

Appendix

A. Demo of P&ID digitization pipeline

In this section, we demonstrate the steps in our P&ID digitization pipeline: starting with an input diagram, detecting symbols and text, detecting connections between symbols, and producing an asset hierarchy table.

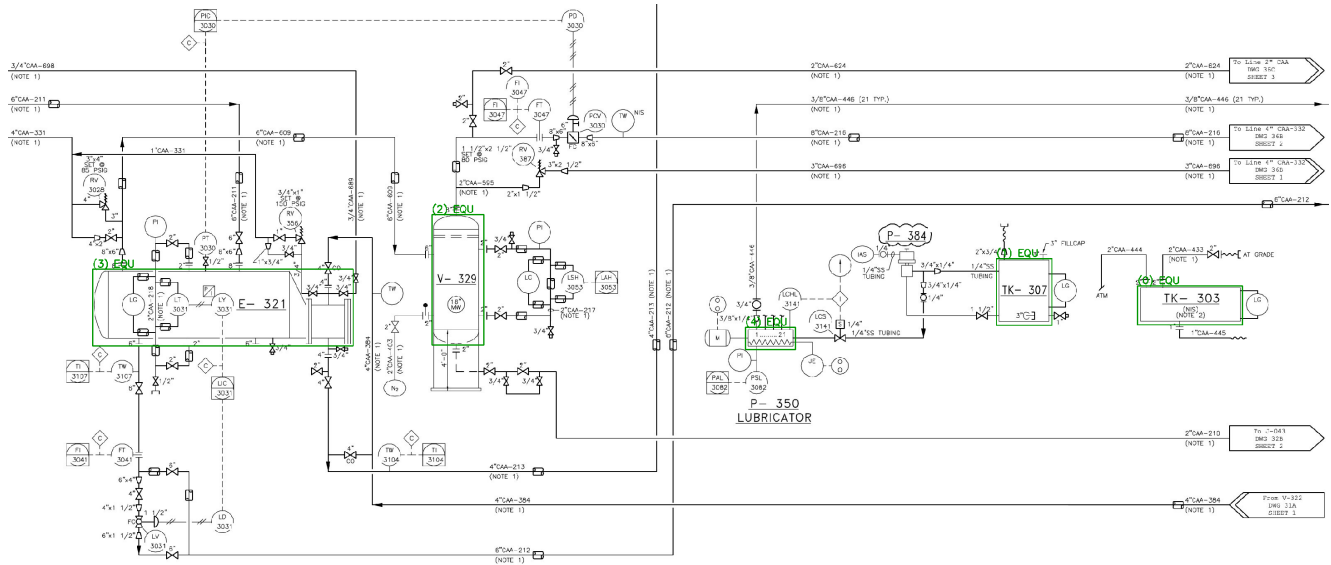


Figure 1. An input diagram to the pipeline with some manually-labeled symbols (green).

