TOPOPID: RESIDUAL-GUIDED EQUIPMENT RECOGNITION AND GRAPH COMPLETION IN P&IDS

Anonymous authors
Paper under double-blind review

ABSTRACT

Digitizing a piping and instrumentation diagram (P&ID) requires an auditable process graph, not just detected symbols. Project-specific equipment often lacks visual exemplars: closed-vocabulary pipelines preserve structure but miss such devices, whereas whole-page vision–language models (VLMs) face dense visual search and cannot directly authorize graph edits. We formulate recovery as residual-guided constrained semantic completion of a partially observed graph. In TopoPID, unresolved topology localizes equipment-scale candidates, a project roster constrains local VLM verification, and deterministic rules govern evidence-linked graph updates. On 45 held-out industrial P&IDs, with target-equipment images excluded from detector development, prompt demonstrations, and candidate tuning, identity-aware recognition F1 rises from 19.44% for whole-page inference to 93.77%; 489/524 devices are correctly recovered. Without adaptation, the same configuration recovers 54/62 task-compatible devices (86.40% F1) on 12 OPEN100 drawings. These results establish device recovery and useful graph association retrieval while retaining an auditable write-back path.

1 INTRODUCTION

Piping and instrumentation diagrams (P&IDs) encode equipment, lines, instruments, and connectivity. Recovering these elements supports relation retrieval and hazard and operability (HAZOP) preparation (Lee et al., 2026; Elhosary & Moselhi, 2026b;a). Missing equipment breaks process paths; unsupported edges can fabricate them. Both recovered relations and their visual or rule provenance must therefore remain inspectable.

Modular digitization pipelines detect symbols and text, recover lines, and assemble relations with explicit rules (Rahul et al., 2019; Kang et al., 2019; Kim et al., 2021; 2022; Paliwal et al., 2021a). Their structure is inspectable, but recognition is bounded by visual classes. Project-specific equipment may occur once or as a large composite outline. Here, *unseen equipment* means a roster-enumerated identity without a target-specific visual exemplar in detector development or prompt demonstrations. Whole-page VLMs can interpret unfamiliar names and appearances, but dense sheets require simultaneous search, disambiguation, and localization; public graph supervision is limited (Stürmer et al., 2025), and blueprint symbols can confuse general VLMs (Shteriyanov et al., 2025).

TopoPID separates these responsibilities. A conservative partial graph leaves unexplained line endpoints as *topology residuals*, which select equipment-scale image regions. A project roster bounds local semantic alternatives; deterministic rules govern how an accepted device alters the graph (Figure 1). Thus topology determines *where to look*, the roster *what may be present*, and graph rules *how to write back*.

With a fixed VLM, roster, prompt, decoding, and matching rule, correctly recognized enterprise devices rise from 93/524 for whole-page inference to 383/524 for endpoint crops and 489/524 for TopoPID. The corresponding identity-aware F1 scores are 19.44%, 80.38%, and 93.77%. Without adaptation, TopoPID reaches 86.40% F1 on compatible OPEN100 devices; its completed enterprise graphs retrieve 1,685/1,763 reviewed instrument–equipment associations.

Our contributions are:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

- We formulate unseen, roster-enumerated equipment recovery as graph completion with explicit spatial, semantic, and update constraints.
- We combine topology-residual localization, roster-grounded local verification, and provenance-preserving deterministic write-back.
- We evaluate identity-correct device recovery under fixed-verifier comparisons, zero-adaptation public transfer, and an engineering association query.

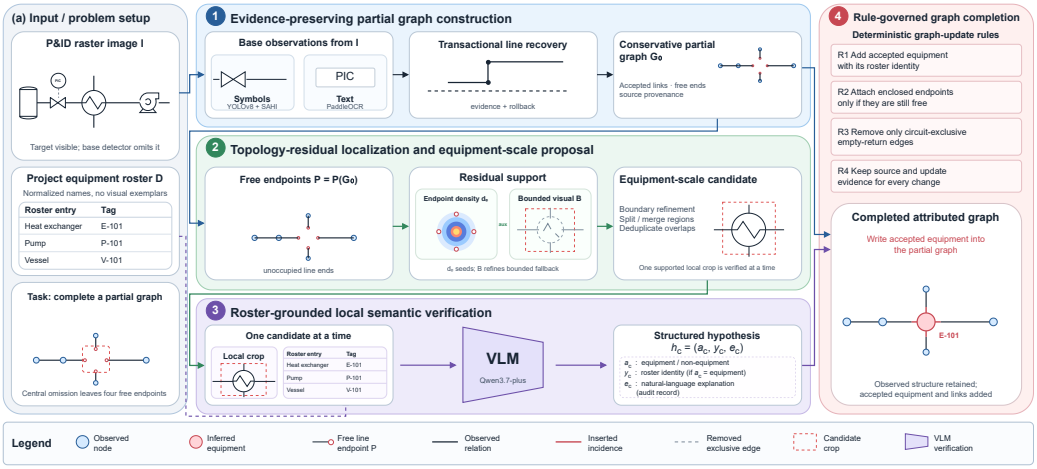

Figure 1: TopoPID flow. A partial graph exposes free-endpoint residuals; bounded visual evidence refines equipment-scale candidates. A VLM verifies local crops against a project roster. Only accepted candidates enter deterministic, evidence-linked node and edge insertion with circuit-exclusive cleanup. The illustrated roster labels are synthetic.

