and control abilities [36]. As a prerequisite for accomplishing complex manipulation tasks, robotic perception is critical and required to perceive DLOs’ shape, topology, deformation, and other physical properties before and during the robotic manipulation [35, 37, 49]. Robotic DLO manipulation essentially involves visual and tactile perception through visual, sound, force, tactile, and range sensing, independently or jointly [14, 36, 37, 39]. Mainly in existing research, visual perception was employed to obtain global information

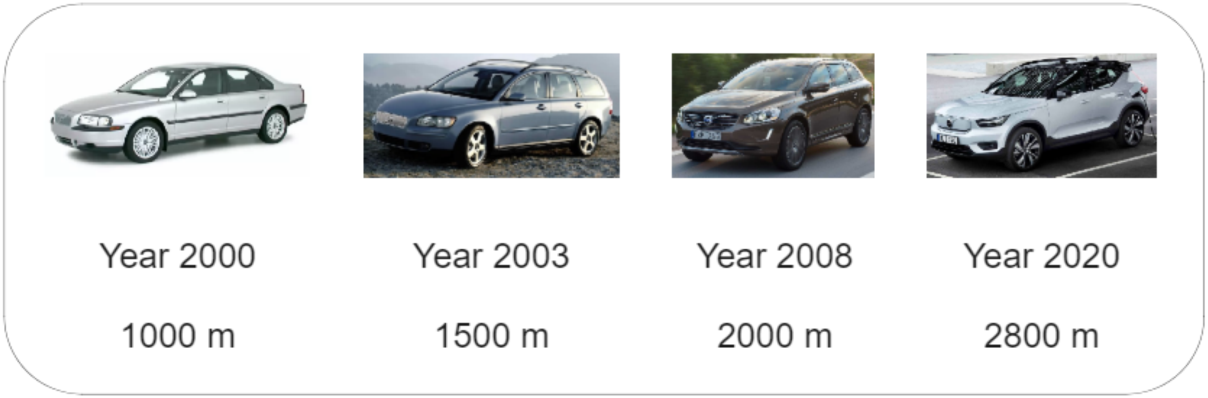


Figure 3: The remarkably increasing length of wires in passenger cars over time. (Courtesy of Volvo Car Corporation) **Color should be used for this figure in print and size should be calibrated for the camera-ready version (This sentence will be removed for the final submission).**

about DLOs' shapes on a large scale, especially when objects exhibit deformation on a large scale [50, 51]. The tactile perception was often involved in acquiring shape and contact information on the local level [37].

Regardless of the tremendous advancement of robotics, DLO manipulation remains challenging in the robot-centered flexible automation [52, 53]. In particular, challenges in DLO manipulation remain in object detection, object modeling, deformation state estimation, and robotic operation [2, 35–37, 53]. Additionally, strategies designed for manipulating regular rigid objects cannot be adapted for DLO manipulation directly due to the high degrees of freedom and deformability of DLOs [27, 35, 52]. These challenges call for further research and development in both theory and application aspects.

2.3. General automation challenges in manufacturing

The third industrial revolution initiated the broad adoption of automation in various sectors of industry [11]. However, many challenges still hamper the scale-up of automation applications in manufacturing.

Safety is fundamental in manufacturing, especially when robots are deployed in the production line [54–56]. Typically, different devices, e.g., steel fences and laser curtains, are installed around the working area of robots to guarantee safety by keeping human operators at a safe distance from functioning robots [54, 55]. Introducing new robots demands comprehensive re-consideration of the interaction between humans and robots and re-design of the workspace. This will further pose new safety and risk management challenges within the existing system.

Besides, some production systems are non-stop, e.g., the final assembly line in the automotive industry [10]. This indicates the necessity of automation systems handling the assembly during the movement of products. Thus, it is inevitable that the development of automation systems in such scenarios will consider how to synchronize the robotic manipulators with the moving assembly line while executing the assembly operations.

Moreover, multiple variants of products are commonly produced on the same production line. This increases the complexity of the design of automation control systems. The multiple variants also challenge the adaptiveness and agility of automation systems regarding different product variants. Besides, the requirement for automation in actual production varies among different sub-sections within an industry. For example, in the automotive industry, passenger vehicle production differs from that of heavy vehicles in terms of the required production rate and assembly environment. Diverse production requirements demand heterogeneous automation solutions and increase the workload of designing automation solutions.

In addition, a never-ending challenge in manufacturing is fulfilling the demand on takt time and maintaining and improving productivity. This everlasting task requires automation systems to operate reliably within a limited time.

2.4. Automation challenges in wire harness assembly

Automated assembly has been adopted in the automotive industry for years to fulfill the continuously increasing production requirements [57, 58]. However, among other assembly operations, most of the operations of wire harness assembly in current production remain manual and challenging to automate [3].

Challenges in the robotic manipulation of DLOs exist in automating wire harness assembly. Like robotic DLO manipulation, automating the assembly of wire harnesses is challenging due to the flexibility of wire harnesses. It is complicated to recognize the long and uneven shape, estimate the state of a deformable wire harness, and control the force for manipulation [59, 60]. Usually, it is also complex to design the moving paths of robots to avoid the formation of knots and entanglements that could block the assembly process or even break the harness [60]. Besides the deformability of the cables, the complex structures and non-rigid materials of other wire harness components exacerbate the challenge of robotic manipulation of wire harnesses [61]. The requirement from the actual production for extremely tight position accuracy in some assembly scenarios and the availability of precise contactless measurement to the state of the target wire in real-time further makes many proposals challenging to implement and unreliable in actual production [62].

Furthermore, wire harness assembly is more complex than generic robotic manipulation of DLOs. Considering the tree-like structure of wire harnesses, wire harnesses consist of a bundle of DLOs (e.g., wires) and a group of rigid objects (e.g., connectors and clamps) [25]. With this, previous research has further considered wire harnesses as semi-deformable linear objects [52] or branched deformable linear objects [27]. Hence, the interaction and constraint among different wire harness branches demand further investigation.

2.5. Current manual assembly of automotive wire harnesses

Based on current work instructions at Volvo Car Corporation and empirical data collected during visual inspection of a production line in a car manufacturing plant, the current manual assembly of automotive wire harnesses in the passenger cabins of automobiles can be summarized into the following five procedures: 1) prepare, 2) transport, 3) untangle, 4) route, and 5) assemble.

Prepare Initially, the wire harnesses arrive at the assembly station tied and packed in a plastic bag or box. The wire harness is too stiff to manipulate by human operators in later assembly processes manually. Therefore, the pack of wire harnesses is warmed in an oven first to soften them so that human operators can manipulate them manually. Since the wire harnesses arrive as tied and packed in sequence, and the positions of wire harness delivery and the oven are fixed, this procedure could be automated by implementing automated conveyors and conventional industrial robots.

Transport After getting warmed, the wire harness is transported by a lifting machine operated by a human operator and dumped in the cabin. In this stage, the wire harness remains tied in a whole chunk. Therefore, conventional industrial robots could be programmed to pick up the warmed wire harnesses from the oven and release them into the vehicle body.

Untangle After placing the wire harness, human operators bend into the car body to untie and untangle the wire harness manually. To automate this procedure, robots require fast and accurate perception of various properties of the wire harnesses, such as shape and topology, before they start manipulating them. During manipulation, robots need to track the deformation and adapt their control flexibly and promptly, which requires more intelligent robotic perception and control capabilities.

Route After disentangling, a wire harness is routed in the car body manually so that different branches of a wire harness reach the mating area based on functionality. Similar to automating the untangling procedure, robots require the capability to manipulate DLOs before robotizing this procedure.

Assemble Lastly, human operators manually mate the clamps and connectors of the wire harness to the counterparts in the car. Robotizing this procedure requires robotic perception of the positions and orientations of manipulating objects and more advanced robotic planning and control capability.

3. Methodology

The systematic literature review is important for comprehensively understanding a subject's state of the art and identifying the gaps requiring future research [63–65]. The systematic literature review in this study followed the methodology for planning and conducting a review suggested by [63]. The co-authors of this article also followed the methodology of investigator triangulation [66], peer debriefing [67], and expert review [68] to strengthen the research quality. A review protocol was developed first to ensure a systematic and reproducible review method, as shown in Table 1.

Table 1: Review protocol for a systematic literature review on computer vision applications in the robotized wire harness assembly.

Review criteria	
Database	*Scopus*
Search string	(wir* OR cabl*) AND (harness* OR bundl*) AND assembl*
Search field	Article title, Abstract, Keywords
Subject area	Engineering; Computer Science; Multidisciplinary; Business, Management and Accounting; Decision Sciences
Article language	English
Inclusion criteria	Proposing computer vision-based algorithm and/or technology for robotized wire harness assembly
Exclusion criteria	Not about wire harnesses; not about robotic assembly; about the manufacturing of wire harnesses
Search date	September 6, 2023

3.1. Literature search

The literature search was conducted on the *Scopus* database. *Scopus* was selected as the online database for searching scientific peer-reviewed articles, considering it a de facto reference standard for the engineering community [69] and its adequate coverage of publications provided by various publishers. It is preferred because of its higher inclusiveness in terms of contributions over the *Web of Science* database [70] and its higher reliability of collected peer-reviewed sources over the *Google Scholar* database [71].

Three researchers deliberated the keywords and the string for searching and the inclusion and exclusion criteria for scrutinizing to ensure the identification of as many relevant articles as possible. After several search trials with different combinations of keywords, the following string was defined for the search within the field of *Article title*, *Abstract*, *Keywords* on *Scopus*: (wir* OR cabl*) AND (harness* OR bundl*) AND assembl*. The asterisk character was used in the string to retrieve more results based on term variations. The words “cable” and “bundle” were included as synonyms for “wire” and “harness”.

The initial search returned a set of 1022 articles. A filter on the subject area was conducted on the initial search to limit the subject areas to *Engineering*, *Computer Science*, *Decision Sciences*, *Multidisciplinary*, and *Business, Management and Accounting*. The intention was to exclude studies referring to irrelevant subjects, which reduced the number of items to 728. The language of the article was also limited to English. In addition, no filter regarding the year of publication was implemented, i.e., all past research works were kept for screening. Finally, 662 articles were identified on September 6, 2023.

3.2. Literature selection

After the literature search, a two-step screening was conducted by three researchers regarding the inclusion and exclusion criteria shown in Table 1 to select the literature for later data synthesis and analysis. The article selection process is reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [72] in Figure 4.

Following the selection criteria in Table 1, a study would be excluded if it did not regard wire harnesses, as described in Section 2.1, as the object of interest in the study. A study would also be excluded if no robot system was involved in the proposed solution for wire harness assembly. Then, as clarified at the end of Section 1, articles that proposed computer vision-based solutions for robotizing wire harness assembly would be qualified for the analysis in this study, while articles presenting studies that focused on manufacturing wire harnesses would be excluded from the analysis.

In the first round of screening, the title and abstract of the 662 articles were initially examined by three researchers independently based on the inclusion and exclusion criteria to minimize subjective bias during screening. After screening individually, all three researchers synchronized their opinions and agreed on disagreements. Firstly, 452 articles were excluded because they were about something other than the wire harnesses focused in this review. Then, 93 articles were excluded because no robotic assembly was involved. Furthermore, 83 articles were excluded because they addressed the robotized assembly process in manufacturing wire harnesses. Lastly, 12 articles were excluded because they addressed robotized wire harness assembly without proposing vision-based solutions. After the first screening on title and abstract, 640 articles were excluded; thus, 22 articles qualified, whose full texts were downloaded and evaluated in the second round of screening.

In the second round of screening, the full texts of downloaded articles were scrutinized meticulously according to the inclusion and exclusion criteria to filter the articles for later analysis, which resulted in 9 records being further

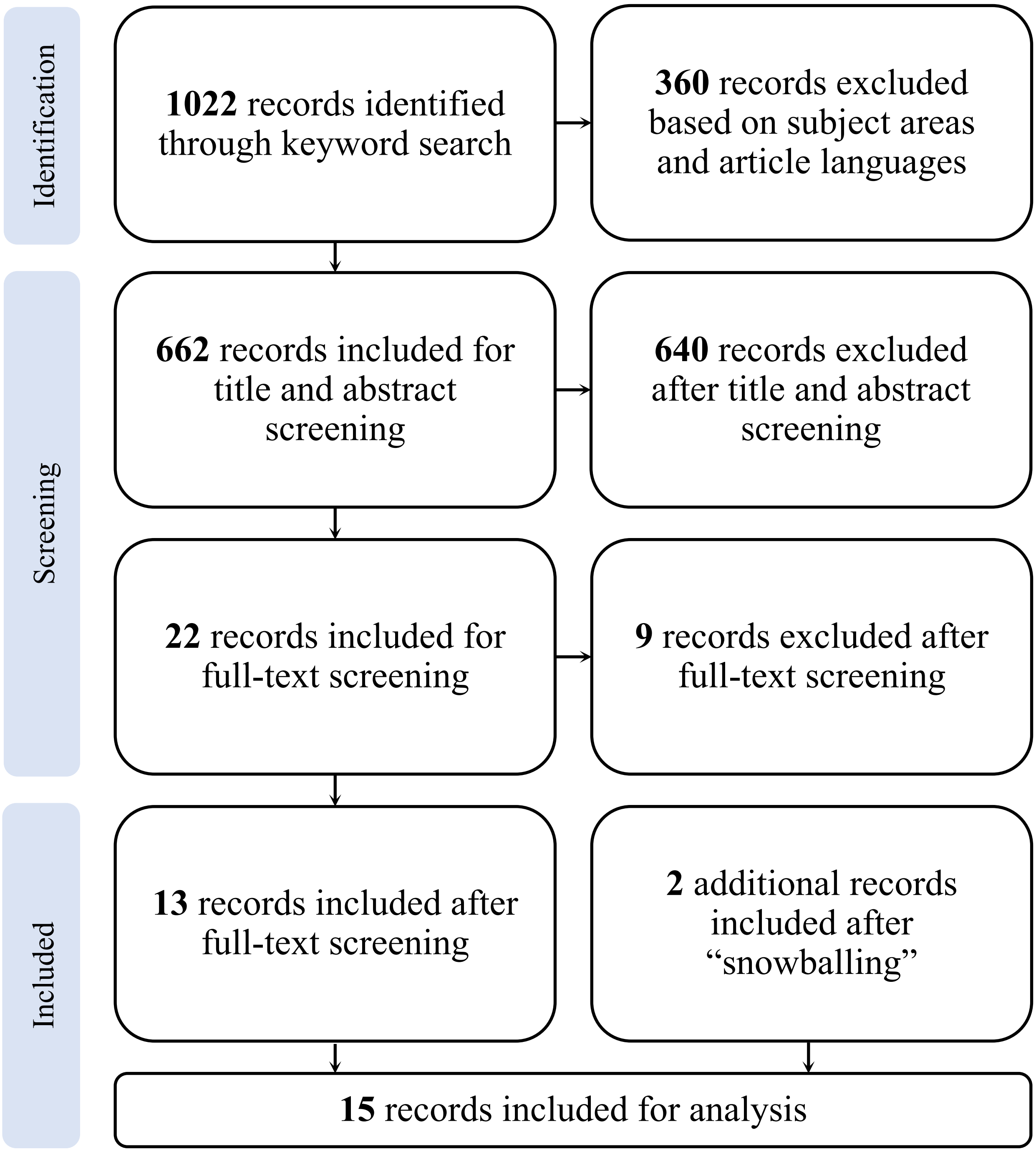


Figure 4: PRISMA [72] flow diagram of review process. **Color should be used for this figure in print and size should be calibrated for the camera-ready version (This sentence will be removed for the final submission).**

excluded due to failing to fulfill different criteria. Hence, after the second round of screening, there were 13 articles left for full-text analysis.

Moreover, "snowballing" [73], including *reference tracking* and *citation tracking*, was implemented on the selected 13 articles for analysis on *Scopus* to identify other relevant articles missed in the original search, which returned 2 more articles. Finally, 15 peer-reviewed scientific articles were identified for further extensive, qualitative, descriptive, and quantitative data synthesis and analysis.

3.3. Research quality and limitations

There are two significant perspectives for research quality evaluation: validity and reliability [74–78]. Validity suggests 1) to what extent the study prevents systematic errors appearing in designing and conducting the research (internal validity) and 2) to what extent the research findings can be generalized and applied to similar populations outside the study (external validity) [63, 75, 77]. Reliability measures the consistency of the usage of research approaches in similar studies [79]. A publicly available and widely adopted systematic review methodology, [63], was followed to strengthen the quality of this systematic review. This systematic review also referred to the Database of Abstracts of Reviews of Effects (DARE) criteria[1] from the University of York, Centre for Reviews and Dissemination [80] as a guideline to self-check the quality of this systematic review. The DARE criteria [80] qualifies a systematic review that addresses the first three criteria and at least one of the fourth and the fifth criteria:

1. The existence of appropriate inclusion/exclusion criteria description
2. The adequate search on relevant literature
3. The existence of study synthesis
4. The existence of quality assessment on the included studies
5. The existence of sufficient details of the included studies in the review

Researcher bias is generally unavoidable in qualitative studies. Researchers' backgrounds and prior knowledge inevitably influence the objectiveness of the research. However, this systematic review required prior computer vision, robotics, automation, and production knowledge. It was used as a positive driver to design the research methods, analyze and interpret the findings, and propose high-quality future research directions. Nevertheless, as suggested in [63], a review protocol, including the methods for literature searching, screening, and selection, was calibrated beforehand to minimize the impact of researcher bias and enhance the objectiveness of this review. Investigator triangulation [66], peer debriefing [67], and expert review [68] were also adopted through the study, where experts in relevant subjects from academia and industry (some of them as co-authors) were involved in calibrating the research methods and cross-validate the findings and interpretation.

The database is another crucial aspect of systematic review and literature search. There have already been various bibliometric databases, e.g., *Web of Science*, *Scopus*, *Google Scholar*, *Microsoft Academic*, and *Dimensions* [71], which makes traversing all databases for a systematic review arduous. Thus, database selection is necessary to improve operability while guaranteeing the adequacy of the search for relevant literature. Though gaps exist among databases' coverage, *Scopus* was selected in this systematic review, considering the consistency within the engineering community [69], its higher inclusiveness over *Web of Science* [70], and its higher reliable collection of peer-reviewed articles over *Google Scholar* [71]. This systematic review also adopted "snowballing" [73] to mitigate the impact and complement the search result.

The identified studies are summarized, synthesized, and analyzed in the following sections to fulfill the rest of DARE criteria [80].

