15th IFK
Proc. of the 15th International Fluid Power Conference <Symposium/Conference>
<DOI>

AI Assisted Engineering: Information Extraction from Various Formats for Engineering Decisions

Vaishnavi Kulkarni, Anwesha Sen, Anik Kumar Samanta*, Shrinivas Kulkarni

Danfoss Power Solutions, Danfoss India Innovation Centre, Pune, India.,

vaishnavi.kulkarni@danfoss.com,

anwesha.sen@danfoss.com,

anik.samanta@danfoss.com,

shrinivas.kulkarni@danfoss.com

Abstract

Engineering drawings and 3D (3-Dimensional) models store vital information that is essential for any industrial product life cycle – right from designing and manufacturing to order placement and inventory management. Accurate retrieval of the information requires intensive manual efforts and is prone to human bias and error. In this paper, we present two novel methods to extract information from 3D models and 2D engineering drawings. This information is contained in the models and drawings with different file formats such as portable document format (PDF) and computer aided design (CAD) models. This makes it difficult to integrate the drawing data into the company's data management system. Hence, we have used optical character recognition (OCR) to gather dimensions and specifications from the drawings and models of different file formats. The information is also present in different fonts and locations across the drawing or models. Although these documents contain a lot of relevant data, they also include unnecessary information. The fine tuning of character recognition models is done in the context of mechanical drawings, which improves the quality of data reading. Additional quality checking parameters such as nearest character matching and template matching of required text areas are also integrated to avoid a trade-off between speed and accuracy. The accuracy of the correctly read tokens using a combination of above-mentioned OCR strategies gives an accuracy of approximately 93%, while the detections with low confidence are automatically filtered out. This leads to the error proof output of the data in a format that is expected by the end user. An OCR based strategy for information extraction in this manner significantly reduces manual efforts and the same methodology is widely scalable across different types of drawings and models.

Keywords: 2D-drawings, 3D-models, Assembly Drawings, Drawing information retrieval, EasyOCR, machine learning, Tesseract OCR, Optical character recognition, text recognition.

1. INTRODUCTION

Hydraulic drawings contain critical information that is used from the design and development stage to manufacturing and ordering. In this era of digitalization, it is an organizational prerogative to arrange, clean, label, and store data for inferences and insights. However, this requires tremendous human effort, and the repetitive manual entry is prone to errors. The drawings and documents have accumulated over the years through internal innovative developments and external mergers. The format of the data is varied, and the numbers are huge. The drawings contain information about different child parts, dimensions, specifications, and material details. Drawings / any relevant fluid power circuitry are stored as three-dimensional (3D) models or two-dimensional (2D) images. Moreover, some of the drawings are scanned and stored in document formats like PDF as discussed in Jameison et. al. [1]. The scanned documents are noisy and extraction of useful information from such documents is challenging. Despite many advances in OCR-based document analysis, significant challenges remain in industrial engineering contexts. A large portion of legacy mechanical drawings exists as old, noisy PDF files, many of which are scanned images containing faded print, and handwritten annotations. Such documents have low contrast, background noise, and inconsistent text orientation, making text extraction difficult for standard OCR systems. The current methods reviewed in the literature cannot be directly applied to accurately extract structured engineering information from these documents, highlighting the need for a robust, domain-specific OCR framework.

The interpretation of 3D model textual data on a large scale without metadata is a hugely time-consuming task. OCR has not been used in the popularly available literature to interpret area, mass and other data for 3D models with high accuracy, this is something that we have attempted for the first time in our use case. The paper presents two OCR-based frameworks to extract engineering information from existing 2D drawings and 3D CAD models. For 2D drawings, we convert pages to images, split them into quadrants, run OCR, and match the text against internal databases to identify materials and surface finishes. For 3D models, we have automated Creo View Lite to open files, capture the “Summary” window, apply OCR, and parse values relevant to the model’s design. OCR in our use case has been finetuned to handle text, numeric and specials symbols data as high accuracy is a requirement for our case. The goal is to reduce manual work and errors in feeding data into company systems.

Information extraction from 2D Drawings: Mechanical drawings often contain critical information that must be accurately identified for further analysis in manufacturing industries. They have critical specifications, dimensions, and material details that directly influence design and production quality. Traditionally, these drawings have been manually interpreted by engineers. However, a vast number of legacy drawings exists only as PDFs or scanned images, making automated extraction of textual information tricky.

Based on literature, very recent research introduces eDOCr2 [2], a hybrid framework that combines traditional OCR, image processing, and Vision-Language (VL) models to extract structured and semantic information from mechanical engineering drawings (EDs). To improve semantic understanding and reasoning, eDOCr2 integrates Vision Language models (Qwen2-VL-7B and GPT-4o) after segmentation to verify, filter, or retrieve information [3]. Some literature proposes an OCR-based system using Tesseract OCR [4] to extract text from general digital images through detection, segmentation, and binarization. It separates the text from the background and then detects the text. Although these techniques extract text by applying different methods like text localization, binarization, etc., the method operates on entire drawings and does not explicitly consider sub-region OCR to boost recognition performance. Unlike standard documents, mechanical drawings contain multi-scale annotations, symbols, non-standard fonts. Optical character recognition (OCR) systems trained on printed or handwritten text often underperform on these noisy, skewed, and symbol-heavy images. Consequently, there is a need for robust and domain-specific extraction pipelines that can bridge the gap between raw images and structured databases.

In this work, we propose a quadrant-based OCR technique for mechanical PDF drawings. The approach converts each drawing page into an image, divides it into four quadrants, applies OCR independently, and then maps extracted text to a structured excel database. Dividing image into more than four quadrants can result into system processing overhead, which will take long time for processing. The resulting information is stored in Excel along with the drawing names, creating a structured dataset. This workflow is lightweight and adaptable, provides improved accuracy by localizing OCR to smaller image regions.

