Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.080624

ARTICLE

Robust Analog Gauge Reading via Virtual Point-Based Geometric Rectification and P2-YOLO-Pose

Jaekyung Lee[1,2], Youngjun Kim[2], Byungsung Ko[2], Taewon Kim[2], Jaeheon Park[2], Jiwon Lee[2] and Wonhee Kim[1,*]

[1]School of Energy Systems Engineering, Chung-Ang University, Seoul, Republic of Korea
[2]KEPCO Research Institute, Daejeon, Republic of Korea

*Corresponding Author: Wonhee Kim. Email: whkim79@cau.ac.kr

Received: 13 February 2026; Accepted: 30 March 2026; Published: 27 April 2026

ABSTRACT: Automated reading of analog gauges in industrial environments is essential for predictive maintenance and safety monitoring. However, conventional computer vision approaches encounter two fundamental bottlenecks: polar unwrapping techniques induce severe nonlinear scaling distortions under oblique viewing angles and axis-aligned bounding boxes (AABBs) are geometrically inefficient for encapsulating high-aspect-ratio rotating needles. To overcome these limitations, this paper proposes a novel end-to-end framework that innovatively redefines gauge reading as a structural pose estimation task. We model each gauge as a topological five-keypoint skeleton ($k_{start}, k_{mid}, k_{center}, k_{end}, k_{tip}$), and localize these landmarks using a customized P2-YOLO-Pose architecture. By integrating a high-resolution P2 feature layer (stride 4) while excising the macro-scale P5 layer, the network yields a 40% enhancement in small-gauge detection recall with a negligible (<1%) frame-rate degradation. Furthermore, to address the intrinsic lack of salient vertices in circular dials, we introduce a Virtual Point (VP) generation algorithm. This algorithm exploits the point symmetry of the detected keypoints to autonomously synthesize four spatial correspondences, thereby enabling markerless, homography-based perspective rectification for corner-free objects. An adaptive control mechanism based on aspect ratio analysis ($AR \leq 1.5$) dynamically regulates the geometric warping to prevent algorithmic over-correction. Extensive evaluations on an 11,000-image field dataset acquired from an operational power data center demonstrate a Pose mAP50 of 99.45% and an mAP50-95 of 99.37%. Under severe vertical tilt conditions, the VP-based rectification curtails the absolute reading error from 3.5% to 0.6% compared to the uncorrected baseline, attaining measurement precision commensurate with physical ArUco marker-based ground truths. Operating in real-time at 25.9 FPS, the proposed system is currently deployed within an integrated inspection platform coupled with an autonomous quadruped robot (Boston Dynamics SPOT), facilitating reliable, perspective-invariant visual inspections across 10 distinct classes of analog gauges in an active industrial facility.

KEYWORDS: Analog gauge; deep learning; keypoint detection; geometric rectification; Industrial Internet of Things (IIoT); pose estimation

1 Introduction

The uninterrupted monitoring of pivotal physical parameters, including pressure, temperature, and flow rate in both analog and digital formats, underpins the operational safety and systemic efficiency of modern industrial infrastructures, most notably power plants and power data management centers [1]. Notwithstanding the rapid proliferation of embedded digital sensors and the emergence of the Industrial Internet of Things (IIoT), external analog gauges remain ubiquitously deployed across diverse industrial

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

sectors. This enduring presence is primarily attributed to their exceptional durability, autonomy from external power sources, and robust reliability in hazardous environments where digital sensors may succumb to electromagnetic interference or extreme temperatures [2]. Typical examples of widely used industrial gauges are shown in Fig. 1.



Figure 1: Typical examples of analog gauges used in industrial facilities, illustrating diverse scale configurations and environmental conditions.

The inherent analog characteristics of these instruments induce a profound systemic gap; specifically, the absence of native data transmission capabilities restricts the acquisition of indicated physical values to visual inspection.

Consequently, industrial digitalization presents a systemic paradox: whereas the shift toward smart infrastructure is an industry-wide mandate, the persistent reliability of analog instruments in harsh environments creates a substantial bottleneck. Absent native digital interfaces, these gauges necessitate a continued reliance on manual inspection, which is a process characterized by heavy labor requirements and susceptibility to human error. This human-centric data acquisition cycle introduces significant temporal gaps between measurement and analysis, ultimately hindering the implementation of real-time monitoring and robust predictive maintenance systems.

1.1 Challenges in Analog Gauge Reading

The persistent reliance on analog instruments creates a critical systemic bottleneck in the digitalization of industrial maintenance. Manual inspection routines, characterized by periodic physical patrols, are inherently labor-intensive and susceptible to human-induced inaccuracies. This dependency not only incurs high operational costs but also introduces significant temporal gaps between data acquisition and analysis, ultimately hindering the realization of real-time monitoring and advanced predictive maintenance systems [3].

To overcome these limitations, automatic gauge reading systems leveraging autonomous agents—such as quadruped robots (Boston Dynamics SPOT [4]), Unmanned Aerial Vehicles(UAVs), and fixed CCTV cameras—have been actively investigated [2]. In this context, computer vision (CV)-based automatic analog gauge reading has evolved rapidly from classical image processing to advanced deep learning techniques [5].

However, realizing field gauge reading through autonomous agents requires addressing several key challenges:

- Environmental variability: industrial sites present diverse conditions including non-uniform illumination, protective glass reflections, dust and fog interference, and complex backgrounds (piping, cables), all of which severely impair the stable operation of classical image processing techniques.
- Viewing angle diversity: image capture via robots or CCTV cameras does not always occur from the frontal position, making acute oblique viewing angles inevitable. Under such conditions, circular dials appear as ellipses, and conventional polar unwrapping methods introduce nonlinear scaling errors.
- Object representation limitations: gauge needles are elongated objects with an extremely small width-to-length ratio. Conventional Axis-Aligned Bounding Boxes (AABBs) cause the background noise ratio to surge when the needle rotates, impeding precise orientation learning.
- Value conversion accuracy: obtaining the final reading requires precise conversion of the needle's visual position to a physical value (pressure, temperature), which in turn presupposes accurate distortion correction and faithful scale structure reconstruction.

To systematically address these challenges, this study focuses on two fundamental geometric limitations.

1.2 The Limitation of Axis-Aligned Bounding Boxes

The second challenge lies in a fundamental limitation of standard object detection frameworks. Previous studies often treat gauge reading as a vanilla object detection task [6], employing Axis-Aligned Bounding Boxes (AABBs) to localize the Region of Interest (ROI). While AABBs are effective for prominent block-shaped objects, they are structurally ill-suited for representing thin, rotating objects such as gauge needles.

As illustrated in Fig. 2, when a needle of length L and width w ($L \gg w$) rotates by angle θ, the enclosing AABB area expands substantially while the actual object area remains constant. This geometric mismatch causes the Intersection over Union (IoU) between the predicted and ground truth boxes to drop sharply, as the regression targets are inherently coupled with the pointer's spatial tilt. Mathematically, the AABB area and the resulting IoU can be approximated as:

$$Area_{aabb} \approx (L \cdot |\cos\theta|) \times (L \cdot |\sin\theta|) \tag{1}$$

$$IoU \approx \frac{L \times w}{L^2 \cdot |\sin\theta\cos\theta|} \to 0 \tag{2}$$

At a 45° inclination, the ratio of object-relevant pixels to irrelevant environmental noise within the ROI reaches its global minimum, compelling the convolutional filters to extract features from a high-entropy region. This dilution of structural information introduces significant "feature pollution", where the stochastic textures of the background (e.g., dial markings, shadows, or dust) dominate the feature maps. Consequently, the gradient signals during backpropagation become increasingly noisy as the spatial overlap decreases, hindering the convergence of orientation-sensitive layers. This structural inefficiency often results in a "systemic decoupling", where the model achieves accurate gauge localization (high recall) but fails to maintain the precision required for fine-grained needle angle estimation.

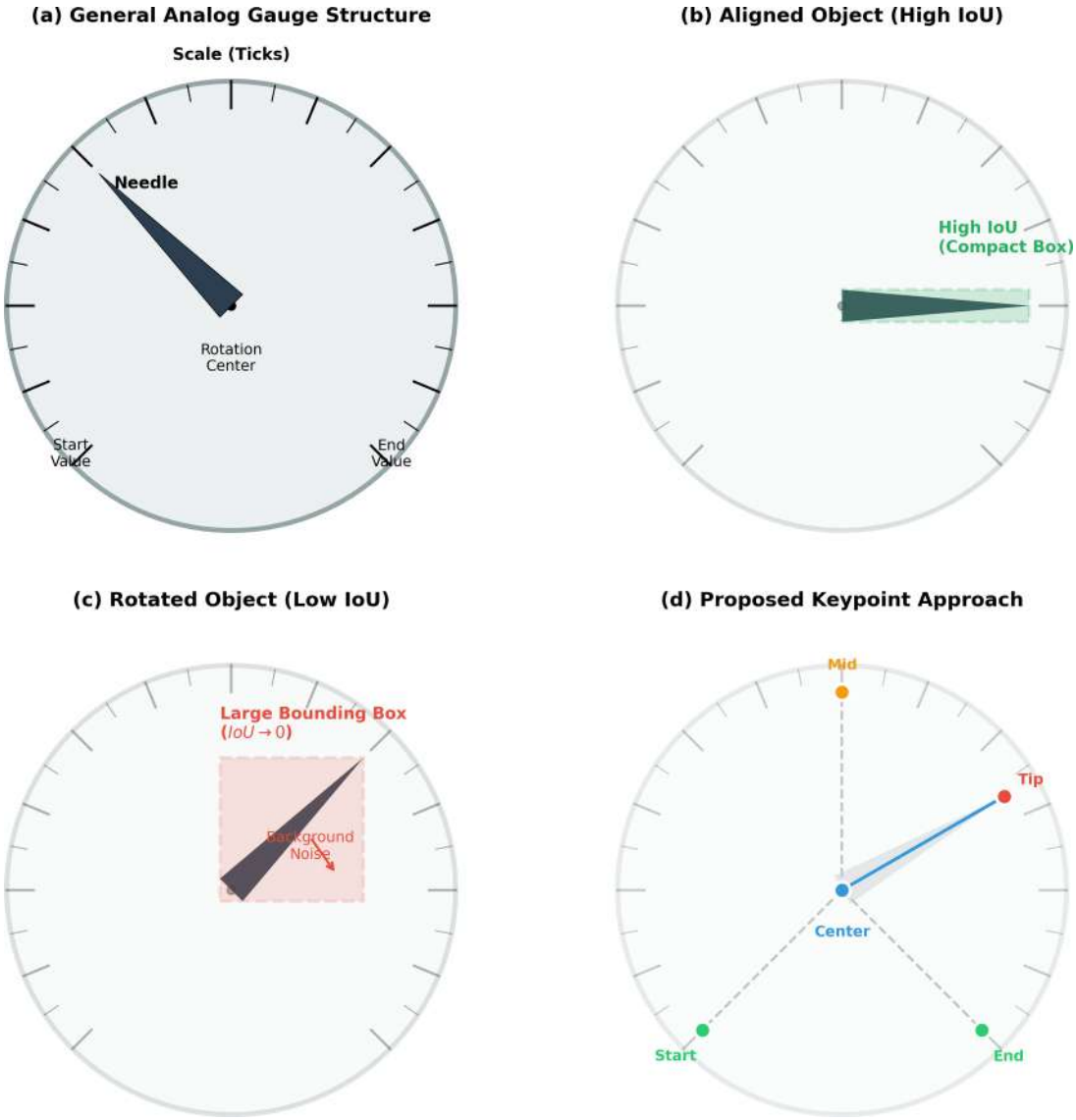


Figure 2: Comparison of object representation methods for analog gauges. (**a**) Typical structure of an analog gauge. (**b**) AABB is efficient for axis-aligned needles. (**c**) As the needle rotates, the AABB expands to include substantial background noise, reducing IoU. (**d**) The proposed keypoint-based approach captures explicit geometric structure regardless of rotation.

1.3 The Challenge of Perspective Distortion

The second critical challenge concerns the frontal viewing angle assumption. Most existing systems rely on polar-to-Cartesian unwrapping or vector direction detection for gauge reading. The fundamental limitation of polar unwrapping is that it is valid only when the gauge appears as a perfect circle. In practical scenarios involving mobile robots or fixed CCTV cameras, gauges are frequently captured at oblique angles, causing the circular dial to appear elliptical. Applying standard polar coordinate transformation to elliptical images introduces severe nonlinear scaling errors, as non-uniform arc lengths on an ellipse are treated linearly.

