

Miikka Pohja

WORKPIECE DETECTION USING MACHINE VISION

Bachelor's Thesis
Faculty of Engineering and Natural Sciences
Examiner: Luis Gonzalez
December 2024

ABSTRACT

Miikka Pohja: Workpiece detection using machine vision

Bachelor's Thesis

Tampere University

Bachelor of Science (Technology), Degree Programme in Engineering Sciences, Automation Engineering

December 2024

This bachelor's thesis explores the use of machine vision to detect workpieces in industrial settings, focusing on identifying the optimal techniques and requirements for this task.

The study begins with an introduction to machine vision, emphasizing its role in automating inspection and identification processes within intelligent factories. Key research questions address what the essential components of machine vision systems are and how they work, what are the requirements for a machine vision system used for detecting workpieces and what are the object detection techniques behind these systems and how do they compare to each other. The research methodology is based on a literature review of academic literature.

The document outlines both classical and modern object detection techniques. Classical methods include edge detection, HOG (Histogram of Oriented Gradients), and SIFT (Scale-Invariant Feature Transform), which rely on lower computational requirements but lack the accuracy and precision. Modern deep learning approaches, such as R-CNN (Region-based Convolutional Neural Network) and YOLO (You Only Look Once), offer improved accuracy when detecting workpieces, with YOLO achieving real-time detection capabilities crucial for industrial applications.

The study evaluates these techniques based on accuracy, precision speed, and ease of implementation. YOLO is identified as the best method on average as it has a great balance of speed and accuracy. Though classical methods remain viable in low-cost or resource-constrained environments. RCNN should not be considered in a factory environment as it can not meet real-time requirements and is costly to develop.

In conclusion, this study advocates for YOLOs usage in real-time industrial applications while acknowledging the limitations of the study, such as the exclusion of newer model versions and the need for consistent testing hardware. Future research suggestions include exploring advanced versions of these models and conducting direct comparisons using standardized datasets.

Keywords: machine vision, object detection, workpiece detection

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Miikka Pohja: Työkappaleiden tunnistaminen käyttäen konenäköä
Kandidaatintutkielma
Tampereen yliopisto
Teknisten tieteiden kandidaattiohjelma, Automaatiotekniikka
Joulukuu 2024

Tässä kandidaatin tutkielmassa tutkitaan konenäön käyttöä työkappaleiden tunnistamiseen teollisessa ympäristössä. Keskittyy optimaalisten tekniikoiden ja vaatimusten määrittämiseen tätä tehtävää varten. Tutkimus alkaa johdannolla konenäköön, jossa korostetaan sen roolia tarkastus- ja tunnistamisprosessien automatisoinnissa älykkäissä tehtaissa. Keskeiset tutkimuskysymykset koskevat konenäköjärjestelmien olennaisia komponentteja, toiminnallisia ja epä-toiminnallisia vaatimuksia sekä erilaisten kappaleiden tunnistustekniikoiden vertailua. Tutkimusmenetelmä perustuu akateemisten lähteiden kirjallisuuskatsaukseen.

Tässä dokumentissa tutkitaan sekä klassisia että nykyaikaisia kohteiden tunnistustekniikoita. Klassisiin menetelmiin kuuluvat edge detection, HOG (Histogram of Oriented Gradients) ja SIFT (Scale-Invariant Feature Transform), jotka tarvitsevat vähemmän laskentatehoa, mutta niiden tarkkuus voi olla riittämätön tietyissä sovelluksissa. Nykyaikaiset syväoppimismenetelmät, kuten R-CNN (Region-based Convolutional Neural Network) ja YOLO (You Only Look Once), tarjoavat parempaa tarkkuutta ja YOLO täyttää reaaliaikaisuusvaatimuksen, joka on ratkaiseva kriteeri teollisissa ympäristössä.

Tutkimuksessa arvioidaan näitä tekniikoita tarkkuuden, täsmällisyyden, laskennallisen nopeuden ja toteutuksen helppouden perusteella. YOLO-menetelmää suositellaan nopeuden ja tarkkuuden tasapainon vuoksi, vaikka klassiset menetelmät ovat edelleen käyttökelpoisia tilanteissa, jossa alhaiset kustannukset ja resurssien puute ovat keskeisiä tekijöitä.

Yhteenvetona tässä tutkimuksessa suositellaan YOLO:n käyttöä reaaliaikaisissa teollisuussovelluksissa, mutta samalla kiinnitetään huomiota tutkimuksen rajoituksiin, kuten uudempien versioiden poissulkeminen. Tulevien tutkimuksien suositellaan tarkastelevan kehittyneempiä versioita ja käyttävän keskenään samanlaisia testauspaketteja suorien vertailujen toteuttamiseksi.

Avainsanat: konenäkö, työkappaleiden tunnistaminen, kohteen tunnistaminen

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

CONTENTS

| | | |
|-------|---------------------------------------------------------------------------------------------------|----|
| 1. | Introduction | 1 |
| 1.1 | Background | 1 |
| 1.2 | Research objectives and scope | 1 |
| 1.3 | Research methodology | 2 |
| 1.4 | Structure of the study | 2 |
| 2. | Machine vision and Object detection | 3 |
| 2.1 | Introduction to Machine Vision | 3 |
| 2.2 | The Concept Of Object Detection | 5 |
| 3. | Functional and Nonfunctional Requirements of Machine Vision Systems in Object Detection | 8 |
| 3.1 | Functional Requirements | 8 |
| 3.2 | Non-functional Requirements | 9 |
| 4. | Classical and Deep Learning-Based Object Detection Techniques | 10 |
| 4.1 | Classical Object Detection Techniques | 10 |
| 4.1.1 | Classical Feature Extraction Techniques | 10 |
| 4.1.2 | Using Support Vector Machine for classification | 12 |
| 4.2 | Deep Learning-Based Techniques | 13 |
| 4.2.1 | Region-based Convolutional Neural Network (R-CNN) | 14 |
| 4.2.2 | You Only Look Once (YOLO) | 14 |
| 5. | Assessment of Object Detection Techniques | 16 |
| 5.1 | Methodology for Assessment | 16 |
| 5.2 | Analysis of Object Detection Techniques | 16 |
| 5.3 | Discussion of Results | 18 |
| 6. | Conclusion | 20 |
| 6.1 | Summary of Findings | 20 |
| 6.2 | Evaluation of study | 20 |
| 6.3 | Ideas for future research | 21 |
| | References | 22 |

ABBREVIATIONS

| | |
|-------|-------------------------------------------|
| CCD | Charge-coupled devices |
| CMOS | Complementary metal-oxide semiconductors |
| HOG | Histogram of oriented gradients |
| IoU | Intersection over union |
| LED | Light-emitting diode |
| mAP | Mean average precision |
| mAR | Mean average recall |
| QC | Quality control |
| R-CNN | Region-based convolutional neural network |
| SIFT | Scale-invariant feature transform |
| YOLO | You only look once |

1. INTRODUCTION

1.1 Background

In today's intelligent and fully automated factories, machine vision, a field of artificial intelligence, has become one of the most important fields of technology due to it enabling machines to understand visual information from the surrounding environments. (Anand & Priya, 2019a, p. 1) Machine vision systems are made to automate inspection, measurement, and identification tasks, which traditionally have relied on human interference. This change not only enhances efficiency and accuracy but also minimizes human error and labor costs. (Smith et al., 2021, p. 1)

Machine vision revolves around the use of sensors, optics, and computing technology to capture and process images of objects. One of the most important applications of machine vision is object detection, which involves identifying and locating specific items within an image. (Merkulova et al., 2019, pp. 79–80) Object detection in machine vision involves several critical components and steps that work together to achieve accurate identification and localization of objects (Kwon & Ready, 2015, pp. 17-26).