Information extraction from 3D models: 3D models across Danfoss and other multinational corporations dealing in mechanical models and assemblies require the area, mass, volume and other data of the parts to be saved across different product lifecycle management software such as SAP, Siemens Teamcenter, etc. The manual entry of a large number and variety of components consumes longer man hours and is a non-value-added activity. Additionally, the handling of different types of files such as step (STP), part (PRT) and assembly (ASM) may employ different software tools and strategies to ensure the correct file is used and its subcomponents are all loaded properly. Repeating this monotonous process leads to human error, especially since area, volume and other values are required to be in single point precision and a single wrong entry may affect the system data in the long run. The quality control of such processes additionally needs to be performed manually, which again may contain human bias and discrepancies.

Hence, there is the requirement of an automation system that can process many files with high accuracy, extract relevant information from the same and return the digitized version of the results. All of this should be done in a fraction of the time required to repeat the same manually. Different software packages were that were popularly available were compared for analysis [5]. Most software either did not contain metadata for viewing or displayed incorrect values. To address this lacuna, a non-parametric method was needed. Thus, CREO View and Python based OCR dependent logic was developed to access and digitize area, volume, mass, and envelope information for STP, PRT and ASM type mechanical model files. This can return accurate information without any human intervention and segregate healthy and unhealthy files on its own, thus producing additional insights.

The main contributions made in this paper can be listed as following: (1) Automated information extraction from old, noisy, and sometimes scanned mechanical drawings in PDF format, (2) proposed a quadrant-based OCR technology to improve the accuracy of text extraction of existing methods, (3) applying OCR to 3D model snapshots, overcoming challenges of graphical overlays and extracting engineering-specific dimensions and annotations, (4) proposed a model viewer based software for 3D model data extraction, (5) combining two OCR engines to achieve optimally best results for small fonts and numeric data of 3D models.

2. INFORMATION EXTRACTION FROM OLD, NOISY DOCUMENTS AND IMAGES

Any leading manufacturing industry player always strives to make their products environment friendly and sustainable. So, the materials that are used to make mechanical products play a pivotal role. Mechanical CAD drawings contain the data of dimensions and tolerances along with critical information regarding which materials are to be used, and surface finishes required for each component in product manufacturing. Accurate identification and evaluation of constituent materials and finishing ensure Danfoss products are aligned with current sustainability goals. We developed a robust OCR approach to identify material specifications and surface finishes. We have used Tesseract OCR as it is open source and all the computation happens locally, which helps to maintain the confidentiality of data.

The proposed solution is implemented in four major steps:

i. Image segmentation and optical character recognition
ii. PDF to Image conversion
iii. Text consolidation
iv. Pattern matching and classification

2.1. Image Segmentation and Optical Character Recognition

The first step involves opening a PDF document with the help of a default PDF viewer of the operating system. The system attempts to identify and activate the system's default PDF viewer window. Now the system's screen dimensions are calculated to identify the best capture regions. The screenshot capture process implements a quadrant-wise approach for capturing the screenshots of PDF content. Now based on the document's width and height, which is extracted from PDF metadata, it calculates the optimal zoom level. This is done to maintain the resolution of the text present in the PDF. The viewport manipulation works in the clockwise direction starting from the top-left quadrant. Now the system applies to the calculated zoom level and captures the screenshot. Following this, it captures the screenshots of all the quadrants ending at the top-right corner. The captured screenshots are then saved with the detailed name of the file followed by the quadrant name for further analysis as seen in Fig. 1.

As an initial step, OCR is applied separately to each quadrant image. Each image maintains its resolution loaded using image processing libraries that provide image format support and image manipulation techniques. Based on the image quality, different methods are used like denoising and contrast enhancements to improve text definition. The images are converted into high contrast black and white images which are optimal for OCR. The Tesseract OCR engine is configured with required parameters for technical document processing, along with including language models appropriate for engineering terminology and character recognition settings. In this process, accuracy is preferred over speed. The extracted text from each screenshot is stored in a temporary text file, which is further used to post-process the data to extract the relevant information.

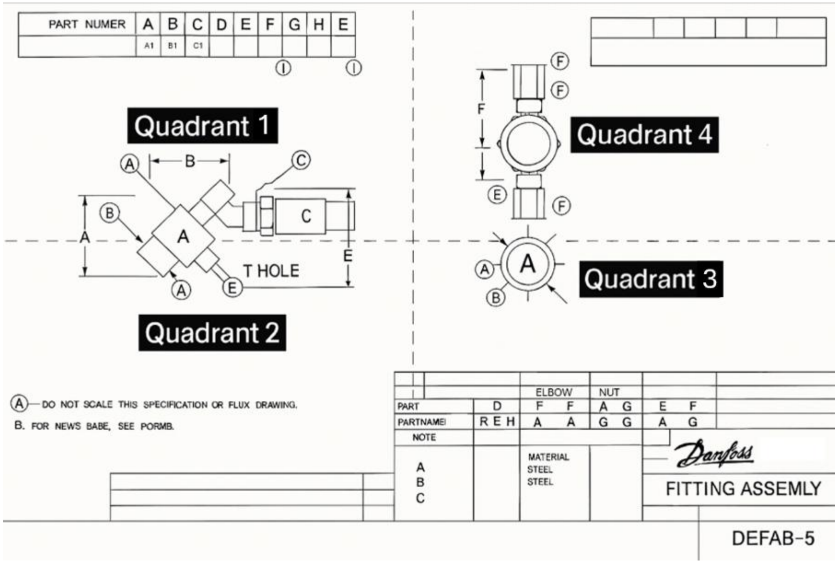


Fig. 1. Representative Dummy Drawing with quadrant references

2.2. PDF to Image Conversion

This step includes direct conversion of PDFs into images. This process makes use of specialized PDF-to-image conversion libraries that interpret the PDF format and generate bitmap representations of the document. The conversion process converts all pages within each PDF document, creating joint photographic experts' group (JPEG) images at a resolution suitable for OCR analysis. The output images are organized in such a way that it maintains the directory hierarchy which has clear association between image name and original source PDF name. Now all JPEG images are converted to portable network graphic (PNG) image format. PNG is a lossless format that preserves image quality throughout the process [6].

