CHALMERS
UNIVERSITY OF TECHNOLOGY

From text to meaning: Semantic interpretation of non-standardized metadata in piping and instrumentation diagrams

Downloaded from: https://research.chalmers.se, 2026-10-16 00:08 UTC

Citation for the original published paper (version of record):
Shteriyanov, V., Dzhusupova, R., Bosch, J. et al (2026). From text to meaning: Semantic interpretation of non-standardized metadata in piping and instrumentation diagrams. Computers and Chemical Engineering, 204. http://dx.doi.org/10.1016/j.compchemeng.2025.109436

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Computers and Chemical Engineering 204 (2026) 109436

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/cace

From text to meaning: Semantic interpretation of non-standardized metadata in piping and instrumentation diagrams

Vasil Shteriyanov [a,b,*], Rimma Dzhusupova [a,b], Jan Bosch [b,c], Helena Holmström Olsson [d]

[a] McDermott, Engineering, The Hague, The Netherlands
[b] Eindhoven University of Technology, Mathematics and Computer Science, Eindhoven, The Netherlands
[c] Chalmers University of Technology, Computer Science and Engineering, Gothenburg, Sweden
[d] Malmö University, Computer Science and Media Technology, Malmö, Sweden

ARTICLE INFO

Keywords:
Information extraction
Hybrid AI systems
Document analysis
Engineering drawings
Engineering automation

ABSTRACT

The extraction of structured metadata from Piping and Instrumentation Diagrams (P&IDs) is a major bottleneck for digitalization in the process industries. Existing methods, based on Optical Character Recognition (OCR), stop at raw text extraction, failing to interpret critical engineering information encoded within variable-format identifiers like pipeline numbers. This paper bridges this semantic gap by introducing a system for the format-aware interpretation of P&ID pipeline metadata. Our hybrid system architecture integrates deep learning for text recognition with domain interpretation rules that allow the system to adapt to new project formats without model retraining. These rules perform validation, error correction, and semantic mapping of raw text to structured data. We validated our system on a challenging dataset of real-world P&IDs from four distinct industrial projects, each with a unique and complex pipeline number format. Our method achieved 91.1% end-to-end accuracy, demonstrating a significant leap in performance over standard OCR tools, which proved insufficient for the task. This work presents a robust solution that unlocks valuable data from non-standardized engineering documents, providing a practical pathway for creating reliable digital twins and supporting plant lifecycle management in the chemical engineering sector.

1. Introduction

The Engineering, Procurement, and Construction (EPC) industry is responsible for delivering large-scale infrastructure projects within the process industries, including energy, chemicals, and pharmaceuticals (Berends, 2007). These complex, high-stakes projects operate under immense pressure to meet fixed budgets and aggressive timelines. In this environment, operational efficiency is paramount, and EPC contractors continually seek innovative ways to streamline processes, mitigate risks, and reduce costs. Thus, the EPC industry could leverage Artificial Intelligence (AI) to enhance productivity and deliver projects with greater precision and speed. Example uses of AI in the EPC sector are the extraction and analysis of technical risks in bidding specifications (Park et al., 2021), error detection in engineering diagrams (Dzhusupova et al., 2022a), and automated material procurement and cost estimation using deep learning and regression techniques (Dzhusupova et al., 2025).

A critical barrier to this AI-driven efficiency lies in the nature of engineering data itself. Much of this data is locked within technical documents that lack consistent formats. This lack of standardization across projects, companies, and decades of work presents a fundamental challenge in the development of scalable automation solutions for the engineering industry. Therefore, developing methods to intelligently extract and interpret information from these non-standard documents is important to unlock significant productivity gains across the engineering sector.

An example of this challenge is the extraction of metadata from essential engineering drawings. Among the most fundamental of these documents is the Piping and Instrumentation Diagram (P&ID), a cornerstone of plant design and process systems engineering. A P&ID illustrates the connection between piping and process equipment, and the instrumentation devices used for process control, forming the basis for critical activities like process hazard analysis, control strategy development, and plant lifecycle management (Theisen et al., 2023). A synthetic simplified example of a P&ID from a public dataset (Paliwal et al., 2021) is shown in Fig. 1.

* Corresponding author at: Eindhoven University of Technology, Mathematics and Computer Science, Eindhoven, The Netherlands.
E-mail addresses: v.shteriyanov1@tue.nl (V. Shteriyanov), rdzhusupova@mcdermott.com (R. Dzhusupova), j.bosch1@tue.nl (J. Bosch), helena.holmstrom.olsson@mau.se (H.H. Olsson).

https://doi.org/10.1016/j.compchemeng.2025.109436
Received 8 July 2025; Received in revised form 8 September 2025; Accepted 3 October 2025
Available online 8 October 2025
0098-1354/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

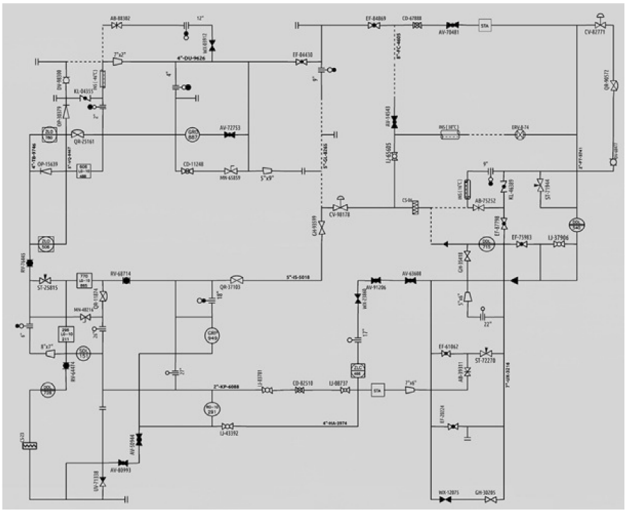

Fig. 1. A synthetic simplified P&ID from a public dataset (Paliwal et al., 2021).

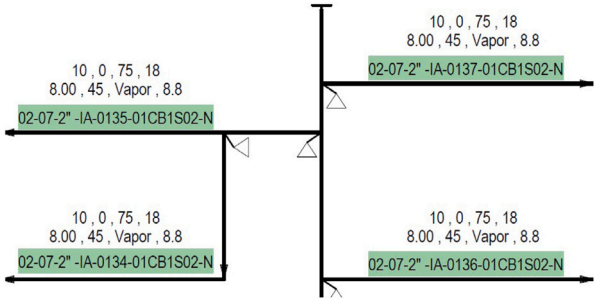

Fig. 2. A region of a real-world P&ID with highlighted pipeline numbers.

These drawings feature pipeline numbers, which encode important metadata about the pipelines represented in the drawings, such as the pipeline's dimensions, the material it is made from, fluid type, operating, and design temperature and pressure, as well as other important process information (Baron, 2010). The pipeline numbers are included in the P&IDs as text close to the pipelines, where the equipment is positioned. Fig. 2 illustrates a section of a real-world P&ID, with pipeline numbers colored in green. Furthermore, pipeline numbers are compiled into an engineering document called the "Process Line List" (Baron, 2010). This document aggregates all pipeline numbers from each P&ID within a project.

During the tender phase of EPC projects, P&ID diagrams are often shared as non-editable PDFs to protect intellectual property (Dzhusupova et al., 2022b). This turns the creation of the Line List into a manual and time-intensive task. Another common scenario involves P&IDs authored in software, which does not support metadata extraction. Additionally, brownfield projects, particularly prevalent in Europe due to space constraints that limit new construction, frequently involve scanned versions of legacy drawings created decades ago, which contain no digital text or formatting. Across all these cases, pipeline numbers are not readily available when the diagrams are received, highlighting the need for automated approaches. Deep learning-based methods can help close this gap by accelerating pipeline number extraction, reducing manual workload, and supporting more efficient project workflows.