1.2 Research objectives and scope

This study is focuses on the elements that do not affect the object detection capabilities of the system. As such the first objective of this study is to determine which elements are necessary to build working a workpiece detection system. The first research question is formulated as follows:

RQ1: What are the key elements of a machine vision system in the context of workpiece detection?

As this study only covers the object detection of workpieces the requirements of this study are going to be very limited. Understanding the requirements of these systems is important for their effective implementation and optimization. This study seeks to delineate both the functional and nonfunctional requirements essential for the successful operation of machine vision systems in detecting workpieces, ensuring they meet performance, reliability, and usability standards. The second research question is formulated as follows:

RQ2: What are the functional and nonfunctional requirements of a machine vision system in workpiece detection?

After finding out the requirements for the workpiece detection system this study moves

forward to study the technologies that are used to locate and identify the objects. However, in the field of machine vision there are many different techniques, and everyone has different pros and cons. The third research question is formulated as follows:

RQ3: What are the most common workpiece detection techniques and how do they compare?

This study has some limitations regarding its scope. Firstly, the research is going to be limited to a factory environment where this technology is the most prevalent. This study focuses solely on the detection of workpieces, specifically aiming at their localization and identification. While object detection can support other tasks, such as quality assurance, this research is limited to identifying and locating objects, excluding quality evaluation aspects.

1.3 Research methodology

This exploratory study is conducted through a literature review. Primary data is gathered from academic sources in the fields of machine vision, computer vision, and industrial automation. Key sources include IEEE Transactions on Industrial Electronics, Computer Vision and Image Understanding, Machine Vision and Applications, Sensors, and Journal of Real-Time Image Processing. These journals provide foundational and current insights into classical and deep learning-based object detection techniques for workpiece detection in industrial settings.

1.4 Structure of the study

The remainder of this literature study starts with chapter two, where the concept of machine vision is explained, and then the key elements of machine vision in terms of object detection are introduced. Then in the third chapter the study focuses on the functional and non-functional requirements of the object detection system. The fourth chapter provides a study on classical and deep learning-based object detection techniques, offering a concise overview of each approach. Afterwards, the fifth chapter takes the requirements from the third chapter and through them analyses different object detection technologies and determines what might be the most ideal for our task. Finally, the sixth chapter provides the conclusion for this study.

2. MACHINE VISION AND OBJECT DETECTION

2.1 Introduction to Machine Vision

Nowadays the term of machine vision and computer vision are sometimes used interchangeably. But they are two different fields of study. Computer vision focuses more on the theoretical aspect of how the machines understand and interpret visual data. It emphasizes developing software algorithms and models. In contrast, machine vision involves the design and integration of both software and hardware needed to create a fully functional machine vision system. (Davies, 2012, p. 13)

Machine vision has many applications across different fields. One of the most common tasks that a machine vision system can perform is quality control (QC). The definition of it is a process that focuses on detecting and monitoring the manufacturing process so that any defects are not let through and quality is kept up to the set standard (Anand & Priya, 2019b, pp. 86-88). This is one of the main ways of keeping up QC in many fields. For example machine vision is used in automobile, food, drug and pharmaceutical industries extensively (Anand & Priya, 2019b, pp. 85-91). But this is not the only way to take advantage of machine vision. It is a very versatile tool that can be used in for example 3D plane reconstruction, 3D pick and place system and pose verification of resistors (Anand & Priya, 2019b, pp. 371-459). This just a few examples but the potential for machine vision is vast.

The concept of computer vision is a technology that attempts to imitate human visions qualities like perception and understanding of what they see (Sonka et al., 2013, p. 1). The primary goal of computer vision is to reconstruct and interpret natural scenes by analysing the content of an image (Peters, 2017). On the other hand of the most important things in machine vision is to implement the task of image acquisition. There are multiple different factors which can affect the quality of the perceived image, and even minor deficiencies can cause problems in the analysis of it. (Davies, 2012, p. 718) This is a reason why one of the first challenges of developing a machine vision system is choosing the right components for it.

A machine vision system consists of several essential components: a camera, a lens, a lighting source, a communication interface, and image processing software (Anand & Priya, 2019a, p. 45). But as this study is only concentrated workpiece detection this chapter is only going to focus on the camera, light source and the lens.

In a machine vision system, the camera is the component which captures the perceived image. One of the most important components inside the camera is the image sensor

which carries out the image acquiring task. Image sensors can be categorized into two main groups: charge-coupled devices (CCD) and complementary metal-oxide semiconductors (CMOS). CCD type have a much better image quality but use more power and are more costly, and on the other hand CMOS sensors make worse quality images but are more affordable. (Anand & Priya, 2019a, pp. 45-47) These are important factors to consider when choosing the sensor.

The image sensor determines the resolution of an image, while the camera's lens defines the focal length (Anand & Priya, 2019a, p. 47). This means the sensor decides how many pixels are inside the image, but the lens determines what the image contains. If the image is of a higher quality there is no need to zoom very close to the image as it can be done digitally when there are more pixels (Anand & Priya, 2019a, p. 55). The opposite is also true if your sensor does not capture great quality there is need for a longer focal length to get the quality needed because the image does not have that many pixels to work with.

The image sensors work by capturing light then converting it to a digital data. Because of this the area where image is captured is often illuminated with a light source to increase the contrast of the image to highlight the characteristics of the analysed object. (Anand & Priya, 2019a, pp. 46-47). As the light is reflected from the observed object, it is important that it is not too bright or too dark (Anand & Priya, 2019a, pp. 67-68). Too little light may cause characteristics like high noise, lousy illumination and bad contrast making it more difficult for the program to identify objects inside the image (Chen & Shah, 2021, p. 1).

There are multiple different light sources each having their strengths and weaknesses. Light-emitting diodes (LED) last a long time and are very efficient, but they are not great for lighting big areas. Lasers are occasionally used, though their application is relatively uncommon. They are very precise at illumination but are not widely used due to them being expensive and dangerous when not used correctly. In the past fluorescent and halogen lights were more popular but nowadays as LEDs are very cost effective and that's why they are most commonly used. (Anand & Priya, 2019a, pp. 68-69)

The captured image is rarely without any distortion or noise in a factory environment. After the image has been acquired it is sent to the image processing unit where the step of image preprocessing can start, which is a task that consist of the manipulation of numerical information inside of the captured image (Sinha, 2012, pp. 3-4). Image processing unit also contains the object detection software which then finds the object inside the image (Anand & Priya, 2019a, p. 88).

Preprocessing is done to increase the quality of the data by highlighting the edges and content inside the image so that the analysis can be done more smoothly (Anand & Priya, 2019a, p. 108). There are multiple different techniques used to enhance the image in preprocessing. The different techniques this section goes over are filtering, scaling and histogram generation. Filtering is an image processing method used to alter or improve

an image.