Now the image cropping phase is implemented, which applies strategic segmentation to full-page images, focusing on relevant content that is present in the PDF. By leveraging domain specific knowledge, we targeted the selection of regions where most critical information lies. The regions include a title block, material specifications, and annotations block. The cropping method begins by analyzing the dimensions of the full-page image to set proportional dimensions for the interested regions. The primary extraction focuses on the upper half of the document, which commonly contains detailed specification table and material information. After that, it extracts focused regions like the bottom-right area of the document, which typically has title blocks and primary material specifications in mechanical drawings.

Now each cropped image is given a descriptive name as "original filename + region-identifier" and saved as seen in Fig. 2 in the regions highlighted by quadrants 3 and 4. This segmentation approach allows for region specific OCR, which not only reduces computational overhead but also improves the detection accuracy.

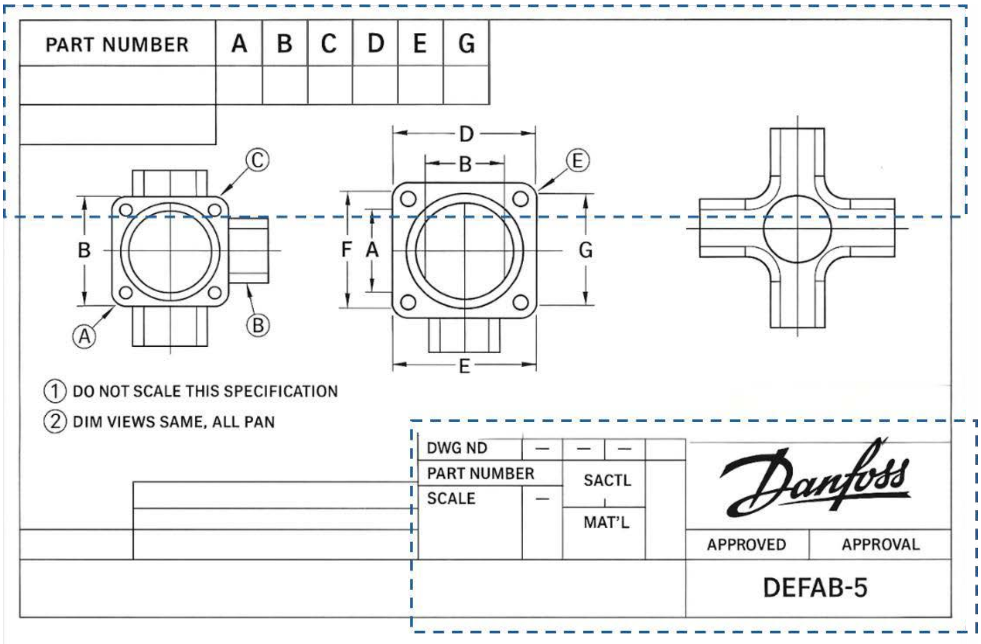


Fig. 2. Representative Dummy Image of Drawing with Identified Region

2.3. Text Consolidation

Text consolidation phase includes concatenation of different text extraction outputs in analyzable collection of text data. This process begins with collecting text data from three different methods:

i. Native PDF text extraction using PyPDF2 library of Python.
ii. From quadrant-based screenshot OCR results from the first step.
iii. Direct image conversion OCR results from the second step.

Each text data source provides unique characteristics. Native PDF text extraction usually provides high accuracy for digitally created documents, preserving exact characters. The screenshot-based OCR results are important, as it extracts text from specific information in dense regions of the drawings. The integration process concatenates these different text sources while maintaining source identification markers that make analysis of data easier.

The consolidated text undergoes initial cleaning process that removes whitespace patterns, normalizing line break characters. Additionally, the code generates an uppercase variant of the text to reduce case insensitive pattern matching operations. This duplicate representation ensures that material specifications written in various case combinations can be identified.

After consolidating and removing inconsistencies from the text, now comes character level error correction phase. This step is specifically performed to handle common misidentifications. This process applies domain-specific correction rules, that account for the visual similarities between certain characters, in the context of industrial specifications.

The most fundamental confusion is between letters and numbers that are visually the same. The letter ‘O’ is replaced by number ‘0’ in context where numeric value is expected. Lowercase ‘l’ characters are converted into numeric ‘1’ when there are alphanumeric specification codes. Also, some technical specification needs corrections for misidentification such as ‘AIST’ replaced by ‘AISI’ to correct common misreading of ‘American Iron and Steel Institute’ abbreviation. Apart from this character level correction, this solution also takes care of spacing irregularities, hyphens, dashes, and punctuation marks that commonly occur in technical document processing. Although rule-based standardization resolves many character-level ambiguities, it does not fully capture contextual intent. Future mitigation strategies should integrate contextual AI models and pattern-based validation to disambiguate visually similar characters using semantic cues and domain knowledge, thereby improving accuracy in technical document interpretation.

2.4. Pattern Matching and Classification

2.4.1 Material Name Recognition

The material recognition process implements a database-driven approach to correctly identify the materials. This process begins with accessing the material column from the predefined database of excel format. Material names are normalized to lowercase format to avoid case-insensitive matching. The matching algorithm works on exact string matching within collected text from different phases. For each material name in the database, the code searches for its presence in the collection of data.

Some domain specific logic rules are implemented to handle ambiguous material terminology that may appear to be the same. For example, when the material name ‘anneal’ is found, the algorithm applies additional check to ensure that ‘annealing’ is not present in the same document, as these terms represent different material specifications and should not be conflicted. Valid matches are stored into result set along with its document name. The identified materials are stored in the form of comma-separated values.

