Thesis for The Degree of Doctor of Philosophy

Computer vision for non-rigid object assembly automation

With applications in automotive wire harness assembly

Hao Wang

Department of Mechanical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2026

Computer vision for non-rigid object assembly automation
With applications in automotive wire harness assembly

Hao Wang

ISBN 978-91-8103-423-3

Acknowledgments, dedications, and similar personal statements in this thesis, reflect the author's own views.

During the preparation of this work, the author used Grammarly and OpenAI's GPT-5 in order to improve the readability and language of the manuscript. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published thesis.

This thesis has been prepared using LaTeX.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5880
ISSN 0346-718X
https://doi.org/10.63959/chalmers.dt/5880

Department of Mechanical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone: +46(0)31 772 1000

Printed by Chalmers Digital Printing,
Gothenburg, Sweden 2026

To my grandparents,
in loving memory.

Computer vision for non-rigid object assembly automation
With applications in automotive wire harness assembly

HAO WANG
Department of Mechanical Engineering
Chalmers University of Technology

Abstract

This thesis examines the role of computer vision in facilitating robotic automation for non-rigid object assembly, with a particular focus on wire harness assembly tasks during the automotive final assembly stage. Despite extensive research in this field, industrial adoption has remained limited. Accordingly, the thesis analyzes the challenges associated with applying computer vision to wire harness assembly automation and explores strategies to enable practical deployment in industrial environments.

Employing both qualitative and quantitative methods, the research progresses from problem identification to artifact design, demonstration, and evaluation in laboratory and industrially relevant environments. The studies identify challenges at the object, scene, data, and operational levels. To address these challenges, three primary artifacts were developed and evaluated. First, a learning-based perception pipeline enables markerless detection of wire harness components. This demonstrates the feasibility of deep learning-based component recognition and highlights limitations when components possess highly similar or occluded visual features. Second, a robot-assisted pipeline for automated multi-view data acquisition and multimodal annotation substantially accelerates the preparation of computer vision datasets. This pipeline also supports the training and evaluation of learning-based perception methods for industrial applications. Third, a vision-based human–robot collaboration framework for wire harness installation significantly reduces localized physical discomfort while maintaining task success. However, this approach increases mental demand and cycle time, with the majority of the additional time attributable to robot execution.

In summary, this thesis provides deployable methods and practical guidance for data-centric development, interaction design, and takt-time-oriented workflow optimization in non-rigid object assembly automation. It also demonstrates that, given current technological constraints, computer vision is most effective as a human-centered enabler of robot-assisted assembly rather than as a direct pathway to fully autonomous robotic assembly of non-rigid objects.

Keywords

Artificial intelligence, computer vision, robotic perception, flexible automation, human–robot collaboration, assembly, automotive industry, non-rigid object

List of publications

Appended publications

This thesis is based on the work contained in the following publications:

[**Paper 1**] B. Johansson, M. Despeisse, J. Bokrantz, G. Braun, H. Cao, A. Chari, Q. Fang, C. A. González Chávez, A. Skoogh, H. Söderlund, **H. Wang**, K. Wärmefjord, L. Nyborg, J. Sun, R. Örtengren, K. A. Schumacher, L. Espinal, K. C. Morris, J. Nunley Jr., Y. Kishita, Y. Umeda, F. Acerbi, M. Pinzone, H. Persson, S. Charpentier, K. Edström, D. Brandell, M. Gopalakrishnan, H. Rahnama, L. Abrahamsson, A. Ö. Rönnbäck and J. Stahre, "Challenges and opportunities to advance manufacturing research for sustainable battery life cycles," *Frontiers in Manufacturing Technology*, vol. 4, 2024. DOI: 10.3389/fmtec.2024.1360076.

H. Wang primarily contributed to the section on human-centric production, particularly automation and human–robot collaboration aspects.

[**Paper 2**] O. Salunkhe, W. Quadrini, **H. Wang**, J. Stahre, D. Romero, L. Fumagalli and D. Lämkull, "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.

H. Wang primarily contributed to the sections on computer vision applications in automated automotive wire harness assembly and, together with O. Salunkhe and W. Quadrini, was responsible for research design, data collection, and analysis.

[**Paper 3**] **H. Wang**, O. Salunkhe, W. Quadrini, D. Lämkull, F. Ore, M. Despeisse, L. Fumagalli, J. Stahre and B. Johansson, "A systematic literature review of computer vision applications in robotized wire harness assembly," *Advanced Engineering Informatics*, vol. 62, p. 102596, 2024. DOI: 10.1016/j.aei.2024.102596.

H. Wang initiated the work and prepared the initial draft of the manuscript. All authors contributed to and reviewed the text. H. Wang, O. Salunkhe, and W. Quadrini collaboratively designed the research, collected data, and performed the analysis.

[**Paper 4**] **H. Wang** and 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.

H. Wang initiated the work, collected and analyzed the dataset, designed and conducted the experiments, and prepared the initial draft of the manuscript. B. Johansson contributed to and reviewed the manuscript.

[**Paper 5**] **H. Wang**, G. Urbanos Uriel, K. El-Nahass, S. Ekered and B. Johansson, "Accelerating industrial vision: Systematic robot-assisted dataset preparation for object detection and pose estimation," *Engineering Applications of Artificial Intelligence*, vol. 176, p. 114741, 2026. DOI: 10.1016/j.engappai.2026.114741.

H. Wang initiated the work, co-designed the robot and vision system with G. Urbanos Uriel, K. El-Nahass, and S. Ekered, designed and conducted the experiments, and prepared the initial draft of the manuscript. All authors contributed to and reviewed the text.

[**Paper 6**] **H. Wang**, O. Salunkhe, A. Hartmann, S. Ekered, P. Bründl, J. Franke, J. Stahre and B. Johansson, "Vision-based human–robot collaboration for wire harness assembly in automotive manufacturing," *Submitted to a scientific journal (under revision)*, 2026.

H. Wang initiated and prepared the initial draft of the manuscript, with contributions and reviews provided by the other authors. The hardware setup was configured by H. Wang and S. Ekered. Research design and data collection were performed by H. Wang and O. Salunkhe. Data analysis and visualization were conducted by H. Wang, O. Salunkhe, and A. Hartmann.

Other publications

The following publications were published during the author's PhD study or are currently under submission or revision. These publications are not included in this thesis because their content overlaps with the appended publications or is not directly relevant to the thesis.

[**a**] X. Zhu, **H. Wang**, H. Fei, Z. Lei and S. Z. Li, "Face forgery detection by 3d decomposition," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 2928–2938. DOI: 10.1109/CVPR46437.2021.00295.

[**b**] X. Zhu, C. Yu, D. Huang, Z. Lei, **H. Wang** and S. Z. Li, "Beyond 3dmm: Learning to capture high-fidelity 3d face shape," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1442–1457, 2023. DOI: 10.1109/TPAMI.2022.3164131.

[**c**] **H. Wang**, O. Salunkhe, W. Quadrini, D. Lämkull, F. Ore, B. Johansson and J. Stahre, "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.

[**d**] M. Despeisse, B. Johansson, J. Bokrantz, G. Braun, A. Chari, X. Chen, Q. Fang, C. A. González Chávez, A. Skoogh, J. Stahre, N. Theradapuzha Mathew, E. Turanoglu Bekar, **H. Wang** and R. Örtengren, "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*, 2023, pp. 521–535. DOI: 10.1007/978-3-031-43688-8_36.

[**e**] O. Salunkhe, C. A. González Chávez, **H. Wang**, A. Syberfeldt, D. Romero and J. Stahre, "Developing code agents for robot programming: Technical and managerial perspectives," in *Advances in Production Management Systems. Cyber-Physical-Human Production Systems: Human-AI Collaboration and Beyond*, 2026, pp. 134–147. DOI: 10.1007/978-3-032-03515-8_10.

This thesis builds upon the previously published thesis for the degree of Licentiate of Engineering by the author [1].

Acronyms

2D	Two-Dimensional
3D	Three-Dimensional
6DoF	Six-Degree-of-Freedom
AI	Artificial Intelligence
BDLO	Branched Deformable Linear Object
CAD	Computer-Aided Design
CIRP	Collège International pour la Recherche en Productique (The International Academy for Production Engineering)
CNN	Convolutional Neural Network
DARE	Database of Abstracts of Reviews of Effects
DETR	Detection Transformer
DINO	DETR with Improved Denoising Anchor Boxes
DL	Deep Learning
DLO	Deformable Linear Object
DLON	Deformable Linear Object Network
DMLO	Deformable Multi-Linear Object
DOM	Deformable Object Manipulation
DOO	Deformable One-dimensional Object
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
ISA	International Society of Automation
ISO	International Organization for Standardization
KPI	Key Performance Indicator
LiDAR	Light Detection And Ranging
mAP	mean Average Precision
MSD	Musculoskeletal Disorder
NASA	National Aeronautics and Space Administration
NASA-TLX	NASA Task Load Index
OSH	Occupational Safety and Health
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
SLR	Systematic Literature Review
SPPNet	Spatial Pyramid Pooling Network
SSD	Single-Shot Detector
TRL	Technology Readiness Level
VOC	Visual Object Classes
YOLO	You Only Look Once

Acknowledgment

This thesis would not have been possible without the support, guidance, and encouragement of many individuals and organizations.

I am deeply grateful to my examiner, Johan Stahre, my main supervisor, Björn Johansson, and my assistant supervisor, Anders Skoogh, for their invaluable mentorship and support throughout my doctoral studies. Their high standards, constant encouragement, and insightful feedback have profoundly shaped my development as a researcher and sustained my motivation.

I would also like to thank my colleagues and friends in the Production Systems division, as well as everyone else at the Department of Mechanical Engineering, which incorporating the former Department of Industrial and Materials Science. I am especially grateful to those in the Stena Industry Innovation Lab (SII-Lab) and on the fifth floor of M-huset, with whom I have shared many joyful moments. A special mention, in alphabetical order, goes to Adam, Anna, Arpita, Clarissa, Dan, Ebru, Elisa, Ellinor, Greta, Henrik, Huizhong, Mohan, Mélanie, Ninan, Omkar, Per, Qi, Sandra, Silvan, Sven, Tina, and Xiaoxia. The stimulating discussions, daily camaraderie, and inspiring work environment they help created have greatly fostered both innovation and fruitful collaboration.

I am also grateful to my academic and industry collaborators, especially the co-authors and reviewers of the publications included in this thesis, whose expertise has greatly enriched this work. A special mention, in alphabetical order, goes to Annalena, Dan, David, Fredrik, Gonzalo, Karim, Patrick, Patrick, and Walter.

Furthermore, I acknowledge our project partners, whose practical insights and engagement have grounded this research in real-world impact. I sincerely thank Chalmers Production Area of Advance, Vinnova, Produktion2030, Västra Götalandsregionen, and the European Commission for their financial and institutional support.

Finally, I express my deepest gratitude to my family for their unwavering support. To my parents, thank you for instilling in me the courage and positivity to pursue my dreams. To my wife, Shan, thank you for your steadfast love, patience, and reassurance. You have brightened every step of this journey and my life beyond it.

Hao Wang
Gothenburg, Sweden
May 2026

Contents

I Summary **1**

1 Introduction **3**
1.1 Background . . . 3
1.2 Problem description . . . 4
1.3 Vision, aim, and research questions . . . 5
1.4 Scope and delimitation . . . 5
1.5 Thesis outline . . . 6

2 Frame of reference **7**
2.1 Assembly automation . . . 7
2.1.1 From manual to automated assembly . . . 7
2.1.2 Robotics in industrial automation . . . 8
2.1.3 Challenges in automating complex assembly tasks . . . 9
2.2 Computer vision in assembly automation . . . 9
2.2.1 Overview of computer vision techniques . . . 10
2.2.2 Computer vision techniques in manufacturing . . . 12
2.2.3 Vision systems in assembly . . . 13
2.3 Wire harness assembly . . . 13
2.3.1 Current manual assembly process . . . 14
2.3.2 Why automation is needed . . . 14
2.3.3 An instance of deformable object manipulation . . . 15
2.3.4 Automation challenges in wire harness assembly . . . 16

3 Research approach **19**
3.1 Philosophical worldview . . . 19
3.2 Research design . . . 20
3.3 Research methods . . . 22
3.4 Research quality assurance . . . 23

4 Research contributions **25**
4.1 Contributions of appended publications . . . 25
4.2 Summary of appended publications . . . 27
4.2.1 Paper 1 . . . 27
4.2.2 Paper 2 . . . 28
4.2.3 Paper 3 . . . 29
4.2.4 Paper 4 . . . 31
4.2.5 Paper 5 . . . 32
4.2.6 Paper 6 . . . 34

5 Discussion **37**
5.1 Answers to research questions . . . 37
5.1.1 The answer to RQ1 . . . 37
5.1.2 The answer to RQ2 . . . 38
5.2 Positioning of this thesis . . . 39
5.3 Contributions of this thesis . . . 41
5.4 Limitations of this thesis . . . 42
5.5 Reflections on research quality . . . 43
5.6 Reflections on research ethics . . . 45
5.7 Reflections on the aspect of sustainability . . . 46
5.8 Future work . . . 48

6 Conclusion **51**

Bibliography **53**

II Appended publications **75**

Paper 1 - Challenges and opportunities to advance manufacturing research for sustainable battery life cycles

Paper 2 - Review of current status and future directions for collaborative and semi-automated automotive wire harnesses assembly

Paper 3 - A systematic literature review of computer vision applications in robotized wire harness assembly

Paper 4 - Deep learning-based connector detection for robotized assembly of automotive wire harnesses

Paper 5 - Accelerating industrial vision: Systematic robot-assisted dataset preparation for object detection and pose estimation

Paper 6 - Vision-based human–robot collaboration for wire harness assembly in automotive manufacturing

Part I

Summary

Chapter 1

Introduction

This chapter begins by presenting the background and central problem addressed by the research. It then outlines the vision and aim of the study, followed by the formulation of the research questions. The scope and delimitation are subsequently defined. The chapter concludes with an overview of the thesis structure.

1.1 Background

The widespread adoption of automation has transformed various sectors of modern industry and delivered substantial gains in quality and productivity [2]. However, expanding automation to address persistent production challenges remains a primary objective, particularly in response to the continuous demand for improved quality, efficiency, and sustainability [3]. A significant challenge exists in assembly processes that require extensive manual operations, which contribute to production inefficiencies and raise ergonomic concerns [4, 5]. Among these, the assembly of non-rigid objects stands out as particularly difficult to automate [6, 7].

Specialized machinery has been designed to automate specific operations involving non-rigid objects, such as agricultural harvesters and textile machinery. While these solutions enhance efficiency, their dependence on highly specialized equipment designed for specific tasks limits adaptability to the growing variability of products and processes in manufacturing [8]. In contrast, industrial robots, as defined by the International Organization for Standardization (ISO) as "automatically controlled, reprogrammable multipurpose manipulator, programmable in three or more axes, which can be either fixed in place or fixed to a mobile platform for use in automation applications in an industrial environment" [9], represent a promising general-purpose alternative for non-rigid object assembly automation.

Effective automation of these tasks depends on deformable object manipulation (DOM), which requires industrial robots to have sufficient autonomy to function in unstructured and dynamic environments [10]. However, the deformable properties of non-rigid objects present significant challenges for robotic perception, modeling, learning, and control [11]. Among these challenges, perception is fundamental because it allows robots to interpret their environment, model object states, and plan appropriate actions [12].

Precise perception of object position and orientation is critical for effective manipulation, particularly when handling product variants or unpredictable object configurations [13]. In human–robot collaboration (HRC) scenarios, perception is also crucial for detecting human presence and interpreting intent, thereby ensuring safe and effective interaction [14].

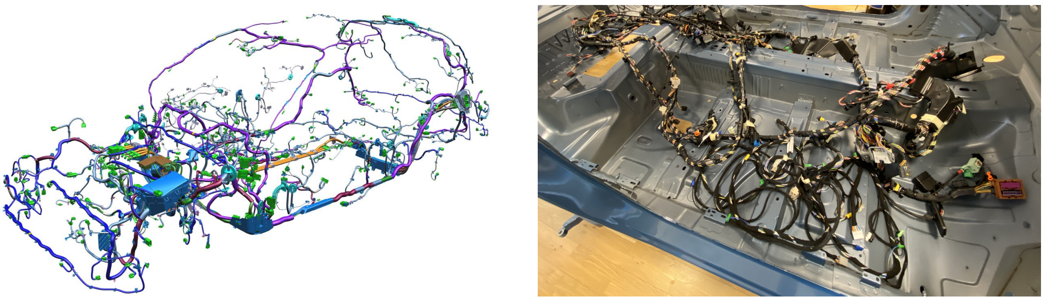
Among sensory modalities, vision provides comprehensive information about the external environment [15]. This capability renders vision indispensable for critical perception tasks in robotic manipulation such as object recognition, pose estimation, and tracking [16]. Enabling robotic visual perception is therefore essential for automating the assembly of non-rigid objects and addressing associated production challenges [7].

Robotic visual perception is closely associated with advancements in artificial intelligence (AI) and computer vision [17]. Research in these fields has produced promising techniques that may enhance the perceptual capabilities of industrial robots and enable autonomous execution of non-rigid object assembly tasks [18].

1.2 Problem description

Previous studies indicate that universal solutions for DOM are impractical because of the heterogeneity among classes of deformable objects [7, 18]. Therefore, this thesis addresses a specific object class and a well-defined industrial task to generate technically robust insights that are broadly applicable to manufacturing contexts.

Wire harnesses represent a sub-category of deformable linear objects (DLOs). They are widely utilized in products with electrical systems, such as automobiles, aircraft, and consumer electronics [19]. Figure 1.1 demonstrates both their prevalence and structural complexity in modern passenger vehicles.



(a) (b)

Figure 1.1: (a) Electrical infrastructure of a passenger vehicle. Source: Volvo Car Corporation. (b) Example of wire harnesses to be installed on the chassis of a passenger vehicle.

Wire harness assembly exemplifies DOM and is regarded as one of the most complex tasks in industrial automation, with direct implications for productivity, quality, safety, and ergonomics [20]. The flexible geometry and diverse configurations of wire harnesses present significant challenges for robotic perception, modeling, learning, and control [11]. Although the importance of visual machine perception in advancing robotic autonomy is widely recognized in both academic and industrial contexts, vision-driven robotic assembly has not yet achieved widespread adoption [21]. Despite extensive research proposing automation solutions for wire harness assembly, practical implementations are still lacking in production environments [22]. Furthermore, the potential of computer vision in this context remains largely unexplored and unrealized [7].

In this context, this thesis investigates wire harness assembly automation as a representative and industrially significant example of non-rigid object assembly [7]. Specifically, it explores the application of computer vision to enable robotic systems to perform wire harness assembly tasks during the final assembly stage.