Meanwhile, purely vector-based methods that detect only pointer direction lack the mechanisms to reconstruct the scale structure and convert directions into physical values. They also lack projective distortion correction, causing the detected needle direction to be severely distorted under oblique viewing angles.

In summary, the common limitations are:

- Reliance on circularity causing geometric errors under oblique capture.
- Lack of physical value conversion mechanisms.
- Inability to geometrically rectify circular objects lacking distinct corners.

1.4 Proposed Approach and Contributions

This study focuses on achieving high accuracy and robustness under extreme viewing angles and challenging illumination conditions. To simultaneously address the limitations identified above—the non-linearity of polar coordinate conversion, the absence of value conversion in vector methods, and the lack of corner-free rectification—we redefine gauge reading not as a simple object detection task but as a *structural pose estimation* problem. Inspired by the Human Pose Estimation (HPE) paradigm, we propose a novel framework that treats each gauge as a skeleton composed of key structural points: five keypoints corresponding to the center, needle tip, and scale start/mid/end positions.

The main contributions of this paper are summarized as follows:

- A robust structural keypoint approach is introduced that defines gauges through structural relationships among specific points (center, needle tip, scale start, scale end). Unlike AABBs, this approach ensures robustness to thin needle geometries and partial occlusion while effectively excluding background noise for precise reading.
- A high-resolution P2 architecture is proposed through a modified YOLOv11-Pose model that integrates a high-resolution P2 layer (stride 4). This architectural enhancement preserves fine spatial information, substantially improving the detection recall for small needles and fine scale markings that are often lost during the downsampling process of standard models.
- A Virtual Point (VP)-based geometric rectification method is proposed, specifically designed for circular objects lacking distinct corners. By exploiting the point symmetry of detected keypoints, the algorithm automatically constructs four correspondence pairs and restores elliptically distorted gauges to a mathematically frontal circle via homography transformation. Unlike the polar coordinate transformations used by GAUREAD or Under Pressure, this approach enables precise angle-to-value conversion without nonlinear distortion even under oblique capture.

The remainder of this paper is organized as follows. Section 2 provides a systematic review of prior work on analog gauge reading, covering traditional image processing, deep learning approaches, pose estimation models, and the limitations of existing methods. Section 3 details the proposed methodology, including problem formulation, the virtual point-based geometric rectification algorithm, and the P2-YOLO-Pose architecture. Section 4 describes the data collection process and experimental setup in a real-world industrial setting. Section 5 presents evaluation metrics, quantitative and qualitative analyses, comparisons with state-of-the-art methods, and ablation study results. Section 6 provides an in-depth discussion of the strengths, limitations, failure cases, and practical deployment considerations. Section 7 presents use cases in industrial monitoring, smart manufacturing and IIoT, and the energy and utility sector. Finally, Section 8 summarizes the contributions and proposes future research directions.

2 Related Work

Automatic analog gauge reading is a core challenge in industrial automation. A broad spectrum of methodologies has been proposed, ranging from traditional image processing techniques to deep learning-based approaches, pose estimation model applications, and recent end-to-end frameworks. This section systematically categorizes prior work, analyzes the contributions and limitations of each approach, and clarifies the motivation and contributions of the present study.

2.1 Traditional Image Processing Methods for Gauge Reading

Early research on analog gauge reading relied primarily on classical computer vision techniques, employing handcrafted features to detect the structural elements of gauges.

The Hough Transform proposed by Duda and Hart [7] is a classical method for detecting lines and circles in images, and has been widely applied to detect the circular contours of gauge dials [8]. Alegria and Serra [9] extended this approach by extracting the center and radius of the dial using the Circle Hough Transform (CHT), computing the needle angle based on these parameters, and proposing the first automated gauge reading system. Their work is regarded as establishing the foundation of the automatic analog gauge reading field.

Canny's [10] edge detection algorithm has been used to extract the contours of needles and scales by detecting abrupt brightness changes, while Otsu's [11] thresholding technique has played a central role in separating the foreground (needle) from the background. Chi et al. [12] combined these preprocessing techniques into a complete pipeline: edge detection followed by Hough Transform for circular dial detection, binarization for needle segmentation, and angle computation for reading. Ma and Jiang [13] subsequently experimented with various preprocessing combinations based on similar principles to improve reading accuracy.

The camera calibration and multiple view geometry frameworks systematized by Zhang [14] and Hartley and Zisserman [15] provide the mathematical foundation for correcting projective distortion in gauge images. In particular, homography-based projective transformation, which estimates a view transformation matrix from planar correspondences, serves as the key tool for restoring obliquely captured gauges to frontal views.

While classical techniques offer computational efficiency and deterministic behavior, they possess inherent limitations. First, they are extremely sensitive to environmental variables such as non-uniform illumination, protective glass reflections, dust, and complex backgrounds (piping, cables), requiring manual parameter tuning for the Hough Transform on a per-environment basis. Second, binarization-based needle detection produces significant errors when contrast between the background and needle colors is insufficient—for example, a black needle on a dark dial face. Third, most of these methods assume frontal capture and lack automatic correction mechanisms for the elliptical distortion caused by oblique viewing angles. These limitations substantially restrict practical deployment in uncontrolled industrial settings.

2.2 Deep Learning Approaches for Analog Instrument Recognition

Advances in deep learning have significantly alleviated the environmental sensitivity problems of traditional methods. Through large-scale data and learnable feature extraction, deep learning-based systems demonstrate more robust gauge detection and reading performance across diverse conditions. This subsection provides a detailed analysis of recently proposed methods.

Milana et al. [2] proposed GAUREAD, an end-to-end gauge reading system comprising YOLOv5-based gauge detection, Circle Hough Transform for circular dial detection, ellipse fitting for shape estimation, and

polar-to-Cartesian unwrapping for scale/needle detection. GAUREAD achieved a processing time of 800 ms on an NVIDIA Jetson Nano, demonstrating the feasibility of edge-device deployment. However, the system exhibits a reading error of 3% within a 20° viewing angle from the frontal position, which escalates to 9% at 50°. This degradation stems from the structural limitation that polar unwrapping treats non-uniform arc lengths on an ellipse linearly. Additionally, the Circle Hough Transform can fail on gauges with unclear circular contours, such as semi-circular or fan-shaped gauges.

Dong et al. [16] proposed the Vector Detection Network (VDN), which models gauge pointers as two-dimensional vectors. In VDN, the initial point of the vector corresponds to the needle tip, and the direction follows tail-to-tip. The network estimates a confidence map to determine the initial point (peak pixel) and extracts direction components from a two-layer scalar map at each peak. Evaluated on the self-constructed Pointer-10K dataset, VDN demonstrated strong generalization performance and real-time processing speed across various gauge forms, including circular, semi-circular, and multi-pointer types. However, VDN detects only pointer direction without reconstructing scale structure (start point, end point, range) or providing a mechanism to convert direction information into physical values (pressure, temperature). Furthermore, the absence of projective distortion correction means that the needle direction itself becomes distorted under oblique viewing angles.

Most recently, Reitsma et al. [17] proposed the Under Pressure framework at ETH Zurich ASL. The system follows a step-by-step pipeline of gauge detection, notch detection with ellipse fitting, needle segmentation, scale marker recognition, and unit extraction. A notable advantage is that each stage's potential failures can be diagnosed in an interpretable manner. The system operates without prior knowledge of gauge type or scale range and provides automatic unit extraction. Experimental results achieved relative error below 2%. However, this performance was primarily measured under near-frontal viewing angles, and robustness under obscure notch conditions or severe projective distortion remains unvalidated.

Leon-Alcazar et al. [3] proposed training robust reading models using large-scale synthetic data for diverse gauge forms and conditions. While synthetic data presents a promising approach to reducing data collection costs, domain gaps between synthetic and real-world field data persist, particularly in reproducing subtle geometric distortions and site-specific interference (glass reflections, condensation). Tian et al. [5] proposed GaugeTracker, a hybrid system combining template matching with deep learning, achieving improved reading precision but lacking flexibility for gauge types without predefined templates or severely distorted images. The Programmable Gradient Information (PGI) concept introduced by Wang et al. [18] in YOLOv9 enhances feature learning efficiency for small object detection and represents an architectural advancement applicable to gauge reading technologies.

2.3 Pose Estimation Models in Visual Measurement

Human Pose Estimation (HPE) is one of the most actively researched topics in computer vision, aiming to estimate joint positions and reconstruct skeletal structures from images. This study innovatively applies the HPE paradigm to the industrial measurement domain.

Pose estimation is broadly classified into two paradigms. Top-down approaches first detect each object and then estimate keypoints within each detected instance, while bottom-up approaches first detect all keypoints and subsequently group them into individual objects. YOLO-Pose [19] is a representative model that integrates the top-down approach into a single network, simultaneously performing object detection and keypoint regression at real-time speed. This model introduces the Object Keypoint Similarity (OKS) loss function to incorporate structural relationships among keypoints into the learning process. He et al. [20] demonstrated with Mask Regions with Convolutional Neural Networks features (R-CNN) that P2-level

feature maps from the Feature Pyramid Network (FPN) are essential for precise keypoint localization, experimentally establishing the importance of high-resolution feature maps.

The Feature Pyramid Network (FPN) proposed by Lin et al. [21] is a key architecture that hierarchically fuses multi-scale feature maps to effectively detect objects of various sizes. YOLO-family models typically use a three-level pyramid comprising P3 (stride 8), P4 (stride 16), and P5 (stride 32). However, for small objects such as gauge needles and fine scale markings, P3 may not preserve sufficient spatial resolution. Adding a P2 (stride 4) layer provides 4× higher spatial resolution, though the trade-off between computational overhead and memory consumption must be considered. This study achieves a balance between high-resolution keypoint detection performance and computational efficiency through a structural optimization that integrates the P2 layer while simultaneously removing the P5 layer.

The keypoint-based structural recognition paradigm established in HPE extends naturally to the industrial measurement domain. Just as human joints define the physical structure of arms, legs, and torso, gauge keypoints (center, needle tip, scale start/mid/end) define the geometric structure of the circular dial-needle system. Based on this analogy, the proposed method encodes the structural relationships among five key gauge keypoints into the loss function and maximizes relative positional accuracy through OKS-based training. This enables precise needle direction estimation and complete scale structure reconstruction that were impossible with AABB-based methods. Furthermore, industrial defect detection benchmarks such as MVTec AD provided by Bergmann et al. [22] underscore the importance of rigorous evaluation methodologies in industrial visual inspection, a principle that this study applies to its experimental design.

2.4 Limitations of Existing Methods

Synthesizing the common limitations of the prior work reviewed above, current automatic analog gauge reading technology faces three critical open challenges.

Most methods, including GAUREAD [2] and Under Pressure [17], perform correction based on polar coordinate transformation (polar unwrapping) or ellipse fitting. However, these approaches assume circular or quasi-circular dials, and nonlinear errors increase rapidly as elliptical distortion from oblique viewing intensifies. GAUREAD reports errors of 3% within 20° and 9% at 50°; such angle-dependent error growth constitutes a critical constraint in real-world robotic patrol scenarios, where it is infeasible for the robot to always stop directly in front of each gauge.

While VDN [16] provides a flexible and generalizable method for pointer direction detection, the pipeline does not include mechanisms to convert detected directions into actual physical values (pressure, temperature). Since the ultimate objective of gauge reading in industrial settings is to obtain quantitative measurements, directional information alone has limited practical utility. Value conversion requires knowledge of scale start points, end points, and the range between them—structural information that VDN does not estimate.

The optimal mathematical solution to ensure measurement invariance regardless of the camera viewpoint is homography rectification; however, this requires securing at least four discrete point-to-point correspondence pairs. Unlike rectangular objects with identifiable salient vertices, circular gauges are intrinsically "corner-free" objects lacking prominent corners. Previous studies attempted to bypass this limitation through curve extraction via CHT or shallow notch matching, but the detection reliability of such alternative features drops precipitously in real-world environments characterized by partial occlusion from piping or irregular notch patterns across manufacturers. This chronic inability to autonomously establish reliable correspondences without external fiducial markers remains the most significant algorithmic barrier to achieving flawless planar rectification.

In conclusion, realizing autonomous inspection in uncontrolled, unstructured industrial environments requires completely departing from the fragmented approaches that treat analog gauges merely as simple bounding boxes or isolated line segments (vectors). To concurrently resolve the three major open challenges that existing methodologies have failed to overcome, (1) the nonlinearity of oblique distortion, (2) the absence of quantitative value conversion, and (3) the inability to rectify corner-free objects, a novel paradigm organically integrating structural topology estimation and geometric rectification is imperative.

