Coil ID Legibility Assessment of Hot-Rolled Coils Using Image Processing, Scene Text Detection and Deep Learning at Tata Steel DSP

by

Etiosa Samson Raymond

GRADUATION REPORT

Submitted to
Hanze University of Applied Sciences

in partial fulfillment of the requirements for
the degree of

Master Smart Systems Engineering

Confidential document

Groningen
202

ABSTRACT

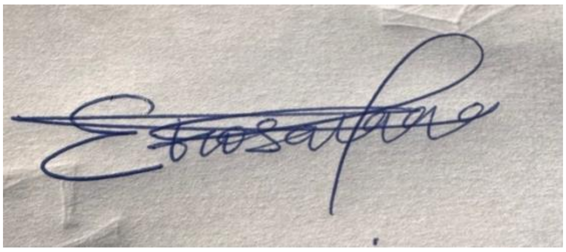
This thesis explores the automation of Coil ID readability assessment at Tata Steel's Direct Sheet Plant (DSP) using machine learning and image processing. A system was developed, combining EAST for localization with YOLO for digit detection and classification to reduce manual effort. While digit detection performed well, achieving 99.4% precision and 99.5% recall, legibility classification faced challenges due to data imbalance. On a balanced dataset, the model performed strongly in both recall and precision, showing its full potential. However, when applied to the highly imbalanced industrial dataset, precision dropped to 44% despite maintaining 73% recall. Data augmentation helped improve recall but also introduced more false positives, making classification less reliable.Additionally, EAST failed in 21 images, all of which had readability issues. Rather than a shortcoming, this confirmed that EAST struggles with poor print quality, making it a useful signal for identifying hard-to-read Coil IDs. Despite these challenges, the system successfully flagged most unreadable Coil IDs and provided insights into factors affecting print clarity. These results highlight the need for better classification strategies. Future work should explore active learning, improved preprocessing, and an authentication step like template matching or SSIM to improve precision in difficult cases.

Keywords: Coil ID, Optical Character Recognition, machine learning, YOLO, EAST, Tata Steel, image processing

DECLARATION

I hereby certify that this report constitutes my own product, that where the language of others was set forth, quotation marks so indicate, and that appropriate credit was given where I have used the language, ideas, expressions or writings of another.

I declare that the report describes original work that has not previously been presented for the award of any other degree of any institution.



Signed,

Etiosa Samson Raymond

ACKNOWLEDGEMENTS

I express my heartfelt gratitude to **Nicolas Pipard**, my supervisor at Tata Steel, and **Mehdi Sedgidi**, my university supervisor, for their invaluable guidance and support throughout this project. Their insights, encouragement, and constructive feedback have been instrumental in shaping the direction and outcomes of my work. This project would not have been possible without their expertise and dedication. I also extend my appreciation to the entire Tata Steel team for their support and collaboration, which greatly enriched my learning experience during this journey.

Table of Contents

List of Figures

List of Tables

Chapter 1

Rationale

1.1 Introduction

Steel manufacturing is central to many industries, supporting sectors such as construction, automotive, and consumer goods [1]. Modern plants rely on precise data and automation to manage production volumes and reduce errors [2]. In large operations, identifying each steel coil accurately is vital for workflow organization and correct deliveries, yet environmental and mechanical factors can compromise the legibility of coil markings which can lead to major safety and sustainability issues.

1.2 Problem Statement

Figure 1 Overview of the automatic printing robot in the Direct Sheet plant

A spray painted print on each coil conveys essential information known as Coil IDs which has seven digits per coil, operators rely on this for inventory control [3].Coil IDs are applied by robotic nozzles, during application as seen in Figure 1, a continuous flow of paint is ejected, ensuring that each character has roughly uniform prints.

Tata Steel's Direct Sheet Plant (DSP), This facility specializes in producing steel sheets for various industries, shows how big this challenge is, producing around 1.4 million tons annually equivalent to about 70,000 coils the DSP must maintain clear, consistent prints[4]. When printing quality declines, which is a very rare but the consequence are significant and it can be very costly.

Environmental conditions affects the legibility of the prints[5]. When temperatures are low, the paint takes longer to dry and does not stick well. Steel coils also come in different sizes and are often held together with straps. Sometimes, the Coil ID gets printed on the strap instead of the coil itself. This can create a problem of legibility, and if the strap comes loose or breaks later during transport, parts of the digits can go missing completely, making the ID unreadable.Additionally, paint nozzle blockages can degrade legibility, when over time, nozzles can get clogged with dried paint, leading smears after prints.

Alternative identification methods like barcodes, QR codes, and RFID tags are impractical in steel plants due to environmental and operational challenges.[5], [6]. Coil IDs are utilized because they are relatively quick to apply, cost effective, and better at resisting surface damage. Their weak point remains the legibility problem which can be subjective based on person to person.

Although poorly readable Coil IDs are rare, the consequences can be severe and costly. When a Coil ID is unclear, tracking becomes difficult, increasing the risk of mix-ups. If a coil is misplaced or misidentified, reproducing it may be necessary, leading to additional production costs and delays, ultimately affecting customer satisfaction. A more serious issue arises if the wrong coil is delivered to a customer. If the material does not match the specification such as a harder steel grade, it could lead to safety concerns in the final product or even damage industrial machinery, resulting in repair costs that could reach hundreds of thousands of euros. Beyond financial losses, such incidents can significantly harm Tata Steel’s reputation, as customers rely on precise material specifications. The rarity of these errors does not reduce their impact, which is precisely why this project was initiated to prevent even the smallest readability issue from turning into a major operational and financial problem.

.

1.3 Scope and constraints

Tasks to be Performed:

- **Develop the Coil ID Legibility Assessment System:**
  Implement a system that evaluates the legibility of Coil IDs using image processing and machine learning techniques.
- **Test and Validate Performance**:
  The system will be tested and validated based on the recall and precision of the readable and unreadable coil IDs.
- **Data Augmentation and Preprocessing:**
  Augment the initial dataset of about approximately 6,000 Coil ID images to simulate various real world conditions. Preprocessing techniques such as thresholding and edge detection will be applied to improve the dataset quality.

Tasks Not Included:

- **Integration with Tata Steel's Systems:**
  The project will not include the integration of the developed system into the company's operational processes, this will be handled by another department, however it will be tested with production data.
- **Improving Coil ID Legibility:**
  The focus was on evaluating Coil ID legibility, not on improving or redesigning the Coil IDs themselves.

Constraints:

- **Limited dataset**: Limited data available, particularly for unreadable Coil IDs.
- **Time constraints:** Four months for completing the research which was limited

Evaluation of End Results:

- The system's performance will be evaluated based on how well it meets the precision and recall targets under real world conditions.
- Multiple model comparison

- The system's ability to use image processing techniques and Region of Interest detection will be evaluated to ensure it enhances image quality and focuses on the areas of the Coil ID that matter most (the digits).

Acceptance Criteria:

- The system must meet the technical requirements and performance metrics.
- The system must provide useful alerts with minimal false positives, ensuring that operators can trust its output.
- A final report must be prepared, summarizing the project, results, and future recommendations.

1.4 Quality expectations and acceptance criteria

This Technical Requirements are listed on a table with justification and validation.

Req ID	**Requirement Description**	**Justification**	**Validation Method**
R1	Detect unreadable Coil IDs with ≥80% recall, ≥70% precision.	Prevents logistical errors	Precision, recall and F1 score
R2	Data Augmentation.	To address dataset imbalance	Compare with non augmented data set.
R3	Localize Coil ID ROIs.	Focus on Coil ID region on the image.	Success and Failure rate.
R4	Recognize Coil IDs.	Confirms presence of valid Coil ID	Precision, recall, F1 score
R5	Process within 5 minutes	To not Disrupt the production flow.	Compare evaluation time to 5 minute limit

Table 1: Technical Requirement

Chapter 2

Situational Analysis & Theoretical Analysis

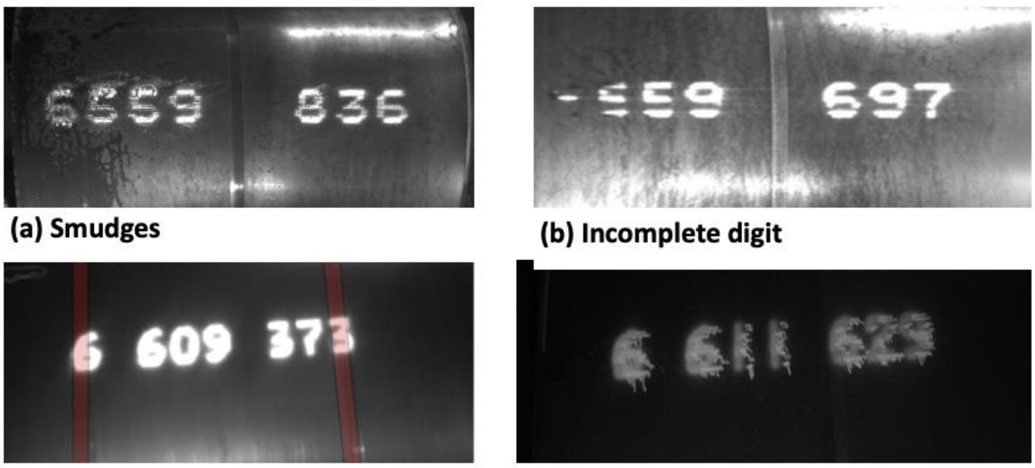
This chapter examines the Definition, challenges, and the theoretical concepts applied to address the legibility of the Coil IDs.

2.1 Coil IDs Legibility Definition at the DSP

After analysis, the R&D department at Tata Steel's DSP concluded that Coil IDs might be considered unreadable if any digit is incomplete or smudged. This was defined after reviewing several defective prints that were not legible. When the Coil ID is located on top of a strap with the digits still visible, it should be classified as moderately readable, as the key information remains accessible but will indicate future readable issues if the strap comes off. Coil IDs should only be regarded as fully readable when the digits are clear and pristine, ensuring accuracy in identification. This definition is specific to the DSP.

Challenges at DSP

Coil ID legibility is usually compromised by three key factors at the DSP:



(a) Smudges **(b) Incomplete digit**

c) Strap Obstruction **(d) Printing Robot Nozzle problem**

Figure 2: Some Common Print Legibility Defects

Paint Application Issues:

Robotic spray painting is prone to nozzle blockages, distorting characters and causing smearing as seen in Figure 2 (d) and (b).

Environmental Issues:

Low temperatures may cause the Coil IDs failure to dry properly and thereby leading to smudges as seen in Figure 2 (a).

Coil Specific Obstructions:

On the coil, straps are usually attached, in the process step just before the robot applies the coil id, as seen in Figure 2 (c) and depending on the size, one or two straps can be attached. This can affect the print quality when it's printed on the strap. Furthermore, if for any reason the strap gets lose or break that digit printed on the strap can become completely unreadable

2.2 Theoretical Analysis

This chapter explains the theory behind the project by introducing the key concepts needed to understand the models, methods, and algorithms used throughout the project.

2.2.1 Machine learning

Artificial intelligence is a subfield that uses data to learn without being programmed explicitly[7]. Supervised, unsupervised, and semi supervised learning algorithms can generally be classified into three categories [8]. Supervised learning: In this type of learning, a model is trained on labelled data, where a given input is accompanied by the output value, and the task is to learn the relationship between the input and output. The model trains on the underlying characteristics of the input images and their corresponding values while training, so it can predict on new or unseen data[9]. On the contrary, unsupervised machine learning techniques are used to discover patterns and structures in observations that are not labelled [6]. Semi supervised uses both labelled and unlabeled data to train the model, and it is common when labelled data is lacking or insufficient[10].

2.3.2 Artificial neural network

Artificial neural networks (ANNs) are computational systems modeled after the functioning of biological brains. This area of study falls within the broader domain of machine learning. ANNs are made up of interconnected units known as neurons, which can be trained to learn and execute a diverse range of tasks. The foundational idea of ANNs was introduced by Warren McCulloch and Walter Pitts in 1943[11]. Interest in this topic has fluctuated over the years, but there has been a significant surge in both application and investigation in the past decade. This increase was largely attributed to the achievements of deep learning algorithms across various applications, including language translation, speech recognition, and object detection. Currently, the predominant type of neural networks in use are feed forward networks, where input is fed into the system and processed through the network to generate an output. Typically, neurons are organized into layers as seen in Figure 3, and the term deep learning refers to neural networks with multiple layers. These deep learning architectures frequently contain millions of nodes and trainable parameters.

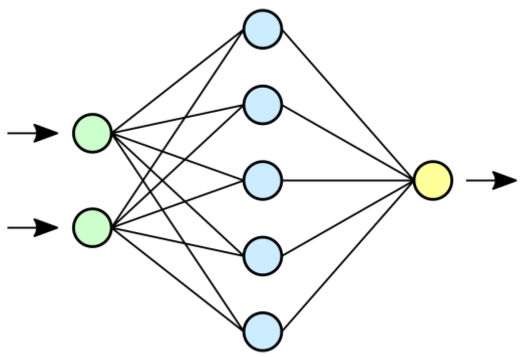


Figure 3 An ANN with one layer

2.3.3 Convolutional neural networks

There are various architectures available for building artificial neural networks (ANNs). A prominent architecture commonly employed in image analysis tasks is the convolutional network[12] In this setup, rather than assigning a unique weight to each connection, the weights are organized within a kernel of a predetermined size, which is then applied across the layer. An illustration of this can be seen in Figure 4. This architectural approach offers the benefit of aggregating information from adjacent nodes, enabling the detection of local patterns within the image[12].