1.3 Vision, aim, and research questions

This thesis envisions a future manufacturing industry characterized by high efficiency and sustainability, with minimal operational disruptions. Achieving this vision necessitates significant technological and organizational advancements, with symbiotic collaboration among humans, robots, and automated machinery as a critical component. In this context, robots function as intelligent agents that autonomously execute tasks that are non-value-adding or ergonomically demanding for human workers. Furthermore, they adapt flexibly to a variety of tasks and dynamic production scenarios with minimal human intervention.

Attaining this level of autonomy requires robots to possess advanced intelligence and contextual awareness. Robots must perceive and interpret their tasks, understand the surrounding environment, and account for the presence and actions of humans and other machines to respond appropriately. Advanced visual perception is essential for these capabilities, enabling robots to interpret complex scenes and interact effectively with objects exhibiting diverse properties. Within this broader vision, the thesis addresses a critical perception challenge in industrial automation: enabling robots to perform non-rigid object assembly using computer vision.

Given this context, the research aims to propose technical solution that leverage computer vision to facilitate robotic automation for non-rigid object assembly tasks. Wire harness assembly is selected as the primary application domain due to its industrial relevance and inherent complexity. Nevertheless, the proposed approaches are designed to be transferable to other non-rigid object assembly contexts.

To achieve this aim, the research first examines the problem space by identifying industrial constraints that complicate computer vision-based automation and by clarifying the capabilities required of vision systems to support wire harness assembly automation. Establishing this understanding is essential for developing effective solutions and motivates the first research question (RQ):

RQ1: What are the challenges in applying computer vision to wire harness assembly automation?

Building on the identified challenges, the research investigates and develops methods to address them and examines how these methods can be integrated into robotic systems and assembly workflows. This leads to the second RQ:

RQ2: How can computer vision be applied to wire harness assembly automation?

1.4 Scope and delimitation

This thesis examines the integration of existing computer vision techniques into robotic systems to enable automation of non-rigid object assembly, with particular emphasis on wire harness assembly during the automotive final assembly stage. The primary objective is to address perception challenges in wire harness assembly by developing practical methods that allow robots to accurately recognize and interpret relevant assembly components. Guided by the two defined RQs, the proposed solutions are developed and evaluated in controlled laboratory environments, with consideration of their potential applicability in industrial settings. The approaches and insights are intended to be transferable to additional assembly tasks that involve non-rigid objects.

To maintain a focused scope, specific delimitations are established. This thesis does not introduce novel computer vision algorithms, nor does it assess the extent to which improved perception enhances overall robotic autonomy. Performance and practicality are not evaluated in actual production environments using system-level production key performance indicators (KPIs). Given the primary focus on visual perception, this work does not introduce new robotic planning, control, or grasping strategies, nor does it investigate multi-sensor perception, such as force, tactile, or acoustic sensing, for estimating deformable states.

This thesis specifically focuses on visual perception of key rigid wire harness components, including connectors and clamps, and its respective roles in assembly subtasks. It does not address end-to-end automation of deformable cable routing, topology estimation, or manipulation control. Consequently, generalizability is pursued at the level of methods and workflows, rather than through direct transfer of trained models across all wire harness variants, production lines, or industrial domains.

1.5 Thesis outline

This thesis comprises six chapters, each contributing to the development and presentation of the research.

Chapter 1 Introduction introduces the research background, motivation, vision, and aim. This chapter formulates the RQs and defines the scope and delimitation. It concludes with an overview of the thesis structure.

Chapter 2 Frame of reference reviews relevant literature on assembly automation, the application of computer vision in this context, and previous research on wire harness assembly automation.

Chapter 3 Research approach describes the philosophical worldview underpinning the research, the adopted research design, and the methods for data collection and analysis. It also explains the strategies implemented to ensure research quality.

Chapter 4 Research contributions presents the contributions of the appended publications in relation to the RQs. Each publication is briefly summarized, highlighting its core research problem, methodology, and contribution.

Chapter 5 Discussion synthesizes the main research findings from the research and addresses the RQs formulated in Chapter 1. It discusses the positioning and contributions of the thesis, reflects on its limitations and research quality, and considers ethical and sustainability aspects. The chapter concludes with suggestions for future research and the potential for industrial implementation.

Chapter 6 Conclusion concludes the thesis with a summary of its key findings and contributions.

Chapter 2

Frame of reference

This chapter establishes the frame of reference for the thesis by introducing core background concepts relevant to the research context, including automation, assembly, and robotic assembly. It then outlines computer vision techniques, emphasizing their roles and challenges in robotic assembly. Finally, the chapter reviews previous research on wire harness assembly automation.

2.1 Assembly automation

2.1.1 From manual to automated assembly

Automation is integral to contemporary manufacturing and has been conceptualized through diverse frameworks within both research and industrial contexts. The International Academy for Production Engineering (CIRP) defines automation as "the conversion of a procedure, a process, or equipment to automatic operation, i.e., without intervention by a human operator" [23]. The International Society of Automation (ISA) instead describes it as "the creation and application of technology to monitor and control the production and delivery of products and services" [24]. Collectively, these definitions underscore the potential of automation to minimize dependence on human intervention within manufacturing operations.

Research on production automation typically distinguishes between physical automation and cognitive automation [25]. Physical automation involves the mechanization of physical tasks [26], while cognitive automation addresses the cognitive activities presently performed by human workers [27]. This thesis examines the enablers of physical automation and investigates technical solutions that may transform manual assembly operations in contemporary production environments.

Assembly constitutes a fundamental aspect of production and can be analyzed from two perspectives: as a process and as a product [28]. The present study adopts the process perspective, with a specific focus on automotive final assembly. Within this context, sub-components such as engines, body frames, wire harnesses, windows, and wheels are integrated to create the final product.

From the process perspective, assembly may be conducted manually [13], semi-automatically [29], or fully automatically [30]. The selection of assembly mode is influenced by factors including product design, required production rate, labor availability, and product market life [31]. Lien [13] proposed a decomposition of assembly work into four recurring subtasks: (1) identifying and gripping parts,

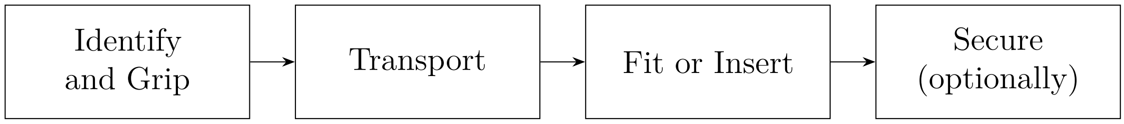


Figure 2.1: Sub-procedures of an assembly task, as adapted from Lien [13].

(2) transporting parts to the target location, (3) fitting or inserting parts, and (4) optionally securing parts, as illustrated in Figure 2.1.

In assembly research, assembly automation denotes the use of automated machinery to transform manual assembly processes into operations that eliminate the need for human intervention [30]. Implementing assembly automation can result in benefits such as enhanced quality, more efficient resource utilization, and increased productivity [32, 33]. Assembly automation may also improve occupational safety and ergonomics by reducing exposure to physically demanding or hazardous tasks [34, 35].

2.1.2 Robotics in industrial automation

Ongoing research initiatives and the integration of automation have led to substantial progress in the field of robotics [36]. Over the past several decades, robotics has evolved into an interdisciplinary field centered on the science and technology of robots and automated systems, encompassing the design, construction, operation, and application of robotic systems [37].

Since the inception of the field, industrial robots have constituted a primary focus of robotics research [38]. A robot is commonly defined as a physical agent that interacts with its environment via effectors and is directed by information obtained from its sensors [39]. Within manufacturing contexts, the terms "industrial robot" and "robot" are frequently used interchangeably [40]. This convention is maintained throughout this thesis unless otherwise specified.

The Unimate, developed by Unimation, a company established by George Devol and Joseph Engelberger in the late 1950s, is widely recognized as the first industrial robot [41]. In 1961, the Unimate was deployed to unload finished castings at a General Motors facility in Trenton, New Jersey, United States [41]. Following this initial deployment, industrial robots have played a central role in factory automation and have contributed to substantial transformations in manufacturing [42]. Compared to other automation technologies, robots are widely adopted due to their ability to perform repetitive and ergonomically demanding tasks with speed, precision, and reprogrammability [43, 44]. Common applications of robotic automation include spot welding, spray painting, part handling, packaging, and palletizing [40].

Industrial robots are also being increasingly utilized in assembly processes. Compared to fully manual work, robotic assembly can improve product quality and productivity while enabling safer, more ergonomic processes with increased precision, repeatability, transparency, and comprehensibility [44]. Improvements in degrees of freedom and payload capacity have further enabled industrial robots to perform tasks that may be hazardous for human operators, such as work in hazardous or unsanitary environments [45], repetitive operations [16], labor-intensive and monotonous tasks [46], and activities that present ergonomic challenges [44].

Despite these advantages, robotic assembly still constitutes a relatively small proportion of industrial robot applications [10, 47].

2.1.3 Challenges in automating complex assembly tasks

Although automation offers significant potential to enhance production efficiency, its implementation in assembly processes remains limited, especially in final assembly [48]. One major barrier is the inherent complexity of assembly operations, which complicates the design and deployment of reliable automation [49]. Progress is further constrained by the limited capacity of current automated systems to manage complex tasks and their insufficient flexibility to accommodate diverse product and production variants [50].

Traditional industrial robots demonstrate optimal performance in repetitive and highly structured environments. These robots typically rely on predefined programming, which makes them well suited for tasks where object positions and orientations can be explicitly specified by programmers [51]. Such assumptions become problematic when assembly operations require greater flexibility, involve increased product variety, or occur in unstructured or dynamic environments [16]. Furthermore, the deployment of traditional robots in assembly is hindered by the adaptability required for complex operations [47, 52] and by stringent safety requirements when humans and robots share workspaces [53].

With the shift toward mass customization in manufacturing, assembly systems are required to accommodate a broader range of product configurations in smaller batch sizes [54]. This challenge is further amplified in operations involving non-rigid objects, as their positions and orientations cannot be predetermined due to deformation. This results in an effectively unbounded number of configurations that would otherwise require explicit programming [7]. Consequently, increasing production demands necessitate industrial robots with enhanced adaptability and autonomy [55].

In response, robotics research has increasingly shifted focus from primarily mechanical capabilities, such as kinematic calibration, motion planning, and control laws, toward enhancing robot intelligence to improve flexibility and autonomy [38]. As noted by Siciliano and Khatib [42], “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.” Achieving this vision requires robots not only to execute tasks effectively but also to perceive their environment, learn from experience, and reason about their actions [16].

2.2 Computer vision in assembly automation

Computer vision is closely connected to automation, robotics, and manufacturing [56]. Computer vision research has two main objectives: developing computational models that emulate human vision from a biological perspective, and designing autonomous systems that perform tasks traditionally managed by human vision, potentially surpassing human capabilities, from an engineering perspective [57].

The study presented in this thesis focuses on the engineering perspective by analyzing challenges and proposing solutions for implementing computer vision techniques to improve robotic visual perception. A robot equipped with a vision system can convert visual data from its environment into an internal symbolic representation [58]. This representation allows the robot to interpret its surroundings and select appropriate subsequent actions [21].

2.2.1 Overview of computer vision techniques

Computer vision is a rapidly evolving field encompassing a broad range of topics [56, 59]. Within this field, two-dimensional (2D) object detection, three-dimensional (3D) object detection, and six-degree-of-freedom (6DoF) object pose estimation are particularly significant for robotics and robotic assembly [60]. Figure 2.2 illustrates the expected results of these three tasks when provided with a red-green-blue (RGB) visual input.

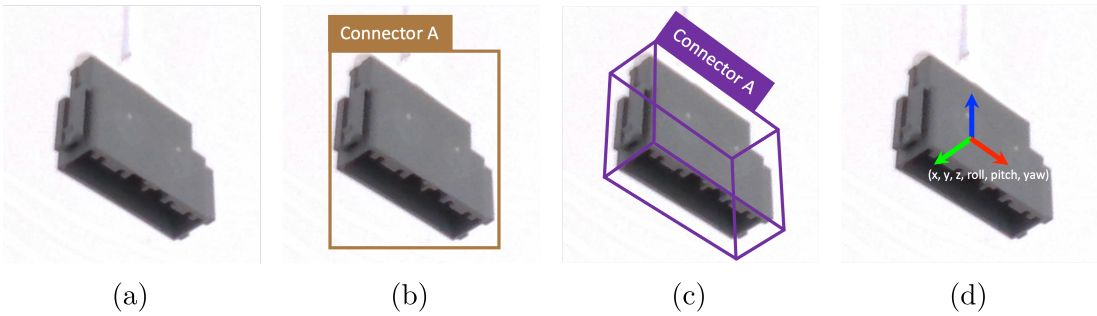


Figure 2.2: (a) An RGB visual input. (b) 2D object detection result. (c) 3D object detection result. (d) 6DoF object pose estimation result.

2D object detection

In 2D space, object recognition generally comprises three primary tasks: image classification, object localization, and object detection. Image classification assigns a class label to an image by identifying and categorizing its principal object [61]. Object localization identifies the spatial position of one or more instances of a target category, typically by predicting rectangular bounding boxes aligned with the image axes [62]. Object detection integrates these two tasks by simultaneously predicting object categories and their locations [63]. Figure 2.2(b) demonstrates how detection yields both object identities and their spatial extents within an image [64]. Research on 2D object detection is typically categorized into traditional approaches and deep learning (DL)-based approaches [65].

Traditional methods depend on handcrafted features and feature descriptors [64]. These methods generally involve three stages: (1) generating informative region proposals, (2) extracting features, and (3) performing classification and bounding-box regression [66]. Due to limitations in learning effective feature representations, traditional methods often require complex feature engineering and precisely tuned optimization strategies [65]. Representative examples include the Viola–Jones detector [67, 68], the Histogram of Oriented Gradients (HOG) detector [69], and the Deformable Part Model [70]. Traditional approaches often exhibit limited accuracy and generalizability, slow processing speeds, and require substantial manual feature engineering [63], which has motivated the transition to deep learning-based methods.

The introduction of the Region-based Convolutional Neural Network (R-CNN) [71] in 2014 marked the beginning of the deep learning (DL) era for 2D object detection [65]. DL-based detectors utilize deep convolutional neural networks (CNNs) and currently dominate both research and practical applications [63]. These detectors are generally classified as either two-stage or one-stage methods [65].

Two-stage detectors initially generate candidate regions and subsequently refine predictions within those regions, a process loosely analogous to attentional mechanisms in the human brain that focus processing on salient parts of a scene [64]. Notable examples include R-CNN [71, 72], Fast R-CNN [73], Faster R-CNN [74], SPPNet [75], and Feature Pyramid Networks (FPN) [76].

One-stage detectors were designed to enhance speed and computational efficiency by performing localization and classification in a single forward pass through the backbone network [65]. The YOLO family [77–81] represents a prominent example, with comprehensive reviews available in Diwan *et al.* [82] and Terven *et al.* [83]. Other widely adopted one-stage detectors include SSD [84], RetinaNet [85], CornerNet [86], and CenterNet [87].

Recent advancements in Transformer models [88] have facilitated the development of Transformer-based detectors [89], including DETR [90], Deformable DETR [91], DINO [92], and Mask DINO [93].

3D object detection

3D object detection presents greater challenges than 2D object detection. Figure 2.2(c) shows a representative output of 3D detection. Unlike 2D detection, which predicts bounding boxes within the image plane, 3D object detection seeks to estimate an amodal bounding box for each object instance in 3D space. An amodal bounding box refers to the smallest 3D bounding box that completely encloses the object of interest [94]. Consequently, 3D object detection involves estimating an object's position, size, and orientation in 3D space [95].

Previous research has explored 3D object detection using various input modalities, such as RGB images [63], point clouds [96], RGB-D data [94], and multimodal sensor combinations [97]. RGB-based approaches are typically categorized as monocular, stereo, pseudo-LiDAR, or multi-view methods [63]. Point cloud-based techniques are generally divided into region proposal-based and single-stage approaches [96]. Methods utilizing RGB-D data often build upon 2D detection or implement explicit data fusion strategies [94]. More broadly, multimodal methods are classified according to their fusion paradigm, including early, late, and deep fusion [97].

6DoF object pose estimation

6DoF object pose estimation seeks to determine an object's position in 3D space, typically represented by Cartesian coordinates (x, y, z), as well as its orientation, which is commonly described by roll, pitch, and yaw [98]. Figure 2.2(d) presents a representative 6DoF object pose estimation result for an RGB visual input. Existing research typically distinguishes between 6DoF object pose estimation at the instance level and at the category level, based on the availability of a 3D model (for example, a CAD model) for each specific object instance [99].

Instance-level 6DoF pose estimation addresses objects for which 3D models are available [94]. Analogous to 2D object detection, instance-level methods are generally categorized as traditional or DL-based approaches. Traditional techniques generally depend on CAD models or rendered 2D views derived from these models. In contrast, DL-based methods are often classified by input modality as RGB-based, depth-based, point-cloud-based, or RGB-D-based approaches [99].

RGB-based approaches are appealing because of the widespread availability and low cost of RGB cameras. However, these methods are sensitive to practical challenges, including occlusion, variations in illumination, limited distinctive visual features for certain objects, real-time processing constraints, and restricted generalizability [99]. Depth-based and point-cloud-based approaches utilize geometric information from 3D scanners or depth cameras. This information is particularly beneficial for objects that lack distinctive texture. However, these approaches are often limited by the expense of acquiring high-quality depth data, the labor required for manual annotation of training datasets, and significant computational demands, particularly for point-cloud processing. RGB-D-based methods aim to enhance robustness by integrating complementary RGB and depth cues. Nevertheless, achieving efficient feature fusion remains a significant challenge [99].

Category-level 6DoF object pose estimation is designed to generalize across instances within a category, including those not seen during training, when estimating object poses [94]. Both traditional and DL-based approaches have been explored in this context. Traditional methods, which frequently require extensive data collection for natural objects, have become less prevalent due to their labor-intensive and skill-demanding nature [100]. In contrast, DL-based methods have attracted considerable attention and are typically classified as regression-based or prior-based techniques [99]. Although category-level methods do not require highly accurate CAD models, their effectiveness is constrained by the limited availability of ground-truth data and significant intra-class variation in appearance and shape.

Despite notable advancements at both the instance and category levels, current methods continue to face challenges in generalizing to previously unseen object categories [101]. This limitation has prompted recent research on novel object pose estimation from monocular images [102–105]. Common strategies encompass the use of support views [106, 107], template matching with rendered views [108, 109], and the application of foundation models [110–112].

2.2.2 Computer vision techniques in manufacturing

Since the early 1970s, computer vision techniques have been utilized across various industrial contexts [113]. Despite initial advancements, widespread commercialization in manufacturing did not occur until the 1990s, primarily due to limited computing resources [114]. Research concerning industrial applications is frequently categorized as either *machine vision* [115] or *industrial vision* [116]. Within this framework, computer vision research is generally linked to methodological development, while machine vision focuses on practical deployment, especially the implementation of vision systems in industrial environments [115].