Accordingly, in the subsequent Section 3 (Proposed Methodology), this study details our uniquely integrated end-to-end framework. This framework seamlessly connects the extraction of keypoint skeletons based on high-resolution P2-YOLO-Pose, the autonomous generation of Virtual Points (VPs) for corner-free objects by mathematically leveraging point symmetry to perform homography rectification, and error-free quantitative value conversion within a linear coordinate system completely devoid of projective distortion.

3 Proposed Methodology

3.1 Problem Definition and System Overview

3.1.1 Problem Formulation

The analog gauge reading problem is formally defined as estimating the true physical value V_{result} indicated by a gauge from a raw 2D input image I. Mathematically, this corresponds to optimizing a composite mapping function f parameterized by learnable weights Θ:

$$V_{result} = f(I;\Theta) = \mathcal{V} \circ \mathcal{R} \circ \mathcal{D}(I;\Theta) \tag{3}$$

where $\mathcal{D}$ denotes the structural pose detection function that extracts the topological keypoint skeleton of the gauge, $\mathcal{R}$ represents the geometric rectification function that neutralizes perspective distortion via homography, and $\mathcal{V}$ is the deterministic vector-based value computation function. This tripartite formulation is explicitly designed to resolve the geometric bottlenecks of existing models: the spatial constraints of bounding boxes, the nonlinear scaling errors of polar unwrapping, and the lack of salient vertices in circular dials.

3.1.2 Overall System Architecture and Workflow

The overall architecture of the proposed system is illustrated in Fig. 3. The system consists of three major stages:

- Stage 1—High-Resolution Keypoint Extraction via P2-YOLO-Pose: the P2-enhanced YOLOv11-Pose model simultaneously extracts five keypoints ($k_{start}, k_{mid}, k_{center}, k_{end}, k_{tip}$) and bounding boxes from the input image.
- Stage 2—Virtual Point-Based Adaptive Geometric Rectification: after multi-stage validation of detected keypoints, virtual points are generated using the point symmetry principle, and the distorted elliptical gauge is restored to a frontal circle via homography transformation.
- Stage 3—Vector-Based Value Computation in Canonical Metric Space: vector-based angle calculation and a circular distance function are applied in the rectified coordinate system to convert the needle position into a physical value.

The operational workflow proceeds as follows. An autonomous robot (Boston Dynamics SPOT) approaches inspection targets along predefined waypoints and acquires high-resolution images through its mounted optical payload. The captured images are processed by P2-YOLO-Pose for keypoint extraction, and only data passing multi-stage validity verification (confidence, structural completeness, physical constraints)

proceeds to subsequent processing. Virtual points are generated from verified keypoints, and the aspect ratio (AR) analysis determines whether rectification is performed. Homography rectification is applied when AR ≤ 1.5; otherwise, values are computed directly from the original image when AR > 1.5.

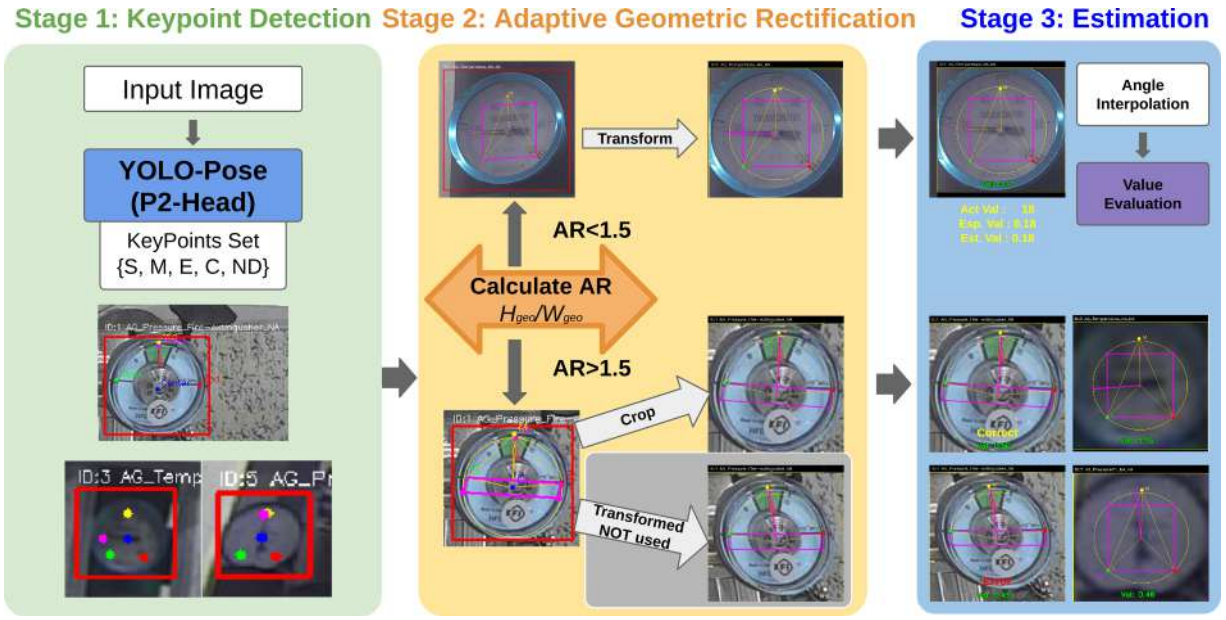


Figure 3: System architecture of the proposed analog gauge reading framework, illustrating the complete pipeline from keypoint detection through geometric rectification to final value computation.

3.2 Stage 1: High-Resolution Keypoint Extraction via P2-YOLO-Pose

To establish a robust geometric foundation for subsequent perspective correction, precise pixel-level localization of the gauge components is paramount. Therefore, we shift the detection paradigm from regional bounding boxes to topological keypoints.

3.2.1 Keypoint Detection for Gauge Needle and Scale

We mathematically model the analog gauge as a rigid topological skeleton comprising five semantically distinct keypoints: $K = \{k_{start}, k_{mid}, k_{center}, k_{end}, k_{tip}\}$. The semantic role of each keypoint is explicitly assigned to bypass the noise-susceptible nature of conventional AABBs:

- Scale points: k_{start}, k_{mid}, k_{end}—corresponding to the physical minimum, midpoint, and maximum positions on the gauge dial. These three points establish the reference for the angular span of the reading range and serve as the foundational data for virtual point generation.
- Rotation center: k_{center}—represents the mechanical pivot of the needle. It functions as the absolute geometric origin for all trigonometric computations and the invariant anchor for point-symmetric reflection.
- Needle tip: k_{tip}—denotes the terminal extremity of the indicator needle that dictates the temporal measured value, demanding the highest sub-pixel localization precision within the network.

This keypoint-based approach differs from AABB-based detection in three fundamental ways. First, *structural context*: the model infers the needle tip position in relation to the center and scale points, enabling robust localization even for thin needles or under partial occlusion. Second, *background noise suppression*: by focusing on specific coordinates rather than the entire bounding box, background elements such as piping

and cables are effectively ignored. Third, *geometric robustness*: the skeleton formed by five keypoints provides complete geometric information for homography rectification.

3.2.2 YOLO-Pose Architecture with P2 Feature Layer Enhancement

YOLO-Pose [19] is a real-time pose estimation framework that simultaneously performs object detection and keypoint regression in a single network. Unlike conventional top-down approaches, it does not require a separate human detector and introduces the Object Keypoint Similarity (OKS) loss function to incorporate structural relationships among keypoints into the learning process.

This study adopts Ultralytics' YOLOv11 [23,24] as the base architecture. YOLOv11 consists of Backbone (feature extraction), Neck (multi-scale feature fusion), and Head (detection and keypoint regression), with the Neck combining Feature Pyramid Network (FPN) [21] and Path Aggregation Network (PAN) structures for efficient multi-scale information integration.

By default, YOLOv11-Pose outputs a three-level pyramid comprising P3 (stride 8), P4 (stride 16), and P5 (stride 32) [21,24]. However, for the few-pixel-wide needles and fine scale markings required in gauge reading, spatial information may not be sufficiently preserved even at P3 (stride 8). He et al. [20] demonstrated with Mask R-CNN that high-resolution feature maps are essential for precise keypoint localization; this study applies this principle to the industrial measurement domain by introducing the P2 layer.

The precision of analog gauge reading depends directly on the accurate localization of the needle tip and scale markings. Since the minimum stride of standard YOLOv11 is 8, the P3 feature map resolution is only 160 × 160 when the input image is 1280 × 1280 pixels. At this resolution, a needle only a few pixels wide spans 1–2 grid cells, making precise keypoint localization difficult.

This study adds a P2 layer with stride 4, doubling the feature map resolution to 320 × 320. This enables needle tip regression on a 4× finer grid, significantly reducing keypoint localization error.

Simultaneously, since the detection of large objects (buildings, background elements) is unnecessary for gauge reading, the P5 (stride 32) layer is removed. This structural optimization (P2 addition + P5 removal) provides the following benefits:

- High-resolution keypoint detection: the P2 layer preserves spatial information for fine structures (needles, scale markings)
- Computational efficiency: removing P5 offsets the computational overhead introduced by adding P2
- Domain optimization: elimination of large-object detection capability focuses the model on small-to-medium objects relevant to gauge reading

3.2.3 Loss Functions and Optimization Strategy

The P2-YOLO-Pose model is trained using the following multi-task loss function:

$$\mathcal{L}_{total} = \lambda_{box} \cdot \mathcal{L}_{CIoU} + \lambda_{cls} \cdot \mathcal{L}_{BCE} + \lambda_{dfl} \cdot \mathcal{L}_{DFL} + \lambda_{pose} \cdot \mathcal{L}_{OKS} + \lambda_{kobj} \cdot \mathcal{L}_{kobj} \tag{4}$$

The role of each component is as follows:

- $\mathcal{L}_{CIoU}$: Complete Intersection over Union loss, which simultaneously regresses the position, size, and aspect ratio of bounding boxes. Compared to standard IoU, it additionally considers center distance and aspect ratio consistency, improving convergence speed and accuracy.
- $\mathcal{L}_{BCE}$: Binary Cross-Entropy classification loss for learning gauge class probabilities.

- $\mathcal{L}_{DFL}$: Distribution Focal Loss, which represents bounding box coordinates as discrete distributions, enhancing flexibility and precision compared to single-point regression.
- $\mathcal{L}_{OKS}$: Object Keypoint Similarity-based keypoint loss that maximizes the similarity between predicted and ground truth keypoints. OKS considers per-keypoint scale, evaluating precision relative to object size rather than absolute pixel error.
- $\mathcal{L}_{kobj}$: Keypoint objectness loss that learns the visibility of each keypoint, providing the basis for judgment under partial occlusion conditions.

A pivotal architectural decision in our optimization strategy is the imposition of an asymmetric weight distribution, where the Object Keypoint Similarity (OKS) loss weight (λ_{pose}) is prioritized substantially over the bounding box loss weight (λ_{box}). In visual measurement tasks, a multi-pixel deviation in a bounding box is functionally negligible; conversely, a sub-pixel coordinate perturbation of the needle tip (k_{tip}) directly induces fatal angular translation errors. Thus, this formulation forces the network's gradient updates to overwhelmingly prioritize structural keypoint fidelity.

3.3 Stage 2: Virtual Point-Based Geometric Rectification

3.3.1 Perspective Distortion Correction

Deep learning-based pose estimation is susceptible to hallucinating coordinates under extreme specular reflections or severe partial occlusions. Prior to executing geometric transformations, the detected skeleton undergoes rigorous deterministic filtering to ensure physical plausibility:

- Confidence Threshold: Detections with output confidence below a preset threshold are eliminated to exclude background noise and false positives.
- Topological Completeness: The inference is rejected if any of the five cardinal keypoints are undetected, as incomplete skeletons preclude exact geometric reconstruction.
- Radius Ratio Consistency: Let d_s, d_m, d_e denote the Euclidean radial distances from k_{center} to the respective scale boundaries $(k_{start}, k_{mid}, k_{end})$. Under valid projection assumptions, these radii must satisfy a geometric concentricity tolerance:

$$\text{if} \quad \frac{\min(d_s, d_m, d_e)}{\max(d_s, d_m, d_e)} < 0.4 \quad \Rightarrow \quad \text{reject} \tag{5}$$

  Since the three radii should theoretically be equal in a circular gauge, a ratio below 0.4 indicates physically implausible keypoint locations (severe deformation or hallucination). This threshold was established from field data analysis showing that legitimate gauges exhibit minimum ratios above 0.45, while erroneous detections mostly fall below 0.3.
- Boundary Constraints: Detections where k_{center} is located within a marginal threshold (δ = 10 pixels) of the image boundaries are discarded:

$$\text{if} \quad x_c < \delta \text{ or } x_c > W - \delta \text{ or } y_c < \delta \text{ or } y_c > H - \delta \quad \Rightarrow \quad \text{reject} \quad (\delta = 10) \tag{6}$$

  This is because symmetric point generation from a truncated gauge projects outside the image boundary, causing numerical instability in the homography transformation.