2 Related Work

P&ID digitization and graph recovery. P&ID systems commonly decompose recognition into symbols, text, lines, and relation assembly (Yu et al., 2019; Kang et al., 2019; Yun et al., 2020; Kim et al., 2021; Moon et al., 2021; Kim et al., 2022). Integrated symbol–text recognition has also been studied (Byun et al., 2025). Other work combines deep recognition with graph search (Mani et al., 2020; Paliwal et al., 2021a), or predicts objects and relations more jointly (Shit et al., 2022; Stürmer et al., 2025). Prasad & Mahapatra (2026) train a graph predictor on topology-preserving synthetic drawings and evaluate it on OPEN100, using edge mAP rather than our identity-aware equipment metric. Related process-flow-diagram systems recover equipment and connections (Theisen et al., 2023). These recognition pipelines primarily assume representative supervision for the object inventory. Zhu et al. (2026) instead separate multimodal tag extraction from language-model topology inference using a tag list and process knowledge. TopoPID starts from a conservative partial graph: unexplained endpoints select local regions before semantic verification, and accepted identities enter the graph only through evidence-linked deterministic updates.

Line semantics and evidence preservation. Generic line and wireframe methods recover segments and junctions (Grompone von Gioi et al., 2012; Zhou et al., 2019; Xue et al., 2020; Xu et al., 2021), while engineering drawings also encode semantics through dash patterns, width, arrows, and conventions (Moon et al., 2021; Han et al., 2024; 2025). This distinction matters because an unsupported line merge can erase a useful free endpoint, whereas a false gap crossing can create a seemingly valid route. TopoPID therefore retains observed segments, inferred group spans, accepted relations, and rejected alternatives as different evidence types.

Open semantics in structured visual recognition. One- and few-shot engineering-symbol methods still require a target contour or support image (Paliwal et al., 2021b; Jamieson et al., 2024). Vision–language pretraining and grounded open-vocabulary detectors broaden category transfer

(Radford et al., 2021; Li et al., 2022; Zhou et al., 2022; Minderer et al., 2022; 2023; Liu et al., 2024; Cheng et al., 2024), while category-agnostic segmentation provides general spatial proposals (Kirillov et al., 2023). Engineering identity and connectivity require further evidence before a proposal enters the graph. Context-aware recognition uses relations to disambiguate unseen objects (Luo et al., 2020), while open-vocabulary scene-graph models predict novel relations (Liu et al., 2025). TopoPID uses structure as an inference constraint: partial topology contracts the spatial search space before semantic prediction, and graph rules constrain the structural consequences after prediction.

3 Problem Formulation and Method

3.1 Constrained Completion Interface

Let I be a raster P&ID and D a project equipment roster containing normalized equipment names but no visual exemplars. A conservative base pipeline produces

$$G_0 = (V_0, E_0, \Pi_0), \tag{1}$$

where V_0 contains known symbols, text observations, line groups, and accepted junctions; E_0 contains accepted continuity and incidence relations; and Π_0 maps every graph element to its visual or rule evidence. The graph is partially observed because equipment categories outside the base detector remain latent.

The task is to infer a completion $\widehat{G}$ that adds localized equipment nodes and supported incidence relations while preserving the inspectability of G_0. Target-equipment images are excluded from detector training, candidate-parameter selection, prompt demonstrations, and project-level development data. The upstream symbol detector, OCR, and line recovery remain supervised or engineered; the open-semantic requirement applies specifically to the missing equipment that must be completed from the drawing and roster.

Let $P = P(G_0)$ be free endpoints of non-artifact line groups; rejected relations are also retained as residual evidence. Let $B = B(I, G_0)$ collect bounded visual cues: unexplained closed regions and curved, oblique, or fragmented foreground, together with known-symbol cores and endpoint-free blank regions used to suppress spurious support. These cues refine proposals and supply lower-priority fallbacks but cannot independently authorize insertion. For candidate c, write $h_c = \mathcal{V}(I[c], D)$, with a_c its acceptance field. Candidate generation $\mathcal{C}(P, B)$, verification $\mathcal{V}$, and deterministic write-back $\mathcal{W}$ compose as

$$\widehat{G} = \mathcal{W}(G_0, \{(c, h_c) : c \in \mathcal{C}(P, B),\ a_c = 1\}). \tag{2}$$

Here $I[c]$ is the crop and h_c has the schema in Equation 5. Only accepted hypotheses enter $\mathcal{W}$, whose provenance-preserving transitions are defined in Section 3.5. Topology supplies proposals, the roster supplies semantic scope, and rules commit graph edits.