Two complementary strategies are commonly identified to address challenges in industrial vision [117, 118]. The first strategy aims to develop general-purpose systems applicable to multiple problems, whereas the second focuses on creating application-specific (ad-hoc) solutions tailored to particular tasks and settings [116].

Over the past several decades, computer vision has become increasingly integral to the advancement of industrial manufacturing systems, facilitating informatization, digitalization, and intelligentization [21, 119]. In manufacturing, vision applications may be classified according to various criteria, such as task type [120] or the stage of the manufacturing life cycle in which they are implemented [21].

2.2.3 Vision systems in assembly

The application of computer vision techniques to the assembly process in manufacturing has been a longstanding focus of research [21, 113]. Previous studies have examined topics such as automated assembly, quality control, and other related applications [21]. Researchers have explored visual perception and learning-based methods to enhance industrial robot performance in assembly tasks, particularly within unstructured environments [12]. Additional research has sought to improve robot adaptability to novel scenarios by leveraging computer vision [121]. While computer vision has traditionally been applied mainly for quality control, there is growing support for broadening its application within assembly processes [122]. Nonetheless, the transition of computer vision techniques from laboratory environments to industrial practice remains challenging because of limitations in algorithm robustness, data accessibility, and the lack of suitable benchmarking resources [21].

Challenges and limitations

Accurate estimation of object position and orientation is essential in assembly tasks, as it underpins reliable robotic grasping and subsequent manipulation. However, traditional computer vision techniques used in manufacturing often fail to meet these stringent requirements [21]. Contemporary DL-based methods have improved performance substantially, yet they still face notable challenges in precise position and orientation estimation [60]. In real production environments, performance is further degraded by factors such as occlusion [123], variable illumination [20], and camera movement [124].

Datasets are fundamental for developing learning-based vision systems [125] and are critical for achieving scalable industrial deployment [126]. However, domain-specific datasets frequently lack sufficient coverage or quality [127]. The collection of high-quality data in manufacturing environments presents significant challenges, while preprocessing and labeling processes are typically labor-intensive [128, 129]. Furthermore, datasets are crucial for benchmarking the performance of various computer vision and robotic systems [130]. Consequently, in numerous manufacturing scenarios, domain-specific benchmark datasets are necessary to facilitate meaningful and comparable evaluations of computer vision techniques [21].

Additional challenges for deploying AI-driven vision systems in industry include the cost of integrating new solutions into existing infrastructure [21], limited trust stemming from interpretability and explainability concerns [7], and the need to ensure safety in HRC settings [21].

2.3 Wire harness assembly

A wire harness bundle generally comprises cables and wires organized in a tree-like structure and includes components such as terminals, connectors, clamps, and wrapping materials [131]. Wire harnesses are extensively used in products with electrical systems, including automobiles, aircraft, and consumer electronics [19]. Although there has been considerable research interest in wire harness assembly, automation in this domain has not yet been successfully achieved [22].

2.3.1 Current manual assembly process

The manual assembly of automotive wire harnesses within passenger cabins typically comprises five sequential steps: preparation, transportation, untangling, routing, and assembly, as illustrated in Figure 2.3.

Figure 2.3: Overview of manual procedures for automotive wire harness assembly.

Preparation Wire harnesses are delivered to the assembly area as bundled units and stored in plastic bags or boxes. The initial rigidity of the bundle complicates subsequent manual handling. To enhance pliability for installation, the bundle is heated in an oven before manipulation. Since wire harnesses are delivered tied and packed in a defined sequence, and the logistics between delivery and the oven are predetermined, this step is well suited for automation with conveyors and conventional industrial robots.

Transportation After heating, an operator uses a dedicated tool to transfer the wire harness bundle into the vehicle cabin. At this stage, the wire harness bundle remains intact as a single unit. Traditional industrial robots can, in principle, be programmed to unload the heated bundle from the oven and position it within the vehicle body.

Untangling Once positioned, operators manually untie and separate the wire harness branches inside the vehicle. Automating this step requires robots to rapidly and accurately perceive key wire harness characteristics, such as shape and topology, both before and during manipulation. As the harness deforms continuously, robots must monitor deformation and adjust their actions accordingly. This process demands advanced perception and adaptive control.

Routing After untangling, operators route the harness through the vehicle body so that each branch reaches its designated mating area according to functional requirements. Similar to untangling, automating routing requires robots capable of manipulating deformable cables while continuously tracking their evolving configuration.

Assembly Finally, operators secure the harness to the chassis by inserting clamps and connect wire harness connectors to their corresponding components. Automating this step requires reliable perception of relevant objects, including their positions and orientations, as well as advanced robotic planning and control to execute insertion and mating operations.

2.3.2 Why automation is needed

The quality of wire harness assembly is crucial, as wire harnesses serve as fundamental components within electronic systems. Wire harnesses facilitate quality-dependent functions, such as engine control units and energy transmission, and support safety-critical operations, including maneuvering, driver assistance, and safety systems [132]. Proper connection of wire harnesses is essential to ensure reliable transmission of signals and electrical current among distributed components and to maintain overall system performance [133]. However, exclusive reliance on manual assembly impedes the maintenance of consistent quality due to the inherent variability of manual processes [134].

Efficient wire harness assembly has become increasingly important due to the rising prevalence of wire harnesses in modern vehicles and the demand for reduced production cycle times. The utilization of wire harnesses in automobiles has increased substantially in recent years [135], and industry projections indicate that this trend will persist [136]. This growth is driven by the expanding integration of electronic devices and broader transitions toward autonomous driving, electrification, and sustainable mobility [137]. Automotive manufacturers face ongoing pressure to enhance competitiveness and productivity; however, dependence on manual assembly constrains the extent of achievable productivity gains [32].

Operator safety and ergonomics continue to be central concerns in manufacturing. Many manual tasks in wire harness assembly require skilled labor and impose significant ergonomic strain, including heavy lifting, high-force pressing, and extended reaching. Such work can contribute to musculoskeletal disorders (MSD) and increase occupational safety and health (OSH) risks [138]. Exoskeletons and other assistive technologies can augment physical capability [139], yet they do not address the broader challenges associated with manual operations. Moreover, high-voltage wire harnesses, particularly those used in electric vehicles (EVs), require careful material handling to ensure safety, assembly quality, and reliability [140, 141].

Therefore, the need to maintain stringent quality and safety standards while enhancing productivity, ergonomics, and resource utilization motivates the adoption of automation, particularly robotic assembly, as a strategic approach to wire harness assembly automation [32, 34, 35].

2.3.3 An instance of deformable object manipulation

In robotic manipulation, objects are typically classified as either rigid or non-rigid, based on whether external forces induce changes in their shape [18]. Within the literature, *non-rigid objects* are also termed *flexible materials* [142] or *deformable objects* [11, 18, 143]. Consistent with prior work [7, 144], the terms *non-rigid objects*, *flexible materials*, and *deformable objects* are used interchangeably throughout this thesis unless otherwise specified.

Deformable objects are frequently classified by their dimensionality as one-dimensional, two-dimensional, or three-dimensional [142]. Two-dimensional deformable objects are further categorized as planar or cloth-like, according to their physical properties [18].

Robotic deformable object manipulation (DOM) provides significant advantages across a range of applications, such as handling flexible printed circuit boards in manufacturing [145], grasping tomatoes in the food industry [146], performing wound suturing in medical procedures [147], and manipulating garments in daily life [148]. Despite substantial advancements in robotics, DOM remains less developed compared to rigid object manipulation [18]. Manipulation of rigid objects primarily involves altering their pose and ensuring collision avoidance [149]. In contrast, manipulation of non-rigid objects necessitates consideration of shape changes, which may alter geometry or topology and result in manipulation failures [150]. Consequently, methods developed for rigid object manipulation are not directly applicable to DOM [18].

Robotic manipulation of deformable linear objects

Wire harnesses are typically categorized as deformable linear objects (DLOs) [151, 152] or, equivalently, as deformable one-dimensional objects (DOOs) [153, 154]. Wire harness assembly constitutes a prominent industrial example of DLO manipulation [142] and is a subset of the broader field of automated DLO assembly [7].

Robotic DLO manipulation represents a significant research focus across various industries [7, 18, 142]. Representative applications encompass wire insertion in the electrical sector [155, 156] and cable assembly within the automotive industry [19, 157]. In general, DLO manipulation demands integrated capabilities in modeling, perception, and task execution [11, 18, 144]. Recent research has investigated diverse modeling approaches and the integration of multiple sensors and AI in robotic systems, facilitating rapid, accurate, and multimodal perception [158] as well as adaptive modeling and control [11]. However, DLO manipulation continues to present significant challenges in flexible automation settings [159–161], including object detection, deformation-state estimation, object modeling, motion planning, and reliable execution of manipulation [11, 18, 143, 161].

Robotic perception in deformable linear object manipulation

Robotic perception serves as a fundamental requirement for the reliable execution of complex manipulation tasks [158]. In robotic manipulation of DLOs, accurate perception of key physical properties, such as geometry, topology, deformation, and strain, both prior to and during manipulation, is essential for effective modeling, motion planning, and manipulation strategies [18, 143, 149].

Robotic perception for DLO manipulation can be accomplished through single or multimodal sensing approaches, which may include visual, auditory, force, tactile, and range modalities [11, 143, 144, 162]. Tactile, force, and auditory sensing modalities are frequently employed to capture local contact and shape information [143, 162]. In contrast, global DLO properties, including overall geometry, topology, and deformation state, are typically estimated using visual perception methods [163, 164].

2.3.4 Automation challenges in wire harness assembly

The adoption of automation in automotive assembly has accelerated in response to increasing production demands [55, 165]. Assembly operations are primarily concentrated in the body shop and final assembly. Currently, body-in-white processes within the body shop exhibit a significantly higher degree of automation compared to tasks performed on the final assembly line [55].

In contrast, wire harness assembly during final assembly remains predominantly manual, resulting in ongoing production challenges. Consequently, both academic and industrial sectors are actively exploring methods to automate all or portions of the wire harness assembly process [166]. However, effective solutions have yet to be realized in actual production settings, as current equipment and technologies lack the flexibility and agility necessary for practical automation [30, 43, 55, 167].

The high degree of customization and deformability inherent to wire harnesses necessitates robotic systems capable of enhanced adaptability and autonomy during assembly. Although robots currently utilized in production execute specialized tasks efficiently, their capacity to manage variation is restricted by limited perception and cognitive abilities [43]. The deformable characteristics of wire harnesses further complicate robotic perception during the assembly process [7]. In addition to initiatives aimed at simplifying wire harness architecture and reducing assembly complexity [168], enhancing the intelligence and autonomy of industrial robots is a critical factor in advancing automation of wire harness assembly.

Intrinsic challenges in wire harness assembly automation

Robotic wire harness assembly involves challenges analogous to those encountered in robotic DLO manipulation, particularly in perception, modeling, and control. Robotic systems must accurately perceive wire harness geometry and topology, estimate manipulation states, and track deformation with high precision [11]. These capabilities provide the foundation for precise harness modeling and support the development of control strategies that accommodate the inherent flexibility of wire harnesses [169]. The deformability of wire harnesses increases the complexity of robotic motion planning [169] and manipulation task execution [149].

Furthermore, wire harness assembly is intrinsically more complex than the manipulation of canonical DLOs. The tree-like bundle structure imposes further constraints during manipulation [131], particularly when multiple DLOs are bundled and effective management of branch interactions is necessary. This branched structure has resulted in more specific classifications of wire harnesses, including branched deformable linear objects (BDLOs) [170, 171], DLO networks (DLONs) [172], and deformable multi-linear objects (DMLOs) [173]. Wire harnesses also combine deformable cables with rigid components, such as connectors and clamps, resulting in the specific classification as semi-deformable linear objects (SDLOs) [160]. Despite advancements in robotic manipulation of rigid objects, the small size and complex geometry of certain rigid harness components continue to pose significant challenges for automation [174].

Additionally, wire harness assembly often occurs in mass customization environments, where production lines must simultaneously accommodate multiple product variants. Consequently, numerous harness variants must be installed to support customized vehicle functionalities, which increases the demands on the adaptability and agility of robotic systems.

Extrinsic challenges in wire harness assembly automation

Beyond effective robotic manipulation, practical implementation requires automation solutions to withstand the rigorous conditions encountered in real-world production environments. To attain industrial viability, automation systems must exhibit operational effectiveness, reliability, and resilience. Jiang *et al.* [175] notes that many proposed approaches to wire harness assembly automation face practical challenges, including strict positioning accuracy requirements and the lack of reliable contactless measurement systems for real-time wire harness state estimation. Moreover, variable and dynamic shop-floor conditions present additional challenges to system robustness. Consequently, automation systems must function consistently and efficiently to meet production targets and sustain manufacturers' competitiveness.

The implementation of robotic systems introduces further challenges related to safety and risk management. In industrial settings, physical safeguards such as steel fences and laser curtains are routinely employed to protect workers [176]. As human–robot collaboration (HRC) becomes more widespread, safety considerations become increasingly critical [176]. Integrating robots into existing production sites may therefore require significant workspace redesign and comprehensive assessment of human–robot interactions.

Final assembly introduces further constraints, as automotive assembly lines typically move forward continuously and require robots to perform tasks in synchrony with the moving line [47]. These operational demands raise requirements for robot mobility and synchronization, thereby increasing the complexity of developing effective robotic assembly solutions.

Universal automation solutions are difficult to implement because of the significant variation in production demands across sectors and production types. This diversity results in differing criteria for automation, such as efficiency, effectiveness, reliability, and robustness, and frequently requires customization. The consequent need for adaptation increases development complexity and adds to the overall challenge of implementing automation systems in practice [177].

Robotic perception in robotized wire harness assembly

Robotic perception serves as a foundational element in the automation of wire harness assembly. Effective assembly operations require robots to identify and estimate the relevant physical and geometric properties of wire harnesses and their critical components. This requirement has driven research on robotic perception utilizing a range of sensing modalities [162, 175, 178, 179].

Visual perception is a crucial non-contact measurement technique, offering comprehensive and scalable information about wire harnesses [163, 164]. Vision enables essential robotic functions, including object recognition, tracking, and scene understanding [16]. The accessibility of visual data [15, 180], the pivotal role of computer vision in robotic manipulation [181], and its wide-ranging applications in manufacturing [21] collectively underscore the potential of computer vision to advance wire harness assembly automation. Vision-based approaches have also demonstrated effectiveness in the robotic manipulation of DLOs more broadly [163, 164, 182]. As a result, vision-based wire harness assembly continues to attract significant research attention [157, 172]; however, achieving robust implementation in actual production environments remains a significant challenge [22, 171].

Previous studies have also investigated the estimation of local physical properties and contact information through tactile sensing [175, 183] and acoustic sensing [162, 179]. Although these sensing modalities are important, they fall outside the scope of this thesis.

Chapter 3

Research approach

This chapter provides an overview of the research approach adopted in this thesis. It begins by outlining the philosophical worldview that informs the research, followed by a description of the research design and the methods implemented across the studies. The chapter concludes with a discussion of the strategies employed to ensure research quality.

3.1 Philosophical worldview

Philosophical worldviews, though often implicit, significantly influence research practice and should be explicitly stated [184]. These worldviews encompass foundational assumptions regarding reality and knowledge, offering a rationale for methodological decisions [185]. Explicitly articulating these assumptions enhances transparency, enabling readers to identify underlying perspectives and potential biases, and to interpret research findings with greater accuracy.

A researcher's worldview is shaped by external factors, including disciplinary traditions, research communities, and academic mentorship, as well as internal factors such as educational background and cultural experiences [185]. This thesis investigates AI-driven computer vision in manufacturing. The author's academic background in electrical engineering, computer science, AI, and computer vision has informed an empirical postpositivist orientation.

Postpositivism extends positivism, which posits a mind-independent reality that can be studied through observation and reasoning [186, 187] and emphasizes empirical grounding [188]. While postpositivism maintains that an objective reality exists, it rejects the possibility of perfectly apprehending that reality [189]. It further acknowledges the fallibility and theory-ladenness of observation, thereby motivating critical reflection on potential sources of bias [190]. Accordingly, this research prioritizes empirical evidence, systematic reasoning, and the falsification of claims in the investigation of technical solutions for computer vision applications [191].

However, research practice is seldom limited to a single worldview and often incorporates multiple perspectives based on the specific problem context [185]. Accordingly, this thesis employs both constructivism and interpretivism to facilitate inductive theory development [192]. This approach is implemented through qualitative synthesis of existing literature to assess the current state of knowledge and identify research gaps. Additionally, it integrates human experience and production requirements into the design and development of computer vision-enabled solutions.

3.2 Research design

A research design establishes a systematic framework for inquiry [185] and informs decisions related to data collection, analysis, and interpretation [193]. The research design adopted in this thesis incorporates three complementary frameworks: a positivist research design as the structural foundation, design science research (DSR) as the overarching paradigm for artifact development and evaluation, and a multiple methods design that integrates both qualitative and quantitative approaches. Collectively, these frameworks facilitate the identification of challenges and the development of technical solutions for computer vision applications in wire harness assembly automation.

Given the postpositivist worldview and its roots in positivism, this thesis adopts a positivist research design framework [187]. Consistent with the hypothetico-deductive model, this framework emphasizes a systematic progression from defining a topic of interest to theory framing [194]. The process commences with comprehensive literature studies to identify key challenges and knowledge gaps. These findings inform hypotheses and objectives that guide subsequent investigations into technical solutions. Solutions are developed and evaluated through rigorous data collection, analysis, and interpretation, enabling hypotheses to be tested and refined. This structure ensures that each stage contributes to a coherent understanding of computer vision applications in wire harness assembly automation.

Within this foundation, DSR offers a solution-oriented paradigm in which researchers contribute to scientific knowledge by designing artifacts that are both theoretically significant and practically useful [195]. These artifacts may take the form of constructs, models, methods, or instantiations that transform a problem situation from its current state to a desired state [196, 197]. The research is operationalized through the design science research methodology (DSRM) proposed by Peffers *et al.* [198], which comprises six iterative activities: problem identification and motivation, objective definition, design and development, demonstration, evaluation, and communication.

In accordance with DSRM [198], the research commenced with *problem identification and motivation* (also referred to as "awareness of the problem" [199]). This stage incorporated a cross-domain perspective on industrial constraints and human-centric automation to frame production-relevant vision-based automation (Paper 1), followed by an examination of automation challenges and fundamental issues related to vision systems in wire harness assembly (Paper 2 and Paper 3). These studies established the relevance of the research domain and clarified principal barriers to industrial adoption. The research then advanced to *objective definition*, where the scope and goals for vision-based solutions were articulated based on the challenges and gaps identified in the literature (Paper 3). These objectives informed the *design and development* of technical artifacts that enhance robotic perception using DL-based computer vision (Paper 4), facilitate scalable learning through automated dataset preparation (Paper 5), and support HRC to address ergonomic and operational concerns (Paper 6). The artifacts were subsequently *demonstrated* and *evaluated* in controlled laboratory environments and contexts representative of industrial practice, enabling assessment against established criteria and critical analysis of strengths and limitations [200]. Finally, the outcomes were disseminated through scholarly publications and integrated into this thesis as part of the *communication* activity.