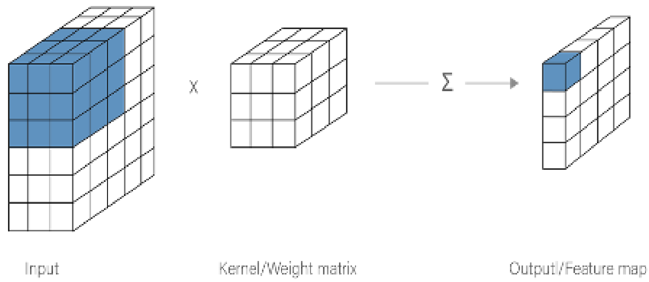


Figure 4 A convolutional operation

The selected input which are the shaded nodes values are multiplied point-by-point with the convolution kernel, and adding these gives us the highlighted output value. Convolutional layers are mainly described by the height and width measurements of their kernels.. Additionally, the kernel includes a third dimension, which corresponds to the number of channels within the network. The spatial dimensions of the output layer, given that both the input layer and kernel possess quadratic spatial dimensions, can be determined using the formula:

$$o = \frac{i-k+2p}{s} + 1 \tag{1.0}$$

where *i* represents the size of the input layer,

k denotes the kernel size,

p indicates the quantity of zero padding surrounding,

i and *s* is the step size.

2.3.4 PVANet

PVANet, as introduced in[13], was developed as a lightweight deep neural network aimed at lowering computational expenses when compared to the commonly utilized ResNets[13]. PVANet reduces complexity with 1x1 convolutions while maintaining performance using C.ReLU and inception modules. C.ReLU halves channels by pairing outputs with their negations, improving early stage efficiency. Inception modules enhance deeper layers, outperforming standard ReLU networks. [14]. So, half of the convolution channels are used, and their negative values are combined with the original output. Using these C.ReLU layers early in the network works better than regular ReLU layers that other networks use. [14], alongside a reduction in computational demands. Inception modules facilitate parallel convolutions using various kernel sizes, enabling the network to capture spatial features across different scales within each layer throughout its architecture [15], [16]. An illustration of the inception modules utilized in PVANet can be found in Figure 5.

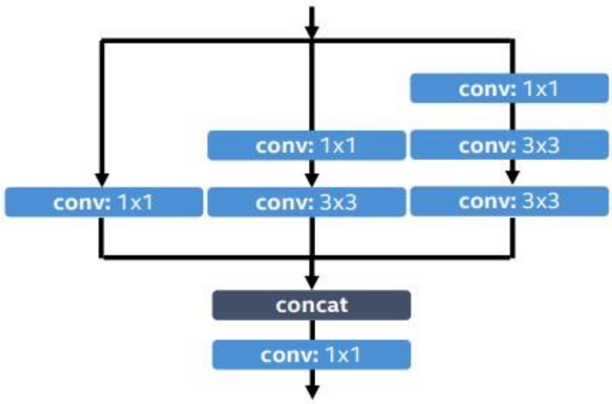


Figure 5 PVANet uses inception blocks with parallel paths operating at the same network depth, though each path applies a different kernel size.

2.4 EAST

The Efficient and Accurate Scene Text Detector (EAST) model, presented by Zhou et al [17], establishes a simple yet efficient pipeline for detecting text in natural scene images. EAST directly outputs text regions from full images using a fully convolutional network (FCN), producing two types of bounding box geometries rotated rectangles (RBOX) and quadrangles.

2.4.1 Network Architecture

Backbone (Feature Extraction Stem), EAST employs PVANet as its backbone. PVANet as defined in section 2.3.4 was chosen for its lightweight design and relatively few trainable parameters while still delivering competitive performance. The backbone network processes the image through multiple layers, creating feature maps at different scales.

Feature Merging Branch, the multi scale feature maps extracted from convolution blocks 2-5 of PVANet are then merged in a hierarchical manner. Due to the difference in spatial dimensions with consecutive feature maps differing by a factor of 2, the smaller maps are upscaled to match the dimensions of the larger ones. These upscaled maps are concatenated and subsequently processed by a 1×1 convolution layer to reduce the number of channels, followed by a 3×3 convolution layer that maintains spatial resolution. This process is repeated iteratively until a single, unified feature map is obtained, which is then fed into an additional 3×3 convolution layer serving as the final merged feature map.

Output Layer, the final output of the network consists of six channels. One channel produces a pixel level score map indicating the likelihood of text presence. Four channels encode the distances from each pixel within a detected text region to the four edges top, right, bottom, and left of the bounding box, and the remaining channel encodes the rotation angle of the bounding box. Each output channel is generated via a 1×1 convolution applied to the final merged feature map.

2.4.2 Loss Function

The loss function in EAST is designed to optimize both the score map and the geometry predictions. It is formulated as:

$$L = L_s + \lambda_g L_g \tag{1.1}$$

where:

- L_s is the loss for the score map,

- L_g is the loss for the geometry,
- λ_g is a hyperparameter balancing the two terms.

For the score map loss L_s, [17]employs class balanced cross entropy. This approach addresses the imbalance between the abundant background pixels and the relatively few text pixels by weighting the loss contributions from positive and negative pixels accordingly.

The geometry loss L_g comprises two components:

1. **Bounding Box Overlap Loss (*L_{AABB}*)**

This term uses the Intersection over Union (IoU) loss, computed as the negative logarithm of the ratio of the intersection area to the union area between the predicted and ground truth bounding boxes. Here, the bounding boxes are assumed to be axis aligned.

2. **Rotation Angle Loss (L_ϑ)**

Next, the loss of rotation angle is computed as

$$L_\vartheta(\vartheta, \hat{\vartheta}^*) = 1 - \cos(\hat{\vartheta} - \vartheta^*). \tag{1.2}$$

where $\hat{\theta}$ is the prediction to the rotation angle and $\theta*$ represents the ground truth. Lastly, the total geometric loss combines the AABB loss and angle loss with assigned weights.[17], given by

$$L_g = L_{AABB} + \lambda_\vartheta L_\vartheta. \tag{1.3}$$

Where λ_ϑ serving as the weight for the angle loss.

2.4.3 Label Generation for EAST

Ground truth labels for the output layers are generated through a two step process involving the score map and the geometry maps.

Score Map Generation:

The score map is generated by shrinking the annotated bounding boxes. Each bounding box is shrunk by moving its four corners inward along the outgoing edges by a factor ***r*** set to 0.3 of the length of the shortest edge. All pixels inside the shrunk box are labeled with 1 (indicating text), while pixels outside are labeled with 0.

Geometry Map Generation:

For pixels within the shrunk bounding box, the ground truth geometry is computed by calculating the distances from the pixel to the four boundaries of the original unshrunk bounding box. Additionally, an angle map is generated by assigning each pixel within the shrunk box the rotation angle in radians of the lower edge relative to the horizontal axis; the angle is constrained between −45 and 45 degrees.

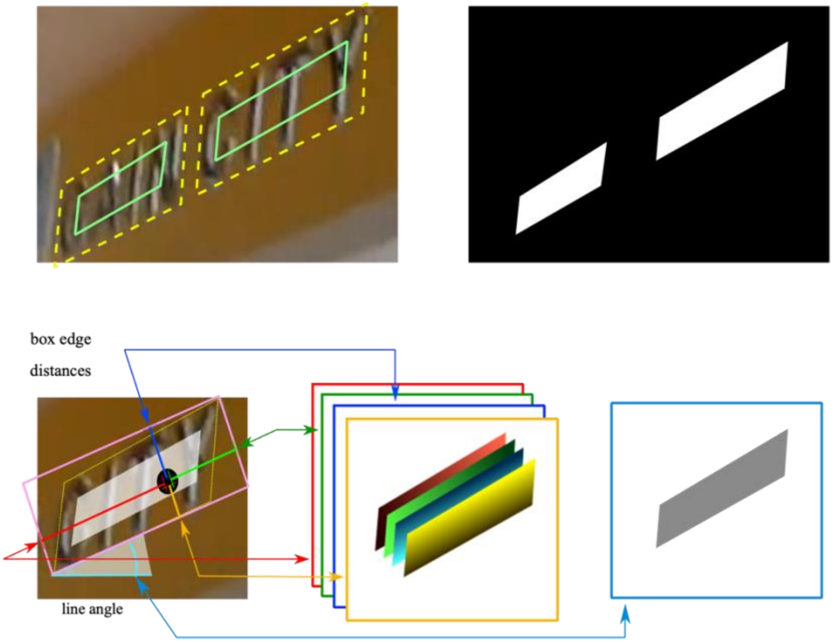


Figure 6 EAST label generation process [17]

Figure 6 illustrates the label generation process, Label generating process for EAST:(a) A Yellow dashed line and a solid green reduced box. (b) Score map in relation to the text; (c) The map showing the geometry for the rotated rectangle; (d) our channels displaying the distances from each pixel to the boundaries of the rectangle; (e) angle of rotation of the rectangle[17].

2.4.4 Post processing with NMS

After the network produces its predictions, post processing is applied to derive the final text bounding boxes. Initially, the score map is thresholded using a threshold of 0.8 to remove low confidence predictions. Given the dense nature of the predictions, which can yield thousands of candidate geometries, the paper introduces a locality aware non maximum suppression

(LANMS) algorithm. This method assumes that geometries from nearby pixels are defect correlated and merges overlapping geometries by averaging their coordinates, weighted by their scores, rather than simply suppressing them. Finally, standard non maximum suppression is applied to the merged geometries to yield the final text bounding boxes.

2.5 YOLO

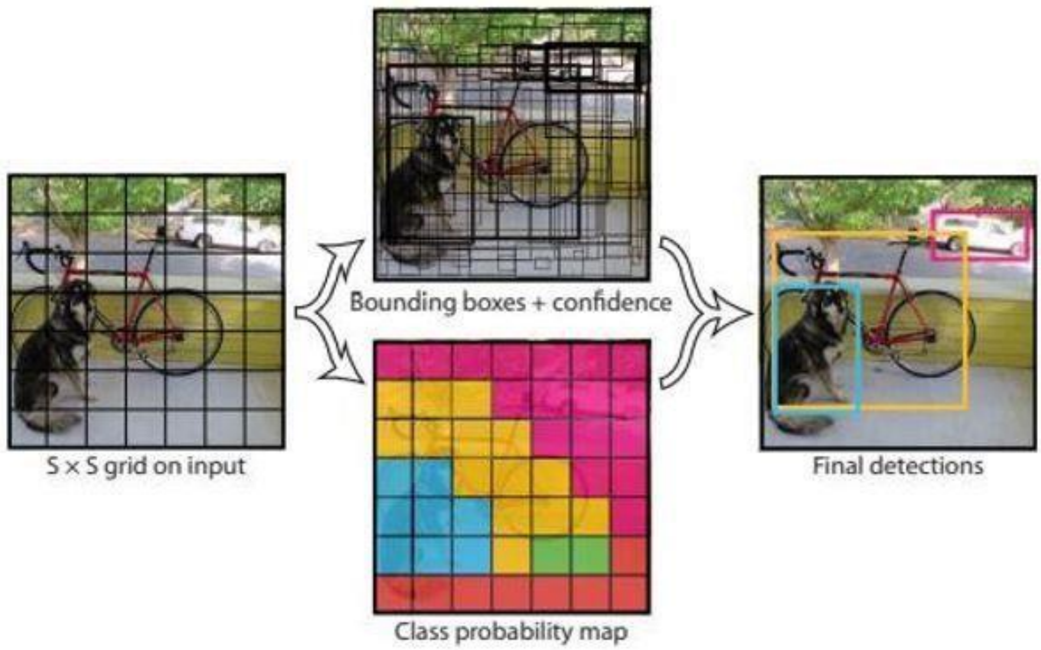


Figure 7 Pipeline of the Yolo algorithm[18]

The **"You Only Look Once" (YOLO)** algorithm is a deep learning model that belongs to the category of single stage detectors utilizing Convolutional Neural Networks (CNN) for object detection. YOLO is characterized by its rapid and precise architecture, making it particularly effective for applications requiring real time object detection[18].

This algorithm employs a grid-based approach akin to CNNs to facilitate swift processing within the network. It works by partitioning input images into grid cells of size S x S, where each cell is tasked with predicting bounding boxes alongside object classifications [18]. The processing occurs through a single forward pass, meaning the input data is run through the neural network from start to finish, resulting in a prediction. This is the rationale behind the name "You Only Look Once."

Each bounding box is defined by four parameters: (x, y, width, and height), in addition to a confidence score that reflects the likelihood of the presence of an object within the box. The network also estimates class probabilities. The final layer's output nodes dictate the number of classes detectable by YOLO, with each node corresponding to a unique class and outputting the likelihood that an object in the input image falls into that category.It is common to encounter multiple bounding boxes for the same object; to address this, YOLO selects the box with the highest probability score among the detections, employing the Intersection over Union (IoU) metric[19]. Consequently, the algorithm can discard unnecessary bounding boxes that do not align with the object's characteristics or dimensions, as illustrated in Figure 7 [18]. Each accountable grid represents generalized value vectors. In equation 1.4, Pc denotes the probability of a class, with Bx and By corresponding to the box's coordinates in relation to the grid. Bw and Bh signify the dimensions of the box.

$$C_{1,1} = (P_c, B_X, B_y, B_w, B_h, C_1, C_2, C_3, C_4) \tag{1.4}$$

Building on this foundation, YOLOv11 introduces further advancements in speed and accuracy which will be explained on the next section.

.