3.2 Evidence-Preserving Partial Graph

The partial graph is the structural state on which constrained inference operates. It exposes accepted observations, unresolved endpoints, rejected relations, and provenance so that a missing device can be distinguished from an upstream merge or an unsupported connection. The particular recognizers are replaceable provided that they preserve this interface. Our implementation constructs the state as follows.

Base observations. A YOLOv8 detector with slicing-aided hyper inference (SAHI) recognizes frequent compact symbols at native scale (Akyon et al., 2022); PaddleOCR detects, orients, and recognizes text (Liao et al., 2020; Du et al., 2022). Their boxes and polygons are stored as observations and masks. Skeletonization with horizontal/vertical traversal extracts axial line seeds from a binary image in which symbol and text interiors are suppressed. An unmasked binary image is retained separately as evidence for testing continuity across a mask.

Transactional line recovery. Each seed retains its identity, orientation, span, and robust width band. Tracing extends compatible foreground, crosses a symbol or OCR mask only when the unmasked image contains a compatible core, records aligned blank intervals as possible dash gaps, and reconnects a single perpendicular interruption only when balanced and pixel-supported. Ambiguity stops the trace. A successful extension atomically commits geometry, group membership, and its evidence record; failure restores the previous state and records a reason. Inferred group spans are never relabeled as observed seeds. Groups are classified as continuous, dashed, or artifact from run/gap lengths, fill ratio, width stability, and recorded gap causes.

Conservative relations. Non-artifact line groups and known symbols form G_0. Pixel-supported orthogonal candidates become turns or T-junctions; unmarked interior crossings remain disconnected. Directional endpoint projection proposes line–symbol incidence, supplemented by side support and short raw-pixel links. Distance, alignment, pixel support, and endpoint occupancy resolve conflicts. An endpoint consumed by a turn cannot simultaneously attach to a symbol; weaker alternatives remain explicit rejections. OCR suppresses residual text strokes but cannot trigger an equipment candidate. This conservative policy deliberately prefers an unresolved endpoint to an unsupported edge that would hide a missing-device hypothesis.

3.3 Topology-Residual Localization

The structural-support operator $\mathcal{C}$ is realized by residual densities and bounded completion evidence. We quantize the free endpoints P on a regular image grid and compute

$$d_e(x) = \sum_{p \in P} \mathbf{1}[\|x - p\|_\infty \le r], \tag{3}$$

where r is a fixed image-space neighborhood radius. Long process lines contain many pixels but few unoccupied endpoints; an omitted equipment body often terminates several recovered lines or internal strokes within one neighborhood. Connected high-density basins of d_e therefore provide the candidate backbone without matching an unknown outline to a template.

Sparse-port devices and drawing furniture require limited visual support. We define

$$\begin{aligned} d_+(x) &= \min\{d_c(x) + d_s(x), \alpha\}, \\ d_-(x) &= \min\{d_y(x) + d_b(x), \beta\}, \\ d_u(x) &= [d_e(x) + d_+(x) - d_-(x)]_+, \end{aligned} \tag{4}$$

Here $[z]_+ = \max\{z, 0\}$. The terms d_c and d_s encode symbol-free closed regions and residual curved, oblique, or fragmented foreground; d_y and d_b encode known-symbol cores and large endpoint-free blank regions. OCR polygons suppress text strokes. Caps α and β prevent visual residuals from dominating topology. Candidate growth uses dense basins of d_e; d_u ranks proposals and supports a lower-priority fallback. This asymmetry prevents a change in residual pixels from erasing a supported topology basin.

Compact basins yield padded endpoint boxes. Local maxima split oversized basins, while paired long boundaries may merge separated ports of one large device. Residual foreground may enlarge an existing box only under a bounded area-growth rule and after long through-lines are removed. Rule-checked rectangle or vertical-U completion can close an open outline, again only after a topology-supported candidate exists. Endpoint-free bands split overextended regions; known-symbol boundaries limit padding. Deduplication stores which endpoints, density cells, residual components, completion rules, fallbacks, and suppressions produced every candidate.

3.4 Roster-Grounded Local Verification

The semantic operator $\mathcal{V}$ is a local verifier. Given a candidate crop, a fixed instruction, and roster D, it returns

$$h_c = (a_c, y_c, e_c), \tag{5}$$

where a_c is the equipment/non-equipment decision, y_c is a roster identity hypothesis for a positive case, and e_c is explanation material retained for audit. The crop preserves nearby strokes but removes most page-scale clutter, and the roster supplies semantic scope rather than a visual template. We use

Qwen3.7-Plus for local verification and all comparison methods. Model version, instruction, roster representation, and decoding settings are fixed for evaluation.

The verifier input comprises the crop, fixed instruction, and roster; OCR strings and local graph descriptors contribute through upstream suppression and localization. Every enterprise equipment instance has an engineer-confirmed roster identity. A positive candidate is semantically correct when it is accepted as equipment and its normalized output name matches that identity; a negative candidate is correct when it is rejected as non-equipment. Candidate semantic-verification accuracy evaluates these two outcomes jointly.

