USING COMPUTER VISION AND NEURAL NETWORKS FOR OBJECT DETECTION

Lappeenranta–Lahti University of Technology LUT

Master's Thesis in Software Engineering and Digital Transformation

2024

Saku Arho

Examiner(s): Associate Professor Ari Happonen, Professor Jari Porras

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Engineering Sciences

Software Engineering

Saku Arho

Using computer vision and neural networks for object detection

Master's thesis

2024

58 pages, 9 figures and 5 tables

Examiner(s): Associate Professor Ari Happonen, Professor Jari Porras

Keywords: Object detection, SMEs (Small and Medium-sized Enterprises), technology adoption in SMEs, YOLO methodology

In this thesis, the focus is on evaluating object detection technologies with a specific application in small and medium-sized enterprises (SMEs). The research conducts a comparative analysis of various object detection methods, notably Haar Cascades, Faster R-CNN, and YOLOv8, within the context of detecting objects from electrical diagrams. The study aims to determine the balance between technological performance and the practical considerations of resource management for SMEs. By assessing the capabilities and limitations of each method, the thesis presents a structured approach to selecting appropriate object detection technologies that align with the operational and financial constraints of SMEs.

The outcomes of this research clarify that YOLOv8, among the methods analysed, provides an optimal balance of accuracy and processing efficiency suitable for SMEs. This finding enables SMEs to make informed decisions on adopting object detection technologies that are not only technologically advanced but also feasible within their operational and financial constraints. By demonstrating YOLOv8's superiority in terms of both performance and resource efficiency, the study lays the groundwork for SMEs to leverage object detection technology effectively, thus potentially enhancing their operational efficiency and competitive position in the market. This contribution is significant to the ongoing discourse on digital transformation within SMEs, offering a clear direction for the strategic implementation of object detection technologies based on empirical evidence.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUTin insinööritieteiden tiedekunta

Tietotekniikka

Saku Arho

Konenäön ja neuroverkkojen hyödyntäminen kohteen tunnistuksessa

Ohjelmistotuotannon Diplomityö

2024

58 sivua, 9 kuvaa ja 5 taulukkoa

Tarkastaja(t): Apulaisprofessori Ari Happonen, Professori Jari Porras

Avainsanat: Objektien tunnistus, pk-yritykset (pienet ja keskisuuret yritykset), teknologian omaksuminen pk-yrityksissä, YOLO-menetelmä

Tässä tutkielmassa keskitytään objektien tunnistusteknologioiden arviointiin keskittyen erityisesti pieniin sekä keskisuuriin yrityksiin (pk-yrityksiin). Tutkimuksessa suoritetaan vertaileva analyysi eri objektien tunnistusmenetelmistä, erityisesti Haar Cascades, Faster R-CNN ja YOLOv8, sähköpiirustusten komponenttien tunnistamisen kontekstissa. Tutkimuksen tavoitteena on määrittää teknologisen suorituskyvyn ja pk-yritysten resurssienhallinnan käytännön näkökohtien välinen tasapaino. Arvioimalla kunkin menetelmän kyvykkyyksiä ja rajoituksia, tutkielma esittää lähestymistavan sopivien objektien tunnistusteknologioiden valitsemiseksi, jotka vastaavat pk-yritysten toiminnallisia ja taloudellisia rajoitteita.

Tutkimuksen tulokset osoittavat, että YOLOv8 tarjoaa parhaimman tasapainon tarkkuuden ja prosessointikyvyn välillä pk-yritysten käyttöön soveltuvien menetelmien joukossa. Tämä tulos mahdollistaa pk-yritysten tehdä tietoon perustuvia päätöksiä objektientunnistusteknologioiden omaksumiseen liittyen. Nämä teknologiat eivät ole ainoastaan teknisesti edistyneitä, vaan myös toteutuskelpoisia pk-yritysten toiminnallisten ja taloudellisten rajoitteiden puitteissa. YOLOv8:n ylivoimaisuuden osoittaminen sekä suorituskyvyn että resurssitehokkuuden näkökulmista luo vankan perustan pk-yrityksille objektientunnistusteknologian tehokkaaseen hyödyntämiseen. Tämä voi potentiaalisesti parantaa niiden operatiivista tehokkuutta sekä vahvistaa yritysten kilpailuasemaa markkinoilla. Tämä kontribuutio on merkittävä lisä pk-yrityksissä käytävään digitaalisen transformaation keskusteluun, tarjoten selvän suunnan objektien tunnistusteknologioiden strategiseen käyttöönottoon empiirisen todistusaineiston pohjalta.

ACKNOWLEDGEMENTS

At this juncture, I extend my heartfelt appreciation to the examiners, especially Ari Happonen. Without his tenacity, his relentless inquiries over multiple years about how my thesis was progressing, and his constant readiness to help with writing, I probably would never have finished it. I would also like to thank my family which throughout these years never doubted me, and my friends. Mostly the friends with which I went through our studies with at LUT, even more specifically those I played Kyykkä with of which three boast master's degrees and only one had the manners to not gloat about it. However, the playful gloating of the two, provided much of the motivation needed in the final stages of my thesis. Moreover, it cannot be overlooked how significant role our dog played in writing the thesis. Their presence not only brought joy and companionship but inadvertently supported my mental well-being. The forced breaks for walks in nature were not just physical diversions, but emotional lifelines, offering moments of tranquillity amidst the chaos of varying mental pressures.

Table of contents

1 Introduction

The rise of machine learning has already revolutionized most industries, as evidenced by its applications in healthcare and data recognition (Usmani, Happonen, & Watada, 2024), digital process automation (Ylä-Kujala et al., 2023), anomaly detection (Usmani, Happonen, & Watada, 2022), and autonomous machinery (Abdelsalam et al., 2022), influencing decision-making and operational processes. Central to this wave of innovation is object detection, which plays a crucial role by providing critical insights and enhancing automation across different sectors. This technology has shown effectiveness in the challenging tasks of analysing and interpreting complex schematics, where accuracy and reliability are crucial. The advancement of object detection technologies, especially through the development of Convolutional Neural Networks (CNNs), has marked significant milestones in the field (Zou et al., 2019). Such progress has not only improved the performance of object detection systems but also led to the development of efficient models, including the "You Only Look Once" (YOLO) methodology, known for its speed and accuracy.

Small and medium-sized enterprises (SMEs) form an integral part of the economic landscape and stand to gain significantly from the adoption of object detection technologies. Yet, they face distinct challenges due to their constrained resources and expertise. The careful selection of an object detection method is crucial, as it can dramatically affect SMEs' operational efficiency and competitive advantage. This thesis will explore how SMEs can strike a balance between the performance benefits of object detection technologies and the practical considerations of resource management, addressing the nuances of implementing technologically advanced solutions within the tighter budgets and limited computational capacities.

The challenge in object detection stems from a lack of clear academic guidance on selecting algorithms, primarily due to the vast array of use cases and the rapid evolution of technology. This thesis aims to bridge this gap by introducing a foundational framework that categorizes various object detection methods, detailing their strengths, limitations, and conditions for optimal application, thereby providing a structured approach to algorithm selection. This approach provides researchers and practitioners with a starting point for a more in-depth exploration of object detection methodologies. The thesis aims to provide easily

understandable and actionable insights, particularly focusing on the needs and constraints faced by SMEs, thereby aiding in making informed decisions about the choice of object detection techniques in various scenarios.

1.1 Significance of the Study

The process of selecting an object detection algorithm is complex, involving considerations such as computational demand, processing velocity, and accuracy. This study focuses on Small and Medium-sized Enterprises (SMEs), defined as companies employing fewer than 250 people (OECD, n.d.). It aims to provide a clear analysis of object detection technologies within the operational and resource constraints typical for SMEs. The significance of this research lies in its attempt to bridge the knowledge gap for SMEs, facilitating informed technological choices that align with their specific needs. This contribution is intended to support SMEs in enhancing their operational processes and maintaining relevance in a technology-centric market environment.

1.2 Research Goals and Objectives

The central aim of this research is to offer a comparative analysis of object detection methodologies, highlighting those that are most relevant and advantageous for SMEs. To navigate this analysis, the study is structured around the following research questions:

1. What are the comparative strengths and weaknesses of prevalent object detection algorithms when applied within the resource constraints of SMEs?
2. How do different object detection methods perform in practical SME scenarios, particularly in terms of processing speed, accuracy, and resource consumption?

These questions will guide the critical evaluation of various algorithms and facilitate the demonstration of their applications through a case study, ultimately informing SMEs about the strategic selection of technology suited to their unique circumstances.

1.3 Scope and Limitations

This thesis evaluates four distinct object detection methods, chosen from a broad spectrum of available algorithms for their relevance to the specific challenges faced by SMEs, which often operate under resource constraints. The selected methods are Template Matching, Haar Cascades, YOLO (You Only Look Once), and Faster R-CNN. These methods were chosen because they represent different levels of resource intensity and algorithmic approaches. Template Matching and Haar Cascades are less resource-intensive but offer limited capabilities, whereas YOLO provides a balance with its efficient single-stage approach. In contrast, Faster R-CNN, though more resource-intensive, showcases the advantages of a two-stage method. The diverse selection is crucial to highlight the balance between performance and resource consumption, based on SMEs' limited resources and efficiency needs. This consideration stems from understanding that SMEs often operate under constraints that necessitate optimizing for both performance and minimal resource use.

The limitations of this study include its focus on a select number of object detection methods, which may not include the entire spectrum of technologies available in the field. Additionally, the findings are based on the specific application to electrical diagrams, which may not fully represent the diverse range of potential use cases in various SME settings. Aside from this, the study’s particular focus and methodology limit the results’ applicability to other SME scenarios or industry sectors.

1.4 Methodology

The study uses a mixed-methods approach, combining both quantitative and the qualitative analyses to examine object detection techniques within SMEs.

Quantitative Methods:

- Performance Evaluation: Assesses object detection algorithms' accuracy, speed, and computational demands using precision, recall, and processing time metrics, vital for determining their fit for SMEs.

- Computational Analysis: Analyses CPU and GPU efficiency of algorithms, crucial for SMEs with limited tech resources.

Qualitative Methods:

- Case Study: Empirically examines the YOLO methodology in an SME setting, offering insights into its real-world adaptability and effectiveness.
- Literature Review: Reviews existing research on object detection technologies, their evolution, and SME applications, providing a theoretical background and identifying research gaps.

1.4.1 Case Analysis

Central to the thesis' empirical data collection is a case study that tests object detection methods under SME constraints. Initially, Template Matching was experimented with, but was quickly found unsuitable due to its lack of accuracy and inability to handle variations in the same component types. According to a review on advances and applications in template matching (Hashemi et al., 2016), while template matching is a fundamental method used in computer vision, its application is limited as a pre- or post-processing step (Hashemi et al., 2016). After further research and experimentation, this led to the exclusion of Template Matching in favour of the YOLO (You Only Look Once) method. YOLO, as described by Redmon et al. (2016), offers a significant balance between detection performance and computational efficiency, processing images in real-time at 45 frames per second and achieving more than twice the mean average precision of other real-time systems (Redmon et al., 2016). This balance is critical for SMEs operating under resource constraints.

1.4.2 Alternative Object Detection Methods

Although the empirical focus is on the YOLO method, since it was eventually used in the finished product, the study also includes a theoretical evaluation of alternative object detection methods, namely Haar Cascades and the R-CNN variant Faster R-CNN, in the context of the case study. Haar Cascades was considered due to their simplicity and lower

computational load, as well as it being recommended by the client, making them a potential fit for projects with limited resources. Faster R-CNN was selected for its high precision and two-stage detection process, offering a contrast to YOLO's single-stage approach. These choices were informed by their specific attributes and project's unique requirements, where Haar Cascades' efficiency and Faster R-CNN's accuracy provided valuable insights into the trade-offs and application scenarios for different detection methods (Ren et al., 2015).

1.4.3 Integrative Analysis

By combining insights from the literature with practical examination, the study analyses specific performance metrics of these object detection methods, such as precision, recall, processing time and Mean Average Precision (mAP). Through this integrated approach, the study not only assesses the theoretical capabilities of these methods but also their practical viability, ensuring that the analysis is both comprehensive and aligned with the study's goals.

2 Background

In the field of object detection, the transition from early techniques to advanced methods showcases the dynamic evaluation of neural networks and computer vision. Initially, object detection relied on conventional methods such as template matching, a staple of early computer vision, and Haar Cascades, significantly enhanced by the Viola-Jones detector in 2001 (Viola & Jones, 2001). These foundational approaches, while pivotal, often lacked accuracy and adaptability. The introduction of Convolutional Neural Networks (CNNs) marked a turning point, steering the field towards more refined and efficient algorithms (Krizhevsky, Sutskever & Hinton, 2012). This shift led to the development of groundbreaking models like R-CNN and its successors, as well as the YOLO series.