This thesis also adopts a multiple methods design, defined as the use of several self-contained studies employing diverse methods to address the same research question or different components of a broader research program [201]. As summarized in Table 3.1, the research commenced with qualitative studies that identified industrial constraints influencing production-relevant vision-based automation (Paper 1 and Paper 2). Subsequently, the state of the art was assessed and future research directions for computer vision in wire harness assembly automation were proposed (Paper 2 and Paper 3). These studies established the theoretical foundation of the thesis, primarily addressing RQ1 by clarifying key challenges and barriers, and informing RQ2 by defining objectives for subsequent methodological development. Building on this foundation, quantitative experimental studies (Paper 4 and Paper 6) and DSR-driven artifact development (Paper 5) were conducted to address the identified challenges. These efforts primarily contributed to RQ2 through the design, demonstration, and evaluation of technical solutions using established metrics, while also generating insights that further informed RQ1 where limitations emerged.

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

Paper	Research design	Research method	Research quality assurance
1	Literature study Qualitative approach Descriptive design	Narrative literature synthesis - Thematic analysis - Integration of literature and industry requirements - Cross-domain synthesis of automation and HRC implications	Investigator triangulation [202] Peer debriefing [203] Expert review [204]
2	Literature study Qualitative approach Descriptive design	Systematic literature review [205] - Review protocol development - Database searching - Literature selection - Inductive coding - Inductive reasoning	PRISMA [206] flow diagram DARE criteria [207] Investigator triangulation [202] Peer debriefing [203]
3	Literature study Qualitative approach Descriptive design	Systematic literature review [205] - Review protocol development - Database searching - Literature selection - "Snowballing" [208] - Deductive coding - Inductive reasoning	PRISMA [206] flow diagram DARE criteria [207] Investigator triangulation [202] Peer debriefing [203] Expert review [204]
4	Experimental study Quantitative approach Experimental design	Empirical validation of DL models - Custom dataset creation - Model training - Performance benchmarking	Controlled lab environment Stratified sampling for dataset split Reproducible hyperparameters Statistical metrics Peer debriefing [203]
5	Design science study Quantitative approach Experimental design	Artifact development and evaluation - Systematic viewpoint generation - Multimodal data acquisition - Automatic data annotation - Performance benchmarking - Time-cost analysis	Modular workflow Systematic viewpoint planning Stratified sampling for dataset split Reproducible hyperparameters Statistical metrics Peer debriefing [203]
6	Experimental study Quantitative approach Experimental design	Prototype development Validation in laboratory and industrially relevant environments Convenience sampling [209] Quantitative measurements Statistical tests	Counterbalancing Validation aligned with TRL [210] Reporting effective sizes and confidence intervals Peer debriefing [203]

3.3 Research methods

A variety of research methods were employed in the studies associated with the appended publications, as summarized in Table 3.1.

Literature study

The research first established a broad industrial motivation by synthesizing cross-domain perspectives on automation and human-centered deployment (Section 3.1 of Paper 1). Building on this foundation, two systematic literature reviews (SLRs) were conducted to assess the current state of knowledge on wire harness assembly automation, with particular focus on robotic implementation in automotive final assembly and computer vision techniques supporting robotic automation. In addition to mapping existing solutions, the reviews aimed to identify key challenges limiting the adoption of computer vision and to highlight opportunities for developing vision-based approaches that could advance automation in industrial contexts.

The literature studies were guided by a constructivist worldview and adopted a qualitative, descriptive design grounded in SLR methodology [205]. SLR is widely recognized as a rigorous approach for synthesizing existing knowledge, identifying research gaps, and informing future inquiry [205, 211, 212]. The two reviews were complementary in scope: Paper 2 examined the state of the art in wire harness assembly automation, including collaborative and semi-automated solutions that have received limited attention, while Paper 3 focused specifically on computer vision applications in robotized wire harness assembly.

For each review, a protocol was established to define the search strategy, inclusion and exclusion criteria, and screening procedures. Literature was collected through structured database searches using predefined search strings, followed by a two-stage selection process comprising title-and-abstract screening and full-text assessment. Selection decisions were validated through consensus among multiple researchers. To minimize the risk of omitting relevant work, the "snowballing" strategy [208], which incorporates both backward and forward reference tracking, was also applied. The final set of literature was analyzed using a combination of deductive and inductive coding to identify recurring themes and patterns.

Experimental study

Building upon insights from literature reviews, experimental studies were designed to evaluate technical solutions that apply computer vision to enhance robotic perception in wire harness assembly. A quantitative approach, grounded in an empirical postpositivist perspective, was employed throughout the research. While prior research demonstrated the potential of computer vision, empirical examination and analysis of their performance under controlled conditions remained necessary [116].

To address this objective, a quantitative experimental design was implemented. The studies aimed to verify the effectiveness of the proposed solutions through systematic evaluation using statistical metrics and to analyze factors that influence performance. Numerical data were collected according to predefined experimental protocols and analyzed using established quantitative measures. All experiments were conducted under controlled conditions using standardized procedures to ensure reproducibility and reliability.

Specifically, Paper 4 presents an experimental study evaluating deep learning-based object detectors for wire harness connector detection, utilizing a custom dataset and standard metrics such as mean Average Precision (mAP). Paper 5 introduces a robot-assisted pipeline developed for systematic dataset generation and multimodal annotation. The pipeline is validated by benchmarking object detection and 6DoF pose estimation algorithms on a dataset generated with the pipeline, using standard metrics such as ADD-0.1d, 5cm-5deg, and Prj-5. Paper 6 examines a vision-based HRC framework for wire harness assembly, validated in both laboratory and industrially relevant environments. Validation metrics include success rate, cycle time, discomfort rating, and workload assessment using NASA-TLX [213], consistent with a hypothesis-driven experimental design [214].

Design science study

The research described in Paper 5 was conducted as a design science study that incorporated experimental evaluation. The primary research activity involved the design and development of an artifact, specifically a robot-assisted pipeline for systematic dataset preparation, following the principles of DSRM [198]. The pipeline comprises a robotic data acquisition module and an automated data annotation module. The experimental component represented the demonstration and evaluation phase of DSRM [198], validating the artifact's effectiveness under controlled conditions. A dataset of automotive wire harness connectors was collected using the developed pipeline to support demonstration and evaluation. The annotated dataset was subsequently used to train deep learning models for object detection and 6DoF pose estimation. Quantitative analysis involved measuring dataset generation time and computing standard evaluation metrics to assess the accuracy and robustness of the trained models. This approach facilitated evaluation of the practical effectiveness and efficiency of the robot-assisted dataset preparation pipeline.

3.4 Research quality assurance

Validity and reliability are essential criteria for evaluating research quality [215]. Reliability denotes the consistency of research methodology, particularly in relation to methods used in comparable studies [216]. Validity addresses the extent to which a study's findings accurately represent reality for populations beyond the immediate research context [217]. Cook and Campbell [218] distinguished between internal and external validity. Internal validity involves minimizing systematic errors during the design and implementation of a study, which establishes the foundation for external validity [219]. External validity concerns the generalizability and applicability of research findings beyond the specific study setting [205]. Table 3.1 presents the quality assurance strategies implemented across the studies in this thesis to maintain methodological rigor.

For qualitative research

Qualitative research must address both validity and reliability [185]. Reliability requires a consistent research approach across studies and researchers, while validity involves implementing procedures to ensure the accuracy of findings [216]. Several strategies were implemented to enhance the validity and reliability of the qualitative components within these studies.

To enhance reliability, the literature studies followed the systematic review methodology proposed by Kitchenham [205], which has been widely applied in systematic reviews in computer science and engineering [220–228].

Additionally, the quality of systematic reviews was assessed using the Database of Abstracts of Reviews of Effects (DARE) criteria [207], which evaluate reviews based on: (1) reporting of inclusion/exclusion criteria; (2) adequacy of the search; (3) synthesis of included studies; (4) assessment of study quality; and (5) provision of sufficient details on individual studies. A review qualifies under the DARE criteria if it meets the first three and at least one of the last two. Consistent with prior work [229, 230], these criteria were adopted to ensure rigor in the systematic literature reviews.

To mitigate subjective bias during data collection, analysis, and interpretation, investigator triangulation [202] was employed, complemented by peer debriefing [203] to incorporate feedback from researchers familiar with the topic. Protocol design, literature selection, and data analysis were collaboratively conducted by multiple researchers to maintain methodological rigor. Furthermore, an expert review [204] was conducted with specialists from academia and industry. Several of these experts also served as co-authors, contributing to the calibration of research methods and the cross-validation of interpretations.

For quantitative research

Quantitative research is subject to several potential threats to validity, including procedural inconsistencies, treatment effects, and participant experiences, all of which may compromise internal validity [231]. Furthermore, difficulties in generalizing results beyond the specific study context may negatively impact external validity [185]. The quantitative experimental studies presented in this thesis adhered to the assessment criteria established by Hammersley [231] to ensure reliability and validity in three key areas: accuracy and consistency of measurement procedures, generalizability of findings to broader populations, and adequacy of variable control.

Measurement reliability was improved by employing computer-based data collection systems that utilized predefined experimental protocols based on prior research. In experimental studies involving the training and evaluation of deep learning models with custom datasets (Paper 4 and Paper 5), stratified sampling structured the datasets according to data distribution and sample ratios to enhance the generalizability of the findings. In the within-subjects experimental study described in Paper 6, counterbalancing was applied to mitigate systematic bias resulting from order-related confounds to strengthen internal validity. Statistical analyses were conducted to minimize subjective bias in evaluating treatment performance. Peer debriefing [203] was utilized to further enhance the overall quality of the study.

Chapter 4

Research contributions

This chapter presents a summary of the contributions of the appended publications in relation to the research questions introduced in Chapter 1. An overview of each appended publication is provided, while the complete manuscripts are available in Part II of this thesis.

4.1 Contributions of appended publications

The appended publications collectively advance the theoretical understanding and practical implementation of computer vision in wire harness assembly automation. Table 4.1 presents an overview of the purpose and contributions of each publication in relation to the RQs.

The research first delineates the problem space and identifies key challenges (Paper 1, Paper 2, and Paper 3), then develops and validates technical solutions for robotic perception (Paper 4 and Paper 5), and ultimately integrates vision into HRC for real-world assembly tasks (Paper 6). This progression reflects a coherent trajectory from conceptual foundations to applied methodologies, addressing both theoretical and practical dimensions of the RQs.

Paper 1 situates the research within a broader manufacturing context by discussing automation and HRC as human-centered strategies for deploying advanced technologies in production environments. This publication offers minor insights to both RQs by clarifying cross-domain industrial constraints that influence the feasibility of vision-based automation and by providing a system-level rationale for integrating perception-enabled automation through human-centered approaches.

Building on this context, Paper 2 investigates automation in wire harness assembly and its associated challenges. This work contributes to both RQs by highlighting the significance of robotic visual perception in automation and by delineating key objectives and challenges for vision systems in this context.

Paper 3 makes major contributions to both RQs through a systematic literature review of the state of the art in computer vision applications for wire harness assembly automation. This paper synthesizes existing knowledge, identifies technical and practical challenges in applying computer vision to robotized wire harness assembly, and proposes future research directions. Collectively, these outcomes establish a strategic roadmap for the development of vision-based solutions to support robot-assisted assembly.

Drawing on these insights, Paper 4 evaluates the feasibility of deep learning-

Table 4.1: Overview of the purpose and contributions of each appended publication to addressing the RQs.

Paper	Purpose	Contribution to RQ1	Contribution to RQ2
1	Identify industrial constraints that complicate computer vision-based automation	**Minor** Clarifying cross-domain industrial constraints, such as variant richness and human-centered integration considerations, that influence the feasibility and requirements of deploying vision-based automation in production settings	**Minor** Providing a system-level rationale for applying vision-based automation through human-centered deployment and HRC
2	Identify objectives and challenges for vision systems in wire harness assembly automation	**Minor** Highlighted challenges in visual recognition exploiting intrinsic object features and ensuring robustness in production	**Minor** Suggested vision tasks: - Robot guidance - Process monitoring - HRC safety - Quality assurance
3	Review state of the art and propose research directions for computer vision in robotized wire harness assembly	**Major** Identified two core challenges: - Effective recognition using intrinsic features - Robustness under production conditions	**Major** Proposed directions: - Learning-based methods - Transfer learning - Practical evaluation - Vision-based HRC - Design-for-vision
4	Develop and evaluate DL-based computer vision models for wire harness component detection	**Minor** Identified issues with occlusion, ambiguous object features, and dataset limitations	**Major** - Demonstrated feasibility of DL for wire harness component detection - Highlighted need for multi-view strategies and data augmentation
5	Automate dataset preparation for developing DL-based machine vision	**Minor** Experimentally identified and validated challenges in 6DoF pose estimation for small, texture-less objects	**Major** Developed a robot-assisted pipeline for scalable, multimodal dataset generation
6	Develop a vision-based HRC method for wire harness installation	**Minor** Highlighted the challenge of maintaining assembly speed and addressing increased mental demand when introducing HRC	**Major** Integrated computer vision to facilitate safe and ergonomic HRC

based computer vision for detecting wire harness components, making a major contribution to RQ2. It also contributes to RQ1 by identifying further challenges that affect detection performance, such as similar product designs and occlusions.

To address the lack of benchmark datasets identified in Paper 3 and the data preparation challenges indicated in Paper 4, Paper 5 introduces a robot-assisted pipeline for automated dataset generation. This work makes a major contribution to RQ2 by enabling scalable creation of multimodal training data for object detection and 6DoF pose estimation, thereby accelerating the development of industrial vision systems. Additionally, it offers a minor contribution to RQ1 by experimentally identifying and validating the challenges of accurate 6DoF pose estimation for small, texture-less objects commonly encountered in wire harness assembly and other production scenarios.

Finally, building on insights from Paper 1 and Paper 3, Paper 6 advances the research by proposing a vision-driven HRC approach for wire harness installation. This study makes a major contribution to RQ2 by demonstrating that computer vision can facilitate safe, efficient, and ergonomic collaboration between humans and robots in wire harness assembly. Additionally, this study highlights the challenge of maintaining assembly speed and addressing increased mental demand when introducing HRC, thereby contributing to RQ1.

4.2 Summary of appended publications

4.2.1 Paper 1

B. Johansson, M. Despeisse, J. Bokrantz, G. Braun, H. Cao, A. Chari, Q. Fang, C. A. González Chávez, A. Skoogh, H. Söderlund, H. Wang, K. Wärmefjord, L. Nyborg, J. Sun, R. Örtengren, K. A. Schumacher, L. Espinal, K. C. Morris, J. Nunley Jr., Y. Kishita, Y. Umeda, F. Acerbi, M. Pinzone, H. Persson, S. Charpentier, K. Edström, D. Brandell, M. Gopalakrishnan, H. Rahnama, L. Abrahamsson, A. Ö. Rönnbäck and J. Stahre, "Challenges and opportunities to advance manufacturing research for sustainable battery life cycles," *Frontiers in Manufacturing Technology*, vol. 4, 2024. DOI: 10.3389/fmtec.2024.1360076.

Background

The increasing product customization and variant diversity in contemporary production, including battery production, are driving a greater demand for flexible and agile automation solutions [8]. This shift exposes the limitations of conventional industrial robotics, particularly for complex assembly and disassembly tasks in constrained workspaces, and underscores the necessity of advanced robotic perception as a critical capability [15].

Methodology

This study synthesizes existing research and industry requirements to identify key research themes relevant to sustainable battery life cycles. Within the human-centric production theme, Section 3.1 integrates findings from the literature and industry perspectives on automation and HRC to identify industrial constraints that complicate computer vision-based automation and clarify required capabilities of vision systems for supporting agile and flexible automation.

Contribution

This paper, particularly Section 3.1, offers a cross-domain analysis that elucidates industrial constraints affecting the feasibility and requirements for implementing vision-based automation in production environments. Specifically, the analysis examines how variant-rich manufacturing and the need for human-centered integration shape expectations for automation solutions, including flexibility, robustness, and safe operation in actual production settings.

The study connects the increasing diversity of product variants to the necessity for agile robotic manipulation and identifies computer vision as a key enabler for advancing intelligent robotic automation. Furthermore, the paper provides a system-level rationale for implementing vision-based automation via human-centered deployment. The analysis highlights HRC as a practical approach to integrating robotic repeatability with human adaptability in complex assembly and disassembly tasks. It also specifies the requirements for safe and efficient close-proximity interaction enabled by vision-based functions, such as advanced scene understanding.

4.2.2 Paper 2

O. Salunkhe, W. Quadrini, H. Wang, J. Stahre, D. Romero, L. Fumagalli and D. Lämkull, "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.

Background

Wire harness assembly within automotive final assembly processes remains predominantly manual and involves physically demanding tasks that present significant production and ergonomic challenges [20, 22]. These factors can reduce productivity and compromise assembly quality, while also negatively impacting operator safety and well-being.

Although there is considerable interest from both academia and the automotive industry, particularly regarding robotic automation, practical implementation of automation in actual production environments is limited [170]. No fully or partially automated solutions have been deployed on assembly lines. This gap highlights the necessity to systematically identify barriers to automation and to propose research directions that support effective wire harness assembly automation in practice.

Methodology

A systematic literature review was conducted according to the methodology proposed by Kitchenham [205]. Research quality was strengthened through investigator triangulation [202], in which three co-authors collaborated on research design, data collection, analysis, and interpretation.

A review protocol was established at the outset to promote transparency and reproducibility (see Section 2.1 in Paper 2). The search string `(wir* OR cabl*) AND (harness* OR bundl*) AND assembl*` was applied in Scopus, targeting the *Article Title*, *Abstract*, and *Keywords* fields on August 8, 2022. The searching were limited to the subject areas of *Engineering*, *Computer Science*, *Decision Sciences*, *Multidisciplinary*, and *Business, Management and Accounting*.

The initial search identified 959 records, which were reduced to 695 following subject filtering. A two-stage screening procedure was implemented. Title and abstract screening yielded 77 articles, and subsequent full-text screening resulted in a final set of 16 articles. The selection process was documented in accordance with PRISMA guidelines [206] (see Fig. 1 in Paper 2). The included studies were analyzed using inductive reasoning to identify recurring themes and research gaps.

