On P&ID Symbol Recognition using Transformers

Leonardo K. T. da Cunha[1], Eduardo F. Leite[1], Lucas Eduardo C. Morais[1], Alessandra Harrison[2], João F. Valiati[1]

[1]SENAI Innovation Institute for Sensing Systems (ISI-SIM)
São Leopoldo – RS – Brazil

[2]Brazil Technology Digital Hub – SHELL Brazil
Rio de Janeiro – RJ – Brazil.

{leonardo.treichel,eduardo.leite,lucas.morais}@senairs.org.br

alessandra.harrison@shell.com,joao.valiati@senairs.org.br

Abstract. *Machine-interpretable graphs of P&IDs are essential to advance automation and digital twins; however, most diagrams remain in vector formats like DWG, limiting automated use. Conversion is further hindered by symbol variability, scarce annotations, and the lack of openly available datasets. This work addresses these challenges by evaluating transformer-based detection of P&ID symbols with DETR, comparing backbones pretrained on ImageNet, COCO, and a domain-specific synthetic data. The study analyzes their generalization from synthetic to real diagrams, with evaluations on real world datasets (OPEN100, IPD) showing that domain-specific pretraining consistently improves performance of about 1 mAP point, underscoring its value for robust industrial diagram interpretation.*

1. Introduction

Piping and Instrumentation Diagrams (P&IDs) are fundamental engineering documents that encode equipment, process flow, instrumentation, and control logic throughout the lifecycle of industrial facilities [Toghraei 2019]. An example is shown in Figure 1.

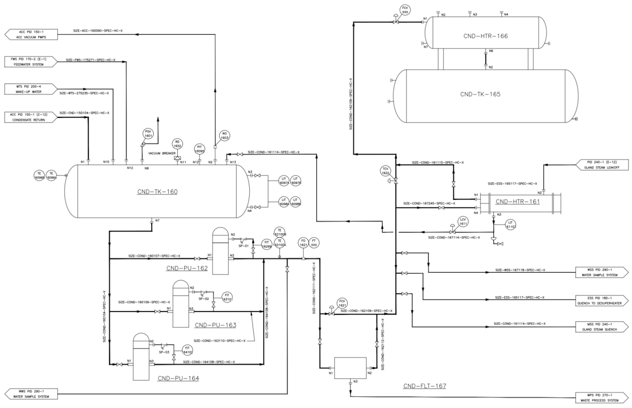

Figure 1. P&ID 160 *"Condensate System"*, from OPEN100

Although modern diagrams are created in CAD environments and recent standards such as DEXPI, together with frameworks like pyDEXPI [Goldstein et al. 2025], aim to improve interoperability, large archives remain locked in raster or vectorized formats

(PDF, DWG). Converting these into structured representations is challenging but critical for enabling automation. This task requires robust symbol detection, text recognition, and connection extraction, yet progress is hindered by inconsistent standards, symbol variability, and the lack of public annotated datasets [Paliwal et al. 2021, Stürmer et al. 2024].

Earlier approaches to P&ID digitization typically employed modular pipelines, combining CNN-based symbol detectors with OCR and rule-based heuristics [Paliwal et al. 2021, Kim et al. 2022], or template-matching and edge-detection methods [Ghadekar et al. 2021]. More recently, transformer-based models such as Relationformer [Stürmer et al. 2024] have shown promise in jointly extracting symbols and their relationships. Given the scarcity of public datasets and the data requirements of transformer models, recent work has turned to synthetic generation to produce large-scale training corpora [Stürmer et al. 2023, Rupprecht et al. 2025, Gao et al. 2025].

Building on these advances, this work explores transformer-based object detection for symbol recognition in P&IDs, focusing on the DEtection TRansformer (DETR) [Carion et al. 2020]. Unlike conventional detectors that rely on region proposals, DETR frames detection as a direct set prediction problem, leveraging a CNN backbone with a transformer encoder–decoder to capture global context. The proposed work investigates an unexplored phenomena on literature, how different backbone pretraining strategies—ImageNet, COCO, and PIDClassify—affect generalization from synthetic to real diagrams, systematically evaluating transformer-based detection in this industrial context.

2. Methodology

This study investigates DETR with different pretraining strategies for symbol recognition in P&IDs, using synthetic data for training and real-world diagrams for testing. The synthetic dataset was generated using two layout patterns: grid-based symbol placement and automatic diagram layouts via the HOLA algorithm [Kieffer et al. 2016], yielding 300 grids and 700 complete P&ID sheets across 40 symbol classes, sampled with uniform frequency. The dataset was randomly split into 60% training, 20% validation, and 20% test subsets, with each sheet divided into $640\,\mathrm{px} \times 640\,\mathrm{px}$ crops at 50% overlap. Two real-world datasets were reserved exclusively for evaluation: the PID2Graph OPEN100 subset [Stürmer et al. 2024], containing 12 P&IDs, and IPD, an internal set of two sheets.

The base model follows the original DETR [Carion et al. 2020], combining a ResNet-50 backbone [He et al. 2016] for feature extraction with a transformer encoder–decoder for context modeling. Three backbone pretraining strategies were compared, each trained for 300 epochs with data augmentations including random flips, rotations, translations, scaling, Gaussian blur, color jitter, contrast adjustments, and elastic transforms: (i) ImageNet pretraining [Russakovsky et al. 2015], (ii) fine-tuning a COCO-pretrained DETR [Lin et al. 2014], and (iii) PIDClassify, referring to pretraining the backbone as a classifier for the 40 symbols used in the synthetic dataset. Performance was evaluated using mean Average Precision (mAP) averaged over 50% to 95% intersection-over-union thresholds on the proposed test sets.