3.5 Rule-Governed Graph Completion

An accepted crop creates one equipment node referencing candidate and verifier evidence. Every still-free endpoint enclosed by that region receives an incidence edge citing its source. These rules, not the VLM, determine the graph transition.

An accepted region can enclose portions of the equipment outline that line extraction represented as a line-only return circuit. An *empty equipment-returning circuit* starts and ends at the new equipment and contains only line or junction nodes, with no recognized symbol. Let E_{ins} be inserted incidence edges, E_{empty} line edges on such circuits, and E_{valid} edges also used by another recognized route. Write-back computes

$$E_{\mathrm{drop}} = E_{\mathrm{empty}} \setminus E_{\mathrm{valid}}, \qquad \widehat{E} = (E_0 \cup E_{\mathrm{ins}}) \setminus E_{\mathrm{drop}}. \tag{6}$$

Only circuit-exclusive edges are removed; shared routes remain. The final graph separately labels observed edges, inferred incidence, and cleanup decisions. Thus semantic acceptance permits write-back, but never selects edges to add or remove.

4 Experimental Setup

4.1 Data and Evaluation Protocol

Table 1 separates the prerequisite detector corpus from equipment evaluation. Thirty enterprise source drawings yield 360 base images with 39 frequent symbol classes and 1,140 instances; augmentation produces 1,080 images under a 70/15/15 split. None of these source drawings enters the 45-drawing enterprise equipment test. That held-out test contains 524 equipment instances, each assigned an engineer-confirmed normalized identity from the enterprise equipment list. The roster contents and normalization dictionary are contractually confidential and are not reproduced in the paper. Engineering review additionally provides 1,763 associations from distributed control system (DCS), safety instrumented system (SIS), or conventional instruments to connected equipment.

Table 1: Dataset roles and evaluation sizes.

Role	Drawings	Instances
Detector prerequisite[†]	30	1,140
Enterprise evaluation	45	524
OPEN100 transfer	12	62

[†]360 base images augmented to 1,080; 70/15/15 split.

Target equipment is absent from detector training, candidate-parameter selection, prompt demonstrations, and project-level development data. Thresholds and prompts are fixed before evaluation. For public transfer, we use PID2Graph OPEN100 (Stürmer et al., 2025), which contains 12 complex P&ID drawings with graph-aligned content. Task compatibility requires an equipment category mappable to a normalized reference identity and excludes background regions. The complete eligible subset comprises 62 instances. We evaluate it without OPEN100 adaptation and report public and enterprise results separately. Reference identities were established by OCR plus manual reading and confirmation of the drawings; instance boxes, instance counts, and edge annotations are not provided to the VLM.

4.2 METRICS

A candidate or final prediction p_i is geometrically eligible for target g_j only when $\mathrm{IoU}(p_i, g_j) > 0.50$. Candidate recall is the fraction of annotated devices with at least one geometrically eligible candidate before verification. For final equipment recognition, a true positive additionally requires the normalized predicted equipment name to equal the engineer-confirmed target identity. Maximum-cardinality one-to-one matching is applied to these joint geometry-and-identity pairs; duplicate, incorrectly named, and unmatched outputs are false positives, while unmatched targets are false negatives. Precision, recall, and F1 are therefore identity-aware equipment-recognition metrics. Candidate semantic-verification accuracy remains a secondary implementation diagnostic over all proposed regions: a negative must be rejected, whereas a positive must be accepted with the correct normalized identity. All percentages are micro-aggregated from the counts shown in the tables.

For graph utility, shortest-path search begins at each DCS, SIS, or conventional instrument and returns the first reachable equipment. Accuracy compares that result with the engineering-reviewed association, directly evaluating retrieval of an equipment relation used in HAZOP preparation.

4.3 FIXED-VERIFIER STRUCTURAL-CONDITIONING COMPARISON

All four methods use Qwen3.7-Plus with the same roster, instruction, decoding settings, and final matching rule: (i) whole-page inference; (ii) whole-page inference after a separate known-symbol mask; (iii) fixed crops around individual unconnected endpoints; and (iv) full TopoPID. Whole-page methods return equipment boxes and identities; crop-based methods verify proposed regions through the local hypothesis schema. The endpoint baseline omits density aggregation, bounded visual residuals, boundary completion, and candidate deduplication. The sequence compares page-scale search, exclusion by masking, point-localized crops, and equipment-scale residual support under a fixed semantic model.

5 RESULTS

5.1 CONTRACTING THE SPATIAL HYPOTHESIS SPACE IMPROVES COMPLETION

Table 2 contains the central comparison. The whole-page VLM baseline correctly recognizes 93/524 devices (19.44% F1). Masking known symbols raises this to 106/524 (22.75% F1), indicating that exclusion alone does not resolve page-scale search and density. Local endpoint crops recover 383/524 devices (80.38% F1), but individual endpoints can truncate or duplicate large equipment.

Table 2: Progressive structural conditioning under one fixed verifier. A true positive requires IoU > 0.50 and the correct normalized equipment identity; P/R/F1 are identity-aware recognition metrics.

