Multi-sensor fusion for inventory monitoring

Ruofan Lin

Department of Automatic Control and Systems Engineering
University of Sheffield

This dissertation is submitted for the degree of
Doctor of Philosophy

September 2024

Acknowledgements

First and foremost, I would like to extend my deepest gratitude to my supervisor, Professor Ashutosh Tiwari, whose support, encouragement, and patience have been instrumental throughout my Ph.D. journey. Each encounter with him provided valuable insights and constructive suggestions that significantly enriched my research. His vast knowledge across many domains and his assistance in both academic and non-academic matters have been a boon to my doctoral pursuit. I am incredibly fortunate to have had the opportunity to complete my academic journey under his guidance, which has enabled me to view problems from a fresh perspective.

I also wish to express my heartfelt appreciation to all my colleagues in Professor Ashutosh Tiwari's group, who have contributed to a conducive research environment and were always ready to provide necessary resources. A special thanks goes to Dr. Boyang Song, who provided support during the initial, yet bewildering, stages of my journey.

I cannot overstate my appreciation for my parents, Hanhong Peng and Changyong Lin, whose unconditional and wholehearted support has been my bedrock at all times. The values they instilled in me growing up have illuminated my path, both in academia and in life. Moreover, my deepest thanks go to my partner, Dr Manyu Zhang whose companionship, understanding, and sacrifices during this period have been nothing short of a sanctuary.

Moreover, I consider myself fortunate to have spent a significant portion of my life in Sheffield, a place I hold dear second only to my hometown city Shenzhen. I have lived through my youthful years here, encountering many remarkable individuals along the way. The warmth, openness, and friendliness of the people here resonate with the hospitable nature of my hometown, making this a treasure trove of unforgettable memories.

In conclusion, my sincere appreciation goes to all who have contributed to making this thesis a reality. Your impact has not only facilitated my academic journey but also shaped my character, and for that, I am eternally grateful.

Abstract

A large number of parts are used in the production activities of the assembly plant. The quantity and quality of these parts in the process of transportation, distribution, use, and recycling in the factory have an impact on production. With a large amount of parts and tools, how to provide the right types and quantity of parts at the time and right location becomes the key to business success. Especially in recent years, with the advent of flexible manufacturing systems, various parts or tools are being customised according to individual preferences, leading to heightened requirements for material supply and process tracking. In order to increase inventory visibility and traceability in internal logistics hence improving the efficiency of the supply chain. The traditional method uses manpower which is costly and time-delayed. Previous single-sensor monitoring has its own limitations. For example, weight recognition is hard to determine when multiple type items are mixed. Visual identification is susceptible to problems such as occlusion and difficulty in identifying dense areas and every item needs to be tagged by using Radio-Frequency identification (RFID).

Hence, this research proposes the use of the multi-sensor method, to address three key topics. Firstly, we examine how to gather information from shelves and identify which sensors are suitable for such inventory checks, design a weight and vision detection method in the rack to identify the object in the rack. Secondly, we explore the use of data fusion from two types of sensors to enhance the accuracy of detections, compared to using a single sensor. In this context, we discuss two fusion techniques: one based on conventional methods and deep learning approaches. The concluding segment underscores the utilization of real-time data to refine strategies concerning parts replenishment within the workstation area. This dynamic improvement replenishment, in turn, augments dispatching and delivery performance. This research aims to develop a multi-sensor data fusion method capable of acquiring inventory information in assembly line storage racks and enabling inventory monitoring with high precision and accuracy, leading to improved intra-logistics performance.

This thesis commences with a literature review on the storage devices for materials, and their characteristics, and revisits the applications of both manual and single sensor methods in this domain. A novel dual-sensor design is proposed for the detection of types and quantities of components. Furthermore, the detection capability of a singular sensor was validated, followed by an exploration of its detection capacity under various interference factors. Due to the limitations of a singular sensor, and considering that no prior work has utilised both methods for mixed object inventory detection, a comparison between traditional methods and deep learning was undertaken. This included considerations of comparison of accuracy, implementation difficulty, feature engineering, and prior knowledge. The traditional method uses a mathematical approach to fuse weight and visual data to predict different types and quantities of parts in small containers. In the realm of deep learning, attempts were made with both feature-level fusion and decision-level fusion.

This research also presents a case study on the simulated classification and enumeration of Aircraft General Standard (AGS) parts, predominantly fasteners. These parts are diminutive, expensive, and pose challenges for identification by single-type sensors. Results suggest that the deployed algorithms exhibit efficacy in scenarios with limited variety. Conventional fusion techniques demonstrate an average accuracy of up to 80%, marking an enhancement compared to 73% for single vision detection and 50% for weight-based detection. When employing deep learning-based fusion at the decision-making level, an accuracy of 96% is achieved, and at the feature level, it ascends to 88%. In the dynamic replenishment model based on the Quantity, Recency, and Frequency (QRF) model, the time of out-of-stock events is reduced at the same inventory level; or at the same out-of-stock event time, the required inventory level is lowered.

Table of contents

List of figures

List of tables

List of Equations

Nomenclature

Roman Symbols

μ	Mean
σ	Standard deviation

Acronyms

AGS	Aircraft General Standard parts
ARIMA	Autoregression Integrated Moving Average Model
CDF	Customer Define Freeze
CNN	Convolutional Neural Network
CRM	Customer Relationship Management
DAI-DAO	Data In-Data Out
DAI-FEO	Data In-Feature Out
DEI-DEO	Decision In-Decision Out
DEI-DEO	Decision-In-Decision-Out
DES	Discrete Event Simulation
DL	Deep Learning
FEI-DEO	Feature In-Decision Out
FEI-FEO	Feature In-Feature Out
FIFO	First-In-First Out
FN	False Negative

FP	False Negative
GMMs	Gaussian Mixture Models
IP	Identification Probability
JWOT	Just Walk Out Technology
LIFO	Last-In-First Out
LSTM	Long Short-Term Memory
MIP	Mixed Integer Programming
NCM	Noisy Channel Model
NLP	Natural Language Processing
QRf	Quality, Recency, and Frequency
RFID	Radio-Frequency identification
RMF	Recency, Monetary, and Frequency
RNN	Recurrent Neural Networks
SAM	Segment Anything Model
SSD	Single Shot Multibox Detector
SVM	Support Vector Machines
TN	True Negative
TP	True Positive
YOLO	You Only Look Once

Chapter 1

Introduction

1.1 Background and Motivation

In recent years, digital factories have attracted increasing attention from manufacturing companies at both domestic and international levels. In the domain of traditional, intensive assembly manufacturing, inventory management frequently emerges as a time-consuming and laborious task. With the advent of Industry 4.0 [1, 2], a significant proportion of enterprises have identified the necessity to enhance their existing assembly lines. As a pivotal component of the supply chain, internal logistics streamline the transfer of raw materials, semi-finished goods, and finished products within factory confines. Automated inventory management, an innovative concept underpinned by Industry 4.0, offers manufacturers the means to forestall out-of-stock situations. To curtail both out-of-stock incidents and concomitant inventory expenses, an in-depth understanding of stock quantity and location has become increasingly paramount [3]. The erstwhile evident benefits of manual inventory checks, in terms of cost and efficiency, may soon become nebulous, especially given the non-real-time updating of inventory levels. Conversely, a myriad of rack designs is ubiquitously employed in warehouses, manufacturing plants, retail hubs, and other storage and distribution facilities, lauded for their spatial efficiency and ergonomic configuration.

Using traditional methods, significant manpower is requisitioned to undertake periodic inventory audits or to enumerate each dispatch individually – an endeavour that is both tedious and costly. While singular sensors might be apt for specific tasks, they frequently manifest constraints, notably the ability to detect or categorise solely one type of object. In stark contrast, each item under inspection requires a radio frequency identification (RFID) tag [4]. Applying these to a plethora of non-recyclable components is not only labour-intensive but might also compromise the external aesthetics of the product.

Fortunately, with the advancement of the Internet of Things (IoT) and sensor detection technologies, the potential exists to integrate multi-sensors for inventory detection into assembly line flow racks [5–7].

On the conventional factory floor, a vast array of parts are consumed during the manufacturing assembly process. For optimal internal logistics outcomes, it is imperative that parts are available in the right quantity, at the right location, and at the opportune moment. Especially in the aerospace manufacturing sector, the production environment demands high standards and precision, involving complex components and assembly processes. Typical aerospace manufacturing facilities are equipped with advanced machinery, automated assembly lines, and specialized storage systems to handle a variety of materials and components. To glean insights into the logistics and performance, monitoring such activities is indispensable. Counting and classification emerge as the primary methodologies. In environments where parts are dispatched to the workstation by a 'shop', there's a tendency among operators to overstock, leading to inefficiencies. For such settings, inventory management is crucial to ensure the availability of parts while minimizing waste and costs.

To surmount these challenges, this thesis will introduce detection rooted in multi-sensor fusion, amalgamating the strengths of multiple sensors to bolster predictive efficacy. Noteworthy studies in this domain include unmanned supermarkets such as Amazon's 'Just Walk-Out Technology' [8, 9] and AIM3S cashier-less convenience stores [10]. These pioneering establishments alleviate the perennial issue of queuing at checkouts and furnish more accurate real-time inventory insights, minimising out-of-stock events. For example, common convenience stores encounter shelf stockout rates of around 5-10%, resulting in a possible sales loss of up to 4% [11]. The merits of automated inventory monitoring discerned in retail settings could be transposed to manufacturing contexts: meticulous inventory oversight results in fewer stock shortages, diminished wastage, reduced assembly line interruptions, and enhanced profitability. Unlike retail environments where products are predominantly displayed on shelves, assembly line components are frequently stashed in storage devices, such as Euro containers. These components are often stacked or positioned in rack compartments for spatial and ergonomic reasons. Manufacturing assembly parts, which often bear similar morphological characteristics (e.g., nuts with varying diameters) and denser detection zones (e.g., a substantial number of parts in a box), present distinct challenges. Thus, relying solely on location-based information, as is the norm in unmanned supermarkets, may not be efficacious for this study.

1.2 Aim, objectives and contribution

This research aims to develop a sensor data fusion method that can acquire the inventory level of the storage racks on the assembly line, and achieve inventory monitoring with high precision and accuracy, therefore, improving the internal logistics performance.

1.2.1 Research objectives

1. Developing a novel architecture sensor system to optimise dispatching schedules and liaise with the logistics team, while implementing a sensor network for in-rack inventory monitoring. The architecture of the system includes two main parts:
   (a) Weight-based inventory monitoring
   (b) Vision-based inventory monitoring inside the rack.
2. To devise a mathematical algorithm for the integration of inventory data from two distinct dimensions: weight-based and vision-based, capitalising on prior knowledge. Additionally, to demonstrate the feasibility and ascertain the accuracy of conventional fusion methods for inventory monitoring.
3. To develop two sensor fusion techniques based on deep learning networks and transfer learning, aiming to enhance accuracy in monitoring both the quantity and type of inventory.
4. Development of a Predictive Quantity, Recency, and Frequency (QRF) Analysis Method to Enhance Logistics Performance through Dynamic Replenishment Dispatching policy.

1.3 Research Methodology

Objective 1 (Chapter 3): Designing a sensor system for improved part traceability.

To implement automatic inventory detection, a review of various sensors and their principles is needed. Comparing the placement, type, and number of sensors will guide the inventory detection approach. The chosen sensor should be cost-effective, highly stable, capable of detecting multiple categories, and due to factory edge device computational limitations, not too numerous. Additionally, these sensors should complement each other to capture as much inventory information as possible. Simulations of sensor placements in racks followed by experiments have demonstrated that weight and visual sensors can detect parts on shelves. However, their recognition capability may diminish in the presence of multiple interferences.

To ensure that sensors can effectively monitor inventory levels, it is paramount to first enable the sensors to receive valid data. Initially, a comparison of different sensor characteristics was conducted to determine the appropriate sensor selection. Following this, a comparative analysis was carried out by placing sensors at various positions on the shelves to evaluate the efficacy of the information collected. A setup that aimed at maximising data collection while minimising the number of sensors was chosen. This, in turn, reduced the complexity of the algorithm and computational demands.

Subsequently, IoT technology was employed to transmit the information from the sensors, via the connected edge devices, to computers for further analysis. By simulating potential challenges such as noise/occlusion from various positions and insufficient lighting, a series of experiments were conducted. This was done to validate the feasibility of the proposed sensor types and placement positions in inventory monitoring.

Objective 2 (Chapter 4): Data Analysis and Fusion for Enhanced Detection Accuracy.

Building on the foundation of the first objective, once sensor output data is available, it's possible to analyse and merge this data, leveraging the fused information to boost detection accuracy. Generally, there are two methodologies: traditional data fusion and the recently popular deep learning approach. Either method should enhance accuracy. A purpose-designed dataset is introduced, with different detection categories and quantities selected to ensure comparability. When attempting to traditionally merge weight and visual data, this approach allows for a limited combination and quantity of fusion. A comparative study is conducted to quantify the accuracy enhancement brought about by the fusion, analysing how much data is needed, the capability to overcome obstructions, and the required conditions and prior knowledge. This comparison aims to benchmark the performance of each algorithm in detecting inventory levels.

In order to achieve Objective 2, advanced visual models were selected, employing decision-level fusion. Through the image models, image results are outputted. Weights are determined by integer programming, outputting a set of possible combinations. Based on prior knowledge and boundaries, a large portion of the possibilities are eliminated. Lastly, the remaining options are subjected to probability calculations through the Noise Channel model, thereby facilitating the prediction of object types and quantities.

Objective 3 (Chapter 5): Creation of an Optimised Algorithm for Simplicity and Reduced Dependency on Feature Engineering.

The third objective is to develop an optimised algorithm, reducing the complexity of manual calculations and minimising the need for extensive feature engineering. Deep learning techniques are explored here, to circumvent reliance on manual prior knowledge, extensive observations, and numerous other requirements. This will introduce two levels of fusion: feature level and decision level. Using previously established testing methods, the performance of the model is assessed. This endeavour is geared towards a more efficient large-scale application, both in terms of ease of implementation and future model enhancement. Comparisons are made with previous algorithms to conclude the study.

The design of this objective is to extract features through the latest object detection models followed by data fusion. One advantage is the utilisation of the most advanced detection algorithms without the necessity of expending substantial time and effort on primary feature extraction, with updates to more advanced models as image models iterate.

Within typical feature-level fusion, there generally exist three methods: multiplication, addition, and concatenation. The choice of method depends on the information content of the fused data, and also on empirical attempts to ascertain which method is most effective. Employing multiplication can exponentially increase the computational difficulty of the model, while using the concatenation method can increase the data dimensions, thereby augmenting the model's complexity, which is unfavourable for operation on factory edge devices.

Regarding a data set with a weight scaler, an initial attempt was made to incorporate it into image features using addition. Here, a noise map was employed, as the weight, being a numerical value, may have a relation with all image features. Unlike semantic understanding where all vocabularies have relevance with the image features thus requiring multiplication for vector fusion. The accuracy of the aforementioned can be verified through experimentation, details of which are elaborated in Chapter 5.

Objective 4 (Chapter 6): Enhancing Internal Logistics through Simulation-based Forecasting via a dynamic change in replenishment policy.

This computational model is engineered to simulate and juxtapose the operational dynamics of a traditional supply chain against a real-time data-driven supply chain within the context of aerospace manufacturing. By meticulously analysing the simulated outcomes, this thesis aims to unravel the potential advantages imbued within a flexible, data-centric supply system. Through this investigative lens, it envisages elucidating how such a system can reduce waste, reduce out-of-stock events, and henceforth usher the manufacturing process towards an enhanced level of efficiency and predictability.

Specifically, a model was designed that can process real-time data, establish a database, and then analyse its QRF (Quantity, Reliability, Frequency) score. A higher score indicates a higher consumption trend of a particular component, thereby adjusting the replenishment quantity and replenishment threshold in real time. Conversely, a lower index would indicate a lower consumption trend, which would also lead to adjustments in replenishment levels accordingly.

1.4 Thesis outline

This thesis first investigates and demonstrates the possibility of collecting information from factory flowrack, in the ability to detect internal inventory using sensors, and explores the shortcomings of each sensor, enabling monitoring to be used for sensor fusion tasks to obtain highly accurate inventory data. This includes the use of prior knowledge, observations of internal logistics, and supply characteristics, and deep learning.

- **Chapter1** Introduce the subject and research inquiry, along with outlining the aim and objectives for this research.

- **Chapter2** A review of the literature pertaining to automatic detection techniques in manufacturing and industry, with a specific focus on the manufacturing sector, as well as sensor fusion and other fields where similar use cases are applied.

- **Chapter3** Development of an inventory monitoring system inside the flow rack for detecting and classifying manufacturing parts

- **Chapter4** This one utilises a mathematical approach to sensor fusion based on prior knowledge to fuse information from weight and vision sensors,

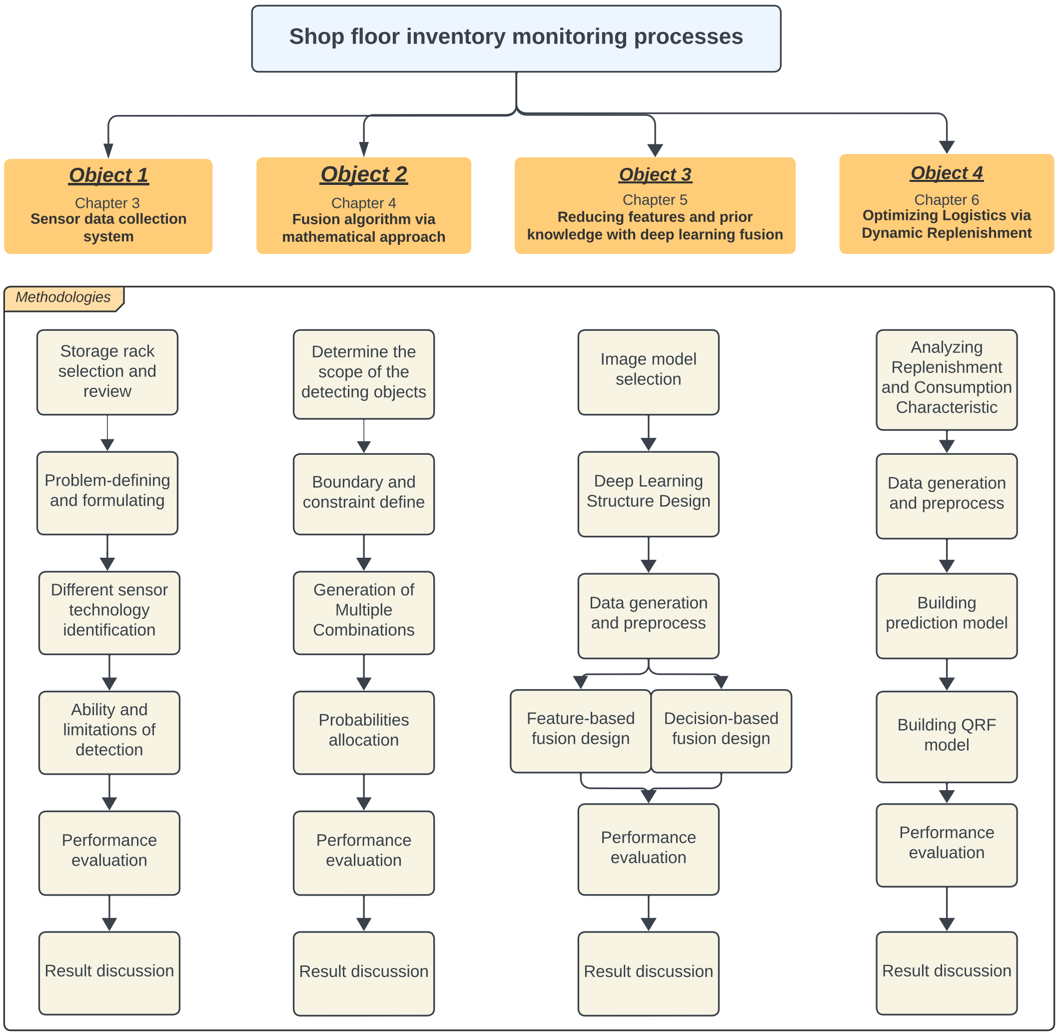


Fig. 1.1 Research methodology

which is based on a noisy channel model, eliminating the need for fusion using devices such as GPUs and enhancing the simplicity and ease of use of fusion on the edge device.

- **Chapter5** This chapter develops a novel FEI-DEO and DEI-DEO-based transfer learning solution that extracts features from state-of-the-art models as a novel way to join weight sources for fusion. There are also attempts to benefit from the adaptive capability of deep learning by introducing probabilistic deep learning through the database to address the challenges in sensor fusion and reducing the reliance on time-consuming industrial observation and prior knowledge.

- **Chapter 6** In this chapter, a consumption assessment methodology based on QRF (Quantity, Reliability, Frequency) has been developed. This approach harnesses real-time data to predict future consumption rates, enabling a swift response to variations in consumption. This reduces the reliance on large amounts of high-quality data upfront, increases flexibility and reduces computational consumption.

- **Chapter7** Discussion and conclusions based on the completed research are presented, alongside an overview of the research's limitations and potential avenues for future work.

1.5 Associated publications

Conference Proceedings

1. Lin, R., Tiwari, A., & Oyekan, J. (2022, September). Counting assistant through multi-sensor fusion for inventory monitoring. In 2022 27th International Conference on Automation and Computing (ICAC) (pp. 1-6). IEEE.

Chapter 2

Literature review

2.1 Introduction

In today's increasingly complex manufacturing landscape, monitoring and managing inventory is paramount to efficient production [12]. The evolution of inventory control and its underlying technologies has become a subject of critical importance, especially in sectors characterized by high-value and intricate products, such as aerospace manufacturing. Yet, despite advancements in other fields, the integration of automated monitoring within manufacturing has lagged [13]. This literature review aims to provide a comprehensive overview of the current research on inventory monitoring techniques within the manufacturing industry [14, 15]. It seeks to understand the placement methods and internal flows of materials, with an emphasis on their application in the high-value aerospace sector.

Additionally, the review aims to explore why automatic monitoring has not been as widely implemented in other industries. The following themes will be addressed: an introduction to the most common methods of material stacking and internal logistics; an examination of prevalent inventory detection methods, including implementation techniques and visual models; an exploration of various sensor detection methods, including both single and multiple types, with a focus on the technology of sensor fusion; and finally, an analysis of the current applications of these technologies within both the manufacturing and commercial sectors. It then reviews predictive methodologies aimed at reducing inventory levels while enhancing replenishment efficiency. By dissecting these aspects, this review aims to contribute valuable insights into the state of the art in inventory monitoring technology, paving the way for future advancements and wider applications.

2.2 Overview of Automated Inventory Detection Systems and Techniques

Inventory management is a multifaceted process encompassing forecasting, requirement determination, data abstraction, and decision-making. A key objective within this domain is to mitigate execution errors such as over-picking, stock-outs, inaccurate inventory counts, and misdelivery of parts, as these shortcomings can deleteriously affect logistics performance, precipitate retail sales losses or interrupt manufacturing operations [16]. Therefore, regular manual stock inspections are a quintessential exercise for companies.

Commonly employed stock counting methodologies are grounded in periodic accounting principles, notably cycle counting and residual balance counting [17]. These methods facilitate the abstraction of essential data concerning the type and quantity of items, and in more expansive terms, the temporal and spatial aspects related to these items. Such data is instrumental not only for adept inventory control and production planning but also for prognosticating inventory levels, thereby aiding informed decision-making. However, the prevalent practice of labour-centric inspection regimes is often neither cost-effective nor time-efficient, leading to delays in critical decision-making processes. Consequently, the proposal to leverage sensor-based detection mechanisms emerges as a salient solution to the aforesaid human resource constraints.

A scrutiny of various inventory detection technologies reveals their utilisation across a diverse array of environments, albeit for analogous purposes. The information gleaned encompasses, among others, quantity, product name, location, and temporal data. Ergo, from a manufacturing standpoint, this discourse delineates the employment of certain detection techniques to obtain the requisite information. Below is a review of typical inventory monitoring methods, elucidating their operational paradigms and contributory significance to the broader inventory management milieu.

2.2.1 Single-sensor based detection

A single type of sensor can detect, classify, and localize objects with some limitations, and these are already widely used in factory automation.

Weight-based detection

Weight-based identification is a cost-effective, mature, and robust technology extensively employed in inventory monitoring, owing to its resistance to external interference such as light and spatial density. This technology primarily utilises load cells, as shown in Figure 2.1, placed on shelves or tables, employing weight variance as a key metric for estimating container availability [18, 19]. However, it falters when an object's weight is a multiple of another.

Piezoresistive textile sensors have been deployed in smart retail settings to ascertain object count, placement, and displacement time on shelves for loads ranging between 0.5 to 1.5 kg [20]. Given the sensor's position beneath the detected object, it can approximate the object's location, facilitating a rudimentary estimation of object type (albeit requiring pre-input information and the assumption that identical object types occupy a given position).

Additionally, capacitive sensing mat technology, when aligned vertically to an object, can infer the object's shape and orientation, thereby deducing its type or quantity [21, 22]. Techniques such as the Four Legs Method have been employed to ascertain tabletop product locations [23], comparing pre-input location data to infer object type and quantity [10]. Nevertheless, neither textile nor capacitive sensing adequately addresses scenarios involving containers, e.g., boxed items, warranting the incorporation of alternative sensing technologies.

Earlier studies have explored capacitive weight-sensing mats, which gauge capacitance alterations between two parallel plates to analyse item presence, absence, or shape, thus identifying and enumerating them [24–26]. Moreover, cost-effective controllers often accommodate weight-based identification systems. Utilising a capacitive mat or textile to capture object imprint shape facilitates object recognition [27, 28]. However, this technology, prone to permanent deformation, continuous impact, and liquid interference, presents longevity challenges.

The application of weight detection in smart homes, particularly in refrigerators through IoT, is well-documented [29]. Nonetheless, it necessitates prior object location prediction. Noteworthy attempts to surmount this include [30], proposing a four-bed setup on support points to simultaneously ascertain object position and quantity.

Comparatively, load cell technology, as depicted in [31–33], often proves more viable than weight-sensing mats. Typically configured in a "Z" shape to apply torque to steel bars, a load cell's four strain gauges measure bending deformation—two gauging pressures and two tension. When integrated within a Wheatstone bridge structure, minute resistance variations in the strain gauges are precisely measured, enabling accurate weight change determination.

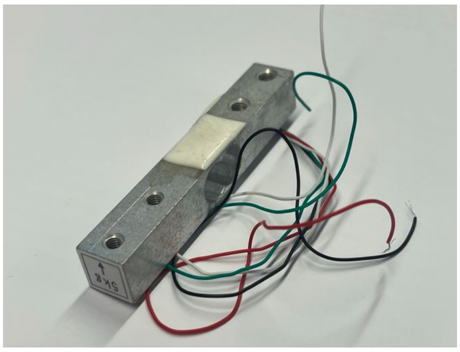


Fig. 2.1 One of many kinds of load cells

Vision-base detection

Vision is a potent sense that enables interaction with the physical world sans direct physical contact. In the recent wave of machine learning advancements, computer vision has emerged as a popular detection modality. Notable applications include autonomous vehicles [34], product quality inspection [35], face recognition [36], and shelf inspection.

Despite significant strides in vision identification, limitations persist, particularly concerning neural networks and light propagation. The constraints are as follows:

- The efficacy of vision technology is compromised by lighting conditions and installation angles, and it is prone to occlusion.
- Blind spots may occur due to the camera's installation angle.
- Image occlusion is exacerbated when numerous boxes are placed on racks.
- The risk of over-training the neural network or the inadequacy of the training set.
- Noise intrusion during signal capture, conversion, transmission, and processing.

Vision-based identification excels when packaging differentiation exists. However, the convenience retail sector often presents visually similar items with distinct content.

Adopting Deep Learning techniques alongside a data-driven methodology has shown promise. Models such as Mask R-CNN [37], Fast R-CNN [38], Faster R-CNN [39], and Yolo [40] have made significant strides. Nonetheless, challenges like the necessity for hyper-parameter tuning and the requirement of thousands of training images under varied conditions persist. Moreover, the similarity in appearance across products (same type with different sizes) may compromise vision accuracy. In light of the rapid evolution in this field, currently, YOLO V8 stands as one of the superior models. However, a method that can swiftly adapt to the state-of-the-art is preferable for our inventory testing pursuits.

The graph-structured visual imitation approach leverages computer vision and machine learning to emulate human finger movements using robotic hands, grounded on visual entities (objects, parts or points) and spatial relationships [41]. This technology requires merely a few object mask examples to ascertain the key points of the human hand, displaying robustness to background clutter and generalising across different object instances [41]. Experimental validations include actions like pushing, stacking, and pouring.

RFID detection

RFID presents an alternative solution worthy of consideration. As noted in the German patent [42], RFID technology has already found applications on flow racks as illustrated in Figure 2.2. While it allows for the detection of box presence, discerning the contents within remains a challenge.

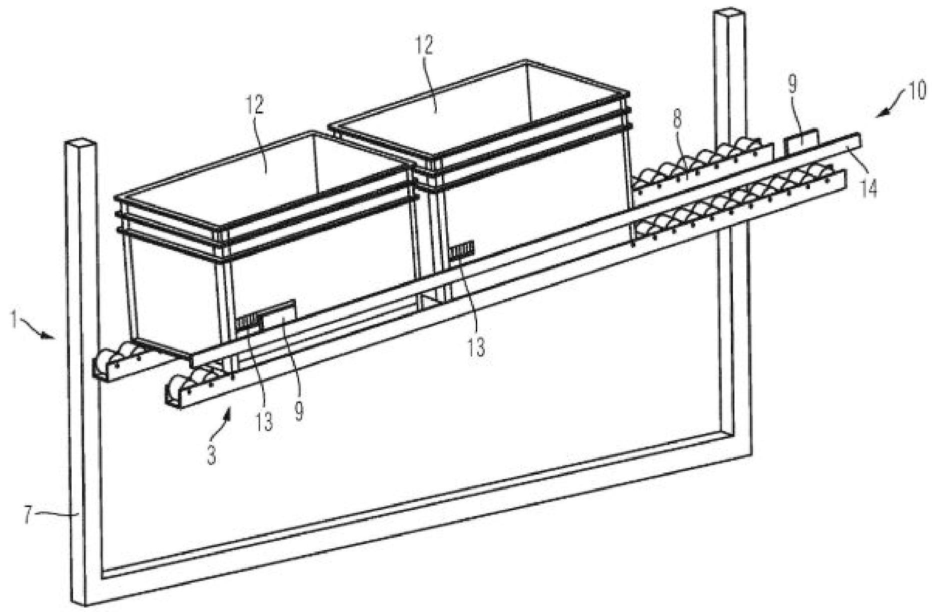


Fig. 2.2 RFID application on flow rack
[42]

Typically, RFID is employed at the case and pallet level rather than at an item level. The high cost and suboptimal read rate hinder its applicability for small items within a factory setting. Furthermore, the abundant presence of metallic objects within the factory can disrupt the RFID signals. During object localisation, the system struggles to ascertain specific locations due to metal-induced reflections and interference from other tags, culminating in positional reading errors [19].