4. Results

After searching and screening following the methodology described in Section 3, 15 articles were identified for analysis, including 5 articles and 10 conference papers regarding the document type. The contributions of these 15 articles regarding computer vision applications are summarized in Table 2.

[1] `https://www.crd.york.ac.uk/CRDWeb/`

Table 2: Identified works and their contributions regarding computer vision applications.

Article	Year	Component	Contributions in the aspect of computer vision applications
[59]	2008	Clamp	This paper proposed to use two stereo vision systems mounted on each end effector of two robot arms to recognize the designed markers on the designed cubic clamp covers using Scale Invariant Feature Transform (SIFT) [81, 82]. The experimental results indicated that the vision system can provide enough precision for gripping clamp covers.
[83]	2009	Connector	This paper proposed to use one charge-coupled-device (CCD) camera for connector grasping error detection and quality control based on basic pattern matching.
[84]	2010	Clamp	This paper improved [59] by using three stereo vision systems mounted on each end effector of three robot arms and ten fixed cameras surrounding the work cell to recognize the AR-ToolKit [85] markers on the designed cylinder-like-shape clamp covers.
[61]	2010	Connector	This paper improved [83] by using two mutually perpendicular CCD cameras to detect magnitudes of tilt angles and horizontal displacements from each side by pattern matching.
[86]	2011	Clamp	This paper extended [84] with more details and discussions.
[87]	2012	Connector	This paper focused on the monitoring task for the connector mating process. Two mutually perpendicular cameras were used to observe the relative and online motions between the two connectors based on pattern matching.
[88]	2013	Connector	This paper proposed to use a high-speed vision system to acquire the categories, orientations, and positions of connectors at a frame rate of 500 frames per second (FPS) via connector corner detection.
[62]	2015	Clamp	This paper proposed to use a wrist camera on the right robot arm to recognize the AR-ToolKit [85] markers on the clamp cover, whose position was estimated beforehand based on tracing trajectory.
[89]	2017	Connector	This paper proposed a method for monitoring the mating process of electric connectors. A hand-eye camera was used for locating the connector headers via visual servoing with markers.
[90]	2020	Connector	This paper proposed using an red-green-blue-depth (RGB-D) camera for the recognition of plug-in cable connectors based on image processing. The positions of connectors were detected based on red-green-blue (RGB) images. The three-dimensional (3D) information of the detected connectors were then acquired by registering the depth information to the RGB-based detection results.
[52]	2020	Connector	This paper proposed to acquire the position and orientation of a connector via learning-based rough locating and shape-based fine positioning. The proposal used three cameras: 1) a fixed global camera for rough locating; and 2) a hand-eye camera per robot arm (two robot arms in total) for fine positioning.
[91]	2021	Wire	This paper focused on the interpretable classification of wire harness branches. The interpretability was visualized using saliency maps based on class activation mapping (CAM) [92]. The experimental results demonstrated the best classification based on a late prediction fusion [93] of RGB and depth modalities; the deteriorating performance with the network pre-trained on the inpainting task; and the positive effect of elastic transform for data augmentation. The saliency maps promoted the interpretability of the experimental results.
[94]	2022	Wire	This paper proposed a multi-branch wire harness object recognition with segmentation and estimation based on point clouds acquired using an RGB-D camera.
[60]	2023	Wire	This paper explored the industrial bin-picking problem on wire harnesses. The study proposed learning a bin-picking policy to infer an optimal grasp and a post-grasping action based on a top-down depth image of the cluttered wire harnesses captured by a Photoneo PhoXi 3D scanner M. This paper also suggested visual noise and heavy occlusion as two major challenges leading to failure results.
[1]	2023	Wire harness bag	This paper focused on RGB-based deformable wire harness bag segmentation. A bag instance segmentation dataset generation pipeline based on the *copy-and-paste* technique [95–97] and geometric and photometric image data augmentation techniques [98] was proposed to address the lack of annotated datasets of task-specific objects of interest.

Figure 5 illustrates the development of research in computer vision applications in the robotized wire harness assembly throughout the past years regarding the total number of publications and citations. The data was retrieved from *Scopus* on September 6, 2023. The statistics indicate a persistent long-term and increasing effort to facilitate the robotized wire harness assembly with vision systems.

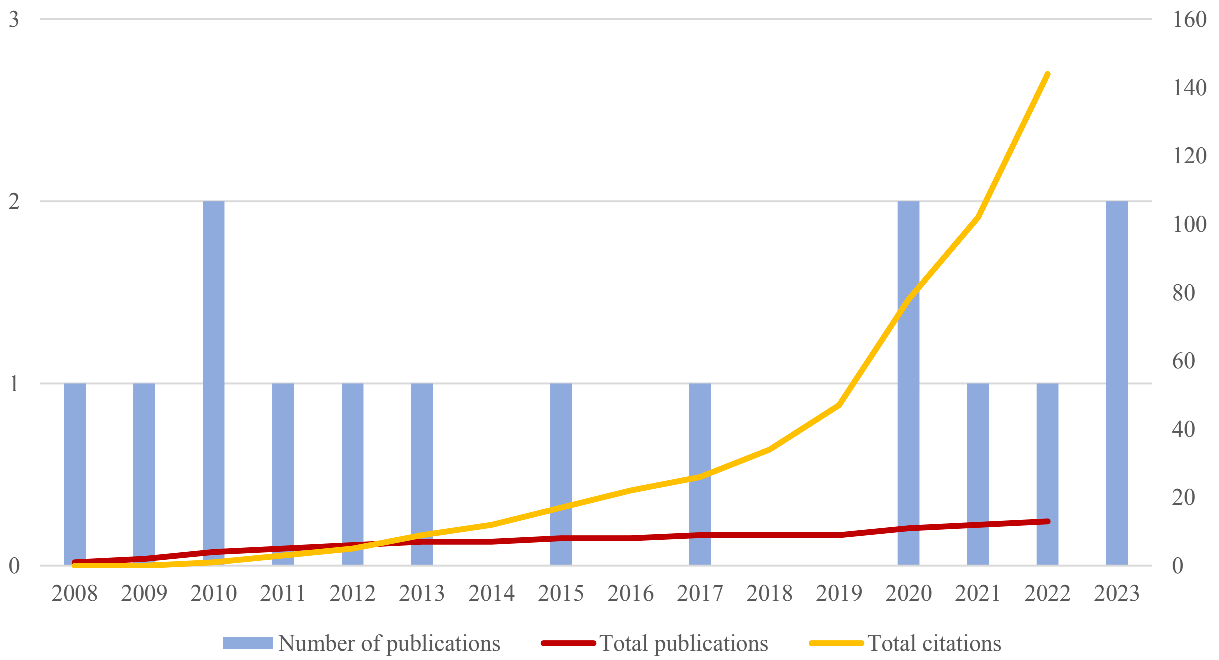


Figure 5: The number of publications per year and the total number of publications and citations by each year (data retrieved from *Scopus* on September 6, 2023). The total number of publications and citations for the year 2023 are excluded considering the incomplete statistics at the time of literature searching. **Color should be used for this figure in print and size should be calibrated for the camera-ready version (This sentence will be removed for the final submission).**

Figure 6 presents the citation relationships among the identified 15 studies. These relationships illustrate that the research in vision-based robotized wire harness assembly began with the recognition and manipulation of wire harness components, such as clamps and connectors. More recent studies have initiated new directions involving understanding the overall structure of wire harnesses. However, the research regarding different components of wire harnesses ceased at different times.

In Figure 6, each study is represented by a colorized circle in the two-dimensional (2D) Cartesian coordinate system regarding the year of publication, the first affiliation of the paper, and the wire harness component in focus. The red, orange, yellow, and teal circles represent previous studies focusing on clamps, connectors, wires, and wire harness bags. Correspondingly, the red, orange, yellow, and teal bars by the x-axis indicate the period of previous studies focusing on clamps, connectors, wires, and wire harness bags, respectively. Solid and dashed arrows visualize the citation relationships. These arrows represent the citation between two studies discussing the same and different wire harness components. Each arrow represents a citation-reference relationship between two previous studies. For example, “A”→“B” means that “A” was cited by “B”. Solid arrows represent the citation between two studies focusing on the same component. Dashed arrows represent the citation between two studies focusing on different components. The double solid line between [84] and [86] represents the extension on details and discussions in [86] from [84].

4.1. Research regarding wire harness components

The studies identified for this systematic literature review were experimental studies in laboratory settings. This review categorizes these studies into four main groups regarding the wire harness component in focus: 1) for clamp manipulation (4 studies); 2) for the mating of connectors (7 studies); 3) for wire harness recognition (3 studies); and 4) for wire harness bag segmentation (1 study).

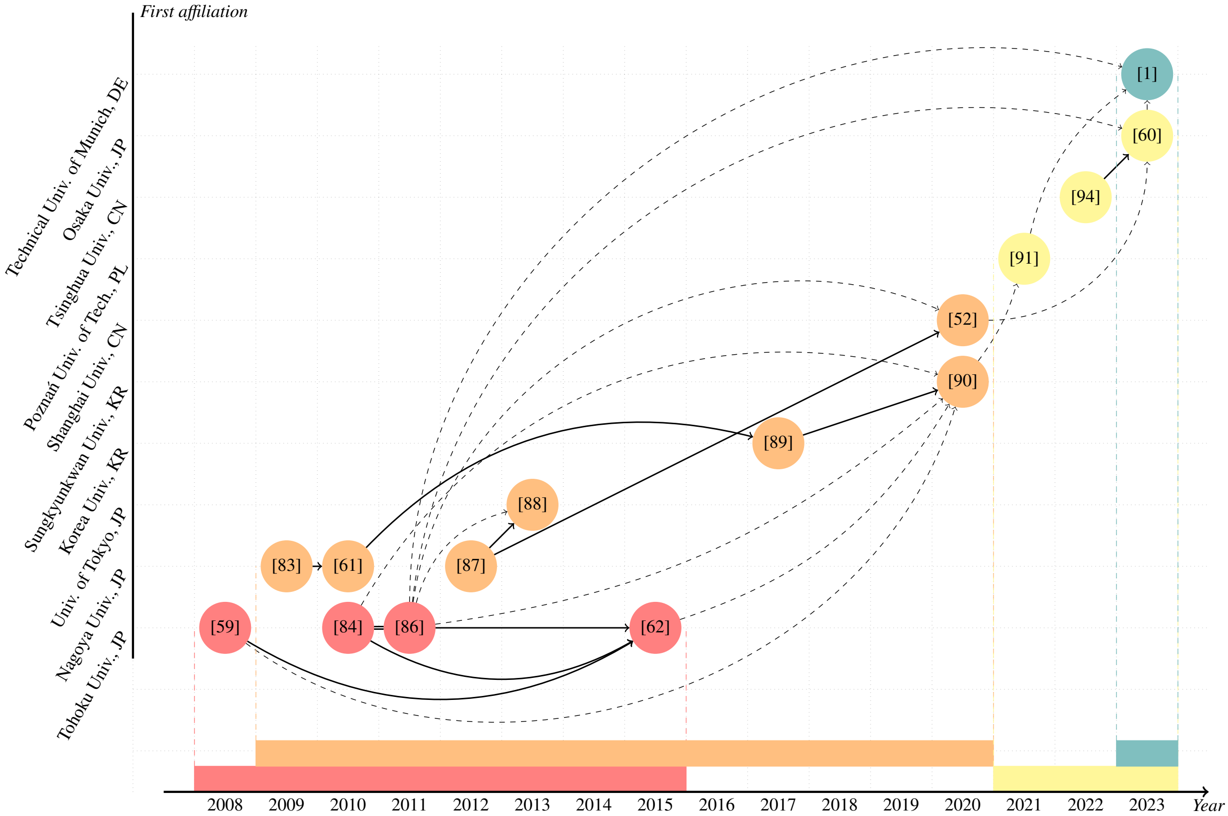


Figure 6: The citation map among the identified studies. The red, orange, yellow, and teal circles and bars represent previous studies focusing on clamps, connectors, wires, and wire harness bags. Solid and dashed arrows represent the citation between two studies discussing the same and different wire harness components. The double solid line represents the extension of the core content. (Univ. = University; Tech. = Technology) **Color should be used for this figure in print and size should be calibrated for the camera-ready version (This sentence will be removed for the final submission).**

4.1.1. Clamp manipulation

Clamps are bound on wire harnesses for attaching wire harnesses to target positions on the final product. Four reviewed articles addressed the manipulation of clamps on wire harnesses achieved with computer vision-based clamp recognition, as summarized in Table 3.

This group of articles focused on mating wire harnesses onto a car body, e.g., inserting the clamps on a wire harness into an instrument panel frame. In these proposals, clamps were recognized using CCD camera [59, 84, 86] and complementary-metal–oxide–semiconductor (CMOS) camera [62]. With the recognition results, the gripper or end effector on a robot arm could reach the clamps for further manipulations. All four studies implemented hand-eye vision systems by mounting cameras on the end effectors of robot arms [59, 62, 84, 86]. Two of them also adopted global vision systems with multiple cameras fixed around the operation area [84, 86].

Although the clamps can be regarded as rigid objects, their relatively small sizes and complex shapes make them difficult to grasp by a robot gripper directly [59]. The complexity of control program development and the insufficiency of the payload makes multi-finger robot hands challenging to adapt in practical applications [59]. Thus, [59] proposed to install cubic clamp covers to facilitate the manipulation of clamps with small sizes and complex shapes. Then, [59] introduced stereo vision systems mounted on the end effectors of two robot arms to recognize the designed markers on the clamp covers based on SIFT [81, 82]. The stereo vision system in [59] contained two CCD cameras with different focal lengths for observing far and near objects, respectively. Based on the recognition results, the robot arms moved to the clamp covers for further manipulation. The experimental results in [59] qualitatively indicated that the implemented

Table 3: Computer vision-based clamp recognition for clamp manipulation. (LC = location of cameras; TV = type of vision; NC = number of cameras)

LC	Article	TV	NC	Method
End effector	[59]	Stereo	4	SIFT [81, 82]
	[84, 86]	Stereo	6	ARToolKit [85]
	[62]	Monocular	1	ARToolKit [85]
Work cell	[84, 86]	Monocular	10	ARToolKit [85]

vision system could visually recognize clamp covers with enough precision for robots to grip clamp covers [59].

Later, [84] and [86] made further improvements to [59] by: 1) deploying one more robot arm; 2) using newly designed cylinder-like-shape clamp covers with more markers from ARToolKit [85] for recognition; 3) implementing a global vision system (ten fixed cameras surrounding the work frame from various directions) besides the three hand-eye vision systems for clamp cover detection to address the occlusion problem on clamps (this increased the number of cameras to sixteen); and 4) adding a laser head on the end effector of each robot arm in case precise measurement of a wire segment was demanded.

Instead of directly detecting the clamp covers on a wire harness visually, [62] proposed to adopt a mechanical tracing operation to identify the position of clamp covers on a wire harness first. Then, the markers on the clamp cover were recognized with one wrist CMOS camera (Point Grey Firefly MV) on the right robot arm to estimate the pose of the identified clamp cover so that the robot could manipulate it in later procedures [62].

4.1.2. Mating of connectors

Mating of connectors is critical in the final assembly of wire harnesses in terms of function and quality. Table 4 summarizes the seven articles focusing on different stages of the mating of connectors, which can be grouped into three sub-procedures: 1) pre-assembly, 2) mating, and 3) post-assembly.

Table 4: Computer vision applications in articles addressing the mating of connectors. (TV = type of vision; TC = type of cameras; LC = location of cameras; NC = number of cameras; GF = globally fixed; HE = hand-eye; 2.5D = two-and-a-half dimensional)

Process	Purpose	TV		TC	LC	NC	Method
		2D	2.5D				
Pre-assembly	Connector detection	[88]	-	MC1362	GF	1	Image processing
		-	[90]	RealSense D435	HE	1	Image processing
		[52]	-	Industrial cameras	GF+HE	1+2	Learning- and model-based
	Pose estimation	[52]	-	Industrial cameras	GF+HE	1+2	Learning- and model-based
	Fault detection	[83]	-	In-Sight 5100	GF	1	Pattern matching
		[61]	-	CCD cameras	GF	2	Pattern matching
Mating	Relative position detection	[87]	-	CCD cameras	GF	2	Pattern matching
		[88]	-	MC1362	GF	1	Image processing
		[89]	-	FL2G-13S2C-C	HE	1	Visual servoing with pattern matching
Post-assembly	Fault detection	[83]	-	In-Sight 5100	GF	1	Pattern matching

Pre-assembly is the initial process where the vision system acquires the necessary information to guide a robot's movement toward connectors and the later robotized mating operation. There were five articles discussing various tasks in pre-assembly, including connector detection [52, 88, 90], pose estimation [52], and fault detection [61, 83].

Perceiving the 3D geometric information of connectors, e.g., position and orientation, is essential for robotized mating connectors on flexible wire harnesses. [88] proposed to use a high-speed camera (EoSens series, MC1362, Mikrotron) fixed above the workbench to distinguish the types of connectors and acquire their positions and orientations by detecting the corners of the connectors at a frame rate of 500 FPS. Eight-bit grayscale images were captured and processed through a sequence of image processing methods from OpenCV library[2]. The image processing included

[2] `https://opencv.org`

image smoothing, binarization, contour detection, and straight-line approximation to acquire the corners of each connector [88]. The recognized connector corners were further used for obtaining the orientation of each connector by calculating each connector's principal axis of inertia [88]. [88] processed 2D grayscale images only as they designed that the connectors handled by the robot hand were placed on a table surface to make the z-coordinate of every connector constant and the orientation of every connector various in the yaw direction only.

Targeted on obtaining the precise positions of the plug-in cable connectors on a workbench, [90] proposed to adopt an Intel RealSense D435 depth camera for the recognition process based on image processing. The captured RGB images were first processed through a sequence of image processing methods from the OpenCV library, including color space conversion, color thresholding, and image moment calculation, to acquire connectors' 2D contours on XY-plane [90]. Then, the obtained 2D contours were mapped to the captured depth image, which was registered to the RGB image in advance, to obtain the 3D information of connectors on z-direction for further robotized manipulation [90].

