

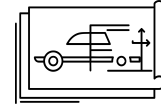
DATA SEER

Digitization of Industrial Diagrams using
Machine Learning & Computer Vision

March 28, 2022

Why is data extraction from industrial diagrams so hard?

Diagrams for brownfield projects are typically only available in scanned PDFs or paper form, not CAD formats



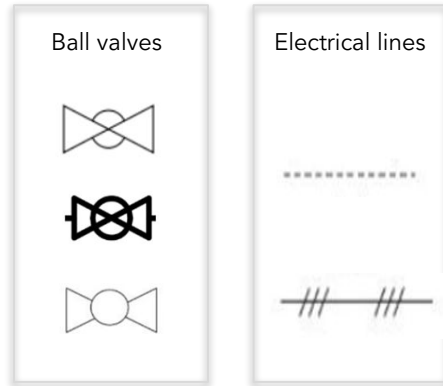
Engineering standards (e.g., ANSI, ISO, ISA, PIP, DIN, BS) are not hard rules and evolve over time



Extraction & digitization of legacy diagrams is highly manual, time-consuming and vulnerable to human error



Mistakes in industrial systems can be fatal when safety is at stake

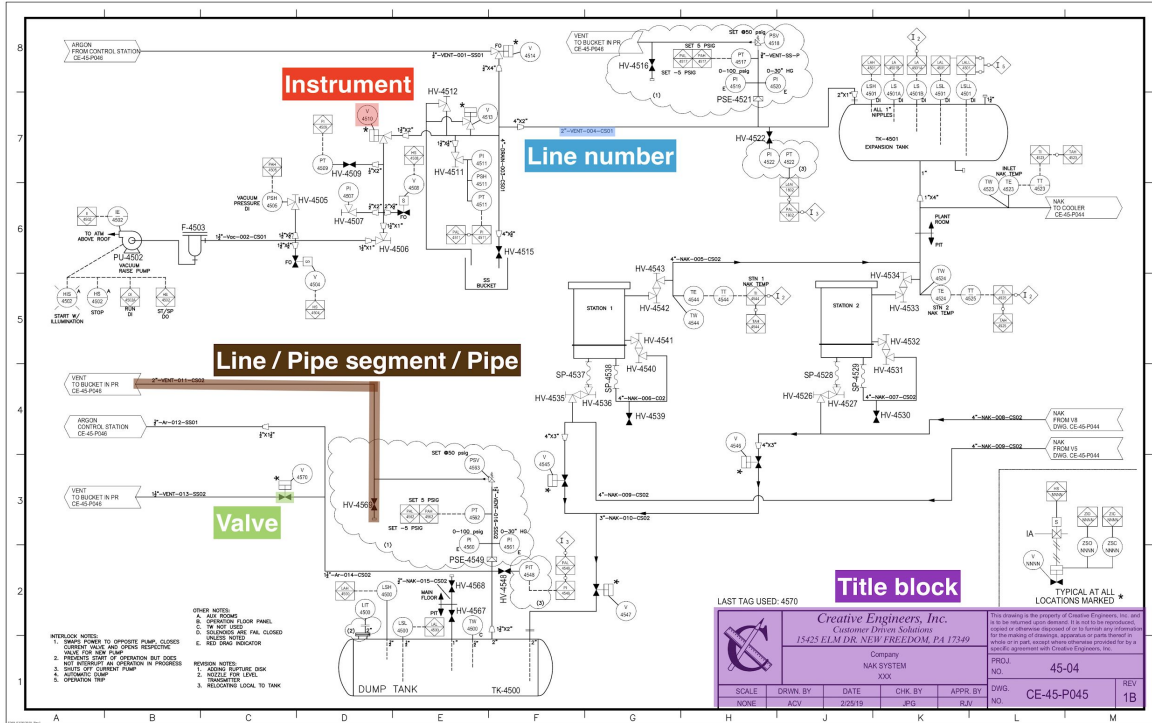


Representations can be ambiguous so we need to account for context and customization when extracting & interpreting information

Extracted data needs to be verified & reviewed

A human-in-the-loop approach is essential to allow engineers to override & guide ML systems

Example: Detection, recognition & segmentation in a P&ID



Items of interest:

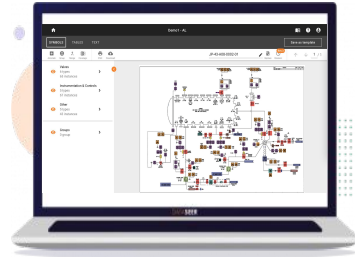
- Symbols
- Text attributes
- Piping
- Tables
- Notes
- Markups
- Continuation flags
- ...

We can use both automatic & semi-automated ML approaches

Fully Automatic



- No human intervention
- Data distribution remains same
- Suitable if rough estimate is sufficient

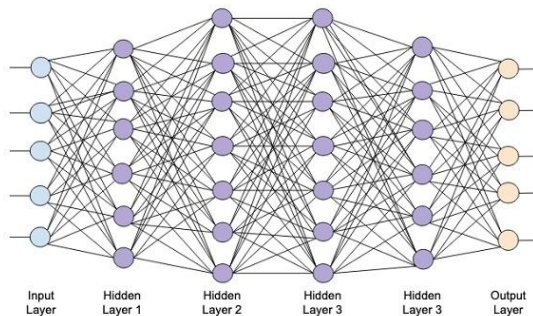


Human-in-the-loop



- Human expert input solicited
- Data distribution is changing over time
- Combines feedback & verification process

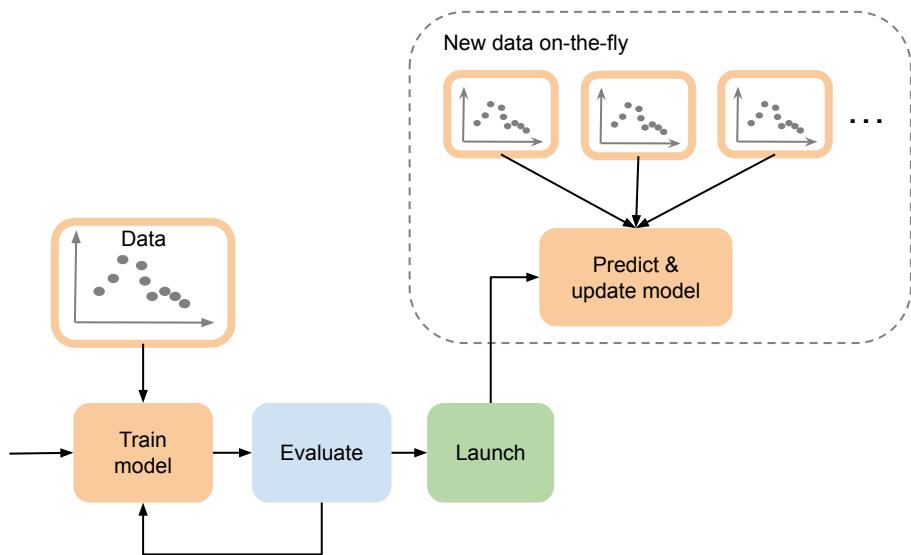
Mode 1: Deep net model to recognize a default symbol library



- ✓ Far superior detection performance
- ✓ Can also recognize (classify)

- ✗ Typically require lots of labeled training data
- ✗ Lengthy training times
- ✗ Learning parameters are tricky to tune

Mode 2: Online learning of user models for custom symbols



Local models are updated in real-time based on user feedback on right / wrong predictions

Some memory of recent history is maintained in local buffers

Users can save these models to templates in order to re-use / continue training them



DATASEER

info@dataseer.digital