The emergence of R-CNN and its enhanced versions, Fast R-CNN and Faster R-CNN, revolutionised object detection by balancing speed with precision, though at the expense of higher computational demands (Grishick, 2015; Ren e tal., 2015). The YOLO (You Only Look Once) model further advanced the field with its real-time detection capabilities,

striking a balance between rapid processing and accuracy (Redmon et al., 2016). The latest innovations include state-of-the-art models like YOLOv7, which surpasses all known object detectors in both speed and accuracy (Wang et al., 2023), These developments along with ongoing progress in deep learning, are transforming object detection and its application in areas such as healthcare, autonomous driving, and smart surveillance.

Table 1: Evolution of Object Detection Methods.

Method	Source(s)	Description and Source Citations
CNN	Krizhevsky, Sutskever & Hinton, 2012	Introduced a paradigm shift in object detection. Uses convolutional layers for feature extraction. (Krizhevsky, Sutskever & Hinton, 2012)
R-CNN	Grishick, 2015	Region-based CNN. Enhances accuracy in object detection but increases computational complexity. (Grishick, 2015)
Fast R-CNN	Grishick, 2015	Improved version of R-CNN. Faster and more efficient, but still with high accuracy. (Grishick, 2015)
Faster R-CNN	Ren et al., 2015	Further improvement on R-CNN, enhancing both speed and accuracy. Introduces Region Proposal Networks. (Ren et al., 2015)
YOLO	Redmon et al., 2016	Stands for "You Only Look Once." Renowned for real-time object detection capabilities. Balances speed and accuracy. (Redmon et al., 2016)
EfficientDet	Tan, M., et al. (2020)	EfficientDet: Scalable and Efficient Object Detection. Focuses on optimizing efficiency and scalability in object detection. (Tan, M., et al., 2020)
YOLOv7	Wang, J., et al. (2023)	YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Presents significant advancements in speed and accuracy. (Wang, J., et al., 2023)

However, challenges such as achieving high precision in complex environments and the computational intensity of deploying these models at scale continue to be areas of focus. Zou et al. (2020) discuss these challenges in their survey of modern deep learning-based object detection models, highlighting the ongoing efforts to address these issues. Future trends in the field may include further integration with augmented and virtual reality, as well as

considerations regarding the ethical implications in the development and deployment of object detection technologies.

In the broad arena of object detection techniques, this study concentrates on exploring various methods including Template Matching, Haar Cascades, a R-CNN variant (Faster R-CNN), and YOLO, especially in the context of identifying electrical components. Template Matching was initially considered for its lightweight approach but was eventually dismissed due to its insufficient handling of variations in components. Haar Cascades, though suggested by our client, was also dismissed because it is not effective with line-intense images typical of electrical schematics. ON the other hand, the R-CNN variant, Faster R-CNN, was selected for its two-stage detection process, which is a good contrast for YOLO's single-stage process, and potentially offers higher accuracy but with a slower rate compared to YOLO. YOLO was ultimately chosen for the project due to its one-stage detection process, balancing speed with accuracy. This selection aligns with the project's goals of creating an effective web-based application for processing and analysis, without the need for continuous intensive computation.

3 Advancements in Object Detection: A Focus on SME Applications

In this section, we conduct a detailed exploration of object detection technologies, focusing on their evolution and current relevance to SMEs. Starting with the fundamentals like template matching and Haar cascades, we navigate through the progressions in CNNs and their variants, leading up to the latest methodologies such as YOLO and SSD. The section goes through the technological advancements but also examines their practical applicability in the SME context. By doing so, this part of the thesis aims to bridge the gap between theoretical innovation and real-world utility, addressing the specific challenges faced by SMEs. Through this analysis, the thesis highlights the transformative impact of these developments on object detection, providing insights into their potential to enhance SME operations within the constraints of limited resources and operational realities.

3.1 Evolution of Object Detection Technologies

In the rapidly evolving field of computer vision, object detection technologies have undergone a remarkable transformation over the past few decades. This chapter delves into the chronological advancements of object detection methods, tracing the journey from early techniques to the latest innovations in the field. Starting with simple template matching and Haar cascades, we will explore the shift towards more sophisticated methods like handcrafted features and the ground-breaking advent of Convolutional Neural Networks (CNNs). The chapter will further discuss the development of region-based CNNs and their variants, as well as the emergence of efficient single-stage detectors like YOLO and SSD. This historical overview not only highlights the technological leaps in object detection but also sheds light on their practical implications, particularly for small and medium-sized enterprises (SMEs). Through this exploration, we aim to provide a comprehensive understanding of how object detection technologies have evolved and the potential they hold for future applications.

Table 2: Evolution of Object detection methods

Year	Method	Summary
2001	Haar-like AdaBoost	Introduced simple feature-based detection with a boosting algorithm for performance.
2005	Histogram of Oriented Gradients (HOG)	Analyzed image gradients for better detection capabilities.
2008	Deformable Part Model (DPM)	Used parts of objects and their deformable configurations for detection.
2009	Integral Channel Features (ICF)	Enhanced Haar features with additional channel types.
2010	Aggregate Channel Features (ACF) / LDCF	Improved channel features for more efficient detection.
2014	R-CNN	Combined deep learning with region proposals for accurate detection.
2014	Spatial Pyramid Pooling (SPP-net)	Processed variable image sizes through spatial pyramid pooling.
2015	Fast R-CNN	Made R-CNN faster with streamlined processing and training.
2015	Faster R-CNN	Achieved real-time detection using neural networks for region proposals.
2016	YOLO (You Only Look Once)	Implemented a single neural network pass for speedy detection.
2017	SSD (Single Shot Detector)	Balanced speed and accuracy with one deep neural network.
2018	YOLOv3	Enhanced YOLO with multi-scale detection and better features.
2020	DETR (Detection Transformer)	Utilized transformers to model object relationships for detection.
2021	YOLOv7	Advanced the YOLO series focusing on real-time application speed and accuracy.
2022	AdaMixer	Combined convolutional and transformer architectures for object detection.

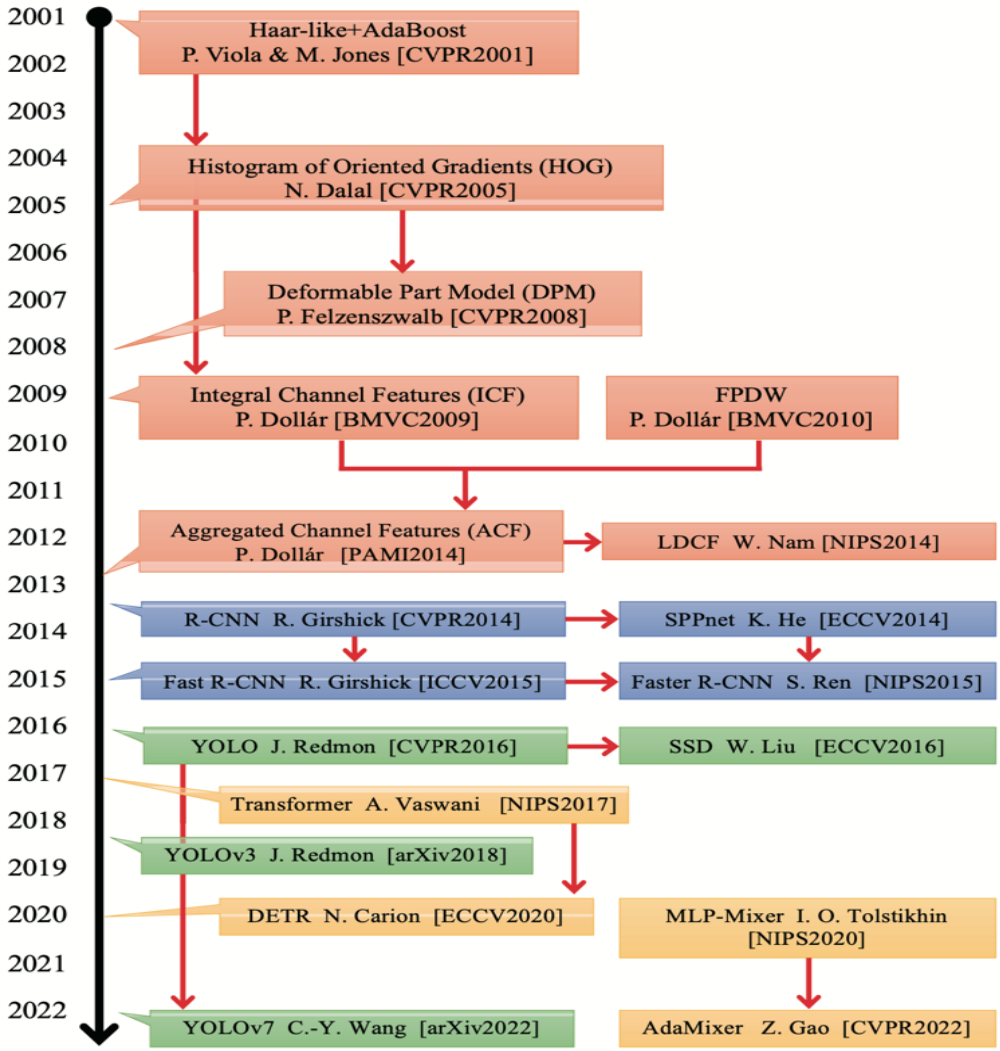

Figure 1: This timeline shows the evolution of object detection technologies from 2001 onwards, hihglighting key developments like Haar-like features, HOG, DPM, and the various versions of R-CNN and YOLO (Deguchi & Murase 2023).

Figure 1, from Deguchi and Murase (2023), showcases the significant milestones in the field of object detection from the early 2000s to the present. This timeline begins with the introduction of Haar-like features in 2001, a critical juncture that underscored the shift by offering a methodological foundation for more complex and effective object detection approaches. Following this, the development of Histogram of Oriented Gradients (HOG) in 2005 and Deformable Parts Model (DPM) in 2008 represented major advancements in feature extraction and part-based models. The timeline then highlights the revolution brought about by the introduction of R-CNN in 2014, which integrated deep learning into object detection, setting the stage for subsequent innovations like Fast R-CNN, Faster R-CNN, and

YOLO variants. Each of these developments, marked on the timeline, indicates a leap forward in accuracy and efficiency, culminating in the latest YOLO versions, which are particularly relevant to SME applications due to their speed and real-time processing capabilities (Deguchi & Murase, 2023).

3.1.1 Template Matching to Haar Cascades and HOG

Object detection technologies have undergone significant advancements, tracing back to early methods such as Template Matching. This technique, which involves identifying matches between a template and a target image, faced challenges due to object appearance variations. Evolution in the field introduced Haar Cascades by Viola and Jones, leveraging feature selection and cascading classifiers for efficient real-time detection, notably in facial recognition. Subsequently, the development of the Histogram of Oriented Gradients (HOG) further refined object detection by analysing the distribution of gradient directions in image sections. This method improved accuracy for detecting objects under various poses and conditions. Each of these methodologies represents a key advancement in the domain of object detection, highlighting the progression towards increasingly sophisticated, accurate, and versatile detection techniques.

Template Matching is portrayed in Figure 2, where a predefined template is methodically slid over the image to identify matching areas based on pixel intensity values. This figure illustrates the process with non-dotted lines representing strong matches and dotted lines indicating matches with lower confidence. This deterministic approach is particularly effective in controlled environments where the target object's appearance remains relatively constant (Binjie & Hu, 2014).

Haar Cascade Classifiers, depicted in Figure 3, signify a leap to a machine learning-based approach in object detection. Developed by Viola and Jones (2001), this method employs feature selection and cascading classifiers for detecting objects with varying appearances. The image demonstrates the Haar features through a series of black and white blocks, highlighting the method's ability to aggregate pixel intensities within specified regions and compute their differences. This innovative approach allows for a detailed analysis of an

image's textural and structural characteristics, enhancing detection capability in complex real-time applications.

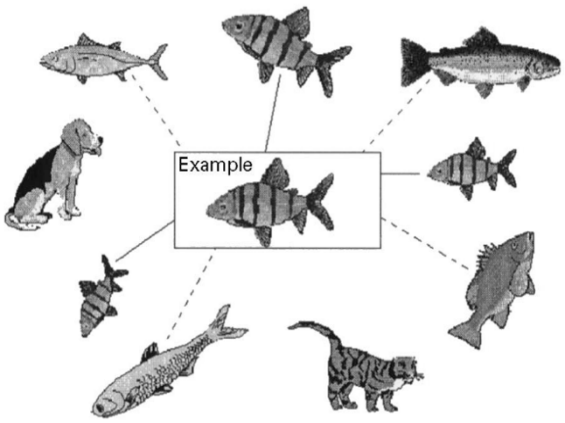

Figure 3: Example of template matching. This figure illustrates the process of template matching where a template (highlighted in the box) is slid over the image to find matching areas. Adapted from Hashemi, N. S., Aghdam, R. B., Ghiasi, A. S. B., & Fate.