It determines the value of every individual pixel in the output image by considering the values of the surrounding pixels in the corresponding area of the input image. (Anand & Priya, 2019a, p. 108)

Scaling is a geometric transformation that adjusts the size of an image to a standard dimension, ensuring consistency for subsequent processing tasks. It can reduce image quality, especially when resizing down and then back up. (Anand & Priya, 2019a, p. 114) An image histogram visually represents the distribution of pixel intensities in an image, showing how many pixels correspond to each specific brightness level. The x-axis represents intensity levels (0-255), and the y-axis represents how often they occur. It aids in analysing image frequency and adjusting brightness and contrast. (Anand & Priya, 2019a, p. 116). The general components of a machine vision system are shown in Figure 2.1.

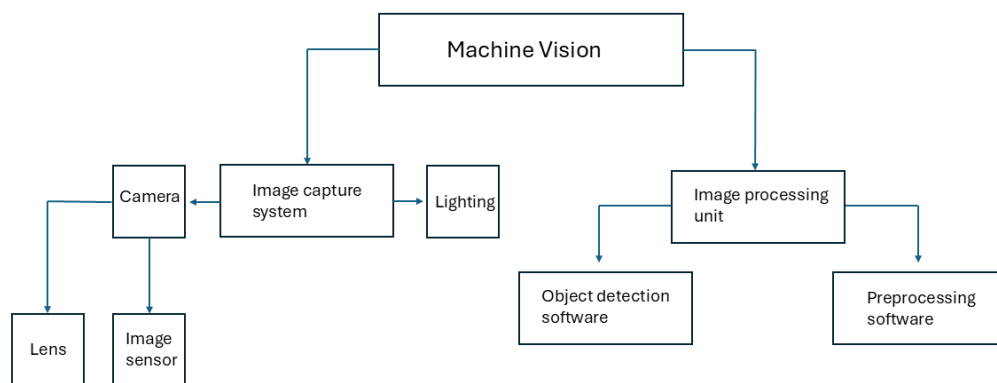


Figure 2.1. The components of a basic machine vision system

2.2 The Concept Of Object Detection

Object detection seeks to determine whether a specific object is present in an image. Occasionally, it is also important to know the object's current appearance and location. (Cyganek, 2013, p. 1). Object detection often follow a structured process to identify and categorize objects in an image. Object detection generally involves three stages: preprocessing, classification of detected objects, and evaluation of results (Merkulova et al., 2019, p. 80).

This workflow is further refined with three technical stages from data acquisition to classification: segmentation, feature extraction, and finally, classification with localization (Amjoud & Amrouch, 2023, pp. 35480-35481). Segmentation is a step where the inputted image is separated into different regions where the model thinks an object is located.

There are multiple different techniques for this such as max-margin object detection and selective search algorithm (Amjoud & Amrouch, 2023, pp. 35479-35481). After regions have been proposed the model can move into feature extraction in which the digital image is transformed into a form where its patterns are more apparent. It simplifies the frame, so it is easier for computers to understand and in turn makes the process more efficient. (Ruano-Ordás, 2024, p. 1) When all the data has been altered into the wanted form the model can begin to analyse the segments and their features. (Amjoud & Amrouch, 2023, p.35481) Then starts the localization step where the objects exact position is located and identified with a bounding box. After that the object is assigned to a specific category in a process called classification. (Zhao et al., 2019, pp. 3212-3213) A generalized illustration of an object detection pipeline is expressed in the figure 2.2.

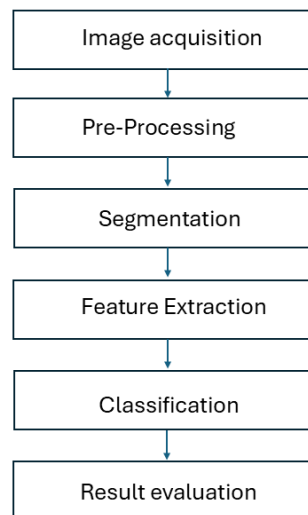


Figure 2.2. *The basic steps of an object detection pipeline*

Bounding boxes are the way in object detection to visualize the location of the object inside the image. Figure 2.3 Presents an example of a bounding box. They are rectangles with a pair of coordinates that help the model to define the scale and location of the object. (Nogueira et al., 2024, pp. 5-6) Confidence score is a threshold used to filter out false positives by ensuring that only bounding boxes with a minimum required score are retained. This score represents the model's certainty that a detected object is present within the bounding box. By setting a confidence score threshold only predictions that meet or exceed this score are let through helping to improve detection accuracy by reducing false detections. (Wenkel et al., 2021, p. 1) Classes explain how detected object are categorized into predefined classes. For example, a class could be a dog, cat or a car (Amjoud & Amrouch, 2023, p. 35479).



Figure 2.3. Example of a bounding box outlining an object

After it has been established that an object is indeed inside the image it must be evaluated based on the given data so the model can be improved. There are multiple different metrics that could be utilized when evaluating the model's performance by comparing it to the data that is known to be the truth. The most common ones are intersection over union (IoU), mean average recall (mAR) and mean average precision (mAP). (Paniego et al., 2022, pp. 3-4) IoU is one of the most important as a detected object can have multiple different bounding boxes. IoU compares these boxes to the information that is known to be real or true. With this the elimination of the boxes that are not accurate enough can start. The IoU can be modified by setting different threshold values. Meanwhile mAP calculates the average precision across different IoU thresholds and object classes, giving a single score that reflects both the model's ability to correctly identify and accurately localize objects. mAR measures the fraction of true positive detections over all actual objects in the image for a given IoU threshold. High average recall (AR) indicates the model successfully identifies most of the objects. (Amjoud & Amrouch, 2023, pp. 35483-35484).

3. FUNCTIONAL AND NONFUNCTIONAL REQUIREMENTS OF MACHINE VISION SYSTEMS IN OBJECT DETECTION

3.1 Functional Requirements

This study aims to identify optimal solutions to the problem of how to locate and classify workpieces efficiently. Clear requirements are needed to enable a proper comparison of the strengths and weaknesses of different techniques. In this chapter requirements are defined as the combination of non-functional and functional requirements. The requirements are prioritized as high, moderate, or low based on their importance within a factory environment. Each priority level is defined to ensure optimal alignment with real-world operational needs.

