Character: 17.352 Figures: 3

Digitization of Raster Drawings with Deep Learning

Framework outperforms OCR software in extracting data from mechanical drawings

Xiao Zhao, BCT and Offenburg University of Applied Sciences, Marko Weber, BCT Technology AG, Jan Schöffmann, BCT and Daniela Oelke, Offenburg University of Applied Sciences

Autoren

Dr. Xiao Zhao works as a postdoc at the Institute for Machine Learning & Analytics (IMLA) of Offenburg University and as a PLM solution architect for AI applications at BCT Technology AG.

Marko Weber, DI is Head of Software Development and Product Management at BCT Technology AG

Jan Schöffmann, BS works as a software developer at BCT Technology AG.

Prof. Daniela Oelke is a Professor for Machine Learning at Offenburg University, Germany and a member of the Institute for Machine Learning & Analytics (IMLA).

Kontakt

xiao.zhao@bct-technology.com
https://bct-technology.com/en/index

Automatic extraction of Product Manufacturing Information (PMI) from mechanical CAD drawings is a prerequisite for manufacturing and production quality control. Because of the special style of CAD drawings and the limited availability of training and test data, digitizing CAD drawings in raster images remains a challenge for Optical Character Recognition (OCR) systems. This work presents a novel deep learning-based framework to address this problem, which localizes and recognizes Geometrical Dimensioning and Tolerancing (GD&T) and dimensions in CAD drawings. The framework is composed of a centralized localization module and several subsequent pipelines for individual classes of PMI. The performance of the localization module, the text recognition network and the individual pipelines is evaluated using real data sets. Their performance is compared with the performance of the OCR software Tesseract.

Mechanical drawings are detailed, precise 2D depictions of mechanical workpiece or systems that are essential tools for the design, development and manufacturing of mechanical parts. Mechanical drawings provide a visual guide that conveys not only geometric information, such as shapes and geometry, but also textual information, e.g. dimensions, tolerances, applicable regulatory framework (e.g. ISO or DIN standards) and heat and surface treatment. Although there is a trend of integrating such Product Manufacturing Information (PMI) into 3D Computer-Aided Design (CAD) models, 2D mechanical drawings are still used in the current industrial practices, serving as a blueprint for the engineering, manufacturing and quality control tasks for large-scale machinery.

Character: 17.352 Figures: 3

Three types of mechanical drawings exist: vectorized drawings, high-quality raster drawings, and low-quality raster drawings [1]. Advances in CAD software allow engineers to create, modify, and analyze vectorized drawings digitally. High-quality raster drawings contain only pixel information. They are generated by CAD software, though additional information e.g. tolerances and norms, must be added manually [2]. Low-quality raster drawings are older, paper-based drawings scanned into digital format. This work focuses on high-quality raster drawings, excluding vectorized drawings, as their contents can be extracted using traditional IT technology without applying Optical Character Recognition (OCR) tools.

Due to legacy reasons, there is still a considerable number of mechanical drawings in the form of raster images that are used in industrial practice. According to [3], 250 million new drawings are generated each year and millions of legacy drawings are still used in circulation. There, mechanical drawings contain only pixel information and it is still a challenge for the typical CAD software to retrieve geometric and textual information, such as dimensions, tolerances, and heat treatment from them. A solution that allows the extraction of this information from mechanical drawings can automate the entire production process, including measuring and quality control [2].

Digitization of raster drawings

An examination of the construction of these mechanical drawings is required to understand what the digitization of raster drawings means. As mentioned earlier, the drawings contain Product Manufacturing Information (PMI) in addition to geometric information. Digiting mechanical drawings therefore means that both the type of information object (e.g. dimensions, tolerances, information about materials, etc.) and its content must be extracted. The aim is to convert the mechanical drawing into a structured format (e.g. json file or xml file) that can then be loaded and processed by down-streaming CAD/CAM software tools.