Contribution

This paper presents an overview of existing research on robot-assisted wire harness assembly within automotive final assembly processes. As summarized in Table 1 in in Paper 2, the reviewed literature is categorized into two primary areas: studies addressing automation principles, such as modeling and control, and studies focusing on robotic systems, including robotic handling and trajectory planning for wire harness assembly. This synthesis identifies key challenges and outlines future research directions, emphasizing the systematic evaluation of existing methods and tools, as well as the transition toward human-centered automation enabled by HRC. The importance of behavioral studies of human operators in confined spaces is highlighted to support the safe and efficient integration of HRC.

Vision systems are identified as key enablers of robotic wire harness assembly, serving three primary functions. First, vision systems support visual servoing through object recognition and sub-process monitoring, including the detection of rigid components, interpretation of wire harness topology, alignment of paired parts, and tracking of moving assemblies. Second, in HRC environments, vision systems enable scene monitoring by enhancing safety and interaction through the detection of human and robot behaviors. Third, vision systems facilitate quality assurance by identifying faults during or after assembly. However, persistent challenges remain in achieving robust visual recognition based on intrinsic object features. Artificial fiducial markers are therefore suggested as a practical interim solution.

In addition, the paper highlights the need to evaluate vision system reliability under real production conditions, thereby informing future research on advanced computer vision techniques for wire harness assembly automation.

4.2.3 Paper 3

H. Wang, O. Salunkhe, W. Quadrini, D. Lämkull, F. Ore, M. Despeisse, L. Fumagalli, J. Stahre and B. Johansson, "A systematic literature review of computer vision applications in robotized wire harness assembly," *Advanced Engineering Informatics*, vol. 62, p. 102596, 2024. DOI: 10.1016/j.aei.2024.102596.

Background

Enhancing ergonomics, optimizing resource utilization, and improving assembly quality while maintaining safety are critical objectives in wire harness installation during the final assembly process [131, 133, 136]. Robotic assembly is a key enabler, providing advantages in replicability, transparency, and explainability compared to manual methods [43].

Despite these advantages, automating wire harness assembly remains highly challenging in real-world production environments [20, 174]. This complexity results from the deformability and variability of wire harnesses, which require robots to perceive their environment and adaptively manipulate objects with high precision using sensory input.

Vision systems are essential as the primary source of information for object localization and recognition [16]. Although vision-based robotic assembly has been explored in other sectors [21, 232–235], practical solutions for wire harness assembly in automotive manufacturing remain limited.

This gap highlights the need to identify the specific challenges associated with robotic visual perception in wire harness assembly and to outline future research opportunities for advancing vision-based robotic assembly solutions.

Methodology

The systematic review methodology described by Kitchenham [205] was employed. Research quality was enhanced through continuous collaboration among three co-authors using investigator triangulation [202].

Three RQs guided the 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?

A review protocol was established to ensure reproducibility (see Table 1 in Paper 3). The article selection process adhered to PRISMA guidelines [206] (see Fig. 4 in Paper 3). A Scopus search was conducted on September 6, 2023, using the predefined string `(wir* OR cabl*) AND (harness* OR bundl*) AND assembl*` within the Article Title, Abstract, and Keywords fields, yielding 1022 articles. After applying restrictions on subject area (*Engineering*, *Computer Science*, *Multidisciplinary*, *Business, Management and Accounting*, and *Decision Sciences*) and article language (English), 662 articles remained.

A two-step screening process was then implemented. The initial screening, based on titles and abstracts, resulted in the selection of 22 articles. The second screening, involving full-text review, further reduced the selection to 13 articles.

A "snowballing" technique [208] was employed to identify two additional articles, yielding a total of 15 studies for analysis. The selected studies were categorized according to assembly operation, object of interest, vision system type and location, and number of cameras. Inductive reasoning was applied to synthesize the findings and to identify research gaps. The quality of the review was enhanced by adhering to the DARE criteria [207].

Contribution

This systematic review identified 15 studies that investigated diverse computer vision techniques for automating wire harness assembly. Table 2 in Paper 3 provides a summary of each study's contributions regarding the application of computer vision techniques. Analysis of these 15 studies revealed that previous research primarily focused on the following areas:

- Robotic guidance enabled by object recognition of both rigid and deformable components

- Monitoring of sub-processes during robotic manipulation

- Quality assurance achieved through fault detection

Most studies concentrated on rigid components; however, interest in deformable parts has been increasing, as discussed in Section 4.1 of Paper 3. The review further observed that research has predominantly targeted routing and assembly tasks following wire harness placement in the vehicle body, as detailed in Section 4.2 of Paper 3.

Based on these findings, Paper 3 identifies two primary challenges in applying computer vision to wire harness assembly automation:

1. Achieving effective and efficient visual recognition by utilizing the intrinsic physical properties of wire harness components

2. Ensuring robustness that matches human visual performance under actual production conditions

Additionally, five future research directions have been proposed:

1. Adapting learning-based computer vision techniques to leverage intrinsic features and multimodal data[1]

2. Transferring knowledge gained from wire harness manufacturing automation

3. Evaluating vision-based solutions in practical production environments[2]

4. Investigating vision-based HRC and developing solutions for additional assembly operations[3]

5. Exploring modifications in product design to facilitate visual recognition

4.2.4 Paper 4

H. Wang and 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.

Background

As discussed in Paper 3, the development of rule-based object detection and pose estimation systems for wire harness components is challenging due to the components' varied and intricate geometries. Manual feature engineering is labor-intensive and often impractical when addressing complex component geometries. Previous studies have utilized artificial fiducial markers to facilitate object recognition [20, 236]. However, this approach necessitates additional procedures for attaching and removing markers, thereby complicating the assembly process. Learning-based object detection methods have consistently demonstrated superior performance compared to rule-based approaches [65, 237]. Nevertheless, as highlighted in Paper 3, research on DL-based solutions for wire harness components remains limited.

This study aims to validate the applicability of general-purpose DL-based object detectors for wire harness assembly, focusing on connector detection within wire harness connections as the primary application scenario. Furthermore, access to high-quality datasets is essential for the effective training and evaluation of learning-based object detection methods [238–241]. However, the lack of publicly available datasets for connectors presents a significant barrier, necessitating the development of a dedicated dataset for training and evaluating DL-based detection models.

[1] Paper 4 advances this direction
[2] Paper 6 advances this direction
[3] Paper 6 advances this direction

Methodology

A dataset comprising 360 images was created to represent 20 distinct connector types (see Fig. 2 in Paper 4). The dataset contains 60 images featuring mixed connectors and 300 images of individual connectors. For each individual connector, images were acquired from 6 standard views (front, back, left, right, top, bottom) and 9 random perspectives. Fig. 3, Fig. 4, and Fig. 5 in Paper 4 present representative examples from the dataset.

Image annotation was performed according to the methodology established in the PASCAL Visual Object Classes (VOC) Challenge 2007 [238]. Dataset statistics are presented in Table I and Figure 6 of Paper 4.

Two DL models were evaluated in this study: Faster R-CNN [74, 242], which is a two-stage detector, and YOLOv5 [243], which is a one-stage detector. Data augmentation techniques, including color jitter, scaling, and flipping, were applied to enhance model robustness. Model performance was assessed using precision and mAP metrics. Results are reported in Table II and Table III of the paper.

Contribution

This study demonstrated the feasibility of utilizing general-purpose DL-based detectors for connector detection within wire harness assembly processes. The primary contributions are the development of a dedicated connector dataset for benchmarking object detection models and the provision of empirical evidence supporting the effectiveness of DL-based detectors.

Nevertheless, several limitations were identified. The performance of the model was limited by data scarcity, highlighting the necessity for scalable methods of dataset generation. Detection accuracy was further diminished by ambiguous visual features, such as similarities among product designs and the occurrence of occlusions. These challenges emphasize the need to adopt multi-view or video-based detection strategies, develop advanced models that capture fine-grained features, and implement design-for-vision principles to enhance part distinguishability.

Collectively, these findings inform future research directions, such as automated dataset preparation[4] and the integration of robust perception capabilities into collaborative assembly systems[5].

4.2.5 Paper 5

H. Wang, G. Urbanos Uriel, K. El-Nahass, S. Ekered and B. Johansson, “Accelerating industrial vision: Systematic robot-assisted dataset preparation for object detection and pose estimation,” *Engineering Applications of Artificial Intelligence*, vol. 176, p. 114741, 2026. DOI: 10.1016/j.engappai.2026.114741.

Background

Paper 3 and Paper 4 emphasized the significance of learning-based computer vision for automating wire harness assembly and identified the limited availability of domain-specific datasets as a primary challenge. Conventional data collection and annotation methods are labor-intensive, time-consuming, and require specialized

[4]Paper 5 advances this direction
[5]Paper 6 advances this direction

expertise. These factors can compromise label quality [244] and hinder rapid model development [128]. This challenge is particularly acute in 6DoF pose estimation, where manual annotation is substantially more complex than in 2D labeling.

Synthetic data generation based on 3D CAD models offers a scalable alternative [245]. However, reducing the simulation-to-reality gap and creating highly accurate CAD models typically demand significant engineering resources and validation on real-world data [246]. Real-world automation methods that utilize gantry-based acquisition systems generally require extensive hardware installations and complex multi-camera calibration, thereby limiting scalability and adaptability [130, 247].

Conversely, employing a robotic arm with an eye-in-hand camera provides a flexible and cost-effective approach for multi-view data acquisition [248–250]. This method reduces hardware complexity, allows the robot to be repurposed after data collection, and enables pose annotation using robotic kinematics. Nevertheless, existing robot-assisted approaches often rely on heuristic viewpoint selection and do not provide an integrated workflow for systematic viewpoint planning or comprehensive multimodal annotation [248–250]. To address these gaps, this study examines a robot-assisted pipeline for automated real-world dataset preparation, enabling scalable and systematic viewpoint generation as well as detailed annotations for object detection and 6DoF pose estimation.

Methodology

The proposed pipeline consists of two primary components: robotic data acquisition and automated data annotation (see Fig. 2 in Paper 5). The data acquisition module employs a robotic arm with an eye-in-hand camera to systematically capture multi-view observations. Camera poses are generated within a spherical coordinate system to ensure scalable and uniform viewpoint coverage (detailed in Section 5.1 in Paper 5). The workspace is enclosed with a white background to facilitate chroma-key-based segmentation.

The annotation component utilizes the acquired data to generate multimodal ground-truth labels through a structured workflow. The target object is first segmented by removing the background using chroma keying, optionally supported by depth thresholding, to produce object masks. These masks are then used to derive 2D bounding boxes, as detailed in Section 5.2 in Paper 5. 6DoF object poses are subsequently computed using the predefined camera-view design and refined through bounding-box-based alignment to improve centroid consistency, as detailed in Section 5.3 in Paper 5. Finally, 3D bounding boxes are reconstructed from 2D boxes obtained from three orthographic views [251] and oriented using the computed 6DoF object poses, as detailed in Section 5.4 in Paper 5.

To validate the pipeline, an example system was implemented utilizing a UR5 collaborative robot and an Intel RealSense D435 RGB-D camera in an eye-in-hand configuration. An automotive wire harness connector dataset was collected and employed to train representative DL baselines for 2D object detection (Faster R-CNN [74, 242] and YOLOv5 [243]) as well as 6DoF pose estimation (DenseFusion [252] and YOLOv5-6D [253]). Performance was evaluated using standard metrics, including precision and mAP for detection, as well as ADD-0.1d, 5cm-5deg, and Prj-5 for pose estimation. Additionally, the efficiency of the pipeline was quantified by measuring the end-to-end dataset generation time, encompassing both robotic acquisition and automated annotation.

Contribution

A systematic, robot-assisted pipeline for real-world dataset preparation is introduced. This pipeline enables scalable data acquisition by providing uniform multi-view coverage with a single robot-mounted camera, thereby eliminating the need for large gantry systems and minimizing complex multi-camera calibration. It further offers automated multimodal annotation, generating ground-truth labels including 2D and 3D bounding boxes, segmentation masks, 6DoF poses, and point clouds. The pipeline also achieves significant efficiency improvements, reducing the average acquisition time to 2.390 seconds per image and annotation time to 0.254 seconds, which is approximately 150 times faster than conventional manual labeling.

The resulting dataset supports end-to-end training and evaluation of DL models for object detection and pose estimation. Experimental results indicate that, despite accelerated dataset generation, accurate 6DoF pose estimation remains challenging for small, texture-less wire harness components. This finding aligns with the challenges reported in Paper 3 and underscores the necessity for further research on pose estimation methods specifically designed for small, low-texture objects, as well as hybrid training strategies that integrate real-world and synthetic data to enhance robustness.

4.2.6 Paper 6

H. Wang, O. Salunkhe, A. Hartmann, S. Ekered, P. Bründl, J. Franke, J. Stahre and B. Johansson, "Vision-based human–robot collaboration for wire harness assembly in automotive manufacturing," *Submitted to a scientific journal (under revision)*, 2026.

Background

The development of fully autonomous solutions is limited by the inherent flexibility and variability of wire harnesses and the restricted workspace within vehicles. Paper 1 identifies HRC as a practical approach for integrating robotic repeatability with human adaptability in complex assembly and disassembly processes. Paper 2 proposes that HRC is an effective strategy for automating wire harness assembly in the final stages of automotive manufacturing. Paper 3 further recommends the integration of vision systems to enhance HRC in automating wire harness assembly. These findings motivated a study to investigate the implementation of vision-based HRC in non-rigid object assembly, such as wire harness assembly, and to assess its potential impact.

The installation of wire harnesses onto vehicle bodies is recognized as a promising application for HRC [254]. This task requires repetitive, force-intensive actions that generate high localized loads on the fingers and hands. As a result, this process significantly contributes to operator discomfort and increases the risk of MSDs. The task can be divided such that the robot executes the repetitive, force-intensive pressing actions, while the operator manages dexterous handling, precise positioning, and task sequencing.

Despite this division of labor, achieving robust robotic perception remains a primary challenge. Wire harnesses deform unpredictably due to gravity and physical contact, which results in frequent occlusions and significant pose variability at the component level. Additionally, rigid wire harness components are small and difficult to localize reliably within cluttered assembly environments. These factors complicate visual recognition and 6DoF pose estimation, both of which are essential for accurate robotic alignment and insertion. This challenge is further intensified in confined spaces where sensor viewpoints are limited.

Methodology

A vision-based HRC framework for wire harness assembly was developed and validated using the wire harness installation task. Within this framework, the operator manages clamp selection, positioning, and sequencing, whereas the robot executes repetitive and force-intensive pressing operations. A reusable clamp cap equipped with an ArUco marker [255–257] enables reliable detection of texture-less clamps in complex wire harness configurations. The estimated 6DoF pose of the marker is mapped to the clamp pose using a predefined cap-to-clamp relationship. The clamp pose is then transformed to the robot base frame via hand–eye calibration. To improve robustness in the presence of occlusions and workspace constraints, marker detection is combined with a hand-triggered strategy. In this approach, the operator positions the cap beneath the camera, restricting detection to a region of interest (ROI) defined by hand landmarks.

The workflow follows a sequential process: the operator attaches the cap to the target clamp, the robot estimates the pose using the marker, aligns the tool, performs the insertion, and returns to a standby position. Subsequently, the operator removes the cap and repeats the procedure for each subsequent clamp.

Corresponding with TRL [210] levels 4 to 6, the approach is validated through two-phase within-subjects experiments conducted in both a controlled laboratory environment and an industrially relevant car-body setting. Performance metrics for the HRC system comprise insertion success rate and assembly time. Human-centered outcomes are assessed using self-reported localized hand discomfort in both setups, general physical discomfort in the industrial setup (both on a 0 to 10 scale), and perceived workload measured by the NASA-TLX [213].

Contribution

Paper 6 introduces a vision-based HRC framework for wire harness assembly. This approach utilizes a detachable markerized cap to achieve reliable pose estimation of visually ambiguous wire harness components and to facilitate consistent robot-assisted clamp insertion. The framework is assessed through wire harness installation tasks conducted in both a controlled laboratory setting and an industrially relevant car-body environment, with environmental realism progressively increased according to the TRL scale [210]. Experimental results indicate that the HRC approach significantly reduces localized physical discomfort and physical demand while preserving task effectiveness, although it also leads to increased mental demand and longer cycle times. These results provide empirical support for the practical feasibility of the collaboration framework and elucidate the trade-offs among ergonomic benefits, cognitive demand, and throughput in non-rigid object assembly tasks. The paper further discusses implications for broader industrial adoption and proposes future directions, such as enhanced workspace monitoring and continued system refinement.

Chapter 5

Discussion

This chapter synthesizes the principal findings of the thesis in relation to the research questions introduced in Chapter 1. A critical evaluation of the contributions to academic research and industrial practice is provided, along with an assessment of research quality, limitations, and ethical and sustainability considerations. Future research directions are outlined to advance computer vision for the automation of wire harness and other non-rigid object assembly.

5.1 Answers to research questions

This section integrates the research findings with respect to the two RQs outlined in Chapter 1. The findings establish a foundation for identifying barriers (RQ1) and strategies (RQ2) pertinent to advancing wire harness assembly automation using computer vision.

5.1.1 The answer to RQ1

RQ1: What are the challenges in applying computer vision to wire harness assembly automation?

The answer in a nutshell

The complexity of objects, variability in assembly conditions, and strict industrial requirements present significant challenges for implementing computer vision in wire harness assembly automation. Wire harnesses comprise both rigid and deformable elements, each of which imposes unique perceptual requirements on vision systems. Rigid components are often small, lack distinctive texture, and exhibit high visual similarity, which limits reliable detection and pose estimation when relying exclusively on intrinsic appearance features. Deformable cable segments undergo continuous shape changes and possess branched topologies, which complicates structural recognition and tracking tasks. These inherent challenges are intensified by occlusions, cluttered environments, dynamic conditions in confined workspaces, variable lighting, and restricted camera placement options. Additionally, the scarcity of domain-specific datasets and the inefficiency of training data preparation impede the development and scalability of learning-based methods. Production requirements, especially takt-time constraints and safety considerations, further complicate the deployment of robust vision systems in real-world assembly environments.

Interpretation and implications

Wire harness assembly automation relies on vision systems for three primary functions: robot guidance via object recognition and pose estimation, safety assurance through workspace monitoring, and quality control through fault detection. Recent research identifies challenges in implementing computer vision for wire harness assembly automation at the object, scene, data, and operational levels.

At the object level, wire harnesses consist of both rigid and deformable elements. Rigid components, including connectors and clamps, require high-precision detection and pose estimation. However, their small size, limited texture, and often similar designs complicate markerless visual recognition. Artificial markers can facilitate perception but introduce additional steps or costs. In addition, deformable cables and wires necessitate robust topology recognition and tracking.

At the scene level, production environments introduce additional complexity. Cluttered backgrounds and variable illumination require robustness under inconsistent operating conditions, motivating the use of strategies such as multi-view imaging, depth sensing, and multimodal data fusion. Confined workspaces, such as vehicle bodies in automotive final assembly, further restrict camera placement and increase the likelihood of occlusions.