A two-step connector detection algorithm was proposed in [52] and verified with a dual-arm robot using three cameras (one fixed global camera and one hand-eye camera on each robot arm). The first step in [52] was a learning-based rough locating, where a 2D grayscale image of the wire harness captured by the fixed global camera was processed by MobileNet-SSD [99] to locate the connectors roughly. Then, one of the hand-eye cameras on the robot arms reached the top of the located connector to capture images for a shape-based fine positioning to obtain the 6-degrees-of-freedom (6-DOF) pose of the detected connector for further manipulation based on computer-aided design (CAD) models and multi-view image matching [52].

Previous studies have also discussed quality assurance in collecting connectors by the gripper on a robot arm. [83] proposed to use an In-Sight 5100 camera for vision-based connector grasping error detection and quality control to confirm that the connector was caught correctly and conveyed to the correct position. The relative translational and rotational displacements between the gripper and the connector were examined based on basic pattern matching with image processing [83]. The pattern matching algorithm processed a cross-shape pattern drawn on the gripper, a circle-shape pattern drawn on the female connector, and the intrinsic design of the male connector [83]. Besides, [83] implemented this fault detection after inserting the connectors to examine the operation result. However, the experiment results of [83] indicated the insufficient training samples and the only single camera used as the main reasons behind the misclassifications. [61] further improved [83] by adding one more camera perpendicular to the first one to detect magnitudes of tilt angles and horizontal displacements on each side based on pattern matching.

After being collected by a gripper, the connector is transported and mated to the target counterpart. Previous studies proposed vision-based solutions for guiding and monitoring the mating process [87–89]. [87] focused on the monitoring task and implemented two mutually perpendicular cameras to observe all the relative and online motions between the two connectors by pattern matching, whose design was akin to the pattern matching algorithms adopted by [83] and [61]. To locate the connector headers, [89] adopted a hand-eye camera (FL2G-13S2C-C, Point Grey Research) and proposed to implement visual servoing with markers [100] to track the headers of connectors. In addition to connector detection in pre-assembly, [88] also proposed to use the high-speed vision system to monitor the mating process in real-time by detecting the distance between the in-hand and the target connector.

4.1.3. Wire harness recognition

Besides clamps and connectors, the wire part of a wire harness is also critical to address. The recognition of the wire part can provide a better perception of the overall structure of a wire harness. There are three studies discussing wire harness recognition, as listed in Table 5, focusing on the interpretable classification of branches [91], wire recognition [94], and grasp detection [60], respectively.

Table 5: Computer vision applications in articles for wire harness recognition. (TC = type of cameras; LC = location of cameras; NC = number of cameras; GF = globally fixed)

Article	Purpose	TC	LC	NC	Method
[91]	Interpretable classification	RealSense D435	GF	(1)	Deep learning-based
[94]	Visual recognition	RGB-D	-	-	Machine learning-based
[60]	Grasp detection	Photoneo PhoXi M	GF	1	Fast Graspability Evaluation [101]

[91] focused on the interpretable classification of wire harness branches. Specifically, [91] proposed a dataset and several convolutional neural networks (CNN) [102] for classification based on different data modalities. The proposed dataset contained RGB-D images of four branches of an automotive wire harness captured by an Intel RealSense D435 depth camera mounted 50 cm above the ground level [91]. The CNNs proposed in [91] shared the same Downsample layer from ERFNet [103]. Data augmentation with elastic transformation and a network pre-trained with the inpainting task were also evaluated in [91]. The experiment results of [91] on various input modalities demonstrated the best accuracy achieved by classifying with a sum of logits of models taking RGB data and depth information as input, respectively, following the late prediction fusion [93]. Further experiments dealing with small datasets in [91] indicated a significant drop in the performance with the classification network pre-trained on the inpainting task but a performance improvement with data augmentation using elastic transform. [91] also visualized the classification results using a saliency map based on class activation mapping (CAM) [92] to interpret the experimental results better visually.

[94] proposed a multi-branch wire harness object recognition with sequential segmentation and probabilistic estimation in aircraft assembly. In the proposal [94], the raw point cloud data was first acquired using an RGB-D camera. Then, the raw data was pre-processed to supervoxels by over-segmentation. Based on the supervoxels, wires were segmented considering the Cartesian distance, color similarity, and bending continuity. After segmentation, there were inevitable gaps in the segmentation result due to sheltering or occlusion. The gaps were further remedied by estimation with Gaussian Mixture Model (GMM) [104] to obtain the complete segmented point cloud of wires [94].

Focusing on the industrial bin-picking problem on wire harnesses, [60] proposed learning a bin-picking policy to infer an optimal grasp and a post-grasping action from a top-down depth image of the cluttered wire harnesses. The proposal enabled the system to prioritize grasping the untangled objects, avoid grasping at the wrong positions, and reason the extracting distance to reduce the execution time for a successful picking [60]. The vision system in [60] included a Photoneo PhoXi 3D scanner M fixed directly above the workbench. Given a top-down depth image as an observation of the current entanglement situation, a model-free grasp detection based on Fast Graspability Evaluation (FGE) [101] was adapted in [60] to detect collision-free grasps. Specifically, the solution outputted a set of grasp candidates ordered by their FGE scores [60]. Then, [60] trained an action success prediction module via a proposed active learning scheme to predict the success possibilities of the disentangling actions. Specifically, [60] encoded and processed the captured depth image, the set of grasp candidates, and the pre-defined action candidates using different CNNs. The action-grasp pair to conduct the operation was then selected based on the FGE [101] score and the action complexity [60]. However, the experiment results of [60] indicated visual noise and heavy occlusion as two significant challenges leading to failure pickings.

4.1.4. Wire harness bags segmentation

Manufacturers have started organizing the overall wire harness into multiple sub-groups with deformable bags to simplify the assembly operations [1]. To promote automating wire harness assembly operations, [1] inquired about RGB-based deformable wire harness bag segmentation. [1] identified the lack of an annotated dataset of specific objects of interest as a significant constraint in developing required vision systems. Hence, [1] proposed a dataset generation pipeline relying on minimal human effort. Specifically, the pipeline adopted the *copy-and-paste* technique [95–97] to generate images with diverse combinations of objects and backgrounds [1]. A list of geometric and photometric image data augmentation techniques [98] was also adopted to address the insufficient data of task-specific objects of interest, the deformable wire harness bags [1].

First, 56 foreground images with a single instance of a wire harness bag in each were captured and manually annotated using an annotation tool, *Segments.ai*[3], to collect bag instances. Then, indoor scene images in Massachusetts Institute of Technology (MIT) indoor scenes dataset [105] were selected as background images for good generalization in real-world scenarios, considering the expected application environment in factories or industrial plants. High-Resolution Salient Object Detection (HRSOD) dataset [106] and a "complex" dataset collected following the data collection methodology in [107] were also adopted for comparison in evaluation. A set of geometric and photometric image data augmentation techniques [98] were implemented on foreground and background images, respectively, before implementing *copy-and-paste* [95–97] to enable greater diversity and variance among data samples in the final overall dataset. In implementing *copy-and-paste* [95–97], an image of a wire harness bag was randomly selected and

[3] `https://segments.ai/`

pasted n times at random locations on a background image selected randomly. The number n of foreground instances in each training image was between 1 and 4. In total, 5000 samples with a resolution of 640×480 pixels were obtained, with the 90%-10% training-validation split for training the segmentation model.

ResNet101 [108], SwinS [109], and ConvNeXtS [110] were selected as backbone architectures for the evaluation of the obtained dataset on the segmentation task of wire harness bags. The trained models were evaluated with Dice coefficient and Intersection-over-Union (IoU) on a test set comprising 75 accurately annotated images grouped into three scenarios (25 images each), including the laboratory scenario, the beginning of assembly operations, and during the operations. The experimental results present the best performance achieved by ConvNeXtS [110] trained with backgrounds from MIT indoor scenes dataset [105] and demonstrate the validity of the obtained dataset on training deep neural networks for deformable wire harness bag segmentation under practical configurations in the actual production.

4.2. Research regarding wire harness assembly operations

The identified studies can also be synthesized regarding which operation of wire harness assembly the proposed vision systems contributed to, as shown in Table 6. The majority of previous research discussed the application of computer vision techniques for facilitating the assembly procedure, including four studies on the fixing of clamps [59, 62, 84, 86] and seven studies on the mating of connectors [52, 61, 83, 87–90]. The interpretable classification of wire harness branches explored in [91] and the recognition method for wires proposed in [94] can support robotic routing with a better understanding of the topology of wire harnesses. [59], [84], and [86] proposed to install covers on clamps to facilitate the detection and manipulation of clamps on wire harnesses, which, on the other hand, can also assist the robotic routing of wire harnesses by locating the positions of clamps.

Table 6: Contributions of computer vision techniques in previous studies to different assembly operations.

Assembly operation	Articles
Prepare	[1]
Transport	-
Untangle	[60]
Route	[27, 59, 84, 86, 91, 94, 111]
Assemble	[52, 59, 61, 62, 83, 84, 86–90]

However, as summarized in Section 2.5, there are also other operations within the complete wire harness assembly onto vehicles but gaining less attention, such as preparation, transport, and disentanglement [1, 60]. Since the wire harness is processed as a pack in preparation and transport, the vision system's task will mainly be identifying the pack of wire harnesses and the location for dropping it. The untangling of wire harnesses leads to more problems requiring future research. The robot system needs a dynamic robotic manipulation strategy to react to wire harness deformation, which stresses the significance of real-time wire harness tracking.

4.3. Vision system evaluation in previous literature

All the analyzed articles contained experimental studies in laboratory configurations and conducted various evaluations on their proposals [1, 52, 59–62, 83, 84, 86–91, 94]. Specifically, there were various metrics for evaluating the performance of the vision systems in some of the proposals qualitatively and quantitatively, as summarized in Table 7 and Table 8, respectively. However, some studies adopted vision systems without reporting the evaluation of their vision systems explicitly [59, 61, 62, 84, 86–89].

[59], [84], [86], [62], [89], and [60] qualitatively indicated the technical feasibility of their vision systems for facilitating the robotized manipulation of clamp covers and estimating the relative position of the connector header, respectively, by the execution of robotic manipulation. [90] demonstrated the qualitative evaluation of their vision system by presenting the detection results with bounding boxes around detected connectors and position references of detected connectors in a simulation environment. [83] reported the fault detection rate to quantitatively reflect their proposed vision system's performance. [60] demonstrated the effectiveness of the proposed learning efficient policy for picking entangled wire harnesses with action-grasp pair prediction images.

Table 7: Qualitative evaluation on vision systems in previous studies.

Article	Component	Metrics
[90]	Connector	Bounding box, position reference
[52]	Connector	Bounding box, pose
[91]	Wire	CAM [92]-based class-agnostic saliency map
[94]	Wire	Recognition result
[60]	Wire	Action-grasp pair
[1]	Wire harness bag	Segmentation mask

Table 8: Quantitative evaluation on vision systems in previous studies.

Article	Component	Metrics
[83]	Connector	Fault detection rate
[52]	Connector	Average Precision (AP); evaluation of repeatability; average and max-min error
[91]	Wire	Classification accuracy; confusion matrix
[94]	Wire	Accuracy; time cost
[1]	Wire harness bag	Dice coefficient; Intersection-over-Union (IoU)

Besides, [52], [91], [94], and [1] reported both qualitative and quantitative experiment results. [52] demonstrated the performance of detection qualitatively by presenting a bounding box and a pose frame around the detected connector and illustrated the performance and reliability of the proposed algorithm quantitatively with Average Precision (AP), repeatability evaluation, and average and max-min error. In addition to the qualitative evaluation result with saliency map, [91] also reported classification accuracy and confusion matrix to demonstrate the performance of their proposal quantitatively. [94] evaluated the practicality of their proposal by examining the time cost of each block of their proposed recognition method besides demonstrating the performance of the proposed methods with recognition result images. Similarly, [1] evaluated the proposed approach regarding the Dice coefficient and IoU and illustrated the performance with segmentation result images.

5. Discussion

In general, previous research has been devoted to enabling robot systems' better autonomy to estimate the state of assembly autonomously for various tasks in wire harness assembly [3]. As a significant aspect in robotics, enabling robotic visual perception has attracted researchers' efforts on robotizing wire harness assembly [21]. Previous studies explored the application of various computer vision techniques to enable robots to have better visual perception capabilities at different levels of the constituent structure of wire harnesses. Specifically, previous research mainly focused on the manipulation of different components of wire harnesses [1, 52, 59, 61, 62, 83, 84, 86–90], monitoring sub-processes of the assembly [60, 83, 87, 89, 91, 94], and fault detection during the assembly [83]. These solutions were mainly achieved with vision-based classification and detection on various components of wire harnesses, e.g., clamps [59, 62, 84, 86], connectors [52, 61, 83, 87–90], wires [60, 91, 94], and wire harness bags [1]. The proposed vision systems also contributed to different operations regarding current work instructions for wire harness assembly in practical production, including preparation [1], disentanglement [60], routing [59, 84, 86, 91, 94], and assembly[52, 59, 61, 62, 83, 84, 86–90]. Nevertheless, these existing studies imply challenges and opportunities for future research toward more efficient and practical computer vision-enabled robotized wire harness assembly.

5.1. Computer vision techniques

Various industrial manufacturing systems have implemented computer vision techniques to promote informatization, digitization, and intelligence [20]. Increasing research efforts in computer vision and robotics have paid attention to the complex DLO manipulation, addressing both theoretical research [27, 111–114] and engineering practices [35, 62, 83, 86, 94]. However, for robotized wire harness assembly, a practical challenge among DLO manipulation problems, it is necessary to recognize wire harnesses successfully within a limited processing time to plan and control robotic manipulation under a constrained production rate [94]. Besides, estimating the shape of wire harnesses by a pure vision-based algorithm in practical applications in an actual production plant indicates extracting image features from an intricate background [59]. On the other hand, the automotive industry has yet to identify a satisfactory vision-based automation solution that can recognize different components of wire harnesses as robustly as humans. The inconsistent production configuration in practical manufacturing exacerbates the difficulty of visual recognition. This has attracted plenty of studies on wire harness component recognition in the past few years [1, 52, 59–62, 83, 84, 86–91, 94]. Nevertheless, further research on object recognition and detection algorithms is needed to achieve a more robust visual machine perception performance for more efficient and practical robotized wire harness assembly. Specific problems include object detection and pose estimation on wire harness components, topology matching on wire structure, and wire deformation tracking while manipulating wire harnesses in production.

5.1.1. Utilizing features of wire harnesses

Novel object recognition and detection algorithms, which harness the intrinsic features of wire harnesses instead of relying on artificial fiducial markers, hold the promise of revolutionizing robotized wire harness assembly, making it more efficient and practical. Previously, clamp detection was achieved with the assistance of clamp covers with unique designs and markers [59, 62, 84, 86]. Some previous works with connector detection also utilize artificial fiducial markers for object identification and location [61, 83, 87].

The clamp covers can facilitate recognizing and manipulating clamps with relatively small sizes and complex structures. The added artificial fiducial markers can simplify feature engineering when designing object recognition algorithms. However, while effective in recognizing and manipulating clamps, clamp covers present significant challenges in the practical assembly of wire harnesses, particularly in compact installation areas. Their presence not only occupies valuable space but also complicates the process when they need to be removed, highlighting the need for a more streamlined approach. It is also not feasible considering an increasing number of wire harnesses assembled in future products, e.g., electric vehicles, as adding artificial fiducial markers would introduce extra complexity, workload, and cost for the assembly operation in actual production [16]. In addition, the effectiveness of adding artificial fiducial markers may be impaired due to their various orientations and potential occlusions in actual production environment [111]. Therefore, it is desirable to recognize the components of wire harnesses and exploit their intrinsic features without additional attachments.

5.1.2. Implementing learning-based algorithms

Earlier research analyzed in this review article mainly achieved 2D image recognition on clamp covers and connectors with traditional rule-based image processing methods [59, 61, 62, 83, 84, 86–90]. By contrast, more recent studies on wire recognition took advantage of learning-based algorithms [1, 52, 60, 91, 94], which enabled more robust recognition of objects with a more complex structure, e.g., the flexible and deformable wires and wire harness bags.

The recent renaissance of CNN [102] and the successful development of deep learning in computer vision research [115] have demonstrated the superior performance of learning-based object recognition than traditional rule-based methods [116]. This booming development of applying deep learning techniques in computer vision research promoted various learning-based designs and solutions for object recognition, e.g., Regions with CNN features (R-CNN) [117], Fast R-CNN [118], Faster R-CNN [119], and You Only Look Once (YOLO) [120]. Applying these learning-based techniques will facilitate better visual machine perception in the future robotic assembly of wire harnesses.

The booming development of AI-supported tools, such as ChatGPT, demonstrates the superiority of large language models (LLM) and generative pre-trained transformers (GPT) in enabling machine intelligence. The robotics community speedily seized the potential opportunities due to the advent of ChatGPT and initiated the discussion on "RobotGPT" to take advantage of LLM and GPT in robotics [121]. Previously, researchers investigated the strengths of language/text parsing of LLM and proposed various approaches to improving the robotic capability on task planning [121] and human-robot interaction [122]. Nevertheless, with the advent of large multimodal models (LMM) competent in some visual problems (e.g., GPT-4 with Vision[4]), the impact of LMM in addressing visual problems in robotic assembly of wire harnesses is worth further investigations [123].

[4] `https://platform.openai.com/docs/guides/vision`

It is also promising to promote the implementation of depth or other 3D cameras to acquire spatial information and process 3D or multi-modality data directly with learning-based methods, considering the increasingly advancing and affordable photography technology and recent development of learning-based object recognition and detection algorithms in the computer vision community [116, 124, 125], Therefore, further research is worth conducting on adapting learning-based detection and pose estimation of wire harness components based on multi-modality data.

Nonetheless, the dataset is essential for learning-based object detection [126–128] and scalable deep learning-based computer vision applications in manufacturing [20, 129]. Additionally, benchmarks are critical to assess and compare the performance across different computer vision and robotic systems [130]. Though increasingly popular within the robotics community, benchmarks for various tasks have yet to gain sufficient research for robotized wire harness assembly [131]. Thus, in addition to learning networks, more studies are needed on the datasets, benchmarks, and metrics to promote the performance of learning-based computer vision techniques implemented for future robotized wire harness assembly and evaluate them consistently and rigorously.

