AI-Based Engineering and Production Drawing Information Extraction

Christoph Haar[(✉)], Hangbeom Kim, and Lukas Koberg

Fraunhofer Institute for Manufacturing Engineering and Automation IPA,
Stuttgart, Germany
`christoph.haar@ipa.fraunhofer.de`

Abstract. The production of small batches to single parts has been increasing for many years and it burdens manufacturers with higher cost pressure. A significant proportion of the costs and processing time arise from indirect efforts such as understanding the manufacturing features of engineering drawings and the process planning based on the features. For this reason, the goal is to automate these indirect efforts. The basis for the process planning is information defined in the design department. The state of the art for information transfer between design and work preparation is the use of digital models enriched with additional information (e.g. STEP AP242). Until today, however, the use of 2D manufacturing drawings is widespread. In addition, a lot of knowledge is stored in old, already manufactured components that are only documented in 2D drawings. This paper provides an AI(Artificial Intelligence)-based methodology for extracting information from the 2D engineering and manufacturing drawings. Hereby, it combines and compiles object detection and text recognition methods to interpret the document systematically. Recognition rates for 2D drawings up to 70% are realized.

Keywords: Drawing · Manufacturing drawing · Image recognition · Text recognition · Process planning

1 Introduction

Most components have been manufactured on the basis of 2D drawings up to now. In addition to the pure geometry, these 2D drawings contain a lot of additional information like e.g. surface tolerances, dimensional tolerances, and a heat treatment that have to be read out and described semantically. This information is called Product Manufacturing Information (PMI) [1].

State-of-the-art transmission of information from design to process planning and on to production are enriched data formats from software such as CATIA, INVENTOR, Pro/E, SolidWorks, NX, etc., or software-independent exchange formats like STEP, JT, 3D PDF, and STL [2]. Some of these formats like STEP AP 242 include the documentation of the described additional information based on the ISO 10303-242:2014. On this basis, the exchange between disciplines is

theoretically possible today. Despite these prerequisites, many manufacturing companies use or receive 2D manufacturing drawings. This is because the reuse of existing drawings, the use of 2D CAD tools, and the not standardized annotation of manufacturing-specific information on 3D models are demanded. The digitalization and semantic description of 2D drawings are the basis for the automation of work planning and digitalized test procedures in production. The combination of this information and technology data form a knowledge database that can be used to derive rules for the automation of the work planning process. This paper focuses on AI (Artificial Intelligence)-based methodology for the extraction of non-geometric information which contains a high information content that is essential for rough pricing and work planning. In further work, it is planned to combine non-geometric and geometric information extracted from 2D manufacturing drawings. This combined information can be merged with 3D models to build a sufficient base for fully automated detailed work planning.

2   State of the Art

A literature review shows that there are several approaches to extracting non-geometric information from technical drawings. Prabhu et al. [3] propose a system, called the AUTOFEAD algorithm, that is able to extract non-geometric information from manufacturing drawings using Natural Language Processing (NLP) techniques. To search for dimensions and their attributes, a heuristic search procedure is developed. Scheibel et al. [4] describe a method to extract dimensional information from pdf manufacturing drawings. The text and position are extracted into HTML format. By clustering multiple text elements by position, the dimensional information is extracted. The authors suggest that the extracted information can be used to optimize a quality control system. Elyan et al. [5] develop an end-to-end framework to process and analyze engineering drawings. To interpret the drawings deep learning methods are used to detect and classify symbols.

To recognize text from images Optical Character Recognition (OCR) technology can be used. In the past years, a lot of research has been done on OCR methods. In [6,7], a rule-based algorithm is developed for text and graphics separation from engineering drawings. OCR method is used to recognize text from the separated areas. Jamieson et al. [8] propose a deep learning-based approach for text detection and recognition from engineering diagrams. The model is capable to recognise horizontal and vertical text.

Object detection is one of the most fundamental problems in computer vision techniques for locating instances of objects in images. A deep convolutional neural network is able to learn robust and high-level feature representations of an image. In the deep learning field, object detection can be categorized into two main groups "one-stage detection" (e.g. YOLO [9], SSD [10]) and "two-stage detection" (e.g. Faster R-CNN [11], Mask R-CNN [12]), where the former is regarded as "complete in one step" while the latter is called as a "coarse-to-fine" process [13]. The object detection technology is applied to understand the class and the location of symbols in [5].

The challenge to transfer OCR and AI-based object detection to manufacturing drawings remains and is addressed in this paper.