2.6 Additional Key Concepts

2.6.1 Bounding Box Co-ordinates:

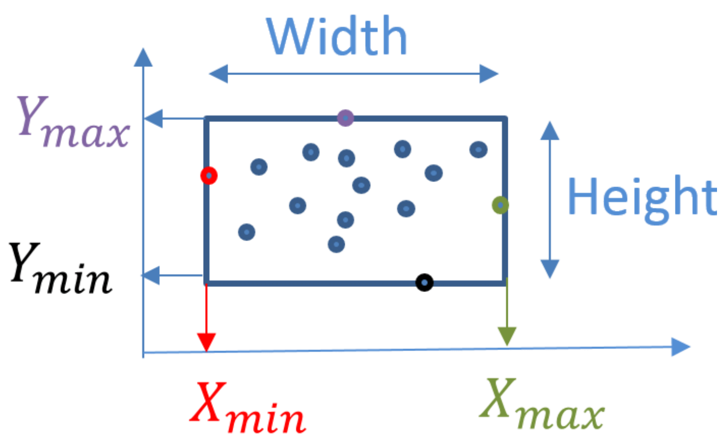


Figure 8: Bounding box with Ymax, Ymin, Xmin and Xmax co-ordinates.

Firstly, Bounding boxes are a crucial component of object detection algorithms, which are used to identify and classify objects within an image or video. A fixed bounding box is a rectangle that surrounds an object of interest in the image and is typically represented by a set of coordinates as seen in Figure 8 [20] that define the box's position and size. A key limitation of fixed Bounding box is that it is not dynamic to the object change of position[21].

2.6.2 Perspective Transformation

Perspective transformation is a crucial technique in image preprocessing for correcting geometric distortions caused by the angle of capture or camera perspective. By remapping the coordinates of an image to a target plane, this method adjusts the spatial arrangement of pixels to simulate a frontal or orthogonal view of the object, as illustrated in Figure 9. This transformation is particularly vital in computer vision applications where accurate geometric representation is essential, such as mitigating distortions like foreshortening, pose variations, or depth-related scaling that arise from non-ideal camera angles.[22]

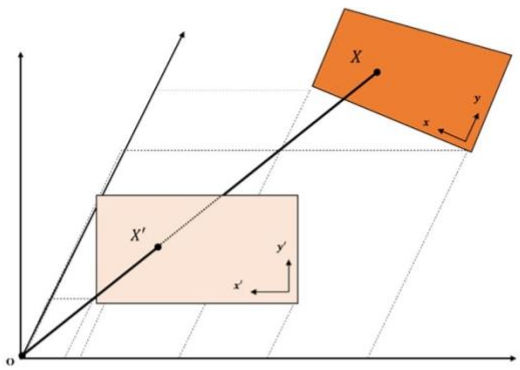


Figure 9: Coordinate system transformation showing translation and rotation.

2.6.3 Augmentation

Data augmentation is a set of techniques designed to enhance the size and diversity of datasets by generating additional data samples through various transformations and processes applied to the original data. This approach is crucial for improving the generalization capabilities of AI models, particularly in scenarios involving scarce or imbalanced datasets[23].Some data augmentation techniques include rotation, zooming, translation, brightness/contrast adjustment, noise injection, and scaling.

2.7 Related Work

After extensive search, literature and work done in this specific field seemed scarce, where only a paper by Caldeira et al [24]that specifically appeared to address similar printing issues was found and according to[5] accurately recognizing and monitoring steel coils has been a major concern in manufacturing industries. Harsh conditions such as dust, heat and mechanical wear have made traditional options including barcodes and RFID tags unpopular even though they have been commonly used.

The author [24] presented an efficient and effective OCR system for steel coil identification to meet the requirements of industrial conditions the proposed method applies advanced image processing methods together with a CNN to perform character recognition. The system starts by extracting the Region of Interest (ROI) from the coil using laser markers to capture the curved surface. This step was crucial because the printed IDs are often distorted due to the camera's fixed position and the coil's shape. The system applies a cascaded filtering framework to tackle lighting variations, beginning with the Top Hat filter to remove glare while enhancing contrast. A filter serves as the first line of defense but becomes active only when the initial filter fails to deliver results which are then enhanced through application of secondary filters that include Homomorphic and UMCLAHE for accommodating varying lighting situations.

Their method contains a key innovation with the geometric transformation step which performs the task of correcting the elliptical distortion that the camera's angle creates. Through a vertical scaling operation and addition of padding the system repositions the characters so that they become horizontal making their segmentation and recognition easier. The final segmentation was carried out through connected component analysis where individual characters are isolated based on their size and aspect ratio. The characters are then classified with the help of a CNN which has been trained on 17 classes, including digits, letters and a noise category to account for non text elements.

To improve the system's strength during post processing stage, the system compares the recognized ID with a database. The system performs retry with different filters or tags the image for human assessment if discrepancies occur. The system can thus tackle many printing errors through its multiple levels yet keep high accuracy. When evaluating the system on 20,000 test images it attained accuracy above 98% indicating its suitability for industrial application.

With these advancements, however, the work by [24] was standing out for its tailored approach to steel coil identification. The specific challenges of steel manufacturing environments are addressed by their system which combines cascaded filtering geometric correction and CNN based classification. The system can process images in under 0.36 seconds which makes it suitable for real time applications, reducing production delays and errors. The example provided by [24]demonstrates how both traditional image processing methods and current deep learning technologies can be used to develop powerful solutions which address actual world challenges.

2.7.1 Proposed pipeline improvements:

To address these challenges, the following enhancements have been integrated into the Pipeline.

1. **Coil ID Localization:**

For the initial text localization, the EAST (Efficient and Accurate Scene Text Detector) model is used. The ability of the model to detect text in arbitrary orientations makes up for the geometric distortions and complex scenarios. As a fallback when the EAST model provides unreliable results because of extreme occlusion or degradation, a fixed coordinate points is used to ensure region of interest extraction

2. **Recognition and Segmentation:**

As an alternative to [5] approach, a unified deep learning framework based on YOLO (You Only Look Once) is employed for both the detection and extraction of digits, rather than using traditional segmentation followed by CNN classification. YOLO directly localizes and extracts the digits and their bounding boxes, which helps in avoiding the error that may be introduced due to segmentation as used in previous works.

3. **Print Legibility Assessment:**

In the detection stage for print quality assessment, a three-class classification module of YOLO will be utilized.. A legibility logic then aggregates these classifications based on defect patterns observed at Tata Steel DSP. Building on the solid foundation created by Caldeira et al. [4], the proposed method reduces manual segmentation needs and may decrease complexities of reproductivity.

2.7.2 Comparison of Alternative Methods

Several methods were evaluated for detecting and assessing Coil ID legibility, considering accuracy, computational efficiency, and adaptability to Tata Steel's industrial setting. The chosen approach builds on existing research while addressing key limitations.

Method	Alternative	Justification
YOLO for Digit Detection	CNN-based classification	YOLO detects and localizes digits simultaneously, eliminating segmentation errors and enabling faster inference.
EAST ROI Localization	Fixed bounding boxes or manual cropping	For dynamic ROI localiztion to avoid hard coding fix bounding boxes coordinates.
Perspective Correction	Raw image	Correct angled or skewed captured images.
Data Augmentation	Non-augmented datasets	Enhances model adaptability to lighting and positional changes, reducing failure cases.
YOLO-Based Readability Assessment	Custom OCR + Database matching	Direct classification of legibility during detection, eliminating post-processing complexity.

Table 2: Comparison of Alternative Methods

Chapter 3

Responsible Engineering

3.1 Contributing to Sustainable Innovation

The Coil ID legibility assessment system when developed and implemented for (DSP) increases efficiency and strengthens sustainability across environmental, social and economic levels. Thus, the system assists in reducing waste, saving resources and boosting workforce productivity by taking over the typically manual process. Furthermore, its design principles are in line with internationally recognized standards, thus sustainability is an integral part of the solution.

3.1.1 Environmental Impact

Resource Efficiency:

Automating the legibility assessment process reduces the reliance on manual inspections, thereby lowering energy consumption and minimizing repetitive human tasks. The system decreases energy usage by reducing the need for prolonged machine operation during manual inspections and eliminating human dependent inefficiencies This approach aligns with the energy management principles of ISO 50001, which encourages organizations to improve energy performance through systematic monitoring and optimization[25].

Material Conservation:

By enhancing the accuracy of Coil ID detection, the system reduces the risk of misidentified coils preventing unnecessary scrapping, reprocessing, or misrouted shipments. This outcome supports the environmental objectives of ISO 14001[26], which emphasizes reducing waste and optimizing material use in production processes.

3.1.2 Societal Impact

Quality Assurance:

Enhanced Coil ID assessment improves the tracking and identification of coils, leading to better product quality and higher customer trust. This reliability supports Tata Steel's reputation and is consistent with the social responsibility guidance found in ISO 26000[27], which encourages ethical business practices and quality assurance.

3.1.3 Economic Impact

Cost Reduction:

Minimizing errors in Coil ID legibility translates into fewer production delays, reduced downtime, and lower costs related to manual rework. Improved inventory management further prevents overproduction and reduces storage expenses, contributing to overall cost efficiency.

3.2 Opportunities for Sustainable Innovation

Integration of Energy Efficient Hardware:

Implementing energy efficient cameras, sensors, and processors can significantly lower the system's energy footprint. Exploring low power machine learning models further optimizes energy usage in line with the recommendations of ISO 50001.

Data Driven Process Optimization:

The system's ability to collect and analyze data on Coil ID legibility trends allows Tata Steel to pinpoint recurring issues (such as nozzle blockages or inconsistent paint adhesion). This data driven approach facilitates continuous process improvement and waste reduction, echoing the environmental management goals of ISO 14001.

Predictive Maintenance:

Leveraging predictive analytics to forecast equipment failures e.g., nozzle blockages can enable proactive maintenance. By reducing unplanned downtime and extending equipment lifespan, this strategy not only lowers operational costs but also aligns with sustainable maintenance practices.

3.3 Ethical and Responsible Engineering Considerations

Developing the Coil ID Legibility Assessment System requires careful attention to ethical aspects such as AI transparency, reliability, and bias prevention.

AI Transparency

Transparency means making the system's operations and decision making clear. This involves thorough documentation of algorithms and data sources, ensuring that the processes are understandable to all stakeholders. Implementing explainable AI (XAI) techniques can further clarify how the system reaches its conclusions, enhancing trust and accountability[28].

Reliability

Reliability ensures the system performs consistently under various conditions. This is achieved through extensive testing across different scenarios to confirm proper functionality. Regular maintenance and updates are also essential to address any emerging issues and adapt to changes in the production environment[29].

Bias Mitigation

To prevent the system from adopting and perpetuating biases present in training data, it's crucial to use diverse and representative datasets during development. Continuous monitoring for biased behavior and implementing corrective measures when biases are detected are also critical steps. Involving a diverse team in the development process can further help in identifying and addressing potential biases[30].

Accountability

Establishing clear accountability mechanisms ensures that there is a defined process for addressing any issues arising from the system's deployment. This includes setting up protocols for monitoring performance, reporting anomalies, and implementing corrective actions when necessary. Assigning responsibility to specific individuals or teams helps maintain the system's integrity and builds trust among stakeholders[31].

By proactively addressing these ethical considerations, the Coil ID Legibility Assessment System can be developed and deployed in a manner that is transparent, reliable, and fair, aligning with Tata Steel's commitment to responsible engineering practices.

Chapter 4

Conceptual Model

4.1 Research Question

How can image processing and machine learning be used for Coil ID localization, recognition, and extraction to assess print legibility with at least 80% recall and 70% precision, while accounting for environmental variations and completing evaluation within 5 minutes?

4.1.1 Sub Research Questions

Proposed Sub Questions:

1. What technique is effective for localizing Regions of Interest (ROIs) in Coil ID images?
2. Which image recognition are most effective for processing Coil ID images under varying environmental conditions?
3. How can machine learning algorithms be applied to assess Coil ID print legibility, and which models offer the best performance in terms of recall and precision?
4. What are the common environmental factors affecting Coil ID print legibility, and how can the evaluation system be adapted to tackle it?
5. How can the evaluation system be optimized to deliver accurate legibility assessments within a 5 minute timeframe?

4.2 Conceptual Model

The conceptual model consists of four main blocks as seen in Figure 10: Data Acquisition, Data Processing, ROI Localization and Extraction, and Classification and Evaluation.

The **Data Acquisition Block** consists of two sub blocks: camera setup and multi exposure application. The camera setup sub block is responsible for configuring the image data collection. The multi exposure sub block captures images under different lighting conditions, and by combining the different exposure images, obtaining a more optimal image lighting.