Method	Correct/524	FP	FN	P/R/F1 (%)
Whole page	93/524	340	431	21.48/17.75/19.44
Whole page + mask	106/524	302	418	25.98/20.23/22.75
Endpoint crops	383/524	46	141	89.28/73.09/80.38
Full TopoPID	**489/524**	30	35	**94.22/93.32/93.77**

Full TopoPID correctly recognizes 489/524 devices and obtains 94.22% precision, 93.32% recall, and 93.77% F1. Relative to endpoint crops, it recovers 106 additional reference devices, reduces false negatives from 141 to 35, raises recall by 20.23 points, and raises F1 by 13.39 points. Under the fixed verifier, the full proposal pipeline outperforms the endpoint-crop baseline.

The gain combines higher completeness with fewer incorrect outputs. Relative to endpoint crops, false positives fall from 46 to 30 while precision and recall both increase. This matters for graph completion because accepted predictions supply equipment-node hypotheses for write-back. With the semantic model, roster, decoding, and matching rule held fixed, the comparison supports the effectiveness of equipment-scale residual context for local verification in this setting.

Figure 2 traces one evaluated OPEN100 case through proposal, verification, and graph update.

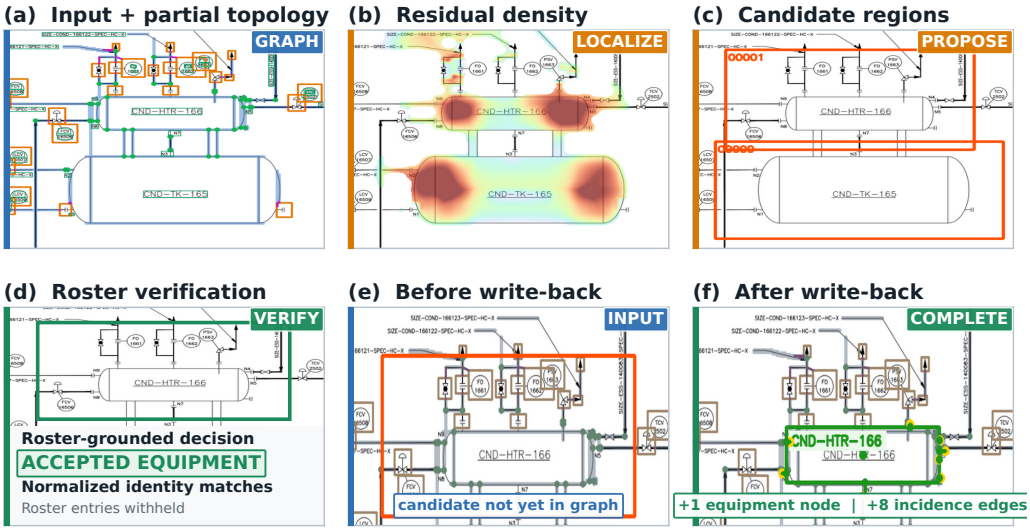

Figure 2: Representative OPEN100 case. (a–c) Residual density separates adjacent candidates. (d) Roster-grounded verification accepts the upper device with the correct normalized identity; roster entries are withheld. (e–f) Write-back adds one equipment node and eight evidence-linked incidence edges. The lower candidate is processed separately.

Table 3 separates candidate coverage from final correct recognition. On enterprise data, candidates geometrically cover 512/524 devices, and the final system correctly recognizes 489/524. The 35 misses comprise 12 devices without an IoU-qualified candidate and 23 localized devices that do not reach a correct final output.

Table 3: Stage-wise equipment recovery. Correct recognition requires IoU > 0.50 and the correct normalized identity.

Dataset	GT	Cand.-covered	Correct	FP/FN	P/R/F1 (%)
Enterprise	524	512/524	489/524	30/35	94.22/93.32/93.77
OPEN100	62	59/62	54/62	9/8	85.71/87.10/86.40

The "Correct" column counts real devices recovered with the right identity. As secondary diagnostics, 1,327/1,380 enterprise and 155/169 OPEN100 candidate decisions satisfy the joint acceptance-and-identity rule, including correct negative rejection.

The stage counts connect proposal availability to the final identity-aware outcome. Among reference devices covered by an IoU-qualified candidate, 489/512 enterprise and 54/59 OPEN100 devices reach a correct final output. These are counts of recovered reference devices: correctly rejecting a non-equipment candidate cannot increase them. Consequently, final recovery reflects both localization and identity agreement, while the auxiliary candidate-decision accuracy describes a separate verification task.

5.2 Transfer Across Drafting Standards Without Adaptation

With thresholds, budget, and prompt unchanged, TopoPID covers 59/62 compatible OPEN100 devices and correctly recognizes 54/62 (54/9/8 TP/FP/FN; 86.40% F1). The result measures transfer to the complete task-compatible subset of these 12 public drawings.

5.3 Engineering Associations in the Completed Graph

Completed enterprise graphs return 1,685/1,763 reviewed instrument–equipment associations (95.58%) in the specified shortest-path query.