3   Methodology

Information to be extracted out of manufacturing drawings can be categorized into 5 categories. The dimensions, geometry, tolerances, general information, and additional manufacturing information (Fig. 1).
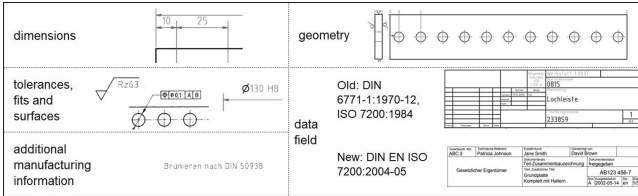

Fig. 1. Information in production drawings

The focus of the work is to recognize the non-geometric information. Figure 2 describes the process of AI-based drawing information extraction. The input drawing is divided into text and symbol information. The OCR method is used to interpret the text from the drawing image. Simultaneously, the object detector allows delivering of the classified objects with bounding boxes. Then, the extracted information is handled by matching algorithms. Based on this, PMIs are read and visualized.
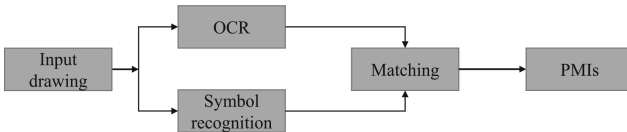

Fig. 2. Phases of the AI-based information extraction system

In the following three sections, the text recognition, the symbol recognition, and the compilation of the information are described in more detail.

3.1   Text Recognition

There are numerous open-source libraries and cloud solutions available for text recognition, also called OCR. The most commonly used systems are MMOCR, EasyOCR, Google Vision, Keras-OCR, and so on. Multiple evaluations of different tools show different results for this system. These results also arise from

the use of a wide variety of image data [14]. A basic analysis of some available solutions showed the superiority of the Google Vision system. Due to the goal of open source and python integration, the Python EasyOCR library is chosen. It is composed of 3 main components: feature extraction (Resnet) and VGG, sequence labeling (LSTM), and decoding (CTC). Due to the poor recognition of vertically oriented text, the images are rotated at intervals of 90° for text recognition.

3.2   Symbol Recognition

You Only Look Once(YOLO), a well-known and single-stage target detection algorithm, is used as our basic architecture [9]. The network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted with predicted probabilities. As a result, YOLO achieves high detection performance and outstanding inference speed.

Dataset Generation. Synthetic data is an approach to producing datasets that meet specific needs [15]. It enables humans to lessen labor for labeling or generating data manually. The synthetic data is generated to make up for the lack of data. 15 basic 2D drawing documents are used to enlarge the dataset with the information relating to the class and the location of symbols for the symbol recognition. For this purpose, 17 symbols like surface, edge, arrow, and tolerance are cropped from the basic drawings and randomly added to the empty background of basic drawings with different rotations and sizes (Fig. 3). The associated information about the class and the location of the symbols is stored in YOLO labeling format. The YOLO model is trained and tested with 1000 synthetic images. 80% of the dataset is used for training and validation and the rest of them is utilized for the test set.
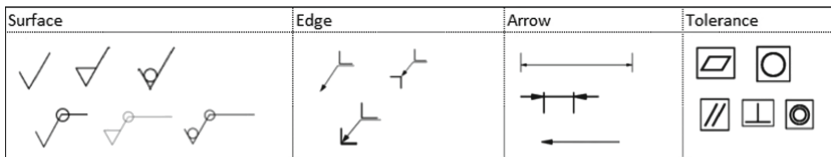

Fig. 3. 17 different extracted symbols, arranged by super-classes, for synthetic data

Object Detection Method. The latest version: YOLOv5 consists of multiple varieties of pre-trained models such as YOLOv5s, YOLOv5m, YOLOv5x, and so on. The difference between them is the size of the model. The lightweight model version YOLOv5s is not for accurate predictions but for fast inference time. Therefore, the YOLOv5x model is considered the main architecture since accuracy is the most significant factor to analyze 2D drawings. The SGD optimizer is used for training with the 1e-2 initial learning rate. Then the model is trained with 16 batch sizes and 640 image sizes. Figure 4 illustrates the symbol recognition sample with an actual 2D drawing document.
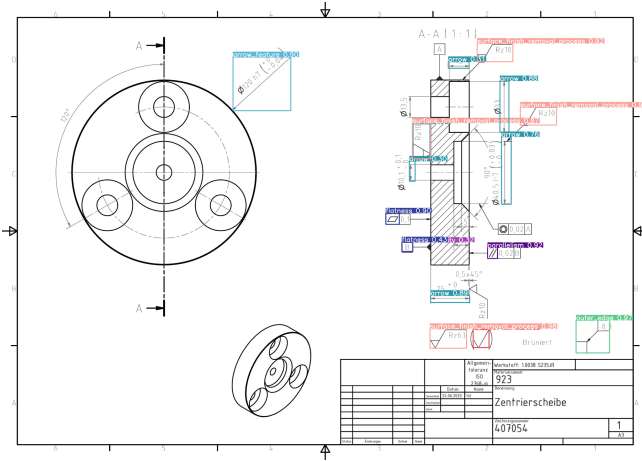
Fig. 4. Inference image using the trained model on the test set of the actual data