Functional requirements specify what the system should be able to achieve. They describe the system's behavior in response to specific inputs and outline the expected outputs (Chung et al., 2012, p. 6). They serve as a blueprint for developers and are crucial for ensuring that the system behaves as intended. These functional requirements are expressed in the Figure 3.1.

| HIGH | MODERATE | LOW |
|---------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| <p>Processing speed (The speed of the system is very important in a high-throughput environment) (Wu et al., 2013)</p> | <p>Cost (A good requirement to think as to not go overboard when designing the system) (Batchelor, 2012)</p> | <p>Security (Crucial to ensure no errors from outside factors) (Yi & Jeong, 2022)</p> |
| <p>Reliability (Machine vision system should not have down time) (Ceccarelli & Montecchi, 2023)</p> | | |

Figure 3.1. Functional requirements for a machine vision system

The most important requirements that should be considered are that the machine vision system should be fast, cheap and reliable (Batchelor et al., 2001, p. 3). Machine vision applications generally focus on real-time applications (Batchelor, 2012, p. 1104). Object detection system should have good recognition and localization accuracy as having the wrong object or the location of it will cause errors (Hoiem et al., 2012). For example, a robot's end effector may not find the workpiece if it has the wrong coordinates. Also

classifying the object in a wrong category, can cause issues with uptime (Hoiem et al., 2012). But as the accuracy requirements are highly dependent on the applications it should be considered case by case basis (Batchelor, 2012, pp. 52). In a machine vision system, factors like the user interface (UI) and Data exchange are also important, though they may not be as critical as other core components. In order to reduce human related error some special attention should be used when developing the UI (Järvenpää & Lanz, 2015, pp. 1-2). Also, real-time data collection is important for quick response between entities. (Kumar et al., 2022, pp. 1-2)

3.2 Non-functional Requirements

In the previous subchapter functional requirements gives the system task that it will need to achieve. Non-functional requirements outline the methods and criteria needed to achieve system objectives, defining how a system should operate rather than what it should accomplish. These non-functional requirements are expressed in the figure 3.2.

| HIGH | MODERATE | LOW |
|---------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| <p>Recognition Accuracy (Incorrect object recognition may cause downtime) (Hoiem et al., 2012)</p> | <p>Object Localization (Incorrect object coordinates can cause downtime) (Hoiem et al., 2012)</p> | <p>Data Exchange (Important when integrating with other systems) (Kumar et al., 2022)</p> |
| <p>Real-Time Processing Requirements (Meeting real-time requirements is crucial for most applications) (Batchelor, 2012)</p> | <p>Object Classification (Essential for some applications, but a simple detection system may suffice in others) (Hoiem et al., 2012)</p> | <p>User Interface (Some considerations should be used to reduce human errors) (Järvenpää & Lanz,2015)</p> |

Figure 3.2. Non-functional requirements for a machine vision system

The speed requirement for most system is that it should achieve real-time requirements for most applications (Wu et al., 2013, p. 4096). The definition of reliability in this case means the continuity of correct service. Good machine vision system should not be predicting false positives as that can lead to more interruptions (Ceccarelli & Montecchi, 2023, p. 44957). The cost of implementing a machine vision system varies significantly based on the specific application. But it is always an important consideration to take into. System should only achieve the results that are needed. (Batchelor, 2012, pp. 51-53) When exporting the data to other machines it is important to have a reliable channel where there is no data leakage or other issues that can cause errors (Yi & Jeong, 2022, pp. 4625-4626).

4. CLASSICAL AND DEEP LEARNING-BASED OBJECT DETECTION TECHNIQUES

4.1 Classical Object Detection Techniques

Object detection techniques have significantly evolved since their inception. Understanding these mature solutions is essential to recognize the progression and advancements of these approaches. Even so these classical vision techniques have their uses in different industries.

There are a lot of different "classical techniques" present in machine vision, but in this study, the focus is on techniques such as Canny edge detection, Histogram of Oriented Gradients (HOG), and Scale-Invariant Feature Transform (SIFT). These methods use low-level features meaning things like edges, colors and textures to detect objects. In contrast more high-level methods use object shapes, relationships and categories (Zou, 2019, pp. 3212-3219). Even though they are mature solutions they still have their applications. For example, HOG is still used for human detection (Solunke & Gengaje, 2023, p. 2). As it was previously discussed in the previous chapter, cost is a huge part of developing a solution. It may be sometimes quite overkill to develop these high-level solutions, and these classical methods can solve the problem in a more streamlined manner (O'Mahony et al., 2020, pp. 5-6). These methods rely on complex mathematical foundations and this study provides an overview to convey the essential ideas without getting into the depths of it.

4.1.1 Classical Feature Extraction Techniques

Edge detection refers to a collection of mathematical techniques designed to identify points in a digital image where gradient (a vector that represents the direction and rate of the steepest increase of a function) of the image drastically changes (R et al., 2019, p. 28). As there are multiple different methods this within the concept of edge detection only focuses on the canny edge detection method. Edge detection transforms an image into a simpler structure while trying to preserve its properties which are in this case the edges (R et al., 2019, p. 28). As with almost any technique it is important that the image does not have any noise or abnormalities which could interfere with the algorithm. Canny technique starts by first eliminating these hindering factors by smoothing the image. Then it locates the areas inside the image where there are sudden or fast changes to intensity of the pixel's values. After the selected areas have been established the algorithm looks

for the peak in the pixels and discard the rest so only a clear edge remains. (R et al., 2019, p. 28) Illustration of how Edge detection works is shown in the Figure 4.1

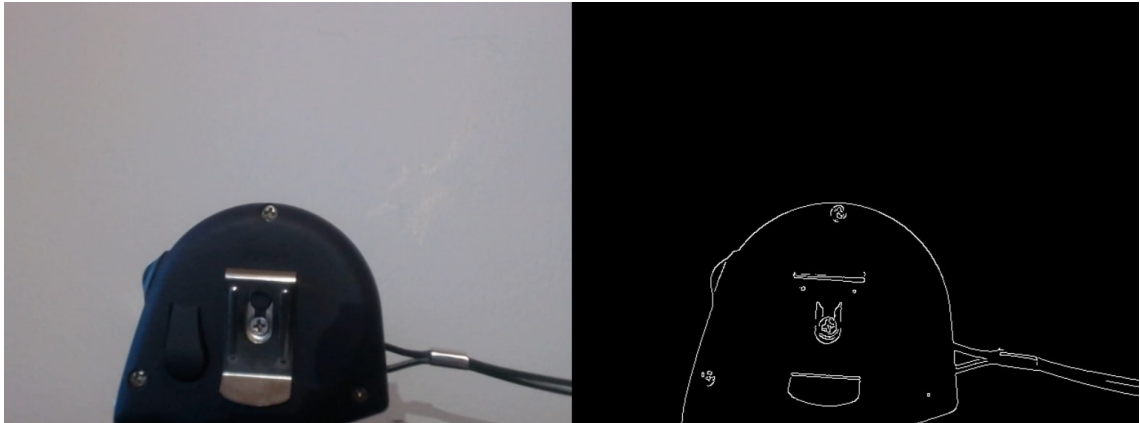


Figure 4.1. A side-by-side visualization showcasing the original clean image alongside the processed image after edge detection made using python cv2 library

Another classical method is HOG, which captures the shape and appearance of objects by analysing patterns in intensity changes across the image, without needing to pinpoint their exact location. HOG features collect information about the direction of changes in pixel brightness around important spots in the image and then combine that information to form a detailed representation of the image. By focusing on the direction of these changes, HOG features are less affected by changes in lighting, since creating a histogram (a summary of the directions) makes the method more flexible. This approach is particularly useful for identifying objects with intricate textures or variable shapes, and it is computationally efficient since the histograms can be calculated quickly. (Shu et al., 2011, p. 217) But HOG may struggle when there are viewpoint or rotation changes inside the image. (Liu et al., 2014, p. 217) Illustration of how HOG works is shown in the Figure 4.2.