Figure 2: Example of Haar cascade object detection. This figure illustrates the various stages of Haar cascade detection on a single image, demonstrating the method's systematic approach to feature identification (Hako, n.d.).

Expanding on the foundation set by Haar Cascades, the Histogram of Oriented Gradients (HOG), proposed by Dalal and Triggs in 2005, introduced an advanced approach to object detection by analysing the distribution of edge directions across localized sections of an image. Primarily utilized for pedestrian detection, HOG leverages gradient information for feature descriptors. It provides a complementary method to Haar-based techniques, as reflected in the combined impact of these methods on the evolution of object detection. The synergy of Haar Cascades and HOG has been essential in creating the foundation for later developments in the field, leading to the development of more sophisticated models. The progression from Template Matching to Haar Cascades and HOG illustrates the dynamic nature of object detection technology. This progression from Template Matching to Haar Cascades and HOG demonstrates the adaptable nature of object detection technology, offering businesses of varying sizes, especially SMEs with limited resources, a scalable range of solutions tailored to diverse operational needs and budgetary constraints.

3.1.2 Advent of Convolutional Neural Networks (CNNs)

Since the onset of deep learning's integration into object detection around the early 2010s, methodologies have seen significant advancements, notably Convolutional Neural Networks (CNNs) and transformer-based architectures. This transition has led to measurable improvements in accuracy and efficiency in object detection tasks, as highlighted by Tasnim & Qi (2023). For small and medium-sized enterprises (SMEs), these developments have expanded practical applications, including surveillance and quality control, as documented by Zhou et al. (2021).

Central to this evolution in object detection were the backbone architectures of CNNs, which evolved from simpler designs like AlexNet and VGG to more complex and efficient structures. This progression included multi-path designs such as Inception and the introduction of residual learning in ResNet, which helped address issues like training degradation and enabled the creation of deeper, more effective networks (Li et al., 2020). Additionally, the development of lightweight networks, such as SqueezeNet, responded to the constraints of memory and computing resources, a consideration particularly relevant for SME applications (Zou et al., 2019; Li et al., 2020).

3.1.3 Development of R-CNN and Its Variants

Building on the CNN advancements discussed in Section 1.1.2, the development of the R-CNN family of algorithms has been a pivotal aspect of object detection evolution. Initially proposed by Girshick et al., R-CNN combined CNNs with region proposal algorithms to improve detection accuracy. This methodology evolved into Fast R-CNN and Faster R-CNN, which enhanced processing speed and accuracy, establishing new standards in the field of object detection (Girshick, Donahue, Darrell & Malik, 2014; Girshick, 2015; Ren, He, Girshick & Sun, 2015).

Figure 4 illustrates the incremental improvements in two-stage object detection methods, starting from the foundational R-CNN to the more advanced Faster R-CNN. Each iteration in this family of algorithms brought key enhancements, such as spatial pyramid pooling in

SPPnet, RoI pooling in Fast R-CNN, and the introduction of the Region Proposal Network in Faster R-CNN. These improvements significantly accelerated the object detection process, demonstrating the rapid advancements in CNN-based methodologies, crucial for real-time applications in SMEs (Zhang & Hong, 2019).

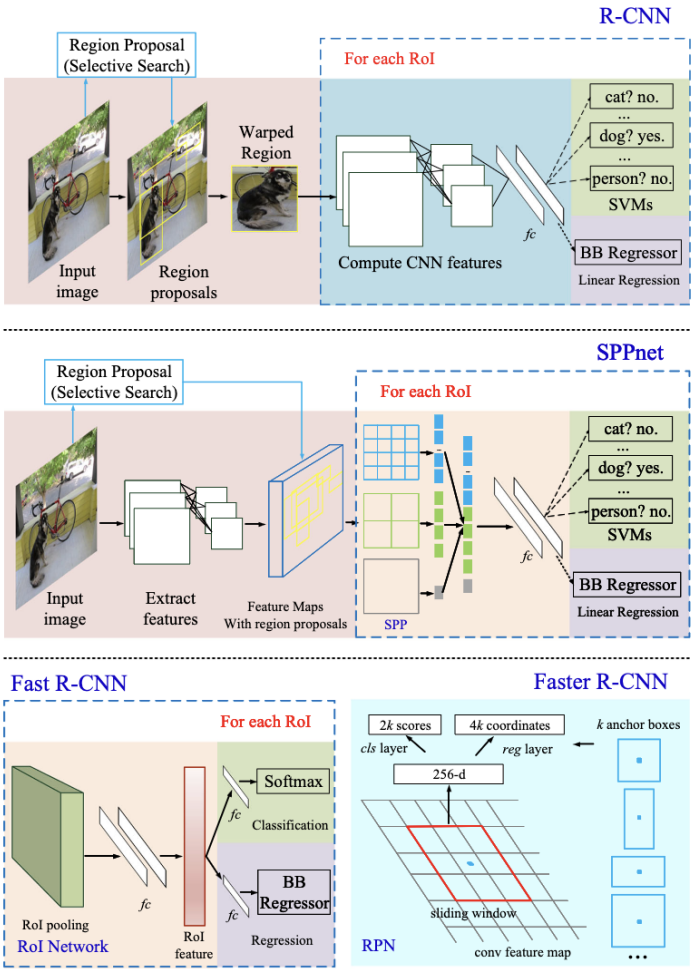

Figure 4: Evolution of Two-Stage Object Detectors. This figure compares the architectures of R-CNN, SPPnet, Fast R-CNN, and Faster R-CNN, showing the incremental improvements in two-stage object detection methods. Adapted from (Zhang & Hong, 2019).

The selection of backbone architectures plays a crucial role in improving feature extraction processes, with a spectrum from traditional Convolutional Neural Networks (CNNs) to newer transformer-based models affecting the efficacy of object detection systems. The shift towards transformer-based architectures, as evidenced by leading methods on the COCO dataset, marks a discernible transition in the feature extraction process of these models. This

progression has expanded the differences in performance between transformers and traditional CNNs, underlining the criticality of choosing suitable backbone architectures for enhancing both accuracy and efficiency in object detection tasks, particularly in the context of the R-CNN framework (Zou et al., 2019).

3.1.4 Emergence of Single-Stage Detectors: YOLO and SSD

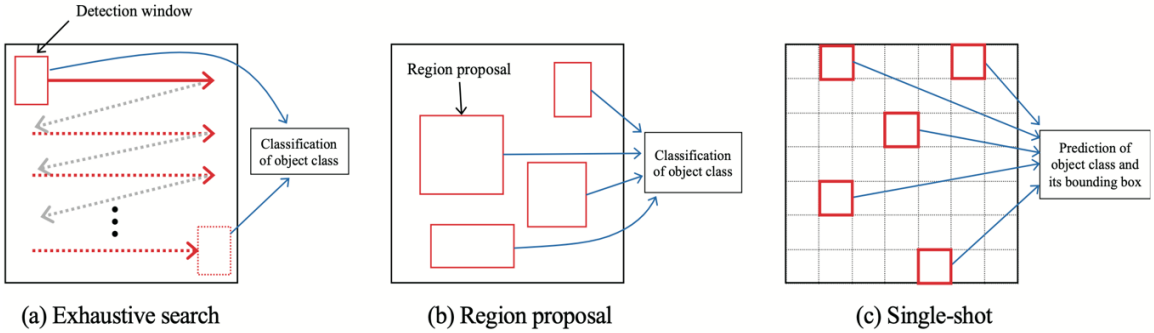

Figure 5: This figure illustrates three main strategies for object detection: exhaustive search, region proposal, and single-shot detection, each depicted through a flow diagram. (Deguchi & Murase, 2023).

The advancements in backbone architectures have notably enhanced the capabilities of single-stage detectors such as YOLO and SSD. Incorporating more efficient and robust backbone models into these detectors has facilitated improvements in both speed and accuracy, elements essential for effective real-time object detection, as reported by Li et al. (2020).

Figure 5, as delineated by Deguchi and Murase (2023), effectively illustrates the primary strategies employed in object detection. The exhaustive search strategy involves scanning the entire image with varying window sizes to detect objects, often resulting in high computational costs. In contrast, the region proposal method focuses on identifying potential regions of interest before classifying them, thereby reducing the computational load. The single-shot detection approach, on the other hand, streamlines the process by combining detection and classification in a single step, enhancing speed but potentially compromising accuracy in complex scenarios (Deguchi & Murase, 2023).

The introduction of single-stage detectors like YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector) marked a significant shift in object detection methodologies. These frameworks, characterized by their efficiency in real-time processing, have been pivotal in advancing object detection. YOLO, developed by Redmon et al. (2016), revolutionized object detection with its unique architecture that enables simultaneous predictions of bounding boxes and class probabilities, significantly increasing detection speed and making it suitable for real-time applications (Zaidi et al., 2022). Similarly, SSD, introduced by Liu et al. (2016), balanced speed and accuracy effectively, offering a robust alternative to two-stage detectors (Zaidi et al., 2022). The practical implications of these technologies, especially for SMEs, are profound, providing efficient and accurate detection capabilities essential for modern-day applications.

Figure 6 illustrates the distinct approaches taken by one-stage detectors like YOLO and SSD. YOLO's unique grid-based detection and SSD's utilization of multi-scale feature maps exemplify the innovative strategies that bypass the need for region proposals, thereby accelerating the detection process (Zhang & Hong, 2019).

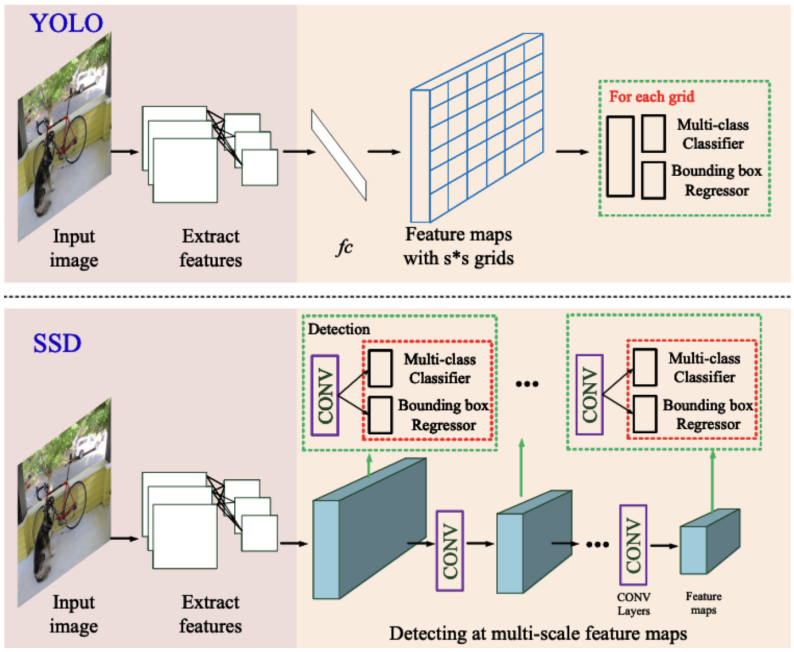

Figure 6: Operational pipelines of YOLO and SSD models. This figure contrasts the grid-based detection strategy of YOLO with the multi-scale feature map approach of SSD. (Zhang & Hong, 2019).

3.1.5 Implications for SMEs

The development of object detection technologies provides small and medium-sized enterprises (SMEs) with a range of adaptable tools suited to various operational needs. Specifically, SMEs can choose between two-stage detectors, which excel in accuracy and are ideal for applications requiring detailed image analysis, and single-stage detectors, which offer faster processing speeds for real-time applications. The advancements in deep learning methodologies enable SMEs to leverage these technologies for efficient and precise object detection, even when operating with constrained resources. This opens possibilities for SMEs to enhance their capabilities in areas such as inventory management, surveillance, and quality control without significant investment.

3.2 Impact of YOLO Methodology

The introduction of YOLO (You Only Look Once) by Redmon et al. (2016) introduced a novel approach to real-time object detection by integrating the detection process into a single neural network evaluation. This method significantly diverged from previous techniques that relied on separate stages for detecting regions and classifying objects. YOLO's innovation lies in its ability to directly predict both bounding boxes and class probabilities from full images, streamlining the detection workflow. This integration resulted in a substantial increase in processing speed, enabling it to run at 45 frames per second, and up to 155 frames per second for the optimized version, Fast YOLO. This enhancement in speed, without compromising accuracy, represented a pivotal advancement in enabling real-time object detection applications.

Figure 6 shows YOLO Detection Pipeline. Illustration of YOLO's grid-based detection strategy, enabling real-time object detection by predicting bounding boxes and class probabilities directly from full images in a single network evaluation.

YOLO's architecture was inspired by the GoogLeNet model for image classification, consisting of 24 convolutional layers followed by 2 fully connected layers. The design allowed the network to reason globally about the image, leading to fewer background errors

compared to methods like Fast R-CNN. Additionally, YOLO demonstrated strong generalizability, performing well when trained on natural images and tested on different domains such as artwork (Redmon et al., 2016).