3.3   Matching of Symbol and Text

After the text recognition and the symbol recognition are completed, the result data is merged to extract the relevant information in each symbol. Tests have shown that any text with less than 50% recognition accuracy was mostly not recognized correctly, these are all filtered out. When it comes to symbol recognition, confidence and intersection over union(IoU) thresholds are set at 0.25 and 0.45, respectively. Then intersections of bounding boxes of text and symbols are searched for. If there is an intersection between two bounding boxes, the text is regarded as a related text for the symbol.

Specific characteristics of the drawing are used to analyze the title block. First, all text fields are extracted that are in the area of the title block. Then the text fields are assigned according to their geometric position or based on their format. Table 1 describes the rules we define:

Table 1. Rules for title block-matching

Feature	Rule
Title	Next to "Benennung"
Drawingnumber	Number with 6 digits
Material	Next to "Werkstoff"
Materialnumber	Next to "number"
Date	Search for data format

4   Experimental Results and Discussion

In this chapter, the results of the implemented method are presented and discussed. The strengths and weaknesses of the methods used are shown.

4.1   Text Recognition

Due to the focus of OCR systems on horizontal texts, the text images are analyzed both horizontally and rotated by 90°. Only texts with a confidence level over 0.5 are included in the analysis. The score for the evaluation of the text recognition is created per correctly recognized character and correctly recognized text field. 5 drawings with together 278 text fields are analyzed.

– 68% of the characters are recognized correctly
– 62% of the conveyors are recognized correctly

The difficulties in recognition are with mathematical special characters and with texts that, are positioned at an angle and position close to other forms (Fig. 5).
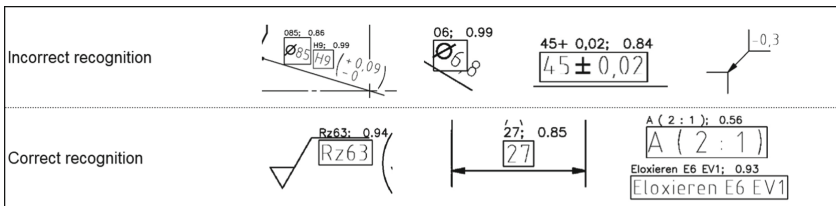

Fig. 5. Examples for incorrect and correct recognized characters, above the bounding boxes the recognized text, and the confidence level are shown

4.2   Symbol Recognition

We use the test set of the synthetic data and the test set of actual 2D drawings to estimate the trained model. The test set of synthetic data is created from the same 15 drawings as the training set but the symbols are positioned in different positions. When it comes to the actual 2D drawings test set, they are original documents and unseen for training the model. The model is evaluated by the detection mean average precision (mAP) since this is a common evaluation metric for object detection. The average precision is calculated with two different IoU thresholds: mAP and AP50. The intersection over union is a similarity measure between the bounding box of ground truth and the predicted detection. The mAP indicates the average of all 10 average precisions with the increment of the 0.05 IoU threshold steps from 0.5 to 0.95. And AP50 is the average precision at the IoU 0.5. The model achieves 0.927 accuracies with AP50 and 0.876

Table 2. The mAP values of the YOLOv5 on the test set of synthetic and actual data

Dataset	Test set (synthetic)	Test set (actual)
all	0.876	0.364
surface_all (6 classes)	0.870	0.537
edge_all (4 classes)	0.911	0.480
arrow_all (2 classes)	0.792	0.215
tolerance_all (5 classes)	0.888	0.207

accuracies with mAP in the test set of the synthetic data. We list the mAP of the final results of the model in each category in Table 2.