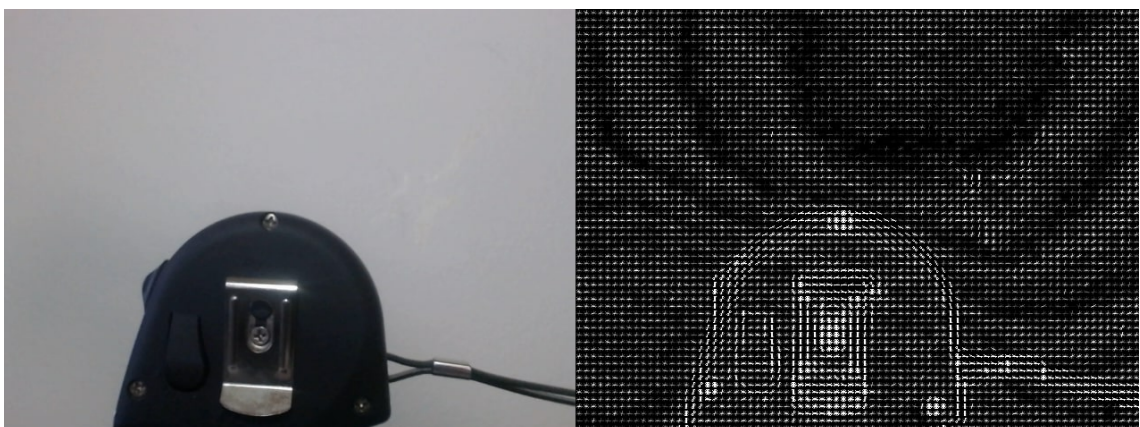


Figure 4.2. A side-by-side visualization showcasing the original clean image alongside the processed image after HOG method made using python CV2 library

Another classical approach is SIFT, which identifies unique descriptors within an image that are resistant to changes in translation, rotation, and scaling (Otero, 2015, p. 370). SIFT works by identifying distinctive points, known as keypoints, in an image at various levels of blur. Each keypoint is essentially a "blob" with a defined location, size, and main orientation. SIFT then creates a small, normalized region around each keypoint, making these points resistant to changes in position, rotation, and scale. Finally, SIFT generates a unique 128-dimensional vector for each keypoint, capturing the gradient patterns in the surrounding area. This vector enables reliable matching of keypoints across different images, making SIFT highly valuable for tasks like object recognition and image stitching. (Otero, 2015, p. 371) SIFT handles rotation and scale variations effectively but struggles with affine transformations (Wu et al., 2013, pp. 124-130). Additionally, due to its computational complexity, SIFT is less suitable for real-time applications unless accelerated by a graphics processing unit (GPU) (Acharya et al., 2018, p. 267). Illustration of how SIFT works is shown in the Figure 4.3.



Figure 4.3. A side-by-side visualization showcasing the original clean image alongside the processed image after SIFT method made using python CV2 library

4.1.2 Using Support Vector Machine for classification

There are a lot of different classification algorithms, but this study is going to focus on Support Vector Machines (SVM) in the classical technique section as it is often combined with SIFT and HOG (Xiao et al., 2020, p. 23732).

SVM is most often used for classification regression analysis and for also novelty detection making it a very versatile algorithm (Awad & Khanna, 2015, p. 11).

SVM is a supervised learning algorithm used for binary classification tasks. It works by transforming the input data into a space with more dimensions and then finding the best hyperplane a flat surface that divides the space to separate the two classes. This hyperplane is selected to increase the margin which is the gap between the hyperplane and the nearest data points from each class. These data points are referred to as support vectors,

and they are critical because they define the hyper plane's position and orientation. By making the gap between the hyperplane and the nearest data points as wide as possible the SVM aims to improve its ability to accurately classify new unseen data. The concept of SVM is expressed in the figure 4.4. The decision made by SVM is non probabilistic assigning new data points to a category based solely on which side of the hyperplane they fall under. (Awad & Khanna, 2015, p. 11)

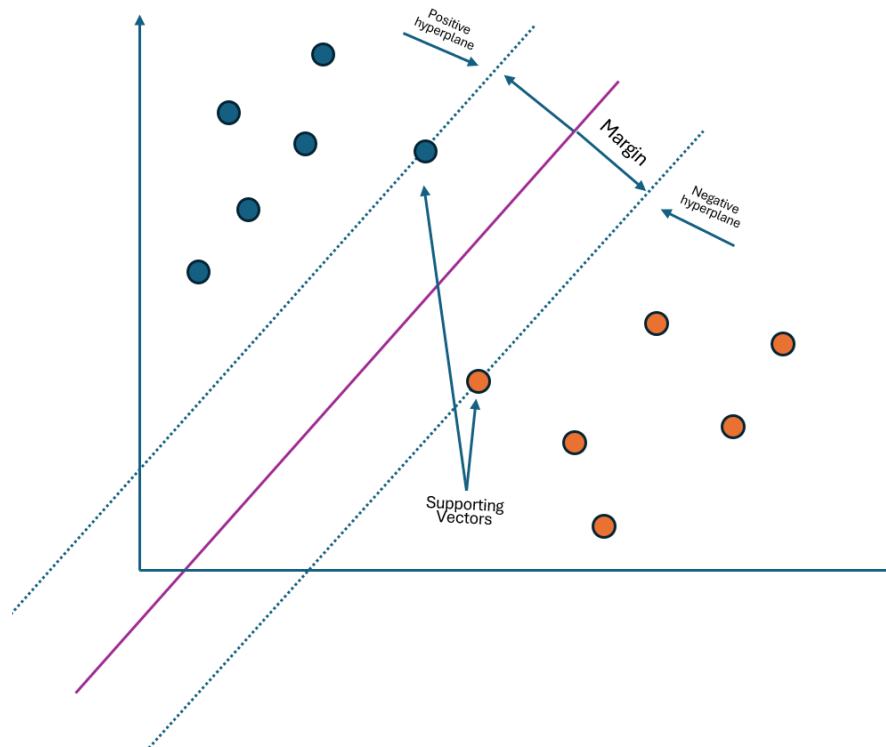


Figure 4.4. Visualization of SVM classification by separating data into two distinct classes with a clear margin and supporting vectors

4.2 Deep Learning-Based Techniques

The main differences between classical methods and deep learning-based methods are that classical methods rely on low-level features and do not require training, whereas deep learning-based methods depend on training and utilize both high-level and low-level features for more comprehensive analysis. (O'Mahony et al., 2020, pp. 1-6) In general the workflows of deep learning methods have notable variance.

There are a lot of different techniques present in the current state of object detection, but this chapter is only going to focus on Region-Based Convolutional Neural Network (R-CNN) and You Only Look Once (YOLO). There are multiple iterations of R-CNN, such as Fast R-CNN and Mask R-CNN, as well as several versions of YOLO, ranging from YOLOV1 to YOLOV8. However, this study focuses exclusively on the original versions of

these techniques.