Missing equipment nodes leave instrument paths without their intended equipment target, making device recovery relevant to association retrieval. Our focus is to recover unseen equipment with the correct normalized identity and a drawing region that satisfies the localization criterion, then insert an evidence-linked equipment node through the specified write-back rules. The association query evaluates whether the completed graphs return the engineering reference equipment from instrument nodes. Together, device recognition and retrieval assess localized node recovery and the relational utility of the completed graph.

6 Discussion

6.1 Interpretation and Scope

The fixed-verifier comparisons support the complete structural-conditioning pipeline: masking alone scarcely improves page-scale search, while endpoint crops localize but fragment large devices. Equipment-scale proposals recover more identity-correct devices. The comparison does not isolate density, bounded visual evidence, or individual proposal rules; the OPEN100 result covers one 12-drawing compatible subset. "Zero-shot" refers to the absence of target-equipment exemplars in the downstream protocol.

Association retrieval measures query utility rather than port- or edge-level graph fidelity. A misplaced candidate or misread circuit can still lead to an incorrect edit. We therefore claim an auditable update procedure and demonstrated association retrieval; the current evaluation does not measure individual graph edits or per-drawing uncertainty.

6.2 Failure Modes and Responsible Use

Missed or occupied endpoints weaken proposals; density basins may merge devices or grow along frames, while tight crops may omit distinguishing features. Similar roster entries or unfamiliar conventions can cause wrong acceptance. Provenance makes these decisions inspectable. Qualified engineers must inspect inferred nodes, attachments, and deletions before safety-related use.

7 Conclusion

TopoPID combines topology-residual proposals, roster-grounded verification, and deterministic graph updates to recover unseen, roster-enumerated equipment. It correctly recognizes 489/524 enterprise and 54/62 compatible OPEN100 devices, raising enterprise identity-aware F1 from 19.44% for whole-page inference to 93.77%. Completed graphs support 95.58% instrument–equipment association retrieval.

AI Use Statement

Beyond the VLM inference evaluated in this work, generative AI assisted literature discovery, manuscript organization and editing, translation, figure layout, code drafting, and interpretation of the reported results. The authors checked the resulting text, citations, figures, code, and claims against the experimental records. These tools were not used to fabricate or alter annotations, measurements, or reported counts. The authors are responsible for the final manuscript.

Ethics Statement

This work uses industrial engineering drawings. Incorrect inferred relations could affect downstream engineering analysis, so safety-related use requires source-drawing review by qualified engineers. Enterprise materials are reported only as disclosure-approved aggregates.

Reproducibility Statement

Sections 3–5 describe the method, evaluation protocol, and aggregate counts; public PID2Graph OPEN100 supports the stated transfer test. Enterprise drawings, rosters, identity mappings, prompts, implementation code, and per-drawing artifacts cannot be released under contract, so the enterprise results cannot be reproduced independently from this submission.

References

Fatih Cagatay Akyon, Sinan O. Altinuc, and Alptekin Temizel. Slicing aided hyper inference and fine-tuning for small object detection. In *Proceedings of the IEEE International Conference on Image Processing*, pp. 966–970, 2022. doi: 10.1109/ICIP46576.2022.9897990.

Junhyung Byun, Bonggu Kang, Duhwan Mun, Gwang Lee, and Hyungki Kim. Optimizing image format piping and instrumentation diagram recognition: Integrating symbol and text recognition with a single backbone architecture. *Journal of Computational Design and Engineering*, 12 (6):55–72, 2025. doi: 10.1093/jcde/qwaf053. URL https://doi.org/10.1093/jcde/qwaf053.

Tianheng Cheng, Lin Song, Yixiao Ge, Wenyu Liu, Xinggang Wang, and Ying Shan. YOLO-World: Real-time open-vocabulary object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16901–16911, 2024. doi: 10.1109/CVPR52733.2024.01599. URL https://doi.org/10.1109/CVPR52733.2024.01599.

Yongkun Du, Zhineng Chen, Caiyan Jia, Xiaoting Yin, Tianlun Zheng, Chenxia Li, Yuning Du, and Yu-Gang Jiang. SVTR: Scene text recognition with a single visual model. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pp. 884–890, 2022. doi: 10.24963/ijcai.2022/124. URL https://doi.org/10.24963/ijcai.2022/124.

Ehab Elhosary and Osama Moselhi. Knowledge-augmented large language models to support automated HAZOP report generation. *Process Safety and Environmental Protection*, 213:108997, 2026a. doi: 10.1016/j.psep.2026.108997. URL https://doi.org/10.1016/j.psep.2026.108997.

Ehab Elhosary and Osama Moselhi. AI-driven similarity assessment between P&ID nodes in support of HAZOP report generation. *Expert Systems with Applications*, 325:132744, 2026b. doi: 10.1016/j.eswa.2026.132744. URL https://doi.org/10.1016/j.eswa.2026.132744.

Rafael Grompone von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: A line segment detector. *Image Processing On Line*, 2:35–55, 2012. doi: 10.5201/ipol.2012.gjmr-lsd. URL https://doi.org/10.5201/ipol.2012.gjmr-lsd.