2.4.2 Material Specification

Material specification is one of the most complex aspects of pattern matching. It requires a valid pattern recognition capability to identify and validate the material specification codes commonly used in industrial applications. This process identifies various material specification codes like AISI steel standards, W-codes, and generic specifications.

The American Iron and Steel Institute (AISI) specification identification starts with pattern matching, which will first identify the characteristics format of ‘AISI’ prefix followed by numeric codes ranging from three to four digits. Each AISI code undergoes the verification of material specification codes against the database, ensuring that only valid AISI designations are stored in the results.

W-codes are processed based on the pattern matching which begins with letter ‘W’ followed of alphanumeric characters. The W-code processing algorithm implements a comprehensive character correction system that addresses common misreads. The character correction process replaces frequently confused characters with the most likely intended value in the context of material specification codes. The letters 'o' and 'O' are standardized to '0', because the material specification does not contain letter ‘o’ or ‘O’. The W-code validation process includes length verification to ensure that processed codes are of five-character format.

2.4.3 Surface Finish

Surface finish recognition focuses on identifying surface treatment and finishing specifications. The surface finish defines the final appearance, texture, and protective characteristics of materials. This process begins with text normalization performed to address the common OCR misinterpretation of finish specifications.

These include the replacement of 'ES _' with 'ES', and the systematic replacement of punctuation marks that are frequently misread during OCR processing. Exclamation marks are converted to '1' when they appear in specification contexts, addressing a common OCR confusion pattern. For each finish specification in the database, the system applies string normalization and searches within the processed OCR text using both exact matching and the character-substituted variants.

Special case handling addresses industry-specific terminology. For instance, when the term 'protective' is identified within the document text, the system automatically adds 'protective oil' to the finish specifications, recognizing this as a standard industry term. This attention to terminological precision reflects the critical importance of accurate finish specification in manufacturing and compliance contexts.

All extracted information related to material names, material specifications, and surface finish details is stored in a structured Excel-based database. The process begins by uploading a folder containing the required PDF mechanical drawings, after which the automated extraction workflow is initiated. Each PDF drawing is processed sequentially, and the extracted results are systematically stored in the Excel database. The structured database maintains key attributes for every processed drawing, including the PDF drawing name, material name, material specifications, and associated surface finish information. This organized representation of extracted data enables efficient retrieval and analysis. As a result, the proposed system significantly supports product engineers in making faster and more informed design and material-related decisions, while reducing the need for manual interpretation of engineering drawings.

3. INFORMATION EXTRACTION FROM 3D MODELS

The 3D model parts of any hydraulic or relevant industry are primarily of 3 types – STEP, PRT, and ASM. These files may be generated by different software and contain different metadata and compositions. The STP files are normally generated through the Autodesk platform software and are widely usable across different cross platform 3D model handlers. However, being a proprietary data format, use with other software requires export and conversion into another suitable format. The PRT and ASM file types are more common and can be generated using a wider variety of software, such as PTC Creo Parametric and Siemens NX.

In terms of complexity, the STP files are easiest to handle as they contain suppressed parts and hence can be directly visible and accessed. In most cases, these types of files contain metadata as well, as they are built using a standardized method and hence require some information related to area and material to be already present in the model. The PRT files are the second easiest type of file to handle as they contain all required sub parts arranged under the parent part. However, the issue occurs when the subparts have failures during building or are damaged as part of the migrations from one platform to another. This can occur either if the model was not downloaded or imported incorrectly, or if the model itself has errors while building. The ASM files are the most difficult to handle as directly opening the ASM alone provides no information other than envelope; it requires the individual PRTs inside it to be handled as separate files. Also, in case of nested ASMs it should be ensured that the incorrect PRT is not chosen as seen in Fig. 3. Additionally, any damaged or broken ASMs should be automatically filtered out and not allowed to proceed with data extraction.

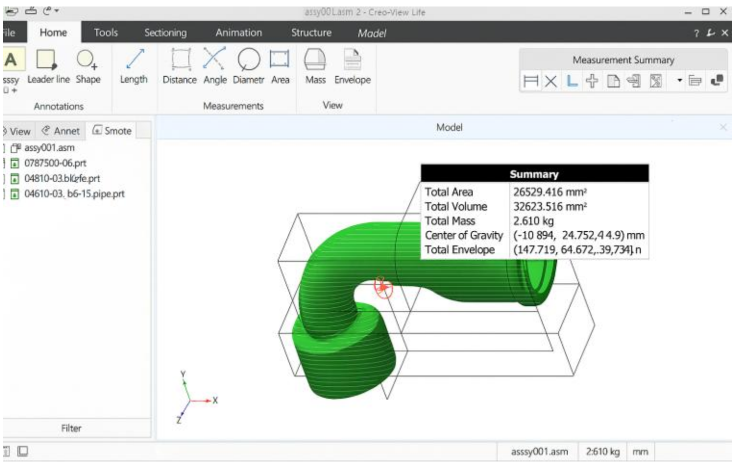


Fig. 3. Representative dummy structure of assemblies and Summary window

3.1. Software to view 3D model data

To get the area, volume, and other descriptive information about the 3D models, a software is required where one can access and open every type of file and additionally, give indicators of damaged cases without any false positives. Generic options such as Creo Parametric 3.5.0, SolidWorks and Autodesk Inventor were explored and referred to through literature. Creo Parametric allows easy access to PRT and ASM files; however, it requires manually intensive conversions to open STP files. Additionally, it offers viewing of metadata parameters, provided that the built file has metadata. For SolidWorks, the file formats are not always handled properly and give a lot of launching errors. For the Autodesk Inventor, it is good for viewing and accessing STP files, but for PRT and ASM files it does not contain a lot of the information needed for data entry, like volume, mass, and more [7].

The reason that these software do not work with optimal efficiency is because they are predominantly parametric software, i.e., they work best with models which already have predefined settings and parameters as a part of their design. However, since most models made in a corporate environment are a result of the interaction of multiple software and their versions, the parametric measures often are hidden or removed with user interaction. Hence, there was a requirement for software that could not damage the entities of the model itself but just give the information in a collected way.

