End-to-End Deep Learning-Driven Automation for Enhanced Inspection in Industry

by

Philopatear Salib

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Mechanical and Mechatronics Engineering

Waterloo, Ontario, Canada, 2024

© Philopatear Salib 2024

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my readers.

I understand that my thesis may be made electronically available to the public.

Abstract

This study explores the integration of machine learning into industrial automation, specifically focusing on transformer inspection to address labor shortages and enhance operational efficiencies. It develops a specialized defect detection system using ResNet architectures with various depths, thereby significantly advancing fault detection capabilities within industrial environments. This research integrates an end-to-end system for automatic inspection, encompassing data collection, augmentation, and labeling of objects for inspection, alongside leveraging ResNet architectures for in-depth training and hyperparameter tuning.

The research aims to reduce reliance on skilled labor while increasing the accuracy of inspections. A Raspberry Pi-based monitoring system is designed, implemented, and evaluated, revealing substantial improvements in both the precision and efficiency of transformer inspections. Achieving an accuracy rate exceeding 90% stands out as a major accomplishment, emphasizing the robustness of the machine learning model and the effectiveness of its training and optimization processes.

Comprehensive reliability and repeatability tests are conducted under real-time conditions, with multiple users adjusting the orientation and placement of objects in varied lighting and location settings. Despite these challenges, the system consistently and accurately determines the status of the objects, demonstrating its ability to operate effectively in diverse and unpredictable environments. This consistent performance confirms the system's readiness for industrial use and its reliability, making it a critical solution for environments where high accuracy and consistency are paramount. The potential of deep learning to revolutionize industrial inspection processes is affirmed, paving the way for future enhancements and broader applications of the technology.

Acknowledgements

I extend my deepest gratitude to everyone who supported me throughout the completion of this thesis. First and foremost, I would like to express my heartfelt appreciation to my family, whose unwavering support and encouragement have been a cornerstone of my journey. Your belief in my abilities and constant love have fueled my dedication and perseverance.

I am equally thankful to my friends, who have provided not only moral support but also moments of relief and laughter during the most challenging periods of my research. Your camaraderie has been invaluable.

Special thanks go to my partner, who has been a source of endless inspiration and motivation. Your patience, understanding, and love have made all the difference in balancing the demands of my research with life's other responsibilities.

I would also like to express my profound gratitude to my supervisor, Prof. William Melek, for his guidance and expertise. Prof. Melek, your mentorship has been instrumental in shaping both the direction of my research and my growth as a scholar. Your critical insights and encouragement have been crucial in helping me navigate the complexities of this work.

Finally, I thank all the readers for engaging with my work. Your interest and feedback are not only appreciated but also motivate continued exploration and improvement.

This thesis would not have been possible without the collective support and belief of each one of you, and for this, I am eternally grateful.

Dedication

To my beloved wife, Marian,

Your unwavering support and partnership have been invaluable throughout this challenging journey. This thesis stands as a testament to our shared commitment and the inspiration you have provided. Thank you for being a constant source of strength and encouragement.

Also to my father, Ashraf,

In loving memory of your wisdom and unwavering support. Your legacy of strength and kindness continues to inspire and guide me.

Table of Contents

Author's Declaration ii

Abstract iii

Acknowledgements iv

Dedication v

List of Figures ix

List of Tables x

1 Introduction 1

1.1 Motivation 1

1.2 Thesis Contributions 3

1.3 Thesis Outline 4

2 Background and Related Work 5

2.1 Background 5

2.1.1 Introduction to Defect Detection in Industrial Products 5

2.1.2 Challenges in Deep Learning-Based Defect Detection 6

2.1.3 The Role of Residual Networks in Enhancing Deep Learning for Automated Inspection 7

2.2 Related Work . . . 9
2.2.1 Traditional Machine Vision vs. Deep Learning Approaches . . . 9
2.2.2 Deep Learning Approaches . . . 10
2.2.3 Case Studies and Applications of ResNet for Defect Detection . . . 12
2.2.4 Multi-Camera Systems for Industrial Inspection in Conveyor Belt Systems . . . 19
2.2.5 Specifications for Conveyor Belt Inspection System . . . 20

3 System Design **22**
3.1 Introduction and Problem Description . . . 22
3.2 Hardware Selection and Design Considerations . . . 24
3.2.1 Raspberry Pi . . . 24
3.2.2 Camera Configuration for Image Acquisition . . . 24
3.2.3 Development of the Camera Mount . . . 25
3.2.4 Commercial Conveyor and Belt Adaptation . . . 26
3.3 Electrical Configuration and Integration . . . 29
3.4 Expandability and Reconfigurability . . . 29

4 Machine Learning Framework **31**
4.1 Data Acquisition . . . 31
4.1.1 Data Description and Statistics . . . 31
4.1.2 Acquisition Process . . . 35
4.2 Data Labeling . . . 35
4.2.1 Binary Class Labeling . . . 36
4.2.2 Multi-Class Labeling . . . 38
4.3 Data Preprocessing . . . 38
4.3.1 Best Focus Selection . . . 39
4.3.2 Data Augmentation . . . 39

4.4 Model Training and Testing . . . 40
4.5 Results and Discussion . . . 41
4.5.1 Initial Experiments . . . 41
4.5.2 Balancing Datasets with Augmentation . . . 42
4.5.3 Multi-Class Classification . . . 43
4.6 Real-time Implementation and Testing . . . 44
4.6.1 Implementation Details . . . 44
4.6.2 Model Repeatability and Adaptability . . . 45

5 Conclusion and Future Work **48**
5.1 Conclusion . . . 48
5.2 Future Work . . . 49
5.2.1 Automation Enhancements . . . 49
5.2.2 Advanced Learning Architectures . . . 51

References **53**

List of Figures

2.1 Residual Block. 8
2.2 Different types of steel surface defects. Reprinted from [1]. 13
2.3 Railway track fasteners defects. Reprinted from [2] 14
2.4 Classification of component control screws. 15
2.5 Types of defects in PCBs. Reprinted from [3] 17
2.6 Types of defects in ICs. Reprinted from [4] 18

3.1 Automatic Inspection System Setup. 23
3.2 Camera Mount Design Overview 27
3.3 Camera Mount Overview. 28

4.1 An illustration of the types of defects in the dataset 33
4.2 Various object orientations 34
4.3 Challenges in data labeling. 37
4.4 Realtime model predictions. 46

5.1 Stopping mechanism design. 50

List of Tables

4.1 Summary of initial experiments with binary classification. 42
4.2 Summary of main experiments. 43

Chapter 1

Introduction

1.1 Motivation

In manufacturing processes, even with small batch production and limited variety, defects remain a significant concern. Despite careful attention to detail and quality control measures, human error and variability can lead to inconsistencies in the final products. These defects compromise the quality and reliability of products, resulting in substantial costs for manufacturers, including rework, recalls, and potential damage to brand reputation.

Manual inspection processes are inherently prone to oversight and subjectivity, increasing the likelihood of defects going undetected. By automating the inspection process, manufacturers can mitigate these risks and ensure a higher level of quality assurance. Automated inspection systems offer consistent and precise evaluation, reducing error margins and enabling timely defect identification and rectification. This enhances product quality and customer satisfaction while optimizing operational efficiency and minimizing financial losses associated with defective products. Therefore, the automation of inspection is imperative for manufacturers seeking to uphold high-quality standards and maintain a competitive edge in today's dynamic market landscape.

In recent years, the task of maintaining and monitoring electrical transformers has become increasingly important [5]. Transformers are crucial for distributing electricity across wide areas, and any faults in these devices can lead to significant disruptions in power supply, affecting homes, businesses, and essential services. Traditionally, the inspection and fault detection of transformers have been performed manually by skilled workers. However, this approach has several drawbacks.

Recent data from Statistics Canada [6] highlights the acute challenges faced by the labor market, particularly in sectors requiring specialized skills, such as electrical maintenance. The first quarter of 2023 saw a slight ease in labor shortages with job vacancies decreasing to 815,295 in March, the lowest since July 2021, yet the unemployment rate has remained steady at 5.0% since December 2022. This paradox highlights an underlying issue of matching skilled labor with job vacancies, further complicated by an average weekly earnings increase of 1.8 % year-over-year in February 2023, amidst a 4.4% year-over-year increase in the Consumer Price Index in April 2023.

The numbers shared by Statistics Canada are justifiable for the following reasons. Firstly, manual inspections are time-consuming and labor-intensive. With the increasing demand for electricity and the expansion of electrical networks, the workload for maintenance teams has grown substantially. This is particularly challenging in current labor market trends, which show a persistent shortage of skilled labor across various sectors, including the electrical maintenance industry.

The development of an automatic inspection system, as opposed to traditional manual inspection methods, is increasingly critical in light of Canada's demographic trends. According to Statistics Canada, the number of Canadians aged 65 and older has reached 7.3 million as of July 2022, comprising 18.8 % of the population. This segment is projected to exceed 20 % by 2025, significantly impacting the labor force in sectors that rely heavily on manual labor, with projections indicating that the proportion of those aged 65 and over will continue to rise markedly in the coming decades [7]. This shift presents a dual challenge for the industrial sector: a diminishing workforce capable of performing complex manual tasks and an increased burden on industries to maintain productivity with fewer human resources. The aging demographic significantly motivates the need for robust, efficient, and less labor-intensive systems such as the one developed in this thesis, which can sustain industrial operations and mitigate the impact of labor shortages due to demographic changes.

Secondly, human inspections are prone to errors. Detecting faults in transformers requires a high level of precision and attention to detail. Human inspectors, despite their expertise, can miss subtle signs of malfunction or degradation, leading to delayed repairs and, in some cases, catastrophic failures. The accuracy of fault detection is critical not only for the reliability of the power supply but also for the safety of the maintenance personnel and the general public.

To address these challenges, our project proposes an innovative solution: a 4-camera system connected to a microcontroller for automated fault detection in transformers. This system utilizes cameras positioned in four directions—up, down, left, and right—to thor-

oughly inspect the transformers. The cameras capture detailed images from multiple angles, allowing for a comprehensive assessment of the transformer's condition. The microcontroller processes these images using advanced algorithms to identify any signs of faults or damage accurately.

The advantages of this approach are twofold. Firstly, it significantly reduces the reliance on manual labor for inspections, thereby mitigating the impact of labor shortages on maintenance operations. In the context of Canada's current labor market, where the scarcity of skilled workers is a pressing issue, adopting technology-driven solutions can help industries maintain high productivity levels. Secondly, the use of automated systems enhances the accuracy of fault detection. By minimizing human errors, we can ensure that transformers are maintained in optimal condition, reducing the risk of power outages and ensuring a reliable electricity supply for consumers.

In summary, the motivation behind this project is to leverage computer vision to overcome the challenges faced by the electrical maintenance industry. By introducing a deep learning-based automated fault detection system, we aim to enhance the efficiency and reliability of transformer inspections, contributing to a more stable and secure electrical infrastructure.

1.2 Thesis Contributions

This thesis presents several contributions to the field of automatic inspection using computer vision:

- Developed an end-to-end system for automatic inspection, integrating multiple components: data collection, augmentation, and labelling of objects for inspection, ensuring a robust dataset for training machine learning models.
- Leveraged ResNet for training and hyperparameter tuning resulting in significant improvements in accuracy and reliability.
- Achieved a computer vision solution with an accuracy of 90%, which surpasses the performance of manual and existing automated solutions in the industry.

1.3 Thesis Outline

The rest of this thesis is organized as follows. After introducing the problem and discussing the motivation behind the study, Chapter 2 presents an extensive background and literature review on current inspection systems that utilize deep-learning-based models.

Chapter 3 outlines the design and integration of the necessary hardware components critical to supporting the machine learning framework. It details the selection and configuration of sensors, cameras, and computational units, and discusses the challenges faced during the hardware setup along with the solutions implemented to address these issues.

Chapter 4 dives into the machine learning framework, covering the entire process from data collection and model training to real-time testing. It emphasizes the use of ResNet for training, details on hyperparameter tuning for improved results, and demonstrates the integration of hardware and software for effective real-world applications.

Finally, Chapter 5 concludes the thesis with a summary of the findings and the significant contributions of the research. It also discusses the limitations encountered during the study and outlines potential future work, suggesting ways to enhance the system's capabilities and adapt the technology for broader applications.

Chapter 2

Background and Related Work

2.1 Background

The widespread adoption of modern digital technologies has significantly changed the landscape of industrial monitoring and maintenance methods. Among these, the integration of compact but powerful computing devices such as the Raspberry Pi [8], which is equipped with many high-resolution cameras, has created new opportunities for real-time surveillance and diagnostic systems. This thesis project uses a new assembly of four Raspberry Pi cameras strategically placed on the left, right, top, and bottom of electric transformers to provide thorough fault detection. The system's analytical capability is provided by ResNet-101 [9], a deep residual network noted for its efficiency in complicated picture identification tasks. This Background section investigates the relevant literature, technology, and theoretical frameworks that support the design and implementation of the unique defect detection system. It aims to contextualize the project within the broader scope of smart monitoring systems, highlighting the significance of leveraging deep learning techniques and edge computing for enhanced predictive maintenance in the power distribution sector.

2.1.1 Introduction to Defect Detection in Industrial Products

In the continuously changing environment of industrial maintenance and quality control, the use of modern technologies to detect structural defects and imperfections in key infrastructure has emerged as a critical topic of research. Small electric transformers stand

out as assets that require comprehensive quality assurance processes due to their critical function in electrical distribution networks [10]. The reliability and safety of the power supply depend primarily on the transformers' integrity [5].

Traditional inspection methods, often manual and labor-intensive, are becoming inadequate for the demands of modern efficiency and precision [11]. These manual methods, while potentially effective, are prone to human error, time-consuming, and not suited for high-volume environments [6]. In response, the adoption of artificial intelligence (AI) and machine vision in the inspection process represents a promising solution [12].

Defect detection plays a vital role in the industrial sector by ensuring that the materials used in construction, automotive, aerospace, and consumer goods meet rigorous quality standards [10]. The early identification of defects not only enhances the safety and reliability of products but also significantly reduces costs associated with recalls and repairs, thereby protecting brand reputation and consumer trust. Furthermore, efficient defect detection processes contribute to the optimization of production lines, reducing waste and improving overall manufacturing efficiency [13].

This thesis project centers on the development of a novel inspection system designed to automate the identification of fractures and faults in small electric transformers. The system's foundation is based on supervised learning using ResNet-101 [9]. ResNet-101 was chosen because of its ability to achieve high accuracy in various domains, including defect detection [14, 15, 16, 1, 17]. By exploiting the capabilities of this advanced algorithm within a carefully built hardware arrangement, the project hopes to improve inspection precision and efficiency and considerably cut the time and labor associated with previous approaches.

2.1.2 Challenges in Deep Learning-Based Defect Detection

The variability and quality of the data used in training deep learning models for defect detection present considerable challenges. Industrial settings generate an extensive variety of defects, frequently observed in distinct lighting conditions and angles, which complicates the task of acquiring a comprehensive dataset that accurately reflects every possible defect scenario [10].

An additional noteworthy obstacle corresponds to the assurance that machine learning models, regardless of their complexity, demonstrate the ability to effectively generalize to unseen data. While models may exhibit outstanding accuracy on the training dataset, their ability to accurately identify defects in real-world applications may be compromised as a result of overfitting or underfitting [12] [18].