The substantial cost entailed in setting up RFID system infrastructure poses a significant drawback, with several limitations yet to be fully explored. For instance, difficulties arise when reading stacked items [43]. Besides, labelling each item induces additional labour costs. In one instance, RFID was utilised to locate goods within a building [44]. While RFID proves advantageous with larger products, it encounters disturbances when dealing with stacked small items. Existing technology is striving to mitigate cross-reading issues and reduce the number of required readers by integrating antennas [45]. However, this solution compromises flexibility as it necessitates fixed antennas.

The research by [46] highlights that, although RFID could resolve supermarket inventory challenges, maximising its efficiency generally necessitates optimising shelf space. Given these considerations, the scope and feasibility of RFID application, particularly within an aerospace manufacturing factory setting, are brought into question. The practicality of labelling hundreds or thousands of parts appears to be significantly diminished.

Barcode-Based Inventory Detection Systems

Upon its invention, barcode technology quickly garnered widespread adoption, demonstrating effective identification capabilities across various complex environments. Employing barcode technology for inventory management and monitoring has emerged as a common practice, presenting a range of benefits and drawbacks. On the positive side, barcode systems notably enhance efficiency, reduce errors, save time, improve data reliability, prove cost-effective over time, allow smooth integration with other IT systems and enable inventory monitoring [32].

Conversely, there are significant shortcomings linked with the implementation of barcode systems. A primary concern is the system's rigidity, which might pose challenges when dealing with non-standard or specialised inventory items, thus showing a lack of adaptability to specific needs. Another major issue is the heavy reliance on barcode labels. The accuracy and efficiency of inventory data hinge on the preservation and readability of the barcodes, rendering the system susceptible to disruptions if labels are damaged or lost.

Furthermore, the necessity to affix labels to all items represents another limitation. This mandate introduces an additional procedural step, potentially heightening operational complexities, particularly in scenarios where items are excessively small or have surfaces unsuitable for barcode labelling. In such cases, the labelling process can become cumbersome, especially on manufacturing parts, and it will also damage the property of the item.

Inductive Detection

Inductive sensors operate by detecting changes in inductance caused by the presence of an object, making them suitable for counting metallic objects. These sensors can measure the position, speed, and distance of metal objects, and are commonly used in applications such as monitoring vehicle traffic [47]. However, they are susceptible to electromagnetic interference, which may not be suitable in large and can affect their performance.

Ultrasonic sensors and and Laser scanners

Ultrasonic technology calculates distance by measuring the time it takes for ultrasonic waves to travel to a surface and reflect back [48]. Inventory systems utilising this technology are particularly effective for detecting smooth liquid surfaces, as ultrasonic sensors are one-dimensional and can only estimate the level from a single point on the surface.

In contrast, laser scanners use laser beams to scan objects and calculate the reflected light signals, allowing for precise measurement of an object's size and shape. This makes them suitable for scenarios requiring high precision and the ability to handle complex shapes [49]. They are ideal for bulk solid inventory management and industries where maintaining a distance from the inventory is essential for hygiene and preventing damage to products or sensors. Laser scanners are also more effective at scanning larger areas compared to deploying numerous scanners in smaller regions, making them particularly useful in food processing, cement, and chemical processing industries. However, laser scanners are generally more expensive than cameras and generate larger amounts of three-dimensional data, requiring significant computational and storage resources for processing.

2.2.2 Multi-sensor based detection

Based on previous reviews of single-sensor technology, object recognition methods can be primarily divided into four categories: size and weight measurement, electromagnetic characteristics, optical scanning, and sound analysis.

None of these solutions can fully tackle the autonomous inventory monitoring issue on their own, due to the sensory modality limitations and environmental complexity. There is still a lot of work to be done to identify objects by weight and appearance characteristics, and most of the cases focus on only one sensing method. Although the previous technology is effective, it has the disadvantages of low accuracy, long delay in obtaining results and cost prohibitive. This automatic monitoring can reduce user input errors, which are common in other popular applications such as expensive tools failure detection or defects of the products. Therefore multi-sensor-based approaches are starting to be invented.

Several studies have ventured to improve food classification through multi-sensor integration, targeting applications such as grocery re-identification or expired item detection within refrigerators [50–52]. Even within restricted and highly controlled environments, these studies underscore the utility of multi-modal sensing for item detection. A notable challenge in sensor fusion is the accurate estimation of the joint conditional probability. Addressing this, Ruiz et al. [53] proposed a higher relevance weighting for product prediction, highlighting that the performance of weight sensors remains stable amidst occlusion. They advocated for additive modalities prediction over multiplication to mitigate the adversities posed by light and angle on vision, thereby averting a null detection probability outcome.

Despite the proprietary nature of industry applications like Amazon Go [54], there has been a dearth of similar techniques concerning sensor fusion via deep learning for inventory acquisition. Nonetheless, other domains such as workpiece repair prediction and process inspection exhibit potential in this realm.

Within a confined scope, certain investigations have delved into multi-sensor implementations in supermarkets, exploring the autonomous shopping experience through the integration of weight and visual sensors. However, the placement of weight sensors at the base of objects, facilitating the capture of object information via location assistance [10, 55], presented challenges. Techniques like capacitive sensing and quadrilateral localisation have been employed to ascertain object positioning [30, 56], yet when applied to flow racks housing multiple boxes within each column, these techniques disrupt the flow and substantially increase the demand for weight sensors, consequently quadrupling the information flow.

This surge imposes a significant demand on invaluable industrial computational resources. Additionally, challenges arise from the intricacies of estimating joint conditional probability within this sensor fusion paradigm.

2.3 Review of Image Detection Model

Image detection, a pivotal subset of computer vision, has markedly evolved with the advent of deep learning algorithms [57]. The principal objective of image detection is to identify and locate objects within images. Over the years, a myriad of models has been developed, each exhibiting distinct advantages and limitations. This section endeavours to elucidate some of the seminal models in image detection, offering a succinct critique of their performance and applicability in diverse settings. By comparing them, we selected 3 cases that are suitable for us, preferably real-time, lightweight, and less computationally intensive models.

Convolutions neural network (CNN)

Historically pivotal in object detection, Convolutional Neural Networks (CNNs) are now witnessing a diminished prevalence due to several emerging trends [58]. A performance plateau has been reached with CNNs in certain domains, prompting a turn towards alternative models that promise enhanced performance or better computational efficiency, crucial in real-world applications with resource constraints. The latency exhibited by CNNs in scenarios demanding real-time object detection is also a factor driving this shift [59]. Furthermore, the rise of multi-modal learning models, which efficiently handle various types of data in a unified manner, has overshadowed the somewhat uni-modal nature of conventional CNNs [60]. Additionally, newer models showcasing better generalisation across diverse tasks and datasets, evolving research focus towards better interpretability and robustness, along with a trend towards hybrid models integrating CNNs with other architectures, like Transformers, represent the contemporary shifts in the object detection realm, steering away from traditional CNN-centric approaches.

You Only Look Once (YOLO) Model

You Only Look Once (YOLO) is a real-time object detection model introduced in 2015 by Joseph Redmon et al [40]. Since its inception, the model has undergone rapid evolution, now being in its eighth iteration.

YOLO, along with the Single Shot Multibox Detector (SSD), has significantly advanced the field by offering real-time detection capabilities. YOLO is particularly acclaimed for its swiftness, achieving object detection in a single forward pass through the network. However, it tends to trade off accuracy for speed, especially in scenarios involving small or overlapping objects [61].

Despite this, YOLO maintains a respectable level of accuracy while keeping the model size compact. It can be trained on a single GPU, making it accessible to a wide range of developers. The affordability of deploying it on edge hardware or cloud platforms further adds to its appeal.

In terms of object detection within an image, YOLO generates thousands of candidate anchor boxes throughout the image to localize potential objects. For each anchor box, an offset is predicted to form a candidate box approximating the true object's shape and size. A loss function is then computed relative to the ground truth, measuring the error between the predicted candidate box and the actual object location. Concurrently, a probability is assessed, quantifying the likelihood of overlap between the offset candidate box and a genuine object within the image. If this probability exceeds a threshold of 0.5, it is factored into the loss function, thereby actively influencing the error measurement. Through this iterative process of rewarding and penalising predicted boxes, the model is steered towards accurately localising true objects within the image.

EfficientNet

EfficientNet, a robust neural network architecture, was proposed by Google researchers Mingxing Tan and Quoc V. Le [62]. This network distinguishes itself by achieving a high level of accuracy while maintaining a reduced count of parameters and computational demand, accomplished through a balanced scaling of network depth, width, and resolution.

A more recent advancement, the EfficientDet model, presents a harmonious blend of accuracy, efficiency, and scalability, making it a suitable candidate for various real-world applications. Its innovative architecture, encompassing a bi-directional feature pyramid network and a weighted feature fusion mechanism, bestows superior detection performance compared to its predecessors.

2.4 Inventory Management in Aerospace Manufacturing

2.4.1 Characteristics of Inventory in Aerospace Manufacturing

Inventory management in aerospace manufacturing is characterized by unique challenges due to the industry's stringent requirements and complex processes. This section outlines the critical aspects involved in managing inventory within this sector.

Precision and Reliability

Aerospace components necessitate highly precise manufacturing and assembly processes. Inventory systems must meticulously track each part's location, condition, and quantity, ensuring compliance with stringent quality standards. Components are often tagged with unique identifiers such as batch or serial numbers, facilitating quick identification and recall if needed [63]. These unique identifiers allow real-time control over the product value chain throughout the production process. This capability enables dynamic and efficient production path planning, and immediate retrieval of component information such as origin, storage status, and location, enhancing production efficiency and quality management.

Unlike the high-volume, repetitive operations in the automotive industry, aerospace manufacturing involves lower production volumes and more varied processes. This often necessitates manual operations, introducing potential variability. However, the industry is gradually integrating automation to support these manual processes. For example, while hand layup is labour-intensive and suited for complex, low-volume parts, automated processes like filament winding and tape lamination ensure higher production rates and consistent quality[64].

High Value

Due to complex designs, advanced materials, and high precision requirements, aerospace components are typically high-value items. The integration of smart manufacturing and Industry 4.0 technologies, such as robotics and artificial intelligence, is increasingly prevalent in aerospace production to enhance efficiency and precision. These technologies optimise production processes, reduce errors, and improve overall component quality [65]. Incorporating cyber-physical systems, additive manufacturing, and product traceability enhances flexibility and responsiveness, crucial for managing small-batch and customized production demands.

Complexity and Diversity

Aircraft are assembled from thousands of unique components, each with specific requirements. Managing such extensive inventory necessitates meticulous tracking and forecasting to ensure availability and regulatory compliance. For instance, the diverse nature of maintenance, repair, and overhaul (MRO) operations means inventories must include a wide range of parts, from frequently used items to those needed for rare or unscheduled repairs [66].

To address these challenges, the aerospace industry is increasingly adopting advanced technologies. AI is used for predictive analytics, RFID for real-time inventory tracking and blockchain for ensuring supply chain transparency and traceability. These technologies optimise inventory levels, reduce costs, and enhance supply chain responsiveness [66].

The quality of aircraft components must strictly adhere to aeronautical standards, as even minor defects can significantly impact safety. Using incorrect parts, even with a small error rate but large volumes, can cause significant defects, production delays, and damage to the company's reputation. By integrating Industry 4.0 technologies, manufacturers can better meet the demand for personalized and customized products while achieving higher production efficiency and flexibility. These technologies not only help manufacturers reduce costs but also improve their competitiveness in a demanding market.

2.4.2 Unidirectional flow of parts

Unlike conventional production lines, aerospace manufacturing demands higher quality standards. While there have been attempts to implement automated production such as [67] and material allocation technologies [68], the industry still primarily relies on labour-intensive processes. This distinction arises from the fact that aerospace production varies in scale and customization compared to industries such as automotive manufacturing. Aircraft often require a diverse array of specialized components, leading to the need for skilled workers to manually assemble a large number of parts. These factors, compounded by the unique nature of the products, result in considerable waste of materials in internal logistics.

One quintessential example of this is when pickers estimate the quantity of parts needed, they often select more than required, leading to an oversupply.

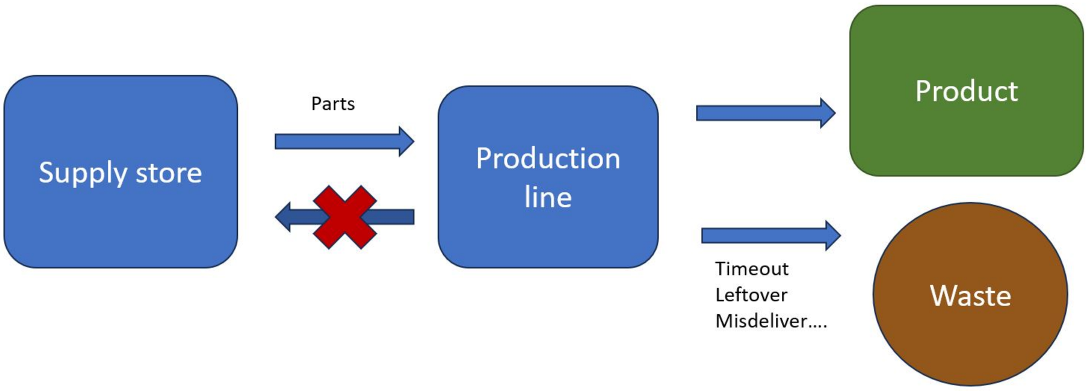


Fig. 2.3 Unidirectional flow of parts on the manufacturing floor

This excess has serious consequences; once placed on the production line, these parts cannot be directly returned to the warehouse due to quality and safety concerns. Unused parts must then be scrapped, resulting in considerable waste. This problem is compounded by the fact that some parts have a shelf life and therefore cannot be reused. Moreover, mistakes in the picking process further contribute to the waste, as wrongly picked parts are rendered unusable.

2.4.3 Storage System in Manufacturing2.

Flow rack

This research is predicated on the monitoring of inventory within gravity flow racks. A gravity flow rack comprises an inclined roller frame, wherein stock is loaded from the rear and withdrawn from the front, as demonstrated in Figure 2.6. This system operates on a First-In-First-Out (FIFO) basis, which enhances inventory density and ergonomics relative to traditional shelving systems due to its depth and the perpetual foreground positioning of inventory for facile retrieval [69]. Given that gravity propels the inventory movement, there is no requisite for additional electricity or other forms of power.

While various types of racking systems may be present within an assembly line—such as the Last-In-First-Out (LIFO) push back rack depicted in Figure 2.4—this research exclusively focuses on the gravity flow rack and its operating principle, as illustrated in Figure 2.5 and 2.6.

Flow rack roller

In the flow rack depicted in Figure 2.7, three types of rollers are identified. The selection of rollers is contingent upon the size and the intended use case of the

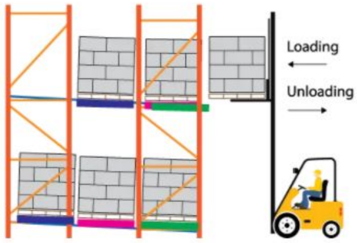


Fig. 2.4 Push back rack [70]

Fig. 2.5 Gravity rack in warehouse [71]

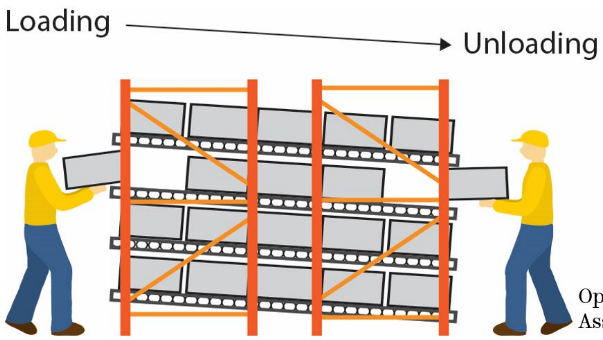


Fig. 2.6 FIFO flow rack

boxes. The full-bed polycarbonate wheels are adept at accommodating varying box widths due to the absence of fixed lanes. Conversely, full-width steel rollers and polycarbonate wheels are more suited for storing boxes of consistent widths as they possess fixed lanes.

Additional accessories for the flow rack are available to enhance its functionality. For instance, guides can be employed to fashion tilt trays at the discharge end, thereby facilitating easier access to the box or tote bag tops for piece-picking tasks.

The two-column downward roller, illustrated in Figure 2.8, is noteworthy due to its provision of a more flexible loading scheme.

Fig. 2.7 Examination of various types of flow Rollers
[72]

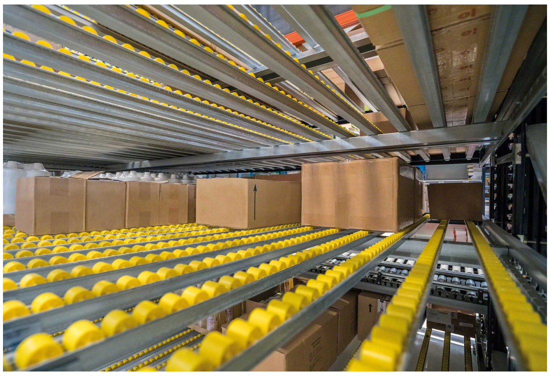


Fig. 2.8 The flow rack roller
[73]

Euro container (flow rack box)

The box housed within the flow rack is designated for accommodating Euro containers (Figure 2.9). These containers, characterized by their solid bases and sides, are optimally suited for small load carriers. Additionally, they are equipped with a top handle, facilitating easy transportation. A noteworthy feature is their capability for vertical stacking across multiple levels. Supplementary components such as top covers and dividers may be employed, the utility of which is governed by the specific requirements at hand. The dimensions of these containers are inherently dependent on the spacing between the rollers of the flow rack, with a recommended maximum weight of 12 kg per box, epitomizing typical industrial applications. It is pertinent to mention that the scope of this research does not extend to the stacking or crossover of boxes within the compartments.

Fig. 2.9 Box applied in flow rack

2.5 Multi-Sensor fusion technique

Sensor fusion is the process of combining data from multiple sensors to provide a more comprehensive view of a system or environment. This can be achieved through techniques such as Kalman filtering [74], Bayesian estimation [75], and decision-making algorithms [76]. Sensor fusion is commonly used in robotics, autonomous vehicles, and other applications where a single sensor may not provide enough information to make decisions. By combining multiple sources of information, the resulting data can be more accurate and reliable, leading to improved decision-making and overall system performance [77].

2.5.1 Classification of Sensors Based on Inter-relationships

Data fusion is an interdisciplinary domain encompassing various fields, making it challenging to delineate a precise and rigid classification. The methods and techniques utilised can be segregated based on the subsequent criteria:

- Considering the relations among the input data sources as delineated by Durrant-Whyte [78] shown in Figure 2.11. These relations are categorized as:

  1. **Complementary**: Obtaining more complete global information by merging different parts of a scene from multiple input sources, e.g., two cameras with distinct fields of view in visual sensor networks.
  2. **Redundant**: Incrementing confidence by fusing overlapping or identical data from multiple sources, e.g., overlapped areas in visual sensor networks.
  3. **Cooperative**: Generating more complex new information by combining provided data, e.g., multi-modal (text and video) data fusion.

- Based on the input and output data source and their inherent nature as delineated by Dasarathy [79] The classifications are illustrated in Figure 2.10, broadly encompassing the following categories:

    1. **Data In-Data Out (DAI-DAO):** This basic fusion method takes in raw data and outputs raw data too, usually with increased reliability. It's an immediate process post data collection, primarily utilising signal and image processing algorithms.
    2. **Data In-Feature Out (DAI-FEO):** At this stage, raw data is used to extract descriptive features of an observed entity.
    3. **Feature In-Feature Out (FEI-FEO):** Here, features are both input and output, aiding in refining or generating new features. It's also referred to as feature fusion or information fusion.
    4. **Feature In-Decision Out (FEI-DEO):** Features are processed to produce decisions. Most classification systems working with sensor data operate at this level. This architecture is prevalently employed in sensor fusion for monitoring purposes, with an extensive body of literature within this domain substantiating its widespread utilisation [80, 81].
    5. **Decision In-Decision Out (DEI-DEO):** Known as decision fusion, it refines or forms new decisions by combining initial decisions [82].

- Classification based on the architecture style [77]:

    1. **Centralized** a central processor hosts the fusion node, receiving and processing data from all input sources. This setup optimally centralises all fusion processes, given proper data alignment, association, and negligible data transfer time. However, real systems may suffer due to time delays in data transfer and the extensive bandwidth required for raw data transmission, especially in visual sensor networks. Within the context of a flow rack scale, centralising the connection of all sensors to a single host emerges as a more judicious option. Establishing a dedicated processing host for each sensor load cell is deemed impractical due to the associated logistical and economic implications.
    2. **Decentralized** This setup forms a network of nodes, each with processing capability, but no singular fusion point. Every node independently fuses its local data with data from its peers. Despite facilitating autonomous data fusion, this architecture may face scalability issues as

node numbers increase due to the necessity of peer-to-peer communication.

3. **Distributed** Each node processes its data before sending it to a fusion node. The fusion node then aggregates this pre-processed information from all nodes, creating a unified view. With options for single or multiple fusion nodes, this architecture allows for initial local processing, reducing the burden on the central fusion process. Upon an examination of performance, considering one rack as an entity, it appears that this architecture is better suited to the shop floor scale but it will also bring computational cost.

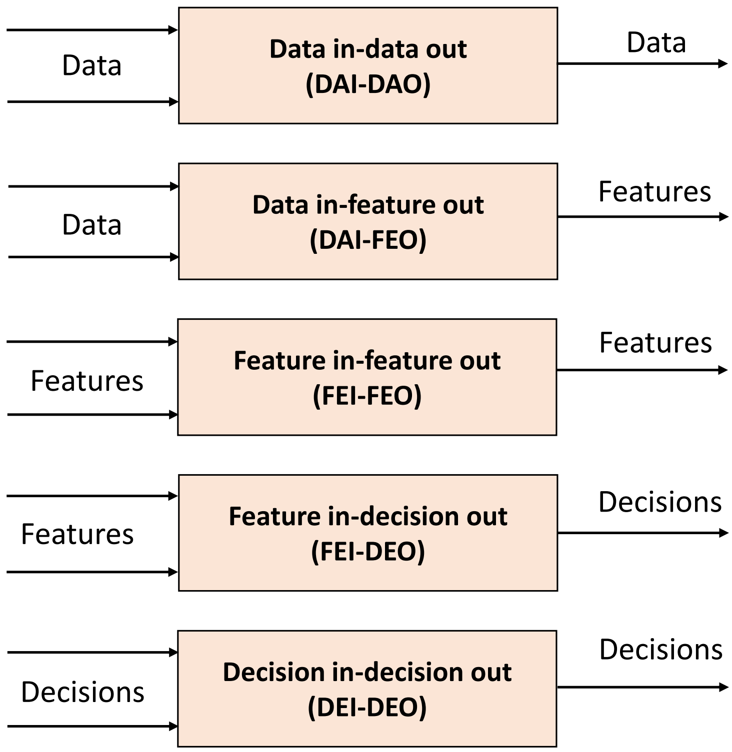


Fig. 2.10 Dasarathy's classification

Whyte's classification is based on distinguishing between various data sources based on the relations among them. This classification system can be a useful way of structuring and understanding the data gathered in qualitative or mixed-methods research. Here's a more detailed breakdown of the key aspects and a review of Whyte's classification system.

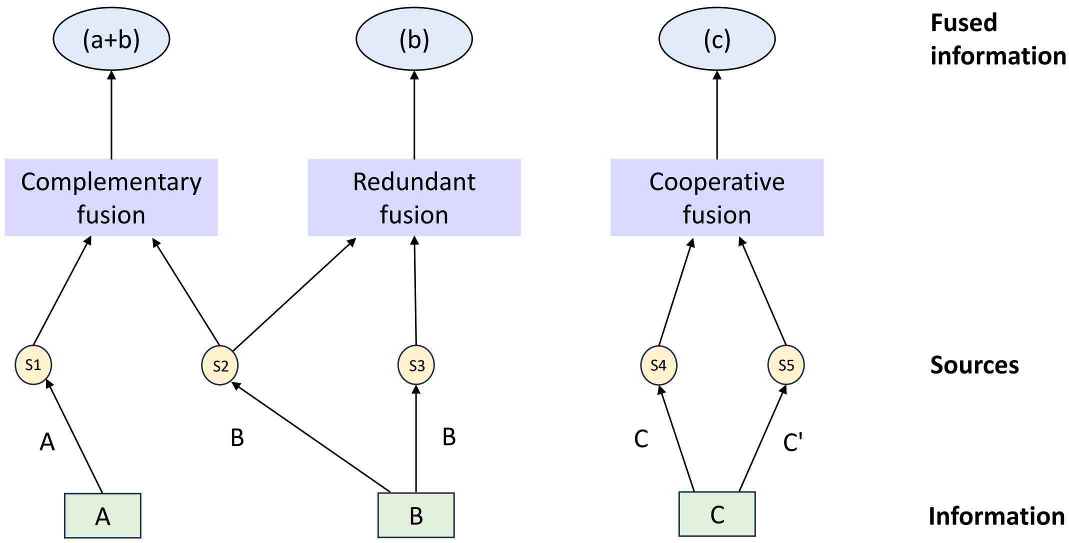


Fig. 2.11 Classification by Whyte is based on the relationships among the sensor sources

2.5.2 Traditional Mathematical Fusion

Kalman Filter The Kalman filter, proposed by Rudolph in 1960 [83], is a prominent sensor fusion algorithm. It fuses data from different sources to estimate unknown values and is predominantly used in navigation and control technology. In dynamic systems, where uncertainty prevails, Kalman filters offer superior accuracy over predictions made using a single measurement method. Their recurrent nature enables the anticipation of present and future conditions, given the previously recognised position and speed of an object, like a car [84]. Despite their usefulness, they come with inherent limitations:

- Assumption of Linearity and Normality: Assumes linearity in the modelled process and normal distribution for errors. Real-world systems often violate these assumptions.
- Initialization Sensitivity: Initial state estimates greatly influence the filter's accuracy and convergence speed.
- Stationarity Assumption: Assumes time-invariant statistical properties which are unsuitable for inventory time series.
- Handling of Outliers: Vulnerability to outliers due to Gaussian distribution assumption.
- Limited Multi-modal Capability: Cannot represent multi-modal probability distributions.

Bayesian Inference and Dempster-Shafer theory

The domain of statistical inference has witnessed the emergence and evolution of various frameworks aimed at handling uncertainty and facilitating decision-making processes in a plethora of fields. Among these, Bayesian Inference and the Dempster-Shafer Theory stand as significant paradigms with distinct mathematical foundations and practical implementations. Bayesian Inference and Dempster-Shafer Theory provide two compelling, albeit different, methodologies for reasoning under uncertainty.

Bayesian Inference is grounded in the theory of probability. It epitomises a quantitative approach to expressing uncertainty by employing probability distributions [85]. The key to Bayesian Inference lies in Bayes' theorem, which in its essence, provides a mechanism for updating beliefs in light of new evidence [86].

Mathematically, Bayes' theorem is expressed as:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

where $P(A|B)$ is the posterior probability of event A given event B, $P(B|A)$ is the likelihood of event B given event A, $P(A)$ is the prior probability of event A, and $P(B)$ is the marginal likelihood of event B.

Contrasting Bayesian Inference, Dempster-Shafer Theory does not necessitate probabilistic assumptions, thus providing a more flexible framework for managing uncertainty. It is built upon the notion of belief functions, which allow for the representation and combination of partial and conflicting evidence.

The mathematical foundation of Dempster-Shafer Theory is rooted in two primary elements: the belief function Bel(.) and the plausibility function Pl(.).

These are defined as:

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B)$$

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\neg A)$$

where m is a basic probability assignment function defined on the power set of the frame of discernment.

The comparative analysis between Bayesian Inference and Dempster-Shafer Theory unveils a divergence in the handling of uncertainty and the associated mathematical frameworks. While Bayesian Inference employs a probabilistic approach, the Dempster-Shafer Theory offers a belief-function based methodology, which could potentially render it more suitable in scenarios laden with ambiguous or incomplete information [87].

2.5.3 Machine Learning based Fusion

As the volume of data burgeons, the utilisation of deep learning-based fusion methodologies has witnessed a surge in recent years. They offer numerous advantages such as [88–90]:

- **Handling High-Dimensional Data**: Uncovering patterns in high-dimensional data is a forte of machine learning algorithms.
- **Feature Learning**: The ability to autonomously learn and extract pertinent features from raw sensor data obviates the necessity for manual feature engineering, which is often laborious and demands expert knowledge.
- **Robustness to Noise**: Proffering robustness to noise and outliers, probabilistic models are adept at managing real-world, disordered sensor data.
- **Scalability**: Machine learning algorithms exhibit proficiency in efficiently managing copious amounts of data, a critical attribute when assimilating data from a multitude of sensors.
- **Predictive Capabilities**: Predictive maintenance or forecasting environmental conditions are among the plethora of applications benefitting from the predictive prowess of machine learning algorithms.
- **Adaptability**: The amenability of machine learning models to retraining and adaptation with new data accrual is pivotal for accommodating alterations in the environment or sensor configuration.

The fusion algorithm can process myriad sensor data channels in unison, yielding classification results grounded on image recognition. For instance, employing convolution neural network-based algorithms, a robot discerns traffic signs from a distance, while in healthcare applications, a sensor fusion system employing a convolutional neural network (CNN) monitors transitional movements [91].

Decision Trees

Decision trees engender decisions based on multiple sensor outputs by evaluating the relationships between sensors and the desired output [92]. Their simplicity and real-time execution efficiency are commendable, albeit their propensity for overfitting and struggle with complex relationships pose challenges.

Random Forests

An extension of decision trees, random forests amalgamate outputs from multiple decision trees for sensor fusion. They exhibit lesser susceptibility to overfitting, can manage complex relationships, and are adept at real-time execution [93]. However, their sensitivity to hyperparameter selection and interpretability challenges due to the multiplicity of trees used are notable drawbacks.

Support Vector Machines (SVMs)

SVMs categorise multiple sensor outputs into distinct classes aiding system state decisions. They excel with high-dimensional data and non-linear relationships but may falter with extensive feature sets and real-time computational efficiency. Their performance is contingent upon apt kernel function selection, and they are susceptible to overfitting [94].

Gaussian Mixture Models (GMMs)

GMMs model the probability distributions of multiple sensor outputs, amalgamating them for a more precise system state estimate. Their flexibility and easy interpretability are advantageous, yet they demand ample data, are computationally intensive in real-time, and are sensitive to hyperparameter and initialization parameter selection [95].

2.6 Predictive replenishment process

In light of evolving market demands, practitioners and scholars alike have explored various models for adept inventory management, especially focusing on predictive analytics within warehouse and retail sectors. A paramount goal within this sphere is to accurately forecast the consumption of each item on the shelf, facilitating timely replenishments and reducing carrying costs. Despite the advancements, predicting the demand for smaller objects remains a relatively under-explored domain. Predominantly, enterprises rely on elementary methods such as spreadsheet analysis for inventory prediction, which frequently fall short in terms of precision and foresight. The primary methodologies employed for projecting future sales or consumption by analysing historical data encompass the following:

2.6.1 Time series analysis

AR, Autoregressive Model

The application of autoregressive models in inventory prediction manifests as a nuanced, yet potent, method to forecast inventory levels. By leveraging the temporal dependencies inherent in inventory data, autoregressive models meticulously capture the dynamics of inventory changes over time. The model's emphasis on analyzing past observations to predict future inventory levels offers a robust mechanism to accommodate the temporal fluctuations in demand and supply, thereby aiding in efficient inventory management [96]. In the case of inventory prediction in manufacturing, using an AR model may not be the most ideal choice because AR models predict future consumption based on linear relationships and past usage data, but in our scenario, usage has significant non-linear characteristics (e.g., the presence of outliers). However, to create a model based on time series analysis to solve this problem.