3.2.1 Advancements in YOLO Versions

YOLOv2, released in 2017, made several improvements over the original YOLO. It introduced batch normalization on all convolutional layers, enhancing convergence and reducing overfitting. The architecture transitioned to a fully convolutional one and employed anchor boxes for bounding box prediction. Moreover, YOLOv2 utilized dimension clusters and direct location prediction to enhance accuracy. It achieved an average precision of 78.6% on the PASCAL VOC2007 dataset, significantly improving upon YOLOv1's 63.4% (Redmon & Farhadi, 2017).

Introduced in 2018, YOLOv3 featured a more extensive architecture with significant updates to match state-of-the-art performance while ensuring real-time processing. It employed a new backbone, Darknet-53, with 53 convolutional layers and residual connections. YOLOv3 made advancements in bounding box prediction and class prediction, adopting binary cross-entropy for independent logistic classifiers. Additionally, it included spatial pyramid pooling (SPP) and multi-scale predictions, enabling better detection of small objects. YOLOv3 achieved an average precision of 36.2% and an AP50 of 60.6%, which indicates its precision for detecting objects with at least half overlap with the perfect bounding box, at 20 FPS on the Microsoft COCO dataset (Redmon & Farhadi, 2018).

These advancements in YOLO's methodology, particularly in real-time processing and accuracy, have significantly impacted the field of object detection. The continuous improvements in YOLO versions demonstrate the evolving nature of object detection technology, with each iteration offering enhanced performance and efficiency suitable for various applications, including those relevant to SMEs.

3.2.2 Limitations and Trade-offs

Despite its strengths, YOLO does have limitations. It struggles with precise localization of small objects and imposes strong spatial constraints on bounding box predictions, which can limit the detection of nearby objects (Redmon et al., 2016).

3.2.3 Performance Metrics and Evaluation

YOLO's performance, when compared to other real-time detection systems like DPM and R-CNN variants, underscores its efficiency. It outperforms these systems in terms of speed and mean average precision (mAP), though it lags slightly behind in accuracy due to higher localization errors. However, YOLO's lower rate of false positives on background objects and its ability to generalize to different domains like artwork highlight its versatility and robustness (Redmon et al., 2016).

3.3 Recent advances in Object Detection Technologies

The field of object detection has seen transformative changes in recent years, leading to notable improvements in accuracy and efficiency. This section delves into the latest innovations that are reshaping the field, such as attention-based models, one-stage fully convolutional detectors, and cutting-edge network frameworks including PANet. These developments represent a departure from conventional methodologies, steering towards more sophisticated, effective, and accurate object detection strategies. The integration of attention mechanisms and the trend towards fully convolutional systems are indicative of the evolving nature of object detection, focusing on adaptive and streamlined solutions.

3.3.1 Innovations in Multiscale Feature Aggregation for Object Detection

Multiscale feature aggregation plugins have revolutionized the field of object detection by dramatically improving accuracy and efficiency. Feng (2022) highlighted the importance of these plugins in remote sensing object detection, emphasizing their role in reducing semantic gaps and improving feature representation (Feng, 2022). Similarly, Rajput et al. (2020)

discussed the effectiveness of these plugins in improving accuracy, particularly through the integration of feature-fusion and attention mechanisms like CBAM and SFCM in various models (Rajput, Mittal, & Narayan, 2020).

A significant advancement in this domain is the Bi-YOLOX network, an enhancement over the traditional YOLOX model. Zhang et al. (2022) innovated with the introduction of the Tri-Head module and the BiNet feature fusion into this network, which significantly improved small object detection in UAV aerial images. These enhancements, including a focus on densely contiguous small objects and complex background scenarios, have been pivotal in advancing the performance of object detection systems (Zhang et al., 2022).

Table 3 provides a comparative analysis of various object detection models, including the baseline SSD and its enhancements through CBAM and other plugins. The table effectively demonstrates the varied applications of each plugin, as evidenced by the mAP and FPS metrics, which respectively measure accuracy and efficiency, showcasing their adaptability to meet specific requirements. These recent advancements underscore the vital role of sophisticated feature aggregation in modern object detection algorithms, setting new benchmarks in the field by effectively addressing the unique challenges of small object detection.

Table 3: Performance comparison of object detection models, detailing parameters, mAP scores, and frame rates, emphasizing multiscale feature enhancements. (Rajput et al., 2020).

Model	Number of parameters	mAP	Frame-rate (FPS)
SSD (baseline)	22 million	77.20	56
SSD-CBAM	23 million	78.14	47
SSD-Fusion-CBAM	28 million	78.78	46
SSD-SFCM	25 million	78.82	50
SSD-SFCM-CFE	28 million	78.94	41
SSD-Fusion-SE	28 million	78.54	48
SSD-Fusion-DANet	31 million	78.29	44

3.3.2 Omni-Supervised Object Detection with Transformers

Recent advancements in object detection have been marked by the adoption of transformer-based models, like Omni-DETR, which utilize a range of annotations from fully labelled to weakly labelled data. Omni-DETR, a pivotal development in omni-supervised object detection (OSOD), leverages weak labels such as image tags, counts, and points, to generate accurate pseudo labels through a unified framework. This approach, which employs a bipartite matching-based filtering mechanism, has demonstrated superior results on multiple datasets, indicating its effectiveness in improving detection performance while offering a better balance between annotation cost and accuracy (Wang et al., 2022).

The use of weak annotations in OSOD, particularly through Omni-DETR, addresses the high costs and scalability challenges associated with complete and accurate detection annotations. For example, fully annotating an image in datasets like MS-COCO can be prohibitively time-consuming (Wang et al., 2022). Omni-DETR's framework, grounded in recent advancements in semi-supervised object detection and end-to-end detection architectures, allows for the effective use of weak ground truth labels. This strategy enhances learning processes and facilitates more cost-effective approaches to object detection (Wang et al., 2022).

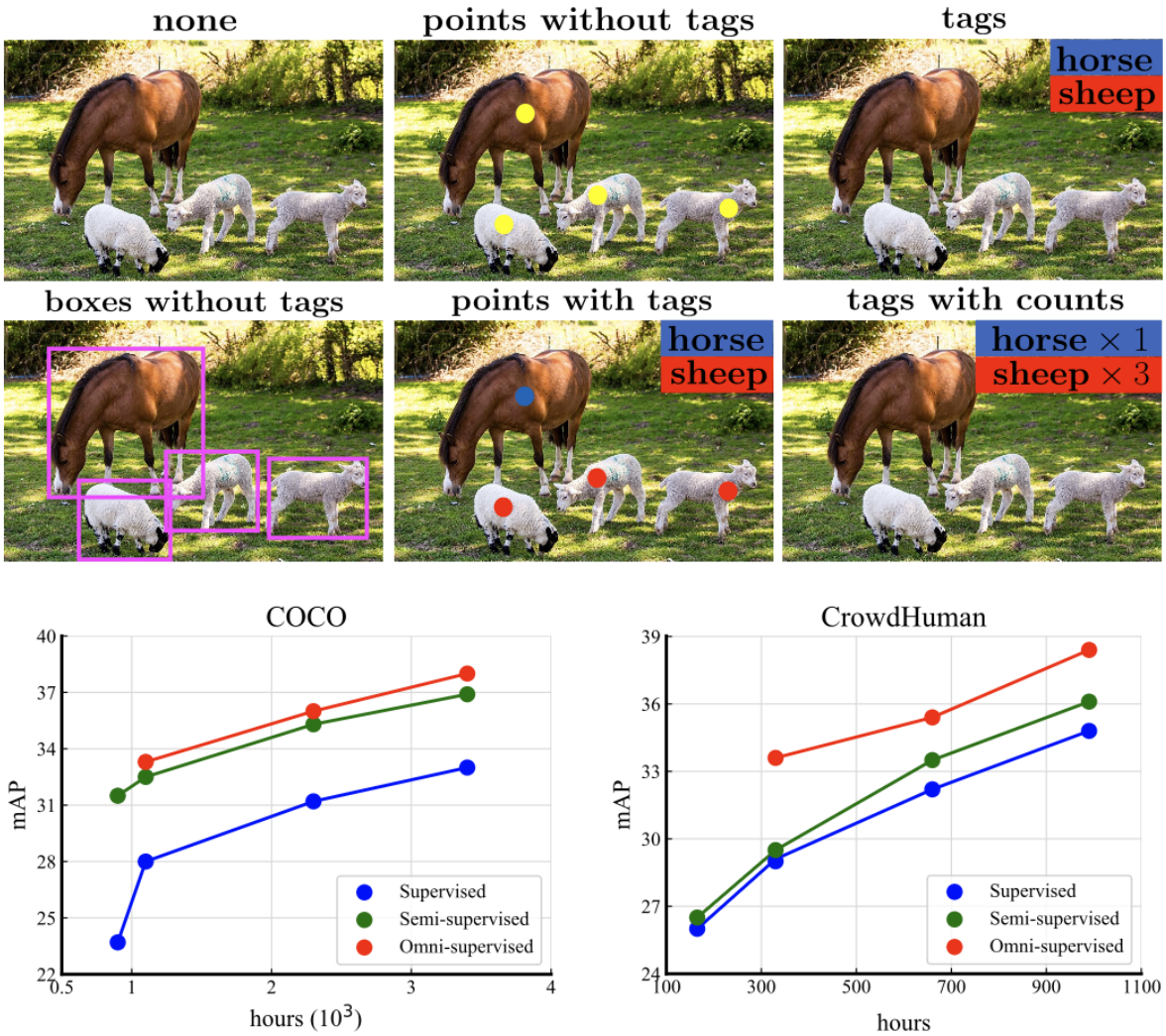

Figure 7: Top row is the visualization of different forms of weak annotations. The bottom row is the trade-off comparison (accuracy vs. annotation cost) of supervised, semi-supervised and omni-supervised detection. (Wang et. al., 2022).

Figure 7 presents a comparison between the annotation costs and detection accuracy of various object detection methods. On the x-axis, annotation time is measured in hours, indicating the investment required to train each model. The y-axis displays mean Average Precision (mAP), a key indicator of detection accuracy, with higher values signifying better performance. The graphs illustrate that omni-supervised learning strikes an effective balance, achieving high mAP with less annotated data compared to fully supervised methods. This suggests that omni-supervised models, like Omni-DETR, can deliver precise object detection with less annotation effort, making them attractive for large-scale applications where efficiency is crucial.

3.3.3 Advancements in Attention Mechanisms for Object Detection

Recent advances in attention mechanisms, when integrated into Convolutional Neural Networks (CNNs), have concretely enhanced the field of object detection. Mei (2020) incorporated a light attention mechanism called the attended residual module into an object detection backbone. This module, combined with a cascade region proposal network (RPN) and a criss-cross attention module, resulted in a tangible improvement in performance. Specifically, it yielded a 43.6 Average Precision (AP) score on the COCO test-dev, which is a standardized dataset used to benchmark object detection algorithms. This AP score represents a quantified improvement, reflecting the precise impact of attention mechanisms on the accuracy of object detection models.

Song (2021) further enhanced attention mechanisms in CNNs by proposing the Cross-Scale Non-Local (CS-NL) attention and exhaustive Self-Exemplars Mining (SEM), which allowed for the full excavation of self-similarity in images. They achieved state-of-the-art performance in image restoration tasks, especially when integrated into the EDSR network, outperforming the original EDSR and the SAN network, which was the best model in 2019 (Mei et al., 2021).

Wang (2019) proposed a pyramid attention structure and salient edge detection module for Salient Object Detection (SOD), achieving state-of-the-art performance (Wang et al., 2019). These innovations hold significant potential for practical applications, particularly in defence technologies, by improving the accuracy, efficiency, and speed of object detection systems.

3.3.4 Innovations in Path Aggregation for Feature Learning

Object detection systems have had notable enhancements in feature learning due to recent strides in path aggregation networks. These advancements include the use of a Gated Path Aggregation Network (GPA) for remote sensing image detection (Zheng, 2022), and a class-constrained spatial-temporal relation network coupled with a correlation-based feature alignment module for video object detection (Han et al., 2020).

Zheng et al. (2022) addressed the unique challenges in remote sensing object detection, such as small geometries and variable orientations of targets. They proposed a GPA network integrating path enhancement and information filtering, using soft switchable atrous convolution in the topmost feature layer. This method demonstrated considerable improvements over traditional Feature Pyramid Network structures, achieving state-of-the-art performance on the NWPU VHR-10 dataset.

For Video Object Detection (VOD), Han et al. (2020) tackled issues like occlusion and motion blur through feature aggregation from local or global support frames. They developed a class-constrained spatial-temporal relation network, which operates on object region proposals to learn dependencies among same-class objects from support frames and spatial relations among different objects in the target frame. Their correlation-based feature alignment module aligns the support and target frames for feature aggregation in the temporal domain. This method significantly improved the accuracy of single-frame detectors, outperforming previous temporal or spatial relation networks, and achieved an impressive 84.80% accuracy with ResNet-101 on the ImageNet VID dataset, without any post-processing methods.