5.1.3. Learning from multi-modality data

Previous research efforts summarized in Table 3, Table 4, and Table 5 indicate the dominant adaptation of 2D vision in previous research on facilitating robotized wire harness assembly with computer vision techniques. For example, for connectors, recognition was mainly used to provide geometric information to the control system so that the robot could flexibly perceive and manipulate the connectors [52, 61, 83, 87–90]. Previous research mainly adapted 2D image recognition to acquire the positions or orientations of connectors [52, 61, 83, 87–89]. For example, [88], using a 2D camera, assumed that all connectors were placed on a flat workbench to reduce the degrees of freedom of connectors. However, considering the actual manufacturing scenario, wire harnesses are not fixed on a flat workbench before or during assembly, making the connectors distributed randomly in the 3D space. Hence, the degrees of freedom of connectors cannot be reduced directly, which results in limited practicality of the vision system proposed by [88] in actual production. [61] implemented two 2D cameras to ensure the correct positioning of connectors in 3D space, which increases the complexity of vision system control. Therefore, it is desired to collect 3D information directly, e.g., depth information or point clouds of objects of interest, to better perceive wire harness components.

The recent acceleration in the development of depth cameras and 3D scanners makes it more achievable to acquire and process depth or 3D visual information in addition to 2D data, e.g., RGB or grayscale images. However, though an RGB-D camera was adopted, [90] acquired the positions of connectors based on 2D image processing and mapped the identified positions to the registered depth image to further obtain the 3D measurement of connectors instead of processing depth or other 3D information independently or together with 2D color information. The more recent studies on recognizing the wire part explored implementing RGB-D camera or 3D scanner and learning from depth images or 3D data and have demonstrated the performance of 3D vision applications on facilitating a better perception of the structure of wire harnesses [60, 91, 94]. On the one hand, these results enable robotized preparation, transport, disentangling, and routing of wire harnesses in the future assembly station. On the other hand, they provide references to address computer vision applications on the other components of wire harnesses based on depth images, point clouds, or other 3D data.

5.1.4. Adapting studies for wire harness manufacturing

It is noteworthy that there are also studies referring “wire harness assembly” to the operations in the manufacturing process of wire harnesses [14, 22–24]. Manufacturing wire harnesses is out of the scope of this systematic review. Nevertheless, some sub-problems are shared among computer vision applications in the assembly of wire harnesses and the assembly of wire harnesses onto other products. For example, in both scenarios, detecting components of wire harnesses and state estimation during wire manipulation is required to provide robots with the necessary information before conducting the following robotic grasping and manipulation.

Notably, previous research has explored vision-based solutions for robotic manufacturing of wire harnesses or wire harness assembly processes. These innovative designs on vision systems could serve as a valuable reference for the robotized wire harness assembly discussed in this article, hinting at promising avenues for future exploration. Therefore, it is imperative that future research focuses on validating the effectiveness of vision systems proposed for facilitating the robotic manufacturing of wire harnesses in the robotic installation of wire harnesses onto other products.

5.2. Practical concerns regarding actual production

The previous research efforts in vision-based robotized wire harness assembly remained experimental studies in controlled laboratory settings. However, vision systems must address various practical conditions to be implemented in industrial production, which poses practical concerns for developing vision-based solutions in future research.

5.2.1. Evaluating vision systems regarding production requirements

The analyzed articles have demonstrated various vision-based solutions for accomplishing different tasks of robotized wire harness assembly under laboratory settings [1, 52, 59–62, 83, 84, 86–91, 94]. However, assessing the proposed vision systems in actual production remains obscure but necessary before they are integrated. Table 7 and Table 8 present the qualitative and quantitative evaluations conducted on vision systems in different studies. Nevertheless, only a few studies considered the practical issues, e.g., repeatability of the proposal [52] and time cost on the vision system [94].

The evaluation of the proposed vision systems is critical to selecting vision systems, both hardware and software, in practical industrial applications [132]. Quantitatively evaluating proposed vision systems regarding specific production requirements can provide practitioners with valuable suggestions on selecting the appropriate vision-based solutions according to their respective needs. As shown in Table 3, Table 4, and Table 5, various types of vision systems were adopted in previous research in robotized wire harness assembly, including stereo vision, monocular vision, and the combination of them, for different purpose, e.g., obtaining positions and orientations of wire harness components or compensating occlusion from specific views. Different types of cameras were also discussed on the level of devices, e.g., RGB cameras, depth cameras, and 3D scanners. However, as shown in Table 8, a few studies conducted quantitative evaluations on the performance of proposed vision systems. This indicates a need for more consideration of the practicality of vision systems in actual industrial applications. In conclusion, the research underscores the need for comprehensive evaluations of different vision systems in the context of practical production requirements. This type of evaluation is essential to provide practitioners with more robust and practical advice on selecting vision systems for their specific industrial applications. Therefore, further research in this direction is highly recommended.

5.2.2. Considering actual production environment

In a practical manufacturing environment, it is inevitable for the vision system to cope with inconsistent conditions, e.g., background [59] and illumination conditions [86]. [59] revealed the complex background as a non-negligible hindrance for a pure vision-based wire shape estimation in an actual plant. [86] indicated the lack of vision and laser processing robustness as one of the significant reasons behind experiment failures, e.g., the variation of illumination conditions. Thus, it is critical to evaluate the practicality, robustness, and reliability of vision systems regarding different perspectives, e.g., recognition accuracy, processing speed, and physical facilities in the assembly station, under actual production configurations with various background and illumination conditions. [1] initiated the consideration of practical operation background from actual production in evaluating the proposed vision-based solution, and more attention from the research community should be paid to this perspective.

Another practical issue that has not been discussed previously is the fact that some wire harnesses are installed onto the final product on a moving production line [16], for which [16] proposed a method for a mobile robot manipulator assembling wire harnesses to track a moving vehicle with visual servoing. Nevertheless, the vision system's processing time and feature engineering process are still challenging considering actual production requirements [16]. Hence, more studies are required to address the robotized wire harness assembly on a moving production line.

5.2.3. Fulfilling economic requirements

From the business perspective, there is also a demand for the processing time of robotic assembly to fulfill the productivity in practical production, which requires the vision system to perceive the state of the manipulating object fast enough to allow for a following robot action [133]. [84] demonstrated the technical feasibility of robotized wire harness assembly, but the average speed and reliability were still far from the requirement of practical application. [94] also identified the necessity of promoting time efficiency. Therefore, evaluating the vision system from the perspective of processing time is desired and necessary, which, however, was involved little in previous research [94]. Regarding sustainability from the economic and social perspectives, it is also critical for practical application to consider the pay-off of automating the overall or part of current manual wire harness assembly by implementing vision systems [25, 130, 134].

5.3. Human-robot collaboration

The deformability of wire harnesses has been identified as a significant problem for automating the assembly of wire harnesses regarding the perception and manipulation [59]. Automation has been widely adopted in manufacturing since the third industrial revolution [11]. Nevertheless, the lack of flexibility and cognitive ability in robots motivates the study of human-robot collaboration (HRC) [135], where the system can benefit from the synergy of humans' strength in perception and flexibility and robots' superiority in payload, accuracy, and repeatability. Recently, Industry 5.0 has also been initiated with a focus on sustainability, human-centricity, and resilience [136–138]. This further promotes the discussions on implementing human-robot collaboration in industrial applications towards human-centric automation [135].

There have been studies on human-robot collaboration driven by computer vision techniques in industrial applications [139] and particularly in handling deformable objects [140]. [141] also proposed to optimize wire harness assembly based on human-robot collaboration. However, previous research on computer vision-based robotized wire harness assembly focused on fully robotic assembly [52, 59–62, 83, 84, 86–91, 94]. More research is needed to investigate human-centered automation and human-robot collaboration for the robotic assembly of wire harnesses.

To introduce human-robot collaboration to robotized wire harness assembly, several challenges demand further research, e.g., task allocation and addressing a safe human-robot interaction [142]. This makes it crucial to design the workspace to maximize the efficiency of human-robot collaboration while minimizing the risk of accidents. Conventionally, industrial robots are surrounded by physical fences or laser curtains in practical automated production lines to ensure the safety of the operation [54, 55]. However, in human-robot collaboration, a human operator and a robot work at a closer distance than the one for current industrial robots, which significantly promotes the priority of ensuring the safety of human operators [56, 143, 144]. Collaborative robots (or cobots) have been identified as the key enabling technology for better performing various automation tasks by combining the skills of humans and robots while maintaining safety and efficiency [145, 146]. Therefore, applying cobots can be a promising solution for automating wire harness assembly. Various computer vision techniques for facilitating an efficient and safe human-robot collaboration have also been discussed previously, such as collision detection, environment perception, human action or gesture recognition, and proactive human-robot collaboration [147–150]. These studies pave the way for robotized wire harness assembly with computer vision-driven human-robot collaboration. Recent advance on "RobotGPT" also suggests novel directions for developing efficient, effective, and safe human-robot collaboration with LLM or LMM for robotized wire harness assembly [122, 151].

5.4. Product design of wire harnesses

Previously, [59] identified that, though the clamps could be regarded as rigid objects, their relatively small sizes and complex shapes made clamps challenging to visually recognize and grasp directly by a robot gripper. Therefore, clamp covers with markers were installed to facilitate the detection of clamps [59, 62, 84, 86]. However, affixing clamp covers will be impractical in future production, where an increasing number of wire harnesses will be installed in tight areas. Besides, various intrinsic properties of current wire harnesses, e.g., the same color of clamps and taped wires and the small radial scale and complex structure of wire harnesses with irregular curves and crossings, also make it considerably challenging to recognize wire harnesses from the complex background and assemble automatically [59, 94]. Hence, inspired by the "Design for X (DfX)" philosophy [8, 152], novel designs of wire harness components are desired to facilitate visual detection without any additional parts attached to wire harnesses.

5.5. Industrialization barriers

Consciously, the research in robotized wire harness assembly is task-specific and application-oriented. Thus, the research envisions the industrialization of proposed technical solutions. However, implementing automation solutions in wire harness assembly remains scarce in production. This situation could be caused by 1) the lack of interest in industry and/or 2) the unsatisfied research results and/or 3) barriers to transferring research results to industrial practice. The industry has exhibited remarkable interest in robotizing the assembly of wire harnesses [3, 21]. However, as discussed in previous sections of this article, the research is insufficient, and the proposed technical solutions are not powerful enough, so this review advocates future research directions. Nevertheless, several aspects related to industrial implementation should also be addressed besides improving robotic systems to facilitate industrialization, e.g., cost, workforce competence, trustworthiness, and standardization [14, 37].

Cost is fundamental in researching, developing, commercializing, and operating technologies. To reach a satisfactory level, vast investment in funding and human efforts is required to study and develop vision-based solutions. Meanwhile, it is critical to balance the performance and cost of technical solutions in practical operations. This requires controlling the price of hardware and software and keeping operations' complexity to manageable levels. The engineering cost of commercializing the technology is also significant. Thus, substantial efforts are required to investigate how to do so efficiently.

Considering workforce competence, training operators on new skills is inevitable when introducing new technologies [153]. This poses requirements on motivating and training operators fast and with high quality, where research in related challenges and treatments are in need [154]. As a prerequisite, trust in new technologies is crucial to accepting technologies, especially for AI-empowered automation solutions. This socio-ethical issue drives academia to promote the explainability and trustworthiness of AI to address the lack of trust in AI-based solutions in the industry. The growing implementation of AI-based technologies also spawns concerns about cybersecurity and data management, which requires research in relevant fields [155, 156].

Additionally, standardization is elemental to new technologies. Although academia has initiated related research, there is no widely applicable standard or benchmark for vision-based robotic manipulation of DLOs in practical industrial production [130]. This situation calls for further research on methods for benchmarking the performance and evaluation of vision-based solutions in academia. In the meantime, researchers, practitioners, and policymakers should also discuss legislating the certification of existing research results with standards and determining requirements to guide future research and engineering.

6. Conclusion

Robotized wire harness assembly is desired in the automotive industry, considering its strength in promoting assembly quality, productivity, safety, and ergonomics. However, it is challenging due to the flexibility and deformability of wire harnesses, the small sizes and complex shapes of wire harness components, and the demanding production requirements.

Through a systematic literature review, this article revealed that previous research explored various computer vision techniques for facilitating robotized wire harness assembly with better robotic visual perception of the wire harness components to be manipulated and the manipulation operations. Different vision-based solutions were proposed for various sub-tasks of robotized wire harness assembly, including manipulating clamps and connectors, identifying wire harness bags, monitoring the mating process of connectors, fault detection during assembly, and bin-picking problem by detecting different wire harness components, including clamps, connectors, wires, and wire harness bags.

Based on past research, this article identified two major challenges for computer vision applications in the robotic assembly of wire harnesses:

1. The robustness of vision systems in actual production has not achieved the compatible level as humans, especially considering the demanding production rate and intricate production environments;
2. Intrinsic physical properties of different wire harness components were identified as hindrances to robotic visual perception.

Furthermore, this article proposed prospective research directions toward more efficient and practical computer vision applications in robotized wire harness assembly:

1. Developing and implementing learning-based computer vision techniques to exploit intrinsic features and multi-modality data of wire harnesses
2. Investigating the adaptation of computer vision techniques proposed for the robotized assembly process of manufacturing wire harnesses
3. Evaluating the practicality, robustness, reliability, and sustainability of the proposed vision systems regarding practical manufacturing scenarios
4. Inquiring novel vision system designs considering human-robot collaborations and different assembly operations
5. Exploring new wire harness designs for facilitating more efficient visual recognition

This article also advocates addressing barriers to the industrialization of technologies, which demands investigation and discussion among researchers and practitioners from various backgrounds.

Acknowledgements

This work was supported by the Swedish innovation agency, Vinnova, and the strategic innovation program, Produktion2030, under grant number 2022-01279. The work was carried out within Chalmers Production Area of Advance. The support is gratefully acknowledged.

References

[1] B. L. Žagar, A. Caporali, A. Szymko, P. Kicki, K. Walas, G. Palli, A. C. Knoll, Copy and paste augmentation for deformable wiring harness bags segmentation, in: 2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), 2023, pp. 721–726. doi:10.1109/AIM46323.2023.10196168.

[2] T. Hermansson, R. Bohlin, J. S. Carlson, R. Söderberg, Automatic assembly path planning for wiring harness installations, Journal of Manufacturing Systems 32 (3) (2013) 417–422. doi:10.1016/j.jmsy.2013.04.006.

[3] O. Salunkhe, W. Quadrini, H. Wang, J. Stahre, D. Romero, L. Fumagalli, D. Lämkull, Review of current status and future directions for collaborative and semi-automated automotive wire harnesses assembly, Procedia CIRP 120 (2023) 696–701. doi:10.1016/j.procir.2023.09.061.

[4] European Commission, Directorate-General for Communication, European green deal - Delivering on our targets, Publications Office of the European Union, Luxembourg, 2021. doi:10.2775/373022.

[5] European Environment Agency, Decarbonising road transport - The role of vehicles, fuels and transport demand, Publications Office of the European Union, Luxembourg, 2022. doi:10.2800/68902.

[6] S. Fankhauser, S. M. Smith, M. Allen, K. Axelsson, T. Hale, C. Hepburn, J. M. Kendall, R. Khosla, J. Lezaun, E. Mitchell-Larson, M. Obersteiner, L. Rajamani, R. Rickaby, N. Seddon, T. Wetzer, The meaning of net zero and how to get it right, Nature Climate Change 12 (1) (2022) 15–21. doi:10.1038/s41558-021-01245-w.

[7] F. J. Riley, Assembly automation: a management handbook, 2nd Edition, Industrial Press Inc., New York, NY, USA, 1996.

[8] G. Boothroyd, Assembly automation and product design, 2nd Edition, CRC Press, Boca Raton, FL, USA, 2005. doi:10.1201/9781420027358.

[9] S. J. Hu, J. Ko, L. Weyand, H. A. ElMaraghy, T. Lien, Y. Koren, H. Bley, G. Chryssolouris, N. Nasr, M. Shpitalni, Assembly system design and operations for product variety, CIRP Annals 60 (2) (2011) 715–733. doi:10.1016/j.cirp.2011.05.004.

[10] H. Chen, B. Zhang, G. Zhang, Robotic assembly, in: A. Y. C. Nee (Ed.), Handbook of Manufacturing Engineering and Technology, Springer, London, UK, 2015, pp. 2347–2401. doi:10.1007/978-1-4471-4670-4_105.

[11] J. Leng, W. Sha, B. Wang, P. Zheng, C. Zhuang, Q. Liu, T. Wuest, D. Mourtzis, L. Wang, Industry 5.0: Prospect and retrospect, Journal of Manufacturing Systems 65 (2022) 279–295. doi:10.1016/j.jmsy.2022.09.017.

[12] M. T. Mason, Toward robotic manipulation, Annual Review of Control, Robotics, and Autonomous Systems 1 (1) (2018) 1–28. doi:10.1146/annurev-control-060117-104848.

[13] M. Saadat, P. Nan, Industrial applications of automatic manipulation of flexible materials, Industrial Robot: An International Journal 29 (5) (2002) 434–442. doi:10.1108/01439910210440255.

[14] S. Makris, F. Dietrich, K. Kellens, S. J. Hu, Automated assembly of non-rigid objects, CIRP Annals 72 (2) (2023) 513–539. doi:10.1016/j.cirp.2023.05.003.

[15] M. Rambow, T. Schauß, M. Buss, S. Hirche, Autonomous manipulation of deformable objects based on teleoperated demonstrations, in: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp. 2809–2814. doi:10.1109/IROS.2012.6386002.

[16] J. Shi, B. Hamner, R. Simmons, S. Singh, Mobile robotic assembly on a moving vehicle, in: International Symposium on Flexible Automation, Vol. 45110, 2012, pp. 117–123. doi:10.1115/ISFA2012-7193.

[17] A. Billard, D. Kragic, Trends and challenges in robot manipulation, Science 364 (6446) (2019) eaat8414. doi:10.1126/science.aat8414.

[18] E. Martinez-Martin, A. P. del Pobil, Vision for robust robot manipulation, Sensors 19 (7) (2019) 1648. doi:10.3390/s19071648.

[19] M. T. Shahria, M. S. H. Sunny, M. I. I. Zarif, J. Ghommam, S. I. Ahamed, M. H. Rahman, A comprehensive review of vision-based robotic applications: Current state, components, approaches, barriers, and potential solutions, Robotics 11 (6) (2022) 139. doi:10.3390/robotics11060139.

[20] L. Zhou, L. Zhang, N. Konz, Computer vision techniques in manufacturing, IEEE Transactions on Systems, Man, and Cybernetics: Systems 53 (1) (2023) 105–117. doi:10.1109/TSMC.2022.3166397.