The primary obstacle to automating this task is the profound lack of standardization in pipeline number formats in the EPC industry. There is no single universally adopted convention. Different projects may adhere to different pipeline number formats in the P&IDs, leading to variations in the encoding of metadata in the pipeline numbers. These variations may encompass differences in the length of the pipeline number, its encoded properties, and the sequence of its properties. Such inconsistency across projects and companies poses a major challenge to generalizable automation. Table 1 provides examples of line numbers from three distinct industrial EPC projects, highlighting the variations in properties and their sequences.

While previous studies have explored information extraction from engineering drawings (Paliwal et al., 2021; Saba et al., 2023; Jamieson et al., 2020; Francois et al., 2022; Kim et al., 2022; Lin et al., 2023; Villena Toro et al., 2023; Schlagenhauf et al., 2023), they consistently stop at the point of converting images to text. This approach leaves a critical semantic gap: while a system might recognize the string 150-12144-PA-031-ACB2B02SN04-N, it cannot interpret that 150 represents the nominal diameter and PA is the fluid code. The value encoded within these identifiers remains locked, and existing methods provide no mechanism for converting such strings into structured, meaningful data according to project-specific rules. This failure to bridge the gap between raw text and its engineering context is the primary reason that scalable automation has remained out of reach.

This paper bridges this semantic gap by introducing a system architecture for format-aware semantic interpretation of pipeline metadata in P&IDs. We move beyond simple OCR to create a solution that not only extracts text but also understands it in the context of variable engineering formats. Our hybrid system integrates state-of-the-art deep learning for recognition with domain-aware semantic interpretation rules. These rules are empowered by an expert-guided configuration module that enables it to perform context-based validation, correction, and semantic mapping. This manuscript is an extended version of our previous research (Shteriyanov et al., 2024), which focused solely on the detection component. Here, we present the complete, end-to-end system and rigorously validate its performance, addressing the overarching research question: **"How can AI be leveraged to automate the extraction and semantic interpretation of pipeline metadata from engineering drawings, overcoming the inherent variability of industry formatting standards to produce actionable data?"**

Main Contributions:

The main contributions of this research are the following:

- **Bridging the Semantic Gap in P&ID Data Extraction**: We are the first to bridge the critical semantic gap between raw text recognition from P&IDs and their engineering context. While prior work stops at OCR, our system performs format-aware semantic interpretation, converting non-standardized pipeline numbers into structured, actionable data.
- **A Novel, Adaptable Hybrid AI System Architecture for Semantic Understanding of Engineering Data**: We introduce an end-to-end system architecture that decouples perception from understanding, designed for the complexities of industrial P&IDs. It comprises a high-recall Pipeline Number Detector (PLN-Detector) for text localization and a unique Pipeline Number Recognizer (PLN-Recognizer) that integrates deep learning with expert-configurable rules. This design enables the recognizer to perform format validation, context-based error correction, and semantic mapping, resulting in higher accuracy than a standalone deep learning model could achieve. This architecture allows the system to adapt to new P&ID pipeline formats without any model retraining.
- **Demonstrated Real-World Viability and Economic Impact**: We validate our system on P&IDs from four distinct industrial engineering projects, each with unique formatting rules, achieving 91.1% end-to-end accuracy. Furthermore, we provide a concrete cost–benefit analysis, demonstrating the system's potential to deliver substantial financial savings and efficiency gains in large-scale EPC projects.

Table 1
Comparison of pipeline numbers from three distinct industrial EPC projects, highlighting the variations in properties and their sequences.

Line Number Format	1st column	2nd column	3rd column	4th column	5th column	6th column
P21010901-3"-A2A1-SM	Fluid Code, Unit Code, P&ID Sequence No., Line Sequence No.	Line size	Pipe Class	Insulation	N/A	N/A
150-12144-PA-031-ACB2B02SN04-N	Nominal Diameter	Unit Number	Fluid Code	Line Sequence Number	Pipe Class	Insulation
250-P-810195-03CM2SAA-HI	Nominal Diameter	Fluid Code	Unit Number, Line Sequence No.	Pipe Class	Insulation, Acoustic Insulation	N/A

This research was executed at McDermott, a global provider of engineering and construction solutions to the energy industry. The methodology utilized in the study was Action Research (Easterbrook et al., 2008), which is well-suited for investigating and solving practical problems within an organizational setting.

The structure of this paper is as follows: Section 2 offers a comprehensive review of prior studies concerning engineering drawings information extraction and identifies the research gap addressed by this study. Section 3 presents our novel solution in detail. Section 4 provides the experimental setup. Section 5 presents the results validating the performance of our solution on diverse industrial data. Section 6 discusses the broader implications and limitations of our findings. Finally, Section 7 concludes the paper.

2. Related work

In recent years, researchers and engineering companies have recognized the potential of Artificial Intelligence (AI) to revolutionize the engineering industry. In particular, AI has been applied to read or digitalize engineering drawings and automate manual tasks.

Saba et al. (2023) utilize the pre-trained Efficient and Accurate Scene Text Detector (EAST) network (Zhou et al., 2017) to detect text within P&IDs. Furthermore, the authors use EasyOCR's CRNN (Convolutional Recurrent Neural Network) model (Jaided.AI, 2021; Shi et al., 2016) to recognize the text in the resulting bounding boxes. The authors' approach achieves a detection precision of 96% and a detection recall of 95% on their testing dataset. However, they do not provide an extensive evaluation of the recognizer model.

Jamieson et al. (2020) use the same EAST text detector along with Tesseract v4 (Smith, 2007) to extract text from engineering drawings. The authors report a text detection accuracy of 90% on five P&IDs reserved for testing. Furthermore, the text recognizer is reported to capture 86% of the detected text instances. The authors also report that the EAST text detection method may truncate pipeline numbers, but have not performed an extensive evaluation.

Francois et al. (2022) employ a fine-tuned EAST network for text detection, as well as a Tesseract-based method with post-recognition correction for text recognition. The authors do not use the standard Non-maximum suppression (NMS) algorithm (Neubeck and Van Gool, 2006) typically employed by EAST methods, as it can cut off long text. Instead, they merge interlocking detections. The text detection method is reported to achieve 82% precision and 86% recall on all text instances in the authors' evaluation data, although the authors state that the method can group several text instances as a single detection. They also report that the text recognition method recognizes 82% of detected tag texts.

Kim et al. (2022) and Paliwal et al. (2021) study the digitalization of P&IDs. In both their studies, they detect text on P&IDs using a pre-trained Character-Region Awareness For Text (CRAFT) model (Baek et al., 2019). The detection method splits the P&IDs into patches, which are passed to the CRAFT detector. Afterwards, the detected texts are passed to a Tesseract OCR. Kim et al. evaluate the text extraction pipeline on 5 test industrial P&IDs and report 97.27% precision and 90.47% recall on text detection and 93.86% precision and 91.75% recall on text recognition. Paliwal et al. evaluate their text extraction method on synthetic P&IDs and report an overall text detection accuracy of 87.18% and text recognition accuracy of 79.21%. Nevertheless, it should be noted that synthetic P&IDs are less complex compared to industrial documents.

Lin et al. (2023) develop a system to reduce manual interpretation time and accurately identify basic categories of dimensions, tolerances, and functional controls in engineering drawings. The authors utilize the YOLO (You Only Look Once) object detector (Redmon, 2016) to detect various objects in 2D engineering drawings, such as symbols and text. Furthermore, they recognize the text objects using a Tesseract OCR. Overall, the system is reported to achieve nearly 70% accuracy in recognition.

Villena Toro et al. (2023) develop an OCR system capable of recognizing and differentiating between different types of information in assembly and production drawings. They utilize the CRAFT text detector along with a pre-trained CRNN text recognizer from Keras-OCR (Chollet et al., 2015). The system achieves a precision and recall of 90% in detection, and an F1-score of 94% in recognition.

Schlagenhauf et al. (2023) use a Faster-RCNN model (Ren et al., 2016) to detect text and the Kears-OCR text recognizer to reliably recognize dimensions, positions, and shape tolerances on technical drawings. Using artificially generated images in the training data, the authors achieve 81.87% detection accuracy and 79.33% recognition accuracy.