3.3.2 Autonomous Virtual Point (VP) Generation via Point Symmetry

Computing a planar homography matrix M mathematically requires a minimum of four distinct correspondence pairs. However, circular gauges are intrinsically "corner-free." To establish reliable correspondences autonomously without relying on ambiguous curve features or unpredictable notches, we

exploit the rotational invariance of the gauge center. By mapping the verified scale boundaries k_{start} and k_{end} through point-symmetric reflection across k_{center}, two mathematically sound Virtual Points (VPs) are synthesized:

$$k'_{start} = 2k_{center} - k_{start}, \quad k'_{end} = 2k_{center} - k_{end} \tag{7}$$

The resulting discrete spatial set $P_{src} = \{k_{start}, k_{end}, k'_{start}, k'_{end}\}$ forms an intrinsic "rhombus grid" parallel to the physical dial plane. This synthesis circumvents the absence of salient vertices by algorithmically inferring exact geometric correspondences.

3.3.3 Adaptive Rectification Control via Aspect Ratio Analysis

Indiscriminate application of forced perspective warping to extreme oblique angles induces aggressive pixel interpolation (stretching), which corrupts fine needle morphology. To mitigate algorithm-induced over-correction, an adaptive geometric control logic evaluates the Aspect Ratio (AR) of the synthesized rhombus grid:

$$AR = \frac{\max(\|k_{start} - k_{end}\|, \|k_{end} - k'_{start}\|)}{\min(\|k_{start} - k_{end}\|, \|k_{end} - k'_{start}\|)} \tag{8}$$

A high AR indicates that the angle between the camera and the gauge plane is very acute, causing the four correspondence points to converge toward collinearity.

An AR threshold of >1.5 is established as the criterion for bypassing rectification. This value is justified as follows:

- Geometric rationale: AR = 1.5 corresponds to a 3:2 major-to-minor axis ratio, equivalent to a camera capturing the gauge plane from approximately 48.2°. Beyond this angle, the condition number of the homography matrix increases sharply, and numerical stability of the resulting matrix cannot be guaranteed.
- Experimental rationale: forced warping at AR > 1.5 induces pixel stretching—expanding a limited number of pixels into a larger space—which was observed to degrade the fine angular information of the needle.

Consequently, the algorithm adopts an adaptive strategy: homography rectification is performed when AR ≤ 1.5, and values are computed directly from the original image when AR > 1.5.

3.3.4 Homography Transformation

From the validated keypoints and generated virtual points, the homography matrix between the source and target coordinate systems is computed.

The source coordinates P_{src} are the rhombus grid $\{k_{start}, k_{end}, k'_{start}, k'_{end}\}$ in the actual image, while the target coordinates P_{dst} are designed so that the gauge forms a perfect frontal circle in the restored image. Specifically, the target coordinates are symmetrically placed on a circle of radius R while preserving the span angle between scale points:

$$\theta_{span} = \arccos\left(\frac{\vec{v}_s \cdot \vec{v}_e}{\|\vec{v}_s\| \cdot \|\vec{v}_e\|}\right), \quad \vec{v}_s = k_{start} - k_{center}, \quad \vec{v}_e = k_{end} - k_{center} \tag{9}$$

where θ_{span} is the angular extent of the arc between the scale start and end points. The target k_{start} and k_{end} are placed symmetrically at $\theta_{span}/2$ from a reference angle, which is set to $-\pi/2$ (upper) or $+\pi/2$ (lower) depending on the gauge orientation (arc-up or arc-down).

The homography matrix M between the two correspondence point sets is defined in homogeneous coordinates as:

$$\mathbf{x}_i' = M\mathbf{x}_i \quad \Rightarrow \quad \begin{bmatrix} w'x_i' \\ w'y_i' \\ w' \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{10}$$

M has 8 independent degrees of freedom and is uniquely determined by the 8 linear equations provided by 4 correspondence pairs. The Direct Linear Transformation (DLT) algorithm is used for matrix computation.

Unlike simple cropping, homography-based restoration mathematically compensates for the "asymmetric displacement of scale markings" caused by the viewing angle. Under lateral capture, scale markings closer to the lens appear compressed while those farther away appear stretched; the matrix M reverses this nonlinear pixel density variation, realigning all scale intervals to be proportional to their actual angles. This process is equivalent to restoring the distorted rhombus grid to a rectangular structure, recovering the orthogonality of coordinate axes and ensuring the geometric linearity required for subsequent vector-based angle computation.

3.4 Stage 3: Coordinate Transformation and Value Calculation

Once the geometric integrity of the dial is mathematically restored (or conditionally bypassed when $AR > 1.5$), the ultimate conversion from visual data to physical quantities is executed within this distortion-free linear coordinate system.

3.4.1 Coordinate Translation and Circular Distance Metric

The top-left origin of the standard digital image coordinate system is translated to the gauge rotation center (k_{center}), and the y-axis is inverted to conform to standard Cartesian and trigonometric conventions. The absolute orientation angle θ for any valid target point $P(x, y)$ is extracted via:

$$\theta = \text{atan2}\left(-(y - y_c),\ x - x_c\right) \cdot \frac{180}{\pi} \tag{11}$$

The computed angles are normalized to the range $[0°, 360°)$ using modular arithmetic. Because analog dials frequently traverse the 360° boundary (e.g., an abrupt transition from 350° to 10°), simple arithmetic subtraction yields unphysical discontinuities. We circumvent this by defining a continuous, clockwise relative circular distance function:

$$\Delta\theta(a, b) = (a - b) \pmod{360} \tag{12}$$

3.4.2 Final Physical Value Interpolation

Given the prerequisite that homography transformation has strictly restored geometric linearity—where equivalent angular segments correctly correspond to uniform physical scale ticks—the definitive continuous physical reading V_{result} is deterministically interpolated using the pre-registered physical operational bounds $[V_{min}, V_{max}]$ of the specific gauge:

$$V_{result} = V_{min} + \frac{\Delta\theta(\theta_{start}, \theta_{tip})}{\Delta\theta(\theta_{start}, \theta_{end})} \cdot (V_{max} - V_{min}) \tag{13}$$

Because this interpolation occurs entirely within the mathematically rectified planar space, the resulting measurement is strictly immunized against perspective-induced nonlinear errors, seamlessly translating unstructured visual capture into an exact quantitative diagnostic.

4 Experiments

This section describes the experiments conducted to validate the performance and field applicability of the proposed system. The data collection process reflecting diverse environmental variables in a real industrial setting is presented first, followed by the characteristics of the constructed dataset. Subsequently, the training configuration for the P2-YOLO-Pose model and the rationale for each hyperparameter selection are provided. Finally, the ArUco marker-based control group experimental design for objectively verifying the accuracy of the geometric rectification algorithm is described.

4.1 Data Collection and Dataset Construction

4.1.1 Experimental Environment

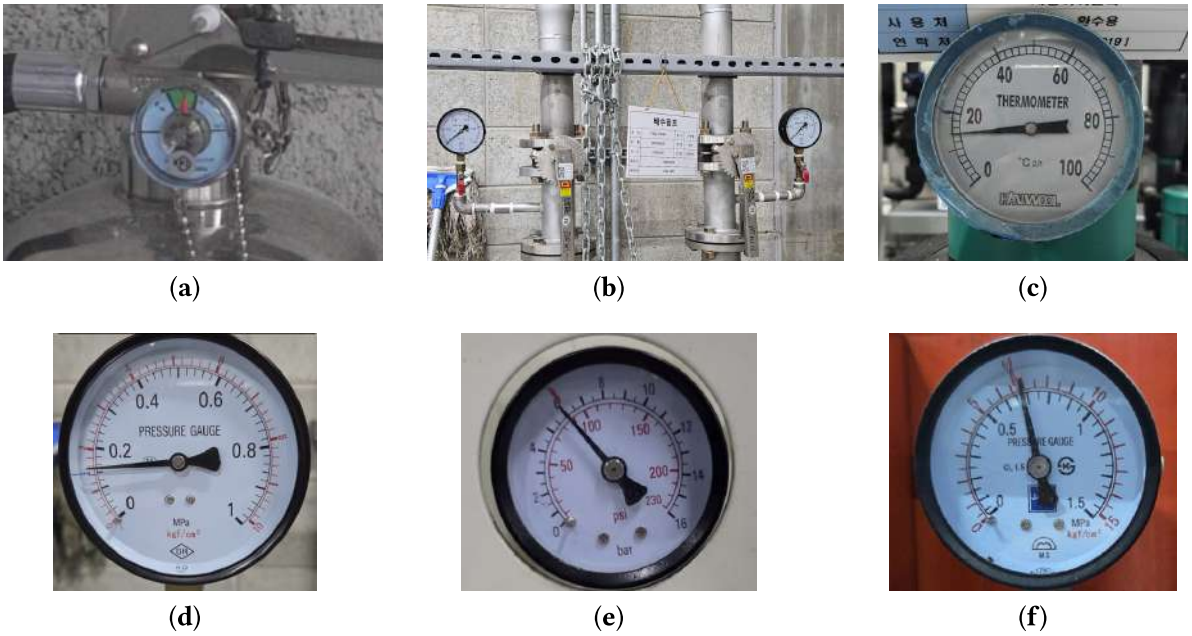
Field experiments were conducted in the underground infrastructure facilities of an operational power data center for deep learning model training and validation. The experimental sites were selected from facilities subject to regular manual inspections to ensure stable data center operation. The data center's underground infrastructure comprises 17 patrol inspection compartments, of which 9 are accessible to robots. The total number of inspection points is approximately 1200, of which roughly 670 were deemed accessible via robotic inspection. The machine rooms, where analog gauges are most extensively distributed, fire extinguisher facilities, emergency diesel generators, reservoirs, geothermal systems, and drainage pumps. The dataset was constructed from 10 types of analog gauges (7 pressure gauges, 1 ammeter, 1 thermometer, and 1 thermo-hygrometer) found in these facilities (Fig. 4).

Fire extinguisher pressure gauges, which constitute the largest proportion of data center infrastructure inspections, are located near fire extinguisher signage. Fire extinguishers are individually distributed, and the pressure gauges attached to agent storage and actuation vessels are extremely small, demanding precise small-object detection capability. Additionally, numerous pumps related to geothermal heat pump systems and water supply pressurization are present, requiring periodic pressure inspection of pumps and piping under conditions where visual occlusion frequently occurs due to complex piping and cylinder configurations.

4.1.2 Mobile Platform and Image Acquisition

A Boston Dynamics SPOT quadruped robot [4] was employed as the mobile platform for data collection, capable of stable navigation within complex and confined underground facilities. A camera module supporting high-resolution optical zoom and PTZ (Pan-Tilt-Zoom) control was mounted on the robot's back, enabling gauge image capture from various heights (low/high angle) and orientations during autonomous or remote navigation.

A total of 500 high-resolution raw images were acquired through field exploration using the robot.



(a) (b) (c)

(d) (e) (f)

Figure 4: Data collection environment and target analog gauge types. (**a**) Fire extinguisher pressure gauge; (**b**) drainage pump room pressure gauge; (**c**) geothermal recovery thermometer; (**d**) pressure gauge Type A; (**e**) pressure gauge Type B; (**f**) pressure gauge Type C.

4.1.3 Data Augmentation and Dataset Split

A two-stage augmentation strategy was adopted to construct a dataset of sufficient scale from the 500 raw images. Stage 1 expands the dataset itself through offline augmentation, while Stage 2 introduces additional diversity during training through online augmentation within the YOLO training pipeline.