From a data-infrastructure perspective, learning-based methods provide a potential solution for markerless recognition. However, their effectiveness relies on access to large, high-quality datasets, which are scarce in this domain. Manual data collection and annotation are labor-intensive and incompatible with agile manufacturing, especially given the substantial number of product variants. These challenges underscore the need for scalable and automated pipelines for training data generation.

Operational constraints impose strict takt-time requirements, allowing minimal tolerance for computational delays or prolonged error recovery periods. Consequently, vision-based automation systems are required to sustain high efficiency while adhering to safety standards within collaborative environments. This requirement necessitates context-aware perception systems that can identify relevant objects and interpret human actions to enable safe and efficient HRC. Moreover, the diversity of wire harness designs increases the need for perception solutions that generalize effectively across different variants and operational environments without requiring extensive redevelopment.

5.1.2 The answer to RQ2

RQ2: How can computer vision be applied to wire harness assembly automation?

The answer in a nutshell

Computer vision facilitates wire harness assembly automation by enabling learning-based object detection and pose estimation, which leverage intrinsic appearance features to localize key components. To enhance scalability in industrial contexts, training data preparation may be optimized through robot-assisted dataset acquisition and automated annotation pipelines. Furthermore, computer vision can be incorporated into HRC workflows to enable safe and efficient task allocation as well as context-aware robot execution. Ultimately, validation in production-relevant industrial environments is necessary to evaluate the robustness, reliability, and scalability of these proposed solutions.

Interpretation and implications

Vision systems are essential for enabling adaptive robotic behavior and ensuring safe HRC during automated wire harness assembly. In practical applications, vision systems are required to detect small, texture-less components, accurately estimate their poses, monitor human actions, and identify faults to maintain assembly quality. Analysis of RQ2 identifies four complementary strategies for effectively applying computer vision.

The first strategy involves learning-based visual recognition. Although artificial markers simplify feature extraction, their use is often impractical at scale due to additional process steps and increased costs. Learning-based methods that leverage intrinsic appearance features provide a more scalable solution by enabling markerless detection and pose estimation. These methods are further enhanced by multi-view, multimodal, or video-based recognition, which increases robustness in dynamic and cluttered environments.

The second strategy focuses on automated dataset preparation. The effectiveness of learning-based methods depends on access to large, high-quality datasets, which remain limited within the wire harness domain. Manual data collection and annotation are time-consuming and incompatible with agile manufacturing processes. Robot-assisted acquisition and annotation pipelines address this challenge by enabling systematic multi-view data capture and efficient generation of multimodal ground-truth labels. This approach accelerates algorithm development while reducing human effort and associated costs.

The third strategy involves vision-based HRC. HRC provides a practical approach to wire harness assembly and other non-rigid object tasks that are difficult to fully automate. Vision systems facilitate collaboration by localizing relevant objects, monitoring operator actions, and supporting safety through context-aware perception. This integration improves ergonomics and increases flexibility and responsiveness in assembly operations.

Industrial validation is also crucial for ensuring robustness and reliability. Vision systems must maintain consistent performance despite variations in lighting conditions, workspace constraints, and takt-time. Advancing these technologies along the TRL scale [210] requires iterative testing and refinement within industrially relevant environments, preferably in collaboration with industry partners.

5.2 Positioning of this thesis

Automation serves as a foundational element in modern manufacturing, delivering significant gains in productivity and quality while minimizing human involvement in structured processes [2]. However, extending automation to resolve ongoing production challenges remains a key objective, driven by persistent demands for greater efficiency, improved quality, and sustainability [3]. This need is particularly pronounced in assembly processes, where manual operations persist and contribute to inefficiencies and ergonomic hazards [4, 5].

The assembly of wire harnesses in the automotive final assembly stage exemplifies this challenge. Despite its essential role in ensuring product functionality and safety, this process remains largely manual, thereby increasing production demands and imposing considerable ergonomic burdens on operators [169]. As wire harness variants proliferate and manufacturing shifts toward mass customization, automation solutions must accommodate diverse configurations and operate reliably under real production constraints [50, 54].

Prior research demonstrates that conventional industrial robots are most effective in repetitive, highly structured environments where object positions and orientations are predefined [40, 51]. These assumptions do not hold in complex assembly contexts, where flexibility and autonomy are required and safety considerations restrict deployment [47, 52, 53].

The complexity is further amplified when non-rigid objects are involved. Research in DOM, and specifically in DLO manipulation, indicates that effective automation requires robots to perceive and reason about object geometry, topology, and deformation to enable modeling, motion planning, and execution [11, 18, 143, 144, 149]. Wire harnesses are generally classified as DLOs; however, they present additional complexity due to their branched structure and hybrid composition, which includes deformable cables and rigid components such as connectors and clamps [7, 131].

In this context, robotic perception assumes a critical role. Although tactile and force sensing yield local contact information, visual perception is indispensable for acquiring global properties and facilitating object recognition and scene understanding at scale [16, 144, 163, 164]. Traditional and learning-based computer vision solutions encounter persistent challenges, such as occlusion, lighting variation, camera movement, requirements for production-level robustness, and limited availability of domain-specific datasets for training and benchmarking [20, 21, 128, 130].

Within this research landscape, this thesis extends prior work by examining computer vision as a practical enabler for robot-assisted wire harness assembly in final assembly processes. This thesis identifies industrial constraints that complicate computer vision-based automation and delineates the capabilities required of vision systems to support wire harness assembly automation, as established through literature reviews (Paper 1, Paper 2, and Paper 3). The thesis further addresses perception challenges that are both foundational and industrially actionable, with a specific focus on the recognition and localization of critical rigid components such as connectors and clamps, and the integration of visual perception into production-relevant workflows (Paper 4, Paper 5, and Paper 6). This focus complements broader DLO manipulation research by addressing the hybrid nature of wire harnesses, in which reliable automation depends on both reasoning about deformation and robust recognition and pose estimation of small, visually ambiguous rigid components essential for key assembly steps.

The thesis also addresses a central barrier to learning-based industrial vision—dataset scarcity—by developing automated solutions for scalable and systematic data generation and benchmarking (Paper 5). Additionally, in alignment with the increasing emphasis on human-centered automation in final assembly, this thesis examines how vision can facilitate safe and effective HRC by enabling context-aware interaction and task initiation in confined workspaces. This approach ensures that technical development remains aligned with real production constraints and safety requirements [176] (Paper 6).

5.3 Contributions of this thesis

This thesis systematically identifies the primary challenges associated with applying computer vision to wire harness assembly automation and proposes technical solutions to overcome these obstacles. Collectively, these contributions advance both academic research and industrial practice by offering theoretical insights, empirical evidence, and methodologies oriented toward practical implementation. By addressing critical barriers to vision-based automation and emphasizing applicability in production settings, this thesis facilitates the transition from research developments to scalable adoption within manufacturing environments.

To academia

This thesis outlines a strategic roadmap for overcoming barriers to vision-based automation in complex assembly, with a focus on non-rigid object assembly. It advances academic understanding of robotic visual perception in DLO manipulation and evaluates the application of these capabilities within the industrially significant context of wire harness assembly.

Literature reviews consolidate the current state of the art and identify critical research gaps in vision-driven wire harness assembly automation. These reviews underscore the central role of vision systems, delineate key technical and practical challenges, and propose future research directions for developing computer vision-based solutions in this domain. The resulting insights assist researchers in prioritizing fundamental challenges and establish a structured foundation for continued inquiry.

Addressing the identified gaps, this thesis advances knowledge on the application of learning-based computer vision in wire harness assembly automation. Quantitative and qualitative analyses characterize the performance and limitations of deep learning-based methods for 2D detection and 6DoF pose estimation of small, texture-less wire harness components. The findings underscore the necessity for continued development of robust visual recognition algorithms and highlight the importance of domain-specific benchmark datasets. Additionally, the thesis emphasizes the significance of product design considerations for enhancing part visual distinguishability.

To mitigate dataset scarcity and reduce the effort required for training data preparation, the thesis proposes a robot-assisted pipeline for systematic real-world dataset generation. This pipeline enables efficient multi-view data acquisition and automated ground-truth annotation, thereby reducing manual workload and expediting research and development of learning-based industrial vision. Utilizing this pipeline, the thesis also contributes a connector dataset to facilitate research on object detection and pose estimation within the wire harness domain.

In addition to fully automated solutions, the thesis advances vision-based HRC for wire harness assembly. It presents empirical evidence demonstrating the effectiveness of vision-enabled collaborative approaches and their potential benefits for ergonomics and flexibility in industrial environments. The thesis also identifies the necessity for future research to address the mental workload and operational time introduced by the robotic system.

To industry

This thesis provides practical guidance for applying computer vision to automate wire harness assembly and related manual tasks involving DLOs. The findings support industrial stakeholders in evaluating the opportunities and challenges associated with integrating vision-based robotics into production environments.

To convey technology maturity, the thesis utilizes the TRL scale, a standardized framework for assessing development progress in complex systems [210]. At the outset of this research, wire harness assembly in automotive final assembly was primarily conducted manually. Although academic research on automation and enabling vision technologies was available, its impact on industrial practice remained limited. Therefore, computer vision for wire harness assembly automation was initially evaluated at approximately TRL 2 to 3. The studies presented in this thesis advance technology maturity by developing and validating key enabling technologies. Deep learning-based detection and pose estimation of rigid wire harness components reach approximately TRL 4 through laboratory validation. The robot-assisted training data preparation pipeline is validated and demonstrated in both laboratory and industrially relevant environments, achieving approximately TRL 6. Similarly, the vision-based HRC solution for wire harness clamp insertion is validated and demonstrated in both laboratory and industrially relevant environments, also achieving approximately TRL 6. Additional validation in actual production settings is required to advance these technologies toward industrial deployment.

The thesis also introduces a cost-effective approach to training data preparation that supports the development of deep learning-based vision systems. The proposed robot-assisted dataset generation method is simple to implement and adaptable to small production batches. Employing robotic arms for data acquisition alongside production tasks can improve equipment utilization, reduce overall costs, and enhance operational flexibility.

In addition to technical results, the thesis presents implementation-oriented workflows, including procedures for dataset preparation, model training and evaluation, and strategies for integrating computer vision into HRC scenarios. These workflows offer actionable examples for practitioners aiming to implement vision-based automation in assembly environments.

The thesis further provides recommendations for design and process improvement. The findings demonstrate that product design, assembly environment conditions, and production requirements can adversely affect robotic visual perception in wire harness and other non-rigid object assembly. To address these challenges, the thesis emphasizes the importance of optimizing product and production design to improve part visual distinguishability and enable reliable vision-based automation.

5.4 Limitations of this thesis

To enhance transparency and support interpretation of the findings, several limitations are identified that may constrain the scope of applicability, particularly concerning the subject matter and the industrial context in which the results were obtained.

First, the research scope was deliberately limited to automotive wire harnesses intended for installation within the passenger cabins of passenger vehicles. This focus enabled detailed investigation of critical assembly tasks, such as connector mating and clamp insertion; however, it also restricts the generalizability of the findings. The appended studies primarily addressed vision-based detection and pose estimation of connectors (Paper 4 and Paper 5) and clamps (Paper 6). Although these are essential subtasks in wire harness assembly, they do not represent the complete set of wire harness assembly operations.

Second, the scale of experimental validation was constrained by available resources. The datasets for the DL-based computer vision experiments included 20 connector types (Paper 4 and Paper 5), which are representative of common automotive wire harness components. Nevertheless, they do not capture the full diversity of rigid wire harness component designs across vehicle models and manufacturers. Therefore, conclusions regarding algorithm performance may not be directly generalizable to certain industrial contexts. Additionally, the vision-enabled HRC experiments (Paper 6) were conducted in controlled laboratory and industrially relevant environments, rather than on fully operational assembly lines. This limitation restricts the evaluation of long-term reliability and impact of the vision-based method, as well as the takt-time compliance under actual production conditions.

Third, wire harnesses constitute a subset of DLOs that include both rigid and deformable elements [7]. The technical contributions of this thesis primarily address perception challenges associated with rigid components, such as connectors (Paper 4 and Paper 5) and clamps (Paper 6), while providing only partial solutions for deformable cable segments (Paper 6). Consequently, the applicability of these methods to other non-rigid object assembly tasks, such as cable routing, hose installation, or textile handling, remains unvalidated and requires further investigation.

Finally, the contextual boundaries of this research should be acknowledged. Wire harness assembly is present in multiple industries, including automotive, aerospace, and electronics, each characterized by distinct design standards, production volumes, and ergonomic requirements [19]. Even within the automotive sector, wire harness configurations vary significantly across vehicle categories, influencing assembly complexity and workflow design. Although Paper 1 aimed to provide cross-domain insights, the problem framing in Papers 2 and 3 is primarily informed by literature and application scenarios specific to wire harness assembly in automotive final assembly, as well as the production assumptions relevant to that context. As a result, the identified challenges, priorities, and research directions are most directly relevant to this context. The findings presented in this thesis also primarily reflect conditions within the Swedish automotive industry and the passenger vehicle segment. Application of the proposed solutions in other industrial contexts or production environments may require further adaptation.

5.5 Reflections on research quality

Validity and reliability constitute essential criteria for evaluating research quality [185, 215, 217, 219, 258]. This research integrates qualitative and quantitative studies. Section 3.4 outlines the procedures implemented in each study to improve transparency, reproducibility, and practical relevance within industrial contexts. This section further discusses strategies for research quality assurance, methods for assessing validity and reliability, and measures to identify and address limitations in both qualitative and quantitative studies.

Research quality of qualitative studies

In accordance with established guidance for software and engineering reviews [229], both systematic literature reviews were conducted under *a priori* review protocols, reported according to the PRISMA statement [206], and appraised using the DARE criteria [207]. These measures enhance methodological reliability by ensuring transparency, consistency, and rigor. The review protocols, search strategies, and selection procedures are documented in detail to improve transparency and facilitate reproducibility, as outlined in Section 2 of Paper 2 and Section 3 of Paper 3. Both reviews provide justification for the adequacy of their search strategies, assess the quality of included studies during selection, and synthesize findings using a structured approach. Additionally, the purpose, methodology, and findings of each included study are summarized (see Table 1 in Paper 2 and Table 2 in Paper 3).

To minimize potential bias in study selection, analysis, and interpretation, the reviews employed *investigator triangulation* [202]. Three researchers collaborated throughout the planning, searching, and selection phases, resolving disagreements by reaching consensus. Additionally, research validity was strengthened through *peer debriefing* [203] and *expert review* [204].

Research quality of quantitative studies

The reliability and validity of the quantitative studies are evaluated according to the framework proposed by Hammersley [231], focusing on three key aspects: measurement, generalization, and control of variables.

Measurement reliability is established through explicit specification of acquisition hardware, datasets, baselines, evaluation metrics, and statistical procedures. Paper 4 introduces a multi-class connector dataset with detailed documentation of data collection and annotation, stratified dataset splits, and clearly defined training settings. Results are reported at both class and aggregate levels. Paper 5 broadens the measurement scope by introducing a robot-assisted pipeline for systematic multi-view acquisition and multimodal ground-truth annotation. Object detection and 6DoF pose estimation models are evaluated using standard metrics, and efficiency is quantified using per-image and per-class timing measures. However, limitations in camera quality and residual manual annotation can reduce measurement reliability due to image noise and annotation deviations. Additionally, illumination and background conditions during capture may affect data quality. Paper 6 operationalizes task-level measurement in the HRC workflow using success rate (counting), cycle time (stopwatch), perceived workload (NASA-TLX) [213], and discomfort ratings on a 0–10 scale. Conclusion validity is enhanced by the use of statistical tests, confidence intervals, and effect sizes, as well as transparent reporting of experimental design, sample size, and analysis procedures. However, subjective self-reports can introduce bias into workload and discomfort measures.

The studies are primarily designed to demonstrate feasibility and performance in automotive wire harness assembly, rather than to establish broad theoretical generalization limits. Paper 4 supports external validity by including diverse connector classes and reporting per-class performance, thereby exposing failure modes linked to visually similar part exteriors that may be obscured by aggregate metrics. Paper 5 enhances ecological plausibility by employing systematic multi-view coverage and multimodal labels that more accurately reflect the spatial reasoning requirements of industrial settings. The viewpoint planning and labeling procedures are parameterized, supporting adaptation to additional components in related contexts. Paper 6 increases ecological validity by evaluating the collaborative workflow in both laboratory and industrially relevant car-body environments under semi-realistic constraints such as limited visibility, occlusion, and restricted access. The study also includes participants with varying levels of prior experience in robotics and assembly. At the same time, Paper 6 uses convenience sampling [209], which is common in HRC research but reduces representativeness compared to probabilistic sampling and therefore limits generalizability. In combination with stratified dataset splits and randomization where applicable (Paper 4 and Paper 5), these methodological choices support cautious generalization to similar products, sensors, and production conditions, while clarifying the boundaries of each study.

All three studies implement explicit controls to reduce confounding variables and support reproducibility. Paper 4 fixes detector configurations, data augmentation policies, and training protocols; reports the compute environment; and analyzes decision-threshold effects, helping to separate algorithmic variation from reporting artifacts. Paper 5 standardizes observation distance and employs identical, logged viewpoints for each object, ensuring that performance differences are attributable to model capability rather than viewpoint drift. Background and lighting are also controlled to support consistent segmentation and annotation. In the deep learning experiments (Paper 4 and Paper 5), deterministic data preprocessing, fixed random seeds for augmentation, and documented hardware specifications further reduce uncontrolled variance. Paper 6 controls the interaction state through hand-triggered robot execution, constrains robot motion using collaborative speed settings, and decomposes cycle time to isolate contributions from the operator, vision system, and robot. The within-subjects design also employs counterbalancing and standardized task scripts to limit order and operator effects.

The quantitative studies further incorporate peer debriefing [203] and expert review [204] to strengthen research quality through feedback from domain specialists and academic peers.

5.6 Reflections on research ethics

Ethical responsibility is fundamental to research integrity, especially in studies involving advanced automation and HRC. This research adopted a proactive strategy to identify and address potential ethical concerns [259–262]. Transparency and fairness were ensured through pre-established authorship agreements that clarified individual contributions and responsibilities, as well as through clear communication with experiment participants regarding the study's purpose and procedures. Privacy risks during data collection and analysis were minimized by primarily utilizing publicly accessible data, such as published literature and product images obtained with explicit provider consent. For studies involving human participants, written informed consent was obtained prior to experimentation, and all data were anonymized using participant identification codes before storage. No sensitive personal data were collected at any stage.

In the literature review, articles were selected exclusively based on relevance and quality, which limited bias toward specific authors or organizations. Experimental studies were reported comprehensively, including both favorable and unfavorable outcomes, to mitigate confirmation bias [263]. Each appended paper details the research design and methodology sufficiently to enable independent assessment of credibility [264]. Furthermore, original datasets, code, and supporting materials were securely archived to promote reproducibility and facilitate future verification.