Another major challenge in industrial settings is the need for real-time defect detection with models that are both accurate and fast. Processing speed is a significant bottleneck when dealing with complex models and high-resolution images. It is essential to optimize both the model architecture and computational resources to achieve the necessary speed for real-time processing without sacrificing accuracy [18].

Additionally, the integration of defect detection systems into existing manufacturing processes requires careful planning around factors such as lighting conditions, camera positioning, and real-time processing capabilities [10]. These considerations are crucial for the successful deployment and operation of advanced defect detection systems in industrial environments.

2.1.3 The Role of Residual Networks in Enhancing Deep Learning for Automated Inspection

Addressing the aforementioned challenges of real-time defect detection and integration of complex systems into manufacturing processes, Residual Network (ResNet) [9] offers a powerful solution in the landscape of automated inspection systems. As a key advancement in deep learning, ResNets significantly enhance the training and functionality of neural networks by effectively managing the complexities associated with deep learning architectures [10]. In the realm of defect detection in industrial products, ResNets excel by facilitating the learning of complex features essential for high accuracy, all while mitigating the problem of vanishing gradients—a frequent issue in deeper networks [19]. This subsection explores the working principles of ResNets and their crucial role in elevating the capabilities of automated inspection technologies.

ResNets introduce a novel architectural innovation called "residual blocks" or "skip connections," (illustrated in Figure 2.1) which directly address the vanishing gradient problem. In traditional deep networks, as the gradient of the loss function is backpropagated to earlier layers, it tends to become increasingly smaller, which significantly slows down the learning process during training and can lead to poor performance during inference. ResNets combat this issue by incorporating shortcut connections that skip one or more layers, allowing the gradient to flow directly through the network without transforming every layer.

Each residual block in a ResNet consists of a sequence of layers—typically convolutional layers followed by batch normalization layers and ReLU activation functions. The key feature of these blocks is the addition of the input of the block to its output, facilitating an identity mapping that encourages the network to learn residual functions regarding the

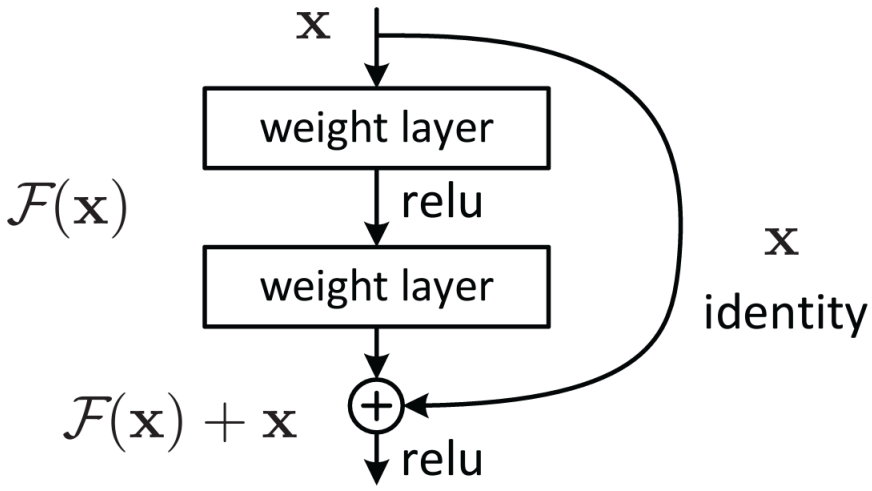


Figure 2.1: Residual Block. The building unit of ResNets. Reprinted from [9].

layer inputs. This means that some original data from the input is directly added to the output without any modification, helping to maintain the flow of information through the network. This design principle not only helps in preserving the strength of the gradient, ensuring that the learning signals remain strong but also simplifies the optimization of the network, enabling the training of substantially deeper networks than was previously possible. Essentially, this approach makes it easier to adjust and improve the network, allowing for the construction of much larger and more complex networks without compromising their effectiveness.

In the context of automated inspection for defect detection, the ability of ResNets to effectively train deeper networks is invaluable. Deeper networks can learn more complex, hierarchical features [9] from images of industrial products, which are essential for recognizing subtle and complex surface irregularities. The enhanced feature representation capabilities of ResNets thus significantly improve the accuracy and reliability of defect detection systems.

Moreover, various adaptations of the original ResNet architecture, such as ResNet-50 and ResNet-101 [9], offer flexibility in modeling complexity, allowing researchers and practitioners to tailor the network depth according to the specific requirements and computational constraints of their inspection systems. These models have set new benchmarks in image classification tasks, which translate effectively to the domain of defect detection

due to the similar nature of the image-based learning tasks [20].

As we transition into the subsequent section on related work, it becomes apparent that the adoption of ResNets has not only revolutionized our approach to training deep networks but has also paved the way for more sophisticated and efficient automated inspection systems. The discussions that follow will investigate how the foundational principles of ResNets, along with comparable deep learning approaches, have been tailored and expanded upon. We will specifically examine their incorporation into systems that demand high precision and reliability, which are quintessential for industrial applications.

2.2 Related Work

2.2.1 Traditional Machine Vision vs. Deep Learning Approaches

Bridging from the innovation of ResNets to a broader spectrum of technologies, it is important to contrast these advancements with traditional machine vision systems. Traditional systems follow a series of steps for defect detection on industrial products—commencing with image acquisition and followed by preprocessing, feature extraction, and finally, defect classification. At the core of these processes lies feature extraction, where specific characteristics such as texture, color, and shape are pivotal in differentiating between defective and non-defective areas [12].

Texture Features. In classical machine vision, one of the most popular methods for defect detection is texture analysis. Texture characteristics can be used to characterize a material's surface characteristics and patterns. To quantify texture and detect anomalies like dents, scratches, or uneven patterns, methods including Fourier Transforms, Local Binary Patterns (LBP)[21], and Gray Level Co-occurrence Matrix (GLCM) [10] are implemented. However, these techniques could be noise-sensitive and require extensive preprocessing to separate the relevant texture elements [12].

Color Features [22]. Color features are essential for identifying flaws such as discolorations, stains, and uneven material composition. To record the distribution and intensity of colors throughout the surface, techniques like color moments, color histograms, and color coherence vectors are used. Variations in illumination can pose a problem to color-based approaches, even when they work well for some types of faults. This can result in inconsistent color perception and possibly false positives or negatives in defect detection[23].

Shape Features [12]. Shape-based detection involves analyzing the geometric attributes of defects, such as edges, contours, or specific shapes indicating chips, cracks, or

holes. Techniques like Hough Transform for line detection or blob analysis for identifying irregular shapes are common. While powerful for certain defect types, shape-based methods may struggle with complex or subtle defects that do not have a well-defined geometric form[24].

Traditional machine vision techniques can perform effectively in certain scenarios, but they have many drawbacks that limit their application in a variety of industrial settings. These techniques usually require specific algorithms and threshold settings for different kinds of defects and goods, which limits their ability to be applied broadly across a large variety of industrial applications without requiring significant manual adjustments [12].

Furthermore, the performance of traditional machine vision methods is highly sensitive to environmental factors like lighting, shadows, and reflections, necessitating extremely regulated settings to get consistent results [25]. Additionally, extensive reconfiguration and human tweaking are frequently required to adapt these conventional techniques to new fault kinds or product lines, which hinders their scalability and adaptability in dynamic industrial environments.

The field of defect detection has been significantly transformed by the development of deep learning, which provides models that develop feature representations automatically from the data. This eliminates the need for challenging manual feature design and extraction, which is a defining characteristic of conventional machine vision approaches. Convolutional Neural Networks (CNNs) [26], in particular, have exhibited exceptional efficacy in the detection of defects spanning a diverse array of materials and surfaces.

2.2.2 Deep Learning Approaches

The field of defect detection has undergone a substantial revolution since the advent of deep learning models, which possess the capability to extract features from data automatically. This advancement significantly diminishes the reliance on human feature generation and extraction, a pivotal attribute of traditional machine vision methodologies. For example, CNNs [27] [28] have emerged as frontrunners in the field when it comes to detecting a multitude of defects across diverse materials and surfaces.

In the arena of defect detection, several deep-learning techniques have taken center stage:

- CNNs [29]: These models have become the cornerstone for image-related tasks, including defect detection. Various architectures like ResNet, AlexNet, and VGG have

been tailored to pinpoint defects by mastering the spatial hierarchy of features [27] [26].

- Transfer Learning [30]: This technique utilizes a model pre-trained on a comprehensive dataset, such as ImageNet, fine-tuned to cater to specific defect detection tasks. It's particularly effective for training robust models with limited labeled data in a given target area [31].

- Generative Adversarial Networks (GANs) [32] : Employed in anomaly detection, GANs can either augment datasets with synthetic defect images or learn the normative patterns of appearance to identify anomalies through the reconstruction of input images.

- Semi-Supervised [27] and Unsupervised [33] Learning: These methodologies prove invaluable when labeled data is sparse, utilizing unlabeled data to train models to spot anomalies by recognizing deviations from established normal patterns

Defect detection has been significantly revolutionized through the implementation of deep learning models, which acquire features autonomously from the data. This advancement substantially diminishes the reliance on manual feature generation and extraction, a defining characteristic of traditional machine vision methodologies. Due to its layered architecture, deep learning is capable of autonomously complex features from unprocessed photographs. This capability obviates the necessity for manual feature engineering by enabling the models to differentiate between inconsequential characteristics such as edges and textures and more conspicuous attributes like patterns and forms, which are critical for defect detection.

A significant advantage of deep learning models is their exceptional capacity for generalization [34]. When trained on vast and diverse datasets, these models demonstrate the ability to effectively process novel, unseen images, exhibiting resilience in the face of illumination, orientation, and background disturbances. This adaptability is of the utmost importance in industrial settings characterized by volatile conditions.

Furthermore, a transformative factor is the scalability of deep learning models [19]. Following the preliminary training stage, these models can be readily implemented and adjusted to various production environments and defect detection obstacles, thereby optimizing deployment procedures and substantially reducing financial and time expenditures.

In the electrical equipment manufacturing industry, such as the production of small transformers, integrating deep learning models with existing inspection systems can significantly enhance defect detection. For instance, traditional inspection methods might involve

visual checks for issues such as insulation breakdown or structural damage in transformers. By adopting deep learning, these systems can improve their accuracy and efficiency.

For example, a deep learning model could be trained on a dataset of transformer images to identify a range of defects, from minor surface cracks to internal component misalignments, which are not easily detectable by the human eye or conventional imaging techniques. This model could then work alongside the existing inspection framework, analyzing images from cameras on the production line in real-time. The result is a more reliable inspection process that quickly identifies potential failures, ensuring that only transformers in perfect condition reach the market. This approach leverages the precision of deep learning to complement and extend the capabilities of traditional inspection methods, ensuring higher quality standards in the industry's output.

2.2.3 Case Studies and Applications of ResNet for Defect Detection

Deep learning models, such as ResNet [9], significantly enhance the capabilities of visual inspection systems in industrial settings by processing images through multiple layers that progressively recognize more complex features. Initially, these models detect simple elements like edges and textures, but as the layers deepen, they become capable of identifying intricate defects such as cracks, corrosion, or irregular patterns. This functionality is crucial for detecting issues in critical components like electrical transformers.

ResNet, and specifically its variants like ResNet-101 [9], represents a pivotal architecture in deep learning, discussed earlier in Section 2.1.3. It is particularly effective for managing very deep networks due to its use of skip connections or shortcuts, which jump over some layers. These connections are vital as they address the issue of vanishing gradients—a prevalent problem in deeper networks where the gradients of the loss function decrease exponentially with the number of layers, complicating the training process.

As previously mentioned in this section, the integration of ResNet models into industrial inspection systems has led to significant improvements in defect detection accuracy. These models excel at recognizing complex defects in electrical transformers, providing robust quality assurance that effectively safeguards against both known and newly emerging defects. This ensures that the inspection processes maintain high standards of reliability and efficiency, vital for industries where precision is mandatory. The remainder of this section will present related work that substantiates ResNet's effectiveness in achieving these improvements, illustrating the broad application and impact of this technology across various industries.



(a) Corner crack (b) Sliver (c) Sponge crack

Figure 2.2: Different types of steel surface defects. Reprinted from [1].

Detection of Surface Defects in Steel Manufacturing. A study by Y. Huang et al. (2020) [1] employed ResNet models to identify surface defects in steel products (shown in Figure 2.2). The research demonstrated that ResNet, with its deep residual learning framework, could accurately classify different types of surface anomalies such as scratches, dents, and rust from images captured during the manufacturing process. This application underscored ResNet's ability to handle the high variability in the appearance of defects on steel surfaces, significantly improving the efficiency of quality control in steel production.

Railway Maintenance and Safety. The application of ResNet in railway track maintenance exemplifies how deep learning can enhance the detection and prevention of structural damages crucial to railway safety. Notable studies have explored this capability, demonstrating the proficiency of ResNet models in identifying and classifying track anomalies.

A significant contribution in this domain is detailed in a study on "Railway Track Fault Detection with ResNet Deep Learning Models," which showcases the model's effectiveness in recognizing track faults from image data [16]. This work highlights the adaptability of ResNet to the complex and varied visual patterns presented by railway tracks, enabling it to discern subtle yet critical structural anomalies that could lead to significant safety hazards.

Further elaborating on the utility of ResNet, a paper [14] investigates the use of these models to monitor railway tracks in real-time. The research emphasizes ResNet's capability to process high-resolution images swiftly, thus providing timely and accurate detections that are vital for maintaining continuous safety checks in railway operations.

In another study [2], ResNet was adapted for inspecting railway tracks to detect struc-



(a) Broken fastener



(b) Missing fastener

Figure 2.3: Railway track fasteners defects. Reprinted from [2]

tural damages, including cracks and breakages, as shown in Figure 2.3. The model was trained on a dataset comprising images of railway tracks under various conditions. ResNet's deep learning architecture excelled in recognizing minute, yet critical, anomalies that could potentially lead to safety hazards. This application highlighted the model's contribution to maintaining railway safety and preventing accidents.

Together, these studies affirm the critical role of ResNet in the realm of railway safety, highlighting its impact on preventive maintenance strategies. By leveraging the deep learning architecture of ResNet, railway maintenance teams can improve their inspection processes, ensuring higher safety standards and preventing accidents before they occur. This application of ResNet not only enhances operational efficiency but also supports the sustainability of safe railway systems.

Automotive Industry for Component Inspection. The automotive sector has seen applications of ResNet in inspecting car parts for defects. A study [15] demonstrated how ResNet could be trained to identify manufacturing defects as presented in Figure 2.4. By automating the defect detection process, manufacturers could ensure higher quality standards and reduce the manual labor typically required for such tasks. The application of ResNet in this context illustrated its adaptability to diverse industrial environments and its role in upholding product quality.

Electronics Manufacturing for PCB Inspection. Printed Circuit Board (PCB) manufacturing is another area where ResNet has been effectively applied. Identifying defects such as missing components, soldering errors, and short circuits, depicted in Figure 2.5 is crucial to ensuring the functionality of electronic devices. Researchers have used



Figure 2.4: Classification of component control screws in four zones. Each zone is categorized into four classes: Flexitol, Rivstud, Rivnut, and missing, representing the possible scenarios of screw presence or absence within the specified regions of interest. Reprinted from [15]

ResNet to automate the inspection of PCBs [3], training the model on high-resolution images to detect and classify various types of defects. This application of ResNet significantly enhanced the accuracy and speed of PCB inspections, contributing to the reliability of electronic products.