For this purpose, the most optimal software to be found was Creo View Lite 8.1. This software comes as a part of the PTC suite of software and is a viewer that can open STP, PRT and ASM type files. Additionally, it can indicate which subparts of these models are damaged using warning marks and it can indicate "broken" assemblies in red as well as observed in Fig. 4. The Creo View window has an interesting feature known as the Summary feature under the markups section, that shows information about the model regarding the Area, Volume, Mass, Centre of Gravity and Envelope. This information is highly accurate and useful to us as it contains data about all the parameters that can be needed as part of the SAP data entry. However, this tool does not save the results of the summary anywhere directly. Hence, an approach was required to save the summary data in some format for use later.

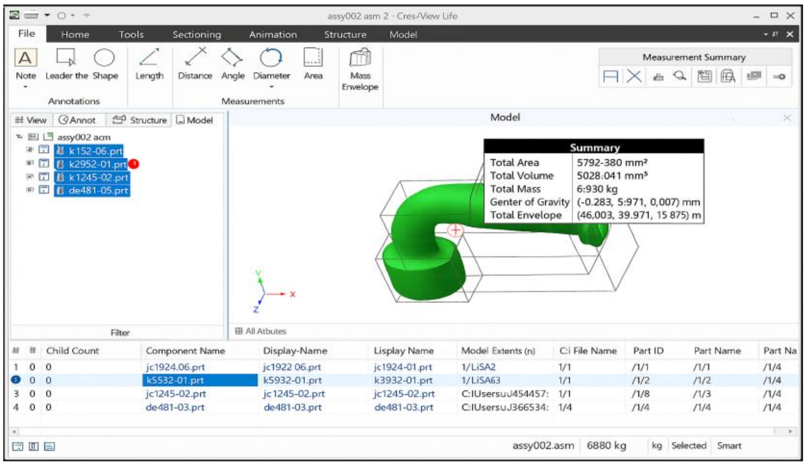


Fig. 4. Representative locations of Summary window with broken parts marked in red

3.2. *Using Metadata Extraction*

An initial attempt to extract the data was done using Creo Parametric's metadata extraction. This involved linking Creo Parametric with a Creoson API server that could intermittently ping the software in the background to access metadata on its own. However, this process observed multiple limitations. The metadata is only present for around 10% of total files (30,000 STP and 20,000 ASM files that we processed), which amounts to a very less percentage. Also, the API can only handle up to 15 fields at once instance, requiring the instances to be stopped and restarted after every 15 files.

The metadata was affected by certain software defined settings which might not be correct for every model. For example, the automatic material assigned for a lot of the models was stainless steel; however, this was the wrong material and hence manipulated the mass value from the actual value.

Owing to all these issues, the OCR-based approach was chosen as the best method to extract the Summary information from the Creo View lite software's summary window.

3.3. The Process - Extracting files

The first step involves choosing the correct file for information extraction. The arrangement of the files is in a nested folder–inside-folder arrangement. The correct file is the one that is inside the folder starting with the keyword "MD" and contains the latest revision. At times, the file may be of a different name than the one inside the folder, hence a smart adaptive logic is needed to find the naming inside which is closest to the given part name. For this purpose, match-by-similar name logic was used, and this was verified with the domain experts to ensure the correct file was being loaded.

The file loads up and opens in Creo View. The first thing to do is to tick the checkbox next to the filename, which makes the file visible.
This is where the OCR process starts.

3.4. Optical Character Recognition Techniques for Different Texts

Optical character recognition finds different ways of working in different use cases. For the clicking of the checkbox at the first box, an approximate coordinate region is given to the system within which it looks for the text entry. In Creo View, the checkbox is at the upper left quadrant region of the window, so the OCR is made to act in that region initially only.
The text is of the font Calibri, with a font size of approximately 12-14 pixels. As per Park and Shin [8], for such conditions, Tesseract OCR or EasyOCR provide the best results. For recognizing small Calibri text in paragraph form, Tesseract OCR generally offers better accuracy than EasyOCR due to its optimized handling of clean, printed fonts and structured layouts. Multiple methods of OCR were also compared, such as TrOCR (Transformer OCR, used by Microsoft Tools) [9], and Mistral OCR [10]. A comparison was done between all the different possibilities using a variety of parameters.

Feature	**Tesseract (v4/5)**	**EasyOCR**	**TrOCR (Base/Large)**	**Mistral OCR (LLM)**
Architecture	LSTM + Statistical	CRNN (Deep Learning)	Vision Transformer	Large Language Model
Best For	Scanned Docs, High Res	Noisy/Scene Text	Handwritten/Curved	Semantic Understanding

Inference Speed	Extremely Fast (<0.2s)	Fast (~0.5s - 1s)	Slow (~3s - 8s)	Slowest (API/GPU heavy)
Model Size	~20 MB	~80 MB	~1.5 GB+	~4 GB+ (or API call)

While EasyOCR is faster and easier to set up, it tends to struggle with small font sizes and produces more errors in such cases where there are multiple words close to each other. Tesseract, though slower, benefits from extensive customization options and consistently delivers more precise results for small, well-formatted text like Calibri. This is due to the differences in the way they target individual frames, as seen in Fig. 5.