4.2.1 Region-based Convolutional Neural Network (R-CNN)

Region-based Convolutional Neural Networks or R-CNN is a machine learning model that proposes by regions where an object might be contained, in and then using convolutional neural networks to pick up features from those regions. How it differs from classical approaches is that rather than looking at a massive number of regions, it separates the frame and proposes only a few regions compared to classical approaches (around 2000 regions). This is called a selective search, it starts by segmenting the image into small regions, then merges these regions based on similarities in color, texture, size, and shape, forming potential object boundaries. (Hmidani & Alaoui, 2022, pp. 1-2) Rather than using techniques like HOG and SIFT, RCNN uses deep convolutional neural networks (DCNN) for feature extraction (Xiao et al., 2020, p. 23748). As with deep learning-based techniques R-CNN's feature extraction DCNN needs to be trained for it to be functional (Hmidani & Alaoui, 2022, p. 2). Training a RCNN is a slow and a complicated task (Xiao et al., 2020, p. 23749). Even though there have been a lot of studies that achieve great performance on different networks it is still a challenge to use and train these models (Zhang et al., 2023, p. 90). For classification like more classical methods RCNN also uses SVM (Hmidani & Alaoui, 2022, pp. 2-3). R-CNN can be very accurate method, but it takes along time to make a prediction for a frame, making it very hard to implement a real time solution (Hmidani & Alaoui, 2022, p. 3).

4.2.2 You Only Look Once (YOLO)

YOLO or You Only Look Once is a framework that since its inception in 2015 has been found to be having great balance of speed and accuracy (Terven et al., 2023, p. 1680). It gets its name from the way the technique works. Opposed to other methods YOLO gets all its information within a single pass of the image. It was one of the first real time end-to-end approaches object detection techniques. End-to-end meaning that YOLO unifies the feature extraction and the classification step. It achieves this by dividing the input frame into a grid that contains the probability and confidence of each object. (Terven et al., 2023, p. 1685) When comparing YOLO to R-CNN it reduces to region proposal count (meaning how many regions does the method propose to be analysed) from 2000 to around 300 (Xiao et al., 2020, p. 23753).



Figure 4.5. Visualization of YOLO object detection process

Even though YOLO can be a very accurate method it has problems detecting dense and small objects. With sacrifice to the accuracy, the faster algorithm makes YOLO be able to achieve real time requirements (Xiao et al., 2020, p. 23749).

5. ASSESSMENT OF OBJECT DETECTION TECHNIQUES

5.1 Methodology for Assessment

This study aims to investigate what is the best method for detecting workpieces through machine vision. This chapter is going to go through the previous techniques and assess them through analysis. The requirement priority was expressed in the third chapter and from the chosen variables to be studied are accuracy, precision, processing speed and difficult in adoption. With these meters this study tries to give a clear view on their strengths and weaknesses in a factory setting. The techniques being analysed are as follows: EDGE Detection, HOG + SVM, SIFT + SVM, R-CNN and YOLOV1.

5.2 Analysis of Object Detection Techniques

Firstly, it is important to understand the pros and cons of each in method in the categories given in the previous section. This subsection will provide a brief overview of the categories identified in chapter three, which include: accuracy, precision, processing speed, and ease of adoption and implementation for each object detection method.

The factors on how techniques perform is entirely dependent on the conditions, dataset and hardware used for testing. For this analysis, the average across all conditions will be assumed as the final accuracy value. Next, a table 5.1 is going to be created to rank each method across the specified categories, with scores assigned to reflect performance in each category. Then radar chart 5.1 is formed based on the values given in the 5.1 table.

Starting with edge detection which can be considered having the substandard accuracy even in ideal conditions it is highly sensitive to noise (Yuksel, 2007, p. 83). HOG and SIFT can attain a comparable level of accuracy depending on the condition and dataset used (Ahmed et al., 2022, p. 21755). Then comes the deep learning method YOLO which while better than the classical methods gets beat out by RCNN (Xiao et al., 2020, pp. 23749-23753).

Due to the reason of edge detection methods being highly sensitive to noise it is not going to very precise while in not ideal situations (Yuksel, 2007, pp. 83). While SIFT and HOG may have comparable accuracy, but their precision differs as HOG struggles with viewpoint and rotation changes (Liu et al., 2014, p. 217). Due to that reason SIFT is more robust than HOG. But RCNN is more precise due to average precision (AP) value being lower on YOLO while using the COCO dataset (Patel, 2023, pp. 6-7). Determining

whether YOLO or HOG is more precise is challenging due to a lack of studies supporting either approach definitively, resulting in both being considered equally.

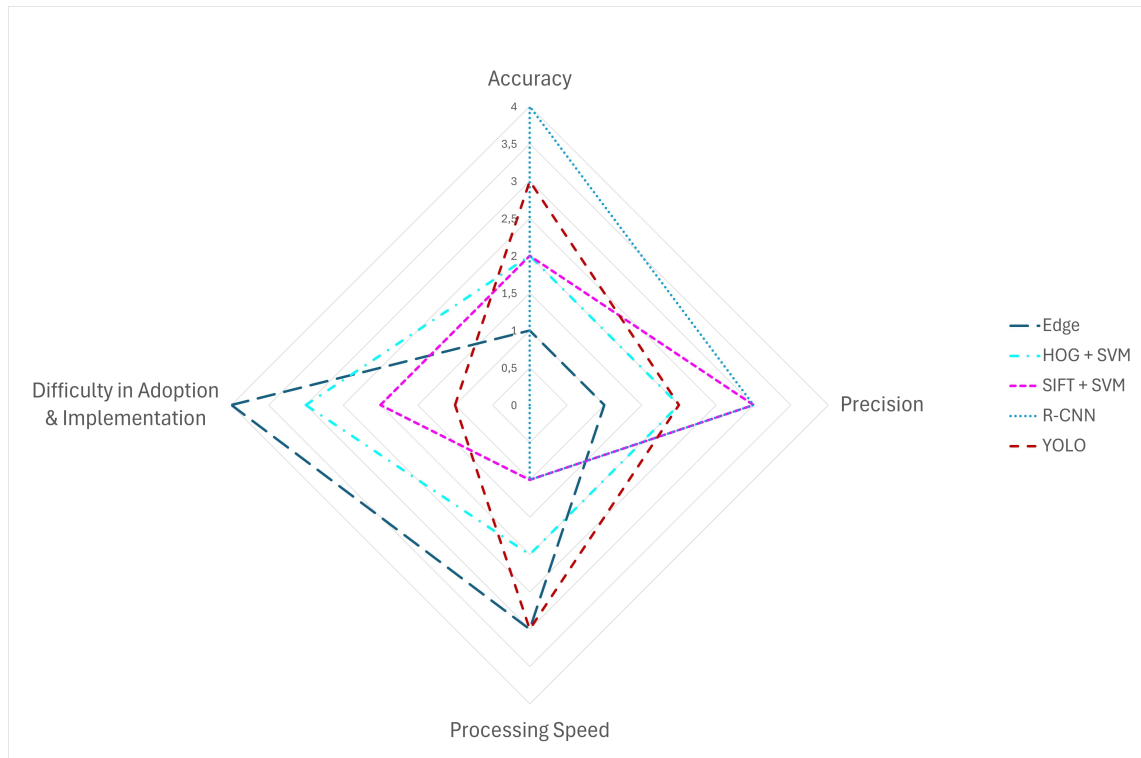
In this study the way processing speed is going to be measured is with frames per second (FPS). It measures how many frames the method can process within one second. Measuring FPS in this category presents challenges, as it heavily depends on the specific hardware used during testing. Most studies reporting FPS likely involve varying hardware configurations, making direct comparisons difficult. When talking about slowest technique RCNN can be considered the less than ideal, as in some studies it achieves about 0.5 FPS (Patel, 2023, p. 7) which can not be considered real time. The method SIFT has been observed to have substandard performance of one to two fps (Ahmed et al., 2022, p. 21755). HOG has been observed to achieve around 25 FPS in certain facial recognition applications (Kortli et al., 2018, p. 6). YOLO can achieve FPS of 45 making it the second-best method (Alif & Hussain, 2024, p. 6). Since there are currently no studies specifying the exact FPS achievable with canny edge detection, it is difficult to provide a precise figure. However, different edge detection methods that are related to canny are known to reach at least 50 FPS (Pham et al., 2024, p. 1).