Furthermore, A paper[35]showcases the application of ResNet in identifying and classifying minute defects in PCB manufacturing, such as misalignments and incomplete circuits. The study demonstrates how ResNet's deep learning capabilities can be harnessed to enhance the accuracy of defect detection, significantly reducing the rate of false positives and false negatives in automated inspection systems.

Another pivotal study[17], focuses on the integration of ResNet models in the real-time analysis of PCB images. This application leverages the robust feature extraction capabilities of ResNet to process high-resolution images and accurately identify defects such as missing components, soldering errors, and short circuits, which are critical for ensuring the functionality of electronic devices.

Moreover, an article from Semi Engineering [36]delves into the use of a deep learning single-step detection model based on ResNet to streamline the PCB inspection process. This approach minimizes the need for multiple inspection stages and manual intervention, thereby speeding up the production line while maintaining high standards of quality assurance.

These collective research efforts reveal the transformative impact of ResNet in the realm of PCB manufacturing. By applying ResNet's sophisticated image analysis and pattern recognition capabilities, manufacturers can significantly enhance the precision and efficiency of their inspection processes, ensuring that the PCBs meet rigorous quality standards essential for the reliability of electronic products

Integrated Circuit Surface Defect Detection. In the work of Xiaobin Wang and colleagues (2024) [4], ResNet was integrated with Vision Transformers to create a hybrid architecture tailored for detecting surface defects in integrated circuits. Figure 2.6 illustrates the various defects present in the utilized dataset. This model, known as the Surface Defect Detection Model (SDDM) [4], excels at handling uneven information densities within datasets, which is a common challenge in industrial applications. The SDDM model significantly improves defect detection capabilities, achieving an accuracy rate of 98.6% and reducing computational costs, which is critical for scalability and practical application in the electronics manufacturing industry.

The effectiveness of ResNet in enhancing defect detection, as evidenced by the case studies, significantly impacts industrial product quality and safety. These advancements in

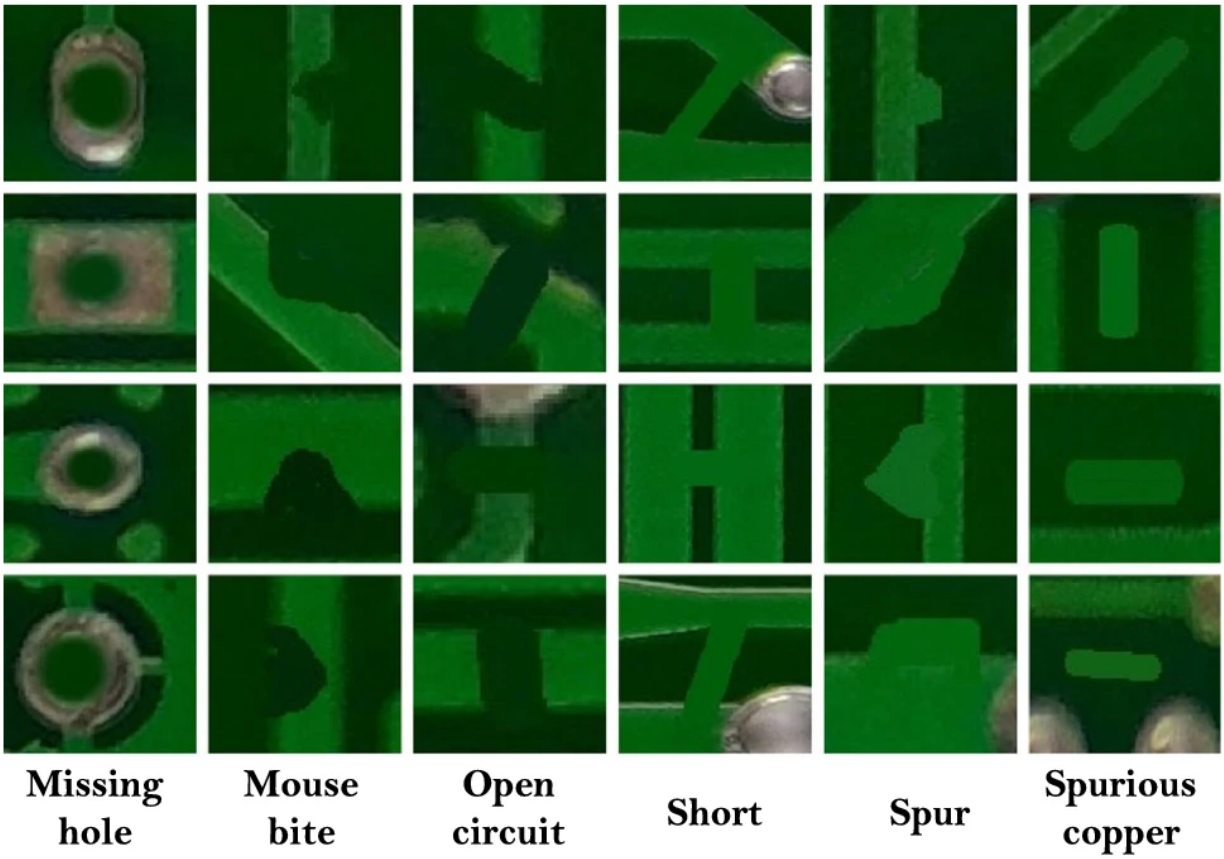


Figure 2.5: Types of defects in PCBs. Reprinted from [3]

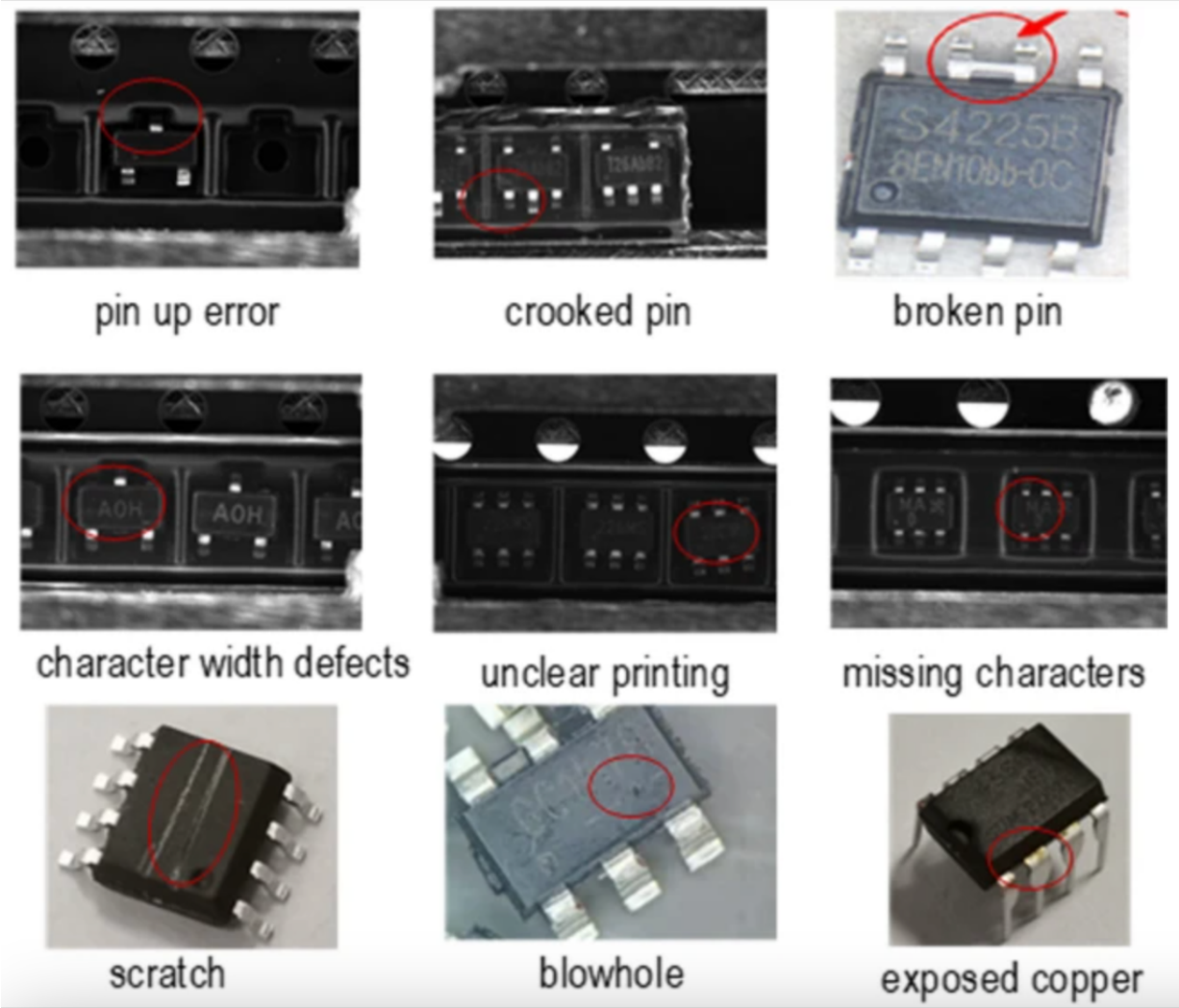


Figure 2.6: Types of defects in ICs. Reprinted from [4]

deep learning technology mark a crucial shift towards more reliable and accurate inspection methods.

2.2.4 Multi-Camera Systems for Industrial Inspection in Conveyor Belt Systems

Transitioning from single-camera to multi-camera systems represents a natural evolution in industrial inspection, further building on the reliability brought by deep learning technologies like ResNet. Multi-camera systems offer enhanced reliability due to the integration of data from multiple viewpoints [24], which greatly increases the robustness and accuracy of defect detection algorithms [37]. This multi-faceted input allows for cross-verification of detected defects, where anomalies identified by one camera are confirmed by others, thereby significantly boosting confidence in the results. Such redundancy is essential in environments demanding high precision, minimizing the risk of missed or misidentified defects.

The ability to capture multiple viewing angles is a significant advantage of multi-camera systems, especially in the context of conveyor belt systems. These systems can switch between different perspectives or generate virtual views to ensure a comprehensive inspection process along the conveyor belt [38]. For instance, a system equipped with four cameras around an object can offer top, bottom, front, and side views, uncovering defects that might remain hidden from a single viewpoint. This extensive coverage is crucial for thoroughly inspecting all aspects of products as they move along the conveyor belt, especially those with complex geometries and detailed features.

Furthermore, the versatility of multi-camera systems extends their utility beyond providing multiple views. Each camera can be specialized for particular tasks within the same inspection setup, such as one camera focusing on surface flaws and another on internal details using different lighting or filtering techniques. This multi-functionality allows for tailored inspection strategies that meet specific needs, enhancing the overall effectiveness of the inspection process by addressing a wider range of defect types and product complexities [38].

Transparent belts are crucial in improving the efficiency and precision of flaw identification in visual inspection systems integrated with conveyor systems. Transparent belts enable clear and uninterrupted vision of the goods being inspected, making it easier to apply imaging technologies from various perspectives. This is especially beneficial when incorporating sophisticated imaging equipment, such as high-resolution cameras, and sensors[24] [38], that depend on unobstructed vistas to precisely capture and evaluate the physical

attributes of goods as they progress along the manufacturing line. The belt's transparency eliminates any shadows or reflections that could hide faults, allowing the automated systems to conduct thorough inspections without any physical or visual obstacles [39] [40]. This configuration not only enhances the dependability of the examination procedure but also considerably diminishes the probability of faulty items surpassing quality control, thus upholding elevated benchmarks of product quality and safety[41].

In summary, multi-camera systems exemplify a significant advancement in industrial inspection, particularly in conveyor belt systems, combining enhanced reliability, versatility in viewing angles, and adaptability for multiple use cases. Their comprehensive capabilities make them invaluable assets in the ongoing effort to improve industrial inspection standards.

2.2.5 Specifications for Conveyor Belt Inspection System

In the realm of industrial inspection, the optimization of efficiency and accuracy is crucial for ensuring product quality and operational excellence. Effective conveyor belt systems equipped with advanced features are instrumental in meeting the diverse needs of industrial applications.

Furthermore, one of the primary specifications for an effective conveyor belt inspection system is achieving a level of accuracy that is at least equivalent to manual inspection. This requirement is fundamental to maintaining trust in automated processes and ensuring that the inspection system can reliably identify defects without human intervention.

Another critical specification is the integration of high-resolution cameras designed to meet the demands of diverse component sizes encountered in industrial settings. These cameras are vital for capturing fine details and optimizing image clarity, thereby ensuring accurate defect detection across a broad range of product types.

A key requirement for these systems is the incorporation of autofocus capability, essential for adapting to the varied assortment of parts moving along the conveyor belt. This functionality allows the system to dynamically adjust focus in response to changes in part sizes and shapes, ensuring clear and precise imaging throughout the inspection process.

Additionally, a comprehensive conveyor belt inspection system should integrate multiple camera views to provide thorough coverage of parts. This multi-camera setup captures various perspectives, facilitating complete defect detection regardless of part orientation and contributing to a higher accuracy inspection process.

To address the issue of occlusions, it is beneficial to employ a transparent conveyor belt design. Such a design guarantees unobstructed visibility of the parts during an inspection, thereby eliminating potential hindrances caused by belt components or shadows and enhancing the overall effectiveness of the inspection.

By focusing on these specifications—autofocus capability, multiple camera views, a transparent conveyor belt, high-resolution imaging, and accuracy equivalent to manual inspection—a well-designed conveyor belt inspection system can significantly enhance product quality and operational efficiency in industrial environments.

Chapter 3

System Design

3.1 Introduction and Problem Description

This chapter follows up on the requirements discussed in Chapter 2 by diving into how we designed our inspection system. We first discuss how we developed and implemented an automated system designed for efficiently inspecting items of various sizes and weights. A key feature of this system is an innovative conveyor with a transparent belt. This choice was the result of a careful evaluation to enhance the system's ability to facilitate easy image capture from multiple angles. Additionally, we discuss the incorporation of a camera system capable of simultaneously capturing images from several angles, significantly enhancing the inspection process.

Furthering our system's design, the selection of a Raspberry Pi [42] equipped with a quad-camera array as the foundational element of our imaging setup was driven by its compact size, processing capability, and cost-effectiveness. This setup achieves comprehensive coverage of parts for the current project stage. However, the potential for incorporating additional cameras remains an important consideration for future research.

The system's enclosure, designed and visualized in CAD before its physical construction, is crucial for protecting the components and ensuring enhanced imaging conditions. This chapter provides an in-depth overview of the design process and the functionality of the hardware through CAD illustrations and real-world photographs of the setup. Additionally, images captured by each of the cameras offer insights into the system's proficiency in conducting detailed inspections from various perspectives, further illustrating the system's capacity for multi-angle examination. An image of the entire system is presented in Figure 3.1.