Both Zheng et al. (2022) and Han et al. (2020) represent significant strides in object detection, showcasing the evolution of feature aggregation techniques. Zheng et al.'s GPA network effectively handles hierarchical convolutional layers, reducing redundancy in feature fusion. Han et al.'s (2020) approach to VOD underscores the importance of class-constrained networks and correlation-based alignment, leading to improved detection performance across various scenarios.

3.3.5 Breakthroughs in Single-Stage Object Detection

Recent developments in single-stage object detection, exemplified by FCOS (Fully Convolutional One-Stage object detection), have improved both the computational efficiency and the localization accuracy of these models. Chu et al. (2022) put forward EfficientFCOS, which enhances the original FCOS by requiring fewer computational resources for operation and providing more accurate object localization within images. This

model leverages EfficientNet for feature extraction, utilizing its fewer parameters and superior performance compared to ResNet. EfficientFCOS employs scaling approaches to uniformly scale the number of channels in the model's backbone network, feature fusion network, and shared head network based on image resolution. It integrates geometric factors (centre point distance, overlap rate, and scale) into the regression loss of target prediction, making the regression of the target prediction box more stable. This results in a 2.8% increase in mAP value on the Pascal VOC dataset, while having a 4.3x smaller network size and improving both GPU and CPU latency by 12.2% and 20.0%, respectively (Chu, Yan, Guo, Jianpeng, Shan, Wen, & Wang, Zhengkui, 2022).

FCOS emerges as a significant simplification in detection frameworks, eliminating the need for anchor boxes and reducing hyper-parameter tuning. Hwang et al.'s (2022) HISFCOS enhances this by improving feature utilization, reflected in a notable accuracy gain on the PASCAL VOC dataset. These developments underscore a push towards more streamlined and effective object detection models.

Tian et al. (2019) introduced FCOS, an anchor-free and proposal-free detector, as a breakthrough in simplifying the detection framework while achieving improved accuracy. FCOS eliminates the need for pre-defined anchor boxes, reducing the complexity associated with anchor-based methods and avoiding hyper-parameter sensitivity, which is often a challenge in object detection models.

Further improvements to FCOS were made by Hwang, Lee, and Lee (2022), who introduced HISFCOS. This model uses a half-inverted stage block to minimize feature loss and reconstruct the feature pyramid. HISFCOS's innovative design leads to a notable 3.0 AP increase in detection accuracy on the PASCAL VOC dataset. The diverse detection capabilities of HISFCOS are further emphasized in the performance table (Table 4), which details the accuracy improvements across various object categories. This illustrates the ongoing trend in object detection models towards achieving higher efficiency and performance through innovative architectural modifications.

Table 4: Comparative accuracy results for FCOS and HISFCOS models on the PASCAL VOC dataset, detailing performance across various object categories, with and without the lightweight detection head (Adapted from Hwang, Lee, & Lee, 2022).

FCOS	**aero**	**bike**	**bird**	**boat**	**bottle**	**bus**	**car**	**cat**	**chair**	**cow**
	83.1	85.4	81.3	72.4	60.6	83.2	87.6	91.7	57.8	81.8
	table	**dog**	**horse**	**mbike**	**person**	**plant**	**sheep**	**sofa**	**train**	**tv**
	64.5	88.1	87.0	82.3	83.8	53.8	81.9	73.7	87.9	79.4
HISFCOS (w/o) Lightweight detection head)	**aero**	**bike**	**bird**	**boat**	**bottle**	**bus**	**car**	**cat**	**chair**	**cow**
	81.0	87.3	84.5	74.6	66.8	85.3	88.7	93.4	60.9	82.8
	table	**dog**	**horse**	**mbike**	**person**	**plant**	**sheep**	**sofa**	**train**	**tv**
	68.7	90.6	87.5	87.2	84.9	55.4	83.1	77.0	90.2	79.2
HISFCOS	**aero**	**bike**	**bird**	**boat**	**bottle**	**bus**	**car**	**cat**	**chair**	**cow**
	85.4	88.6	83.8	76.1	65.7	88.2	89.0	93.3	58.9	84.8
	table	**dog**	**horse**	**mbike**	**person**	**plant**	**sheep**	**sofa**	**train**	**tv**
	72.4	89.9	90.0	86.9	85.1	56.3	85.2	74.1	91.3	81.6

In Table 4, the PASCAL VOC dataset serves as the benchmark for evaluating the object detection accuracy of two models: the original FCOS and its more advanced iteration, HISFCOS. The table breaks down performance by object category, offering a clear view of the improvements HISFCOS brings to the table, especially when it incorporates a lightweight detection head. This data is critical for understanding the specific advancements in detection accuracy that HISFCOS contributes to the field.

These advancements highlight the ongoing evolution of single-stage object detection models. The move to more streamlined architectures such as EfficientFCOS, along with strategies to boost accuracy while cutting down on computational demands, signifies considerable progress in the domain. The consistent enhancement of these models is progressively expanding the limits of object detection capabilities.

3.3.6 Technological Advancements Overview

This section aimed to concisely summarize recent advancements in object detection technologies, focusing on the objectives and outcomes associated with multiscale feature aggregation, attention-based models, one-stage fully convolutional detectors, and advanced network frameworks like PANet. The primary purpose of this section was to highlight how these technological innovations collectively mark a significant advancement in the field, notably enhancing the accuracy, efficiency, and sophistication of detection systems. By exploring transformer-based models such as Omni-DETR, detailing advancements in attention mechanisms within CNNs, and examining the evolution of models like FCOS and its variants, this summary underscored the substantial progress achieved. It showcased the transition from traditional methods to cutting-edge approaches, paving the way for future innovations. Ultimately, this section demonstrated the dynamic progression of object detection research and its promising potential to transform various applications, from autonomous driving to advanced surveillance systems.

3.4 SME Object Detection: Resource Constraints and Tech Choices

In the rapidly evolving technological landscape, small and medium-sized enterprises (SMEs) grapple with obstacles such as financial constraints, limited access to cutting-edge technological expertise, and challenges in adopting flexible object detection methods suited to their unique operational contexts. These challenges have been further intensified by the COVID-19 pandemic, which has demanded rapid adaptation and innovation from SMEs under tight resource constraints. This section explores how SMEs manage the adoption of object detection technologies amidst these specific challenges: navigating financial limitations that restrict their ability to invest in new technologies, bridging the knowledge gap in a field that is both specialized and rapidly advancing, and selecting object detection technologies that are sufficiently adaptable to fit within their operational constraints. The discussion aims to provide insights into effective strategies that enable SMEs to integrate advanced object detection technologies into their operations, overcoming these distinct challenges to leverage the benefits of technological advancements.

3.4.1 Introduction to SME Challenges in Technology Adoption

The COVID-19 pandemic dramatically impacted the global economy, with small and medium-sized enterprises (SMEs) among the hardest hit (Sarker et al., 2022). These businesses, integral to economic development and employment, have been compelled to reassess their operational strategies. The pandemic's enforced lockdowns and subsequent economic downturn exposed the vulnerability of SMEs, leading to significant financial losses and, in many cases, closures (Mishrif & Khan, 2023). In response, SMEs have increasingly turned to technology and innovation as essential tools for survival and competitiveness. This technological pivot, however, has been fraught with challenges due to SMEs' inherent limitations in resources and expertise.

3.4.2 The Impact of Limited Resources on Technology Selection

Despite the recognized necessity for technological adaptation, SMEs face significant barriers due to their limited resources. These constraints are not merely financial but also encompass restricted access to advanced technological infrastructure and a scarcity of in-house technical expertise. The challenge is twofold: SMEs must identify technologies that are both affordable and aligned with their operational capabilities. During the pandemic, the importance of digital skills and innovation was emphasized, yet a significant research gap remains concerning technology adoption in developing countries, where the majority of SMEs are located (Mishrif & Khan, 2023). The choice of technology, particularly in the realm of object detection, must therefore consider these limitations while also providing the potential for enhanced performance and a competitive edge. The normative literature has focused on the identification of factors such as benefits, barriers, and costs affecting the integration technologies (e.g., Electronic Data Interchange (EDI) and Enterprise Resource Planning (ERP)) adoption in SMEs. However, it is unclear whether these factors efficiently explain SMEs’ decision-making process related to emerging integration technologies like EAI and Web Services adoption (Themistocleous, 2003).

3.4.3 The Role of Technological Know-How in SMEs

The lack of technological knowledge within SMEs (Kareem et al., 2021; Valdez-Juárez, García-Pérez de Lema, & Maldonado-Guzmán, 2022) further complicates the adoption of advanced technologies like object detection. Previous research highlights the necessity for SMEs to substantially improve their technological innovation capabilities to ensure growth and sustainability (Mishrif & Khan, 2023). This need for improvement is not just in the adoption of new technologies but also in understanding and leveraging them effectively within their business operations. The pandemic has accelerated the technological transformation, prompting SMEs to quickly adapt to new technologies for managing their operations under crisis conditions. This rapid adaptation, however, has often been hamstringed by a lack of in-depth technological understanding and expertise within these organizations. Whipp and Rosenfeld (1989) and Caldeira and Ward (2003) emphasize the internal and external resources to analyse the IS/IT implementation in SMEs. They state that the internal resources include financial resources, human resources, management perspectives and attitudes, IS/IT competences, organizational structure, power relationships, and user attitudes. The external resources cover external expertise (e.g., vendors' support or consultant effectiveness), technology available, and business environment (e.g., clients and suppliers' pressure to adopt IS/IT) (Themistocleous, 2003).

3.4.4 Adaptability of Object Detection Methods for SMEs

In the context of SMEs, selecting and implementing object detection technologies comes with specific challenges. A critical aspect is the choice of technology. SMEs often face difficulties in choosing technologies that align with their operational requirements without overburdening their resources (Aura Technology, n.d.). Moreover, maintaining data security, especially when managing large volumes of data required for training object detection (OD) models, presents a considerable challenge. This is particularly relevant for SMEs, which may become prime targets for cybercriminals due to their typically less developed security infrastructures (Aura Technology, n.d.). Therefore, ensuring the security of object detection technologies is vital. Cost is another substantial barrier for SMEs. A study indicates that financial constraints are among the biggest hurdles SMEs face in adopting new

technologies (Chen et al., 2016). Exploring cost-effective solutions, such as open-source tools or cloud-based services, can be a practical approach for SMEs. Moreover, the integration of these technologies into existing systems is a challenge, emphasizing the need for compatibility with SMEs' current infrastructure. Finally, SMEs frequently lack the technical expertise required for deploying and managing new technologies effectively. This challenge underscores the importance of selecting user-friendly object detection technologies and ensuring the availability of training and support services (Chen et al., 2016). Addressing these challenges requires a multi-faceted approach that balances performance, security, cost, and ease of integration.

To effectively address the implementation challenges of object detection solutions, several platforms offer turnkey solutions tailored for Small and Medium-sized Enterprises (SMEs). Notably, the Google Cloud Vision API exemplifies a scalable, cloud-based service that simplifies object detection for organizations with limited technical resources. Alongside Amazon Rekognition, it underscores the importance of data security with robust measures. Similarly, Microsoft Azure Computer Vision and IBM Watson Visual Recognition provide SMEs with advanced tools for image and video analysis, accompanied by comprehensive support and training resources to alleviate the technical expertise barrier.

Implementing Google Cloud Vision API, for example, entails a series of straightforward steps, illustrating the platform's accessibility:

1. Create a Google Cloud project, enabling the organization to access Google's cloud-based services.
2. Enable the Cloud Vision API for the project, granting access to object detection capabilities.
3. Generate an API key for authentication, ensuring secure access to the service.
4. Optionally set up a service account for enhanced security, offering more granular control over API access.

These steps demonstrate the seamless integration of advanced object detection technologies with SMEs' existing infrastructure and operational demands, ensuring both compatibility and cost-effectiveness.

3.5 Future Trends and Challenges

The landscape of object detection technology is rapidly evolving, (Wang, Jiao, & Xu, 2021) driven by advancements in deep learning and increasing computational capabilities. This evolution presents a range of opportunities and challenges, particularly for Small and Medium-sized Enterprises (SMEs) that seek to leverage these technologies for enhanced operational efficiency and competitive advantage. While the advancements offer unprecedented accuracy and capabilities in object detection, they also bring forth challenges in terms of computational intensity and the need for high precision in diverse application scenarios. This section delves into the emerging trends in object detection technologies, examining their potential impact on SMEs. It also explores the ongoing challenges these businesses face, particularly in achieving high precision in complex environments and managing the computational demands of advanced object detection models. The discussion is anchored on recent academic research and studies, providing a comprehensive view of the future trajectory of object detection technology and its implications for SMEs.