The way the ease of adoption and implementation can be measured is by looking at the attributes of each technique such as, does it need to be trained, how difficult is implementation part and what kind of equipment are the minimum requirements for processing speed. Both deep learning and SVM based methods need to be trained but SVM requires less data and computational power to achieve comparable results (Billichová et al., 2024, pp. 1-5).

RCNN is the most challenging to implement, as it relies on a deep learning architecture that demands high computational resources and advanced hardware. Like classical methods RCNN also uses SVM for classification (Hmidani & Alaoui, 2022, pp. 2-3). But due to the reason it uses DCNNs for feature extraction which make it hard to train (Xiao et al., 2020, p. 23749). YOLO, though also deep learning-based, is more efficient in terms of computational load (Alif & Hussain, 2024, p. 6). SIFT, while computationally intensive (Acharya et al., 2018, p. 267) it uses SVM, which is relatively easy to train. HOG, combined with SVM, is simpler to run and requires less computational power than SIFT (Kortli et al., 2018, p. 6). Finally edge detection does not use machine learning and is very easy to run, making it easy to implement and adopt.

Table 5.1. Strengths and weaknesses of each method across the different categories

| Method | Accuracy | Precision | Processing Speed | Difficulty in Adoption & Implementation |
|-------------------|---------------|--------------|------------------|-----------------------------------------|
| EDGE | Low (1) | Low (1) | High (3) | Low (4) |
| HOG + SVM | Moderate (2) | Moderate (2) | Moderate (2) | Moderate (3) |
| SIFT + SVM | Moderate (2) | High (3) | Low (1) | Moderate to High (2) |
| RCNN | Very High (4) | High (3) | Low (1) | Very High (0) |
| YOLO | High (3) | Moderate (2) | High (3) | High (1) |

**Figure 5.1.** Radar chart based on the values presented in Table 5.1

5.3 Discussion of Results

All the techniques have their niches in some kind of field. There are several considerations that must be contemplated when choosing the optimal object detection technique. These factors are: data availability for training the model, accuracy requirements, real-time performance, computational resources, and development constraints are essential considerations when choosing the ideal model.

For most applications YOLO is the best when considering average use. It only has a few negatives, and other methods should be only approached when there are very limiting factors. These factors are when detecting very small or dense objects, no training data can be gathered and resources to develop this solution is not worth the payoff. Method should not be judged based on accuracy or other criterion alone (Batchelor, 2012, p.

769). The different categories have varying weights depending on the application. But if the previous reasons are an eliminating criterion for YOLO, other options should be considered. If the problem is the accuracy next method to be consider should be SIFT. If the process needs to be completely real time HOG is the good choice. RCNN should not be considered as it is too hard to use efficiently in the task of workpiece detection. While Edge detection is the simplest form of object detection it is only useful when the task is detecting the boundary of the object and not to identify accurately the object inside the frame.

Even choosing the method can depend on the application being build. As different situation may require different specifications. Going overkill on a task where there is a need to see if an object entered the monitored frame does not need a machine learning algorithm and is a waste of resources. The biggest challenge is finding the balance of cost and efficiency. Acquiring training data to develop a deep learning model needs a lot of care and time. Nowadays there are multiple different coding libraries such as CV2 and PyTorch that help with development step of the model. Hence building a good object detection system is not an arduous task anymore as it was before these libraries existed.

6. CONCLUSION

6.1 Summary of Findings

This thesis explores the use of machine vision for detecting workpieces in industrial environments, aiming to identify optimal techniques and the system components for efficient object detection usage. The study is grounded in three key research questions: determining elements of a machine vision system, understanding the functional and non-functional requirements and evaluating various object detection techniques.

The first RQ addresses the fundamental components of a machine vision system for workpiece detection in industrial contexts: "What are the key elements of a machine vision system for detecting workpieces?" Essential components identified include cameras, lenses, lighting and image processing unit which collectively form the foundation for capturing high-quality images crucial to the detection process.

The second RQ poses the question of what this system should be able to achieve and how it achieves this: "What are the functional and nonfunctional requirements of a machine vision system in workpiece detection". Functional requirements focus on achieving real-time processing, high accuracy, and system reliability to ensure smooth operations in industrial settings. Non-functional requirements highlight the importance of speed, cost-effectiveness, and resilience to environmental noise, which help maintain consistent performance in potentially challenging conditions.

Third RQ addresses the problem of "what are the most common workpiece detection techniques and how do they compare". The study compares classical methods like edge detection, HOG and SIFT with deep learning techniques RCNN and YOLO. Classical methods are cost-effective but can lack the accuracy needed for complex industrial tasks. YOLO is identified as the most efficient and balanced approach, providing real-time capability suitable for industrial applications, despite some limitations with dense, small objects. YOLO's real-time performance is favored over R-CNN's higher precision but slower processing. Classical techniques remain viable in cost-sensitive scenarios where high precision is less critical.

6.2 Evaluation of study

This study provides analysis of machine vision techniques for workpiece detection, assessing both classical and deep learning approaches through criteria such as accuracy, precision, speed, reliability, and ease of implementation. A key strength is the study at-

tention to the difficulty of implementation and the cost of it.

However there are some limitations associated with this study such as not going through the other iterations of deep learning methods. Also, there are some issues with the sources as they do not specify what hardware they are using which will cause some concerns regarding the assessment section. Deep learning-based methods are highly complex, making it challenging to fully explain them within the scope of this bachelor's thesis. It also appears that deep learning-based methods have significantly more recent studies compared to classical methods.

6.3 Ideas for future research

This study only looked at the first versions of R-CNN and YOLO. Since newer versions of these models have been developed, future research should include these updated versions to see if they offer improvements in speed, accuracy, or efficiency. Additionally, there have been no studies directly comparing classical methods to deep learning techniques using the same dataset and hardware. Future research should use identical datasets and hardware for both types of methods to allow a clearer comparison of their strengths and weaknesses.