Figure 1 gives examples of two important types of PMI: size dimensions and Geometric Dimensioning and Tolerancing (GD&T). As can be seen, the look of dimensional information may differ depending on whether default or explicit tolerances apply. GD&T, on the other hand, is always presented in a table-like format, which may consist of single or multiple rows and columns. Cells in GD&T contain typing symbols, nominal values and tolerances, and sometimes a reference. This article focuses on these two types of PMI. However, the proposed framework can be extended to recognize other types of PMI as well.

Figure 1: Examples of two important types of Product Manufacturing Information.

Conventional OCR software does not work well

An obvious solution would be to simply use off-the-shelf OCR software for the task. However, OCR software is normally trained on normal text data such as books, newspapers, handwritten documents etc. and not on mechanical drawings. Is it plausible to expect good results with mechanical drawings?

To investigate this question, experiments were conducted with OCR software. In a preliminary test, Tesseract OCR [4] performed best, which is why it was determined to investigate this software solution in detail for the task.

Typically, there are two sub-steps that need to be performed when digitizing drawings: First, the document must be segmented, i.e. the areas containing objects of interest (e.g. a GD&T or a dimension) must be localized. Subsequently, the individual characters depicted in this area must be recognized and extracted.

Tesseract OCR follows this two-stage approach. This evaluation therefore tests the software using three different tasks:

1. Localization: How well does Tesseract manage to find the areas of GD&T or dimensions, two of the most frequently used Product Manufacturing Information items in quality control tasks?

2. Line recognition: How well does Tesseract succeed at recognizing the characters inside the corresponding image areas?

3. Localization + line recognition: How well does Tesseract perform on the overall task?

A total of 14 real mechanical drawings were available for the evaluation. For each task, a ground truth dataset was created by human annotators. As Tesseract does not have separate classes for the GD&T or dimensions specific to drawings, a heuristic rule had to be applied in order to compare Tesseract's result with the ground truth. The heuristic rule checks the Intersection over Union (IoU) of each predicted bounding box with all ground truth bounding boxes. IoU is a metric to measure the localization accuracy of predicted bounding boxes. If the maximal IoU is bigger than a threshold value (set to 0.05 here), the predicted bounding box takes the class of the ground truth bounding box, which has the maximal IoU. Otherwise, the predicted bounding box is considered to be "background".

Figure 2 summarizes the evaluation results. For Task 1 (localization), precision, recall, F1 and IoU on the pixel level of localization results are used as metrics to measure the performance. Precision measures what proportion of the assigned labels is correct, while recall measures what proportion of the objects that should receive a certain label have actually received this label. Because there is a trade-off between these two measures, the F1 measure is also calculated, which considers both measures in combination. As can be seen, Tesseract achieves a relatively high precision and a low recall. This means that areas that are detected are correct in many cases, but at the same time many areas containing relevant information cannot be found at all. The low scores of the IoU value reflect the low overlap between the ground truth bounding boxes and the bounding boxes predicted by Tesseract.

For Task 2 (line recognition), the test data set consists of 1411 line images (a single sequence of characters in one row). Tesseract's performance was evaluated based on a set of line images containing only GD&T information and another set of all line images containing any textual information. As the task is to extract the characters shown on the images, Character Error Rate (CER) and Word Error Rate (WER) are used as evaluation measures. While Tesseract performs quite well on a test data set containing all types of line images, its performance drops significantly when being tested specifically for GD&T information.

The overall test (Task 3, which includes over 3700 characters) is unsatisfactory across all scores. This is to be expected due to the relatively low performance of the individual tasks, namely localization and line recognition. Remember the low recall of task 1. If an area of interest is not recognized at all, its information cannot be extracted, which leads to poor performance in subsequent tasks.

Figure 2: Evaluation results.