3.5.1 Emerging Technologies and SME Applications

From the late 2010s to the early 2020s, advancements in object detection have been significantly influenced by deep learning (Safdar et al., 2022; Wu, Sahoo, & Hoi, 2020). Hector et al. (2021) discusses the high accuracy of deep neural networks (DNNs) in complex detection scenarios, making them suitable for smart city applications. However, the deployment of these technologies in SME settings is challenging due to their high computational demands. To counter this, elastic neural networks have been proposed as a resource-adaptive solution, effectively balancing computational complexity and detection accuracy (Hector et al., 2021).

In addition to these technological advancements, the field is witnessing a shift towards more holistic and integrated approaches. A study by Kaur et al. (2021) outlines several anticipated areas of focus within the object detection field, including expressive learning, quick training, and the development of universal object detectors. Furthermore, federated learning and brain-inspired computing are identified as promising areas for enhancing object detection capabilities, particularly in resource-constrained environments like those faced by SMEs (Kaur et al., 2021).

3.5.2 Challenges in High Precision and Computational Intensity

In addressing computational efficiency and precision in object detection, Kaur et al. (2021) point to the adoption of transformers and improved hyper-parameter optimization as keys to enhancing model performance. For SMEs, where both precision and computational resources are critical, these advancements could lead to more effective object detection without disproportionately increasing computational demands.

Aligning with these insights, the scatter plot in Figure 5 from Vaidya and Paunwala (2019) presents the empirical relationship between an algorithm's processing speed and its accuracy. The plot serves as a valuable reference for SMEs to assess object detection algorithms, emphasizing the necessity to evaluate the performance implications of adopting newer models that promise to balance speed with accuracy. Thus, for an SME prioritizing quick and accurate object detection, understanding this trade-off is essential for selecting a system that aligns with their operational needs and resource capacities.

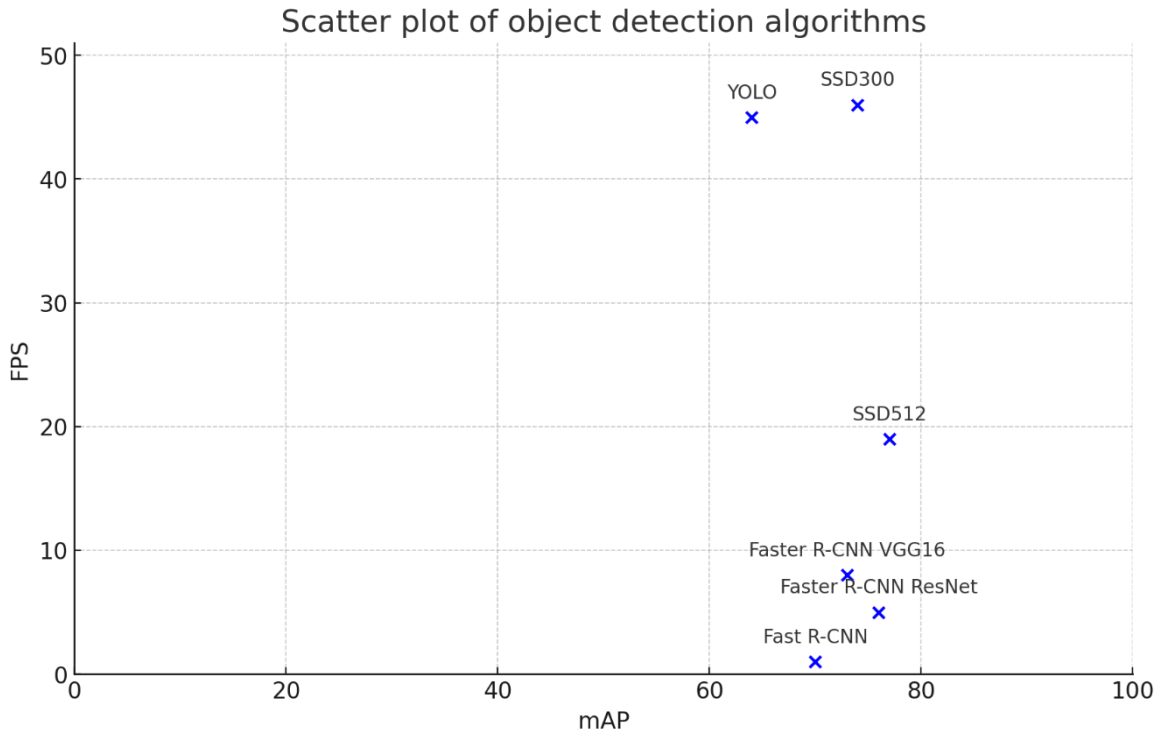

Figure 8: Performance Trade-offs in Object Detection Algorithms. The plot illustrates balance between speed (FPS) and accuracy (mAP) across different object detection algorithms, highlighting the variations in perform suitable for SME applications (Adapted from Vaidya and Paunwala (2019).

Moreover, the study by Zhou et al. (2021) underscores the importance of multi-scale object detection and addressing intra-class variance, which are critical for security applications. The ability to detect objects of varying sizes and similar class types with high precision is essential for SMEs engaged in activities like surveillance and quality control (Zhou et al., 2021).

3.5.3 Future Directions and Challenges in Object Detection for SMEs

The field of object detection continues to evolve, with researchers exploring new directions to address open challenges. Integrating object detection with tasks like segmentation and pose estimation can lead to more comprehensive scene representations, beneficial for applications in robotics, autonomous vehicles, and virtual reality. Developing few-shot and zero-shot learning methods can improve scalability and adaptability in object detection

(Tasnim & Qi, 2023). Additionally, ensuring fairness, accountability, and transparency in object detection algorithms is essential for their ethical deployment.

3.6 Key Insights and Future Directions

The literature review highlights significant advancements in object detection technologies, especially in their application to SMEs, underlining the importance of digital technologies for competitiveness and growth in the modern era (Chen et al., 2016). This field, a crucial aspect of computer vision, has seen remarkable developments in enhancing accuracy, efficiency, and capabilities in real-time and video object detection, crucial for broad applications including autonomous vehicles and surveillance systems ("Object Detection in 2023: The Definitive Guide," 2023). Yet, challenges persist, particularly the need for greater precision in complex environments and the management of computational demands crucial for resource-limited SMEs, alongside the underexplored potential of integrating emerging technologies like Blockchain and Distributed Ledger Technologies (DLT) to offer significant benefits to SMEs (Chen et al., 2016).

Looking forward, the field is set to continue its focus on enhancing the efficiency and accuracy of object detection algorithms, with future trends indicating new avenues for practical applications that promise to revolutionize key industries further. Despite the rapid evolution of object detection technologies presenting both opportunities and challenges, the importance of ongoing research and development is underscored, especially in leveraging these advancements within the unique operational constraints and market dynamics of SMEs. This dynamic environment emphasizes the critical need for SMEs to stay abreast of technological advancements to harness their potential effectively ("Object Detection in 2023: The Definitive Guide," 2023; Chen et al., 2016).

4 Advanced Technologies for Electrical Diagram Object Detection

In this section, we compare the performance of Haar Cascade, Faster R-CNN, and YOLOv8 technologies on electrical diagram datasets. Object detection, a crucial component of

automated diagram analysis, has made significant advancements recently. The research aims to benchmark and compare these methods' effectiveness in real-life situations, contributing to the broader context of object detection technology in the field.

Haar Cascade, Faster R-CNN, and YOLOv8 were selected for their historical significance, adaptability, and diverse complexity and efficiency levels. Haar Cascade is known for its compact size and effectiveness in basic applications. Faster R-CNN, recognized for its precision and state-of-the-art performance, utilizes convolutional neural networks to achieve remarkable accuracy in object identification. YOLOv8 represents the pinnacle of real-time detection technology, offering unmatched speed and accuracy for tasks requiring rapid processing.

The analysis is based in the specifics of electrical diagrams, employing a detailed methodology that includes data collection, preparation, model optimization, and evaluation. With a custom-annotated dataset of nearly 500 PDFs, our study directly relates to and applies within the sector.

Evaluation will be based on a comprehensive set of metrics, including accuracy and processing time, to provide an in-depth understanding of each method's strengths and weaknesses. This approach allows for a detailed assessment, highlighting not only performance but also practical aspects such as computational expense and ease of implementation.

The research plan anticipates and addresses key challenges, particularly focusing on computational constraints, to ensure the study's findings are both relevant and actionable for scholarly discussion and practical application. The study aims to outline the strengths and weaknesses of each object detection technique, providing critical analysis and recommendations to the field. By bridging the gap between theoretical knowledge and practical application, this research contributes to the future development and implementation of object detection technologies in electrical diagram analysis, with a keen focus on optimizing computational efficiency and model performance.

4.1 Methodology Overview and Data Preparation

This section details the methodologies and preparation steps for evaluating the three distinct object detection technologies in electrical diagram analysis. The core of the methodology is designed around a dual approach: preparing a dataset annotated in the YOLO format to suit the requirements of each technology and adapting each technology to the specific challenges of accurately detecting and classifying objects within complex electrical diagrams. This preparation involves a process of converting these YOLO annotations, which are normalized coordinates, into formats compatible with the HAAR Cascade's precise pixel coordinates requirements and the XML format requisite for Faster R-CNN and ensuring that the dataset accurately represents the diversity of components found in electrical diagrams.

Central to the methodology is the management of software dependencies and optimization of environment setups for model training. Recognizing the limitations posed by OpenCV 4.x in classifier training, we employed Docker to create a compatible training environment with OpenCV 3.x for HAAR Cascade, chosen for its support of essential training tools. Similarly, for both Faster R-CNN and YOLOv8, we employed Conda to simplify the setup process, ensuring consistency and efficiency in software across these techniques. This purposeful integration of Docker and Conda emphasizes the crucial role of efficient dependency management in conducting evaluations of different models. Through this comprehensive strategy, the research aims to create a detailed comparative analysis of each method's performance, offering valuable insights into their practical utility and efficiency in real-world applications.

4.2 Technology-Specific Implementations

4.2.1 HAAR Cascade Classifier

To adapt our dataset for HAAR Cascade training, we first converted the annotations from their original YOLO format, which uses normalized coordinates, into a form suitable for HAAR Cascade. This adaptation was crucial for creating a training dataset that HAAR Cascade could effectively learn from. Subsequently, the dataset was organized into positive

samples (images containing the objects of interest) and negative samples (images without the objects), a necessary step for HAAR training which enables the generation of .vec files for positive samples and the use of plain images for negative samples. The creation of these files is essential for training the classifier to distinguish between objects of interest and the background.

Leveraging Docker containers was necessary in this process, providing a consistent and controlled environment that mitigated potential issues arising from different legacy OpenCV versions. This approach ensured the efficiency and effectiveness of the HAAR Cascade training process, allowing us to maintain a stable setup across various computational environments and OpenCV configurations.

4.2.2 Faster R-CNN

For the implementation of Faster R-CNN, we adapted its architecture to the specific demands of electrical diagram object detection. This adaptation involved customizing the Faster R-CNN ResNet50 FPN V2 architecture, leveraging ResNet50's deep learning capabilities and FPN V2's efficient handling of varying object scales. The dataset's annotations, initially formatted for YOLO, were converted into the XML format required by Faster R-CNN. This conversion was accomplished using a customized script, which streamlined the process without necessitating further adjustments to individual annotations. The Conda environment was helpful in this process, providing a seamless setup and ensuring consistency across the software stack.

4.2.3 YOLOv8

Given the dataset was already formatted in YOLO annotations, no additional adjustments were necessary. As it is with Faster R-CNN there was no need for separating the dataset into different categories for objects of interest and background or negative samples, as YOLO inherently distinguishes between annotated objects and the rest of the image as background. Including images without any objects (or annotations) as explicit negative samples can enhance the model's accuracy by teaching it to identify scenes without target objects, thereby

reducing false positives. However, such negative samples should be used sparingly, maintaining a balance to prevent bias in the model training process.

4.3 Model Training

This passage provides a more detailed overview of the training and ongoing refinement processes for three distinct object detection methods (Haar Cascade, Faster R-CNN, and YOLOv8) as applied to electrical diagram datasets. It delves into the tailored training approaches, the optimization of parameters, and the iterative testing phases specific to each method, emphasizing the critical role of continuous refinement based on testing results and evaluation metrics.

4.3.1 Haar Cascade Classifier

In training a cascade classifier for object detection with OpenCV 3.x, two critical steps were executed. The first involved using `opencv_createsamples` to generate a vector file from positive samples—images featuring the target object. This required specifying an input file (`positives.txt`), the quantity of samples (515), and their dimensions (24x24 pixels), resulting in `positives.vec` for the next phase. Notably, larger sample dimensions induced core dump errors due to limited RAM, resolved by reducing sample size or alternatively, by increasing RAM to avoid compromising model accuracy. Processing images in smaller batches was also recommended to manage memory usage efficiently.

```
opencv_createsamples -info positives.txt -num 515 -w 24 -h 24 -vec positives.vec
```

Subsequently, opencv_traincascade was utilized for training, leveraging both the vector file and negative samples—images absent of the target. The command:

```
opencv_traincascade -data ./ -vec ./positives.vec -bg ./negatives.txt -numPos 500 -numNeg 122 -numStages 10 -w 24 -h 24
```

included parameters for the save directory, positive vector file, negative sample list, number of positive (500) and negative (122) samples, training stages (10), and sample dimensions, aligning with the initial phase. Adjusting training stages impacted model accuracy, processing time, and false positive rates, highlighting the balance between efficiency and performance.