ARMA, Autoregressive Moving Average Model

A robust predictive model necessitates a thorough analysis of historical data encompassing various parameters like operation time, daily or weekly stock levels, and promotional or seasonal demand among others. For instance, retail inventory prediction often deals with products having differing sales patterns—some are sold consistently throughout the year while others exhibit seasonal demand.

The foundation of a robust predictive model is laid on a meticulous analysis of historical data, embracing a variety of parameters such as operational time, daily or weekly stock levels, and the impact of promotional or seasonal demand, to name a few [97]. For instance, retail inventory prediction often wrestles with products manifesting disparate sales patterns—some maintain a steady sales trajectory throughout the annum, while others are beholden to seasonal demand fluctuations.

2.6.2 Quantity, Recency, and Frequency(QRF) Model

Quantitative models are invaluable tools for understanding customer behaviours and forecasting various aspects within a business realm, especially in inventory management and demand forecasting. The models' QRF [98], and Recency, Monetary, and Frequency (RMF) [99] have found their relevance not only in marketing analysis but also in predicting stock consumption to a certain extent.

The QRF model, which focuses on the Quantity, Frequency, and Recency of purchases, can be applied in inventory consumption forecasting. The 'Quantity' aspect can help businesses understand the volume of products customers tend to purchase over a specific period. The 'Frequency' aspect helps in understanding how often these purchases are made, and the 'Recency' aspect can give insights into the latest buying trends. By analyzing these three dimensions, businesses can anticipate demand and thus plan their inventory levels accordingly to minimize holding costs and avoid out-of-stock or overstocking situations.

On the other hand, the RMF model, with its focus on Recency of purchase, Monetary value, and Frequency of purchases, can also offer valuable insights for inventory consumption forecasting. The 'Monetary' aspect could help in understanding the financial value of the inventory consumed, while the 'Frequency' and 'Recency' aspects can also provide insights into buying patterns [100].

These models can help in identifying customer segments that contribute significantly to inventory consumption. High-frequency buyers or high monetary value buyers can significantly impact the rate at which inventory is consumed. Understanding the buying behavior of such segments can aid in better demand forecasting and in turn, better inventory management. This may also be reflected in trends in the consumption of manufacturing spare parts.

- **Quantity (Q)**: The total amount purchased by the customer, typically the total purchase amount over a specific period.
- **Frequency (F)**: The frequency of the customer's purchases, typically the number of purchases made over a specific period.
- **Recency (R)**: The time since the last purchase, typically measured in days from the present.

The basic expression for Customer Value can be written as:

$$\text{QRF score} = a \cdot Q + b \cdot F + c \cdot R \tag{2.1}$$

where a, b, and c are weight parameters that can be adjusted based on different business objectives and strategic needs.

Drawing parallels with the paradigm of Customer Relationship Management, the QRF mode, analogous to the Recency, Monetary, and Frequency (RMF) model delineated furnishes invaluable insights for inventory management. Both frequency and recency depend on time series analysis. However, for assembly factories, Monetary is not a major factor. The consumption of parts is influenced not only by the demands of Recency and Frequency but is more affected by noise.

2.6.3 Discrete Event Simulation

Discrete Event Simulation (DES) serves as a potent tool for modelling the inventory system, thereby prognosticating its behaviour and performance under a miscellany of conditions. DES facilitates the examination of complex interactions within inventory systems, making it an indispensable instrument for predictive replenishment strategy development [101].

2.6.4 Commonly Used Deep Learning Based Inventory Prediction

Deep learning (DL) has been increasingly adopted for inventory and sales forecasting due to its capability to handle vast amounts of data and unveil complex relationships therein [102, 103]. The method can usually handle non-linear and complex data patterns without the need for a priori statistical assumptions.

Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs) are a class of neural networks designed to handle sequential data by incorporating a form of memory through cyclic connections [104]. Unlike traditional feedforward neural networks such as CNN mentioned earlier, RNNs maintain a hidden state that captures information from previous time steps, allowing them to process sequences of variable length.

The fundamental operation of an RNN can be expressed mathematically as follows:

$$h_t = \sigma(W_h \cdot h_{t-1} + W_x \cdot x_t + b), \tag{2.2}$$

where h_t represents the hidden state at time step t, h_{t-1} is the hidden state from the previous time step, x_t is the input at time step t, W_h and W_x are weight matrices, b is the bias, and σ is the activation function (such as tanh or ReLU) [105].

Despite their capabilities, RNNs suffer from significant limitations, particularly the problems of vanishing and exploding gradients. These issues arise during the backpropagation process when dealing with long sequences, making it difficult to learn long-term dependencies [106].

To address these limitations, Long Short-Term Memory (LSTM) networks were introduced by Hochreiter and Schmidhube [107]. LSTMs incorporate a memory cell and gating mechanisms to better manage the flow of information through the network, thereby mitigating the vanishing gradient problem.

An LSTM cell consists of three primary gates: the forget gate, the input gate, and the output gate, each controlling different aspects of the cell's operation. The key equations governing an LSTM cell are as follows:

Forget Gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{2.3}$$

Input Gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{2.4}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{2.5}$$

Cell State Update:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \tag{2.6}$$

Output Gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{2.7}$$

$$h_t = o_t \cdot \tanh(C_t), \tag{2.8}$$

where σ is the sigmoid activation function, W_f, W_i, W_C, W_o are weight matrices, b_f, b_i, b_C, b_o are biases, h_t is the hidden state, and C_t is the cell state [108].

RNNs and their variant LSTM networks are pivotal in time-series prediction tasks. For instance, [109] utilised LSTM networks to predict the consumption of materials in inventory management. Their results show that the average inventory demand prediction accuracy of the deep inventory management (DIM) method exceeds 80%, can reduce inventory costs by approximately 25%, and can quickly detect abnormal inventory behaviours. Although the paper mentions the ability to quickly detect abnormal inventory behaviours, it does not discuss how to handle these anomalies. In practical applications, effectively managing sudden anomalies and making real-time adjustments is a question that requires further exploration.

Gated Recurrent Units (GRUs), introduced by [110], are a type of RNN designed to address the vanishing gradient problem inherent in standard RNNs. GRUs achieve this by incorporating gating mechanisms, specifically the update gate and the reset gate, which regulate the flow of information.

The architecture of GRUs is simpler compared to LSTMs, as GRUs combine the functionalities of the forget and input gates of LSTMs into a single update gate. This reduction in complexity results in fewer parameters and faster training times, making GRUs computationally more efficient.

Studies such as [111] have demonstrated the effectiveness of GRUs in capturing short-term dependencies in material consumption data. However, GRUs may struggle with long-term dependencies due to their relatively simpler gating mechanisms compared to LSTMs.

GRUs present a compelling alternative to LSTMs for predicting material consumption due to their computational efficiency and ability to capture short-term dependencies. However, their shortcomings in handling long-term dependencies and high variability in data highlight the need for careful consideration when choosing long-term prediction.

Transformer model

The Transformer model, introduced by Vaswani et al. in 2017 [112], is based on a self-attention mechanism that allows the model to weigh the importance of different elements in the input sequence dynamically. This capability is particularly advantageous for time series forecasting, where capturing long-range dependencies and intricate patterns is essential. Unlike previously mentioned RNNs and LSTMs, Transformers do not rely on sequential data processing, enabling parallel computation and thus, greater efficiency in handling large datasets.

In the context of material consumption prediction, several studies have explored the effectiveness of Transformer models. For instance, Li et al. [113] demonstrated that the self-attention mechanism of the Transformer could effectively capture the complex temporal dynamics of material usage data, outperforming traditional RNN-based approaches. Their experiments on a large manufacturing dataset showed improvements in forecasting accuracy and computational efficiency.

Furthermore, Wu et al. [114] proposed a hybrid model combining Transformer networks with Convolutional Neural Networks (CNNs) to enhance feature extraction and temporal pattern recognition. This hybrid approach leverages the strengths of both models, where CNNs capture local patterns and spatial features, and Transformers handle the temporal dependencies. The study reported gains in predictive performance and robustness, particularly in scenarios with highly volatile material consumption patterns.

Another contribution is the work by Zhou et al. [115], who introduced the Informer, an extension of the Transformer model designed for long-sequence time-series forecasting. The Informer employs a ProbSparse self-attention mechanism to reduce the computational complexity and memory usage, making it suited for real-time material consumption prediction in industrial settings.

Nevertheless, despite their strengths in capturing long-range dependencies and handling large datasets, Transformers have certain limitations when it comes to predicting material usage in industrial applications. One key limitation is their requirement for substantial computational resources due to the self-attention mechanism, which scales quadratically with the input sequence length. This can be particularly challenging when dealing with high-resolution time series data or very long sequences typical in industrial contexts. Furthermore, transformers can struggle with extrapolation outside the range of the training data, making them less effective when future material usage patterns deviate significantly from historical data [113]. Another challenge is the need for extensive hyperparameter tuning to achieve optimal performance, which can be resource-intensive and time-consuming. Consequently, these limitations can hinder the practical implementation of transformers in material usage prediction for industrial settings.

Challenges of Deep Learning for inventory prediction

Although there are benefits from above, several challenges that practitioners may encounter when leveraging deep learning for forecasting purposes in inventory include:

- **Data Dependency:** Deep learning models heavily rely on large volumes of high-quality data for training. The predictive performance of these models can significantly degrade if they are trained on insufficient or poor-quality data [116].
- **Lack of Interpretability:** Often dubbed as "black-box" models, deep learning models suffer from a lack of interpretability. It becomes exceedingly difficult to understand or explain the decision-making process of these models, which could be a critical shortfall in business scenarios that require transparency and accountability [117]. This issue is particularly important in sectors where decision rationale must be clear for compliance and trust reasons [118].
- **Computational Resource Intensiveness:** The training and execution of deep learning models require substantial computational resources. This could translate to high computational costs, especially for large datasets and complex models.

- **Model Complexity:** Designing and optimising deep learning models necessitate a high level of expertise and a considerable time investment. The tuning process of these models can be tedious and time-consuming[119].
- **Risk of Overfitting:** Deep learning models can easily fall into the trap of overfitting, especially in scenarios where the available data is limited or the data distribution is skewed.
- **Real-time Performance Challenges:** In scenarios necessitating real-time forecasting, the computational complexity of deep learning models may result in response latency, failing to meet the real-time requirements.

2.7 Case Studies in Manufacturing and Retail Industry

2.7.1 M4 Intelligent Manufacturing

A typical example is the assembly of high-value, low-volume production because they are not sufficient to set up automatic assembly lines. Flexible manufacturing can save working space and assemble more products in one place. Such as the assembly of aviation components. A similar case is Meggitt Modular Modifiable Manufacturing (M^4), where they assemble diverse aerospace components in one workstation [120]. A Multi-sensor system has the ability to detect the location and quantity of objects. Information obtained from the sensor system will feedback to the operator. They used lasers and video guiding the operator that the racks were selected in order and placed on the table, and machine vision and AR technology would continue to be integrated in the future [121]. On the other hand, [120] also proposed an architecture that uses multiple sensors to fuse information. The information flow of each component can be interoperable to achieve scheduling optimisation, KPI visibility and traceability, thereby reducing costs.

2.7.2 Just Walk-Out Technology

Just walk out technology (JWOT) is a new technology that is for cashless checkout in order to enhance customer experience queue time. This technology is developed and implemented by Amazon.com [122]. When they pick up goods, three technologies consisting of computer vision, sensor fusion, and deep learning algorithms will work together to increase the reliability and accuracy of the results. RFID was used in early versions [123]. But more is that it uses a large number of cameras to scan the code as soon as the customer enters the store to confirm the identity. Tracking with a large number of cameras installed on the ceiling [29]. These cameras not only detect products and their location but also track shoppers. However, JWOT is very expensive and it is not real-time, customers need to wait 10 minutes to a few hours according to the customer report in order to receive them receive the bill, this waiting time is unacceptable in the assembling line. In addition, the system can only accommodate up to 20 people in store [29].

2.8 Research Gap

Drawing upon the literature reviewed and the challenges discussed in the preceding section regarding sensor application for inventory monitoring on flow racks, this section delineates the identified research gaps:

- **Research gap 1** (Objective 1): In industrial production facilities, the systematic arrangement of materials is a commonplace practice; however, extracting data from these configurations proves challenging, particularly when diverse types and quantities of components are compactly organised within a confined area. Traditionally, single sensors have found utility across various detection domains, albeit with a limitation to recognising only one type of object, demonstrating a heightened sensitivity to complex environmental conditions. While there have been sporadic forays into multi-sensor based methodologies, they typically necessitate auxiliary inputs such as positional data during the detection phase. However, acquiring the positional data of an individual component on a flow rack can be an arduous task, which complicates the distinction among multiple types of objects. As such, the first objective posits the novel application of multi-sensor technology on a flow rack to accrue valid data, which could later be harnessed for fusion purposes, obviating the need for positional information.
- **Research gap 2** (Objective 2): The exploration of fusion methods, tailored for multi-sensors — notably those concerning weight and vision sensors — demands a meticulous investigation, particularly focusing on the fusion of information post data acquisition. Within the academic discourse, two predominant approaches have emerged: the conventional fusion methods and those grounded in deep learning algorithms. The literature, as of yet, does not present a universally accepted method to fuse specific data, thus rendering it a challenge to ascertain an optimal solution amidst the prevalent methods. Traditional fusion methods possess certain merits, such

as the minimal requirement for extensive data sets, a robust interpretative framework, and a lesser reliance on intensive computational resources for effective detection. Object 2 endeavours to scrutinise these claims, evaluating the efficacy and performance of traditional fusion methods in comparison to their deep learning counterparts. This research aims to contribute a nuanced understanding towards determining a more effective fusion method.

- **Research gap 3** (Objective 3): In traditional fusion methodologies, a certain degree of accuracy can be attained; however, deep learning algorithms stand out as significantly advantageous and have demonstrated success across various domains. Despite this advancement, the public availability of deep learning-based inventory detection remains limited. One of the defining attributes of this approach is its independence from domain-specific knowledge, which consequently lowers the potential introduction of high levels of uncertainty. Deep learning possesses the ability to autonomously learn and output features, thereby expanding the scope for exploring various fusion strategies including feature-level fusion and decision-level fusion. In addition, the comparative performance of deep learning models versus traditional models warrants a thorough investigation. Particularly, an analytical comparison at the decision level and feature level may provide invaluable insights.

- **Research gap 4** (Objective 4): Examining the potential for predicting future consumption inventory levels using real-time inventory data presents an uncharted territory. Models hinged on deep learning necessitate high-quality data and substantial computational prowess. On the contrary, empirical methodologies often yield inaccurate forecasts, thus failing to establish precise replenishment strategies grounded in real-time inventory assessments.

  While literature exists on forecasting inventory to address issues of stock depletion, they invariably require data to some extent and find it challenging to rapidly respond to current consumption fluctuations. Regarding the anticipation by retailers of the sales potential of certain goods, the concepts of QRF and QFM have been utilised to evaluate a customer's potential purchase intent. However, the application of these concepts has not yet been explored in terms of forecasting factory parts consumption. Objective 4 aims to bridge this gap while ensuring the model's interpretability.

2.9 Summary

In this review, initially, the working principles of various types of racks and their accessories are introduced. Following that, the discourse shifts to an examination of technologies potentially applicable to our inventory monitoring efforts, encompassing both single-sensor and multi-sensor applications. Subsequently, methods of data fusion and prediction are scrutinised, illustrating how different data sources can be coalesced to form actionable insights. Furthermore, recent research examples from two distinct industrial applications are cited to offer a real-world perspective on the discussed theories and technologies. Lastly, the review delves into the realm of demand forecasting, examining how it informs parts consumption, contrasting it with traditional replenishment strategies to potentially unveil a pathway for enhanced inventory management.

Chapter 3

Sensor network equipped flow rack

3.1 Introduction

In workstations along the assembly line, the continual engagement in repetitive production activities necessitates a timely supply of raw materials by the logistics team, as the inventory of parts steadily decreases. In traditional factory settings, reliance is placed on operators to report inventory levels and replenish stock based on past experience. Out-of-stock scenarios during periods of high production activity can cause delays across the entire assembly line, culminating in unnecessary cost expenditures and potential delays in delivery. Automated inventory management solutions present a viable resolution to this issue.

In Chapter 2, we summarised the monitoring of inventory and solutions offered by existing research, noting that a majority of the research employs a single sensor. The prevailing research endeavours are still centred on transitioning from traditional methods to digital ones to address real-time and flexible requirements.

The objective of utilising data fusion in multi-sensor environments is to reduce the probability of detection error and enhance reliability by harnessing data from various distributed sources. This is evidenced in industrial cases such as unmanned supermarkets [10] and Just Walk Out technology (JWOT) [8].

According to [53], the integration of weight sensor data and vision information remains a challenging feat, having been executed only a handful of times. The current state of the art involves monitoring shelf inventory using vision and weight sensors, achieving a remarkable accuracy rate of 93.2%. This method was simulated in a convenience store and relies on sensor information and location-based knowledge for product identification. In addition, within the rack setting, objects are not placed vertically on the sensor. The force exerted on the sensor is at an inclined angle to the horizontal, thereby potentially reducing the accuracy of detection.

On the other hand, the stacking of inventory on the ground is impractical, thus manufacturing operations typically employ shelves or racks to store inventory, aiming to optimise storage space utilisation, alongside improving safety and ergonomics. Raw materials, being constantly consumed in manufacturing processes, are better suited to flow rack storage solutions. This is due to their ability to utilise gravity for loading, organising, and retrieving parts to the forefront without necessitating additional manpower or electricity, as compared to other rack types.

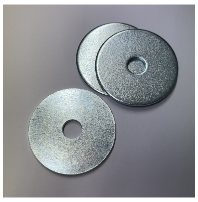
3.2 Methodology

3.2.1 Selection of the object to be detected

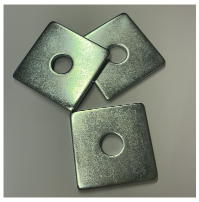
This research also provides a case study, simulated classification and counting of Aircraft General Standard (AGS) parts, AGS parts encompass a category of components chiefly characterized as fasteners, crucial to the aerospace industry. These diminutive elements, albeit small in size, bear a significant expense due to their specialized nature, conforming to stringent quality and performance benchmarks requisite in aviation. The identification and classification of AGS parts pose a unique challenge, particularly to single sensor systems, given the subtle variations in their physical attributes, such as being lightweight, having similar shapes, colours, and sizes, and existing in large quantities. Their critical function of ensuring structural integrity and the need for accurate identification and management underscore the importance of a sophisticated tracking and monitoring system within the aerospace sector. In this thesis, Round washers, Flat square washers, External coach screws, and Studding connectors are utilised to simulate AGS, as they embody all the characteristics of AGS. Figure 3.1 illustrates their appearance and its detail below:

- Object 1 is a round washer with a weight of 38g;
- Object 2 comprises flat square washers with a weight of 60g;
- Object 3 consists of external coach screws with a weight of 14g;
- Object 4 is represented by studding connectors with a weight of 41g.

These simulated AGS parts were then tested to determine the classification results of single sensors and fused model results under different types and quantities.



(a) Round washers



(b) Flat square washers

(c) External coach screw

(d) Studding connectors

Fig. 3.1 Different AGS parts were used for the experiment

3.2.2 A Selection of sensors

Automatic inventory monitoring primarily involves obtaining information from one or more sensors. High-quality visual sensors can provide both raw data and contextual information, being readily amenable to analysis and utilisation. However, they are susceptible to environmental influences and obstructions from objects. In contrast, weight sensors require only physical contact to receive information, are not affected by object obstructions, and are less likely to be influenced by environmental conditions, yet they can only provide limited information. Therefore, experiments are needed to ascertain the impact of the environment on the material inventory detection capabilities and whether the sensors are capable of extracting valid data under such conditions.

Load cells, often utilised as weight sensors, are principally grounded in strain gauge technology. They are pivotal in a plethora of domains encompassing industrial, commercial, and scientific research sectors, mainly to accomplish precise weight measurement and monitoring. Among the critical features for employing load cells as weight sensors are:

- **Accuracy**: Load cells are renowned for their high precision in measuring load or weight. This precision is indispensable in myriad applications such as commercial weighing, industrial manufacturing, and experimental research, where accurate weight data are paramount.

- **Linear Outpu**: Load cells provide a linear output signal when within the elastic limit, signifying a directly proportional relationship between the output signal and the measured weight. This linearity facilitates straightforward readings or uncomplicated calculations to ascertain the actual load values.

- **Stability and Reliability**: Characterized by robust design and stable operation, load cells can reliably function over extended periods across a variety of manufacturing environments.

- **Cost-effectiveness**: The relative affordability of load cells, coupled with their low long-term maintenance costs.

- **Digitization**: In the wake of technological advancement, contemporary load cells can embody digitization and intelligence features. In this research, the load cells are connected with Raspberry Pi 3 model B+. When connected with the Pi, they require minimal computation overhead to enable automatic monitoring.

The camera has the following characteristics as a non-contact detection method:

- **Recognition Ability**: Visual sensors can recognize various attributes of items such as type, colour, and shape, facilitating more complex inventory management tasks like categorization and sorting. Conversely, weight sensors only provide weight information and cannot support multi-dimensional item recognition and categorization.

- **Non-Contact Measurement**: Visual sensors are non-contact, posing no risk of damaging the items being monitored, making them particularly suitable for fragile or high-value items. In contrast, weight sensors require direct contact with the items, which may potentially affect the integrity of the items.

- **Compatibility**: Visual sensors are easy to integrate with other intelligent systems. In most cases, the camera is provided with an interface that can be easily connected and implemented to the computer or edge device, like Figure 3.2.

Nonetheless, visual sensors have their constraints, such as being influenced by lighting conditions, occlusions, and image processing algorithms, and may demand heightened computational resources and advanced algorithmic backing. In summary, both visual sensors and weight sensors hold their respective advantages in the realm of inventory monitoring and management. Given that visual sensors can be employed more effectively for multi-classification, harnessing the merits of both presents becomes promising.

Fig. 3.2 Raspberry Pi camera module

3.2.3 Coordinate of the rack

To discuss the use of sensors to monitor inventory, the location of the inventory on the shelf is important because each sensor focuses on a different area. Hence, use the common positioning coordinate system to represent the position of inventory as shown in Figure 3.3.

There are three rows and four columns for this rack. The corresponding row 3 and column 2 are a compartment. We first construct each compartment to place 3 boxes and count the first box from the exit.

3.2.4 Boundary and constrain

It is not easy to perceive densely packed parts, because they may not be effectively covered by the camera when placed inside the rack, or the weight is too small to coincide with the weight of the box. A single sensor has its own limitations, as clearly seen in Figure 3.7. Hence, inspections based on multiple information sources may have the potential to be improved to an acceptable level. The first step of the project is to determine the scope of the research and verify that the sensors at the designated location can provide valid inventory information. The following assumptions have been enumerated, and determining the scope of the research is essential:

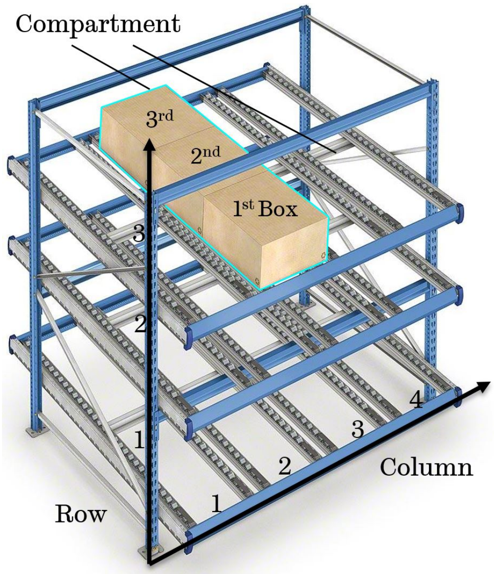


Fig. 3.3 Flow rack and its coordinate

- At the initial stage, each box will contain objects of the same type; however, as the detection difficulty escalates, they will house a diverse range of objects.
- A rack may accommodate a varying number of compartments per row.
- Ideally, a single flow rack should consist of 3-5 rows.
- A maximum of three boxes is allowed per compartment.
- The properties of the objects within each box will remain unaltered throughout the course of the research.
- Items can only traverse from the entrance to the exit. Any item picked beyond demand will be categorised as used.
- No unidentified items are to be placed on the rack.

3.2.5 A sensor network equipped flow rack

This research mainly considers the use of two types of sensors simultaneously, weight and vision sensors. The weight sensor is used as a mature and low-cost technology. The weight sensor is installed in each column on each level of the rack or installed in four corners of each column. On the other hand, The camera can extract various features, such as the colour, shape, texture and text of the object. Cameras are positioned at the rack exits to monitor the stock conditions within the outermost boxes in each compartment.

In order to monitor the inventory level, it is important to ensure that the sensor can obtain effective information at a specific location. In the previous chapter, common weight-based identification sensors are generally installed at the bottom of the items, but this does not apply to flow racks, since sensors must be installed in each compartment at the point of support. The load cell equipped at the exit of the rack measuring compartment weight as shown in Figure 3.6, the entire compartment only requires one sensor would be a better solution.

The comparison of different positions of the camera is shown in Table 3.1. Position of the camera 1, 2, and 3 are at the entrance of the rack, at the exit of the rack and above the rack at the exit side. Compare the above advantages and disadvantages choose positions 1 and 3 can have the most information. However, at this stage, we are still in the process of verifying the sensors' feasibility; more importantly, our primary focus is on the inventory within the rack rather than any other information. Installing at positions 1 and 3 would require additional technologies such as human pose estimation to determine which compartment the parts are located in, which is beyond the scope of this research. Therefore, the decision was made to install the visual sensor at position 2 for experimental purposes.

Table 3.1 Comparison of camera installation at different positions

Position	Strengths	Weaknesses
1	Objects of different types can be monitored Replenishment shift time can be recorded	Picture quality may be poor due to object movement Current inventory levels are unavailable
2	Can monitor the inventory of the box at the exit	Requires full box at entrance Objects in one compartment must be identical Can only observe the box at the exit
3	Monitors pick-up activities by sensing human pose Requires fewer cameras	Multiple items picked up at once may cause occlusion Accuracy may be lower than installation at exit

Another solution is to get more information from the flow rack as well as from the shop floor, cameras can be installed around the rack exit. The visual system at the exit is tasked with recording the items selected by the operator, thereby deducing what remains. However, a downside exists; it is unable to ascertain the state of incoming goods. In more complex scenarios, as depicted in Figure 3.4, discerning the contents of the top box on the shelves becomes challenging.

To overcome this issue. Figure 3.4 is a side view of the rack. If the cameras on both sides are intended to be installed outside the rack. The advantage of not being installed inside the rack (above the box) is that it requires fewer cameras to get enough information, but the replenishment may require logistics team cooperation, that is, after the recognition is successful, it is put into the box and

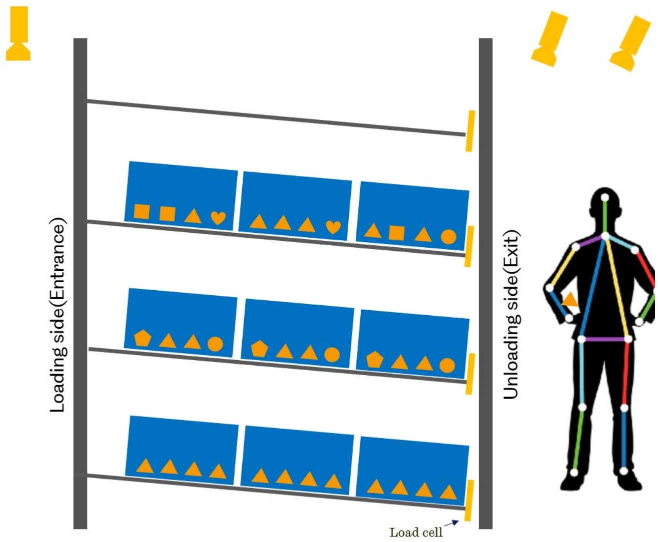


Fig. 3.4 Side view of rack sensor installation location

enters the rack. Once at the exit, the operator's posture will be sensed first, and then the bounding box estimation is set according to the position of the wrist to analyze the items taken.

3.3 Experiment Result and discussions

After creating two methods to collect information, a simulated experiment was planned, that is, the camera was placed at position 2 (Figure 3.5), as we are still in the process of verifying the sensors' feasibility, we shall commence with the implementation and experiment of simpler cases. Figure 3.6 and Figure 3.5 both will illustrate the sensor locations.

The experiment should be separated into two phases. Firstly, verify the feasibility of monitoring inventory in a single type of sensor (weight and vision). Secondly, measuring the accuracy subsequent to the fusion of two types of sensor signals, will be discussed in the following chapter.

3.3.1 Weight-base identification

Throughout the experiment, a 2 kg load cell will be utilised due to the lightweight nature of the object being detected. All load cells employed will be of the model mentioned above, with the experimental variables for the sensors being meticulously controlled, as illustrated in Table 3.2.