Based on the investigated literature for both general drawings and P&IDs, there is a focus on text detection and recognition, with no emphasis on semantic interpretation. These methods successfully extract a string of characters but do not provide a mechanism to parse this string into structured, meaningful data according to project-specific rules. This interpretation step is crucial for any practical engineering application, yet it is consistently overlooked.

While the engineering literature has not focused on format-aware interpretation, research in other domains, like business and legal document processing, has explored it using hybrid systems. These approaches, however, are generally designed for a pre-defined set of information and are not architected for the ad-hoc format variability common in engineering. For instance, some systems use rules to pre-select candidate text for an Large Language Model (LLM) to process (Nan et al., 2024), while others apply a fixed set of hard-coded rules to refine DL-based predictions for a specific document type, like purchase documents (Arroyo et al., 2022). A common characteristic of these hybrid systems is that their logic is static and not designed for flexible modifications.

This analysis reveals a dual research gap. Existing research in engineering drawing information extraction lacks methods for semantic interpretation, while existing hybrid systems in other domains lack the declarative adaptability required for high-variability environments. An extracted text string is only useful if it can be parsed into its constituent metadata fields (e.g., pipe size, fluid code), which is impossible without knowledge of the specific format being used. While our previous study developed a high-recall detector for this metadata (Shteriyanov et al., 2024), it did not bridge this critical interpretation gap.

To our knowledge, this paper is the first to directly address this void. We introduce a system built on a format-aware architecture that combines deep learning perception with a configurable interpretation

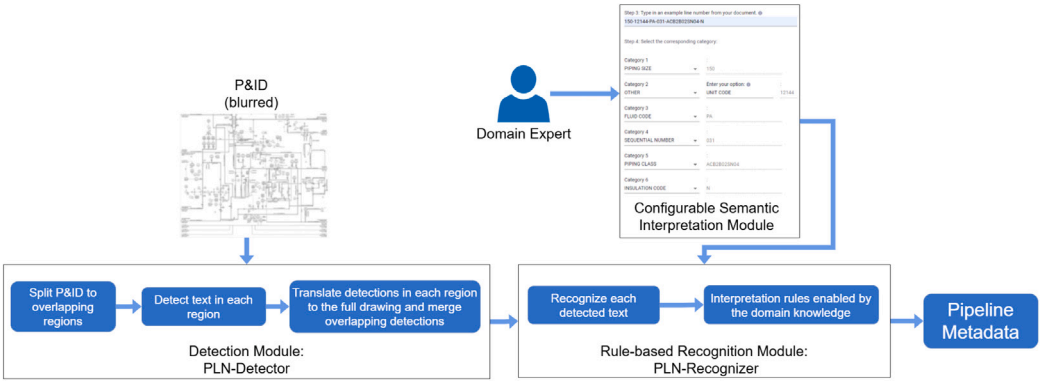

Fig. 3. Method for extraction and semantic interpretation of P&ID pipeline numbers.

module. This approach yields a solution that can understand the engineering context and is dynamically adaptable to handle the format variability, which is common in the engineering industry.

3. A method for semantic interpretation of engineering metadata: The case of P&ID pipeline numbers

A key challenge in automating the use of engineering drawings is the semantic interpretation of non-standardized metadata. This problem is particularly acute for P&IDs, where pipeline numbers encode critical data but their format can change for each new project. This variability makes a single, monolithic AI model unviable, as it would require constant, costly retraining.

The process of digitizing text from technical diagrams like P&IDs is typically a two-stage process involving text detection and text recognition. First, a text detection model analyzes the image to locate all instances of text, creating a bounding box around each one. Second, a text recognition model processes the image patch within each bounding box to transcribe the pixels into a character string. However, this standard pipeline is often insufficient for P&IDs. While it can extract raw text strings, it provides no understanding of their engineering context. For instance, a standard OCR system might correctly recognize the string "PL-1001-C03", but it would not know that this refers to a pipeline. This semantic gap is the primary challenge our work addresses.

To solve this, we propose a hybrid AI system architecture that explicitly separates perception from understanding. The system takes a full P&ID in an image format as its input. Initially, a detection module, called PLN-Detector, localizes all potential text candidates. These candidates are then processed by a recognition module, called PLN-Recognizer, which transcribes the raw text. Finally, the PLN-Recognizer interprets this text by applying a set of rules that leverage the semantic information captured by the configurable semantic interpretation module. This decoupled design is the key to adaptability, allowing the system to learn new formats without modifying the core deep learning models. The entire end-to-end process is visualized in Fig. 3.

3.1. Detection module: The PLN-detector

The process begins by finding the location of all potential pipeline number candidates on a P&ID image. To maximize recall and ensure no true pipeline numbers are missed, this operation is performed by our detection module called PLN-Detector. PLN-Detector utilizes the Progressive Scale Expansion Network (PSENet) (Wang et al., 2019). We chose PSENet due to its demonstrated strength in distinguishing closely positioned text instances and detecting text in various orientations. Its progressive scale expansion algorithm effectively separates closely packed text instances, a feature crucial for accurately isolating line numbers from adjacent symbology and annotations in engineering drawings. While alternative models like Differentiable Binarization Network (DBNet) (Liao et al., 2020) and Fast Oriented Text Spotting (FOTS) (Liu et al., 2018) are well-regarded, comparative studies in other domains have shown PSENet to achieve superior recall performance (Yao et al., 2025; Lu et al., 2022), which is the primary metric for our task to minimize missed pipeline numbers.

To further enhance performance on full-sized P&IDs, the PLN-Detector incorporates an overlapping tiling technique (Ozge Unel et al., 2019). During preprocessing, each P&ID is split into overlapping tiles, which are then processed by the detector. In post-processing, the detected text regions from each tile are translated back to their original coordinates on the full P&ID, and overlapping detections are merged into single bounding boxes. The output of the PLN-Detector is a list of bounding boxes, where each box is represented by a set of coordinates (x_min, y_min, x_max, y_max) corresponding to a detected text candidate on the original P&ID. The PSENet model within this module was fine-tuned on a large corpus of industrial P&IDs to learn the specific visual characteristics of text in this domain.

3.2. Configurable semantic interpretation module

Our method's adaptability is powered by a standalone component that allows a domain expert to define the interpretation logic for a new project. The configurable semantic interpretation module, a web-based GUI shown in Fig. 4, is designed for minimalist, "one-shot" configuration. To configure the system, the user provides two pieces of information:

- An Example Format: The user provides a single, representative pipeline number (e.g., 250-P-810195-03CM2SAA-HI). From this, the system automatically derives a structural template, including the number of components and their expected character patterns.
- Semantic Labels: The user assigns a meaning (e.g., piping size, fluid code, piping class) to each component of the format.

This minimal input is the crucial link that empowers the recognition module. The output from this module is a "semantic template" structured as a JavaScript Object Notation (JSON) object. This template contains the example format, the number of expected components, and for each component, its assigned semantic label. By externalizing the project-specific logic into this configuration step, the core deep learning modules remain generic and reusable, while the overall method achieves high accuracy and flexibility across diverse and non-standardized project requirements.