In conclusion, the performance of the off-the-shelf OCR software Tesseract is not high enough for the task of digitizing mechanical drawings. One reason for this may be the fact that mechanical drawings contain special objects such as GD&T, dimensions, tolerances, and special characters, such as ∠ ↗ ○◎⊥⌖⌒⌓⌰ⒻⓁⓂⓅⓈⓉⓊ, are not present in normal documents, nor are they used in the training task of normal OCR software.

Mechanical drawings differ significantly from ordinary documents in that they feature extensive white space, geometric shapes, special fonts, and purely binary black-and-white pixels. It is likely that Tesseract's model did not see these features during training and was therefore unable to adapt to them.

Deep learning for the digitization of drawings

In order to overcome the problems of off-the-shelve OCR software, a new framework was developed to digitize mechanical drawings with the focus on GD&T and dimensions. The framework is composed of a centralized localization module and several down-streaming pipelines. The localization module segments objects of interest from an entire mechanical drawing. Depending on the type of segmented object, a corresponding downstream pipeline is invoked. Each pipeline takes the cropped image generated by the localization module as input and recognizes the text-based content of each segmented image. In the last step, the results from different pipelines are merged, which forms the final recognition result of the entire mechanical drawing. The proposed framework originates from the work [1]. Compared to their work, however, a centralized deep learning-based localization step is proposed and we update the GD&T pipeline by a cell detection step and a text recognition step. Figure 3 illustrates the framework.

Figure 3: Proposed framework for extracting Product Manufacturing Information from drawings.

The localization module is implemented with a U-Net [5], a semantic segmentation network, which classifies the image on the pixel level. A post-processing step is implemented to estimate the bounding box and its orientation. The estimated bounding boxes are allowed to have angles to ensure that non-horizontal PMI can be localized as well. The network is trained with an internal data set, which contains real CAD drawings. The locations of the objects of interest were labeled by human annotators. Instead of using other common object detection networks like YOLO or Faster R-CNN, which are used in other work [6–9], U-Net comes with the advantage that images with arbitrarily large dimensions can be directly processed. Note that comparing the performance of implemented networks with others published in the literature for typical tasks in computer vision, e.g. object detection, is beyond the scope of this paper.

Analogous to [10], the GD&T pipeline is composed of a cell detection step and a text recognition step. The separation of the GD&T pipeline into two steps is motivated by the observation that table borders inside GD&T objects decrease the performance of the recognition results. For the cell detection step, a U-Net [5] trained on generated data is employed. Similarly, the dimension pipeline starts with a segmentation step as well which localizes the positions of nominal values and tolerances. In the dimension pipeline Tesseract OCR is used for this step.

In both cases, the preprocessing step results in line images (segments containing a single sequence of characters) which are then sent to the shared line recognition model to recognize the content. Off-the-shelf OCR software is not used for this step; rather, an R-CNN [11] is trained from scratch using a synthetic dataset of approximately 8 million image lines. The reason for this is the large set of unusual characters that appears in mechanical drawings and the low performance of off-the-shelf OCR software. This framework has the advantage that the localization stage and individual pipelines can be designed and implemented in a modularized way. Individual pipelines can be updated and replaced, if more advanced technology exists in the future. Additional pipelines can also be added to the existing framework, if additional types of objects are to be recognized.

Evaluation of the deep learning framework

To allow comparison, the proposed deep learning framework is evaluated on the same tasks as the ones for Tesseract. The evaluation results are also shown in Figure 2.

As can be seen, the precision value for Task 1 (localization) deteriorates slightly with the framework by 3.0% (GD&T) and 2.1% (dimension). At the same time, however, there is a significant improvement in recall (from 29.4% to 86.2% for GD&T, from 25.8% to 56.5% for dimension) and IoU (from 27.9% to 72.6% for GD&T, from 23.4 to 45.3% for dimension). A higher recall is important for the task, as nothing can be extracted in the subsequent steps from segments that are not found at all. This information would therefore be missing in the digitized version of the mechanical drawing at the end. As expected, this is also reflected in higher values for the IoU score and the F1 measure. While the precision and recall values for GD&T are satisfactory, we still aim to improve the localization performance for dimensions further.