[21] H. Wang, O. Salunkhe, W. Quadrini, D. Lämkull, F. Ore, B. Johansson, J. Stahre, Overview of computer vision techniques in robotized wire harness assembly: Current state and future opportunities, Procedia CIRP 120 (2023) 1071–1076. doi:10.1016/j.procir.2023.09.127.

[22] J. Trommnau, J. Kühnle, J. Siegert, R. Inderka, T. Bauernhansl, Overview of the state of the art in the production process of automotive wire harnesses, current research and future trends, Procedia CIRP 81 (2019) 387–392. doi:10.1016/j.procir.2019.03.067.

[23] H. G. Nguyen, M. Kuhn, J. Franke, Manufacturing automation for automotive wiring harnesses, Procedia CIRP 97 (2021) 379–384. doi:10.1016/j.procir.2020.05.254.

[24] G. E. Navas-Reascos, D. Romero, J. Stahre, A. Caballero-Ruiz, Wire harness assembly process supported by collaborative robots: Literature review and call for r&d, Robotics 11 (3) (2022) 65. doi:10.3390/robotics11030065.

[25] E. Aguirre, B. Raucent, Economic comparison of wire harness assembly systems, Journal of Manufacturing Systems 13 (4) (1994) 276–288. doi:10.1016/0278-6125(94)90035-3.

[26] J. Tilindis, V. Kleiza, The effect of learning factors due to low volume order fluctuations in the automotive wiring harness production, Procedia CIRP 19 (2014) 129–134. doi:10.1016/j.procir.2014.05.019.

[27] M. Wnuk, C. Hinze, M. Zürn, Q. Pan, A. Lechler, A. Verl, Tracking branched deformable linear objects with structure preserved registration by branch-wise probability modification, in: 2021 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), 2021, pp. 101–108. doi:10.1109/M2VIP49856.2021.9665147.

[28] S. M. Research, Automotive wiring harness market: By material (copper, aluminum, others), by vehicle type (passenger cars, two-wheelers, commercial vehicles), by application (chassis harness, engine harness, body & lighting harness, dashboard/ cabin harness, sunroof harness, battery harness, hvac harness, seat harness, door harness), by geography, size, global share, forecast, 2021-2030, Tech. rep., Strategic Market Research (2022).
[29] N. Sylla, V. Bonnet, F. Colledani, P. Fraisse, Ergonomic contribution of able exoskeleton in automotive industry, International Journal of Industrial Ergonomics 44 (4) (2014) 475–481. doi:10.1016/j.ergon.2014.03.008.
[30] D. Romero, J. Stahre, T. Wuest, O. Noran, P. Bernus, Å. Fast-Berglund, D. Gorecky, Towards an operator 4.0 typology: A human-centric perspective on the fourth industrial revolution technologies, in: CIE 2016: 46th International Conferences on Computers and Industrial Engineering, 2016, pp. 1–11.
[31] C. Fischer, J. Bönig, J. Franke, M. Lušić, R. Hornfeck, Worker information system to support during complex and exhausting assembly of high-voltage harness, in: 2015 5th International Electric Drives Production Conference (EDPC), 2015, pp. 1–7. doi:10.1109/EDPC.2015.7323211.
[32] S. Olbrich, J. Lackinger, Manufacturing processes of automotive high-voltage wire harnesses: State of the art, current challenges and fields of action to reach a higher level of automation, Procedia CIRP 107 (2022) 653–660. doi:10.1016/j.procir.2022.05.041.
[33] K. Galassi, G. Palli, Robotic wires manipulation for switchgear cabling and wiring harness manufacturing, in: 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS), 2021, pp. 531–536. doi:10.1109/ICPS49255.2021.9468128.
[34] N. Lv, J. Liu, Y. Jia, Dynamic modeling and control of deformable linear objects for single-arm and dual-arm robot manipulations, IEEE Transactions on Robotics 38 (4) (2022) 2341–2353. doi:10.1109/TRO.2021.3139838.
[35] J. Sanchez, J.-A. Corrales, B.-C. Bouzgarrou, Y. Mezouar, Robotic manipulation and sensing of deformable objects in domestic and industrial applications: a survey, The International Journal of Robotics Research 37 (7) (2018) 688–716. doi:10.1177/0278364918779698.
[36] H. Yin, A. Varava, D. Kragic, Modeling, learning, perception, and control methods for deformable object manipulation, Science Robotics 6 (54) (2021) eabd8803. doi:10.1126/scirobotics.abd8803.
[37] J. Zhu, A. Cherubini, C. Dune, D. Navarro-Alarcon, F. Alambeigi, D. Berenson, F. Ficuciello, K. Harada, J. Kober, X. Li, J. Pan, W. Yuan, M. Gienger, Challenges and outlook in robotic manipulation of deformable objects, IEEE Robotics & Automation Magazine 29 (3) (2022) 67–77. doi:10.1109/MRA.2022.3147415.
[38] F. Gu, Y. Zhou, Z. Wang, S. Jiang, B. He, A survey on robotic manipulation of deformable objects: Recent advances, open challenges and new frontiers (2023). doi:10.48550/arXiv.2312.10419.
[39] F. Nadon, A. J. Valencia, P. Payeur, Multi-modal sensing and robotic manipulation of non-rigid objects: A survey, Robotics 7 (4) (2018) 74. doi:10.3390/robotics7040074.
[40] X. Li, X. Su, Y.-H. Liu, Vision-based robotic manipulation of flexible pcbs, IEEE/ASME Transactions on Mechatronics 23 (6) (2018) 2739–2749. doi:10.1109/TMECH.2018.2869147.
[41] M. C. Gemici, A. Saxena, Learning haptic representation for manipulating deformable food objects, in: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014, pp. 638–645. doi:10.1109/IROS.2014.6942626.
[42] L. Cao, X. Li, P. T. Phan, A. M. H. Tiong, H. L. Kaan, J. Liu, W. Lai, Y. Huang, H. M. Le, M. Miyasaka, K. Y. Ho, P. W. Y. Chiu, S. J. Phee, Sewing up the wounds: A robotic suturing system for flexible endoscopy, IEEE Robotics & Automation Magazine 27 (3) (2020) 45–54. doi:10.1109/MRA.2019.2963161.
[43] F. Zhang, Y. Demiris, Learning garment manipulation policies toward robot-assisted dressing, Science Robotics 7 (65) (2022) eabm6010. doi:10.1126/scirobotics.abm6010.
[44] S. Javdani, S. Tandon, J. Tang, J. F. O'Brien, P. Abbeel, Modeling and perception of deformable one-dimensional objects, in: 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 1607–1614. doi:10.1109/ICRA.2011.5980431.
[45] A. Keipour, M. Bandari, S. Schaal, Deformable one-dimensional object detection for routing and manipulation, IEEE Robotics and Automation Letters 7 (2) (2022) 4329–4336. doi:10.1109/LRA.2022.3146920.
[46] C. Chen, Y. Zheng, Deformation identification and estimation of one-dimensional objects by using vision sensors, in: Proceedings. 1991 IEEE International Conference on Robotics and Automation, Vol. 3, 1991, pp. 2306–2311. doi:10.1109/ROBOT.1991.131538.
[47] W. Wang, D. Berenson, D. Balkcom, An online method for tight-tolerance insertion tasks for string and rope, in: 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015, pp. 2488–2495. doi:10.1109/ICRA.2015.7139532.
[48] J.-E. Byun, T. i Nagata, Determining the 3-d pose of a flexible object by stereo matching of curvature representations, Pattern Recognition 29 (8) (1996) 1297–1307. doi:10.1016/0031-3203(95)00165-4.
[49] M. A. Lee, Y. Zhu, P. Zachares, M. Tan, K. Srinivasan, S. Savarese, L. Fei-Fei, A. Garg, J. Bohg, Making sense of vision and touch: Learning multimodal representations for contact-rich tasks, IEEE Transactions on Robotics 36 (3) (2020) 582–596. doi:10.1109/TRO.2019.2959445.
[50] A. Nair, D. Chen, P. Agrawal, P. Isola, P. Abbeel, J. Malik, S. Levine, Combining self-supervised learning and imitation for vision-based rope manipulation, in: 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 2146–2153. doi:10.1109/ICRA.2017.7989247.
[51] M. Yan, Y. Zhu, N. Jin, J. Bohg, Self-supervised learning of state estimation for manipulating deformable linear objects, IEEE Robotics and Automation Letters 5 (2) (2020) 2372–2379. doi:10.1109/LRA.2020.2969931.
[52] H. Zhou, S. Li, Q. Lu, J. Qian, A practical solution to deformable linear object manipulation: A case study on cable harness connection, in: 2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM), 2020, pp. 329–333. doi:10.1109/ICARM49381.2020.9195380.
[53] D. Andronas, Z. Arkouli, N. Zacharaki, G. Michalos, A. Sardelis, G. Papanikolopoulos, S. Makris, On the perception and handling of deformable objects - a robotic cell for white goods industry, Robotics and Computer-Integrated Manufacturing 77 (2022) 102358. doi:10.1016/j.rcim.2022.102358.
[54] International Organization for Standardization, Iso 10218-1:2011 robots and robotic devices — safety requirements for industrial robots — part 1: Robots (2011).
[55] International Organization for Standardization, Iso 10218-2:2011 robots and robotic devices — safety requirements for industrial robots —

part 2: Robot systems and integration (2011).
[56] International Organization for Standardization, Iso/ts 15066:2016 robots and robotic devices — collaborative robots (2016).
[57] G. Michalos, S. Makris, N. Papakostas, D. Mourtzis, G. Chryssolouris, Automotive assembly technologies review: challenges and outlook for a flexible and adaptive approach, CIRP Journal of Manufacturing Science and Technology 2 (2) (2010) 81–91. doi:10.1016/j.cirpj.2009.12.001.
[58] S. Heydaryan, J. Suaza Bedolla, G. Belingardi, Safety design and development of a human-robot collaboration assembly process in the automotive industry, Applied Sciences 8 (3) (2018) 344. doi:10.3390/app8030344.
[59] K. Koo, X. Jiang, K. Kikuchi, A. Konno, M. Uchiyama, Development of a robot car wiring system, in: 2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 2008, pp. 862–867. doi:10.1109/AIM.2008.4601774.
[60] X. Zhang, Y. Domae, W. Wan, K. Harada, Learning efficient policies for picking entangled wire harnesses: An approach to industrial bin picking, IEEE Robotics and Automation Letters 8 (1) (2023) 73–80. doi:10.1109/LRA.2022.3222995.
[61] B. Sun, F. Chen, H. Sasaki, T. Fukuda, Robotic wiring harness assembly system for fault-tolerant electric connectors mating, in: 2010 International Symposium on Micro-NanoMechatronics and Human Science, 2010, pp. 202–205. doi:10.1109/MHS.2010.5669533.
[62] X. Jiang, Y. Nagaoka, K. Ishii, S. Abiko, T. Tsujita, M. Uchiyama, Robotized recognition of a wire harness utilizing tracing operation, Robotics and Computer-Integrated Manufacturing 34 (2015) 52–61. doi:10.1016/j.rcim.2014.12.002.
[63] B. Kitchenham, Procedures for performing systematic reviews, Tech. Rep. TR/SE-0401, Keele University, Keele, Staffs, UK (2004).
[64] J. Rowley, F. Slack, Conducting a literature review, Management research news 27 (6) (2004) 31–39. doi:10.1108/01409170410784185.
[65] D. Denyer, D. Tranfield, Producing a systematic review, in: D. A. Buchanan, A. Bryman (Eds.), The SAGE Handbook of Organizational Research Methods, Sage Publications Ltd., 2009, pp. 671–689.
[66] N. K. Denzin, The research act: A theoretical introduction to sociological methods, 1st Edition, Routledge, New York, NY, USA, 2009. doi:10.4324/9781315134543.
[67] V. J. Janesick, Peer debriefing, in: G. Ritzer (Ed.), The Blackwell Encyclopedia of Sociology, John Wiley & Sons, Ltd, 2015. doi:10.1002/9781405165518.wbeosp014.pub2.
[68] M. W. Tracey, Design and development research: a model validation case, Educational Technology Research and Development 57 (4) (2009) 553–571. doi:10.1007/s11423-007-9075-0.
[69] P.-N. Chou, A comparison study of impact factor in web of science and scopus databases for engineering education and educational technology journals, Issues in Informing Science and Information Technology 9 (2012) 187–194. doi:10.28945/1615.
[70] A.-W. Harzing, S. Alakangas, Google scholar, scopus and the web of science: a longitudinal and cross-disciplinary comparison, Scientometrics 106 (2) (2016) 787–804. doi:10.1007/s11192-015-1798-9.
[71] A. Martín-Martín, M. Thelwall, E. Orduna-Malea, E. Delgado López-Cózar, Google scholar, microsoft academic, scopus, dimensions, web of science, and opencitations' coci: a multidisciplinary comparison of coverage via citations, Scientometrics 126 (1) (2021) 871–906. doi:10.1007/s11192-020-03690-4.
[72] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, L. A. McGuinness, L. A. Stewart, J. Thomas, A. C. Tricco, V. A. Welch, P. Whiting, D. Moher, The prisma 2020 statement: an updated guideline for reporting systematic reviews, BMJ 372 (n71) (2021). doi:10.1136/bmj.n71.
[73] T. Greenhalgh, R. Peacock, Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources, BMJ 331 (7524) (2005) 1064–1065. doi:10.1136/bmj.38636.593461.68.
[74] R. Adcock, D. Collier, Measurement validity: A shared standard for qualitative and quantitative research, American Political Science Review 95 (3) (2001) 529–546. doi:10.1017/S0003055401003100.
[75] J. W. Creswell, Research design: qualitative, quantitative, and mixed methods approaches, 4th Edition, SAGE Publications, Inc., Thousand Oaks, CA, USA, 2014.
[76] H. K. Mohajan, Two criteria for good measurements in research: Validity and reliability, Annals of Spiru Haret University 17 (3) (2017) 58–82. doi:10.26458/1746.
[77] C. M. Patino, J. C. Ferreira, Internal and external validity: can you apply research study results to your patients?, Jornal brasileiro de pneumologia 44 (3) (2018) 183. doi:10.1590/S1806-37562018000000164.
[78] L. Sürücü, A. Maslakci, Validity and reliability in quantitative research, Business & Management Studies: An International Journal 8 (3) (2020) 2694–2726. doi:10.15295/bmij.v8i3.1540.
[79] G. R. Gibbs, Analyzing Qualitative Data, SAGE Publications, Ltd, London, UK, 2007. doi:10.4135/9781849208574.
[80] M. Petticrew, F. Song, P. Wilson, K. Wright, Quality-assessed reviews of health care interventions and the database of abstracts of reviews of effectiveness (dare), International Journal of Technology Assessment in Health Care 15 (4) (1999) 671–678. doi:10.1017/S0266462399015469.
[81] D. G. Lowe, Object recognition from local scale-invariant features, in: Proceedings of the Seventh IEEE International Conference on Computer Vision, Vol. 2, 1999, pp. 1150–1157. doi:10.1109/ICCV.1999.790410.
[82] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110. doi:10.1023/B:VISI.0000029664.99615.94.
[83] P. Di, J. Huang, F. Chen, H. Sasaki, T. Fukuda, Hybrid vision-force guided fault tolerant robotic assembly for electric connectors, in: 2009 International Symposium on Micro-NanoMechatronics and Human Science, 2009, pp. 86–91. doi:10.1109/MHS.2009.5352078.
[84] X. Jiang, K.-m. Koo, K. Kikuchi, A. Konno, M. Uchiyama, Robotized assembly of a wire harness in car production line, in: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010, pp. 490–495. doi:10.1109/IROS.2010.5653133.
[85] H. Kato, M. Billinghurst, Marker tracking and hmd calibration for a video-based augmented reality conferencing system, in: Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99), 1999, pp. 85–94. doi:10.1109/IWAR.1999.803809.
[86] X. Jiang, K. mo Koo, K. Kikuchi, A. Konno, M. Uchiyama, Robotized assembly of a wire harness in a car production line, Advanced Robotics 25 (3-4) (2011) 473–489. doi:10.1163/016918610X551782.
[87] P. Di, F. Chen, H. Sasaki, J. Huang, T. Fukuda, T. Matsuno, Vision-force guided monitoring for mating connectors in wiring harness assembly

systems, Journal of Robotics and Mechatronics 24 (4) (2012) 666–676. doi:10.20965/jrm.2012.p0666.

[88] T. Tamada, Y. Yamakawa, T. Senoo, M. Ishikawa, High-speed manipulation of cable connector using a high-speed robot hand, in: 2013 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2013, pp. 1598–1604. doi:10.1109/ROBIO.2013.6739695.

[89] H.-C. Song, Y.-L. Kim, D.-H. Lee, J.-B. Song, Electric connector assembly based on vision and impedance control using cable connector-feeding system, Journal of Mechanical Science and Technology 31 (12) (2017) 5997–6003. doi:10.1007/s12206-017-1144-7.

[90] F. Yumbla, M. Abeyabas, T. Luong, J.-S. Yi, H. Moon, Preliminary connector recognition system based on image processing for wire harness assembly tasks, in: 2020 20th International Conference on Control, Automation and Systems (ICCAS), 2020, pp. 1146–1150. doi:10.23919/ICCAS50221.2020.9268291.

[91] P. Kicki, M. Bednarek, P. Lembicz, G. Mierzwiak, A. Szymko, M. Kraft, K. Walas, Tell me, what do you see?-interpretable classification of wiring harness branches with deep neural networks, Sensors 21 (13) (2021) 4327. doi:10.3390/s21134327.

[92] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2921–2929. doi:10.1109/CVPR.2016.319.

[93] M. Bednarek, P. Kicki, K. Walas, On robustness of multi-modal fusion-robotics perspective, Electronics 9 (7) (2020) 1152. doi:10.3390/electronics9071152.

[94] J. Guo, J. Zhang, Y. Gai, D. Wu, K. Chen, Visual recognition method for deformable wires in aircrafts assembly based on sequential segmentation and probabilisitic estimation, in: 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC), Vol. 6, 2022, pp. 598–603. doi:10.1109/ITOEC53115.2022.9734432.

[95] D. Dwibedi, I. Misra, M. Hebert, Cut, paste and learn: Surprisingly easy synthesis for instance detection, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1310–1319. doi:10.1109/ICCV.2017.146.

[96] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, B. Zoph, Simple copy-paste is a strong data augmentation method for instance segmentation, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 2917–2927. doi:10.1109/CVPR46437.2021.00294.

