
CASE STUDY: TRANSFORMER-BASED SOLUTION FOR THE AUTOMATIC DIGITIZATION OF GAS PLANTS

**Bailo I. Buonora F. Ciarfaglia G. Consoli L.T. Evangelista A. Gabusi M.
Ghiani M. Petracca Ciavarella C. Picariello F. Sarcina F. Tuosto F. Zullo V.**
Eng AI&Data @Engineering Group

Airoldi L. Bruno G. Gobbo D.D. Pezzenati S. Tona G.A.
Snam S.p.A.

ABSTRACT

The energy transition is a key theme of the last decades to determine a future of eco-sustainability, and an area of such importance cannot disregard digitization, innovation and the new technological tools available. This is the context in which the Generative Artificial Intelligence models described in this paper are positioned, developed by Engineering Ingegneria Informatica SpA in order to automate the plant structures acquisition of SNAM energy infrastructure, a leading gas transportation company in Italy and Europe. The digitization of a gas plant consists in registering all its relevant information through the interpretation of the related documentation. Currently, this process is carried out manually using the Modulo Grafico Misura (MGM) application and is a laborious task that requires a lot of resources. The aim of this work is therefore to design an effective solution based on Artificial Intelligence techniques to automate the extraction of the information necessary for the digitization of a plant, in order to streamline the daily work of MGM users. The developed solution receives design data, the equipment list and the Piping & Instrumentation Diagram (P&ID) of the plant as input, each one in pdf format, and uses OCR, Vision LLM, Object Detection, Relational Reasoning and optimization algorithms to return an output consisting of 2 set of information: a structured overview of the relevant design data and the hierarchical framework of the plant. To achieve convincing results, we extend a state-of-the-art model for Scene Graph Generation introducing a brand new Transformer architecture with the aim of deepening the analysis of the complex relations between the plant's components. The synergistic use of the listed AI-based technologies allowed to overcome many obstacles arising from the high variety of data, due to the lack of standardization. An accuracy of 91% has been achieved in the extraction of textual information relating to design data. Regarding the plants topology, 93% of components are correctly identified and the hierarchical structure is extracted with an accuracy around 80%.

Keywords P&ID · Scene Graph Generation · Deep Learning · GenAI · OCR

1 Introduction

In recent decades, the energy transition has become a central topic in shaping an environmentally sustainable future. Such a critical field cannot overlook the role of digitization, innovation, and emerging technological tools. Within this context, the Generative Artificial Intelligence models discussed in this paper, developed by Eng AI&Data, have the key role of automating the acquisition of plant structures belonging to the energy infrastructure of SNAM, a leading gas transportation company in Italy and across Europe.

With a fleet of about 11,000 Regulating and Metering plants (ReMi) distributed on the Italian gas transportation network, efficient control and management is essential to manage the paperwork related to the creation of new plants or the changes made on existing plants. This activity is focused on extracting relevant information from plants documents and schematics, such as P&IDs (Piping & Instrumentation Diagrams) and components lists.

The high heterogeneity and complexity of plant types and related documentation have traditionally imposed highly

manual and specialized activities by SNAM experts, who receive the documentation from plant owners, analyze it, extract the necessary information (plant structures, lines, sections, components and related characteristics), and enter the extracted details into SNAM internal IT systems. However, this manual registration is a laborious task and requires a lot of resources. Moreover, it could lead to typos or general inaccuracies, not ensuring consistency.

The goal of the solution described in this paper is therefore to make such a data extraction process automatic through Generative Artificial Intelligence tools, specifically by processing plant schematics of varying complexity and type and their accompanying documentation. Time and resources to be spent on this task would decrease drastically and automatic systems are generally accurate in repetitive tasks, allowing to reduce the error rate in the registration.

An automatic digitization system should therefore receive the plant documents and analyse them to extract the relevant information. The main obstacle stems from the lack of rigorous standardisation in the drafting of key elements of the documentation, which are therefore characterised by a high degree of variability. As a result, the information to be extracted appears within the documents in different orders, formats and styles, making it impossible for simple systems to read. Solutions must therefore be able to analyse the documentation in a flexible manner, simulating the type of reasoning that a human operator would apply in interpreting the documents.

This study proposes an end-to-end system for the digitization of gas plants based on the relevant documentation. This system consists of several progressive steps that analyse the main elements and combine the extracted information, each of which uses advanced computer vision and optimization techniques. Compared to existing solutions, which are based on the analysis of a few key elements compared to those available, this study aims to maximise the accuracy of the system by leveraging the entire documentation.

2 Background

Digitization of plants documentation could involve several analysis tasks, such as text, tables and diagram recognition and interpretation. Text and table reading is possible through consolidated and evolving techniques like OCR or Vision LLMs. [1] [2].

An important reference element of plant documentation is the P&ID, a detailed diagram showing installed equipment and their connections, used in all the processes on the plant, from installation to maintenance. The transformation of the P&ID in a structured representation brings many advantages for the whole plant analysis. A widespread and reasonable method is the encoding of the diagram in a Scene Graph containing information on present components and their relations.

Existing methods to perform Scene Graph Generation (SGG) from an image are divided into two-stage and one-stage methods: two-stage methods [3] first identify entities in the scene to then predict relations between them, while one-stage methods [4] [5] [6] directly generate subject-predicate-object candidate triplets. The former are less efficient, as they are more onerous, slower and more prone to error propagation, thus the latter represent the state of the art in this field.

Within the plant digitization field, several works used combinations of the previously described techniques to develop smart solutions.