Tesseract has great difficulty with the second task, which is to recognize characters in each line segment, especially for the ones in GD&T segments. Here, the self-developed R-CNN for text recognition task, which is trained on the specially generated drawing data, reduces the

CER from 51.2% to 13.5%, and the WER from 67.4% to 24.3%. At the same time, a slight deterioration in the word recognition rate for the overall dataset has to be accepted.

The achieved improvements are also reflected in the evaluation results of Task 3, in which the overall system is tested. Significant improvements were achieved for both types of product manufacturing information. Nevertheless, it must be said that the performance for the dimension type is still below expectations. One explanation is that the new system is not yet good enough at segmenting elements of this type. This may be due to an insufficient amount of training data, which makes it difficult for the deep neural network to recognize the peculiarities of this variant-rich type of Product Manufacturing Information.

The key to success

The digitization of legacy mechanical drawings is and remains an important task in the automation of production processes. Conventional off-the-shelf software often does not achieve good results, presumably because the underlying models have been trained on different data and are therefore not well adapted to the special characteristics of drawing data. It is therefore worth developing and using specialized software for this task. A multi-level framework, in which different PMI types are handled separately, provides the opportunity to address the specific characteristics of each PMI and thus significantly improves quality. In addition, the proposed modular structure makes it possible to replace individual components at any time if better technical solutions become available.

The project underlying this paper was funded by the Federal Ministry of Education and Research (BMBF) and the State of Baden-Württemberg under grant number 03FHP127B. The authors are responsible for the content of this publication.

Bibliography

[1] Toro, J.; Wiberg, A.; Tarkian, M.: Optical Character Recognition on Engineering Drawings to Achieve Automation in Production Quality Control. In: Frontiers in Manufacturing Technology 3 (2023).
[2] Scheibel, B.; Mangler, J.; Rinderle-Ma, S.: Extraction of Dimension Requirements from Engineering Drawings for Supporting Quality Control in Production Processes. In: Computers in Industry 129, pp. 103442 (2021).
[3] Henderson, T.: Analysis of Engineering Drawings and Raster Map Images. Springer 2014.
[4] Kay, A.: Tesseract: An open-source optical character recognition engine. In: Linux Journal 159 (2007).
[5] Ronneberger, O.; Fischer, P.; Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention 9351, pp. 234–241 (2015).
[6] Schlagenhauf, T.; Netzer, M.; Hillinger, J.: Text detection on technical drawings for the digitization of brown-field processes. In: 16th CIRP Conference on Intelligent Computation in Manufacturing Engineering 118, pp. 372–377 (2023).

[7] Haar, C.; Kim, H.; Koberg, L.: AI-based engineering and production drawing information extraction. In: Flexible Automation and Intelligent Manufacturing: The Human-Data-Technology Nexus, Springer International Publishing, pp. 374–382 (2023).
[8] Nguyen, T. et al.: Object detection and text recognition in large-scale technical drawings. In: Proceedings of the 10th International Conference on Pattern Recognition Applications and Methods. SciTePress, pp. 612–619 (2021).
[9] Elyan, E.; Jamieson, L.; Ali-Gombe, A.: Deep learning for symbols detection and classification in engineering drawings. In: Neural Networks 129, pp. 91–102 (2020).
[10] Hashmi, K. A. et al.: Current status and performance analysis of table recognition in document images with deep neural networks. In: IEEE Access 9, pp. 87663–87685 (2021).
[11] Shi, B.; Bai, X.; Yao, C.: An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. In: IEEE Transactions on Pattern Analysis and Machine Intelligence 39, pp. 2298–2304 (2017).

Keywords:
deep learning, optical character recognition (OCR), mechanical drawings, quality control, object detection, Product Manufacturing Information