



ARTICLE

Enhancing Object Detection in Electrical Substations through Post-Processing Module with Spatial Contexts

Jordán Pascual Espada¹, Lucía Alonso Virgós^{2,*}, Juan Luis Carús³ and Miguel Ángel Fernández Fernández³

¹Faculty of Computer Science and Engineering, Universidad de Oviedo, Oviedo, Asturias, Spain

²Faculty of Engineering and Technology, Universidad Internacional de la Rioja, Logroño, La Rioja, Spain

³Faculty of Engineering and Technology, TSK Group, Gijón, Asturias, Spain

*Corresponding Author: Lucía Alonso Virgós. Email: lucia.alonso.virgos@unir.net

Received: 09 December 2025; Accepted: 31 March 2026; Published: 13 August 2026

ABSTRACT: This research focuses on multi-object detection for interrelated industrial components, such as substation parts. Traditional vision models like YOLO rely primarily on visual features, which limits their ability to: (1) disambiguate visually similar objects using contextual cues like relative position or size, and (2) reinforce low-confidence detections that are spatially plausible given neighboring elements. For instance, a visually similar but semantically incorrect object may be misclassified, or a valid but partially occluded component may be discarded due to a low appearance-based score. Our approach addresses these issues by integrating formalized rules based on relative positions and proportional sizes, enabling contextual reasoning that improves detection accuracy. These challenges are common in industrial inspection scenarios, where components may be occluded, poorly lit, or surrounded by visually similar distractors. Experimental results show that the proposed spatially informed refinement significantly improves precision (from 95.1% to 98.1%) and recall (from 86.5% to 88.9%), leading to a higher F1 score (from 0.905 to 0.933). The system effectively reduces false positives and recovers missed detections, particularly in complex or cluttered scenes, demonstrating the value of integrating spatial logic into object-detection workflows. Although the experiments were conducted on a curated dataset of 714 fixed-camera images from electrical substations, the proposed methodology is scalable and adaptable to other structured domains where spatial relationships are critical for accurate object recognition.

KEYWORDS: Object detection; electrical substations; convolutional neural networks; YOLOv8; rule-based systems

1 Introduction

Object detection has become a key enabling technology for automated inspection and monitoring in industrial environments. In safety-critical infrastructures such as electrical substations, accurate identification of components is essential to support preventive maintenance, fault detection, and operational safety. False positives may trigger unnecessary inspections or interventions, while false negatives can delay the detection of critical anomalies, increasing the risk of equipment failure and personnel exposure. As a result, reliability in object detection is not only a matter of accuracy but also a prerequisite for safe and efficient industrial operation.

Despite the remarkable progress achieved by modern convolutional neural networks (CNNs), purely appearance-based detectors still exhibit important limitations in highly structured industrial settings. State-of-the-art models such as YOLOv8 excel at learning visual patterns from data, yet they typically process objects in isolation and lack explicit mechanisms to encode spatial and structural constraints. In electrical substations, components follow strict physical and functional layouts: connections are usually located near the upper regions of the structure, radiators are partially contained within the main body, and expansion tanks occupy well-defined positions relative to the core equipment. Consequently, standard detectors cannot reduce confidence for visually similar objects that violate these positional rules, nor boost it for low-confidence detections that align with the expected structural context. This rigidity results in recurring false positives and missed detections, revealing a gap between visual recognition capabilities and the spatial logic inherent to industrial infrastructure.

Several recent works have attempted to address this limitation by incorporating contextual cues into object detection through attention mechanisms, feature aggregation strategies, or learned contextual priors [1,2]. While these approaches improve robustness in complex scenes, they typically rely on implicit representations learned from data and remain largely generic. In environments governed by strong physical constraints, such as electrical substations, the absence of explicit, interpretable spatial reasoning limits the ability of these models to reject structurally implausible detections or recover components that are visually ambiguous but spatially coherent. This motivates the exploration of hybrid approaches that combine data-driven visual detection with domain-specific spatial logic [3–6].

A key enabler of this advance is our novel, curated dataset of 714 fixed-camera images of electrical substations, manually annotated into five classes (connections, expansion tanks, bodies, radiators, electrical substations). Fixed-camera imagery provides stable perspective and geometry, which is crucial to learn and validate spatial regularities and to evaluate a spatially informed post-processing stage with minimal confounding from viewpoint drift. Beyond standard benchmarking, we include a diagnostic subset of 40 challenging scenes characterized by occlusions, clutter, and out-of-context regions to stress-test how spatial reasoning corrects typical CNN failure modes.

This paper’s contributions are threefold:

Methodological: We introduce a context-aware post-processing stage that quantitatively re-weights CNN detections using domain-specific spatial constraints (position, containment, proportion). This design preserves the probabilistic nature of detector outputs—improving robustness compared to hard logical filters—while injecting interpretable structure that CNNs do not encode explicitly.

Data and evaluation: We release a well-specified dataset protocol for fixed-camera substation imagery with class-level annotations aligned to operationally relevant components, together with a diagnostic evaluation protocol that isolates and measures performance on the most safety-critical failure cases.

Empirical evidence with operational relevance: On real-world substation images, our framework consistently improves precision, recall, and F1 over the CNN baseline (e.g., precision 95.1% $\rightarrow$ 98.1%, recall 86.5% $\rightarrow$ 88.9%, F1 0.905 $\rightarrow$ 0.933), while visually reducing out-of-context false positives and recovering occluded components. These gains translate into clear practical benefits for substation monitoring—fewer false alarms, better maintenance targeting, and improved worker safety due to higher confidence in component localization.

Improving precision detection in electrical substations has direct operational and safety implications. More accurate identification and localization of connections, radiators, and expansion tanks enable earlier to address this challenge, we propose a hybrid detection framework that integrates a strong CNN baseline, YOLOv8, with a context-aware post-processing module based on explicit spatial rules. Instead of applying

rigid accept-or-reject filters, the proposed system refines detection confidence scores using deterministic, yet interpretable constraints derived from the physical semantics of electrical substations. These constraints capture three key aspects of component layout: relative position within the structure, containment and overlapping relationships between elements, and proportional size consistency. By operating as a confidence-aware refinement stage, the framework preserves the probabilistic nature of CNN outputs while injecting spatial knowledge that is not explicitly encoded during neural training.

The proposed approach is evaluated using a curated dataset of 714 fixed-camera images of electrical substations, annotated across five operationally relevant classes. Fixed viewpoints provide stable geometry, enabling both the learning and validation of spatial regularities. In addition to standard benchmarking, we introduce a diagnostic evaluation focused on challenging failure cases characterized by occlusions, clutter, and out-of-context detections, offering insight into how explicit spatial reasoning complements purely visual detectors. The main contributions of this work can be summarized as follows:

- (i) a context-aware post-processing framework that improves object detection reliability through explicit spatial reasoning without modifying the underlying CNN architecture.
- (ii) an evaluation strategy emphasizing safety-critical detection failures in real industrial imagery; and
- (iii) empirical evidence demonstrates significant gains in precision, recall, and F1 score, with direct operational relevance for electrical substations and other structured industrial environments.

2 Background

The use of convolutional neural networks (CNNs) for object detection has expanded across numerous industrial applications. However, in infrastructures with defined structural layouts, such as electrical substations conventional models may produce errors by misidentifying visually similar elements or failing to detect certain components due to changes in lighting, equipment arrangement, or the presence of obstructing objects. These issues can compromise the reliability of automated monitoring systems.

This section reviews recent advances in object detection using CNNs, with particular emphasis on methods that incorporate contextual information to improve accuracy. It also explains why the use of spatial rules can be an effective solution to reduce errors and enhance detection reliability in structured industrial environments.

2.1 Object Detection

Convolutional neural networks have enabled significant progress in automatic object detection. Models such as YOLO and Faster R-CNN have demonstrated strong performance across a range of tasks, and newer versions of YOLO continue to improve speed and accuracy [7–9].

The authors propose an optimized deep learning pipeline using TensorFlow Lite on Android to achieve efficient on-device inference, thereby addressing latency and portability constraints that are critical in assistive technologies. Furthermore, the paper provides a comparative evaluation against established object detection models such as Faster R-CNN and SSD, demonstrating improved detection accuracy and computational efficiency. By integrating state-of-the-art deep learning techniques with practical mobile deployment and real-world testing, the study advances existing work in intelligent assistive systems and highlights the feasibility of lightweight, real-time computer vision solutions for accessibility applications.

This research work provides a comprehensive review of real-time object detection techniques based on YOLO and CNN models, analyzing their architectural foundations, performance trade-offs, and practical applications. It systematically compares regression-based one-stage detectors such as YOLO with

classification-based and two-stage approaches, highlighting YOLO's advantages in speed, unified architecture, and real-time performance. The study also discusses improvements across different YOLO versions. The paper offers a structured overview of the evolution of YOLO-based detection frameworks and outlines future research directions, making it a valuable reference for understanding current trends and challenges in real-time deep learning-based object detection [10].

However, these models typically process objects independently from their surrounding context, treating each detection as an isolated event rather than part of a cohesive structural system. This limitation can lead to significant errors in highly structured environments, where component placement follows rigorous engineering standards. Misclassifications may arise when components share similar shapes or colors, causing the model to confuse non-target objects with critical infrastructure. Furthermore, when their expected spatial arrangements are not considered, the system fails to leverage relational cues, such as relative position and scale [11–13]. Without this contextual reasoning, the model cannot suppress false positives that violate spatial norms nor reinforce true positives that are visually ambiguous but contextually plausible, ultimately limiting its robustness in complex industrial scenes.

Furthermore, many existing detectors do not explicitly leverage the predictable organization of substation components, which is dictated by strict engineering standards and safety regulations. Specific structural priors such as the consistent positioning of connectors near the upper regions of the structure are often ignored during inference [13–15]. Without this spatial awareness, models operate without crucial contextual constraints, making them prone to incorrectly classifying background patterns or clutter as equipment when visual features align superficially. Conversely, they may fail to recognize partially occluded components or those under poor lighting conditions, as they cannot boost confidence based on the logical presence of an object in a specific location. This lack of structural reasoning limits their ability to distinguish between plausible infrastructure and visual noise, ultimately reducing reliability in complex operational scenarios.