Most of the symbols are detected as their original label on the test set of synthetic data. Especially, the model shows outstanding predictions for edge classes with 0.911 accuracies. The average of the arrow classes results in relatively low accuracy at 0.792. This is because some lines are drawn across the arrow symbols and numbers are placed inside the bounding box of the arrows. On the other hand, the model shows difficulties in predicting the symbols on the actual dataset since the model relies on only small amounts of sample symbols and drawings. For this reason, we recommend training the model with abundant data for precise detection of symbols (Fig. 6).
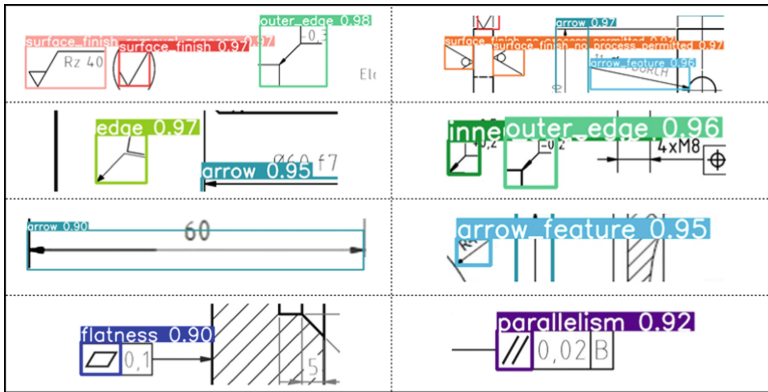

Fig. 6. Samples of detected symbols in 2D drawing documents

4.3   Matching of Symbol and Text

To find a score for the matching algorithm, the number of correct matches from the recognized texts and symbols is calculated. Only recognized features are used, so that recognition issues are not included in the score. 21 drawings are analyzed, and a total of 72% correct assignments are found. Problems arise

especially when the bounding box is too small and the text is further away from the symbol. The actual quality of the assignments depends to a large extent on the text and symbol recognition. From the title box, 88% of the information could be extracted correctly. The defined rules are reliable but must be adapted to other drawing types.

5   Conclusion

In this work, the flexibility of machine learning-based systems is adapted to the use case of production drawings recognition. Thus a system could be developed, which is able to read out information from production drawings. Based on 15 test drawings an accuracy of more than 70% is achieved. There is still potential for optimization in each of the described fields text recognition, symbol recognition, and merging. In the field of text recognition, the orientation of the texts and the recognition of special characters is a weakness, which can be eliminated by training new models. In the area of symbol recognition, the greatest potential lies in the extension of the training data set, especially to non-standard drawings. In the area of merging, the extension of the set of rules for semantic processing of the recognized texts and symbols has great potential. All in all, the realized approach represents an expandable basis for the recognition of information from 2D drawings. The biggest advantage of the presented method is the easy extension of the logic due to the use of machine learning-based approaches.

References

1. Lipman, R., Lubell, J., Hedberg, T., Freeney, B., Frechette, S.: MBE PMI validation and conformance testing project-NIST. Technical report, NIST (2013)
2. Nzetchou, S., Durupt, A., Remy, S., Eynard, B.: Review of CAD visualization standards in PLM. In: Fortin, C., Rivest, L., Bernard, A., Bouras, A. (eds.) PLM 2019. IAICT, vol. 565, pp. 34–43. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-42250-9_4
3. Prabhu, B., Biswas, S., Pande, S.: Intelligent system for extraction of product data from CADD models. Comput. Ind. **44**, 79–95 (2001)
4. Scheibel, B., Mangler, J., Rinderle-Ma, S.: Extraction of dimension requirements from engineering drawings for supporting quality control in production processes. Comput. Ind. **129**, 103442 (2021)
5. Elyan, E., Jamieson, L., Ali-Gombe, A.: Deep learning for symbols detection and classification in engineering drawings. Neural Netw. **129**, 91–102 (2020)
6. Kulkarni, C.R., Barbadekar, A.B.: Text detection and recognition: a review. Int. Res. J. Eng. Technol. (IRJET) **4**(6), 179–185 (2017)
7. Lu, Z.: Detection of text regions from digital engineering drawings. IEEE Trans. Pattern Anal. Mach. Intell. **20**(4), 431–439 (1998)

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

9. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: Yolov4: optimal speed and accuracy of object detection, arXiv preprint arXiv:2004.10934 (2020)

10. Liu, W., et al.: SSD: single shot multibox detector, CoRR, vol. abs/1512.02325 (2015)

11. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks, CoRR, vol. abs/1506.01497 (2015)

12. He, K., Gkioxari, G., Dollár, P., Girshick, R.B.: Mask R-CNN, CoRR, vol. abs/1703.06870 (2017)

13. Zou, Z., Shi, Z., Guo, Y., Ye, J.: Object detection in 20 years: a survey, CoRR, vol. abs/1905.05055 (2019)

14. Smelyakov, K., Chupryna, A., Darahan, D., Midina, S.: Effectiveness of modern text recognition solutions and tools for common data sources. In: CEUR Workshop Proceedings, pp. 154–165 (2021)

15. Nikolenko, S.I., et al.: Synthetic Data for Deep Learning. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-75178-4