[97] F. Wang, H. Wang, C. Wei, A. Yuille, W. Shen, Cp^2: Copy-paste contrastive pretraining for semantic segmentation, in: Computer Vision – ECCV 2022, Springer, Cham, Switzerland, 2022, pp. 499–515. doi:10.1007/978-3-031-20056-4_29.

[98] C. Shorten, T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of Big Data 6 (1) (2019) 1–48. doi:10.1186/s40537-019-0197-0.

[99] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications (2017). doi:10.48550/arXiv.1704.04861.

[100] F. Chaumette, S. Hutchinson, Visual servo control. i. basic approaches, IEEE Robotics & Automation Magazine 13 (4) (2006) 82–90. doi:10.1109/MRA.2006.250573.

[101] Y. Domae, H. Okuda, Y. Taguchi, K. Sumi, T. Hirai, Fast graspability evaluation on single depth maps for bin picking with general grippers, in: 2014 IEEE International Conference on Robotics and Automation (ICRA), 2014, pp. 1997–2004. doi:10.1109/ICRA.2014.6907124.

[102] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, Cambridge, MA, USA, 2016.

[103] E. Romera, J. M. Álvarez, L. M. Bergasa, R. Arroyo, Erfnet: Efficient residual factorized convnet for real-time semantic segmentation, IEEE Transactions on Intelligent Transportation Systems 19 (1) (2018) 263–272. doi:10.1109/TITS.2017.2750080.

[104] D. Reynolds, Gaussian mixture models, in: S. Z. Li, A. Jain (Eds.), Encyclopedia of Biometrics, Springer, Boston, MA, USA, 2009, pp. 659–663. doi:10.1007/978-0-387-73003-5_196.

[105] A. Quattoni, A. Torralba, Recognizing indoor scenes, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 413–420. doi:10.1109/CVPR.2009.5206537.

[106] Y. Zeng, P. Zhang, Z. Lin, J. Zhang, H. Lu, Towards high-resolution salient object detection, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 7233–7242. doi:10.1109/ICCV.2019.00733.

[107] R. Zanella, A. Caporali, K. Tadaka, D. De Gregorio, G. Palli, Auto-generated wires dataset for semantic segmentation with domain-independence, in: 2021 International Conference on Computer, Control and Robotics (ICCCR), 2021, pp. 292–298. doi:10.1109/ICCCR49711.2021.9349395.

[108] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. doi:10.1109/CVPR.2016.90.

[109] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 9992–10002. doi:10.1109/ICCV48922.2021.00986.

[110] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A convnet for the 2020s, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 11966–11976. doi:10.1109/CVPR52688.2022.01167.

[111] M. Zürn, M. Wnuk, A. Lechler, A. Verl, Topology matching of branched deformable linear objects, in: 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023, pp. 7097–7103. doi:10.1109/ICRA48891.2023.10161483.

[112] M. Saha, P. Isto, Motion planning for robotic manipulation of deformable linear objects, in: Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006., 2006, pp. 2478–2484. doi:10.1109/ROBOT.2006.1642074.

[113] J. Schulman, A. Lee, J. Ho, P. Abbeel, Tracking deformable objects with point clouds, in: 2013 IEEE International Conference on Robotics and Automation, 2013, pp. 1130–1137. doi:10.1109/ICRA.2013.6630714.

[114] T. Tang, C. Wang, M. Tomizuka, A framework for manipulating deformable linear objects by coherent point drift, IEEE Robotics and Automation Letters 3 (4) (2018) 3426–3433. doi:10.1109/LRA.2018.2852770.

[115] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436–444. doi:10.1038/nature14539.

[116] Z. Zou, K. Chen, Z. Shi, Y. Guo, J. Ye, Object detection in 20 years: A survey, Proceedings of the IEEE 111 (3) (2023) 257–276. doi:10.1109/JPROC.2023.3238524.

[117] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587. doi:10.1109/CVPR.2014.81.

[118] R. Girshick, Fast r-cnn, in: 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448. doi:10.1109/ICCV.2015.169.

[119] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 28, Curran Associates, Inc., 2015, pp. 91–99.
[120] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788. doi:10.1109/CVPR.2016.91.
[121] Y. Jin, D. Li, Y. A, J. Shi, P. Hao, F. Sun, J. Zhang, B. Fang, Robotgpt: Robot manipulation learning from chatgpt, IEEE Robotics and Automation Letters 9 (3) (2024) 2543–2550. doi:10.1109/LRA.2024.3357432.
[122] B. Xie, X. Xi, X. Zhao, Y. Wang, W. Song, J. Gu, S. Zhu, Chatgpt for robotics: A new approach to human-robot interaction and task planning, in: H. Yang, H. Liu, J. Zou, Z. Yin, L. Liu, G. Yang, X. Ouyang, Z. Wang (Eds.), Intelligent Robotics and Applications, Springer, Singapore, 2023, pp. 365–376.
[123] I. Hou, O. Man, S. Mettille, S. Gutierrez, K. Angelikas, S. MacNeil, More robots are coming: Large multimodal models (chatgpt) can solve visually diverse images of parsons problems, in: Proceedings of the 26th Australasian Computing Education Conference, Association for Computing Machinery, New York, NY, USA, 2024, pp. 29–38. doi:10.1145/3636243.3636247.
[124] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, M. Bennamoun, Deep learning for 3d point clouds: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (12) (2021) 4338–4364. doi:10.1109/TPAMI.2020.3005434.
[125] T. Zhang, Q. Zhang, E. Lim, J. Sun, Deep learning based 3d point clouds recognition for robotic manufacturing, in: 2022 27th International Conference on Automation and Computing (ICAC), 2022, pp. 1–5. doi:10.1109/ICAC55051.2022.9911175.
[126] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, International Journal of Computer Vision 88 (2009) 303–308. doi:10.1007/s11263-009-0275-4.
[127] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014, Springer, Cham, Switzerland, 2014, pp. 740–755. doi:10.1007/978-3-319-10602-1_48.
[128] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, Imagenet large scale visual recognition challenge, International Journal of Computer Vision 115 (2015) 211–252. doi:10.1007/s11263-015-0816-y.
[129] H. G. Nguyen, R. Habiboglu, J. Franke, Enabling deep learning using synthetic data: A case study for the automotive wiring harness manufacturing, Procedia CIRP 107 (2022) 1263–1268. doi:10.1016/j.procir.2022.05.142.
[130] K. Kimble, J. Albrecht, M. Zimmerman, J. Falco, Performance measures to benchmark the grasping, manipulation, and assembly of deformable objects typical to manufacturing applications, Frontiers in Robotics and AI 9 (2022). doi:10.3389/frobt.2022.999348.
[131] H. Wang, B. Johansson, Deep learning-based connector detection for robotized assembly of automotive wire harnesses, in: 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE), 2023, pp. 1–8. doi:10.1109/CASE56687.2023.10260619.
[132] L. Pérez, Í. Rodríguez, N. Rodríguez, R. Usamentiaga, D. F. García, Robot guidance using machine vision techniques in industrial environments: A comparative review, Sensors 16 (3) (2016) 335. doi:10.3390/s16030335.
[133] L. Bodenhagen, A. R. Fugl, A. Jordt, M. Willatzen, K. A. Andersen, M. M. Olsen, R. Koch, H. G. Petersen, N. Krüger, An adaptable robot vision system performing manipulation actions with flexible objects, IEEE Transactions on Automation Science and Engineering 11 (3) (2014) 749–765. doi:10.1109/TASE.2014.2320157.
[134] E. Aguirre, L. Ferrière, B. Raucent, Robotic assembly of wire harnesses: Economic and technical justification, Journal of Manufacturing Systems 16 (3) (1997) 220–231. doi:10.1016/S0278-6125(97)88890-5.
[135] L. Wang, S. Liu, H. Liu, X. V. Wang, Overview of human-robot collaboration in manufacturing, in: L. Wang, V. D. Majstorovic, D. Mourtzis, E. Carpanzano, G. Moroni, L. M. Galantucci (Eds.), Proceedings of 5th International Conference on the Industry 4.0 Model for Advanced Manufacturing, Springer, Cham, Switzerland, 2020, pp. 15–58. doi:10.1007/978-3-030-46212-3_2.
[136] European Commission, Directorate-General for Research and Innovation, J. Müller, Enabling Technologies for Industry 5.0: results of a workshop with Europe's technology leaders, Publications Office of the European Union, Luxembourg, 2020. doi:10.2777/082634.
[137] European Commission, Directorate-General for Research and Innovation, M. Breque, L. De Nul, A. Petridis, Industry 5.0: towards a sustainable, human-centric and resilient European industry, Publications Office of the European Union, Luxembourg, 2021. doi:10.2777/308407.
[138] European Commission, Directorate-General for Research and Innovation, A. Renda, S. Schwaag Serger, D. Tataj, A. Morlet, D. Isaksson, F. Martins, M. Mir Roca, C. Hidalgo, A. Huang, S. Dixson-Declève, P. Balland, F. Bria, C. Charveriat, K. Dunlop, E. Giovannini, Industry 5.0, a transformative vision for Europe: governing systemic transformations towards a sustainable industry, Publications Office of the European Union, Luxembourg, 2022. doi:10.2777/17322.
[139] A. Mohammed, B. Schmidt, L. Wang, Active collision avoidance for human-robot collaboration driven by vision sensors, International Journal of Computer Integrated Manufacturing 30 (9) (2017) 970–980. doi:10.1080/0951192X.2016.1268269.
[140] D. Kruse, R. J. Radke, J. T. Wen, Collaborative human-robot manipulation of highly deformable materials, in: 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015, pp. 3782–3787. doi:10.1109/ICRA.2015.7139725.
[141] P. Heisler, D. Utsch, M. Kuhn, J. Franke, Optimization of wire harness assembly using human-robot-collaboration, Procedia CIRP 97 (2021) 260–265. doi:10.1016/j.procir.2020.05.235.
[142] O. Salunkhe, J. Stahre, D. Romero, D. Li, B. Johansson, Specifying task allocation in automotive wire harness assembly stations for human-robot collaboration, Computers & Industrial Engineering 184 (2023) 109572. doi:10.1016/j.cie.2023.109572.
[143] R. Gerbers, K. Wegener, F. Dietrich, K. Dröder, Safe, flexible and productive human-robot-collaboration for disassembly of lithium-ion batteries, in: A. Kwade, J. Diekmann (Eds.), Recycling of Lithium-Ion Batteries: The LithoRec Way, Springer, Cham, Switzerland, 2018, pp. 99–126. doi:10.1007/978-3-319-70572-9_6.
[144] E. Kim, Y. Yamada, S. Okamoto, M. Sennin, H. Kito, Considerations of potential runaway motion and physical interaction for speed and separation monitoring, Robotics and Computer-Integrated Manufacturing 67 (2021) 102034. doi:10.1016/j.rcim.2020.102034.
[145] A. Hentout, M. Aouache, A. Maoudj, I. Akli, Human-robot interaction in industrial collaborative robotics: a literature review of the decade 2008-2017, Advanced Robotics 33 (15-16) (2019) 764–799. doi:10.1080/01691864.2019.1636714.

[146] S. Proia, R. Carli, G. Cavone, M. Dotoli, Control techniques for safe, ergonomic, and efficient human-robot collaboration in the digital industry: A survey, IEEE Transactions on Automation Science and Engineering 19 (3) (2022) 1798–1819. doi:10.1109/TASE.2021.3131011.

[147] S. Li, R. Wang, P. Zheng, L. Wang, Towards proactive human-robot collaboration: A foreseeable cognitive manufacturing paradigm, Journal of Manufacturing Systems 60 (2021) 547–552. doi:10.1016/j.jmsy.2021.07.017.

[148] S. Li, P. Zheng, J. Fan, L. Wang, Toward proactive human-robot collaborative assembly: A multimodal transfer-learning-enabled action prediction approach, IEEE Transactions on Industrial Electronics 69 (8) (2022) 8579–8588. doi:10.1109/TIE.2021.3105977.

[149] J. Fan, P. Zheng, S. Li, Vision-based holistic scene understanding towards proactive human-robot collaboration, Robotics and Computer-Integrated Manufacturing 75 (2022) 102304. doi:10.1016/j.rcim.2021.102304.

[150] S. Li, P. Zheng, S. Liu, Z. Wang, X. V. Wang, L. Zheng, L. Wang, Proactive human-robot collaboration: Mutual-cognitive, predictable, and self-organising perspectives, Robotics and Computer-Integrated Manufacturing 81 (2023) 102510. doi:10.1016/j.rcim.2022.102510.

[151] Y. Ye, H. You, J. Du, Improved trust in human-robot collaboration with chatgpt, IEEE Access 11 (2023) 55748–55754. doi:10.1109/ACCESS.2023.3282111.

[152] M.-C. Chiu, G. l. E. Okudan, Evolution of design for x tools applicable to design stages: A literature review, in: Proceedings of the ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 6, 2010, pp. 171–182. doi:10.1115/DETC2010-29091.

[153] K. Singer-Coudoux, G. Braun, J. Stahre, Managing change towards the future of work - clustering key perspectives, in: E. Alfnes, A. Romsdal, J. O. Strandhagen, G. von Cieminski, D. Romero (Eds.), Advances in Production Management Systems. Production Management Systems for Responsible Manufacturing, Service, and Logistics Futures, Vol. 689, Springer, Cham, Switzerland, 2023, pp. 513–527. doi:10.1007/978-3-031-43662-8_37.

[154] G. Braun, M. Järvinen, J. Stahre, R. Hämäläinen, Motivational challenges of engineers participating in an online upskilling program, Proceedings of the 21st European Conference on e-Learning - ECEL 2022 21 (1) (2022) 25–31. doi:10.34190/ecel.21.1.594.

[155] J. Lee, H. Davari, J. Singh, V. Pandhare, Industrial artificial intelligence for industry 4.0-based manufacturing systems, Manufacturing Letters 18 (2018) 20–23. doi:10.1016/j.mfglet.2018.09.002.

[156] N. Tuptuk, S. Hailes, Security of smart manufacturing systems, Journal of Manufacturing Systems 47 (2018) 93–106. doi:10.1016/j.jmsy.2018.04.007.

Paper 3

Deep Learning-Based Connector Detection for Robotized Assembly of Automotive Wire Harnesses

Hao Wang, Björn Johansson

Presented at IEEE 19th International Conference on Automation Science and Engineering (CASE 2023), Auckland, New Zealand, 26-30 August 2023

Published in: *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*

DOI: 10.1109/CASE56687.2023.10260619

Deep Learning-Based Connector Detection for Robotized Assembly of Automotive Wire Harnesses*

Hao Wang[1] and Björn Johansson[1]

2023 IEEE 19th International Conference on Automation Science and Engineering (CASE) | 979-8-3503-2069-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/CASE56687.2023.10260619

***Abstract*— The shift towards electrification and autonomous driving in the automotive industry results in more and more automotive wire harnesses being installed in modern automobiles, which stresses the great significance of guaranteeing the quality of automotive wire harness assembly. The mating of connectors is essential in the final assembly of automotive wire harnesses due to the importance of connectors on wire harness connection and signal transmission. However, the current manual operation of mating connectors leads to severe problems regarding assembly quality and ergonomics, where the robotized assembly has been considered, and different vision-based solutions have been proposed to facilitate a better perception of the robot control system on connectors. Nonetheless, there has been a lack of deep learning-based solutions for detecting automotive wire harness connectors in previous literature. This paper presents a deep learning-based connector detection for robotized automotive wire harness assembly. A dataset of twenty automotive wire harness connectors was created to train and evaluate a two-stage and a one-stage object detection model, respectively. The experiment results indicate the effectiveness of deep learning-based connector detection for automotive wire harness assembly but are limited by the design of the exteriors of connectors.**

I. Introduction

Electrification and autonomous driving have driven a paradigm shift in the current automotive industry, making the electronic system increasingly critical in modern automobiles. Numerous automotive wire harnesses have been installed in current vehicles as an essential infrastructure for supporting signal transmission within the electronic system. Meanwhile, more and more wire harnesses are expected to be installed, considering the increase of automotive wire harnesses in vehicles in the past decades and the paradigm shift in the industry. Thus, it is crucial to guarantee the quality of the assembly of automotive wire harnesses.

However, the current final assembly of automotive wire harnesses into vehicles remains mostly manual and skill-demanding, which makes it challenging to control and improve the quality and productivity of the assembly. Some manual operations also involve heavy lifting (for example, approximately 40 kg for some automotive wire harnesses) and high-pressure manual manipulations on different components of automotive wire harnesses, which poses severe ergonomic problems to human operators. In particular, the mating of connectors is one of the sub-process relating to ergonomic issues due to the repetitive high-pressure manual pressing in the assembly line. Fig. 1 demonstrates an example of an automotive wire harness, where red rectangles highlight connectors on the automotive wire harness.

*This work was supported by the Swedish innovation agency, Vinnova, and the strategic innovation program, Produktion2030, under grant no. 2022-01279. The work was carried out within Chalmers' Area of Advance Production. Wiretronic AB provided the connectors used in this study. The support is gratefully acknowledged.

[1]Hao Wang and Björn Johansson are with the Division of Production Systems, Department of Industrial and Materials Science, Chalmers University of Technology, Hörsalsvägen 7A, SE-412 96 Gothenburg, Sweden haowang@chalmers.se

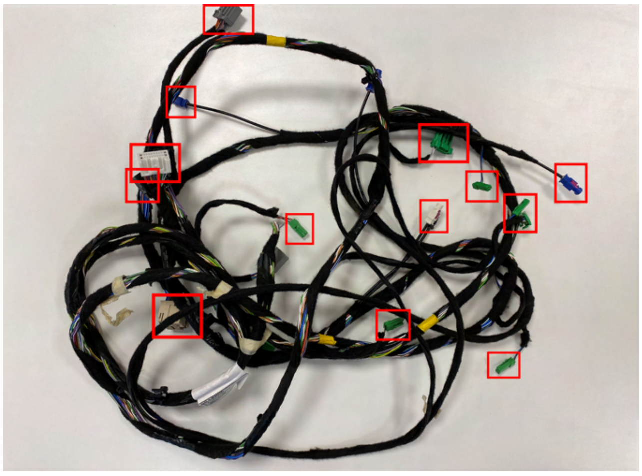


Fig. 1. An example of an automotive wire harness, with connectors highlighted by red rectangles.