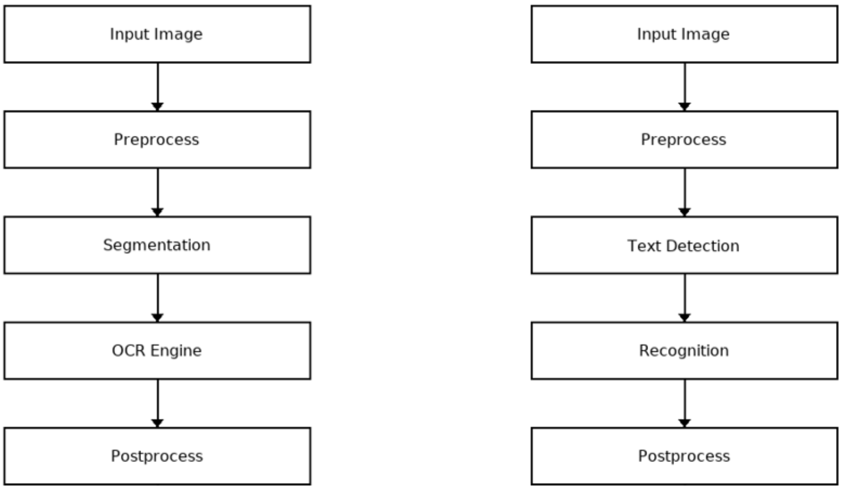


Fig. 5. Comparison of Tesseract OCR and EasyOCR in terms of working

Thus, a combination of EasyOCR and Tesseract OCR is used. First, Tesseract OCR detects smaller text like names of file and subparts, and EasyOCR detects the bigger text like markup subtab and summary option.

Tesseract OCR includes a set of configurable modes known as page segmentation modes (PSMs), which are essential for tailoring how the engine interprets the layout of text in an image. These modes are particularly useful when dealing with different types of documents or image structures, such as full pages, single lines, or isolated words. For example, PSM 3 is the default mode and works well for general documents with multiple blocks of text, while PSM 11 is ideal for images containing a single, uniform block of texts such as a cropped paragraph in Calibri font. These modes help Tesseract decide whether to look for multiple columns, lines, or just a single word, which directly impacts the accuracy of OCR results. Selecting the right PSM is crucial, especially when working with structured fonts like Calibri, as it allows the engine to better understand the spatial arrangement of characters and lines, reducing recognition errors and improving overall performance [11].

For ASM files, the OCR can find files which have broken PRTs as present in it as well. Often, the name of the PRT may contain a drop-down triangle next to it, indicating the presence of 1 or more nested ASMs. In such nests, the name of the PRT and the ASM may be the same, hence a 50 to 100-pixel margin to the bottom of the last entry needs to be reread for confidence. The broken ASM is found by applying a red mask over the images of the summary. If a red region is identified in the region of the ASM, it means it is a broken file.

3.5. OCR for Data Extraction from Snapshots

To extract the data, pattern matching is performed to get entries of area, volume, and other information from the summary snap. Generally, the data follows a fixed structure of ~ symbol followed by the value followed by 3 decimal points. However, it may also contain exponential marks like +/- e^0, which need to be interpreted as scientific symbols. This is done using Tesseract OCR, which finds good application in this case as it works for mathematical text.

Additionally, the text might be present in multiple regions like top right, middle, bottom right, etc. Hence, a fallback logic is added starting in anticlockwise order across the image to read if not found in any region. For the envelope, it may contain data such as "Enve", "ENVELOP" and other formats, hence all possible combinations of case and letter mappings are considered. Once the data is extracted for each file it is automatically added to an excel. This principle can also be extended to cases involving fluid power drawings specifically – involving complex wiring and/or component symbols. The relevant character / symbol matching in this case simply changes but follows the similar generic principles of searching by character in multiple locations of a document as fall back, or mapping the mis detected characters to their nearest matches based on historical interpretation of that character based on testing with a sample dataset.

4. RESULTS AND DISCUSSION

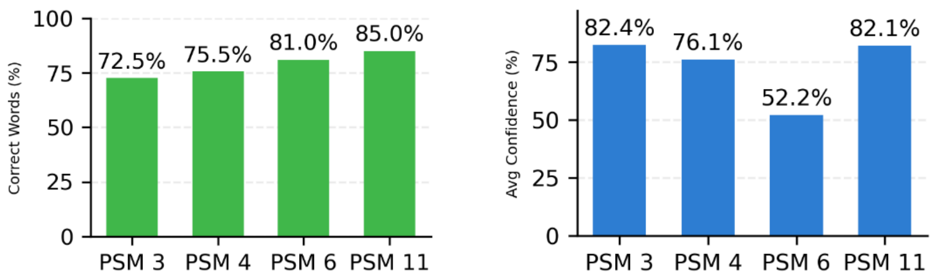
The entire pipeline of the OCR process for both 2D PDF drawings and snapshots of 3D models involves the need of a dedicated system with stable to high configuration settings. A PC with a high configuration of not less than 32GB RAM and a GPU was used to perform the OCR for the drawings and to capture the snapshots for the 3D model information extraction. The pipeline follows the mass download of all required files onto the dedicated system, followed by the identification of the relevant files to be processed using a naming convention as specified by a product engineering team.

In case of the PDF files, they are converted into images (as discussed in Section 2) and processed. Overall, each file takes around 20-30 seconds to be processed from end to end. The proposed system for text extraction from PDF drawings was tested on a collection of PDFs mechanical drawings. This solution was tested on a sample of 30 engineering drawings in PDF format. Material information was successfully extracted from 90% of the drawings with 79% accuracy. For material finish information, data was extracted from 83% of the drawings with 72% accuracy. By implementing the proposed solution, manual effort can be reduced by approximately 76%, resulting in significant time savings and improved efficiency in processing engineering drawings. Unsuccessful extraction of information from PDF was due to low-quality and unreadable drawings. The database matching step helped filter out irrelevant or misrecognized text, ensuring high-quality results. The final structured output in excel format enabled engineers to take decisions related to the material specifications.

As per the observations and discussed Section 3.4, for the 3D models text extraction, EasyOCR performs better in individual character cases, but Tesseract OCR performs better in cases of repeated words. For example, for occurrences like "Volume: 12345.67", EasyOCR is better at interpreting the relevant data; however, when it comes to compound words such as "Total Volume", Tesseract OCR performs better. Hence, EasyOCR is first used to quickly extract text from the snapshot for an initial overview and to give a general idea about locations of words. Then, Tesseract OCR with different PSM settings is applied to the same file for more detailed results.

