2. Related literature

The automated detection of symbols in engineering drawings and architectural drawings is an active research area with many proposed solutions. Most involve some heuristics (Ablameyko et al. 2007), and more recent approaches mainly use deep learning methods (Moreno-García et al. 2019).

In the field of engineering, piping and instrumentation diagrams (P&IDs) are commonly used to represent complex systems. Elyan et al. (2018) combine various heuristic methods to localize symbols in P&IDs. Their approach includes thresholding, blurring, circle Hough transform, and text/graphics separation. The identified symbols are then classified with Convolutional Neural Networks (CNN), random forests, and support vector machines. In Elyan et al. (2020), the same task is addressed with YOLO (Redmon et al. 2016), which yields a robust bounding box-based symbol detector for the document type at hand (cf. Figure 1). While Nurminen et al. (2020) (cf. Figure 1) and Gupta et al. (2022) also use YOLO versions for symbol detection in P&IDs, they present approaches for enlarging the training data set with synthetically generated images.

A related research field is the detection of symbols in architectural drawings, which mainly represent building components or interior. A prominent public data set in this regard is SEYSD (Delalandre et al. 2010) which is used for evaluation in various studies. For instance, Rezvanifar et al. (2020) leverage YOLOv2 (Redmon & Farhadi 2017) to detect doors, windows, appliances, and furniture in the drawing files (cf. Figure 1). Mishra et al. (2021) train a Cascaded Mask R-CNN (Cai & Vasconcelos 2018) for symbol detection in floor plans and propose their dataset for the task called Synthetic Floor Plan Images (SFPI). Another study published by Ziran et al. (2020) suggests using Faster R-CNN (Ren et al. 2017) for the task.

It is evident that, while object detection in engineering and architectural drawings is a research field often addressed, most existing studies suggest the use of bounding box-based

3. Methodology

The proposed approach is evaluated by comparing the performance of different neural network architectures. Therefore, two state-of-the-art architectures are used: Keypoint R-CNN, a specialization of Mask R-CNN, and YOLO-Pose, based on YOLOv7 (Wang et al. 2022). The performances of these architectures are compared to a baseline model consisting of the object detection network Faster R-CNN and a custom regression model. The networks are trained on a set of synthetically generated training images and tested on a set of real-world drawings. Figure 2 summarizes the methodology of this paper.

3.1 Data set

A symbol usually comprises two components: first, a marker, which indicates the position and orientation of the symbol, and second, the accompanying text, which contains additional information such as dimensions or references. To accurately assess a symbol, the marker and the text component must be reliably recognized and linked to each other. Therefore, to

Therefore, in this study, state-of-the-art neural networks are trained for keypoint detection. To

Table 1: Object detection performance scores (left) and keypoint detection performance scores (right) achieved by the models on the real-world bridge construction drawings. Scores are given in mean Average Precision (mAP). Details about the score computations are given in (Padilla et al. 2022).

6. Conclusion

In conclusion, this paper demonstrates the efficacy of keypoint-based object detection models, such as Keypoint R-CNN and YOLOv7-Pose, for symbol pose estimation in technical drawings. The study compares these models with a two-stage baseline approach. The results indicate that the keypoint-based models outperform the baseline approach in terms of both keypoint detection accuracy and bounding box detection accuracy. In addition, the keypoint detection also allows for the automatic linkage of symbol marker and reference, while determining the exact position of the symbol. This improves upon other object detection-based approaches, such as Faltin et al. (2022), which require detecting three objects per instance and the rule-based linkage. However, the proposed method and its results are limited to one drawing style and may not generalize to other domains or drawing collections.