Recent studies have specifically evaluated YOLOv8-based object detection under diverse real-world and deployment-oriented conditions, including scenarios characterized by environmental variability, clutter, and partial occlusion [16]. For example, recent analyses in signal and image processing contexts have shown that YOLOv8 achieves strong baseline performance across a range of challenging operational settings, particularly due to its improved feature aggregation and decoupled detection head. However, these studies also report that YOLOv8, like other modern one-stage detectors, remains fundamentally appearance-driven and lacks explicit mechanisms to encode structural or relational constraints between objects, which limits its robustness in highly organized environments governed by physical layout rules (e.g., component co-location, containment, or proportional size relationships). As a result, while architectural refinements and training strategies can improve visual discrimination, they do not address recurring failure modes caused by spatial implausibility or contextual inconsistency. This gap motivates complementary approaches—such as the one proposed in this work—that enhance YOLOv8 detections through explicit, domain-specific spatial reasoning applied as a post-processing stage, without modifying the underlying neural architecture.

2.2 Integration of Contextual Information

To address these limitations, several approaches have begun integrating contextual cues, attention mechanisms, or structural priors into CNN-based detectors. These methods demonstrate the potential of combining visual features with contextual reasoning, but most remain general-purpose and are not tailored to the specific spatial semantics of electrical infrastructure. This motivates the development of approaches that incorporate domain-specific spatial constraints, particularly in environments where component placement follows strict physical logic [17–19].

Previous described limitations addressed in this work reflect a fundamental challenge shared by state-of-the-art object detection frameworks, including Faster R-CNN, DETR, EfficientDet, and all YOLO variants. These models, regardless of architecture, rely predominantly on appearance-based features and lack native mechanisms to encode domain-specific contextual constraints such as absolute or relative object locations, inter-class spatial relationships, proportional size expectations, or semantic overlap rules. Consequently, integrating contextual reasoning requires an external filtering mechanism or applied as a post-processing layer. YOLOv8 was selected as the baseline architecture for this study due to its improved accuracy, enhanced feature aggregation, and decoupled head design, which currently make it a strong reference for building robust computer vision systems in industrial settings [18,19]. However, the proposed contextual refinement strategy is architecture-agnostic: if successfully applied, these rules would enhance any detection model, as none of them currently incorporate this type of structural reasoning by design. Several recent benchmarks report its superior performance compared with other architectures, particularly in scenarios involving small or partially occluded [19].

2.3 Applications of Spatial Rules in Other Sectors

In other industrial sectors, spatial rules have proven effective in object detection tasks. In machinery inspection, spatial rules are applied to ensure that parts are correctly positioned within an assembly system. In the petrochemical industry, for example, visual monitoring systems based on spatial rules are used to identify pipeline components, detecting whether they are properly aligned or showing deviations that could compromise safety [18]. Similarly, in the automotive industry, neural networks trained with contextual information about part positioning can ensure that vehicle components are correctly placed for assembly [13].

This study proposes combining convolutional neural networks with spatial rules based on three principles:

- Proximity: Verifies whether objects are logically positioned relative to other elements, reducing erroneous detections.
- Containment: Confirms that each object is located within the expected area, minimizing confusion.
- Relative Size: Assesses whether the proportions between objects are appropriate, filtering out incorrect identifications.

Additionally, the dataset used for model training will be expanded to improve its ability to recognize objects under varying conditions and reduce detection bias.

These enhancements aim to reduce the errors affecting current models and develop a more reliable system for component identification in electrical substations. This approach not only improves automated monitoring of such infrastructures but could also be applied to other environments where object distribution is key to accurate detection—such as pipeline supervision in industrial plants or machinery inspection in automated factories.

3 Methodology

The proposed system follows a two-stage object detection pipeline that combines data-driven visual recognition with explicit spatial reasoning. In the first stage, a convolutional neural network identifies candidate objects in each image and assigns an initial confidence score to every detection. In the second stage, a context-aware post-processing module refines these confidence scores by applying spatial and structural rules derived from the physical layout of electrical substations. This modular design allows spatial reasoning to be incorporated without modifying the underlying neural architecture.

Throughout the manuscript, the term “electrical substation” is used consistently to refer to the overall equipment system. Alternative expressions previously used in the text (such as “substation structure” or “main structure”) have been standardized to avoid ambiguity. The term “body” is retained only as a dataset class label.

To establish a robust baseline for the first stage, several state-of-the-art detection architectures were evaluated, including Faster R-CNN, RetinaNet, EfficientDet, DETR, and YOLO variants (v5, v7, and v8) using a preliminary assessment on the curated substation dataset. While all models demonstrated competitive visual recognition capabilities, YOLOv8 was ultimately selected as the base detector due to its superior balance of detection accuracy, inference speed, and architectural improvements such as a decoupled detection head, enhanced feature pyramid network, and stable optimization dynamics. These properties make YOLOv8 a robust starting point for industrial object detection tasks, particularly in environments where components may appear at different scales or be partially occluded. Nevertheless, the choice of baseline architecture is not central to the core contribution of this work: the proposed contextual refinement mechanism is designed to be architecture-agnostic and can be integrated with any modern object detector. The key innovation lies not in the visual backbone itself, but in the external rule-based filtering layer that enhances detection reliability by encoding domain-specific spatial constraints, thereby improving performance regardless of the underlying vision model.

To ensure a fair comparison in the ablation study, all training conditions—including hyperparameters, initialization weights, input resolution, data splits, augmentation settings, and optimization schedule—were kept strictly identical between the Base Model (YOLOv8) and the Full System (YOLOv8 + spatial rules). The rule-based module was applied only after inference and was evaluated independently for each object class to quantify its specific contribution to true positives, false positives, and false negatives on a per-class basis.

The dataset comprises 714 images of outdoor electrical substations, combining real photographs captured from fixed inspection viewpoints with stock images. The five annotated classes: connections, expansion tanks, bodies, radiators, and main structure (body), are inherently interrelated, representing functional components of a cohesive electrical substation system. This structural interdependence is critical, as detection errors often arise not only from visual ambiguity but also from violations of expected spatial relationships between parts. Additionally, variable lighting conditions, shadows, and color similarities can cause visually indistinguishable elements to be confused, further complicating appearance-based recognition. A recurring challenge is the presence of objects highly similar to substation components, such as pipes and other electrical elements located outside the actual infrastructure, which can trigger false positives when context is ignored. Annotated with bounding boxes and stratified into training (70%), validation (20%), and testing (10%) splits to preserve class proportions, the dataset ensures methodological rigor while reflecting real-world complexity. A comprehensive description of the dataset construction, annotation protocol, and illustrative examples of both standard and diagnostic evaluations is provided in detail in [Section 3.1](#).

To train the neural network for detecting substation elements, the YOLOv8 model was trained using a dataset defined in *data.yaml* containing labeled images of substations and their relevant components. Training was conducted for 200 epochs with an input resolution of 800 pixels, using stochastic gradient descent (SGD) as the optimizer with a momentum of 0.937 and a weight decay of 0.0005 to prevent overfitting. The initial learning rate was set to 0.001 and gradually decreased by a factor of 0.1 to stabilize the training process. This configuration was selected to ensure stable convergence while preventing overfitting.

To improve generalization, various data augmentation techniques were applied, including horizontal and vertical flipping (0.5 probability), rotations up to 10 degrees, random translations (0.1), scaling (0.5), shearing (2 degrees), and mixup augmentation (0.2). Mosaic augmentation (1.0) was also incorporated to enhance robustness against image variability. An early stopping patience of 20 epochs was set to halt

training if no significant performance improvements were observed. These augmentations simulate realistic variability in substation imagery and help the model learn more stable representations.

To achieve optimal performance, multiple training sessions were conducted exploring different hyperparameter combinations. Grid search was used to define candidate values for each parameter and systematically evaluate all configurations. The setup yielding the best balance between accuracy and stability was selected. This tuning ensures that the baseline CNN model is strong and well-calibrated before applying the spatial refinement stage.

This study argues that incorporating the rule-based system helps reduce such misclassifications by validating detections based on their location, size, and relationship to the electrical substations, discarding those that do not conform to the expected spatial logic.

As shown in Fig. 1, the model exhibits a positive evolution in its performance metrics. The loss curves for bounding boxes (box_loss), classification (cls_loss), and distribution focal loss (dfl_loss) show a progressive decrease during both training and validation phases, indicating that the model was successfully adapting to the data.

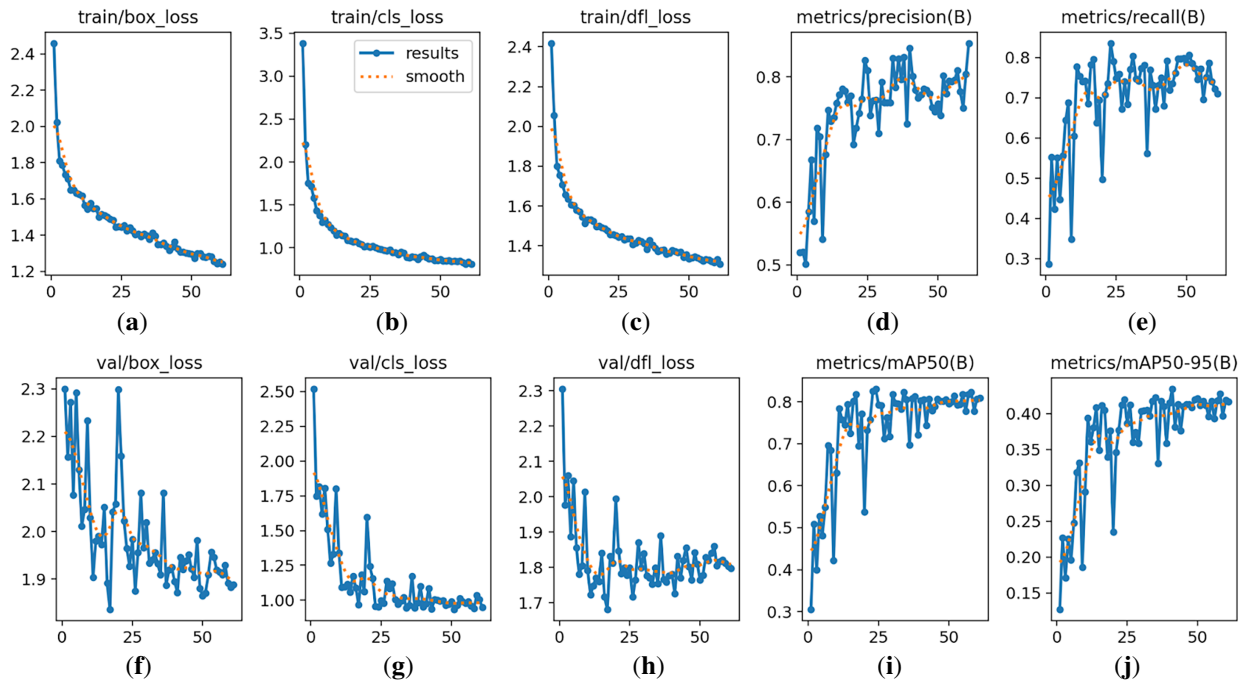


Figure 1: Evolution of YOLOv8 metrics and losses. (a) train/box_loss: training box-regression loss. (b) train/cls_loss: training classification loss. (c) train/dfl_loss: training DFL loss. (d) metrics/precision(B): precision metric as reported in the logs. (e) metrics/recall(B): recall metric as reported in the logs. (f) val/box_loss: validation box-regression loss. (g) val/cls_loss: validation classification loss. (h) val/dfl_loss: validation DFL loss. (i) metrics/mAP50(B): mAP@50 metric as reported in the logs. (j) metrics/mAP50-95(B): mAP@50–95 metric as reported in the logs.