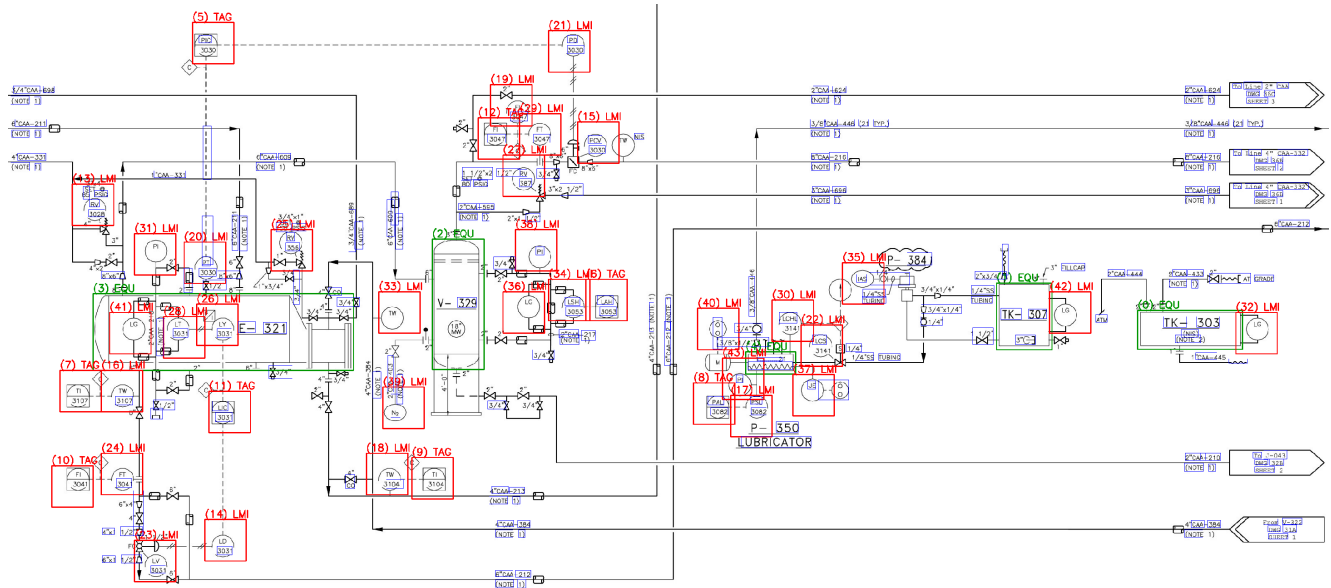


Figure 2. Detected symbols (red), detected text (blue), and manually-labeled symbols (green) in the input diagram. This is the intermediate result of the symbol detection and text detection steps.

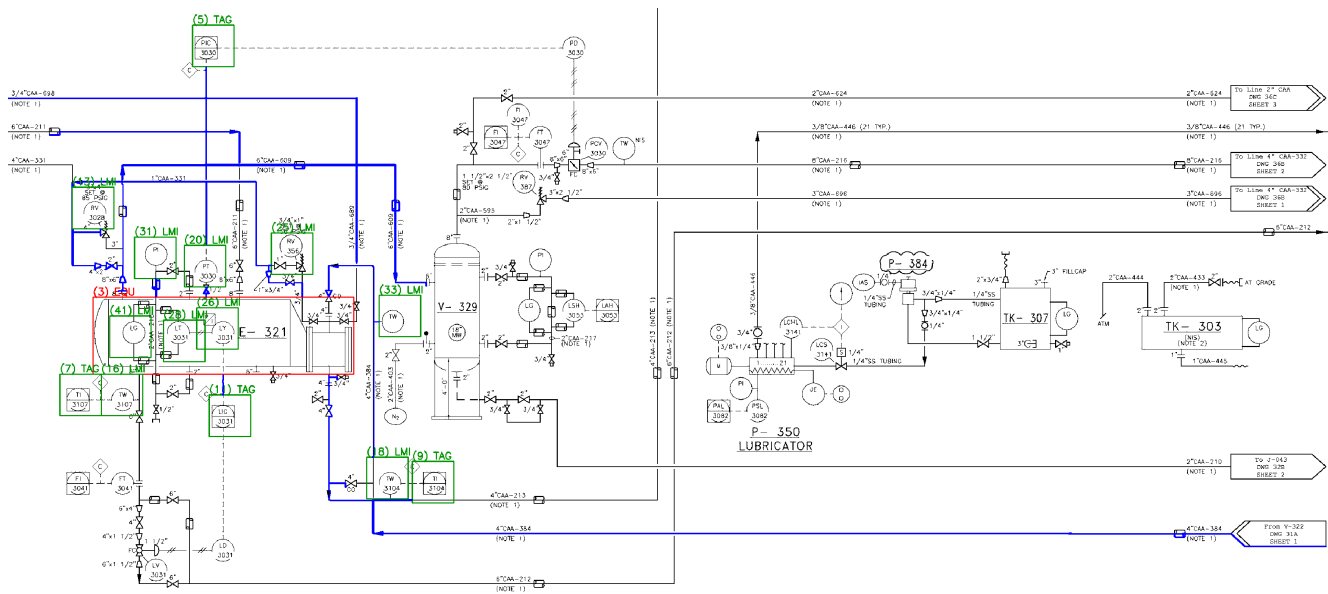


Figure 3. Connections detected between a source symbol (red) and connected symbols (green) in the diagram through lines traversed by depth-first search (blue).