S.O. Kang et.al. [7] proposed a tool for the conversion of paper P&ID into a digital format. The system is based on a pipeline that includes OCR algorithms for text recognition and CNN for symbols detection. A distinctive aspect of this work is its integration with industrial components databases, allowing the mapping of the detected elements with the associated technical specifications.

S.Mani et.al. [8] combined computer vision and graph research techniques. They use a CNN for main symbols detection, merge text and symbols information and later leverage a graph research algorithm to identify symbols connections through diagram's lines.

B.C. Kim et.al. [9] proposed an end-to-end framework for P&ID digitization combining deep learning techniques and graphic analysis to obtain structured representations of the diagrams. The system has been tested on industrial datasets and proved robustness in components detections and it has the capability of building digital formats compatible with plants-management softwares.

S. Paliwal et.al. [10] proposed an automatic digitization system based on Computer Vision and Machine Learning techniques to extract structured information from P&ID. It uses a combination of CNN for symbols detections and rule-based models to identify connections, allowing diagrams conversion in a standard digital format.

S.M. Gajbhiye et.al. [11] set YOLOv5, the fifth edition of one of the most advanced architectures for object detection, at the center of their work. The system has been trained on annotated P&ID datasets and optimized to reach high-precision symbols detection. Compared to traditional methods, YOLOv5 stands out for its capability of quickly process images and identify precisely even small symbols, making it very suitable for real-time industrial applications.

J.M. Stürmer et.al. [12] introduced the use of Transformers, a class of deep learning models typically employed in Natural Language Processing, for P&ID digitization. Specifically, the system leverages the Vision Transformer architecture to analyse diagrams structure and identify symbols and connections. Thanks to the Transformers capability

of catching spatial relations between elements, the model extracts information with high accuracy, improving performance compared to traditional CNN-based models.

Despite significant progress in this field, current technologies still have some limitations. Although advanced models such as YOLOv5 and Transformers have greatly improved symbol recognition, the accuracy of identification can be compromised by variations in design, such as the presence of manual annotations or differences in style. Challenges also remain regarding the management of connections between components, the ability to generalise to different industry standards, and the integration of the solution with existing software. Finally, most solutions only consider the P&ID, which, although it contains a lot of condensed information, may not be sufficiently comprehensive.

3 Methods

An overview of the designed pipeline of our solution is provided in Figure 1. The solution aims to achieve 2 main goals. The first is to extract and organize the relevant design data of the ReMi plant, which is performed through OCR methods. The second and most challenging is to build a hierarchical representation of the plant, which is done using a combination of OCR and LLM methods, SGG models and optimization algorithms in order to maximize the generalization capabilities of the system given the great variability of data. Specifically, the hierarchy typically used to represent a plant is characterized by the following levels, from the specific to the general:

- Level I: characteristic or **type**, which serves to define a quantitative or qualitative aspect of a device;
- Level II: component or **sub-type**, which is defined on the basis of a set of mandatory and optional characteristics;
- Level III: **section**, i.e. a coherent grouping of several components that perform a specific function;
- Level IV: **line**, i.e. a set of sections, which may be for regulation or measurement.

3.1 Input Data

The main input data, examples of which are shown in Figure 2, consist mainly of PDF documents, each corresponding to an aspect of the plant description and including:

1. **Design data** (Figure 2a): a page, usually in tabular or similar format, containing general information about the plant, such as its main flow rates and pressures and its height above sea level. Its content must be reported in full in the output;
2. **Equipment list** (Figure 2b): one or more pages in tabular format which, with reference to the P&ID, list the equipment and components present, specifying their identification in the diagram, category, manufacturer, description and main technical specifications;
3. **Piping & Instrumentation Diagram** (Figure 2c): a page containing a schematic drawing of the plant, showing the components and their connections. Some of the information that can be extracted from the P&ID and specified in the equipment list constitutes the second part of the output

The plant documentation dataset, extracted and provided by SNAM, consists of approximately 1500 ReMi.

Other relevant data that is not provided as input to the pipeline but that has been used for its design and for training some of its models is that concerning the Regulations and the Plant Registry.

The Regulations consist of a set of Excel files that translate the information in the UNI 9167 standard, the main regulatory reference for the design and management of a plant. These contain, for example, the catalogue of possible equipment, which contains information on the brand, model and characteristics of each possible component that can be used, and the catalogue with the rules for hierarchically dividing the plant into information units. The information contained in the Regulations was used to define sets of rules and constraints that the solution's output must comply with. The Plant Registry is a set of documents containing manually recorded information on all plants installed up to the present. It specifies the characteristics of each plant's components and the hierarchical structure of the plant itself, data that has been used to extract a statistical representation of the composition of the plants.

3.1.1 Synthetic Data Generation Using Diffusion Models and Algorithmic Techniques

The availability of P&ID diagrams represents a significant obstacle to the optimization of the system capabilities, especially that of the SGG model. The amount of original data is insufficient to ensure adequate generalisation capacity, given the high level of diversification of the diagrams. To achieve satisfactory performance, it was therefore necessary