At the same time, the precision and recall metrics increase steadily, suggesting that the model not only learned to identify objects more accurately but also managed to detect a greater proportion of relevant elements within the images.

This behavior—characterized by decreasing losses and improving metrics—confirms that the tuning process using Grid Search was effective in identifying an optimal configuration. The resulting model generalizes well across different scenarios and delivers consistent performance in object detection tasks.

Fig. 2 provides a visual representation of the model's performance in classifying five categories: connector, electrical substations, radiator, expansion tank, and background. Diagonal values indicate correct classifications, while off-diagonal values represent misclassifications. The model shows strong precision levels for the main classes, but there is a noticeable tendency to confuse the background with objects such as connectors (17%) and radiators (16%). These types of errors can lead to false positives in areas where no detections should occur.

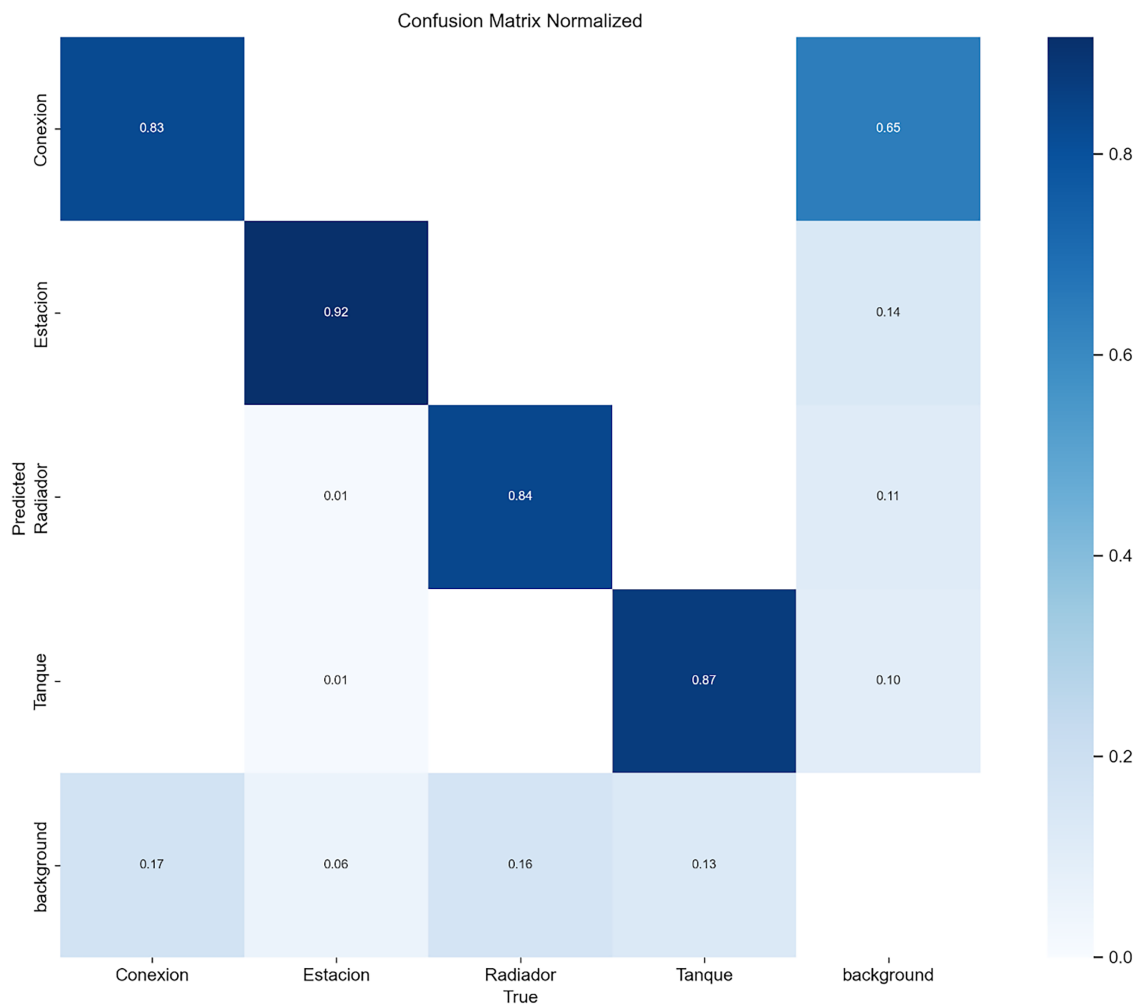


Figure 2: Normalized confusion matrix.

The YOLOv8 computer vision system was trained and validated on the curated substation dataset to establish a strong appearance-based detection baseline. Following training, a detailed diagnostic analysis was conducted to identify specific failure patterns. Two recurring error types were observed (Fig. 3): (1) False positives from visual ambiguity, where objects sharing similar visual features with target components (e.g., structural elements or background clutter) were incorrectly classified as substation infrastructure, despite violating domain-specific spatial or contextual constraints. (2) False negatives due to low confidence,

where genuine substation components were either missed or assigned insufficient confidence scores, typically caused by partial occlusion, unfavorable lighting, or small object scale; this systematic error characterization informed the design of the subsequent rule-based refinement module, which specifically targets these limitations by integrating contextual reasoning into the detection pipeline.

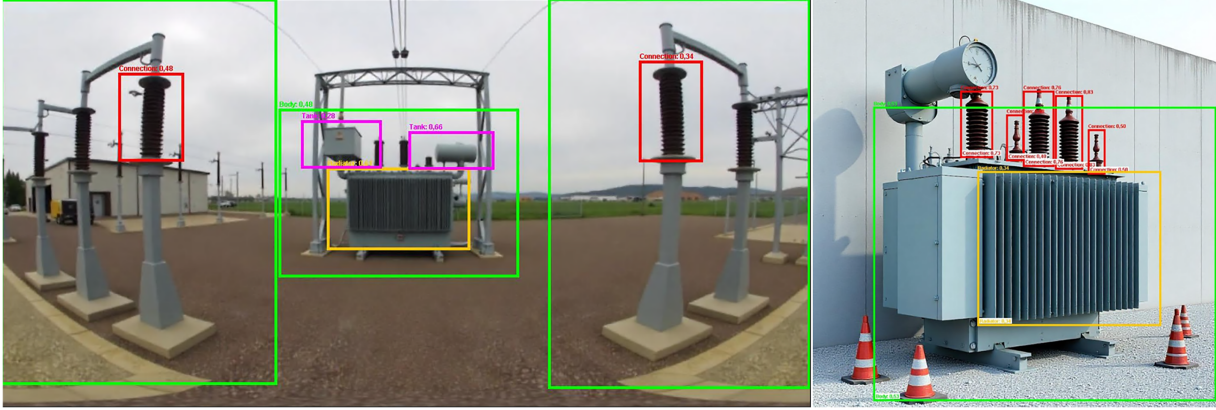


Figure 3: Illustrative examples of the two primary failure modes in appearance-based detection.

Left: False positive detection. Connector-like objects located outside the electrical substations are incorrectly identified as valid connections due to visual similarity. This occurs despite violating domain-specific spatial constraints, which dictate that connectors should be positioned near the upper regions of or overlapping with the main body.

Right: False negative detection. An expansion tank is missed or assigned a low confidence score due to challenging visual conditions. Its spatial relationship to neighboring components, such as the expected position relative to the main body, connectors, and radiator, strongly suggests that it is a valid instance.

To address the limitations of traditional appearance-based systems, we propose a rule-based module that modifies detection confidence based on contextual relationships between elements within the substation. This rule-based module acts on the detections produced by the neural network and is grounded in three core principles:

- (i) **Relative Size:** Components such as connectors, radiators, or expansion tanks must exhibit appropriate proportions relative to the electrical substations.
- (ii) **Relative Position:** Certain elements, such as connectors or expansion tanks, are expected to appear in specific regions of the structure according to their functional role.
- (iii) **Containment:** Objects such as radiators are required to be partially or fully contained within the electrical substations to be considered physically valid detections.

The definition of these spatial rules commenced with a systematic visual analysis of the dataset to identify recurring patterns of absolute and relative relationships among components, including size ratios and overlap percentages. Initially, a broad set of candidate rules was generated based on these observations, followed by a pruning phase to eliminate those contributing negligible value. Specific coefficients were then determined to quantify the remaining spatial constraints, such as proportional size thresholds and positional structures. The impact of each rule on the final confidence score was evaluated using multiplicative factors for either increment or reduction. To optimize these factors, an automated Python script was developed to perform an exhaustive maximum exploration, iteratively testing coefficient values on the validation and test sets. This calibration process continued until the configuration maximizing True Positives (TP)

while minimizing False Negatives (FN) and False Positives (FP) was identified, with all trials and results systematically logged for analysis.

To provide a clearer understanding of the strictness of these constraints, we include here a concrete numerical example from the calibration process. A radiator detection was considered spatially valid only when it exhibited at least an 80% overlap ($\text{IoU} \geq 0.80$) with the body and its area did not exceed roughly 50% of the body's area. Similarly, a connector was treated as contextually plausible when its relative size fell within 30%–50% of the body's size and it overlapped the upper region of the structure. These numerical thresholds illustrate the typical level of rigidity applied during rule formalization.