Seung-Tae Han, Yoochan Moon, Hyunsik Lee, and Duhwan Mun. Rule-based continuous line classification using shape and positional relationships between objects in piping and instrumentation diagram. *Expert Systems with Applications*, 248:123366, 2024. doi: 10.1016/j.eswa.2024.123366.

Seung-Tae Han, Yoochan Moon, Ji-Beob Kim, Hyunsik Lee, and Duhwan Mun. Graph neural network-based method for classifying continuous lines in piping and instrumentation diagram. *Advanced Engineering Informatics*, 66:103457, 2025. doi: 10.1016/j.aei.2025.103457.

Laura Jamieson, Eyad Elyan, and Carlos Francisco Moreno-García. Few-shot symbol detection in engineering drawings. *Applied Artificial Intelligence*, 38(1):2406712, 2024. doi: 10.1080/08839514.2024.2406712.

Sung-O Kang, Eul-Bum Lee, and Hum-Kyung Baek. A digitization and conversion tool for imaged drawings to intelligent piping and instrumentation diagrams (P&ID). *Energies*, 12(13):2593, 2019. doi: 10.3390/en12132593. URL https://doi.org/10.3390/en12132593.

Byung Chul Kim, Hyungki Kim, Yoochan Moon, Gwang Lee, and Duhwan Mun. End-to-end digitization of image format piping and instrumentation diagrams at an industrially applicable level. *Journal of Computational Design and Engineering*, 9(4):1298–1326, 2022. doi: 10.1093/jcde/qwac056.

Hyungki Kim, Wonyong Lee, Mijoo Kim, Yoochan Moon, Taekyong Lee, Mincheol Cho, and Duhwan Mun. Deep-learning-based recognition of symbols and texts at an industrially applicable level from images of high-density piping and instrumentation diagrams. *Expert Systems with Applications*, 183:115337, 2021. doi: 10.1016/j.eswa.2021.115337.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3992–4003, 2023. doi: 10.1109/ICCV51070.2023.00371. URL https://doi.org/10.1109/ICCV51070.2023.00371.

Junseo Lee, Sunhwa Park, Sehyeon Oh, and Byungchol Ma. Can large language models automate the HAZOP process without human intervention? *Safety Science*, 194:107039, 2026. doi: 10.1016/j.ssci.2025.107039. URL https://doi.org/10.1016/j.ssci.2025.107039.

Liunian Harold Li, Pengchuan Zhang, Haotian Zhang, Jianwei Yang, Chunyuan Li, Yiwu Zhong, Lijuan Wang, Lu Yuan, Lei Zhang, Jenq-Neng Hwang, Kai-Wei Chang, and Jianfeng Gao. Grounded language-image pre-training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10955–10965, 2022. doi: 10.1109/CVPR52688.2022.01069. URL https://doi.org/10.1109/CVPR52688.2022.01069.

Minghui Liao, Zhaoyi Wan, Cong Yao, Kai Chen, and Xiang Bai. Real-time scene text detection with differentiable binarization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 11474–11481, 2020. doi: 10.1609/aaai.v34i07.6812. URL https://doi.org/10.1609/aaai.v34i07.6812.

Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Qing Jiang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, and Lei Zhang. Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. In *European Conference on Computer Vision*, pp. 38–55, 2024. doi: 10.1007/978-3-031-72970-6_3. URL https://doi.org/10.1007/978-3-031-72970-6_3.

Tao Liu, Rongjie Li, Chongyu Wang, and Xuming He. Relation-aware hierarchical prompt for open-vocabulary scene graph generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 5576–5584, 2025. doi: 10.1609/aaai.v39i5.32594.

Ruotian Luo, Ning Zhang, Bohyung Han, and Linjie Yang. Context-aware zero-shot recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 11709–11716, 2020. doi: 10.1609/aaai.v34i07.6841.

Shouvik Mani, Michael A. Haddad, Dan Constantini, Willy Douhard, Qiwei Li, and Louis Poirier. Automatic digitization of engineering diagrams using deep learning and graph search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 673–679, 2020. doi: 10.1109/CVPRW50498.2020.00096.

Matthias Minderer, Alexey Gritsenko, Austin Stone, Maxim Neumann, Dirk Weissenborn, Alexey Dosovitskiy, Aravindh Mahendran, Anurag Arnab, Mostafa Dehghani, Zhuoran Shen, Xiao Wang, Xiaohua Zhai, Thomas Kipf, and Neil Houlsby. Simple open-vocabulary object detection. In *European Conference on Computer Vision*, pp. 728–755, 2022. doi: 10.1007/978-3-031-20080-9_42. URL https://doi.org/10.1007/978-3-031-20080-9_42.

Matthias Minderer, Alexey Gritsenko, and Neil Houlsby. Scaling open-vocabulary object detection. In *Advances in Neural Information Processing Systems*, volume 36, pp. 72983–73007, 2023. doi: 10.52202/075280-3191. URL https://doi.org/10.52202/075280-3191.