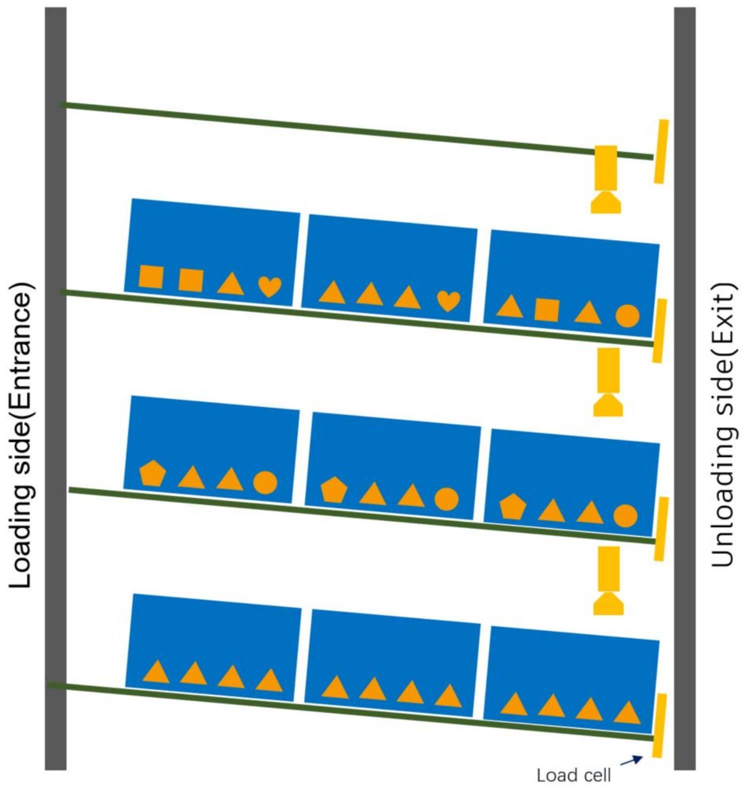


Fig. 3.5 Side view of the rack and location of the vision sensor

Fig. 3.6 Stereoscopic view of rack weight sensor installation location

Fig. 3.7 Two examples of image occlusion

Table 3.2 Experiment setup and response variables

Category	Description
Hardware	• Raspberry Pi 3 Model B+ • Load Cell Amplifier HX711 • DIYmalls Load Cell 2 kg • LarmTek W3 web camera
Setup	• Calibrated with a 1kg weight every time the Pi is restarted. • Sampling rate for both weight and vision is 2Hz. • Power is steadily connected to the Pi (AC) and HX711. • Deep transfer learning models are used simultaneously for image identification. • The camera position is fixed.
Response Variables	• Item class • Number of items

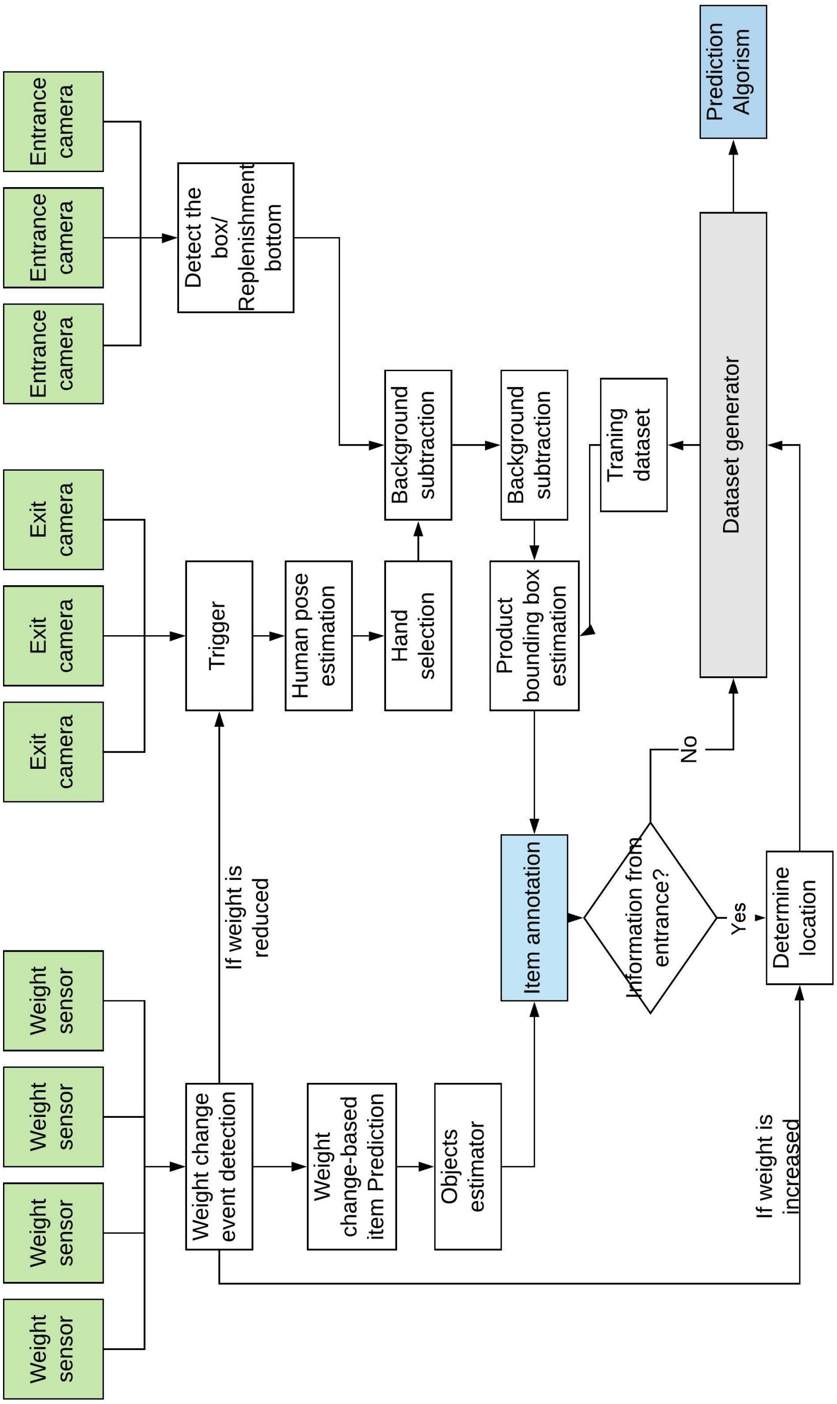


Fig. 3.8 Overview of the inventory monitoring system

Since the object is flow lowered from the top, the load cell is installed at the bottom of each column to calculate the weight through the angle and friction coefficient.

A load cell sensor, amplifier, Raspberry Pi, fixtures, and flow rack have been employed for this research. The load cell is installed at the exit of the

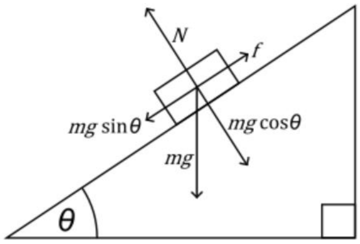


Fig. 3.9 Force applied on load cell

rack to measure gravitational force. The angle between the compartment and the horizontal line is assessed prior to the experiment and remains unchanged throughout the process.

Incline forces physics will be used in this experiment. Since the box carries the object is dropped from the top. The inclined force received by the load cell without friction could represent as:

$$F = mg\sin\theta$$

If consider surface friction:

$$F = mg\sin\theta - f$$

Where F is force received by the load cell without friction

F_f is the force received by the load cell with friction

f is the friction force causing the roller, calculated by $f = \mu mg\cos\theta$

μ is the dimensionless rolling resistance coefficient or the coefficient of rolling friction (CRF) for roller rolling.

There are a number of objectives for this experiment:

1. Check the accuracy on different levels of weight of load cell

   This ensures that the load cell is reliable initially and conducive to future error calculation. The weight is applied vertically, with a force ranging from 10%, 20%, up to the 100% maximum force limit of the load cell. This procedure is repeated thrice.

2. Determine the roller friction coefficient on the rack The roller friction coefficient can be calculated:

$$\mu = \frac{f}{mg\cos\theta} \tag{3.1}$$

Table 3.3 Accuracy on different levels of weight on the load cell

Percentage (%)	Weight (g)	Average Error (g)
0	0	0
10	200	1
20	400	-1
30	600	1
40	800	0
50	1000	0
60	1200	1
70	1400	1
80	1600	2

and

$$F - F_f = f \tag{3.2}$$

In such a system, pulleys or ball tracks are commonly employed to facilitate the smooth sliding of items. The rolling resistance may vary significantly due to different designs, materials, and operational environments. The rolling resistance of the pulleys is influenced by a multitude of factors including the material of the pulleys or balls, the weight of the load, the angle of the slide, the material and the surface roughness of the slide path. In some instances, manufacturers or suppliers might provide certain information regarding the rolling resistance; however, this information often lacks representativeness across a broad range of scenarios. Therefore, to simplify calculations and provide a reliable basis for the experiment, a typical value of **0.01** is adopted, as it falls within the common range of 0.005 to 0.02, balancing practicality and accuracy [124].

3. Determine the accuracy of monitoring inventory under the same parts (type) in a compartment

   In this part, no external force was applied except for the decreased weight of the box. At the beginning, place one full box in one compartment, and continue to pick up parts from the full box until the last empty box is left. The speed taken (each box can be taken 5 times) should be equivalent to the weight of each object, at the same time record each F_f and the F_f of the empty box, and compare the calculated change of F_f to judge the accuracy of detecting objects.

   The results, conducted twice and illustrated by Table 3.4 demonstrate that the weight sensor is essentially capable of fully identifying items of a singular

Table 3.4 Results of weight detection for various objects

Weight(g) \ Quantity / Item	5	4	3	2	1	0
Objects1	191	152	113	75	38	0
Objects2	300	241	180	120	60	0
Objects3	70	56	42	28	14	0
Objects4	205	163	122	82	42	0

category, corroborating our initial hypothesis. The colour highlights in the table indicate the boundary ranges for the overlap condition, demonstrating that weight sensing is significantly limited by the overlap as the case becomes more complicated. Even with only 20 possible scenarios, there are already 8 data points that cannot be effectively identified. Although load cells are economical, less susceptible to external interference and less computationally intensive, a single cell alone is not sufficient to recognise mixed cases.

3.3.2 Vision-base identification

The vision-based identification experiment was conducted under controlled factory conditions to ensure consistent lighting and minimal external interference. The sensor setup included a white background, LED lighting, and a web camera positioned directly above the material flow rack. The camera was set to default mode with the flash turned off to maintain uniform image capture conditions.

The experiment aimed to test the vision sensor's ability to identify objects accurately in three different storage scenarios within a flow rack shown in Figure 3.5:

- Bottom Row: All boxes contain identical objects.
- Middle Row: Each box contains a combination of the same objects.
- Top Row: Each box contains a different combination of objects.

These scenarios represent varying levels of complexity in object identification, from single-type uniformity to mixed and varied item configurations. To clearly define the complexity levels, we categorized the scenarios as follows (shown in 3.10):

- Single Type and Quantity Classification: Homogeneous items with a fixed quantity.

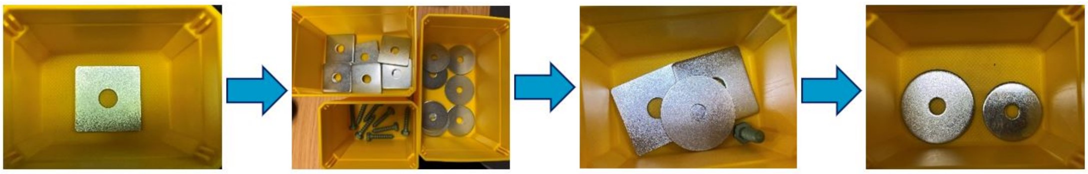


Fig. 3.10 Complexity level from left to right (a)Single type and quantity-classification (b)Multiple items but the same type (c)Different types and quantity (d)Different type(similar item like M10 and M12 round plate washers) and quantity

- Multiple Items of the Same Type: Different quantities but the same type of items.
- Different Types and Quantities: Heterogeneous items with varying quantities.
- Similar Items (e.g., M10 and M12 round plate washers) and Quantities: Items that are similar but not identical, posing a higher challenge for identification.

Image recognition requires data for training, and the content of the database needs to be roughly consistent with the proposed environment. A more comprehensive general database is ImageNet, which has a total of 14,197,122 pictures. From this database, images of four specific object categories were extracted through a search process. However, some photos did not meet the requirements of the experiment, so they were manually screened. Consequently, a new dataset was created for training purposes. Each object category contains 100 pictures. The dataset was then divided into a training set and a validation set in a 4:1 ratio, resulting in a total of approximately 125 pictures per object category.

The EfficientNet training model was employed to process the images. This model is known for its balance of accuracy and computational efficiency, making it suitable for factory conditions where quick and reliable identification is crucial. The experiment focused on assessing the model's performance under different environmental conditions.

During the experiment, the following conditions were maintained:

- Camera Positioning: The camera was positioned directly above the items to capture top-down images.

Conditions such as lighting and partial display were recorded in a binary format (1 for true, 0 for false), and the results showing the object's number were

Table 3.5 Vision experiment result

Object	Lightning	Partial display	Result	Correct
1	1	0	1	Y
1	1	0	1	Y
1	0	0	4	F
1	0	0	1	Y
1	1	1	1	Y
2	1	0	2	Y
2	1	0	2	Y
2	0	0	2	Y
2	0	0	1	F
2	1	1	2	Y
3	1	0	3	Y
3	1	0	3	Y
3	0	0	3	Y
3	0	0	3	Y
3	1	1	1	F
4	1	0	4	Y
4	1	0	3	F
4	0	0	4	Y
4	0	0	4	Y
4	1	1	1	F

identified and compiled in Table 3.5. The model achieved an overall accuracy of approximately 75%, with most errors occurring under inadequate lighting and occlusion scenarios.

The experimental results indicate that under optimal conditions, the vision sensor can achieve satisfactory accuracy. However, performance drops significantly when lighting is poor or objects are occluded. This highlights the impact of changing conditions on visual recognition in a factory environment, but these unfavourable factors for visual sensing are unlikely to have an impact on weight sensors.

While a single type of sensor is insufficient for complex identification tasks, a multi-sensor approach combined with advanced machine learning models like EfficientNet can significantly improve accuracy and reliability in real-world applications. The detailed parameters and methods for the multi-sensor setup will be further discussed in Chapters 4 and 5.

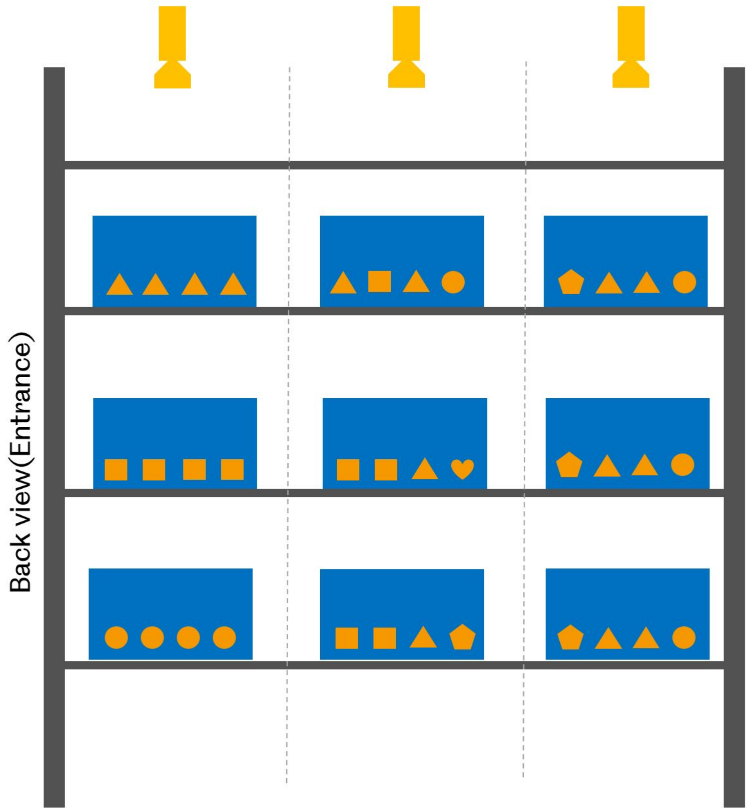


Fig. 3.11 Installation position of the camera at the entrance side

3.4 Summary

In summary, this chapter presents a solution for internal inventory monitoring within a flow rack, utilising both weight and vision sensors for real-time inventory management. Factors such as worker misdirection and other anomalies are not considered within this scope. Experiments were conducted using individual sensors to compare the impact of different mounting locations and external factors. The results indicate that both types of sensors are applicable in this context. However, due to the limitations inherent to each sensor type, a complementary approach, such as installing sensors at the back of the rack shown in Figure 3.11, appears to be a more effective solution. While hand-detection methods mentioned in the text are beyond the scope of this chapter, incorporating additional hand-related data could indeed enhance real-time accuracy by calculating inventory reductions. This will be considered in future research plans. Moreover, this technology is not restricted to flow racks and has the potential for broader application, which can be achieved by modifying the weight sensors or camera positions on alternative rack systems.

Chapter 4

Utilising Mathematical Methods Based on the Noisy Channel Model for Factory Monitoring

4.1 Introduction

There are roughly three typical technical routes for utilising multi-sensor data to Extract useful information. The first is a model-based approach [125]. Utilising this method requires significant domain expertise to enable investigators to accurately mirror the target system's behaviour using mathematical and physical descriptions [126]. Nevertheless, constructing an exact model for a complicated manufacturing operation or machinery is challenging due to the insufficient physical specifics, coupled with uncertainties inherent to certain processes. The alternative method relies on a statistical foundation [127]. This strategy deduces sensor information from past system conditions and examination of pertinent sensor readings, necessitating a significant amount of computation and knowledge in the domain. This requires a large amount of high-quality observed sensor data and a correctly estimated noise distribution [128]. These methods usually require manual feature extraction [129, 130], such as eigendecomposition and principal component analysis [82]. Performance can be affected by how representative the extracted features are and how well they estimate the noise distribution. The third approach involves the emerging field of artificial intelligence, especially deep learning. This approach does not require detailed domain knowledge like model-based approaches.

On a typical factory floor, a large number of parts are consumed in the manufacturing assembling process, to achieve successful internal logistics results, while parts are shown at the right place, right amount and at the right time. To understand the logistics and performance, monitoring those activities is necessary. Counting and classification are two methods for achieving this. In a factory where parts are sent to the workbench by a 'shop', to avoid running out of stock on the assembly line, operators tend to take more parts than required, which results in waste. We have therefore designed a counting system to help the 'shop' to give the right number of parts and types when they are sent out, thus improving material utilisation.

Traditional methods require manpower to repeat inventory checks at certain intervals or to count each dispatch at once, which is time-consuming and costly.

Although there are some cases where a single sensor can be used for specific tasks, it generally comes with limitations. For example, only a single type of object can be detected or classified based on weight. On the other hand, RFID needs to be marked on every object inspected. Marking a large number of unrecyclable parts is also labour-intensive and may cause changes to the external shape of the part.

To overcome these challenges, detection based on multi-sensor fusion will be introduced in this chapter. This method combines the advantages of two or more sensors to improve predictive performance. There have been similar studies on unmanned supermarkets, such as Amazon's Just Walk-Out Technology [54] and AIM3S cashier-less convenience stores [55]. These unmanned supermarkets solve the pain point of queuing for checkout and provide more accurate real-time inventory to reduce out-of-stock scenarios.

The advantages of automated inventory monitoring applications in supermarkets could also be applied in manufacturing scenarios. Precise inventory control leads to fewer out-of-stock events, reduced waste, less assembly line downtime, and higher profits. Unlike retail, where most products are displayed on shelves, parts in the assembly line are stored in some kind of storage device, such as a Euro container. These containers are stacked or placed in rack compartments to save space and for ergonomic purposes. Manufacturing assembly parts often come in similar shapes and sizes (e.g., nuts with different diameters) and are stored in denser detection areas (e.g., a large number of parts in a box). Hence, location-based information, like that of unmanned supermarket products being fixed in a shelf location, may not be efficient for this study.

Therefore, this chapter introduces the multi-sensor method, which uses a mathematical approach to fuse weight and visual data to predict different types and quantities of parts in small containers.

4.2 Methods and Materials

4.2.1 An Examination of the Mathematical Parallels between Sensor Fusion Techniques and Word Spelling Correction Methods

While each type of sensor has its advantages and disadvantages and a single sensor can only collect limited information. This thesis uses the combination of vision and weight sensors to compensate for the occlusion of vision due to light, dense stacking of parts, and the error caused by the inconsistent test environment; as the weight sensor makes it difficult to effectively identify different types of products at one time, but its hard effect by the disadvantage factor of vision. This research will explore the fusion of these two types of sensors, by using mathematical fusion methods to archive better inventory monitoring results. This automatic method is used to detect what is in the box placed on the table, so as to monitor picking activities and hence inventory level.

4.2.2 Boundary and constrain

An automated detection system that can deal with extreme cases of various situations is very challenging. Since this research is still in the early stage, some boundaries and constraints are set in this stage to avoid overcomplicating the system, and all operations should be made like normal human behaviour e.g. not to fool the system, moving the parts or boxes at high speed, maintain a stable external environment, focus on the monitoring parts type and quantity. To explain the scope of this paper, we make the following assumptions:

1. Upper and lower weight boundary
   Due to the maximum weight limit of the weight sensor, the total weight of each test case must be less than or equal to the maximum weight limit, e.g. 5 kg, quantity for each type of part o will be limited. Moreover, the total weight would include the detected object's weight plus the box's weight while the minimum total weight is equal to zero i.e. empty shelf, no box.

2. According to reality, no more than 10 types of objects appear in each identification.

3. The maximum quantity of a single type item is limited by the box's volumes. The volume of each object is different, only a limited number of objects can

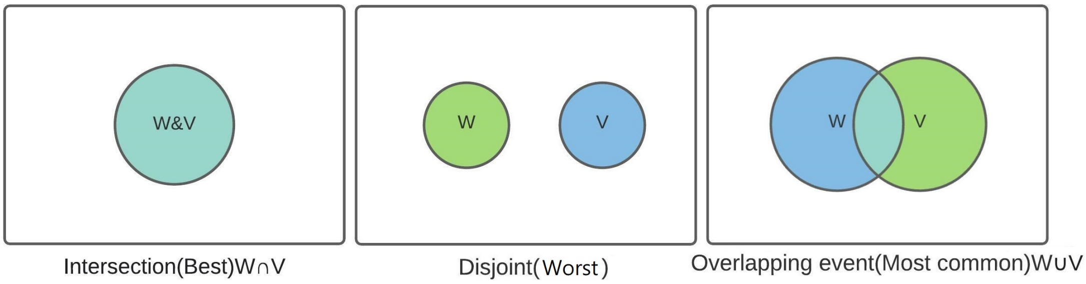


Fig. 4.1 Different event situations when the sensor output

be placed in each box, limiting the maximum number of objects based on prior knowledge. Currently, a maximum of 20 parts are allowed in one box.

4. Total weight during the detection is a constant. The signal received by the weight sensor is constantly changing, we will take the average value as a reference during a unit time.

Through the collection of sensors and the properties of the detection target input in advance, the system will determine the objects within the detection range based on the following previous information.

- Total weight (W): Measured by weight sensor including the weight of all boxes and the objects;
- Number of types of item in the data set (up to 4);
- Item's properties, including weight, volume as well and vision properties (pretrained model);
- RGB picture images from the camera.

This study adopts Decision-In-Decision Out (DEI-DEO) when using the fusion method of mathematics. The output would show the type and number of objects. In the case of receiving multiple sensor outputs at the same time, there will be three situations in Figure 4.1.

Where W represents the output signal of the weight sensor, V represents the model of the visual output, which can be seen from Figure 4.1.

If the weight and vision outputs are identically defined as intersections, implying that both types of sensors confirmed the same result, the result will be output directly. Disjoint is the opposite, the two types of sensors will not have any identical results, this situation will not be discussed in this paper due to conflicting messages from the two sensors. The overlapping is the most common

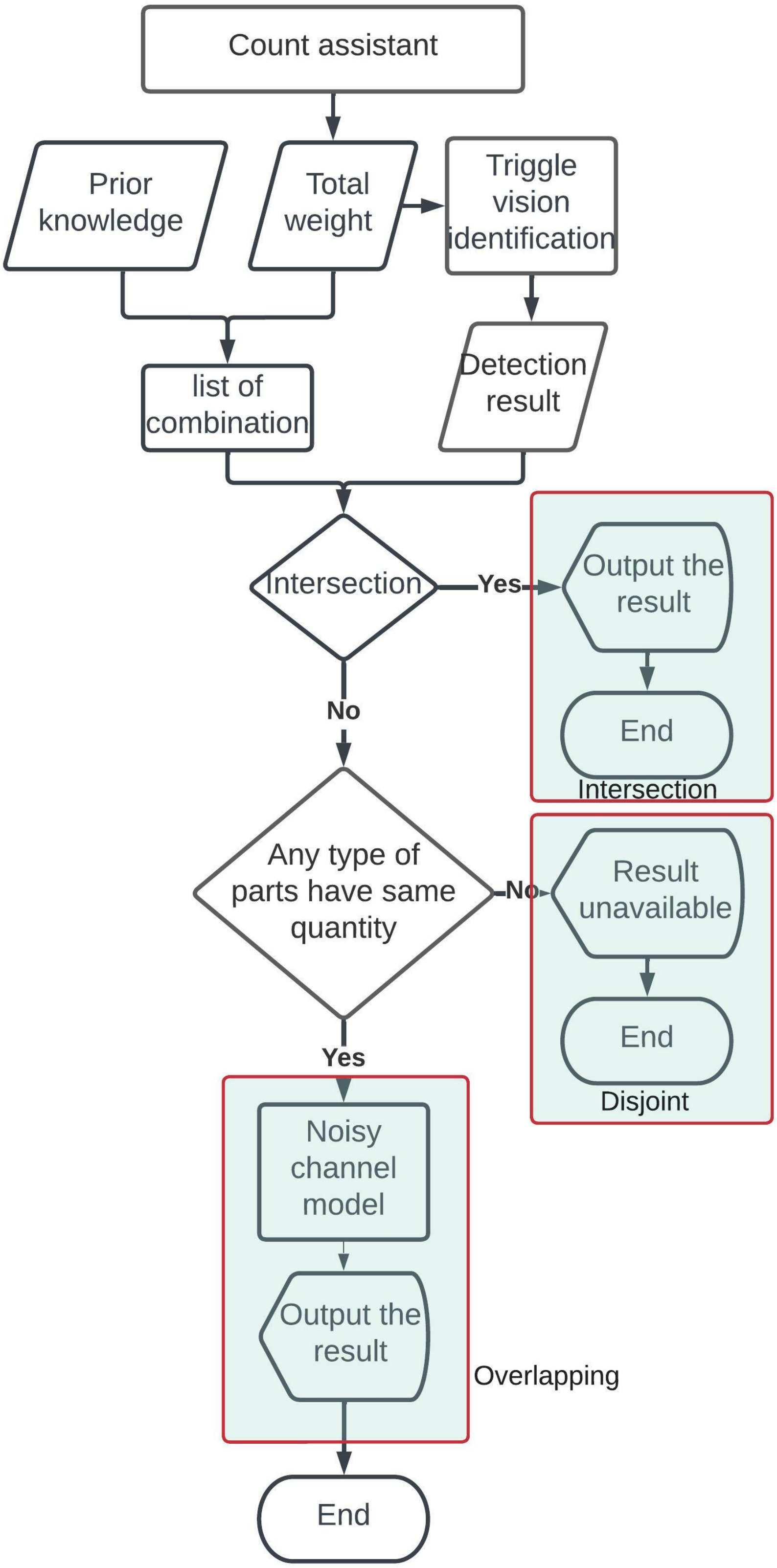


Fig. 4.2 A flowchart of the system framework for items classification

case, some of the output from vision and weight are identical, and the final output will be determined by the noisy channel model. See Figure 4.2 for the algorithms corresponding to different situations.

4.3 Results and Discussion

As fasteners were used in the experiment since they have different visual characteristics but similar size and weight. In the experiment, three different combinations of parts were simulated, as detailed in Figure 3.1:

- Object 1 is a round washer with a weight of 38g;
- Object 2 comprises flat square washers with a weight of 60g;
- Object 3 consists of external coach screws with a weight of 14g;
- Object 4 is represented by studding connectors with a weight of 41g.

Table 4.1 Possible weight for all combination

Weight in g / Quantity per type / Number of types	1	2	3	4
1	38	76	114	152
2	98	196	294	392
3	112	224	336	448
4	153	306	459	612

Initially, both the weight and vision sensors underwent calibration. The weight sensor was tested with standard weights to ensure an accuracy tolerance within ±5g, while the vision sensor (camera) was calibrated for colour accuracy, resolution, and frame rate, ensuring it captured clear images at 10 fps in a controlled lighting environment. The data collection process was designed to minimize environmental variability, with each part weighed five times to calculate an average, thus mitigating transient errors. Multiple angles of images were captured under consistent lighting conditions to ensure comprehensive visual data.

To account for variability and ensure statistical validity, each combination of parts was tested 50 times, shaking the box to reset it each time. The tests were randomized to prevent sequence effects, and this large number of repetitions allowed for the assessment of the system's consistency and the detection of outliers or anomalies. The experiments were conducted in a controlled environment, minimizing external factors such as vibration, temperature fluctuations, and camera position, ensuring sensor readings were influenced solely by the parts being tested and not by external noise.

The experiment design involved selecting four different types of parts. Combinations of these parts were created to test the system's ability to identify varying types and quantities. The mathematical model for fusing sensor data was based on the noisy channel model, chosen for its ability to handle probabilistic data and effectively combine multiple data sources. The experiment aimed to determine the type and quantity of parts using this model, leveraging both weight and visual data.

The system used for the experiments included Windows 10 OS, 16GB RAM, NVIDIA GTX3070 GPU, and Intel I7-9700 CPU, with Torch 1.12.1 + cu113 and Anaconda managing computational tasks and EfficientNet models.

4.3.1 Weight only classification

Given that weight is merely a one-dimensional metric, it does not provide comprehensive insights in isolation. Fortunately, as the weights of various parts are known, we can transform this into multidimensional data, facilitating correlation with image information.

As the measure object properties are prior knowledge, it can be transferred to a pure integer programming problem. It is a specialized branch of mathematical optimisation that seeks the best possible outcome under a given set of constraints, with the unique distinction that the decision variables must assume integer values. In particular, Pure Integer Programming restricts all decision variables to be integers, as opposed to some variables being allowed continuous values, as seen in Mixed Integer Programming (MIP). The mathematical formulation typically entails maximizing or minimizing a linear objective function, subject to a set of linear constraints.

In the realm of optimisation, Pure Integer Programming stands out as a pivotal technique for scenarios demanding discrete decision-making. Unlike its counterpart, Linear Programming, which allows for continuous decision variables, Pure Integer Programming insists on solutions that are whole numbers, ensuring that decisions are explicit and tangible, as often required in real-world contexts. This distinction is paramount in areas in our case, where, parts cannot be divided arbitrarily; the weight output is invariably a positive integer, or, in the absence of any placement on the scale, it is zero. Despite the computational challenges of polynomials, computer aid would help solve the problem.

Mathematically, the total weight measured by the weight sensor is W. Where the weight of each object h_n is known. Due to the natural fact, all numbers will be positive integers or zero. Given that all variables, including the weight of the object h, belong to the set of non-negative integers, $\mathbb{N}_0$, Equation (4.1) became:

$$h_1x_1 + h_2x_2 + ... + h_nx_n = W_n \tag{4.1}$$

The main object of this thesis is item identification, for which correctness is defined as each combination coinciding with the actual type and quantity.

Using integer programming for object quantity h_n in the equation(4.1) and computer assistance, all the combinations can be calculated. In general, the number of attempts will increase exponentially while the number of type objects increases. Since assumption (1) (2) (3) the quantity for each item x_n is limited, with ten or fewer, the number of attempts will be significantly reduced. This would give the number of each object (quantity) $x_1, x_2, ..., x_n$.

Consider the error from the load cell sensor, which, according to experience, for load cells with a carrying capacity of 2-5kg, the error is within ±5g, and the tolerance was set to be 5g. Thus, obtaining the closest set or sets of results. Similarly, this significantly reduces the number of combinations. Additionally, since the weight-sensor system not only includes the load cell but also a connected Raspberry Pi, power amplifier, and components for converting analogue data to digital data, its error may be amplified. This phenomenon has been validated in preliminary experiments.