These principles encode domain knowledge about the physical and functional organization of electrical substations and guide the design of explicit spatial constraints used to refine detection outputs. The formal rules are applied independently to each image and adjust detection confidence according to specific spatial criteria. In particular, the following confidence modifications are performed:

- Connector not overlapping the top of the electrical substations → reduce confidence by 0.83
- Connector overlapping the top of the electrical substations → increase confidence by 0.42
- Connector at least 50% smaller than the electrical substations and contained or overlapping → increase confidence by 0.31
- Connector with approximately 50% of the electrical substations's size → reduce confidence by 0.76
- Radiator not partially contained within the substation → reduce confidence by 0.70
- Radiator fully or partially contained within the substation → increase confidence by 0.47
- Radiator larger than 80% of the electrical substations and contained or overlapping → reduce confidence by 0.76
- Electrical substations overlapping with connectors, radiators, or tanks → increase confidence by up to 0.62, with 0.12 added per detected element
- Electrical substations without overlaps with connectors, radiators, or tanks → reduce confidence by 0.20
- Expansion tank not overlapping or located near the top of the substation → reduce confidence by 0.34
- Expansion tank overlapping the top of the electrical substations → increase confidence by 0.54

**A complete description of the mathematical formulations and algorithmic implementation of the spatial rules is provided in [Section 3.2](#) (Algorithms).

These refinements help suppress physically implausible detections while reinforcing detections with strong spatial coherence relative to the expected substation layout.

The proposed system uses the neural network to identify objects and assign an initial confidence score to each detection. These scores are then modified using the spatial-context constraints summarized above to refine the final confidence levels. This two-stage workflow combines the strengths of data-driven detection with deterministic spatial reasoning, improving robustness and reliability in structured industrial scenes.

Comparative Baselines and Proposal

To isolate and quantify the contribution of the proposed spatial-context refinement module, we adopted YOLOv8 as the sole baseline architecture for comparison. This design choice allows us to directly attribute performance changes to the post-processing stage, without confounding effects from variations in detector architecture or training dynamics. YOLOv8 was trained using the 70/20/10 train/validation/test split defined in [Section 3.1](#), with ImageNet-pretrained weights, a unified input resolution of 800 px on the long side, and minimal, consistent hyperparameter tuning to ensure reproducibility.

In addition to the standard test partition, we constructed a dedicated evaluation set of 40 images specifically designed to stress-test contextual reasoning. This set combines real photographs with AI-generated samples that emulate challenging operational conditions, such as partial occlusion, clutter, visually similar distractors, and out-of-context placements while preserving the spatial semantics of substation layouts. Both the baseline YOLOv8 outputs and the proposed spatial-context filter were evaluated on this extended set to assess robustness under realistic failure-mode scenarios.

We report standard object detection metrics, including per-class Precision, Recall, F1-Score, Average Precision (AP), mean Average Precision at IoU 0.5 (mAP@0.5), mean Average Precision following the COCO protocol (mAP@ [0.5:0.95]), and inference speed in frames per second (FPS). The proposed post-processing module is applied exclusively to YOLOv8 detections; all reported improvements reflect the incremental gain achieved by integrating spatial reasoning on top of an already strong appearance-based detector. This evaluation protocol ensures a clear, detector-agnostic assessment of how contextual constraints enhance detection reliability in structured industrial environments.

3.1 Dataset

The dataset is primarily used to perform standard object-detection evaluation on a held-out test set using conventional metrics such as precision, recall, and F1-score. In addition to this standard benchmarking protocol, a subset of challenging images is used for diagnostic analysis. This targeted evaluation is not intended to replace full-dataset testing, but rather to analyze typical failure modes of appearance-based detectors and assess the specific contribution of the proposed spatial post-processing module.

We constructed a dataset comprising 714 images to support development and evaluation in outdoor electrical substation scenarios. The dataset includes a total of 5173 annotated object instances, with an average of 7.25 annotations per image. Four operationally relevant object classes are considered: connections, bodies, radiators, and tanks.

The dataset combines real-world photographs captured during actual construction projects using fixed cameras, along with stock and AI-generated images depicting substations under controlled conditions, such as clean white backgrounds. This deliberate combination of authentic construction environments and visually isolated object representations was designed to balance real-world complexity—such as varying illumination, occlusions, and background clutter—with simplified scenes that allow the model to focus on the intrinsic visual characteristics of substation components rather than spurious contextual cues.

Our data collection followed a two-phase rationale. First, real images were gathered to capture the typical visual and geometric regularities of substation components under stable viewpoints. Subsequently, AI-generated images were introduced to expand coverage of rare or safety-critical edge cases that were under-represented in the real corpus, including severe occlusion, ambiguous backgrounds, and visually similar distractors. This mixed-origin strategy prioritizes scenario coverage and failure-mode stress testing over fixed source ratios; for this reason, no single percentage split between real and synthetic sources is reported. Instead, methodological rigor is ensured through stratified data splits that preserve class proportions and prevent cross-split duplicates and near-duplicates, particularly among AI-generated variants.

Due to strict security and privacy regulations governing construction sites, photographs of real construction environments cannot be publicly released, and the complete dataset cannot be made openly available. However, to support reproducibility and downstream use, the trained detection model will be released via GitHub.

All images were resized to 800×800 pixels while preserving their original aspect ratio to avoid geometric distortion. Annotations were created using the LabelImg tool and underwent double independent review to ensure consistency and accuracy. Annotation guidelines specified class boundaries, handling of partial containment (e.g., radiators partially embedded within bodies), and expected spatial priors relevant to substation layouts.

The dataset was split using a stratified procedure that preserves class proportions, with 70% of images used for training (500 images), 20% for validation (142 images), and 10% for testing (72 images). Unless otherwise stated, all quantitative results reported in the manuscript are computed on the test partition. In addition, a strictly held-out evaluation set of 43 images, not used during training or validation, was employed for final model assessment. This set includes 16 real images and 27 AI-generated images, ensuring balanced coverage of real and synthetic conditions.

Table 1 summarizes the class distribution across the entire dataset. Connections represent the majority class, reflecting their frequent occurrence in substation layouts, while tanks and radiators appear less frequently, introducing moderate class imbalance.

Table 1: Class distribution and annotation summary.

Class	Count	Percentage	Annotation Criteria
Connection	3030	58.7%	Must be visibly attached to the main substation body
Body	7770	15%	May include partially contained radiators or tanks
Radiator	766	14.8%	Fully visible and unoccluded
Tank	594	11.5%	Must be in the expected top region relative to the main body
Total	5173	–	–

To further characterize the dataset, Table 2 reports the mean relative size statistics for each class, including width, height, area, and aspect ratio, expressed as percentages of the image dimensions. These statistics highlight the significant scale variation across object types and underscore the challenge of detecting small or elongated components such as connections.

Table 2: Mean relative object size statistics.

Class	Width	Height	Area	Aspect Ratio
Connection	4.9%	13.6%	0.87%	0.39
Body	54.3%	50.2%	33.03%	1.10
Radiator	25.0%	24.9%	8.43%	1.02
Tank	22.3%	14.3%	3.93%	1.68

Together, these characteristics provide a detailed and transparent description of the dataset composition, annotation protocol, and inherent challenges, supporting reproducibility and enabling informed interpretation of the experimental results.

- Image 5* (reproduced in Fig. 4): The model incorrectly classifies a helmet as an expansion tank, assigning it a high confidence score. The rule-based system eliminates this detection by verifying that the object is not located in the expected area and does not have the appropriate size relative to the electrical substations.

- Fig. 4 shows image 43: The model detects tanks, electrical substations, and connectors outside the structural body. The rule-based system discards all these incorrect detections by validating their spatial relationship and containment within the expected layout.
- In Image 6 (see Fig. 5), the YOLOv8 model fails to detect an expansion tank that is correctly positioned at the top of the electrical substations. The rule-based system successfully recovers this detection by validating its location and overlap with the substation. Image 16 (see Fig. 6) presents the opposite case, where electrical substations are detected outside the actual structural body. The rule-based system eliminates these detections by confirming the absence of valid spatial relationships. Fig. 5 illustrates a case where the YOLOv8 model detects tanks and connectors on the lateral structures surrounding the transformer. Although these components visually resemble elements of the substation, they are located outside the structural body of the equipment. The rule-based system removes these detections by verifying spatial containment, ensuring that tanks and related components must lie within the expected bounding region of the electrical substation. This behavior is consistent with the scenario described in Image 16 (see Fig. 6), where detections outside the structural body are eliminated by validating spatial relationships.
- In Image 18 (referenced in Fig. 7), the YOLOv8 model detects a radiator that is not linked to any structural element. The rule-based system discards this detection by applying containment and overlap criteria, preventing the validation of components that are not properly integrated into the substation layout.
- Image 25 (reproduced in Fig. 7) shows the opposite situation: the model fails to detect a partially occluded radiator. The rule-based system recovers this detection by validating its overlap with nearby structures and substation bodies. Additionally, it corrects a false detection of an expansion tank. Image 18 (referenced in Fig. 7) also illustrates a similar case in reverse, where a radiator is incorrectly detected, and the spatial-context constraints eliminate it.
- In Image 29 (see Fig. 8), the system displays one of the most evident errors: the YOLOv8 model detects multiple elements—such as tanks, electrical substations, and connectors—in empty areas unrelated to the actual infrastructure. The rule-based system eliminates all these detections by applying criteria based on position, containment, and relative size. The result is a clear improvement in both precision and spatial coherence.
- In Image 40 (reproduced in Fig. 8), the YOLOv8 model detects an electrical substation and a connector outside the actual structural body. The rule-based system corrects these detections by verifying the absence of overlap and logical positioning.
- Image 41 (see Fig. 8) presents a similar error, where electrical substations are detected outside the main body. The rule-based system discards this detection for not meeting the defined spatial criteria.
- In Image 41 (see Fig. 8), the YOLOv8 model detects tanks, electrical substations, and connectors in areas outside the structural body—locations where these elements should not appear. The rule-based system eliminates all these incorrect detections by applying criteria based on position, containment, and relative size. This reinforces the system's ability to filter out components that do not align with the physical logic of a substation.