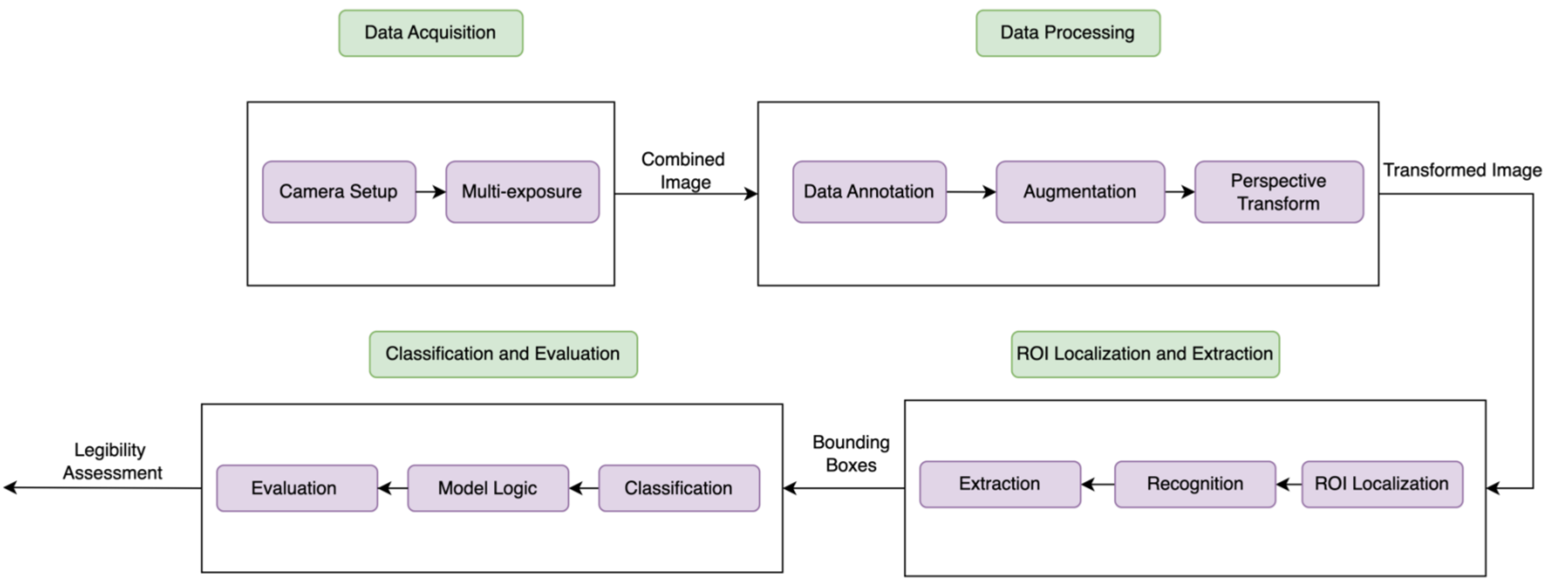


Figure 10 Proposed Conceptual model architecture

The **Data Processing Block** consists of three sub blocks: data annotation, augmentation, and perspective transformation**.** The data annotation sub block involves accessing the collected images, visualizing them, and establishing ground truth labels. The augmentation sub block applies transformations to increase dataset diversity, improving the model's generalization to different conditions. The perspective transformation sub block corrects misalignments in the images.

The **ROI Localization and Extraction Block** consists of three sub blocks: ROI localization, recognition, and extraction. The ROI localization sub block identifies and isolates the Region of Interest (ROI), ensuring the focus is on the Coil ID. The recognition sub block processes the localized ROI to detect and interpret digit information. The extraction sub block isolates individual digits for further legibility analysis.

The **Classification and Evaluation Block** consists of three sub blocks: classification, model logic, and evaluation. The classification sub block assesses the legibility of detected digits, categorizing them based on predefined legibility criteria.

The **model logic** sub block integrates classification results to provide a final decision on Coil ID legibility. The **evaluation** sub block measures model performance using metrics such as recall, and precision, comparing classification results against ground truth annotations.

The main blocks of the conceptual model are elaborated in detail in Chapter 5: Research Design, with Data collection discussed in Section 5.1, Data Processing in Section 5.2, ROI localization and extraction block in Section 5.3 and Classification and Evaluation Block in Section 5.4.

Chapter 5

Research Design

This chapter was organized in the following way. First, the test setup and data collection are detailed in Section 5.1. The data processing steps are outlined in Section 5.2, followed by ROI localization and extraction in Section 5.3. Finally, the classification and evaluation methodology was described in Section 5.4.

5.1 Test Setup & Data Collection

Figure 11: DSP Overview(left) and the process of data collection(right).

Description of the Data Collection Environment

The data collection was conducted at DSP, this facility specializes in producing steel coils where each steel coil was assigned a Coil ID for identification.

A Basler Scout 1400 30gm camera with an 8mm lens was used, mounted 2.5 meters above the coil surface and positioned 1 meter from the spray painting robot as seen in Figure 11(right). The fixed camera positioning ensured consistency in image capture while allowing adaptability across different coil types and lighting conditions. To minimize overlays and shadows, a multi exposure technique was employed to capture images under different lighting conditions.

Setup Components

- **Camera Setup**: Basler Scout 1400 30gm.
- **LED lighting**

Scenario of Data Collection

Figure 12: Coil ID under different lighting conditions at DSP

At the onset of data collection, each coil moved along the production line, positioned directly under the camera to ensure a complete view of the printed Coil ID. The system captures images at predefined intervals, ensuring consistency across different coils and environmental conditions. Every coil is recorded under varying lighting and surface conditions to test robustness against real world factory challenges.

The camera was configured to capture images continuously as the coil moved through these conditions. Each recording session lasted approximately 5 minutes, with additional images captured at the beginning and end to account for initialization and saving delays. This ensured that no critical Coil ID information was lost.

Collected Data:

- Images from the Basler Scout 1400 30gm camera
- Annotated Coil ID bounding boxes for ground truth validation

Data Description

The dataset consists of images of Coil IDs collected under different environmental conditions. It includes annotated samples for digit detection and legibility assessment, with additional augmented images to improve model performance across varying conditions.

Table 1: Overview of Datasets Used in the Study

This table summarizes the different datasets used in digit detection and legibility classification. The datasets were split into training (70%), validation (20%), and test (10%) sets, except for the **Legibility Assessment Test Set**, which was used exclusively for final evaluation.

Dataset Name	**Total Images**	**Train (70%)**	**Validation (20%)**	**Test (10%)**
Digit Baseline (Non Augmented)	4,794	3,347	965	482
Digit Augmented	11,488	8,041	2,298	1,149
Legibility Baseline(Non Augmented)	265	125	75	40
Legibility Augmented	1,008	704	202	102
Legibility Assessment Test Set (Final Evaluation)	**5,834**			

Table 3: This table provides a breakdown of the datasets utilized for digit detection and legibility classification, including their total size and the distribution into training, validation, and test sets. The Legibility Assessment Test Set was reserved exclusively for final evaluation.

Table 2: Class Distribution for Digit Recognition Datasets

This table presents the number of images per class for the Digit Baseline (Non Augmented) and Digit Augmented datasets.

Class	Digit Baseline	Digit Augmented	Class	Digit Baseline	Digit Augmented
0	3,679	8,741	**5**	4,899	11,727
1	1,558	3,746	**6**	8,949	21,339
2	1,290	3,098	**7**	4,080	9,758
3	2,531	6,047	**8**	3,401	8,103
4	1,575	3,795	**9**	1,593	3,794

Table 4: This table presents the number of images per class for both the Digit Baseline and Digit Augmented datasets, showing the distribution of digits (0–9) used in model training and evaluation

Table 3: Class Distribution for Legibility Datasets

This table shows the distribution of legibility classes (High, Medium, Low) in both non augmented and augmented datasets.

Classes	Legibility Baseline (Non Aug)	Legibility Augmented
High	239	478
Medium	256	512
Low	290	580

Table 5: This table presents the distribution of legibility classes (High, Medium, and Low) in both the non-augmented and augmented datasets used for training and evaluation

5.2 Data Processing

Data Annotation

Two separate annotation tasks were completed for digit recognition and print legibility classification. For the digit recognition step, data was processed with the EAST and cropped to include only the relevant digit areas region as seen in Figure 13.

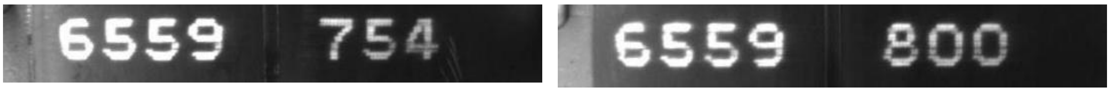


Figure 13: Coil IDs ROI after EAST localization

A small set of images underwent manual annotation to create an initial model, which was then used to annotate the remaining images that contained classes (0-9). The bounding box coordinates were saved in a Pascal VOC XML format. A similar approach was applied to the second dataset, where common defects were categorized into three groups: high, medium, and low. The high category covered clear digits, the medium included digits with visible straps, and the low encompassed smudged digits. Peer review by a colleague helped maintain consistency throughout the process.

Each Coil ID was further categorized based on digit legibility:

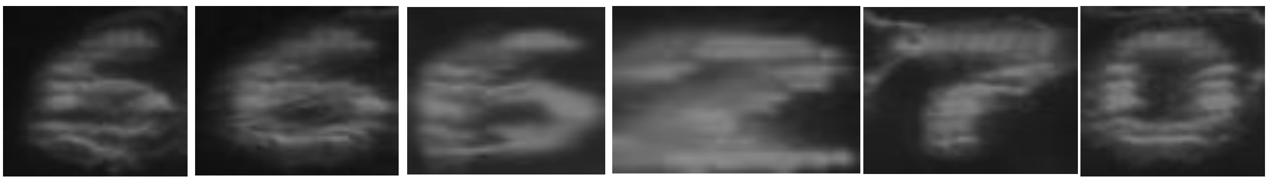


Figure 14: A Visualization of the Low class

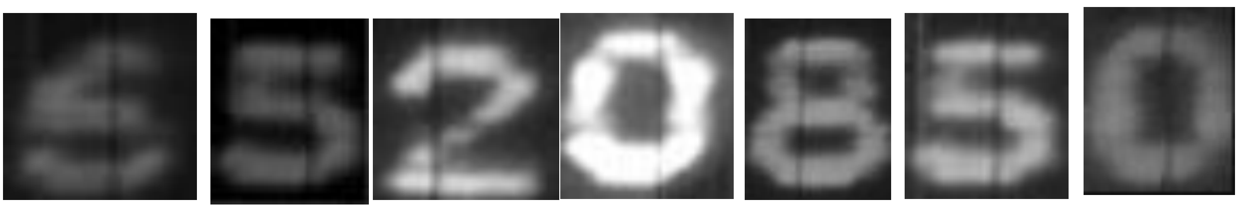


Figure 15: A visualization of the Medium class

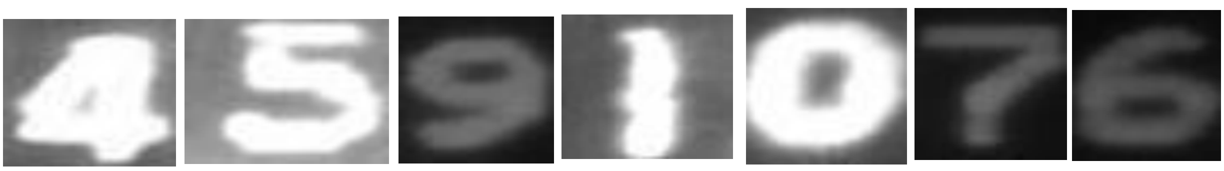


Figure 16: A visualization of the High class

Legibility Class	Description
High	Clearly printed digits with sharp edges and minimal occlusion.
Medium	Straps attached and still legibile
Low	Heavily smudged or partially missing digits.

Table 6: This table defines the criteria used to categorize Coil ID digits into High, Medium, and Low legibility classes based on print clarity and occlusion factors.

Augmentation

Data augmentation was applied to increase dataset diversity and enhance generalization across different conditions. Transformations included:

- +5 and -5 Rotation to simulate angular variations.
- Brightness and contrast adjustments to account for fluctuating lighting conditions.
- Blurring and noise addition to simulate real world smudging and printing inconsistencies

Perspective Transformation

To mitigate distortions arising from skewed image captured as seen in figure 17, perspective transformation was implemented to standardize the orientation of the input image. This process was critical for ensuring accurate digit detection and subsequent legibility analysis. The transformation involves defining corresponding source and destination points, computing a transformation matrix, and applying a warp perspective. Specifically, four source points are manually defined to represent the corners of a quadrilateral in the original image requiring correction.These points are mapped to corresponding destination points that represent the desired, corrected quadrilateral corners.

Figure 17: Coil ID before (left) and after (right) perspective transformation.

5.3 ROI Localization and Extraction

Text Detection Using EAST

The Efficient and Accurate Scene Text Detector (EAST) was employed to localize the Coil ID ater perspective transformation was done correcting any skew introduced during capture. The transformed image was then resized to a standardized resolution of 640×640 pixels to maintain consistency across detections. The EAST model was loaded via OpenCV's deep learning module, utilizing the trained weights, with two primary output layers: one for confidence scores and another for geometry calculations. A strict confidence threshold of 0.9 ensures that only high certainty detections are retained, minimizing false positives. Bounding boxes are computed using EAST's geometry data, factoring in offsets and angle corrections to refine positioning. If no bounding box meets the confidence threshold, a predefined fallback region of interest (ROI) was applied to ensure continuity. This structured approach guarantees that high confidence text regions are prioritized while still accommodating low visibility scenarios through a robust fallback mechanism.

Recognition and Extraction

After identifying areas likely to contain text, the process moved on to isolating individual digits within those regions. A pre-trained digit detector, built on a YOLOv11 model, was used to recognize digits from 0 to 9. The detector scanned each text area and drew boxes around every digit it found, assigning a confidence score to each one to indicate how certain it was about the detection. These boxes, along with their coordinates, confidence levels, and the predicted digits, were then organized and sorted from left to right to preserve the natural order. This detailed mapping of each digit's location and reliability laid the groundwork for the next step, which assessed how legible each digit was.

5.4 Classification and Evaluation

Once the digits were extracted and organized, the next step was to determine their legibility. This required classifying each detected digit into predefined legibility categories: High, Medium, or Low. To achieve this, a deep learning-based classification model was employed to assess print quality, distinguishing between clear, partially degraded, and unreadable digits.