Symbol id	Method	Symbol type	Associated text	Connected symbols
0	Manually-labeled	EQU	MEYE, NIS, 303, CAA—A, TK—,	32
1	Manually-labeled	EQU	TUBING, FILLCAF, 307, MA, 2X34E, 14SS, E, TK-,...	42
2	Manually-labeled	EQU	NOYE, 329, 342, 1	34, 36, 38, 15, 19, 21, 27
3	Manually-labeled	EQU	3030, 321, CO, E“X6F, XVZ, E., XJA, BX6F, 303,...	33, 5, 7, 9, 41, 11, 13, 16, 18, 20, 25, 26, 2...
4	Manually-labeled	EQU	KVVVVVVX, TM AT] T, PL, {8, 34?, JE, 7	37, 8, 40, 43, 17, 22, 30
5	Detected	TAG	PLC	3, 13, 20, 21, 25
6	Detected	TAG	UFIÉ, 73, 3053, -2171, LSU, 3053, 2CAA	34
7	Detected	TAG	3I07	16, 3
8	Detected	TAG	JJBRICATOR, P _i I, 3082, 3082, KVVVVVVX, ; PAL, PL	17, 43, 4
9	Detected	TAG	NO	18, 3
10	Detected	TAG	4“	18, 3
11	Detected	TAG	7L?, XJA, 303J	3, 14
12	Detected	TAG	PÉLC, RV, W, 3047, 387., 3047, 3047, ARM *2, 34X	27, 19, 29
13	Detected	LMI	3023, RV, 1251C, 3%	25, 3, 20, 5
14	Detected	LMI	3031	24, 11, 23
15	Detected	LMI	CH], 3030, NIS, 34X, 212	19, 2, 27, 21
16	Detected	LMI	3I07, XVI, MOE.9, 1	3, 7, 41, 24, 26, 28
17	Detected	LMI	JJBRICATOR, P _i I, 3082, 3082, 350, ; PAL, PL	8, 43, 4
18	Detected	LMI	¿iCO, NO	9, 3, 33
19	Detected	LMI	3047, 3047, 3047	2, 12, 15, 21, 27, 29
20	Detected	LMI	3030, XVZ, BX6F, UFOZ, E, J91, E, RIÉ.00	3, 5, 13, 25, 31
21	Detected	LMI	PD, 3039	2, 5, 15, 19, 27
22	Detected	LMI	4SS, KVVVVVVX, FL, 14, {8, ¿0, JE, 7	4, 37, 30
23	Detected	LMI	3031, 6“, 4“, WZV VW, 1ZB, 3%	24, 14
24	Detected	LMI	3%, 30	16, 14, 23

Table 1. The asset hierarchy table (showing the first 25 rows) populated with information extracted from the diagram by the pipeline.

B. Symbol detection: detected symbols vs. ground truth

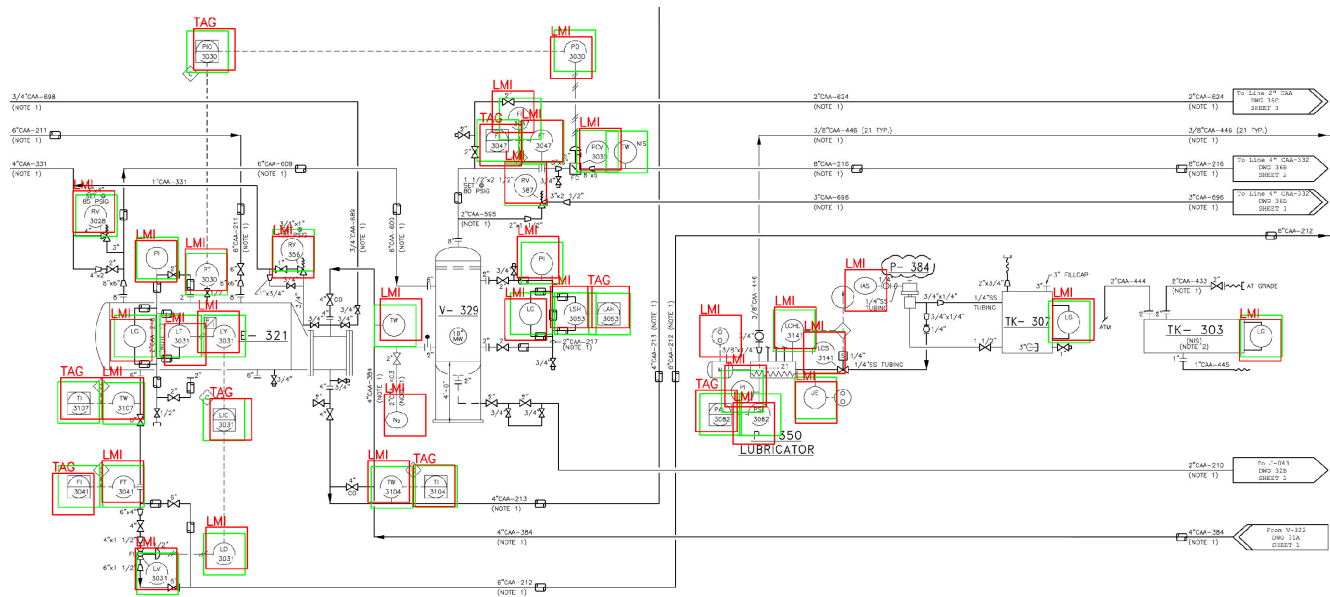


Figure 4. A comparison between symbols detected by the pipeline (red) and ground truth symbols (green) in the diagram.