Connectors are essential components on automotive wire harnesses, among the others, such as clamps and cables. Automotive wire harnesses are connected to the target unit or other automotive wire harnesses via connectors so the signal can be transmitted continuously within the electronic systems responsible for various functions of automobiles, which are safety-critical in particular. Thus, ensuring the quality of mating connectors in the final assembly of wire harnesses into vehicles is critical. However, the current manual process of mating connectors constrains the productivity and quality of assembly and generates ergonomic problems for human operators. To relieve the problems regarding productivity, assembly quality, and ergonomics, robotized wire harness assembly is of great interest to the automotive industry, considering its better replicability, transparency, and comprehensibility, and has been discussed in different studies previously [1], [2], [3], [4]. Nevertheless, the robotized mating of connectors is non-trivial as the robotic operation needs to address not only high manipulation accuracy but also intricate structures and non-rigid materials of connectors [5]. It is also fundamental to retrieve the geometrical information of connectors beforehand so that a robot arm can flexibly reach, grasp, and manipulate the perceived connector.

Computer vision has demonstrated a significant potential on the robotized assembly to the manufacturing industry in solving ergonomic issues while increasing quality and productivity [6]. There have been studies on computer vision techniques for robotized manipulation of wire harness connectors [5], [7], [8], [9], [10], [11], [12]. However, a few studies discussed connector detection [9], [11], [12], which mainly explored methods based on basic image processing techniques [9], [11]. Considering the various designs of connectors on automotive wire harnesses, such as colors, shapes, and sizes, it is intricate to manage the manual feature engineering on connectors for flexible robotized manipulation. The recent advancement in implementing convolutional neural networks (CNN) and deep learning in computer vision research has demonstrated the remarkable effectiveness of learning-based solutions for object detection compared to traditional image processing-based solutions [13]. Previously, Zhou et al. [12] explored deep learning-based connector detection for the robotized wire harness connection, but the proposal mainly focused on one-connector detection. The learning-based detection on multiple connectors remained unsolved but is required for the robotized assembly of automotive wire harnesses in actual production.

This paper presents a study on the deep learning-based connector detection for the robotized mating of connectors on automotive wire harnesses and discusses the feasibility and potential problems of implementing deep learning-based object detection on the task of mating connectors in robotized automotive wire harness assembly under laboratory conditions. As there is no publicly available dataset on automotive wire harness connectors, a dataset comprising twenty different types of connectors was collected initially. Then, two different detection models, a two-stage object detection model, Faster R-CNN [14], and a one-stage object detection model, YOLOv5 [15], were adopted for the training and inference. The experiment results demonstrate the effectiveness of deep learning-based connector detection as both detection methods achieved remarkable detection outcomes with various combinations of connectors presenting in the scene. Yet, detection performance can be improved further, and a more extensive dataset comprising more connectors and more images per connector is needed. Some detection errors on classes and positions of connectors in inference results further reflect the effect of the design of the exterior of connectors, which motivates the future connector detection based on multi-view images of connectors and with new exterior design of connectors so that more visually distinguishable features of the connector can be extracted.

This paper is organized in the following structure: Section II introduces the related research in connector detection and deep learning-based object detection. Section III introduces the data collection and annotation strategy and the statistics of the collected dataset of connectors. Section IV introduces the experiment setups of two-stage and one-stage connector detection, whose results are presented and further discussed in Section V. The study is concluded in Section VI with an outlook on the future work of this study.

II. Related Work

A. Connector Detection for Robotized Mating of Connectors

Connector detection is needed to acquire the position and categories of connectors so that the robot can flexibly reach, grasp, and manipulate connectors. Although some vision-based solutions have been proposed for facilitating different sub-tasks in robotized mating of connectors [5], [7], [8], [9], [10], [11], [12], connector detection has yet gathered few attention in previous studies [9], [11], [12], where the basic image processing-based methods are dominant [9], [11].

Tamada et al. [9] proposed to recognize the types and poses of connectors using a high-speed vision system. An image processing method was adopted in Tamada et al. [9] to detect the positions of connectors via detecting the corners of connectors, which was further processed to calculate the orientations of connectors. Yumbla et al. [11] later proposed a basic image processing-based method to detect multiple connectors, including converting color space and applying color thresholding. However, the task in Yumbla et al. [11] was a one-class detection, where all connectors were considered the same class. Deep learning-based connector detection has also been discussed in a recent study [12], which proposed to roughly locate the position of a connector and then zoom in to the detected connector to acquire the finer pose of the connector. Nevertheless, the proposal in Zhou et al. [12] mainly focused on manipulating one pair of connectors instead of multi-connector manipulation, which is more common in actual production.

B. Deep Learning-Based Object Detection

The rebirth of convolutional neural networks (CNNs) in 2012 [16] initiated the research on introducing deep learning [13] to object detection [17], which further promoted the remarkable development of two major groups of detectors for object detection based on deep learning in previous years: two-stage detection and one-stage detection [17].

Similar to the attentional mechanism of the human brain, the two-stage detection model first scans the whole scenario coarsely and then focuses on regions of interest (ROIs) to distinguish the object [18]. The region-based convolutional neural network (R-CNN) proposed by Girshick et al. [19], [20] symbolized the inauguration of two-stage object detection. In R-CNN [19], [20], a set of object proposals were extracted and fed into a CNN model to extract features for classification. However, the redundant feature computations due to many overlapped proposals made the detection speed extremely slow, which was improved later by Spatial Pyramid Pooling Networks (SPPNet) [21]. A Spatial Pyramid Pooling (SPP) layer was introduced in SPPNET [21] to enable a CNN to generate a fixed-length representation to avoid re-scaling. Nevertheless, SPPNET [21] remained multi-stage training and only fine-tuning fully-connected layers. To improve R-CNN [19] and SPPNet [21], Fast R-CNN [22] was proposed later, where the detector and the bounding box regressor could be trained under the same network configurations simultaneously. Furthermore, Faster

R-CNN [14] was proposed to accelerate the detection further by introducing a Region Proposal Network (RPN), but the problem of computation redundancy remained at the subsequent detection stage. Besides the R-CNN family, Lin et al. [23] proposed Feature Pyramid Networks (FPNs), which can be integrated into other detectors to enable high-level semantics building at all scales besides the feature maps of the networks' top layer.

Though able to attain high-precision detection, two-stage detection methods are constrained by their ponderous detection speed and computation, stimulating the research on one-stage detection. You Only Look Once (YOLO) [24] was the first deep learning-based one-stage detection that simultaneously predicted bounding boxes and probabilities for each sub-region of an image. Although the detection speed was improved significantly, the localization accuracy dropped remarkably compared to two-stage detection, especially for some small objects, which was enhanced in YOLO's subsequent versions [25], [26], [27], [28]. There were also other one-stage detection methods besides the YOLO family proposed to improve the detection accuracy while maintaining the advantage of high detection speed, including Single-Shot Multibox Detector (SSD) [29], RetinaNet [30], and CornerNet [31].

Recent years have also witnessed the profound influence of Transformer models [32] in deep learning and computer vision [33], which has spawned DEtection TRansformer (DETR) [34] and its successors, such as Deformable DETR [35], DINO [36], and Mask DINO [37], and promoted deep learning-based object detection to higher performance.

III. The Connector Dataset

The dataset is essential for learning-based object detection [38], [39], [40] and scalable deep learning-based solutions in industry [41]. However, to the best of the authors' knowledge, no publicly available benchmark dataset is dedicated to automotive wire harness connector detection. Thus, to facilitate the study of deep learning-based connector detection for the robotized assembly of automotive wire harnesses, a dataset was collected and annotated first, consisting of 20 types of connectors from four different product series commonly occurring on automotive wire harnesses installed in passenger vehicles. Fig. 2 demonstrates one example image for each of the 20 connectors. The following subsections will introduce the strategy for image collection and annotation and summarize the statistics of the dataset used in the experiments.

A. Image Collection Procedure

Connectors are placed on a white workbench for image acquisition using the main camera of an iPhone 11. The original image format is RGB, and each image has a size of 4032×3024 pixels. The distance between the camera and connectors varied between $20\,cm$ and $40\,cm$, considering the various locations of connectors in the three-dimensional (3D) space in actual assembly situations.

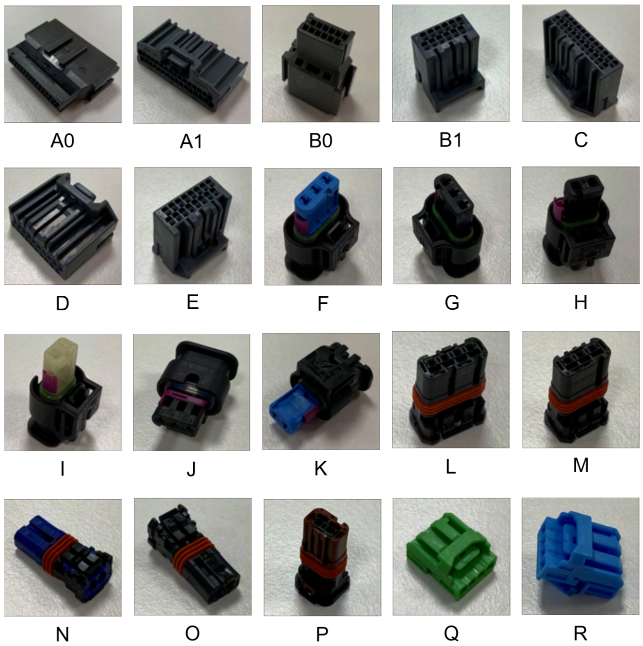


Fig. 2. The twenty types of connectors collected for dataset creation. The class of each connector is simplified and labeled below images.

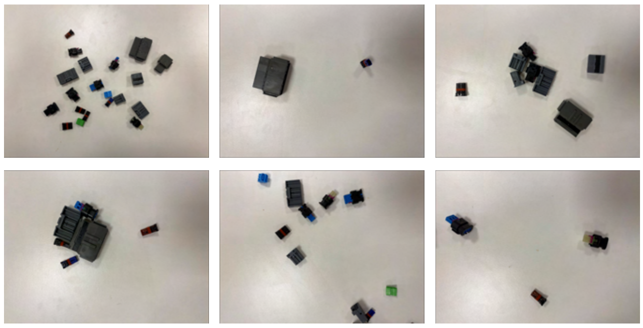


Fig. 3. Examples of images with different combinations of connectors.

The collected dataset of connectors contains 360 images. Initially, 60 birds'-eye view images of various combinations of connectors with random poses were collected to simulate the random distribution of connectors in the actual assembly scenario. Fig. 3 demonstrates some examples of these 60 images. For clarification, the distribution of connectors in these 60 images does not represent the actual distribution of connectors on practical automotive wire harnesses or in the final assembly of automotive wire harnesses.

In addition, images of each of the 20 connectors were also collected to train the detector with more features of respective classes, considering the positive contribution of batch-training on single-object scenes to detection performed on multi-object scenes [19]. For each connector, 15 images were captured from different views, including six images captured from the front, back, top, down, left, and right of the connector, and nine images captured from random views, as an example of class A0 shown in Fig. 4 and Fig. 5.

TABLE I
THE NUMBERS OF ANNOTATED OBJECT INSTANCES IN THE COLLECTED CONNECTOR DATASET.

Series	MX34							MCON 1.2 LL						MX80					MX81	
Class	A0	A1	B0	B1	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Train	39	38	29	36	33	35	33	34	36	32	32	33	38	40	26	37	29	36	36	28
Validation	1	1	3	1	2	1	1	3	3	3	3	2	1	5	2	2	2	2	2	3
Test	4	4	2	3	4	2	2	2	3	2	1	2	2	1	3	3	4	3	1	5
Total	44	43	34	40	39	38	36	39	42	37	36	37	41	46	31	42	35	41	39	36

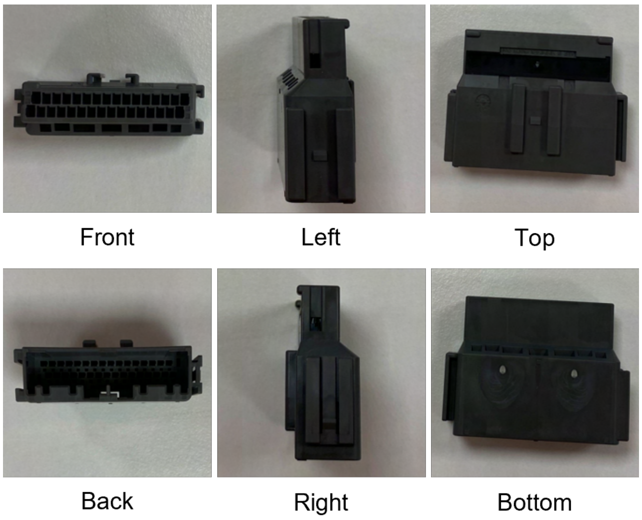


Fig. 4. The six images captured from the front, back, top, down, left, and right of A0. These images are cropped from the raw data for demonstration.

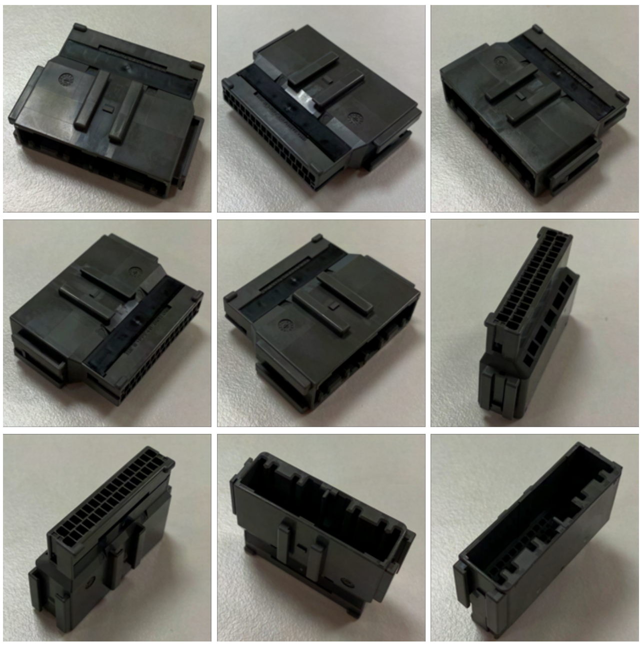


Fig. 5. The other nine images of A0 were captured from random perspectives. These images are cropped from the raw data for demonstration.

B. Image Annotation Procedure

The image annotation procedure of the dataset of connectors followed the methodology implemented in the PASCAL visual object classes (VOC) challenge 2007 [38].

The image annotation includes the **class** and the **bounding box** for every connector in the target set of classes. As shown in Fig. 2, this study simplified the 20 classes of connectors into A0, A1, B0, B1, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, and R, which can be easily mapped to the actual types of connectors in practical applications. An axis-aligned rectangular bounding box surrounding the connector was drawn for each connector visible in each image in the dataset. Though relatively quick to annotate, choosing an axis-aligned rectangular bounding box for the annotation is a compromise. Some connectors in images fit well because of their rectangular or approximately rectangular profiles, for example, class A0 shown in Fig. 4. However, for other connectors presented in images, an axis-aligned bounding box can be a poor fit because either they are not axis-aligned, for example, been captured from random perspectives (Fig. 5) or placed randomly (Fig. 3), or the connector is not in the shape of a box, for example, class I shown in Fig. 2.

The actual image annotation was conducted using an annotation platform, “Labelme” [42]. It was trivial to annotate images with a single connector due to the structured storage of images. For images with multiple connectors, a list of visible connectors in each image was documented first during the image collection procedure. Then, each connector visible in the images was compared with the original physical counterpart and annotated exhaustively. The annotation results were compared to the documented list to guarantee the consistency and accuracy of the image annotation.

C. Dataset Statistics

The total number of annotated images is 360. The data is primarily divided into three main subsets: training data (Train), validation data (Validation), and test data (Test), with a ratio of 90%/5%/5%. The images in the validation set and test set were selected randomly. For each subset of the connector dataset and class of connectors, the number of object instances is shown in TABLE I. In the collected dataset, the most frequent class is “L”, with 46 object instances, and the least frequent class is “M”, with 31 object instances. Fig. 6 illustrates a histogram of the number of object instances presented in different subsets of the collected connector dataset for each class of connectors.

Fig. 6. Histogram of the numbers of object instances shown in the collected connector dataset. The classes and the corresponding counts are shown on the x-axis and the y-axis, respectively.

IV. Experiment Settings

This study investigated a two-stage detector and a one-stage detector for automotive wire harness connector detection. The experiment on two-stage detection was conducted based on Faster R-CNN [14], and the one-stage detection was achieved based on YOLO [24]. Both models were trained using the union of the train and validation set of the collected connector dataset and evaluated on the test set using an NVIDIA GeForce RTX 4090. The following subsections introduce the detailed implementation of the two-stage and one-stage detection, respectively.

A. Data augmentation

In general, better deep learning models are achievable with larger datasets [43], [44]. Data augmentation is a technique to inflate existing training data with artificially-generated training data and has drawn many research efforts to address applications with limited datasets [45]. This study implemented several image data augmentation techniques to increase the amount of training data, including image HSV-Hue augmentation (0.015, fraction), image HSV-Saturation augmentation (0.7, fraction), HSV-Value augmentation (0.4, fraction), image translation (± 0.1, fraction), image scaling (± 0.5, gain), horizontal flipping (0.5, probability), and mosaic technique (1.0, probability).

B. Two-Stage Detection

This study investigated the two-stage detection based on Faster R-CNN [14] and implemented Faster R-CNN [14] with ResNet [46] plus Feature Pyramid Network (FPN) [23] as the backbone. The overall baselines and hyper-parameters followed Faster R-CNN [14] provided in the publicly available code of Detectron2 [47]. Specifically, the model was trained with a learning rate of 0.00025 using Stochastic Gradient Descent (SGD) as the optimizer. The batch size was 8. The weights of the model were initiated with the pre-trained checkpoint, *faster_R-CNN_R_101_FPN_3x*, provided by Detectron2 [47].

Several sub-stages of processing are involved in the Faster R-CNN [14] algorithm. In one of these sub-stages, which classifies regions of an image as either object or background, a threshold value is required to be set by the user to determine the confidence score needed for a region to be considered as an object, i.e., a region is considered as background and discarded if its confidence score is below the threshold value, otherwise, an object and retained. This study set two threshold values, 0.1 and 0.5, to evaluate the Faster R-CNN-based model.