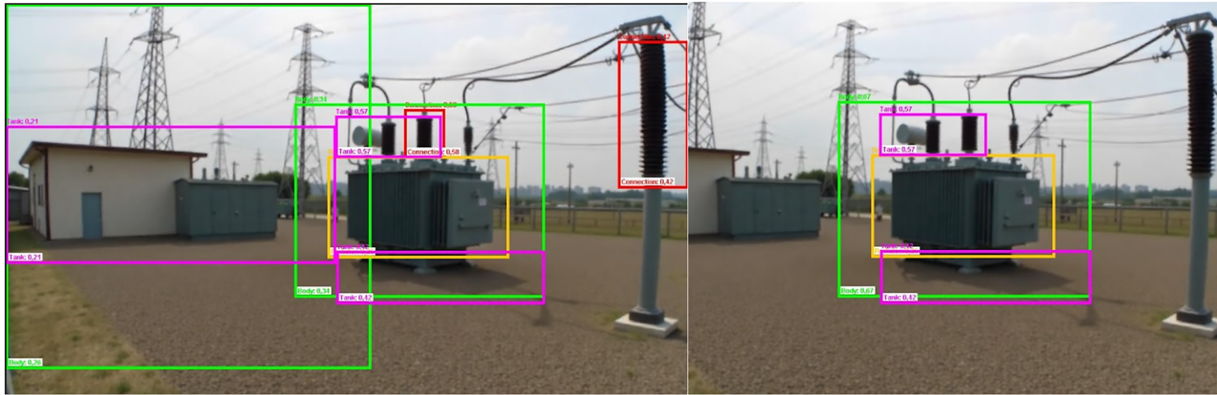


Figure 4: Image 43 YOLOv8 vs. rule-based system.

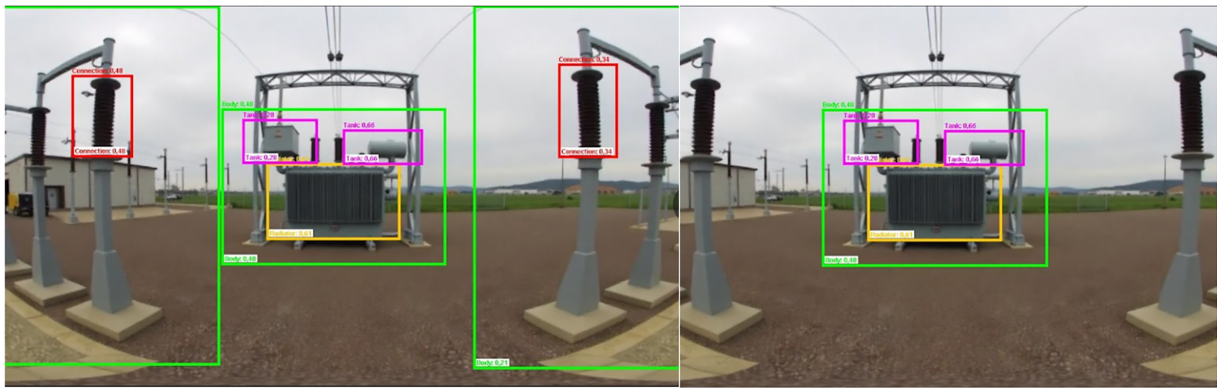


Figure 5: Image 7. YOLOv8 vs. rule-based system.

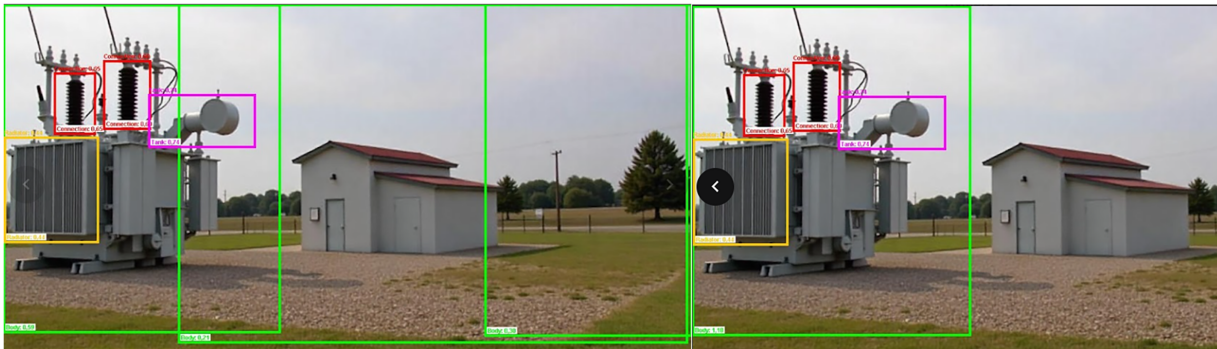


Figure 6: Image 16. YOLOv8 vs. rule-based system.

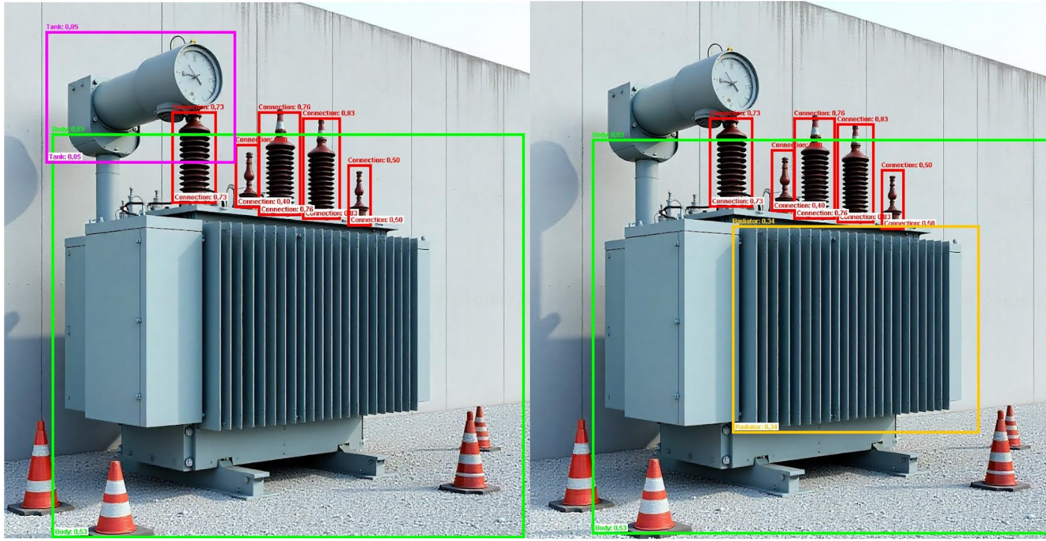


Figure 7: YOLOv8 vs. rule-based system.

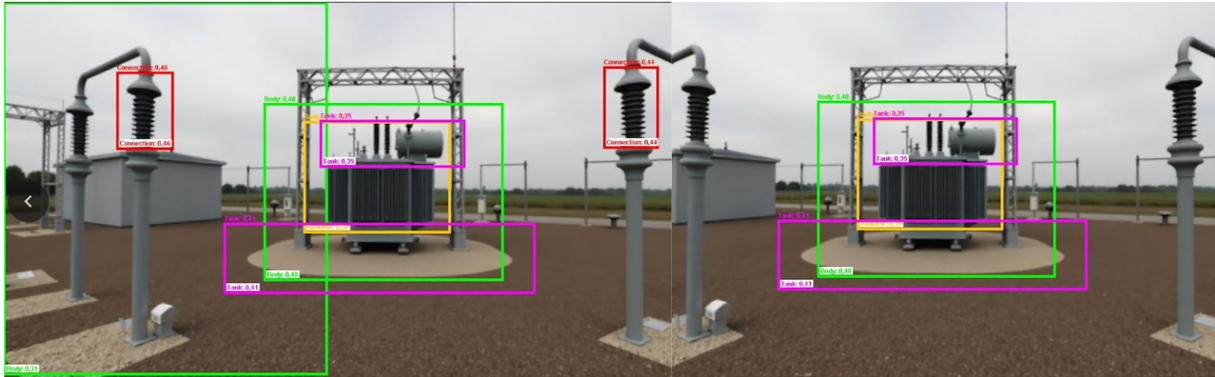


Figure 8: YOLOv8 vs. rules-based system.

3.2 Algorithms

This section details the algorithmic implementation of the spatial-context refinement rules introduced previously. Each algorithm operationalizes a specific type of domain constraint, such as relative positions (e.g., connectors near the top of the structure), absolute locations (e.g., tanks in upper regions), center-to-center distances with tolerance margins, overlap percentages between components (e.g., radiators partially contained within bodies), proportional size relationships, and confidence-based conditional logic. These rules are applied as a post-processing layer to the baseline detector outputs, adjusting confidence scores through multiplicative factors that either reinforce spatially coherent detections or suppress contextually implausible ones. The algorithmic rules are applied sequentially in the exact order presented in this section, ensuring consistent and deterministic refinement of detection outputs. The following figures and pseudocode illustrate the core algorithms, demonstrating how structural priors are systematically encoded to enhance detection reliability without modifying the underlying neural architecture.

Algorithm 1 illustrates the algorithm used to adjust the detection probability of a Radiator object when its initial confidence is low (<0.2). The algorithm iterates through all detected objects in the image and checks whether the Radiator significantly overlaps ($\geq 80\%$) with any Body object that has a detection confidence greater than 0.4. If such an overlap is found, the Radiator is considered to be correctly positioned within a valid structure, and its confidence score is doubled. This mechanism allows the system to recover weak detections that exhibit strong spatial coherence with other relevant components.

Algorithm 1: Penalty adjustment for a Connection object near the center of a low-confidence Body

Input: An object with class *Connection*, its original probability *adjustedProbability*, coordinates *left*, *top*, *right*, *bottom*, list of detected objects *allItems*, and scale factors *scaleX*, *scaleY*

Output: Adjusted probability of the *Connection* object

```

1 if the object's class is Connection then
2   foreach other in allItems do
3     if other is of class Body and its probability  $< 0.5$  then
4       bodyWidth  $\leftarrow$  width of other  $\times$  scaleX;
5       tolerance  $\leftarrow$  bodyWidth  $\times 0.35$ ;
6       centerDistance  $\leftarrow$  distance between the center of the Connection object and the center of Body;
7       if centerDistance  $\leq$  tolerance then
8         adjustedProbability  $\leftarrow$  adjustedProbability  $\times 0.3$ ;
9         break;
```

Algorithm 2 presents an algorithm that adjusts the detection probability of a Radiator object based on its overlap with Body objects. Regardless of its initial confidence score, the algorithm checks whether the Radiator significantly overlaps ($\geq 80\%$) with any Body object whose confidence is greater than 0.2. If no valid overlap is found, the Radiator's confidence is penalized by multiplying it by 0.3. This approach helps reduce false positives by discarding radiators that are not properly integrated into the substation structure.