To accurately detect word combinations like 'Area', 'Total Area', and 'Volume', analysis of different PSM modes of Tesseract was done. Fig. 6(a) shows that higher PSM values generally result in more correctly detected words for a given file. However, as seen in Fig. 6(a) when comparing based on confidence of the individual PSM modes independently, it was observed that PSM 3 and PSM 11 gave good results in terms of detection confidence. The only case that seemed to act poorly in this case was the PSM 6, with a confidence of around 52%. This did not correlate with the accuracy-based findings in Fig. 6(a). This meant that no PSM configuration alone would give accurate and confident results.

Thus, the best approach was using the average result of all the different modes. When tested on a sample batch of 100 files, it was observed that this method gave balanced results. The logic was finetuned and scaled by mentioning the exact possible combinations of words that could be seen in the screen capture and estimating fallback regions across the image to determine which region contained the numeric information. For example, the image could contain the information in either the middle region or the upper right region of snapshot, and hence the code was arranged to look at those regions in order of possibility and determine the final reading from only one.

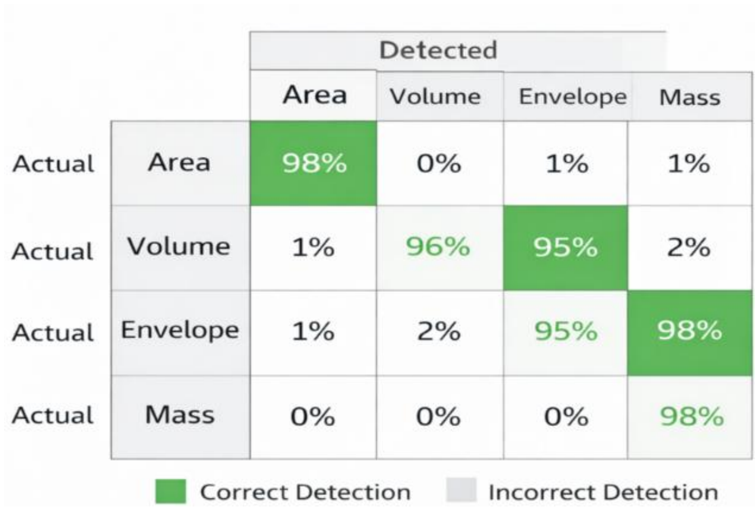


(a) Accuracy of detecting correct words for different modes

(b) Corresponding confidence score for each mode in detecting the words

Fig. 6. Performance of Tesseract OCR page segmentation for accuracy of detecting correct words and corresponding confidence score for different modes

Once the information was extracted, it was seen that the extracted text and numeric values for the different classes like area, volume etc., have some amount of overlap due to their occurrences appearing very close to each other and the OCR not being able to understand which unit at the end of the numeric value belongs to which class. For example, as seen in Fig. 7, some of the values of Area are interpreted as Volume and some values of Volume are often interpreted as Envelope. The cases with more than 2 overlaps are highlighted in green. However, they are filtered out on the basis of the error handling strategies which involves extracting the values at least twice in one iteration to check if the value extracted belongs to that class actually or not. Coupled with code to declare some preset formats of the text (E.g., volume values start with ~, integral value of the number will always be followed by 3 decimal points, etc), it highly improves the accuracy of text extraction from the snapshots irrespective of the location of the text.



		Detected			
		Area	Volume	Envelope	Mass
Actual	Area	98%	0%	1%	1%
Actual	Volume	1%	96%	95%	2%
Actual	Envelope	1%	2%	95%	98%
Actual	Mass	0%	0%	0%	98%

Fig. 7. Confusion Matrix indicating the accuracy of detection of words as per category

The logic was finalized on a batch of 10,000 ASM files, of which only 4700 were healthy unbroken files whose information we could enter into the system. Out of a total of 23,500 words/phrases (4,700 healthy file entries * 5 columns), 21,900 were detected correctly. This was verified using an automation script checking generic conditions (incomplete words, white spaces, etc) and manual quality checking. The same method when scaled to all ~50,00 files led to an overall accuracy of 93.1% for the text interpretation logic, which was in-line with the accuracy seen on the batch of 10,000 files. While the paper shows samples of dummy parts, the logic has been tested on actual industrial partner parts across multiple divisions and types.

5. CONCLUSIONS

This system demonstrates a practical approach for extracting textual information from mechanical drawings using OCR-based techniques. The methodology of converting PDFs into images, segmenting them into four quadrants, and applying OCR provides significant improvements in recognition accuracy. By matching extracted results with a database, the system ensures that only relevant and valid information is retained. This helps in reducing errors commonly associated with OCR in technical drawings. The data extraction from 3D models offers an opportunity to expedite a process that would otherwise require extensive manual intervention with high chances of error. Additionally, a large database of readily available dimension and weight possibilities might also help in creating models for interpolation or estimation in cases of those models wherein a stable build is not present. In the field of fluid power and its related disciplines, the adoption of methods that minimize repetitive manual tasks and automate data collection offers significant advantages. By streamlining these processes, organizations can not only improve the efficiency of data management but also enhance their ability to estimate and identify optimal design solutions for future projects. Automation reduces human error, accelerates workflows, and ensures that valuable engineering data is captured consistently and accurately.

Importantly, the scope of this approach is not confined solely to technical drawings within the fluid domain. The same strategy can be extended to a wide range of fluid power documentation, including power circuitry diagrams, hydraulic schematics, pneumatic layouts, and design review charts. Because the methodology does not rely on project-specific retraining or exhibit bias toward a particular type of drawing, it remains flexible and broadly applicable across different design contexts.

This universality means that engineers and designers can apply automated data handling techniques to diverse aspects of fluid power systems without the need for specialized customization.