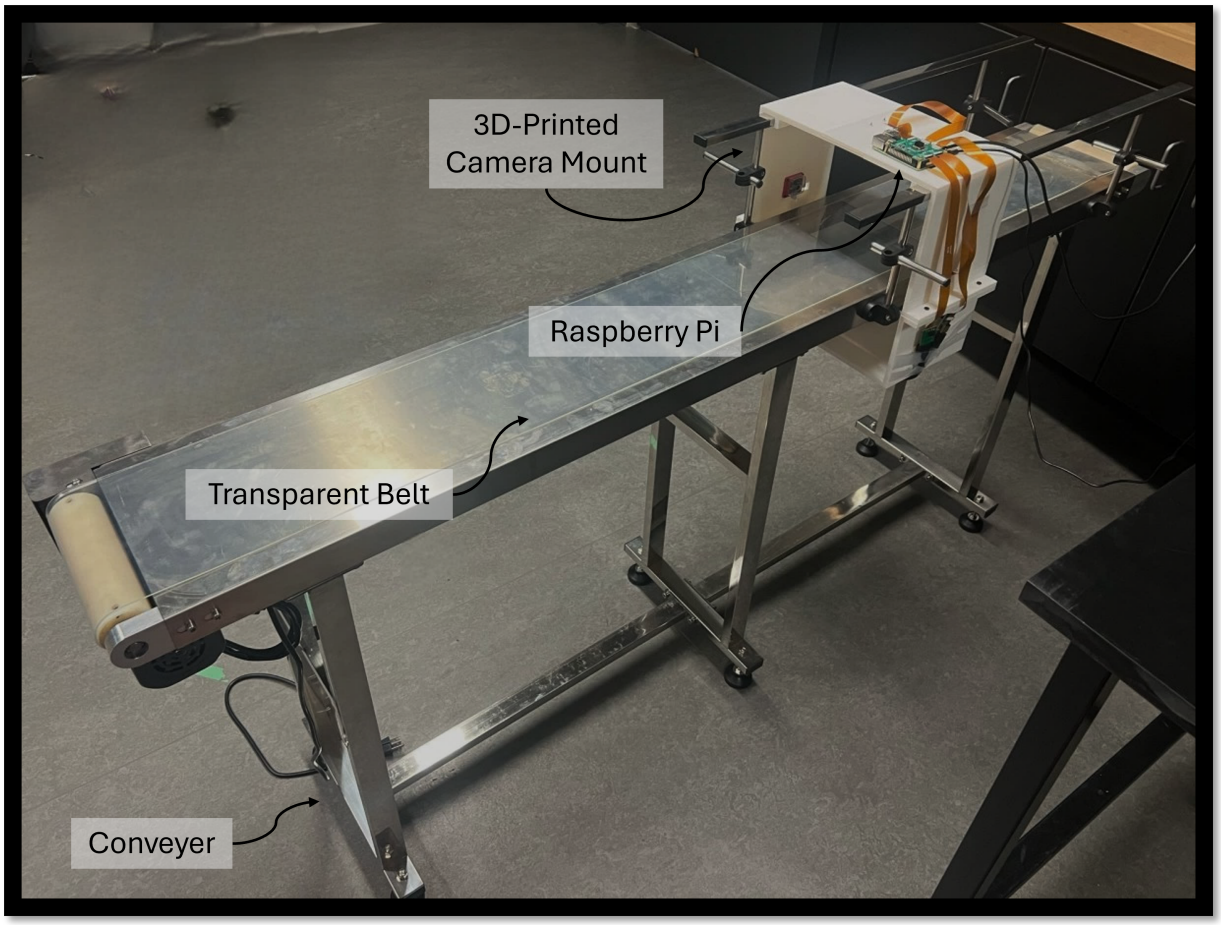


Figure 3.1: Automatic Inspection System Setup.

3.2 Hardware Selection and Design Considerations

3.2.1 Raspberry Pi

The Raspberry Pi 4b [42] was chosen as the foundation of our system for several compelling reasons. Its extensive documentation and general-purpose CPU support a wide range of Python scripts, including those for querying AI models as discussed in Chapter 4. Its compatibility with various image capture systems, ranging from USB cameras to more specialized options like our project's setup, offers significant flexibility. Furthermore, Raspberry Pi's rich ecosystem of image processing software libraries and tools, including OpenCV [43] and TensorFlow [44], provides a solid platform for our image acquisition needs. The built-in WiFi and Ethernet capabilities also facilitate remote monitoring, control, and data analysis, making it an ideal choice for our system.

Moreover, the hardware specifications of the Raspberry Pi 4B make it an exceptionally robust platform for our needs. It features dual Micro-HDMI ports supporting 4K resolution, which is beneficial for displaying the 4 1080p video streams from the cameras, simultaneously without scaling or compression. At its core, the Raspberry Pi 4B is powered by a Broadcom BCM2711, Quad-core Cortex-A72 (ARM v8) 64-bit SoC clocked at 1.5GHz, offering substantial computational power for multimedia processing and real-time data analysis. The device also includes a CSI camera port and a DSI display port, facilitating the connection of a Raspberry Pi camera and touchscreen display, respectively. This is particularly useful for our project, which utilizes these ports for image capture and user interface interactions. Furthermore, the 40-pin GPIO header on the Raspberry Pi 4B meets our current project requirements, providing sufficient connectivity for camera power, signal transmission, and colour-coded LED indicators, as discussed in Section 3.3. It also offers the flexibility to accommodate future expansions, such as controlling power for the conveyor and additional devices and integrating servo motors for object feeding mechanisms, highlighting the Raspberry Pi 4B's capability to evolve with the project's growing needs.

3.2.2 Camera Configuration for Image Acquisition

In the process of identifying an appropriate camera that meets our specific needs, our primary focus was on selecting a model that not only delivers images of a quality consistent with industry standards but also ensures compatibility with Raspberry Pi programming. A crucial aspect of our decision-making was the camera's ability to easily connect in a

multi-camera configuration without the need for extra space or complex programming adjustments for Raspberry Pi connectivity. We encountered several cameras that were unsuitable due to poor image quality, inappropriate size, or their inability to function simultaneously.

After a thorough evaluation, the Arducam 16MP IMX519 Autofocus Synchronized Quad-Camera Kit [45] was chosen as it satisfies our requirements effectively. It provides a high resolution of 4656x3496 pixels, capable of varying frame rates depending on the platform and resolution settings—up to 40 frames per second (fps) at 1280x720 and 4.5fps at the maximum resolution with Raspberry Pi. The camera is equipped with autofocus functionality, capable of focusing from 10cm to infinity. Its optical format is 1/2.53", and it features an integrated IR-cut filter that allows only visible light to pass through, enhancing image quality for our intended use.

The decision to incorporate four cameras into our system was influenced by the geometric principle that any object can be enclosed within a cube, which inherently has six faces. Although no formal optimization analysis was conducted to ascertain the ideal number of cameras, a notable avenue for future investigation, we opted for four cameras to streamline the system's complexity. This configuration was chosen with the prospect of future enhancements in mind, specifically the potential expansion to a six-camera array to more fully encompass the cubic perspective.

To ensure the production line operates efficiently without compromising on inspection times or computational demands, we employed a technique to adjust the camera focus over a defined range of distances. This was followed by the application of specialized software, as introduced in Chapter 4, to analyze these focus variations and select the best focus for each camera. This approach enables us to adapt to changes in the size of the objects and their proximity to the cameras, thus assuring consistently high-quality images under a range of scenarios.

3.2.3 Development of the Camera Mount

The foundational concept behind the camera mount was to establish a platform capable of accommodating four cameras positioned to capture images from four distinct perspectives: top, bottom, left, and right. This configuration was intentionally designed to encompass the target object comprehensively, enabling detailed image acquisition from every essential angle. Additionally, the mount's width was carefully considered to minimize background noise in the images, enhancing clarity and focusing on the subject matter.

Upon evaluating the initial design, we identified significant issues related to stray light interference, which compromised the image quality. These problems arose from unintended light reflections and diffraction within the camera mount's structure, leading to distortions in the captured images. This insight necessitated a re-evaluation and subsequent re-design of the camera mount to mitigate the adverse effects of stray light.

To address these challenges, the second design iteration (shown in Figure 3.2) incorporated strategic modifications to minimize light distortion. The revised design involved modifications to the dimensions of the structure's contours in addition to alterations in material density, accomplished by changing the printing infill settings. These changes effectively reduced reflections and diffraction. By focusing on these aspects, we were able to significantly diminish the impact of stray light on the image quality, resulting in clearer and more accurate captures from each of the four designated angles. Figure 3.3 depicts the practical realization of the final design configuration, showcasing the assembled cameras.

3.2.4 Commercial Conveyor and Belt Adaptation

The VEVOR Conveyor [46], integral to our system's design, is equipped with a 90W motor and boasts a loading capacity of up to 55 lbs (25 kg), alongside a variable speed control ranging from 0-18 m/min. Its dimensions, with a belt size of 71" x 7.8" (1800 x 200 mm) and an overall size of 71" x 9.8" x 29.5" (1800 x 250 x 750 mm), are tailored to meet the demands of our visual inspection setup. This conveyor has been specifically chosen for its capacity to handle items of various sizes and weights, facilitating a smooth and adjustable inspection process.

To meet the unique needs of our visual inspection system, we customized the conveyor by modifying its metal sheet bed. A rectangular section was cut out, allowing a camera positioned below the belt to have an unobstructed view of the items passing above. This modification ensures the capture of clear images for inspection purposes. Additionally, the decision to use a transparent material for the conveyor belt, inspired by visibility needs in the food industry, marks a departure from conventional belt materials. This choice was critical in avoiding the complexities and costs of implementing extra mechanical components or robotic arms for item repositioning. Our innovative design not only simplifies the inspection process but also significantly enhances its efficiency by allowing continuous, multi-angle image capture without manual intervention.

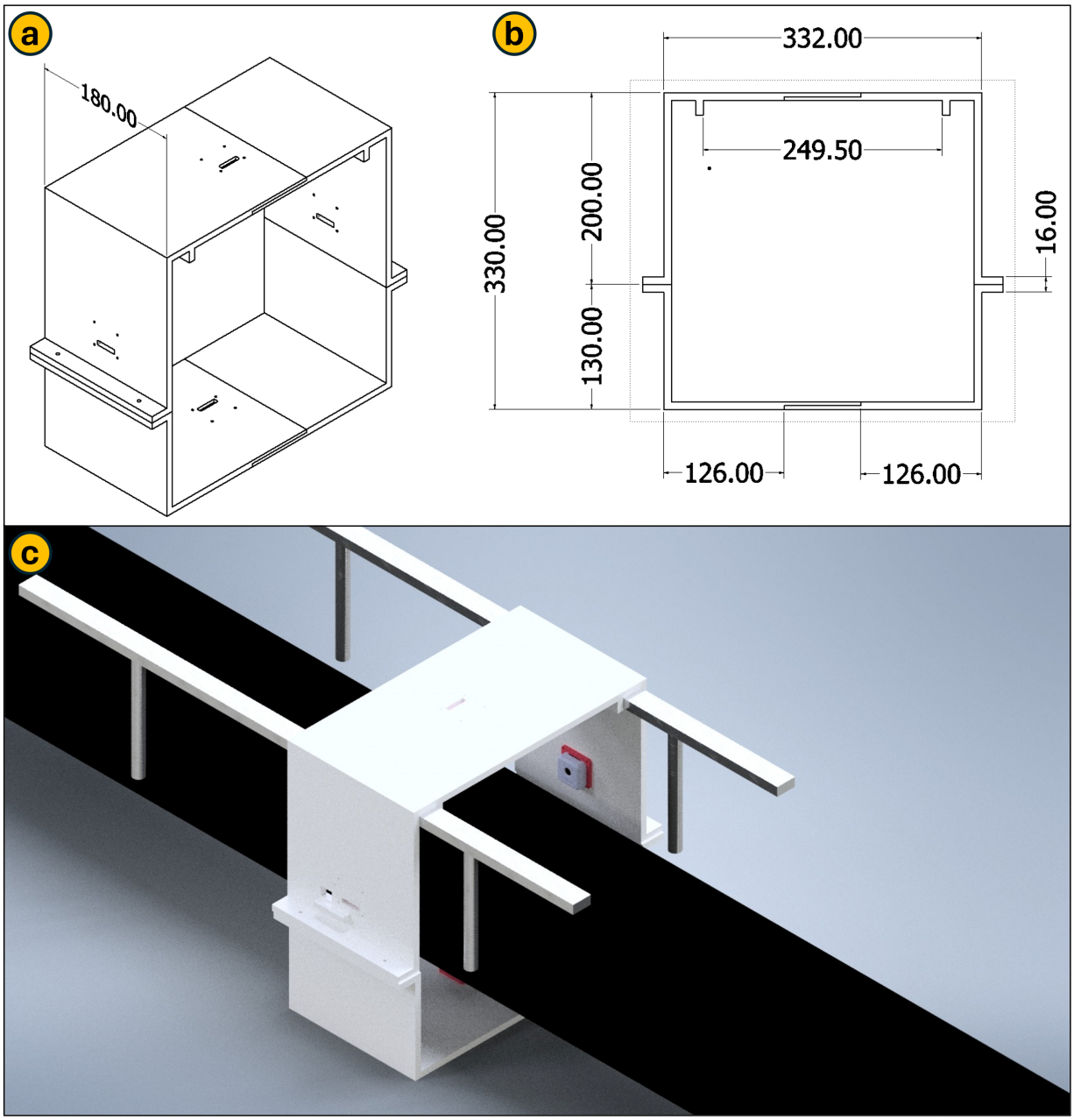


Figure 3.2: Camera Mount Design Overview. (a) Isometric view of the mount design. (b) Front view of the mount design. (c) Rendered 3D model. All dimensions are in millimetres.

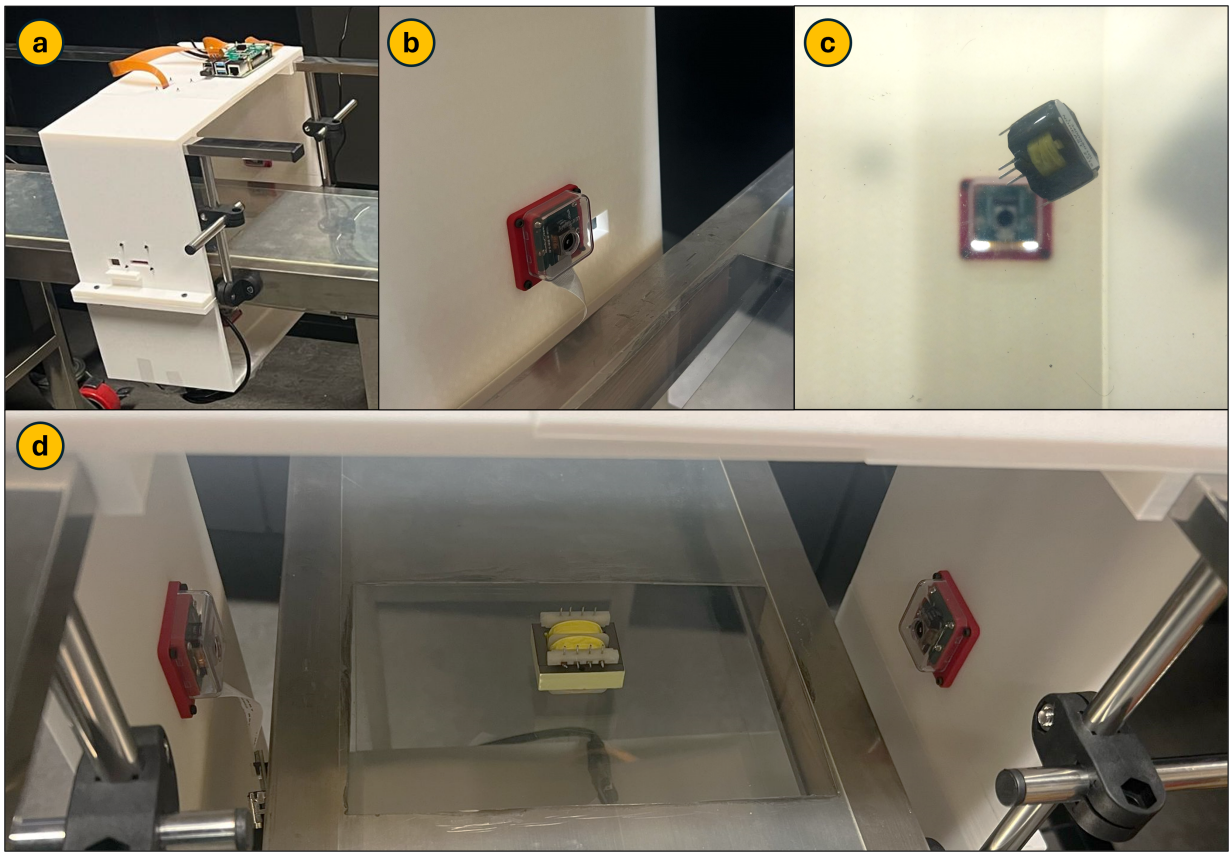


Figure 3.3: Camera Mount Overview. (a) Complete assembly view of the camera mount. (b) Left-side view of the mounted camera. (c) Below-belt perspective, showcasing the camera through the transparent conveyor belt. (d) An object centred within the camera mount, highlighting the conveyor's rectangular cutout and providing a view of both the left and right cameras.