Algorithm 2: Penalize Radiator if it does not significantly overlap with a confident Body

Input: An object of class *Radiator* with probability *adjustedProbability*, coordinates *left*, *top*, *right*, *bottom*, a list of all detected objects *allItems*, and scaling factors *scaleX*, *scaleY*

Output: Possibly modified *adjustedProbability*

```

1 if the object's class is Radiator then
2   hasValidOverlap  $\leftarrow$  false;
3   foreach other in allItems do
4     if other.className is Body and other.probability  $> 0.2$  then
5       overlap  $\leftarrow$  overlap percentage of current Radiator in other;
6       if overlap  $\geq 0.8$  then
7         hasValidOverlap  $\leftarrow$  true;
8         break;
9   if hasValidOverlap is false then
10    adjustedProbability  $\leftarrow$  adjustedProbability  $\times 0.3$ ;
```

Algorithm 3 presents the algorithm that adjusts the detection probability of a Connection object when its initial confidence is low (<0.2). The algorithm iterates through all detected objects in the image and checks whether the Connection significantly overlaps ($>80\%$) with any Body object that has a high confidence

score (>0.6). If such an overlap is found, the Connection is considered to be correctly positioned within a valid structure, and its confidence is increased by multiplying it by 1.5. This adjustment helps recover weak detections that, due to their spatial context, are likely to be correct.

Algorithm 3: Boost Connection probability if it significantly overlaps with a confident Body

Input: An object of class Connection with probability `adjustedProbability` < 0.2 , coordinates `left`, `top`, `right`, `bottom`, a list of all detected objects `allItems`, and scaling factors `scaleX`, `scaleY`

Output: Possibly updated `adjustedProbability`

```

1 if object's class is Connection and adjustedProbability  $< 0.2$  then
2   boosted  $\leftarrow$  false;
3   foreach other in allItems do
4     if other.className is Body and other.probability  $> 0.6$  then
5       overlap  $\leftarrow$  overlap percentage of current Connection in other;
6       if overlap  $\geq 0.8$  then
7         adjustedProbability  $\leftarrow$  adjustedProbability  $\times 1.5$ ;
8         boosted  $\leftarrow$  true;
9       break;
```

Algorithm 4 presents an algorithm that adjusts the detection probability of a Body object when its initial confidence is greater than 0.1. The goal is to reinforce the detection if the Body significantly overlaps ($\geq 80\%$) with at least one Tank or Radiator object, and also with a Connection object, provided these components have a minimum confidence of 0.3. If both conditions are met, the Body is considered to be correctly structured, and its confidence score is multiplied by 2. This adjustment helps consolidate detections that exhibit spatial coherence with other components, improving reliability in environments where a logical arrangement of elements is essential.

Algorithm 4: Boost Body probability if it overlaps with both a Tank or Radiator and a Connection

Input: An object of class Body with `adjustedProbability` > 0.1 , coordinates `left`, `top`, `right`, `bottom`, list of all detected objects `allItems`, and scaling factors `scaleX`, `scaleY`

Output: Possibly updated `adjustedProbability`

```

1 if object class is Body and adjustedProbability  $> 0.1$  then
2   hasTankOrRadiator  $\leftarrow$  false;
3   hasConnection  $\leftarrow$  false;
4   foreach other in allItems do
5     if other.className is Tank or Radiator and other.probability  $> 0.5$  then
6       overlap  $\leftarrow$  overlap percentage of current Body in other;
7       if overlap  $\geq 0.5$  then
8         hasTankOrRadiator  $\leftarrow$  true;
9     if other.className is Connection and other.probability  $> 0.5$  then
10      overlap  $\leftarrow$  overlap percentage of current Body in other;
11      if overlap  $\geq 0.5$  then
12        hasConnection  $\leftarrow$  true;
13   if hasTankOrRadiator and hasConnection then
14     adjustedProbability  $\leftarrow$   $\min(\text{adjustedProbability} \times 1.5, 1.0)$ ;
```

Algorithm 5 presents the algorithm that adjusts the detection probability of a Connection object based on its overlap with a Body object. The algorithm first checks whether the Connection significantly overlaps ($\geq 60\%$) with any Body object whose confidence is greater than 0.1. If no such overlap is found, the Connection is considered to be incorrectly positioned or isolated, and its confidence score is penalized by multiplying it by 0.3. This adjustment helps reduce false positives and ensures that connections are only validated when they are properly integrated into a recognized structure.

Algorithm 5: Reduce the probability of Connection if it does not overlap sufficiently with a Body

Input: An object of class Connection and coordinates left, top, right, bottom, list of all detected objects allItems, and scaling factors scaleX, scaleY

Output: Possibly updated adjustedProbability

```

1 hasSufficientOverlap ← false;
2 foreach other in allItems do
3   if className is Connection and other.className is Body and other.probability > 0.1 then
4     overlapArea ← overlap percentage of Connection in Body;
5     connectionArea ← (right-left) × (bottom-top);
6     if connectionArea > 0 and overlapArea ≥ 0.6 then
7       hasSufficientOverlap ← true;
8       break;
9 if className is Connection and hasSufficientOverlap is false then
10  adjustedProbability ← adjustedProbability × 0.3;

```

Algorithm 6 presents the procedure used to adjust the detection probability of a Tank object based on its spatial overlap with a Body object. First, the algorithm verifies whether the Tank overlaps sufficiently ($\geq 60\%$) with any Body whose detection confidence is equal to or greater than 0.29. If no significant overlap is found, the Tank is considered to be incorrectly positioned or not associated with a valid substation structure. Consequently, its detection confidence is penalized by multiplying it by 0.2. This refinement reduces false positives by ensuring that Tank detections are only retained when they are spatially consistent with a valid Body.

Algorithm 6: Penalize the probability of Tank if it does not overlap sufficiently with a valid Body

Input: An object of class Tank, coordinates left, top, right, bottom, list of all detected objects allItems, and scaling factors scaleX, scaleY

Output: Possibly updated adjustedProbability

```

1 tankInsideValidBody ← false;
2 foreach other in allItems do
3   if other.className is Body and other.probability > 0.29 then
4     overlap ← overlap percentage of Tank in Body;
5     if overlap ≥ 0.6 then
6       tankInsideValidBody ← true;
7       break;
8 if tankInsideValidBody is false then
9   adjustedProbability ← adjustedProbability × 0.2;

```

Algorithm 7 shows the algorithm that penalizes the detection probability of a Connection object when it significantly overlaps ($\geq 80\%$) with a Body object that has low confidence (< 0.2). If this condition is met, the Connection is considered potentially misclassified due to its association with an unreliable structure, and its confidence score is reduced by multiplying it by 0.3.

Additionally, the algorithm includes print statements to log the overlap percentage and the Body object's confidence, facilitating system behavior analysis and debugging. This adjustment improves model precision by preventing doubtful connections from being mistakenly validated.

Algorithm 7: Penalize the probability of Connection if it overlaps with a low-confidence Body

Input: An object of class Connection, coordinates left, top, right, bottom, list of all detected objects allItems, and scaling factors scaleX, scaleY

Output: Possibly updated adjustedProbability

```

1  if className is Connection and adjustedProbability < 0.65 then
2    foreach other in allItems do
3      if other.className is Body and other.probability < 0.2 then
4        overlap ← overlap percentage of Connection in Body;
5        print("overlap:" + overlap);
6        print("Body Probability:" + other.probability);
7        if overlap ≥ 0.8 then
8          adjustedProbability ← adjustedProbability × 0.3;
9          print("Rule 9 applied: Connection penalized by overlap with
              low-confidence Body") break;
10       end
11     end
12   end
13 end

```

4 Evaluation

To evaluate the impact of our proposal, we used a dataset of 714 images of electrical substations obtained from public databases and real-world photographs. A YOLOv8 model was trained under different configurations, and a confidence threshold of 0.6 was selected. Subsequently, we identified 40 images in which the neural network produced incomplete or erroneous detections. Two evaluation approaches were applied:

- MIO-YOLO: performs detections directly using the neural network.
- MIO-Rules: adjusts those detections by applying the spatial and structural criteria introduced in this study.

Both models were evaluated across four types of components: connectors, tanks, bodies, and radiators.

YOLOv8 demonstrates strong performance in detecting components within electrical substations, successfully identifying a large number of objects with high precision and recall across all evaluated classes. As shown in [Table 3](#):

- For connectors, the model correctly identifies most elements, achieving a precision of 95.8% and a recall of 88.3%, although some false positives and omissions slightly affect overall balance.
- For expansion tanks, the model reaches a precision of 90.9% and a recall of 85.7%, indicating that while it performs well in most cases, it tends to confuse tanks with other components or miss them when they are partially occluded.

- For bodies, the model shows its best performance, with a perfect precision of 100% and a recall of 94.9%, reflecting high reliability in detecting these structural elements.
- For radiators, the results are also strong, with a precision of 97.6% and a recall of 88.9%, although some errors are observed in recovering elements that were not initially detected.

Table 3: MIO-YOLO baseline.

Class	TP	FP	FN	Precision	Recall	F1 Score
Connections	154	8	24	0.958	0.883	0.919
Tanks	30	4	12	0.909	0.857	0.882
Bodies	37	3	5	1	0.949	0.974
Radiators	40	1	5	0.976	0.889	0.93

MIO-YOLO proves to be an effective model, but with room for improvement in spatial interpretation, particularly in scenarios where the logical arrangement of components is crucial for accurate detection.

The MIO-Rules system builds upon YOLO detections and refines them by applying spatial and structural rules. These rules validate or discard objects based on their location, size, and relationship to the electrical substations.