3.3. Recognition module: The PLN-recognizer

The third component, the PLN-Recognizer, receives three inputs: the P&ID image, the list of candidate bounding boxes from the PLN-Detector, and the semantic template generated by the configuration

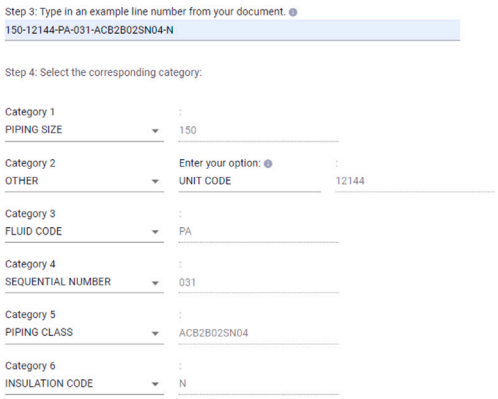

Fig. 4. The configurable semantic interpretation module. By providing a single example, an expert can define the interpretation rules for an entire project.

module. First, it crops the candidate text regions from the image. Afterwards, it performs text recognition on each region to get a raw string. Then, it applies a set of project-specific rules to interpret that string. The PLN-Recognizer executes the following steps for each candidate region it receives:

- Core Text Recognition: The core recognition engine is built upon a pre-trained backbone from PP-OCRv4 (PaddlePaddle-Optical Character Recognition version 4), a practical OCR toolkit (Li et al., 2022). Specifically, we use the SVTR-LCNet model, which combines the Scene Visual Text Recognition (SVTR) architecture with a Lightweight CPU Network (LCNet) as its visual backbone for efficient feature extraction (Du et al., 2022; Cui et al., 2021). The primary role of this engine is to produce the initial textual transcription for each candidate bounding box. This architecture was selected for its strong baseline performance on complex and varied character sets found in technical documents.
- Rule-based Interpretation: The recognized text is then refined by the recognizer's interpretation rules involving: (a) Filtering any text strings that do not match the expected pipeline number format; (b) Semantic Mapping into meaningful fields (e.g., "Nominal Diameter", "Fluid Code"); (c) Correction rules to fix common OCR errors based on mapped semantic meaning (e.g., mistaking O for 0, inserting missing hyphens);

The final output of this module, and the system as a whole, is an Excel file, listing each pipeline number and its semantically interpreted components.

3.3.1. *Rule-based interpretation*

The PLN-Recognizer includes a rule-based interpretation stage that is implemented in Python and employs an algorithmic, example-based template matching approach. The process begins with filtering false positives. To achieve this, all non-alphanumeric characters are first stripped from both the OCR string and the user-provided example format. A candidate string is considered a structural match if its stripped version matches an example's stripped version in two aspects: (1) it has the same length, and (2) it satisfies a character-type plausibility check. This check is designed to be tolerant of common OCR ambiguities (e.g., O/0 or I/1), which are addressed in a later semantic correction step.

If a valid structural match is found, the system uses the corresponding example as a template to reconstruct the OCR string. If the recognized pipeline number lacks separators (e.g., "12A5" instead of "12-A-5"), the recognizer inserts hyphens according to the expected segment lengths. Furthermore, we apply a set of character-level correction rules derived from common OCR misrecognitions observed during

Ground-Truth: 035-50-BD-0012-01CB1S01-N
Prediction: 035-50>BD-001201CB1S01-N

Fig. 5. Example of common errors produced by the SVTR-LCNet network, where missed and misrecognized hyphens in the prediction are highlighted in the ground-truth.

analysis by domain experts. These rules leverage structural knowledge of pipeline numbers. The ">" symbol, which is often misrecognized, is replaced with "-" where hyphens are expected. This confusion may be due to the presence of occlusions or font style. Fig. 5 shows an example of both of these errors produced by the base SVTR-LCNet network in "PLN-Recognizer", highlighting character differences between the ground truth and the prediction. In the prediction shown, the character ">" should be corrected to a "-", and an additional "-" should be inserted between the "2" (at position 14) and the "0" (at position 15) in the predicted string. Misrecognized or missing separators are among the issues addressed by the correction rules in our approach. A possible hypothesis for some of these errors is the potential presence of visual elements appearing on top of the pipeline number (occlusions) and limitations of the base SVTR-LCNet network.

Furthermore, the strings' components are mapped to specific pipeline metadata types. The supported component types for pipeline numbers include: "sector code", "system code", "piping size", "fluid code", "sequential number", "piping class", "insulation code", "site", "main area", "sub area", "other" (custom-defined). Based on the selected type and representative examples, the recognizer can infer the expected format of each component (numeric, alphabetic, or alphanumeric).

In components that are expected to contain only digits, such as piping size or sequential number, misrecognized letters are corrected by replacing them with the most likely digit. For example, replacing "O" with "0". A similar correction is applied to fields that contain only letters, where digits are replaced with their alphabetic counterparts based on context. The system also allows certain special characters (e.g., #) in digit-based fields when appropriate, particularly in piping size.

For components with ambiguous alphanumeric content, such as piping class or fluid code, where no consistent structure can guide corrections, the system does not apply any modifications.

These rules are designed not only to improve the accuracy of pipeline number recognition but also to remain flexible across different P&ID formats and conventions. Because the correction logic is driven by the injected domain knowledge, it can be adapted to any project structure.

The rules are summarized in Table 2. These rules were developed through a systematic empirical error analysis process. This involved running the base recognition model on a development set of P&IDs and meticulously comparing the raw OCR output against the ground truth labels. Recurring error patterns, such as the consistent misrecognition of > for – or the appearance of O in numeric-only fields like piping size, were identified. These empirical observations were then formalized into deterministic rules in collaboration with domain experts, who provided the contextual knowledge to confirm which corrections were safe and universally applicable within this domain.

4. Experimental setup

To validate our method, this section details the experimental setup used to evaluate our detection, recognition, and end-to-end performance. We outline the baseline methods used for comparison, the datasets, and the model training to ensure a rigorous assessment via real-world industrial P&IDs.

Table 2
PLN-Recognizer filtering and correction rules.

Rule type	Example	Correction	Applicability
Replace '>' separators with '–'	12>A-5 instead of 12-A-5	Replace '>' with '–' where hyphens are expected	All projects (general rule)
Missing separator insertion	12A5 instead of 12-A-5	Insert - based on expected segment lengths	All projects (general rule) given the example format
Filter on the number of components	Detection: 5 components Expected: 6 components	Discard if not matching expected number of components	All projects (general rule) given the example format
Letter-only component fix	1 in alphabetic field	Replace digit with most likely letter (e.g., replace 1 with I)	e.g., insulation code
Digit-only component fix	O in a numeric field	Replace letter with most likely digit (e.g., replace O with 0)	e.g., piping size, sequential number
Special character allowance	" or # in otherwise digit-only fields	Accept ", #, x in otherwise digit-only fields	e.g., piping size
No correction possible	Ambiguous alphanumeric text	Skip correction if no structure can guide it	e.g., piping class, fluid code, other

Table 3
Overview of project and P&ID distribution for training and evaluation.

Purpose	Number
Total Projects Used	21
Training Projects	17
Evaluation Projects	4
Training P&IDs	355
Evaluation P&IDs	40

4.1. Baselines

To benchmark our method's components, we selected baseline methods representing common approaches in the literature for text extraction from engineering drawings. For detection, we compared against two methods inspired by existing P&ID research (Francois et al., 2022; Kim et al., 2022). The first, EAST with Fusion NMS, is based on the approach in Francois et al. (2022) and uses an industrial-trained EAST model with an adjusted Non-Maximum Suppression (NMS) algorithm (Neubeck and Van Gool, 2006) to fuse bounding boxes. The second, CRAFT with Tiling, follows the method in Kim et al. (2022), which divides the P&ID into patches and processes each with a pre-trained CRAFT text detector.

For recognition, we benchmarked against two widely used standard OCR tools (Paliwal et al., 2021; Saba et al., 2023; Jamieson et al., 2020; Francois et al., 2022; Kim et al., 2022; Lin et al., 2023) to assess the performance of general-purpose models on this specialized task. We selected TesseractOCR, a popular engine based on Long Short-Term Memory (LSTM) neural networks (Smith, 2007; Hochreiter, 1997), and easyOCR, which implements a CRNN model composed of a Residual Network (ResNet) feature extractor and an LSTM network (Jaided.AI, 2021; He et al., 2016).

4.2. Data collection and preparation

To train and evaluate our method, we used a large dataset of industrial P&IDs from 21 past EPC projects executed by McDermott. These documents were converted from their original PDF format to images. The distribution of projects and P&IDs is summarized in Table 3.

4.2.1. Training data for text detection

The training set for the detection module consisted of 355 P&IDs from 17 distinct projects. Domain experts annotated the bounding box coordinates for all text instances within these documents.