3.3 Electrical Configuration and Integration

For the core of our electrical setup, we selected the Raspberry Pi 4b as the primary microcontroller due to its robust processing capabilities and versatility. To facilitate the connection of the IMX519 Quad-Camera Kit, a UC-512 Cam-array HAT was employed. This combination was chosen for its reliability and compatibility with the required camera array, ensuring seamless integration and optimal performance.

The camera system's communication infrastructure primarily utilizes standardized ribbon cables chosen for their compactness and efficiency in data transmission. However, ribbon cables have limitations regarding their maximum length for maintaining data integrity. To overcome this, we adapted our system to use HDMI cables, known for their capacity to support high-speed data transfers over longer distances. We ensured all cameras could be reliably connected by employing a commercially available ribbon connector to an HDMI adapter. It is important to note that not all HDMI cables met our requirements for this setup. After testing ten different HDMI cables from various manufacturers, we selected the most reliable and consistent connections, optimizing our system's data transfer capabilities.

Finally, a feature was developed to improve operational effectiveness involving color-coded LED indicators. The system employs three distinct colours to represent the status of each part as determined by the model's decision: a green LED indicates that a part has been classified as standard, a red LED signals a defective part that fails to meet the required specifications, and a yellow LED denotes a part that necessitates manual inspection due to the automated system's inability to classify it definitively as standard or defective. This approach simplifies the decision-making process for operators by providing clear, at-a-glance notifications about the quality assessment of parts.

3.4 Expandability and Reconfigurability

To fully maximize the practical application and efficiency of our inspection system, we have envisioned its potential for growth and adaptability to meet emerging needs. A critical aspect of this is the automation of object-handling processes. This entails integrating a feeder mechanism to systematically introduce objects into the inspection area. Additionally, deploying a conveyor system equipped with sensors would ensure precise positioning of objects at the camera station by enabling accurate conveyor stopping.

In addition to automating object handling, we propose the integration of a sorting mechanism before the inspection station. This sorting mechanism would classify objects

based on size, ensuring they align appropriately with the system's inspection requirements. Following inspection, an additional sorting stage could categorize objects based on their detected status, thereby optimizing the process and increasing throughput.

Furthermore, we recognize the importance of continuously improving the system's machine learning capabilities, presented in Chapter 4. By retraining it to recognize a broader range of objects and defects, we ensure its adaptability to evolving inspection requirements. This ongoing development ensures that the system remains responsive to changes in the industrial landscape, allowing for seamless integration into diverse production environments.

In summary, our inspection system is designed for expandability and reconfigurability, ensuring it not only meets current needs but also adapts effectively to future challenges in a rapidly changing industrial landscape. It addresses the complexities of detecting defects across objects captured by four cameras, resulting in a dataset with unique challenges. This unique configuration necessitates specialized labeling and training strategies to effectively interpret and utilize the data. Moreover, the limited size and specialized requirements of our dataset mean traditional machine learning models are inadequate. These constraints make it essential to develop a custom deep learning model, not yet documented in existing literature, which will be the focus of the next chapter.

Chapter 4

Machine Learning Framework

The integration of machine learning into industrial automation and quality control marks a pivotal advancement in enhancing the precision and efficiency of classifying objects as either standard or defective. This chapter presents the development and deployment of a robust machine learning framework tailored to overcome the distinct challenges encountered in industrial settings. Emphasizing the complete machine learning workflow, we detail the progression from initial data collection and labeling through to model training, evaluation, and real-time implementation. This framework is designed to achieve reliability and scalability, adhering to the typical requirements of these environments.

The subsequent sections detail the various phases within the machine learning framework, starting with the gathering of a diverse dataset and the application of two distinct labeling techniques. Attention then shifts to preprocessing efforts intended to improve the dataset for machine learning, outlining the training and testing strategies devised to confirm the model's effectiveness in real-world environments. This comprehensive approach highlights the synergy between theoretical concepts and their practical application, spotlighting the strategic thought process behind each step of the framework's development.

4.1 Data Acquisition

4.1.1 Data Description and Statistics

The dataset under examination comprises images focused explicitly on industrial electrical transformers. These images are systematically categorized into two principal groups: those

depicting 'standard' conditions and those illustrating 'defective' states. The collection process involved capturing multiple high-resolution images for each transformer, employing various angles and orientations to ensure a detailed visual record.

Given the natural variance in the sizes of industrial electrical transformers, our research strategically narrows its focus to those measuring approximately 5x5x5 cm. This deliberate limitation allows for including various shapes, such as cylindrical, cubic, spherical, and irregular geometries, while maintaining size consistency across the dataset. Such uniformity is critical for enhancing our specialized focus algorithm, which adeptly selects the optimal focus from a series of images, as elaborated in Section 4.3.1. By confining the objects within a predetermined size range, we streamline the data collection process by applying specific focus settings and lay a foundational framework conducive to the efficiency of real-time applications.

The dataset is further enriched by cataloging many defects, including scratches, missing structures, cracks, and structural deformities. Figure 4.1 illustrates these for reference. Each defective object is paired with its regular counterpart to offer a direct point of comparison for the learning algorithm. This pairing is pivotal for the model to distinguish between defective and non-defective features effectively.

Specific constraints were methodically applied during its constitution to maintain the integrity and balance of the dataset. As previously mentioned, we excluded objects exceeding the dimensions of 5x5x5 cm. Moreover, objects that lacked a standard or defective counterpart were omitted. The absence of paired examples would hinder the model's ability to learn distinctive features relevant to the classification task and could introduce bias or uncertainty into the predictive process. This exclusion criterion ensures that the dataset remains balanced and the learning is grounded in direct, comparative analysis between matched pairs of standard and defective items.

Following the presented selection criteria, the dataset was compiled to include a collective sum of 16 objects, with an equal number of 8 standard items and 8 corresponding defective counterparts. To capture a comprehensive array of visual data for each object, we employed a technique of altering the object's orientation and utilizing four cameras to take images from multiple perspectives, as depicted in Figure 4.2, showcasing different orientations for a single object. This method yielded approximately 100 images for each object in its standard state and around 130 for each in its defective state, resulting in a total collection of 2000 images. This deliberate and extensive data-gathering approach served as our primary data augmentation strategy, significantly enriching the dataset with a wide range of visual information.

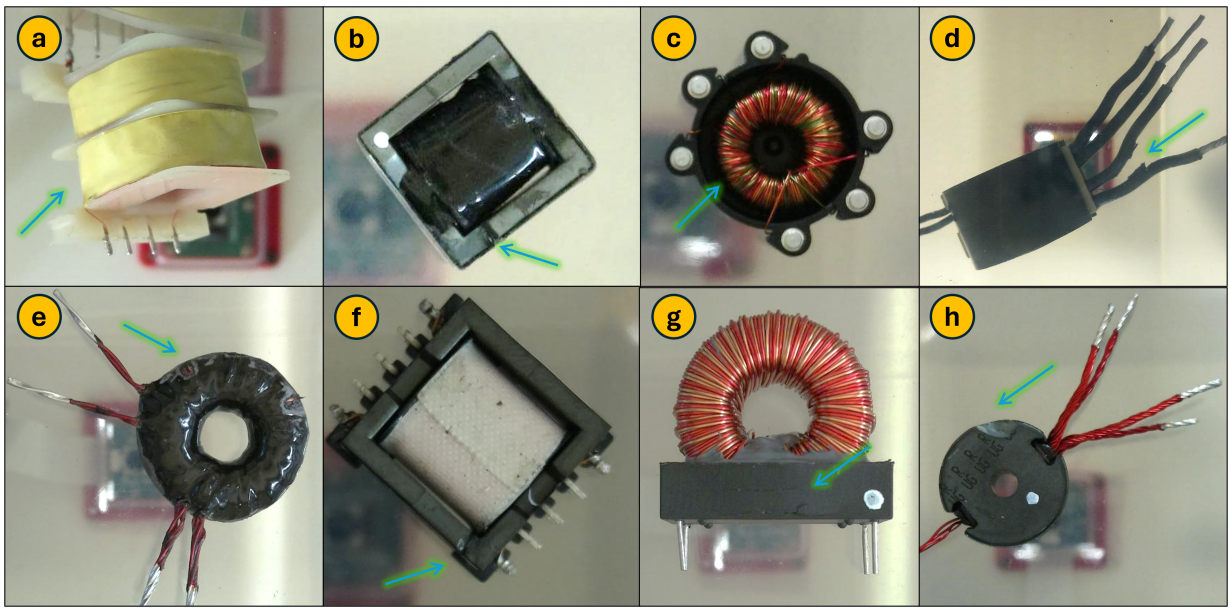


Figure 4.1: An illustration of the types of defects in the dataset with green arrows indicating the specific defects on each object. (a) Absence of the outer casing. (b) A crack located at the object's corner. (c) Lack of epoxy surrounding the wiring. (d) A cut within the cables. (e) A scratch on the plastic casing. (f) A crack on the external surface of the object. (g) A crack on the object's front portion. (h) A crack on the top area of the object. The brightness of the image has been enhanced to highlight the defects.

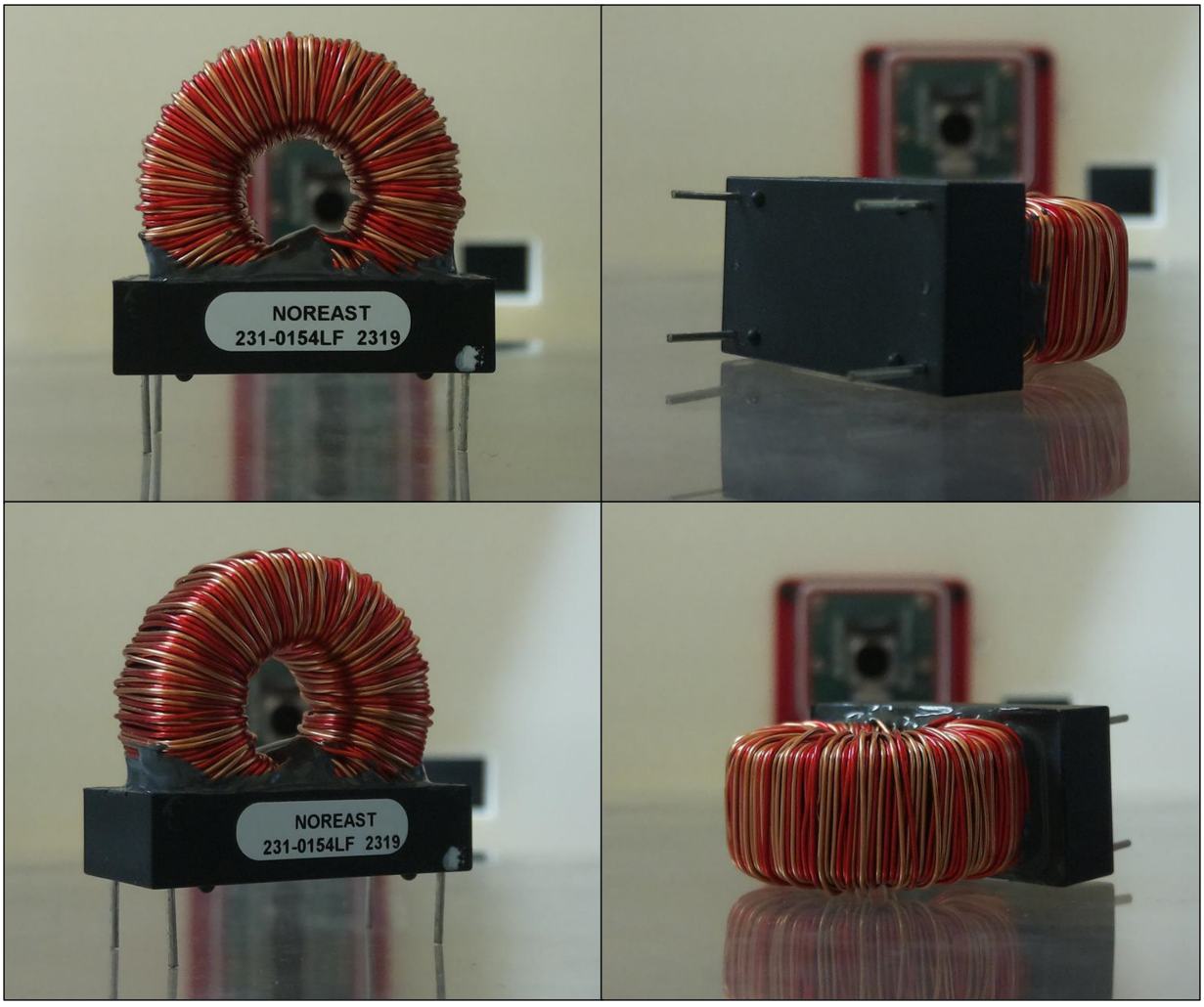


Figure 4.2: A visual representation showing various orientations of an object both for data augmentation and to accommodate the range of possible orientations the object may have when cameras capture it.

4.1.2 Acquisition Process

The data collection method is specifically designed to manage the variability in how an object might be oriented when it appears in front of the cameras, with the consistent premise that the object will be centred. This necessity to capture multiple orientations addresses the unpredictable nature of the object's presentation and serves as a crucial strategy for data augmentation. With the finite number of objects available for our study, maximizing the informational value of each by documenting it in a wide range of visual contexts is imperative. This approach aims to expand the dataset's diversity and informational richness. It comprehensively covers each object through approximately 11 different orientations on average, utilizing four distinct faces that align with our camera setup.