3. Results and Discussion

Training each DETR variant on the synthetic dataset took approximately 70 hours on an NVIDIA GeForce RTX 3060 GPU. The mAP on the test sets is presented in Table 1.

Table 1. Test mAP(%) for different pretraining strategies (higher is better)

Backbone pretrain	Synthetic	IPD	OPEN100
COCO	90.5	33.8	4.0
ImageNet	52.4	25.1	1.9
PIDClassify	**91.6**	**35.3**	**4.9**

The results show that PIDClassify initialization slightly outperforms on every test dataset, with COCO-pretrained performing marginally worse and ImageNet-pretrained lagging far behind. These findings suggest that using a pretrained backbone within the target domain improves generalization in symbol detection. Moreover, the performance gap between COCO and ImageNet pretraining indicates that leveraging models trained on domains with greater similarity may yield superior results, highlighting the potential of large-scale, diverse training. Figure 2 illustrates qualitative detections from the PIDClassify-pretrained model on the OPEN100 dataset.

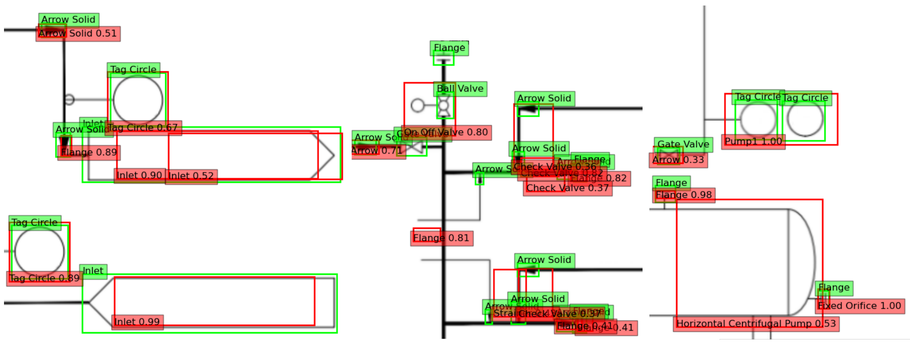

Figure 2. Predictions (red) compared with ground-truth (green) on OPEN100

While the model accurately detects isolated symbols, its performance degrades under complex conditions. Background elements, high proximity, overlap, low-resolution regions and visually similar components significantly disrupt detection. The resulting errors, such as partial detections, false positives, and missed symbols, seem to be related to the under-representation of these challenges in the synthetic training data. Notably, even correctly localized symbols are frequently misclassified due to inter-class confusion.

4. Conclusion

This work evaluated transformer-based object detection for symbol recognition in P&IDs, focusing on the DETR model with three backbone pretraining strategies: COCO, ImageNet, and PIDClassify. Training was performed on a synthetic P&ID dataset and testing on the OPEN100 and IPD benchmarks to study the synthetic-to-real transfer gap. Results underscore the impact of pretraining choice on detection performance, with domain-adapted features yielding improvements of approximately 1 mAP point over all three test datasets. Limitations include the realism of synthetic data, the considerable gap between domains, and challenges in detecting fine-grained symbol variants. Future work will explore improved synthetic generation, hybrid synthetic-real dataset strategies, domain adaptation techniques, and integration into automated pipelines for P&ID graph extraction. Additionally, it is envisioned the possibility of making the IPD publicly available, which could foster benchmarking, reproducibility, and further progress in this domain.

Acknowledgments

This research was carried out in association with the ongoing R&D project registered as ANP nº 24520-9 “3D APT: 3D Automated Perception and Tagging” (SENAI/Shell Brazil/ANP), sponsored by Shell Brasil Petróleo Ltda under the ANP R&D levy as “Compromisso de Investimentos com Pesquisa e Desenvolvimento”. We also thank all contributors from SENAI which are not named authors, but helped develop tasks related to the results shown.

References

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). End-to-end object detection with transformers. In *Computer Vision – ECCV 2020*.

Gao, Q., Yang, H., Theisen, M. F., and Schweidtmann, A. M. (2025). Accelerating process synthesis with reinforcement learning. *Computers & Chemical Engineering*.

Ghadekar, P., Joshi, S., Swain, D., Acharya, B., Pradhan, M. R., and Patro, P. (2021). Automatic digitization of engineering diagrams using intelligent algorithms.

Goldstein, D. P., Balhorn, L. S., Alimin, A. A., and Schweidtmann, A. M. (2025). pyDEXPI. In *The 35th European Symposium on Computer Aided Process Engineering*.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Kieffer, S., Dwyer, T., Marriott, K., and Wybrow, M. (2016). HOLA: Human-like orthogonal network layout.

Kim, B. C., Kim, H., Moon, Y., Lee, G., and Mun, D. (2022). End-to-end digitization of image format piping and instrumentation diagrams at an industrially applicable level.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO. In *Computer Vision – ECCV 2014*.

Paliwal, S., Jain, A., Sharma, M., and Vig, L. (2021). Digitize-PID: Automatic digitization of piping and instrumentation diagrams. In *Trends and Applications in Knowledge Discovery and Data Mining*. Springer International Publishing.

Rupprecht, S., Hounat, Y., Kumar, M., Lastrucci, G., and Schweidtmann, A. M. (2025). Text2model.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*.

Stürmer, J. M., Graumann, M., and Koch, T. (2023). Demonstrating automated generation of simulation models from engineering diagrams. In *2023 International Conference on Machine Learning and Applications (ICMLA)*.

Stürmer, J. M., Graumann, M., and Koch, T. (2024). Transforming engineering diagrams: A novel approach for P&ID digitization using transformers.

Toghraei, M. (2019). *Piping and Instrumentation Diagram Development*. Wiley.