4.2.2. Evaluation data

The evaluation dataset consists of 40 industrial P&IDs originating from 4 distinct projects. The four projects were excluded from the sampling process when selecting P&IDs for the training data. The projects will be referred to as projects A, B, C, and D in the rest of this manuscript. We randomly chose 10 P&IDs from each of the four projects. Each project employed a unique pipeline number format. The size of this evaluation set is consistent with or exceeds that used in related studies (Jamieson et al., 2020; Francois et al., 2022; Kim et al., 2022).

For evaluation, two sets of ground-truth annotations were created. For the detection task, the bounding boxes of all pipeline numbers were annotated. For the end-to-end recognition task, individual image crops of each pipeline number were generated, each paired with a text file containing its ground-truth string.

While public synthetic P&ID datasets such as the one from Paliwal et al. (2021) are available, we deliberately chose to use this real-world dataset for our evaluation. Our analysis indicated that the pipeline numbers in the synthetic data are generally shorter and exhibit less structural complexity, and therefore would not adequately test the robustness of our format-aware semantic interpretation module. The curated industrial data, in contrast, provides a more challenging and meaningful benchmark for the specific task of interpreting long, multi-component pipeline numbers.

4.3. Training details

Both the PLN-Detector (our method) and the EAST with Fusion NMS (baseline) were trained on the industrial P&ID dataset using a single GPU with 16 GB of VRAM (Video Random-Access Memory). For the PLN-Detector, the training data was processed using a tiling technique, generating 4338 tiles of 860 × 860 pixels with a 200-pixel overlap. Its underlying PSENet model was trained for 40 epochs with the Adam optimizer, a learning rate of 0.001, and a batch size of 8. The baseline EAST model was trained on the full images for 35 epochs, also with the Adam optimizer and a batch size of 8, but with a learning rate of 0.0001. These hyperparameters were chosen empirically for optimal convergence, with the batch size limited by available GPU memory.

4.4. Evaluation method

The method's performance was evaluated at three stages: detection, recognition, and end-to-end. First, the detection recall was calculated to measure the percentage of pipeline numbers correctly located by the detection module, as shown in Eq. (1). A detection was considered correct only if its bounding box fully enclosed the pipeline number without truncating characters or including adjacent text.

Our focus on recall for this initial stage is a deliberate choice rooted in our system's two-stage architecture, which explicitly separates the responsibilities for recall and precision. The detection module is optimized to act as a comprehensive candidate generator, maximizing recall to ensure no true pipeline numbers are missed. The subsequent rule-based interpretation logic is then designed to act as a powerful filter, enforcing precision by discarding false positives that do not match the project-specific format template. Consequently, the precision of the detection stage alone is not an informative measure of the system's final ability to reject incorrect detections.

$$Recall = \frac{\text{Number of Correctly Detected Pipeline Numbers}}{\text{Number of Pipeline Numbers in Ground Truth}} \quad (1)$$

Second, the recognition accuracy of the standalone OCR models was measured using the ground-truth image crops of pipeline numbers. This metric evaluates the raw transcription performance of the recognizers in isolation. For this evaluation, a pipeline number was considered correctly recognized if the transcribed text string was an exact, character-for-character match with the corresponding ground-truth label. The accuracy was then calculated as shown in Eq. (2).

$$\text{Accuracy} = \frac{\text{Number of Correctly Recognized Pipeline Numbers}}{\text{Number of Pipeline Numbers in Ground Truth}} \quad (2)$$

Finally, the end-to-end accuracy of the complete, integrated method was calculated using the original 40 evaluation P&IDs, as shown in Eq. (3). For this end-to-end evaluation, a pipeline number is counted as correct only if it is successfully detected on the drawing and its final, semantically corrected text string is an exact, character-for-character match with the ground-truth label.

$$\text{End-to-End Accuracy} = \frac{\text{Correctly Interpreted Pipeline Numbers}}{\text{Total Pipeline Numbers in Ground Truth}} \quad (3)$$

5. Results

This section presents the empirical results of our study. We first benchmark the performance of the core detection (PLN-Detector) and recognition (PLN-Recognizer) modules against established baselines. We then evaluate the end-to-end accuracy of the complete method to demonstrate its practical effectiveness on our multi-project industrial dataset.

5.1. Evaluation of the detection module

The results for each pipeline number detection method across different projects are presented in Fig. 6. Overall, our proposed method, "PLN-Detector" achieved an impressive overall 95.14% recall across all projects, indicating its reliable capability to detect pipeline numbers in various formats. In contrast, "EAST with Fusion NMS" and "CRAFT with Tiling" delivered overall recall rates below 50%. Specifically, "EAST with Fusion NMS" missed most of the pipeline numbers in project B, while "CRAFT with Tiling" missed most in project C. In projects A and D, these two methods correctly identified approximately half of the ground-truth pipeline numbers.

Upon analyzing the errors, we found that the "EAST with Fusion NMS" method produced truncated bounding boxes. In Project D, this method produced bounding boxes, capturing other text or symbology. Fig. 7 illustrates detections by "EAST with Fusion NMS" (a) and the corresponding detections by our proposed method, "PLN-Detector" (b). Moreover, the "CRAFT with Tiling" method generated bounding boxes for a single pipeline number. Fig. 8 shows detections by "CRAFT with Tiling" (a) and corresponding detections by "PLN-Detector" (b).

Our proposed method, PLN-Detector, occasionally truncated the edges of pipeline numbers or produced multiple bounding boxes for missed pipeline numbers. However, these instances were limited, and the method successfully captured most target pipeline numbers. Based on these findings, the PLN-Detector method proved to be reliable for pipeline number detection.

Table 4
Accuracy results of the end-to-end system on the four evaluation projects.

	Project A	Project B	Project C	Project D	Overall
Accuracy	83.20%	97.80%	91.30%	88.20%	91.10%

5.2. Evaluation of the recognition module

The recognition accuracy for each pipeline number recognition method across the various projects is illustrated in Fig. 9. As illustrated, our proposed PLN-Recognizer method exhibited the highest accuracy in all projects. Importantly, it achieved 95.7% overall accuracy, assuming a perfect detector, which showcases the strong generalization of the method to the pipeline number formats across the four diverse projects.

The 4.3% inaccurate recognition cases from the proposed PLN-Recognizer with correction rules are attributed to character errors in pipeline number components, such as piping class and fluid code, where it is challenging to infer the correct character, and correction rules cannot be applied.

An example of an error is given in Fig. 10. In the example below, a "0" was predicted as "Q". The component "WR00006" consist of both letters and digits. In such components we do not apply correction rules.

Conversely, the CRNN recognizer implemented in EasyOCR yielded the worst results. A high frequency of errors was observed in almost all the evaluation pipeline numbers. These errors included omitted hyphen (-) characters and incorrectly substituted characters (such as 8 instead of B, 0 instead of D, O instead of 0, I instead of 1, and G instead of 6). Similar errors were observed with the use of Tesseract OCR, although they were less prevalent. Examples of these errors are given in Fig. 11, where the missed and misidentified characters are highlighted in the ground-truth.

Importantly, even without applying correction rules, the PLN-Recognizer significantly outperformed Tesseract and EasyOCR in terms of accuracy, demonstrating the robustness of the underlying model. The no-rules variant consistently achieved higher recognition accuracy across most projects. However, in Project A, the accuracy of the no-rules variant dropped significantly to 40%, largely due to systematic misrecognitions involving hyphen characters, such as misreading "-" as ">" or omitting them altogether. This highlights the value of correction rules in normalizing such recurring formatting issues and recovering the correct structure. When these rules are applied, the accuracy for Project A increases dramatically, demonstrating their effectiveness in handling project-specific conventions.

The additional benefit of applying correction rules is evident when comparing both versions of our method. Across all projects, the accuracy improved from 81.4% (without correction rules) to 95.7% (with rules). These rules are particularly valuable in correcting common recognition errors.