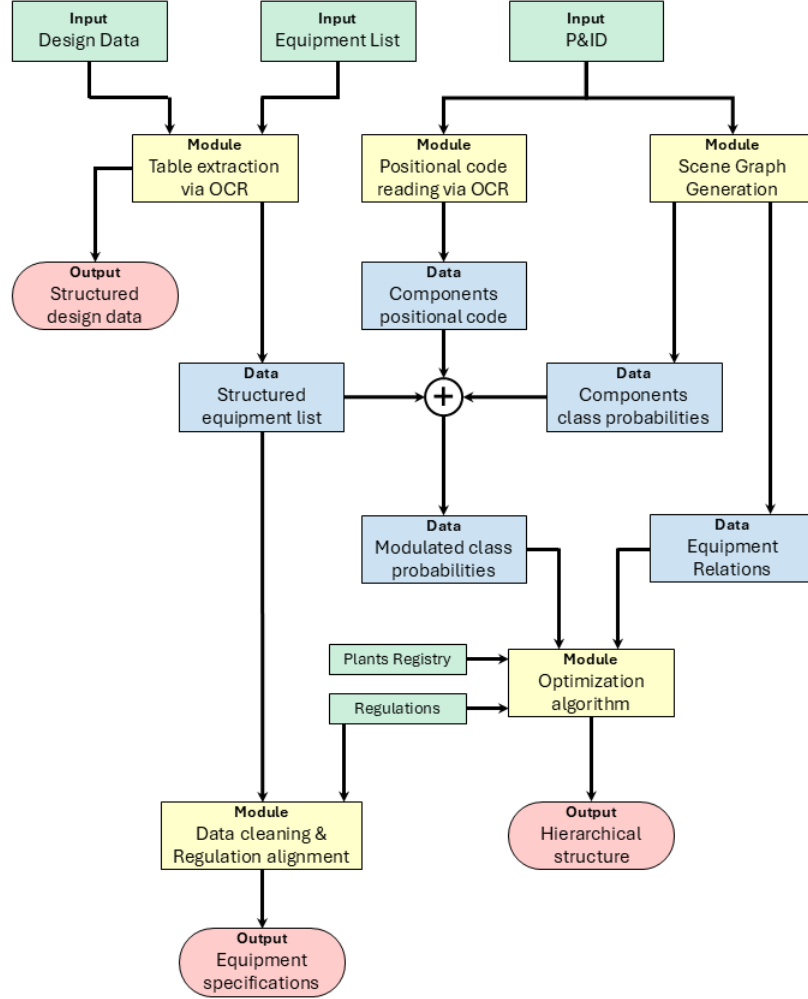


Figure 1: Diagram of the developed solution. All the input data are in pdf format

to use data augmentation techniques to increase the cardinality of the available data. Synthetic data was added using both algorithmic approaches and based on Generative AI.

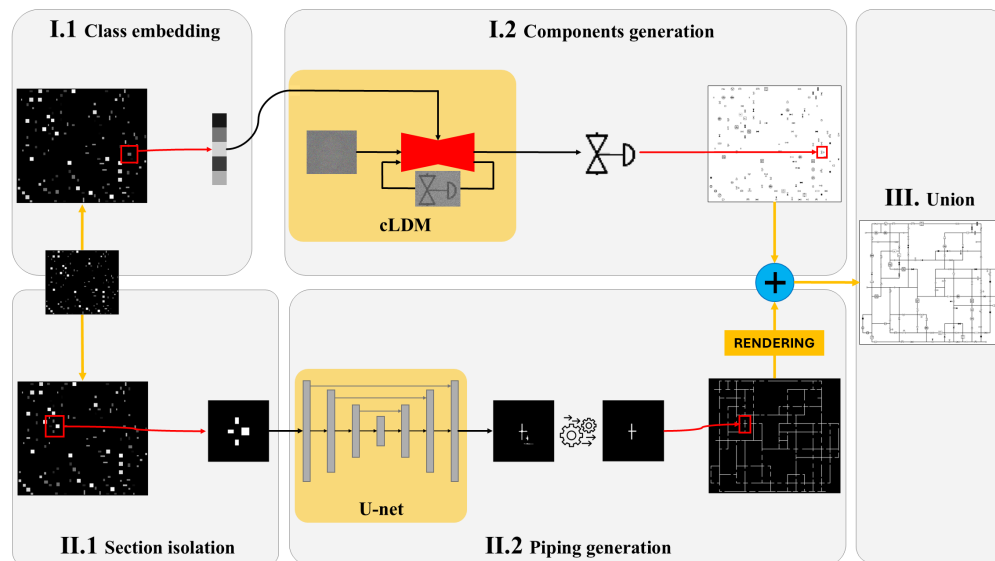
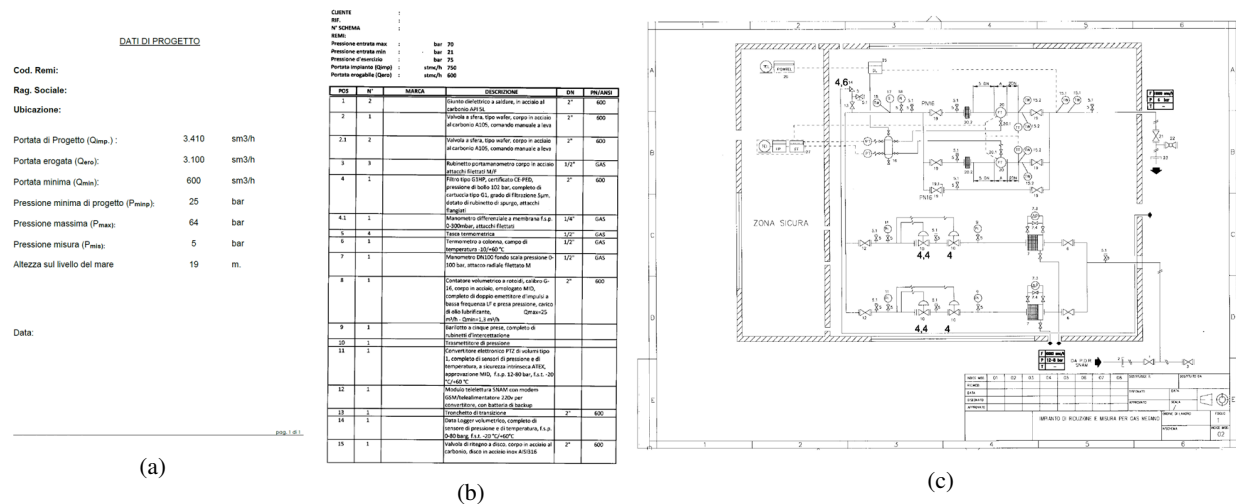
The first approach consists of the "manual" production of images containing symbols of components present in the original diagrams, positioned within the image after applying random transformations such as distortions, rotations and the addition of noise. The image is then modified to increase its noise, for example by adding text, tables and figures or setting random backgrounds. The images produced using this method were mainly used in the pre-training phase of the SGG model, focused on the object detection task.