An offline augmentation pipeline based on the Albumentations [25] library was constructed using a custom-developed data management tool. A key feature of this pipeline is the simultaneous transformation of image and five-keypoint coordinates, ensuring geometric consistency between augmented images and labels. The following augmentation techniques were applied:

- Random perspective transform: perspective transformation was applied by randomly displacing image corners to simulate the robot's diverse approach angles. This provides a more realistic simulation of actual oblique capture distortion compared to simple 2D rotation.
- Color jittering: brightness, contrast, saturation, and hue were simultaneously randomized to simulate color variations arising from time-of-day and lighting conditions.
- Gaussian blur: blur effects were applied to simulate out-of-focus conditions of the PTZ camera.
- Gaussian noise: signal noise arising from low-light environments or sensor characteristics was simulated.
- Occlusion: to simulate partial gauge occlusion by piping, cables, and other elements common in industrial settings, random rectangular regions were generated within the keypoint bounding area and filled with black masking or blur. The occlusion area was set to approximately 5% of the object area to maintain geometric relationships among keypoints while learning robustness to partial occlusion.

Through offline augmentation, a final dataset of 11,000 images was constructed. For reliable model training and evaluation, the dataset was split into Training (70%), Validation (15%), and Test (15%) subsets.

4.2 Model Training Configuration

This subsection describes the specific hyperparameters used for P2-YOLO-Pose model training and the rationale for each setting.

4.2.1 Model Architecture and Input Settings

The P2 layer-integrated and P5 layer-removed YOLOv11-Pose was used as the base architecture, initialized with pre-trained weights from the COCO-Pose dataset [26].

The input image resolution was set to 1280 × 1280 pixels to maximize detection capability for small objects such as fire extinguisher gauges, providing 4× the pixel space compared to the 640 × 640 typically used in standard YOLO training. When combined with the P2 layer (stride 4), the feature map resolution reaches 320 × 320, enabling stable regression of needle tips only a few pixels wide. Furthermore, to evaluate the maximum bounds of small-object detection capability, inference was performed not only at the native training resolution (1280 px) but also at an upscaled resolution (1920 px).

4.2.2 Loss Function Weight Configuration

The weights for each component of the multi-task loss function are configured as shown in Table 1.

Table 1: Loss function weight configuration and rationale.

Loss Function	Weight (λ)	Rationale
$\mathcal{L}_{CIoU}$ (Box)	7.5	Stable gauge region localization
$\mathcal{L}_{BCE}$ (Cls)	0.5	Reduced classification burden (10 classes)
$\mathcal{L}_{DFL}$	1.5	Default weight for coordinate distribution learning
$\mathcal{L}_{OKS}$ (Pose)	31.0	Keypoint precision prioritized
$\mathcal{L}_{kobj}$	2.0	Auxiliary keypoint visibility judgment

Notably, $\lambda_{pose} = 31.0$, approximately 4.1× that of $\lambda_{box} = 7.5$. Although the dataset covers 10 types of analog gauges, all classes share an identical five-keypoint skeleton (k_{start}, k_{mid}, k_{end}, k_{center}, k_{tip}) and are read through the same geometric pipeline regardless of gauge type; consequently, accurate class identification contributes minimally to the final reading value, justifying a relatively low classification loss weight ($\lambda_{cls} = 0.5$). Conversely, since the final reading accuracy is directly determined by the pixel coordinate precision of the needle tip (k_{tip}), the highest weight is assigned to the keypoint loss. This ratio was determined through a grid search over $\lambda_{pose} \in \{15, 17, 19, \ldots, 43\}$ (step of 2) in preliminary experiments, selecting the value yielding the highest Pose mAP50-95.

4.2.3 Optimizer and Learning Rate Schedule

The optimizer and schedule settings used for training are presented in Table 2.

AdamW was adopted because it decouples weight decay from the learning rate, providing more stable regularization for precise regression tasks such as keypoint coordinate prediction. The initial learning rate $\eta_0 = 4.305 \times 10^{-3}$ was determined through Ultralytics' auto-tuning mechanism, and cosine annealing gradually reduces it to 1% of η_0 for fine-grained adjustments in later training stages.

Table 2: Summary of model training configuration.

Parameter	Value	Rationale
Model Configuration		
Model	YOLOv11-Pose (P2+, P5−)	Small keypoint detection optimization
Pre-training	COCO-Pose Pretrained	Accelerated feature extraction
Input resolution	1280 × 1280	Small gauge accommodation
Optimizer Settings		
Optimizer	AdamW (auto)	Adaptive LR + decoupled weight decay
Initial LR (η_0)	4.305×10^{-3}	Ultralytics auto-tune
Final LR ratio (η_f/η_0)	0.01	Late-stage fine adjustment
LR Schedule	Cosine Annealing	Convergence stability
Momentum	0.937	Gradient inertia retention
Weight Decay	5×10^{-4}	Overfitting prevention
Training Strategy		
Warm-up	3 epochs, momentum 0.8 → 0.937	Initial instability prevention
Total epochs	50	Convergence verified (plateau after epoch 30)
Batch size	12	GPU VRAM limit (at 1280 px)
Early Stopping	Patience = 15	Early overfitting prevention

A batch size of 12 is the maximum feasible setting considering the 1280 × 1280 high-resolution input and the additional computational burden of the P2 layer. At 1280 px input with the P2-included model, GPU memory usage reaches approximately 20 GB (on a 24 GB VRAM GPU), and increasing the batch size beyond 12 resulted in OOM (Out-of-Memory) errors. The total training epochs were set to 50, based on learning curve analysis confirming validation loss saturation after approximately epoch 30. Early stopping was employed to prevent overfitting.

4.2.4 Training-Time Data Augmentation Strategy

In addition to offline augmentation (Section 4.1), the specific parameters for online augmentation applied in real-time during training are presented in Table 3.

Table 3: Training-time data augmentation parameter configuration.

Augmentation	Parameter	Simulated Environmental Variable
Mosaic	1.0 (disabled last 10 epochs)	Complex backgrounds (piping, cables)
MixUp	$\alpha = 0.5$	Decision boundary regularization
Copy-Paste	1.0 (flip mode)	Gauge appearance contextual diversity
HSV-Hue	±0.015	Illumination color temperature
HSV-Saturation	$\times[0.3, 1.7]$	Saturation (glass reflections, dust)
HSV-Value	$\times[0.6, 1.4]$	Brightness (shadows, direct sunlight)
Scale	$s \sim U(0.5, 1.5)$	Capture distance variation
Horizontal Flip	$P = 0.5$	Left-right placement differences
RandAugment	Automatic policy	General invariance

Mosaic augmentation is effective for inducing rapid convergence by quadrupling per-batch contextual diversity during early training. However, artificially composed images may hinder fine-grained keypoint regression in later stages [27]. Therefore, Mosaic is deactivated during the last 10 epochs, allowing the model to focus on precise keypoint localization in individual images.

The Hue transformation range is conservatively set because gauge scale color differentiation (normal = green, danger = red) may be utilized for initial reading verification. Conversely, Saturation and Value ranges are set relatively wide to ensure robustness against field saturation and brightness variations caused by glass reflections, dust, and shadows.

Copy-Paste augmentation [28], which copies and pastes gauge instances with flipping, is applied at 100% probability. This simulates the characteristic of industrial sites where identical gauge types are installed against various backgrounds, effectively learning background invariance for keypoint detection.

4.3 Geometric Rectification Accuracy Validation

To objectively validate the accuracy of the proposed keypoint-based geometric rectification method, a precision comparison experiment was designed using ArUco markers as a control group.

4.3.1 Control Group Setup

During data collection, ArUco markers (dictionary: `DICT_ARUCO_ORIGINAL`, size: 30 mm) were attached adjacent to selected target gauges (within 5 cm on the floor or side). When multiple markers are present in a single image, the Euclidean distance between the detected gauge bounding box center and each marker center is computed, and the nearest marker is automatically selected as the reference marker for that gauge.

The 6-Degrees of Freedom (6-DoF) pose is estimated using the four corner points of the marker, and the 3×3 rotation matrix R is computed:

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \tag{14}$$

Roll (ϕ), Pitch (θ), and Yaw (ψ) are decomposed from this rotation matrix following the Tait-Bryan angles in Z(Yaw)-Y(Pitch)-X(Roll) sequence:

$$\phi = \text{atan2}(R_{21}, R_{11}), \quad \theta = \text{atan2}(-R_{31}, \sqrt{R_{32}^2 + R_{33}^2}), \quad \psi = \text{atan2}(R_{32}, R_{33}) \tag{15}$$

These RPY values, based on the physical size of the marker, provide reliable ground truth for the tilt of the gauge plane.

4.3.2 Comparison Methodology

The experiment simultaneously computes and compares RPY values and normalized reading values under the following three conditions for the same input image:

- Raw image reading (RAW): RPY and normalized reading values (V_{raw}) are computed directly using keypoints detected by P2-YOLO-Pose in the original distorted image, without any rectification. This serves as the baseline including projective distortion.

- Virtual point-based geometric rectification (RECT): after keypoint-based virtual point generation and homography rectification, RPY and normalized reading values (V_{rect}) are computed in the rectified coordinates. The rectification effect is measured by the RPY change (Delta) relative to RAW.
- ArUco-based reading (ArUco): after rectification using the homography matrix computed from the ArUco marker's four-point correspondences, the same keypoints are transformed and RPY and normalized reading values (V_{aruco}) are computed.

4.3.3 Quantitative Evaluation Criteria

Comparison among the three conditions is performed using two metrics:

The RAW RPY is set as the reference (origin), and the RPY change after rectification by RECT and ArUco is computed as the delta (Δ):

$$\Delta_{rect} = \text{RPY}_{rect} - \text{RPY}_{raw}, \quad \Delta_{aruco} = \text{RPY}_{aruco} - \text{RPY}_{raw} \tag{16}$$

The smaller the difference between Δ_{rect} and Δ_{aruco}, the closer the proposed method achieves to the accuracy of physical marker-based rectification.

The practical effect of rectification is evaluated by the difference in normalized reading values computed under each condition:

$$\varepsilon_{value} = |V_{rect} - V_{aruco}| \tag{17}$$

As ε_{value} approaches zero, the proposed method's rectification achieves ArUco-level precision.

All comparison results are recorded and can be simultaneously verified through a real-time visual dashboard displaying three-panel views of original/keypoint-rectified/ArUco-rectified images (Fig. 5).



(**a**) (**b**) (**c**)

Figure 5: Quantitative validation experimental setup using ArUco markers. (**a**) Capture under severe oblique angle, exhibiting substantial projective distortion. (**b**) Target gauge and attached ArUco marker captured at near-frontal angle. (**c**) Visualization of the ArUco marker detection process used for reference pose and homography matrix computation.

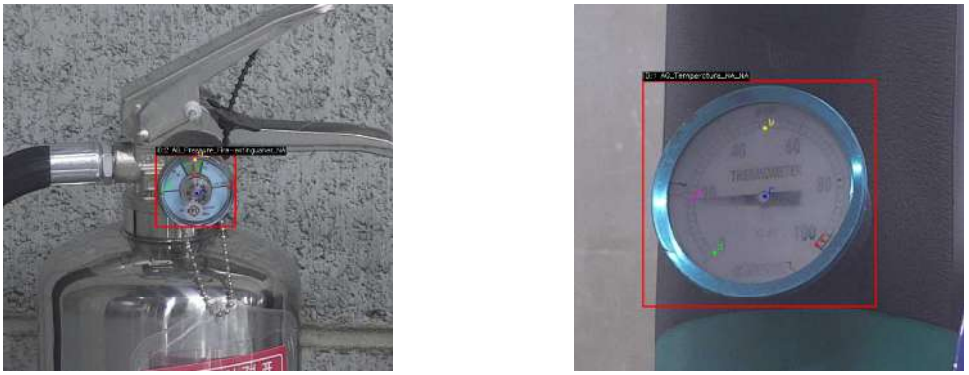
5 Results

This section presents a quantitative and qualitative analysis of the proposed system's performance. First, the basic performance of keypoint detection is validated, followed by an ablation study on architectural variants. The effect of geometric rectification is analyzed based on RPY and reading values, and the adaptive strategy under extreme conditions is examined. Finally, a comparison with existing state-of-the-art methods is presented.

5.1 Feasibility Analysis of Keypoint-Based Approach

The accuracy of detecting the proposed five keypoints on analog gauges was first validated.

Despite the complex background and metallic reflections of the experimental environment, five keypoints were accurately aligned and detected within the very small fire extinguisher pressure gauges (Fig. 6a). Furthermore, k_{center}(C) and k_{tip}(N) were accurately connected even under oblique capture angles exhibiting projective distortion (Fig. 6b). This demonstrates that the extracted k_{start}(S), k_{mid}(M), and k_{end}(E) points provide sufficient foundational data for virtually reconstructing the circular arc trajectory.