C. One-Stage Detection

YOLO [24] was selected as the backbone of the one-stage detection in the experiment. The overall baselines and hyper-parameters of the one-stage detection in this study followed the publicly available code of YOLOv5 [15]. Specifically, the model was trained with an initial learning rate of 0.01 using SGD as the optimizer. The weight decay was 0.0005, and the momentum was 0.937. The batch size was 16. The model weights were initiated with the pre-trained checkpoint, *yolov5x*, provided by YOLOv5 [15]. An early-stop module was adopted to control the end of the training process, which terminated the training if there was no improvement after 300 consecutive epochs.

V. Results and Discussion

The initialization, training, and evaluation of the two-stage detection model based on Faster R-CNN [14] and the one-stage detection model based on YOLOv5 [15] were conducted following the experiment protocol explained in section IV. Fig. 7 demonstrates some inference results of Faster R-CNN [14] with two threshold values and YOLOv5 [15] as well as the corresponding ground-truth images with original bounding boxes and labels.

The detection results with both threshold values demonstrate the deep learning-based two-stage detection model's effectiveness in detecting automotive wire harness connectors. As shown in the second row of Fig. 7, all connectors are located with the threshold value of 0.1, but there are many detection errors in the classes of connectors and uncertain detection of the positions of connectors in the detection results. By increasing the threshold value to 0.5, the inference results present a more precise detection on different connectors, as shown in the third row of Fig. 7, but some bounding boxes are excluded from the final detection results, which left some connectors in images not detected, leading to a deteriorated recall rate. Hence, more data on connectors is desired to train a better detection model with higher precision and recall rates. Further study on selecting the appropriate threshold value is also critical to make the detection more accurate and robust for practical applications.

The last row in Fig. 7 demonstrates some detection results from the one-stage detection model based on YOLOv5 [15]. The detection results indicate the effectiveness of the deep learning-based one-stage detection model on connector detection. However, there are also some detection errors on the positions and classes of connectors presented in the detection results, where the augmentation of the dataset [48] can be helpful for better training and inference.

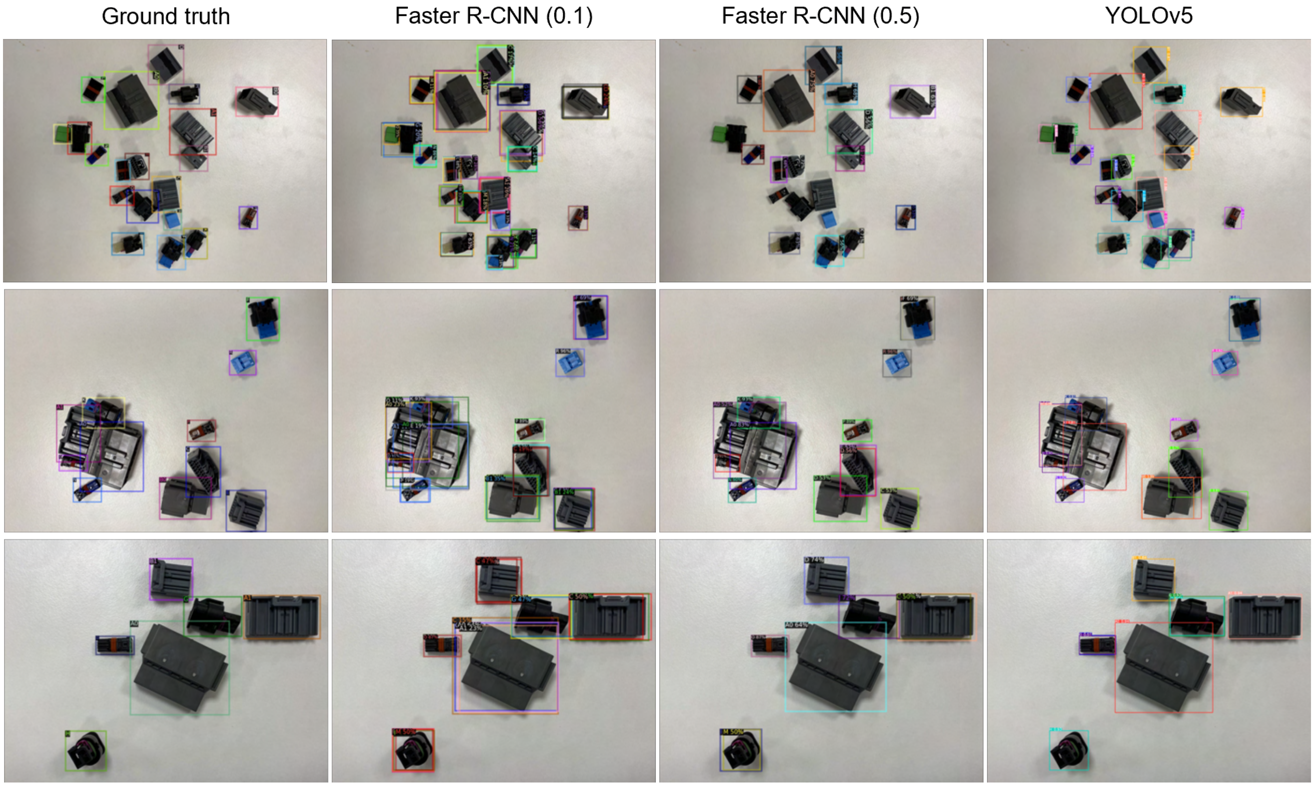


Fig. 7. Detection results of Faster R-CNN [14] with different threshold values and YOLOv5 [15] with bounding boxes around detected connectors and inferred classes on the upper left corners of bounding boxes. Images were cropped to remove the blank areas surrounding the groups of connectors. The first row demonstrates the corresponding ground truth. The second row demonstrates the Faster R-CNN [14] detection results with a threshold of 0.1. The third row demonstrates the Faster R-CNN [14] detection results with a threshold of 0.5. The last row demonstrates the YOLOv5 [15] detection results.

TABLE II
THE PRECISION (%) OF FASTER R-CNN [14] WITH THRESHOLD VALUES OF 0.1 AND 0.5 AND YOLOV5 [15] AMONG CLASSES.

Class	A0	A1	B0	B1	C
Faster R-CNN [14] (0.1)	**83.2**	70.3	0.0	57.3	32.8
Faster R-CNN [14] (0.5)	**83.2**	48.0	0.0	48.4	0.0
YOLOv5 [15]	76.4	**79.2**	**100.0**	**73.8**	**69.7**

Class	D	E	F	G	H
Faster R-CNN [14] (0.1)	**13.2**	6.7	70.1	**52.4**	40.4
Faster R-CNN [14] (0.5)	0.0	0.0	70.1	30.3	40.4
YOLOv5 [15]	0.0	**47.2**	**100.0**	30.9	**90.7**

Class	I	J	K	L	M
Faster R-CNN [14] (0.1)	45.0	48.6	**80.2**	80.0	60.0
Faster R-CNN [14] (0.5)	0.0	0.0	**80.2**	80.0	30.3
YOLOv5 [15]	**88.5**	**53.0**	63.2	**85.5**	**100.0**

Class	N	O	P	Q	R
Faster R-CNN [14] (0.1)	**93.3**	63.2	72.4	35.0	78.8
Faster R-CNN [14] (0.5)	**93.3**	50.4	72.4	0.0	67.2
YOLOv5 [15]	91.6	**96.4**	**94.6**	**93.0**	**96.8**

TABLE III
THE MEAN AVERAGE PRECISION (%) OF FASTER R-CNN [14] WITH THRESHOLD VALUES OF 0.1 AND 0.5 AND YOLOV5 [15].

	mAP_{50}	mAP_{50-95}
Faster R-CNN [14] (0.1)	65.7	54.1
Faster R-CNN [14] (0.5)	47.1	39.7
YOLOv5 [15]	**88.5**	**82.1**

Quantitatively, TABLE II and TABLE III present the rate of precision and mean Average Precision (mAP) of the Faster R-CNN-based model with threshold values of 0.1 and 0.5 and the YOLOv5-based model. In general, the YOLOv5-based model outperforms the Faster R-CNN-based model regarding mAP on the collected connector dataset under the experiment settings in this study. However, there are several rates of precision in TABLE II lower than 50%, including the ones of both detection model on class D and E, the ones of the Faster R-CNN-based model on class C, and the one of the YOLOv5-based model on the class G.

By observing the exteriors of the connectors in the collected dataset, this study found that similar designs of some connectors may affect detection performance. For example, class G and J have identical appearances but different seal ring colors inside the connectors, which are occluded when the images are captured from specific views, as shown in Fig. 8. Besides, the widths of classes A1, B1, C, D, and E are different, but their left and right profiles are highly similar, as shown in Fig. 9. These observations indicate that if some connectors share similar exterior designs and are placed with specific poses, their distinguishable features can be occluded, making it hard to recognize them. Nonetheless, similar exteriors motivate two feasible strategies to relieve this detection problem: 1) conducting further connector detection based on

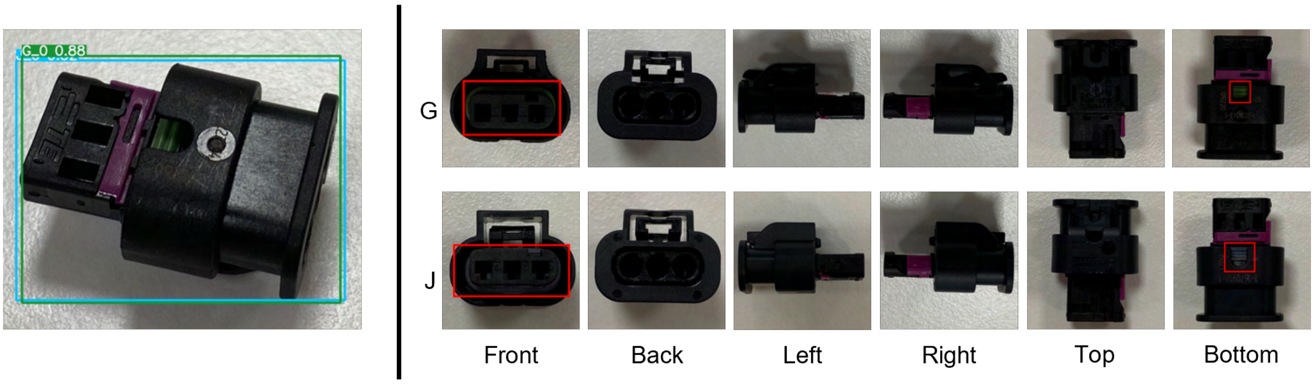


Fig. 8. Inference result by YOLOv5 on class G and J (left), whose exteriors (right) are highly similar, but the colors of seal rings inside are different (highlighted by red rectangles).

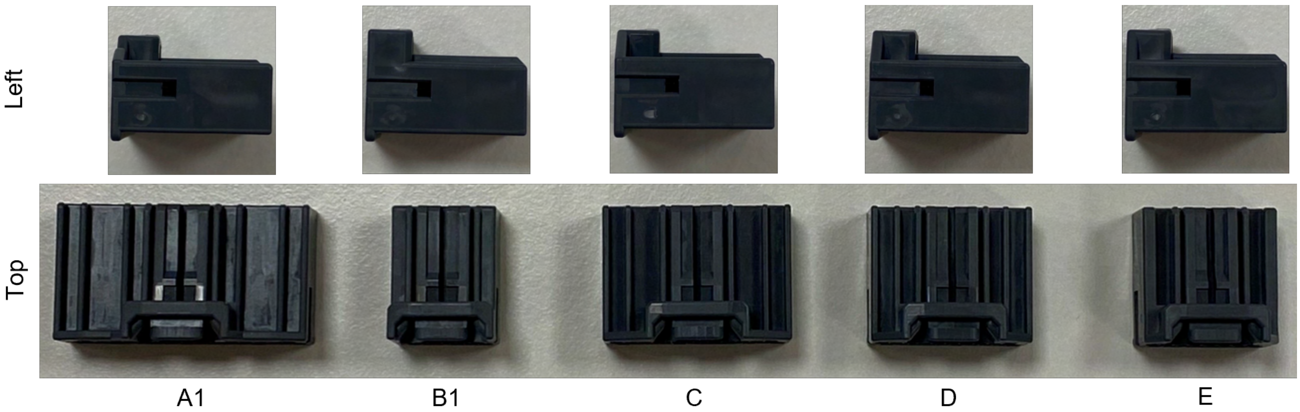


Fig. 9. Class A1, B1, C, D, and E with highly similar profiles. The first and the second row show the pairs of left-top profiles of each class.

multi-view images or videos; 2) re-design the exteriors of connectors with more distinguishable features. Specifically, for the former solution, if the inference of the class of a connector is uncertain, multi-view images of the connector or a video capturing different views of the connector can be acquired for further classification. And for the latter solution, changing the design of the exteriors of connectors, for example, changing the color of the whole connector or part of the connector, may substantially facilitate the detection, which calls for collaboration with the manufacturers of connectors.

VI. Conclusions and Future Work

This study collects a dataset with twenty connectors that are commonly used on automotive wire harnesses and trains a two-stage Faster R-CNN-based detector and a one-stage YOLOv5-based detector to validate the feasibility of deep learning-based connector detection for robotized automotive wire harness assembly. The experiment results indicate the effectiveness of both types of object detection methods and demonstrate the better performance achieved by the one-stage YOLOv5-based model, but also reflect problematic detection outcomes that require further study with other detection algorithms and more data, which will be investigated in future research. Besides, the fact that connectors are attached to cables and wires in actual scenarios inspires future research to consider taking advantage of the part of a cable on one side of a connector. In addition, observations on collected connectors motivate the problematic detection potentially affected by the similar designs of some connectors, especially the exteriors, which encourages future studies on multi-view image-based and video-based connector detection as well as on new exterior designs of connectors.

References

[1] K. mo Koo, X. Jiang, K. Kikuchi, A. Konno, and M. Uchiyama, "Development of a robot car wiring system," in *2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2008, pp. 862–867.

[2] K. Koo, X. Jiang, A. Konno, and M. Uchiyama, "Development of a wire harness assembly motion planner for redundant multiple manipulators," *Journal of Robotics and Mechatronics*, vol. 23, no. 6, pp. 907–918, 2011.

[3] X. Jiang, K.-M. Koo, K. Kikuchi, A. Konno, and M. Uchiyama, "Robotized assembly of a wire harness in a car production line," *Advanced Robotics*, vol. 25, no. 3-4, pp. 473–489, 2011.

[4] X. Jiang, Y. Nagaoka, K. Ishii, S. Abiko, T. Tsujita, and M. Uchiyama, "Robotized recognition of a wire harness utilizing tracing operation," *Robotics and Computer-Integrated Manufacturing*, vol. 34, pp. 52–61, 2015.

[5] B. Sun, F. Chen, H. Sasaki, and T. Fukuda, "Robotic wiring harness assembly system for fault-tolerant electric connectors mating," in *2010 International Symposium on Micro-NanoMechatronics and Human Science*, 2010, pp. 202–205.

[6] L. Zhou, L. Zhang, and N. Konz, "Computer vision techniques in manufacturing," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 1, pp. 105–117, 2023.

[7] P. Di, J. Huang, F. Chen, H. Sasaki, and T. Fukuda, "Hybrid vision-force guided fault tolerant robotic assembly for electric connectors," in *2009 International Symposium on Micro-NanoMechatronics and Human Science*, 2009, pp. 86–91.

[8] P. Di, F. Chen, H. Sasaki, J. Huang, T. Fukuda, and T. Matsuno, "Vision-force guided monitoring for mating connectors in wiring harness assembly systems," *Journal of Robotics and Mechatronics*, vol. 24, no. 4, pp. 666–676, 2012.

[9] T. Tamada, Y. Yamakawa, T. Senoo, and M. Ishikawa, "High-speed manipulation of cable connector using a high-speed robot hand," in *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2013, pp. 1598–1604.

[10] H.-C. Song, Y.-L. Kim, D.-H. Lee, and J.-B. Song, "Electric connector assembly based on vision and impedance control using cable connector-feeding system," *Journal of Mechanical Science and Technology*, vol. 31, no. 12, pp. 5997–6003, 2017.

[11] F. Yumbla, M. Abeyabas, T. Luong, J.-S. Yi, and H. Moon, "Preliminary connector recognition system based on image processing for wire harness assembly tasks," in *2020 20th International Conference on Control, Automation and Systems (ICCAS)*, 2020, pp. 1146–1150.

[12] H. Zhou, S. Li, Q. Lu, and J. Qian, "A practical solution to deformable linear object manipulation: A case study on cable harness connection," in *2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM)*, 2020, pp. 329–333.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015.

[15] "Yolov5," https://github.com/ultralytics/yolov5, accessed: 2023-02-07.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012.

[17] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023.

[18] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[19] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[20] ——, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, 2016.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[22] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.

[23] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 936–944.

[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[25] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018.

[26] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020.

[27] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6517–6525.

[28] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 7464–7475.

[29] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 21–37.

[30] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2999–3007.

[31] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, pp. 765–781.

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017.

[33] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Comput. Surv.*, vol. 54, no. 10s, 2022.

[34] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham: Springer International Publishing, 2020, pp. 213–229.

[35] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable detr: Deformable transformers for end-to-end object detection," in *International Conference on Learning Representations*, 2021.

[36] H. Zhang, F. Li, S. Liu, L. Zhang, H. Su, J. Zhu, L. M. Ni, and H.-Y. Shum, "Dino: Detr with improved denoising anchor boxes for end-to-end object detection," 2022.

[37] F. Li, H. Zhang, H. Xu, S. Liu, L. Zhang, L. M. Ni, and H.-Y. Shum, "Mask dino: Towards a unified transformer-based framework for object detection and segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 3041–3050.

[38] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, pp. 303–308, 2009.

[39] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 740–755.

[40] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, pp. 211–252, 2015.

[41] H. G. Nguyen, R. Habiboglu, and J. Franke, "Enabling deep learning using synthetic data: A case study for the automotive wiring harness manufacturing," *Procedia CIRP*, vol. 107, pp. 1263–1268, 2022, leading manufacturing systems transformation – Proceedings of the 55th CIRP Conference on Manufacturing Systems 2022.

[42] "Labelme," https://github.com/wkentaro/labelme, accessed: 2023-02-16.

[43] A. Halevy, P. Norvig, and F. Pereira, "The unreasonable effectiveness of data," *IEEE Intelligent Systems*, vol. 24, no. 2, pp. 8–12, 2009.

[44] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 843–852.

[45] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.

[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[47] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.

[48] M. Everingham, S. A. Eslami, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International journal of computer vision*, vol. 111, pp. 98–136, 2015.