Initially, a range of foci was captured for each camera, from which an algorithm described in Section 4.3.1 was employed to select the image with the optimal focus. As repetitive data acquisition was conducted, it was observed that, due to the fixed approximate size of the objects, the algorithm consistently selected the exact focus for each camera. Therefore, we adjusted our acquisition strategy accordingly. To boost efficiency and ensure the system is more conducive to real-time applications, we opted to set a fixed focus for each camera. This change meant we could forgo the process of cycling through multiple focus settings.

A significant challenge encountered in the dataset's composition was the imbalance, which could potentially lead to biased results. An imbalance, with significantly more defective items than standard ones, could increase the false positives rate, while the opposite could inflate the false negatives rate [47]. We undertook additional data collection efforts to mitigate this problem and ensure a balanced dataset. Furthermore, the data augmentation techniques described in Section 4.3.2 were applied, which were crucial due to the limited number of objects available. These measures collectively ensure that our dataset is not only extensive and varied but also balanced, thereby supporting the development of a more accurate and reliable detection model.

4.2 Data Labeling

The approach to labelling the dataset was primarily influenced by the nature of the objects we are studying and the intended application of the machine learning model. A fundamental consideration was whether to label an object's entirety, encompassing all four faces in one comprehensive annotation, or label each face of the object independently. The

latter method was chosen, and here we discuss the rationale behind this decision and its implications for data augmentation and machine learning efficacy.

Assigning a single label to the entire object provides a varied perspective of each item in one instance. However, this approach results in limited data points or examples for machine learning algorithms. In addition, labelling an object in its entirety would necessitate combining or concatenating images of each object from four different angles into a single, unified representation. This process introduces considerable challenges because of the technical complexities of merging these perspectives into a coherent and usable format for machine learning purposes. Moreover, given the limited number of objects available in our dataset, such an approach would amplify the risk of overfitting, as the model might learn to recognize specific configurations of concatenated images rather than generalize from the defects' characteristics.

Our approach acknowledges that object defects are not visible from every angle, as illustrated in Figure 4.3, emphasizing the importance of a machine learning model's ability to detect these imperfections from the most revealing perspectives. This visibility challenge introduces complexity into the process of labelling objects in their entirety, as errors might occur due to misunderstandings about the presence or appearance of defects from various viewpoints. Opting to focus on the individual faces of the objects simplifies data processing and sharpens the focus on critical characteristics for identifying defects. Meanwhile, labelling the whole object as a single entity presents a promising avenue for future research, suggesting the possibility of developing new methods to navigate and surmount these existing challenges.

4.2.1 Binary Class Labeling

The intuitive approach to categorizing the faces of objects is through binary classification, labelling them as either "standard" or "defective." In this context, "standard" labels correspond to faces of standard, defect-free items, whereas "defective" labels are assigned to faces of objects with visible defects. For items without any defects, all faces are uniformly labelled as "standard." However, the scenario is more complex for items identified as defective. These may exhibit both "defective" faces with defects and "standard" faces without, indicating that imperfections don't uniformly affect all sides of an item.

The binary classification system introduces a significant learning challenge due to the variability in defects and objects. Defects can differ significantly in appearance, size, and severity, while objects vary in shape, size, and material. This variability demands that the

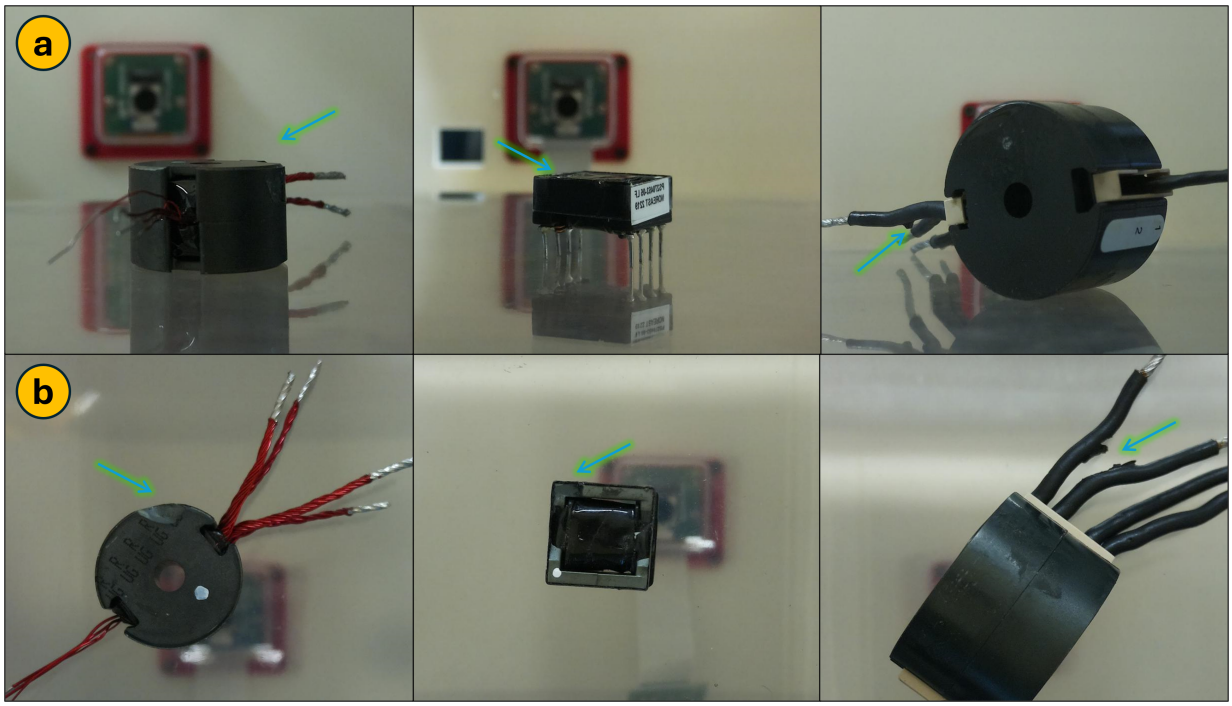


Figure 4.3: Challenges in data labeling. Row (a) presents faces with ambiguous indications of defects, making it difficult to determine their presence, while row (b) displays images of the same defect but on different faces, where the imperfections are unmistakably visible.

model learn to accurately identify and differentiate between "standard" and "defective" faces amidst a spectrum of defect manifestations and object characteristics.

A graphical user interface (GUI) was developed to simplify the labelling process. It displays each object's face for efficient classification. Users can quickly label faces as "standard" if they are defect-free, "defective" if defects are present, or "uncertain" if further assessment is needed. The "standard" and "defective" labels are automatically sorted into respective folders, creating an organized dataset. Faces marked as "uncertain" are subject to a more thorough review to either ensure their categorization into the correct folder or to exclude them from the dataset if a precise classification cannot be determined.

4.2.2 Multi-Class Labeling

In our pursuit of refining the labelling process for enhanced model performance, we explored the concept of multi-class labelling. This approach involved labelling faces not just as "standard" or "defective" but coupling these conditions with the object's name, effectively creating a label that combines the object's identity with its condition (e.g., "ObjectX_standard" or "ObjectY_defective"). The rationale behind this modification was to investigate whether associating the defect status directly with a specific object would simplify the model's task of distinguishing between classes.

The potential for improved model performance motivated the shift towards multi-class labelling. This method labels data by combining each object's identity with its condition (standard or defective), theorizing that a model could more easily differentiate between specific cases. Unlike binary labelling, which groups all "standard" and all "defective" conditions without distinguishing between different objects or defects, multi-class labelling aims to create distinct categories for each object and its state.

4.3 Data Preprocessing

In preparing our dataset for analysis, we focused on two preprocessing steps: selecting the best focus and implementing data augmentation. Choosing the optimal focus includes cropping the images to ensure each object is displayed with excellent clarity in preparation for the focus algorithm. Data augmentation then broadens the variety of our dataset, helping our models learn from a more diverse set of images. These initial steps are crucial as they refine our dataset, making it more suitable for the detailed analysis and training that follows.

4.3.1 Best Focus Selection

Choosing the proper focus is essential to improve the sharpness and detail of images, especially when inspecting objects. To ensure the main object is the focal point, images are first cropped to reduce background noise and distractions. We then experimented with various focus settings, capturing images at each one. An autofocus algorithm reviewed these images to identify the focus setting that yielded the clearest picture. This method ensures that the focus is precisely adjusted for optimal object clarity.

The focus algorithm employs a thresholded gradient method inspired by Santos et al. technique [48]. Labelled 'GRAT,' this approach evaluates the image's focus by analyzing gradients, which highlight variations in brightness often seen at edges and textures, crucial for determining image focus. Here's how it works:

- Initially, we set a threshold (Th) to zero, ignoring gradient values that barely affect the focus assessment.
- We calculate the horizontal (I_x) and vertical (I_y) gradients by measuring the differences in intensity across these directions.
- The focus measure (FM) is identified by selecting the larger absolute values between I_x and I_y, disregarding any below the threshold.
- To quantify focus, the FM is normalized by the total of all significant gradient values, creating a numerical focus indicator. This process ensures the autofocus algorithm can precisely gauge the sharpness and clarity of the image based on structural details.

Through repeated experiments, we observed that the constant distance between the objects and the cameras and the objects' similar sizes led the algorithm to choose the exact focus for each camera regularly. To make the process more efficient, we set a fixed focus for each camera, using the focus setting the algorithm most often selected. This decision simplifies our setup and guarantees a uniform focus in all imaging sessions, enhancing the speed and dependability of our inspection workflow. Moreover, this method significantly enhances our capability for immediate, on-the-spot inspections. It increases the volume of items we can process in a given time-frame, making our overall operations more efficient.

4.3.2 Data Augmentation

Data augmentation is critical in training machine learning algorithms, especially in image classification, where datasets may be limited or unbalanced. The primary objective of

data augmentation is to artificially increase the size and diversity of the training dataset by applying various transformations to the original images. This process helps prevent the model from overfitting and enhances its ability to generalize well to new, unseen images by simulating a wider variety of scenarios and variations that the model might encounter in real-world applications.

In the context of training a ResNet model [9] for classifying defective versus standard objects, we initially employed a wide range of image augmentation techniques using OpenCV [43]. These included transformations such as center crop, noise addition, blur, sharpness adjustments, dimming, brightening, rotation, and flipping. However, upon evaluating the model's performance, it became evident that not all augmentation techniques contributed positively. We hypothesize that this discrepancy arises because certain augmentations, such as center cropping and adding noise, do not accurately reflect the variations observed in real-world settings. Such transformations might introduce artificial patterns or remove critical information from images, decreasing the model's ability to learn meaningful features from the data.

Given these insights, we decided to refine our augmentation strategy by focusing on two techniques: rotation and flipping. These transformations were selected based on their relevance to the real-world scenarios in which the objects might be presented in various orientations and directions. This targeted approach to data augmentation is expected to facilitate more effective learning and improve the model's generalization ability, thereby achieving superior classification performance on both defective and standard objects.

4.4 Model Training and Testing

In formulating our model training and evaluation strategy, we deployed ResNet [9] architectures, exploring versions from ResNet-18 to ResNet-101. Initial tests, which were limited to a small set of classes, showed that ResNet-101 achieved the highest accuracy. Consequently, we selected ResNet-101 for subsequent experiments to capitalize on its enhanced performance.

ResNet stands out among other deep learning models for its efficacy in enhancing the capabilities of visual inspection systems within industrial contexts, as presented in Section 2.2.3. Its ability to process images through multiple layers, progressively recognizing increasingly complex features, is crucial for detecting complex defects like cracks and corrosion, particularly in critical components such as electrical transformers. Additionally, the integration of ResNet models into industrial inspection systems significantly

improves defect detection accuracy, ensuring robust quality assurance and high reliability in precision-dependent industries.

With respect to hyperparameter selection, our choices were carefully aligned with our specific use case requirements. A batch size of 4 was adopted to mirror the operational scenario where four images are captured for each object for immediate classification analysis. This setup enables concurrent processing of all images of an object, facilitating efficient and prompt decision-making. The training regimen was configured to last up to 500 epochs, with an early stopping criterion to terminate training when no further enhancements in model performance were detected. This approach helps in maximizing training efficiency while ensuring model efficacy.

Regarding data preparation, images were adjusted to a resolution of 512x512 pixels, and the dataset was partitioned using a split of 70% for training, 20% for validation, and 10% for testing. The learning rate was determined to be 0.0001, a decision based on initial experiments to balance convergence speed and learning process stability. The training was executed using Python 3.10.12 and PyTorch 2.0.1, with CUDA 12 support to utilize GPU acceleration, thereby boosting the training process's efficiency. Essential libraries incorporated in our workflow include PyTorch [49] and OpenCV [43].

The computational resources for training included a 13th Gen Intel® Core™ i9-13900HX processor and an NVIDIA® GeForce RTX™ 4070 GPU with 8 GB GDDR6 memory, supported by 32 GB of DDR5 RAM. This powerful combination provided a solid foundation for the demanding task of training the deep learning model. For real-time testing purposes, a Raspberry Pi (described in Section 3.2.1) was utilized.

4.5 Results and Discussion

In this section, we present the outcomes of various experiments designed to evaluate the performance of deep learning models, specifically ResNet, across its varying depths in image classification tasks. These experiments span from binary to multi-class classification and explore the impact of data augmentation techniques in addressing dataset imbalance.

4.5.1 Initial Experiments

In the preliminary phase of our study, we focused on binary classification tasks involving a small set of six objects, divided equally between three defective and three standard

types. We applied basic data augmentation techniques such as flipping and rotation to augment the limited dataset and introduce variability. These steps enhanced the dataset's robustness, aiding our models' training process.

To establish a baseline for further research, we assessed the performance of various ResNet architectures, including ResNet-18, ResNet-34, ResNet-50, and ResNet-101, selected for their range in complexity and depth. Recorded accuracies were 48% for ResNet-18, escalating to 80% for ResNet-101, indicating that models with greater depth are more adept at discerning the nuanced distinctions required for binary classification in a limited dataset. A detailed comparison is provided in Table 4.1 that summarizes accuracy, model depth, number of epochs, and training loss, elucidating the relationship between network depth, data augmentation, and performance, thereby enhancing our understanding of their collective impact on outcomes.

Table 4.1: Summary of initial experiments with binary classification.

Model	**Accuracy (%)**	**Depth**	**Epochs**	**Training Loss**
ResNet-18	48	18	11	0.19
ResNet-34	68	34	11	0.13
ResNet-50	76	50	21	0.04
ResNet-101	80	101	21	0.03

These initial findings are insightful, indicating a positive correlation between model depth and binary classification accuracy. This observation is particularly relevant for the design of future experiments and for applications requiring nuanced data interpretation. As our research progresses towards more complex classification challenges and refined augmentation techniques, these initial experiments serve as a valuable reference point, demonstrating the importance of model depth in achieving high classification performance.

4.5.2 Balancing Datasets with Augmentation

In addressing the challenge of dataset imbalance in our experiments, we adopted a data augmentation strategy not only to increase the representation of defective objects but also to expand the dataset of standard objects. The primary augmentation technique employed was rotation. We systematically rotated images by a series of angles where the selection of these angles was directly proportional to our objective of expanding the dataset to achieve balance. This method not only augmented the number of images but also diversified the representation of object feature appearances, aiming to simulate real-world conditions under which a defect might manifest in different orientations.