The second approach consists of using a pipeline that produces a fictitious diagram based on its layout and information regarding the division of symbols into sections. The generative process, schematized in Figure 3, is divided into two steps. In the first step, using the classes and positions of each component identifiable from the layout, a conditional Latent Diffusion Model is used to generate plausible versions of the symbols. In the second step, a U-net receives the image portions with the components divided into sections and generates the connection map for each group, which is then rendered.

3.2 Plant's Design Data and Equipment List Extraction

The first functional block of the entire pipeline consists of extracting design data from the relative documentation page and the specifications of the components present in the plant as listed in the equipment list.

The first step in the process is to use an OCR model to extract all tables from the documentation in text format. The obtained output is then processed using an LLM to get ordered and structured tables. The operations performed by the LLM include:



- The procedure described allows the automatic generation of the plant design data table, which also includes all the components and their respective characteristics required by the Regulations, specifically regarding levels I e II of the hierarchy. This approach overcomes the difficulties associated with the heterogeneity of formats and table structures in technical plant documents.

3.3 Plant's Hierarchical Structure Reconstruction

In this phase, the analysis of the plant's P&ID plays a central role, and it is divided into three phases. In the first phase, an SGG model used to extract an indirect graph from the diagram. This graph identifies the symbols corresponding to the equipment of interest, including their locations and types, and traces the connections between them, allowing for an initial grouping into sections. In the second phase, OCR techniques are used to find the numerical codes associated with each symbol within the diagram. By subsequently identifying the correspondence in the equipment list table, extracted in the task described in section 3.2, it is possible to correct any inaccuracies in the SGG model and finally merge the results. In the last phase, an optimization algorithm uses the merged output to reconstruct the whole hierarchical structure of the plant.

Given a set of components $\mathcal{C} = \{c_1, c_2, \dots, c_N\}$, it's possible to define sections $\mathcal{T} = \{T_1, \dots, T_K\}$ as a partition of $\mathcal{C}$, i.e. as subsets of $\mathcal{C}$ so that:

$$T_k \subseteq \mathcal{C}, \quad T_i \cap T_j = \emptyset \quad \forall i \neq j, \quad \bigcup_{k=1}^K T_k = \mathcal{C}.$$

Similarly, it's possible to define lines $\mathcal{L} = \{L_1, \dots, L_M\}$ as a partition of sections $\mathcal{T}$:

$$L_j \subseteq \mathcal{T}, \quad L_j \cap L_{j'} = \emptyset \quad \forall j \neq j', \quad \bigcup_{j=1}^M L_j = \mathcal{T}.$$

Each component $c_i \in \mathcal{C}$, section $T_k \in \mathcal{T}$ and line $L_j \in \mathcal{L}$ is then associated a class that defines its type, $y_i \in \mathcal{Y}_C$, $y_{T_k} \in \mathcal{Y}_T$, $y_{L_j} \in \mathcal{Y}_L$ respectively. The hierarchical structure of the plant is then defined by the triplet $(\mathcal{T}, \mathcal{L}, \mathbf{y})$, where $\mathbf{y} = \{y_i\}_{i=1}^N \cup \{y_{T_k}\}_{k=1}^K \cup \{y_{L_j}\}_{j=1}^M$ is the complete assignment of classes.

3.3.1 Scene Graph Generation

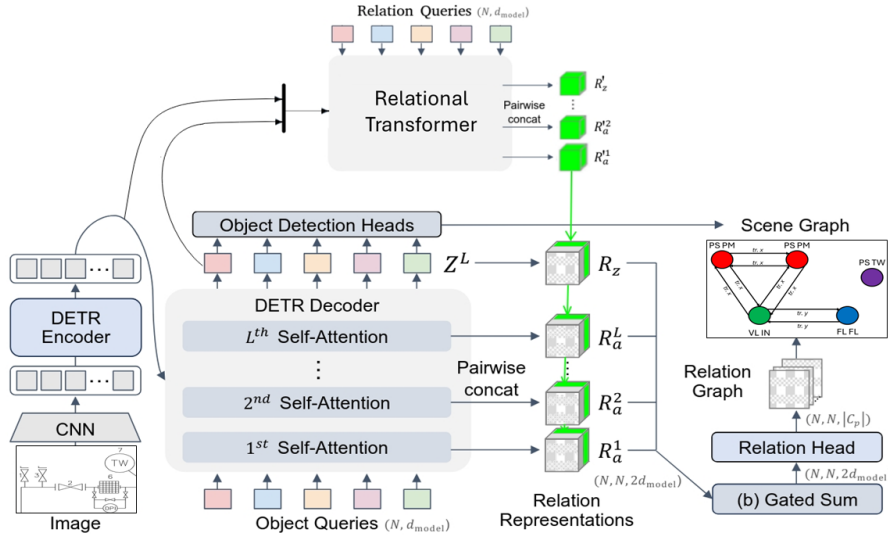


Figure 4: EGRTR architecture

The extraction of a Scene Graph from the plant's P&ID provides a first representation of the plant based on the diagram topology. To address this challenge, the EGRTR (Extracting Graph from Relational Transformer) model was developed, an innovative architecture for SGG that extends the EGTR framework [4] by introducing advanced contextual reasoning capabilities. The architecture, which belongs to the class of one-stage methods and is shown in Figure 4, combines two approaches: EGTR extracts local relationships from the self-attention layers of the DETR decoder, while the Relational Transformer captures global relationships using specialised queries that observe the entire scene through the encoder's features.