Ethical considerations were particularly emphasized in the HRC experiments described in Paper 6, where human participants worked in close proximity to a robot. Participants were provided with comprehensive information regarding study objectives, procedures, and potential risks, and written consent was obtained prior to participation. The experimental setup complied with relevant ISO safety requirements for industrial robots [176]. The robot operated within reduced speed and force limits, and an emergency stop mechanism was available at all times. No personally identifiable information or personal video data were recorded. Ergonomic factors were addressed to minimize physical strain and cognitive load, and participants retained the right to withdraw at any time without penalty. Collectively, these measures ensured that the research was conducted with attention to safety, transparency, and respect for participants, while enabling valuable insights into collaborative assembly scenarios.

5.7 Reflections on the aspect of sustainability

Sustainable development is widely acknowledged as a global priority, and incorporating this dimension into research enhances both inclusivity and long-term relevance [265]. This section positions the research within the framework of sustainable development, with particular emphasis on the triple bottom line: economic, social, and environmental aspects [266].

Economic sustainability

Many assembly tasks involving non-rigid objects, such as wire harnesses, continue to be performed manually because of their inherent complexity and the limited autonomy of existing automation systems. This reliance on manual labor results in operational inefficiencies and inconsistent product quality. The integration of computer vision technologies can enhance robotic autonomy by enabling industrial robots to perceive both the task and the surrounding environment, thereby supporting complex assembly operations typically performed by human workers. Enhanced perception capabilities can improve process consistency and reduce error rates, leading to higher assembly quality, shorter cycle times, and increased throughput. Collectively, these improvements can strengthen economic sustainability by increasing cost efficiency and enhancing competitiveness.

However, potential benefits must be weighed against the associated costs and organizational challenges of implementing new technologies. Advanced vision and robotics systems generally require significant capital and operational investments, research and development to tailor solutions to specific production environments, and the acquisition of new competencies for system development, deployment, operation, and maintenance. A balanced evaluation should therefore adopt a holistic perspective, considering both benefits and costs across the entire system life cycle. In practice, such an assessment involves evaluating productivity gains, quality-related savings such as reduced scrap and rework, and flexibility benefits, alongside the total cost of ownership, which includes hardware, software, integration, calibration, training, and ongoing support. This comprehensive assessment provides a robust foundation for determining the net economic value of vision-based automation in wire harness assembly and other tasks involving DLO.

Social sustainability

The results presented in this thesis have significant implications for social sustainability, especially regarding worker health, job quality, and the responsible integration of automation within wire harness and other non-rigid object assembly.

Manual wire harness assembly frequently requires repetitive movements, awkward postures, and force-intensive actions, all of which are linked to increased risks of MSD and fatigue. Vision-driven automation enables robots to perceive components and process states with greater reliability, facilitating the transfer of the most physically demanding and monotonous tasks from human workers to machines. This approach can enhance ergonomics, lower injury risk, and promote longer, healthier working lives, thereby advancing socially sustainable manufacturing.

Simultaneously, increased robotic autonomy can alter task allocation and the competencies required of workers. Rather than eliminating human involvement, vision-enabled systems typically shift work toward supervision, exception handling, quality verification, and continuous improvement. Without effective management, this transition may lead to concerns regarding job displacement, work intensification, or unequal access to the skills necessary for new roles. Consequently, social sustainability relies not only on technical performance but also on governance practices, such as transparent communication about role changes, worker participation in system design, and consideration of psychosocial factors including trust, workload, and perceived control in HRC.

Although workforce development is beyond the immediate scope of this thesis, the findings highlight the necessity for complementary measures to facilitate a just transition. Such measures include structured upskilling and reskilling pathways, such as training in basic robot operation, troubleshooting, and data-driven quality practices, as well as cross-functional collaboration among researchers, educators, unions or worker representatives, and industry stakeholders. Responsible implementation should also involve risk assessments that address both physical and psychosocial impacts, along with post-deployment monitoring to ensure that safety and ergonomic improvements are achieved without unintended social consequences.

Environmental sustainability

Vision-based automation that enhances assembly quality can contribute to environmental sustainability by minimizing scrap, rework, and the associated consumption of energy and materials. Increased precision in wire harness assembly improves first-pass yield and reduces the occurrence of misrouted, damaged, or incorrectly terminated components. Consequently, a lower defect rate decreases material waste, the use of additional consumables, and the need for repeated processing. Enhanced process reliability thus promotes resource efficiency and can reduce the overall environmental footprint of manufacturing operations.

However, the implementation of AI-driven vision systems introduces environmental trade-offs due to increased computational demands. Training and operating deep learning models can elevate electricity consumption, and large-scale computing infrastructure often requires additional cooling and supporting equipment. The resulting indirect emissions are influenced by the underlying energy mix [267]. Thus, environmental sustainability depends on balancing resource savings in manufacturing with the life-cycle impacts of the digital infrastructure that supports automation. Future research should explicitly evaluate these trade-offs by comparing reductions in waste and rework with the additional computational and hardware requirements. Further investigation is also needed into more environmentally sustainable approaches for industrial deployment.

5.8 Future work

This thesis evaluated the feasibility of computer vision-enabled automation for wire harness assembly by introducing learning-based perception, automated training data generation, and an initial vision-based HRC prototype. However, a significant gap persists between laboratory demonstrations and robust industrial deployment. Future research directions are as follows: (i) reinforce the scientific foundations of visual perception for small, texture-less, and partially occluded components; (ii) develop reusable datasets and benchmarks to support advanced algorithm development; (iii) expand from isolated functions to comprehensive end-to-end assembly with integrated error handling; (iv) advance vision-based HRC toward seamless and safe collaboration; and (v) validate and adapt the proposed methods under industrial constraints and in related non-rigid object assembly domains.

Robust visual perception under production constraints

A primary research direction is to enhance the robustness and time efficiency of perception systems in realistic assembly conditions, since these factors continue to impede practical deployment. For wire harness component perception, sensitivity to confidence thresholds and limited training data necessitate expanding datasets and exploring calibrated decision strategies and uncertainty management to minimize false positives and false negatives in production environments. Misclassification among visually similar part types underscores the need for multi-view or video-based recognition approaches and the integration of additional contextual cues, such as cable geometry and local assembly context. Exploring attention-based architectures and transformer models for fine-grained feature extraction may further improve performance in environments characterized by clutter and frequent occlusions.

More broadly, perception pipelines must remain reliable under varying illumination, cluttered backgrounds, partial occlusions, and stringent latency constraints in actual production environments. These challenges motivate research on multi-modality fusion, including RGB-D data integration, uncertainty-aware inference, and evaluation protocols that jointly consider both accuracy and latency.

Benchmark datasets and synthetic data

Advancement in learning-based methods requires access to comprehensive and reusable datasets. Future research should expand domain-specific benchmark datasets to include a broader range of wire harness components, configurations, and environmental conditions, while ensuring sufficient sample diversity within each category for meaningful evaluation. Benchmarks should address tasks beyond 2D detection and 6DoF pose estimation in industrially representative scenes. These should include grasp detection, wire harness segmentation, and wire harness topology estimation, thereby enabling more comprehensive perception and robotic manipulation of wire harnesses.

Acquiring corresponding CAD models for dataset objects would enable precise 3D alignment and scalable synthetic data generation. This approach would facilitate research on integrating real and synthetic data and on comparing various sensing modalities. Simultaneously, automated real-world data acquisition and annotation pipelines should be extended to accommodate cluttered multi-object scenes. These pipelines should also incorporate refinement mechanisms to minimize systematic labeling errors and address non-ideal backgrounds. Adopting standardized data formats and metadata conventions will further enhance reproducibility and facilitate data reuse across different sites and platforms.

End-to-end automation and error handling

In addition to perception, achieving industrial viability requires robust integration with manipulation and task-level control. Future research should progress from addressing isolated vision system subtasks to supporting extended assembly sequences. It should also address online state estimation, anomaly detection, such as mis-insertions or missing parts, and the development of recovery behaviors that minimize downtime. Emphasis should be placed on evaluating cumulative error propagation across sequential operations and on designing recovery strategies that explicitly account for perception uncertainty and actuation constraints. Comprehensive end-to-end studies are essential for translating component-level performance into system-level reliability.

Vision-based HRC and workstation intelligence

HRC offers a practical solution for operations that are challenging to fully automate, particularly those involving deformable cable segments and frequent product variants. Future research should focus on developing adaptive vision systems capable of monitoring collaborative workspaces, recognizing human actions and intentions, and adjusting robot behavior to ensure safety and improve task fluency. Interaction design should prioritize ergonomic benefits and reduced cognitive load by implementing intuitive coordination mechanisms and delivering real-time feedback. Furthermore, the demonstrated HRC principles should be applied to additional wire harness assembly steps to identify tasks that benefit most from shared autonomy and to determine the optimal allocation of responsibility between humans and robots.

Industrial validation and domain transfer

Bridging the gap between prototypes and deployment requires prioritizing validation and demonstration in industrially relevant environments, with particular attention to robustness against production variability, adherence to takt-time constraints, and sustained reliability during extended operations. This objective requires the development of standardized protocols for performance benchmarking and safety compliance, as well as integration into production systems such as traceability, monitoring, maintainability, and interfaces to quality management. Transfer studies should also examine the extent to which the developed methods generalize to other non-rigid object assembly domains, such as cable routing, hose installation, and textile handling. The use of cross-domain datasets and simulation frameworks is essential for quantifying transferability and distinguishing between reusable and domain-specific perception and control components.

Chapter 6

Conclusion

This thesis examines computer vision as a critical enabler for robotic automation of non-rigid object assembly, with wire harness assembly in automotive final assembly serving as the primary application. Vision systems play a central role in robotic assembly by enabling object detection, pose estimation, and process monitoring. These systems further support quality assurance and enable safe collaboration between humans and robots. Despite their significance, vision-driven wire harness assembly automation has not yet achieved widespread adoption in industry.

This thesis identifies challenges inherent to non-rigid object assembly, such as object complexity, high variability, and frequent occlusion. These characteristics impede robust visual recognition and accurate object pose estimation, particularly for small, texture-less components. Although artificial markers can simplify perception tasks, their use introduces additional manual steps that reduce overall process efficiency. Learning-based computer vision provides a promising markerless alternative. However, its effectiveness depends on the availability of large, domain-specific datasets and efficient dataset preparation, both of which are often limited in industrial settings. In addition to challenges at the object and data levels, technology deployment in production environments introduces requirements concerning practicality, reliability, robustness, and sustainability.

To address these challenges, this thesis demonstrates the feasibility of deep learning-based object detection and pose estimation for wire harness components and presents a robot-assisted pipeline for generating training datasets. Integration of robotic data acquisition with automated multimodal labeling enables the pipeline to operate at a speed approximately 150 times greater than manual methods. This advancement facilitates efficient training and benchmarking of vision models. Additionally, this thesis examines vision-based human–robot collaboration for wire harness installation. The findings indicate that collaborative automation significantly reduces physical load, but introduces trade-offs in cognitive demand and cycle time. These results advocate takt-time-oriented workflow design and underscore the necessity of optimizing robot execution time for practical deployment.

Overall, this thesis conceptualizes computer vision as a human-centered integrator that connects robust perception, scalable data generation, and collaborative workflows within the domain of non-rigid object assembly automation. By addressing challenges in visual recognition, dataset preparation, and collaborative integration, this research offers foundational knowledge and practical tools to advance intelligent, adaptive, and human-centered robotic systems in next-generation manufacturing environments.

Bibliography

[1] H. Wang, “Toward enabling robotic visual perception for assembly tasks,” Licentiate thesis, Chalmers University of Technology, 2024. [Online]. Available: https://research.chalmers.se/en/publication/540720.

[2] X. Xu, T. Ji, P. Zheng, and L. Wang, “Human-centric manufacturing: Rethinking, re-justifying, and re-envisioning,” *Journal of Manufacturing Systems*, vol. 84, pp. 259–268, 2026. DOI: 10.1016/j.jmsy.2025.12.001.

[3] 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*. Publications Office of the European Union, 2020. DOI: 10.2777/082634.

[4] A. Giampieri, J. Ling-Chin, Z. Ma, A. Smallbone, and A. Roskilly, “A review of the current automotive manufacturing practice from an energy perspective,” *Applied Energy*, vol. 261, p. 114 074, 2020. DOI: 10.1016/j.apenergy.2019.114074.

[5] J. Liu, A. H. P. Tan, Y. Zheng, Y. Luo, and J. L. Y. Tuan, “Enhancing efficiency and profit growth in automotive final assembly: A system dynamics approach to automated feeding and fastening systems,” in *Selected Proceedings from the 2nd International Conference on Intelligent Manufacturing and Robotics, ICIMR 2024, 22-23 August, Suzhou, China*, 2025, pp. 221–233. DOI: 10.1007/978-981-96-3949-6_17.

[6] J. Heyn, P. Gümbel, P. Bobka, F. Dietrich, and K. Dröder, “Application of artificial neural networks in force-controlled automated assembly of complex shaped deformable components,” *Procedia CIRP*, vol. 79, pp. 131–136, 2019. DOI: 10.1016/j.procir.2019.02.027.

[7] 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.

[8] 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*, Springer, 2023, pp. 137–148. DOI: 10.1007/978-3-031-15602-1_11.

[9] International Organization for Standardization, *ISO 8373:2021 Robotics – Vocabulary*, 2021. [Online]. Available: https://www.iso.org/standard/75539.html.

[10] L. Rozo *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.

[11] 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.

[12] 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.

[13] T. K. Lien, "Manual assembly," in *CIRP Encyclopedia of Production Engineering*, Springer, 2014, pp. 825–828. DOI: 10.1007/978-3-642-20617-7_6624.

[14] S. Asif *et al.*, "Exploring tasks and challenges in human-robot collaborative systems: A review," *Robotics and Computer-Integrated Manufacturing*, vol. 97, p. 103 102, 2026. DOI: 10.1016/j.rcim.2025.103102.

[15] 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.

[16] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, eaat8414, 2019. DOI: 10.1126/science.aat8414.

[17] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach* (Pearson series in artificial intelligence), 4th ed. Pearson, 2021.

[18] 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.

[19] 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.

[20] 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. DOI: 10.1163/016918610X551782.

[21] 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.

[22] C. Kienle, B. Alt, F. Schneider, T. Pertlwieser, R. Jäkel, and R. Rayyes, "Ai-based framework for robust model-based connector mating in robotic wire harness installation," in *2025 IEEE 21st International Conference on Automation Science and Engineering (CASE)*, 2025, pp. 2444–2449. DOI: 10.1109/CASE58245.2025.11164054.

[23] 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*, Springer, 2020, ch. 1, pp. 1–59. DOI: `10.1007/978-3-662-53334-5_1`.

[24] International Society of Automation. "What is automation?" (2009), [Online]. Available: `https://www.isa.org/about-isa/what-is-automation` (visited on 05/29/2025).

[25] 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`.

[26] 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`.

[27] 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`.

[28] S. J. Hu, "Assembly," in *CIRP Encyclopedia of Production Engineering*, Springer, 2014, pp. 50–52. DOI: `10.1007/978-3-642-20617-7_6616`.

[29] 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`.

[30] G. Reinhart, "Assembly automation," in *CIRP Encyclopedia of Production Engineering*, Springer, 2014, pp. 52–54. DOI: `10.1007/978-3-642-20617-7_6617`.

[31] G. Reinhart, "Assembly line," in *CIRP Encyclopedia of Production Engineering*, Springer, 2014, pp. 55–60. DOI: `10.1007/978-3-642-20617-7_8`.

[32] F. J. Riley, *Assembly automation: a management handbook*, 2nd ed. Industrial Press Inc., 1996.

[33] 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`.

[34] G. Boothroyd, *Assembly automation and product design*, 2nd ed. CRC Press, 2005. DOI: `10.1201/9781420027358`.

[35] S. J. Hu *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`.

[36] 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`.

[37] B. Siciliano and O. Khatib, Eds., *Springer Handbook of Robotics*, 2nd ed. Springer, 2016. DOI: 10.1007/978-3-319-32552-1.

[38] 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.

[39] S. J. Russell and P. Norvig, "Robotics," in *Artificial intelligence: a modern approach*, ser. Pearson series in artificial intelligence, 4th ed., Pearson, 2021, ch. 26, pp. 925–980.

[40] T. K. Lien, "Robot," in *CIRP Encyclopedia of Production Engineering*, Springer, 2014, pp. 1068–1076. DOI: 10.1007/978-3-642-20617-7_6628.

[41] 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.

[42] B. Siciliano and O. Khatib, "Robotics and the handbook," in *Springer Handbook of Robotics*, 2nd ed., Springer, 2016, ch. 1, pp. 1–6. DOI: 10.1007/978-3-319-32552-1_1.

[43] 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.

[44] 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*, Springer, 2020, pp. 15–58. DOI: 10.1007/978-3-030-46212-3_2.

[45] S. Yun *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.

[46] 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.

[47] H. Chen, B. Zhang, and G. Zhang, "Robotic assembly," in *Handbook of Manufacturing Engineering and Technology*, Springer, 2015, pp. 2347–2401. DOI: 10.1007/978-1-4471-4670-4_105.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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, vol. 272, Springer, 2020, pp. 55–63. DOI: 10.1007/978-3-030-37841-7_5.

[54] 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.

[55] 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.

[56] R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed. Springer, 2022. DOI: 10.1007/978-3-030-34372-9.

[57] T. S. Huang, "Computer vision: Evolution and promise," in *1996 CERN School of Computing*, 1996, pp. 21–25. DOI: 10.5170/CERN-1996-008.21.

[58] B. K. P. Horn, *Robot vision*. MIT press, 1986.

[59] S. J. D. Prince, *Computer Vision: Models, Learning, and Inference*. Cambridge University Press, 2012. DOI: 10.1017/CBO9780511996504.

[60] 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.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] 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.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

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

[74] 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*, vol. 28, Curran Associates, Inc., 2015, pp. 1–9.

[75] 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.

[76] 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.

[77] 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.

[78] J. Redmon and A. Farhadi, *Yolov3: An incremental improvement*, 2018. DOI: 10.48550/arXiv.1804.02767.

[79] 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.

[80] 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.

[81] 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.

[82] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using yolo: Challenges, architectural successors, datasets and applications," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, 2023. DOI: 10.1007/s11042-022-13644-y.

[83] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023. DOI: 10.3390/make5040083.

[84] W. Liu *et al.*, "Ssd: Single shot multibox detector," in *Computer Vision – ECCV 2016*, Springer, 2016, pp. 21–37. DOI: 10.1007/978-3-319-46448-0_2.

[85] 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.