The term Identification Probability (IP) is used to measure the likelihood of correctly identifying objects based on their weights. Given that the analysis deals with discrete integer combinations, the Identification Probability represents the probability of correctly identifying the objects among all possible combinations. It intuitively conveys the success rate of object identification through weight data. Specifically, the Identification Probability is calculated as the reciprocal of the number of possible combinations, as shown in the following equation 4.2:

$$IP = \frac{1}{\text{Number of combinations}}(\%) \tag{4.2}$$

The experiment is designed to identify objects based on their weights and corroborate the findings with image data to enhance accuracy and reliability. The primary goal is to transform weight data into multidimensional information and use it in a pure integer programming framework to solve for object quantities. This is achieved by measuring the total weight of various combinations of objects and applying integer programming to determine the possible quantities of each object type. The process includes calculating the Identification Probability (IP) to evaluate the success rate of object identification through weight data. To further validate the results, a camera is activated to trigger image recognition, cross-checking the weight-based identifications with visual data

Four different types of parts with varying weights are selected. According to Table 4.2, the complexity increases with the number of types and quantities. For instance, one type with one quantity corresponds to one object1, while four types with four quantities mean four objects, each having four quantities. Each part is weighed five times to calculate an average, thereby determining the confidence level for each measurement. The experimental setup includes a load cell sensor, a Raspberry Pi for data processing, and a camera for image recognition. The weights are measured with an error tolerance of ±5g.

In the case of weight identification, it can also be seen from Table 4.2 that the accuracy decreases rapidly with the increase of type and quantity.

Table 4.2 Confident level for weight identification

Accuracy / Quantity per type / Number of types	1	2	3	4
1	50%	25%	16%	6.6%
2	25%	14%	9%	1.1%
3	25%	11%	6.7%	0.86%
4	5.6%	1.6%	0.82%	0.61%

4.3.2 The reason for using EfficientNet

EfficientNet, a state-of-the-art deep learning architecture, achieves higher accuracy and efficiency by employing a compound scaling method, which harmoniously scales network depth, width, and resolution. Designed with computational efficiency in mind, it offers a compelling solution for tasks on mobile and edge devices. However, training these networks, especially their larger versions, may demand intricate hyperparameter tuning and significant computational resources. In our study, the project is of a modest scale, thus making it feasible to utilise EfficientNet.

Furthermore, due to the integration of the DEI-DEO fusion, we are afforded the flexibility to select from a variety of state-of-the-art models. This model,

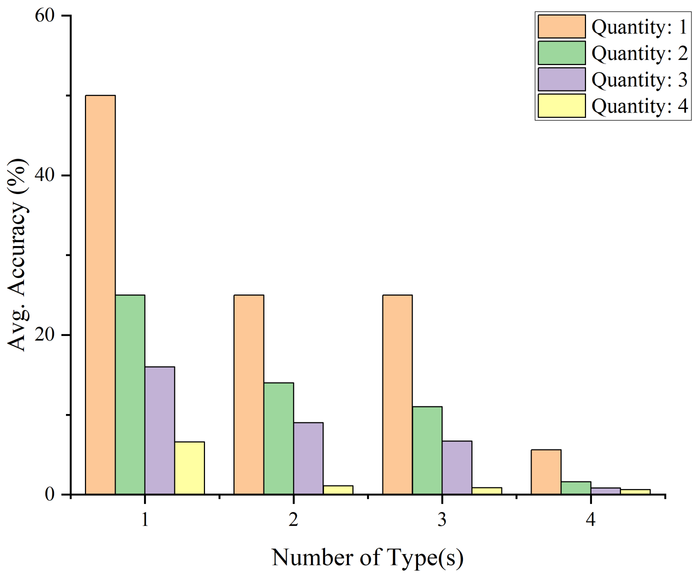


Fig. 4.3 Weight only identification accuracy for different types of quantity

possessing commendable accuracy for its time, has been extensively applied and is easily deployable on various edge devices. Through the model, we can determine both the type and quantity of objects. Combining this with weight, hence enhances detection results.

4.3.3 Vision only classification

When designing visual identification, camera angles and camera types may impact the result, as neural networks have improved in recent years, the resolution of the camera may not have to be high-end. From previous experience with small objects 500-720 pixels is more common for training and computation. As for the frame rate, since the object is placed on the static weighing sensor and not taken down until the end of the detection, 10 fps or more is sufficient to obtain a clear view. In addition, sufficient light was maintained in the study, the type and placement of the cameras were constant in the experiment and it is beyond the scope of this thesis to optimise these factors.

We define the accuracy of the vision identification as follows:

$$V_{accuracy} = \frac{\text{Correct identification per type}}{\text{True quantity}}(\%), \tag{4.3}$$

Table 4.3 Confident level for vision identification

Accuracy — Quantity per type / Number of types	1	2	3	4
1	100%	100%	83%	87%
2	100%	87%	58%	56%
3	83%	66%	56%	58%
4	75%	56%	54%	49%

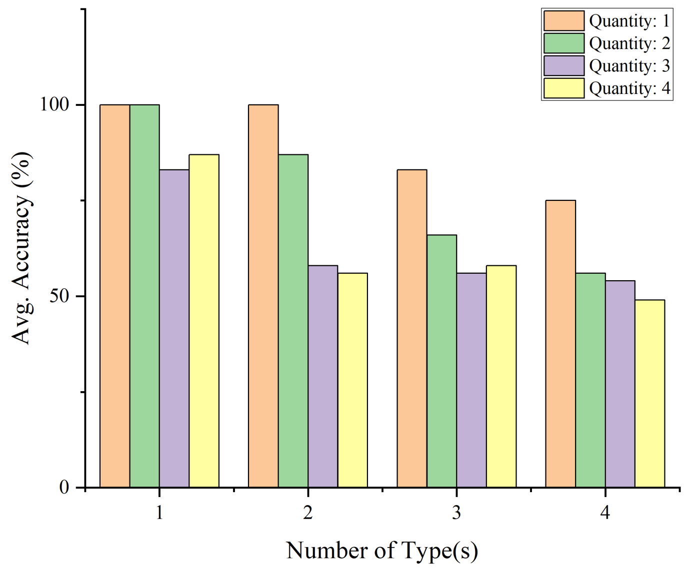


Fig. 4.4 Vision only identification accuracy for different types of quantity

4.3.4 Combination of weight and vision via Noisy channel model algorithms

The results of a single weight and visual classification were described previously, in this section it is described how prior knowledge and these two sensor modalities can be combined to estimate a final prediction.

The noisy channel model is a classic model in Natural Language Processing (NLP) and Information Theory that describes the process of transmitting information over a noisy channel. The core idea of the model is that the source message is distorted or corrupted when it passes through a noisy channel, and the receiver receives a distorted version of the message. This algorithm is frequently employed to address spelling errors. In the prior study on the noisy channel model, the primary objective was to correct spelling by generating candidate choices with maximum probabilities from the corpus [131]. In the current study, the true combination is discerned from the camera-derived combination and through the weight sensing of these combinations. As in Eq. (4.1), one or more sets of h are obtained from the weight sensor; then another set of h is obtained from the camera.

In this chapter, we utilise the noisy channel model as a probabilistic framework to deduce the most likely genuine information. The model is employed for prediction by aligning the predictive input derived from weight sensing with visual data.

The mathematical formulation and formally described using Bayes' Theorem can be represented as below:

$$P(s|t) = \frac{P(t|s) \times P(s)}{P(t)} \tag{4.4}$$

Where:

- $P(s|t)$ represents the posterior probability of the source combination s given the received combination t.
- $P(t|s)$ is the likelihood, representing the probability of receiving combination t given the source combination s. It captures the nature of the noise in the channel.
- $P(s)$ is the prior probability of the source combination s.
- $P(t)$ is the marginal probability of the received combination t, effectively a sum over all possible source combination s of $P(t|s) \times P(s)$.

Series number	A	B	C	D	Total weight
2200	0	0	2	2	110
1500	0	0	5	1	111
800	0	0	8	0	112
111	1	1	1	0	112
3	3	0	0	0	114
1110	0	1	1	1	115
410	0	1	4	0	116
1002	2	0	0	1	117
302	2	0	3	0	118

Table 4.4 The output, given an input weight of 114, produces the number of possible combinations

In general, information derived from the weight modality serves as a more robust product predictor, primarily because it is less influenced by occlusions; consequently, we attribute higher probabilities to it. Based on observation and reality, the following factors were used to obtain the closest results when applying the model.

- Higher probabilities are given to the combination with the largest quantity of type in visual detection;
- Distance between the guess and the correct answer
- If the distances are too large, compare with the original group by increasing the number of guessed parts to see if more results are obtained;
- The closer the measured weight is to the calculated combination, the more probabilities are given;
- Selecting from weight calculation combinations where the quantity is equal to larger than the vision identified;
- An equal number of parts of two or more types allows for greater possibilities.

On the other hand, since only combinations within situations are detected, it is possible to filter out most of the combinations for the weight sensor. For example, an example with an input of 114 would result in nine different combinations when the condition is met, but only two are possible. Refer to Table 4.4, an example with an input of 114 would result in nine different combinations when the condition is met, but only two are possible.

As four items are in the inventory, A, B, C, and D. Each item has a corresponding weight (which we call a coefficient), h_a, h_b, h_c, and h_d. Our goal is to estimate these weights from the given table. In addition, the given constraint is that the total weight must be in the form of one of the following:

$$
\begin{aligned}
&h_a \\
&h_a + h_b \\
&h_a + h_b + h_c \\
&h_a + h_b + h_c + h_d
\end{aligned}
\tag{4.5}
$$

From the table 4.4, only the rows with series numbers 3 and 111 satisfy this condition.

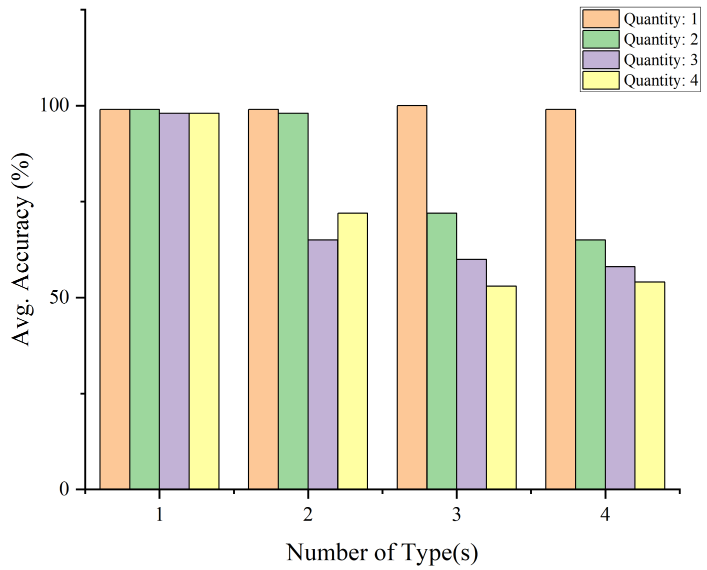


Fig. 4.5 Fusion identification accuracy for different types of quantity

Similarly, define the fusion accuracy as the fusion result being intersection with the actual situation, the item's quantity and types match the correct combination. The results are shown in Figure 4.5, with the fused data showing an improvement over the single type sensor (Figure 4.3 and 4.4).

In an experimental setting identical to the one used for the previous single vision and weight analysis, it is evident that as the variety and quantity of parts increase, there is a notable decline in accuracy. This can be attributed to the fact that with an increase in the number of parts, the coverage in the specified area not only results in diminished visual outcomes but also leads to a significant reduction in fusion. Nevertheless, the results remain superior to those obtained using a single sensor.

The limitations of the study were similarly disclosed when it applied the noisy channel model to combinatorial optimisation problems.

1. **Computational Overhead**: The NCM necessitates the estimation of multiple probability distributions. This can lead to a surge in computational demands, especially in scenarios with large combinatorial spaces.

2. **Accuracy and Stability**: The efficacy of the NCM hinges on the accuracy of the model definition and the chosen probability distributions. An improper choice can result in unstable or imprecise outcomes, diminishing the reliability of the model.

3. **Over-Engineering Risks**: For certain problems, deploying the NCM might be an over-sophisticated approach. In several cases, more straightforward methods such as direct mathematical solutions or heuristic algorithms might offer quicker and more intuitive solutions.

4. **Dependency on Prior Knowledge**: Bayesian approaches, often associated with NCM, rely on prior probability distributions. Bias may be introduced in assigning probabilities to them through actual situations and observations

4.4 Summary

To summarize, while the NCM can offer insights into various domains, it still requires a great deal of prior knowledge and manual probabilistic calculations, making the use of this model a hindrance. Moreover, the difficulty rises further as the variety and number of parts and situations increase in complexity.

In this chapter, an algorithm was proposed for the automatic detection of the type and quantity of parts in a dense area. A mathematical noisy channel model is used to fuse the list of combinations from weight difference and vision identification. This fusion results in an accuracy of up to 80%, which is higher than either single vision (73%) or weight detection (up to 50%) alone. Future work will expand the system to allow for the detection of more types and quantities of parts, as well as optimising the image recognition model. Furthermore, experiments with machine learning methods will be conducted to compare fusion performance.

Chapter 5

Developing sensor fusion for higher accuracy dense objects identification via ML method

5.1 Introduction

Industry 4.0, also part of smart manufacturing, epitomises a distinct trend in the evolution of industrial technology [132, 133]. This evolution is propelled by advanced communication technologies along with sophisticated data analysis techniques [134]. Given the preceding discussion in previous chapters, a significant expansion in the diversity and volume of data generated from sensors within production activities, products, and manufacturing apparatuses is anticipated. Such data encompasses aspects like temperature, light, vision, vibration, and pressure through physical contact.

In Chapter 4, the statistical method is employed. While this explored the feasibility of the statistical approach in fusing sensor and weight data, it also has significant limitations, such as a strong reliance on prior knowledge.

The main contribution of this chapter is a pioneering deep learning approach that accommodates both weight and image inputs, enabling the model to adeptly fuse data from these two sensors. This research was conducted from two perspectives:

- A novel deep transfer learning approach utilising YOLOv8 is presented, which extracts features and noise and then weights them through an FEIDEO fusion process.

- Subsequently, the efficacy of the model is validated through experimental procedures.

A pivotal question arises: How can one effectively harness sensor data from varied modalities to extract valuable insights, ultimately enhancing productivity and diminishing time expenses? Within industrial settings, the data procured from a myriad of sensors is inherently intricate, encompassing multiple modalities, disparate measured physical quantities, and functioning at diverse scales.

Consequently, there is a pressing need to explore methods that do not heavily rely on computational power. Within the realm of data fusion, prevalent statistical methods include [74–76]. These techniques have been expounded upon in the literature review. In recent years, the swift progression of chip technology has propelled deep learning to the forefront of attention. Within the field of multi-sensor fusion, deep learning techniques are particularly prominent. In contrast to conventional approaches, those rooted in deep learning typically possess the capability to autonomously discern and assimilate intricate features, dependencies, and patterns, circumventing assumptions regarding noise. However, the construction of neural networks mandates specific expertise and familiarity, especially when tailored to a precise objective, such as merging weight data with sensors for the purposes of classification and counting. The literature review delineates four fusion techniques, namely, Feature in-Decision out (FEI-DEO), Decision In-Decision (DEI-DEO), and so forth. Fusion at the data level predominantly favours homogenous perceptions and tends to generate a substantial volume of input widgets, while concurrently grappling with the challenges of incomplete measurements. As such, the other two methods, specifically FEI-DEO and DEI-DEO, appear more apt. Yet, it remains a challenge to ascertain which technique will offer superior accuracy, ease of implementation, or necessitate less foundational knowledge prior to its actual deployment. As such, investigations should be initiated from these two perspectives.

In the context of FEI-DEO, features are extracted from the raw data of sensors and then input into the model for fusion. By independently analysing these features, an approach to detect objects is achieved. Conversely, DEI-DEO requires the processing of raw data to output results from each sensor before undergoing fusion for prediction. While both methodologies aim to identify the quantity and type of objects, their methods are distinct, leading to the development of two separate approaches. Each method was tested across nine categories, operating under the assumption that each input algorithm would have

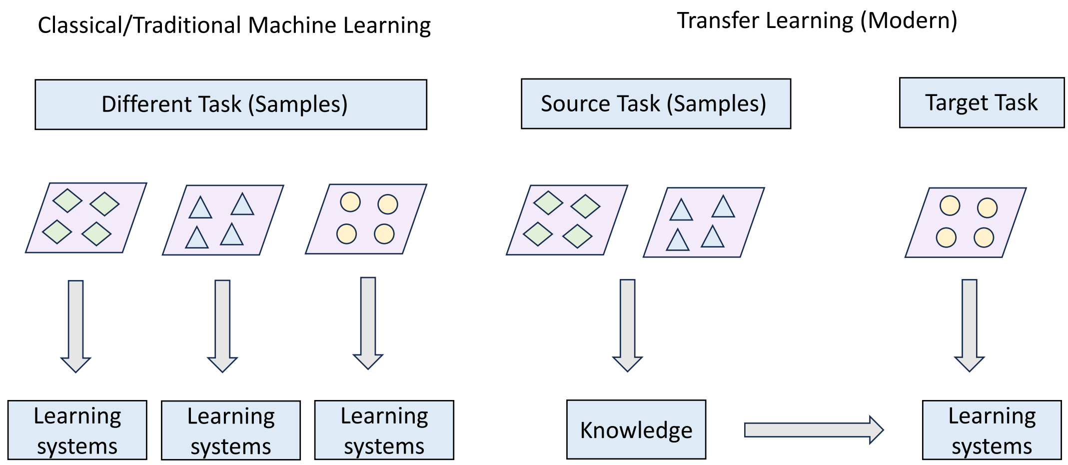


Fig. 5.1 Comparison between new fusion mode design and transfer learning that can train itself

a certain degree of error during detection. Their accuracy was monitored, and limitations were identified.

The main contribution for this chapter is A pioneering deep learning approach is presented which accommodates both weight and image inputs, enabling the model to adeptly fuse data from these two sensors. This research is conducted from two perspectives:

- A novel deep transfer learning approach utilising YOLO v8 is presented, which extracts features and noise and then weights them through an FEI-DEO fusion process.
- A novel approach harnessing deep learning has been devised for decision-making, integrating weight and visual sensors. The refinement of this model was undertaken by utilising historical noise data for training and verification purposes.
- Subsequently, the efficacy of the model is validated through experimental procedures.

5.2 Problem statement

Deep learning-based approaches have become increasingly common in data fusion. It is an effective tool for processing and interpreting huge amounts of data, but it also poses computational, data and model interpretability challenges. It is difficult to determine intuitive to determine which approach is best suited for fusion. inputting all features into a single model via FEI-DEO may allow the model to capture more complex, richer information. However, better feature engineering would also be required, and model uncertainty would increase.DEI-DEO is human-understandable for each data source, so a specific, targeted model could be used for each data source. But it may also result in partial loss of information.

However, when dealing with large and complex sensor data, different dimensions will bring challenges and require a lot of human effort for feature extraction. On the other hand, the consumption of GPU and the artificial feature extraction time cannot be ignored. It is not realistic to redesign a model for every industrial case. Therefore one should avoid training a completely new model, start with a similar model, and make changes.

In the preceding chapter, we delved into the capabilities of the DEI-DEO fusion through a mathematical lens. This approach facilitates the switching between image models but concurrently demands comprehensive probability estimation. In numerous instances within the manufacturing plant, obtaining the necessary prior knowledge and probabilities through mathematical methods proves challenging. To enhance inventory monitoring and material traceability within the factory, we shall introduce a sensor fusion technique rooted in deep learning. Given its proficiency in managing high-dimensional, noisy, and expansive datasets, alongside its capacity to discern valuable features and make precise predictions, machine learning emerges as a viable solution for sensor fusion challenges. Specifically, a deep fusion approach grounded on the DEI-DEO and FEI-DEO sensor fusion methods will be presented, with its efficacy appraised relative to both techniques via empirical assessment. Deploying these methods can facilitate the flow of components within the factory, thereby refining the overall internal logistics.

Specifically, a deep fusion approach grounded on the DEI-DEO and FEI-DEO sensor fusion methods will be presented, and its efficacy relative to both techniques will be appraised via empirical assessment. This chapter also aims to develop a feature extractor based on YOLO v8 that extracts features and noise from the input data and then weights them through an FEI-DEO fusion process. The deployment of these methods can facilitate the flow of components within the factory, thereby refining the overall internal logistics.

5.3 Methodology

5.3.1 Selection of image detection model

The 'You Only Look Once' (YOLO) methodology signifies a paradigm shift in computer vision, particularly in object detection. Unlike traditional methodologies that rely on iterative scanning techniques, YOLO, as the name implies, necessitates only a single pass through the image to detect multiple entities. This unique approach ensures real-time object detection, which has proven invaluable for applications necessitating prompt responses such as autonomous vehicles and surveillance systems. YOLOv8 is compatible with all previous versions of YOLO and allows seamless transitioning between them. Additionally, it operates efficiently on a range of hardware platforms, including both CPUs and GPUs, demonstrating significant versatility. Furthermore, YOLO operates as an integrated neural network, offering end-to-end training that directly predicts bounding boxes and class probabilities from the input image. This obviates the conventional two-step detection process, which first proposes regions and subsequently classifies them.

In the implementation presented, standard YOLOv8 is utilised for the initial feature extraction and object detection due to its high accuracy and efficiency. These modifications include custom layers for noise injection and a specialized fusion process to combine image features with weight data.

Both approaches require the use of an image recognition technique to extract features from the image. Since the design is such that the image model can be switched at will, the accuracy of the various YOLO models is shown in Figure 5.2. We can see that the eighth version has the highest accuracy, even though it also has some certainty, as will be mentioned below.

The strength of this algorithm lies in its enhanced precision for detecting small-sized objects; moreover, it guarantees that the detection accuracy for each object size matches, if not surpasses, that of current algorithms.

This section provides an overview of the most prominent algorithms in recent years, with a particular focus on the enhancements introduced in YOLOv8. YOLO stands as the leading real-time object detector for several reasons, which include: (a) its lightweight network architecture, (b) efficient feature fusion techniques, and (c) enhanced accuracy in detection results.

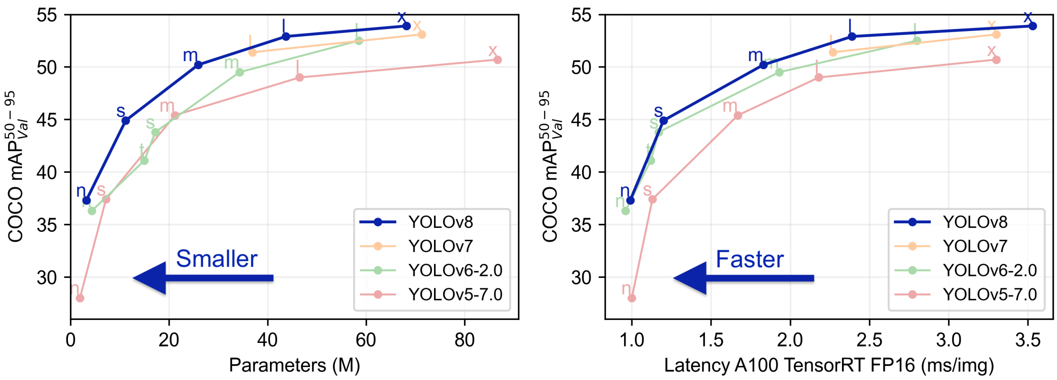


Fig. 5.2 Performance comparison of YOLO Versions [135]

Investigating small or large objects

YOLO's design also presents certain challenges. Given its intrinsic mechanism of segmenting input images into an SxS grid and assigning object detection to the grid cell containing the object's centre, YOLO can sometimes falter in accurately detecting overlapping or smaller objects.

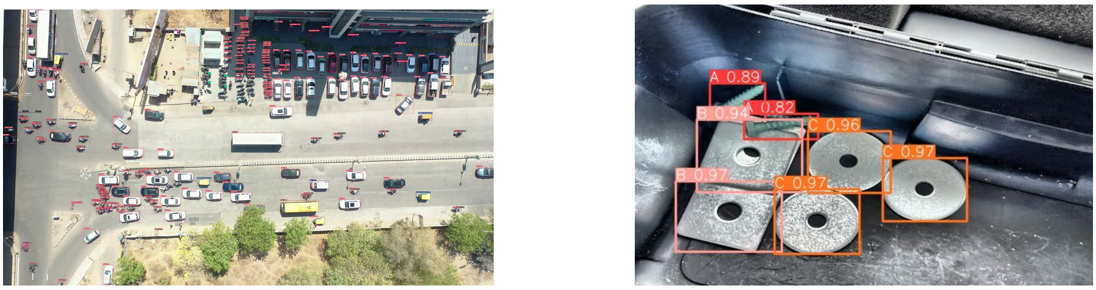
In image detection, distinguishing between large and small objects is often based on their area or scale in the input image. The definition of 'large' and 'small' is relative, depending on the application and task requirements.

Taking into account the size of objects, in this case, it's not akin to drones detecting 'large' targets like cars on the ground (which appear quite small in images). Rather, our scenario involves detecting 'small' targets like screws, which, in reality, appear prominently in the images. Therefore, our case primarily focuses on larger targets.

The aforementioned is based on the absolute scale of pixel count and the relative scale of the proportion of the image. Additionally, due to YOLOv8 employing scale pyramids to handle multi-scale images, there are three detectors designed to handle different sizes. Therefore, we can start with these three detectors and try to find a solution.

5.3.2 Data generation and preprocessing process

These data principally comprise two components: one part pertains to images, while the other is concerned with weight. In the realm of deep learning, the quality of preprocessed data bears a significant relationship to the model, hence the closer the data aligns with reality, the more advantageous it becomes. Observations from previous experiments indicate that sensors invariably exhibit a certain degree of



(a) Small object (b) Larger object

Fig. 5.3 Comparison of large and small object detection

error. For the load cell aspect, the error in weight sensors manifests as a normal distribution within the bounds of their elastic limit. Hence, normal distribution weights were labelled to mimic the measurement error generated by the load cell.

The function (5.1) and (5.2) show the probability distribution of the value of zero. Mean $\mu = 0$ determines the location of the symmetry axis of the function, which is the centre of the distribution, and standard deviation $\sigma = 1$ determines the width of the distribution, which indicates how far the data of the distribution is discrete. For the standard normal distribution, approximately 68% of the data lie in the interval [-1, 1], 95% in the interval [-2, 2], and 99.7% in the interval [-3, 3]. When the mean is zero and the standard deviation is 1, it is known as the standard normal distribution [136]. The mathematical properties of this parameter are precisely set to characterise the weight error. Most data points cluster approximately 1 unit away from the mean value, as graphically illustrated in Figure 5.4. From Figure 5.4, in rare cases of 0.3%, the weight was distributed outside of ± 3. These weights are excluded as anomalies and are not trained. Drawing upon the Gaussian distribution, the weights of various categories have been processed in preparation for the subsequent step of input into the FEI-DEO model.

$$f(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5.1}$$

When $\mu = 0$ and $\sigma = 1$, simplify:

$$f(x \mid 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{5.2}$$

On the side of imaging, distinct components are arranged within a container according to a prior combination, and subsequently subjected to random oscillation

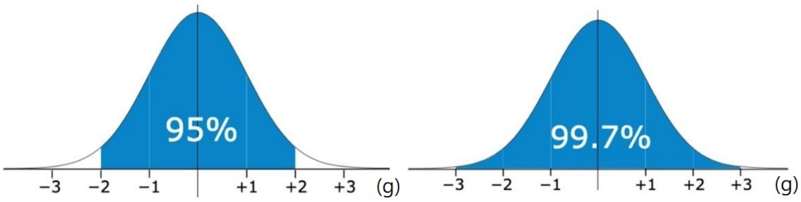


Fig. 5.4 Weight allocation via normal distribution

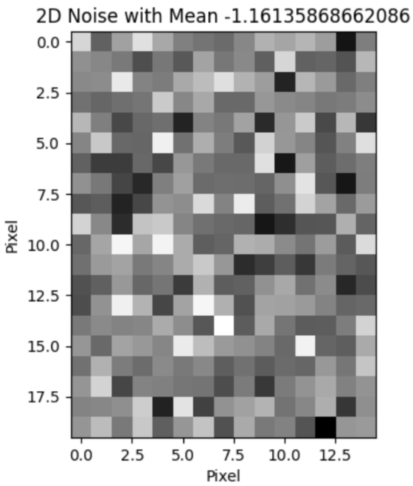


Table 5.1 Weight data transfer to Noise map

to generate images, with approximately one hundred iterations per combination. These images are then resized to a 640 × 640 resolution and undergo a series of enhancements, including contrast enhancement, as well as transformations such as clockwise, counter-clockwise rotations and grayscale alteration at 9%. Additionally, to bolster the model's resistance to occlusion, cutouts are employed, specifically using five boxes with 7% size each. Through this methodology, there is an approximately threefold increase in the number of images. In the second phase, all images are annotated using the Segment Anything Model (SAM), with the input image exemplified in Figure 5.5

5.3.3 Data alignment

Timestamp alignment is a crucial procedure when dealing with data that carries time-related labels. This ensures that data is properly synced according to its associated time. Due to inconsistencies in the sampling rates between weight sensors and cameras, during each detection, the average weight is derived from ten data points and is then fused with the image data proposed. Each detection

(a) AAA (b) AAA (c) AABBCC (d) AAABBBCCC

Fig. 5.5 Image after segmentation

is carried out through the 'trigger' process; hence, at the timestamp when the last weight detection concludes, image information is extracted.

If the task wishes to fuse contextual information from different feature maps, such as merging images from different sources or resolutions, then joining along width or height may be more appropriate. If the persona wishes to share spatial information between different feature maps, but each feature map represents a different feature or channel, then a connection along the channel dimension may be more appropriate.

5.3.4 Feature in decision out fusion

A quintessential case of feature extraction and fusion is elucidated by [88, 89], wherein features are initially segregated prior to undergoing a training phase. This process entails extracting two distinct types of features from images and weight respectively. The first extraction leverages the YOLOv8 algorithm, while the second entails a transformation of weights into a particular feature representation. Utilising the You Only Look Once version 8 (YOLOv8) model, a significant advancement has been made in the field of object detection and classification within images. By meticulously dissecting this model, we have engineered a novel approach that accommodates two distinct inputs: weight and image. The weight input serves as a vital parameter to modulate the detection sensitivity, providing a unique adaptive response tailored to different object characteristics. Concurrently, the image input undergoes an intricate process of feature extraction and contextualisation through the YOLOv8 model. This innovative bifurcation of inputs fosters a more nuanced and precise analysis, potentially revolutionising the way computer vision systems are integrated and applied across various industrial and technological domains. Fusion of images with other data. Embedding images and language into an embedding vector (analogous to our feature map) This allows the dimensions of the feature representations of visual and other data to be

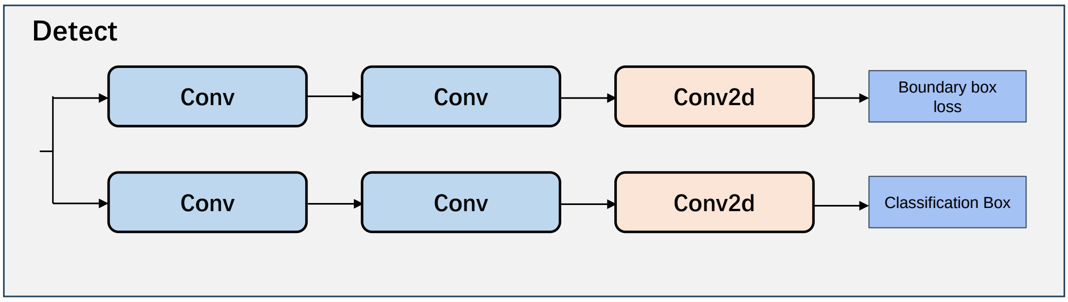


Fig. 5.6 Yolo v8 detection head that outputs the image feature

in the same format for feature-level fusion [137]. Through the YOLOv8 structure Figure 5.7, can be split into two modules as follows:

- Backbone represents the foundational architecture of object detection models, generally comprising deep convolutional neural networks. Functionally, the Backbone is responsible for extracting salient features from an input image. Capture hierarchical features ranging from low-level textures to mid-level patterns. These extracted features are integral for downstream tasks and set the stage for more refined processing [138, 139].