The rule-based system improves the overall precision of the model and reduces YOLO's most common errors. Although it slightly sacrifices recall in some cases, the result is a more reliable model that aligns better with the spatial logic of substation environments. As shown in [Table 4](#):

- **Connections.** The proposed spatial rules significantly reduced false positives (from 8 to 3) while maintaining a high number of true positives, resulting in perfect precision (1.0). This indicates that the contextual constraints, such as requiring connectors to overlap with the upper regions of the electrical substations. The slight reduction in recall suggests that a small number of valid connectors in atypical positions were conservatively suppressed, but the net gain in F1-Score ($0.919 \rightarrow 0.932$) confirms an overall improvement in detection reliability.
- **Tanks.** The application of spatial rules eliminated all false positives (FP: $4 \rightarrow 0$) and improved precision ($0.909 \rightarrow 0.931$), demonstrating that the constraint requiring expansion tanks to be located near the top of the structure successfully rejected out-of-context detections. The modest decrease in recall reflects the conservative nature of the rule-based filter, which prioritizes precision in safety-critical scenarios. The F1-Score remains competitive ($0.882 \rightarrow 0.871$), indicating a favorable trade-off for operational deployment.
- **Bodies.** As the largest and most visually distinctive component, the main structure already achieved perfect precision in the baseline. The proposed rules further reduced false negatives ($5 \rightarrow 2$) while maintaining zero false positives, slightly improving recall and preserving the high F1-Score ($0.974 \rightarrow 0.964$). This suggests that spatial consistency checks reinforce confidence in structurally central elements without introducing unnecessary filtering.
- **Radiators.** This class benefited most substantially from the spatial refinement: false negatives dropped from 5 to 1, and the F1-Score increased from 0.93 to 0.967. The rule requiring radiators to be partially contained within the main structure allowed the system to recover valid detections that were initially assigned to low confidence due to occlusion or lighting, while maintaining high precision. This result highlights the strength of the proposed approach in handling components with strong spatial priors but challenging visual conditions.

Table 4: Proposal MIO-rules.

Class	TP	FP	FN	Precision	Recall	F1 Score
Connections	152	3	19	1	0.873	0.932
Tanks	27	0	10	0.931	0.818	0.871
Bodies	27	0	2	1	0.931	0.964
Radiators	29	1	1	0.967	0.967	0.967

Across all classes, the integration of domain-specific spatial rules consistently reduced false positives and, in most cases, improved or maintained recall. The results confirm that contextual reasoning complements appearance-based detection by suppressing implausible detections and reinforcing contextually coherent ones, leading to more robust and trustworthy performance in structured industrial environments.

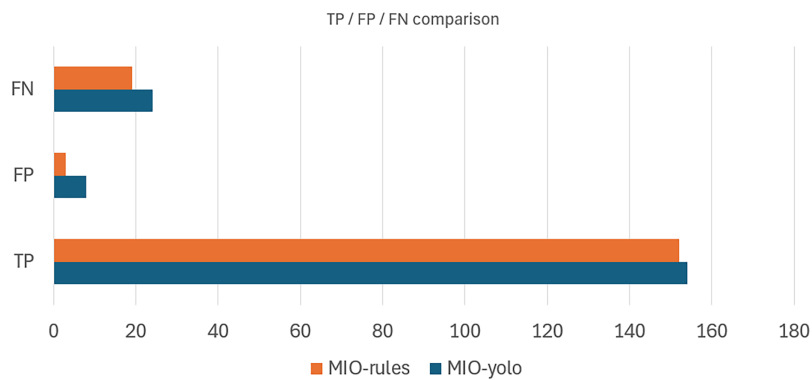
To evaluate the impact of the proposed rules, we present a table with the total values of True Positives (TP), False Positives (FP), Precision, Recall, and F1 Score for each class in the original dataset (resultado020).

Figs. 9–11 show the comparison between MIO-YOLO baseline y MIO-rules the proposal which includes a post processing filter which improves the quality of the objects detections. As shown in Tables 5 and 6:

- Precision: Increases from 95.1% to 98.1%, indicating a reduction in false positives. The system becomes more reliable. For example, in Image 20, YOLO incorrectly detected a connector and a body with high confidence. The spatial-context constraints successfully eliminated these errors.
- Recall: Improves from 86.5% to 88.9%, showing the model's enhanced ability to detect relevant elements and reduce false negatives. In Image 41, YOLO detected an electrical substations outside the main body. The spatial-context constraints corrected the location and validated it properly.
- F1 Score: Rises from 0.905 to 0.933, reflecting a better balance between precision and recall. In Image 37, both metrics improved thanks to rules that validated the structural logic of the substation.

These results suggest that the new rules help adjust incomplete or erroneous detections, especially in images where the neural network alone struggles to interpret the spatial arrangement of components.

Moreover, by applying spatial criteria, the system filters out spurious detections that do not conform to the physical logic of a substation. Therefore, the rule-based model is considered better suited to variations in real-world images, making it more useful in operational environments.

**Figure 9:** Comparative analysis of true positives, false positives, and false negatives: original model vs. proposal.

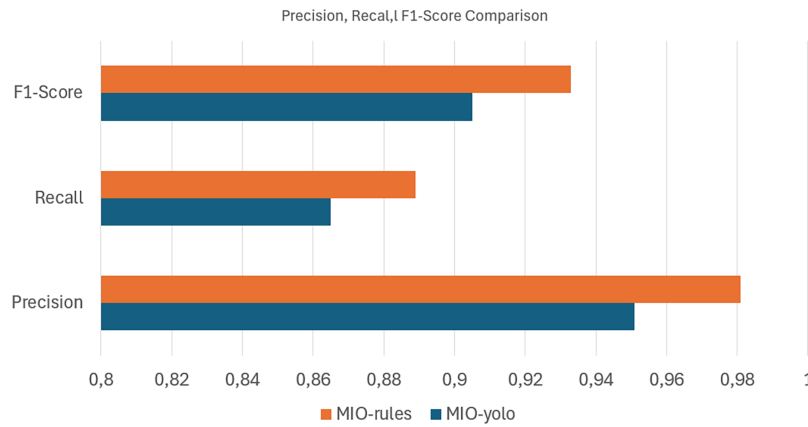


Figure 10: Comparison of precision, recall, and F1-score between the original and proposed models, showing improved performance across all metrics.

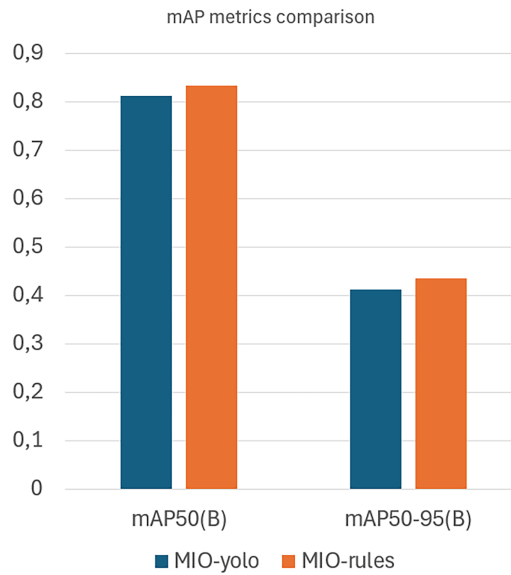


Figure 11: Comparison of mAP@50 and mAP@50–95 metrics between the original and proposed models, demonstrating improved detection accuracy across IoU thresholds.

Table 5: Overall performance comparison between MIO-YOLO and MIO-rules.

	TP	FP	FN	Precision	Recall	F1-Score	mAP50(B)	mAP50–95(B)
MIO-yolo	154	8	24	0.951	0.865	0.905	0.813	0.412

Table 6: Overall performance comparison of the proposed MIO-rules method.

	TP	FP	FN	Precision	Recall	F1-Score	mAP50(B)	mAP50–95(B)
MIO-rules	152	3	19	0.981	0.889	0.933	0.834	0.436

The proposal improvements in mean Average Precision across both evaluation thresholds. Specifically, mAP@50 increases from 0.813 to 0.834 (+2.6%), while the more stringent mAP@50–95 averaged across IoU thresholds from 0.5 to 0.95 rises from 0.412 to 0.436 (+5.8%). These gains, though modest in absolute terms, are significant given the strictness of the COCO-style metric and reflect the system's enhanced ability to produce spatially coherent, high-confidence detections. The improvements align with the observed reduction in false positives and false negatives, confirming that integrating domain-specific spatial rules not only refines per-class performance but also elevates overall detection quality across varying localization tolerances.

After incorporating the proposed rules, a second table is presented with the same metrics for the adjusted dataset (resultadoReglas).

5 Discussion

While the proposed rule-based refinement results in a modest recall reduction for certain classes, it consistently yields a substantial increase in precision across the dataset. This precision–recall trade-off is particularly relevant in industrial inspection scenarios, where false positives may lead to unnecessary interventions, incorrect maintenance planning, or increased exposure of personnel to hazardous environments. In such contexts, higher precision is often preferable to marginal gains in recall, as it improves trust in automated detections and reduces the operational cost associated with erroneous alarms.

The post-processing layer explicitly suppresses detections that violate expected spatial and structural constraints, which explains the reduction of out-of-context false positives commonly observed in appearance-only models. At the same time, the system can recover visually weak but spatially coherent detections—especially when partial occlusion or illumination changes degrade appearance signals—mitigating recall losses in challenging scenes. This balance ultimately favors reliability and consistency, key requirements for safety-critical workflows in electrical substations.

Class-wise behavior reflects the underlying spatial priors of each component. Structural elements with strong containment and proportion relationships to the electrical substation body (e.g., radiators) benefit more from spatial rules, improving both precision and recall in several cases. In contrast, components with greater positional variability (e.g., connectors, expansion tanks) may exhibit slight recall fluctuations under stricter constraints. Nevertheless, the overall precision gains for these classes outweigh limited recall losses, leading to improved F1 scores and more spatially coherent detections.

Beyond metric improvements, the integration of explicit spatial reasoning offers operational advantages. The qualitative corrections observed—such as eliminating detections in implausible regions or reinforcing occluded but contextually valid components—can reduce the cognitive burden on human supervisors. Rather than adjudicating ambiguous outputs, operators receive detections pre-filtered for structural consistency, which aligns the system more closely with day-to-day inspection conditions and supports decision-making in maintenance scheduling, risk assessment, and preventive interventions.

Conceptually, the results suggest that a portion of YOLOv8's misclassifications in electrical substations originates not from feature extraction limitations per se, but from the absence of explicit structural expectations about how components relate in space. The rule-based layer complements the appearance model by reinstating such expectations as auditable, interpretable constraints that can be revised or extended by domain experts. This property is uncommon in purely end-to-end approaches and facilitates alignment with industry safety standards.