[86] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in *Computer Vision – ECCV 2018*, Springer, 2018, pp. 765–781. DOI: 10.1007/978-3-030-01264-9_45.

[87] X. Zhou, D. Wang, and P. Krähenbühl, *Objects as points*, 2019. DOI: 10.48550/arXiv.1904.07850.

[88] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017.

[89] 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.

[90] 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*, Springer, 2020, pp. 213–229. DOI: 10.1007/978-3-030-58452-8_13.

[91] 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.

[92] H. Zhang *et al.*, "Dino: Detr with improved denoising anchor boxes for end-to-end object detection," in *ICLR2023 - 11th International Conference on Learning Representations*, 2023.

[93] F. Li *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.

[94] 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.

[95] 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.

[96] 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.

[97] 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.

[98] 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.

[99] 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.

[100] 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.

[101] S. Thalhammer, D. Bauer, P. Hönig, J.-B. Weibel, J. García-Rodríguez, and M. Vincze, "Challenges for monocular 6-d object pose estimation in robotics," *IEEE Transactions on Robotics*, vol. 40, pp. 4065–4084, 2024. DOI: 10.1109/TRO.2024.3433870.

[102] Y. Li, G. Wang, X. Ji, Y. Xiang, and D. Fox, "Deepim: Deep iterative matching for 6d pose estimation," *International Journal of Computer Vision*, vol. 128, no. 3, pp. 657–678, 2020. DOI: 10.1007/s11263-019-01250-9.

[103] B. Okorn, Q. Gu, M. Hebert, and D. Held, "Zephyr: Zero-shot pose hypothesis rating," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 14 141–14 148. DOI: 10.1109/ICRA48506.2021.9560874.

[104] J. Sun *et al.*, "Onepose: One-shot object pose estimation without cad models," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 6815–6824. DOI: 10.1109/CVPR52688.2022.00670.

[105] Q. Gu, B. Okorn, and D. Held, "Ossid: Online self-supervised instance detection by (and for) pose estimation," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3022–3029, 2022. DOI: 10.1109/LRA.2022.3145488.

[106] Y. Liu *et al.*, "Gen6d: Generalizable model-free 6-dof object pose estimation from rgb images," in *Computer Vision – ECCV 2022*, Springer Nature Switzerland, 2022, pp. 298–315. DOI: 10.1007/978-3-031-19824-3_18.

[107] Z. Fan, P. Pan, P. Wang, Y. Jiang, D. Xu, and Z. Wang, "Pope: 6-dof promptable pose estimation of any object, in any scene, with one reference," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2024, pp. 7771–7781. DOI: 10.1109/CVPRW63382.2024.00773.

[108] V. N. Nguyen, Y. Hu, Y. Xiao, M. Salzmann, and V. Lepetit, "Templates for 3d object pose estimation revisited: Generalization to new objects and robustness to occlusions," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 6761–6770. DOI: 10.1109/CVPR52688.2022.00665.

[109] I. Shugurov, F. Li, B. Busam, and S. Ilic, "Osop: A multi-stage one shot object pose estimation framework," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 6825–6834. DOI: 10.1109/CVPR52688.2022.00671.

[110] W. Goodwin, S. Vaze, I. Havoutis, and I. Posner, "Zero-shot category-level object pose estimation," in *Computer Vision – ECCV 2022*, Springer Nature Switzerland, 2022, pp. 516–532. DOI: 10.1007/978-3-031-19842-7_30.

[111] S. Thalhammer, J.-B. Weibel, M. Vincze, and J. Garcia-Rodriguez, "Self-supervised vision transformers for 3d pose estimation of novel objects," *Image and Vision Computing*, vol. 139, p. 104 816, 2023. DOI: 10.1016/j.imavis.2023.104816.

[112] B. Wen, W. Yang, J. Kautz, and S. Birchfield, "Foundationpose: Unified 6d pose estimation and tracking of novel objects," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 17 868–17 879. DOI: 10.1109/CVPR52733.2024.01692.

[113] 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.

[114] 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.

[115] 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.

[116] J. Jolion, "Computer vision methodologies," *CVGIP: Image Understanding*, vol. 59, no. 1, pp. 53–71, 1994. DOI: 10.1006/ciun.1994.1004.

[117] H. Freeman, Ed., *Machine vision: algorithms, architectures, and systems* (Perspectives in Computing). Academic Press, Inc., 1988, vol. 22. DOI: 10.1016/B978-0-12-266720-6.X5001-3.

[118] H. Freeman, Ed., *Machine vision for inspection and measurement* (Perspectives in Computing). Academic Press, Inc., 1989, vol. 24. DOI: 10.1016/B978-0-122-66719-0.X5001-5.

[119] 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.

[120] 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.

[121] 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, ASME, 2018, V002T02A093. DOI: 10.1115/IMECE2018-86720.

[122] 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.

[123] 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.

[124] 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.

[125] L. Alzubaidi *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.

[126] 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.

[127] 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. DOI: 10.1109/ICCV.2017.97.

[128] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: A big data - ai integration perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328–1347, 2021. DOI: 10.1109/TKDE.2019.2946162.

[129] H. Chen, W. Wan, M. Matsushita, T. Kotaka, and K. Harada, "Automatically prepare training data for yolo using robotic in-hand observation and synthesis," *IEEE Transactions on Automation Science and Engineering*, pp. 1–17, 2023. DOI: 10.1109/TASE.2023.3304420.

[130] 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.

[131] 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.

[132] 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.

[133] 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.

[134] 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.

[135] B. L. Žagar *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.

[136] Strategic Market 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.

[137] S. Fankhauser *et al.*, "The meaning of net zero and how to get it right," *Nature Climate Change*, vol. 12, no. 1, pp. 15–21, 2022. DOI: 10.1038/s41558-021-01245-w.

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

[139] D. Romero *et al.*, "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.

[140] 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.

[141] 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.

[142] 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.

[143] J. Zhu *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.

[144] 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.

[145] 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.

[146] 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.

[147] L. Cao *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.

[148] 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.

[149] 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.

[150] 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.

[151] 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.

[152] A. Monguzzi, T. Dotti, L. Fattorelli, A. M. Zanchettin, and P. Rocco, "Optimal model-based path planning for the robotic manipulation of deformable linear objects," *Robotics and Computer-Integrated Manufacturing*, vol. 92, p. 102 891, 2025. DOI: 10.1016/j.rcim.2024.102891.

[153] 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.

[154] 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.

[155] 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.

[156] 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.

[157] 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.

[158] M. A. Lee *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.

[159] K. E. Stecke and R. P. Parker, "Flexible automation," in *Encyclopedia of Production and Manufacturing Management*, Springer, 2000, pp. 213–217. DOI: 10.1007/1-4020-0612-8_343.

[160] 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.

[161] D. Andronas *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.

[162] A. Hartmann, Z. Liu, S. Lamprecht, P. Bründl, and J. Franke, "Ai-driven multisensor quality inspection: A focus on robotic wire harness assembly," in *Advances in Production Management Systems. Cyber-Physical-Human Production Systems: Human-AI Collaboration and Beyond*, 2026, pp. 349–363. DOI: 10.1007/978-3-032-03538-7_25.

[163] A. Nair *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.

[164] 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.

[165] 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.

[166] O. Salunkhe *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.

[167] M. Despeisse *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*, 2023, pp. 521–535. DOI: 10.1007/978-3-031-43688-8_36.

[168] S. Chandra, I. J. Chung, A. Esmail, M. Blum, and R. Bhandari, *Wiring system architecture*, 2022.

[169] 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.

[170] 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.

[171] 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.

[172] 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," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 12 585–12 591. DOI: 10.1109/ICRA57147.2024.10611551.

[173] A. Caporali, K. Galassi, R. Zanella, and G. Palli, "Gnn topology representation learning for deformable multi-linear objects dual-arm robotic manipulation," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 14 738–14 751, 2025. DOI: 10.1109/TASE.2025.3562231.

[174] 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.

[175] 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.

[176] International Organization for Standardization, *Iso 10218-1:2025 robotics — safety requirements*, 2025.

[177] W. ElMaraghy, H. ElMaraghy, T. Tomiyama, and L. Monostori, "Complexity in engineering design and manufacturing," *CIRP Annals*, vol. 61, no. 2, pp. 793–814, 2012. DOI: 10.1016/j.cirp.2012.05.001.

[178] K.-m. 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.

[179] 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.

[180] 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.

[181] 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.

[182] 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.

[183] 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.

[184] B. D. Slife and R. N. Williams, *What's Behind the Research? Discovering Hidden Assumptions in the Behavioral Sciences.* SAGE Publications, Inc., 1995. DOI: 10.4135/9781483327372.

[185] J. W. Creswell and J. D. Creswell, *Research design: qualitative, quantitative, and mixed methods approaches*, 6th ed. SAGE Publications, Inc., 2022.

[186] J. Paley, "Positivism," in *The SAGE Encyclopedia of Qualitative Research Methods*, vol. 2, SAGE Publications, Inc., 2008, pp. 647–650. DOI: 10.4135/9781412963909.

[187] 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, 2nd ed., Chandos Publishing, 2002, ch. 2, pp. 25–47. DOI: 10.1016/B978-1-876938-42-0.50009-5.

[188] 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.

[189] J. G. Ponterotto, "Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science," *Journal of Counseling Psychology*, vol. 52, no. 2, pp. 126–136, 2005. DOI: 10.1037/0022-0167.52.2.126.

[190] D. C. Phillips and N. C. Burbules, *Postpositivism and educational research.* Rowman & Littlefield Publishers, Inc., 2000.

[191] E. G. Guba, Y. S. Lincoln, and S. A. Lynham, "Paradigmatic controversies, contradictions, and emerging confluences," in *The SAGE Handbook of Qualitative Research*, 5th ed. SAGE Publications, Inc, 2017, pp. 108–150.

[192] T. E. Costantino, "Constructivism," in *The SAGE Encyclopedia of Qualitative Research Methods*, vol. 2, SAGE Publications, Inc., 2008, pp. 116–120. DOI: 10.4135/9781412963909.

[193] E. Bell, A. Bryman, and B. Harley, *Business research methods*, 6th ed. Oxford University Press, 2022.

[194] Y. S. Park, L. Konge, and A. R. J. Artino, "The positivism paradigm of research," *Academic Medicine*, vol. 95, no. 5, pp. 690–694, 2020. DOI: 10.1097/ACM.0000000000003093.

[195] 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, Springer, 2010, ch. 2, pp. 9–22. DOI: 10.1007/978-1-4419-5653-8_2.

[196] 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.

[197] 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.

[198] 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.

[199] W. K. Vijay K. Vaishnavi, *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*, 2nd ed. CRC Press, 2015. DOI: 10.1201/b18448.

[200] J. van Aken, A. Chandrasekaran, and J. Halman, "Conducting and publishing design science research," *Journal of Operations Management*, vol. 47-48, no. 1, pp. 1–8, 2016. DOI: 10.1016/j.jom.2016.06.004.

[201] J. Morse, "Procedures and practice of mixed method design: Maintaining control, rigor, and complexity," in *SAGE Handbook of Mixed Methods in Social & Behavioral Research*, 2nd ed., SAGE Publications, Inc., 2010, ch. 14, pp. 339–352. DOI: 10.4135/9781506335193.n14.

[202] N. K. Denzin, *The research act: A theoretical introduction to sociological methods*, 1st ed. Routledge, 2009. DOI: 10.4324/9781315134543.

[203] V. J. Janesick, "Peer debriefing," in *The Blackwell Encyclopedia of Sociology*, John Wiley & Sons, Ltd, 2015. DOI: 10.1002/9781405165518.wbeosp014.pub2.

[204] 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.

[205] B. Kitchenham, "Procedures for performing systematic reviews," Keele University, Tech. Rep. TR/SE-0401, 2004.

[206] M. J. Page *et al.*, "The prisma 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, n71, 2021. DOI: 10.1136/bmj.n71.

[207] 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.

[208] 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.

[209] S. J. Stratton, "Population research: Convenience sampling strategies," *Prehospital and Disaster Medicine*, vol. 36, no. 4, pp. 373–374, 2021. DOI: 10.1017/S1049023X21000649.

[210] 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.

[211] J. Rowley and F. Slack, "Conducting a literature review," *Management research news*, vol. 27, no. 6, pp. 31–39, 2004. DOI: 10.1108/01409170410784185.

[212] D. Denyer and D. Tranfield, "Producing a systematic review," in *The SAGE Handbook of Organizational Research Methods*, Sage Publications Ltd., 2009, pp. 671–689.

[213] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Human Mental Workload*, ser. Advances in Psychology, vol. 52, North-Holland, 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9.

[214] G. Hoffman and X. Zhao, "A primer for conducting experiments in human–robot interaction," *ACM Transactions on Human-Robot Interaction*, vol. 10, no. 1, pp. 1–31, 2020. DOI: 10.1145/3412374.

[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.

[216] G. R. Gibbs, *Analyzing Qualitative Data*. SAGE Publications, Ltd, 2007. DOI: 10.4135/9781849208574.

[217] 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.

[218] T. D. Cook and D. T. Campbell, *Quasi-experimentation: Design & analysis issues for field settings*. Houghton Mifflin, 1979.

[219] 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.

[220] 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.

[221] 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.

[222] 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.

[223] J. Dalzochio *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.

[224] 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.

[225] 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.

[226] 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.

[227] 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.

[228] 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.

[229] 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.

[230] 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.

[231] M. Hammersley, "Assessing validity in social research," in *The SAGE Handbook of Social Research Methods*, SAGE Publications Ltd, 2008, ch. 4, pp. 42–53. DOI: 10.4135/9781446212165.

[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.

[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.

[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.

[235] 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.

[236] 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.

[237] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015. DOI: 10.1038/nature14539.

[238] 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.

[239] T.-Y. Lin *et al.*, "Microsoft coco: Common objects in context," in *Computer Vision – ECCV 2014*, Springer, 2014, pp. 740–755. DOI: 10.1007/978-3-319-10602-1_48.

[240] 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.

[241] O. Russakovsky *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.

[242] 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.

[243] G. Jocher, *Yolov5 by ultralytics*, version 7.0, 2020. DOI: 10.5281/zenodo.3908559.

[244] H. Lyu *et al.*, “Farpls: A feature-augmented robot trajectory preference labeling system to assist human labelers’ preference elicitation,” in *Proceedings of the 29th International Conference on Intelligent User Interfaces*, 2024, pp. 344–369. DOI: 10.1145/3640543.3645145.

[245] N. Patki, R. Wedge, and K. Veeramachaneni, “The synthetic data vault,” in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2016, pp. 399–410. DOI: 10.1109/DSAA.2016.49.

[246] N. Jakobi, P. Husbands, and I. Harvey, “Noise and the reality gap: The use of simulation in evolutionary robotics,” in *Advances in Artificial Life*, 1995, pp. 704–720. DOI: 10.1007/3-540-59496-5_337.

[247] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, vol. 1, 2006, pp. 519–528. DOI: 10.1109/CVPR.2006.19.

[248] R. Jensen, A. Dahl, G. Vogiatzis, E. Tola, and H. Aanæs, “Large scale multi-view stereopsis evaluation,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 406–413. DOI: 10.1109/CVPR.2014.59.

[249] P. Koch, M. Schlüter, S. Thill, and J. Krüger, “Towards robot-assisted data generation with minimal user interaction for autonomously training 6d pose estimation in operational environments,” *Procedia CIRP*, vol. 120, pp. 249–254, 2023. DOI: 10.1016/j.procir.2023.08.045.

[250] M. Zürn, M. Dzubba, C. Reiff, S. Ajdinović, A. Lechler, and A. Verl, “Cobot for automated vision data: Streamlining production with automated annotations for machine learning,” in *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, 2024, pp. 1–6. DOI: 10.1109/ACDSA59508.2024.10468034.

[251] W. Hu, J. Zheng, Z. Zhang, X. Yuan, J. Yin, and Z. Zhou, “Plankassembly: Robust 3d reconstruction from three orthographic views with learnt shape programs,” in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 18 449–18 459. DOI: 10.1109/ICCV51070.2023.01695.

[252] C. Wang *et al.*, “Densefusion: 6d object pose estimation by iterative dense fusion,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 3338–3347. DOI: 10.1109/CVPR.2019.00346.

[253] C. G. A. Viviers, L. Filatova, M. Termeer, P. H. N. de With, and F. van der Sommen, “Advancing 6-dof instrument pose estimation in variable x-ray imaging geometries,” *IEEE Transactions on Image Processing*, vol. 33, pp. 2462–2476, 2024. DOI: 10.1109/TIP.2024.3378469.

[254] O. Salunkhe, J. Stahre, D. Romero, D. Li, and B. Johansson, "Specifying task allocation in automotive wire harness assembly stations for human-robot collaboration," *Computers & Industrial Engineering*, vol. 184, p. 109 572, 2023. DOI: 10.1016/j.cie.2023.109572.

[255] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014. DOI: 10.1016/j.patcog.2014.01.005.

[256] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using mixed integer linear programming," *Pattern Recognition*, vol. 51, pp. 481–491, 2016. DOI: 10.1016/j.patcog.2015.09.023.

[257] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Speeded up detection of squared fiducial markers," *Image and Vision Computing*, vol. 76, pp. 38–47, 2018. DOI: 10.1016/j.imavis.2018.05.004.

[258] 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.

[259] B. L. Berg, *Qualitative research methods for the social sciences*, 7th ed. Allyn & Bacon, 2009.

[260] S. N. Hesse-Biber, *The Practice of Qualitative Research*, 3rd ed. SAGE Publications, Inc, 2016.

[261] K. F. Punch, *Introduction to social research: quantitative and qualitative approaches*, 2nd ed. SAGE Publications, Inc, 2005.

[262] J. E. Sieber, "Planning ethically responsible research," in *The SAGE Handbook of Applied Social Research Methods*, 2nd ed., SAGE Publications, Inc., 2009, ch. 4, pp. 106–142. DOI: 10.4135/9781483348858.

[263] 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.

[264] W. L. Neuman, *Social research methods: qualitative and quantitative approaches*, 7th ed. Allyn & Bacon, 2011.

[265] M. Mishra *et al.*, "A bibliometric analysis of sustainable development goals (sdgs): A review of progress, challenges, and opportunities," *Environment, Development and Sustainability*, vol. 26, no. 5, pp. 11 101–11 143, 2024. DOI: 10.1007/s10668-023-03225-w.

[266] J. Elkington, "The triple bottom line," in *Environmental management: Readings and cases*, 2nd ed., SAGE Publications, Inc., 1997, pp. 49–66.

[267] C.-J. Wu *et al.*, "Sustainable ai: Environmental implications, challenges and opportunities," in *Proceedings of Machine Learning and Systems*, vol. 4, 2022, pp. 795–813. [Online]. Available: https://proceedings.mlsys.org/paper_files/paper/2022/file/462211f67c7d858f663355eff93b745e-Paper.pdf.