This methodological approach stresses the importance of parameter optimization, reflecting the challenge of training in resource-limited environments and suggesting solutions to maintain training quality without compromising model integrity.

4.3.2 Faster R-CNN

In this study, the Faster R-CNN model was enhanced by integrating the ResNet50 FPN V2 architecture, leveraging the principles of transfer learning to augment the model's object detection capabilities. Transfer learning, as described by Pan and Yang (2010), involves applying knowledge gained from one domain to a different but related problem domain. This approach significantly streamlines the training process by using a model pre-trained on a large dataset, such as ImageNet, thus enabling the adaptation of the model to our specific object detection tasks with less computational overhead and data requirement.

The employment of the COCO dataset further exemplifies the application of transfer learning, wherein the model, through exposure to a diverse set of annotated images, benefits from the wide-ranging visual information, thereby enhancing its detection performance across various object types and scenarios (Lin et al., 2014).

Following the integration of the ResNet50 FPN V2 architecture and the COCO dataset, the actual implementation of the model training was executed via a specific command-line instruction. The command:

```
python train.py --model fasterrcnn_resnet50_fpn_v2 --config data_configs/diagrams.yaml --epochs 10 --project-name fasterrcnn_resnet50_fpn_v2_diagrams --use-train-aug --no-mosaic
```

was constructed to define the model architecture (`fasterrcnn_resnet50_fpn_v2`), specify the custom dataset via (--config `data_configs/diagrams.yaml`) which contained instructions of paths to directories and number of classes present, and set the training duration (`--epochs 10`). The epoch count was chosen to grant each model approximately the same amount of time for development, ensuring a fair comparison. Using the `--use-train-aug` parameter means incorporating variations in the training data through methods like rotating, scaling, or cropping images. This enhances the model's robustness by teaching it to recognize objects under various conditions, improving its ability to generalize. The `--no-mosaic` parameter was used to exclude mosaic augmentation from the training process. This decision was made to simplify training and focus on augmentation methods more directly beneficial to the specific dataset being used, aiming to improve the model's accuracy and computational efficiency by avoiding the complexity mosaic augmentation can introduce.

YOLOv8

During our study on training the YOLOv8 object detection model, we specifically utilized Metal Performance Shaders (MPS) on Apple's M1 chip to ensure compatibility and performance consistency across different detection methods. Although YOLO models typically show enhanced performance on NVIDIA GPUs due to CUDA optimization, the utilization of MPS on the M1 chip was essential for achieving a fair comparison within our research context, despite the conventional preference for CPU or NVIDIA GPU training in other scenarios. This decision allowed us to maintain a balanced evaluation framework for the object detection methods under review.

Training the YOLOv8 model using MPS on the M1 chip, given its optimization for CUDA and NVIDIA GPUs, posed certain challenges, especially regarding the model's efficiency and accuracy. The training process was initiated by loading a pre-trained `yolov8n.pt` model, aiming to capitalize on the model's pre-existing knowledge base. Our training command, adapted for MPS, looked as follows:

The command `yolo detect train data=dataset.yaml model=yolov8n.pt epochs=100 imgsz=640 device=mps` was used in our study to train the YOLOv8 model on the M1 chip using Metal Performance Shaders (MPS). This section breaks down the elements of the command and reflects on the training performance without repeating previous information.

- `data=dataset.yaml`: Specifies the dataset configuration file, which includes paths to the training and validation datasets, class names, and other dataset-related parameters.

- `model=yolov8n.pt`: Indicates the YOLOv8 Nano model file to be used for training.

- `epochs=100`: Sets the number of training cycles through the entire dataset. In this case, the model will be trained for 100 cycles, allowing for sufficient learning without overfitting.

- `imgsz=640`: Defines the input image size as 640x640 pixels, which is a standard size for YOLO models that balances between speed and accuracy.

- `device=mps`: Directs the training to utilize the MPS backend on the M1 chip, optimizing the process for Apple's hardware architecture.

Training the YOLOv8 model on Apple's M1 chip with MPS showcased the model's flexibility across different hardware but also underscored the importance of selecting optimal hardware for deep learning. Moreover, leveraging NVIDIA GPUs for training can significantly improve performance, enabling the trained models to be efficiently deployed even on less powerful devices. This approach not only optimizes the training phase but also ensures versatile deployment capabilities across various platforms. Hardware-aware training and co-designing model architectures with an understanding of hardware capabilities have been shown to further enhance deep learning model performance (Yang et al., 2022; Spoon et al., 2021).

In conclusion, while training on the M1 chip using MPS offers insights into hardware adaptability, using NVIDIA GPUs with deep learning optimizations presents clear advantages in terms of efficiency and performance. This is crucial for developing strategies to train complex models like YOLOv8 in various computational environments.

4.4 Evaluation and Comparative Analysis

This section goes thorough evaluation and comparative analysis of performance metrics derived from the application of Haar Cascades, Faster R-CNN, and YOLOv8 on datasets of electrical diagrams. It scrutinizes essential metrics including accuracy (precision, recall, F1 score), processing time, model training duration, model size, and complexity.

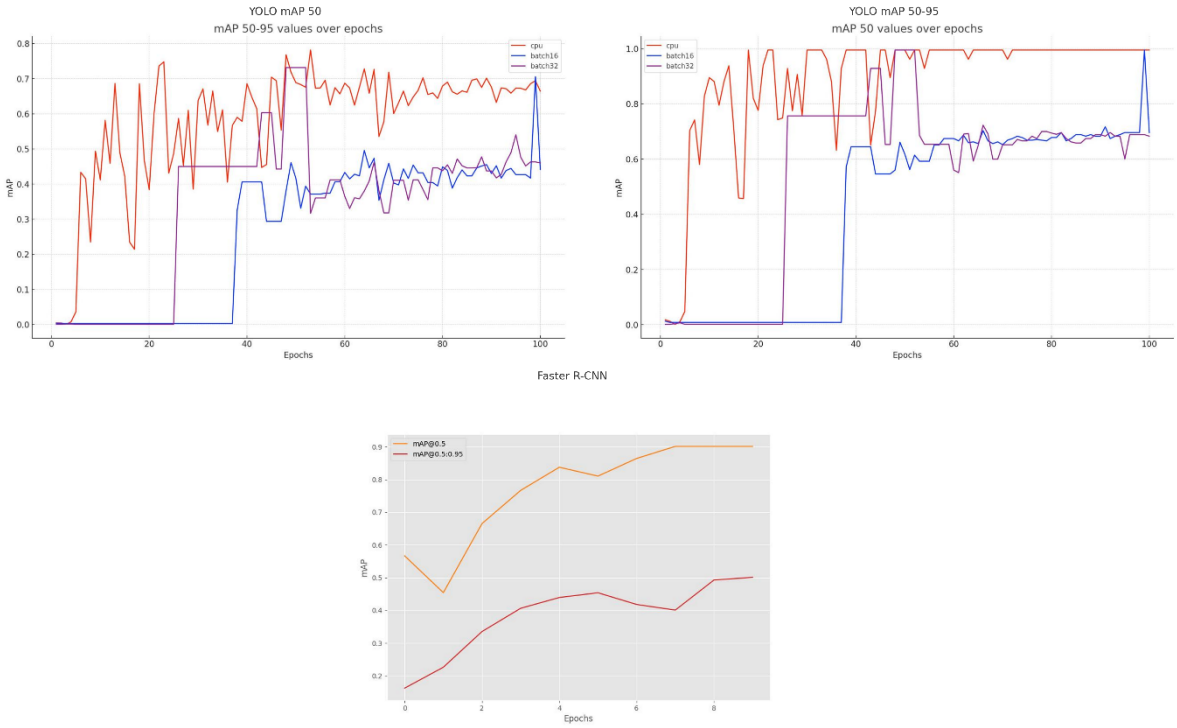

Figure 9: Training performance of Faster R-CNN and YOLOv8 on electrical diagram datasets. The mAP50 and mAP50-95 scores over epochs illustrate Faster R-CNN's high precision and YOLOv8's balance between accuracy and processing speed, highlighting fluctuations during training.

Table 5: Performance Metrics of Haar Cascades, Faster R-CNN, and YOLOv8 in Object Detection

Method	Precision/Score	Recall/mAP50-95	F1 Score	Inference Time/Image	Training Time	Model Size	Remarks
Haar Cascades	Precision: 0.3125	Recall: 1	0.4762	3.576 sec	2 min 44 sec	4.4K	Moderate setup complexity, limited accuracy
Faster R-CNN	mAP50: 0.900	mAP50-95: 0.492	-	1.34 sec	1.635 hours	173.4 MB	High complexity, high accuracy
YOLOv8	mAP50: 0.645	mAP50-95: 0.421	-	-	1.094 hours	6.3 MB	High efficiency, compact size, adaptable

In the evaluation of object detection algorithms, Haar Cascade, Faster R-CNN, and YOLOv8 each present distinct characteristics and constraints. Specifically, Haar Cascade is noted for its operational efficiency and low computational requirements. However, it cannot employ the mean Average Precision (mAP) as a metric for performance assessment due to it not predicting with confidence score and instead just saying if something is there or not without saying how sure it is.

Conversely, object detection models like Faster R-CNN and YOLOv8 rely on mAP as a key performance indicator. As illustrated in Figure 9, while Faster R-CNN consistently demonstrates high precision in its mAP scores, the YOLOv8's mAP scores, when trained on a Metal Performance Shaders (MPS) chip, exhibit an initial phase of stagnation. This plateau early in the training suggests an issue with MPS's training optimization for YOLOv8, potentially due to the complex nature of the electrical diagram data. It contrasts with the behaviour observed when training on a CPU, where YOLOv8 does not exhibit this early performance lag, underlining the need for specific tuning or methodological adjustments to fully leverage MPS capabilities from the start.

These models predict bounding boxes and their associated class labels, enabling detailed assessments of precision at various confidence levels across different classes. Such evaluations are essential for understanding each model's effectiveness in differentiating between object classes and accurately localizing objects within images (Mahendrakar, Ekblad, Fischer, White, Wilde, Kish, & Silver, 2023).

When training on a Metal Performance Shaders (MPS) chip, the distinction among object detection methods such as Haar Cascade, Faster R-CNN, and YOLOv8 becomes evident, each with its own operational focus. Haar Cascade is noted for its applicability to basic detection tasks with its low computational demand. Conversely, Faster R-CNN is tailored for tasks that demand greater precision in object detection. YOLOv8 offers a balance between detection accuracy and processing speed, making it suitable for various applications, including those constrained by computational resources. This comparison underscores the unique advantages and applications of each detection method, particularly in the context of training on an MPS chip, highlighting their adaptability and performance in specific computational environments.

4.5 Challenges and Solutions

Throughout the exploration of object detection algorithms applied to electrical diagram datasets, this research encountered a range of challenges inherent to the methods utilized: Haar Cascade Classifier, Faster R-CNN, and YOLOv8. The methodological accuracy and innovative solutions developed in response to these challenges are central to the research findings.

The initial deployment of the Haar Cascade Classifier encountered a core dump error due to the dimensional limitations of the .vec file. This issue, along with a high rate of false positives—a common problem in object detection tasks—significantly hindered the refinement process. To overcome these obstacles, the dimensions for .vec file creation were decreased to meet the requirements of the OpenCV tool. Employing OpenCV 3.x tools within a Docker container facilitated addressing version incompatibilities, enabling successful generation of .xml files. Moreover, adjusting parameters such as scaleFactor and minNeighbors led to a notable improvement in detection accuracy.

The implementation of Faster R-CNN presented the dual challenges of ensuring computational efficiency and achieving high precision, despite the model's complexity and computational demands. Through hyperparameter tuning and data preparation, the model attained mAP scores of 0.900 (mAP50) and 0.492 (mAP50-95), which underscore its

precision. However, this came at the cost of increased processing times and model size, highlighting its suitability for contexts where computational resources are not a limiting factor.

The performance gap between YOLOv8 and Faster R-CNN, particularly in mAP scores, highlights a critical challenge: the impact of training environments on model effectiveness. YOLOv8's lower mAP scores, when compared to Faster R-CNN, suggest training on Multi-Processor Systems (MPS) rather than on NVIDIA GPUs with at least 8GB of memory and CUDA support may compromise training efficiency and model accuracy. NVIDIA GPUs facilitate faster and more efficient parallel processing crucial for deep learning model training. Transitioning YOLOv8's training to the recommended NVIDIA hardware could enhance its learning capability, potentially improving accuracy and performance. Once trained on the appropriate system, the model can be deployed across various setups, maintaining flexibility in application.