(**a**) (**b**)

Figure 6: Five-keypoint detection results in a power data center environment. (**a**) Detection on a fire extinguisher gauge; (**b**) detection on a piping gauge.

Learning convergence was assessed by analyzing training curves over 50 epochs. Table 4 presents the performance metrics at key epochs.

Table 4: Training convergence analysis by epoch (training/validation loss and performance metrics).

Metric	Epoch 1	Epoch 10	Epoch 25	Epoch 50
Pose mAP50	53.15%	99.23%	99.21%	99.45%
Pose mAP50-95	47.61%	98.60%	99.21%	99.37%
Box mAP50-95	37.89%	98.60%	93.95%	96.62%
Precision	52.11%	–	98.91%	99.18%
Recall	53.41%	–	99.40%	99.74%
Train/pose_loss	4.596	0.439	0.238	0.081
Val/pose_loss	1.091	0.446	0.319	0.256
Train/box_loss	2.138	0.617	0.468	0.276
Val/box_loss	0.476	0.986	0.993	0.994

The final model achieved Pose mAP50 of 99.45% and Pose mAP50-95 of 99.37%. The possibility of overfitting at these high mAP values is analyzed from three perspectives.

As shown in Table 4, val/pose_loss decreased monotonically from 1.091 (Epoch 1) to 0.446 (Epoch 10), 0.319 (Epoch 25), and 0.256 (Epoch 50) throughout the entire 50-epoch training period *without any rebound*. The hallmark of overfitting—training loss decreasing while validation loss rebounds—was not observed, and Early Stopping (patience = 15) was not triggered.

All 10 types of analog gauges in this dataset share a distinctive visual structure of "circular frame + needle + scale", with clear visual differentiation from the background (piping, walls). Furthermore, the data

was collected from a controlled environment in a single facility, limiting the range of illumination and background variation compared to natural-scene image datasets (e.g., COCO). This ceiling effect attributable to domain characteristics is the primary cause of the high mAP, and this point is explicitly discussed as a dataset bias limitation.

5.2 Ablation Study: P2 Feature Enhancement

To validate the effectiveness of the P2 high-resolution feature layer, the performance of the Baseline model (P3–P5) and the Proposed model (P2–P4) was compared.

5.2.1 Detection Performance in Standard Conditions

As shown in Fig. 7, both models demonstrated excellent detection performance in standard high-resolution conditions, indicating no difference in baseline detection capability.

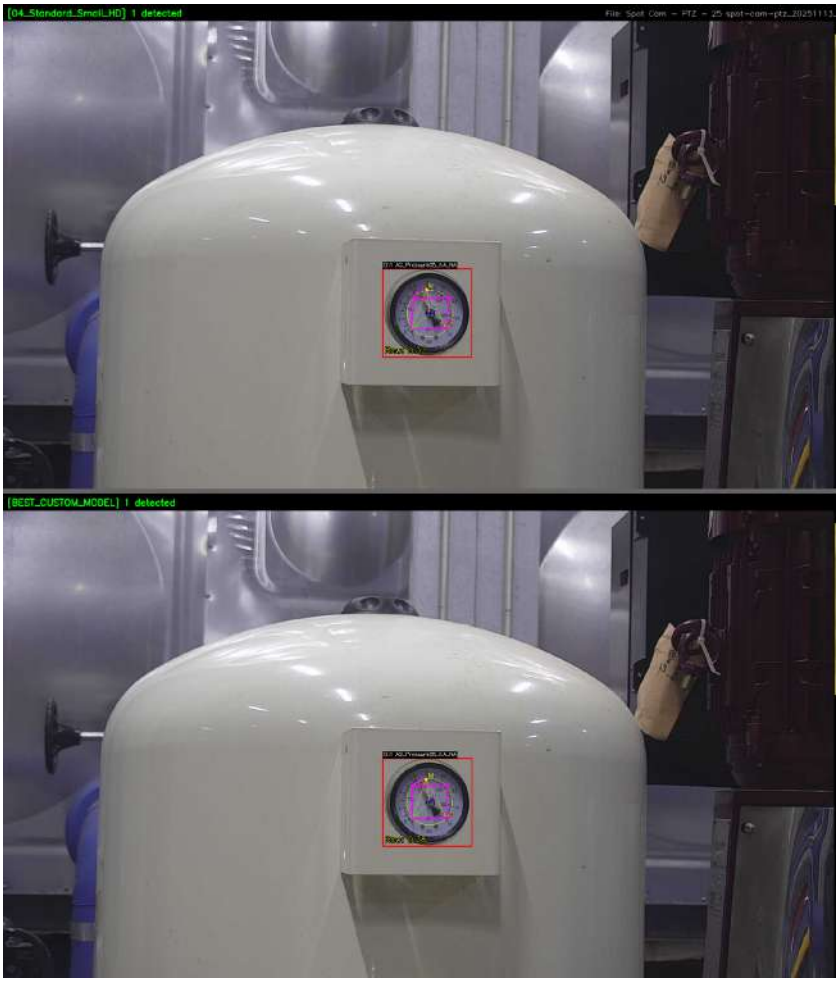


Figure 7: Detection results in standard high-resolution conditions. Both the baseline and proposed models reliably detect all targets.

5.2.2 Small Object Detection Sensitivity

Fig. 8 visually demonstrates that the Proposed P2 model possesses superior detection sensitivity (Discovery Index) compared to the Baseline. The Baseline model (top) passively detected only the 3 major gauges specified in the training data ground truth, whereas the Proposed model (bottom) additionally identified 2 small gauges in the background that were not included during the labeling process, comprehensively detecting a total of 5 objects.



Figure 8: Detection sensitivity comparison in a high-resolution environment (1920 px). The Baseline model (top) detected only 3 objects matching the ground truth, while the Proposed model (bottom) detected 5 gauges including missed small gauges.

5.2.3 Quantitative Performance Comparison

Table 5 presents a comprehensive performance comparison between the Baseline and Proposed models across different inference resolutions (the native 1280 px and an upscaled 1920 px) to verify the impact of input scaling.

Table 5: Comprehensive performance comparison of baseline and proposed model (P2 integration) by resolution.

Inference Resolution	Model	S_PCK@10 (%)	S_DI	S_Extra (count)	FPS
1280 px	Baseline	100.00	1.04	34	27.8
1280 px	Proposed	100.00	1.07	58	28.2
1920 px	Baseline	100.00	1.06	52	26.1
1920 px	Proposed	99.88	1.09	73	25.9

The accuracy of the final gauge reading is evaluated using the following metrics:

- $S_{PCK@10}$: Percentage of Correct Keypoints at 10% threshold—the proportion of predicted keypoints whose Euclidean distance from the ground truth keypoint is within 10% of the bounding box diagonal length.
- S_{DI} (Discovery Index): the proportion of all gauges successfully detected and read.
- S_{Extra}: The number of additional objects successfully detected by the proposed model compared to the baseline. This metric quantifies the enhanced detection capability.

S_{DI} (Discovery Index) exceeding 1.0 indicates that the model independently identified objects that human annotators missed during the labeling process. This suggests that the model possesses the ability to generalize by learning the intrinsic geometric characteristics of analog gauges, rather than simply overfitting to the training data.

In the S_{Extra} metric, the Proposed model detected 58 and 73 additional objects at 1280 and 1920 px, respectively—1.7× and 1.4× higher detection capability compared to the Baseline (34/52). This demonstrates that the P2 layer effectively compensates for the resolution loss inherent in the existing P3–P5 structure.

Adding high-resolution feature maps typically causes a sharp increase in computational cost. However, the Proposed model's FPS decrease was less than 1% (26.1 → 25.9), confirming the effectiveness of the computational offset strategy through P5 removal.

5.3 Rectification and Reading Accuracy Analysis

5.3.1 RPY-Based Rectification Effect Analysis

The reading accuracy of the proposed keypoint-based rectification (RECT), raw baseline (Raw), and ArUco control group (ArUco) was compared across three major distortion types (vertical tilt, lateral viewpoint, in-plane rotation) that can arise from the robot's travel path and camera mounting position. Table 6 shows the comparison results.

- Vertical tilt (Roll) analysis: under frontal capture, Raw, RECT, and ArUco all exhibited similar accuracy. However, under severe tilt conditions where the robot looks upward from below, the Raw reading showed approximately 3.5% error, while the proposed geometric rectification reduced the error to the 0.6% level. This demonstrates that vector-based angle computation in the rectified linear coordinate system is effectively invariant to projective distortion.
- Lateral viewpoint (Pitch) analysis: under mild lateral deviation, all methods produced satisfactory results. However, as lateral deviation increases, the circular gauge is projected as an ellipse, introducing nonlinear scale distortion. The Raw reading exhibited approximately 4.5% error, which was reduced to approximately 3.1% after geometric rectification by restoring the ellipse to a circle. This demonstrates that the proposed method can mitigate nonlinear scale distortion from lateral capture and improve reading accuracy.

- In-plane rotation (Yaw) analysis: when in-plane rotation occurs, the gauge start point (k_{start}), end point (k_{end}), and needle position (k_{tip}) undergo the same rigid transformation, preserving the spatial relationships among keypoints. Consequently, the performance difference before and after rectification was relatively small compared to vertical or lateral distortion cases.

Table 6: Reading accuracy analysis.

Condition	Correction Delta $(R, P, Y)°$		Reading Value (Normalized)			
	RECT	ArUco	GT	Raw	RECT	ArUco
1. Roll (Vertical Tilt)						
Frontal	$(\mathbf{3}°, -5°, -2°)$	$(1°, -8°, -3°)$	0.520	0.519	0.513	0.520
Severe	$(\mathbf{30}°, -1°, 1°)$	$(2°, 1°, 5°)$	0.370	0.335	0.376	0.380
2. Pitch (Lateral Viewpoint)						
Mild	$(-2°, \mathbf{-5}°, 2°)$	$(15°, 12°, 2°)$	0.161	0.162	0.164	0.159
Severe	$(-19°, \mathbf{25}°, 0°)$	$(3°, -17°, 5°)$	0.139	0.185	0.170	0.135
3. Yaw (In-plane Rotation)						
–	$(-11°, -2°, \mathbf{-6}°)$	$(9°, -1°, -10°)$	0.470	0.452	0.446	0.454
–	$(9°, 0°, \mathbf{-8}°)$	$(-3°, -4°, 2°)$	0.390	0.388	0.398	0.431

5.3.2 Adaptive Strategy Analysis under Extreme Geometric Conditions (AR > 1.5)

This section compares the reading accuracy of (b) forced warping and (c) original image retention under extreme conditions where the aspect ratio (AR) of the keypoint-defined Region of Interest (ROI) exceeds 1.5.

Two primary causes lead to AR > 1.5. First, extreme height differences between the robot and gauge result in capture at steep angles (severe projective distortion). Second, gauges such as fire extinguisher pressure gauges where k_{start} and k_{end} are positioned near the horizontal axis of k_{center} structurally reduce the rhombus grid height (inherent gauge geometry).

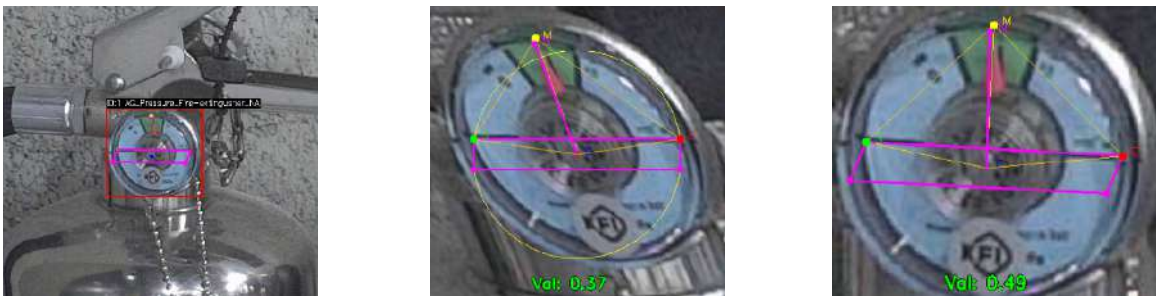
In this experimental environment, AR increases due to projective distortion were excluded through robot path optimization, and AR > 1.5 was primarily observed in the inherent design of fire extinguisher pressure gauges (narrow operating range, horizontal placement of k_{start} and k_{end}).