Given the structured nature of Coil ID digits and the need for fast, real-time processing, YOLOv11 was selected as the primary model for this task. YOLO's ability to detect and classify objects in a single pass made it a strong candidate for digit recognition and legibility assessment.

YOLOv11 for Legibility classification

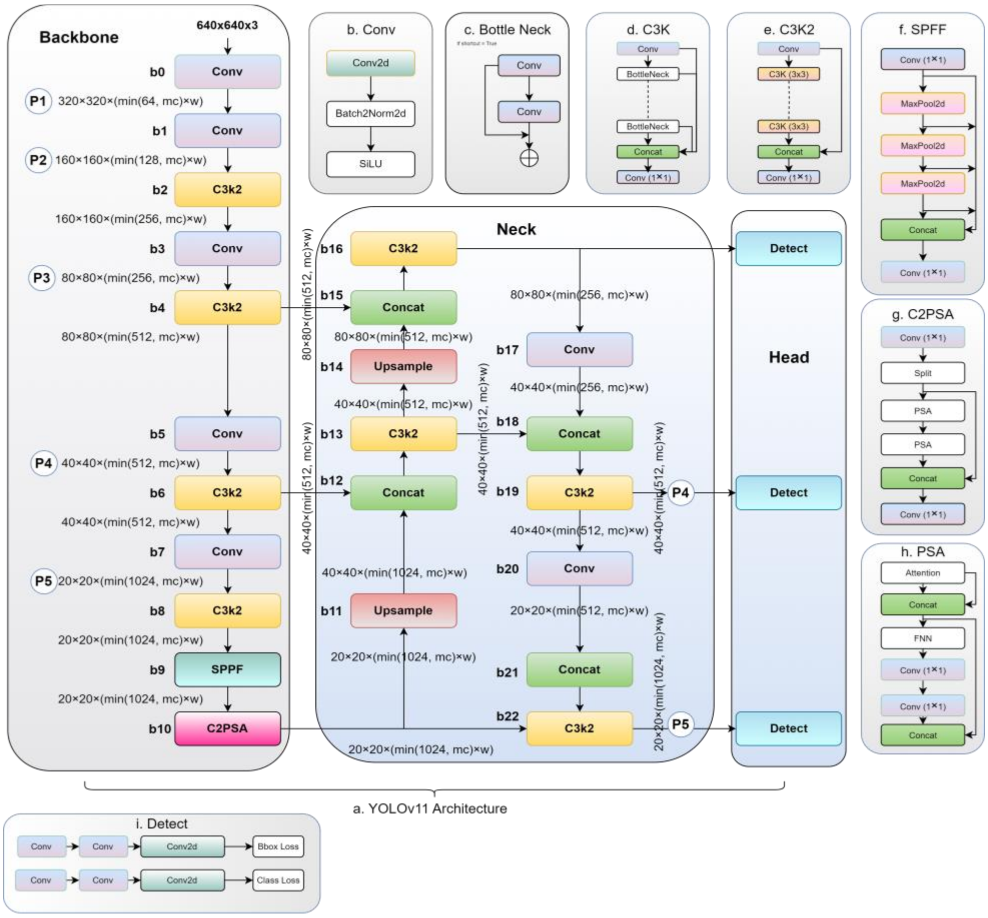


Figure 18 YOLOv11 Architecture[32]

Figure 18 presents YOLOv11, a modern iteration of the YOLO series, optimized for improved accuracy and efficiency in object detection tasks. Earlier versions, such as YOLOv3, were widely used for general-purpose object detection but struggled with small object recognition. YOLOv5 introduced optimizations in training speed and model scaling, making it more efficient for real-time applications.

YOLOv11 builds on these advancements with the architecture includes residual blocks (C3 blocks), which help improve training efficiency by allowing information to bypass certain layers, reducing the risk of vanishing gradients. It also incorporates spatial pyramid pooling (SPPF), a technique that captures multi-scale features by applying pooling operations at different scales. The final stage of the model is a multi-scale detection head, which helps in detecting objects of various sizes within an image[33].

In its fused configuration, the YOLOv11 network consists of 303 layers, with approximately 20 million parameters (trainable weights of the network), and a computational complexity of 67.7 GFLOPs (Giga Floating Point Operations Per Second), which measures how many billion operations the model performs per second.

Module	**Filter Configurations & Operations**
Stem	Conv: 3×3, 64, stride=2; Conv: 3×3, 128, stride=2
Block 1	C3 (C3k2): 1×1, 256, stride=1, dropout=0.25
Block 2	Conv: 3×3, 256, stride=2; C3 (C3k2): 1×1, 512, stride=1, dropout=0.25
Block 3	Conv: 3×3, 512, stride=2; C3 (C3k2): 1×1, 512, stride=1
Block 4	Conv: 3×3, 512, stride=2; C3 (C3k2): 1×1, 512, stride=1
Neck & Head	SPPF (5×5 pooling); C2PSA: 1×1, 512; Upsample & Concat; Detection head with 4 classes (using feature scales: 256, 512, 512)

Table 7: YOLO11m Architecture for Print Legibility Detection

Training was performed using the Ultralytics framework with automatic mixed precision (AMP) enabled, and an AdamW optimizer was employed with learning rate and momentum parameters determined automatically. The network was trained for 100 epochs with a batch size of 16 and an input resolution of 640×640 pixels.

Parameter	Value
Epochs	100
Batch Size	16
Input Resolution	640 × 640
Optimizer	AdamW (auto-selected)
Learning Rate	~0.00125
Momentum	0.9
Weight Decay	0.0005
Warmup Epochs	3
Mixed Precision (AMP)	Enabled
Dataloader Workers	8
Loss Function	YOLO detection loss (box, class, objectness)
Environment	Ultralytics 8.3.40, PyTorch 2.5.1, Python 3.11.11
GPU	Tesla T4

Table 8:Key Training Hyperparameters for YOLO11m

Table 8 summarizes the key hyperparameters and training settings for print legibility detection using a YOLO-based model. Training lasted 100 epochs with a batch size of 16 and an input resolution of 640×640 pixels. The AdamW optimizer was automatically selected, operating with a learning rate of ~0.00125, momentum of 0.9, and weight decay of 0.0005, with a 3 epoch warmup. Mixed precision (AMP) was enabled and data loading used 8 workers. The YOLO detection loss function, incorporating box, class, and objectness components, was applied. The environment comprised Ultralytics 8.3.40, PyTorch 2.5.1, Python 3.11.11, and a Tesla T4 GPU Overall.

Print Legibility Model Algorithm

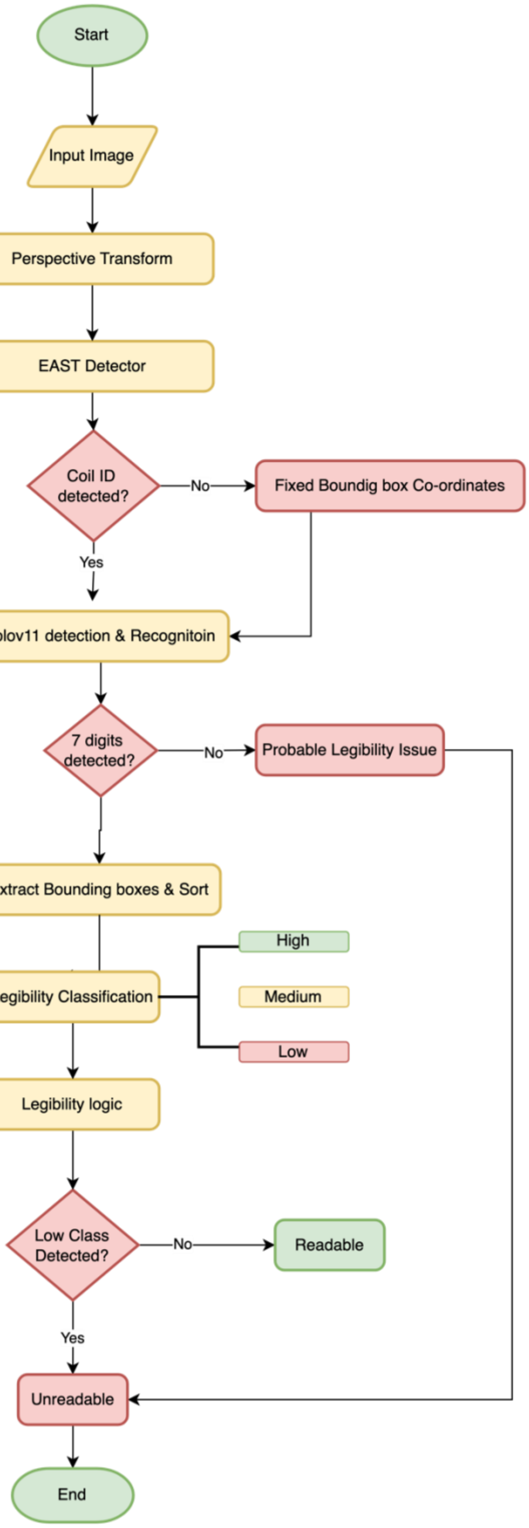


Figure 19: Flow chart of the Print accessment algorithm

The final Coil ID legibility score was determined by If any digit was labeled **Low**, the entire Coil ID was classified as **Unreadable**

Evaluation Metrics

Classification in object detection entails the task of assigning a specific label or class to each identified object. As previously discussed, the YOLO algorithm generates a distribution of multiple bounding boxes, each associated with a probability of belonging to a certain class. The class with the highest probability was selected as the predicted class for the object.

Recall

a performance metric in object detection, quantifies the ability of a classifier to accurately identify all positive instances, irrespective of any false positives [34]. The formula for recall was expressed as:

$$Recall = \frac{True\ positives}{True\ postives + False\ Negatives} \tag{1.5}$$

Precision

The classifier's capability to accurately classify positive instances when both target and non target classes are assessed together, as outlined in equation 1.6 [35]. Precision was calculated as the ratio of the number of correct 'target' decisions when true target objects are present to the total number of 'target' decisions made, regardless of class identity. The class identity in this context can be either a target class or a non target class.

$$precision = \frac{true\ positives}{true\ postives + false\ positves} \tag{1.6}$$

Mean average precision

Mean average precision (mAP) was derived as the mean of the AP values computed for all classes within the dataset [36]. It serves as a standard evaluation metric for assessing the overall accuracy of an object detection system across all classes by averaging the AP values from each object class, as illustrated in equation 1.7.

$$mAP = \frac{1}{n}\sum_{k=1}^{n} AP(k) \tag{1.7}$$

F1 Score

The F1 score is the harmonic mean of precision and recall, providing a single measure that balances both metrics. It is useful when dealing with imbalanced datasets, ensuring that neither precision nor recall is favored disproportionately as seen in equation 1.8.

$$2 \times \frac{precision \times Recall}{Precision + Recall} \tag{1.8}$$

Confusion Matrices

A confusion matrix is a table that summarizes the performance of a classification model by showing true positives, false positives, true negatives, and false negatives. It helps identify misclassification patterns and assess model reliability across different classes.

Precision-Recall Curve

The Precision-Recall curve illustrates how precision and recall vary across different classification thresholds. It is particularly useful for evaluating models in scenarios where one class is significantly more common than the other.

Inference Time

Inference time refers to the duration a model takes to process an input and produce an output. It is crucial for real-time applications where fast decision-making is required. Optimizing inference time involves reducing computational complexity while maintaining accuracy as seen in equation 1.9.

$$inference\ time = \frac{total\ processing}{Number\ of\ processed\ images} \tag{1.9}$$

CHAPTER 6

Research Results

This chapter presents the experimental outcomes of the custom OCR system developed for Coil IDs identification and explains what these results mean. The system was divided into three parts: digit detection, legibility classification, and legibility assessment. Performance was measured by, recall, and the F1 score. In addition, confusion matrix results for the legibility classification and assessment modules are included. The results are compared with previous work and evaluated against the following project requirements:

- **R1:** Detect unreadable Coil IDs with at least 80% recall and 70% precision.
- **R2:** Use data augmentation to address dataset imbalance.
- **R3:** Localize Coil ID regions (ROIs) reliably.
- **R4:** Recognize Coil IDs accurately.
- **R5:** Process each image in less than 5 minutes.

Common tuning methods, such as empirical hyperparameter adjustments and sufficient training epochs (detailed in Chapter 5), were applied in all experiments. The following sections present the results for each module and then evaluate the overall system against the project requirements.