To further augment the diversity of our dataset, we incorporated both horizontal and vertical flipping. This approach effectively doubled the dataset's size, introducing mirror variations of each image. Lastly, to simulate the effect of real-world environmental conditions that could affect the appearance of objects, we added Gaussian noise to the pictures. This type of noise introduction mimics the impact of factors like sensor noise or variations in lighting conditions that might occur during the operational use of the classification model.

The results of the experiment demonstrated how balancing datasets affects model accuracy. We trained two models: one with a balanced dataset containing about 3000 images and the other with an unbalanced set of roughly 2200 images, each using a 70-20-10 training, validation, and testing split. The balanced model reached an 80% accuracy, while the unbalanced one achieved only 72%, as summarized in Table 4.2. Significantly, the unbalanced model showed a much higher rate of false positives, emphasizing the importance of dataset balance in improving model precision and reliability.

Table 4.2: Summary of main experiments using balanced and unbalanced dataset, with binary and multi-class classification, using ResNet-101.

Dataset Type	**Dataset Size**	**Accuracy (%)**
Unbalanced	2200	72.0
Balanced	3000	80.0
Binary	7000	69.2
Multi-Class	7000	**90.8**

4.5.3 Multi-Class Classification

After achieving a balanced dataset through augmentation, we transitioned from binary to multi-class classification, broadening our framework to include a diverse range of objects and their conditions. This strategic expansion aimed to support the model's ability to discriminate effectively between standard and defective items across multiple object types.

In the comparative analysis, we used a dataset of approximately 7000 images with a consistent 70-20-10 split for training, validation, and testing. This applied to binary (one standard and one defective) and multi-class classification, encompassing 16 classes (eight standard and eight defective). Despite the identical dataset size and split, the outcomes were markedly different. The binary classification model achieved a 69.2% accuracy rate on the test data, whereas the multi-class model notably excelled, reaching a 90.8% accuracy rate, as illustrated in Table 4.2.

This contrast underscores the impact of classification strategy on model efficacy. By differentiating among sixteen distinct categories, the multi-class model provided a more refined understanding of the dataset. This approach enhanced the model's discriminative capability and demonstrated the advantages of employing a detailed classification scheme, particularly when the dataset supports such granularity. The results affirm that a comprehensive classification framework can lead to superior model performance, even when operating within the confines of the same dataset size and distribution.

4.6 Real-time Implementation and Testing

4.6.1 Implementation Details

This research studies our machine learning model's real-world deployment and evaluation, showcasing the fundamental differences from traditional model testing approaches. A notable divergence in our approach is the employment of a Raspberry Pi (presented in Section 3.2.1) for real-time operations, which relies on a CPU instead of a GPU for processing. This change requires modifications in the model's deployment and running due to the CPU's distinct processing power and efficiency. Considering the superior performance of the multi-class classifier over the binary classifier, we chose it for our real-time application.

A notable challenge in real-time testing is the absence of predefined test data and ground truth, which traditionally serve as benchmarks for measuring model accuracy. To navigate this, we implemented a system where each model decision triggers a specific message output. In our setup, four images are simultaneously processed, each undergoing initial preprocessing steps such as cropping and resizing. The model, configured with a batch size of four, evaluates each image to determine whether an object is defective or standard and identifies the object type.

To further refine our system, we developed a script consolidating the four decisions associated with each object. If any of the four faces is classified as defective, the classifier outputs a "Defective" message and activates a red LED indicator, as detailed in Section 3.3. Conversely, a "Standard" alert is generated if all faces are standard, alongside a green LED signal. We also integrated a feature to address instances where the model's confidence level is below a certain threshold, which we set at 75%, a decision shaped by the practical needs of our industry partner. This threshold was selected to align with real-world applications where our partner deems it acceptable, and a lower threshold would be impractical. If the model's confidence in its decision exceeds this threshold, the system proceeds with the

standard or defective classification. If not, it issues a "Further inspection needed" message accompanied by a yellow LED, signalling the need for manual object inspection.

4.6.2 Model Repeatability and Adaptability

It's noteworthy to highlight that our trained model demonstrated exceptional performance in real-time scenarios. Through extensive experiments encompassing all available objects and their variants, the model consistently and accurately identified each object's type and condition, emphasizing its robustness and reliability. Further validation of the system's reliability was confirmed through extensive real-time testing involving multiple users. These users placed objects at random orientations and positions within the inspection setup, encompassing various lighting conditions and setup locations. Despite these variations, the model consistently and accurately predicted the status of the objects every time. This consistent performance across diverse user interactions highlights the system's readiness for industrial application and its capability to operate effectively under varied and unpredictable conditions. Such repeatability ensures the system's reliability, rendering it a dependable solution for industrial environments where accuracy and consistency are paramount.

We introduced an unfamiliar object to explore the model's adaptability and see how it would respond. Interestingly, the model treated each face of this novel object as if it belonged to a different known object, a reasonable reaction given its lack of prior exposure. What was particularly enlightening was that the model's predictions were not just random guesses but closely aligned with specific objects, with varying confidence levels, some as high as over 80% for particular faces and others lower, reflecting a lesser degree of feature resemblance.

Visual evidence of this is provided in Figure 4.4, where the top row displays various faces of the unfamiliar object. In contrast, the bottom row presents the objects the model predicted these faces to correspond with. Notably, images (e) and (f), both associated with high confidence levels, share a prominent yellow colour with the unfamiliar object, suggesting that colour is a highly influential feature for the model. Conversely, image (g), which features words or letters, resulted in moderate confidence levels, indicating lesser importance of textual features. The prediction for image (h), which was around 30%, points to size being the least significant feature for the model.

These observations reflect the model's capacity for high accuracy in familiar scenarios and its ability to apply learned patterns to novel situations. Certain features, like colour, are crucial in its decision-making process.

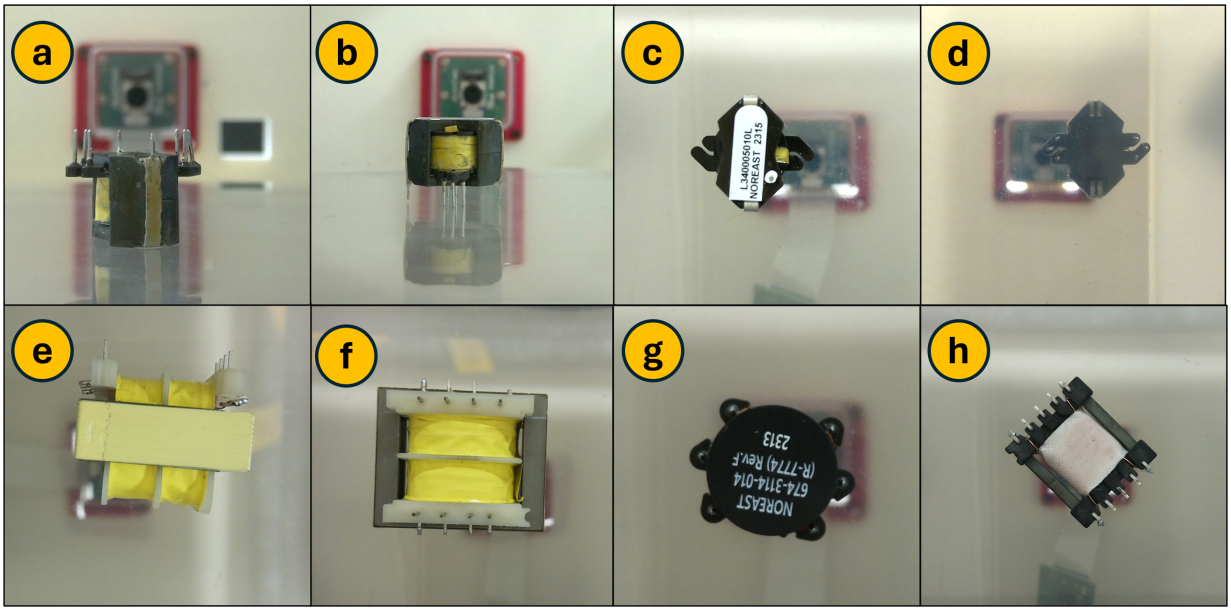


Figure 4.4: Realtime model predictions for an unseen object. Panels (a) to (d) display different facets of the unfamiliar object. Panels (e) to (h) illustrate the model's predictions, aligning each face with a known object. High confidence in (e) and (f) indicates a significant influence of colour. Moderate confidence in (g) hints at the lesser importance of textual features. Minimal confidence in (h) implies size as the least impactful characteristic.

The framework presented in Chapters 3 and 4 was put into practice with our industrial partner. It takes the system 1 minute and 6 seconds to capture all the required images and render a decision for automatic inspection. Our industrial partner has noted that this timeframe is quite reasonable for the intended application, particularly compared to the existing inspection rates.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

This thesis has conducted a comprehensive exploration into integrating machine learning into industrial automation, with a particular focus on transformer inspection. Beginning with a motivation to address labor market constraints and enhance operational efficiency, the project developed an end-to-end system for automatic inspection that includes data collection, augmentation, and labeling of objects. Utilizing various depths of ResNet architectures, this work has advanced automation technologies within industrial settings, significantly improving fault detection.

The need to reduce reliance on skilled labor and improve inspection precision was a driving force behind this research. The design, implementation, and evaluation of a Raspberry Pi-based monitoring system, combined with strategic applications of deep learning models, were discussed, demonstrating significant improvements in the accuracy and efficiency of transformer inspections.

In the evaluation of the system, a notable achievement was the attainment of an accuracy rate exceeding 90%. This high level of precision significantly indicates the robustness of the machine learning model developed within this thesis, illustrating its effectiveness in handling complex defect detection tasks. The high level of precision attained not only demonstrates the sophisticated capabilities of the ResNet architecture but also reflects the thorough training and optimization processes that were implemented.

Further validation of the system's reliability was confirmed through extensive real-time testing, where multiple users participated by placing objects at random orientations and

positions within the inspection setup. Regardless of these variations, including different lighting conditions and locations of the setup, the model consistently and accurately predicted the status of the objects every time. This consistent performance across diverse user interactions highlights the system's readiness for industrial application and its capability to operate effectively under varied and unpredictable conditions. This repeatability ensures the system's reliability, making it a dependable solution for industrial environments where accuracy and consistency are critical.

5.2 Future Work

This section outlines several strategic directions and enhancements to improve the practical application and efficiency of the inspection system. These enhancements aim to streamline operations, accommodate a wider variety of objects, and expand the system's learning capabilities to dynamically adapt to new challenges.

5.2.1 Automation Enhancements

To fully automate the object handling processes, integrating a feeder mechanism to systematically feed objects into the inspection area is essential. A conveyor system equipped with a mechanical actuation mechanism could ensure objects are accurately positioned at the camera station. The precise stopping of the conveyor would allow each object to be properly aligned for inspection. After inspection, introducing a sorting stage could categorize objects based on their detected status, thereby streamlining the process and increasing throughput.

To address the aforementioned system requirements for automation, we proposed an initial design (illustrated in Figure 5.1) for a stopping mechanism consisting of a servo with a rack and pinion setup. This addition serves two critical purposes in the system's design and functionality. Firstly, the stopping mechanism ensures that parts are halted precisely at the location for inspection, aligning accurately within the field of view of the four cameras.

Secondly, the stopping mechanism acts as a staging area for sorting products into 'standard' and 'defective' categories based on the inspection results. This functionality is integral to the quality control process, ensuring that only parts meeting quality standards proceed to the next stage of manufacturing, while defective items are rerouted or discarded.

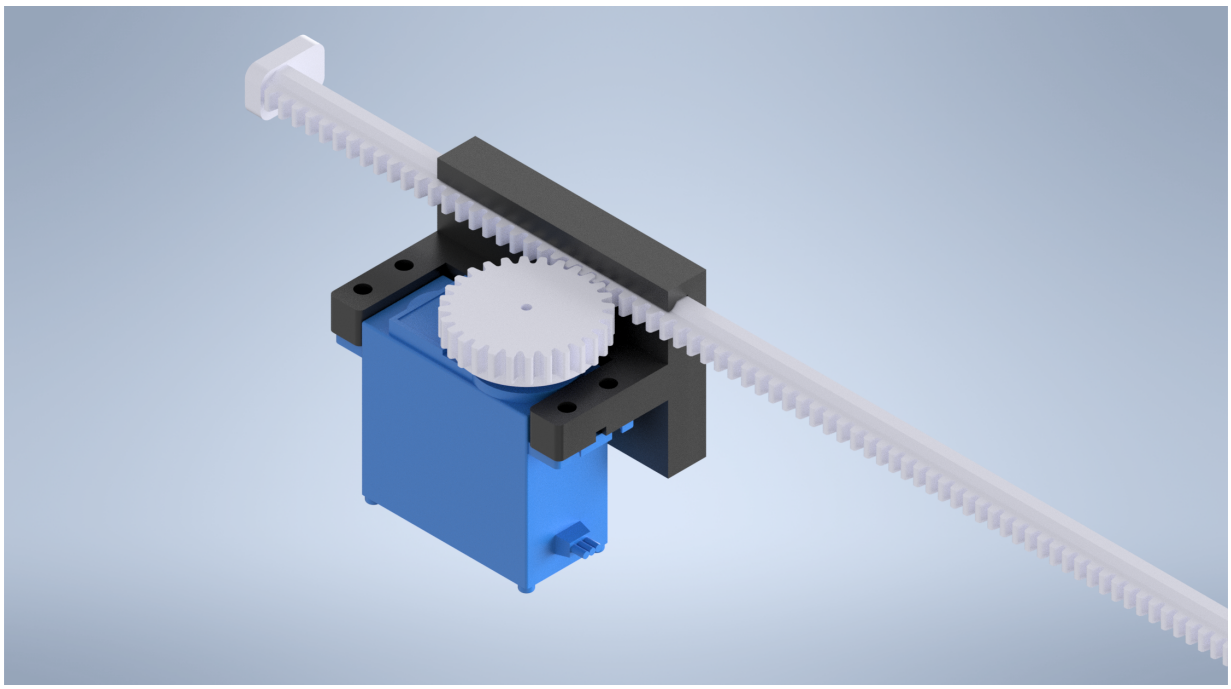


Figure 5.1: Stopping mechanism design. The design consists of a servo with a rack and pinion setup.

Looking forward, the stopping mechanism's design presents several opportunities for further enhancements and applications in subsequent stages of the production line. Future modifications could include refining the mechanism's precision and responsiveness to accommodate a wider range of part sizes and shapes, enhancing the system's adaptability and efficiency. Additionally, integrating advanced sensors or feedback systems could further optimize the sorting process, reducing human oversight requirements and increasing throughput.

The concept and design of this stopping mechanism could also be adapted and applied to other stages of production where precise positioning and sorting are necessary. Such applications could extend beyond quality inspection to include operations like assembly, packaging, or more complex sorting tasks that require high accuracy.

Further enhancements of the conveyor system should focus on accommodating different sizes and shapes of objects and incorporating features like autofocus to handle variations in object dimensions. Refining the sorting capabilities to manage a diverse range of inspected items would contribute significantly to the scalability and adaptability of the system, ensuring it remains effective across various industrial applications.

5.2.2 Advanced Learning Architectures

Exploring more advanced deep learning architectures could push the boundaries of accuracy and generalization capabilities beyond the current implementations. A vital area for development is the expansion of the training dataset and the incorporation of zero-shot learning (ZSL) techniques [50]. ZSL would facilitate the easier expansion of the dataset to include a wider array of conditions and variables, allowing the model to adapt more swiftly and effectively to new or unseen defect types without extensive retraining.