REFERENCES

- Acharya, K. A., Venkatesh Babu, R., & Vadhiyar, S. S. (2018). A real-time implementation of sift using gpu. *Journal of Real-Time Image Processing*, 14, 267–277.
- Ahmed, K., Gad, M. A., & Aboutabl, A. E. (2022). Performance evaluation of salient object detection techniques. *Multimedia Tools and Applications*, 81(15), 21741–21777.
- Alif, M. A. R., & Hussain, M. (2024). Yolov1 to yolov10: A comprehensive review of yolo variants and their application in the agricultural domain. *arXiv preprint arXiv:2406.10139*.
- Amjoud, A. B., & Amrouch, M. (2023). Object detection using deep learning, cnns and vision transformers: A review. *IEEE Access*, 11, 35479–35516.
- Anand, S., & Priya, L. (2019a). *A guide for machine vision in quality control* (1st ed.). CRC Press LLC.
- Anand, S., & Priya, L. (2019b). *A guide for machine vision in quality control*. Chapman; Hall/CRC.
- Awad, M., & Khanna, R. (2015). *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*. Springer nature.
- Batchelor, B., Waltz, F., Batchelor, B., & Waltz, F. (2001). Machine vision for industrial applications. *Intelligent Machine Vision: Techniques, Implementations and Applications*, 1–29.
- Batchelor, B. G. (2012). *Machine vision handbook*. (No Title).
- Billichová, M., Coan, L. J., Czanner, S., Kováčová, M., Sharifian, F., & Czanner, G. (2024). Comparing the performance of statistical, machine learning, and deep learning algorithms to predict time-to-event: A simulation study for conversion to mild cognitive impairment. *Plos one*, 19(1), e0297190.
- Ceccarelli, A., & Montecchi, L. (2023). Evaluating object (mis) detection from a safety and reliability perspective: Discussion and measures. *IEEE Access*, 11, 44952–44963.
- Chen, W., & Shah, T. (2021). Exploring low-light object detection techniques. *arXiv preprint arXiv:2107.14382*.
- Chung, L., Nixon, B. A., Yu, E., & Mylopoulos, J. (2012). *Non-functional requirements in software engineering* (Vol. 5). Springer Science & Business Media.
- Cyganek, B. (2013). *Object detection and recognition in digital images: Theory and practice*. John Wiley & Sons.
- Davies, E. R. (2012). *Computer and machine vision: Theory, algorithms, practicalities*. Academic Press.
- Hmidani, O., & Alaoui, E. I. (2022). A comprehensive survey of the r-cnn family for object detection. *2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet)*, 1–6.
- Hoiem, D., Chodpathumwan, Y., & Dai, Q. (2012). Diagnosing error in object detectors. *European conference on computer vision*, 340–353.

- Järvenpää, E., & Lanz, M. (2015). Guidelines for designing human-friendly user interfaces for factory floor manufacturing operators. *Advances in Production Management Systems: Innovative Production Management Towards Sustainable Growth: IFIP WG 5.7 International Conference, APMS 2015, Tokyo, Japan, September 7-9, 2015, Proceedings, Part II 0*, 531–538.
- Kortli, Y., Jridi, M., Al Falou, A., & Atri, M. (2018). A comparative study of cfs, lbp, hog, sift, surf, and brief techniques for face recognition. *Proceedings of SPIE - The International Society for Optical Engineering*, 10649, 106490M-106490M-7.
- Kumar, P., Singh, D., & Bhamu, J. (2022). Machine vision in industry 4.0: Applications, challenges and future directions. In *Machine vision for industry 4.0* (pp. 263–284). CRC Press.
- Kwon, K.-S., & Ready, S. (2015). *Practical guide to machine vision software: An introduction with labview*. John Wiley & Sons.
- Liu, K., Skibbe, H., Schmidt, T., Blein, T., Palme, K., Brox, T., & Ronneberger, O. (2014). Rotation-invariant hog descriptors using fourier analysis in polar and spherical coordinates. *International Journal of Computer Vision*, 106, 342–364.
- Merkulova, I. Y., Shavetov, S. V., Borisov, O. I., & Gromov, V. S. (2019). Object detection and tracking basics: Student education. *IFAC-PapersOnLine*, 52(9), 79–84.
- Nogueira, C., Fernandes, L., Fernandes, J. N., & Cardoso, J. S. (2024). Explaining bounding boxes in deep object detectors using post hoc methods for autonomous driving systems. *Sensors*, 24(2), 516.
- O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., & Walsh, J. (2020). Deep learning vs. traditional computer vision. *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1*, 128–144.
- Otero, I. R. (2015). *Anatomy of the sift method* [Doctoral dissertation, École normale supérieure de Cachan-ENS Cachan].
- Paniego, S., Sharma, V., & Cañas, J. M. (2022). Open source assessment of deep learning visual object detection. *Sensors*, 22(12), 4575.
- Patel, H. (2023). A comprehensive study on object detection techniques in unconstrained environments. *arXiv preprint arXiv:2304.05295*.
- Peters, J. F. (2017). Basics leading to machine vision. In *Foundations of computer vision* (pp. 1–85, Vol. 124). Springer International Publishing AG.
- Pham, H. V., Tran, T. G., Le, C. D., Le, A. D., & Vo, H. B. (2024). Benchmarking jetson edge devices with an end-to-end video-based anomaly detection system. *Future of Information and Communication Conference*, 358–374.
- R, R., Saklani, N., & Verma, V. (2019). A review on edge detection technique “canny edge detection”. *International journal of computer applications*, 178(10), 28–30.
- Ruano-Ordás, D. (2024). Machine learning-based feature extraction and selection.

- Shu, C., Ding, X., & Fang, C. (2011). Histogram of the oriented gradient for face recognition. *Tsinghua Science and Technology*, 16(2), 216–224.
- Sinha, P. K. (K. (2012). *Image acquisition and preprocessing for machine vision systems*. SPIE Press.
- Smith, M. L., Smith, L. N., & Hansen, M. F. (2021). The quiet revolution in machine vision—a state-of-the-art survey paper, including historical review, perspectives, and future directions. *Computers in Industry*, 130, 103472.
- Solunke, B. R., & Gengaje, S. R. (2023). A review on traditional and deep learning based object detection methods. *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 1–7.
- Sonka, M., Hlavac, V., & Boyle, R. (2013). *Image processing, analysis and machine vision*. Springer.
- Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolonas. *Machine Learning and Knowledge Extraction*, 5(4), 1680–1716.
- Wenkel, S., Alhazmi, K., Liiv, T., Alrshoud, S., & Simon, M. (2021). Confidence score: The forgotten dimension of object detection performance evaluation. *Sensors*, 21(13), 4350.
- Wu, J., Cui, Z., Sheng, V. S., Zhao, P., Su, D., & Gong, S. (2013). A comparative study of sift and its variants. *Measurement science review*, 13(3), 122–131.
- Xiao, Y., Tian, Z., Yu, J., Zhang, Y., Liu, S., Du, S., & Lan, X. (2020). A review of object detection based on deep learning. *Multimedia Tools and Applications*, 79, 23729–23791.
- Yi, K. J., & Jeong, Y.-S. (2022). Smart factory: Security issues, challenges, and solutions. *Journal of Ambient Intelligence and Humanized Computing*, 1–14.
- Yuksel, M. E. (2007). Edge detection in noisy images by neuro-fuzzy processing. *International journal of electronics and communications*, 61(2), 82–89.
- Zhang, Y., Cai, R., Chen, T., Zhang, G., Zhang, H., Chen, P.-Y., Chang, S., Wang, Z., & Liu, S. (2023). Robust mixture-of-expert training for convolutional neural networks. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 90–101.
- Zhao, Z.-Q., Zheng, P., Xu, S.-t., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11), 3212–3232.
- Zou, X. (2019). A review of object detection techniques. *2019 International conference on smart grid and electrical automation (ICSGEA)*, 251–254.