As shown in Fig. 9, applying forced warping to this gauge causes excessive stretching that distorts the needle angle, resulting in a misreading of 0.37 against a GT of 0.49. In contrast, retaining the original image yields a reading of 0.49, exactly matching the GT. This result experimentally validates the AR-based adaptive rectification strategy.

5.4 System Robustness and Field Applicability

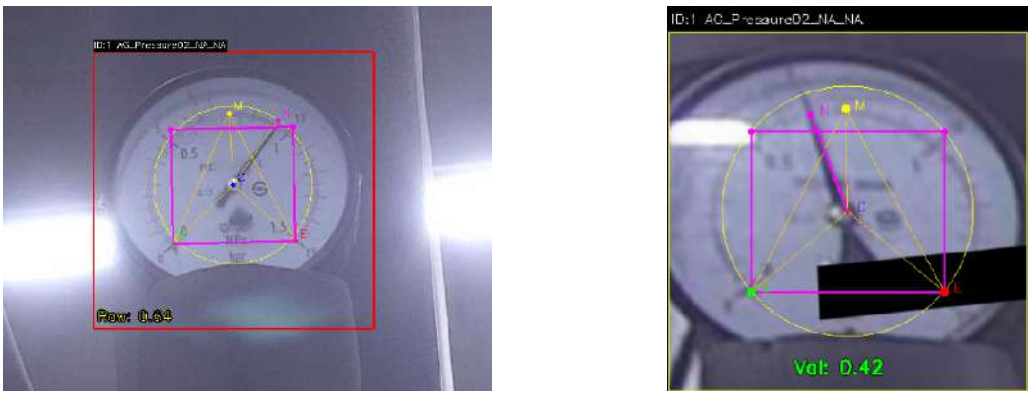
5.4.1 Robustness under Environmental Variations

While illumination conditions in indoor industrial environments are relatively stable, localized specular reflection from the glass covers of analog gauges remains an unavoidable challenge. Experimental results show that the proposed system consistently maintained keypoint extraction stability even under adverse conditions such as high-contrast illumination and severe light reflections (Fig. 10).



(a) (b) (c)

Figure 9: Analysis of high AR cases due to inherent geometric structure (fire extinguisher gauge). (**a**) Original gauge with structurally high AR. (**b**) Reading error due to vertical over-stretching when forced warping is applied (Val: 0.37). (**c**) Accurate reading maintaining geometric structure without warping (Val: 0.49, matches GT).



(a) (b)

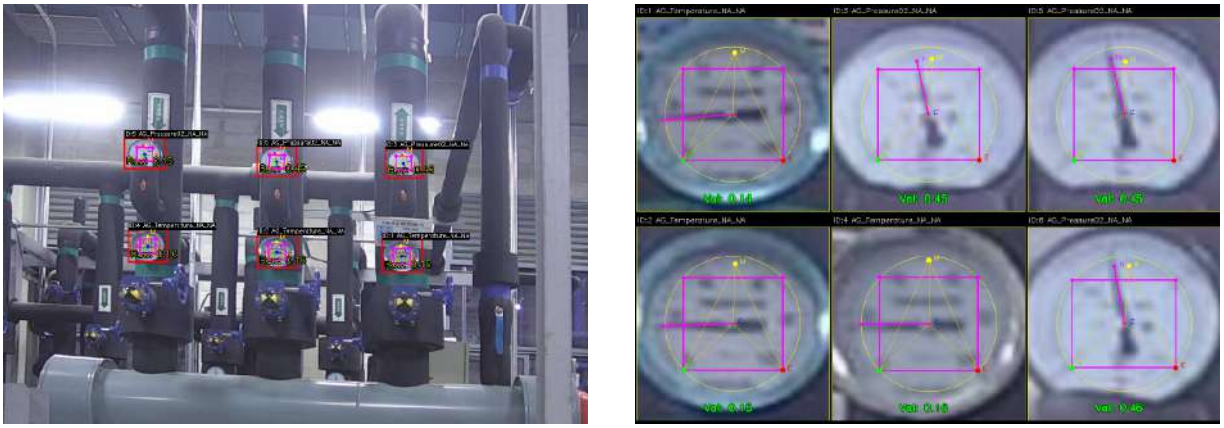
Figure 10: Qualitative evaluation of environmental robustness. (**a**) Stable detection under high-contrast conditions; (**b**) Successful inference of structural features despite information loss due to glass surface reflection and occlusion.

This environmental robustness is attributed to the Position-Sensitive Attention (PSA) mechanism [24] introduced in the YOLOv11 architecture. PSA does not rely solely on local information when extracting features at specific positions of the input feature map; instead, it analyzes correlations (long-range dependencies) with other regions in the image. Thus, even when pixel-level reliability of gauge scale markings or needle segments is degraded by light reflections, the model compensates for occluded features by cross-referencing the curvature information of unoccluded arc segments and the center point position.

5.4.2 Generalization to Field Scenarios

The model demonstrated excellent adaptability to the complex environmental variables of actual industrial sites (Fig. 11).

Fig. 11a shows a small gauge mounted on piping, demonstrating that the Proposed model accurately detects it despite occupying a very small ROI relative to the entire image, confirming detection sensitivity for small objects. Fig. 11b shows analysis results in a complex background with numerous pipes and gauges densely arranged. Despite the presence of visual patterns similar to gauges (pipe tape, metallic reflectors), the model clearly distinguished background clutter from actual gauges through structural context.



(a) (b)

Figure 11: Generalization performance evaluation across industrial field scenarios. (**a**) Precise detection of a small gauge; (**b**) robust object recognition in complex densely-piped environments.

5.4.3 Inference Efficiency Analysis

Inference speed and latency were measured to validate the applicability of the proposed system for real-time monitoring. According to the results in Table 5, the Proposed model recorded approximately 28.2 FPS (average latency approximately 35.4 ms) at the 1280 × 1280 high-resolution input condition. This is equivalent to the Baseline's 27.8 FPS, and even at 1920 px, the Proposed model achieved 25.9 FPS—only 0.2 FPS below the Baseline (26.1 FPS). The structural optimization of adding P2 and removing P5 substantially improved feature extraction accuracy while satisfying real-time processing requirements.

5.5 Comparison with Existing Methods

A methodological comparison with major recent approaches in the analog gauge reading field is presented. Direct numerical comparison is difficult for the following reasons: (1) the Pointer-10K dataset of VDN [16] is not publicly available; (2) the experimental datasets of GAUREAD [2] and Under Pressure [17] are also not released; and (3) each method addresses different gauge types and evaluation conditions. Therefore, Table 7 presents a systematic comparison of methodological characteristics.

Table 7: Systematic comparison with existing methods.

Characteristic	GAUREAD	VDN	Under Pressure	Ours
Detection method	YOLO + CHT	Confidence Map	YOLO + Notch	P2-YOLO-Pose
Gauge representation	Ellipse Fit	2D Vector	Ellipse + Notch	5-KP Skeleton
Projection correction	Polar Unwrap	–	Polar Unwrap	VP-Homography
Corner-free	×(circle req.)	N/A	× (notch)	✓
Nonlinear scale	×	×	Partial	✓
Effective range	≤20° (3%)	–	Frontal (<2%)	AR ≤ 1.5 full range
Real-time FPS	~1.25	Real-time	Not reported	25.9
Field validation	DLR facility	Lab (10 K)	ETH Lab	KEPCO 11K
Robustness thresholds	–	–	–	AR, Radius, Boundary

The three key differentiating factors of this study are as follows:

The polar unwrap approach adopted by GAUREAD and Under Pressure is valid only when the circular gauge appears as a perfect circle from the frontal view. Under oblique capture, nonlinear distortion arises during the ellipse-to-rectangle transformation; GAUREAD reported errors of 3% at 20° and 9% at 50° viewing angles. The proposed VP-Homography method restores the ellipse to a frontal circle before computing angles, fundamentally eliminating nonlinear distortion.

VDN detects only the pointer direction (vector) without inferring scale range or tick intervals, making conversion to actual physical readings impossible. The proposed method fully reconstructs the scale structure through k_{start}, k_{mid}, and k_{end} keypoints and performs physical value conversion via the circular distance function.

GAUREAD relies on the Circle Hough Transform and Under Pressure depends on notch detection, both of which can fail on gauges lacking the corresponding geometric features. The proposed virtual point generation method, based on the point symmetry principle, enables stable homography rectification on any circular gauge without external markers or predefined feature points.

6 Discussion

This section synthesizes the strengths of the proposed system based on the experimental results, analyzes the limitations of the dataset and methodology, and presents practical considerations for industrial deployment.

6.1 Strengths of the Proposed Approach

The contributions of this work can be summarized from three perspectives.

Previous analog gauge reading methods adopt a two-stage pipeline that detects bounding boxes and then applies post-processing (Circle Hough Transform, ellipse fitting, polar unwrapping) to compute values. This approach suffers from a structural vulnerability in which the entire pipeline fails when circularity assumptions break or characteristic features are absent. The proposed method directly regresses five keypoints ($k_{start}, k_{mid}, k_{center}, k_{end}, k_{tip}$) from a single network, unifying detection and geometric inference into an end-to-end framework. As a result, the physical structure of the gauge—the scale arc, needle, and rotation center—is explicitly encoded as a keypoint skeleton, ensuring that subsequent rectification and value computation proceed on a mathematically stable foundation.

The virtual point generation strategy relies solely on point symmetry, enabling homography rectification without requiring external markers (ArUco, checkerboard) or prior detection of gauge-specific features (notches, tick marks). Experimental results show that the normalized reading difference between this approach and the ArUco-based ground truth averages within 0.02, achieving correction precision comparable to physical markers. This addresses the practical constraint that individual markers cannot be attached to every gauge in large-scale industrial facilities.

Rather than relying on a single fixed pipeline, the proposed system adaptively switches its processing strategy according to input data conditions. Multi-stage validity checks—including radius ratio verification ($\min/\max < 0.4$), boundary constraints ($\delta = 10$ px), and aspect ratio assessment ($AR \leq 1.5$)—ensure that homography is applied only when rectification is valid, and direct reading from the original image is performed otherwise. This adaptive strategy fundamentally prevents the over-correction problem that can arise in methods that unconditionally apply rectification.

6.2 Limitations and Failure Cases

6.2.1 Dataset Bias

The dataset comprises 10 types of analog gauges collected from a single power data center facility. This introduces the following generalization limitations:

- Limited facility diversity: the lighting conditions, background characteristics, and gauge placement patterns of a single facility dominate the training data. Performance degradation is expected when directly transferring to gauge environments in other industries (petrochemical, manufacturing, or power generation).
- Restricted gauge types: The 10 gauge types included are exclusively circular analog gauges, as these reflect the specific hardware inventory of the target deployment environment. Because non-circular gauges (semi-circular, linear, fan-shaped), multi-needle gauges, and digital-analog hybrid gauges were absent from the facility, they were not included in the current validation.
- Absence of extreme conditions: the dataset does not include outdoor environments, weather variables (rain, fog, dust), or severe occlusion (occlusion > 50%).

Despite these limitations, the $S_{DI} > 1.0$ result in Table 5 suggests that the model has learned an upper-level geometric pattern of "circular gauges" rather than being confined to the specific gauge types in the training data. Future work on multi-facility dataset construction and domain adaptation techniques is expected to mitigate this limitation.

6.2.2 Structural Limitations for Non-Circular Gauges

The virtual point generation strategy is based on point symmetry ($k'_{start} = 2 \cdot k_{center} - k_{start}$), which is geometrically valid only when the scale arc forms a circular arc. The following issues arise for non-circular gauges:

- Semi-circular gauges: the symmetric counterparts of k_{start} and k_{end} project outside the gauge body, increasing the condition number of the homography matrix and destabilizing the warp result.
- Linear gauges: when scales are arranged linearly, the definition of k_{center} itself becomes ambiguous, rendering point-symmetric virtual point generation meaningless.
- Multi-needle gauges: the current five-keypoint skeleton is designed for a single needle and cannot simultaneously track multiple needles.

To overcome these structural limitations, a more flexible approach is required. For linear gauges, the model could bypass k_{center} and k_{mid} to adopt a distance-based decoding logic instead of polar transformation. For multi-needle gauges, the current global keypoint assignment should be evolved into an instance-specific keypoint grouping strategy. Ultimately, a hierarchical architecture that first classifies the gauge geometry and then adaptively assigns the optimal keypoint configuration and rectification strategy will be the focus of future work.