- The Head of the model functions as the task-specific segment, handling the precise objectives of classification and localisation. Utilising the features channelled from the Neck, the Head is tailored according to the detection task and may encompass structures such as fully connected layers, convolutional layers, and specific activation functions. For instance, in object detection paradigms, the Head typically consists of two distinct branches: one for classification and another for bounding box regression. This orchestrated interplay between the Backbone, Neck, and Head ensures a comprehensive and nuanced approach to object detection, adeptly balancing the requirements of feature extraction, integration, and task-specific interpretation [139, 140].

In the 'Head' segment, three detection heads are situated at the 15th, 18th, and 21st layers. Each of these detection heads is optimised for recognising small, medium, and large targets, respectively. By modifying these layers, the feature maps are extracted in advance without entering the detection and classification heads. These highly extracted features are then fed into the subsequent Convolutional Neural Network (CNN) model for final convolution to integrate weight information. As a result, the experimental attention will primarily centre on the large target detection feature map at the 21st layer. For training purposes, feature maps from all three detection heads are input to the following CNN model.

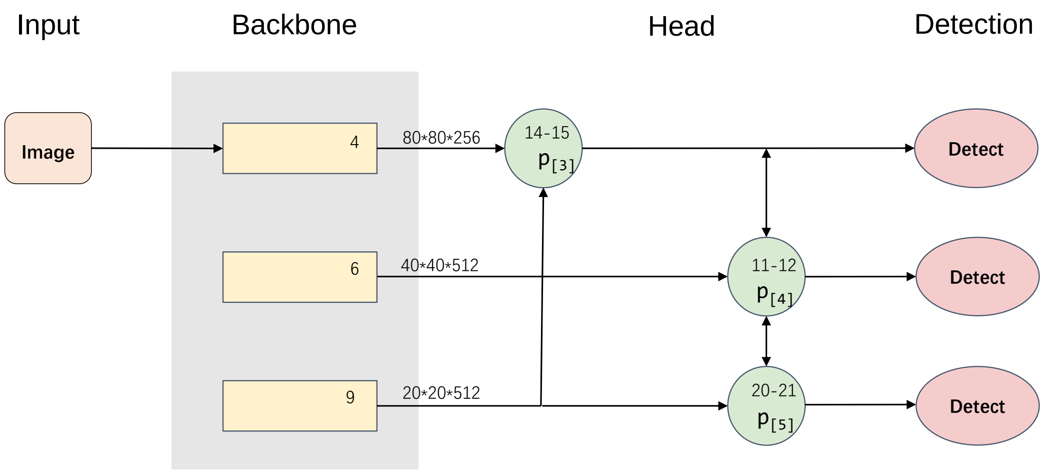


Fig. 5.7 Model structure with weight input

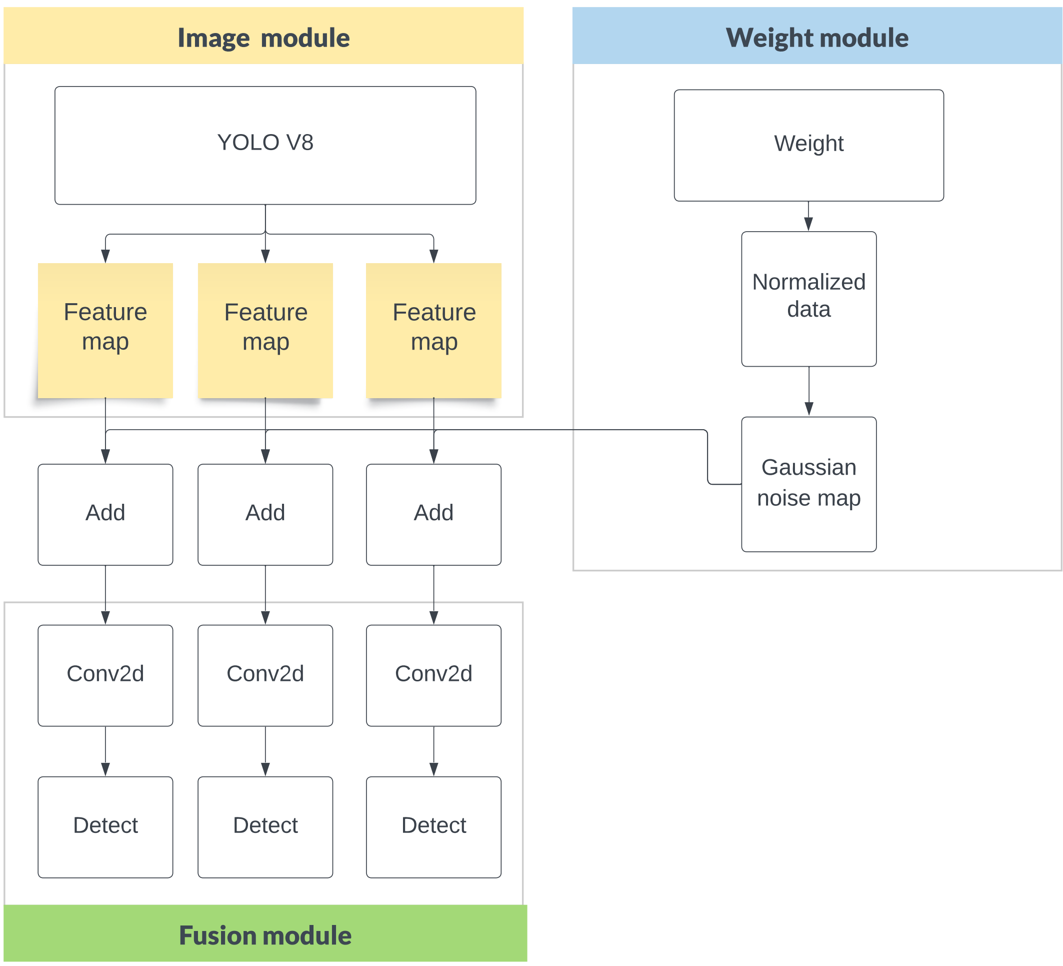


Fig. 5.8 Feature input process for the FEI-DEO fusion

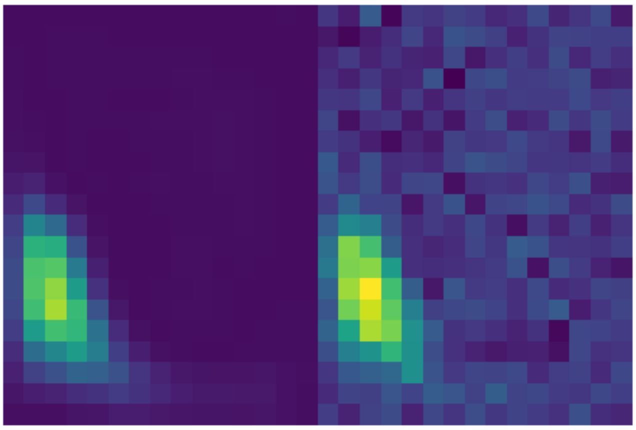


Fig. 5.9 A feature map is augmented with a noise map: On the left is the feature map generated by YOLOv8, and on the right is the combined feature map with the noise map

The feature map produced by V8 is illustrated on the left-hand side of Figure 5.9. In contrast, the image incorporating weighted noise is displayed on the right. This is prior to its introduction to the fusion section. When merging two feature maps along the channel dimension, what we are effectively doing is 'stacking' each channel from the second feature map atop the channels of the first feature map. This action retains the spatial dimensions of the feature map (i.e., both width and height remain unchanged).

The fusion model employed for processing two feature maps is a conventional CNN tailored for image classification tasks. This model is architecturally designed to accommodate images of dimensions 3x128x128.

Since the input is a feature map and the desired features have already been extracted, the fused model starts with two consecutive convolutional layers:

The first convolutional layer takes the 3-channel input and transforms it into a 32-channel feature map using 3x3 filters. This is followed by a ReLU activation function and a max-pooling layer with a 2x2 kernel that downsamples the spatial dimensions by half. The second convolutional layer further processes the 32-channel input from the previous layer and produces a 64-channel feature map, again using 3x3 filters. This layer is similarly followed by a ReLU activation and another max-pooling operation. After these convolutional stages, the network flattens the feature maps into a 1D tensor and feeds it into a series of fully connected layers (also known as dense layers). The first dense layer reduces the dimensions from 64x32x32 to 1024, with a ReLU activation function and dropout regularization to prevent overfitting. The final dense layer maps the

1024-dimensional vector to a vector of 'num classes' dimensions, which represents the probabilities of each class in the classification task.

In essence, this model employs a combination of convolutional layers to extract features from input images and dense layers to classify these features into predefined categories.

5.3.5 Decision in decision out fusion

A primary reason for the necessity of sensor fusion is that the noise from a single sensor can impair its detection capability. Fusion algorithms also serve to reduce the noise impact by integrating data from multiple sensors. Typically, the error found in a single sensor also exists within multi-sensor systems. In this section, a dataset will be established based on historical visual data. Noisy inputs will be based on previous vision data to generate the dataset which will subsequently be fused using a deep learning technique. Ibid. The thesis uses summation. This is due to the fact that the pattern of matrix multiplication is suitable for fusion where there are complex interrelationships between features. Our case has only noisy MEAN and can simply use summation to reduce computational complexity [141].

In terms of the input, recap the equation:

$$h_1x_1 + h_2x_2 + ... + h_nx_n = W_n \tag{4.1}$$

We introduce several different new variables: I_x,I_y, I_z and W_I. Where I represents the mathematically possible all the real numbers, and the subscripts x,y, and z represent the number of parts A, B and C. Convert (4.1) to the equation below:

$$h_1I_x + h_2I_y + h_3I_z = W_I \tag{5.3}$$

$$\begin{aligned} I_x, I_y, I_z \in \mathbb{R}_{\geq 0} &\quad \text{(where } \mathbb{R}_{\geq 0} \text{ denotes the set of all non-negative real numbers)} \\ W_I \in \mathbb{N}_0 &\quad \text{(where } \mathbb{N}_0 \text{ denotes the set of all non-negative integers)} \end{aligned}$$

The methodology in this session is divided into three steps:

(i) Firstly, the method simulates the generation of a dataset where each data contains the number of three parts (A, B and C) plus some random anomalies. The datasets are labelled with real combinations of parts.

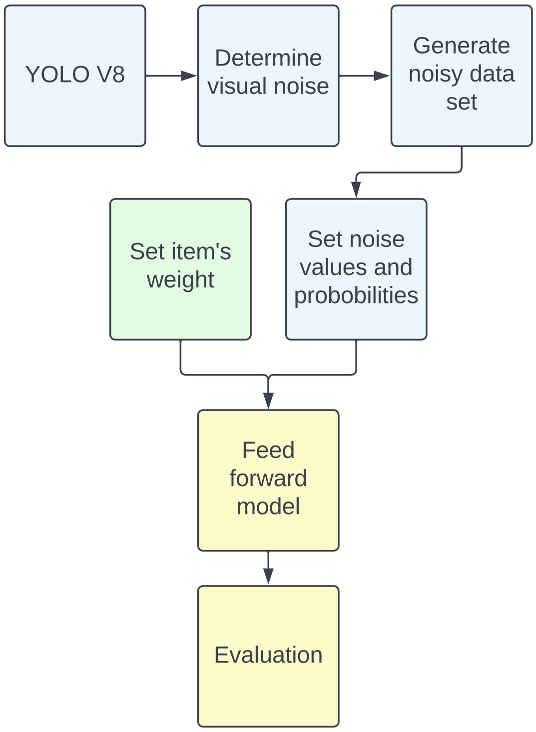


Fig. 5.10 Overview of Anomaly-Infused dataset generation and fusion structure

(ii) Next, a neural network model is defined and trained to learn and predict combinations of parts from the simulated data.

(iii) At the end of each training cycle, the model is evaluated on a test set to see how well it performs.

To be specific, we utilised the previously trained YOLO v8 model, with the test set encompassing scenarios such as overlapping and occlusion. After 500 tests, the probability of errors was derived, with the results appearing to align with a normal distribution; larger errors were less frequent. The probability distribution for both positive and negative values is presented in the table 5.2 to 5.4 under 'Probability of each noise value'. These random anomalies correspond to errors for each category, as depicted in Figure 5.10. Using these probabilities, we generated 500 combinations to form a dataset, which, upon normalisation, was rendered into a format amenable to deep learning, facilitating subsequent classification training.

In terms of model design, after continuous debugging and testing, the structure of the neural network as shown in Figure 5.12 has the highest accuracy.

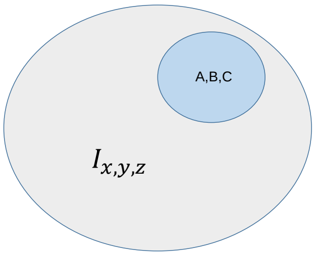


Fig. 5.11 Probability distributions of reality and theory

This model is a deep feed-forward neural network. It is characterized by multiple linear transformations, with ten linear layers mapping the input to a 64-dimensional output space, except for the first layer which takes a 4-dimensional input. Regularization was introduced through two batch normalization layers, which aim to maintain a consistent distribution of inputs during training. The model also incorporates the ReLU activation function for non-linearity at certain points.

For tasks involving classification across multiple noise levels, the essence lies in the accurate extraction of features and the determination of classification boundaries. Feed-forward neural networks, especially deep ones, possess the capability to learn the art of extracting complex features from input data. This model, with its multiple linear layers and nonlinear activation functions, can theoretically approximate any function, making it adept at classifying varied noise patterns.

One of the strengths of this design is its model complexity achieved by stacking several linear layers and nonlinear activations. This allows the model to capture intricate decision boundaries. Moreover, the integration of BatchNorm reduces the risk of overfitting, ensuring that the model performs well on unseen data.

Table 5.2 Parameters and their values case 1

Parameter	**Value**
Possible noise values	[0, 1, -1, 2, -2, 3, 4, 5]
Probability of each noise value	[0.6, 0.1, 0.1, 0.05, 0.05, 0.05, 0.03, 0.02]
Weight of part A	14
Weight of part B	60
Weight of part C	38
Maximum quantity	4
Number of variations	500
Training set	32000
Testing set	4000

Table 5.3 Parameters and their values case 2 (2A=B)

Parameter	**Value**
Possible noise values	[0, 1, -1, 2, -2, 3, -3]
Probability of each noise value	[0.6, 0.1, 0.1, 0.05, 0.05, 0.05, 0.03, 0.02]
Weight of part A	14
Weight of part B	28
Weight of part C	38
Maximum quantity	4
Number of variations	500
Training size	28000
Testing size	4000

Table 5.4 Parameters and their values case 3

Parameter	**Value**
Possible noise values	[0, 1, -1, 2, -2, 3, -3]
Probability of each noise value	[0.6, 0.1, 0.1, 0.05, 0.05, 0.06, 0.04]
Weight of part A	14
Weight of part B	28
Weight of part C	38
Maximum quantity	4
Number of variations	500
Training size	28000
Testing size	4000

5.4 Experiment and Results

5.4.1 Experimental platform

This experiment was conducted in the Google Colab environment using an NVIDIA Tesla V100 PCIe with 16 GB of memory, and the deep learning framework employed was PyTorch.

Tables 5.2 to 5.4 indicate, the hyperparameters of the training in order to compare the results.

5.4.2 Evaluation Methodology of Results

In the experiments, the accuracy of the FEI-DEO and DEI-DEO fusion methods was compared and also compared with the detection ability of a single sensor. There are several ways to assess accuracy and model classification performance:

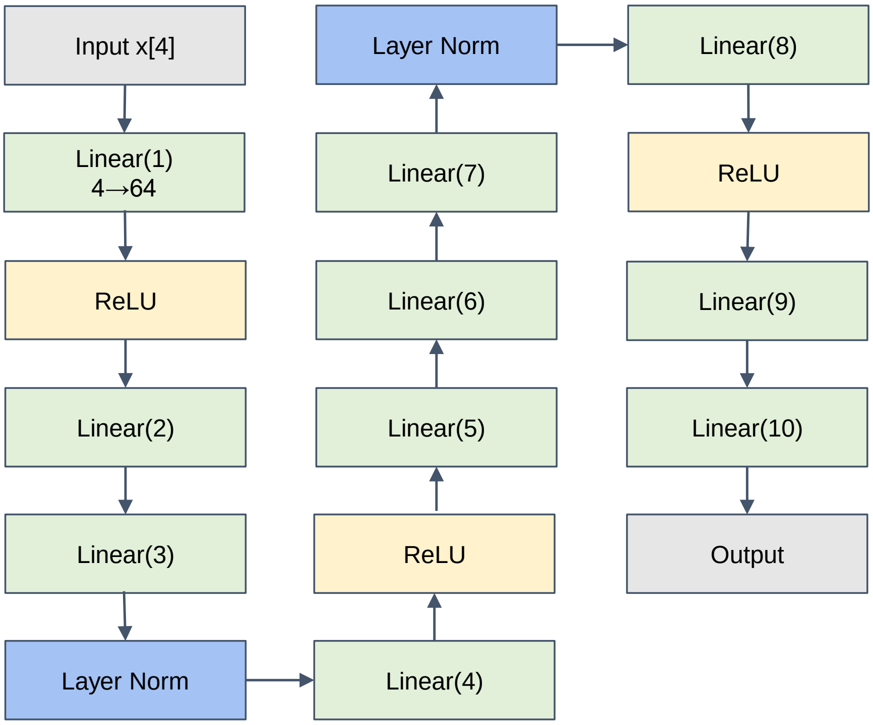


Fig. 5.12 Feed-forward neural network for DEI-DEO fusion

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5.4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5.5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.6}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5.7}$$

In addition to this, a confusion matrix is also common on classification tasks and is a tabular representation of actual versus predicted classifications. In a multi-class classification setting, the matrix dimension is N×N, where N represents the number of distinct classes. Each row of the matrix corresponds to the actual class, while each column corresponds to the predicted class. This matrix aids in visualizing the performance of an algorithm, not only in its successes but also where it tends to make mistakes.

Although confusion matrices are common in the classification field, standard confusion matrices are designed for binary classification problems. It is possible to extend the confusion matrix for multi-classification tasks, its complexity increases with the number of classifications, making interpretation and visualisation more difficult.

In the field of sensor inventory detection, True positive (TP) is the result we want. False Negatives (FN) and False Positives (FP) have a more detrimental impact on the overall performance of the model, whereas the effect of True Negatives (TN) is relatively insignificant. For instance, if there's a need to determine that an object is not A, and it is detected as B but is actually C, this is still considered within the acceptable range. As a result, an accuracy assessment method is introduced. In the context of this study, an entire group of combinations is regarded as a single entity. The formula for accuracy can then be adapted as follows:

$$\text{Accuracy} = \frac{\text{Number of correct prediction}}{\text{Total number of predictions}} \tag{5.8}$$

5.4.3 Results and Discussion

DEI-DEO

The classification task and training setup will remain consistent; however, the input images will vary. In terms of weight, inputs will follow the previous normal distribution to simulate noise. There are a total of nine categories, and each category will be measured 50 times. This means that after measuring all the categories once, the process will be repeated 50 times.

FEI-DEO

With regard to feature-level fusion, the experiment comprises three distinct cases. These cases explore varying weights of A, B, and C, where 2A=B and situations without utilising error probability (details can be found in parameter Table 5.2 - 5.4). The outcomes for the cases are depicted in Figure 5.13.

5.5 Discussion

According to the above, modifications were made to the YOLO v8 algorithm to facilitate its integration with weight data. These modifications include a custom layer for simulating weight sensor noise, which applies a Gaussian noise model to the weight data before fusion with image features. Additionally, a feature fusion layer was introduced to combine the image features extracted by YOLOv8 with the transformed weight features, done by concatenating the feature maps

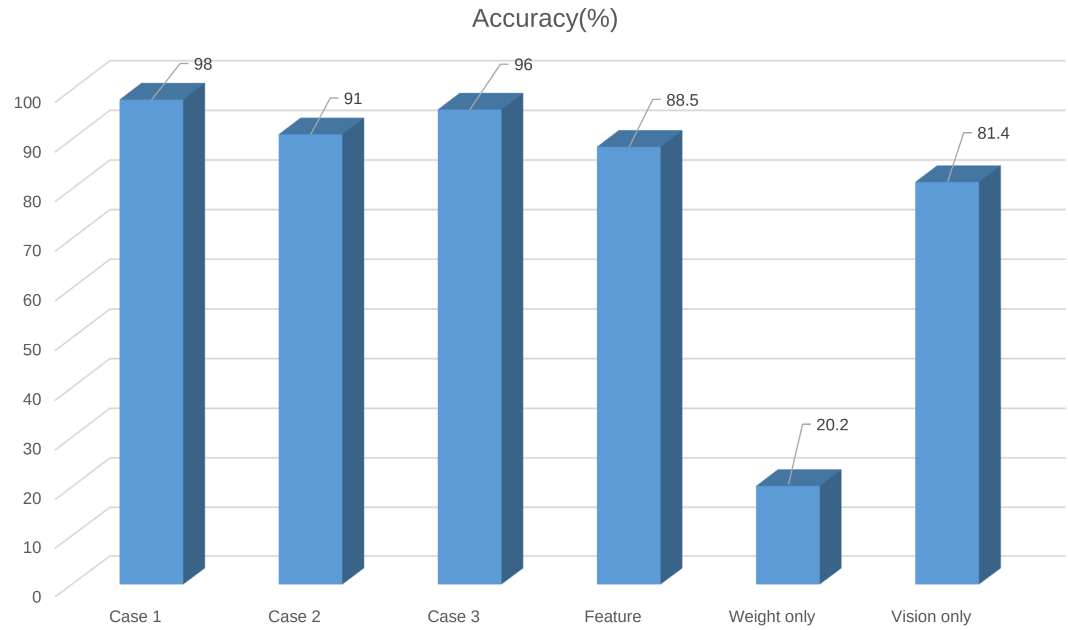


Fig. 5.13 Results of the model's accuracy and comparison across different cases

along the channel dimension to learn complex interactions between visual and weight data. A specialized dual-objective training procedure was also developed, accounting for both object detection accuracy and weight prediction accuracy, ensuring the model's effectiveness in handling both data types.

From Figure 5.13, it can be observed that the accuracy of the feature-based fusion is marginally lower compared to the decision-level fusion. A plausible reason for this could be that some features were not inputted into the model. Given that only one detection head from YoloV8 was utilised, this might be one contributing factor. However, the accuracy still surpasses that of a singular sensor.

For the decision-level fusion, the probability from visual experiments in Case 1 represents the real-world scenario. This dataset includes some amount of anomalous data and severe occlusions. Components are categorised into three classes, with each class being randomly detected from 0 to 4 times, achieving an accuracy of 98%. Owing to the limited number of categories and minimal weight overlap, Case 2 was introduced. This scenario incorporates overlapping weights where weight $2A = B$ and modifies the error to resemble a normal distribution (0, 1, -1, 2, -2, 3, -3), thus testing the model's overall capability. Theoretically, the noise level should be controlled within ± 3, as anything exceeding this could likely result from a suboptimal image recognition model or anomalous data. The outcome for Case 2 displayed an accuracy of 91%. The purpose of designing Case

3 was to compare the accuracy under genuine target noise (screws and washers) with that under normal distribution noise, achieving a 96% probability.

A feature-level fusion should ideally yield superior performance metrics. However, this often goes hand in hand with extensive feature engineering efforts. Feature engineering is an intricate process encompassing the selection, construction, transformation, and optimisation of data features to facilitate better model learning and prediction. This is particularly pronounced in multi-modal or multi-source data contexts. Feature-level fusion becomes paramount, ensuring the harmonious coexistence of features across diverse data sources for optimal learning. The introduction of undue noise by improper feature engineering might explain why the experimental results were slightly below that of decision-level fusion.

On the other hand, decision-level fusion primarily focuses on the output layer of the model. This fusion method obviates the need to directly manipulate raw features and combines the decision results from multiple models to achieve a consolidated and more reliable prediction. A notable advantage of this method is its aptitude for handling diverse model outputs effectively. However, this introduces new challenges. For instance, our weight allocation might not be reflective of the actual situation, and there is a need to ensure that decision-level fusion doesn't amplify biases inherent in any particular model.

It is imperative to acknowledge various constraints when interpreting the outcomes of this thesis. Within this below, the limitations pertinent to each methodology are delineated, and potential remedies are suggested.

In FEI-DEO fusion, the model cannot detect conditions other than labels due to training limitations. In order to solve the above problems, a method is proposed so that whenever we have a complete box (the same as the training one), we can determine what is left in the box in real-time by weight subtraction. The mathematical relationship can be represented as:

$$\Delta W = W_i - W_f \tag{5.9}$$

Let: ΔW represents the difference in weight, where W_i is the initial weight (of the full box) and W_f is the final weight.

$$Q = \frac{\Delta W}{w_(A, B, C)}$$

Therefore, the difference of ΔW is that every item that is taken away is subtracted from the inventory by the method of full inventory before getting the latest real-time inventory.

In the case of DEI-DEO, the model possesses a higher accuracy, but it also proceeds under our previous constraints, i.e., that is, according to Eq. 4.1. That is, the types of combinations are limited, and the same subtraction operation is required when the material is consumed.

5.6 Summary

This chapter introduces a novel machine-learning-based fusion method and subsequently trains a new model based on YOLOv8 to address the sensor fusion part-counting issue in factories. The modifications included custom layers for noise injection and a specialized fusion process to combine image features with weight data. In solving this problem scenario, both FEI-DEO and DEI-DEO approaches were explored, with FEI-DEO achieving 86% and DEI-DEO 98% in accuracy respectively. In theory, the FEI-DEO method which captures more features should be more accurate, but in practice uncertainties such as the extraction of features and processing of noise maps lead to lower results than DIDO. This is also in line with the expectations.

Results indicate that the method is effective, albeit with certain limitations. It is adept at identifying within a limited scope of categories. Additionally, YOLOv8 boasts other detection capabilities, such as pose detection and real-time video monitoring. Serving as a feature extraction engine, it significantly reduces the workload associated with repeatedly training new models. Furthermore, the SAM model, by segmenting training images, ensures that the image input possesses reduced noise, which positively influences the model's enhancement.

Chapter 6

Development of the Quantity, Recency, and Frequency (QRF) Concept Model for Enhancing the Response to Consumption Variations in Replenishment Policies

6.1 Introduction

The ever-evolving landscape of aircraft manufacturing frequently presents with a myriad of model versions, each entailing a distinct set of specifications and requirements. This inherent variation often incubates uncertainty, notably in terms of determining the exact needs for each manufacturing cycle. The volatility is a significant challenge in supply chain management, especially in aerospace, due to factors like changes in product design, equipment failures, customer demand shifts, and economic conditions [142]. On the other hand, as illustrated in Figure 6.1, although a customer places an order for a product, the specific version of the aircraft is often determined post-Customer Definition Freeze (CDF), typically extending beyond a six-month period. The volatile nature of demand and the occasional alterations in design specifications further exacerbate this ambiguity, thus accentuating the necessity for a more flexible supply mechanism. Such a mechanism should not only accommodate the inherent variability but also be

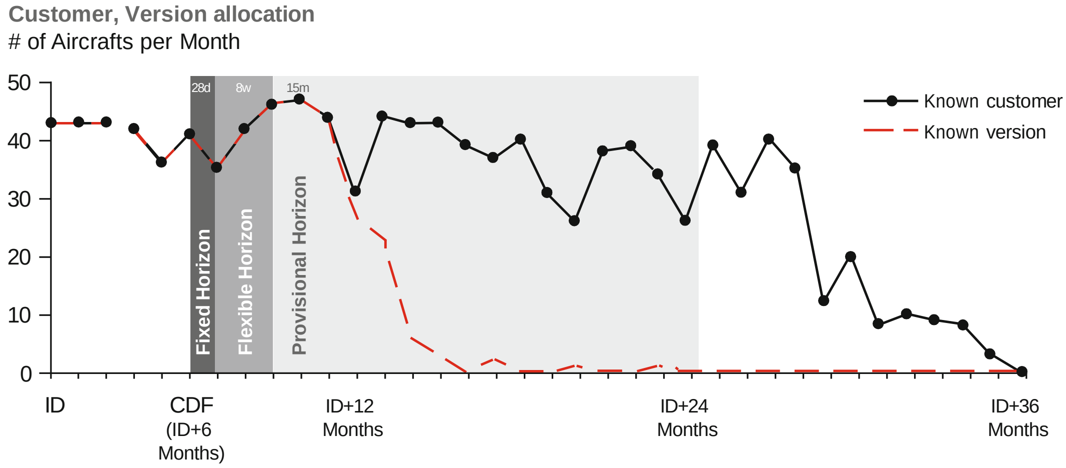


Fig. 6.1 The aircraft model is not assigned even though the customer is identified [143]

adept in responding promptly to system feedback, thereby enabling real-time adjustments based on the newly acquired data.

The aerospace industry is entangled with a myriad of challenges when it comes to managing component consumption and ensuring accurate forecasting. Among the prominent challenges are inaccurate forecasts, untimely responses, and excessive inventory costs which may arise from conventional replenishment strategies based on timing, experience, and usage. These conventional strategies often fail to accurately reflect real-time material demands, leading to either overstocking or stock-outs.

The inadequacy in precise forecasting is primarily rooted in relying on traditional methods that are fundamentally grounded on time, experience, and usage for replenishment. Such methods seldom accurately mirror real-time material needs, paving the way for overstocking or stock shortages. Moreover, the reactionary measures are often found to be lacking in timeliness. The traditional stock replenishment strategies are frequently devoid of ample data support and real-time oversight, leading to less precise and belated decisions which in turn hike up the inventory costs. A visual representation of the rising costs attributed to high and low inventory levels is elucidated in Figure 6.2.