Zero-shot Learning (ZSL) for Enhanced Feature Matching

ZSL, where models perform tasks on data they have not explicitly seen during training, offers a promising method to enhance the adaptability and accuracy of systems in real-world applications. ZSL operates by leveraging a model's ability to infer properties and categories based on learned representations, even when no direct examples of those categories have been presented during the training phase. This capability is particularly valuable in environments where new object types frequently emerge and collecting extensive labeled data for each type is impractical.

Our proposed approach involves extracting image features using a pretrained ResNet model and generating text embeddings with a pretrained BERT model [51]. BERT, or Bidirectional Encoder Representations from Transformers, excels at processing language by considering the context of words from both the left and the right sides of a word within a sentence. This makes BERT particularly effective for generating rich, contextualized embeddings that capture the nuances of language, which are essential for matching text descriptions with visual data accurately.

The semantic mismatch between the visually-oriented image features and contextually-rich text embeddings poses significant challenges, impacting the effectiveness of the system. Initially, we utilized a ResNet-101 model pretrained on ImageNet, which resulted in a low accuracy of 24%. However, when we utilized our ResNet-101 model from Section 4.5.1 trained on six of the object categories (three standard and their corresponding three defective objects) and then tested it on the same six categories, the accuracy significantly improved to 72%. Further testing on a broader set of 16 categories, including ten that the model had not previously seen, resulted in a similar accuracy of 76%. This progression displays the potential of ZSL capabilities in our model.

To further enhance our promising preliminary results, future work will focus on incorporating hybrid models like VisualBERT [52] or VilBERT [53], which are designed to integrate image and text data seamlessly. These models could reduce the semantic mismatch by treating the integration of modalities within a unified framework.

In conclusion, while this thesis represents a substantial step forward in the application of deep learning within industrial monitoring, it also sets the stage for future innovations in automated monitoring and maintenance technologies. The path toward fully automated systems is still unfolding, and the insights gained here pave the way for continued advancements in this dynamic field.

References

[1] W.-Y. Lin, C.-Y. Lin, G.-S. Chen, and C.-Y. Hsu, "Steel Surface Defects Detection Based on Deep Learning," in *Advances in Physical Ergonomics & Human Factors* (R. S. Goonetilleke and W. Karwowski, eds.), vol. 789, pp. 141–149, Cham: Springer International Publishing, 2019. Series Title: Advances in Intelligent Systems and Computing.

[2] X. Wei, Z. Yang, Y. Liu, D. Wei, L. Jia, and Y. Li, "Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study," *Engineering Applications of Artificial Intelligence*, vol. 80, pp. 66–81, Apr. 2019.

[3] A. Bhattacharya and S. G. Cloutier, "End-to-end deep learning framework for printed circuit board manufacturing defect classification," *Scientific Reports*, vol. 12, p. 12559, July 2022. Publisher: Nature Publishing Group.

[4] X. Wang, S. Gao, J. Guo, C. Wang, L. Xiong, and Y. Zou, "Deep Learning-Based Integrated Circuit Surface Defect Detection: Addressing Information Density Imbalance for Industrial Application," *International Journal of Computational Intelligence Systems*, vol. 17, p. 29, Feb. 2024.

[5] A. Valyanskii, I. Kartashev, and Y. Sharov, "The effect of power quality on power-transformer reliability," *Russian Electrical Engineering*, vol. 85, pp. 138–144, Mar. 2014.

[6] S. C. Government of Canada, "Economic performance associated with digitalization in Canada over the past two decades," Feb. 2021. Last Modified: 2021-02-24.

[7] S. Canada, "Canada's population estimates: Age and sex, as of july 1, 2022," 2022. Accessed: 2024-04-15.

[8] "These are the top 10 tech trends that will shape the coming decade, according to McKinsey," Oct. 2021.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

[10] Y. Liu, C. Zhang, and X. Dong, "A survey of real-time surface defect inspection methods based on deep learning," *Artificial Intelligence Review*, vol. 56, pp. 12131–12170, Oct. 2023.

[11] S. C. Government of Canada, "The Daily — Labour Force Survey, December 2023," Jan. 2024. Last Modified: 2024-01-05.

[12] Y. Chen, Y. Ding, F. Zhao, E. Zhang, Z. Wu, and L. Shao, "Surface Defect Detection Methods for Industrial Products: A Review," *Applied Sciences*, vol. 11, p. 7657, Jan. 2021. Number: 16 Publisher: Multidisciplinary Digital Publishing Institute.

[13] "Labour shortage trends in Canada."

[14] M. Saritas, Y. Taspinar, I. Cinar, and M. Koklu, *Railway Track Fault Detection with ResNet Deep Learning Models*. Apr. 2023.

[15] C. E. Hachem, G. Perrot, L. Painvin, and R. Couturier, "Automation of Quality Control in the Automotive Industry Using Deep Learning Algorithms," in *2021 International Conference on Computer, Control and Robotics (ICCCR)*, pp. 123–127, Jan. 2021.

[16] D. Zheng, L. Li, S. Zheng, X. Chai, S. Zhao, Q. Tong, J. Wang, and L. Guo, "A Defect Detection Method for Rail Surface and Fasteners Based on Deep Convolutional Neural Network," *Computational Intelligence and Neuroscience*, vol. 2021, p. e2565500, Aug. 2021. Publisher: Hindawi.

[17] J. Shen, N. Liu, and H. Sun, "Defect detection of printed circuit board based on lightweight deep convolution network," *IET Image Processing*, vol. 14, no. 15, pp. 3932–3940, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1049/iet-ipr.2020.0841.

[18] F. Saeed, M. J. Ahmed, M. J. Gul, K. J. Hong, A. Paul, and M. S. Kavitha, "A robust approach for industrial small-object detection using an improved faster regional convolutional neural network," *Scientific Reports*, vol. 11, p. 23390, Dec. 2021. Number: 1 Publisher: Nature Publishing Group.

[19] Z. Jia, M. Wang, and S. Zhao, "A review of deep learning-based approaches for defect detection in smart manufacturing," *Journal of Optics*, vol. 53, pp. 1345–1351, Apr. 2024.

[20] D. Soukup and R. Huber-Mörk, "Convolutional Neural Networks for Steel Surface Defect Detection from Photometric Stereo Images," in *Advances in Visual Computing* (G. Bebis, R. Boyle, B. Parvin, D. Koracin, R. McMahan, J. Jerald, H. Zhang, S. M. Drucker, C. Kambhamettu, M. El Choubassi, Z. Deng, and M. Carlson, eds.), (Cham), pp. 668–677, Springer International Publishing, 2014.

[21] Z. Ren, F. Fang, N. Yan, and Y. Wu, "State of the Art in Defect Detection Based on Machine Vision," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 9, pp. 661–691, Mar. 2022.

[22] Z. Wang, W. Xie, H. Chen, B. Liu, and L. Shuai, "Color Point Defect Detection Method Based on Color Salient Features," *Electronics*, vol. 11, p. 2665, Jan. 2022. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.

[23] Z. Wang, H. Chen, W. Xie, and H. Wang, "Color Mura Defect Detection Method Based on Channel Contrast Sensitivity Function Filtering," *Electronics*, vol. 12, p. 2965, Jan. 2023. Number: 13 Publisher: Multidisciplinary Digital Publishing Institute.

[24] C. Yang, "Multi-Camera Systems in Embedded Vision: Features and Applications," Aug. 2023.

[25] Dev22Admin, "The Power of Machine Vision: High-Speed Product Counting on Conveyor Belts," Aug. 2023.

[26] E. Cumbajin, N. Rodrigues, P. Costa, R. Miragaia, L. Frazão, N. Costa, A. Fernández-Caballero, J. Carneiro, L. H. Buruberri, and A. Pereira, "A Systematic Review on Deep Learning with CNNs Applied to Surface Defect Detection," *Journal of Imaging*, vol. 9, p. 193, Oct. 2023. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

[27] B. Dey, E. Dehaerne, K. Khalil, S. Halder, P. Leray, and M. A. Bayoumi, "Deep Learning based Defect classification and detection in SEM images: A Mask R-CNN approach," Nov. 2022. arXiv:2211.02185 [cs].

[28] "J. Imaging | Free Full-Text | A Systematic Review on Deep Learning with CNNs Applied to Surface Defect Detection."

[29] M. Ferguson, R. Ak, Y.-T. T. Lee, and K. H. Law, "Detection and Segmentation of Manufacturing Defects with Convolutional Neural Networks and Transfer Learning," Aug. 2018. arXiv:1808.02518 [cs] version: 1.

[30] S. Jain, G. Seth, A. Paruthi, U. Soni, and G. Kumar, "Synthetic data augmentation for surface defect detection and classification using deep learning," *Journal of Intelligent Manufacturing*, vol. 33, pp. 1007–1020, Apr. 2022.

[31] T. Benbarrad, M. Arioua, and H. Anas, "Performance Evaluation of Transfer Learning for Surface Defect Classification," in *Innovations in Smart Cities Applications Volume 5* (M. Ben Ahmed, A. A. Boudhir, R. Karaș, V. Jain, and S. Mellouli, eds.), (Cham), pp. 977–988, Springer International Publishing, 2022.

[32] V. Sampath, I. Maurtua, J. J. Aguilar Martín, A. Iriondo, I. Lluvia, and G. Aizpurua, "Intraclass Image Augmentation for Defect Detection Using Generative Adversarial Neural Networks," *Sensors*, vol. 23, no. 4, 2023.

[33] Y. Wang, Y. Zhang, L. Zheng, L. Yin, J. Chen, and J. Lu, "Unsupervised Learning with Generative Adversarial Network for Automatic Tire Defect Detection from X-ray Images," *Sensors*, vol. 21, p. 6773, Jan. 2021. Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015. arXiv:1512.03385 [cs].

[35] B. Hu and J. Wang, "Detection of PCB Surface Defects With Improved Faster-RCNN and Feature Pyramid Network," *IEEE Access*, vol. 8, pp. 108335–108345, 2020.

[36] T. P. Link, "Identifying PCB Defects with a Deep Learning Single-Step Detection Model," July 2022.

[37] C. Zhu, H. Hong, H. Sun, G. Wang, J. Shen, and Z. Yang, "Real-Time Damage Detection Method for Conveyor Belts Based on Improved YoloX," *Journal of Failure Analysis and Prevention*, vol. 23, pp. 1608–1620, Aug. 2023.

[38] J. Smith, "A Comprehensive Guide to Effective Conveyor Belt Inspection," Mar. 2024.

[39] "Transparent belt with fluoroscopic fabric and nubby surface - Märtens conveyor belts."

[40] "Backlit Conveyors - Dorner Conveyors."

[41] "Industrial cameras for industrial inspection and automation."

[42] Raspberry Pi Foundation, "Raspberry pi 4 model b," 2024. Datasheet.

[43] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[44] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[45] Arducam, "16mp imx519 autofocus synchronized quad-camera kit for raspberry pi, nvidia® jetson nano/xavier nx/agx orin/orin nano/orin nx," 2024. Datasheet.

[46] Vevor, "Vevor canada," 2024. Accessed: 2024-04-24.

[47] H. He and E. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, p. 1263–1284, Sep 2009.

[48] A. SANTOS, C. ORTIZ DE SOLÓRZANO, J. J. VAQUERO, J. M. PEÑA, N. MALPICA, and F. DEL POZO, "Evaluation of autofocus functions in molecular cytogenetic analysis," *Journal of Microscopy*, vol. 188, p. 264–272, Dec 1997.

[49] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, pp. 8024–8035, Curran Associates, Inc., 2019.

[50] W. Wang, V. W. Zheng, H. Yu, and C. Miao, "A survey of zero-shot learning," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, p. 1–37, Jan 2019.

[51] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Association for Computational Linguistics, 2019.

[52] L. H. Li, M. Yatskar, D. Yin, C.-J. Hsieh, and K.-W. Chang, "Visualbert: A simple and performant baseline for vision and language," in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

[53] J. Lu, D. Batra, D. Parikh, and S. Lee, "Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks," in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

[54] "Browse Inspection Conveyors | Backlit & Translucent | mk."

[55] M. Mayr, M. Hoffmann, A. Maier, and V. Christlein, "Weakly Supervised Segmentation of Cracks on Solar Cells using Normalized Lp Norm," Jan. 2020. arXiv:2001.11248 [cs].

[56] G. Li, R. Shao, H. Wan, M. Zhou, and M. Li, "A Model for Surface Defect Detection of Industrial Products Based on Attention Augmentation," *Computational Intelligence and Neuroscience*, vol. 2022, p. e9577096, May 2022. Publisher: Hindawi.

[57] M. Zhang, H. Shi, Y. Yu, and M. Zhou, "A Computer Vision Based Conveyor Deviation Detection System," *Applied Sciences*, vol. 10, p. 2402, Jan. 2020. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

[58] "Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study - ScienceDirect."

[59] M. Minhas and J. Zelek, "Defect Detection using Deep Learning from Minimal Annotations:," in *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, (Valletta, Malta), pp. 506–513, SCITEPRESS - Science and Technology Publications, 2020.

[60] K. Nakashima, F. Nagata, H. Ochi, A. Otsuka, T. Ikeda, K. Watanabe, and M. K. Habib, "Detection of minute defects using transfer learning-based CNN models," *Artificial Life and Robotics*, vol. 26, pp. 35–41, Feb. 2021.

[61] Y. Yang, C.-Y. Miao, K. Kang, and X.-G. Li, "Machine vision inspection technique for conveyor belt deviation," *Zhongbei Daxue Xuebao (Ziran Kexue Ban)/Journal of North University of China (Natural Science Edition)*, vol. 33, pp. 667–671, Dec. 2012.

[62] Z. Li, X. Wei, M. Hassaballah, Y. Li, and X. Jiang, "A deep learning model for steel surface defect detection," *Complex & Intelligent Systems*, vol. 10, pp. 885–897, Feb. 2024.

[63] admin, "Types of Conveyor Belts used for Industrial Purposes," Jan. 2020.

[64] admin, "10 Benefits of Vision-Based Inspection Systems in Manufacturing," Oct. 2023.

[65] C. Yang, "What is an Industrial Inspection Camera? Where is It Used?," Aug. 2023.

[66] "Great Leap Forward Interactive – Desktop."

[67] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019.

[68] "OpenCV · GitHub."

[69] S. C. Government of Canada, "Analysis on labour challenges in Canada, second quarter of 2023," June 2023. Last Modified: 2023-06-22.

[70] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), Lecture Notes in Computer Science, (Cham), pp. 21–37, Springer International Publishing, 2016.

[71] L. Wang, M. Törngren, and M. Onori, "Current status and advancement of cyber-physical systems in manufacturing," *Journal of Manufacturing Systems*, vol. 37, pp. 517–527, Oct. 2015.

[72] R. Mordia and A. Kumar Verma, "Visual techniques for defects detection in steel products: A comparative study," *Engineering Failure Analysis*, vol. 134, p. 106047, Apr. 2022.

[73] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," pp. 2961–2969, 2017.

[74] G. Rosati, S. Minto, and F. Oscari, "Design and construction of a variable-aperture gripper for flexible automated assembly," *Robotics and Computer-Integrated Manufacturing*, vol. 48, pp. 157–166, Dec. 2017.