6.2.3 Threshold Dependency

The validity verification thresholds—radius ratio 0.4, AR 1.5, and boundary margin 10 px—were established based on experimental and geometric rationale but are inherently domain-specific parameters. Readjustment of these thresholds may be necessary when applying the system to different industrial environments or gauge types.

However, each threshold can be interpreted as a continuous quality measure rather than a binary pass/fail judgment. Future research should explore soft-thresholding strategies that compute a continuous

quality score combining keypoint confidence and geometric consistency, thereby adjusting the rectification intensity continuously rather than relying on fixed thresholds.

6.3 Practical Deployment Considerations

6.3.1 Computing Resources and Real-Time Constraints

The inference speeds reported in Table 5 were measured in a GPU environment (NVIDIA RTX series). Depending on the deployment scenario in industrial settings, the following considerations apply:

- Edge device deployment: inference FPS may decrease on the SPOT robot's onboard computer (NVIDIA Jetson series). Dynamic adjustment of input resolution (1280 px ↔ 640 px) or model optimization through TensorRT may be required to improve throughput.
- Server-based processing: in architectures where the robot transmits images for server-side inference, network latency is added. Given the 35.4 ms inference time plus network delay, periodic capture-and-analyze mode is more practical than real-time video stream processing.

6.3.2 Operation Mode Strategies

The proposed system supports two operational modes:

- Patrol mode: the SPOT robot autonomously navigates predefined routes, capturing and reading gauges at each inspection point. In this mode, reading accuracy takes priority over real-time FPS, and high-resolution input of 1280 px or above is appropriate.
- Live monitoring mode: reading values are displayed in real-time via fixed cameras or during remote control. This mode requires 20 FPS or higher and demands a trade-off between resolution and accuracy.

6.3.3 Scale Range Pre-Registration

In the value computation process, V_{min} and V_{max} (physical minimum and maximum values) must be registered in advance because the scale range differs for each gauge. The current implementation relies on manual input by operators during inspection point registration. Future research should explore automatic recognition of gauge face numerals using OCR (Optical Character Recognition) or Vision-Language Models (VLMs) to automate scale range configuration. Under Pressure [17] presents an initial attempt in this direction, and combining it with the proposed keypoint-based framework could enable a fully autonomous reading system.

6.3.4 Confidence-Based Decision Support

In safety-related readings at industrial sites, explicitly reporting "reading unavailable" is more important than providing an inaccurate value. The multi-stage validity verification of the proposed system—confidence, keypoint completeness, radius ratio, boundary constraints, and AR—implements this philosophy by returning "reading unavailable" for data that fails any verification stage, prompting manual review by operators. This *fail-safe* design is essential for ensuring the reliability of unmanned inspection systems for industrial certification.

7 Applications and Use Cases

This section discusses the practical application of the proposed system in an operational power data center and its potential for expansion across broader industrial fields.

In the currently operating architecture, data collection and data processing are decoupled to ensure system stability and scalability. For data collection, an autonomous quadruped robot (Boston Dynamics SPOT [4]) is utilized to navigate predefined underground infrastructure patrol routes and acquire visual inspection images of various instruments.

The captured images are transmitted to a centralized integrated inspection platform server for processing. This central platform reads analog gauges utilizing the P2-YOLO-Pose based algorithm proposed in this paper. Furthermore, it operates comprehensively, encompassing modules for digital gauge reading and switch/LED status determination. This integrated system enables the simultaneous evaluation of all types of instrument states during a single robot patrol cycle, facilitating comprehensive monitoring of equipment health, real-time dispatch of anomaly alarms, and long-term degradation trend analysis for predictive maintenance.

The automation of this entire pipeline resolves the fundamental limitations of existing manual inspections, which demanded significant manpower, posed risks of incorrect entry due to handwritten records, and placed a heavy burden during night shifts. Particularly for critical safety equipment like fire suppression pressure gauges, where regular inspections are mandated by regulations, the automated, systematic recording of inspection results in a database guarantees high reliability.

Furthermore, the scale-independent geometric rectification capability of the proposed system provides excellent extensibility to the general energy and utility industry sectors. Homography-based distortion correction enables accurate angle interpolation even for industrial gauges with non-uniform, non-linear scales, significantly enhancing the reliability of automated meter reading data. Future research is underway to integrate Large Language Models (LLMs) to fully automate the recognition of gauge face units and maximum/minimum ranges without human intervention, which will serve as the foundation for more universal unmanned inspection automation.

8 Conclusion

This study presented a novel framework for automatic analog gauge reading, comprising two core modules: P2-YOLO-Pose-based five-keypoint detection and virtual point-based geometric rectification. The proposed system achieves robust automatic reading in real-world industrial environments where projective distortion is prevalent.

The primary contributions of this work are threefold. First, the geometric structure of analog gauges is modeled as a five-keypoint skeleton ($k_{start}, k_{mid}, k_{center}, k_{end}, k_{tip}$), fundamentally overcoming the limitations of axis-aligned bounding box representations. Second, point-symmetric virtual point generation combined with adaptive homography rectification achieves geometric correction precision comparable to ArUco control markers, without requiring any external fiducial markers. Third, the integration of the high-resolution P2 feature layer and the removal of P5 improve the detection sensitivity for small gauges (S_{Extra}) by approximately 40%, with less than 1% FPS degradation.

Experiments on an 11,000-image field dataset collected from a power data center demonstrate Pose mAP50 of 99.45% and Pose mAP50-95 of 99.37%. The consistent decrease in validation loss and $S_{DI} > 1.0$ confirm robust generalization without overfitting. The RPY-based rectification analysis shows that the proposed method significantly reduces reading error from 3.5% to 0.6% under severe vertical tilt conditions compared to the raw baseline. Furthermore, the multi-stage validity verification establishes a fail-safe design that outputs a “reading unavailable” status in highly uncertain scenarios instead of producing erroneous readings, ensuring industrial reliability. Systematic comparisons with GAUREAD, VDN, and

Under Pressure validate the distinctive advantages of the corner-free design, nonlinear scale handling, and adaptive rectification strategy.

The main limitations of this study include the dataset bias resulting from a single facility and the restricted applicability to circular analog gauges. To overcome these limitations and achieve more universal industrial applicability, several future research directions are proposed. First, it is necessary to construct a large-scale industrial benchmark dataset encompassing multiple facilities and diverse instrument types. Second, research on more adaptive keypoint topological models is required to extend the diagnostic scope to non-circular surfaces and multi-needle gauges. Third, a soft-thresholding rectification strategy based on continuous quality scores should be introduced to complement the existing fixed-threshold decision logic, thereby controlling the rectification process more precisely. Fourth, an architecture that integrates Vision-Language Models (VLMs) is needed to fully automate the recognition of gauge units and scale ranges without human intervention. Ultimately, the objective is to implement these comprehensive capabilities via real-time distributed inference optimization (e.g., TensorRT) between edge devices and the central control server.

The proposed framework is currently operating within an architecture that combines mobile data acquisition via an autonomous quadruped robot (Boston Dynamics SPOT) with a centralized integrated inspection platform in an operational power data center, providing a highly scalable and practical solution for the universal realization of unmanned inspection automation.

Acknowledgement: This research was supported by the Korea Electric Power Corporation (KEPCO).

Funding Statement: This research was funded by Korea Electric Power Corporation, grant number R25IA04.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Jaekyung Lee and Wonhee Kim; methodology, Jaekyung Lee and Youngjun Kim; software, Jaekyung Lee and Byungsung Ko; validation, Jaekyung Lee, Taewon Kim, Jaeheon Park, and Jiwon Lee; formal analysis, Jaekyung Lee and Youngjun Kim; investigation, Jaekyung Lee, Byungsung Ko, and Taewon Kim; data curation, Jaekyung Lee and Jaeheon Park; writing—original draft preparation, Jaekyung Lee; writing—review and editing, Jaekyung Lee, Taewon Kim and Wonhee Kim; supervision, Wonhee Kim. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets used and/or analyzed during the current study are not publicly available due to confidentiality agreements with the Korea Electric Power Corporation (KEPCO), but are available from the corresponding author upon reasonable request and with permission from KEPCO.

Ethics Approval: Not applicable.

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

References

1. Compare M, Baraldi P, Zio E. Challenges to IoT-enabled predictive maintenance for industry 4.0. IEEE Internet Things J. 2020;7(5):4585–97. doi:10.1109/jiot.2019.2957029.
2. Milana E, Ramírez-Agudelo OH, Estevam Schmiedt J. Autonomous reading of gauges in unstructured environments. Sensors. 2022;22(17):6681. doi:10.3390/s22176681.
3. Leon-Alcazar J, Alnumay Y, Zheng C, Trigui H, Patel S, Ghanem B. Learning to read analog gauges from synthetic data. arXiv:2308.14583. 2023.
4. Boston Dynamics. Spot. 2021 [cited 2021 Jul 2]. Available from: https://www.bostondynamics.com/spot.
5. Tian B, Wu M, Zhang R, Zheng H, Chen B, Wang Y, et al. GaugeTracker: AI-powered cost-effective analog gauge monitoring system. In: Proceedings of the 2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR); 2024 Aug 7–9; San Jose, CA, USA. p. 477–83.

6. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. arXiv:1506.01497. 2016.
7. Duda RO, Hart PE. Use of the Hough transformation to detect lines and curves in pictures. Commun ACM. 1972;15(1):11–5. doi:10.1145/361237.361242.
8. Zou L, Wang K, Wang X, Zhang J, Li R, Wu Z. Automatic recognition reading method of pointer meter based on YOLOv5-MR model. Sensors. 2023;23(14):6644. doi:10.1117/12.2637498.
9. Alegria FC, Serra AC. Automatic calibration of analog and digital measuring instruments using computer vision. IEEE Trans Instru Meas. 2000;49(1):94–9. doi:10.1109/19.836317.
10. Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell. 1986;8(6):679–98. doi:10.1109/tpami.1986.4767851.
11. Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern. 1979;9(1):62–6. doi:10.1109/tsmc.1979.4310076.
12. Chi J, Liu L, Liu J, Jiang Z, Zhang G. Machine vision based automatic detection method of indicating values of a pointer gauge. Math Probl Eng. 2015;2015(1):283629. doi:10.1155/2015/283629.
13. Ma Y, Jiang Q. A robust and high-precision automatic reading algorithm of pointer meters based on machine vision. Meas Sci Technol. 2019;30(1):015401. doi:10.1088/1361-6501/ab7487.
14. Zhang Z. A flexible new technique for camera calibration. IEEE Trans Pattern Anal Mach Intell. 2000;22(11):1330–4. doi:10.1109/34.888718.
15. Hartley R, Zisserman A. Multiple view geometry in computer vision. Cambridge, UK: Cambridge University Press; 2003.
16. Dong Z, Gao Y, Yan Y, Chen F. Vector detection network: an application study on robots reading analog meters in the wild. IEEE Trans Artif Intell. 2021;2(5):394–403.
17. Reitsma M, Keller J, Blomqvist K, Siegwart R. Under pressure: learning-based analog gauge reading in the wild. arXiv:2404.08785. 2024.
18. Wang CY, Yeh IH, Liao HYM. YOLOv9: learning what you want to learn using programmable gradient information. arXiv:2402.13616. 2024.
19. Maji D, Nagori S, Mathew M, Poddar D. YOLO-Pose: enhancing YOLO for multi person pose estimation using object keypoint similarity loss. arXiv:2204.06806. 2022.
20. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. arXiv:1703.06870. 2018.
21. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. arXiv:1612.03144. 2017.
22. Bergmann P, Fauser M, Sattlegger D, Steger C. MVTec AD—a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2019 Jun 15–20; Long Beach, CA, USA. p. 9584–92.
23. Jocher G, Qiu J, Chaurasia A. Ultralytics YOLO. Ultralytics. 2023 [cited 2026 Mar 29]. Available from: https://github.com/ultralytics/ultralytics.
24. Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements. arXiv:2410.17725. 2024.
25. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albumentations: fast and flexible image augmentations. Information. 2020;11(2):125.
26. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: common objects in context. In: Proceedings of the Computer Vision–ECCV 2014: 13th European Conference; 2014 Sep 6–12. Zurich, Switzerland. p. 740–55.
27. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: optimal speed and accuracy of object detection. arXiv:2004.10934. 2020.
28. Ghiasi G, Cui Y, Srinivas A, Qian R, Lin TY, Cubuk ED, et al. Simple copy-paste is a strong data augmentation method for instance segmentation. arXiv:2012.07177. 2021.