6.1 Digit Detection and Classification Results

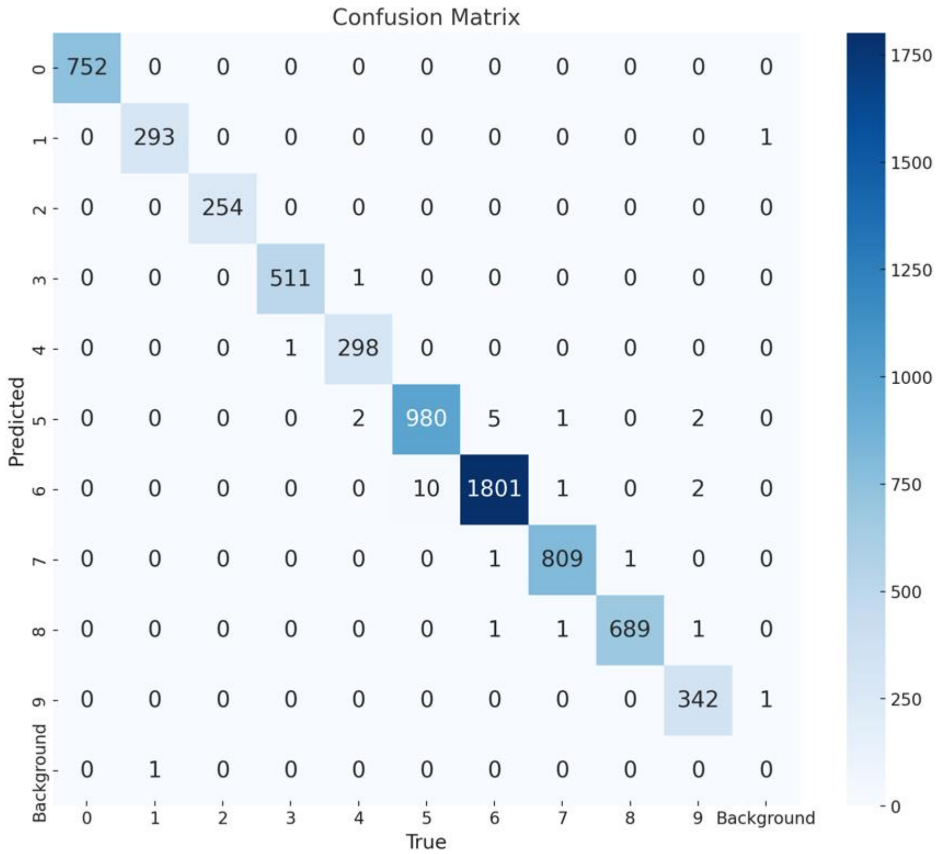


Figure 20: Digit Recognition model Confusion matrix

The digit detection module uses a YOLO based method to extract individual digits from Coil ID images. This module was evaluated using the **Digit Augmented** dataset. Table 1 summarizes its performance.

Accuracy	**Precision**	**Recall**	**F1 Score**	**ROC AUC**	**mAP@0.5**	**mAP@0.5–0.95**
0.855	0.994	0.995	0.995	0.85	0.902	0.866

Table 9: Digit Detection Performance

Error Analysis:

The digit detection classification model shows strong performance, with most digits classified correctly. Misclassifications are minimal, but digit 6 is sometimes mistaken for 5 (10 instances), and digit 9 has some background misclassifications.

6.2 Legibility Classification Results

The legibility classification module evaluates the legibility of each detected digit and classifies it as High, Medium, or Low based on the DSP definiton criteria. Two training models were tested: one using the **Legibility Baseline** (non augmented) dataset and another using the **Legibility Augmented** dataset.

6.2.1 Non Augmented

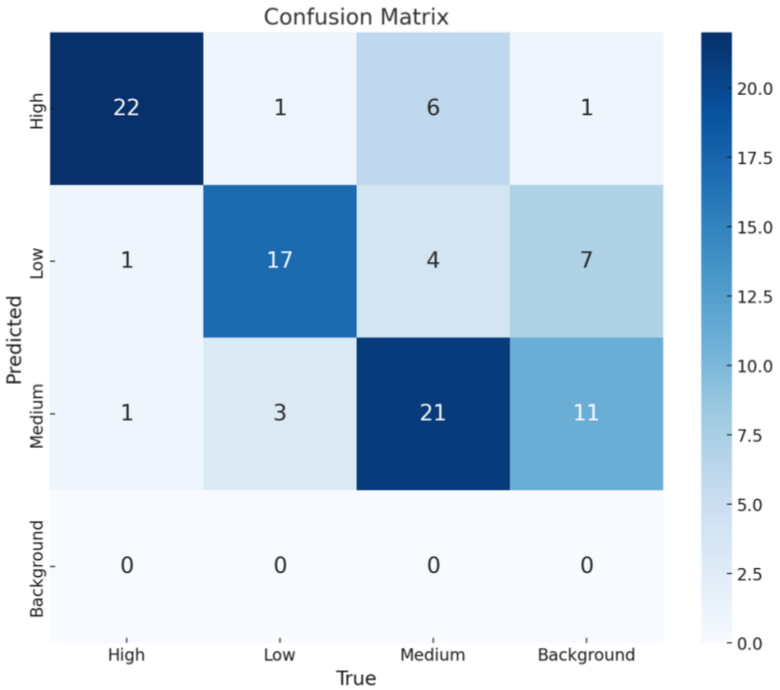


Figure 21: Non-Augmented Confusion Matrix

Using the **Legibility Baseline** dataset, the module achieved:

Table 2. Legibility Classification – Non Augmented Model

Accuracy	Precision	Recall	F1 Score	ROC AUC	mAP@0.5	mAP@0.95
0.79	0.84	0.79	0.80	0.853	0.907	0.867

Table 10: . Legibility Classification – Non Augmented Model performance

Confusion Matrix Summary (Non Augmented):

- **High:** 22 correct; 6 misclassified as Medium, 1 as Low, and 1 as Background.
- **Low :** 17 correct; 4 misclassified as Medium and 7 as Background.
- **Medium :** 21 correct; 11 misclassified as Background, 3 as Low, and 1 as High.

The main error was that many Medium legibility digits are confused with Background.

6.2.2 Augmented Model

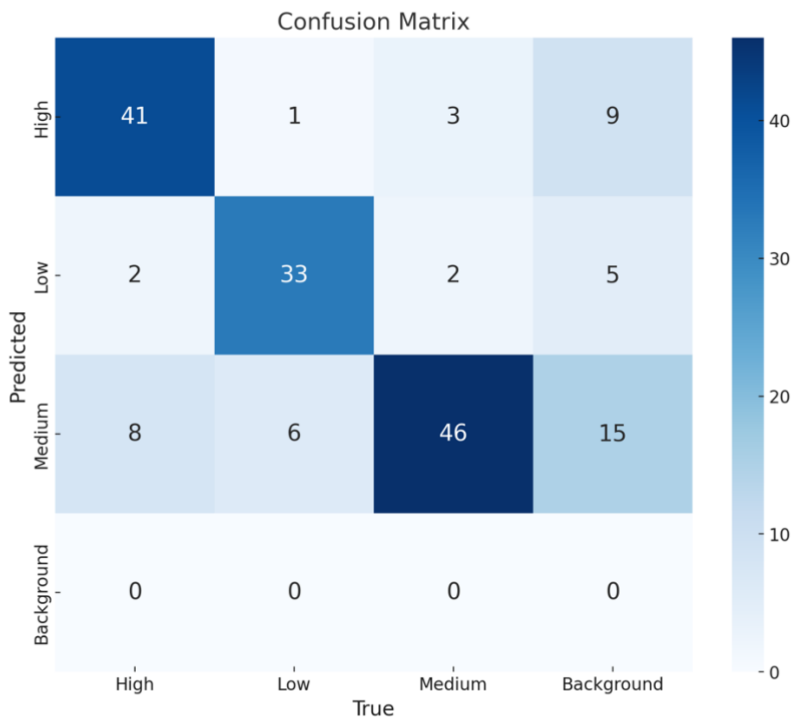


Figure 22: Augmented Confusion Matrix

With the **Legibility Augmented** dataset, the module achieved:

Accuracy	Precision	Recall	F1 Score	ROC AUC	mAP@0.5	mAP@0.5–0.95
0.855	0.702	0.845	0.767	0.85	0.902	0.866

Table 11: Legibility Classification – Augmented Model Performance

Confusion Matrix Summary (Augmented):

- **High Legibility:** 41 correct; 3 misclassified as Medium, 1 as Low, and 9 as Background.
- **Low Legibility:** 33 correct; 2 misclassified as Medium and 5 as Background.
- **Medium Legibility:** 46 correct; 15 misclassified as Background.

The augmented model improves the detection of degraded digits (recall increases) but also leads to more false alarms (precision decreases). This result shows that data augmentation helps to address dataset imbalance (R2).

6.3 Legibility Assessment Results

The legibility assessment module decides if a whole Coil ID was readable based on the algorithm of the overall system as explained in chapter 5. The rule was simple: if any digit was classified as Low, the Coil ID was marked as Unreadable. Firstly, EAST detector was evaluated in this test before the two models were evaluated using the **Legibility Assessment Test Set**.

EAST Performance Evaluation

EAST was tested on the Legibility Assessment Test Set to measure how often it successfully detected the Coil ID. Out of 5,834 images, EAST detected the Coil ID in 5,813 cases, giving a 99.64% success rate. In 21 images, EAST did not detect the Coil ID, so the fallback method was used.

All **21 fallback cases** were later found to have legibility issues. This shows that when EAST fails, it is usually because the Coil ID has an indirect probable legibility issue.

Detection Outcome	**Number of Images**	**Percentage**
Coil ID detected by EAST	5,813	99.64%
Coil ID not detected (fallback used)	21	0.36%

Table 12: EAST Detection Performance on the Legibility Assessment Test Set

EAST Detection Outcome	**Readable Coil IDs**	**Unreadable Coil IDs**	**Total**
Coil ID detected	5,762	51	5,813

Table 13: Legibility of Coil IDs Detected by EAST

Fallback Used	**Readable Coil IDs**	**Unreadable Coil IDs**	**Total**
Yes	0	21	21

Table 14: Legibility of Coil IDs When Fallback Was Used

The results indicate that EAST works well for localizing Coil IDs, and when it does not detect one, the Coil ID is usually unreadable. The fallback ensures that these cases are not missed, and still goes through the pipeline.

6.3.1 Augmented Model Test

The augmented model achieved these results from the test:

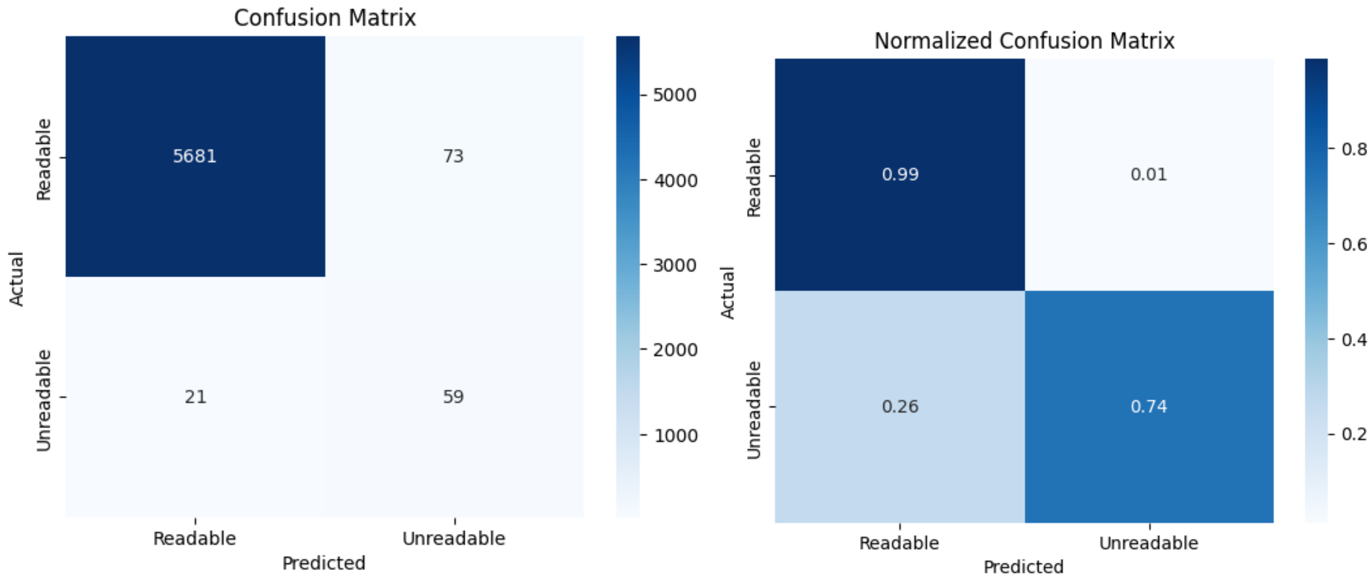


Figure 23: Augmented Assessment matrix showing counts (left) and normalized values (right).

Precision	**Recall**	**F1 Score**
0.44	0.73	0.55

Table 15: Legibility Assessment – Augmented Model

Confusion Matrix Summary (Augmented Assessment):

- For Readable Coil IDs: 5,681 correct and 73 misclassified as Unreadable.
- For Unreadable Coil IDs: 59 correct and 21 misclassified as Readable.
- Normalized confusion matrix results indicate 99% of Readable IDs are correct and 74% of Unreadable IDs are correctly identified.

These results mean that the system catches about 73% of poor quality Coil IDs, but its precision was low (44%), meaning many acceptable IDs are falsely flagged. This falls short of the target (R1) of at least 70% precision.

The Precision-Recall curve

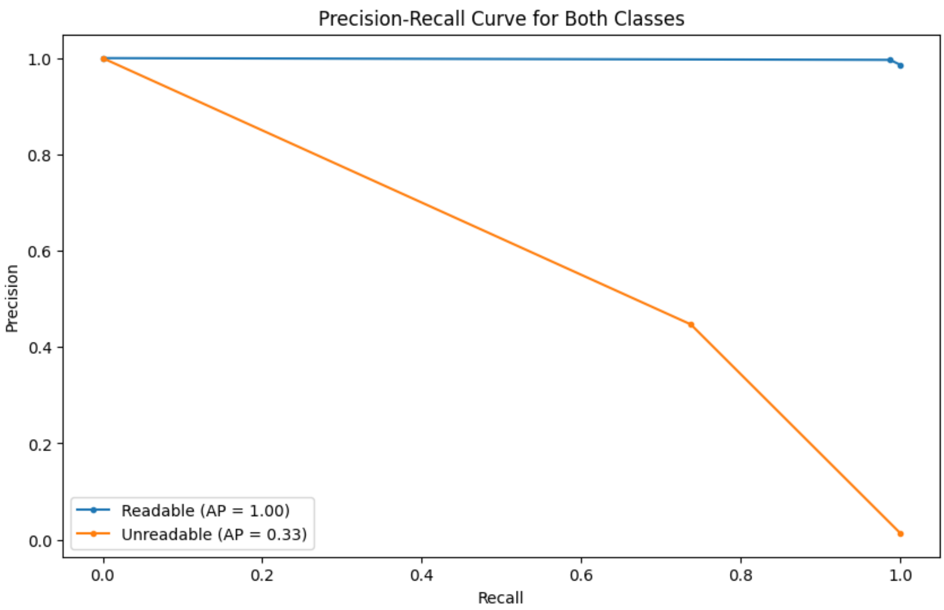


Figure 24: Augmented Model test Precision-Recall curve

The Precision-Recall curve shows perfect performance for the readable class with an AP of 1.00, indicating flawless precision and recall. In contrast, the unreadable class has an AP of 0.33, with precision dropping sharply as recall increases.