5 Conclusion

In this thesis, we have explored the significant advancements in object detection technologies, focusing on their application within the context of small and medium-sized enterprises (SMEs). This investigation reveals how these technologies, which form a crucial part of the broader field of computer vision, have evolved to offer remarkable improvements in accuracy, efficiency, and the capability for real-time detection. The study delves into various methodologies such as Haar Cascades, despite its less advanced nature, Faster R-CNN, and YOLOv8, emphasizing their respective impacts on operational efficiency and the competitive landscape for SMEs. The critical analysis provided aims to guide SMEs in selecting appropriate object detection technologies that align with their operational needs and resource constraints, acknowledging the challenges posed by complex environments and computational demands.

This research directly addresses the main question regarding the strategic selection and implementation of object detection technologies by SMEs. It is evident that the thoughtful

integration of these technologies can significantly influence an SME's operational processes and market positioning. However, this potential benefit is accompanied by the need for a careful approach that considers the balance between technological benefits and the inherent limitations of SMEs, such as resource scarcity. Future research directions highlighted in this thesis, including the potential integration of Blockchain and Distributed Ledger Technologies (DLT) in object detection processes, suggest promising areas that could yield substantial benefits for SMEs. These emerging technologies present an opportunity for innovative applications and efficiencies that have yet to be fully explored in the SME context.

Furthermore, the contributions of this thesis to the academic and practical understanding of object detection technologies in SMEs are manifold. By providing a comprehensive comparison of different object detection methods, this work aims to demystify the technological landscape for SMEs, enabling them to make informed decisions that best suit their specific requirements. The findings underscore the importance of continued research and development in this fast-evolving field to ensure that SMEs are equipped to navigate the challenges and seize the opportunities presented by digital technologies.

In summary, as object detection technology continues to advance rapidly, its application in SMEs poses unique challenges but also opens avenues for transformative growth and innovation. This thesis serves as a resource for SMEs looking to leverage object detection technologies, offering insights into the nuances of various methodologies and their practical implications. The continuous evolution of this field underscores the need for SMEs to stay informed and adaptable, ensuring they can leverage these technologies to enhance their operations and competitive advantage in a digitalized market landscape.

References

Abdelsalam, A., Happonen, A., Kärhä, K., Kapitonov, A., & Porras, J. (2022). Toward Autonomous Vehicles and Machinery in Mill Yards of the Forest Industry: Technologies and Proposals for Autonomous Vehicle Operations. *IEEE Access*, 10, 88234-88250. https://doi.org/10.1109/ACCESS.2022.3199691

Aura Technology. (n.d.). Five technology challenges for SMEs. Retrieved from www.auratechnology.com

Binjie, X., & Hu, J. (2014). Fabric appearance testing. In The Hong Kong Polytechnic University, China. https://doi.org/10.1533/9781845695064.148

Chen, X., et al. (2016). Small and Medium Enterprises (SMEs) facing an evolving technological era: a systematic literature review on the adoption of technologies in SMEs. Emerald Insight. Retrieved from https://www.emerald.com/insight

Chu, Y., Guo, J., Shan, W., & Wang, Z. (2022). EfficientFCOS: An Efficient One-stage Object Detection Model based on FCOS. Proceedings of the IEEE International Conference on Computer Supported Cooperative Work in Design.

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. arXiv:cs/0505040.

Feng, X. (2022). Better Fusion of Multi-scale Features for Remote Sensing Object Detection. In 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE). IEEE. https://doi.org/10.1109/ICCECE54139.2022.9712836

Girshick, R. (2015). Fast R-CNN. arXiv preprint arXiv:1504.08083. Also in Proceedings of the IEEE International Conference on Computer Vision (ICCV).

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. CVPR 2014.

Han, L. (2020). Exploiting Better Feature Aggregation for Video Object Detection. MM '20: Proceedings of the 28th ACM International Conference on Multimedia. https://doi.org/10.1145/3394171.3413927

Hashemi, N. S., Aghdam, R. B., Ghiasi, A. S. B., & Fatemi, P. (2016). Template Matching Advances and Applications in Image Analysis. [PDF document]. Retrieved from arXiv:1610.07231v1

Hector, R., Umar, M., Mehmood, A., Li, Z., & Bhattacharyya, S. (2021). Scalable object detection for edge cloud environments. Frontiers in Sustainable Cities, 3, Article 675889. https://doi.org/10.3389/frsc.2021.675889

Hwang, B., Lee, S., & Lee, S. (2022). HISFCOS: Half-Inverted Stage Block for Efficient Object Detection Based on Deep Learning. J. Imaging.

Kareem, H. M., Aziz, K. A., Maelah, R., Mohd Yunus, Y., Alsheikh, A., & Alsheikh, W. (2021). The Influence of Accounting Information Systems, Knowledge Management Capabilities, and Innovation on Organizational Performance in Iraqi SMEs. *International Journal of Knowledge Management*, April-June 2021. https://doi.org/10.4018/IJKM.2021040104

Kaur, J., Singh, P., & Josan, G. S. (2021). A study on generic object detection with emphasis on future research directions. Journal of King Saud University-Computer and Information Sciences. https://doi.org/10.1016/j.jksuci.2021.08.001

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012.

Li, W., Liu, K., Zhang, L., & Cheng, F. (2020). Object detection based on an adaptive attention mechanism. Scientific Reports, 10(1), 11307.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. Proceedings of the European Conference on Computer Vision (ECCV).

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. ECCV 2016.

Mahendrakar, T., Ekblad, A., Fischer, N., White, R. T., Wilde, M., Kish, B., & Silver, I. (2023). Performance Study of YOLOv5 and Faster R-CNN for Autonomous Navigation around Non-Cooperative Targets. arXiv:2301.09056. https://ar5iv.org/pdf/2301.09056

Mei, Y., Fan, Y., Zhou, Y., Huang, L., Huang, T. S., & Shi, H. (2020). Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5690-5699).

Mei, Y., Fan, Y., Zhang, Y., Yu, J., Zhou, Y., Liu, D., Fu, Y., Huang, T. S., & Shi, H. (2021). Pyramid attention networks for image restoration. arXiv preprint arXiv:2004.13824.

Mishrif, A., & Khan, A. (2023). "Technology adoption as a survival strategy for small and medium enterprises during COVID-19." Journal of Innovation and Entrepreneurship, 12, Article 53. DOI: 10.1186/s13731-023-00317-9.

Object Detection in 2023: The Definitive Guide. (2023). Augmented Startups. Retrieved from https://www.augmentedstartups.com/

OECD. (n.d.). Enterprises by business size. Retrieved from https://data.oecd.org/entrepreneur/enterprises-by-business-size.htm

OpenAI. (2024). ChatGPT (March 2023 version) [3.5]. https://chat.openai.com/chat

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 22(10).

Rajput, P., Mittal, S., & Narayan, S. (2020). Improving Accuracy and Efficiency of Object Detection Algorithms Using Multiscale Feature Aggregation Plugins. In Artificial Neural Networks in Pattern Recognition (ANNPR 2020), Lecture Notes in Computer Science (LNCS, Vol. 12294, pp. 65–76). Springer. https://doi.org/10.1007/978-3-030-58309-5_5

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified Real-Time Object Detection. CVPR 2016.

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. CVPR 2017.

Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. ArXiv.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. NIPS 2015.

Safdar, M. F., Bilal, Z., Sharif, H., Ahmad, S., Rehman, F., & Maqsood, H. (2022). A Review of Recent Advances in Deep Learning for Object Detection. In 2022 3rd International Conference on Innovations in Computer Science & Software Engineering (ICONICS). https://doi.org/10.1109/ICONICS56716.2022.10100486

Sarker, M. R., Rahman, S. M. A., Islam, A. K. M. H., Bhuyan, M. F. F., Supra, S. E., Ali, K., & Noor, K. M. A. (2022). Impact of COVID-19 on Small- and Medium-sized Enterprises. Global Business Review, 1–14. https://doi.org/10.1177/09721509221093489

Song, T., Qin, W., Liang, Z., Qin, Q., & Liu, G. (2021). Research on CNN for Anti-missile Object Detection Algorithm Based on Improved Attention Mechanism. In Proceedings of

the 40th Chinese Control Conference. Xi'an Research Institute of High-Tech, China Academy of Launch Vehicle Technology, Equipment Department of Rocket Force.

Spoon, K., et al. (2021). Toward software-equivalent accuracy on transformer-based deep neural networks with analog memory devices. Frontiers in Computational Neuroscience.

Tan, M., et al. (2020). EfficientDet: Scalable and Efficient Object Detection. arXiv preprint arXiv:1911.09070.

Tasnim, S., & Qi, W. (2023). Progress in Object Detection: An In-Depth Analysis of Methods and Use Cases. European Journal of Electrical Engineering and Computer Science.

Themistocleous, M. (2003). Resource Constraints Related to Emerging Integration Technologies.

Tian, Z., Shen, C., Chen, H., & He, T. (2019). FCOS: Fully Convolutional One-Stage Object Detection.

Training Haar Cascades. Retrieved from https://hako.github.io/dissertation/

Usmani, U.A., Happonen, A., & Watada, J. (2022). A Review of Unsupervised Machine Learning Frameworks for Anomaly Detection in Industrial Applications. In *Intelligent Computing, SAI 2022, Lecture Notes in Networks and Systems*, 507, 158-189. https://doi.org/10.1007/978-3-031-10464-0_11

Usmani, U.A., Happonen, A., & Watada, J. (2024). Enhancing Medical Diagnosis Through Deep Learning and Machine Learning Approaches in Image Analysis. In *Lecture Notes in Networks and Systems*, 825, 449-468. https://doi.org/10.1007/978-3-031-47718-8_30

Vaidya, B., & Paunwala, C. (2019). Deep Learning Architectures for Object Detection and Classification. In M. K. Mishra, B. Majhi, S. Sa, & S. Sahoo (Eds.), Smart Techniques for a

Smarter Planet: Towards Smarter Algorithms (pp. 53–78). Springer. https://doi.org/10.1007/978-3-030-03131-2_4

Valdez-Juárez, L. E., García-Pérez de Lema, D., & Maldonado-Guzmán, G. (2022). Management of Knowledge, Innovation, and Performance in SMEs: Examining the Impact on Organizational Efficiency. Journal of Organizational and Strategic Novelties for Technology (JOASNT), https://doi.org/10.55945/joasnt.2022.1.3.65-72

Viola, P., & Jones, M. J. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, 1, I-511.

Wang, J., et al. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696.

Wang, P., Cai, Z., Yang, H., Swaminathan, G., Vasconcelos, N., Schiele, B., & Soatto, S. (2022). Omni-DETR: Omni-Supervised Object Detection with Transformers.

Wang, W., Zhao, S., Shen, J., Hoi, S. C. H., & Borji, A. (2019). Salient Object Detection with Pyramid Attention and Salient Edges. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Wang, Z., Jiao, B., & Xu, L. (2021). Visual Object Detection: A Review. Proceedings of the 40th Chinese Control Conference, July 26-28, Shanghai, China.

Wu, X., Sahoo, D., & Hoi, S. C. H. (2020). Recent advances in deep learning for object detection. Neurocomputing, 396. https://doi.org/10.1016/j.neucom.2020.01.085

Yang, X., et al. (2022). Tolerating noise effects in processing-in-memory systems for neural networks: a hardware–software codesign perspective. Advanced Intelligent Systems.

Ylä-Kujala, A., Kedziora, D., Metso, L., Kärri, T., Piotrowicz, W., & Happonen, A. (2023). Robotic process automation deployments: a step-by-step method to investment appraisal. *Business Process Management Journal*, 29(8), 163-187. https://doi.org/10.1108/BPMJ-08-2022-0418

Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., & Lee, B. (2022). A survey of modern deep learning based object detection models. Digital Signal Processing, 126, 103514. https://doi.org/10.1016/j.dsp.2022

Zhang, H., & Hong, X. (2019). Recent progresses on object detection: a brief review. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-019-07898-2

Zhang, Q., Zhang, H., Lu, X., & Han, X. (2022). Anchor-Free Small Object Detection Algorithm Based on Multi-scale Feature Fusion. In 2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI) (pp. 1-8). IEEE. https://doi.org/10.1109/PRAI55851.2022.9904251

Zheng, Y., Zhang, X., Zhang, R., & Wang, D. (2022). Gated Path AggregationFeature Pyramid Network for Object Detection in Remote Sensing Images.

Zhou, X., Wang, D., & Krähenbühl, P. (2021). Objects as Points. ArXiv preprint arXiv:1904.07850.

Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). Object Detection with Deep Learning: A Review. arXiv preprint arXiv:1807.05511.

Zou, Z., et al. (2020). A survey of modern deep learning-based object detection models. Digital Signal Processing, 107. ScienceDirect.