5.3. End-to-end performance

The proposed method was evaluated as an end-to-end system on each of the four evaluation projects was carried out by a senior process engineer. The results, combining all these techniques, are shown in Table 4. The engineer investigated the recognized text in the detected pipeline numbers. The accuracy was calculated by finding the percentage of correctly detected and recognized pipeline numbers out of all pipeline numbers in the P&IDs.

The lower performance of our method on Project A can be attributed to a few key challenges related to the filtering of truncated pipeline numbers and character errors by PLN-Recognizer in pipeline number components, where correction rules cannot be applied. There were several instances where the detector truncated pipeline numbers. An example of a filtering error is shown in Fig. 12. Since the number of components in these predictions did not match the expected format,

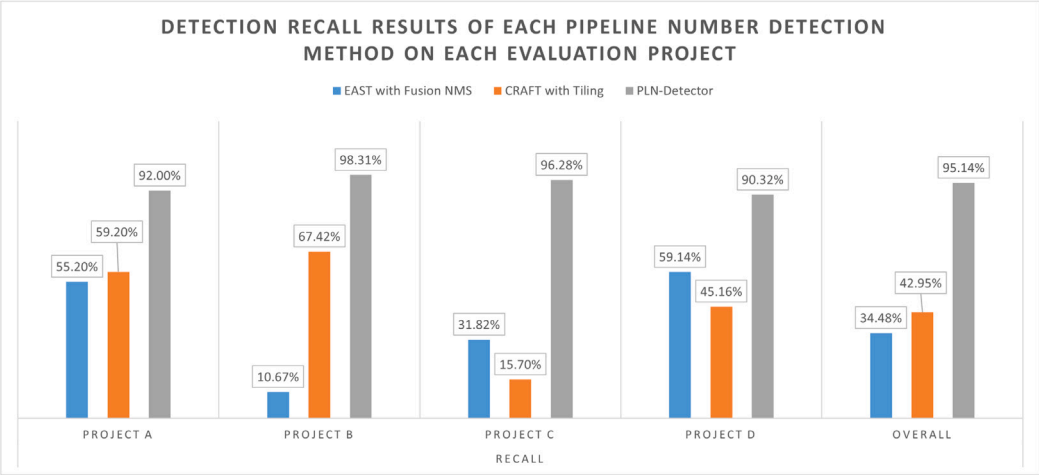

Fig. 6. Recall results of each pipeline number detection method on each evaluation project.

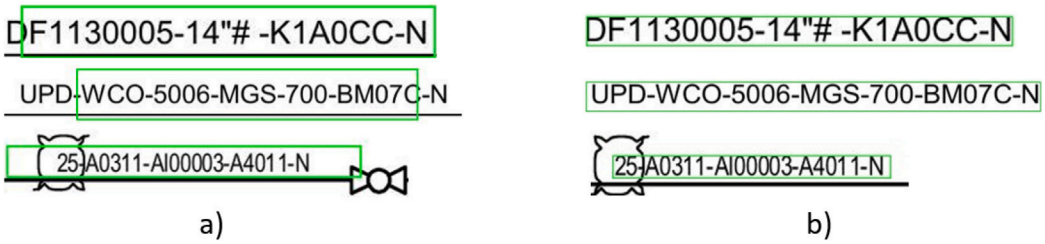

Fig. 7. Examples of detections produced by the "EAST with Fusion NMS" method (a) and the corresponding detections produced by our proposed "PLN-Detector" method (b).

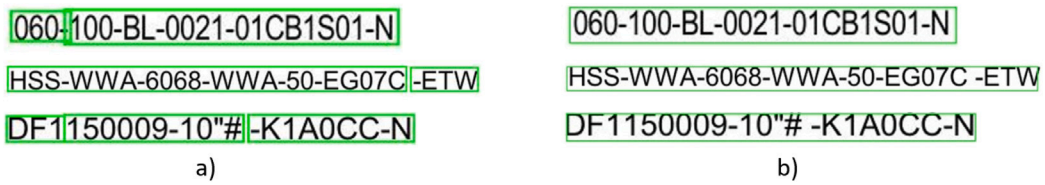

Fig. 8. Examples of detections produced by the "CRAFT with Tiling" method (a) and the corresponding detections produced by our proposed "PLN-Detector" method (b).

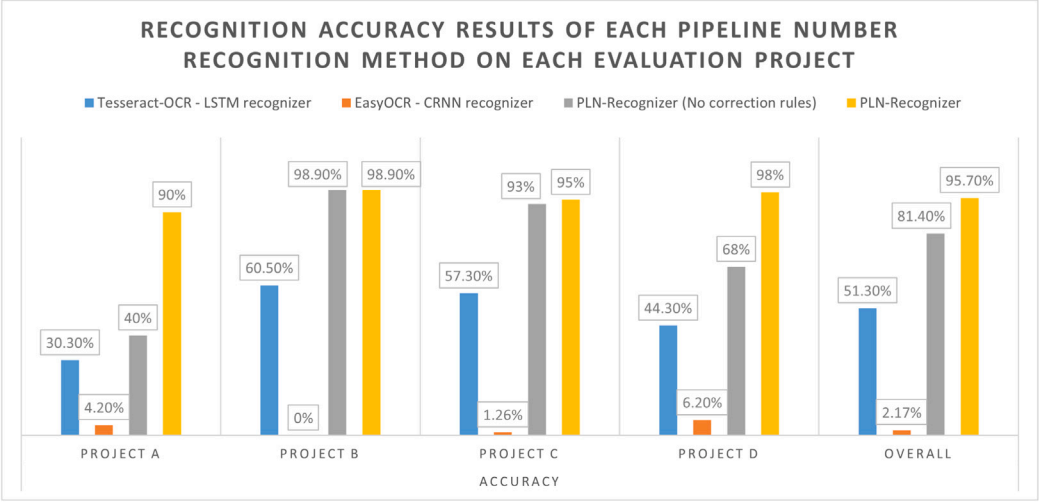

Fig. 9. Accuracy results of each pipeline number recognition method on each evaluation project.

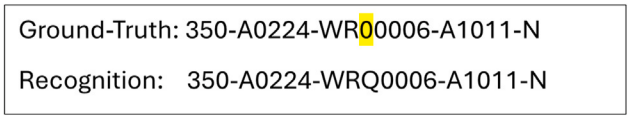

Fig. 10. Example of an error produced by PLN-Recognizer with correction rules.

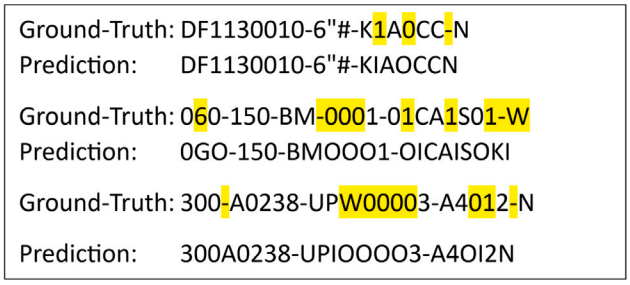

Fig. 11. Example of common errors produced by the "TessearctOCR-LSTM" and "easyOCR-CRNN" recognition methods, where character differences between ground-truth and prediction are highlighted.

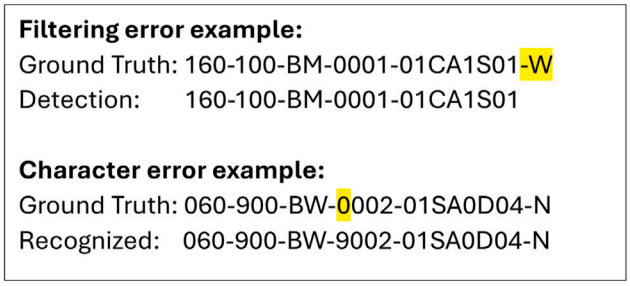

Fig. 12. Example of errors produced by PLN-Recognizer in Project A, related to filtering truncated pipeline number detections and character errors.

part of these results were filtered by the PLN-Recognizer. Additionally, in terms of recognition, there were cases where incorrect characters appeared in components, where correction rules could not be applied as discussed in Section 3.3.1. Fig. 12 also shows an example of a character error. In this case, the third component consisted entirely of digits, so there are no rules to correct the "9" to "0". These issues highlight the importance of considering both the detection and recognition phases during the extraction.