6.3.2 Non Augmented Model Test

The non augmented model produced these results from the test:

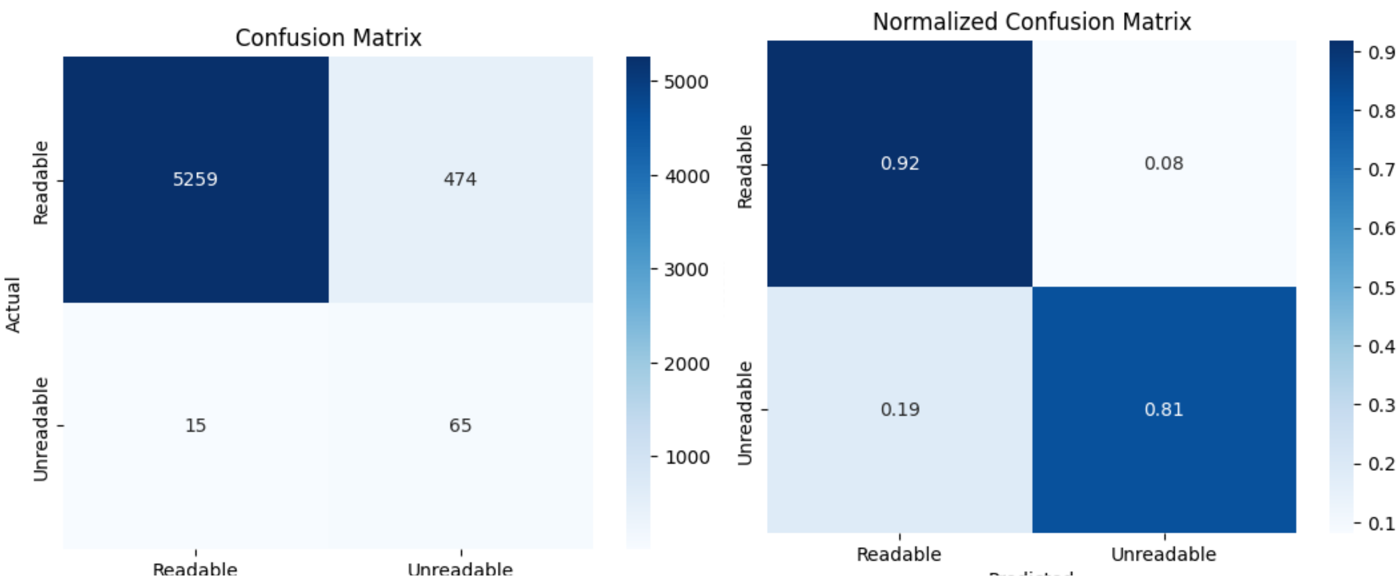


Figure 25: Non-Augmented Assessment matrix showing counts (left) and normalized values (right).

Precision	Recall	F1 Score
0.116	0.812	0.203

Table 16:Legibility Assessment – Non Augmented Model

Confusion Matrix Summary (Non Augmented Assessment):

- For Readable Coil IDs: 5,259 correct and 474 misclassified as Unreadable.
- For Unreadable Coil IDs: 65 correct and 15 misclassified as Readable.
- Normalized results show that 92% of Readable IDs and 81% of Unreadable IDs are correctly classified.

The non augmented model has very low precision (about 11.6%) and was therefore impractical for industrial use. Overall, the augmented model offers a better balance, though it still does not meet the industrial requirement.

The Precision-Recall curve

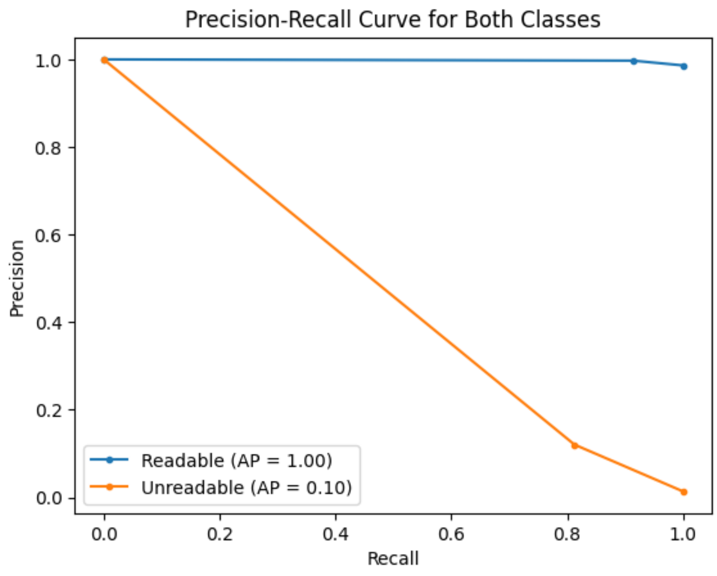


Figure 26: Non-Augmented Model test Precision-Recall curve

The Precision-Recall curve for the non-augmented model shows an average precision (AP) of 1.00 for the readable class, indicating perfect precision across all recall levels. For the unreadable class, the AP is 0.10, with precision sharply declining as recall increases, reflecting poor performance in correctly identifying unreadable instances.

Processing Time

The system's processing time was measured using the Legibility Assessment Test Set (5,834 images). Each image took 6.3 seconds to process, resulting in a total runtime of approximately 10 hours and 12 minutes**.** The table below summarizes the processing time:

Metric	Value
Total Images Processed	5,834
Processing Time per Image (seconds)	6.3
Total Processing Time (seconds)	36,754.2
Total Processing Time (minutes)	612.57

Table 17: total processing time for legibility assessment test set

6.4 Hypotheses Evaluation

This research was guided by five key requirements. Here was a plain explanation of each requirement, what was expected, what was actually observed, and the overall conclusion:

Detecting Unreadable Coil IDs (R1):

The goal was to correctly identify unreadable Coil IDs with at least 80% recall and 70% precision. In simple terms, the system was expected to find at least 80% of the truly unreadable IDs and be correct about 70% of the time when it flagged an ID as unreadable. However, the tests showed that the augmented model only reached 73% recall and 44% precision, and the non augmented model had a similar recall (around 81%) but an even lower precision (about 12%). In plain language, the system missed too many good IDs and incorrectly flagged too many acceptable IDs as unreadable. Thus, this requirement was not met.

Data Augmentation (R2):

Unreadable Coil IDs were rare, making it impossible to collect enough training data quickly. Augmentation was used to artificially increase the dataset, ensuring the model saw more degraded examples.This improved recall from 79% to 84.5%, meaning the model detected more unreadable digits. However, precision dropped from 84% to 70%, increasing false positives cases where readable digits were misclassified as unreadable.Since this model was later used for full Coil ID assessment, these errors carried over. The final assessment reached 73% recall but only 44% precision, meaning too many readable Coil IDs were wrongly flagged as unreadable.Augmentation helped detect more unreadable digits but also led to more mistakes. While it addressed data imbalance, it introduced new challenges, making the result only partially successful.

Localization of Coil ID Regions (R3):

This requirement focused on accurately extracting the Coil ID region from images. The expectation was high detection accuracy with minimal failures.EAST was used as the primary localization method. It successfully detected the Coil ID in 99.64% of cases on the **Legibility Assessment Test Set**, failing in 21 images where the fallback method was applied. All 21 fallback cases were later found to have legibility issues, confirming that EAST failures were mostly due to illegible prints rather than detection errors..This requirement was fully met, with EAST effectively localizing the Coil ID in most cases, and fallback handling the few failures.

Recognition of Coil IDs (R4):

The system was expected to recognize digits (0-9) after EAST localized the Coil ID region. The primary goal was not just classification accuracy but accurately extracting bounding boxes for further pipeline analysis. The YOLO digit recognition model successfully identified and extracted bounding boxes for each digit, preserving their order and structure. This ensured a clean input for the legibility classification stage. Since the extracted bounding boxes enabled the next steps, this requirement was fully met despite occasional detection inconsistencies.

Processing Time (R5):

Finally, the system was required to process each image in less than 5 minutes. In fact, the observed inference time was below 7 seconds per image, which was well within the required limit. This requirement was fully met.

Literature Comparison

A review of related work as written in Chapter 2 provides useful context for these findings. Caldeira et al. [25] The system's digit detection (99.4% precision, 99.5% recall) marginally exceeded [25] 98% accuracy benchmark, though their system processed more images (20,000 vs 5,834) and included letters. Text localization through EAST achieved 99.64% detection success, but processing speed (6.3 seconds/image) was significantly slower than Caldeira's 0.36 seconds. The primary limitation emerged in legibility assessment, achieving only 44% precision against the 70% target. While the system introduced novel legibility assessment capabilities, it traded speed .

Chapter 7

Conclusions and Recommendation

Examining the outcomes presented in Chapter 6, together with the initial Problem Statement in Section 1.2 and Research Questions from Chapter 4, two primary conclusions can be derived. Firstly, the developed system achieved exceptional performance in digit recognition, with 99.4% precision in identification and 99.64% accuracy in locating text regions. However, the legibility assessment component, while detecting 73% of unreadable, demonstrated insufficient precision at 44%, below the requirement of 70%.

The sub-questions stated in Chapter 4 were systematically evaluated, yielding these findings:

- Sub-question 1: The implementation of EAST for text localization proved highly capable, successfully identifying Coil ID regions in 99.64% of 5,834 test cases. Notably, instances where EAST failed to detect text corresponded with genuine print quality issues.
- Sub-question 2: The YOLO algorithm system exhibited robustness under diverse industrial conditions, delivering 99.4% precision and 99.5% recall in digit identification tasks.
- Sub-question 3: The application of machine learning for legibility evaluation revealed mixed outcomes. Despite achieving satisfactory recall rates (73%), the system's precision remained inadequate for industrial deployment at (44%).
- Sub-question 4: The study successfully identified and mitigated various environmental challenges through enhanced data augmentation techniques, improving detection rates from 79% to 84.5%, though this introduced additional precision challenges.
- Sub-question 5: Processing efficiency exceeded expectations, with an average time of 6.3 seconds per image, substantially outperforming the 5-minute requirement and confirming suitability for real-time applications.

7.1 Discussion

Reflection on Digit Detection Results

Analyzing the digit detection experiments from Section 6.1, particularly examining the confusion matrices, the system achieved 99.4% precision and 99.5% recall. This high performance can be attributed to YOLO's robust architecture and the clear distinction between digits. However, some digits showed consistent misclassification patterns, particularly between visually similar pairs like '8' and '6', suggesting the need for more targeted training on these specific cases.

Reflection on Legibility Assessment

Examining the legibility classification results introduced in Section 6.2, the system demonstrated a clear bias toward flagging prints as "unreadable." The precision of 44% versus a target of 70% indicates systematic oversensitivity. This bias likely stems from the training approach, where the cost of missing an unreadable print was weighted higher than false positives. The model's performance improved with augmentation (recall increased from 79% to 84.5%), but at the cost of reduced precision.

Reflection on Text Localization

The EAST detector's 99.64% success rate in text localization proved highly effective, with an unexpected correlation between detection failures and print quality issues. All 21 failed detections corresponded to legitimately problematic prints, suggesting EAST's sensitivity to print quality could serve as an additional quality indicator. However, this correlation requires further validation across different operating conditions.

Reflection on Pipeline and Implementation

The potential bottleneck in the pipeline is processing time (6.3 seconds per image). While meeting the 5-minute requirement, it's significantly slower than existing solutions. This stems from:

- Multiple processing stages (EAST followed by YOLO)
- Perspective transformation overhead
- Additional legibility assessment computations

The system demonstrates robust performance in controlled conditions but may require additional validation for varying industrial environments and print qualities.

7.2 Limitations and Gaps

Despite the system's performance, several limitations remain :

High False Positive Rate in Legibility Classification

The biggest issue is too many readable Coil IDs being marked as unreadable**.** While the model successfully detects unreadable prints, it overcompensates by flagging prints that a human operator would still consider legible. This is a major problem in an industrial setting where false rejections can cause unnecessary false alarms to operators.

Data Augmentation Trade-offs

Augmenting the dataset helped increase recall (catching more bad prints), but it also introduced more mistakes in classification. The model learned to expect certain types of defects that might not always exist in real world conditions. A smarter augmentation strategy is needed one that mimics actual print failures rather than just adding random distortions.

Limited data of unreadable Coil IDs

The dataset lacked enough samples of Coil IDs that were severely degraded, smudged, or partially missing. As a result, the model struggles with extreme cases that happen less often but matter the most when they do.