Finally, although the method's structure appears compatible with other highly organized environments, any broader applicability beyond electrical substations should be regarded as prospective. Verifying transfer to domains such as industrial assembly, warehouse logistics, or medical imaging would require

dedicated datasets, domain-specific rule sets, and rigorous evaluation protocols. As outlined in Future Work, these extensions remain an avenue for subsequent studies rather than evidence established by the present experiments.

6 Conclusions

This work demonstrates that integrating spatial reasoning into object detection significantly enhances reliability in structured industrial environments such as electrical substations. By combining YOLOv8 with a set of explicit rule-based adjustments grounded in positional, containment and proportional constraints, the proposed hybrid system effectively addresses common detection failures that arise when visual cues alone are insufficient. Our experiments show clear improvements across all metrics: precision increased from 95.1% to 98.1%, recall rose from 86.5% to 88.9%, and the overall F1-score improved from 0.905 to 0.933. Qualitative analyses further confirm that the rule-based module reduces false positives in out-of-context regions and recovers missed detections when objects exhibit strong spatial coherence with the electrical substations. These results demonstrate that spatial logic provides a valuable complement to convolutional neural networks, particularly in environments where components follow predictable physical arrangements.

Summary of Achievements

- We introduced a hybrid detection framework that integrates CNN-based detection with explicit spatial rules.
- We developed a novel, annotated dataset of substation components and evaluated the system on challenging real-world scenarios.
- We demonstrated that the rule-based refinement improves both quantitative performance (precision, recall, F1) and qualitative spatial consistency.
- We showed that the system is especially effective in mitigating two key failure modes: visually plausible false positives and missed detections caused by occlusion or structural ambiguity.

6.1 Theoretical Contributions. We demonstrate that integrating explicit, domain-specific spatial rules based on relative positions and sizes as a confidence-aware post-processing stage effectively complements CNN-based object detection in structured environments. To validate this, our ablation study compares the performance of the ‘Base Model’ (standard detection) against the ‘Full System’ (Base + Formalized Rules), showing how spatial reasoning compensates for the limitations of purely appearance-driven models without modifying the underlying neural architecture.

6.2 Practical Contributions. We introduce a deployable hybrid detection framework for electrical substations that significantly improves detection precision and overall reliability. By reducing out-of-context false positives, the method directly addresses key operational requirements of safety-critical inspection workflows and can be integrated as a lightweight post-processing module within existing industrial vision systems.

6.3 Broader Implications. Although our experiments were conducted exclusively on electrical substations, the proposed rule-based framework suggests potential applicability to other structured environments where spatial logic governs object arrangement, such as autonomous driving, industrial assembly lines, warehouse logistics, or medical imaging. However, these broader applications should be considered prospective rather than experimentally validated. The methodology provides a conceptual foundation for generalization, but its effectiveness beyond substations remains to be empirically tested in future work.

7 Future Work

Building on these findings, several avenues merit further exploration. First, automating the discovery and optimization of spatial rules could increase scalability and reduce reliance on expert-defined constraints. Given that the spatial rules used in this work are explicit, interpretable, and text-describable, transitioning toward automated rule induction naturally aligns with the capabilities of modern large language models. One promising direction involves leveraging emerging Large Language Models (LLMs) to analyze detection outcomes, both successful and failed, and infer plausible spatial, size, and overlap constraints directly from structured feedback rather than manual engineering. Additionally, specialized AI agents could be trained to recognize domain-specific relational patterns and propose or refine rules dynamically, reducing the need for extensive human intervention.

Beyond automation, expanding the dataset to include additional substation types, voltage levels, and environmental conditions would help assess the robustness of the proposed approach within the broader substation domain. Deploying the framework in real inspection workflows could further clarify its practical impact on maintenance efficiency, safety compliance, and operator trust.

Finally, while the conceptual structure of this framework suggests potential applicability to other highly organized environments, such as industrial assembly or medical imaging, these directions remain prospective. Future studies would be required to determine whether the spatial-constraint logic demonstrated here transfers effectively to domains beyond electrical substations, and such validation would necessitate domain-specific datasets, constraints, and evaluation protocols.

The proposed hybrid detection framework, which integrates YOLOv8 with explicit rule-based spatial reasoning, demonstrates significant improvements in both quantitative metrics and qualitative consistency in structured industrial environments such as electrical substations. While the rule-based refinement produces modest recall reductions for certain classes, it consistently increases precision, reducing false positives in out-of-context regions and recovering detections under partial occlusion or illumination changes. Class-wise analysis shows that structural components with strong spatial priors benefit most, while elements with higher positional variability may experience slight recall fluctuations, though overall F1 scores improve. The approach provides operational advantages by pre-filtering ambiguous outputs, reducing cognitive load for human inspectors, and aligning system outputs with day-to-day inspection conditions. Experimental results indicate increases from 95.1% to 98.1% in precision, 86.5% to 88.9% in recall, and from 0.905 to 0.933 in F1-score. Conceptually, the method demonstrates that misclassifications in YOLOv8 often arise from missing explicit structural expectations rather than feature extraction limitations, and that incorporating interpretable spatial rules as a post-processing stage complements CNN-based detection without modifying the underlying network. While validated on electrical substations, the framework suggests potential applicability to other structured domains such as industrial assembly, warehouse logistics, autonomous driving, or medical imaging, though broader adoption requires domain-specific datasets and rules. Future work includes automating rule discovery and optimization, leveraging large language models to infer spatial constraints, expanding datasets to cover diverse substation types and conditions, and assessing real-world deployment to evaluate operational impact, safety compliance, and user trust.

Acknowledgement: Some images used in this work were obtained from Freepik and used to generate additional scenario images for the project, increasing the diversity of operational conditions in the dataset.

Funding Statement: The authors received financial support for the research, authorship, and/or publication of this article: “SITETRAK: investigación en nuevas tecnologías para el seguimiento inteligente de proyectos EPC” Projects Program (Actividades cofinanciadas por la Agencia SEKUENS y el Plan de Ciencia, Tecnología e Innovación (PCTI) Asturias, Spain) IDE/2023/000202.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, formal analysis, investigation, data curation, writing—review and editing, and supervision, Jordán Pascual Espada; methodology, validation, resources, writing—original draft preparation, and visualization, Lucía Alonso Virgós; resources, Dr. Juan Luis Carús and Miguel Ángel Fernández Fernández. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The complete dataset used in this study is not publicly available due to security and privacy restrictions associated with real-world electrical substation construction environments. The dataset contains images collected from ongoing industrial projects and is subject to confidentiality constraints imposed by the relevant stakeholders. To support reproducibility, the trained detection model and relevant implementation details will be made available via GitHub.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hussain M. YOLOv1 to v8: unveiling each variant—a comprehensive review of YOLO. *IEEE Access*. 2024;12:42816–33. doi:10.1109/ACCESS.2024.3378568.
2. Kang S, Hu Z, Liu L, Zhang K, Cao Z. Object detection YOLO algorithms and their industrial applications: overview and comparative analysis. *Electronics*. 2025;14(6):1104. doi:10.3390/electronics14061104.
3. Wang X, Liu S, Du P, Liang H, Xia J, Li Y. Object-based change detection in urban areas from high spatial resolution images based on multiple features and ensemble learning. *Remote Sens*. 2018;10(2):276. doi:10.3390/rs10020276.
4. Trigka M, Dritsas E. A comprehensive survey of machine learning techniques and models for object detection. *Sensors*. 2025;25(1):214. doi:10.3390/s25010214.
5. Min L, Fan Z, Lv Q, Reda M, Shen L, Wang B. YOLO-DCTI: small object detection in remote sensing base on contextual transformer enhancement. *Remote Sens*. 2023;15(16):3970. doi:10.3390/rs15163970.
6. Zang Y, Li W, Han J, Zhou K, Loy CC. Contextual object detection with multimodal large language models. *Int J Comput Vis*. 2025;133(2):825–43. doi:10.1007/s11263-024-02214-4.
7. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016 Jun 27–30; Las Vegas, NV, USA. p. 779–88.
8. Reddy S, Pillay N, Singh N. Comparative evaluation of convolutional neural network object detection algorithms for vehicle detection. *J Imaging*. 2024;10(7):162. doi:10.3390/jimaging10070162.
9. Gheorghe C, Duguleana M, Boboc RG, Postelnicu CC. Analyzing real-time object detection with YOLO algorithm in automotive applications: a review. *Comput Model Eng Sci*. 2024;141(3):1939–81. doi:10.32604/cmes.2024.054735.
10. More SS, Bansode R. FCN-YOLOS: an effective deep-learning model for real-time object detection. *J Field Robot*. 2025;42(8):4053–74. doi:10.1002/rob.70001.
11. Putluru AR, Pannuru VR, AshokKumar C, Kumar CJ. An optimized YOLO-based license plate recognition system with integrated privacy safeguards. In: *Proceedings of the 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*; 2025 Mar 10–12; Prawet, Thailand. p. 1053–60.
12. Leinonen M. Monocular 3D object detection and tracking in industrial settings [master's thesis]. Helsinki, Finland: University of Helsinki; 2021.
13. Tulbure AA, Tulbure AA, Dulf EH. A review on modern defect detection models using DCNNs—deep convolutional neural networks. *J Adv Res*. 2022;35(2):33–48. doi:10.1016/j.jare.2021.03.015.
14. Wang K, Liu M. YOLO-Anti: YOLO-based counterattack model for unseen congested object detection. *Pattern Recognit*. 2022;131(108350):108814. doi:10.1016/j.patcog.2022.108814.
15. Zhu Q, Ding Y, Jiang J, Yang SH. Anomaly detection using invariant rules in industrial control systems. *Control Eng Pract*. 2025;154(10):106164. doi:10.1016/j.conengprac.2024.106164.

16. Terven J, Córdova-Esparza DM. A comprehensive review of YOLO: from YOLOv1 to YOLOv8 and beyond. *Mach Learn Knowl Extr.* 2023;5(4):1680–716.
17. Méndez-Gurrola II. *Advances in artificial intelligence*. Mexico City, Mexico: Instituto Politécnico Nacional, Centro de Investigación en Computación; 2024.
18. Zhang G, Fan J, Chen L, Zhang Z, Lei Z, Zhang L. General geometry-aware weakly supervised 3D object detection. *arXiv:2407.13748*. 2024.
19. Tsirtsakis P, Zacharis G, Maraslidis GS, Fragulis GF. Deep learning for object recognition: a comprehensive review of models and algorithms. *Int J Cogn Comput Eng.* 2025;6(8):298–312. doi:10.1016/j.ijcce.2025.01.004.