Although the accuracy score in Project A is lower than in the other projects, engineers will review and potentially correct the results. This is also the norm for manually produced documents. Thus, even in such cases, our pipeline will help automate the initial generation and save engineering hours.

6. Discussion

This study solves the problem of format-aware semantic interpretation of pipeline metadata from P&IDs. Unlike previous studies that focused only on general text extraction (Paliwal et al., 2021; Saba et al., 2023; Jamieson et al., 2020; Francois et al., 2022; Kim et al., 2022; Lin et al., 2023; Villena Toro et al., 2023; Schlagenhauf et al., 2023), our work provides an end-to-end method for achieving format-aware semantic interpretation of the extracted text. Furthermore, in contrast to the hybrid systems reviewed in other domains with static rules pre-programmed for a specific task (Nan et al., 2024; Arroyo et al., 2022), our method allows for user-driven adaptability. The high-level format definition provided by a process engineer via the configurable semantic interpretation module empowers our system to perform format validation and context-aware correction to new ad hoc formats without model retraining.

Our system architecture strategically combines deep learning with a rule-based interpretation enabled by an expert-guided configuration process. This design allows a domain expert to inject project-specific knowledge into the system, which enables the system to process new formats without retraining. This knowledge empowers our format-aware recognition module to perform context-specific validation, correction, and semantic mapping. The 91.1% end-to-end accuracy achieved on P&ID pipeline metadata extraction and interpretation serves as strong validation of this method.

It is important to frame this result within the context of real-world industrial workflows. In safety-critical domains, a final human verification step is mandatory to ensure 100% accuracy in all engineering documentation. Therefore, the goal of an automated system is not full autonomy but human augmentation. A 91% accurate initial draft provides immense value by transforming the engineer's task from laborious data entry to efficient validation and correction. Given that P&IDs can contain hundreds of pipeline numbers, this level of automation significantly reduces manual effort and potential for human error. While our system's scope is focused on pipeline numbers, its performance demonstrates a critical step towards the reliable digitization of entire P&IDs within a practical, human-in-the-loop framework.

Furthermore, our method offers a pragmatic solution to the lack of industry-wide ontologies for many types of technical data. Instead of waiting for universal standards to emerge, this on-demand semantic interpretation is a crucial enabler for digitalization initiatives in the process industries, particularly for creating reliable "as-is" digital twins from vast archives of legacy documentation. This "expert-in-the-loop" design also aligns with emerging regulatory principles, such as those in the EU AI Act (Union, 2024), which emphasize human oversight in high-risk AI applications, particularly within the safety-critical domain of chemical plant design and operation. This deliberate feature ensures both practical utility and responsible deployment.

Our evaluation demonstrates the effectiveness of our system against a strong open-source baseline. We deliberately did not include benchmarks against commercial cloud OCR services, despite their strong general performance. This decision was guided by stringent client confidentiality agreements that prohibit the use of their data with third-party services. Obtaining the necessary client approval for internal research purposes is often not feasible, making cloud-based experiments impractical. For this reason, a foundational requirement was the development of a solution that is entirely developed in-house and hosted within the company's secure infrastructure. Thus, our work presents a practical solution designed to meet these non-negotiable data security and client confidentiality constraints.

It is also important to note that while our rule-based interpretation module is designed to be model-agnostic, the choice of the underlying recognition engine plays a crucial role in the system's overall performance. Our evaluation shows that the base PLN-Recognizer, even without the post-processing rules, achieves a higher accuracy than the Tesseract and EasyOCR baselines on our dataset. This superior baseline is particularly critical for complex alphanumeric components where semantic correction rules offer limited benefit. Therefore, while applying our rule-based module to other OCR engines would likely boost their performance, our system's superior results are attributable to both the advanced recognition backbone and the domain-specific post-processing logic. A detailed ablation study analyzing the interplay between different recognition backbones and our semantic module would be a valuable direction for future research.

6.1. Cost-benefit analysis

The practical value of our system is best understood within the context of industrial engineering workflows. In domains where operational

safety is paramount, final documentation requires 100% accuracy. Consequently, a final human verification and sign-off stage is a mandatory and constant requirement for all deliverables, regardless of whether the initial draft was generated manually or by an automated system. The primary benefit of our method, therefore, is not to replace this crucial review step but to automate the laborious and error-prone initial data entry phase.

To quantify this benefit, we analyze the cost of the manual creation process for a typical engineering office. We use the following conservative parameters: an annual workload of 5 large scale projects, with 1000 P&IDs each undergoing an average of 3 revisions. We estimate the manual labor for the initial data entry, which involves locating, transcribing, and entering all pipeline numbers from a P&ID, to be a blended average of 1 h per P&ID. With an engineering rate of $100 per hour, the resulting annual cost for this manual creation phase is substantial, as calculated in Eq. (4).

$$5projects * 1000P\&IDs * 3revisions * 1h * 100\$ = 1,500,000\$ \quad (4)$$

Our system, with its 91% accuracy, automates the creation of this initial draft. While the subsequent human review time remains constant, the elimination of the initial manual labor, as quantified in Eq. (4), constitutes the primary cost saving. The engineer's task is transformed from one of tedious creation followed by review, to one of efficient validation and correction of a prepopulated draft. This analysis highlights the significant financial incentive for automation by demonstrating the tangible economic impact of reducing manual effort. Furthermore, this estimate does not account for potential secondary benefits, such as providing highlighted P&IDs to the engineer, which could further streamline the unchanged human review process.

6.2. Limitations

A primary limitation is the confidential nature of our industrial dataset, which cannot be made public due to intellectual property restrictions, thereby limiting direct reproducibility. To mitigate this, we have described our method in sufficient detail to allow other organizations to implement and validate our approach on their own proprietary data, assessing its applicability within their specific contexts. Another limitation is the potential for unintentional researcher bias. Our close involvement was necessary to access the proprietary industrial data but presents a risk of subjective influence. Therefore, independent replication of this study is crucial to validate our findings and provide an external check against this potential bias.

6.3. Future work

While this study successfully demonstrates the viability of our method, it also opens several promising avenues for future research. Future research should focus on validating the method's design by applying it to other types of engineering drawings, such as Isometrics or General Arrangement Drawings (GADs), and to entirely different fields like legal or financial analysis, where format variability is also a key challenge. Additionally, the methods's components could be further optimized. For example, a systematic evaluation of alternative state-of-the-art text detection architectures, such as DBNet (Liao et al., 2020) and FOTS (Liu et al., 2018), would provide valuable insights into the optimal choice for processing complex industrial documents.

Another direction for future work is to extend our methodology to other critical information on P&IDs, particularly instrumentation and equipment tags. While this was outside the scope of the current study, which focused on the more complex challenge of long-format pipeline numbers, our preliminary analysis has already confirmed that the base detection and recognition models effectively capture these shorter tags. We are therefore confident that our configurable interpretation module can be readily adapted to interpret their specific formats, demonstrating the broader applicability of our expert-in-the-loop framework for comprehensive P&ID digitization.

7. Conclusion

This study addressed a fundamental challenge in process systems engineering: the automated extraction of metadata from P&IDs that lack consistent, standardized formats. To solve this, we proposed a novel hybrid AI system architecture that strategically decouples perception from understanding. It uses a high-recall detection module to capture text candidates, and a unique recognition module that is empowered by expert-defined rules to perform the semantic interpretation. This design allows the method to be adapted to new formats without costly model retraining, as it decouples the core AI models from the variable, project-specific rules they must operate on. Our proposed system was rigorously evaluated on real-world data from four distinct projects and achieved a high end-to-end accuracy of 91.1%. These results confirm our method is a highly effective and scalable solution for complex process engineering environments. This work provides a practical pathway for unlocking valuable data from legacy documentation, directly supporting digitalization and digital twin initiatives within the process industries.