Yoochan Moon, Jinwon Lee, Duhwan Mun, and Seungeun Lim. Deep learning-based method to recognize line objects and flow arrows from image-format piping and instrumentation diagrams for digitization. *Applied Sciences*, 11(21):10054, 2021. doi: 10.3390/app112110054.

Shubham Paliwal, Arushi Jain, Monika Sharma, and Lovekesh Vig. Digitize-pid: Automatic digitization of piping and instrumentation diagrams. In *Trends and Applications in Knowledge Discovery and Data Mining*, pp. 168–180. Springer International Publishing, 2021a. doi: 10.1007/978-3-030-75015-2_17.

Shubham Paliwal, Monika Sharma, and Lovekesh Vig. OSSR-PID: One-shot symbol recognition in P&ID sheets using path sampling and GCN. In *2021 International Joint Conference on Neural Networks*, pp. 1–8, 2021b. doi: 10.1109/IJCNN52387.2021.9534122. URL https://doi.org/10.1109/IJCNN52387.2021.9534122.

Suraj Prasad and Pinak Mahapatra. SynthPID: P&ID digitization from topology-preserving synthetic data, 2026. URL https://arxiv.org/abs/2604.16513.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Rohit Rahul, Shubham Paliwal, Monika Sharma, and Lovekesh Vig. Automatic information extraction from piping and instrumentation diagrams. In *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, pp. 163–172, 2019. doi: 10.5220/0007376401630172. URL https://doi.org/10.5220/0007376401630172.

Suprosanna Shit, Rajat Koner, Bastian Wittmann, Johannes Paetzold, Ivan Ezhov, Hongwei Li, Jiazhen Pan, Sahand Sharifzadeh, Georgios Kaissis, Volker Tresp, and Bjoern Menze. Relationformer: A unified framework for image-to-graph generation. In *European Conference on Computer Vision*, pp. 422–439, 2022. doi: 10.1007/978-3-031-19836-6_24. URL https://doi.org/10.1007/978-3-031-19836-6_24.

Vasil Shteriyanov, Rimma Dzhusupova, Jan Bosch, and Helena Holmström Olsson. BlueprintSymVL: A discriminative benchmark for VLM symbol recognition in engineering blueprints. *Results in Engineering*, 28:108171, 2025. doi: 10.1016/j.rineng.2025.108171.

Jan Marius Stürmer, Marius Graumann, and Tobias Koch. From engineering diagrams to graphs: Digitizing p&ids with transformers. In *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics*, pp. 1–11, 2025. doi: 10.1109/DSAA65442.2025.11248012.

Maximilian F. Theisen, Kenji Nishizaki Flores, Lukas Schulze Balhorn, and Artur M. Schweidtmann. Digitization of chemical process flow diagrams using deep convolutional neural networks. *Digital Chemical Engineering*, 6:100072, 2023. doi: 10.1016/j.dche.2022.100072.

Yifan Xu, Weijian Xu, David Cheung, and Zhuowen Tu. Line segment detection using transformers without edges. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4255–4264, 2021. doi: 10.1109/CVPR46437.2021.00424. URL https://doi.org/10.1109/CVPR46437.2021.00424.

Nan Xue, Tianfu Wu, Song Bai, Fu-Dong Wang, Gui-Song Xia, Liangpei Zhang, and Philip H. S. Torr. Holistically-attracted wireframe parsing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2785–2794, 2020. doi: 10.1109/CVPR42600.2020.00286. URL https://doi.org/10.1109/CVPR42600.2020.00286.

Eun-Seop Yu, Jae-Min Cha, Taekyong Lee, Jinil Kim, and Duhwan Mun. Features recognition from piping and instrumentation diagrams in image format using a deep learning network. *Energies*, 12(23):4425, 2019. doi: 10.3390/en12234425. URL https://doi.org/10.3390/en12234425.

Dong-Yeol Yun, Seung-Kwon Seo, Umer Zahid, and Chul-Jin Lee. Deep neural network for automatic image recognition of engineering diagrams. *Applied Sciences*, 10(11):4005, 2020. doi: 10.3390/app10114005. URL https://doi.org/10.3390/app10114005.

Xingyi Zhou, Rohit Girdhar, Armand Joulin, Philipp Krähenbühl, and Ishan Misra. Detecting twenty-thousand classes using image-level supervision. In *European Conference on Computer Vision*, pp. 350–368, 2022. doi: 10.1007/978-3-031-20077-9_21. URL https://doi.org/10.1007/978-3-031-20077-9_21.

Yichao Zhou, Haozhi Qi, and Yi Ma. End-to-end wireframe parsing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 962–971, 2019. doi: 10.1109/ICCV.2019.00105. URL https://doi.org/10.1109/ICCV.2019.00105.

Baikai Zhu, Samuel Duong, Javal Vyas, and Mehmet Mercangöz. From P&ID drawings to process graphs: A multimodal language model approach. *Systems and Control Transactions*, 5: 1737–1745, 2026. doi: 10.69997/sct.198584. URL https://doi.org/10.69997/sct.198584.