Starting Pipeline The architecture begins with the standard DETR pipeline:

1. Backbone CNN: receives the P&ID image and extracts a feature map;

2. DETR encoder: processes the feature maps by integrating contextual information;
3. DETR decoder: processes N object queries through L layers of self-attention, generating intermediate $(\{Z^l\}_{l=0}^{L-1})$ and final (Z^L) representations.

Relational Reasoning The extraction of relationships is based on two main flows.

In the EGTR flow, which focuses on local analysis, interactions between objects that emerged during the detection process are captured. For each layer l of the DETR decoder, queries and keys are projected and replicated to form specific R_a^l relationship matrices for each pair:

$$\begin{aligned} Q_{\text{proj}}^l &= \text{Proj}_Q(Q^l), & K_{\text{proj}}^l &= \text{Proj}_K(K^l) \\ Q_{\text{proj}}^l &= \text{Proj}_Q(Q^l), & K_{\text{proj}}^l &= \text{Proj}_K(K^l) \\ R_a^l[i, j, :] &= [Q_{\text{proj}}^l[i, :]; K_{\text{proj}}^l[j, :]] \in \mathbb{R}^{N \times N \times 2d_{\text{model}}} \end{aligned}$$

The final representations are extracted from the last layer (L) of the decoder:

$$R_z[i, j, :] = [\text{Proj}_{\text{sub}}(Z^L[i, :]); \text{Proj}_{\text{obj}}(Z^L[j, :])] \in \mathbb{R}^{N \times N \times 2d_{\text{model}}}$$

The Relational Transformer, focused on global context analysis, processes features from the encoder and decoder. Initially, it enriches them through the projection of sinusoidal and learnable positional encoding (PE):

$$\begin{aligned} F_{\text{enc}}^* &= \text{Proj}_{\text{enc}}(\text{PE}_{\text{enc}}(F_{\text{enc}})) & Z^{L*} &= \text{Proj}_{\text{dec}}(\text{PE}_{\text{dec}}(Z^L)) \\ F_{\text{combined}} &= [F_{\text{enc}}^*; Z^{L*}] \in \mathbb{R}^{(S+N) \times d_{\text{model}}} \end{aligned}$$

Then, a set of N relational queries $Q_{\text{rel}} \in \mathbb{R}^{N \times d_{\text{model}}}$ is processed by a dedicated Transformer Decoder.

$$Q'_{\text{rel}} = \text{RelTransformer}(Q_{\text{rel}}, F_{\text{combined}}) \in \mathbb{R}^{B \times N \times d_{\text{model}}}$$

The Relational Transformer then applies:

- Self attention to specialize each expert: $Q_{\text{rel}} \leftarrow \text{SelfAttn}(Q_{\text{rel}}, Q_{\text{rel}}, Q_{\text{rel}})$;
- Cross attention to extract global patterns: $Q_{\text{rel}} \leftarrow \text{CrossAttn}(Q_{\text{rel}}, F_{\text{combined}}, F_{\text{combined}})$.

From the total N queries processed Q'_{rel} , the first $L + 1$ experts ($\mathcal{E}$) are selected and each is replicated for all pairs of objects

$$R_a'^l[i, j, :] = \mathcal{E}[l, :] \quad \forall i, j \in [1, N], \forall l \in [0, L]$$

Multi-Expert Adaptive Fusion For each layer, the representations extracted from the two flows are concatenated:

$$\begin{aligned} \tilde{R}^l &= [R_a^l; R_a'^l] \in \mathbb{R}^{N \times N \times 3d_{\text{model}}} \quad \forall l \in [0, L-1] \\ \tilde{R}^L &= [R_z; R_a'^L] \in \mathbb{R}^{N \times N \times 3d_{\text{model}}} \end{aligned}$$

For each pair of objects (i, j) and each layer/expert l , a custom gate determines its importance:

$$g_{i,j}^l = \sigma(\text{MLP}_{\text{gate}}(\tilde{R}^l[i, j, :])) \in \mathbb{R}^1$$

All representations are therefore merged through the weights of the relative gates:

$$\mathcal{R}_{\text{fused}}[i, j, :] = \sum_{l=0}^L g_{i,j}^l \odot \tilde{R}^l[i, j, :] \in \mathbb{R}^{N \times N \times 3d_{\text{model}}}$$

Finally, the result is converted into the final graph:

$$\begin{aligned} \hat{G}_{\text{rel}} &= \sigma(\text{MLP}_{\text{rel}}(\mathcal{R}_{\text{fused}})) \in \mathbb{R}^{N \times N \times |\mathcal{Y}_T|} \\ \hat{G}_{\text{conn}} &= \sigma(\text{MLP}_{\text{conn}}(\mathcal{R}_{\text{fused}})) \in \mathbb{R}^{N \times N \times 1} \end{aligned}$$

Benefits In the context of P&ID analysis, EGTR allows certain complexities to be overcome. Experts are able to cover local ($l \sim 1$) and global ($l \sim L$) aspects, capturing both short-range and macroscopic details and relationships. As a result, local experts ensure accuracy in fine details, while the context analysis provided by global experts allows for overcoming non-trivial challenges: for example, it is possible to effectively distinguish graphically similar symbols such as different types of valves and identify long-range connections suggested by pipes.