CRediT authorship contribution statement

Vasil Shteriyanov: Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Rimma Dzhusupova:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Jan Bosch:** Writing – review & editing, Supervision, Conceptualization. **Helena Holmström Olsson:** Writing – review & editing, Supervision.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used the GPT-4o model in order to improve the readability of the text. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Vasil Shteriyanov reports a relationship with McDermott International Inc that includes: employment and travel reimbursement. Rimma Dzhusupova reports a relationship with McDermott International Inc that includes: employment and travel reimbursement. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by McDermott Inc. and Software Center (Gothenburg, Sweden) and conducted in collaboration with Eindhoven University of Technology, The Netherlands.

Data availability

The data that has been used is confidential.

References

Arroyo, R., Yebes, J., Martínez, E., Corrales, H., Lorenzo, J., 2022. Key information extraction in purchase documents using deep learning and rule-based corrections. arXiv preprint arXiv:2210.03453.

Baek, Y., Lee, B., Han, D., Yun, S., Lee, H., 2019. Character region awareness for text detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9365–9374.

Baron, H., 2010. The Oil and Gas Engineering Guide. Editions Technip.

Berends, K., 2007. Engineering and construction projects for oil and gas processing facilities: Contracting, uncertainty and the economics of information. Energy Policy 35 (8), 4260–4270.

Chollet, F., et al., 2015. Keras. https://github.com/fchollet/keras, (Accessed 06 December 2024).

Cui, C., Gao, T., Wei, S., Du, Y., Guo, R., Dong, S., Lu, B., Zhou, Y., Lv, X., Liu, Q., et al., 2021. PP-LCNet: A lightweight CPU convolutional neural network. arXiv preprint arXiv:2109.15099.

Du, Y., Chen, Z., Jia, C., Yin, X., Zheng, T., Li, C., Du, Y., Jiang, Y.-G., 2022. Svtr: Scene text recognition with a single visual model. arXiv preprint arXiv:2205.00159.

Dzhusupova, R., Banotra, R., Bosch, J., Olsson, H.H., 2022a. Pattern recognition method for detecting engineering errors on technical drawings. In: 2022 IEEE World AI IoT Congress. AIIoT, IEEE, pp. 642–648.

Dzhusupova, R., Bosch, J., Olsson, H.H., 2022b. Challenges in developing and deploying AI in the engineering, procurement and construction industry. In: 2022 IEEE 46th Annual Computers, Software, and Applications Conference. COMPSAC, IEEE, pp. 1070–1075.

Dzhusupova, R., Shteriyanov, V., Bosch, J., Olsson, H.H., 2025. Smart material estimation for the engineering, procurement, and construction (EPC) sector. Results in Eng. 105802.

Easterbrook, S., Singer, J., Storey, M.-A., Damian, D., 2008. Selecting empirical methods for software engineering research. Guid. Adv. Empir. Softw. Eng. 285–311.

Francois, M., Eglin, V., Biou, M., 2022. Text detection and post-OCR correction in engineering documents. In: International Workshop on Document Analysis Systems. Springer, pp. 726–740.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hochreiter, S., 1997. Long short-term memory. Neural Computation MIT-Press.

Jaided.AI, 2021. Easyocr. https://www.jaided.ai/easyocr, (Accessed 06 December 2024).

Jamieson, L., Moreno-Garcia, C.F., Elyan, E., 2020. Deep learning for text detection and recognition in complex engineering diagrams. In: 2020 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–7.

Kim, B.C., Kim, H., Moon, Y., Lee, G., Mun, D., 2022. End-to-end digitization of image format piping and instrumentation diagrams at an industrially applicable level. J. Comput. Des. Eng. 9 (4), 1298–1326. http://dx.doi.org/10.1093/jcde/qwac056.

Li, C., Liu, W., Guo, R., Yin, X., Jiang, K., Du, Y., Du, Y., Zhu, L., Lai, B., Hu, X., et al., 2022. PP-OCRv3: More attempts for the improvement of ultra lightweight ocr system. arXiv preprint arXiv:2206.03001.

Liao, M., Wan, Z., Yao, C., Chen, K., Bai, X., 2020. Real-time scene text detection with differentiable binarization. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, (07), pp. 11474–11481.

Lin, Y.-H., Ting, Y.-H., Huang, Y.-C., Cheng, K.-L., Jong, W.-R., 2023. Integration of deep learning for automatic recognition of 2d engineering drawings. Machines 11 (8), 802.

Liu, X., Liang, D., Yan, S., Chen, D., Qiao, Y., Yan, J., 2018. Fots: Fast oriented text spotting with a unified network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5676–5685.

Lu, M., Leng, Y., Chen, C.-L., Tang, Q., 2022. An improved differentiable binarization network for natural scene street sign text detection. Appl. Sci. 12 (23), 12120.

Nan, H., Marx, M., Wolswinkel, J., 2024. Combining rule-based and machine learning methods for efficient information extraction from enforcement decisions. In: Legal Knowledge and Information Systems. IOS Press, pp. 321–326.

Neubeck, A., Van Gool, L., 2006. Efficient non-maximum suppression. In: 18th International Conference on Pattern Recognition. ICPR'06, vol. 3, IEEE, pp. 850–855.

Ozge Unel, F., Ozkalayci, B.O., Cigla, C., 2019. The power of tiling for small object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

Paliwal, S., Jain, A., Sharma, M., Vig, L., 2021. Digitize-PID: Automatic digitization of piping and instrumentation diagrams. In: Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2021 Workshops, WSPA, MLMEIN, SDPRA, DARAI, and AI4EPT, Delhi, India, May 11, 2021 Proceedings 25. Springer, pp. 168–180.

Park, M.-J., Lee, E.-B., Lee, S.-Y., Kim, J.-H., 2021. A digitalized design risk analysis tool with machine-learning algorithm for EPC contractor's technical specifications assessment on bidding. Energies 14 (18), 5901.

Redmon, J., 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: Towards real-time object detection with region proposal networks. arXiv:1506.01497.

Saba, A., Hantach, R., Benslimane, M., 2023. Text detection and recognition from piping and instrumentation diagrams. In: 2023 8th International Conference on Image, Vision and Computing. ICIVC, IEEE, pp. 711–716.

Schlagenhauf, T., Netzer, M., Hillinger, J., 2023. Text detection on technical drawings for the digitization of brown-field processes. Procedia CIRP 118, 372–377.

Shi, B., Bai, X., Yao, C., 2016. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Trans. Pattern Anal. Mach. Intell. 39 (11), 2298–2304.

Shteriyanov, V., Dzhusupova, R., Bosch, J., Olsson, H.H., 2024. Robust detection of line numbers in piping and instrumentation diagrams (P&IDs). In: 2024 International Conference on Machine Learning and Applications (ICMLA). IEEE, pp. 888–893.

Smith, R., 2007. An overview of the tesseract OCR engine. In: Ninth International Conference on Document Analysis and Recognition. ICDAR 2007, vol. 2, pp. 629–633. http://dx.doi.org/10.1109/ICDAR.2007.4376991.

Theisen, M.F., Flores, K.N., Balhorn, L.S., Schweidtmann, A.M., 2023. Digitization of chemical process flow diagrams using deep convolutional neural networks. Digit. Chem. Eng. 6, 100072.

Union, E., 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L 168, 12 July 2024, pp. 1–254, https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

Villena Toro, J., Wiberg, A., Tarkian, M., 2023. Optical character recognition on engineering drawings to achieve automation in production quality control. Front. Manuf. Technol. 3, 1154132.

Wang, W., Xie, E., Li, X., Hou, W., Lu, T., Yu, G., Shao, S., 2019. Shape robust text detection with progressive scale expansion network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9336–9345.

Yao, L., Tang, C., Wan, Y., 2025. Advanced text detection of container numbers via dual-branch adaptive multi-scale network. Appl. Sci. 15 (3), 1492.

Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., Liang, J., 2017. East: an efficient and accurate scene text detector. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5551–5560.