6. FUTURE SCOPE

The strategy discussed is highly scalable and can be replicated with minor fine tuning for different types of drawings and models across industries. OCR-powered vision systems observe and automate conversion systems for electronics components as well, providing detailed information of components at each process step and saving thousands of engineering hours. The collected information can be implemented in generative AI to generate drawings and models of a given configuration on its own. To integrate additional fluid power circuitry and documentation, a modification would be needed to understand symbolic text instead of regular numeric text. With a combination of OCR powered text matching and location-based template matching, the symbols and nearby text can be found in a similar manner. Additionally, conversion of the relevant drawing circuitry into a suitable image or PDF format will also be viable to simplify the complexity of working with different file types.

7. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support from the different Danfoss divisions, whose 3D models and drawings were instrumental in validating this solution. The domain expertise and technical guidance provided by the different divisions were invaluable to the successful validation of this work. The authors thank Satish T., Shilpi S., Ketan M., Bernie B., Hemant S. and Prajakta K. We would also appreciate support from Jeff H., Rahul V., Prashant J. and leadership of Danfoss India Innovation Centre (DIIC). The authors highly appreciate Dr.-Ing. Robert Rahmfeld for his unwavering help and support.

8. REFERENCES

[1] Jamieson, L., Francisco Moreno-García, C. & Elyan, E. A review of deep learning methods for digitisation of complex documents and engineering diagrams. Artif Intell Rev 57, 136 (2024). https://doi.org/10.1007/s10462-024-10779-2

[2] A. Kae, G. Huang, C. Doersch and E. Learned-Miller, "Improving state-of-the-art OCR through high-precision document-specific modelling," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, pp. 1935-1942, doi: 10.1109/CVPR.2010.5539867

[3] L. Li, Y. Wei, Z. Xie, X. Yang, Y. Song, P. Wang, C. An, T. Liu, S. Li, B. Y. Lin, L. Kong, and Q. Liu, "VL-RewardBench: A challenging benchmark for vision-language generative reward models," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025, pp. 24657–24668.

[4] R. Smith, "An Overview of the Tesseract OCR Engine," Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Curitiba, Brazil, 2007, pp. 629-633, doi: 10.1109/ICDAR.2007.4376991.

[5] E. Tyflopoulos and M. Steinert, "A comparative study of the application of different commercial software for topology optimization," Applied Sciences, vol. 12, no. 2, Art. no. 611, 2022, doi: 10.3390/app12020611.

[6] E. Öztürk and A. Mesut, "Performance Evaluation of JPEG Standards, WebP and PNG in Terms of Compression Ratio and Time for Lossless Encoding," 2021 6th International Conference on Computer Science and Engineering (UBMK), Ankara, Turkey, 2021, pp. 15-20, doi: 10.1109/UBMK52708.2021.9558922.

[7] J. Villena Toro and M. Tarkian, "Optimizing text recognition in mechanical drawings: A comprehensive approach," Machines, vol. 13, no. 3, Art. no. 254, 2025, doi: 10.3390/machines13030254.

[8] Y. Park and Y. Shin, "Gradual OCR: An effective OCR approach based on gradual detection of texts," Mathematics, vol. 11, no. 22, Art. no. 4585, 2023, doi: 10.3390/math11224585.

[9] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei, "TrOCR: Transformer-based optical character recognition with pre-trained models,"

arXiv:2109.10282 [cs.CL], Sep. 2021, last revised Sep. 2022. [Online], https://doi.org/10.48550/arXiv.2109.10282

[10] A. Scius-Bertrand, A. Fakhari, L. Vögtlin, D. R. Cabral, and A. Fischer, "Are layout analysis and OCR still useful for document information extraction using foundation models?," in Proc. Int. Conf. on Document Analysis and Recognition (ICDAR)*, 2024, Lecture Notes in Computer Science, vol. 14807. Cham, Switzerland: Springer, pp. 171–186, doi: 10.1007/978-3-031-70546-5_11.

[11] Majumder, Mujibur Rahman, et al. "Offline optical character recognition (OCR) method: An effective method for scanned documents." 2019 22nd International Conference on Computer and Information Technology (ICCIT). IEEE, 2019.

Biographies

Vaishnavi Kulkarni is an Associate Engineer at Danfoss India Innovation Centre at Pune. She had received her B.Tech. degree in Mechanical Engineering from Vishwakarma Institute of Technology, Pune. Her research interests include natural language processing and image processing and more.

Anwesha Sen is an Associate Engineer at Danfoss India Innovation Centre at Pune. She had received her B.Tech. degree in Instrumentation and Controls Engineering from Cummins College of Engineering for Women in Pune. Her research interests include computer vision and image processing.

Anik Kumar Samanta is the Lead Data Science and Machine Learning Engineer at Danfoss India Innovation Centre at Pune. He received his Ph.D. (2021) and Masters (2016) degrees from the Indian Institute of Technology Kharagpur. He serves on the editorial board as associate editor of the IEEE Transactions on Instrumentation and Measurement and is an Executive Committee Member of the IEEE Signal Processing Society, Pune Section. His research interests include high-resolution spectral estimation, signal-based fault diagnosis of induction motors and hydraulic pumps, detection theory, and embedded signal processing.

Shrinivas Kulkarni is leading 'Data Science Technologies' team as Senior Manager at Danfoss India Innovation Centre at Pune. He received his Ph.D. in Machine Learning from the Indian Institute of Technology Bombay, India. He completed his master's (M.Sc.) degree in 'Mechatronics' from University of Duisburg-Essen Germany. He has 17+ years of experience in reputed organizations like, German Aerospace Centre (DLR)- the Institute of Robotics & Mechatronics, Moog GmbH (Germany), Moog India Innovation Centre, Eaton India Innovation Centre and Danfoss. He has expertise in the field of Mechatronics and machine learning techniques. He has patents to his credit and authored many technical papers in international conferences and journals. His research focus includes mathematical modelling of mechatronic components, intelligent mechatronic systems, Autonomous Systems, Control Systems, blending AI & Data Science Technologies with Industrial requirements.