Output EGRTR produces a multi-component output:

- Bounding box and class probability for each identified component;
- A Relational Scene Graph G_{rel} , i.e. an $N \times N \times |\mathcal{Y}_T|$ matrix where $G_{\text{rel}}[i, j, k]$ represents the probability of the k -th relationship between components i and j identified;
- A Connectivity Graph G_{conn} , i.e. an $N \times N \times 1$ matrix where $G_{\text{conn}}[i, j]$ indicates the probability of direct connection between components i and j identified.

The output structured in this way encodes the components identified with their types, positions and characteristics (through the nodes and their attributes), the relationships between them with associated probabilities (through the arcs) and the structure of the P&ID up to level III (through the graph topology), and constitutes a piece of the input for the final optimization algorithms.

3.3.2 P&ID and Equipment List Matching via OCR

This section takes as inputs the plant's P&ID and the structured equipment list extracted in the previous task. Here, a further OCR analysis is performed on the image of the diagram in order to identify the numerical codes related to each present symbols, extracting their values and coordinates in the image. At this point, each identified code is associated with two entities:

- Its correspondence in the previously structured equipment list, therefore a component class specified by a couple type-subtype;
- The nearest symbol identified by the SGG model.

For each symbol of the scene graph will be built a new list of class probabilities, related to the interpretation of the equipment list. This new set of values is used to modulate the output class probabilities of the SGG model, in order to combine topological and descriptive information.

3.3.3 Optimization Algorithm

In order to harmonize the final output of the solution with the information contained in the Regulations, post-processing is performed on the hierarchical representation of the P&ID provided by the SGG model. The Regulations define a set of rules and constraints that the hierarchical structures of the plants must satisfy, and it is therefore necessary for the final output of the solution to be compatible with these constraints. In particular, the hierarchical levels subject to post-processing are levels I, II and III, i.e. the definition of component classes (type and sub-type), their grouping into sections and the types of sections. In addition, at this stage, level IV of the plant's hierarchical structure is reconstructed, i.e. the grouping of sections into lines and the definition of line types (measurement or regulation). However, with regard to the permissible compositions of sections in lines, the Regulation defines minimum requirements, and these often do not allow for the unambiguous identification of the plant's lines. The information contained in the plant registry is therefore also used at this stage. The registry specifies the hierarchical structures of the installed plants, from which it is possible to extract a statistical representation of the composition of the plants and define the likelihood of a given hierarchical structure. To produce the final hierarchical structure of the plant, an optimization algorithm has been implemented. It takes as input the merged information extracted from the previous phases of documentation analysis, the set of rules and constraints in the Regulations, and the overall structural likelihood of the plants learned from the registry. The merged information extracted from the previous phases of documentation analysis consist of:

- Bounding boxes and class probabilities for each identified component. The bounding boxes are provided by the SGG model while the probabilities are weighted averages between those of the SGG model and those of the equipment list matching method;
- The Relational Scene Graph G_{rel} and the Connectivity Graph G_{conn} as they are produced by the SGG model.

The goal of the algorithm is to find partitions $\mathcal{T}, \mathcal{L}$ and classes $\mathbf{y}$ so that:

$$\arg \max_{\mathcal{T}, \mathcal{L}, \mathbf{y}} \mathcal{S}(\mathcal{T}, \mathcal{L}, \mathbf{y}),$$

where the objective function to be maximized is:

$$\mathcal{S}(\mathcal{T}, \mathcal{L}, \mathbf{y}) = \lambda_1 \mathbf{E}_{\text{node}}(\mathbf{y}) + \lambda_2 \mathbf{E}_{\text{edge}}(\mathcal{T}, \mathbf{y}) + \lambda_3 \mathbf{E}_{\text{struct}}(\mathcal{T}, \mathcal{L}, \mathbf{y}) - \lambda_4 \mathbf{E}_{\text{norm}}(\mathcal{T}, \mathcal{L}, \mathbf{y}) - \lambda_5 \mathbf{E}_{\text{reg}}(\mathcal{T}, \mathcal{L}).$$

The terms of the objective function represent the probabilities provided by the previous phases about components and connections classes, the overall structural plausibility learned from the plant registry, the penalties resulting from violations of the constraints defined in the Regulations, and a regularization. The terms $\mathbf{E}_{\text{node}}(\mathbf{y})$ and $\mathbf{E}_{\text{edge}}(\mathcal{T}, \mathbf{y})$ are defined

based on the class probabilities $\hat{p}_i(y_i)$ associated with each component c_i , the connectivity graph $\hat{G}_{\text{conn}} \in [0, 1]^{N \times N}$ representing the probability that two components are connected, and the relational graph $\hat{G}_{\text{rel}} \in [0, 1]^{N \times N \times |\mathcal{Y}_T|}$ representing the probability that a relationship of type $k \in \mathcal{Y}_T$ exists between two components:

$$\mathbf{E}_{\text{node}}(\mathbf{y}) = \sum_{i=1}^N \log \hat{p}_i(y_i), \quad \mathbf{E}_{\text{edge}}(\mathcal{T}, \mathbf{y}) = \sum_{\substack{k=1 \\ i, j \in T_k, i \neq j}}^{|\mathcal{T}|} \log \left(\hat{G}_{\text{conn}}[i, j] \cdot \hat{G}_{\text{rel}}[i, j, y_{T_k}] \right).$$