In addition to the above, existing studies have predominantly focused on discerning sequential patterns [100] based on the frequency of occurrences within data. However, the consumption behaviour of factory components might undergo variations over time. These variations could be intertwined with certain random factors, such as temporary changes in assembly ideas by customers. Therefore, relying solely on temporal analysis proves to be insufficient. It underscores the

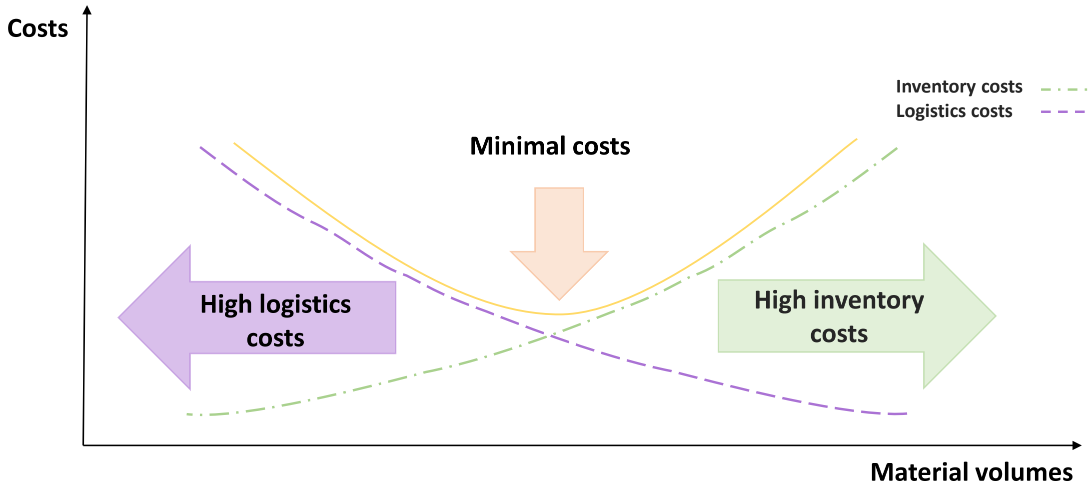


Fig. 6.2 Balancing transaction costs against inventory costs

necessity to consider recent quantities, proximity, and frequency among other factors to identify significant consumption patterns.

In order to transcend these challenges, a fusion of real-time data analytics, prediction algorithms, and a responsive inventory management system could be envisioned. This amalgamation could enable a more accurate prediction of component consumption, timely decision-making, and optimal inventory levels, thereby reducing operational costs and augmenting efficiency in the manufacturing process.

6.2 Methodology and Architecture

6.2.1 Analyzing Replenishment and Consumption Characteristic

Airplanes represent highly complex products with stringent technical requirements. The plethora of diverse components intermingle within the supply chain, where instances of incorrect or over-supply are not uncommon. A prime example of this is the AGS parts discussed in a previous chapter of this thesis. Through a literature review and understanding gained from factory visits, several characteristic features have been identified:

- **High Customisation:** Components necessitate a high degree of customisation to meet varying customer requirements, rendering the process of component replenishment and consumption more intricate than that observed in industries with mass-produced items, such as automotive. As

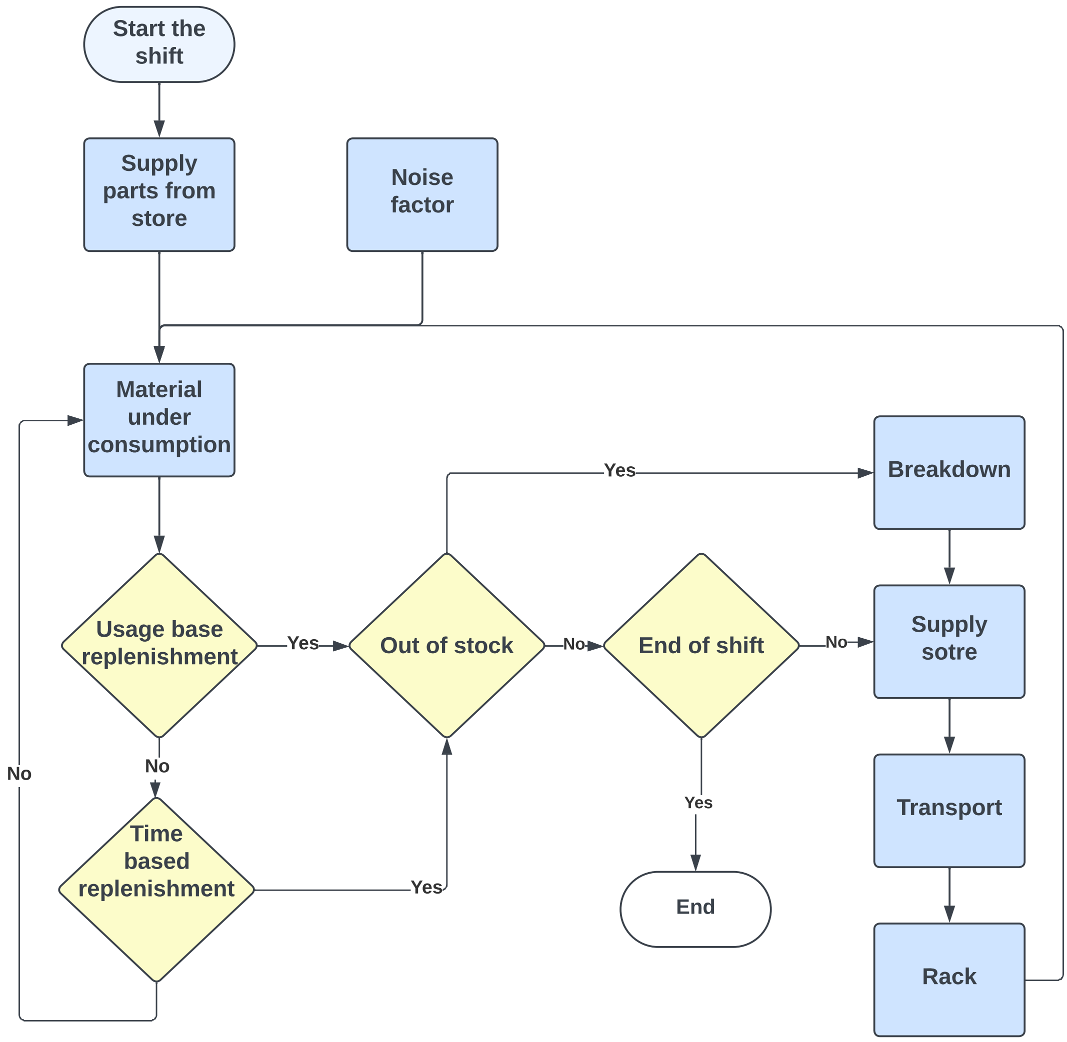


Fig. 6.3 Material flow: Conventional consumption and replenishment on the factory floor

illustrated in Figure 3.5, various components can be assembled into a certain part, facilitating the subsequent production phase [144].

- **Trends change in time** The dynamics of trends are influenced by numerous factors including seasonality and the sequencing of assembly

The selected parameters for generating the simulated inventory consumption data shown in Table 6.1 are designed to reflect the real-world variability and requirements in aerospace material consumption. The use of simulated data allows us to control the experimental conditions precisely, ensuring analysis of different variables' impacts. The standard deviation of 43 captures the moderate fluctuations in daily usage, while the range of 500-2500 units ensures a broad

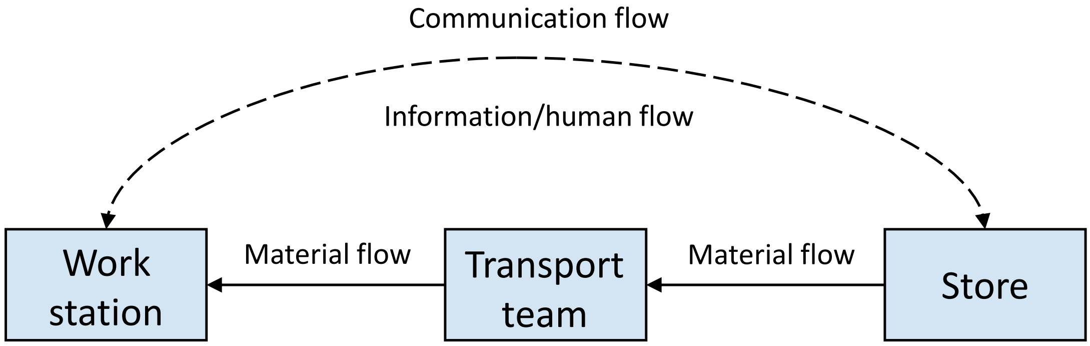


Fig. 6.4 Shop floor logistics: Moving between workstations and the parts issue store

representation of consumption levels, from low to high demand periods. An average value of 1500 units aligns with typical consumption rates observed in the industry. This approach also addresses the challenge of data scarcity, as obtaining real-time consumption data from aerospace manufacturing can be difficult and costly. Additionally, by generating repeatable and controllable data, we can run multiple experiments to verify the stability and consistency of our results. This method is not only cost-efficient but also saves time, enabling faster iterations and refinements in the analysis. Including 10% anonymous data accounts for possible recording errors or incomplete data collection, enhancing the dataset's realism.

Table 6.1 Parameter when generating simulate inventory consumption data

Description	**Parameters**
Standard Deviation	43
Range	500-2500
Average value	1500
Anonymous data	10%
Number of data generated	43200

The dataset records the consumption of resources over time, measured in minutes across multiple days. Each row in the dataset represents the consumption value at a specific minute of a specific day. The columns in the dataset are as follows:

- **Day**: The day on which the consumption is recorded.
- **Minute**: The minute within the day when the consumption is recorded.
- **Consumption**: The amount of resources consumed at that specific minute.

Below is a sample table from the dataset, showing the first few rows and the last few rows, as shown in Table 6.2.

Table 6.2 Sample consumption data

Day	Minute	Consumption
1	1	0
1	2	7
1	...	...
1	1439	0
1	1440	0
...	...	...
10	1	2
10	2	0
10	...	...
10	1439	3
10	1440	0

In order to evaluate the predictive performance, consumption data spanning 30 days was generated, with parameters as illustrated in Table 6.1. Additionally, using the same parameters, data for the subsequent 7 days was generated. To compare the accuracy of different models, past data was utilised for predictions while future data was employed to assess the predictive performance. It was observed that with the progression of time, the predictive ability of the models declined, showcasing a trend towards regression to the mean.

Based on consumption data generated above. Choice two typical inventory algorithm to predict the inventory consumption. The Autoregressive Integrated Moving Average Model (ARIMA) model consistently predicts values that are relatively stable, ranging from approximately 1442.09 to 1646.23 units shown in Table 6.3 and has performed reasonably well in its first 4 days compared to others shown. This model shows a smooth forecast trend with minimal variation, except for a slight increase in the first day's forecast. However, it struggles to capture significant deviations observed in the actual data, such as on the fifth day where the real consumption spikes to 2350 units while ARIMA's forecast remains steady at around 1485.64 units.

On the other hand, the Autoregressive Model presents more variability in its predictions, ranging from 1323.33 to 1652.57 units. This model demonstrates slightly more responsiveness to changes compared to the ARIMA model, but still fails to accurately predict the significant spike on the fifth day, though it does come closer than ARIMA on several days.

Both models show limitations in accurately tracking the real data's volatility, particularly evident on days where there are large swings in inventory consumption. The actual consumption data shows significant fluctuations that neither model adequately captures, indicating a potential gap in the forecasting ability of these classical models.

Table 6.3 Typical forecasting models for inventory consumption

Day	ARIMA	Autoregressive Model	Future Real data
1	1646.23	1652.57	1560
2	1442.09	1366.12	1400
3	1496.178	1323.33	1510
4	1481.847	1507.237	1650
5	1485.644	1443.61	2350
6	1484.638	1460.10	1460
7	1484.834	1477.41	1450
8	1484.852	1468.18	1480
9	1484.847	1472.84	1600
10	1484.847	1473.83	1450

Consumption As assembly progresses, components are gradually depleted. The type and quantity of consumption depend on the nature of the assembly process.

Replenishment: Replenishment is orchestrated based on past consumption estimates and the expertise of the workforce. Primarily, two methods are employed; one is time-based, while the other is event-driven. Replenishment occurs at set intervals or when a particular event triggers the need for restocking, ensuring the quality and availability of components as depicted in Figure 6.3.

The current inventory level is denoted by the equation:

$$\text{Current Inventory} = \text{Initial Inventory} + \text{Replenishment} - \text{Consumption} \quad (6.1)$$

Through the aforementioned formula (see Equation 6.1), it can be discerned that, holding other parameters constant, there exists a linear relationship between the consumption and replenishment rates of inventory. Within a production workshop setting, the consumption rate is often perceived as a proxy for productivity - an element deemed favourable when maximised. However, production activities inherently exhibit fluctuations, thus necessitating the dynamic adjustment of replenishment rates to adeptly match the prevailing consumption rates. Such adjustments are pivotal to ensuring that the curves representing cost and material volumes perpetually linger at their nadir, thereby engendering a reduction in operational costs.

Should the current inventory plummet below a certain threshold (not necessarily zero, for instance, a stockout might occur even if there are 10 units left) due to its inadequacy in manufacturing a particular part, a stockout event occurs. Such events deter production efficiency and induce additional losses. Hence, factories endeavour to forecast future demand to pre-emptively address replenishment needs. Various forecasting methods have been delineated in the previous literature review, primarily relying on historical data for predictions. However, as the nature of material consumption begins to shift, the accuracy of these methods dramatically dwindles. Therefore, investigating a model that requires fewer parameters, operates without reliance on extensive high-quality data, exhibits strong interpretability, demands minimal computational resources, and can dynamically refine replenishment strategies based on real-time inventory monitoring, holds promise.

6.2.2 Similarity between RFM concept sales prediction and parts inventory consumption

In the retail sector, a common practice to gauge and respond to consumer behaviour hinges on the triad of recency, frequency, and monetary (RFM) sequential patterns [99, 100]. This metric serves to capture the essence of consumer purchasing behaviour, subsequently guiding inventory and supply chain decisions. However, transplanting this approach to a manufacturing scenario, particularly that of aircraft production, reveals the inadequacy of the 'Money' component. This shortcoming stems mainly from the insensitivity to material values in the manufacturing environment, rendering the 'Money' metric less informative and therefore unsuitable.

The model based on QRF is more related to the characteristics of part consumption, and its advantage is that it only requires understanding through three parameters to grasp the potential material consumption trends in the future. To circumvent these challenges and stride towards a more robust and responsive supply chain model, this thesis endeavors to develop a QRF model using a Python project within the Google Colab environment.

In the previous review chapter, equation 2.1 demonstrated how to integrate the QRF model into the sales prediction system. In the QRF model, Recency (R) typically denotes the time interval between the occurrence of the most recent event and the present moment. We have amended R to $1/R$ as shown in equation

Table 6.4 Parameters and assumptions used in the simulation experiment

Description	**Parameters**
q_{weight}	0.5
f_{weight}	0.3
r_{weight}	0.2
Total hours	(240 hours, or 10 days)
Initial restocking threshold	50
Parts initial inventory	80
Out-of-stock time for each event	30
Transport time	10

6.2. In certain scenarios, a lower Recency value (i.e., the event occurred more recently) is often considered to be more valuable or significant as it may signify a recent trend or behaviour.

For instance, within Customer Relationship Management (CRM), a customer's recent purchase might be more indicative of the current interest in a product or service than purchases made earlier. Similarly, a component that has been recently in production is likely to have a higher probability of production than one that has not been produced for a long time. Thus, a lower Recency value is preferable in such instances. Consequently, its value increases as Recency decreases, aligning with the direct usage of Quantity (Q) and Frequency (F), where higher values usually represent a stronger trend or a higher level of consumption. This ensures all three indicators (Q, F, and R) move in the same direction (value increase) to represent a positive impact. Simultaneously, this design facilitates a more intuitive and consistent application of weights, aiding in the comprehension and adjustment of the model.

The above reiteration elucidates the strategic alteration of the Recency parameter to 1/R in the QRF model, bringing about a conceptual consistency across the Q, F, and R dimensions. This adjustment not only harmonizes the directional implications of all three metrics but also enhances the intuitive accessibility and ease of model manipulation, which is quintessential in practical applications such as CRM.

The expression of the equation becomes:

$$QRF = Q \cdot q_{\text{weight}} + F \cdot f_{\text{weight}} + \left(\frac{1}{R}\right) \cdot r_{\text{weight}} \tag{6.2}$$

where:

- Q is the total consumption over the recent period,
- F is the ratio of unique time points at which consumption occurred to the total time points,
- R is the time since the most recent consumption event.

The purpose of this formula was to show how to incorporate the QRF model into the replenishment system. In the equation (6.2), the total consumption (Quantity) over a recent period (e.g., the last 60 minutes), the ratio of the unique time points at which consumption occurred to the total time points (Frequency), and the time since the most recent consumption event (Recency) were calculated. These values were then multiplied by the provided weight coefficients (q_{weight}, f_{weight}, r_{weight}) and summed to obtain a QRF score.

6.3 Experiment result and Discussion

In evaluating the efficacy of factory logistics, it is imperative to ensure timely delivery of the accurate quantity to the designated location; any deviation from these parameters could precipitate a decline in operational efficiency. The criteria are paramount to ensuring an efficient inventory management system. A pivotal aspect of this evaluation hinges on the minimization of 'out of stock' instances, paired with the maintenance of lower inventory levels when made same amount of product.

Historical data from various industries, particularly aerospace, demonstrate that the quantity of consumed materials is the most direct indicator of demand. Higher quantities consistently reflect higher consumption needs, making this factor more important [68]. Thus, assigning a weight of 0.5 to the quantity factor q_{weight} ensures that the model is primarily driven by actual consumption levels, which are the most immediate and impactful variable.

The importance of frequency f_{weight} lies in its ability to capture consumption patterns. Frequency measures how often parts are used, providing insights into the regularity and consistency of demand. A weight of 0.3 balances its influence, acknowledging that while frequency is important, it is slightly less critical than the sheer volume of consumption. Frequent usage patterns can indicate stable and predictable demand, aiding in more reliable forecasting and reducing the likelihood of stockouts overstocking.

The relevance of recency r_{weight} reflects the impact of recent consumption events on current demand. Weighted at 0.2, recency captures short-term fluctuations and

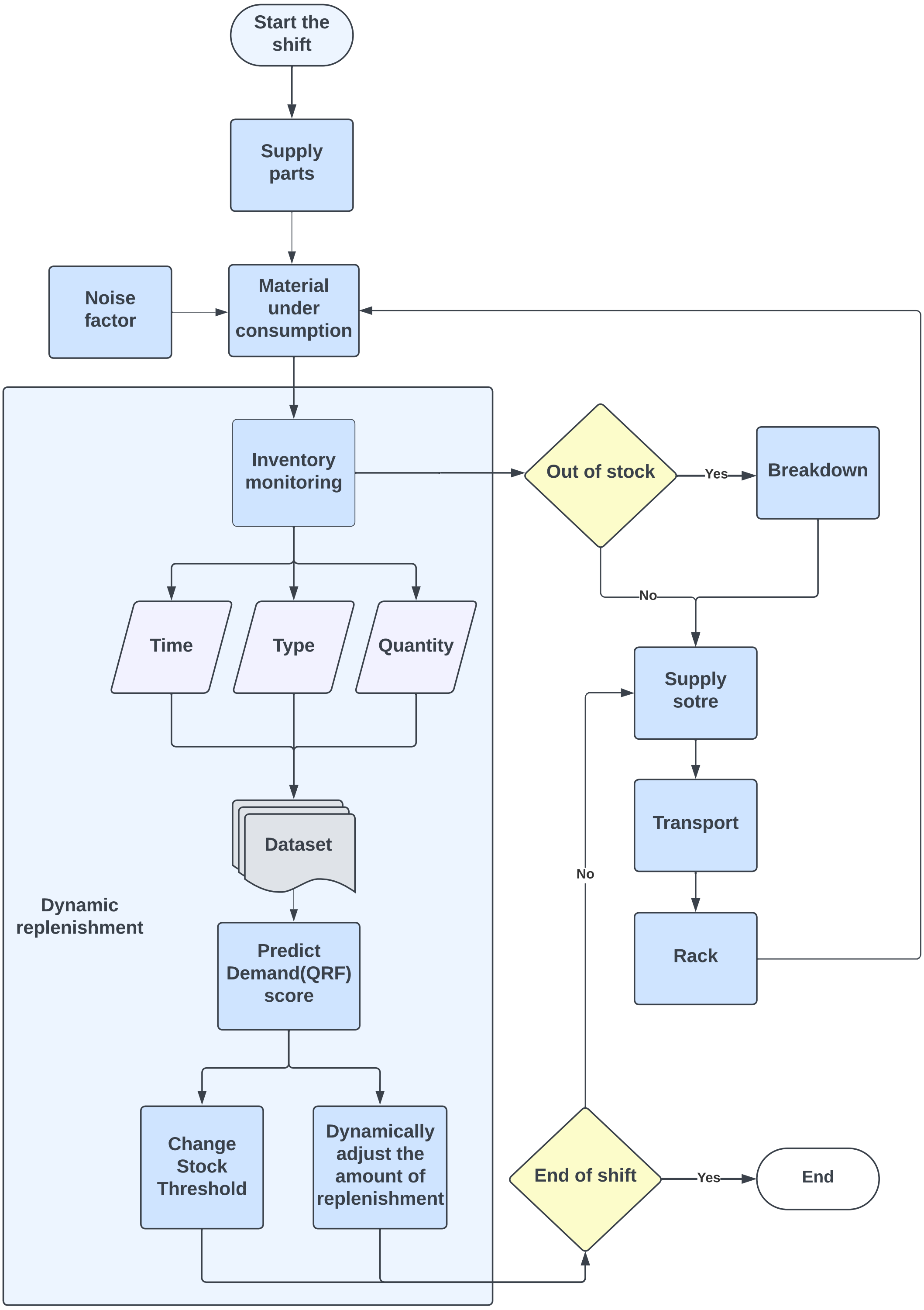


Fig. 6.5 Conceptual model for material flow Dynamics: Consumption and replenishment on the Factory Floor Using the QRF model

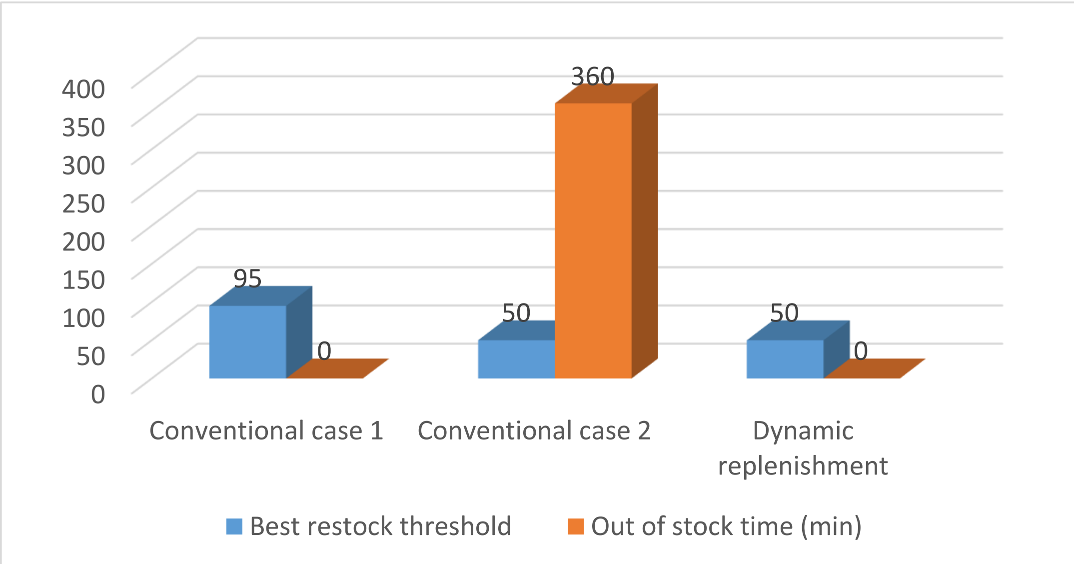


Fig. 6.6 Logistics performance: Dynamic replenishment based on QRF model implementation versus conventional replenishment policy

recent trends, ensuring the model can adapt to sudden changes in consumption. Its lower weight compared to quantity and frequency ensures that the model remains stable and not overly reactive to short-term variations, which could lead to excessive adjustments and inefficiencies.

In the experiment conducted, the results were obtained under the constant parameters specified in Table 6.1 and are illustrated in Figure 6.6. Over a span of 10 days, while processing 15,910 parts, it was observed that if the 'out of stock' time remains consistent, the inventory threshold required by traditional methods is higher. Conversely, with an equivalent required inventory, an out-of-stock situation arises, lasting for 360 minutes. This implies that within this simulated environment, downtime has occurred using traditional methods. However, according to the dynamics replenishment method proposed by the QRF model, a lower downtime can be achieved with the same inventory threshold, or a lower inventory threshold can be maintained with the same downtime.

6.4 Summary

In this chapter, common predictive methods were examined and it was discovered that they have a tendency to inadequately adapt to the variations in part consumption over time. Consequently, a model anchored on Quantity, Recency, and Frequency (QRF) was formulated, which perpetually fine-tunes the quantity of replenishment and inventory levels based on the data input, aligning with real-time consumption whilst averting stockouts and excessive inventory scenarios. The experimental findings illustrated that should the duration of "stockouts" remain unchanged, the inventory requisites would escalate under traditional methods when juxtaposed with the QRF model. On the flip side, given identical requisite inventory levels, stockout occurrences would materialize under conventional replenishment approaches.

Chapter 7

Discussion and Conclusion

7.1 Research Contributions

The contributions of this research are delineated in the subsequent sections, categorised according to both the project objectives and their corresponding chapters.

Contribution 1: A Novel Inventory Data Collection Method for Shop Floor Rack In Chapter 3, the significance of choosing a reliable sensor, which forms the cornerstone of any detection process, is explored. This chapter discusses the suitability of various sensors for detecting parts within a flow rack in a factory environment and examines the impact of different external influences. The chosen detection method predominantly utilises weight and visual sensors, ensuring that weight detection is unaffected by the diversity of objects, whilst vision detection remains uninhibited by obstructions.

Contribution 2: A Mathematically-Based Approach to Sensor Fusion Algorithm Sensor fusion lays the foundation for leveraging the complementary advantages of multiple sensors. By attaining information from within the shelves through Objective 1, a fusion method for decision-level information based on integer programming and a noisy channel model is introduced in Chapter 4 for Objective 2. This method necessitates preliminary preparation to extract features. By training the pre-trained model EfficientNet, vision results with noise are acquired. Through meticulous probability calculations, the combinations of integer programming are matched, achieving decision-level fusion. The accuracy of this method is validated within the defined four categories and up to sixteen combinations.

Contribution 3: A Deep Learning Method Utilising Sensor Fusion to Reduce Reliance on Prior Knowledge and Human Feature Extraction, Thereby Achieving Higher Accuracy In Chapter 4, a substantial amount of prior knowledge and reasoning processes were necessary, which could be time-consuming in practical situations. Implementing this in industrial contexts could prove challenging. Therefore, the study in Chapter 5 utilised a deep learning-based fusion, enabling the computer to autonomously identify and extract features, thereby simplifying the process. By comparing the FEI-DEO and DEI-DEO methods, accuracies of 86% and 98% were respectively achieved, which not only surpass the accuracy in [53] without the need for location information but also present a more feasible implementation.

Development of the Quantity, Recency, and Frequency (QRF) Concept Model for Enhancing the Response to Consumption Variations in Replenishment Policies In previous chapters, we introduced a method for automated inventory detection. With this inventory information, it is possible to better understand the characteristics of parts consumption, thereby improving capture and, consequently, enhancing the performance of internal logistics. Specifically, a model based on Quantity, Recency, and Frequency (QRF) was developed to measure the likelihood of particular parts consumption. This model allows for continuous adjustment of replenishment and inventory levels according to the input data, to match real-time consumption while preventing stockouts and overstock situations. Experimental results demonstrate that under certain conditions, this approach can reduce stockout events and decrease the level of inventory required.

7.2 Limitation of research

The research was conducted through laboratory-based experiments, with no real-world, industrial-grade experimentation. Detection accuracy is largely contingent upon set constraints and boundaries, and the variety and quantity of objects considered are limited. Additionally, factors such as sensor communication, power supply stability, etc., which are pertinent in an actual factory setting, are not addressed. It has been observed that unstable currents can influence detection accuracy. Further limitations for specific chapters are detailed at the end of each section. Beyond the above, there are limitations to each objective:

7.2.1 Limitation for Objective 1

Given our experiments are grounded in a simulated environment, certain scenarios may fall short in mirroring the genuine flow rack dynamics. Our visual sensors are installed at the exit of each column, enabling us to obtain material information from the outermost boxes. However, data regarding previous boxes remain inaccessible unless included in the system during replenishment. To address this limitation, an alternative approach has been proposed: the information flow chart of the complete design system, as depicted in Figure 3.8.

Cameras are strategically placed around the rack to capture the human pose of the operator alongside the objects within the rack. They can mounted at both the entrance and the top exit of the rack. The cameras on the entrance side are tasked with recording the types and quantities of objects placed in the rack as well as the shift time, while the visual system at the exit aims to document the items selected by the operator. The camera at the entrance plays a crucial role in gathering replenishment information and recording the replenishment time and the quantity of items added to the rack.

7.2.2 Limitation for Objective 2

The study delineates several limitations concerning the application of the Noisy Channel Model (NCM) to combinatorial optimisation problems. Primarily, there's a notable dependency on prior knowledge since Bayesian approaches intrinsic to the NCM hinge on prior probability distributions. This dependency could potentially constrain the model's versatility and adaptability. Additionally, the computational overhead incurred due to the necessity of estimating multiple probability distributions cannot be overlooked. This overhead markedly escalates in scenarios encompassing vast combinatorial, thus potentially curbing the model's efficiency and practicality.

Furthermore, the accuracy and stability of the NCM are contingent upon the precision of the model definition and the appropriateness of the chosen probability distributions. Erroneous selections in this regard can culminate in unstable or imprecise outcomes, thereby undermining the model's reliability. Moreover, the NCM might manifest as an over-engineered solution for certain problems. Lastly, the task of assigning probabilities, particularly through actual situations and observations, might introduce bias, thereby further impinging on the model's efficacy and objectivity.

7.2.3 Limitation for Objective 3

In Chapter 5, we delved into the intricacies of feature-based and decision-level fusion, the constraints faced in the FEI-DEO fusion model which is impeded from identifying conditions beyond its trained labels. To circumvent this, a novel method was introduced, whereby, given a fully-populated box akin to the training module, the remaining contents within the box can be real-time identified through a weight subtraction process, represented mathematically.

In assessing the DEI-DEO model, a notable limitation emerges pertaining to its operational constraints as delineated by Equation (4.1). Despite its higher accuracy, the model is bound to a restrictive framework where the variety of combinatory operations is limited. Moreover, a mandatory subtraction operation ensues upon material consumption, further exemplifying the model's bounded flexibility. This rigidity could potentially hinder the applicability of DEI-DEO in more complex or dynamic scenarios, warranting consideration for future model adaptations.

7.2.4 Limitation for Objective 4

The QRF model's effectiveness depends heavily on accurate and complete historical data, which can be difficult to obtain in real-world scenarios. The model's fixed parameter weights may not be universally applicable, limiting its adaptability. Simplistic assumptions in the simulation might not reflect real-world complexities, and the model's response to sudden demand changes might lag. While computationally efficient, scaling to larger datasets remains a challenge. Comparisons with advanced predictive methods are limited, and practical implementation could face resistance and complexity.

7.3 Future work

Future work according to each object can be listed as follows:

- **Objective 1**

  In the preliminary investigations conducted within a controlled laboratory environment, web cameras were utilised to gather data from the simulated shelves; however, these experiments did not extend to a real flow rack setting. Consequently, a significant avenue for future research is to transition this experimental framework to a real-world workshop setting, specifically targeting the shelves for comprehensive testing. Moreover, enhancing the precision of positional identification through the integration of gesture recognition technologies such as [145], alongside the methodologies delineated in Figure 3.8 human pose estimation branch, stands as a subsequent priority. Such advancements are projected to augment the accuracy of data acquisition significantly, thereby contributing a robust foundation for the ensuing phases of this research project.