Unnecessary Complexity in Text Localization Block

The perspective transformation step adds unnecessary complexity, making the system heavily dependent on a fixed adjusted points. This reduces scalability, as the approach may not generalize well to other environments with different camera positions or coil orientations. Additionally, while EAST effectively fails on poor-quality prints, confirming its potential as a readability indicator, its role in text localization adds extra processing steps that could be eliminated. Using a single model like YOLO for both localization and classification would streamline the pipeline, improve reproducibility.

7.3 Recommendations for Future Work

Based on these conclusions, the following recommendations are proposed:

Data Augmentation Techniques:

Investigate alternative methods, such as GAN based synthetic data generation, to create more realistic training samples without excessive variability. This may help improve precision while maintaining high recall.

Eliminate the EAST text localization step:

For easy reproducibility and reducing complexity the text detection step using EAST detector can be eliminated and a recommendation would be to use YOLO for localization and as well recognition.

Improve Post Processing Strategies:

A post processing logic can be utilised using template matching or SSIM as a final evaluation strategy for legitimate assessment, where the value of the score can be utilised.

Expansion of the Dataset:

Increase the variety and number of images, particularly for unreadable Coil IDs, to better capture the full range of conditions encountered in production. This will help the model generalize better and reduce bias.

Active Learning:

Implementing active learning would allow the model to prioritize uncertain predictions and request human annotations only when necessary, reducing the need for fully labeled datasets. This approach enhances scalability by continuously improving model performance with minimal manual intervention, making the system more adaptable to new environments and unseen variations in digit legibility.

References

[1] "(PDF) Toward intelligent manufacturing: label characters marking and recognition method for steel products with machine vision," *ResearchGate*, Oct. 2024, doi: 10.1007/s40436-014-0057-2.

[2] A. Brewer, N. Sloan, and T. L. Landers, "Intelligent tracking in manufacturing," *Journal of Intelligent Manufacturing*, vol. 10, no. 3, Art. no. 3, Sep. 1999, doi: 10.1023/A:1008995707211.

[3] Xiuying Wang, Shengping Yu, Binglin Zheng, and Tianyou Chai, "Intelligent Scheduling System of Steelmaking and Continuous Casting Based On ERP/MES/PCS," in *2006 6th World Congress on Intelligent Control and Automation*, Dalian, China: IEEE, 2006, pp. 7381–7384. doi: 10.1109/WCICA.2006.1714520.

[4] "Plants | Tata Steel." Accessed: Feb. 09, 2025. [Online]. Available: https://www.tatasteelnederland.com/en/How-we-make-steel/Plants

[5] T. Caldeira, P. M. Ciarelli, and G. A. Neto, "Industrial Optical Character Recognition System in Printing Quality Control of Hot-Rolled Coils Identification," *Journal of Control, Automation and Electrical Systems*, vol. 31, no. 1, Art. no. 1, Dec. 2019, doi: 10.1007/s40313-019-00551-1.

[6] M. Liukkonen and T.-N. Tsai, "Toward decentralized intelligence in manufacturing: recent trends in automatic identification of things," *Int J Adv Manuf Technol*, vol. 87, no. 9–12, pp. 2509–2531, Dec. 2016, doi: 10.1007/s00170-016-8628-y.

[7] E. Alpaydin, *Introduction to Machine Learning, fourth edition*. MIT Press, 2020.

[8] C. M. Bishop, *Pattern recognition and machine learning*. in Information science and statistics. New York: Springer, 2006.

[9] Yumpu.com, "(The MIT Press Essential Knowledge) John D. Kelleher, Brendan Tierney - Data Science-The MIT Press (2018)," yumpu.com. Accessed: Feb. 08, 2025. [Online]. Available: https://www.yumpu.com/en/document/view/63817371/the-mit-press-essential-knowledge-john-d-kelleher-brendan-tierney-data-science-the-mit-press-2018

[10] "(PDF) Semi-Supervised Learning edited by O. Chapelle, B. Schölkopf and A. Zien." Accessed: Feb. 08, 2025. [Online]. Available: https://www.researchgate.net/publication/281327930_Semi-Supervised_Learning_edited_by_O_Chapelle_B_Scholkopf_and_A_Zien

[11] "A logical calculus of the ideas immanent in nervous activity".

[12] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," Dec. 02, 2015, *arXiv*: arXiv:1511.08458. doi: 10.48550/arXiv.1511.08458.

[13] K.-H. Kim, S. Hong, B. Roh, Y. Cheon, and M. Park, "PVANET: Deep but Lightweight Neural Networks for Real-time Object Detection," Sep. 30, 2016, *arXiv*: arXiv:1608.08021. doi: 10.48550/arXiv.1608.08021.

[14] W. Shang, K. Sohn, D. Almeida, and H. Lee, "Understanding and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units," Jul. 19, 2016, *arXiv*: arXiv:1603.05201. doi: 10.48550/arXiv.1603.05201.

[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

[16] "Going deeper with convolutions | IEEE Conference Publication | IEEE Xplore." Accessed: Feb. 09, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/7298594
[17] X. Zhou *et al.*, "EAST: An Efficient and Accurate Scene Text Detector," Jul. 10, 2017, *arXiv*: arXiv:1704.03155. doi: 10.48550/arXiv.1704.03155.
[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.
[19] G. Yang *et al.*, "Face Mask Recognition System with YOLOV5 Based on Image Recognition," *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pp. 1398–1404, Dec. 2020, doi: 10.1109/ICCC51575.2020.9345042.
[20] "Varying bounding box coordinates. | Download Scientific Diagram." Accessed: Feb. 04, 2025. [Online]. Available: https://www.researchgate.net/figure/arying-bounding-box-coordinates_fig7_336760546
[21] P. Deepak and S. Suresh, "Design and Utilization of Bounding Box in Human Detection and Activity Identification," in *Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2*, vol. 338, S. C. Satapathy, A. Govardhan, K. S. Raju, and J. K. Mandal, Eds., in Advances in Intelligent Systems and Computing, vol. 338. , Cham: Springer International Publishing, 2015, pp. 59–70. doi: 10.1007/978-3-319-13731-5_8.
[22] A. Bousaid, T. Theodoridis, S. Nefti-Meziani, and S. Davis, "Perspective Distortion Modeling for Image Measurements," *IEEE Access*, vol. 8, pp. 15322–15331, 2020, doi: 10.1109/ACCESS.2020.2966716.
[23] Z. Wang *et al.*, "A Comprehensive Survey on Data Augmentation," 2024, *arXiv*. doi: 10.48550/ARXIV.2405.09591.
[24] T. Caldeira, P. M. Ciarelli, and G. A. Neto, "Industrial Optical Character Recognition System in Printing Quality Control of Hot-Rolled Coils Identification," *J Control Autom Electr Syst*, vol. 31, no. 1, pp. 108–118, Feb. 2020, doi: 10.1007/s40313-019-00551-1.
[25] "(PDF) ISO 50001: 2018 and Its Application in a Comprehensive Management System with an Energy-Performance Focus." Accessed: Feb. 09, 2025. [Online]. Available: https://www.researchgate.net/publication/337878120_ISO_50001_2018_and_Its_Application_in_a_Comprehensive_Management_System_with_an_Energy-Performance_Focus
[26] admin, "The New ISO 14001 Standard | TTBS." Accessed: Feb. 09, 2025. [Online]. Available: https://gottbs.com/2015/09/25/the-new-iso-14001-standard/
[27] "ISO 26000:2010," ISO. Accessed: Feb. 09, 2025. [Online]. Available: https://www.iso.org/standard/42546.html
[28] "AI transparency: What is it and why do we need it? | TechTarget." Accessed: Feb. 09, 2025. [Online]. Available: https://www.techtarget.com/searchcio/tip/AI-transparency-What-is-it-and-why-do-we-need-it?utm_source=chatgpt.com
[29] "AI Risks and Trustworthiness." Accessed: Feb. 09, 2025. [Online]. Available: https://airc.nist.gov/airmf-resources/airmf/3-sec-characteristics/
[30] "What is AI Bias? - Understanding Its Impact, Risks, and Mitigation Strategies." Accessed: Feb. 09, 2025. [Online]. Available: https://www.holisticai.com/blog/what-is-ai-bias-risks-mitigation-strategies?utm_source=chatgpt.com
[31] "Frontiers | Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making." Accessed: Feb. 09, 2025. [Online]. Available:

https://www.frontiersin.org/journals/human-dynamics/articles/10.3389/fhumd.2024.1421273/full?utm_source=chatgpt.com

[32] A. F. Rasheed and M. Zarkoosh, “YOLOv11 Optimization for Efficient Resource Utilization,” Dec. 21, 2024, *arXiv*: arXiv:2412.14790. doi: 10.48550/arXiv.2412.14790.

[33] R. Khanam and M. Hussain, “YOLOv11: An Overview of the Key Architectural Enhancements,” Oct. 23, 2024, *arXiv*: arXiv:2410.17725. doi: 10.48550/arXiv.2410.17725.

[34] D. M. W. Powers, “Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation,” Oct. 11, 2020, *arXiv*: arXiv:2010.16061. doi: 10.48550/arXiv.2010.16061.

[35] J. Davis and M. Goadrich, “The relationship between Precision-Recall and ROC curves,” *Proceedings of the 23rd international conference on Machine learning - ICML ’06*, pp. 233–240, 2006, doi: 10.1145/1143844.1143874.

[36] R. Padilla, S. L. Netto, and E. A. B. Da Silva, “A Survey on Performance Metrics for Object-Detection Algorithms,” *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 237–242, Jul. 2020, doi: 10.1109/IWSSIP48289.2020.9145130.

Definitions	Abbreviations
CLAHE	Contrast Limited Adaptive Histogram Equalization.
Coil ID	Coil Identification Number.
CNN	Convolutional Neural Network.
DSP	Direct Sheet Plant.
EAST	Efficient and Accurate Scene Text Detector.
F1 score	F1 Score.
FP	False Positives.
FN	False Negatives.
HOG	Histogram of Oriented Gradients.
IoU	Intersection over Union.
mAP	Mean Average Precision.
OCR	Optical Character Recognition.
RFID	Radio Frequency Identification.
ROI	Region of Interest.
ROC AUC	Receiver Operating Characteristic Area Under the Curve.
SVM	Support Vector Machine.
TP	True Positives.
VOC	Visual Object Classes.
YOLOv11	You Only Look Once Version 11.

Appendix A

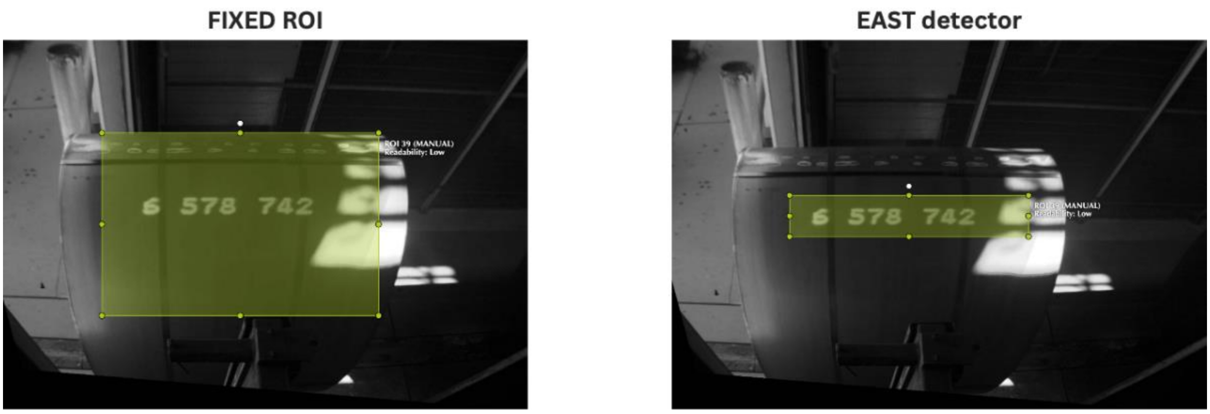


Figure 27:Fixed ROI vs. EAST Detector for Coil ID Localization

Comparison of Fixed ROI vs. EAST Detector for Coil ID Localization

The figure 27 illustrates the difference between using a **Fixed ROI approach** and an **EAST Detector-based dynamic ROI** for extracting Coil ID regions. The EAST detector provides a **more adaptive and precise** localization compared to the manually set Fixed ROI, which often captures unnecessary background information.

Observations:

- **Fixed ROI**: Covers a **larger region**, which includes unnecessary background and potential noise.
- **EAST Detector**: Dynamically adjusts to the **actual text area**, making segmentation and readability assessment more **efficient**.

Impact: The EAST-based method Localizes the region of the Coil ID tighter compared the secondary mechanism of the Fixed bounding box.

Appendix B

```
# Destination points for perspective transform
pts_src = np.float32([[237, 539], [242, 491], [909, 508], [893, 465]])
pts_dst = np.float32([[237, 535], [237, 490], [908, 535], [908, 490]])
```

Table 18: Perspective transformatino points

Perspective Transformation Points

The figure defines **source (pts_src) and destination (pts_dst) points** for a **perspective transformation**. This is used to correct distortions in the coil ID region and align it properly for further processing.

Key Details:

- **pts_src (Source Points)**: These are the original coordinates from the image where the Coil ID appears.
- **pts_dst (Destination Points)**: These are the transformed coordinates, mapping the text into a straight, aligned region.
- **Purpose**: Helps in **rectifying perspective distortions**, ensuring better segmentation and readability assessment.

Usage in the Pipeline:

This transformation ensures that the printed Coil ID image is **correctly aligned**, improving the efficiency of the **deep learning-based readability assessment**.