The term $\mathbf{E}_{\text{struct}}(\mathcal{T}, \mathcal{L}, \mathbf{y})$ is the logarithm of a plausibility score for the hierarchical structure of the plant defined on the basis of the information contained in the plant registry. It was constructed by training a probabilistic discriminative model that approximates a monotonically increasing likelihood score with respect to the probability that the hierarchical structure corresponds to a real plant $P(\mathcal{T}, \mathcal{L}, \mathbf{y} \mid \text{Real plant})$. The term $\mathbf{E}_{\text{norm}}(\mathcal{T}, \mathcal{L}, \mathbf{y})$ penalizes hierarchical structures that are not compatible with the constraints of the Regulations, and is defined as:

$$\mathbf{E}_{\text{norm}}(\mathcal{T}, \mathcal{L}, \mathbf{y}) = -\log \left(0.5 \cdot \frac{\sum_{k=1}^{|\mathcal{T}|} \phi_T(y_{T_k}, \{y_i \mid c_i \in T_k\})}{|\mathcal{T}|} + 0.5 \cdot \frac{\sum_{j=1}^{|\mathcal{L}|} \phi_L(y_{L_j}, \{y_{T_k} \mid T_k \in L_j\})}{|\mathcal{L}|} \right),$$

where ϕ_T and $\phi_L \in [0, 1]$ are continuous scoring functions that evaluate the compliance of segments and lines with the minimum requirements of the Regulation. Finally, the regularization term $\mathbf{E}_{\text{reg}}(\mathcal{T}, \mathcal{L})$ penalizes segments that are too large and an excessive number of partitions:

$$\mathbf{E}_{\text{reg}}(\mathcal{T}, \mathcal{L}) = \alpha_1 |\mathcal{T}| + \alpha_2 |\mathcal{L}| + \alpha_3 \sum_{k=1}^{|\mathcal{T}|} |T_k|^2.$$

4 Results and Discussions

Table 1 shows the performance obtained in extracting textual information from the various documents. The values represent, in order, the accuracy of reading the page containing the design data, reading the tables containing the list of equipment, matching the equipment with the respective records in the regulations, and extracting the technical specifications of the components.

KPI	Design data	Equipment	Registry match	Specifications
Accuracy	91%	80%	84%	84%

Table 1: Accuracy values obtained in the extraction of textual information from documentation

Table 2 shows the scores characterizing the final performance of the SGG model. The relatively low values are justified by the following characteristics of the data used:

- The bounding boxes of the components, generated manually, are highly variable even for the same symbol. This makes it impossible for the model to accurately predict the coordinates of each bounding box, which is heavily penalized in the mAP calculation as the IoU threshold increases.
- Due to the way the dataset was constructed, a single P&ID can contain a large number of relationships between components, even more than 1000. This is directly reflected in the lowering of traditional scores for measuring the performance of SGG models.

However, as far as mAP is concerned, it is more important that the model identifies a component and accurately predicts its class rather than its position. The IoU decision threshold can therefore be relatively low.

Furthermore, with regard to recall on relationships, it can be seen that as the number of predicted relationships taken into consideration increases, the score value increases proportionally. This indicates a high accuracy of the model in this field.

Visual examples of EGRTR's ability to identify components and relate them to each other based on the section they belong to are shown in Figure 5.

The above discussion is reflected in the results shown in Table 3, which relate to the identification of equipment and sections downstream of the optimization algorithm.

mAP@[.5]	mAP@[.75]	mAP@[.5:.95]	R@20	R@50	R@100
0.762	0.428	0.442	0.089	0.225	0.424

Table 2: Main results achieved following EGRTR training on labelled test P&IDs.