In the future, Camera technology can employed to ascertain human pose estimations, facilitating the identification of pick-up activities and product types. Research may delve into leveraging the vision system to discern human pose for pinpointing hand positions, aiming at enhancing object estimation for more precise inventory monitoring. While this technology remains unexplored within the industrial realm, analogous applications have been observed in supermarket settings. Supermarket merchandise typically bears distinct textual imaging, whereas industrial components exhibit unique shapes. A commonality resides in the generally fixed positions of items.

- **Objective 2**

  The range of conditions tested herein is limited, thereby demonstrating certain constraints. There's scope for expanding to a more diverse array of detection categories in future endeavours. Furthermore, through the meticulous research and advancement of more generic and standardised data fusion algorithms, the technical exchange and development across a myriad of applications and scenarios can be significantly promoted.

- **Objective 3**

  In Chapter 5, it is articulated that the disparate sensors are not amalgamated in real-time, signifying that each detection necessitates a form of activation mechanism. This proves to be cumbersome in real-world production activities. A possible avenue for research could be to investigate a dynamic inspection mode, though this proposition also presents a challenge concerning computational load, necessitating empirical validation within real-world production scenarios. On the other hand, fusion at the feature level incurs a certain degree of feature loss. The quest for a method that incurs lesser feature loss is of academic merit, as it could potentially lead to heightened accuracy.

- **Objective 4**

Future research may delve into and juxtapose various predictive methodologies, including artificial intelligence, to discern the most precise method for forecasting sales quantities. This aspect is notably relevant when addressing distinct data patterns, as exemplified in the case study delineated in this research. Generally, not only is component supply forecasting requisite in aircraft manufacturing, but other industries can also utilise this model for dynamic amelioration of replenishment strategies. Subsequently, there's potential to explore additional parameters in the future, thereby broadening the model's applicability.

7.4 Conclusions

Sensor-based automated inventory monitoring has brought numerous benefits to many manufacturing and retail businesses. These benefits include reduced manual errors, optimised inventory levels, increased customer satisfaction, real-time data analysis, lowered inventory costs, faster response times, and a decrease in waste. For industries and enterprises, automated monitoring is a highly valuable investment. This is evidenced by the growing number of industries and businesses undergoing transformations, a trend which continues to rise.

Previous literature has predominantly focused on using single sensors for detecting individual types of objects. However, with the rise of intelligent manufacturing and the emergence of a variety of personalised custom products, it becomes meaningful to store a set of parts in one compartment, which can be assembled into a component within limited storage space.

However, there are limitations to using a single sensor. A solitary sensor can easily be disrupted and fail to capture comprehensive and valid information. By capitalising on the complementary strengths of multiple sensors, these challenges can be addressed. As such, a mathematical method for multi-sensor fusion has been proposed. These conventional methods are transparent and highly interpretable. Their behaviour is predictable, which makes their principles and outputs easier to understand. Additionally, they are cost-effective computationally and do not require large datasets. Yet, while traditional fusion methods address the limitations of single sensors, they also introduce issues like manual feature engineering and computational complexity. As an increasing amount of sensor data has flooded in recent years, using traditional fusion methods has become challenging. Conversely, deep learning offers significant advantages in this domain, and as a result, it has garnered immense industry attention in recent times.

From the research presented in this dissertation, it is evident that both traditional and deep learning approaches can address the fusion of weight and visual sensors. Notably, no prior literature has mentioned using sensor fusion for the classification and counting of multiple objects across various categories.

The traditional approach, using the noisy channel model, corrects in a manner akin to spelling correction. The weight sensor generates a series of possible combinations through prior knowledge and integer rules. This is then matched with the most probable set based on past visual sensor experimental probabilities, aiming to achieve a complementary data fusion of the two. Experimental results indicate that satisfactory precision and accuracy levels can be achieved with certain prior knowledge. One of the strengths of this method is its strong interpretability; each step can be understood, and there's no need for vast computational resources or datasets. This approach has not been attempted in prior literature.

In contrast, the deep learning approach proposed in this paper introduces two distinct deep learning techniques: the FEI-DEO and DEI-DFO methods. Both can self-extract features through the model to obtain classification results. More specifically, both methods utilise state-of-the-art image recognition models for feature extraction. The FEI-DEO method converts weight into a noise map, ensuring a consistent model output dimension before entering a convolutional network to produce classification results. The DEI-DEO method guesses all possibilities based on historical data, enabling a new prediction. When compared, the experiments demonstrate that the deep learning DEI-DEO method boasts the highest accuracy.

The last, highlighted the shortcomings of conventional predictive methods in adapting to variations in part consumption over time, leading to the development of a Quantity, Recency, and Frequency (QRF) model. This model continuously adjusts replenishment and inventory levels according to real-time data, efficiently mitigating stockouts and excessive inventory scenarios. Experimental results demonstrated that, with constant stockout durations, traditional methods increase inventory requirements compared to the QRF model. Conversely, at identical inventory levels, stockouts were more prevalent with conventional replenishment approaches.

a consumption assessment methodology based on Quantity, Reliability, Frequency has been developed. This approach harnesses real-time data to predict future consumption rates, enabling a swift response to variations in consumption. This reduces the reliance on large amounts of high-quality data upfront increases flexibility and reduces computational consumption.

Each method discussed has its advantages and limitations. The constraints of each approach have been deliberated upon in this thesis. However, considering the predetermined scenarios, types, and quantities of materials we've set, these limitations are controllable.

This work can aid the manufacturing sector in more rapidly and accurately counting mixed goods and can provide dependable inventory levels. As future information capture methods expand, such as using human posture to assist in locating item positions, the potential of deep learning models can be further extended by integrating more cameras and information sources.

References

[1] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & information systems engineering*, vol. 6, pp. 239–242, 2014.

[2] A. Rojko, "Industry 4.0 concept: Background and overview.," *International journal of interactive mobile technologies*, vol. 11, no. 5, 2017.

[3] L. Barreto, A. Amaral, and T. Pereira, "Industry 4.0 implications in logistics: an overview," *Procedia manufacturing*, vol. 13, pp. 1245–1252, 2017.

[4] R. Weinstein, "Rfid: a technical overview and its application to the enterprise," *IT professional*, vol. 7, no. 3, pp. 27–33, 2005.

[5] B. Tjahjono, C. Esplugues, E. Ares, and G. Pelaez, "What does industry 4.0 mean to supply chain?," *Procedia Manufacturing*, vol. 13, pp. 1175–1182, 2017. Manufacturing Engineering Society International Conference 2017, MESIC 2017, 28-30 June 2017, Vigo (Pontevedra), Spain.

[6] B. B. Gupta and M. Quamara, "An overview of internet of things (iot): Architectural aspects, challenges, and protocols," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 21, p. e4946, 2020.

[7] S. Madakam, V. Lake, V. Lake, V. Lake, *et al.*, "Internet of things (iot): A literature review," *Journal of Computer and Communications*, vol. 3, no. 05, p. 164, 2015.

[8] B. Ives, K. Cossick, and D. Adams, "Amazon go: disrupting retail?," *Journal of Information Technology Teaching Cases*, vol. 9, no. 1, pp. 2–12, 2019.

[9] A. Polacco and K. Backes, "The amazon go concept: Implications, applications, and sustainability," *Journal of Business and Management*, vol. 24, no. 1, pp. 79–92, 2018.

[10] C. Ruiz, J. Falcao, S. Pan, H. Y. Noh, and P. Zhang, "Autonomous inventory monitoring through multi-modal sensing (aim3s) for cashier-less stores," in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 395–396, 2019.

[11] T. W. Gruen, D. S. Corsten, and S. Bharadwaj, "Retail out-of-stocks: A worldwide examination of extent," *Causes, and Consumer Responses, Grocery Manufacturers of America, Washington DC*, 2002.

[12] P. Reyes and M. S. Raisinghani, "Integrating information technologies and knowledge-based systems: a theoretical approach in action for enhancements in production and inventory control," *Knowledge and Process Management*, vol. 9, no. 4, pp. 256–263, 2002.

[13] C. Xia, Z. Pan, J. Polden, H. Li, Y. Xu, S. Chen, and Y. Zhang, "A review on wire arc additive manufacturing: Monitoring, control and a framework of automated system," *Journal of manufacturing systems*, vol. 57, pp. 31–45, 2020.

[14] G. Priniotakis and P. Argyropoulos, "Inventory management concepts and techniques," in *IOP conference series: Materials Science and Engineering*, vol. 459, p. 012060, IOP Publishing, 2018.

[15] S. Aro-Gordon and J. Gupte, "Review of modern inventory management techniques," *Global Journal of Business & Management*, vol. 1, no. 2, pp. 1–22, 2016.

[16] H. H.-C. Chuang, "Mathematical modeling and bayesian estimation for error-prone retail shelf audits," *Decision Support Systems*, vol. 80, pp. 72–82, 2015.

[17] R. Ballard, "Methods of inventory monitoring and measurement," *Logistics information management*, vol. 9, no. 3, pp. 11–18, 1996.

[18] K. Murao, J. Imai, T. Terada, and M. Tsukamoto, "Recognizing activities and identifying users based on tabletop activities with load cells," in *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, pp. 1–6, 2015.

[19] K. Murao, J. Imai, T. Terada, and M. Tsukamoto, "Activity recognition and user identification based on tabletop activities with load cells," *Journal of Information Processing*, vol. 25, pp. 59–66, 2017.

[20] G. Hossain, I. Z. Hossain, and G. Grabher, "Piezoresistive smart-textile sensor for inventory management record," *Sensors and Actuators A: Physical*, vol. 315, p. 112300, 2020.

[21] C. Han and T. Naemura, "Bumpmarker: a 3d-printed tangible marker for simultaneous tagging, tracking, and weight measurement," *ITE Transactions on Media Technology and Applications*, vol. 7, no. 1, pp. 11–19, 2019.

[22] M. Sergio, N. Manaresi, M. Tartagni, R. Guerrieri, and R. Canegallo, "A textile based capacitive pressure sensor," in *SENSORS, 2002 IEEE*, vol. 2, pp. 1625–1630, IEEE, 2002.

[23] K. Murao, J. Imai, T. Terada, and M. Tsukamoto, "Recognizing activities and identifying users based on tabletop activities with load cells," in *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, pp. 1–6, 2015.

[24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pp. 740–755, Springer, 2014.

[25] C. B. Green, "Weight sensing apparatus and methods for inventory control," Apr. 14 2009. US Patent 7,518,068.

[26] P. Godlewski, "Automated inventory system and method," Jan. 19 2010. US Patent 7,650,298.

[27] C. Metzger, "Indirect object-sensing technology to prevent out-of-stock at retail-level," in *Proceedings of the Workshop on Smart Object at the International Conference on Ubiquitous Computing, Tokyo*, Citeseer, 2005.

[28] C. Metzger, J. Meyer, E. Fleisch, and G. Tröster, "Weight-sensitive foam to monitor product availability on retail shelves," in *Pervasive Computing: 5th International Conference, PERVASIVE 2007, Toronto, Canada, May 13-16, 2007. Proceedings 5*, pp. 268–279, Springer, 2007.

[29] R. Kamoda, M. Ueda, T. Funatomi, M. Iiyama, and M. Minoh, "Grocery re-identification using load balance feature on the shelf for monitoring grocery inventory," in *Proceedings of the Cooking with Computers workshop (CwC)*, pp. 8–18, 2012.

[30] A. Schmidt, M. Strohbach, K. Van Laerhoven, A. Friday, and H.-W. Gellersen, "Context acquisition based on load sensing," in *UbiComp 2002: Ubiquitous Computing: 4th International Conference Göteborg, Sweden, September 29–October 1, 2002 Proceedings 4*, pp. 333–350, Springer, 2002.

[31] B. Nagaria, P. Shroff, and R. Mehrotra, "Iot based inventory system for stock management," *Int. Res. J. Eng. Technol*, vol. 6, no. 4, pp. 4094–4097, 2019.

[32] V. Prithvish, S. Agrawal, and J. S. R. Alex, "An iot-based smart shopping cart using the contemporary barcode scanner," in *Intelligent Embedded Systems: Select Proceedings of ICNETS2, Volume II*, pp. 45–58, Springer, 2018.

[33] J. Velasco, L. Alberto, H. D. Ambatali, M. Canilang, V. Daria, J. B. Liwanag, and G. A. Madrigal, "Internet of things-based (iot) inventory monitoring refrigerator using arduino sensor network," *arXiv preprint arXiv:1911.11265*, 2019.

[34] J. Janai, F. Güney, A. Behl, A. Geiger, *et al.*, "Computer vision for autonomous vehicles: Problems, datasets and state of the art," *Foundations and Trends® in Computer Graphics and Vision*, vol. 12, no. 1–3, pp. 1–308, 2020.

[35] Ø. Sture, E. R. Øye, A. Skavhaug, and J. R. Mathiassen, "A 3d machine vision system for quality grading of atlantic salmon," *Computers and Electronics in Agriculture*, vol. 123, pp. 142–148, 2016.

[36] J. Brownlee, *Deep learning for computer vision: image classification, object detection, and face recognition in python.* Machine Learning Mastery, 2019.

[37] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[38] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

[39] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[40] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[41] M. Sieb, Z. Xian, A. Huang, O. Kroemer, and K. Fragkiadaki, "Graph-structured visual imitation," in *Conference on Robot Learning*, pp. 979–989, PMLR, 2020.

[42] D. Markus and E. Stephan, "Rfid-kommunikationssystem zur elektronischen bestandsüberwachung eines durchlauflagers," 2018. EP 3276552 A1, Germany, European Patent Application.

[43] R. Kumar and U. K. Wiil, *Recent Advances in Computational Intelligence*, vol. 823. Springer, 2019.

[44] B. S. S. Tejesh and S. Neeraja, "Warehouse inventory management system using iot and open source framework," *Alexandria engineering journal*, vol. 57, no. 4, pp. 3817–3823, 2018.

[45] Y. Yuan and D. Yu, "Uhf rfid shelf solution with cascaded reader antenna and positioning capability," in *2012 IEEE International Conference on RFID (RFID)*, pp. 149–156, IEEE, 2012.

[46] F. Thiesse, E. Fleisch, C. Sorensen, and C. Tellkamp, "Rfid-enabled shelf replenishment with case-level tagging: a simulation study," *AMCIS 2007 Proceedings*, p. 112, 2007.

[47] J. J. Lamas-Seco, P. M. Castro, A. Dapena, and F. J. Vazquez-Araujo, "Sidivs: Simple detection of inductive vehicle signatures with a multiplex resonant sensor," *Sensors*, vol. 16, no. 8, p. 1309, 2016.

[48] I. Ihara, "Ultrasonic sensing: fundamentals and its applications to nondestructive evaluation," *Sensors: Advancements in Modeling, Design Issues, Fabrication and Practical Applications*, pp. 287–305, 2008.

[49] K. Singh, "Smart inventory management using electronic sensor based computational intelligence," *Recent Advances in Computational Intelligence*, pp. 3–25, 2019.

[50] J. L. Torralbo-Muñoz, S. Sendra, L. Parra, and J. Lloret, "Smartfridge: The intelligent system that controls your fridge," in *2018 Fifth International Conference on Internet of Things: Systems, Management and Security*, pp. 200–207, IEEE, 2018.

[51] H. Nasir, W. B. W. Aziz, F. Ali, K. Kadir, and S. Khan, "The implementation of iot based smart refrigerator system," in *2018 2nd International Conference on Smart Sensors and Application (ICSSA)*, pp. 48–52, IEEE, 2018.

[52] T. Kwon, E. Park, and H. Chang, "Smart refrigerator for healthcare using food image classification," in *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 483–484, 2016.

[53] C. Ruiz, J. Falcao, S. Pan, H. Y. Noh, and P. Zhang, "Aim3s: Autonomous inventory monitoring through multi-modal sensing for cashier-less convenience stores," in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 135–144, 2019.

[54] B. Ives, K. Cossick, and D. Adams, "Amazon go: disrupting retail?," *Journal of Information Technology Teaching Cases*, vol. 9, no. 1, pp. 2–12, 2019.

[55] C. Ruiz, J. Falcao, S. Pan, H. Y. Noh, and P. Zhang, "Autonomous inventory monitoring through multi-modal sensing (aim3s) for cashier-less stores," in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 395–396, 2019.

[56] W. Wang, J. Yang, L. Huang, D. Proverbs, and J. Wei, "Intelligent storage location allocation with multiple objectives for flood control materials," *Water*, vol. 11, no. 8, p. 1537, 2019.

[57] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman, "Comparative analysis of deep learning image detection algorithms," *Journal of Big data*, vol. 8, no. 1, pp. 1–27, 2021.

[58] R. Chauhan, K. K. Ghanshala, and R. Joshi, "Convolutional neural network (cnn) for image detection and recognition," in *2018 first international conference on secure cyber computing and communication (ICSCCC)*, pp. 278–282, IEEE, 2018.

[59] J. Du, "Understanding of object detection based on cnn family and yolo," in *Journal of Physics: Conference Series*, vol. 1004, p. 012029, IOP Publishing, 2018.

[60] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341–1360, 2020.

[61] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022.

[62] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, pp. 6105–6114, PMLR, 2019.

[63] E. Tekin, "A method for traceability and "as-built product structure" in aerospace industry," *Procedia CIRP*, vol. 17, pp. 351–355, 2014.

[64] P. K. Saha, *Aerospace manufacturing processes.* CRC Press, 2016.

[65] T. Tahmina, M. Garcia, Z. Geng, and B. Bidanda, "A survey of smart manufacturing for high-mix low-volume production in defense and aerospace industries," in *International conference on flexible automation and intelligent manufacturing*, pp. 237–245, Springer, 2022.

[66] M. Helmold, A. Küçük Yılmaz, T. Dathe, and T. G. Flouris, "Scrm in the aviation industry:"risk management strategies to resilience retake-off"," in *Supply Chain Risk Management: Cases and Industry Insights*, pp. 179–220, Springer, 2022.

[67] A. Drouot, L. Irving, D. Sanderson, A. Smith, and S. Ratchev, "A transformable manufacturing concept for low-volume aerospace assembly," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 5712–5717, 2017.

[68] Y.-C. Chou, C.-H. Lu, and Y.-Y. Tang, "Identifying inventory problems in the aerospace industry using the theory of constraints," *International journal of production research*, vol. 50, no. 16, pp. 4686–4698, 2012.

[69] A. Rushton, P. Croucher, and P. Baker, *The handbook of logistics and distribution management: Understanding the supply chain.* Kogan Page Publishers, 2022.

[70] REB, "Flow through racking vs. push back racking: Which is idea for your foodservice operation?." Accessed: 2020.

[71] I. Mecalux, "Embracing cross-docking: 3 success stories," 2019. Accessed: 2020.

[72] M. MANUFACURING, "1 3/8" diameter gravity roller conveyor." Accessed: 2020.

[73] Constrofacilitator, "Warehouse storage systems and material handling equipments," 2022. Accessed: 2022.

[74] C. K. Chui, G. Chen, *et al.*, *Kalman filtering.* Springer, 2017.

[75] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan kaufmann, 1988.

[76] R. A. Ribeiro, A. Falcao, A. Mora, and J. M. Fonseca, "Fif: A fuzzy information fusion algorithm based on multi-criteria decision making," *Knowledge-Based Systems*, vol. 58, pp. 23–32, 2014.

[77] F. Castanedo *et al.*, "A review of data fusion techniques," *The scientific world journal*, vol. 2013, 2013.

[78] H. F. Durrant-Whyte, "Sensor models and multisensor integration," in *Autonomous robot vehicles*, pp. 73–89, Springer, 1990.

[79] B. V. Dasarathy, "Sensor fusion potential exploitation-innovative architectures and illustrative applications," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 24–38, 1997.

[80] S. Zhang, X. Meng, Q. Liu, G. Yang, and W. Sun, "Feature-decision level collaborative fusion network for hyperspectral and lidar classification," *Remote Sensing*, vol. 15, no. 17, 2023.

[81] V. Lekidis, M. Tsakiri, K. Makra, C. Karakostas, N. Klimis, and I. Sous, "Evaluation of dynamic response and local soil effects of the evripos cable-stayed bridge using multi-sensor monitoring systems," *Engineering Geology*, vol. 79, no. 1-2, pp. 43–59, 2005.

[82] A. Stief, J. R. Ottewill, J. Baranowski, and M. Orkisz, "A pca and two-stage bayesian sensor fusion approach for diagnosing electrical and mechanical faults in induction motors," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9510–9520, 2019.

[83] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, pp. 35–45, 03 1960.

[84] F. van Wyk, Y. Wang, A. Khojandi, and N. Masoud, "Real-time sensor anomaly detection and identification in automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1264–1276, 2020.

[85] J. O. Berger, *Statistical decision theory and Bayesian analysis.* Springer Science & Business Media, 2013.

[86] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis.* Chapman and Hall/CRC, 1995.

[87] K. Sentz and S. Ferson, "Combination of evidence in dempster-shafer theory," 2002.

[88] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 54, no. 4, pp. 1–40, 2021.

[89] V. Radu, C. Tong, S. Bhattacharya, N. D. Lane, C. Mascolo, M. K. Marina, and F. Kawsar, "Multimodal deep learning for activity and context recognition," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 1, no. 4, pp. 1–27, 2018.

[90] V. Radu, N. D. Lane, S. Bhattacharya, C. Mascolo, M. K. Marina, and F. Kawsar, "Towards multimodal deep learning for activity recognition on mobile devices," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pp. 185–188, 2016.

[91] N. Dawar and N. Kehtarnavaz, "A convolutional neural network-based sensor fusion system for monitoring transition movements in healthcare applications," in *2018 IEEE 14th International Conference on Control and Automation (ICCA)*, pp. 482–485, IEEE, 2018.

[92] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Information Fusion*, vol. 35, pp. 68–80, 2017.

[93] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pižurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2405–2418, 2014.

[94] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS journal of photogrammetry and remote sensing*, vol. 66, no. 3, pp. 247–259, 2011.

[95] T. Pfeifer and P. Protzel, "Robust sensor fusion with self-tuning mixture models," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3678–3685, IEEE, 2018.

[96] C. Chandra and J. Grabis, "Application of multi-steps forecasting for restraining the bullwhip effect and improving inventory performance under

autoregressive demand," *European Journal of Operational Research*, vol. 166, no. 2, pp. 337–350, 2005.

[97] S. E. Said and D. A. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika*, vol. 71, no. 3, pp. 599–607, 1984.

[98] C.-C. Liang, "Smart inventory management system of food-processing-and-distribution industry," *Procedia Computer Science*, vol. 17, pp. 373–378, 2013.

[99] Y.-L. Chen, M.-H. Kuo, S.-Y. Wu, and K. Tang, "Discovering recency, frequency, and monetary (rfm) sequential patterns from customers' purchasing data," *Electronic Commerce Research and Applications*, vol. 8, no. 5, pp. 241–251, 2009.

[100] R. Gustriansyah, D. I. Sensuse, and A. Ramadhan, "A sales prediction model adopted the recency-frequency-monetary concept," *Indones. J. Electr. Eng. Comput. Sci*, vol. 6, no. 3, pp. 711–720, 2017.

[101] G. R. Clay and F. Grange, "Evaluating forecasting algorithms and stocking level strategies using discrete-event simulation," in *Proceedings of the 29th conference on Winter simulation*, pp. 817–824, 1997.

[102] X. Guo, C. Liu, W. Xu, H. Yuan, and M. Wang, "A prediction-based inventory optimization using data mining models," in *2014 Seventh International Joint Conference on Computational Sciences and Optimization*, pp. 611–615, 2014.

[103] A. Hamedianfar, C. Mohamedou, A. Kangas, and J. Vauhkonen, "Deep learning for forest inventory and planning: a critical review on the remote sensing approaches so far and prospects for further applications," *Forestry*, vol. 95, no. 4, pp. 451–465, 2022.

[104] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[105] J. L. Elman, "Finding structure in time," *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.

[106] S. Hochreiter, "Untersuchungen zu dynamischen neuronalen netzen," *Diploma, Technische Universität München*, vol. 91, no. 1, p. 31, 1991.

[107] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[108] A. Graves, "Generating sequences with recurrent neural networks," *arXiv preprint arXiv:1308.0850*, 2013.

[109] C. Deng and Y. Liu, “A deep learning-based inventory management and demand prediction optimization method for anomaly detection,” *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, p. 9969357, 2021.

[110] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using rnn encoder-decoder for statistical machine translation,” *arXiv preprint arXiv:1406.1078*, 2014.

[111] X. Li, X. Ma, F. Xiao, F. Wang, and S. Zhang, “Application of gated recurrent unit (gru) neural network for smart batch production prediction,” *Energies*, vol. 13, no. 22, p. 6121, 2020.

[112] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[113] S. Li, H. Jin, Y. Xuan, X. Zhou, W. Wang, and Y. Yan, “Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting,” *arXiv preprint arXiv:1907.00235*, 2019.

[114] L. Wu, Q. Zhang, C.-H. Chen, K. Guo, and D. Wang, “Deep learning techniques for community detection in social networks,” *IEEE Access*, vol. 8, pp. 96016–96026, 2020.

[115] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, and J. Geng, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” *arXiv preprint arXiv:2012.07436*, 2021.

[116] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, “Dive into deep learning,” *arXiv preprint arXiv:2106.11342*, 2021.

[117] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature machine intelligence*, vol. 1, no. 5, pp. 206–215, 2019.

[118] Z. C. Lipton, “The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.,” *Queue*, vol. 16, no. 3, pp. 31–57, 2018.

[119] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, “Language models are few-shot learners,” *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[120] K. Jackson, K. Efthymiou, and J. Borton, “Digital manufacturing and flexible assembly technologies for reconfigurable aerospace production systems,” *Procedia Cirp*, vol. 52, pp. 274–279, 2016.

[121] V. J. Mawson and B. R. Hughes, “The development of modelling tools to improve energy efficiency in manufacturing processes and systems,” *Journal of Manufacturing Systems*, vol. 51, pp. 95–105, 2019.

[122] R. Chuawatcharin and N. Gerdsri, “Factors influencing the attitudes and behavioural intentions to use just walk out technology among bangkok consumers,” *International journal of public sector performance management*, vol. 5, no. 2, pp. 146–163, 2019.

[123] K. Wankhede, B. Wukkadada, and V. Nadar, “Just walk-out technology and its challenges: A case of amazon go,” in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 254–257, IEEE, 2018.

[124] T. Stolarski, “Mechanical engineering design: Je shigley and cr mischke,” 1992.

[125] K. Huang, B. Shi, X. Li, X. Li, S. Huang, and Y. Li, “Multi-modal sensor fusion for auto driving perception: A survey,” *arXiv preprint arXiv:2202.02703*, 2022.

[126] M. Ferdowsi, A. Benigni, A. Monti, and F. Ponci, “Measurement selection for data-driven monitoring of distribution systems,” *IEEE Systems Journal*, vol. 13, no. 4, pp. 4260–4268, 2019.

[127] N. Li, N. Gebraeel, Y. Lei, X. Fang, X. Cai, and T. Yan, “Remaining useful life prediction based on a multi-sensor data fusion model,” *Reliability Engineering & System Safety*, vol. 208, p. 107249, 2021.

[128] R.-T. Wu and M. R. Jahanshahi, “Data fusion approaches for structural health monitoring and system identification: past, present, and future,” *Structural Health Monitoring*, vol. 19, no. 2, pp. 552–586, 2020.

[129] A. J. Torabi, M. J. Er, X. Li, B. S. Lim, and G. O. Peen, “Application of clustering methods for online tool condition monitoring and fault diagnosis in high-speed milling processes,” *IEEE Systems Journal*, vol. 10, no. 2, pp. 721–732, 2015.

[130] B. Sun and X. F. Liu, “Significance support vector machine for high-speed train bearing fault diagnosis,” *IEEE Sensors Journal*, 2021.

[131] E. Brill and R. C. Moore, “An improved error model for noisy channel spelling correction,” in *Proceedings of the 38th annual meeting of the association for computational linguistics*, pp. 286–293, 2000.

[132] P. Zheng, H. Wang, Z. Sang, R. Y. Zhong, Y. Liu, C. Liu, K. Mubarok, S. Yu, and X. Xu, “Smart manufacturing systems for industry 4.0: Conceptual framework, scenarios, and future perspectives,” *Frontiers of Mechanical Engineering*, vol. 13, pp. 137–150, 2018.

[133] M. Ghobakhloo, "Industry 4.0, digitization, and opportunities for sustainability," *Journal of cleaner production*, vol. 252, p. 119869, 2020.

[134] G. Aceto, V. Persico, and A. Pescapé, "A survey on information and communication technologies for industry 4.0: State-of-the-art, taxonomies, perspectives, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3467–3501, 2019.

[135] U. LLC, "Ultralytics yolov8," 2024. Accessed: 2024-06-03.

[136] A. M. Mood, "Introduction to the theory of statistics.," 1950.

[137] J. Lu, C. Clark, R. Zellers, R. Mottaghi, and A. Kembhavi, "Unified-io: A unified model for vision, language, and multi-modal tasks," *arXiv preprint arXiv:2206.08916*, 2022.

[138] Y. Li, Q. Fan, H. Huang, Z. Han, and Q. Gu, "A modified yolov8 detection network for uav aerial image recognition," *Drones*, vol. 7, no. 5, p. 304, 2023.

[139] J. Terven and D. Cordova-Esparza, "A comprehensive review of yolo: From yolov1 to yolov8 and beyond," *arXiv preprint arXiv:2304.00501*, 2023.

[140] H. Lou, X. Duan, J. Guo, H. Liu, J. Gu, L. Bi, and H. Chen, "Dc-yolov8: Small-size object detection algorithm based on camera sensor," *Electronics*, vol. 12, no. 10, p. 2323, 2023.

[141] Y. Huo, M. Zhang, G. Liu, H. Lu, Y. Gao, G. Yang, J. Wen, H. Zhang, B. Xu, W. Zheng, *et al.*, "Wenlan: Bridging vision and language by large-scale multi-modal pre-training," *arXiv preprint arXiv:2103.06561*, 2021.

[142] A. A. Alogla, A. Alzahrani, and A. Alghamdi, "The role of additive manufacturing in reducing demand volatility in aerospace: a conceptual framework," *Aerospace*, vol. 10, no. 4, p. 381, 2023.

[143] A. Goré and A. Nathaus, "End-to-end demand management for the aerospace industry," *Supply Chain Integration Challenges in Commercial Aerospace: A Comprehensive Perspective on the Aviation Value Chain*, pp. 105–118, 2017.

[144] A. Zaleski, "Ge's bestselling jet engine makes 3-d printing a core component," 2015. http://fortune.com/2015/03/05/ge-engine-3d-printing/ (Accessed: 2023-09-23).

[145] V. A. Prabhu, M. Elkington, D. Crowley, A. Tiwari, and C. Ward, "Digitisation of manual composite layup task knowledge using gaming technology," *Composites Part B: Engineering*, vol. 112, pp. 314–326, 2017.