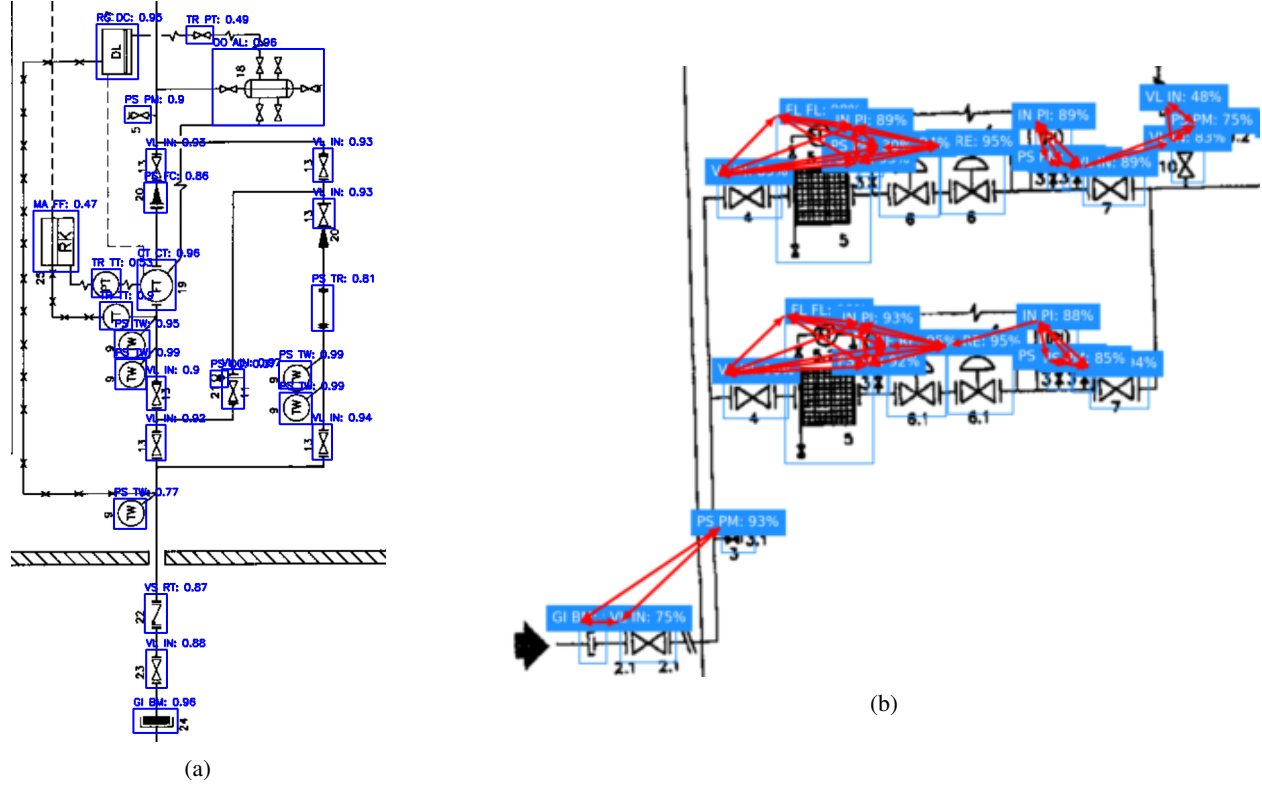


Figure 5: (a) Example of inference by EGRTR for the recognition of components in a test P&ID. (b) Example of inference by EGRTR for the identification of relationships (in red) between recognised components in a test P&ID.

The percentage value for the components identified is the ratio between those correctly identified and those present in the diagram and available in the ground truth. The correct identification of a component is understood as the correct identification not only of its position but also, and above all, of its type and subtype.

The percentage value for the sections is calculated using a weighted average of the individual scores obtained on the sections: this compares how many components in the individual section are correctly associated with the components actually present in the section, weighted by the complexity of the section itself, proportional to the number of components that make it up. The individual scores are then averaged to obtain a single value on the P&ID.

KPI	Identified Components	Components correctly assigned to sections	Identified Regulation sections	Identified Measure sections
Accuracy	95%	90%	87%	85%

Table 3: Accuracy values obtained in the identification of components and sections

5 Limitations and Conclusions

This work proposes an end-to-end system for the digitization of gas plants that optimizes results through the comprehensive use of the relevant documentation that is usually available. The analysis pipeline consists of various sequential blocks based on Artificial Intelligence, OCR and constrained optimisation techniques. Given its performance, the system developed is already applicable as support for MGM users but has room for improvement to make it even more effective. At present, while excellent results have been achieved on simpler systems, performance on complex systems, which are those for which more resources are used, is not excellent. Possible future developments to overcome this limitation would include an increase in the cardinality of the available dataset and the refinement of data augmentation techniques.

In conclusion, the system is an effective resource for the purposes of this work, and its components are potentially scalable to other tasks: methods and models can be used in various other fields of study that involve the analysis of technical documentation in addition to the digitization of plants.

6 Acknowledgments

Part of this work is framed within the Project "AVANT" – Project no. IPCEI-CL_0000005 - Application protocol no. 108421 of 14/05/2024 - CUP B89J24002920005 - Grant decree no. 1322 of August 8, 2024 - financed by the European Union – NextGenerationEU (IPCEI Funding).

References

- [1] Iliia Karmanov, Amala Sanjay Deshmukh, Lukas Voegtli, Philipp Fischer, Kateryna Chumachenko, Timo Roman, Jarno Seppänen, Jupinder Parmar, Joseph Jennings, Andrew Tao, and Karan Sapra. éclair – extracting content and layout with integrated reading order for documents, 2025.
- [2] Yitong Zhou, Mingyue Cheng, Qingyang Mao, Qi Liu, Feiyang Xu, Xin Li, and Enhong Chen. Enhancing table recognition with vision llms: A benchmark and neighbor-guided toolchain reasoner. *arXiv preprint arXiv:2412.20662*, 2024.
- [3] Jianwei Yang, Jiasen Lu, Stefan Lee, Dhruv Batra, and Devi Parikh. Graph r-cnn for scene graph generation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 670–685, 2018.
- [4] Jinbae Im, JeongYeon Nam, Nokyoung Park, Hyungmin Lee, and Seunghyun Park. Egtr: Extracting graph from transformer for scene graph generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24229–24238, 2024.
- [5] Rongjie Li, Songyang Zhang, and Xuming He. Sgtr: End-to-end scene graph generation with transformer. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 19486–19496, 2022.
- [6] Yao Teng and Limin Wang. Structured sparse r-cnn for direct scene graph generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19437–19446, 2022.
- [7] 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.
- [8] 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*, pages 176–177, 2020.
- [9] Byung Chul Kim, Hyunki 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.
- [10] 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: PAKDD 2021 Workshops, WSPA, MLMEIN, SDPRA, DARAI, and AI4EPT, Delhi, India, May 11, 2021 Proceedings 25*, pages 168–180. Springer, 2021.
- [11] Shreya M Gajbhiye, SR Bhamre, LN Teja Tadepalli, MR Pillai, and Deepak Uplaonkar. Advancing p&id digitization with yolov5. In *2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS)*, pages 1–6. IEEE, 2023.
- [12] Jan Marius Stürmer, Marius Graumann, and Tobias Koch. Transforming engineering diagrams: A novel approach for p&id digitization using transformers. *arXiv e-prints*, pages arXiv–2411, 2024.